Available online at www.sciencedirect.com
SCIENCE DIRECT®
Engineering Applications of Artificial Intelligence 16 (2003) 709–716
Engineering Applications of
ARTIFICIAL INTELLIGENCE
www.elsevier.com/locate/engappai

Elicitation of classification rules by fuzzy data mining

Yi-Chung Hu[a,*], Gwo-Hshiung Tzeng[b]

[a] Department of Business Administration, Chung Yuan Christian University, Chung-Li 320, Taiwan, ROC
[b] Institute of Management of Technology, National Chiao Tung University, Hsinchu 300, Taiwan, ROC

Received 25 November 2002; received in revised form 22 May 2003; accepted 18 September 2003

Abstract

Data mining techniques can be used to find potentially useful patterns from data and to ease the knowledge acquisition bottleneck in building prototype rule-based systems. Based on the partition methods presented in simple-fuzzy-partition-based method (SFPBM) proposed by Hu et al. (Comput. Ind. Eng. 43(4) (2002) 735), the aim of this paper is to propose a new fuzzy data mining technique consisting of two phases to find fuzzy if–then rules for classification problems: one to find frequent fuzzy grids by using a pre-specified simple fuzzy partition method to divide each quantitative attribute, and the other to generate fuzzy classification rules from frequent fuzzy grids. To improve the classification performance of the proposed method, we specially incorporate adaptive rules proposed by Nozaki et al. (IEEE Trans. Fuzzy Syst. 4(3) (1996) 238) into our methods to adjust the confidence of each classification rule. For classification generalization ability, the simulation results from the iris data demonstrate that the proposed method may effectively derive fuzzy classification rules from training samples.

Keywords: Data mining; Fuzzy sets; Classification problems; Association rules

1. Introduction

Pattern classification is a problem that partitions a pattern space into classes and then assigns a pattern to one of those classes (Kim and Bang, 2000). In fact, classification problems have played an important role in industrial engineering such as the group technology (Chuang et al., 1999), in engineering applications, such as OCR recognition and facial recognition (Kim and Bang, 2000).

Data mining is the exploration and analysis of the data in order to discover meaningful patterns (Berry and Linoff, 1997). The aim of this paper is to propose a fuzzy data mining method that can automatically find a set of fuzzy if–then rules for classification problems. Actually, data mining problems involving classification can be viewed within a common framework of rule discovery (Agrawal et al., 1993a). The advantage for mining fuzzy if–then rules for classification problems is that knowledge acquisition can be achieved for users by carefully checking these rules discovered from training patterns. Additionally, data mining can also ease the knowledge acquisition bottleneck in building prototype expert systems (Hong et al., 2000) or rule-based systems.

The discovery of association rule is an important topic in data mining techniques. In addition, association rules elicited from transaction databases have been applied to help decision makers determine which items are frequently purchased together by customers (Berry and Linoff, 1997; Han and Kamber, 2001). Initially, Agrawal et al. (1993b) proposed a method to find the frequent itemsets. Subsequently, Agrawal et al. (1996) proposed an influential algorithm named the Apriori algorithm consisting two phases. In the first phase, frequent itemsets are generated, whereas a candidate k-itemset ($k \geqslant 1$) containing k items, is frequent (i.e., frequent k-itemset) if its support is larger than or equal to a user-specified minimum support. In the second phase, association rules are generated by frequent itemsets discovered in the first phase.

Additionally, the comprehensibility of fuzzy representation by human users is also a criterion in designing a fuzzy system. The simple fuzzy partition methods are thus preferable (Ishibuchi et al., 1999). In this method, each attribute, which is used to describe each sample data, is viewed as linguistic variables (Zadeh,

*Corresponding author. Tel.: +886-3-2655130; fax: +886-3-2655199.
E-mail address: ychu@cycu.edu.tw (Y.-C. Hu).

0952-1976/$ - see front matter
doi:10.1016/j.engappai.2003.09.007

Nomenclature

d	number of attributes used to describe each sample data, where $1 \leqslant d$
k	dimension of one fuzzy grid, where $1 \leqslant k \leqslant d$
K	maximum number of various linguistic values defined in each quantitative attribute, where $K \geqslant 3$
$A^{xm}_{K_i,j_m}$	j_mth linguistic value of K_i various linguistic values defined in attribute x_m, where $1 \leqslant m \leqslant d, 3 \leqslant K_i \leqslant K$ for the MTDM, $K_i = K$ for the STDM, and $1 \leqslant j_m \leqslant K_i$
$\mu^{x_m}_{K_i,j_m}$	membership function of $A^{x_m}_{K_i,j_m}$
t_p	pth training sample, where $t_p = (t_{p_1}, t_{p_2}, \cdots t_{p_d})$, and t_{p_i} is the attribute value with respect to the ith attribute.

1975a, b, 1976). Based on the partition methods used in simple-fuzzy-partition-based method (SFPBM) (Hu et al., 2002) this paper proposes a fuzzy data mining method for eliciting fuzzy classification rules for classification problems.

Since the classification performance can be improved by adjusting the grade of certainty of fuzzy rules, the adaptive rules proposed by Nozaki et al. (1996) are incorporated into the proposed method to adjust the fuzzy confidence of each fuzzy rule. For classification generalization ability, the simulation results from the iris data (Anderson, 1935) demonstrate that the proposed method performs well in comparison with other fuzzy classification methods. This shows that applications of the proposed method to engineering problems are feasible.

The rest of this paper is organized as follows. STDM and MTDM are introduced in Section 2. In Section 3, we present definitions of the fuzzy support and the fuzzy confidence, and the two phases of the proposed method is presented in detail. In Section 4, the performance of the proposed method is examined by computer simulation on the iris data. Discussions and conclusions are presented in Section 5.

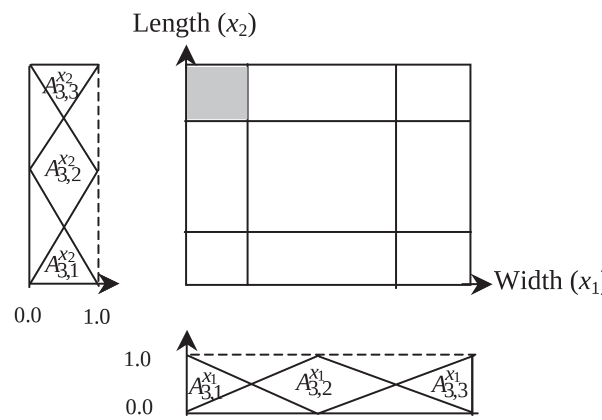

Fig. 1. $K = 3$ for "Width" and "Length".

2. Simple fuzzy partition methods

Concepts of linguistic variables (Zadeh, 1975a, b, 1976). Formally, a linguistic variable is characterized by a quintuple (Pedrycz and Gomide, 1998; Zimmermann, 1996) denoted by (x, $T(x)$, U, G, M), in which x is the name of the variable; $T(x)$ denotes the set of names of linguistic values or terms of x; U denotes a universe of discourse; G is a syntactic rule for generating values of x; and M is a semantic rule for associating a linguistic value with a meaning.

Actually, each attribute can be partitioned by its various linguistic values with pre-specified membership functions, such as triangular shape functions. Simple fuzzy grids or grid partitions (Ishibuchi et al., 1995; Jang and Sun, 1995) in feature space are thus obtained. The advantage of the simple fuzzy partition method is that the linguistic interpretation of each fuzzy set is easily obtained. Fuzzy partition methods have been widely used in pattern recognition and fuzzy reasoning, such as applications to pattern classification by Ishibuchi et al. (1992, 1995, 1999), Ravi and Zimmermann (2000), and Ravi et al. (2000), to fuzzy neural networks (Jang, 1993), and to the fuzzy rule generation by Wang and Mendel (1992).

If both x_1 and x_2 are partitioned by three various linguistic values, then a feature space can be divided into nine two-dimensional (2-dim) fuzzy grids, as shown in Fig. 1. The shaded fuzzy subspace denoted by $A^{\text{Width}}_{3,1} \times A^{\text{Length}}_{3,3}$ stands for a 2-dim fuzzy grid whose linguistic value is "small AND large".

Two partition types used in SFPBM are employed in the proposed method: one is the multiple type division method (MTDM), and the other is the single type division method (STDM). If K is the maximum number of various linguistic values on each quantitative attribute, then MTDM allow us to partition each quantitative attribute into various $(3+4+\cdots+K)$ linguistic values. In other words, we sequentially divide each quantitative attribute into 3, 4, … ,K various linguistic values. As for STDM, only K various linguistic values are defined.

For simplicity, the membership function with triangular shape is used for each linguistic value in the quantitative attributes. However, we emphasize that Pedrycz (1994) had pointed out the usefulness and effectiveness of the triangular membership functions in the fuzzy modeling. A membership function such as $\mu_{K,j_1}^{\text{Width}}$ is represented as follows:

$$\mu_{K,j_1}^{\text{Width}}(x) = \max\{1 - |x - a_{j_1}^K|/b^K, 0\}, \quad (1)$$

where

$$a_{j_1}^K = \text{mi} + (\text{ma} - \text{mi})(j_1 - 1)/(K - 1), \quad (2)$$

$$b^K = (\text{ma} - \text{mi})/(K - 1), \quad (3)$$

where ma is the maximum value of domain, and mi is the minimum value. Each linguistic value is actually viewed as a candidate one-dimensional (1-dim) fuzzy grid in the proposed method. It is clear that the set of candidate 1-dim fuzzy grids generated for a pre-specified K by STDM is contained in that generated by MTDM. For example, if we divide both "Width" (denoted by x_1) and "Length" (denoted by x_2) by four various linguistic values, then $\{A_{4,1}^{\text{Width}}, A_{4,2}^{\text{Width}}, A_{4,3}^{\text{Width}}, A_{4,4}^{\text{Width}}, A_{4,1}^{\text{Length}}, A_{4,2}^{\text{Length}}, A_{4,3}^{\text{Length}}, A_{4,4}^{\text{Length}}\}$ is generated by STDM, and $\{A_{3,1}^{\text{Width}}, A_{3,2}^{\text{Width}}, A_{3,3}^{\text{Width}}, A_{3,1}^{\text{Length}}, A_{3,2}^{\text{Length}}, A_{3,3}^{\text{Length}}, A_{4,1}^{\text{Width}}, A_{4,2}^{\text{Width}}, A_{4,3}^{\text{Width}}, A_{4,4}^{\text{Width}}, A_{4,1}^{\text{Length}}, A_{4,2}^{\text{Length}}, A_{4,3}^{\text{Length}}, A_{4,4}^{\text{Length}}\}$ is generated by MTDM when $K = 4$.

A significant task is how to use the candidate 1-dim fuzzy grids to generate the other frequent fuzzy grids and the fuzzy classification rules. An effective method is thus described in following section.

3. Discovering fuzzy classification rules

In the proposed method, frequent fuzzy grids and fuzzy classification rules are generated by phases I and II, respectively. One fuzzy partition method (i.e., STDM or MTDM) must be specified before performing the proposed algorithm.

The main difference between the proposed method and SFPBM is that SFPBM did not consider all information distributed in the pattern space during the mining process. That is, SFPBM ignored those fuzzy subspaces containing any two linguistic values belonging to different K_i partitions. Thus, SFPBM cannot generate a fuzzy space like $A_{K_1,j_1}^{x_1} \times A_{K_2,j_2}^{x_2} \times \cdots \times A_{K_{k-1},j_{k-1}}^{x_{k-1}} \times A_{K_k,j_k}^{x_k}$. For example, if two quantitative attributes, say x_1 and x_2, are partitioned into 4 linguistic values (i.e., $K = 4$) with MTDM, then fuzzy subspaces or candidate fuzzy grids $A_{K_1,j_1}^{x_1} \times A_{K_2,j_2}^{x_2}$ are not considered by SFPBM when K_1 is not equal to K_2 (e.g., $A_{4,2}^{\text{Width}} \times A_{3,3}^{\text{Length}}$ or $A_{3,1}^{\text{Width}} \times A_{4,1}^{\text{Length}}$). However, it is possible that the ignored subspaces, which are further considered in the proposed method, are useful. It should be noted that since $K_i = 2$ is somewhat coarser, $K_i \geqslant 3$ is considered in SFPBM and the proposed method.

In this section, we describe the individual phase of the proposed method in Sections 3.1 and 3.2.

3.1. Phase I: generate frequent fuzzy grids

Suppose each quantitative attribute, x_m, is divided into K various linguistic values. Without loss of generality, given a candidate k-dim fuzzy grid $A_{K_1,j_1}^{x_1} \times A_{K_2,j_2}^{x_2} \times \cdots \times A_{K_{k-1},j_{k-1}}^{x_{k-1}} \times A_{K_k,j_k}^{x_k}$ which is a fuzzy set, $3 \leqslant K_1, K_2, \cdots, K_{k-1}, K_k \leqslant K$ for the MTDM and $K_1 = K_2 = \cdots = K_{k-1} = K_k = K$ for the STDM, the degree to which t_p belongs to this fuzzy grid (i.e., $A_{K_2,j_2}^{x_2} \times \cdots \times A_{K_{k-1},j_{k-1}}^{x_{k-1}} \times A_{K_k,j_k}^{x_k}(t_p)$) can be computed as $\mu_{K_1,j_1}^{x_1}(t_{p_1}) \cdot \mu_{K_2,j_2}^{x_2}(t_{p_2}) \cdots \mu_{K_{k-1},j_{k-1}}^{x_{k-1}}(t_{p_{k-1}}) \cdot \mu_{K_k,j_k}^{x_k}(t_{p_k})$. To check whether this fuzzy grid to be frequent or not, the fuzzy support (Ishibuchi et al., 2001; Hu et al., 2002) of $A_{K_1,j_1}^{x_1} \times A_{K_2,j_2}^{x_2} \times \cdots \times A_{K_{k-1},j_{k-1}}^{x_{k-1}} \times A_{K_k,j_k}^{x_k}$ with the algebraic product, which is a t-norm operator in the fuzzy intersection, is defined as follows:

$$\begin{aligned} &\text{FS}\left(A_{K_1,j_1}^{x_1} \times A_{K_2,j_2}^{x_2} \times \cdots \times A_{K_2,j_2}^{x_2} \times A_{K_{k-1},j_{k-1}}^{x_{k-1}} \times A_{K_k,j_k}^{x_k}\right) \\ &= \sum_{p=1}^{n} \mu A_{K_1,j_1}^{x_1} \times A_{K_2,j_2}^{x_2} \times A_{K_{k-1},j_{k-1}}^{x_{k-1}}(t_p)/n \\ &= \left[\sum_{p=1}^{n} \mu_{K_1,j_1}^{x_1}(t_{p_1}) \cdot \mu_{K_2,j_2}^{x_2}(t_{p_2}) \cdots \mu_{K_{k-1},j_{k-1}}^{x_{k-1}} \right. \\ &\quad \left. \times (t_{p_{k-1}}) \cdot \mu_{K_k,j_k}^{xk}(t_{p_k})\right]/n). \end{aligned} \quad (4)$$

When $\text{FS}(A_{K_1,j_1}^{x_1} \times A_{K_2,j_2}^{x_2} \times \cdots \times A_{K_{k-1},j_{k-1}}^{x_{k-1}} \times A_{K_k,j_k}^{x_k})$ is larger than or equal to the user-specified minimum fuzzy support (min FS), $A_{K_1,j_1}^{x_1} \times A_{K_2,j_2}^{x_2} \times \cdots \times A_{K_{k-1},j_{k-1}}^{x_{k-1}} \times A_{K_k,j_k}^{x_k}$ is a frequent k-dim fuzzy grid. For any two frequent grids, say $A_{K_1,j_1}^{x_1} \times A_{K_2,j_2}^{x_2} \times \cdots \times A_{K_{k-1},j_{k-1}}^{x_{k-1}} \times A_{K_k,j_k}^{x_k}$ and $A_{K_1,j_1}^{x_1} \times A_{K_2,j_2}^{x_2} \times \cdots \times A_{K_{k-1},j_{k-1}}^{x_{k-1}} \times A_{K_k,j_k}^{x_k} \times A_{K_{k+1},j_{k+1}}^{x_{k+1}}$, since $\mu_{A_{K_1,j_1}^{x_1} \times A_{K_2,j_2}^{x_1} \times \cdots \times A_{K_{k-1},j_{k-1}}^{x_{k-1}} \times A_{K_k,j_k}^{xk} \times A_{K_{k+1},j_{k+1}}^{x_{k+1}}}(t_p) \leqslant \mu_{A_{K_1,j_1}^{x_1} \times A_{K_2,j_2}^{x_2} \times \cdots \times A_{K_{k-1},j_{k-1}}^{x_k} \times A_{K_k,j_k}^{x_k}}(t_p)$ from (4), $A_{K_1,j_1}^{x_1} \times A_{K_2,j_2}^{x_2} \times \cdots \times A_{K_{k-1},j_{k-1}}^{x_{k-1}} \times A_{K_k,j_k}^{x_k} \times A_{K_{k+1},j_{k+1}}^{x_{k+1}} \subseteq A_{K_1,j_1}^{x_1} \times A_{K_2,j_2}^{x_2} \times \cdots \times A_{K_{k-1},j_{k-1}}^{x_{k-1}} \times A_{K_k,j_k}^{x_k}$ thus holds. It is obvious that any subset of a frequent fuzzy grid must also be frequent.

Like SFPBM, Table FGTTFS is implemented to generate frequent fuzzy grids. FGTTFS consists of the following substructures:

(a) Fuzzy grid table (FG): each row represents a fuzzy grid, and each column represents a linguistic value $A_{K_i,j_m}^{x_m}$.
(b) Transaction table (TT): each column represents t_p, and each element records the membership degree of the corresponding fuzzy grid.
(c) Column FS: stores the fuzzy support corresponding to the fuzzy grid in FG.

An initial tabular FGTTFS is shown as Table 1 as an example, from which we can see that there are two samples t_1 and t_2, with two attributes x_1 and x_2. Both x_1

Table 1
Initial table FGTTFS for an example

Fuzzy grid	FG			TT			FS		
	$A^{x_1}_{3,1}$	$A^{x_1}_{3,2}$	$A^{x_1}_{3,3}$	$A^{x_2}_{3,1}$	$A^{x_2}_{3,2}$	$A^{x_2}_{3,3}$	t_1	t_2	
$A^{x_1}_{3,1}$	1	0	0	0	0	0	$\mu^{x_1}_{3,1}(t_{1_1})$	$\mu^{x_1}_{3,1}(t_{2_1})$	$FS(A^{x_1}_{3,1})$
$A^{x_1}_{3,2}$	0	1	0	0	0	0	$\mu^{x_1}_{3,2}(t_{1_1})$	$\mu^{x_1}_{3,2}(t_{2_1})$	$FS(A^{x_1}_{3,2})$
$A^{x_1}_{3,3}$	0	0	1	0	0	0	$\mu^{x_1}_{3,3}(t_{1_1})$	$\mu^{x_1}_{3,3}(t_{2_1})$	$FS(A^{x_1}_{3,3})$
$A^{x_2}_{3,1}$	0	0	0	1	0	0	$\mu^{x_2}_{3,1}(t_{1_2})$	$\mu^{x_2}_{3,1}(t_{2_2})$	$FS(A^{x_2}_{3,1})$
$A^{x_2}_{3,2}$	0	0	0	0	1	0	$\mu^{x_2}_{3,2}(t_{1_2})$	$\mu^{x_2}_{3,2}(t_{2_2})$	$FS(A^{x_2}_{3,2})$
$A^{x_2}_{3,3}$	0	0	0	0	0	1	$\mu^{x_2}_{3,3}(t_{1_2})$	$\mu^{x_2}_{3,3}(t_{2_2})$	$FS(A^{x_2}_{3,3})$

and x_2 are divided into three linguistic values (i.e., $K=3$). Assume that x_2 is the attribute of class labels. Since each row of FG is a bit string consisting of 0 and 1, FG[u] and FG[v] (i.e., uth row and vth row of FG) can be paired to generate certain desired results by applying the Boolean operations. For example, if we apply the *OR* operation on two rows, FG[1] = (1, 0, 0, 0, 0, 0) (i.e., $A^{x_1}_{3,1}$) and FG[4] = (0, 0, 0, 1, 0, 0) (i.e., $A^{x_2}_{3,1}$), then (FG[1] *OR* FG[4]) = (1, 0, 0, 1, 0, 0) corresponding to a candidate 2-dim fuzzy grid $A^{x_1}_{3,1}\times A^{x_2}_{3,1}$ is generated. Then, $FS(A^{x_1}_{3,1}\times A^{x_2}_{3,1}\times A^{x_2}_{3,1}) = A^{x_1}_{3,1}\times A^{x_2}_{3,1}(t_1) + A^{x_1}_{3,1}\times A^{x_2}_{3,1}(t_2) = [\mu^{x_1}_{3,1}(t_{1_1})\cdot\mu^{x_2}_{3,1}(t_{1_2}) + \mu^{x_2}_{3,1}(t_{1_2})\cdot\mu^{x_1}_{3,1}(t_{2_2})\mu^{x_2}_{3,1}(t_{2_2})]/2 = (TT[1]\cdot TT[4])$ is obtained to compare with the min FS. However, any two linguistic values defined in the same attribute cannot be contained in the same candidate k-dim fuzzy grid ($k\geqslant 2$). Therefore, for example, (1, 1, 0, 0, 0, 0) and (0, 0, 0, 1, 0, 1) are invalid.

In the Apriori algorithm, two frequent ($k-1$)-itemsets are joined to be a candidate k-itemset, and these two frequent itemsets share ($k-2$) items. Similarly, two frequent ($k-1$)-dim grids that share ($k-2$) linguistic values can be used to derive a candidate k-dim ($2\leqslant k\leqslant d$) fuzzy grid. For example, if $A^{x_1}_{3,2}\times A^{x_2}_{3,1}$ and $A^{x_1}_{3,2}\times A^{x_3}_{3,3}$ are frequent, then these two grids share $A^{x_1}_{3,2}$ can be used to generate $A^{x_1}_{3,2}\times A^{x_2}_{3,1}\times A^{x_3}_{3,3}$. Then, $A^{x_1}_{3,2}\times A^{x_2}_{3,1}\times A^{x_3}_{3,3}(t_p) = A^{x_1}_{3,2}(t_p)A^{x_2}_{3,1}(t_p)A^{x_3}_{3,3}(t_p)$ is computed.

3.2. Phase II: generate fuzzy classification rules

The general type R of the fuzzy associative classification rule is stated as follows:

$$\text{Rule } R: A^{x_1}_{K_1,i_1}\times A^{x_2}_{K_2,i_2}\times\cdots\times A^{x_{k-1}}_{K_{k-1},i_{k-1}}\times A^{x_k}_{K_k,i_k} \Rightarrow A^{x_\alpha}_{C,i_\alpha} \quad \text{with FC(R)}, \tag{5}$$

where x_α ($1\leqslant\alpha\leqslant d$) is the class label and FC(R) is the fuzzy confidence of rule $A^{x_1}_{K_1,j_1}\times A^{x_2}_{K_2,j_2}\times\cdots\times A^{x_{k-1}}_{K_{k-1},j_{k-1}}\times A^{x_k}_{K_k,j_k}\Rightarrow A^{x_\alpha}_{C,i_\alpha}$. The above rule represents that: if x_1 is $A^{x_1}_{K_1,j_1}$ and x_2 is $A^{x_2}_{K_2,j_2}$ and … and x_k is $A^{x_k}_{K_k,j_k}$, then x_α is $A^{x_\alpha}_{C,i_\alpha}$. The left-hand side of "⇒" is the antecedent part of R, and the right-hand side is the consequent part. FC(R) can be viewed as the grade of certainty of R. Since $(A^{x_1}_{K_1,j_1}\times A^{x_2}_{K_2,j_2}\times\cdots\times A^{x_{k-1}}_{K_{k-1},j_{k-1}}\times A^{x_k}_{K_k,j_k}\times A^{x_\alpha}_{C,i_\alpha})\subseteq(A^{x_1}_{K_1,j_1}\times A^{x_2}_{K_2,j_2}\times\cdots\times A^{x_{k-1}}_{K_{k-1},j_{k-1}}\times A^{x_k}_{K_k,j_k})$ holds, R can be generated by $A^{x_1}_{K_1,j_1}\times A^{x_2}_{K_2,j_2}\times\cdots\times A^{x_{k-1}}_{K_{k-1},j_{k-1}}\times A^{x_k}_{K_k,j_k}\times A^{x_\alpha}_{C,i_\alpha}$ and $A^{x_1}_{K_1,j_1}\times A^{x_2}_{K_2,j_2}\times\cdots\times A^{x_{k-1}}_{K_{k-1},j_{k-1}}\times A^{x_k}_{K_k,j_k}$. We define the fuzzy confidence (Ishibuchi et al., 2001; Hu et al., 2002) of R (i.e., FC(R)) as follows:

$$\text{FC}(R) = \text{FS}\left(A^{x_1}_{K_1,j_1}\times A^{x_2}_{K_2,j_2}\times\cdots\times A^{x_{k-1}}_{K_{k-1},j_{k-1}}\times A^{x_k}_{K_k,j_k}\times A^{x_\alpha}_{C,i_\alpha}\right)/\ \text{FS}\left(A^{x_1}_{K_1,j_1}\times A^{x_2}_{K_2,j_2}\times\cdots\times A^{x_{k-1}}_{K_{k-1},j_{k-1}}\times A^{x_k}_{K_k,j_k}\right). \tag{6}$$

Unlike SFPBM, the proposed method tries to reserve all fuzzy rules because it is not easy to specify an appropriate threshold for users. The user-specified minimum fuzzy confidence (min FC) (Ishibuchi et al., 2001; Hu et al., 2002) is set to zero for simplicity. We still apply Boolean operations to obtain the antecedent part and consequent part of each rule. For example, if there exists FG[u] = (1, 0, 0, 0, 0, 0) and FG[v] = (1, 0, 0, 1, 0, 0) corresponding to frequent fuzzy grids L_u and L_v, where $L_v\subset L_u$, respectively; then FG[u] *AND* FG[v] = (1, 0, 0, 0, 0, 0), corresponding to the frequent fuzzy grid $A^{x_1}_{3,1}$, is generated to be the antecedent part of rule, say R. Then, FG[u] *XOR* FG[v] = (0, 0, 0, 1, 0, 0), corresponding to the frequent fuzzy grid $A^{x_2}_{3,1}$, is generated to be the consequent part of rule R. Then, $FS(A^{x1}_{3,1}\times A^{x2}_{3,1})/FS(A^{x1}_{3,1})$ is easily obtained by (6).

The redundant rules must be further eliminated in order to achieve the goal of compactness (Hu et al., 2002). If there exist two rules R and S, having the same consequent part and the antecedent part of R is contained in that of S, then R is redundant and can be discarded, and S is temporarily reserved. For example, if S is "$A^{x_1}_{K_1,j_1}\times A^{x_2}_{K_2,j_2}\times\cdots\times A^{x_{k-1}}_{K_{k-1},j_{k-1}}\Rightarrow A^{x_\alpha}_{C,i_\alpha}$", then R can be eliminated. This is because that the number of antecedent conditions should be minimized.

On the other hand, for improving the classification performance of fuzzy rule-based systems, Nozaki et al. (1996) proposed the adaptive rules to adjust the grade of certainty of each rule. These useful rules are further incorporated into the proposed methods. The adaptive procedure for adjusting fuzzy confidences is presented as follows:

Set the maximum number of iterations $J_{\max}$.
Set J to be zero.
Repeat
$J = J + 1$
For each training sample t_p do

a. Find the "firing" fuzzy rule R_β.
b. If t_p is correctly classified then FC(R_β) is adjusted as follows:

$$\text{FC}(R_\beta) = \text{FC}(R_\beta) + \eta_1(1-\text{FC}(R_\beta)) \tag{7}$$

otherwise, FC(R_β) is adjusted as follows:

$$\text{FC}(R_\beta) = \text{FC}(R_\beta) + \eta_2\text{FC}(R_\beta) \tag{8}$$

where η_1 and η_2 are learning rates.

End. Until $J = J_{\max}$

The firing rule is found by determining the class label of t_p through the use of fuzzy rules derived by the proposed method. Without losing generality, if the antecedent part of a fuzzy classification rule R_τ is $A^{x_1}_{K_1,i_1} \times A^{x_2}_{K_2,i_2} \times \cdots \times A^{x_\tau}_{K_\tau,i_\tau}$, then we can calculate its firing strength ω_τ for t_p as follows:

$$\omega_\tau = \mu^{x_1}_{K_1,j_1}(t_{p_1})\mu^{x_2}_{K_2,j_2}(t_{p_2})\cdots\mu^{x_\tau}_{K,i_\tau}(t_{p_\tau}) \tag{9}$$

Then t_p can be determined to categorize to the class label which is the consequent part of the "firing" rule, say R_β, if

$$\omega\tau\mathrm{FC}(R_\beta) = \max_j\{\omega_j\mathrm{FC}(R_j)|R_j \in TR\}, \tag{10}$$

where TR is the set of fuzzy rules generated by the proposed method. The adaptive rules are further employed to adjust the fuzzy confidence of R_β. If t_p is correctly classified then $\mathrm{FC}(R_\beta)$ is increased; otherwise, $\mathrm{FC}(R_\beta)$ is decreased. Nozaki et al. (1996) also suggested that the learning rates should be specified as $0<\eta_1 \ll \eta_2 < 1$. Actually, $\eta_1 = 0.001$, $\eta_2 = 0.1$ and $J_{\max} = 500$ are used in the experiment. In the subsequent section, experimental results from the iris data demonstrate the effectiveness of the proposed method. However, the aim of the experiment is to show the feasibility and the problem-solving capability of the proposed method for classification problems. That is, methods about the acquisition of appropriate parameter specifications to obtain higher classification accuracy rates and smaller number of fuzzy if–then rules are not considered in this paper.

4. Experimental results

The classification performances of the proposed method with two fuzzy partition types are examined by computer simulations. We employ the proposed method to find fuzzy classification rules from the iris data that consists of three classes and each class consists of 50 samples. Moreover, class 2 overlaps with class 3. Suppose that the attributes "sepal length", "sepal width", "petal length", and "petal width" are denoted by x_1, x_2, x_3, and x_4 respectively. x_5 denote "class label" (i.e., $d = 5$) to which $t_p = (t_{p_1}, t_{p_2}, \cdots, t_{p_5})$, $(1 \leqslant p \leqslant 150)$ belongs. Only three linguistic values can be defined in x_5; they are $A^{\text{classlabel}}_{3,1}$: "Class 1", $A^{\text{classlabel}}_{3,2}$: "Class 2", and $A^{\text{classlabel}}_{3,3}$: "Class 3" without doubt.

$K = 6$ is first considered for each attribute except x_5. Simulation results with different user-specified minimum supports are shown in Tables 2 and 3 using MTDM and STDM, respectively. Tables 2 and 3 indicate that classification rates are more sensitive to larger min FS (i.e., min FS = 0.18, 0.20). Therefore, the smaller min FS for both MTDM and STDM should be a better choice when all non-redundant rules are reserved. That is, larger min FS may lead to discarding more useful fuzzy grids, thus reducing the effectiveness of fuzzy rules. From Tables 2 and 3, we can see that the best classification rate 100.00% obtained by MTDM is higher than that (i.e., 97.33%) obtained by STDM. In comparison with STDM, MTDM uses more fuzzy if–then rules to classify samples. The best results of SFPBM and the proposed method are also summarized in Table 4. Except for min FS = 0.20, we can see that the best results of the proposed method with MTDM outperforms those of SFPBM with MTDM for each value of min FS. It is obvious that the proposed method with STDM outperforms SFPBM with STDM for each value of min FS.

Simulation results with min FS = 0.05 and different values of K are shown in Tables 5 and 6 with MTDM and STDM, respectively. From Tables 5 and 6, we can see that the classification rates seem not to be sensitive to K for both partition methods. Therefore, it seems that K is not a serious problem from the viewpoint of

Table 2
Simulation results by the proposed method with the MTDM with $K = 6$

Min FS	Classification rate (%)	Number of rules
0.05	100.00	101
0.10	100.00	71
0.15	100.00	48
0.18	96.67	35
0.20	96.00	28

Table 3
Simulation results by the proposed method with the STDM with $K = 6$

Min FS	Classification rate (%)	Number of rules
0.05	97.33	30
0.10	97.33	17
0.15	95.33	11
0.18	93.33	5
0.20	93.33	5

Table 4
Classification rates (%) of SFPBM and the proposed method with $K = 6$ and various min FS

Min FS	Division method			
	MTDM		STDM	
	SFPBM	The proposed method	SFPBM	The proposed method
0.05	96.67	100.00	96.67	97.33
0.10	96.67	100.00	96.67	97.33
0.15	96.67	100.00	92.67	95.33
0.20	96.67	96.00	88.67	93.33

classification rates. The classification rate of the MTDM also outperform that of the STDM for each value of K, and the best result (i.e., 99.33%) from the STDM is slightly worse than that of the MTDM. From Tables 2–6, we can also see that since the min FS and the min FC are not optimized to reduce the number of rules, a large number of rules are generated when the MTDM is used for various K. Although how to set the appropriate values to the min FS and the min FC is a significant work, this topic is not discussed in this paper for simplicity.

Some significant fuzzy if–then rule-based classification systems using simple fuzzy partition methods have been proposed, such as the simple-fuzzy-grid method (Ishibuchi et al., 1992), the multi-rule-table method (Ishibuchi et al., 1992), the pruning method (Nozaki et al., 1996), and the GA-based method (Ishibuchi et al., 1995). In addition, simulation results of the aforementioned methods demonstrated by Nozaki et al. (1996) are summarized in Table 7. The best results of the proposed method with MTDM or STDM are also shown in this table. From the viewpoint of classification rates, we can see that the proposed method with STDM or MTDM works well in comparison with other fuzzy if–then rule-based classifiers. It is noted that the best results of SFPBM with MTDM or STDM can be obtained by setting appropriate values to min FS and min FC (e.g., min FS = 0.10 and min FC = 0.80).

In the above simulation, all 150 samples are used for the training process to generate fuzzy rules. To examine the generalization ability of the proposed method, we perform the leave-one-out technique, which is an almost unbiased estimator of the true error rate of a classifier (Weiss and Kulikowski, 1991). In each iteration of the leave-one-out technique, fuzzy if–then rules are generated from 149 training samples and tested on the single remaining sample. This procedure is iterated until all the given 150 samples are used as a test sample. Now, we try to choose another values of min FS to examine the relationship between min FS and the generalization ability of the proposed method. Simulation results with lower values of min FS (i.e., 0.05, 0.10, 0.15) are shown in Table 8. We can see that the proposed method with MTDM seems not to be sensitive to min FS, and the best classification rate is 96.67%; however, the proposed method with STDM is more sensitive to min FS, and the best classification rate is 95.33%. Therefore, from the viewpoint of the generalization ability, we may conclude that the proposed method with MTDM works more robustly than with STDM does.

Based on the leave-one-out technique, we try to make a comparison between the proposed method and the above-mentioned fuzzy rule-based systems. We summarize the simulation results in Table 9. The best result of the proposed method with MTDM or STDM is also shown in this table. From the viewpoint of classification rates, we can see that the proposed method with MTDM performs well in comparison with other fuzzy if–then

Table 5
Simulation results by the proposed method with the MTDM with various K

K	Classification rate (%)	Number of rules
4	100.00	46
5	100.00	71
6	100.00	101
7	100.00	131

Table 6
Simulation results by the proposed method with the STDM with various K

K	Classification rate (%)	Number of rules
4	97.33	25
5	98.00	25
6	97.33	30
7	99.33	30

Table 7
Simulation results by various fuzzy if-then rule-based classification systems

Method	Classification rate (%)
The proposed method with MTDM	100.00
The proposed method with STDM	99.33
Simple-fuzzy-grid	98.67
Multi-rule-table	95.33
Pruning	100.00
GA-based	99.47
SFPBM with MTDM	96.67
SFPBM with STDM	96.67

Table 8
Classification rates by the leave-one-out technique for MTDM and STDM

Method	Minimum fuzzy support		
	0.05	0.10	0.15
MTDM	95.33	96.67	95.33
STDM	92.67	94.00	95.33

Table 9
Simulation results by the leave-one-out technique for various fuzzy if-then rule-based classification systems

Method	Classification rate (%)
The proposed method with MTDM	96.67
The proposed method with STDM	95.33
Simple-fuzzy-grid	96.67
Multi-rule-table	94.67
Pruning	93.33
GA-based	94.67
SFPBM with MTDM	96.67
SFPBM with STDM	96.67

Table 10
Classification accuracy rates of various fuzzy classification methods for the iris data

Fuzzy methods				
Perceptron criterion 95.33%	Quadratic criterion 96.67%	Minimum operator 96.00%	Fast heuristic search 92.00%	Simulated annealing 91.33%
Fuzzy *k*-nearest neighbor 96.67%	Fuzzy *c*-means 93.33%	Fuzzy *c*-means for histograms 93.33%	Hierarchical fuzzy *c*-means 95.33%	

rule-based classifiers. However, it should be noted that the classification performance of the GA-based method can be highly improved by carefully tuning parameters (e.g., 97.33%). We also find that the best rate (96.67%) of SFPBM with STDM outperforms that of the proposed method with STDM (95.33%). This means that the reservation of all non-redundant rules for the latter method may lead to overfitting.

On the other hand, classification rates of nine fuzzy classification methods, including fuzzy integral with perceptron criterion, fuzzy integral with quadratic criterion, minimum operator, fast heuristic search with Sugeno integral, simulated annealing with Sugeno integral, fuzzy *k*-nearest neighbor, fuzzy *c*-means, fuzzy *c*-means for histograms and hierarchical fuzzy *c*-means, for the iris data estimated by the leave-one-out technique were reported by Grabisch and Dispot (1992). From the summarized results shown in Table 10, we can see that the best result (i.e. 96.67%) was obtained by using the fuzzy integral with quadratic criterion or the fuzzy *k*-NNR method. It is clear that the best result of the proposed method with MTDM (i.e., 96.67%) is equal to the best result of these nine fuzzy methods, whereas the best result of the proposed method with STDM (i.e., 95.33%) is slightly worse than those of the fuzzy integral with quadratic criterion, the minimum operator and the fuzzy *k*-nearest neighbor.

5. Discussions and conclusions

In this paper, we propose a two-phase fuzzy data mining technique that can find fuzzy association rules for classification problems based on SFPBM proposed by Hu et al. (2002). There are three main differences between the proposed method and SFPBM. First, ignored fuzzy subspaces are considered in the proposed method. Second, all non-redundant fuzzy if–then rules take part in the mining process by setting zero to min FC. Specially, adaptive rules proposed by Nozaki et al. (1996) are further incorporated into the proposed method for improving the classification performance. From summarized results shown in Table 4, we can see that the proposed method with STDM or MTDM performs well in comparison with SFPBM with STDM or MTDM.

The generalization ability of the proposed method is examined by the iris data, indicating that best classification rate of the MTDM apparently outperforms that of the STDM. Simulation results with various parameter specifications (i.e., min FS and *K*) also demonstrate that the proposed method may effectively derive fuzzy classification rules.

On the other hand, we do not discuss how to set the appropriate values to the min FS and the min FC for simplicity. Actually, this is a significant work. Since the parameter specification (i.e., min FS and min FC) is not optimized to reduce the number of rules, as we have shown in the previous section, a large number of rules are generated when STDM or MTDM is used for various *K*. Therefore, it is necessary to develop methods such as the genetic algorithms (Goldberg, 1989) to automatically determine the appropriate values of min FS and the min FC to obtain higher classification performances with a compact set of fuzzy if–then classification rules. Then, the proposed method may be viewed as an effective knowledge acquisition tool for classification problems.

Moreover, since fuzzy knowledge representation can facilitate interaction of the expert system and the users (Zimmermann, 1996), it is necessary to extend the proposed method to find other types of fuzzy association rules to ease the fuzzy knowledge acquisition bottleneck in building prototype expert systems or fuzzy rule-based systems. The aforementioned issues are left for future works. Additionally, Hong et al. (2001) discussed the relationship between the computation time and the number of rules for the fuzzy data mining technique. We consider that their study will provide useful suggestions to improve our method.

References

Anderson, E., 1935. The irises of the gaspe peninsula. Bulletin of the American Iris Society 59, 2–5.

Agrawal, R., Imielinski, T., Swami, A., 1993a. Database mining: a performance perspective. IEEE Transactions on Knowledge and Data Engineering 5 (6), 914–925.

Agrawal, R., Imielinski, T., Swami, A., 1993b. Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, Washington, D.C. pp. 207–216.

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A.I., 1996. Fast discovery of association rules. In: Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (Eds.), Advances in Knowledge Discovery and Data Mining. AAAI Press, Menlo Park, pp. 307–328.

Berry, M., Linoff, G., 1997. Data Mining Techniques: For Marketing, Sales, and Customer Support. Wiley, New York.

Chuang, J.H., Wang, P.H., Wu, M.C., 1999. Automatic classification of block-shaped parts based on their 2D projections. Computers and Industrial Engineering 36 (3), 697–718.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading, MA.

Grabisch, M., Dispot, F., 1992. A comparison of some methods of fuzzy classification on real data. In: Proceedings of the Second International Conference on Fuzzy Logic and Neural Networks, Iizuka, Japan, pp.659–662.

Han, J.W., Kamber, M., 2001. Data Mining: Concepts and Techniques. Morgan Kaufmann, San Francisco.

Hong, T.P., Wang, T.T., Wang, S.L., Chien, B.C., 2000. Learning a coverage set of maximally general fuzzy rules by rough sets. Expert Systems with Applications 19 (2), 97–103.

Hong, T.P., Kuo, C.S., Chi, S.C., 2001. Trade-off between computation time and number of rules for fuzzy mining from quantitative data. International Journal of Uncertainty Fuzziness and Knowledge-Based Systems 9 (5), 587–604.

Hu, Y.C., Chen, R.S., Tzeng, G.H., 2002. Mining fuzzy association rules for classification problems. Computers and Industrial Engineering 43 (4), 735–750.

Ishibuchi, H., Nozaki, K., Tanaka, H., 1992. Distributed representation of fuzzy rules and its application to pattern classification. Fuzzy Sets and Systems 52 (1), 21–32.

Ishibuchi, H., Nozaki, K., Yamamoto, N., Tanaka, H., 1995. Selecting fuzzy if–then rules for classification problems using genetic algorithms. IEEE Transactions on Fuzzy Systems 3 (3), 260–270.

Ishibuchi, H., Nakashima, T., Murata, T., 1999. Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems. IEEE Transactions on Systems, Man, and Cybernetics 29 (5), 601–618.

Ishibuchi, H., Yamamoto, T., Nakashima, T., 2001. Fuzzy data mining: effect of fuzzy discretization. In: Proceedings of the First IEEE International Conference on Data Mining, San Jose, USA, pp.241–248.

Jang, J.S.R., 1993. ANFIS: adaptive-network-based fuzzy inference systems. IEEE Transactions on Systems, Man, and Cybernetics 23 (3), 665–685.

Jang, J.S.R., Sun, C.T., 1995. Neuro-fuzzy modeling and control. Proceedings of the IEEE 83 (3), 378–406.

Kim, D., Bang, S.Y., 2000. A handwritten numeral character classification using tolerant rough set. IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (9), 923–937.

Nozaki, K., Ishibuchi, H., Tanaka, H., 1996. Adaptive fuzzy rule-based classification systems. IEEE Transactions on Fuzzy Systems 4 (3), 238–250.

Pedrycz, W., 1994. Why triangular membership functions? Fuzzy Sets and Systems 64, 21–30.

Pedrycz, W., Gomide, F., 1998. An Introduction to Fuzzy Sets: Analysis and Design. MIT Press, Cambridge, MA.

Ravi, V., Zimmermann, H.-J., 2000. Fuzzy rule based classification with FeatureSelector and modified threshold accepting. European Journal of Operational Research 123 (1), 16–28.

Ravi, V., Reddy, P.J., Zimmermann, H.-J., 2000. Pattern classification with principal component analysis and fuzzy rule bases. European Journal of Operational Research 126, 526–533.

Wang, L.X., Mendel, J.M., 1992. Generating fuzzy rules by learning from examples. IEEE Transactions on Systems, Man, and Cybernetics 22 (6), 1414–1427.

Weiss, S.M., Kulikowski, C.A., 1991. Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems. Morgan Kaufmann, Los Altos, CA.

Zadeh, L.A., 1975a. The concept of a linguistic variable and its application to approximate reasoning (Part 1). Information Science 8 (3), 199–249.

Zadeh, L.A., 1975b. The concept of a linguistic variable and its application to approximate reasoning (Part 2). Information Science 8 (4), 301–357.

Zadeh, L.A., 1976. The concept of a linguistic variable and its application to approximate reasoning (Part 3). Information Science 9 (1), 43–80.

Zimmermann, H.-J., 1996. Fuzzy Set Theory and Its Applications. Kluwer, Boston.