



PERGAMON

Computers & Industrial Engineering 43 (2002) 735–750

**computers &
industrial
engineering**

www.elsevier.com/locate/dsw

Mining fuzzy association rules for classification problems

Yi-Chung Hu^a, Ruey-Shun Chen^a, Gwo-Hshiung Tzeng^{b,*}

^a*Institute of Information Management, National Chiao Tung University, Hsinchu 300, Taiwan, ROC*

^b*Institute of Management of Technology, National Chiao Tung University, Hsinchu 300, Taiwan, ROC*

Abstract

The effective development of data mining techniques for the discovery of knowledge from training samples for classification problems in industrial engineering is necessary in applications, such as group technology. This paper proposes a learning algorithm, which can be viewed as a knowledge acquisition tool, to effectively discover fuzzy association rules for classification problems. The consequence part of each rule is one class label. The proposed learning algorithm consists of two phases: one to generate large fuzzy grids from training samples by fuzzy partitioning in each attribute, and the other to generate fuzzy association rules for classification problems by large fuzzy grids. The proposed learning algorithm is implemented by scanning training samples stored in a database only once and applying a sequence of Boolean operations to generate fuzzy grids and fuzzy rules; therefore, it can be easily extended to discover other types of fuzzy association rules. The simulation results from the iris data demonstrate that the proposed learning algorithm can effectively derive fuzzy association rules for classification problems. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Data mining; Knowledge acquisition; Classification problems; Association rules

1. Introduction

Data mining is a methodology for the extraction of new knowledge from data. This knowledge may relate to a problem that we want to solve (Myra, 2000). Thus, data mining can ease the knowledge acquisition bottleneck in building prototype systems (Hong & Chen, 1999; Hong, Wang, Wang, & Chien, 2000). On the other hand, database-mining problems involving classification can be viewed within a common framework of rule discovery (Agrawal, Imielinski, & Swami, 1993). These concepts demonstrate that effective development of data mining techniques to discover knowledge from training samples for classification problems is necessary. Moreover, it is necessary to develop effective methods for classification problems in industrial engineering, such as group technology.

* Corresponding author.

E-mail address: ghtzeng@cc.nctu.edu.tw (G.-H. Tzeng).

Recently, the discovery of association rules from databases has become an important research topic, and association rules have been applied to analyze market baskets to help managers determine, which items are frequently purchased together by customers (Berry & Linoff, 1997; Han & Kamber, 2001; Yilmaz, Triantaphyllou, Chen, & Liao, 2002). Initially, Agrawal, Mannila, Srikant, Toivonen, and Verkamo (1996) proposed the Apriori algorithm to quickly find association rules. Han, Karypis, and Kumar (2000) also proposed parallel data mining techniques implemented in large databases. Generally, there are two phases for mining association rules. In phase I, we find large itemsets, whose supports are larger than or equal to the user-specified minimum support. If there are k items in a large itemset, then we call it a large k -itemset, and the Apriori property shows that any subset of a large itemset must also be large (Han & Kamber, 2001). In phase II, we use large itemsets generated in phase I to generate effective association rules. An association rule is effective, if its confidence is larger than or equal to the user-specified minimum confidence.

In this paper, we propose a learning algorithm to discover fuzzy associative classification rules for classification problems. We define that a fuzzy associative classification rule is a fuzzy if-then rule, whose consequent part is one class label. Since the comprehensibility of fuzzy rules by human users is a criterion in designing a fuzzy rule-based system (Ishibuchi, Nakashima, & Murata, 1999), fuzzy associative classification rules with linguistic interpretation must be taken into account. To cope with this problem, we consider that both quantitative and categorical attributes are divided into many fuzzy partitions by the concept of the fuzzy grids, resulting from fuzzy partitioning in feature space (Ishibuchi, Nozaki, Yamamoto, & Tanaka, 1995; Ishibuchi et al., 1999). Since each fuzzy partition is a fuzzy number, a linguistic interpretation of each fuzzy partition is easily obtained.

Each fuzzy partition distributed in either quantitative or categorical attributes is viewed as a candidate one-dimension (1-dim) fuzzy grid used to generate large k -dim ($k \geq 1$) fuzzy grids. We give the definitions of the fuzzy support and the fuzzy confidence to determine, which candidate fuzzy grids are large and which fuzzy rules are effective, respectively. The proposed learning algorithm also consists of two phases: one to generate large fuzzy grids from training samples by fuzzy partitioning in each attribute, and the other to generate the fuzzy associative classification rules from these large fuzzy grids. The proposed learning algorithm is implemented by scanning training samples stored in databases only once and applying a sequence of Boolean operations to generate fuzzy grids and fuzzy associative classification rules. Therefore, it can be easily extended to discover other types of fuzzy association rules for market basket analysis. The well-known iris data proposed by Fisher (1936) is often used to compare the performance between the proposed learning algorithm and other classification methods, such as the genetic-algorithm-based method (Ishibuchi et al., 1995), and some results are reported by Grabisch and Dispot (1992). The simulation results reported in this paper demonstrate that the proposed learning algorithm works well in comparison with other classification methods. Therefore, the proposed learning algorithm may effectively derive fuzzy associative classification rules; moreover, the goal of knowledge acquisition can also be easily achieved.

This paper is organized as follows. The concepts of fuzzy partitions are introduced in Section 2. In Section 3, we give definitions for the fuzzy support and the fuzzy confidence, and the proposed learning algorithm is also presented in this section. In Section 4, the performance of the proposed learning algorithm is examined by computer simulation on the iris data. Discussions and conclusions are presented in Sections 5 and 6, respectively.

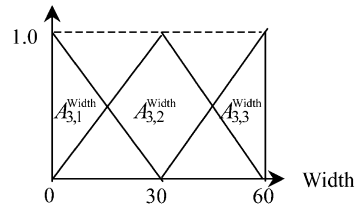


Fig. 1. $K' = 3$ for Width.

2. Fuzzy partitions

The notation used in this paper is as follows:

- C total number of class labels
- d total number of data attributes, where $1 \leq d$
- k dimension of one fuzzy grid, where $1 \leq k \leq d$
- K' maximal number of fuzzy partitions in each quantitative attribute
- $A_{K',i_m}^{x_m}$ i_m -th fuzzy partition of K' fuzzy partitions defined in attribute x_m , where $1 \leq m \leq d$, $3 \leq K' \leq K'$ and $1 \leq i \leq K'$
- $\mu_{K',i_m}^{x_m}$ membership function of $A_{K',i_m}^{x_m}$
- t_p p -th sample or tuple, where $t_p = (t_{p_1}, t_{p_2}, \dots, t_{p_d})$, and $p \geq 1$

Fuzzy sets were proposed by Zadeh (1965), who also proposed the concepts of a linguistic variable and its applications to approximate reasoning (Zadeh, 1975a,b,c, 1976). A linguistic variable is a variable, whose values are linguistic words or sentences in a natural language (Chen & Jong, 1997). The division of the attributes into many fuzzy partitions has been also widely used in pattern recognition and fuzzy reasoning. Examples are the application by Ishibuchi et al. to pattern classification (Ishibuchi, Nozaki, & Tanaka, 1992; Ishibuchi et al., 1995; Ishibuchi, Murata, & Gen, 1998; Ishibuchi et al., 1999), and fuzzy rules generation by Wang and Mendel (1992). In addition, some methods for partitioning an attribute space were discussed by Sun (1994).

The discussions for fuzzy partitioning in quantitative and categorical attributes are introduced in Sections 2.1 and 2.2, respectively.

2.1. Fuzzy partitioning in quantitative attributes

The proposed learning algorithm includes two methods resulting from fuzzy partitions: one is the M -type (multiple type) division method (MTDM), and the other is the S -type (single type) division method (STDM). The MTDM allows us to divide each quantitative attribute into $(3 + 4 + \dots + K')$ fuzzy partitions. That is, we sequentially divide each quantitative attribute into 3, 4, ..., K' fuzzy partitions. As for the STDM, only K' fuzzy partitions are defined. In these two methods, K' is pre-specified before executing the proposed learning algorithm. Triangular membership functions are used for the fuzzy partitions defined in the quantitative attributes. Hence, fuzzy partitions are fuzzy numbers, whereas a fuzzy number is a fuzzy partition in the universe of discourse, that is, both convex and normal (Chen & Jong, 1997). For example, by using the STDM, we describe $K' = 3$ and $K' = 4$ for attribute ‘Width’ (denoted by x_1) that range from 0 to 60 in Figs. 1 and 2, respectively. Then, μ_{K',i_1}^{Width} can be represented as

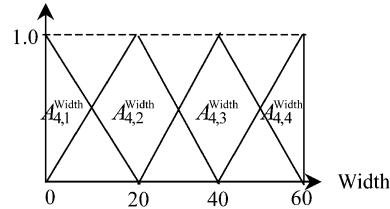


Fig. 2. $K' = 4$ for Width.

follows:

$$\mu_{K,i_1}^{Width}(x) = \max\{1 - |x - a_{i_1}^K|/b^K, 0\} \tag{1}$$

where

$$a_{i_1}^K = mi + (ma - mi)(i_1 - 1)/(K - 1) \tag{2}$$

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

where ma is the maximal value of the domain, and mi is the minimal value. Here, $ma = 60$ and $mi = 0$ for Width. Moreover, if we view Width as a linguistic variable, then the linguistic term A_{K,i_1}^{Width} can be described in the sentences with different i_1 :

$$A_{K,1}^{Width} : \text{Width is small, and below } 60/(K - 1) \tag{4}$$

$$A_{K,K}^{Width} : \text{Width is large, and above } [60 - 60/(K - 1)] \tag{5}$$

$$A_{K,i_1}^{Width} : \text{Width is close to } (i_1 - 1)[60 - 60/(K - 1)], \text{ and between } (i_1 - 2)[60 - 60/(K - 1)] \tag{6}$$

and $i_1[60 - 60/(K - 1)]$, for $1 < i_1 < K$

Clearly, the set of candidate 1-dim fuzzy grids generated for the same K' by the STDm is contained in the one generated by the MTDM. For example, when $K' = 4$, then $\{A_{4,1}^{Width}, A_{4,2}^{Width}, A_{4,3}^{Width}, A_{4,4}^{Width}\}$ is generated by the STDm, and $\{A_{3,1}^{Width}, A_{3,2}^{Width}, A_{3,3}^{Width}, A_{4,1}^{Width}, A_{4,2}^{Width}, A_{4,3}^{Width}, A_{4,4}^{Width}\}$ is generated by the MTDM. If we divide both Width and ‘Length’ (denoted by x_2) into three fuzzy partitions, then a feature space is divided into nine 2-dim fuzzy grids, as shown in Fig. 3. For the shaded 2-dim fuzzy grid shown in Fig. 3, we can use $A_{3,1}^{Width} \times A_{3,3}^{Length}$ to represent it.

Note that, Ishibuchi et al. (1999) proposed another method to define fuzzy sets for discrete values of

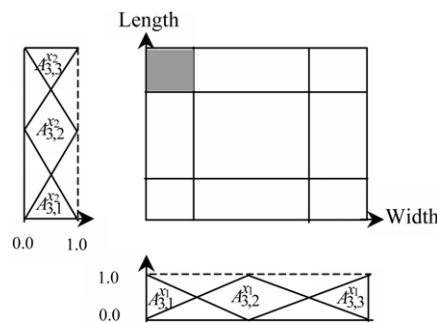


Fig. 3. Divide both Width and Length into three fuzzy partitions.

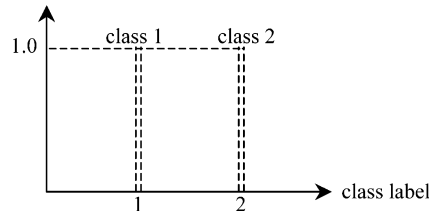


Fig. 4. Class label is divided into two fuzzy partitions.

quantitative attributes. For example, if one quantitative attribute has two attribute values {0, 1}, then we may use the fuzzy sets ‘small’ and ‘large’ with degrees 0.0 and 1.0, respectively.

2.2. Fuzzy partitioning in categorical attributes

If the number of distinctly categorical attribute values is equal to n' (where n' is finite), then this attribute can only be divided into n' fuzzy partitions. At first, we view a categorical attribute as a quantitative attribute. That is, each value of the categorical attribute can correspond to an integer number. This is helpful for us in dividing the categorical attribute. A linguistic term $A_{n',i_m}^{x_m}$ ($1 \leq i_m \leq n'$) is defined in the partition $(i_m - \varepsilon, i_m + \varepsilon)$ ($\varepsilon \rightarrow 0$). The membership function of $A_{n',i_m}^{x_m}$ is 1.

For example, ‘class label’ is a linguistic variable, and suppose its values include ‘class 1’ and ‘class 2’ which correspond to 1 and 2, respectively. The result of partitions is shown in Fig. 4. We can see that there are two fuzzy partitions distributed in class label, one is $(1 - \varepsilon, 1 + \varepsilon)$ and the other is $(2 - \varepsilon, 2 + \varepsilon)$. Linguistic terms $A_{2,1}^{\text{class label}}$ and $A_{2,2}^{\text{class label}}$ can be interpreted as Eqs. (7) and (8), respectively. The membership functions can be described in Eqs. (9) and (10). Sometimes, we can divide a quantitative attribute with discrete values in this way (e.g. number of cars which a person owns)

$$A_{2,1}^{\text{class label}} : \text{class label is class 1} \tag{7}$$

$$A_{2,2}^{\text{class label}} : \text{class label is class 2} \tag{8}$$

$$\mu_{2,1}^{\text{class label}}(x) = 1, \quad 1 - \varepsilon \leq x \leq 1 + \varepsilon, \quad \varepsilon \rightarrow 0 \tag{9}$$

$$\mu_{2,2}^{\text{class label}}(x) = 1, \quad 2 - \varepsilon \leq x \leq 2 + \varepsilon, \quad \varepsilon \rightarrow 0 \tag{10}$$

Initially, each fuzzy partition distributed in either quantitative or categorical attributes is viewed as a candidate 1-dim fuzzy grid. The next important task is how to use the candidate 1-dim fuzzy grids to generate the other large fuzzy grids and fuzzy associative classification rules. Therefore, we propose the learning algorithm described in Section 3.

3. Mining fuzzy associative classification rules

In this section, we first describe the learning model for generating fuzzy associative classification rules in Fig. 5. From this figure, we can see that fuzzy associative classification rules are generated in the two phases of the proposed algorithm. Large fuzzy grids and effective fuzzy associative classification rules are generated in phases I and II, respectively. We describe the individual phases of the learning model in Sections 3.1 and 3.2. The proposed learning approach is presented in Section 3.2.

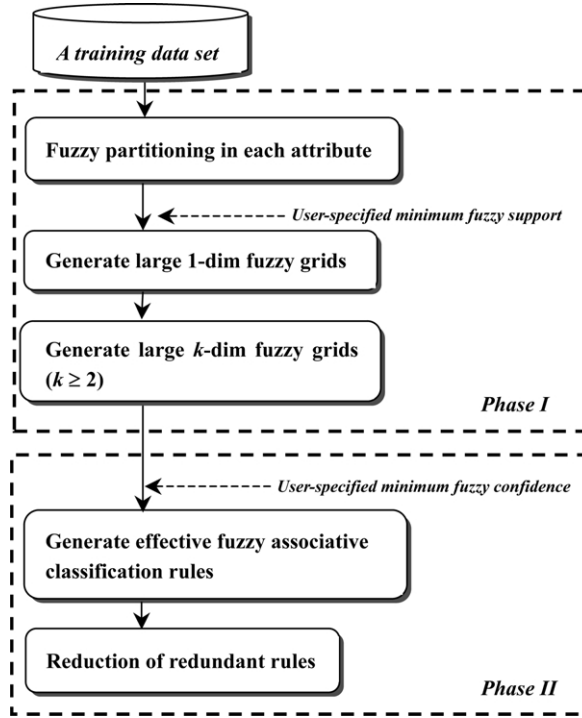


Fig. 5. Learning model for generating fuzzy associative classification rules.

3.1. Phase 1: generate large fuzzy grids

Suppose each attribute, x_m , is divided into K' fuzzy partitions. Without loss of generality, given a candidate k -dim fuzzy grid $A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \dots \times A_{K,i_{k-1}}^{x_{k-1}} \times A_{K,i_k}^{x_k}$, where $1 \leq i_1, i_2, \dots, i_k \leq K$, $3 \leq K \leq K'$ for the MTDM and $K = K'$ for the STDm, the degree to which t_p belongs to this fuzzy grid can be computed as $\mu_{K,i_1}^{x_1}(t_{p_1})\mu_{K,i_2}^{x_2}(t_{p_2}) \dots \mu_{K,i_{k-1}}^{x_{k-1}}(t_{p_{k-1}})\mu_{K,i_k}^{x_k}(t_{p_k})$. To check whether this fuzzy grid is large or not, we define the fuzzy support $FS(A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \dots \times A_{K,i_{k-1}}^{x_{k-1}} \times A_{K,i_k}^{x_k})$ as follows:

$$FS(A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \dots \times A_{K,i_{k-1}}^{x_{k-1}} \times A_{K,i_k}^{x_k}) = \left[\sum_{p=1}^n \mu_{K,i_1}^{x_1}(t_{p_1})\mu_{K,i_2}^{x_2}(t_{p_2}) \dots \mu_{K,i_{k-1}}^{x_{k-1}}(t_{p_{k-1}})\mu_{K,i_k}^{x_k}(t_{p_k}) \right] / n \quad (11)$$

where n is the number of training samples. When $FS(A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \dots \times A_{K,i_{k-1}}^{x_{k-1}} \times A_{K,i_k}^{x_k})$ is larger than or equal to the user-specified minimum fuzzy support (min FS), we can say that $A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \dots \times A_{K,i_{k-1}}^{x_{k-1}} \times A_{K,i_k}^{x_k}$ is a large k -dim fuzzy grid. This is similar to defining a large k -itemset, whose support is larger than the user-specified minimum support.

Table $FGTTFS_K$ is implemented to generate large fuzzy grids for K . $FGTTFS$ consists of the following substructures:

- (a) Fuzzy grids table (FG_K) : each row represents a fuzzy grid, and each column represents a fuzzy partition $A_{K,i_m}^{x_m}$.

Table 1
Initial table FGTTFS₃

Fuzzy grid	FG ₃						TT ₃		FS ₃
	A _{3,1} ^{x₁}	A _{3,2} ^{x₁}	A _{3,3} ^{x₁}	A _{3,1} ^{x₂}	A _{3,2} ^{x₂}	A _{3,3} ^{x₂}	t ₁	t ₂	
A _{3,1} ^{x₁}	1	0	0	0	0	0	μ _{3,1} ^{x₁} (t ₁)	μ _{3,1} ^{x₁} (t ₂)	FS(A _{3,1} ^{x₁})
A _{3,2} ^{x₁}	0	1	0	0	0	0	μ _{3,2} ^{x₁} (t ₁)	μ _{3,2} ^{x₁} (t ₂)	FS(A _{3,2} ^{x₁})
A _{3,3} ^{x₁}	0	0	1	0	0	0	μ _{3,3} ^{x₁} (t ₁)	μ _{3,3} ^{x₁} (t ₂)	FS(A _{3,3} ^{x₁})
A _{3,1} ^{x₂}	0	0	0	1	0	0	μ _{3,1} ^{x₂} (t ₁)	μ _{3,1} ^{x₂} (t ₂)	FS(A _{3,1} ^{x₂})
A _{3,2} ^{x₂}	0	0	0	0	1	0	μ _{3,2} ^{x₂} (t ₁)	μ _{3,2} ^{x₂} (t ₂)	FS(A _{3,2} ^{x₂})
A _{3,3} ^{x₂}	0	0	0	0	0	1	μ _{3,3} ^{x₂} (t ₁)	μ _{3,3} ^{x₂} (t ₂)	FS(A _{3,3} ^{x₂})

- (b) Transaction table (TT_K) : each column represents t_p, while each element records the membership degree of the corresponding fuzzy grid.
- (c) Column FS_K : stores the fuzzy support corresponding to the fuzzy grid in FG.

An initial tabular FGTTFS₃ is shown as Table 1 as an example, from which we can see that there are two tuples t₁ and t₂, and two attributes x₁ and x₂ in a given database. Both x₁ and x₂ are divided into three fuzzy partitions (i.e. K' = 3). In the learning process for K' = 4, this means that FGTTFS₃ and FGTTFS₄ will be used for M-type division and only FGTTFS₄ for S-type division. Assume that x₂ is the attribute of class labels. Since each row of FG is a bits string consisting of 0 and 1, FG_K[u] and FG_K[v] (i.e. u-th row and v-th row of FG_K) can be paired to generate certain desired results by applying the Boolean operations. For example, if we apply the OR operation on two rows, FG₃[1] = (1, 0, 0, 0, 0, 0) and FG₃[4] = (0, 0, 0, 1, 0, 0), then (FG₃[1] OR FG₃[4]) = (1, 0, 0, 1, 0, 0) corresponding to a candidate 2-dim fuzzy grid A_{3,1}^{x₁} × A_{3,1}^{x₂}, is generated. Then, FS(A_{3,1}^{x₁} × A_{3,1}^{x₂}) = TT₃[1]TT₃[4] = [μ_{3,1}^{x₁}(t₁)μ_{3,1}^{x₂}(t₁) + μ_{3,1}^{x₁}(t₂)μ_{3,1}^{x₂}(t₂)]/2 is obtained to compare with the min FS.

However, any two fuzzy partitions defined in the same attribute cannot be contained in the same candidate k-dim fuzzy grid (k ≥ 2). Therefore, (1,1,0,0,0,0) and (0,0,0,1,1,0) are all invalid. To solve this problem, we implement a 1-dim array Group of Fuzzy Grids (GFG_K). From GFG_K, we can easily distinguish, which fuzzy partitions are defined in the same attribute. Each index of GFG_K corresponds to a fuzzy partition, and fuzzy partitions defined in the same attribute must be set to the same integer number. GFG₃ is shown as an example Table 2. For example, since GFG₃[1] = GFG₃[2] = 1, string (1,1,0,0,0,0) generated by FG₃[1] OR FG₃[2] = (1, 0, 0, 0, 0, 0) OR (0, 1, 0, 0, 0, 0) = (1, 1, 0, 0, 0, 0) is invalid.

In the Apriori algorithm (Agrawal et al., 1996), two large (k - 1)-itemsets are joined to be a candidate k-itemset, and these two large itemsets share (k - 2) items. Similarly, a candidate k-dim (3 ≤ k ≤ d) fuzzy grid is derived by joining two large (k - 1)-dim fuzzy grids, and these two large grids share (k - 2) fuzzy partitions. For example, we can use A_{3,2}^{x₁} × A_{3,1}^{x₂} and A_{3,2}^{x₁} × A_{3,3}^{x₂} to generate the candidate 3-dim fuzzy grid A_{3,2}^{x₁} × A_{3,1}^{x₂} × A_{3,3}^{x₂}, because A_{3,2}^{x₁} × A_{3,1}^{x₂} and A_{3,2}^{x₁} × A_{3,3}^{x₂} share the linguistic term A_{3,2}^{x₁}. However, A_{3,2}^{x₁} × A_{3,1}^{x₂} × A_{3,3}^{x₂} can also be generated by joining A_{3,2}^{x₁} × A_{3,1}^{x₂} to A_{3,1}^{x₁} × A_{3,3}^{x₂}. This implies that we must select one of many possible combinations to avoid redundant computations. To cope with this problem, the method we adopt here is that if there exist integers 1 ≤ e₁ < e₂ < ... < e_k ≤ d, such that FG_K[u, e₁] = FG_K[u, e₂] = ... = FG_K[u, e_{k-2}] = FG_K[u, e_{k-1}] = 1 and FG_K[v, e₁] = FG_K[v, e₂] = ... = FG_K[v, e_{k-2}] = FG_K[v, e_{k-1}] = 1, where FG_K[u] and FG_K[v] correspond to large (k - 1)-dim fuzzy grids, then FG_K[u] and FG_K[v] can be paired to generate a candidate k-dim fuzzy grid.

Table 2
One-dimensional array GFG₃ (group of fuzzy grids)

Index					
[1]	[2]	[3]	[4]	[5]	[6]
1	1	1	2	2	2

3.2. Phase 2: generate effective fuzzy associative classification rules

The general type of the fuzzy associative classification rule R is stated as Eq. (12)

$$\text{Rule } R : A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \dots \times A_{K,i_{k-1}}^{x_{k-1}} \times A_{K,i_k}^{x_k} \Rightarrow A_{C,i_\alpha}^{x_\alpha} \text{ with FC}(R) \tag{12}$$

where x_α ($1 \leq \alpha \leq d$) is the class label and $\text{FC}(R)$ is the fuzzy confidence of rule ' $A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \dots \times A_{K,i_k}^{x_k} \Rightarrow A_{C,i_\alpha}^{x_\alpha}$ '. The above rule represents that: if x_1 is $A_{K,i_1}^{x_1}$ and x_2 is $A_{K,i_2}^{x_2}$, ... and x_k is $A_{K,i_k}^{x_k}$, then x_α is $A_{C,i_\alpha}^{x_\alpha}$. The left-hand-side of ' \Rightarrow ' is the antecedent part of R , and the right-hand-side is the consequent part. $\text{FC}(R)$ can be viewed as the grade of certainty of R . R is generated by two large fuzzy grids, one is $A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \dots \times A_{K,i_{k-1}}^{x_{k-1}} \times A_{K,i_k}^{x_k} \times A_{C,i_\alpha}^{x_\alpha}$ and the other is $A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \dots \times A_{K,i_{k-1}}^{x_{k-1}} \times A_{K,i_k}^{x_k}$. We define the fuzzy confidence $\text{FC}(R)$ of R as follows:

$$\text{FC}(R) = \text{FS}(A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \dots \times A_{K,i_{k-1}}^{x_{k-1}} \times A_{K,i_k}^{x_k} \times A_{C,i_\alpha}^{x_\alpha}) / \text{FS}(A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \dots \times A_{K,i_{k-1}}^{x_{k-1}} \times A_{K,i_k}^{x_k}) \tag{13}$$

If $\text{FC}(R)$ is larger than or equal to the user-specified minimum fuzzy confidence (min FC), then R is effective and can be reserved. This is similar to defining an effective association rule, whose confidence is larger than or equal to the user-specified minimum confidence. We still apply Boolean operations to obtain the antecedent part and consequent part of each fuzzy rule. For example, if there exists $\text{FG}_3[u] = (1, 0, 0, 0, 0, 0)$ and $\text{FG}_3[v] = (1, 0, 0, 1, 0, 0)$ corresponding to large fuzzy grids L_u and L_v , where $L_v \subset L_u$, respectively; then $\text{FG}_3[u] \text{ AND } \text{FG}_3[v] = (1, 0, 0, 0, 0, 0)$, corresponding to the large fuzzy grid $A_{3,1}^{x_1}$, is generated to be the antecedent part of rule R . Then, $\text{FG}_3[u] \text{ XOR } \text{FG}_3[v] = (0, 0, 0, 1, 0, 0)$, corresponding to the large fuzzy grid $A_{3,1}^{x_2}$, is generated to be the consequent part of rule R . Then, $\text{FC}(R) = \text{FS}(A_{3,1}^{x_1} \times A_{3,1}^{x_2}) / \text{FS}(A_{3,1}^{x_1})$ is obtained to compare with the min FC to determine R is effective or not.

However, some redundant rules must be eliminated from the viewpoint in order to achieve the goal of compactness. If there exist two rules R and S , having the same consequent part, and the antecedent part of S is contained in that of R , then R is redundant and can be discarded, and S is temporarily reserved. For example, if S is ' $A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \dots \times A_{K,i_{k-1}}^{x_{k-1}} \Rightarrow A_{C,i_\alpha}^{x_\alpha}$ ', then R can be eliminated. This can happen because the number of antecedent conditions must be minimized.

The main difference between the MTDM with the STDM is that the MTDM needs to calculate the total number of fuzzy partitions in each quantitative attribute, but STDM does not. Therefore, the proposed learning algorithms for MTDM and STDM are almost the same. We describe the general algorithm as following.

Algorithm: learning algorithm for mining fuzzy associative classification rules

Input:

- a. A set of training samples selected from the specified classification problem

- b. The minimum fuzzy support and the minimum fuzzy confidence are user-specified
- c. K'

Output:

Phase I: Generate large fuzzy grids

Phase II: Generate effective fuzzy associative classification rules

Method:

Phase I. Generate large fuzzy grids

Step 1. Fuzzy partitioning in each attribute

Divide each quantitative into fuzzy partitions by M -type or S -type division, and the number n' is also pre-determined for the categorical attribute class label of the specified classification problem (e.g. $n' = 3$ for the iris data).

Step 2. Scan the training samples from a database, and then construct the initial table FGTTFS and GFG

Step 3. Generate large fuzzy grids

For FGTTFS $_K$ ($K = 1, 2, \dots, K'$ for M -type division and $K = K'$ for S -type division) do

3-1. Generate large 1-dim fuzzy grids

Set $k = 1$ and eliminate the rows of initial FGTTFS $_K$ corresponding to candidate 1-dim fuzzy grids that are not large.

3-2. Generate large k -dim fuzzy grids

Set $k + 1$ to k . If there is only one $(k - 1)$ -dim fuzzy grid, then go to phase II.

For two unpaired rows, FGTTFS $_K[u]$ and FGTTFS $_K[v]$ ($u \neq v$), corresponding to large $(k - 1)$ -dim fuzzy grids do

Compute (FG $_K[u]$ OR FG $_K[v]$) corresponding to a candidate k -dim fuzzy grid c .

3-2-1. From non-zero elements of (FG $_K[u]$ OR FG $_K[v]$), retrieve all corresponding values from GFG $_K$. If any two values are the same, then discard c and skip Steps 3-2-2, 3-2-3, and 3-2-4. That is, c is invalid.

3-2-2. If FG $_K[u]$ and FG $_K[v]$ do not share $(k - 2)$ linguistic terms, then discard c and skip Steps 3-2-3 and 3-2-4. That is, c is invalid.

3-2-3. If there exist integers $1 \leq e_1 < e_2 < \dots < e_k \leq d$ such that (FG $_K[u]$ OR FG $_K[v])[e_1] =$ (FG $_K[u]$ OR FG $_K[v])[e_2] = \dots =$ (FG $_K[u]$ OR FG $_K[v])[e_{k-1}] =$ (FG $_K[u]$ OR FG $_K[v])[e_k] = 1$, then compute (TT $_K[e_1]$ TT $_K[e_2] \cdot \dots \cdot$ TT $_K[e_k])$ and the fuzzy support fs of c .

3-2-4. Add (FG $_K[u]$ OR FG $_K[v]$) to FG $_K$, (TT $_K[e_1]$ TT $_K[e_2] \cdot \dots \cdot$ TT $_K[e_k])$ to TT $_K$ and fs to FS $_K$ when fs is larger than the minimum fuzzy support; otherwise, discard c .

End

3-3. Check whether any large k -dim fuzzy grid is generated or not

If any large k -dim fuzzy grid is generated, then go to Step 3-2. Note that the final FGTTFS $_K$ stores only large fuzzy grids.

End

Phase II: Generate effective fuzzy associative classification rules

For FGTTFS $_K$ ($K = 1, 2, \dots, K'$ for M -type division and $K = K'$ for S -type division) do

Step 1. Generate effective fuzzy rules

For two unpaired rows, FG $_K[u]$ and FG $_K[v]$ ($u < v$), corresponding to large fuzzy grids L_u and L_v respectively do

1-1. Generate the antecedent part of fuzzy rule R_v .

Compute the number of non-zero elements in FG to $temp$.

If the number of non-zero elements in $(FG_K[u] \text{ AND } FG_K[v])$ is equal to $temp$, then $L_v \subset L_u$ is hold, and the antecedent part of R_v is generated as $(FG_K[u] \text{ AND } FG_K[v]) = L_u$; else skip Steps 1-2 and 1-3.

1-2. Generate the consequent part of fuzzy rule R_v .

Use $(FG_K[u] \text{ XOR } FG_K[v])$ to obtain the consequent part.

If $(FG_K[u] \text{ XOR } FG_K[v])$ contains only one fuzzy partition defined in the class label then generate R_v ; else skip Step 1-3.

1-3. Check whether R can be reserved or not.

$FC(R_v) = FS(L_v)/FS(L_u)$

If $FC(R_v) \geq \min FC$, then mark $FG_K[v]$ to represent that R_v and $FC(R_v)$ are reserved; else discard R_v .

End

Step 2. Reduce redundant rules

For any two marked rows $FG_K[u]$ and $FG_K[v]$ ($u < v$) corresponding to effective fuzzy rules R_u and R_c respectively do

If $FG_K[u] = (FG_K[u] \text{ AND } FG_K[v])$, then unmark $FG_K[v]$.

End

End

Those marked rows of FG_K are used for classification problems. The performance of the proposed learning algorithm is mainly dependent on the size of candidate grids and the size of large grids for phases I and II, respectively. Clearly, the proposed learning algorithm is implemented by scanning training samples stored in a database only once and applying a sequence of Boolean operations to generate fuzzy grids and fuzzy rules. In Section 4, some simulation results are presented to demonstrate the effectiveness of the proposed learning algorithm.

4. Experiments

In this section, the performance of the proposed learning algorithm is examined. We employ the proposed algorithm to discover fuzzy associative classification rules from the well-known iris data proposed by Fisher (1936). The computer programs were coded using the Delphi ver. 5.0 system and were executed on a personal computer with Pentium III-500 CPU and 128 MB RAM running Win98.

The iris data consists of three classes (Class 1: Iris setosa, Class 2: Iris versicolor, and Class 3: Iris virginica) and each class consists of 50 data points (Ishibuchi et al., 1995). Moreover, Class 2 overlaps with Class 3. These data are stored in a relational database. Suppose that attribute x_1 is the sepal length, attribute x_2 is the sepal width, attribute x_3 is the petal length, attribute x_4 is the petal width, and attribute x_5 is the class label (i.e. $n' = 3$ for x_5 is determined) to which $t_p = (t_{p1}, t_{p2}, \dots, t_{p5})$ ($1 \leq p \leq 150$) belongs. The pairs (ma,mi) for x_1, x_2, x_3 , and x_4 are (79,43), (44,20), (69,10), and (25,1), respectively. Although the iris data is a crisp data set, the fuzzy rules can be still extracted from these 150 data points by using the proposed learning algorithm.

Now, we determine the class label of t_p by applying the proposed learning algorithm to classify the iris data. Without losing generality, if the antecedent part of a fuzzy associative classification rule R_τ is $A_{K,i_1}^{x_1} \times A_{K,i_2}^{x_2} \times \dots \times A_{K,i_r}^{x_r}$, then we can calculate the compatibility grade $\mu_\tau(t_p)$ of t_p as $\mu_{K,i_1}^{x_1}(t_{p1}) \mu_{K,i_2}^{x_2}(t_{p2}) \dots \mu_{K,i_r}^{x_r}(t_{pr})$. Then, t_p can be determined to categorize the class label, which is the

Table 3
Classification accuracy rate (%) by M -type division

Min fuzzy confidence	Min fuzzy support			
	0.05	0.10	0.15	0.20
0.50	96.67	96.67	94.00	92.67
0.55	96.67	96.67	94.00	91.33
0.60	96.67	96.67	94.00	91.33
0.65	96.67	96.67	94.00	91.33
0.70	96.67	96.67	94.00	91.33
0.75	96.67	96.67	94.00	94.00
0.80	96.67	96.67	96.67	96.67
0.85	96.67	96.67	96.67	96.67
0.90	96.00	96.67	96.00	66.00

consequent part of R_β , when

$$\mu_\beta(t_p)FC(R_\beta) = \max_j \{ \mu_j(t_p)FC(R_j) | R_j \in TR \} \quad (14)$$

where TR is the set of fuzzy rules generated by the proposed learning algorithm.

First, we consider $K' = 6$ for each attribute except x_5 . No doubt that only three fuzzy partitions can be defined in x_5 ; they are $A_{3,1}^{\text{class label}}$: ' t_p belongs to Class 1', $A_{3,2}^{\text{class label}}$: ' t_p belongs to Class 2', and $A_{3,3}^{\text{class label}}$: ' t_p belongs to Class 3'. Simulation results with different user-specified minimum support and confidence are shown in Tables 3 and 4 using the MTDM and the STDm, respectively. From Tables 3 and 4, we can see that classification accuracy rates are more sensitive to larger min FS. Therefore, the smaller min FS could be a better choice. On the other hand, rules generated by the MTDM are more robust than those generated by the STDm with respect to different parameter specifications. For example, rules generated by the MTDM work well for min FS = 0.15 with different min FC; however, much lower rates are obtained by rules generated by the STDm, when min FC is larger than 0.75.

From Tables 3 and 4, we can see that the best classification accuracy rate 96.67% is simultaneously obtained in both tables with (min FS, min FC) = (0.05,0.85) and (0.10,0.80); hence, these best parameter specifications are used in subsequent simulations. For different values of K , we show the simulation results in Tables 5 and 6 with (0.05,0.85) and (0.10,0.80), respectively. From Tables 5 and 6, we can see that the classification accuracy rates are not sensitive to larger values of K' (i.e. $K' = 6-8$). For comparison, we show simulation results with the same best parameter specifications by the STDm in Tables 7 and 8. From the comparison of Table 7 with Table 5 and Table 8 with Table 6, we can see that the simulation results of the MTDM are more robustly than those of the STDm with respect to larger values of K' (i.e. $K' = 6-8$). By the same parameter specification, the training time of the STDm is shorter than that of the MTDM for all values of K' .

In the above simulation, all 150 data are used as training samples to generate fuzzy rules. To examine the error rate of the proposed learning algorithm for testing samples, we perform the leaving-one-out technique, which is an almost unbiased estimator of the true error rate of a classifier (Weiss & Kulikowski, 1991). In each iteration of the leaving-one-out technique, fuzzy rules are generated from 149 training samples and tested on the single remaining data. This procedure is iterated, until all the

Table 4
Classification accuracy rate (%) by *S*-type division

Min fuzzy confidence	Min fuzzy support			
	0.05	0.10	0.15	0.20
0.50	95.33	94.67	92.67	88.67
0.55	95.33	94.67	92.67	88.67
0.60	95.33	96.00	92.67	88.67
0.65	95.33	96.00	92.67	88.67
0.70	95.33	96.00	92.67	88.67
0.75	95.33	96.00	66.00	66.00
0.80	96.00	96.67	66.67	66.00
0.85	96.67	94.67	66.67	66.00
0.90	94.67	94.67	66.67	66.00

given 150 data are used as a test sample. Classification rates with different parameter specifications are shown in Table 9. From Table 9, we can see that the best result obtained by the MTDM and the STDM are all 96.67%. However, some poor results (e.g. 91.33 and 92.67% for $K' = 7$) are obtained by the STDM resulting from overfitting the training samples.

Based on the leaving-one-out technique, we try to make a comparison between the proposed learning algorithm and other fuzzy classification methods. Previously, Ishibuchi et al. (1995) proposed a genetic-algorithm-based method to select the fuzzy classification rules from the iris data. They demonstrated that the classification rate was 94.67% by the leaving-one-out technique. Some proper parameters, including the stopping condition (1000 generations), population size, and biased mutation probability, were also needed. It is clear that the best result (i.e. 96.67%) of the proposed learning algorithm outperforms that of Ishibuchi et al.'s genetic-algorithm-based method.

Previously, error rates of nine fuzzy classification methods, including fuzzy integral with perception criterion, fuzzy integral with the quadratic criterion, minimum operator, fast heuristic search with the Sugeno integral, simulated annealing with the 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 leaving-one-out technique were reported by Grabisch and Dispot (1992). The best result (i.e. 96.67%) was obtained by using the fuzzy integral with the quadratic criterion or the fuzzy k -NMR method. It is clear that the best result of the proposed learning algorithm is equal to the best result of these nine fuzzy methods. However, because a linguistic interpretation of each fuzzy associative classification rule is

Table 5
Simulation results by the MTDM with min FS = 0.05, min FC = 0.85

K'	Classification accuracy rate (%)
3	66.67
4	92.67
5	94.00
6	96.67
7	96.67
8	96.67

Table 6
Simulation results by the MTDM with min FS = 0.10 and min FC = 0.80

K'	Classification accuracy rate (%)
3	66.67
4	88.00
5	95.33
6	96.67
7	96.67
8	96.67

easily obtained, the goal of knowledge acquisition for users can be achieved by checking the fuzzy rules generated from the proposed learning algorithm.

5. Discussions

Fuzzy associative classification rules with linguistic interpretation discovered by data mining techniques are helpful to build a prototype fuzzy knowledge base of the fuzzy classifier system. For this, the generation of fuzzy classification rules with linguistic interpretation from the training data becomes quite necessary. The proposed learning algorithm can also be viewed as a knowledge acquisition tool for classification problems.

The performance of the proposed learning algorithm is tested on the iris data. From some simulation results, we can see that the classification rates are more sensitive to larger min FS, indicating that a smaller min FS can be a better choice. This result demonstrates that smaller min FS can lead to reserve a larger number of large fuzzy grids; that is, more valuable information can be reserved to generate fuzzy rules. Moreover, the MTDM is not sensitive to larger values of K' (i.e. $K' = 6-8$) and works more robustly than the STD. When examining the generalization ability for testing samples, we find that some poor results (e.g. 91.33 and 92.67% for $K' = 7$) resulted from larger values of K' (i.e. $K' = 7, 8$) are obtained by the STD. That is, the STD can suffer from overfitting the training samples, because larger values of K can result in fine partitions in a feature space. From the viewpoint of improving classification rates, it seems that it might be better for the proposed learning algorithm to use smaller min FS and the MTDM to derive fuzzy rules with high classification capability. The downside is that the number of rules increases as well to verify the generalization of this observation; it is reasonable for the proposed learning algorithm to be further tested on other classification problems. Recently, automation for the classification task of group technology become a significant research topic, e.g. the classification of block-shaped parts by [Chuang, Wang, and Wu \(1999\)](#). After features and data from workpieces are collected, then it is possible to employ the proposed learning algorithm to discover fuzzy rules and classify the data.

The performance of the proposed learning algorithm for phase I and phase II is mainly dependent on the size of candidate grids and the size of large grids, respectively. It seems that it will waste much more time in phase I for generating candidate fuzzy grids. For simplicity, we briefly discuss the time complexity of phase I for operating FG_{TK} . For FG_{TK} , we need $(dK')^2$, $n(dK')$ and (dK') operations for building FG_K , TT_K , and FS_K , respectively. In addition, the worst cases for operating FG_K ,

Table 7
Simulation results by the STDM with min FS = 0.05 and min FC = 0.85

K'	Classification accuracy rate (%)
3	66.67
4	94.67
5	94.00
6	96.67
7	96.00
8	95.33

TT_K and FS_K in generating candidate k -dim ($k \geq 2$) fuzzy grids are roughly measured as $(dK')C_2^{s_k-1}$, $nC_2^{s_k-1}$, and $C_2^{s_k-1}$, where s_i ($i \geq 1$) denotes the number of large i -dim fuzzy grids, respectively. Therefore, the worst case of phase I of the proposed learning algorithm could be roughly measured as $(dK')^2 + n(dK') + (dK') + \sum_{i=1}^{k-1} (n + dK' + 1)C_2^{s_i}$.

We stress the feasibility and the problem-solving capability of the proposed method for classification problems, rather than providing formal methods to find general parameter specifications that can obtain the best classification accuracy rate. That is, it seems relatively difficult to determine appropriate values of K' , minimum fuzzy support and fuzzy confidence. Previously, some tuning methods were proposed, for example, ANFIS by Jang (1993), and design of fuzzy controllers by Homaifar and McCormick (1995). On the other hand, Nozaki, Ishibuchi, and Tanaka (1996) demonstrated that the performance of fuzzy rule-based systems can be improved by adjusting the grade of certainty of each rule. Therefore, to develop methods for determining appropriate membership functions and the fuzzy confidence of each rule to obtain higher classification rates by machine learning techniques, such as genetic algorithms or neural networks, is quite appropriate and left for future research.

6. Conclusions

In this paper, we propose a learning algorithm to discover effective fuzzy associative classification rules. As we have explained earlier, the proposed learning algorithm consists of two phases: one to generate large fuzzy grids from training samples by fuzzy partitioning in each attribute, and the other to generate fuzzy associative classification rules by large fuzzy grids. The proposed learning algorithm is implemented by scanning training samples stored in a database only once and applying a sequence of

Table 8
Simulation results by the STDM with min FS = 0.10 and min FC = 0.80

K'	Classification accuracy rate (%)
3	66.67
4	91.33
5	94.00
6	96.67
7	93.33
8	95.33

Table 9
Classification accuracy rates (%) obtained by the leaving-one-out technique

Min FS	Min FC	Division method	K		
			6	7	8
0.05	0.85	MTDM	96.67	96.67	96.67
		STDM	95.33	93.33	91.33
0.10	0.80	MTDM	96.67	96.00	96.00
		STDM	96.67	92.67	95.33

table operations to generate fuzzy grids and fuzzy rules. Therefore, it can be easily extended to discover other types of fuzzy association rules for market basket analysis that can help managers design different store layouts and help retailers to plan which items to put on sale (Han & Kamber, 2001). Especially, because each fuzzy partition is a fuzzy number, a linguistic interpretation of each fuzzy partition is easily obtained.

The performance of the proposed learning algorithm is tested on the iris data. According to the simulation results, the MTDM is not sensitive to larger values of K' (i.e. $K' = 6-8$) and it works more robustly than the STDM. Based on the leaving-one-out technique, we make a comparison between the proposed learning algorithm and other fuzzy classification methods. It is clear that the best result obtained by the proposed learning algorithm outperforms that of Ishibuchi et al.'s genetic-algorithm-based method. The best result of the proposed algorithm is also equal to the best result reported by Grabisch and Dispot (1992) for nine fuzzy classification methods. However, because a linguistic interpretation of each fuzzy associative classification rule is easily obtained, the goal of knowledge acquisition for users can be achieved by checking the fuzzy rules. The simulation results from the iris data indicate that the proposed learning algorithm may effectively derive fuzzy associative classification rules. On the other hand, from discussions in Section 5, it seems that it might be better for classification problems to use the MTDM with smaller min FS to perform the proposed learning algorithm.

Acknowledgements

We are very grateful to the anonymous referees for their valuable comments and constructive suggestions. This research is supported by the National Science Council under grant NSC-90-2416-H-009-002.

References

- Agrawal, R., Imielinski, T., & Swami, A. (1993). Database mining: A performance perspective. *IEEE Transactions on Knowledge and Data Engineering*, 5(6), 914–925.
- Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, A. I. (1996). Fast discovery of association rules. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy (Eds.), *Advances in knowledge discovery and data mining* (pp. 307–328). AAAI Press, Mewlo Park, USA.
- Berry, M., & Linoff, G. (1997). *Data mining techniques: For marketing, sales, and customer support*. New York: Wiley.

- Chen, S. M., & Jong, W. T. (1997). Fuzzy query translation for relational database systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 27(4), 714–721.
- 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.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals in Eugenics*, 7(2), 179–188.
- Grabisch, M., & Dispot, F. (1992). A comparison of some methods of fuzzy classification on real data. *Proceedings of the Second International Conference on Fuzzy Logic and Neural Networks, Japan*, pp. 659–662.
- Han, J. W., & Kamber, M. (2001). *Data mining: Concepts and techniques*. San Francisco: Morgan Kaufmann.
- Han, E. H., Karypis, G., & Kumar, V. (2000). Scalable parallel data mining for association rules. *IEEE Transactions on Knowledge and Data Engineering*, 12(3), 337–352.
- Homaifar, A., & McCormick, E. (1995). Simultaneous design of membership functions and rule sets for fuzzy controllers using genetic algorithms. *IEEE Transactions on Fuzzy Systems*, 3(2), 129–139.
- Hong, T. P., & Chen, J. B. (1999). Finding relevant attributes and membership functions. *Fuzzy Sets and Systems*, 103(3), 389–404.
- 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.
- 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., Murata, T., & Gen, M. (1998). Performance evaluation of fuzzy rule-based classification systems obtained by multi-objective genetic algorithms. *Computers and Industrial Engineering*, 35(3–4), 575–578.
- 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.
- Jang, J. S. R. (1993). ANFIS: Adaptive-network-based fuzzy inference systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3), 665–685.
- Myra, S. (2000). Web usage mining for web site evaluation. *Communications of the ACM*, 43(8), 127–134.
- Nozaki, K., Ishibuchi, H., & Tanaka, H. (1996). Adaptive fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems*, 4(3), 238–250.
- Sun, C. T. (1994). Rule-base structure identification in an adaptive-network-based fuzzy inference system. *IEEE Transactions on Fuzzy Systems*, 2(1), 64–73.
- 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*. CA: Morgan Kaufmann.
- Yilmaz, E., Triantaphyllou, E., Chen, J., & Liao, T. W. (2002). A heuristic for mining association rules in polynomial time. *Mathematical and Computer Modelling*, in press.
- Zadeh, L. A. (1965). Fuzzy sets. *Information Control*, 8(3), 338–353.
- Zadeh, L. A. (1975a). The concept of a linguistic variable and its application to approximate reasoning. *Information Science (part 1)*, 8(3), 199–249.
- Zadeh, L. A. (1975b). The concept of a linguistic variable and its application to approximate reasoning. *Information Science (part 2)*, 8(4), 301–357.
- Zadeh, L. A. (1976c). The concept of a linguistic variable and its application to approximate reasoning. *Information Science (part 3)*, 9(1), 43–80.
- Zadeh, L. A. (1976). The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences (part 3)*, 9(1), 43–80.