European Journal of Operational Research 218 (2012) 202–210

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Computational Intelligence and Information Management

Mining axiomatic fuzzy set association rules for classification problems ☆

Xin Wang [a,b,*], Xiaodong Liu [a,b], Witold Pedrycz [c], Xiaolei Zhu [b], Guangfei Hu [b]

[a] Research Center of Information and Control, Dalian University of Technology, Dalian 116024, PR China
[b] Department of Mathematics, Dalian Maritime University, Dalian 116026, PR China
[c] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada T6G 2G7

ARTICLE INFO

Article history:
Received 4 April 2010
Accepted 22 April 2011
Available online 29 April 2011

Keywords:
Data mining
Fuzzy association rules
AFS fuzzy logic
Knowledge acquisition
Classification

ABSTRACT

In this paper, we propose a novel method to mine association rules for classification problems namely AFSRC (AFS association rules for classification) realized in the framework of the axiomatic fuzzy set (AFS) theory. This model provides a simple and efficient rule generation mechanism. It can also retain meaningful rules for imbalanced classes by fuzzifying the concept of the class support of a rule. In addition, AFSRC can handle different data types occurring simultaneously. Furthermore, the new model can produce membership functions automatically by processing available data. An extensive suite of experiments are reported which offer a comprehensive comparison of the performance of the method with the performance of some other methods available in the literature. The experimental result shows that AFSRC outperforms most of other methods when being quantified in terms of accuracy and interpretability. AFSRC forms a classifier with high accuracy and more interpretable rule base of smaller size while retaining a sound balance between these two characteristics.

1. Introduction

Classification is one of the most commonly applied supervised data mining paradigms. There exist various methods of classification depending on the information available and the nature of the classification task. Both classification accuracy and interpretability are essential in any classification problem. High accuracy usually leads to complex and hard to understand models, while the existing high interpretable models cannot satisfy the requirement of high accuracy. For example, the "rule-based" classifiers, say C4.5 (Quinlan, 1993), Ripper (Cohen, 1995) and OneR (Holte, 1993) may suffer from limited accuracy. While the other classifiers, such as KNN (Aha and Kibler, 1991) and SVM (Keerthi et al., 2001) are difficult to comprehend.

An important classification technique that has attracted an increasing attention in recent years comes in the form of classification rules based on association rule mining techniques (see e.g. Agrawal et al., 1993; Chen et al., 2006; Li et al., 2001). Association rule discovery originates from market basket analysis and aims at finding interesting relationships hidden in large data sets. The general form of association rule is expressed in terms of the implication $A \Rightarrow B$. Here A is the antecedent and B is the consequent which consists items. The new technique, called "associative classification" (Pach et al., 2008), aims to combine the advantages of both traditional classification and association discovery. Associative classifiers have certain advantages which make them suitable for application to classification problems. The major advantages of associative classifiers include the interpretability, fast training and the power of handling training sets of high dimensionality. Unfortunately, studies also showed that classical associative classifiers have some weaknesses such as a huge number of discovered rules.

Fuzzy predicates have been incorporated into the realm of machine learning and data mining to augment the types of data relationships that can be represented in the form of rules, to facilitate the interpretation of rules, and to avoid Boolean boundaries when partitioning attribute domains (see e.g. Amo et al., 2004; Ravi et al., 2000; Serrurier et al., 2007). The incorporation of fuzzy sets into classification tasks enables us to combine uncertainty handling and approximate reasoning capabilities of the former with the comprehensibility and ease of application of the latter. This combination augments the representation capabilities of rules with the knowledge component inherent to fuzzy logic subsequently leading to their robustness, noise immunity, and substantial applicability level in particular when dealing with situations we encounter a factor of uncertainty.

Different methods were proposed for mining fuzzy association rules from quantitative data (see e.g. Chen and Weng, 2009; Hong and Lee, 2008) where the membership functions of the linguistic terms were specified in advance. The AFS framework proposed by Liu (1998) supports the studies on how to convert the information hidden in databases into the membership functions and their

☆ This work is supported by the Natural science Foundation of China under Grant 60875032 and the Fundamental Research Funds for the Central Universities 2011QN059 and 2011QN057.

* Corresponding author at: Research Center of Information and Control, Dalian University of Technology, Dalian 116024, PR China. Tel.: +86 411 84709380.

E-mail address: xenawang7811@hotmail.com (X. Wang).

0377-2217/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2011.04.022

fuzzy logic operations. In this paper, we use fuzzy sets (membership functions) and underlying logic operations generated by AFS to eliminate potential subjective bias present in the conventional fuzzy rules resulting from the use of different membership functions of the fuzzy terms. In order to realize mining meaningful rules from data with imbalanced classes we define a new concept of a fuzzy class support. In order to avoid the difficulty of determining the minimal confidence level we propose the optimal fuzzy confidence truncation leading to the reduction of the available rule base. The tuning of the optimal fuzzy confidence truncation and the accuracy-driven pruning methods help generate a compact and efficient fuzzy rule-based classifier.

To offer a thorough comparative test, we experimented with the algorithm using twelve well-known real data sets coming from the UCI Repository (Merz and Murphy, 1996), and compared the proposed algorithm with twelve classical classifiers implemented in Weka Witten and Frank (2005) with respect to their accuracy and interpretability (rule size). Among them, eight are rule-based classifiers and the others are "black box" classifiers. We also compared obtained results with those reported for the previously published association rule-based classifiers. Finally, the consistency of the AFS membership functions is illustrated. The experiments show that the proposed model performs well on both interpretability and accuracy compared with other classifier systems. AFSRC offers an understandable and accurate classifier and it can retain sound balance between these efficiency criteria.

The paper is structured as follows. Section 2 recalls the basic notions and properties of the AFS theory that are essential in the framework of our investigations. In Section 3, we proposed the algorithm of AFSRC. Section 4 is concerned with experiments and some comparative analysis. The concluding remarks are presented in Section 5.

2. Selected preliminaries of the AFS theory

In this section, we recall some notations and present the most pertinent results of AFS theory.

2.1. AFS algebra

In Liu (1998), the AFS algebra was defined along with their applications to study the development of membership functions for fuzzy concepts. The following example serves as a brief illustration of the AFS algebra.

Example 1. Let $X = \{x_1, x_2, \ldots, x_{10}\}$ be a set of 10 customers with some features which are described by real numbers, categorical/Boolean values, ranks and the order relations. It is shown in Table 1. The ordinal number *i*th in the"white" column which corresponds to some $x \in X$ implies that the hair color of x has ordered *i*th following our perception of the color. For example, the numbers in the column *"white"* imply the order (>) : $x_4 > x_5 = x_8 > x_3 > x_1 = x_9 > x_7 > x_2 > x_6 > x_{10}$. Here, $x_i > x_j$ states that the hair of x_i is closer to the white color than the color of hair the individual x_j. The relationship $x_i = x_j$ means that the hair of x_i looks as white as that of x_j. The values in the "*work*"column has the relationship according to the work skill level: high > med > prim > none.

Table 1
A set of customers.

Customer	*Age*	*Estate*	*Male*	*Work*	*white*	*credit*
x_1	25	0	1	prim	4th	0
x_2	19	0	0	prim	6th	0
x_3	50	34	0	high	3rd	1
x_4	80	80	1	none	1st	1
x_5	34	2	1	med	2nd	0
x_6	37	28	0	high	7th	1
x_7	45	90	1	med	5th	1
x_8	70	45	1	none	2nd	1
x_9	60	98	0	med	4th	1
x_{10}	3	0	0	none	8th	0

A concept on X may associate to one or more features. For instance, the fuzzy concept "*rich*" associates a single feature "*estate*" and the fuzzy concept "*old white hair males*" associates three features "*age*", "*white*" and "*male*". The concepts associated with a single features are viewed as fuzzy (or numeric) linguistic terms of the corresponding feature. For instance, the fuzzy concepts (fuzzy linguistic terms) "*old*" and "*about 40 years old*" associate to the feature "*age*".

Let $M = \{m_1, m_2, \ldots, m_8\}$ be the set of fuzzy (or numeric) linguistic terms on X and each $m \in M$ associates to a single feature. Where m_1: "*old person*", m_2: "*rich person*", m_3: "*high work skill person*", m_4: "*young person*", m_5: "*the person about 45 years old*", m_6: "*male*", m_7: "*female*"(i.e., *not male*), m_8: "*person with white hair*". The elements of M are viewed as " elementary terms" (or "**simple concept**") of the corresponding features. For each set of fuzzy terms $A \subseteq M, \prod_{m \in A} m$ represents the conjunction of the fuzzy terms in A. For instance, $A = \{m_1, m_7\} \subseteq M, \prod_{m \in A} m = m_1 m_7$ represents a new fuzzy concept "*old woman*" which is associating to the features "*age*" and "*male*". $\sum_{i \in I}\left(\prod_{m \in A_i} m\right), A_i \subseteq M, i \in I$, which is a formal sum of the fuzzy terms $\prod_{m \in A_i} m$, is the disjunction of the conjunctions of the fuzzy terms in A_i represented by $\prod_{m \in A_i} m$'s (i.e., the disjunctive normal form of a formula representing a concept). For example, we may have $\gamma = m_1 m_6 + m_1 m_3 + m_2$ which translates as "*old males*" or "*high work skill old persons*" or "*rich persons*" (the "+" here denotes a disjunction of terms). While M may be a set of fuzzy or two-valued terms, every $\sum_{i \in I}\left(\prod_{m \in A_i} m\right), A_i \subseteq M, i \in I$, has a well-defined meaning such as the one we have discussed above. By a straightforward comparison of the expressions

$$\xi = m_3 m_8 + m_1 m_3 + m_1 m_6 m_8 + m_1 m_3 m_8 \quad \text{and} \quad \zeta = m_3 m_8 + m_1 m_3 + m_1 m_6 m_8.$$

we conclude that the above concepts are equivalent. Considering ξ, for any x, the degree of x belonging to the fuzzy concept represented by $m_1 m_3 m_8$ is always less than or equal to the degree of x belonging to the fuzzy concept represented by $m_1 m_3$. Therefore, the item $m_1 m_3 m_8$ is redundant when forming the fuzzy concept ξ. Let us take into consideration two expressions of the form $\alpha : m_1 m_3 + m_2 m_5 m_6$ and $v : m_5 m_6 + m_6 m_8$. The semantic content of the fuzzy concepts "α or v" and "α and v" can be expressed as follows "α or v" : $m_1 m_3 + m_2 m_5 m_6 + m_5 m_6 + m_6 m_8$ equivalent to $m_1 m_3 + m_5 m_6 + m_6 m_8$. "$\alpha$ and v" : $m_1 m_3 m_5 m_6 + m_2 m_5 m_6 + m_1 m_3 m_6 m_8 + m_2 m_5 m_6 m_8$ equivalent to $m_1 m_3 m_5 m_6 + m_2 m_5 m_6 + m_1 m_3 m_6 m_8$.

The semantics of the logic expressions such as "*equivalent to*", "*or*" and "*and*" as expressed by $\sum_{i \in I}(\prod_{m \in A_i} m), A_i \subseteq M, i \in I$, can be formulated in items of the AFS algebra in the following manner.

Let M be a non-empty set. The set EM^* is defined by $EM^* = \left\{\sum_{i \in I}\left(\prod_{m \in A_i} m\right) | A_i \subseteq M, i \in I, I \text{ is any non-empty indexing set}\right\}$.

Definition 1 Liu (1998)). Let M be a non-empty set. A binary relation R on EM^* is defined as follows. For any $\sum_{i \in I}\left(\prod_{m \in A_i} m\right)$, $\sum_{j \in J}\left(\prod_{m \in B_j} m\right) \in EM^*$, $\left[\sum_{i \in I}\left(\prod_{m \in A_i} m\right)\right] R \left[\sum_{j \in J}\left(\prod_{m \in B_j} m\right)\right] \iff$ $(i) \forall A_i \ (i \in I), \exists B_h (h \in J)$ such that $A_i \supseteq B_h$; $(ii) \forall B_j (j \in J), \exists A_k (k \in I)$, such that $B_j \supseteq A_k$.

It is clear that R is an equivalence relation. The quotient set EM^*/R is denoted by EM. The notation $\sum_{i \in I}\left(\prod_{m \in A_i} m\right) = \sum_{j \in J}\left(\prod_{m \in B_j} m\right)$ means that $\sum_{i \in I}\left(\prod_{m \in A_i} m\right)$ and $\sum_{j \in J}\left(\prod_{m \in B_j} m\right)$ are equivalent under equivalence relation R. Thus the semantics they represent are