International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Building Fuzzy Associative Classifier Using Fuzzy Values

P. Kayal[1], S. Kannan [2]

[1]Department of CSE, Bharathiar University, Coimbatore, India

[2]Department of Computer Science, Madurai Kamaraj University, Madurai, India

Abstract: ***Association rule mining is a popular and well researched method for discovering interesting relations between variables in large databases. Classification based on association rules is considered to be an effective and beneficial approach. However, mining the domain of quantitative attributes leads to a common existing "sharp boundary problem". This can be over-ruled by the use of fuzziness in the variables and the association rule mining. This paper aims at proposing a classification based on fuzzy association rule mining called "Fuzzy Associative Classifier" (FAC) to predict the level of work-force leadership qualities.***

Keywords: Fuzzification, Partitioning, Associative classifier, Fuzzy weight, Skill level.

1. Introduction

A standard association rule is a rule of the form X→Y which says that if X is true of an instance in a database, so is Y true of the same instance, with a certain level of significance as measured by two indicators, support and confidence[3]. The goal of classification is to accurately predict the target class for each case in the data. In our paper, a classification model is used to identify the level of leadership as Not satisfactory, Satisfactory, Good, Excellent which is in turn going to be a fuzzy classification. The best way of classification is classification based on association rule mining which builds an associative classifier. An associative classifier is composed of only those association rules of the form X→C where C is a class label treated as a special case of Y in association Rules[13].

Human Resource application that are embedded with artificial intelligent (AI) techniques can be used to help decision makers to solve unstructured decisions [1].In the advancement of technology many techniques are used to advance the HR application capabilities [4]. Data mining involves extracting knowledge based on patterns of data in very large databases. Yet, data mining goes beyond simply performing data analysis on large data sets [2].With the help of association rules and classification, the model FAC helps to identify the right individual as leader in an organization.

2. Association Rules

2.1 Basics

Association rules are if/then statements that help uncover relationships between seemingly unrelated data in a relational database or other information repository. An example of an association rule would be "If a customer buys a dozen eggs, he is 80% likely to also purchase milk."An association rule has two parts, an antecedent (if) and a consequent (then). An antecedent is an item found in the data. A consequent is an item that is found in combination with the antecedent. Association rules are created by analysing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true.

2.2 Definition

Let $\mathbf{I} = \{i_1, i_2, ..., i_n\}$ be a set of n binary attributes called items. Let $\mathbf{D} = \{t_1, t_2, ..., t_n\}$ be a set of transactions called the database. Each transaction in **D** has a unique transaction ID and contains a subset of the items in **I**. A rule is defined as an implication of the form $\mathbf{X} \rightarrow \mathbf{Y}$ where $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{I}$ and $X \cap Y = \emptyset$. The sets of items **X** and **Y** are called antecedent (left-hand-side or LHS) and consequent (right-hand-side or RHS) of the rule respectively. To select interesting rules from the set of all possible rules, constraints on various measures of significance and interest can be used. The best-known constraints are minimum thresholds on support and confidence [3].The support and confidence of a rule is defined as

$$\text{Supp}(X) = \text{no. of transactions which contain the item set X / total no. of transactions} \quad (1)$$

$$\text{Conf}(X \rightarrow Y) = \text{Supp}(X \cup Y) / \text{Supp}(X) \quad (2)$$

3. Fuzzy Association Rule Mining

3.1 Fuzzy Set

The use of fuzzy approach in our research is well discussed in our previous paper [14].A fuzzy set is a class of objects with a continuum of grades of membership. Such a set is characterized by a membership function which assigns to each object a grade of membership ranging between zero and one. In many cases linguistics terms are of fuzzy nature in concept and usually formulated by fuzzy sets. Let U be the universe of discourse. Then a fuzzy set FS on U is $FS = \{(x, \mu(x)) / x \varepsilon U\}$,where μ is the membership function reflecting a mapping from U to [0,1].

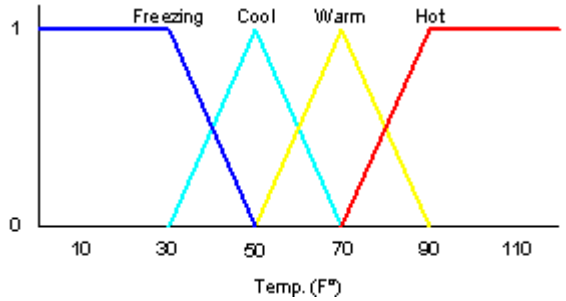

Figure 1: A sample Membership mapping

3.2 Fuzzy Domain Partitioning

Domain partitioning is a technique used with quantitative association rule mining. This type of partitioning describes domain with sharp boundary problem, i.e., the domain elements nearing the boundaries may be ignored or overemphasized .For example, an abstract concept such as "short" to describe a person's height would be distinguished by a precise value such as (150,170cm), often leading to a difficulty to intuitively identify nearby values such as 171 due to such a nature of artificial precisionization. Furthermore, the degrees of the elements belonging to the interval are not distinguishable, such as 150 and 170 being treated equally, which are not considered intuitively appealing. To cope with the problem, a flexible and transitional setting of the interval boundaries is regarded desirable [5]. Hence fuzzy logic is employed.

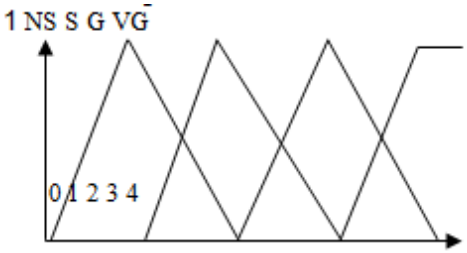

Figure.2: Triangular Membership Function

Table 1: Skill level

NS	Not Satisfactory
S	Satisfactory
G	Good
VG	Very good

3.3 Use of Fuzzy Weight

The attribute weight plays a major role in determining the attribute's importance. Fuzzy weight has value between 0 and 1.Though our initial values are discrete, this fuzzy weight plays its role to convert discrete values into continuous one. Moreover the attributes importance is much necessary to determine the leadership class as well as in the mining rules generated. The weight assigned for each attribute is purely by the domain experts. Here the weight chart for only four attributes is presented. The product of attribute value and the corresponding attribute's weight gives the actual attribute value for our fuzzification process.

Table 2: Attribute weight in fuzzy

Attribute	*Weight(w)*
CS	0.8
MG	0.4
TS	0.4
TD	0.1

4. The Fuzzy Associative Classifier (FAC)

4.1 Building FAC

The process of building the Fuzzy associative classifier (FAC) involves various steps. As first let us define our area of work, attributes and their values. Let us consider four skills (attributes) for the computation. Communication Skill (CS), Magnamity (MG), Team Support (TS), Team dominance (TD).The following table shows the level of employee's skill. The skill level varies from 1 to 4 which is a discrete value meaning 1-NOT SATISFACTORY 2-SATISFACTORY 3-GOOD 4-VERY GOOD. Table 3 shows the actual database D and Table 4 shows database D' after applying weight to the original database D.

Table 3: Database (D)

Emp Id	***CS***	***MG***	***TS***	***TD***
E01	3	2	2	2
E02	1	3	3	1
E03	4	3	2	3
E04	2	2	1	4
…	…	…	…	…

Table 4: Database D '

Emp Id	***CS***	***MG***	***TS***	***TD***
E01	2.4	0.8	0.8	0.2
E02	0.8	1.2	1.2	0.1
E03	3.2	1.2	1.2	0.3
E04	1.6	0.8	0.8	0.4
…	…	…	…	…

4.2 Fuzzification

Fuzzification is the process of changing a real scalar value into a fuzzy value. This is achieved with different types of fuzzifiers called the membership functions. There are several parameterized ways to define membership function.Some popular membership functions are trapezoidal,bell,Gaussian and triangular.A parameterized membership function can be defined in terms of a number of parameter.A triangular membership fuction is defined by three parametes(x,y,z)which are known values .For a given value u , with known x,y,z the membership for u in all defined partitions can be calculated. For example the attribute CS is further defined into four different fuzzy sets on its domain as NS(CS),S(CS),G(CS),VG(CS) with triangular membership function as shown in Figure 2.The mapping of values of quantitative attribute to membership of several new attributes in forms of linguistic terms will help to mine better fuzzy association rules.After fuzzication process the database is now newly constructed with multiple attributes for the base arrtibutes and the database is shown as in Table 5.

Table 5: After Fuzzification

Emp Id	CS				MG				TS				TD			
	NS	S	G	VG	NS	S	G	VG	NS	S	G	VG	NS	S	G	VG
E01	0.2	0.6	0.4	0.1	0.5	0.4	0.2	0	0.5	0.4	0.2	0	0.8	0.2	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

4.3 Normalization

Necessity of normalization arises in this situation. Normalization is fuzzy membership values are used to scale the normal form between 0 and 1. But according to our data in table 4, the sum of values of CS in all domains exceeds 1 which is not desirable. Therefore a normalization action is necessary to scale and it is done as follows:

$$\mu_k^j(d) = \{\mu_k^j(d) / \sum_{j=1}^{qk} \mu_k^j(d) \quad (3)$$

where d is the transaction,$[I_k]$ is the value of k^{th} attribute and $\mu_k^j(d)$ is the degree of $d[I_k]$ in the fuzzy set of I_k[5].

After the above normalization application the data in table 5 is modified as shown below.

Table 6: Part of DB before Normalization

CS			
NS	S	G	VG
0.2	0.6	0.4	0.1

Table 7:Part of DB before Normalization `

CS			
NS	S	G	VG
0.15	0.46	0.30	0.07

4.4 Generation of Associative Rules

Once the normalization process freezes, the next is to identify the frequent item set from the huge amount of data which is required to mine the association rules. Mining interesting rule is the challenge which is carried out by different pruning techniques. These rules will uncover the relationship between the 12 attributes we have already defined. With the help of this relationship, it is easy to identify the degree of leadership qualities in the defined different level of skills.

4.5 Associative Classification of Interesting Rules

Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. A classification model could be used to identify the level of leadership as Not satisfactory, Satisfactory, Good, Very good which is in turn going to be a fuzzy classification. The best way of classification is classification based on association rule mining which build a associative classifier. An associative classifier is composed of only those association rules of the form X→C where C is a class label treated as a special case of Y in association Rules.

5. Conclusion

This paper discussed the idea of empowering classical association rules by combining them with fuzzy set theory. Though this has already been around since several years, the application of fuzzy weight with the new technique FAC is a new approach which brings a better and more accurate result. By using such technique the result obtained will be useful to make decision on leadership issues which is the prime factor necessary for today's corporate [14]. Moreover the application helps HR of an organization to identify the right individual as leader which indirectly reduces organization's risks.

References

[1] Hamidah Jantan, Abdul Razak Hamdan, Zulaiha Ali Othman, "Towards applying data mining techniques for talent management"2009 International Conference on Computer Engineering and Applications IPCSIT vol.2 (2011) .IACSIT Press, Singapore.
[2] Lori K. Long, Marvin D. Troutt Data Mining for Human Resource Information Systems, Kent State University, USA.
[3] J. Han, M. Kamber, "Data Mining", Morgan Kaufmann Publishers, San Francisco, CA, 2001.
[4] Society for human resource management.
[5] ZuoliangChen,Guoqing Chen :"Building an associative classifier based on fuzzy association rules" International Journal of computational Intelligence Systems,Vol.1.No.3(August,2008),262-273.
[6] P.Kayal,S.Kannan, " International Journal of Scientific and Research Publications ",Vol.3,Issue 10, 2013.
[7] http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/sbaa/report.fuzzysets.html
[8] Margaret Wheatley , Debbie Frieze ©2010 published in Resurgence Magazine, Winter 2011]
[9] Timothy J.Ross "Fuzzy Logic with Engineering Application".
[10] Bakk. Lukas Helm "Fuzzy Association Rules, Vienna University of Economics and Business Administration.
[11] Mannila, Heiki; Toivonen, Hannu; Verkamo, A. Inkeri: Discovery of Frequent Episodes in Event Sequences. Data Mining and Knowledge Discovery, Volume 1, Issue 3, 1997.
[12] Kuok, Chan Man; Fu, Ada; Wong, Man Hon: Mining Fuzzy Association Rules in Databases. SIGMOD Record Volume 27,1998.

Author Profile

P. Kayal, M.Sc., M.Tech, Research Scholar, Bharathiar University, Coimbatore, TamilNadu, India.

Dr. S. Kannan, M.Sc., Ph.D., Associate Professor, Department of Computer Applications, Madurai Kamaraj University, Madurai, TamilNadu, India.