ISSN 2250-3153

Efficient Association Rule Mining using Fuzzy Weight on Fuzzy Values

P.Kayal*, Dr.S.Kannan**

* Research Scholar, Department of Computer Science and Engineering, BharathiarUniversity, Coimbatore, TamilNadu, India.
** Associate Professor, Department of Computer Applications, Madurai Kamaraj University, Madurai, TamilNadu, India.

Abstract- One of the most important challenges of HR professional is to manage organization's talents, especially to ensure the right person for the right job. This paper presents an idea of identifying leadership quality in a person, which is a potential contribution that Data Mining (DM) could make within the Human Resource (HR) function in organizations,using association rule mining which is one of the most important techniques in the field of Data Mining. The original idea derives from attempts to deal with quantitative attributes in a database, where subdivision of the quantitative values into crisp sets would lead to over or underestimating values. The concept of Fuzzy will help to overcome the problem by allowing partial memberships to variable values.

Index Terms- Data mining, Association rule, Fuzzy value, Human resource

1. Introduction

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. Data mining may be regarded as an evolving approach to data analysis in very large databases that could become a useful tool to HR professionals. 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].

Today's competitive marketplace requires human resource professionals to have an expanded role in the organization due to increasing importance of social and relationship capital. [7]There are many researches on solving HRM problems that uses Data Mining approach.

High risk requires high control. As situations grow more complex and challenging, power needs to shift to the top with the leaders who know what to do [8]. Such an important role should be played by a leader, for which organizations strives to find the right person for the right place at right time. The following figure shows the roadmap of a leader in which we are going to concentrate on the second part. [7]

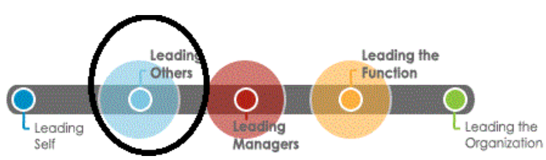

Figure.1. Leader Roadmap

2. Association Rules

2.1 Basics

In data mining, association rule learning is a popular and renowned method for discovering interesting relations between variables in large databases. 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[6].

2.2 Definition

Let $\mathbf{I} = \{\mathbf{i_1, i_2, ..., i_n}\}$ be a set of *n* binary attributes called *items*. Let $\mathbf{D} = \{\mathbf{t_1, t_2, ..., t_n}\}$ 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 ruleis defined as an implication of the form

X→Y where **X, Y** ⊆ **I** and X ∩Y =Ø**.** 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 [6].

1.**Support**-The support of a rule is defined as

Supp(X) **= no. of transactions which contain the itemset X / total no. of transactions** **(1)**

2.**Confidence**-The ***confidence*** of a rule is defined as

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

3. Fuzzy Association Rules

Based on classical association rule mining, a new approach has been developed expanding it by using fuzzy sets. The new fuzzy association rule mining approach emerged out of the necessity to mine quantitative data frequently present in databases efficiently. When dividing an attribute in the data into sets covering certain ranges of values, we are confronted with the sharp boundary problem.

Elements near the boundaries of a crisp set will either be ignored or overemphasized. For example, one can consider a set representing persons of middle age, ranging from 30 to 50 years old (Fig.2). In this example, a person aged 29 years would be a 0% representative and a 31 year old would be 100%. In reality, the difference between those ages is not that great. Implementing fuzziness can overcome this problem. The same problem can occur if one is dealing with categorical data. Sometimes, it is not ultimately possible to assign an item to a category. As an example, one can say that a tomato is a vegetable but also, in a way, a fruit. Crisp sets would only allow assigning the item to one single category, fuzzy sets allow different grades of membership to more than one set [10].

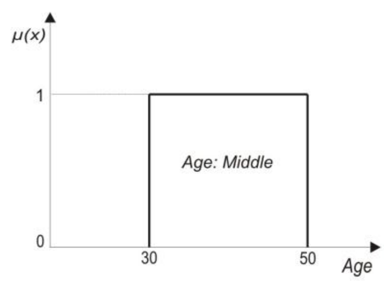

Figure. 2.Crisp Set

3.1 Quantitative Approach

Kuok et al. describe fuzzy association rules as follows [12], “Mining fuzzy association rule is the discovery of association rules using fuzzy set concepts such that the quantitative attribute can be handled” As in classical association rules, $I=\{i_1, i_2, \ldots, i_m\}$ represents all the attributes appearing in the transaction database $T=\{t_1, t_2, \ldots, t_n\}$. I contain all the possible items of a database, different combinations of those items are called item sets. Each attribute $i_k$will associate with several fuzzy sets. In order to represent the fuzzy sets associated with i_k, we use the notion $F_{i_k} = \{f_{i_k}^1, f_{i_k}^2, \ldots, f_{i_k}^l\}$ where $f_{i_k}^j$ is the j^{th}fuzzy set in F_{ik}. Fuzzy sets and their corresponding membership functions have to be defined by domain experts. Each of the fuzzy sets can be viewed as a [0,1] valued attribute, called fuzzy attribute. A fuzzy association rule has the following form: If X is A then Y is B.In this case, $X = \{x_1, x_2, \ldots, x_p\}$ and $Y = \{y_1, y_2, \ldots, y_q\}$ are item sets which are subsets of I . It is important to notice that those two sets must be disjoint and thus do not have any attributes in common. $A = \{f_{x_1}, f_{x_2}, \ldots, f_{x_p}\}$ and $B = \{f_{y_1}, f_{y_2}, \ldots f_{y_q}\}$ contain the fuzzy sets that are associated with X and Y .Known from classical association rules, *X is A* is the antecedent, *Y is B* is the consequent. If a sufficient amount of records approves this rule, we will call it satisfied.

3.2 Item set

The task of mining frequentitem sets deals with discovering frequent episodes in a sequence of events [11]. The data can be viewed as a sequence of events associated with a certain time of occurrence. Speaking of a frequent episode, we mean a collection of events that frequently occurs jointly. A basic problem of the analysis of such episodes is to discover them in the first place. Episodes are partially ordered sets of events. Looking at these episodes enables to discover regularities, for example an event X is followed by an event Y in most of the cases. The crucial problem here is the definition of how close together two items have to be in a timely manner in order to qualify as an episode.

ISSN 2250-3153

4. THE PROPOSED WORK- APPLYING FUZZY WEIGHT ON FUZZY ATTRIBUTE VALUES

Leadership is a topic on which huge amount of literature has been produced over the past decades. Many different management experts have developed theories and concepts on what leadership essentially is all about, what makes a person a good leader, how leaders could become better leaders, etc.[3] In this regard mining association rule with various attributes can determine a person how much he is capable of being a leader. After a vast search 12 attributes are selected as top qualities for leadership. Usually the presence and absence of a quality is represented as a binary value. Using binary value for these quality will not be a right way to determine the result since human being will have the quality in fuzzy[9] nature. The percentage and level of quality may vary but it cannot be determined by the values ‘yes’ or ‘no’. In this situation relying on fuzzy values will be a better approach to determine the attribute values for the leadership qualities derived.

e fuzzy values are determined the next step is to assign weight for those values according to the priority and importance of the quality. For instance Integrity and creativity are most important characteristics for a person to be a leader than the others. So the importance of a quality is represented as weight to determine the result. In this research a vast survey has been made to determine the attributes and its corresponding weight. Having all the data in hand, the next step is to determine how much a person is eligible to be a leader (LQ) is given as follows.

$$LQ = \sum I_i \times W_i \div \text{Total No of I} \qquad (3)$$

is the attribute value,W is the weight assigned for the corresponding I, and i varies from 1 to N(12) which is the number of attributes derived for leadership quality.Table 1 represents the attributes, attribute values in fuzzy, the corresponding fuzzy weights for the attributes which has been calculated by taking the maximum of n number of values and finally the product of attribute value and weight respectively. The frequent itemset identified using apriori algorithm[6] on the following table data when applied with weight(I_iW_i) are {0.3},{0.6},{0.2},{0.7}.In absence of weight the frequent item set found are {0.5},{0.7},{0.9}.

Table 1 .Attribute, value and weight

Attribute	Value(I)	Fuzzy Weight(W)	I_iW_i
Attr1	1.0	1.0	1.0
Attr2	0.5	0.6	0.30
Attr3	0.7	0.9	0.63
Attr4	0.7	0.9	0.63
Attr5	0.5	0.6	0.30
Attr6	0.5	0.5	0.25
Attr7	0.8	0.9	0.73
Attr8	0.6	0.4	0.24
Attr9	0.7	1.0	0.70
Attr10	0.9	0.9	0.81
Attr11	0.9	0.8	0.72
Attr12	0.7	0.7	0.49

Method 1:Applying Fuzzy weight on attribute values.

For the following table data, when the algorithm (equation 3) is applied i.e., with fuzzy weight the result obtained is 0.5, which states that a particular person is capable to be a leader only by a membership value of 0.5 ,which can be further interpreted as his leader-ship may be fruitful of 0.5 and he may fail on his duties by 0.5.Here the risk factor in selecting such a person as leader is 50%.An organization must be very careful in taking such a person whose efficiency is such uncertain.

Method2: No Fuzzy weight applied

s fuzzy weight is not applied the result obtained is 0.7, which states that the same person is capable to be a better leader by 0.7 and he may fail only by 0.3, which implies that the risk factor here is only 30%. When decision making comes in, organization obviously go for this person since the risk is lesser here. But actually he is of high risk to the company according to our algorithm discussed above.

Results:

Method1:When weight applied the calculation goes as follows: the sum of product of attributes and corresponding

weight divided by N=6.6/12=0.5.
Method 2:When weight not applied the calculation goes as follows: the sum of attributes divided byN=8.5/12= 0.7.

In Method 2 the same data is providing a result ignoring the efficiency and accuracy of the person's leadership quality by 0.2 membership value high,which is not at all a negligible factor for an organization's decision making to choose the right person for right job. When the organization takes a decision with Method 2, then they are at risk in choosing the lead. If they go with Method 1 theycould identify that the person is uncertain and cannot be a leader for the team. This will help them to find a better leader who is really fit and suits the job as well.

5. CONCLUSION

Association rule mining has a wide range of applicability such market basket analysis, medical diagnosis/ research, Website navigation analysis, homeland security and so on. Advantage of apriori [5,6] is its easy implementation. In this paper we presented a new approach to mine better association rules by using fuzzy values and appropriate fuzzy weights [9] for the attributes. 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.The idea of empowering classical association rules by combining them with fuzzy set theory has already been around since several years. But the application of fuzzy weight on fuzzy values brings a better and more accurate result for the above application of finding a correct person as leader to lead the team which directly reduces organization's risk.

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] The DevelopmentCompendium 2009 IIX - 19(leadership quality).
[4] R. Agrawal, T. Imielinski, A. Swami. Mining Association Rules between Sets of Items in Large Databases". In Proceedings of the 1993 International Conference on Management of Data (SIGMOD 93), pages 207-216, May 1993
[5] Hilderman R. J., Hamilton H. J., "Knowledge Discovery and Interest Measures", Kluwer Academic, Boston, 2002.
[6] J. Han, M. Kamber, "Data Mining", Morgan Kaufmann Publishers, San Francisco, CA, 2001.
[7] Society for human resource management.
[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.

AUTHORS

First Author – P.Kayal, M.Sc., M.Tech, ResearchScholar, BharathiarUniversity, Coimbatore, TamilNadu, India.
Email: kayalpaddu@gmail.com
Second Author – Dr.S.Kannan, M.Sc., Ph.D.,Associate Professor, Department of Computer Applications, Madurai Kamaraj University, Madurai, TamilNadu, India.

Correspondence Author – P.Kayal, E-mail : kayalpaddu@gmail.com.Mobile No: +91 8341391215.