www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 03 Issue 07 July, 2014 Page No. 6963-6966

A Review on Dealing Uncertainty, imprecision and Vagueness in Association Rule Mining Using Extended and Generalized Fuzzy

[1]Vivek Badhe, [2]Dr. R.S. Thakur, [3]Dr. G.S. Thakur
[1,2,3]MANIT Bhopal
vivekbadhe@yahoo.com
ramthakur2000@yahoo.com
ghanshyamthakur@gmail.com

ABSTRACT

The common approaches to association rule mining focus on generating rule by using correlation among data and finding frequent occurring patterns. The main technique uses support and confidence measures for generating rules. There are few other approaches that have already been induced to improve association rule mining. The purpose of this paper is to review the existing approaches based of fuzzy and its variants and extensions namely Rough Set, Vague Set and Soft Set and their combinations that have been used to increase the effectiveness of association rule mining techniques dealt with the uncertainty, approximation, vagueness, and imprecision theories.

Keywords: Association Rule Mining, Fuzzy Set, Rough Sets, Vague Sets, Soft Sets.

1.INTRODUCTION

Nowadays, large quantity of data is being accumulated in the data repository. Seeking knowledge from massive data is one of the most desired attributes of Data Mining. However, in reality, a substantial portion of the available information is stored in text databases (or document databases), which consists of large collections of documents from various sources, such as news articles, books, digital libraries and Web pages. Since web search engines have become pervasive and search has become integrated, retrieving of information from these search engines consist of three essentials: query, documents, and search results.

The emerging growth of data mining raises the large range of complex applications. It leads the broad study of data mining frequent patterns. Mining frequent sets over data streams present attractive new challenges over traditional mining in static databases. Data mining is generally used for retrieving the desire information to make it into knowledge from the large size databases.

2.DEFINITION

2.1 Association Rule Mining

Association rules discovery is one of the most important technologies which was given by Mr. Agrawal in 1993 [1]. It gives the information like "if-then" statements. These rules are invoked from the dataset. It generates from calculation of the support and confidence of each rule that can show the frequency of occurrence of a given rule. Association Analysis is the process of discovering hidden pattern or condition that occur frequently together in a given dataset. Association Rule mining techniques looks for interesting associations and correlations among data set. An association rule is a rule, which entails probabilistic relationship, with the form $X \Rightarrow Y$ between sets of database attributes, where X and Y are sets of items, and $X \cap Y = \phi$. Given the set of transactions T, we are interested in generating all rules that satisfy certain constraints. These constrains are *support* and *confidence*. The *support* of the rule is the fraction of the transactions in T that satisfy the union of items in X and Y. The probability, measured as the fraction of the transactions containing X also containing Y, is called the *confidence* of the rule.

Support should not be confused with confidence. While confidence is a measure of the rule's strength, support corresponds to statistical significance.

With the help of these constraints, rules are computed from the data and, association rules are calculated with help of probability. Mining frequent itemsets is a fundamental and essential problem in many data mining applications such as the discovery of association rules, strong rules, correlations, multi-dimensional patterns, and many other important discovery tasks.

The first and foremost algorithm that was given to generate association rules was *A priori* [2]. Its proposal used the same two constraints: support and confidence, and forming rules in accordance with these constraints.

2.2 Fuzzy Set Theory (mid 60's)

The most appropriate theory, for dealing with uncertainties is the theory of fuzzy sets developed by Zadeh [3]. The notion of fuzzy sets provides a convenient tool for representing vague concepts by allowing partial memberships.

A *fuzzy set (class) A* in *X* is characterized by a *membership (characteristic) function* $f_A(x)$ which associates with each point in *X* a real number in the interval [0, 1], with the value of $f_A(x)$ at *x* representing the "grade of membership" of *x* in *A*.

$$f_A(x): U \rightarrow [0,1]$$

$$X = \{(f_A(x)/x): x \in U, f_A(x) \in [0,1]\}$$

Thus, the nearer the value of $f_A(x)$ to unity, the higher the grade of membership of *x* in *A*. A fuzzy set can be interpreted by a family of crisp sets, and fuzzy set operators can be defined using standard set operators. The membership values may be interpreted in terms of truth values of certain propositions, and fuzzy set operators in terms of logic connectives in many-valued logic. This provides a formulation of fuzzy set theory based on many-valued logic. The fuzzy set theory deals with the ill-definition of the boundary of a class through a continuous generalization of set characteristic functions.

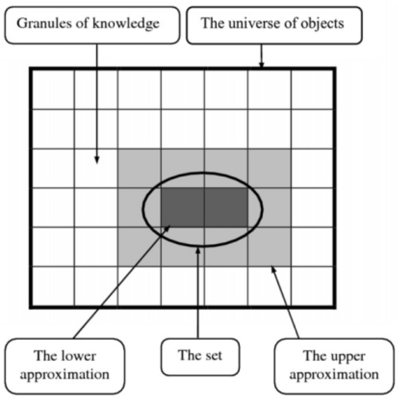
2.3 Rough Set Theory (Early 80's)

Rough set theory, proposed by Pawlak in 1980s [4] can be seen as a mathematical approach to deal with vagueness. The rough set philosophy is founded on the assumption that with every object of the universe of discourse we associate some information (data, knowledge). Objects characterized by the same information are similar in view of the available information about them. The indiscernibility relation generated in this way is the mathematical basis of rough set theory. In the rough set approach indiscernibility is defined relative to a given set of functional (attributes).

Let U be a set called universe and R be an equivalence relation on U, called indiscernibility relation. This pair (U, R) is called an approximation space or Pawlak approximation space. For any X ⊆ U, we call the following two subsets

$$\underline{R}(X) = \{x \in U \mid [x]\underline{R} \subseteq U\},$$
$$\overline{R}(X) = \{x \in U \mid [x]\overline{R} \cap X \neq \emptyset\},$$

the *lower* and *upper* approximation with respect to the approximation space (U, R). Moreover, if $\underline{R}(X) = \overline{R}(X)$, then X



is called a definable set with respect to (U, R). Otherwise, X is called rough set in (U, R). The equivalence relation *R* induces a partition of U, denoted by *U/R*. The subsets contained in *U/R* are called equivalence classes, which are the building blocks to construct rough set approximations.

Therefore, we assume that any vague concept is replaced by a pair of precise concepts – called the *lower approximation* consisting of all objects which surely belong to the concept and the *upper approximation* containing all objects which possibly belong to the concept. The difference between the upper and the lower approximation constitutes the boundary region of the vague concept. Approximations are two basic operations in rough set theory [5].

Any set of all indiscernible (similar) objects is called an elementary set, and forms a basic granule (atom) of knowledge about the universe. Any union of some elementary sets is referred to as crisp (precise) set – otherwise the set is rough (imprecise, vague). Each rough set has boundary-line cases, i.e., objects which cannot with certainty be classified either as members of the set or of its complement, thus assuming that knowledge has granular structure. Due to the granularity of knowledge, some objects of interest cannot be discerned and appear as the same (or similar). As a consequence, vague concepts in contrast to precise concepts cannot be characterized in terms of information about their elements.

2.4 Vague Set Theory (late 80's)

A vague set [6] V in a universe of discourse U is characterized by a *true* membership function, t_v and a *false* membership function, f_v, as follows:

$$t_v: U \rightarrow [0, 1],$$
$$f_v: U \rightarrow [0, 1], \text{ and}$$
$$t_v(u) + f_v(u) \leq 1,$$

where $t_v(u)$ is a *lower bound* on the grade of membership of u derived from the evidence for *u*, and $f_v(u)$ is a lower bound on negation of *u* derived from the evidence against *u*.

2.5 Soft Set Theory (late 90's)

Soft set as a new mathematical tool for dealing with uncertainties that was free from the inadequacy of the parameterization tools [7]. The way of describing any object in the soft set theory principally differs from the way in which we use classical mathematics.

In classical mathematics, we construct a mathematical model and the perception of the exact solution of this model for an object. This model being too complex in nature, no exact solution is established. Thus, we introduce the approximate solution to that object.

Let *U* be an initial universe set and let *E* be a set of parameters. A pair *(F, A)* is called a soft set over *U*, where *F* is a mapping given by

$$F: A \rightarrow P(U) \text{ such that } F(\varepsilon) = \phi \text{ if } \varepsilon \notin A$$

In other words, a soft set over U is a parametrized family of subsets of the universe U. For $\varepsilon \in A$, $F(\varepsilon)$ may be considered as the set of ε-approximate elements of the soft set (F, A).
In the soft set theory, the object has an approximate nature since its initialization, and we do not need to introduce the notion of exact solution. We can introduce any parameterization since there is no restriction on the approximate description of the object.

Analysis of the above set theories:

The indiscernibility between objects is not used in fuzzy set theory. A fuzzy set may be viewed as a class with unsharp boundaries.

Y. Yao [8] illustrates how fuzzy sets and rough sets are generalizations of classical set theory. Both fuzzy set and rough set are applied for modeling vagueness and uncertainty. The rough set theory takes into consideration the indiscernibility between objects. The indiscernibility is typically characterized by an equivalence relation. Rough sets are the results of approximating crisp sets using equivalence classes. The fuzzy set theory deals with the ill-definition of the boundary of a class through a continuous generalization of set characteristic functions. The indiscernibility between objects is not used in fuzzy set theory. A fuzzy set may be viewed as a class with unsharp boundaries, whereas a rough set is a crisp set which is coarsely described.

All these theories can deal with diverse types of uncertainties and imprecision and vagueness but only major problem shared by these theories is their incompatibility with the parameterizations tools. Thus soft set overcomes this problem by allowing parameterization.

3. RELATED WORK

Fuzzy sets and logic are being used in rule mining from very long time, since it provide a simple structure to deal uncertainty. Fuzzy can be applied to both pre-mining and post-mining, i.e., we can either fuzzify the transactions in a dataset, then apply rule mining algorithm to the fuzzified data or we can first generate the rule and then create more fuzzified rule to deal with uncertainty. Some of the techniques perform both [9, 10]. C. Weng et.al [12] proposed method to mine fuzzy association rules based on possibility and deviation along with traditional a priori.

In ARM, support and confidence are the basic measures that have been used since its inception, which define the statistical significance of any rule. Wei-Min Ma et.al [14] provided another measure named *dependency* that assists in finding more interesting rules using fuzzy membership.

Rough set theory has the ability to deal with imprecise, incomplete, missing inaccurate or inconsistence data as is the case. By implementing rough set [15,17], rule induction is much nearer to maximal association rule and more optimized since the attributes can be weighted in upper and lower bounds. Further, Xun Jio et al.[16] confer the advantages of rough sets by creating a decision table and including three steps, such as by eliminating redundant attributes, reducing number of attributes and on scanning decision table once had better produced decision attribute sets.

Herawan and Deris[18] proposed that mining rules can be more specific if used the parameterization, thus applying soft set theory to rule mining. They concluded with the representation of traditional transactional data into soft sets via Boolean valued system. Hence the rules generated were identical to the classical approach but provides more interestingness to regular rules.

An Lu. et.al. [19,22] helps in identifying what is better among Fuzzy sets, Intutionistic fuzzy set and vague sets also giving notion of vague association rules by employing two more measure attractiveness and hesitation of a data item allowing interval-based membership to confine more evidence to an object of the universe.

4. DISCUSSION AND FUTURE SCOPE

So far we have seen that in there lays high probability of uncertainty among data in databases/datasets. There are certain limitations of fuzzy, rough, vague and soft sets [8, 12]. To deal with it, above mentioned set theories which cling to address uncertainty, approximations and vagueness, have been used exclusively to particular application. We can, therefore, use the combination of all to deal with above said notion [17]. As discussed above, we can find certain dependencies and interest measures to find more interesting rules by using either of the theories.

Fuzzy set [9, 10, 11, 12, 13, 14], rough set [15, 16,17], and soft set [18] approaches have already been applied to rule mining. As discussed, coalesce of these theories [20, 21] can be used to discover strong, legitimate, appropriate and interesting rules. Another aspect that could be look upon is using *vague or vague-soft theory*. This allows contrasting measure of how to handle the vagueness of data and up to which certain degree softness can be applied.

5. CONCLUSION

The following conclusion is drawn from above review. The data present in current scenario uses mining technique that can only generate rules with certain statistical bound. Since the data is not crisp throughout, the techniques are inadequate for rules generation. Thus by introducing Fuzzy, Rough, Vague, and Soft set theories we allow flexibility to our approach which can deal with various types of diverse problems pertaining uncertainty, approximation, vagueness, imprecision. Thus improving the way we deduce rules and mine datasets.

6. REFERENCES

[1] Rakesh Agrawal, Tomasz Imielinski & Arun Swami, "Mining Fuzzy Weighted Association Rules" 1993 © SIGMOD ACM

[2] Rakesh Agrawal,Ramakrishnan Srikant, "Fast Algorithms for Mining Association Rules - A Priori", © 1994

[3] L. Zadeh, "Fuzzy Sets", © 1965 Elsevier Inc.

[4] Zdzislaw Pawlak, "Rough sets" © 1982 Elsevier Inc.

[5] Zdzislaw Pawlak, Andrzej Skowron, “Rudiments of rough sets”, Information Sciences 177 (2007) © 2006 Elsevier Inc

[6] D Molodtsov, “Soft Set Theory – First Results”, © 1999 Elsevier Inc.

[7] Gau, W.-L., Buehrer, D.J.: Vague sets. IEEE Transactions on Systems, Man, and Cybernetics23(2), 610–614 (1993)

[8] Yao, Y.Y., “A comparative study of fuzzy sets and rough sets, Information Sciences”, Vol. 109, No. 1-4, pp. 227-242, 1998 © Elsevier Inc.

[9] *Keith C.C. Chan, Wai-Ho Au,* “Mining Fuzzy Association Rules”, 2001 Springer.

[10] Yi-Chung Hu, Ruey-Shun Chen, Gwo-Hshiung Tzeng, “Discovering Fuzzy Association Rules Using Fuzzy Partition Methods” 2002 © Elsevier

[11] Tzung-Pei Hong, Kuei-Ying Lin, Shyue-Liang Wang, “Fuzzy data mining for interesting generalized association rules” Fuzzy Sets and Systems 138 (2003) 255–269 © Elsevier

[12] Cheng-Hsiung Weng, Yen-Liang Chen, “Mining fuzzy association rules from uncertain data” 2009 © Springer

[13] Tzung-Pei Hong, Yeong-Chyi Lee, “Springer An Overview of Mining Fuzzy Association Rules” 2008 © Springer

[14] Wei-Min Ma, Ke Wang, Zhu-Ping Liu, “Springer Mining potentially more interesting association rules with fuzzy interest measure” 2010 © Springer

[15] J. W. Guan, D. A. Bell, D. Y. Liu, “The rough set approach to association rule mining” 2003 © IEEE

[16] Xun Jiao, Xu Lian-cheng, Qi Lin, “Association rules mining algorithm based on Rough Set” 2012 © IEEE

[17] Shu-Hsien Liao, Yin-Ju Chen, “A rough set-based association rule approach implemented on exploring beverages product spectrum” 2013 © Springer

[18] Tutut Herawan, Mustafa Mat Deris, “A soft set approach for association rules mining” 2010 © Elsevier

[19] An Lu and Wilfred Ng, “Vague Sets or Intuitionistic Fuzzy Sets for Handling Vague Data- Which One Is Better?” 2005 © Springer

[20] Chengxi Dong, Dewei Wu, Haibo Liu “Mining Method of Attribute Reduction Based on Rough Fuzzy Set Theory”, 2008 IEEE.

[21] Satya Ranjan Dash, Satchidananda, Uma kant “Usage of Fuzzy, Rough, and Soft Set Approach in Association Rule Mining”, 2012 © IGI Global.

[22] An Lu, Yiping Ke, James Cheng, and Wilfred Ng, “Mining Vague Association Rules” 2007 © Springer