International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. 05 Special Issue 01 | 2020

Revelation of Interesting Association Rules by Incorporating Vague Set Based Measures

Vivek Badhe[1], Terrence Shebuel Arvind[2], Satpal Singh[3]

[1] Assistant Professor, Computer Science, JNKVV Jabalpur, Madhya Pradesh, India
[2] Senior Software Engineer, BVM Microsystems, Indore, Madhya Pradesh, India
[3] Assistant Professor, Department of CSE, Global Engineering College, Jabalpur, Madhya Pradesh, India

ABSTRACT

Uncertainty is a characteristic aspect of human presence. Improbability signifies that whatever can't be purportedly announced with complete certainty has to be addressed using mathematical tools of Set theory, Probability, Fuzziness, and so on. The distinction between the exact and hypothetical techniques should be routed to accumulate strategies that give an extensive report and analyze subtleties in uncertainty. As of late, advancement of progressive systems utilizes Data Mining to give viable business understandings and procedures. Ambiguity impact in neoteric data is high and Association Rules needs some new intriguing concepts that might be rewarding for knowledge discovery. On interlacing classical approach with uncertainty, rules become surer from shady ones found in datasets. The idea to formulate vague association rule (VAR) is novel with a few research accomplished in ARM with vague set theory. Previously, ARM innovation follows customary methodology with measures utilizing support and confidence for rule generation. Since the information has become progressively pervasive, it is essential to discover meaningful methodologies that solve such issues. The relevancy of the term vague association rules employed here is for the rules framed by vague set theory.

Keyword: - Uncertainty, Data Mining, Association Rule Mining, Support, Confidence, Vague Set Theory.

1. INTRODUCTION

In last few decades, there is an explosion in information space leaving us gainsay of how to extract useful knowledge from it. Because of this huge amount of data, database systems are being challenged to manage such a pile. To store information is one thing, but to deal with it is another. To recognize and extract the hidden knowledge and potentially interesting patterns from these large databases is accomplished by data processing. What data mining [1] does is it process large databases and discovers knowledge by application of algorithms that have both statistical and logical significance which is predictive, precise, perceivable, and have profitable outcome. These databases collect information from various sources and are liable to have some magnitude of uncertainty and vagueness in data. To deal with vagueness in databases, vague rule generation is a new direction in finding out the correlations and rules that eventually profits a business as well as helps in decision making process [2].

1.1 Data Mining

The formal definition of data mining may be given as “Data mining or knowledge discovery in the database is the non-trivial extraction of implicit, previously unknown and potentially useful information from data” [1, 2]. Pattern discovery from huge volume of knowledge is one among the foremost desired attributes of knowledge Mining. However, actually, a considerable portion of the available information is stored in text databases, which consists of huge collections of documents from various sources, like news articles, books, digital libraries and sites. Since web search engines became pervasive and search has become integrated, retrieving of data from these search engines contains three essentials: query, documents, and search results. The emerging growth of knowledge mining raises the massive range of complex applications

1.2 Association Rule Mining

Association rules are employed to discover relationships among different items in a transactional database. It was first introduced by Agrawal, Imielinski, and Swami in 1993 [1], and invariably it has been an area of research since then and discovers potentially interesting rules that have certain correlation among data elements of a very large data set.

For any given database D there are a number of complete transactions T. In each transaction there is number of items belonging to itemset I. If n is the different number of items in database D then I = {i1, i2…in} is a set of complete items present in database. Also any transaction tϵT may contain variable set of items over I, i.e., ii, ij, ik $\subset$ I. Each transaction is coupled with a exclusive identifier T_ID. The association rule is of the form of X$\Rightarrow$Y, where X, Y $\subset$ I and X$\cap$Y = ϕ, where X is the consequent of the rule. For finding an association rule

[1, 2], two basic measures are defined; *support* and *confidence*. Any itemset that has statistical significance is associated by its support measure. Then the rule strength generated is calculated by its confidence measure.
Support(s) = support (X⇒Y) = sup(R) = P(XUY).
Confidence(c) = confidence (X⇒Y) = conf(R) = P(Y/X) = support_count(XUY)/support_count(X) = sup(R)/sup(X)

1.3 Vague Set Theory

A vague set [3] is a set of element distributed in a universe that has a grade of membership values in the continuous subinterval of [0, 1]. Hence, such a set can be marked by true membership and false membership functions. The continuous subinterval states both about the evidence that is in favor of the object and also that is opposing it.

Let V be the vague set. If U is the universe of discourse having X objects with x elements than V in U can be defined using the true membership (Vt) and false membership (Vf) functions. Considering that both Vt and Vf consorted as real numbers in the subinterval of [0, 1]. Also Vt is the lower bound on grade of membership of x derived in favor of x, and Vf is the lower bound on grade of membership derived against x, with each element in X where Vt + Vf ≤ 1 and Vt:X→[0,1], Vf:X→[0, 1]. Hence, the grade of membership of x is bounded to a subinterval [Vt(x), 1-Vf(x)] of [0, 1].

1.4 Vague Association Rule

The concept of Vague Association Rules (VARs) is based on four types of support and confidence [9], which applied on AH-Pair Transactions. Given the transactions of the customers, we then aggregate the transactions to obtain the intent of each item. Based on the intent of an item, we next define the attractiveness and hesitation factor of it. (Intent, Attractiveness factor and Hesitation factor, AH-Pair Transactions) The intent of an item x, denoted as intent(x), is a vague value [α(x), 1 − β(x)]. The attractiveness of x, denoted as MA (x), is defined as the median membership of x, i.e., MA (x) = (α(x) + (1 − β(x)))/2. The hesitation factor of x, is denoted as MH (x), is defined as the imprecision membership of x, i.e., MH (x) = ((1 − β(x)) − α(x)). The group of hMA (x), MH (x)i is called the AH-pair of x. An AH-pair transaction T is a tuple <v1, v2, . . . , vm > on an itemset IT = {x1 , x2, . . . , xm }, where IT ⊆ I and vj = hMA (xj), MH (xj)i is an AH-pair of the item xj , for 1 ≤ j ≤ m. An AH-pair database is a sequence of AH-pair transactions. (Vague Association Rule) A Vague Association Rule (VAR), r = (X => Y), is an association rule obtained from an AH-pair database. Based on the attractiveness factor and hesitation factor of an item, we describe four different types of support and confidence of a VAR depending on what kind of knowledge we want to acquire. For clarity, we use A to denote Attractiveness factor and H to denote Hesitation factor.

1.5 Support and Confidence

Given an AH- pair database, D , we can define different types of support and Confidence for an itemset Z or a VAR X => Y ,where X U Y => Z [4].

Support TS = $\sum$ Mp / | D | and Confidence C = TS(Z)/ TS(X)

Where TS = {A-sup, H-Supp,AH-Supp,HA-Supp}; p = {X,Y,XUY} ; M = {MA,MH, MA.MH MHMA}

2. LITERATURE REVIEW AND RELATED WORK

An Lu et al. [4] gave a proposition of incorporating vague set theory within data mining and concluded that every database does have some vagueness in it. But vagueness in itself is indistinguishable and to identify objects that are vague in databases is important where vague set theory can be applied. They formulated a vague functional dependencies (VFDs) concept that finds vagueness by using a measure called Similar Equality (S_{EQ}). The proposed method forms vague relations in the database by using the long-familiar integrity constraint Functional Dependency [5]. The vague relation is an approximation which is determined by the median and imprecision membership functions and *mi-overlap* between vague sets relations. In another related work [6], a method was proposed that mined vague association rule (VARs). Herein, the vague rules are found by employing a notion of three parameters: intent, attractiveness and hesitation which are being derived from the median and imprecision memberships. The AH-pair database is then formed by attractiveness (A) and hesitation (H) which is mined and rules are generated. The paper, which is primarily intended towards e-commerce, proposed that every item carries with itself a hesitation status of whether it'll be sold or not. The hesitation statuses (HS) [7] which is a factor of hesitation taken as a whole ingratiate much information. For instance, any item with hesitation status as zero has the maximum attractiveness which makes mining hesitation information important to find VARs.

Another important research within domain specificity of academics conducted using vague sets theory and association rules is done by Anjana Pandey et al. [8][12]. In their study, the courses taken by students correspond to vague objects and their relation with students and their correlation. In the domain of consumer goods, Terrence Shebuel Arvind et al. also conducted some experiments and highlighted the benefit of using vague sets with data mining [9]. The classical of traditional support and confidence parameters in *apriori*

method was enhanced to use true support and true confidence. These are quite different than the traditional counterparts as they take into consideration the inclusive vagueness. The results were quite good in performing the market-basket analysis with certain products. They defined two types of rules generated: subversive rules and puissant rules. The first being the undermining of rules while the latter being the over mining. Satpal Singh et. al [10] also modified the measures of interest by incorporating weighted feature of profit and inventory as a factor the conventional statistical support and confidence is customize as per the need of business and generate the lucrative rules for business concern. In the area of customer merchandise, Vivek Badhe et al. [11] additionally directed a few examinations and featured the advantage of utilizing obscure sets with information mining. The outcomes were very acceptable in playing out the Commercial investigation with specific items. Arun Kumar Singh et. al [13], Surabhi Pathak et. al. [14] and B Sowkarthika et. al [15] also explores the vague set theory and its incorporation with mining association rule for the discovery of vague association rule mining, hesitation mining and profit pattern discovery.

3. Proposed Approach and Description

The proposed approach is particularly produced for business exchanges where everything having an item code yet for the distinctive pressing the code is contrast for a similar thing. This will make the vagueness and can't be bargain by the ordinary association rule digging for this reason we utilize the vague association rule mining and create those rules which produce the benefit criticalness yet precluded due to statically infringement brought about by vagueness. We utilize a calculation to mine Vague set based Association Rules. As examined in past area, the vague sets have discovered application in association rule mining from numerous points of view. The method we propose comprise two new formula for discovering support and Confidence. The proposed system portrayed in Figure 1 shows the utilization of vague set hypothesis to frame vague association rules by applying not many vague estimates expressed in following definitions:

3.1 Vague set based proposed terminology for association rule mining:

i. Vague Percentage (VP): The vague percentage denotes the amount of vagueness contained in a database for a particular vague item. Since for an item, there exists in a database, a true membership (Vt), a false membership (Vf), and a certain amount of vagueness, termed as vague percentage (VP). Thus, a database consists $|D| = V_t + V_f + V_p \leq 1$, i.e, all the memberships and percentage values. The vague percentage can be denoted as $V_p = |D| - (V_t + V_f)$.

The new type of support values is as follows:

ii. **Vague Support** (Svg): The Vague Support (Svg) of a vague item is defined as: the sum of true membership (Vt) of that item with the vague percentage (Vp) found in the whole database D. It is denoted by: (Vt + VP) / |D| or (1 - Vf)/ |D|. Thus the lower bound on the negation of evidence against an item in whole database is called vague support.

Vague Support, $S_{vg} = \frac{true\ members\ hip\ of\ x_i\ and\ Vague\ Percentage}{full\ description} = \sum_{i=1}^{n} \frac{(V_t(x_i)+V_p)}{|D|} = \sum_{i=1}^{n} \frac{(1-V_f(x_i))}{|D|}$

The new type of confidence value is given as follows:

iii. Vague Confidence (Cvg): The vague confidence is defined as the union of vague support values of any items A and B to either of its individual vague support. It is denoted by:

Vague Confidence, $C_{vg} = \frac{Vague\ \ Support\ \ of\ \ (A \cup B)}{Vague\ \ Support\ \ of\ (A)\ or\ (B)} = \frac{S_{vg}(A \cup B)}{S_{vg}(A)\ or\ S_{vg}(B)}$

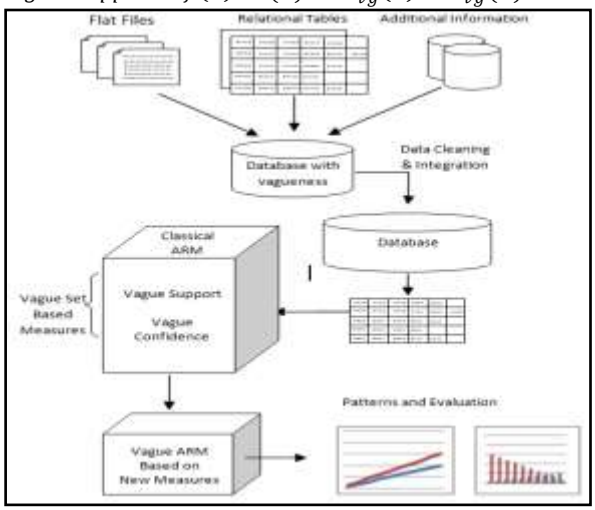

Figure 1: Proposed Approach with Vague Set Based Measures

iv. Variation (Vague Mapping) Table: Once the vague items in the database are identified it is important to make them available during the mining process. For this purpose a variation table or vague table is designed show in Table-1, which consist the number of vague items in its N rows and the variant of that item in its M columns.

Table 1: Variation (Vague Mapping) Table

Item(s)/Variants	Var(1)	Var(2)	Var(3)	…Var(P)	Var(M)
Item(A)	A1	A2	A3	…AP	…
Item(B)	B1	B2		…BP	Bm
…	…	…	…	…	…
Item(N)	N1	N2	N3	…	…

Now here apply the means of proposed approach as given in Figure-1 to test dataset, the standard Apriori algorithm changed with novel proposed Vague Support and Vague Confidence measures creates the particular Vague Association Rules.

4. EXPERIMENTS AND RESULTS

In any traditional FMCG market, there is a certain condition where an item can be represented by many unique codes. This usually happens due to different production and manufacturing batch numbers, the weight of the items, etc. In our experimentation, we collected the data of a grocery store and applied vague set theory to calculate some interesting results shown in Chart-1 and Chart-2. After removing noise from our dataset, we observed that,

- A product with **same name** having **different packaging** in quantity has different product codes and is considered a vague object e.g shown in Table2. Not necessarily every item is vague.

Table2 : Same product, multiple packaging

Product Name	Product Code 1	Product Code 2	Product Code 3
Sugar	ABK057	ABK058	ABK059
Packaging (weight)	500gm	1 kg	5 kg

- Then we classify the dataset by creating two files:
- Task Relevant Attributes File (**TRAF**) containing **Bill Number** and **Product Code** in a transaction.
- Vague Object Identification File (**VOIF**) containing vague item information found in database.

The results computed have an interesting overlapping between items. As the charts below depicts, a numerous vague (v) and non-vague (nv) item combination is possible when vying with vague support and vague confidence.

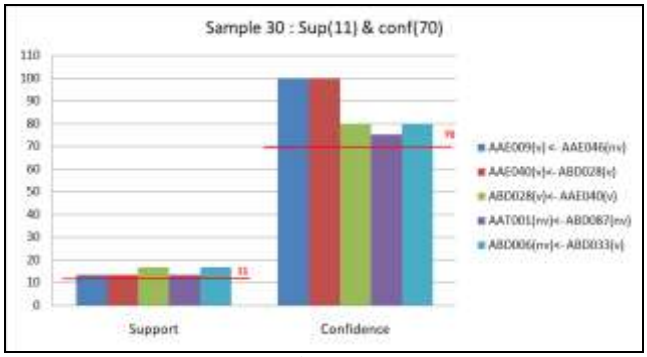

Chart 1 : Rules of Interest in Sample 30

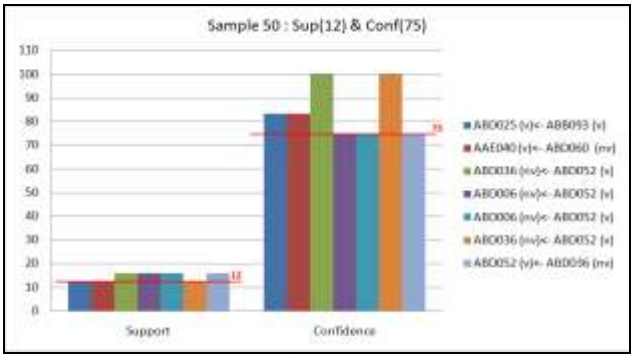

Chart 2: Rules of Interest in Sample 50

5. CONCLUSION

The proposed work gives us an advantage in the direction of association rules by incorporating vague set theory. It allows taking into consideration the vagueness existing in databases and how it can be utilized to mine better results correlating with different objects. The vague support and vague confidence distinctly finds the results by showing an overlap between items. A non-vague item can be correlated to a vague item and vice-versa. There also can be a case where only non-vague items produce result or only vague items. These results herein are better than the traditional approach and suitable as an alternative approach to association rule mining.

6. REFERENCES

[1] Rakesh Agrawal, Tomasz Imielinski & Arun Swami, "Mining Association Rules between Sets of Items in Large Databases" Proceeding SIGMOD '93 Proceedings of the 1993 ACM SIGMOD international conference on Management of data Pages 207-216 ACM New York, NY, USA ©1993
[2] Rakesh Agrawal, Ramakrishnan Srikant, "Fast Algorithms for Mining Association Rules - A Priori", Proceeding VLDB '94 Proceedings of the 20th International Conference on Very Large Data Bases Pages 487-499 Morgan Kaufmann Publishers Inc. San Francisco, CA, USA ©1994
[3] Wen-Lung Gau and Daniel J. Buehrer, "Vague Sets", © 1993 IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, VOL. 23, NO. 2, MARCHIAPRIL 1993
[4] An Lu, Wilfred Ng, "Managing Merged Data by Vague Functional Dependencies" 2004 © Springer.
[5] An Lu, Wilfred Ng, "Handling Inconsistency of Vague Relations with Functional Dependencies" Conceptual Modeling - ER 2007 Lecture Notes in Computer Science Volume 4801, 2007, pp 229-244 © Springer
[6] An Lu, Yiping Ke, James Cheng, and Wilfred Ng, "Mining Vague Association Rules" DASFAA 2007, LNCS 4443, pp. 891–897, 2007 © Springer-Verlag Berlin Heidelberg 2007
[7] An Lu, Yiping Ke, James Cheng, and Wilfred Ng, "Mining Hesitation Information by Vague Association Rules" ER 2007, LNCS 4801, pp. 39–55, 2007 © Springer-Verlag Berlin Heidelberg 2007
[8] Pandey Anjana, Pardasani K.R. ,A Model for Mining Course Information using Vague Association Rule, International Journal of Computer Applications (0975 – 8887), Volume 58– No.20, November 2012
[9] Terrence Shebuel Arvind, Vivek Badhe, "Discovery of Certain Association Rules from an Uncertain Database", International Conference on Computational Intelligence and Communication Networks © 2015 IEEE
[10] Satpal Singh , Vivek Badhe, " Profit Association Rule Mining with Inventory Measures", International Conference on Computational Intelligence and Communication Networks © 2015 IEEE
[11] Vivek Badhe, Dr. R.S Thakur and Dr. G.S Thakur, "Vague Set Theory for Profit Pattern and Decision Making in Uncertain Data" International Journal of Advanced Computer Science and Applications(IJACSA), 6(6), 2015. http://dx.doi.org/10.14569/IJACSA.2015.060625
[12] Pandey Anjana, Pardasani K.R. ," A Model for Vague Association Rule Mining in Temporal Databases, Journal of Information and Computing Science, Vol. 8, No. 1, 2013, pp. 063-074
[13] Arun Kumar Singh, Akhilesh Tiwari, "Vague Set Based Association Rule Mining for Profitable Patterns",International Journal for Science and Advance Research In Technology,ISSN [Online] : 2395-1052, - Volume 2 Issue 2 –FEBRUARY 2016 PP:1-6
[14] Surabhi Pathak and Akhilesh Tiwari. A Survey on Hesitation Information Mining. International Journal of Computer Applications 141(9):14-21, May 2016.
[15] B Sowkarthika, Akhilesh Kumar Tiwari, Uday Pratap Singh, "Elephant Herding Optimization based Vague Association Rule Mining Algorithm, International Journal of Computer Applications (2017), DOI:10.5120/ijca2017913626

BIOGRAPHIES

	Dr. Vivek Badhe, Working as Assistant Professor (Computer Science), at Jawaharlal Nehru Krishi Vishwa Vidyalaya (JNKVV), Jabalpur, The State Agriculture University of Madhya Pradesh. Received M.Tech (Computer Technology and Application) in 2006 From Rajiv Gandhi Proudyogiki Vishwavidyalaya (RGPV), Bhopal, The State Technical University of Madhya Pradesh. Received PhD (Computer Application) from Maulana Azad National Institute of Technology (MANIT) The Institute of National Importance, Bhopal in 2015. Published more than 30 Papers in National and International Conferences and Journals. Areas of Interest are Data Science, Machine Learning and Soft Computing.
	Terrence Shebuel Arvind received his Masters of Technology in Computer Technology and Applications from Rajiv Gandhi Technical University Bhopal, The State Technical University of Madhya Pradesh, India in 2015. Currently he is working as Senior Software Engineer at BVM Microsystems Indore, Madhya Pradesh, India. He is inclined towards research and actively participates in various new technological trends that help businesses grow their strategy and online presence.
	Prof. Satpal Singh, Working as Assistant Professor, Department of Computer Science & Engineering, at Global Engineering College, Jabalpur. Received M.Tech (Computer Technology and Application) in 2016 From Rajiv Gandhi Proudyogiki Vishwavidyalaya (RGPV), Bhopal, The State Technical University of Madhya Pradesh. Published more than 15 Papers in National and International Conferences and Journals. Areas of Interest are Data Mining, Machine Learning and Soft Computing.