Vague Set Based Association Rule Mining for Profitable Patterns

Arun Kumar Singh[1], Akhilesh Tiwari[2]
Department of CSE & IT
[1,2] Madhav Institute of Technology and Science Gwalior, MP

Abstract- *In the realistic situation, there are many areas which contain imprecisely specified data. This imprecise data indicates the presence of vagueness, incompleteness and uncertainty which causes the problem during important decision-making task. Traditional association rule mining has limitations as it only deals with the items that are actually present in the transaction and ignores the items that are almost sold (customer hesitate to purchase such items). Furthermore, these items may be placed with profitable items by imposing attractiveness measure. Profit pattern mining in the scenario of imprecise and vague environment is very difficult which is frequent in recent years. For the effectiveness of retrieved hesitated patterns and rules, the concept of vague set theory is used. For the same consideration, price as weighting factor is used for generation of profitable patterns.*

Keywords- Vague Set Theory, Vague Association Rule, AH-pair Database, profit Patterns

I. INTRODUCTION

The problem of decision making, especially in financial issues is a crucial task in every business. In past years, data mining technology follows the conventional approach that offers only statistical analysis and discovers rules. The main technique uses support and confidence measures for generating rules. But since the data have become complex today it is crucial to find the solution to deal with such problems. Profit pattern mining hits the target, but this job is found very difficult when it depends on the imprecise and vague environment, which is frequent in recent years.

Association rules can effectively use to uncover unknown relationship producing the result that can provide a basis for forecasting and decision making. Traditional association rule mining model [1] considers that items have the same significance without taking consideration of item weight within the transaction and also ignores vague status of item which is not always case. Considering an example [TV →vcd,1%,80%] may be more important than [bread →milk,3%,80%] although former holds a lower support because those items in the first rule usually come with more profit per unit auction, but conventional association rule is disregarded this variance.[21]

The vague set theory can be a basis of modeling hesitation information of items which used interval-based membership that captures three types of evidence with respect to an object in the universe of discourse: support, against, hesitation. Thus, it can naturally model the hesitation information of an item in mining context as the evidence of hesitation with respect to the item.

To study the relationship between the support evidence and the hesitation evidence with respect to an item the concept of attractiveness and hesitation which are based on the median membership and the imprecision membership that are derived from the vague membership in vague sets. An item with high attractiveness means that the item is well sold and has high possibility to be sold next time. An item with high hesitation means that customers are always hesitating to buy the item due to some reason (e.g., the customer is waiting for price (or weight) reduction) but has a high possibility to buy it subsequent time, if the cause of giving up the item is acknowledged and resolved (e.g., some promotion on the item is provided) [6].

II. BACKGROUND AND RELATED WORK

The traditional association rule mining given by Agrawal, Imielinski and Swami (1993) ignores the uncertain situations [1]. Number of studies were done that extends the traditional association rule mining for handling uncertain situations and data in the different application such as mining fuzzy association rules. Fuzzy association rule are proposed to handle quantitative items in the form of “ A is X” ⇒ “ B is Y”, where X and Y is set of items and A and B are fuzzy concepts represented as fuzzy sets.

Gau and Buehrer proposed vague set theory [2] for handling uncertain situations which provide two membership functions instead of single membership function in the fuzzy set thus it provide a natural way of handling hesitation information. An important study made by An Lu and Wilfred Ng which give the comparison between vague set and intuitionistic fuzzy set for handling vague data which one is better[7].

Yanhong Li and Zheng Qin made a comparative analysis of similarity measure between intuitionistic fuzzy and vague set. This research provides benefit of selection and application of similarity measures for intuitionistic fuzzy sets and vague sets in practice [22].

Vague information is common in many database applications due to intensive data dissemination arising from different pervasive computing sources such as high volume data obtained from different resources. Lu and Wilfred Ng give how to maintaining consistency of vague database using data dependencies. They extend the concept of functional dependency (FD) in relational databases by applying vague set theory in order to handle the widely existent vague information and proposed vague functional dependency (VFD) [5].

After this An Lu , James Cheng and Wilfred Ng proposed the notion of vague association Rules (VARs) [6] and devise an algorithm to mine the vague association rules they show that their algorithm for vague association rule capture more specific and richer information than traditional association rules. However the modeling of hesitation information with respect to different hesitations status is new concept given in his paper "Mining Hesitation Information by Vague Association Rules" a basic approach to incorporate the hesitation information into association rules (ARs) is provided in this.

K.R. Pardasani and Anjana in 2012 Pandey gives a model for mining course information using Vague Association rule [12] in which they studies that different university offering different courses of different types over several years and find out biggest issue with that how to get information to make course more effective and solve this issue through vague association rule (VARs). They extend this concept further by giving a model for vague association rule mining in the temporal database [11].

Weighted association rule (WAR) doesn't hinder with the procedure of generating frequent itemset. Relatively, it focuses on how weighted association rules can be generated by examining the weighting factors of the items included in generated frequent itemsets. Therefore, we could classify this type of weighted association rule mining methods as a technique of post-processing or maintaining association rules [15].

III. PRELIMINARIES

The basics of handling uncertainty and vagueness for hesitation information can be explained through vague set theory and intuitionistic fuzzy set theory. The graphical representation of vague set theory is more intuitive in perceiving vague values. The following concept is used in developing the model and algorithm for a weighted vague association rule for hesitation mining.

Let U be a classical set of objects, called the universe of discourse, where an element of U is denoted by u.

A. Vague Set

A vague set V in a universe of discourse U is characterized by a true membership function, α_V, and a false membership function, β_V, as follows: $\alpha_V : U \rightarrow [0,1]$, $\beta_V : U \rightarrow [0,1]$, and $\alpha_V(u) + \beta_V(u) \leq 1$, where $\alpha_V(u)$ is a lower bound on the grade of membership of u derived from the evidence for u, and $\beta_V(u)$ is a lower bound on the grade of membership of the negation of u derived from the evidence against u. Suppose a universe of discourse $U = \{u_1, u_2, u_3 \ldots u_n\}$. A vague set V of the universe of discourse U can be represented by expression

$$V = \sum_{i=1}^{n} [\,\alpha_V(u_i)\,, 1 - \beta_V(u_i)] / u_i \qquad (1)$$

Where $0 \leq \alpha(u_i) \leq \beta(u_i) \leq 1$ and $1 \leq i \leq n$. In other words, the grade of membership of u_i is bounded to a subinterval $[\alpha_V(u_i), 1 - \beta_V(u_i)]$ of $[0,1]$.

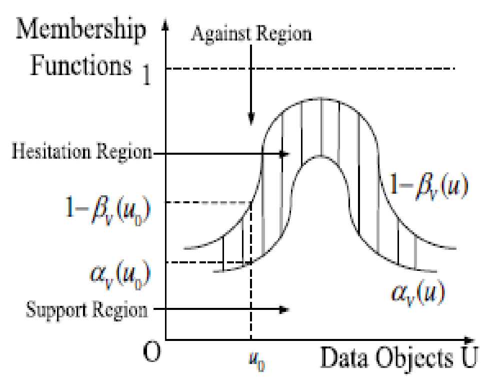

Fig 1. Membership Function of vague set

B. Median and Imprecision Membership

In order to compare vague values we used two memberships: median membership and imprecision membership. It is noteworthy that for a given vague value $[\alpha(x), 1 - \beta(x)]$ we have unique median membership M_m and imprecision membership M_i and vice versa.

Median membership is defined as

$$M_m = \frac{1}{2}\,(\alpha + (1 - \beta)) \qquad (2)$$

which represents the overall evidence contained in a vague value. It can be checked that $0 \leq M_m \leq 1$. Obviously, the vague value [1, 1] has the highest M_m, which means the corresponding object definitely belongs to the vague set (i.e., a crisp value). While the vague value [0, 0] has the lowest M_m, this means that the corresponding object definitely does not belong to the vague set.

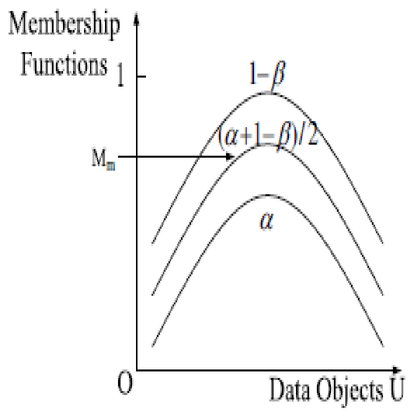

Fig 2 Median Membership of a Vague set

Imprecision membership is defined as

$$M_i = ((1 - \beta) - \alpha) \quad (3)$$

which signifies the overall imprecision of a vague value. It can be veteran that 0 ≤〖 M〗_i≤1. The vague value [p, p] (p ∈ [0, 1]) has the lowest M_iwhich means that the membership of the corresponding object is exact (i.e., a fuzzy value). While the vague value [0, 1] has the highest M_i this means that we do not have any information about the membership of the corresponding object.

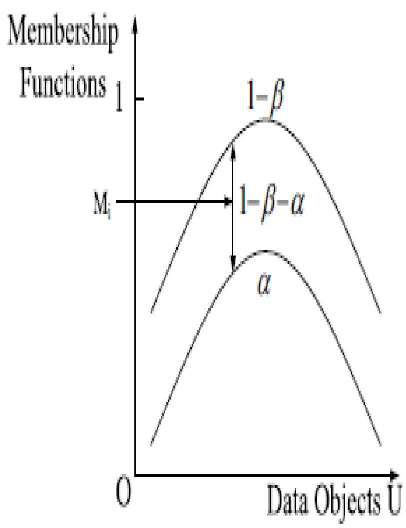

Fig 3 Imprecision Membership of a vague set

The median membership and imprecision membership are employed to measure the attractiveness and hesitation of an item respect to a costumer.

C. Hesitation and overall Hesitation

Given an item $x \in I$ and a set of HSs $S = \{s_1, s_2, \ldots, s_n\}$ with a partial order $\leq$. The hesitation of x with respect to a hesitation status HS $s_i \in S$ is a function $h_i(x): I \to [0, 1]$ such that

$$\alpha(x) + \beta\,(x) + \sum_{i=1}^{n} h_i(x) = 1 \quad (4)$$

where $h_i(x)$ represents the evidence for the HS s_i of x. The overall hesitation of x with respect to S is given by

$$H(x) = \sum_{i=1}^{n} h_i\,(x) \quad (5)$$

This can be easily find from the above definition that

$$H(x) = 1 - \alpha(x) - \beta(x) \quad (6)$$

D. Intent and overall intent

Given a set of HSs (S, ≤), the intent of an item x with respect to an HS $s_i \in S$, denoted as $\mathrm{int}(x, s_i)$ is a vague value $[\alpha_i(x), 1 - \beta_i(x)]$ which is sub interval of $[\alpha(x), 1 - \beta(x)]$ .The overall intent of x denoted as INT(x) is the interval$[\alpha(x), 1 - \beta(x)]$.

E. Attractiveness and overall Attractiveness

The attractiveness of x with respect to an HSs_i, denoted as $\mathrm{att}(x, s_i)$ is defined as the median membership of x with respect to S_i that is $\frac{1}{2}\left(\alpha_i(x) + \left(1 - \beta_i(x)\right)\right)$. The overall attractiveness of x is a function $\mathrm{ATT}(x): I \to [0, 1]$ such that

$$\mathrm{ATT}(x) = \frac{1}{2}\left(\alpha(x) + \left(1 - \beta(x)\right)\right). \quad (7)$$

F. AH-pair transaction and database

An AH-pair database is sequence of AH-pair transactions. An AH-pair transaction T is a tuple $<v_1, v_2, \ldots, v_m>$ on an itemset $I_T = \{x_1, x_2, \ldots, x_m\}$ where $I_T \subseteq I$ and $v_j = < M_A(x_j), M_H(x_j) >$ is an AH-pair of the item x_j with respect to a given HS or the overall hesitation for $1 \leq j \leq m$.

IV. VAGUE ASSOCIATION RULE

Association rule mining can be used to discover unknown or hidden correlation between items originate in the database of the transactions. A vague association rule (VAR) r=(X⇒Y) is an association rule obtained from an AH-pair database. The vague association rules (VARs) has four types of support and confidence which evaluates their quality. Based on the attractiveness and hesitation of an item with respect to HS, the different type of support and confidence of vague association rule is defined [6, 9].

For example if someone have special interest in the association between well-sold items (high attractiveness) and all most sold items (high hesitation) then some analysis between the former and later may make some improvements to boost the sales of the latter. For this purpose Attractiveness-Hesitation (AH) support and confidence of a VAR to evaluate the vague association rule. Here A (or H) can refer to either the overall attractiveness (or Hesitation) of a given HS.

A. Support

For a given AH-pair database, D, four types of support for an itemset Z or a VAR$X \Rightarrow Y$ where $X \cup Y = Z$ as follows:

1. The attractiveness support (A-support) of Z is defined as$\frac{\sum_{T\in D}\prod_{z\in Z} M_A(z)}{|D|}$.
2. The hesitation support (H-support) of Z is defined as$\frac{\sum_{T\in D}\prod_{z\in Z} M_H(Z)}{|D|}$.
3. The attractiveness-hesitation (AH-support) of Z is defined as$\frac{\sum_{T\in D}\prod_{x\in X, y\in Y} M_A(x)M_H(y)}{|D|}$.
4. The hesitation-attractiveness (HA-support)of Z is defined as$\frac{\sum_{T\in D}\prod_{x\in X, y\in Y} M_H(x)M_A(y)}{|D|}$

Z is an A (or H or AH or HA) frequent item FI if the A-support (or H-support or AH-support or HA-support) support of Z is no less than the (respective A or H or AH or HA) minimum support threshold σ.

B. Confidence

For a given AH-pair database four types of confidence for an itemset Z or a VAR, $r = (X \Rightarrow Y)$ where $X \cup Y = Z$ as follows:

1. If both X and Y is A FIs, then the confidence of rule called the A-confidence of rule is defined as $\frac{\text{Asupp }(Z)}{\text{Asupp }(x)}$.
2. If both X and Y is H FIs, then the confidence of rule, called the H-confidence of rule a is defined as $\frac{\text{Hsupp }(Z)}{\text{Hsupp }(x)}$.
3. If X is an A FI and Y is an H FI, then the confidence of rule , called the AH-confidence is defined as $\frac{\text{AHsupp }(Z)}{\text{Asupp }(x)}$
4. If X is an H FI and Y is an A FI, then the confidence of rule , called the HA-confidence is defined as $\frac{\text{HAsupp }(Z)}{\text{Hsupp }(x)}$.

VI. PROPOSED METHODOLOGY

On critical analyzed it is examine that hesitation for an item decrease the attractiveness of an item and hence reduce the probability of selling the item that results decrease in profitability of store in many folds. The cost of an item can be a major factor that increases the hesitation or attractiveness toward an item and hence it is an important factor in mining the hesitation information. The proposed an algorithm mine the vague association rules that is used for the profit patterns generation which if used for to increase the profitability of store.

Proposed Algorithm:

1. Initialize array to store intent, AH-pair and frequent itemset;
2. Initialize favor(α) and against (β) variable with value zero;
3. Increment favor (α) by one when value in database indicates that item is purchased;
4. Increment against (β) by one when value in database indicates that item is not purchased;
5. Generate intent $[\alpha, 1-\beta]$ using favor and against and store in intent array;
6. Generate Attractiveness as a median membership i.e. $\frac{1}{2}(\alpha + (1-\beta))$ and Hesitation as a difference of α and $1-\beta$ using intent;
7. Store the value as attractiveness and hesitation in AH-pair array.
8. Mine all items whose attractiveness and hesitation is greater than minimum support and minimum weighted support.
9. Generate the candidate set of size two as C_{A_2} from A_1, $C_{A_1H_1}$ from A_1 and H_1 and C_{H_2} from H_1;
10. Generate all frequent itemset of size two whose support and weighted support is greater than minimum threshold;
11. Similarly generate frequent itemset for all size and store in array of frequent itemset;
12. Generate subsets of frequent items such that subset is contained in superset of frequent items;
13. If subset is weighted vague frequent itemset, then find rule otherwise obliterate the subset the from the itemset list;
14. Return weighted vague rule;

VII. EXPERIMENT AND RESULT

To evaluate the result experiment is performed on MATLAB 2014b installed on machine having 2.53 GHz intel®core i3 CPU with 3 GB RAM. Dataset used for experimentation purpose is prepared for items under observation for each customer will purchased it or not to their visit in store and status of customer exploration is recorded. Each item is associated to cost of the item which is considered as the weight of the item. The set of hesitation status is given by $S = \{ S_1, S_2, S_3, S_4, S_5 \}$.

Table 1: Experimental result on Fixed Threshold

Customer ID	Total Hesitated pattern	Total Frequent hesitated patterns	Execution Time
1	611	12	4.6792
2	553	9	3.5230
3	558	10	3.2586
4	554	8	4.2583
5	617	12	4.2016

The above result is evaluated on considering minimum support = 0.01, minimum confidence = 0.01 minimum weighted support = 0.2 and weighted confidence = 0.2. Total frequent hesitated patterns are very less in comparison to total hesitated patterns. Thus decision maker only required to make selling strategies to increase the profit.

VII. CONCLUSION

Hesitation information of an item contains valuable knowledge which can use for making the selling strategies of an item. Cost is much-concerning factor for the attractiveness of an item hence we proposed algorithm which is efficient in discovering hesitation information of items based vague set theory. The effectiveness of this algorithm is also revealed by experiments. This algorithm has wide application in real the scenario like temporal occasional mining, weighted ranked based scores together click through data of search result can be modeled as an object having different hesitation status. In this case, vague association rule can be used to reflect different user's preferences. Such models can further be developed and extended to problems involving mining useful information in different situations where hesitation is present.

REFERENCES

[1] Agrawal R., Imielinski T., Swami A.N. "Mining association rules between sets of items in large databases". In Buneman, P., Jajodi a, S., eds.: SIGMOD Conference, ACM Press (1993) 207–216.

[2] Gau W.L., Buehrer D.J. "Vague sets". IEEE Transactions on Systems, Man, and Cybernetics 23 (1993) Pages-610–614.

[3] Zadeh, L.A. "Fuzzy sets". Information and Control 8 (1965) Pages-338–353.

[4] An Lu and Wilfred Ng "Maintaining consistency of vague databases using data dependencies "Data and Knowledge Engineering, Volume 68,2009,Pages 622-641.

[5] Lu A., Ng W. "Managing merged data by vague functional dependencies". In: Atzeni P., Chu W., Lu H., Zhou S., Ling T.-W. (eds.) ER 2004. LNCS, vol. 3288, pp. 259–272. Springer, Heidelberg.

[6] An Lu and Wilfred Ng "Mining Hesitation Information by Vague Association Rules" Lecture Notes in Computer Science ,Springer Volume 4801/,2008,pg 39-55.

[7] Lu A., Ng W. "Vague sets or intuitionistic fuzzy sets for handling vague data": Which one is better? In: Delcambre L.M.L., Kop C., Mayr H.C., Mylopoulos J., Pastor, O. (eds.) ER 2005. LNCS, vol. 3716, pp. 401–416. Springer, Heidelberg

[8] Lu. A., Ng. W: Handling Inconsistency of vague relations with functional dependencies. In: ER (2007). LNCS Vol 4321 pp 301-312.

[9] Lu,A., Ke,Y., Cheng ,J., Ng,W.: Mining Vague association rules.In:DASFAA,pp.891-897(2007)

[10] Atanassov, K.T.: Intuitionistic Fuzzy Sets: Theory and Applications (Studies in Fuzziness and Soft Computing). Springer-Verlag Telos (1999).

[11] Pardasani K.R., Anajan Pandey "A Model for Vague association rule Mining in Temporal Database" in Journal of Information and Computing Science, Vol.8, 2013, ISSN 1746-7659, pp. 063-074.

[12] Pardasani K.R., Anajan Pandey " A Model for Mining Course Information Using Vague Association Rule " in International Journal of Computer Applications , Vol 58, ISSN 0975-8887, November 2012.

[13] Vivek Badhe, Arvind T.S. " Comparative Analysis of Fuzzy, Rough, Vague and Soft set Theories in Association Rule Mining" in International Journal of Scientific Progress and Research (IJSPR) , Vol-2, ISSN 2349-4689, November 2014.

[14] A. Tiwari, R.K. Gupta and D.P. Agrawal "A survey on Frequent Pattern Mining: Current Status and Challenging issues" Information Technology Journal 9(7) 1278-1293, 2010.

[15] W. Wang, J. Yang and P. Yu “Efficient mining of weighted association rules (WAR)”, Proc. Of the ACM SIGKDD Conf. on knowledge Discovery and Data Mining. 270-274,2000.

[16] Feng Tao, “Mining Binary Relationships from transaction data in weighted Setting” PhD Thesis, School of Computer science, Queen’s University Belfast, UK, 2003.

[17] G.D. Ramkumar, Sanjay Ranka, and Shalom Tsur, “Weighted Association Rules: Model and Algorithm” KDD1998, 1998.

[18] N.Pasquier, Y.Bastide, R. Taouil, and L.Lakhal, “Efficient mining of association rules using closed itemset lattices,” Information Systems, Vol 24, No. 1, 1999, pp. 25-46.

[19] Bing Liu, Wynne Hsu, and Yiming Ma, “Mining Association Rules with Multiple Supports”, Proc. Of the ACM SIGKDD Int’1 Conf. On Knowledge Discovery and Data Mining (KDD-99), SanDiego, CA, USA, 1999.

[20] Jiawei Han and Yongjian Fu. “Discovery of Multiple-Level Association Rules from Large Databases” in the Proceedings of the 1995 Int’1 Conf. on Very Large Data Bases (VLDB’95), Zurich, Switzerland, 2002, pp. 420-431.

[21] F. Tao, F. Murtagh and M.Farid, “Weighted Association Rule Mining Using Weighted Support and Significance Framework, “Proc , ACM SIGMOD ’03, pp. 661-666, 2003.

[22] Yanhong Li, David L.Olson, Zheng Qin “Similarity measures between intuitionistic fuzzy (vague) sets: A comparative analysis” Elsevier, Pattern Recognition Letters 28 (2007) pp. 278-285.