

Genetic Based Hesitation Information Mining for Profitability Management

Prateek Shrivastava¹ and Akhilesh Tiwari²

¹*Department of CSE & IT, Madhav Institute of Technology and Science,
Gwalior (M.P.), India*

²*Associate Professor, Department of CSE & IT, Madhav Institute of Technology
and Science, Gwalior (M.P.), India*

¹*prateekshrigwl@gmail.com, ²atiwari.mits@gmail.com*

Abstract

Traditional Association Rule Mining has been extensively used to discover interesting rules or relationships between items in large databases but it has limitations that it solely deals with the items or products that are sold but avoids the items that are nearly sold. These nearly sold things carry hesitation data since customers are indecisive to shop for them. In this paper, with the help of vague set theory, we describe that item's hesitation information is precious knowledge for the design of profitable selling strategies. This work proposed Genetic Algorithm based on evolution principles that has found its strong base in mining or maximize the rules for the items that customers mostly hesitate to purchase or has a high percentage of hesitation because of some reasons like price of an item, quality of an item, etc. Fitness function, crossover, and mutation are the main parameters involved in Genetic Algorithm which we used in our work. This work describes that if the reason of giving up the items is identified and resolved, we can easily remove this hesitation status of a customer and considering newly evolved rules as the interesting ones for boosting the sales of the item.

Keywords: *Data mining, Association rule mining, Vague Sets, Hesitation Information, Genetic Algorithm*

1. Introduction

Data mining is a process of extracting interesting patterns and valuable knowledge and information from huge volume of data. There are several techniques that have been used to discover such kind of specific and richer information, most of them resulting from statistics and machine learning [1]. The jobs performed in the data mining depend on what kind of precious information someone needs to mine. Association Rule Mining is a characteristic of data mining which is used to observe relationships between things that are mostly buying together, [2]. Association rule mining determines interesting valid rules among a large data set of items.

This work describes how to mine useful valid rules from hesitation information using genetic algorithm that capture more richer information than traditional ARs. To describe the information of a customer's interest on items, we use the concepts of median membership and imprecision membership using vague set theory. The hesitation information of an item describes using following two scenarios:

- A customer wants to buy a car but hesitate to buy it because of some reasons like he cannot afford the price of a car. If seller offered him loan for a car then his hesitation percentage decreases, but if seller is ready to provide a loan according to that customer, then his hesitation status removes and comes in to the category of favor and buy it.

- A man needs to purchase both refrigerator and washing machine but he has some limitations and gives preference to refrigerator over washing machine. If seller identified his problem and tell him about good offers like if he purchased both refrigerator and washing machine together, then he can get some discounts, doing this seller can easily removes his hesitation status.

2. Association Rule Mining

Association rule mining is a well-known method to discover interesting rules and relations between various items in large databases. According to the concept of valid rules, Rakesh Agrawal [3] introduced association rules for finding regularities between items in large-scale exchange information recorded. An association rule has two different sections, an antecedent (if) and also a consequent (then). An antecedent is a create item in the knowledge. A consequence is a found item combination with the antecedent [4]. Applications of association rule mining includes market basket analysis, census data, medical diagnosis, protein sequences *etc.*

Association rule generation is generally divided into two step process:-

- To find all frequent item-sets in a database or to form a pattern, a user specified minimum support threshold is used.
- In order to form valid rules in large databases, a user specified minimum confidence threshold is applied to these frequent item-sets.

$$\text{Support}(A \rightarrow B) = \frac{\text{No. of transactions containing both A and B}}{\text{Total no. of transactions}} \quad (1)$$

$$\text{Confidence}(A \rightarrow B) = \frac{\text{No. of transactions containing both A and B}}{\text{No. of transactions containing A}} \quad (2)$$

Numerous business enterprises collect vast amounts of information from their regular operations. For example, large quantities of client purchase information are gathered day by day at the grocery store checkout counters. Association rule mining can be of many forms and various other fields can be easily merged in this.

2.1. Apriori Algorithm

Apriori algorithm was first given by Agrawal in [Agrawal and Srikant 1994]. Apriori is the improvement over association rule mining. It is one of the most commonly used algorithm that generates valid association rules. Traditional Apriori algorithm works on Boolean logic. Because of this reason Apriori algorithm has several drawbacks. It works on only yes (1) and no (0) form. It is inappropriate to handle imprecise and inexact data. In real life domain, there are various vague situations or hesitation information. For example, when we go to mall for shopping, then we put items in our basket, but sometimes at the time of transactions, our mind is changed suddenly and we walkout for some items. This type of situation is not handled by crisp sets.

There are two different measures to find out all huge itemsets from Apriori algorithm. First the candidate itemsets are created then scanned database to check the actual support count of the corresponding itemsets. At the time of first scanning of the database the support count of all item is designed and large 1 –itemsets are created through pruning those itemsets whose supports are below pre-defined threshold as present in Table 1 (a) and (b). In all pass only those candidate itemsets that conclude the similar specified number of items are created and also checked.

The candidate K-itemsets are created after (M-1)th passes over the database through joining frequent M-1 -itemsets. Each the candidate M-itemsets are pruned through check their sub (M-1)-itemsets, if any of its sub (M-1)-itemsets is not in the list of frequent (M-1)-itemsets, this M-itemsets candidate is pruned out because it has no hope to be frequent according Apriori property. The Apriori property says that each sub (M-1)-itemsets of the frequent itemsets must be frequent. Let us consider the candidate 3-itemsets generation as an example. First each candidate itemsets are created through joining frequent 2-itemsets, which include (I1, I2, I5), (I1, I2, I3), (I2, I3, I5), (I1, I3, I5). Such itemsets are then checked for their sub itemsets, since (I3, I5) is not frequent 2-itemsets, the two 3-itemsets are eliminated from the candidate 3-itemsets list as present in Table 1(e). Each those processes are performed iteratively to find each frequent itemsets until the candidates itemsets or the frequent itemsets become empty.

Table 1. Apriori Mining Process

Items	Count No.
I1	7
I2	8
I3	6
I4	2
I5	3
I6	1

(a) C₁

Large 1 Items
I1
I2
I3
I5

(b) L₁

Items	Count No.
I1,I2	5
I1, I3	4
I1, I5	3
I2, I3	4
I2, I5	3
I3, I5	1

(c) C₂

Large 2 Items
I1, I2
I1, I5
I2, I5
I2, I3
I1, I3

(d) L₂

Items	Count No.
I1,I2,I5	3
I1,I2,I3	2

(e) C₃

2.2 Fuzzy Association Rule Mining

Fuzzy association rule mining uses fuzzy logic or fuzzy concept to generate interesting affiliation rules which matches on inexact and obscure statistics. These affiliation rules can help in selection making for diverse hesitation facts. Fuzzy policies are a variation of classical association rule mining.

Fuzzy good judgment is brought to address obscure and inexact information, due to the fact inside the actual, there may be a vague records or information about specific applications which include in sensor databases. Fuzzy set uses point-value selected from the unit c programming language [0,1] which is the restrictions inside the sets to tackle with a few ordinary vague records. From time to time, there's an even more vague situations got here inside the actual life area wherein fuzzy regulations are also inappropriate to examine extra outcomes. Thus, there may be a wide scope of improving principles of affiliation rule mining within the future paintings. As information mining is the maximum important idea in real lifestyles applications like in businesses, stores, *etc.*, there scope has been extensively increased.

3. Vague Sets

In the real world there are vaguely specified data values in many applications, such as sensor information. Fuzzy set principle has been proposed to address such

vagueness with the aid of generalizing the notion of membership in a set. Essentially, in a Fuzzy Set (FS) every detail of an element is associated with a point-value selected from the unit interval [0,1], which is called the grade of membership in the set. A vague Set (VS), as well as an Intuitionistic Fuzzy Set (IFS), is a in addition generalization of an FS. Rather than using point-based membership as in FSs, interval-based membership is used in a vague sets (VS). The interval-based membership value in VSs is greater expressive in capturing vagueness of records than point – based membership values. Fuzzy set idea has lengthly been introduced to deal with inexact and vague facts by using Zadeh’s seminal paper in [5], seeing that within the actual international there’s vague records approximately distinctive applications, together with in sensor databases, we will formalize the measurements from extraordinary sensors to a vague set. In fuzzy set concept, every item $u \in U$ is assigned a single actual price, known as the grade of membership, among zero and one. (here U is a classical set of items, called the universe of discourse.). In [6], Gau *et al.* factor out that the disadvantage of using point-based value in fuzzy set concept is that the proof for $u \in U$ and the proof against $u \in U$ are in reality combined together. With the intention to tackle this problem, Gau *et al.* propose the perception of vague sets (VSs), which allow the use of interval-based membership instead of the use of point-based values as in FSs. The interval-based club generalization in VSs is extra expressive in capturing vagueness of facts. For that reason, the thrilling functions for handling indistinct information which are unique to VSs are largely left out.

Let I be a classical set of items, called the universe of discourse, in which detail of I is denoted by means of x .

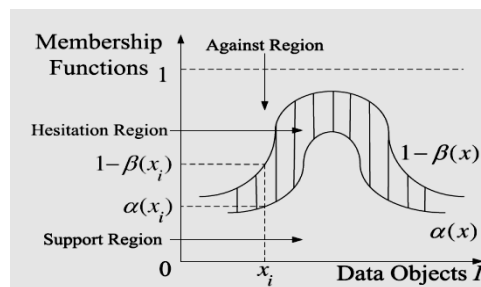


Figure 1. The True (α) and False (β) Membership Functions of a Vague Set

Gau’s and Buehrer added the perception of indistinct sets. A indistinct set V in a universe of discourse I is characterized by way of a true membership function, αV , and a false club function, βV , as follows:

$$\alpha V : I \rightarrow [0,1], \beta V : I \rightarrow [0,1], \text{ where } \alpha V(x) + \beta V(x) \leq 1,$$

$\alpha V(x)$ is a lower bound on the grade of membership of x derived from the evidence for x , and $\beta V(x)$ is a lower bound on the grade of membership of the negation of x derived from the evidence against x . Suppose $I = \{x_1, x_2, x_n\}$. The grade of membership of x is bounded to $[\alpha V(x); 1 - \beta V(x)]$, which is a sub interval of $[0,1]$ as depicted in Figure 1. For brevity, we omit the subscript V from αV and βV . We say that to $[\alpha V(x); 1 - \beta V(x)]/x$ is a vague element and the interval to $[\alpha V(x); 1 - \beta V(x)]$ is the vague value of the object x .

In this section, we discuss the following relationships of vague membership values (vague values for short) in VSs: crisp and imprecision. Remarkably, there are no such meaningful relationships based on IFS membership values. We also show two lattices arising from the crisp and the imprecision orders. In order to compare vague values, we need to introduce two derived memberships for discussion.

- The first is called the median membership, $Mm = (\alpha(x) + (1 - \beta(x)))/2$, which represents the overall evidence contained in a vague value. In addition, the vague value [1,1] has the highest Mm, which means the corresponding object totally belongs to the VS (*i.e.* a crisp value). While the vague value [0,0] has the lowest Mm which means that the corresponding object totally does not belong to the VS (*i.e.* the empty vague value).
- The second is called the imprecision membership, $Mi = ((1 - \beta(x)) - \alpha(x))$, which represents the overall imprecision of a vague value and is shown. In addition, the vague value [a, a], ($a \in [0, 1]$) has the lowest Mi which means that we know exactly the membership of the corresponding object (*i.e.* a fuzzy value). While the vague value [0,1] has the highest Mi which means that we know nothing about the membership of corresponding object.

Definition (Intent, Attractiveness and Hesitation, AH-Pair Transactions): The intent of an item x , denoted as $intent(x)$, is a vague value $[\alpha(x), 1 - \beta(x)]$. The attractiveness of x , denoted as $MA(x)$, is defined as the median membership of x , *i.e.*, $MA(x) = (\alpha(x) + (1 - \beta(x)))/2$. The hesitation of x , denoted as $MH(x)$, is defined as the imprecision membership of x , *i.e.*, $MH(x) = ((1 - \beta(x)) - \alpha(x))$. The pair $\langle MA(x), MH(x) \rangle$ is called the AH-pair of x .

An object with high attractiveness approach that the item is nicely bought and has an excessive opportunity to be bought again next time. An item with high hesitation way that clients are constantly hesitating to shop for the item due to a few purpose (*e.g.*, the client is anticipating rate reduction).

4. Genetic Algorithm

Genetic Algorithms are adaptive heuristic seek set of rules premised at the evolutionary ideas of natural selection and genetic. The fundamental concept of genetic algorithm is designed to simulate procedures in natural device vital for evolution, specifically people who follow the standards first laid down by using Charles Darwin of survival of the fittest. As such they constitute an sensible exploitation of a random seek within a described seek space to clear up a trouble. As such they represent an intelligent exploitation of a random search within a defined search space to solve a problem.

Genetic algorithm are one of the pleasant ways to solve a trouble for which little is understood. They are a totally widespread algorithm and so will work nicely in any search area. The Genetic algorithm [7] become developed by John Holland in 1970. GA is stochastic search set of rules modeled at the procedure of herbal choice, which underlines organic evolution. GA has been effectively implemented in lots of seek, optimization, and device mastering troubles.

GA works in an iterative way by way of generating new populations of strings from vintage ones. Every string is the encoded binary, actual and so on, model of a candidate solution. An evaluation function associates a health measure to each string indicating its fitness for the hassle [8].

Popular GA observes genetic operators such selection, crossover and mutation on to begin with random populace in order to compute a whole generation of new strings. GA runs to generate solutions for successive generations. The possibility of an individual reproducing is proportional to the goodness of the answer it represents. For this reason the first-class of the solutions in successive generations improves. The technique is terminated whilst a suitable or optimum answer is observed. GA is suitable for issues which require optimization, with appreciate to a few computable criterions.

4.1. Some Functions of Genetic Operators

- (a) Selection: Selection deals with the probabilistic survival of the fittest, in that, greater in shape chromosomes are selected to continue to exist. Where in health is a comparable measure of how properly a chromosome solves the problem handy.
- (b) Crossover: This operation is accomplished with the aid of deciding on a random gene alongside the length of the chromosomes and swapping all the genes after that factor.
- (c) Mutation: Alters the brand new solutions so as to add inside the look for higher solutions. This is the risk that a chunk within a chromosome might be flipped (zero becomes 1, 1 becomes zero).

4.2. Pseudo-Code of Genetic Algorithm

Simple genetic algorithm pseudo-code:

- A. [Start] Generate random population of n chromosomes
- B. [Fitness] Evaluate the fitness $f(x)$ of each chromosome x in the population
- C. [New population] Create a new population by repeating the following steps until the new population is complete.
 - i. [Selection] Select two parent chromosomes from a population according to their fitness.
 - ii. [Crossover] with a crossover the parents to form a new offspring (children).
 - iii. [Mutation] Mutate new offspring at each locus (position in chromosome).
- D. [Replace] Use new generated population for a further run of algorithm
- E. [Test] If the end condition is satisfied, stop, and return the best solution in current population
- F. [Loop] Go to step B.

In this work for all generations, the fitness of each individual in the population is evaluated, multiple individuals are selected from the present population (based on their fitness), and modified (recombined and possibly mutated) to form a new population. Fitness function was defined as follows:

$$\text{Fitness value} = 1/(\alpha_1 * r_1 + \alpha_2 * r_2 + \alpha_3 * r_3 + \alpha_4 * r_4 + \alpha_5 * r_5 + \alpha_6 * r_6)$$

Where r_n represents the value of the rules present in the database and coefficients α control effect of each parameter inside fitness function. Here, the values of α are fixed. For e.g. α_1 is 0.15. The selection of the fittest individual is done with the help of various methods.

5. Related Work

A Lu, Yiping Ke, James Cheng, and Wilfred Ng [9] implemented vague set concept inside the context of AR mining as to encompass the hesitation facts into the ARs. Describe the concepts of attractiveness and hesitation of an item, which characterize the overall information of a customer's intent on an item. Depending on these two concepts, proposed the notion of Vague Association Rules (VARs) and designed an effective algorithm to mine the VARs. Experiments demonstrate that the algorithm was effective and the VARs capture more exact and better information in turn to conventional ARs.

Bernecker *et al.* [10], and Yiu *et al.* [11] tested, utilizing of predictable support may additionally perhaps reason the lack of a few important styles. For computing the probability that a pattern is frequent, and presented the notion of PFI. Dynamic programming primarily based solutions become introduced to regain PFIs from

attribute indistinct databases. However their set of rules compute chances, and validate that an item set is a PFI in $O(n^2)$ time. The set of rules evades the utilizing of dynamic programming, and are capable to validate a PFI tons faster.

An Lu *et al.* [12-13] aids in recognition of what is improved within Fuzzy Sets, Intuitionistic Fuzzy Sets and Vague Sets also providing notion of vague association rules (VARs) [14] by utilizing two more measures: attractiveness and hesitation of a data item permits interval-based membership to confine more evidence to an object of the universe. Vague set theory has been applied to education field for mining association rules [15].

Anjana Pandey and K. R. Pardasani [16] presented a vague association rule to make available hesitation information and expand an algorithm to mine the hesitation information. The algorithm was devised to mine the courses and the hesitation of students to attend the courses. Experiments on real datasets confirmed that the algorithm to mine the Vague Association Rule is effective. In contrast to the conventional Association Rule mined from transactional databases, the Vague Association Rule mined from the AH-pair databases are more detailed and capable to capture better information.

Bala Yesu Chilakalapudi, Narayana Satyala and Satyanarayana Menda [17] presented an algorithm for a resolving the issue problem of extracting frequent item sets from a huge vague database, interpreted under the Possible World Semantics (PWS). This issue is strictly difficult since a vague database consists of an exponential number of possible worlds. By examining the mining process can be modeled as a Poisson binomial distribution, an algorithm is implemented which can effectively and exactly determine frequent item sets in a huge vague database. The devised mining algorithm facilitates Probabilistic Frequent Item set (PFI) outcomes to be re-energized. The devised algorithm can maintain incremental mining and provides the precise outcomes on mining the vague database. The broad estimation on real data set to certify the scheme is performed.

6. Proposed Work with Achieved Result

Traditional Association rule (AR) mining has limitations that it only takes the items that are sold but ignores the items or products that are almost sold. In this work, we have proposed genetic algorithm for mining hesitation information of those items that customers are mostly hesitating to buy, by using vague set theory. If the reason of giving up the items is identified and resolved, we can easily remove this hesitation status of a customer and considering new evolved rules as the interesting ones for the design of profitable selling strategies.

The whole proposed algorithm is explained with the help of a flow chart which can be shown below:

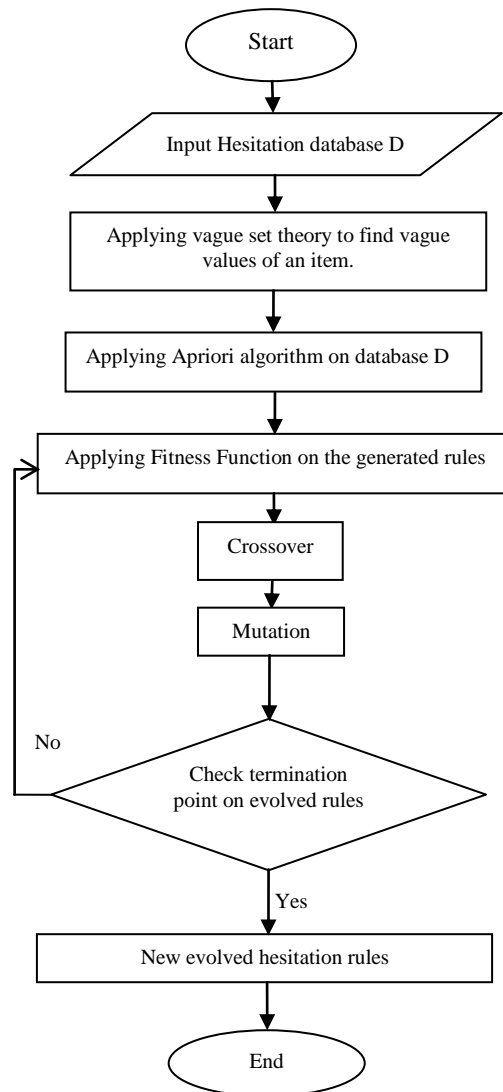


Figure 2. Flowchart of Proposed Algorithm

This flowchart describes that first we take hesitation database as input , then applying vague set theory to find vague values of an items which shows percentage of item’s hesitation. Then applying genetic algorithm on rules generated from Apriori algorithm for generating profitable hesitation rules. We are using two algorithms, Cal_intent () and Cal_AH-pair (intent) in our proposed algorithm which is described as:

Algorithm 1: Cal_intent ()

1. Initialize the intent array to store the intent.
2. For each $i = 0, 1, 2, 3, \dots$ where $i < \text{no. of iid}$, do;
3. Initialize favor(α) and against(β) variable with value zero;
4. For each $j = 0, 1, 2, 3, \dots$ where $j < \text{no. of tid}$, do;
5. Increment favor(α) by one when $D[i][j]$ is equal to one;
6. Increment against(β) by one when $D[i][j]$ is equal to zero;
7. End;
8. Generate intent as $[\alpha, 1-\beta]$;
9. End;
10. Return all intent;

This algorithm is iterative procedure to calculate intent which is first module of framework. This algorithm takes a dataset (D) as input. This dataset consists of rows and column as transaction ID (tid) and item ID (iid) of a supermarket. This algorithm returns an intent array.

The second module of framework is to calculate the AH-pair. The Cal-AH-pair Algorithm is a simple iterative method that calculates AH pair. This algorithm takes intent as input which is calculated in previous algorithm.

Algorithm 2:Cal_AH-pair (intent)

1. Initialize an array AH pair;
2. For each $i=0,1,2,3,\dots$ where $i < \text{no. of iid}$;
3. Attractiveness as a median membership *i.e.* $\frac{1}{2}(\alpha+(1-\beta))$;
4. Hesitation as a difference of α and $1-\beta$ using intent;
5. End;
6. Return all AH-pair

We analyzed that hesitation for an item decrease the attractiveness of an item and hence reduce the probability of selling the item. This reduces the profitability of store in many folds. Here we analyzed the case of supermarket and find the main cause of hesitation in customer's point of view. This algorithm is efficient in increasing selling probability of an item in case of supermarket scenario.

Proposed Algorithm

Input: Hesitation database D, min support, min confidence.

Output: Profitable valid rules.

1. First initiate the hesitation database D that contains the values such as yes, no, and hesitation of selling of products.
2. Applying vague set theory to find vague values and AH-pair values from algorithm Cal_intent () and Cal_AH-pair (intent) respectively, which indicates hesitation percentage of each items.
3. Then applying Apriori algorithm on hesitation database D to generate valid rules, before it taking the value of yes, no, and hesitation in the form of 0 and 1 *i.e.* called binary conversion; $Y=1$, $H \cong 0$, and $N=0$.
4. Now applying genetic algorithm on generated rules.
 - (a) First, find fitness function of each generated rules by the formula:-
$$\text{FitnV} = 1 / (\alpha_1 * r_1 + \alpha_2 * r_2 + \alpha_3 * r_3 + \alpha_4 * r_4 + \alpha_5 * r_5 + \alpha_6 * r_6)$$
 - (b) Selection is done according to fitness values of rules.
 - (c) Single point crossover is applied.
 - (d) Single bit mutation is applied.
 - (e) New evolved hesitation rules.
 - (f) Repeat steps from (a) to (e) until termination condition is reached ($\text{FitnV} < 0.9527$).
5. End.

Steps in the Process Example:

Here we are explaining proposed work with the help of an example.

Step1 – Input Hesitation database D .Given below database is one of the sample databases that we used for our work.

Table 2. Sample Database D

ID	A	B	C	D	E	F
1	Y	Y	Y	H	N	N
2	Y	Y	Y	Y	Y	Y
3	N	N	N	Y	H	N
4	H	N	Y	N	N	N
5	Y	Y	Y	H	H	N
6	Y	Y	Y	Y	Y	H
7	Y	Y	Y	Y	Y	H
8	Y	Y	Y	Y	H	H
9	Y	Y	Y	H	H	H
10	Y	N	H	Y	N	N
11	N	N	N	N	Y	H
12	Y	Y	Y	N	N	Y
13	Y	Y	Y	H	H	Y
14	H	H	Y	N	H	N
15	Y	Y	Y	H	H	N
16	Y	N	N	N	N	H
17	Y	N	H	H	N	N
18	Y	Y	Y	Y	N	N
19	Y	Y	Y	Y	H	N
20	N	N	N	Y	N	H
21	Y	Y	Y	H	Y	Y
22	Y	Y	Y	Y	Y	H
23	Y	Y	H	H	N	N
24	Y	Y	Y	N	N	Y
25	Y	Y	Y	Y	Y	Y
26	Y	Y	Y	H	H	H
27	Y	Y	Y	Y	Y	H
28	Y	Y	Y	Y	Y	Y
29	N	N	N	N	Y	H
30	Y	Y	Y	Y	H	N

In above table, Y=yes (favor), N= no (against), H= hesitation, according to vague set theory which is capable of dealing with vague situations.

And A, B, C, D, E, F denotes the item sets where, A = Utensils, B = Furniture, C = Refrigerator, D= Washing Machine, E= Water Purifier and F= Car.

And ID denotes the number of records we are taken into account. This work is capable of handling large volumes of data.

Step 2 – Calculate vague value or intent value of each items.

A	B	C	D	E	F
[0.800000, 0.866667]	[0.700000, 0.733333]	[0.73333, 0.833333]	[0.466667, 0.766667]	[0.333333, 0.666667]	[0.233333, 0.600000]

Step 3 – Calculate AH (Attractiveness – Hesitation) value of each items.

A	B	C	D	E	F
[0.833333, 0.066667]	[0.716667, 0.033333]	[0.783333, 0.100000]	[0.616667, 0.300000]	[0.500000, 0.333333]	[0.416667, 0.366667]

The above step (3) shows that item F (car) is the item that has higher percentage of hesitation. It means customers are mostly hesitated to buy it

Step 4 – Then applying Apriori algorithm to generate valid rules by taking Support = 8 and Confidence = 30%, before it taking the value of y , n , and h in the form of 0 and 1 *i.e.* called binary conversion. $Y = 1$, $H \cong 0$, $N = 0$. Following rules are generated:

```
{ Utensils -> Furniture,  
  Utensils -> Refrigerator,  
  Utensils -> Washing Machine,  
  Utensils -> Water Purifier,  
  Furniture -> Refrigerator,  
  Furniture -> Washing Machine  
  Furniture -> Water Purifier,  
  Refrigerator -> Washing Machine,  
  Refrigerator -> Water Purifier,  
  Utensils, Furniture -> Refrigerator,  
  Utensils, Furniture -> Washing Machine,  
  Utensils, Furniture -> Water Purifier,  
  Utensils, Refrigerator -> Washing Machine,  
  Utensils, Refrigerator -> Water Purifier,  
  Furniture, Refrigerator -> Washing Machine,  
  Furniture, Refrigerator -> Water Purifier,  
  Utensils, Furniture, Refrigerator -> Washing Machine,  
  Utensils, Furniture, Refrigerator -> Water Purifier }
```

Step 5 – By Applying Genetic algorithm on rules generated in step (4), we get following new rules which we considered as hesitation rules –

```
{ Utensils, Refrigerator -> Car  
  Utensils, Water Purifier -> Car  
  Furniture, Refrigerator -> Car  
  Furniture, Washing Machine -> Car  
  Refrigerator, Washing Machine -> Car  
  Refrigerator, Water Purifier -> Car  
  Utensils, Furniture, Refrigerator -> Car  
  Utensils, Furniture, Water Purifier -> Car  
  Utensils, Refrigerator, Washing Machine -> Car  
  Furniture, Refrigerator, Washing Machine -> Car  
  Furniture, Refrigerator, Water Purifier -> Car  
  Utensils, Furniture, Refrigerator, Washing Machine-> Car }
```

These new hesitation rules show that if the reason of giving up the items is identified and resolved, then we also considered these new evolved rules derived from genetic algorithm as the interesting ones for boosting the sales of the items.

7. Conclusion

In this paper, we have presented a novel genetic based algorithm to mine hesitation association rules in big data sets. We apply the vague set based approach to model the hesitation information of the items. This work giving more valid rules *i.e.* converting the hesitation of higher order into attractiveness and thus giving user an idea about the possibility of the rules to be generated. This concept helps in exposing necessary hesitation rules that may also be considered for profitable decision making process.

References

- [1] J. Han and M. Kamber, "Data Mining: Concepts and Techniques", 2nd ed., the Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor, (2006).
- [2] Frawley W. J., P. Shapiro, G. Matheus and Christopher J., "Knowledge Discovery in Databases", an Overview. AAAI/MIT Press, (1992).
- [3] R. Agrawal, S. Tomasz, I. BalaIyer and A. Swami, "An Interval Classifier for Database Mining Applications", VLDB-92, Vancouver, British Columbia, (1992), pp. 560-573.
- [4] Agrawal R., Imielinski T. and Swami A. N., 1993. "Mining association rules between sets of items in large databases", In Proceedings of the ACM SIGMOD International Conference on Management of Data, (1993), pp. 207-216.
- [5] L. A. Zadeh, "Fuzzy sets inf. Control", vol. 8, (1965), pp. 338-353.
- [6] W. L. Gau and D. J. Buehrer, "Vague sets", IEEE Transactions on Systems, Man and Cybernetics, vol. 23, (1993), pp. 610-614.
- [7] Pei M., Goodman E. D. and Punch F., "Feature Extraction using genetic algorithm", Case Center for Computer-Aided Engineering and Manufacturing W. Department of Computer Science, (2000).
- [8] Goldberg D. E. "Genetic Algorithms in Search optimization and Machine Learning", Addison Wesley, (1989), pp. 41.
- [9] A. Lu, Y. Ke, J. Cheng and W. Ng, "Mining Vague Association Rules", Department of Computer Science and Engineering The Hong Kong University of Science and Technology Hong Kong, China.
- [10] Q. Zhang, F. Li and K. Yi, "Finding Frequent Items in Probabilistic Data," Proc. ACM SIGMOD International Conf. Management of Data, (2008).
- [11] A. Lu and W. Ng, "Vague Sets or Intuitionistic Fuzzy Sets for Handling Vague Data- Which One Is Better?" © Springer, (2005).
- [12] "ACM Usage of Fuzzy, Rough, and Soft Set Approach in Association Rule Mining", (2012).
- [13] A. Lu, Y. Ke, J. Cheng and W. Ng, "Mining Vague Association Rules", © Springer, (2007).
- [14] International Journal of Computer Applications, "A Model for Mining Course Information using Vague Association Rule" Anjana Pandey, UIT RGPV Bhopal, K. R. Pardasani MANIT Bhopal, vol. 58, no. 20, November (2012), pp. 0975 – 8887.
- [15] A. Pandey and K. R. Pardasani, "A Model for Mining Course Information using Vague Association Rule", International Journal of Computer Applications, vol. 58, no. 20, November (2012), pp. 0975–8887.
- [16] A. Pandey and K. R. Pardasani, "A Model for Vague Association Rule Mining in Temporal Databases", Journal of Information and Computing Science, ISSN 1746-7659, England, UK, vol. 8, no. 1, (2013), pp. 063-074.
- [17] B. Y. Chilakalapudi, N. Satyala and S. Menda, "An Improved Algorithm for Efficient Mining of Frequent Item Sets on Large Uncertain Databases", International Journal of Computer Applications, vol. 73, no.12, July (2013).

Authors



Prateek Shrivastava, is currently pursuing the M. tech. degree in computer science and engineering from CSE/IT department, MITS Gwalior (M.P), India. He has received B.E. degree from Institute of Information Technology and Management, Gwalior (M.P), India. His areas of interest are data mining, Association rule mining, frequent pattern mining, vague set and their applications.



Akhilesh Tiwari, has received the Ph.D. degree in Information Technology from Rajiv Gandhi Technological University, Bhopal, M.P. (India). He is currently working as Associate Professor in the Department of CSE & IT, Madhav Institute of Technology & Science (MITS), Gwalior, India. He has guided several students at Master and Under Graduate level. His areas of current research include Knowledge Discovery in Databases and Data Mining, Wireless Networks. He has published more than 20 research papers in the journals and conferences of international repute. He is also working as a reviewer & member in the editorial board of various international journals. He is having the memberships of various Academic/ Scientific societies including IETE, CSI, GAMS, IACSIT and IAENG.

