Elephant Herding Optimization based Vague Association Rule Mining Algorithm

Sowkarthika B.
Department of CSE &IT,
Madhav Institute of Technology and Sciene,Gwalior,M.P,India.

Akhilesh Tiwari
Associate Professor,
Department of CSE &IT,
Madhav Institute of Technology and Sciene,Gwalior,M.P,India.

Uday Pratap Singh
Assistant Professor,
Department of Applied Science,
Madhav Institute of Technology and Sciene,Gwalior,M.P,India.

ABSTRACT
Huge amount of data is being gathered, processed and analyzed in every sector to derive useful information. So, automated tool like data mining has evolved in order to extract information and solve the overhead in manual approach. Association rule mining which is an essential part of data mining fails to address the vague and uncertain situations. In shopping applications, traditional approach estimates rules containing frequently bought items. In real world, data being mined might be vague. The items that are 'almost bought' or considered by customers are also helpful in planning efficient strategy. Identifying the hesitation in buying such items improves the profit drastically. Also, the traditionally used majors are not sufficient in addressing the profitability concern. There is a need to incorporate appropriate parameter changes in the currently used measures to deal with profitability. Additionally, the items being sold on special occasions or on a season are considered interesting only at the time of that occasion or the season respectively and not throughout the year. So, the mining that involves generating patterns that are considered interesting at some point of time or within a time interval are needed. To accomplish the above objectives, vague set theory is used which addresses the uncertain situations, over the temporal database for a particular occasion, followed by vague association rule mining algorithm with measures that combines the statistical data and value-based data for finding association rule that yields maximum profit. Elephant Herding Optimization (EHO) is used in optimizing the obtained resultant rules. The proposed methodology thereby generates optimal profitable seasonal rules that address vague situations and removes hesitation of a product, which is beneficial for any enterprise in the current scenario for effective decision making.

Keywords
Association Rule Mining, Vague theory, Profit Mining, Elephant Herding Optimization, Temporal rules.

1. INTRODUCTION
In this digital world, data is being gathered from various sectors and analyzed to extract necessary details, which helps in obtaining some useful information or evaluating outcome. The gathered data might be large which leads to the need for evolution of an automated tool like Data Mining for mining the data to get the necessary information. Association Rule (AR) mining is the main task associated with data mining which helps in analyzing customer's behavior for a scenario like market-basket analysis.

Here, the domain under discussion is the market-basket analysis where database of customer transactions is considered and each transaction involves buying set of items and relationship between the set of items frequently purchased together have to be estimated.

For set of items $I=\{I_1,I_2,\ldots.I_n\}$ and set of transactions from the transactional database $D=\{T_1,T_2,\ldots.,T_n\}$, the association rule[1] takes the form $A \Rightarrow B$ where A, B are the set of items i.e., $A,B \subseteq I$ and $A \cap B = \phi$. The rule 'Computer ⇒AntiVirus_Software' implies that whenever Computer is purchased, AntiVirus_Software is more likely to be purchased. The set of items on the left hand side of the rule is antecedent while set of items on the right hand side is consequent. Two factors play an important role in measuring a rule's interestingness. They are support and confidence. Support gives an estimate of how frequently two sets of items are purchased together from the given transactions in database D.

$$Support(A \to B) = \frac{n(AUB)}{|D|} = P(AUB)$$

Confidence gives an estimate of finding the possibility of the purchase of the items in consequent given the purchase of set of items in the antecedent.

$$Confidence(A \to B) = \frac{Support(AUB)}{Support(A)} = P(\frac{B}{A})$$

The conventional AR mining always considers the items bought by customers. But in case of shopping markets, some items might have been considered by customers, but they falter to buy them. There can be different reasons in not buying the products like doubt in quality, cost of the product etc,. Once their hesitation is identified and tended to, the profit increases. Every time a user logs in, he is given a unique transactional id. For every item he browses, the information stored in the weblog. This information is later extracted and used for removing the hesitation so that sales increases boosting the profit. The following scenarios illustrate the hesitation status of customers. The hesitation of a product might arise at one of the possible stages.

i. Customer views an item and not includes it in the kart. ii. Customer hesitation status for item browsed in detail. iii. Hesitation on specifying the shipping address. iv. Hesitation on viewing the delivery details. v. Hesitation after entering the coupon code for discount. vi. Customer saves the item for later and not buys it. vii. Customer adds the item to kart but fails to check out. viii. Customer cancels the product after checking the delivery date. This kind of vague data needs some specialized algorithm to be handled in an effective way. To represent and analyze uncertainty and vagueness, specialized algorithm on vague set theory is utilized. AR mining on vague set theory is done and the resultant rules obtained help in planning efficient organizational strategies.

Additionally, there are items that are purchased for certain special events and not considered through the rest of the year. For example, items like 'crackers' are bought only during Diwali and 'colors' only during Holi. Also, few shopping destinations may have sales only on a particular time of the year. The algorithms proposed so far failed to generate rules that might interest the customer and organization at some point of time or within a time interval. The proposed algorithm makes sure that rules generated are interesting at any point of time.

The basic objective of any business organization is to have increased profit. For a given set of transactions T, rules are considered interesting if that satisfy minimum support and confidence constraints. Few other interestingness measures include correlation, lift and spread. Though AR mining has taken the frequent item set into consideration, it fails to address the rules that yields more profit [2]. In spite of the fact that few items won't not be purchased frequently, their benefit could be quite high. For example, selling a laptop might yield a profit of Rs.1000 while a packet of sugar might yield a profit of Rs.10. On a comparison scale, sugar must be bought 100 times so that it gives same profit as one laptop. So, necessary parameter settings have to be done in order to get maximum profit.

Although enough thresholds are applied over the rule, the number of rules obtained is still huge. Optimization algorithms are employed so that the resultant rules are profiting and meets customer's interest. The optimization algorithms used here is EHO (Elephant Herding Optimization).

Section 2 briefly reviews the vague set theory and EHO. Section 3 sheds some light on the issues and challenges in the existing approach. Section 4 abbreviates the proposed approach. Section 5 gives detailed description of the proposed approach. Section 6 gives the working example of the proposed approach in obtaining the optimal rules.

2. LITERATURE SURVEY

Few theories have been proposed to address the uncertain situations. Crisp set fails to represent approximate values. Fuzzy fails because of its usage of single membership function whereas an interval based membership function is necessary for representing the vague scenario and also, the need for precise knowledge on every member in fuzzy is not possible in real life situations [2]-[4]. Rough theory suffers from redundancy problem and it is mainly useful for classification [5]-[8]. These led to the evolution of vague theory. In order to optimize the resultant rules, EHO is proposed which has been proven efficient in more than 15 benchmark functions when compared to other nature inspired algorithms and genetic algorithm.

2.1. Vague Set Theory

To deal with uncertain data, few algorithms have been proposed on vague theory [9]. A. Lu et al [10]-[11] proposed an algorithm to mine the vague data through vague association rule. In a market basket analysis, hesitation at different stages of buying a product is analyzed. Then vague theory is applied and rules are generated by Vague Association Rule mining (VAR). The main advantage of applying vague theory is that it helps in eliminating the cost of decision table unlike rough theory, and it has the continuous membership function unlike fuzzy. Apriori is applied in generating frequent itemset.

The true membership function α(x) of an item is considered as the item being bought and the false membership function β(x) is the item not being bought. If Σ h(x) gives the overall hesitation of an item, then α(x) + β(x) + Σ h(x) = 1. The support region is the region where the customer buys a product and most likely to buy it next time.

The interested region is the hesitation region which lies in the interval [α(x) , 1-β(x)]. This region contains information of a product that is not sold and which could be possibly turned into sold product.

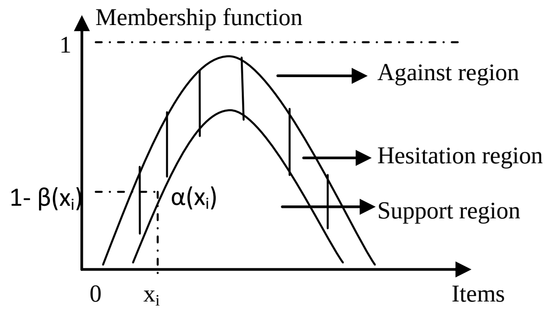

Figure 1. The true (α) and False (β) Membership Functions of a Vague set

Attraction and Hesitation play a key role in designing the rule according to the interest of an organization. The attraction AT(x) which is the median membership of 'x' and Hesitation of 'x', H(x) which is the difference of upper and lower bound are given by

$$AT(x) = \{ \alpha(x) + 1 - \beta(x) \}/2 \;, AT(x) \in [0,1] \quad (1)$$

$$H(x) = \{1 - \beta(x) \} - \alpha(x) \quad (2)$$

A Vague Association Rule X⇒Y is generated based on the AH (Attraction-Hesitation) pair values.

Anjana Pandey et al [12] proposed mining the course information based on the vague association rule so that hesitated courses are identified and necessary parameters are changed to make it interesting. The topics containing different courses are taken and their attendance are viewed which gives the hesitation status of a student in attending the topic. VAR is applied over this data. The rules generated give the idea of making a hesitated course an interesting one.

Vivek Badhe et al [13] proposed profit pattern mining on vague data. The traditional methods have used support and confidence as their measurers. Here, true support and true confidence are calculated. This helps in generating rules of certain profit by eliminating the vagueness.

A. Pandey al [14] proposed the VAR for making courses effective. From time to time, the interestingness in the course varies. So a temporal database is taken and VAR is applied over that to mine the hesitation status at that instant.

Reeti Trikha et al [15] proposed the theory of including profit parameters rather than using only support and confidence in AR mining. The Q-factor and Profit Weighing factors have been included which improves the interest of any business organization since the mined rules gives improved profit.

Vivek Badhe et al [16] proposed genetic algorithm which helps in optimizing the obtained rules from AR mining. The fitness function is calculated in such a way that it helps in improving the profit. The rules with maximum profit are allowed to survive after optimization which gives the profitable rules that interest any enterprise.

Prateek Shrivastava et al [17] proposed an algorithm that not only considers the statistical significance but also value based significant parameters like profit. Genetic algorithm is used in obtaining the profitable patterns.

Though many methodologies have been proposed on profit pattern mining, profit is the only concern in all the cases. The uncertainty part is not taken much into consideration on optimizing the rules. The proposed approach takes all the factors like profit, seasonal effect and uncertainty into consideration while generating necessary optimal rules.

2.2. Elephant Herding Optimization

EHO [18] is the meta-heuristic optimization algorithm based on the characteristics of social behavior of elephants. This has two operations. One is clan updating operation and the other is separation operation.

2.2.1 Clan Updation:

Elephants from different groups live together under the leadership of matriarch. The matriarch is the best fit elephant in the clan. Each elephant positions are updated based on its position and the position of matriarch. The position of matriarch is updated based on the positions of the individuals in the clan.

2.2.2 Clan Separation:

Male elephants leave their group when they grow up. The worst elephant (elephant with worst fitness) is removed and it does random search in the search space for joining the appropriate group.

2.2.3 Assumptions Made:

Following are the assumptions made in EHO.

- The population is divided into clans.
- Each clan has fixed number of elephants.
- Fixed number of male elephants leaves the group and lives far away.
- Elephants in each group live under the leadership of matriarch.

2.2.4 EHO Algorithm:

Initialize population.

Repeat

- Sort all elephants according to their fitness.
- Implement clan updating operation. For each elephant j in the clan C_i, the position is updated as

$$x_{new,c_{i,j}} = x_{c_{i,j}} + \alpha * \left(x_{best,ci} - x_{c_{i,j}}\right) * r$$

where $x_{new,c_{i,j}}$=newly updated position, $x_{c_{i,j}}$=old position, $x_{best,ci}$=position with best fit (matriarch) in clan c_i, α ε [0,1], r=random number ε [0,1] . The position update for best fit in the clan is given by $x_{new,c_{i,j}} = \beta * x_{center,ci}$ where $x_{center,ci}=\sum_{j=1}^{n}\frac{x_{ci,j,d}}{n}$ and βε [0,1],$x_{center,ci}$ = center of the clan c_i and n=number of elephants in clan c_i

- Implement separating operation.

$$x_{worst,ci} = x_{min} + (x_{max} - x_{min} + 1) * rand$$

where rand ε [0,1], x_{min}=lower bound position of the elephant, x_{max}=upper bound position

- Evaluate population by newly updated positions.

Until the required iteration.

3. ISSUES AND CHALLENGES

There are various issues prevailing in the existing approach.

- Huge number of rules is obtained in traditional AR mining. But interesting ones are needed to be extracted from this huge rule set. The rule should be considered as interesting in organization as well as customer point of view. Generation of huge rule set leads to overhead in searching and analyzing rules according to the requirement. So there is a need to involve advanced optimization mechanisms.
- The traditional approach does not consider rules that might be interesting at some point of time or within a time interval. So, the rules that might be interesting for a particular season need to be generated. So, a temporal database is needed.
- Traditional AR fails to capture, represent and analyze vague situations to generate rules.
- Support and Confidence are considered to be the key parameters in generating any association rule. Thus statistical significance is found rather than value-based significance. That is, frequently bought items might yield less profit and highly profiting set of items might not be bought frequently and these items are missed out in the rule leading to rare item problem. Necessary parameters have to be added to make the rules yield more profits and contain frequent items.

4. PROPOSED ALGORITHM

Input: Transactional dataset D, min. support, min. confidence and min. profit thresholds, occasion, min. fitness threshold, clan size, probability of randomness P_{rand}, weights α, β, γ.

Output: Optimal Profitable Seasonal rules that removes hesitation.

Support, Confidence and Profit Parameters: For a given AH-pair database D, four support-types and confidence-types for VAR X->Y where Z= $X \cup Y$ are given by

The A-support of Z and A-confidence of Z, where both X and Y are Attractive (being bought)

$$A_{supp}(Z) = \frac{\sum_{T \in D}\prod_{z \in Z} M_A(z)}{|D|} \quad (3)$$

$$A_{conf}(Z) = \frac{A_{supp}(Z)}{A_{supp}(X)} \quad (4)$$

The H-support of Z and H-confidence of Z, where both X and Y are item sets containing items that are being hesitated to buy

$$H_{supp}(Z) = \frac{\sum_{T \in D}\prod_{z \in Z} M_H(z)}{|D|} \quad (5)$$

$$H_{conf}(Z) = \frac{H_{supp}(Z)}{H_{supp}(X)} \quad (6)$$

The AH-support of Z and AH-confidence of Z, where X is bought and Y is being hesitated to buy

$$AH_{supp}(Z) = \frac{\sum_{T \in D}\prod_{x \in X, y \in Y} M_A(x)M_H(y)}{|D|} \quad (7)$$

$$AH_{conf}(Z) = \frac{AH_{supp}(Z)}{A_{supp}(X)} \quad (8)$$

The HA-support of Z and HA-confidence of Z, where Y is bought and X is being hesitated to buy

(9)

$$HA(Z) = \frac{\sum_{T \in D} \prod_{x \in X, y \in Y} M_H(x) M_A(y)}{|D|}$$

$$HA_{conf}(Z) = \frac{HA_{supp}(Z)}{H_{supp}(X)} \quad (10)$$

The profit for any rule $X \rightarrow Y$ is estimated for ‘n’ items

$$Profit(X \rightarrow Y) = \frac{\sum_{x \in XUY} Profit(x)}{\sum_{i=1}^{n} Profit(i)} \quad (11)$$

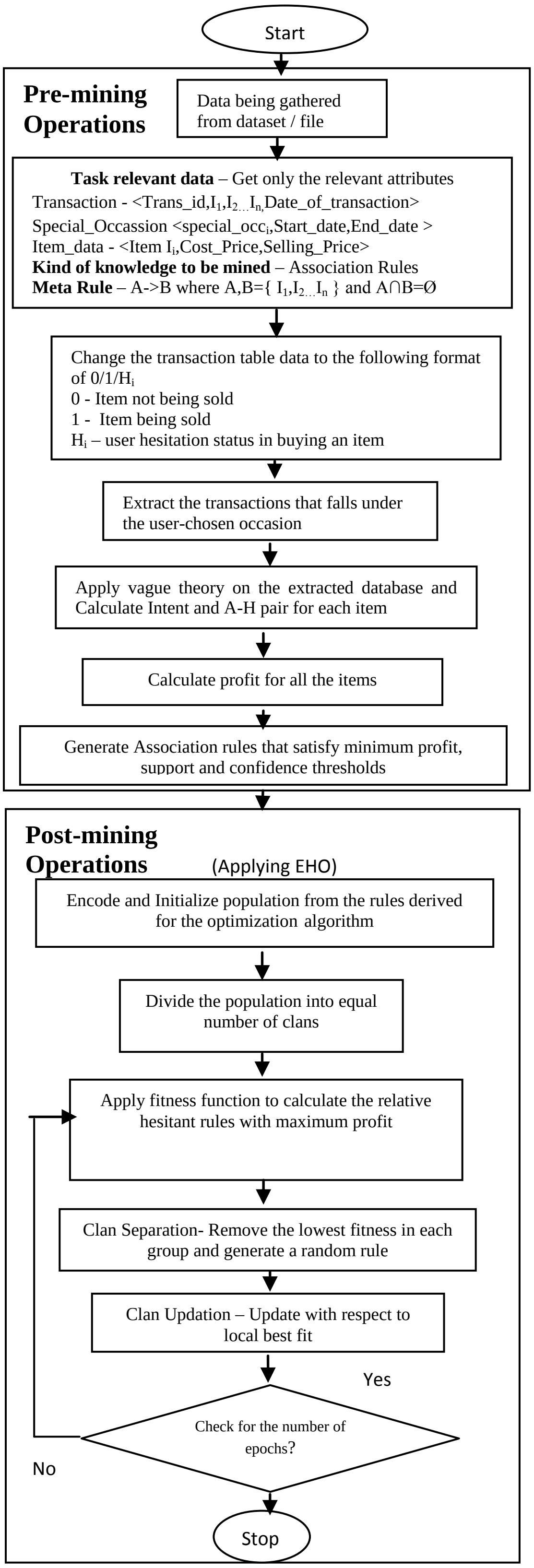

Figure 2. Flow Chart of the Proposed Methodology

Algorithm:

Step 1: Take the transactional dataset D and ignore the irrelevant attributes and extract the transaction database. Item_data is extracted by removing the irrelevant attributes from Item space.

Transaction - <Trans_id,I_1,$I_{2\ldots}$$I_n$,Date_of_transaction>
Special_Occassion-<special_occ$_i$,StartDate,EndDate>
Item_data - <Item I_i,Cost_Price,Selling_Price>

Step 2: Change the transaction table data to the following format of 0/1/H_i

Step 3: Based on the user-specified occasion, extract the database accordingly.

Step 4: Apply vague set theory on the database and calculate intent and A-H pair.

Step 5: Calculate profit from Item_data database.

Step 6: Generate Association rules based on their support, confidence and profit values.

Step 7: Apply EHO on the rules obtained.

- Encode and Initialize the population with the rules generated in step 6.
- Divide population into clans of equal size.(Clan size=odd number)
- Repeat for specified number of iterations.
  Clan separation operator: Remove low fitness individual and choose a random individual.
  Clan updation operator: Update with respect to the best fit individual in each clan.
- Check termination condition.

Step 8: The rules are chosen according to their fitness and their number is based on the user specification of the threshold value. The result gives the profitable optimized rule.

5. DESCRIPTION OF THE ALGORITHM

Step 1: The transactional dataset is taken and the irrelevant attributes are removed and the three databases are formed. The data from customers’ weblog is extracted and gathered for the transactional dataset. Many other hesitation statuses are possible. Here, only 2 hesitation statuses are being considered for the sake of testing the working of the algorithm.

H_1=Customer saves the item for later and not buys it and H_2=Customer adds the item to kart but fails to check out.

Step 2: Based on the transactional dataset, the items are represented as 0/1/H_i in the transaction database along with the date of transaction.

0-Item not being sold, 1-Item being sold, H_i–user hesitation status on buying the item.

Step 3: Based on the special occasion chosen by customer or organization, extract the time period for that occasion from Special_Occasion database and draw the set of transactions that falls under that period from Transaction database.

Step 4: Apply Vague set theory on the extracted transaction set and calculate intent [α(x) 1-β(x)] and AH-pair {(α(x)+1-β(x))/2, 1-β(x)- α(x)} considering every hesitation status of all the items in the database.

Since all hesitation status are independent Chain Group, and also $H_1 \leq H_2$, the following formula is used for estimating the intent for each hesitation status of an item. For an item x, the intent of hesitation status H_i is calculated as $[\alpha_i(x)\ 1-\beta_i(x)]$ and it should be according the following conditions.

- $[\alpha_i(x)\ 1-\beta_i(x)]$ is within the subinterval of $[\alpha(x)\ 1-\beta(x)]$ and $1-\beta_i(x)=H_i(x)+\alpha_i(x)$

- If H_i is in the chain of $H_1 \leq H_2 \leq \ldots \leq H_i \leq \ldots \leq H_n$, then

$$\alpha_i(x) = \frac{\alpha(x)+1-\beta(x)}{2} - \frac{1}{2}\sum_{k=1}^{n} H_k(x) + \sum_{k=1}^{i-1} H_k(x) \quad (12)$$

and $$1-\beta_i(x) = H_i(x) + \alpha_i(x) \quad (13)$$

<u>Step 5:</u> Calculate profit of an item 'x' in the Item_data by using the formula

Profit(x)=Selling_Price(x)-Cost_Price(x) (14)

<u>Step 6:</u> Generate VAR by considering interestingness parameters <support, confidence and profit> by applying equations (3)-(11).

<u>Step 7:</u> Optimize the rules by using nature inspired EHO.

- Encoding

$$Encoding = \begin{cases} \text{'0' if item not present in the rule} \\ \text{'1' if item present in the rule} \\ \text{'2' after antecedent and before consequent} \end{cases} \quad (15)$$

For the rule $A \rightarrow B$, containing I = {A, B, C, D}, the encoding is done as follows. Initially 1100 is obtained where '1' represents the presence of an item in the rule while '0' represents the absence of an item in the rule. Since 'A' is antecedent and 'B' is consequent, a '2' is inserted immediately after that '1' of 'A' and immediately before the '1' of 'B' which leads to the encoding of 122100.

- Clan Separation

When calculating the new rule by clan separation operation, the bits can be changed randomly since multiplication is performed with the random number. If the probability of randomness is given, then the number of bits to be changed can be estimated. If the probability of randomness, P_{rand}=0.25, then one fourth of the encoded bits can be changed randomly (from 0 to 1 and vice versa).

- Clan Updation of an individual

When estimating a rule by clan updation, the following formula is used.

$$x_{new,c_{i,j}} = x_{c_{i,j}}\ OR\ \left(x_{best,ci} \oplus x_{c_{i,j}}\right) \quad (16)$$

1100 ----- $A \rightarrow B$
$\oplus$
0101 ----- $B \rightarrow D$

1001 OR
1100 ----- $A \rightarrow B$

1101 ---- $A, B \rightarrow D$

Figure 3. Update procedure for an Individual

For a rule $A \rightarrow B$, with local best rule in the clan $B \rightarrow D$, the rule $A \rightarrow B$ can be encoded as 1100 and the rule $B \rightarrow D$ can be encoded as 0101 without '2'. $x_{best,ci} \oplus x_{c_{i,j}}$ gives 1001 and $x_{c_{i,j}}\ OR\ \left(x_{best,ci} \oplus x_{c_{i,j}}\right)$, 1100 OR 1001=1101 is the new item-set obtained. Intersection_Of_2(122100 and 012021) = 002000. Inserting this 2 in the 3rd position of the resultant rule, the encoded rule 112021 is obtained which is interpreted as $A, B \rightarrow D$.

- Clan Updation of Matriarch

For the local best fit in the clan the updation is as follows

$$x_{new,c_{i,j}} = Max\_bitValue\left\{\sum_{j=1}^{n} x_{ci,j,d}\right\} \quad (17)$$

where n=clan size. For rules $A \rightarrow B, A \rightarrow D, B \rightarrow D$, the encoded values are 1100, 1001, 0101, the updated rule is given by

1100 ----- $A \rightarrow B$
1001 ----- $A \rightarrow D$
0101 ----- $B \rightarrow D$

1101 ----- $A, B \rightarrow D$

Figure 4. Update procedure for Matriarch

- Calculation of fitness

The fitness function is given by

$$Fitness, f(x) = \frac{\alpha * Comp + \beta * Profit + \gamma * MaxSup\{A, H, AH, HA\}}{\alpha + \beta + \gamma} \quad (18)$$

Where α, β and γ ε [0,1] are the user-defined weights, Comp is the completeness of the rule and MaxSup{A,H,AH,HA} gives the maximum value of support among A,H,AH and HA for the rule 'x'.

- Calculation of Completeness

Completeness of any rule $X \rightarrow Y$ is given by

$$Comp(X \rightarrow Y) = \frac{n(XUY)}{\max(n(z))}\ where\ z\varepsilon X, Y \quad (19)$$

For any particular transaction,

$$n(XUY) += \begin{cases} 1\ if\ \{X, Y = \{1, Hi, Hj\}\}\ \text{where i = j or i} \neq \text{j} \\ 0\ otherwise\ if\ \{either\ X\ or\ Y = 0\}\ or\ if\ \{X, Y = 0\} \end{cases} \quad (20)$$

Scenarios for different requirement specification:

The proposed algorithm can be tuned according to the organization's requirement. The following cases illustrate the ability of the proposed algorithm to handle the organization's requirements (simple vague, profitable, seasonal, optimal rule) by making alterations in few parameters.

<u>Case I:</u> When special_occasion=ϕ and min. profit threshold=0, then the algorithm reduces to normal vague association rule mining that handles hesitation statuses and generates rules that helps in removing hesitation of an item.

<u>Case II:</u> When min. profit threshold=0 and special_occasion='occasion', then the occasionally hesitated rules are obtained rather than taking the entire transaction into consideration.

Case III: When special_occasion=ϕ and min. profit threshold=ϴ, hesitation rules with potential profitability (>=ϴ) is obtained.

6. ILLUSTRATION OF THE ALGORITHM

Following Synthetic dataset is used to illustrate the efficiency of the proposed algorithm.

Table 1. Transactional Dataset

TID	Itemset	Date of Transaction
1	A,B(H2),C,D	31-Dec-16
2	A,B	31-Dec-16
3	A(H1),D	31-Dec-16
4	B,C	1-Jan-16
5	A,C(H1),D(H2)	1-Jan-16
6	A,B(H1),C	2-Jan-16
7	A	2-Jan-16
8	B,C	2-Jan-16
9	A,C(H2)	3-Jan-16
10	B(H2),C	3-Jan-16

Step 1 and 2: Extract the transaction database from the given data set and represent the items accordingly. For any particular transaction, the presence of an item is represented by '1' while absence of an item is represented by '0'. B(H2) represents that the item 'B' is hesitated and its hesitation status for that particular transaction is given by H2.

The extracted Transaction database from the dataset in Table 1 is given in Table 2.

Table 2. Transaction Database

TID	A	B	C	D	Date of Transaction
1	1	H2	1	1	31-Dec-16
2	1	1	0	0	31-Dec-16
3	H1	0	0	1	31-Dec-16
4	0	1	1	0	1-Jan-16
5	1	0	H1	H2	1-Jan-16
6	1	H1	1	0	2-Jan-16
7	1	0	0	0	2-Jan-16
8	0	1	1	0	2-Jan-16
9	1	0	H2	0	3-Jan-16
10	0	H2	1	0	3-Jan-16

Table 3. Special Occasion Database

Special_Occasion	Start_Date	End_Date
End of year sale	31-Dec-16	1-Jan-16

Table 4. Item_Data Database

Item	Cost_Price	Selling_Price
A	40	45
B	70	78
C	15	17
D	100	110

In the above databases, items A-D represent A-Pen, B-Crayon set, C-Pencil, D-Geometry Box.

Step 3: Since 'End of year sale' is considered by organization, the transactions that falls between 31-Dec-16 and 1-Jan-16 are extracted. So, the first 5 transactions are being extracted and analysed further.

Step 4: Calculate intent by applying vague theory by using equations (12) and (13). Calculate AH pair value by (1) and (2). The overall intent and intent for different hesitation status of each item is given by Table 5 where as AH is given by Table 6.

Table 5. Intent of each item

H	A	B	C	D
$H_1(x)$	[0.6,0.8]	[0.4,0.4]	[0.4,0.6]	[0.4,0.4]
$H_2(x)$	[0.8,0.8]	[0.4,0.6]	[0.6,0.6]	[0.4,0.6]
Intent	[0.6,0.8]	[0.4,0.6]	[0.4,0.6]	[0.4,0.6]

Table 6. AH pair for each item

AH	A	B	C	D
$H_1(x)$	<0.7,0.2>	<0.4,0>	<0.5,0.2>	<0.4,0>
$H_2(x)$	<0.8,0>	<0.5,0.2>	<0.6,0>	<0.5,0.2>
Overall-AH	<0.7,0.2>	<0.5,0.2>	<0.5,0.2>	<0.5,0.2>

Step 5: Profit is calculated for each item in Table 7 by using the equation (14) in the Item_Data database

Table 7. Profit of each item

Item	A	B	C	D
Profit	5	8	2	10

Total profit=5+8+2+10=25

Step 6: Apply VAR to extract the rules by taking support=10%, confidence=10%, and profit=50%.The profitable frequent 2-item set in Table 9 obtained from candidate-2 item set in Table 8.

Table 8. Candidate-2 itemset

Rule	A		H		AH		HA	
	H1	H2	H1	H2	H1	H2	H1	H2
A→B	0.28	0.4	0	0	0	0.16	0.08	0
A→C	0.35	0.48	0.04	0	0.14	0	0.1	0
A→D	0.42	0.56	0	0	0	0.16	0.12	0
B→C	0.2	0.3	0	0	0.08	0	0	0.12
B→D	0.24	0.35	0	0.04	0	0.1	0	0.14
C→D	0.3	0.42	0	0	0	0.12	0.12	0

Table 9. Frequent-2 itemset

H	A	H	AH	HA
H1	A→B,A→D,B→D	-	-	-
H2	A→B,A→D,B→D	-	A→B,A→D,B→D	B→D

The candidate 3-item set is given by Table 10 and profitable frequent 3-item set is given by Table 11.

Table 10. Candidate-3 itemset

Type	Rules	Support		Confi(%)	
		H1	H2	H1	H2
A	A→B,D	0.168	0.28	0.24	0.35
	A,B→D	0.168	0.28	0.6	0.7
AH	A→B,D	-	0.032	-	-

Table 11. Frequent-3 itemset

Type	Rules (H1,H2)
A	A→B,D
	A,B→D

Following are the rules generated on items that are being bought frequently and most likely to buy in the future by framing appropriate selling strategies for both hesitation statuses.

Pen→Crayon Set,

Pen→ Geometry Box,

Crayon Set→ Geometry Box,

Pen →Crayon Set, Geometry Box and

Pen, Crayon Set → Geometry Box.

And following are the rules on items which are hesitated and likely to buy if added with some items that are bought frequently thereby helping in making effective selling strategy.

Pen→ Geometry Box for hesitation status H1,

Pen→Crayon Set, Pen→ Geometry Box and Crayon Set→ Geometry Box for hesitation status H2

Step 7: Now, optimize the rules by using nature inspired EHO. Let us assume a clan size of 3. i.e., each clan should contain 3 individuals. Here, 2 clans exist. So, the rules can be represented by setting α=0.15, β=0.60, γ=0.25 and probability of randomness, R=0.27.

Let Profit(%) be P, Completeness calculated according to (19), (20) be C, Maximum Support value among A,H,AH,HA for any hesitation status H_i be S and Fitness function of an individual 'x' be F(x) calculated by (18).

If R<0.25, one of the bits should be changed. If 0.25<=R<=0.5, any two bits should be changed and so on. For epoch of 2, the following steps are executed. Encoding is done according to (15) and clan updation by (16), (17). The highlighted rows of Table 12 and Table 13 are eliminated by clan separation operation and a random rule is generated.

Table 12. Epoch-1 of EHO

X	P	C	S	F(x)	Clan Updation
122100	***52***	***0.5***	***0.35***	***0.475***	***120210***
120021	60	0.75	0.35	0.56	112021
012021	72	0.5	0.25	0.57	112021
122101	92	0.25	0.175	0.633	122101
112021	92	0.25	0.175	0.633	122101

Table 13. Epoch-2 of EHO

X	P	C	S	F(x)	Clan Updation
120210	***28***	***0.5***	***0.35***	***0.331***	***120021***
122101	92	0.25	0.175	0.633	122101
112021	92	0.25	0.175	0.633	122101

Table 14. Condition termination

X	P	C	S	F(x)
120021	60	0.75	0.35	0.56
122101	92	0.25	0.175	0.633

If the user-defined threshold is specified, then rules with fitness greater than that threshold are considered as the optimal ones. The rules that yield more profit and help in removing the hesitation status of an item for framing efficient selling strategy is given by

Pen→ Geometry Box,

Crayon Set→ Geometry Box,

Pen →Crayon Set, Geometry Box,

Pen, Crayon Set → Geometry Box.

Scenarios for different requirement specification:

Following tables illustrate the rules generated on different organizational requirement specification scenarios.

Case I: When special_occasion=ϕ and minimum profit threshold=0, then the rules obtained will be

Table 15. Vague Association Rule

H	A	H	AH	HA
H1	A→B | A→C | A→D | B→C | C→D | A→B,C | A,B →C.	-	-	-
H2	A→B | A→C | A→D | B→C | B→D | C→D | A→B,C | A,B→C | A→C,D | A,C →D,	-	A→B	B→C

Case II: When minimum profit threshold=0 and special_occasion='End of year sale', then the rules generated are as follows.

Table 16. Occasional Vague Association Rule

H	A	H	AH	HA
H1	A→B | A→C | A→D| B→C | B→D | C→D A→B,C | A,B →C |A→C,D | A,C→D |A→B,D | A,B →D	-	A→C	A→C
H2	A→B | A→C | A→D | B→C | B→D | C→D |A→B,C | A,B →C |A→C,D | A,C →D |A→B,D | A,B →D | B→C,D | B,C→D | A,B→C,D | A→B,C,D | A,B,C→D	-	A→B A→D B→D C→D	B→C B→D

Case III: When special_occasion=ϕ and minimum profit threshold=50%, then the profitable rules generated are as follows

Table 17. Profitable Vague Association Rule

H	A	H	AH	HA
H1	A→B | A→D	-	-	-
H2	A→B | A→D | B→D	-	A→B	-

7. CONCLUSION

The proposed algorithm uses VAR Mining and utilizes EHO and is capable of generating more relevant, optimal rule set that is beneficial for effective mining on vague situations. The proposed methodology is capable of

- Accepting, evaluating and analyzing the vague information. The rules generated also helps in removing the hesitation of products and convert the 'almost sold' status of the product to 'sold' with an effective strategy which boosts the sales.
- Generating rules that interests the customer, business organization, retailers. The input parameters are set accordingly based on their interest (profitable/seasonal) in the rule.
- Delivering rules that yield maximum profit.
- Generating rules that are interesting at any point of time or within any time interval. The generated rules will vary to a great extent for different occasions based on the sale of any particular item on that occasion. This helps in stocking the items during the specific occasion accordingly. Also, it helps in removing the hesitation of products that have its maximum sale during that special occasion or event, and have not been considered throughout the year.
- Diversifying the search space and makes sure the optimal rules are generated according to the organization specified proportions. (profitable / complete / frequent rules or mixed proportion of these rules). It also removes sharp boundary and rare item problems and, helps making decisions in a more customized way.

8. REFERENCES

[1] J. Han and M. Kamber, 2006, Data Mining: Concepts and Techniques, 2nd ed., the Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor.

[2] Verlinde, H., De Cock, M., Boute, 2006, Fuzzy Versus Quantitative Association Rules: A Fair Data-Driven Comparison, IEEE Transactions on Systems, Man, and Cybernetics, Volume 36, Page(s): 679-683.

[3] Ashish Mangalampalli, Vikram Pudi, 2009, Fuzzy Association Rule Mining Algorithm for Fast and Efficient Performance on Very Large Datasets, IEEE 2009, Korea, ISSN: 1098-7584, E-ISBN: 978-1- 4244-3597-5, Page(s): 1163 – 1168, August 20-24.

[4] Improved apriori algorithm using fuzzy logic, 6th June 2014, IJARCSSE, Volume 4.

[5] Zdzisław Pawlak ,Rough set, Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Gliwice, Poland.

[6] Thabet Slimani, Class Association Rules Mining based Rough Set Method, Computer Science Department, Taif University College of Computer Science and Information Technology.

[7] Pawlak, Z., Grazymala-Busse, J. W., Slowinski, R., and Ziarko, 1995, Rough sets. Communications of the ACM, 38, 88-95.

[8] Grzymala-Busse, J. W. Slowinski , 1997, A new version of the rule induction system,. Fundamenta Informaticae, 31, 27-39.

[9] Surabhi Pathak and Akhilesh Tiwari, 2016, A Survey on Hesitation information mining, International Journal of Computer Applications, Vol.141, No.9.

[10] A. Lu, Y. Ke, J. Cheng and W. Ng, 2007, Mining Vague Association Rules, Springer.

[11] A. Lu, Y. Ke, J. Cheng and Wilfred. Ng, Mining Vague Association Rules, Department of Computer Science and Engineering The Hong Kong University of Science and Technology Hong Kong, China.

[12] Anjana Pandey and K. R. Pardasani, 2012, A Model for Mining Course Information using Vague Association Rule, International Journal of Computer Applications, vol. 58, no. 20, pp. 0975 – 8887.

[13] Vivek Badhe, Dr.R.S.Thakur and Dr.G.S.Thakur, 2015, Vague Set Theory for Profit Pattern and Decision Making in Uncertain Data, International Journal of Advanced Computer Science and Applications, Vol. 6, No. 6.

[14] A. Pandey and K. R. Pardasani, 2013, A Model for Vague Association Rule Mining in Temporal Databases, Journal of Information and Computer Science, vol. 8, no. 1, pp. 063–074.

[15] Reeti Trikha, Jasmeet Singh, 2014, Improvement in Apriori Algorithm with New Parameters, International Journal of Science and Research (IJSR) ,ISSN: 2319-7064, Volume 3 Issue 9.

[16] Vivek Badhe, Dr.R.S.Thakur and Dr.G.S.Thakur, A Model For Profit Pattern Mining Based On Genetic Algorithm, IJRET: International Journal of Research in

Engineering and Technology, eISSN: 2319-1163 , pISSN: 2321-7308.

[17] Prateek Shrivastava and Akhilesh Tiwari, 2015, Genetic Based hesitation Information Mining for Profitability Management , International Journal of Database Theory and Application, Vol.8, No.6,pp 75-88.

[18] Gai-Ge Wang, Suash Deb, Leandro Coelho, 2015, Elephant Herding Optimization, 3^{rd} International Conference on Computational and Business Intelligence.