

Exploring the Utility of Vague Concept for Uncertainty and Hesitation Management

Arun Kumar Singh¹ and Akhilesh Tiwari²

^{1,2}Department of CSE & IT,
Madhav Institute of Technology and Science,
Gwalior Madhya Pradesh, India
¹arunluh@gmail.com, ²atiwari.mits@gmail.com

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. The present paper focuses on the problem of mining important inference from supermarket basket data (in the presence of vagueness). The paper specifically studies the usefulness of vague set theory for the exploration of hesitation information and vague association rules. The hesitation information of an item plays a vital role in making selling strategies for the exhilaration of business. For this purpose, the vague set concept can be used as an important tool which can assist in the identification of hesitated item. The vague set theory with its two membership function provides more intuitive way to interact with the vague situation that causes the hesitation for any item. The effectiveness of the hesitated pattern and rule provide advanced decision-making capabilities that transform 'almost sold' items to 'sold items'.

Keywords: Vague Concept, Hesitation Information, Vague Association Rule, AH-pair

1. Introduction

Data mining techniques is being highly used for mining the unseen explorative information from the large database which has great potential to examine important information from these large databases. But the data which is accumulated in the data repository is not only large but more imprecise and real indicating presence of vagueness, incompleteness and uncertainty which causes the problem during important decision-making task [4]. Dealing with uncertainty is a major challenge faced in many research areas in data mining that rises in the large range of complex applications. The problem of decision making, particularly in monetary issues, is a critical task in every business.

In past years, data mining technology follows the conservative move that offers only statistical analysis and discovers rules. The two fundamental goals of data mining are prediction and description. Prediction makes use of existing attribute in the database in order to predict unknown or future attribute of interest and description emphasizes on discovering patterns, association, changes, and significant structures from large database which become more crucial in uncertain and vague situations. Various mathematical approaches are used to deal with the uncertainty are probability theory, fuzzy logic approach, rough set, vague set theory and soft set approach [13]. All these theories deal with different kinds of uncertainties in database.

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 same significance without taking consideration of vague status of an item which is not always the case. Association rule mining is conducted on the transactions that consist of the item bought by customers.

There are numerous things that are not purchased but rather client may consider purchasing them. This data on customer's deliberation to purchasing associate degree item is that the hesitation data of the item since the client is hesitated whether or not to shop for it. In implementing and designing of selling strategies the importance of hesitation information of an item play vital role that have potential to turn "under deliberation" item into "well sold" items. As a result this useful information will boost up the sale of item. Conversely, in the past due to difficulty in collection of hesitation information the traditional association rule mining does not consider such items information for mining purposes. Nevertheless, with time as advancement in technology of data dissemination happens obtaining hesitation information of an item has become much easier task than past, as a result necessity of incorporating hesitation information for mining purposes are strongly felt by the data scientist.

Now take an instance of online shopping circumstances on "flipkart.com", it is possible to collect enormous amount of data from web logs and application account information to collect hesitation information. For example in case of online shopping case:

- The items that customers visited most frequently.
- The items that are in customers put into their online shopping cart but not buy eventually.
- The items that customer's favorite list that he considered to buy next time visit in store.
- The unavailability of item in store that is in customer's wishing list and so on.

Therefore based on different browsing patterns for an items hesitation information can identified and categorized with respect to different application. The hesitation information will then be accustomed plan and execute selling strategies that can potentially turn those "under consideration" things into "well sold" things.

There are several statuses of a part of hesitation information called hesitation status (HS) [6]. For motivations consider an example of online shopping scenario that involves various hesitation statuses:

- The hesitation status (s1) for item in which customer browsed only once for that item and leave.
- Hesitation status (s2) for items that are browsed in detail by the customer that is all specifications and figures but not put into their online shopping carts;
- Hesitation status (s3) for items that customers checked out eventually that he put in shopping cart.

The above mention Hesitation statuses are the pieces of hesitation information for items. Some of the Hesitation statuses are comparable based on some measure which means we can describe an order on these Hesitation statuses. For example, the possibility that an item purchased by customer following criterion may possible *i.e.*, $s1 \leq s2 \leq s3$. For better understanding let us consider scenario in which customer bought an item 4 times and almost buy 3 times and did not browse it 3 times out of 10 visit then we can obtain vague membership value, [0.4, 0.7] for that item. When we distinguish dissimilar Hesitation statuses, say the customer hesitated to buy the item 3 times in Hesitation statuses, 2 times in s1 and once in s2, where $s1 \leq s2 \leq s3$. In this situation vague membership value for (s1) is [0.4, 0.6] and for (s2) is [0.6, 0.7] since there is no hesitation evidence for hesitation in (s3) and $s2 \leq s3$, hence vague membership for (s3) is a single point *i.e.*, [0.7, 0.7].

The vague set theory can be a basis of modeling hesitation information of items which capture three types of evidences based on interval-based membership with respect to object in the universe of discourse: support, against, hesitation. Thus, for mining context it naturally models the hesitation information.

The concept of attractiveness and hesitation of an item is used to study the relation between support evidence and hesitation evidence that are based on median and imprecision membership that derived from vague set membership. This can be understand that items with high attractiveness are well selling items and its probability of selling next time is also high and item with high hesitation also have high probability of selling but customer hesitated due to some reason hence identifying and resolving issue relating to this hesitation transform these items to “well sold” items.

Vague association rules (VARs) [9] is defined using attractiveness and hesitation of items an AH-pair database that consist of AH-transactions can model with hesitation information where attractiveness is represented by A and H represents hesitation which captures four types of relationship between two set of items. Four types of support and confidence are defined to appraise the quality of the different types of vague association rules. Also properties of support and confidence can be used to boost up the mining process. The concept of vague associations rules are not limited to online shopping it can be applied to any area where hesitation situation can be found.

From the literature it is apparent that very modest attentions are given for mining Hesitation information. In this paper we are trying to explore the different theories and methods that help in mining hesitation information and there pro and cons regard in this field and trying exploring the utility of vague concept for uncertainty and hesitation management.

2. Preliminaries

2.1. Type of Uncertainty in Data Mining

The traditional data mining approach uses statistical and logical significances to find knowledge from the databases. But databases become more pervasive and heterogeneous which have close resemblance to real world, hence more susceptible to uncertainty. It is impossible to assert any value to an object when modeling is done when uncertainty occurs. Various types of uncertainties are:

- Imprecision occurs if available information is not specific to the desired modeling.
- Ambiguity occurs when object in the model have stringency.
- Inconsistency when more statement in modeling cannot be true at same instant.
- Vagueness occurs when object in modeling is includes intrinsic vague value which is not clearly expressed.

Various soft techniques that does deal uncertainty to some extent and have various uses in research field of association rule mining.

2.2. Vague Set and Intuitionistic Fuzzy Sets

The basics of handling uncertainty and vagueness for hesitations information can be explain through vague set theory and intuitionistic fuzzy set theory we explore basic concept related to these and discover which one is better in handling vague data [7].

2.2.1. Vague Set: [2] 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:

$$\alpha_V : U \rightarrow [0, 1], \quad (1)$$

$$\beta_V : U \rightarrow [0, 1], \quad (2)$$

$$\text{and } \alpha_V(u) + \beta_V(u) \leq 1, \quad (3)$$

Where $\alpha_V(u)$ is derived from the evidence for u , which is lower bound on the grade of membership of u and $\beta_V(u)$ is derived from the evidence against u which is lower bound

on the grade of membership of the negation of u . Suppose a universe of discourse $U = \{u_1, u_2, u_3 \dots u_n\}$. A vague set V of the universe of discourse U can be represented by expression

$$V = \sum_{i=1}^n [\alpha_V(u_i), 1 - \beta_V(u_i)] / u_i \quad (4)$$

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

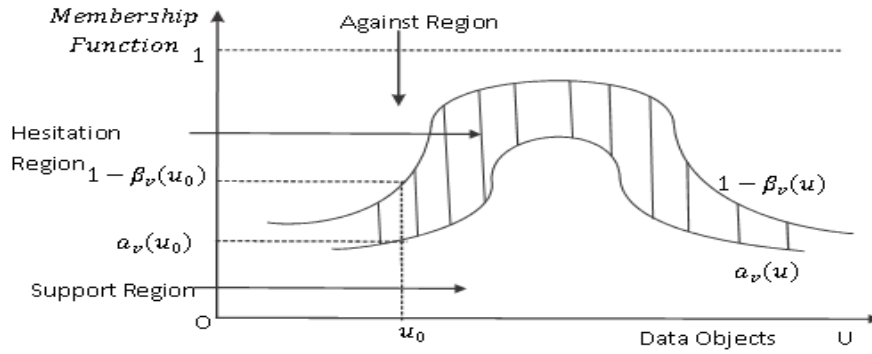


Figure 1. Membership Function of Vague Set

Vague sets are a generalization of Fuzzy sets, since the grade of membership $\mu_V(u)$ may be inexact in vague sets. In intuitionistic fuzzy set the concept of membership generalization through interval has already proposed.

2.2.2. Intuitionistic Fuzzy Sets: [10] an intuitionistic fuzzy set $A = \{ \langle u, \mu_A(u), \nu_A(u) \rangle \mid u \in U \}$ in a universe of discourse U is characterized by a membership function μ_A and a non-membership function ν_A as follows:

$$\mu_A : U \rightarrow [0, 1], \quad (5)$$

$$\nu_A : U \rightarrow [0, 1], \quad (6)$$

$$\text{and } 0 \leq \mu_A(u) + \nu_A(u) \leq 1. \quad (7)$$

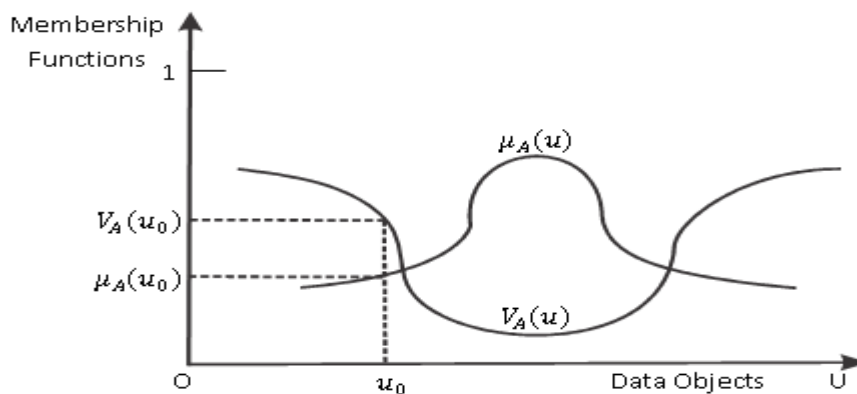


Figure 2. Membership Function of Intuitionistic Fuzzy Set

2.2.3. Views in Measurement of Both Set for Vagueness: What a membership grade means actually there is no consensus in the interpretation given in the literatures that

argue some ideas of how to measure memberships of vague sets and intuitionistic fuzzy sets. Bilgic and Turksen present a review of various interpretations of fuzzy membership function and the ways of obtaining a membership function [17]. Vague sets have similar interpretations of membership grades. Suppose a vague predicates “Arun is tall” is given in interval [0.5, 0.8].

The possible views on how to measure the membership:

- **Likelihood View:** Population that declares Arun is tall is 50% - 80%.
- **Random Set View:** 50% - 80% of given population describes the interval containing Arun’s height is “tall”.
- **Similarity View:** Arun’s height is far from the benchmark objective which is truly “tall” to the degree 0.2-0.5. Here if we presume a benchmark instance of “tall” is 300cm with the full degree [1,1], then Arun’s height is away from 300cm to the degree 0.2-0.5 means his height is between $(1 - 0.5) \times 300$ and $(1 - 0.2) \times 300$ cm, that is, 150-240cm.

In case of IFSs following interpretation is made:

- **Likelihood View:** The population that declares Arun is tall is 50% and 20% declaring Arun is not tall. Another 30% are neutral.
- **Random Set View:** Population describes “tall” as an interval containing Arun’s height is 50% while 20% does not. Another 30% is neutral.
- **Similarity View:** Same as case of vague set.

2.2.4. Comparison between Vague Set and Intuitionistic Fuzzy Set for Handling Vague Data:

The difference between vague sets and intuitionistic fuzzy set is due to the definition of membership intervals. We have $[\alpha_V(u), 1 - \beta_V(u)]$ for u in Vague sets but $\langle \mu_V(u), \nu_A(u) \rangle$ for u in Intuitionistic fuzzy set. Here the semantics of μ_A is the same as with α_V and ν_A is the same as with β_V . Conversely, the boundary $(1 - \beta_V)$ is able to indicate that data value possibly exists. On analyzing graphical view of data sets the subtle difference provide simpler and meaningful information. It can be viewed that, the shaded region formed by the boundaries in “Figure 1” of vague set naturally represents the possible existence of data. Thus, this “hesitation region” corresponds to the perception of representing vague data.

Vague sets are more natural than Intuitionistic fuzzy set for merging fuzzy object. Consider an example if we merge three fuzzy values say 0.5/u, 0.6/u, 0.8/u. then we can directly find the vague value $[.05, .08]/u$ this means that lower bound of the membership of u is the minimum of the fuzzy membership (.05) and that the upper bound is maximum of the fuzzy membership (0.8). But by using intuitionistic fuzzy value we obtain $\langle u, 0.5, 0.2 \rangle$ which is much less intuitive.

2.3. Vague Set Memberships

Vague set captures different notions of incompleteness its membership provides various deductions that are useful in evaluation of hesitation information for vaguely specified data. In order to compare vague values we used two memberships: median membership and imprecision membership [7]. 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.

2.3.1. Median Membership: Median membership is defined as

$$M_m = \frac{1}{2} (\alpha + (1 - \beta)) \quad (8)$$

This signifies a vague value that represents the overall evidence. It can be verified that $0 \leq M_m \leq 1$. Apparently, the vague value $[1, 1]$ has the highest M_m , which means the corresponding object that enormously 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 subsequent object absolutely does not belong to the vague set.

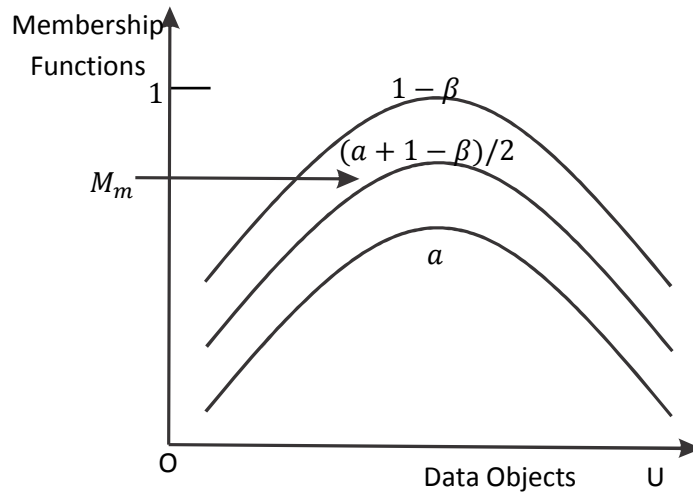


Figure 3. Median Membership of a Vague Set

2.3.2. Imprecision Membership: Imprecision membership is defined as

$$M_i = ((1 - \beta) - a) \quad (9)$$

This signifies a vague value that represent overall imprecision. It can be veteran that $0 \leq M_i \leq 1$. The vague value $[p, p]$ ($p \in [0, 1]$) has the lowest M_i this means that the membership of the corresponding object is precise, *i.e.*, a fuzzy value. Whereas the vague value $[0, 1]$ has the highest M_i that doesn't have any information regarding the membership of corresponding objects.

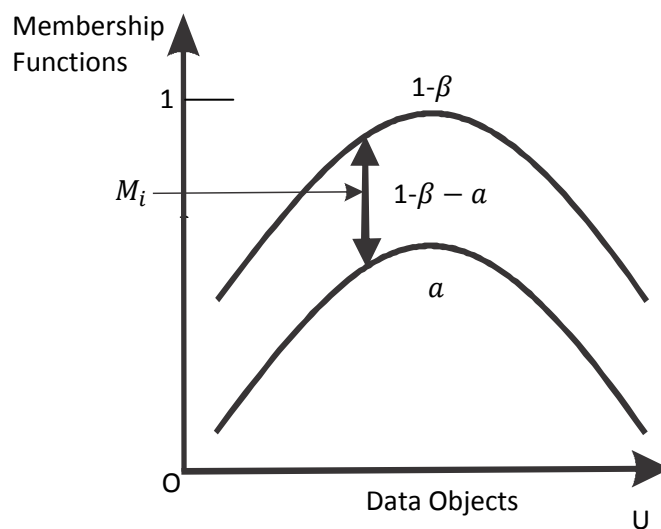


Figure 4. Imprecision Membership of a Vague Set

2.4. Empty Vague Set

A vague set is an empty vague set only in condition when it's true membership function $\alpha = 0$ and false membership function $\beta = 1$ for all u . It is denoted as \emptyset_V . In ordinary set theory the \emptyset_V can be regarded as the generalization of the empty set. A vague set is empty, if only if it's true membership function and false membership function are both 0 (i.e., $\alpha = 0, 1 - \beta = 0$) which means that there is no information about the status of corresponding object whether it is in vague set or not [2]. Further empty vague value or simply empty value is represented $[0, 0]$. There are three kinds of nulls in vague set that can be generalized.

- Unknown (UNK) signifies that the value exists but unknown at represent time. In case of crisp sets all memberships are considered as $[1, 1]$ which is regarded as special case of vague sets. In case of vague set generalized UNK signifies that memberships can be vague values but unknown at the represent time for any possible vague set such that $\forall u_i \in U. \alpha_V(u_i) \in (0, 1]$ and $\beta_V(u_i) \in (0, 1]$.
- Does Not Exist (DNE) signify that "the value is inapplicable". We can view DNE as vague set $V = \sum_{i=1}^n [0, 0]/u_i$ that means u_i does not fit in to the vague set and the indications is totally against it means $\alpha_V(u_i) = 0$ but $\beta_V = 1$.
- No information (NI) represents that "no information is available for the values", i.e., it is either unknown or does not exist.

2.5. Similarity Measures of Vague Sets

Similarity measure between two vague sets which is based on median membership and imprecision membership. Let x and y be vague values such that $x = [\alpha_x, 1 - \beta_x]$ and $y = [\alpha_y, 1 - \beta_y]$ and let ΔM_m as difference between median membership and M_i as difference between imprecision memberships then the similarity measure between two vague values x and y is given as

$$M(x, y) = \sqrt{(1 - \Delta M_m)(1 - \Delta M_i)} \quad (10)$$

$$M(x, y) = \sqrt{\left(1 - \frac{|\alpha_x - \alpha_y| - |(\beta_x - \beta_y)|}{2}\right)(1 - |(\alpha_x - \alpha_y) - (\beta_x - \beta_y)|)} \quad (11)$$

Considering X and Y be two vague sets such that

$$X = \sum_{k=1}^n [\alpha_X(u_k), 1 - \beta_X(u_k)]/u_k \quad (12)$$

$$Y = \sum_{k=1}^n [\alpha_Y(u_k), 1 - \beta_Y(u_k)]/u_k \quad (13)$$

The similarity measure between the vague sets X and Y is evaluated as:

$$M(X, Y) = \frac{1}{n} \sum_{k=1}^n M([\alpha_X(u_k), 1 - \beta_X(u_k)], [\alpha_Y(u_k), 1 - \beta_Y(u_k)]) \quad (14)$$

$$M(X, Y) = \frac{1}{n} \sum_{k=1}^n \sqrt{(1 - \Delta M_{m,k})(1 - \Delta M_{i,k})} \quad (15)$$

For two vague sets X and Y the following statement are true that give good feature for similarity measure between vague sets [4].

- The similarity measure is commutative.
- The similarity measure is bounded to $[0, 1]$.
- $M(X, Y) = 1$ if and only if $X = Y$.
- $M(X, Y) = 0$ if and only if
 1. A vague value in X is $[0, 0]$, then the corresponding vague value in Y is $[1, 1]$
 2. A vague value in X is $[0, 1]$ then he corresponding vague value in Y is $[a, a]$ such that $0 \leq a \leq 1$.

2.6. Vague Relations

The vague relations is set of vague sets that can considered as an extension of classical relations in which all vague values are $[0, 1]$ and fuzzy relations where all vague values $[a, a]$ such that $0 \leq a \leq 1$. The vague relations capture more information about vagueness and defined as follows [7]:

Let $U = \{U_1, \dots, U_m\}$ be a collection of universes of discourse. Let $\text{Dom}(P_i)$ be the domain corresponding to the attribute P_i we define $\text{Dom}(P_i) = \{V \mid V \text{ is a vague set of } U_i\}$. A vague tuple (a_1, a_2, \dots, a_m) over a relation scheme, $R = \{P_1, P_2, \dots, P_m\}$, is an element in $\text{Dom}(P_1) \times \text{Dom}(P_2) \times \dots \times \text{Dom}(P_m)$. A vague relation r over R is a subset of $\text{Dom}(P_1) \times \text{Dom}(P_2) \times \dots \times \text{Dom}(P_m)$.

2.7. Operations on Vague Sets

The basic operation of vague set which include

- **Complement:** The complement of a vague set V is denoted by V' and is defined by

$$\alpha_{V'}(u) = \beta_V(u). \quad (16)$$

$$1 - \beta_{V'}(u) = 1 - \alpha_V(u). \quad (17)$$

- **Containment:** A vague set V_A is contained in another vague set V_B , $V_A \subseteq V_B$, if and only if,

$$\alpha_{V_A}(u) \leq \alpha_{V_B}(u), \quad (18)$$

$$1 - \beta_{V_A}(u) \leq 1 - \beta_{V_B}(u) \quad (19)$$

- **Equal:** Two vague sets V_A and V_B are equal, written as $V_A = V_B$, if and only if, $V_A \subseteq V_B$ and $V_B \subseteq V_A$; that is

$$\alpha_{V_A}(u) = \alpha_{V_B}(u), \quad (20)$$

$$1 - \beta_{V_A}(u) = 1 - \beta_{V_B}(u). \quad (21)$$

- **Union:** The union of two vague sets V_A and V_B is a vague set V_C , written as $V_C = V_A \cup V_B$, whose true membership and false membership functions are related to those of V_A and V_B by

$$\alpha_{V_C}(u) = \max(\alpha_{V_A}(u), \alpha_{V_B}(u)) \quad (22)$$

$$1 - \beta_{V_C}(u) = \max(1 - \beta_{V_A}(u), 1 - \beta_{V_B}(u)) = 1 - \min(\beta_{V_A}(u), \beta_{V_B}(u)). \quad (23)$$

- **Intersection:** The intersection of two vague sets V_A and V_B is a vague set V_C , written as $V_C = V_A \cap V_B$, whose true membership and false membership functions are related to those of V_A and V_B by

$$\alpha_{V_C}(u) = \min(\alpha_{V_A}(u), \alpha_{V_B}(u)) \quad (24)$$

$$1 - \beta_{V_C}(u) = \min(1 - \beta_{V_A}(u), 1 - \beta_{V_B}(u)) = 1 - \max(\beta_{V_A}(u), \beta_{V_B}(u)). \quad (25)$$

3. Hesitation Information

Hesitation information [6] plays a vital role in calculating the customer intent for a particular item or item set in the set of transactions. This is based on hesitation status

which represents a particular state involving two certain situations of “buying and not buying” with admiration to a customer in the process of customer transactions. This information is important and should effectively mine to improve the selling strategies so that items that customer take under consideration can transform to well sold items. Vague set theory plays important role in mining the hesitation information as its interval provide hesitation region. Based on suitable modeling it can apply to various scenarios. Let consider a supermarket scenario in which the possible causes for hesitation from customer’s view point can be:

- The impact of suggestions of other people about an item.
- Hesitation due to the similar old items present (dependency between new and old items)
- Delay factors due to availabilities of item.
- Delay due to old items present.
- Money constraints.
- The attractiveness of similar items.
- Occasional effect (people wait for buying some items) *etc.*,

These factors lead the hesitation from a customer point of view and hence decrease the attractiveness of an item. These factors are analyzed and some proper strategies are made that increase the attractiveness of item to increase the profitability.

3.1. Modeling Hesitation Status

Consider an example of online shopping scenario on “flipkart.com” for placing an order by the customer which illustrates the idea of hesitation status (HS). The following steps show the sequential process for placing orders.

- HS1: Find the item you want to buy
- HS2: Put items to your shopping cart
- HS3: Proceed to checkout
- HS4: Sign in
- HS5: Enter a shipping address
- HS6: Choose a shipping method
- HS7: Provide a payment method information
- HS8: Review and submit your order
- HS9: Check your Order options
- HS10: Product returns options

Where HS1, HS2, HS3.....HSn shows different hesitation status stages. During the progress of any transaction, a customer may drop the ordering process at any of the hesitation status. This may be due to any reason. Furthermore, if a customer is in process of placing an order and is currently exploring some higher level hesitation status stage then this higher level hesitation status stage exploration may result in the successfulness of transaction *i.e.*, increase in the certainty for the purchase of item.

3.2. Hesitation and Overall Hesitation

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

$$\alpha(x) + \beta(x) + \sum_{i=1}^n h_i(x) = 1 \quad (26)$$

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

$$H(x) = \sum_{i=1}^n h_i(x) \quad (27)$$

This can be easily find from the above definition that

$$H(x) = 1 - \alpha(x) - \beta(x) \quad (28)$$

3.3. Intent and Overall Intent

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

3.4. Attractiveness and Overall Attractiveness

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

$$\text{ATT}(x) = \frac{1}{2} (\alpha(x) + (1 - \beta(x))). \quad (29)$$

3.5. Attractiveness Hesitation Pair Database

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

4. Association Rules

An association rule is a rule, that implies convinced association relationships between a set of objects such as occurs together or one implies to other in a database. Association rules recognize relationships between sets of items in a transaction database. Association Rule Mining finds interesting association or correlations among a large set of data items. It provides a useful mechanism for discovering correlations between items belonging to customer transactions in market basket database.

4.1. Traditional Association Rule Mining

An association rule is a rule, that entails probabilistic relationship, with the form $X \Rightarrow Y$ between sets of database attributes, where X and Y are set of items, and $X \cap Y = \Phi$. Given the set of transactions we are interested in generating all rules that satisfy certain constraints. These constraints are support and confidence. The support of the rule is the fraction of the transactions in database D that satisfy the union of the items in X and Y . The probability measured as the fraction of the transactions containing X also contains Y , called the confidence of the rule. Apriori is first algorithm in this context that is based on market basket analysis [1].

The traditional association rule considered all item with similar importance and imposed on transactions that actually happen and ignored all vague situations.

4.2. Vague Association Rule

A vague association rule (VAR) $r = (X \Rightarrow Y)$ 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. 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.

4.3. Support of Vague Associations Rule

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

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

Where Z is an A (or H or AH or HA) frequent itemset if the A-support (or H-support or AH-support or HA-support) support of Z is not smaller than the (respective A or H or AH or HA) minimum support threshold σ .

4.4. Confidence of Vague Association Rule

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

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

To design an efficient algorithm for mining vague association rule some properties of vague association rule are beneficial. The anti-monotonic property for the support defined for a certain itemset with respect to hesitation statues. For example support of an itemset with respect to an hesitation statues to be less than minimum support σ then we can prune the same itemset in mining search space.

5. Mining Vague Association Rule

For mining vague association rule we first mine attractiveness-hesitation pair database in regard of certain hesitation status or overall hesitation in order to get set of all A, H, AH and HA frequent itemset from this we generate vague association rules using obtained frequent itemset. To generate the A, H, AH, HA pair from database first step is to calculate the intent of an item. The intent of an item x , is a vague value $[\alpha(x), 1 - \beta(x)]$.

Intent of an item is calculated from the hesitation based database with respect to any hesitation status or overall hesitation status as follow:

- Initialize the intent array to store the intent;
- Initialize favor(α) and against (β) variable with value zero;
- For each $i = 0, 1, 2 \dots$ where $i < \text{no of attributes id}$, do
- For $j = 0, 1, 2 \dots$ where $j < \text{no of tuple of the attributes}$, do
- Increment favour (α) by one when $D[i][j]$ is equal to one;
- Increment against(β) by one when $D[i][j]$ is equal to zero;
- End of for ;
- Generate intent using favour and against as $[\alpha, 1 - \beta]$;
- End of for ;
- Return all intent

The second step is to calculate the AH-pair which is calculated by a simple iterative method that takes the intent as input and evaluated as:

- Initialize AH-pair array to store AH pair;
- For each $i = 0, 1, 2 \dots$ where $i < \text{no of attributes id}$
- Attractiveness as a median membership *i.e.*, $\frac{1}{2} (\alpha + (1 - \beta))$;
- Hesitation as a difference of α and $1 - \beta$ using intent;
- End of for;

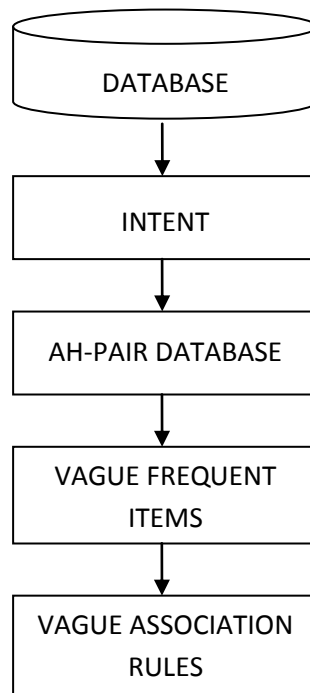


Figure 5. Framework for Mining Vague Association Rules

Now let A_i and H_j be the set of A frequent itemset and H frequent itemset respectively. Consider A_iH_j be the set of AH frequent itemsets containing (i) item with A values and (j) items with H values here A_iH_j is equivalent to H_jA_i and C_w be the candidate frequent itemset from that frequent itemset W is to be generated where W is A_iH_j or H_jA_i . Mining vague frequent item (VFI) is as follows:

- Initialize FIs array to store FI;
- Calculate 1st vague frequent itemset by scanning database mine A_1 and H_1 ;
- Generate C_{A_2} from A_1 , $C_{A_1H_1}$ from A_1 and H_1 and C_{H_2} from H_1 ;
- Verify the candidate frequent item set C_{A_2} , $C_{A_1H_1}$ and C_{H_2} to give A_2 , A_1H_1 and H_2 respectively;
- for each $k = 3, 4, \dots$, where $k = i + j$, do
- Generate C_{A_k} from A_{i-1} and C_{H_k} from H_{j-1} for $i = k$;
- Generate $C_{A_iH_j}$ from $A_{i-1}H_j$, for $2 \leq i < k$ and from A_1H_{j-1} for $i = 1$;
- Verify the candidate frequent items in C_{A_k} , C_{H_k} and $C_{A_iH_j}$ to give A_k , H_k and A_iH_j ;
- If all A, H, AH are greater than σ add into the array;
- return all ;

This method calculates the vague frequent itemsets in which we first mine the set of frequent items A_1 and H_1 from the input AH-pair of database (D). Then generate the candidate frequent itemset. Next we use the frequent itemset containing $(k - 1)$ items to generate the candidate frequent itemset containing k items, for ≥ 3 . The support of the candidate frequent itemset is computed and only those with that greater than minimum threshold are retained as frequent itemset. Terminate the process when no candidate itemset is generated and return all frequent set. After mining the set of all frequent itemsets vague association rule are generated from the frequent itemset. After mining all frequent itemset vague association rule is found from the FIs. There are four type of vague association rule. First using classical association rule for each A or H frequent item Z we can generate vague association rule $X \Rightarrow Y, \forall X, Y$ where $X \cup Y = Z$. Then for each AH (or HA) frequent item set $Z = X \cup Y$, where X is an A frequent itemset and Y is an H frequent itemset we can generate two vague association rule $X \Rightarrow Y$ and $Y \Rightarrow X$.

8. Background and Related Work

In 1993 Agrawal, Imielinski and Swami [1] proposed the Association Rule Mining that identify relationships among sets of items in a transaction database. An association rule is a rule, that entails probabilistic relationship, with the form $X \Rightarrow Y$ between sets of database attributes, where X and Y are set of items, and $X \cap Y = \Phi$. Apriori is the first algorithm in this field which determine the frequent itemsets for Boolean association rules. The name of the rule is predicated on the very fact that the rule uses previous information of frequent itemset properties it is a iterative approach as a level-wise search, where k -itemset are used to explore $(k+1)$ itemset.

In 1993 Gau W.L., Buehrer, *et al.*, [2] proposed the concept of Vague set theory which address the drawback of fuzzy set theory in finding the vague values by providing the two membership function that are more efficient in handling that vague data in real scenario. Their paper on vague set theory provides the bases of exploring the data that are vaguely specified in nature and provide the foundation for mining hesitated items.

In 1999 Atanassov *et al.* [10] proposed the Intuitionistic Fuzzy Sets: Theory and Applications which is similar in nature with vague set theory, having two membership function for handling vague data. Although vague set theory and Intuitionistic Fuzzy Sets Theory are equivalent by basic definition but the main difference between the vague set theory and Intuitionistic Fuzzy Sets Theory is that allow for a more intuitive graphical

representation of vague data, which facilitates significantly improved analysis in data relationships, incompleteness and similarity measures.

In 2005 Lu. A., *et al.*, [7] done a comprehensive study between vague sets or intuitionistic fuzzy sets for behavior vaguely specified data and provide which one is better method in handling vague data. According to Lu. A. et al. vague set theory and Intuitionistic Fuzzy Sets Theory, of Fuzzy sets. Based on interval-based membership, the vague set theory provide much communicative way for capturing vagueness present in data. The notions of vague set theory and intuitionistic fuzzy set theory can treated as comparable in context that an intuitionistic fuzzy sets theory is isomorphic to a vague set theory.

In 2006 Przemyslaw Grzegorzewski [18] proposed the approach of generalization for well-known Kendall's coefficient of concordance in the scenario where order of all elements is not structured. The basic requirement of scenario for measure of concordance is possessed by this coefficient. It can be further viewed that there coefficient is useful in the situation where non-comparable and missing information may occurs frequently such as data mining, statistics etc. The generalization suggested that there is no degeneration in the set of ranking that means the object taken under consideration can uniquely ranked by at least one of the observer. Further exploration of coefficient of concordance is still required for the omission of assumptions.

In 2007 Yanhong Li, *et al.*, [15] proposed Similarity measures between intuitionistic fuzzy (vague) sets and perform comparative analysis. They compared and summarized by their counterintuitive examples in pattern recognition. They also demonstrate positive aspects of each similarity measure along with contradict cases and conversation of the circumstances under which each may not work as desired. His research presented the benefits selection and application of vague set and intuitionistic fuzzy set in context of similarity measure.

In 2007 Lu. A, *et al.*, [5] proposed management of inconsistency in functional dependencies of vague relations. According to him due to intensive data distribution arising from different pervasive computing resources such as high volume data obtained from different resources has frequently found vague information in many databases. Lu and Wilfred Ng provide the approach to maintaining consistency of vague database using data dependencies. They applied vague set theory in relational databases to extend the concept of functional dependencies in order to deal with widely existent vague information and proposed vague functional dependencies.

In 2007 Ng. W., *et al.*, [6, 9] proposed the method of mining vague association rules and devised an algorithm for it that capture more richer and specific information than the original traditional association rule mining. However the modelling of hesitation information with respect to different hesitations status is newly concept that is incorporates the hesitation information into association rules (ARs) is provided in his research.

In 2010 Qing Shen, *et al.*, [19] do basic comparison between Fuzzy Sets, Rough sets and Vague sets according to him all these theory are extension of classical sets theory. They extended the relevance of set theory to vague and uncertain problem and discuss the differences and relationship between them.

In 2011 Shirong Gou, *et al.*, [20] combines the rough set and vague set in fuzzy data processing and provides a novel method for uncertain knowledge acquisition through vague rough set approach. Under the uncertain situations for knowledge extraction vague-rough set is beneficial. They compute lower approximation distribution using vague-rough set for the attribute reduction. The concept of discernability matrix and attribute reduction is utilized in vague decision information system. The result for decision information system extends the corresponding method of classical rough set and provides a new avenue to uncertain knowledge acquisition.

In 2011 P.K. Das, *et al.*, [21] projected the concept vague soft sets in student's evaluation they applied to extend Biswas's method for student's answer script evaluation and a hypothetical case has taken as an example.

In 2012 & 2013 Pardasani K. R., Anajan Pandey [11, 12] Use vague association rule in modelling for mining course information in which they studies that different university offering different courses of different types over several years and find out biggest issue with that is to collect information is utilize to make course more effective and solve this issue through vague association rule (VARs). They extend this concept further by giving model for vague association rule mining in temporal database.

In 2015 Vivek Badhe, R. S. Thakur, *et al.*, [16] used vague set theory for profit pattern and decision making in uncertain data. They perform vague association rule on FMCG database that contains the daily inventory products purchased by user.

9. Discussions

The traditional approach to association rule mining uses the Apriori or similar algorithms that are based on statistical implication of items present in the database. It considered certain data that contains items actually happen in transactions. But as technology evolved the databases are more closed to real world that contains certain amount of uncertainty hence basic measure of support and confidence were proving scarce in finding useful knowledge. Some effective measure and dimensions are needed to be incorporated in support and confidence to handle the uncertain data for finding proficient rules. As there are various type of uncertainty present in database there are many mathematical tools like Fuzzy set, Rough set, Vague set, Soft set that have been used over time to fulfill one requirement with database. Vagueness is the property of an item that is difficult to differentiate and comprehend hence principle of vague set theory is useful in dealing with vagueness. Fuzzy logic is a special case vague logic where both membership functions have same value therefore vague set theory allows binding the existence of items to an interval. Incorporation of vague logic to traditional association algorithm to find more relevant yet vague rules for advanced decision making. There are various fields where these methods can be applied like temporal mining, occasional mining, course selection in academic etc. The vague set theory is more conformable to people's thinking than fuzzy set. It describes unknowns and uncertain data more effectively. Vague set is mainly applied to machine learning, fuzzy decisions, information processing, and expert system.

10. Conclusions

The traditional association rule does not include the uncertain and hesitated information. The vague set theory provide the natural way of modeling hesitation information by providing two membership function thus provide a model for hesitation information which address the limitation in association rule that previously used which not include the hesitation information present in the transactions. Vague associations rule mining combine the statistic based pattern extraction with value based decision making to achieve the commercial goals. Hesitation and attractiveness is used for an item to different hesitation status that reflects the overall information of customer intention for item. The vague set allows us to consider the vague uncertainty existing in the database and to utilize it to mine such rules that ultimately provide better correlation among items. In future the vague association rule and algorithm has wide application for example different ranking scores together with click through data can be modeled as an object having different hesitation status. In this vague association rule can be used to reflect different user's preferences. Such models can further be developed and extended to problems involving mining of hesitation information in different conditions and situations.

References

- [1] R. Agrawal, T. Imielinski and A. N. Swami, "Mining association rules between sets of items in large databases", In Buneman, P., Jajodi A, S., eds.: SIGMOD Conference, ACM Press, (1993), pp. 207–216.
- [2] W. L. Gau and D. J. Buehrer, "Vague sets", IEEE Transactions on Systems, Man, and Cybernetics, vol. 23, (1993), pp. 610–614.
- [3] L. A. Zadeh, "Fuzzy sets", Information and Control, vol. 8, (1965), pp. 338–353.
- [4] A. Lu and W. Ng, "Maintaining consistency of vague databases using data dependencies", Data and Knowledge Engineering, vol. 68, (2009), pp. 622-641.
- [5] A. Lu and W. Ng, "Managing merged data by vague functional dependencies", In: Atzeni P., Chu W., Lu H., Zhou S., Ling T.-W. (eds.) ER, LNCS, vol. 3288, (2004), pp. 259–272. Springer, Heidelberg.
- [6] A. Lu and W. Ng, "Mining Hesitation Information by Vague Association Rules", Lecture Notes in Computer Science, Springer, vol. 4801, (2009), pp. 39-55.
- [7] A. Lu and W. Ng, "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, LNCS, vol. 3716, (2005), pp. 401–416. Springer, Heidelberg.
- [8] A. Lu and W. Ng, "Handling Inconsistency of vague relations with functional dependencies", In: ER, LNCS, vol. 4321, (2007), pp. 301-312.
- [9] A. Lu, Y. Ke, J. Cheng and W. Ng, "Mining Vague association rules", In DASFAA, (2007), pp. 891-897.
- [10] K. T. Atanassov, "Intuitionistic Fuzzy Sets", Theory and Applications (Studies in Fuzziness and Soft Computing), Springer-Verlag Telos, (1999).
- [11] K. R. Pardasani and A. Pandey, "A Model for Vague association rule Mining in Temporal Database", in Journal of Information and Computing Science, ISSN 1746-7659, vol. 8, (2013), pp. 063-074.
- [12] K. R. Pardasani and A. Pandey, "A Model for Mining Course Information Using Vague Association Rule", In International Journal of Computer Applications, ISSN 0975-8887, vol. 58, (2012) November 2012.
- [13] V. Badhe and T. S. Arvind, "Comparative Analysis of Fuzzy, Rough, Vague and Soft set Theories in Association Rule Mining", in International Journal of Scientific Progress and Research (IJSPR), ISSN 2349-4689, vol. 2, (2014) 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, vol. 9, no. 7, (2010), pp. 1278-1293.
- [15] Y. Li, D. L. Olson and Z. Qin, "Similarity measures between intuitionistic fuzzy (vague) sets: A comparative analysis", Elsevier, Pattern Recognition Letters, vol. 28, (2007), pp. 278-285.
- [16] V. Badhe, R. S. Thakur and G. S. Thakur, "Vague set theory for profit pattern and decision making in uncertain data", In International journal of advanced computer science and applications, vol. 6, no. 6, (2015).
- [17] T. Bilgic and I. B. Turksen, "Measurement of membership functions: Theoretical and experimental work", In Dubois D., Prade H., eds.: Fundamental of fuzzy sets, the handbook of Fuzzy Set Series. Kluwer Academic Publishers, Boston, MA, (2000), pp. 195-230.
- [18] P. Grzegorzewski, "The coefficient of concordance for Vague data", Elsevier, Computational Statistics & Data Analysis, vol. 51, (2006), pp. 314-322.
- [19] Q. Shen and Y. Jiang "Fuzzy sets, Rough Sets and Vague sets", 3rd International Conference on Advanced Computer theory and Engineering (ICACTE), vol. 2, (2010), pp. 461-465.
- [20] L. Feng, T. Li, D. Ruan and S. Gou, "A Vague-Rough Set Approach for Uncertain Knowledge Acquisition", Elsevier, Knowledge-Based System, vol. 24, (2011), pp. 837-843.
- [21] P. K. Das and B. Chetia, "Application of Vague Soft Sets in student's evaluation", Pelagia Research Library, Advances in Applied Science Research, ISSN: 0976-8612011, vol. 2, Issue 6, (2011), pp. 418-423.

Authors



Arun Kumar Singh, is M. Tech. Scholar in Information Technology at Madhav Institute of Technology and Science (MITS), Gwalior, M.P. (India). He completed his Bachelor of Engineering in 2011 in Computer Science & Engineering from Institute of Technology and Management (ITM), Gwalior, M.P. (India). His research area includes data mining, databases, fuzzy system.



Akhilesh Tiwari, He has received 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 theses at Master and Under Graduate level. His area of current research includes Knowledge Discovery in Databases and Data Mining, Wireless Networks. He has published more than 60 research papers in the journals and conferences of international repute. He is also acting as a reviewer & member in editorial board of various international journals. He is having the memberships of various Academic/Scientific societies including IETE, CSI.

