A Survey on Hesitation Information Mining

Surabhi Pathak
M. Tech. (CSE)
Madhav Institute of Technology & Science
Gwalior, MP, India

Akhilesh Tiwari
Department of CSE & IT
Madhav Institute of Technology & Science
Gwalior, MP, India

ABSTRACT

The tremendous advancement in technology has given rise to an increasing requirement for the storage of data in files, databases and other data repositories. As a result, decision-makers are required to use new and powerful automated tools for the purpose of the analysis and interpretation of the stored data, as well as for the extraction of interesting patterns in data. Since the stored data is not always exact and precise, some means are required to handle this aspect of data and extract the useful information (e.g. hesitation information) arising from such uncertainties. For this purpose vague set theory has been applied for efficiently modeling the uncertainties that occur in datasets. Vague sets are an extension of the classical set theory. They extend the application of set theory to vague and uncertain problems. Vague set theory has emerged as a new tool to deal with the uncertainties of the data and the parameters attached to the data. This theory exhibits a very promising approach to analyzing uncertain data and deriving some interesting results suitable for use in various applications. This paper discusses the notions of vague sets and vague association rules. Also, optimization techniques have been introduced that would be helpful in optimizing the outcomes. When both are implemented together, a new approach will be created which is expected to generate much improved results. This paper emphasizes on the study of vague sets for a series of applications especially in decision making problems.

Keywords

Fuzzy Sets, Optimization Techniques, Rough Sets, Soft Sets, Vague Sets

1. INTRODUCTION

The term data mining fundamentally refers to extracting knowledge from huge extents of information [1]. The term data mining is aimed at exploiting the data stored in files and databases, so that useful and actionable information can be extracted. One of the typical applications of data mining, as introduced by Agrawal and Srikant (1993), is the supermarket transaction scenario. They called it the Market Basket Analysis [2], which is a modelling technique based upon the theory that if one buys a certain group of items, he/she is more likely to buy another group of items.

The output of the market basket analysis is generally a set of rules that can be used by the decision-maker to make business decisions, for example, decisions related to marketing or product placement. The discovery of such rules, also known as Association Rules (AR), is one of the most important tasks of data mining intended towards decision support. Basically it is the process of finding some relationships that may exist among the various attributes of a huge database. The process of extracting these relationships is termed as Association Rule Mining. These relationships can be represented as an IF–THEN statement.

Association Rule Mining [3] techniques look for interesting associations and correlations among datasets. An association rule is a rule comprised of a probabilistic relationship, with the form $X \Rightarrow Y$ between sets of database attributes, where X and Y are sets of items, and $X \cap Y = \phi$. Given the set of transactions T, the goal is to generate all rules that satisfy certain constraints. These constraints are support and confidence. The support of the rule is the fraction of the transactions in T that satisfy the union of items in X and Y. In other words, it is an evidence of how frequently the items show up in the database. Whereas the probability, measured as the fraction of the transactions containing X also containing Y, is called the confidence of the rule.

Association rules are generated in such a way that they satisfy the user-specified minimum support and confidence. The association rule generation process is typically split up into two different phases:

1. A minimum support threshold is applied to discover all the *frequent item-sets* in a database.

2. A minimum confidence constraint is applied to the frequent item-sets in order to form rules.

While confidence is a measure of the rule's strength, support corresponds to statistical significance. With the help of these constraints, rules are computed from the data and, association rules are calculated with help of probability.

For example, the following association rule represents the information that customers who purchase computers also tend to buy printer at the same time.

Computer $\Rightarrow$ Printer
Support = 20%, Confidence = 80%

Although the traditional association rule mining approach is capable of extracting useful patterns from the data, it has limitations as it fails to capture some of the information in the applications that involve uncertain data. One such application is the scenario of online shopping. In many cases, it is observed that a customer puts some items in his shopping cart, but eventually removes some of those items at the time of checkout. Those items, which can be considered as 'almost sold' items carry hesitation information [4], since customers are hesitating to buy them. The hesitation information of items is valuable knowledge for the design of good selling strategies. However, the conventional approach which involves association rule mining only deals with the items that are sold and pays no attention to the items that are almost sold, and, thus, ignores the hesitation information.

In order to address the limitation in traditional AR mining problem, and to capture the hesitation information of items in transactions, the concept of vague sets has been used. The method used for modeling hesitation information of an item rests on the foundation of vague set theory [5, 6, 7]. The basic advantage of this method is that vague set theory addresses

the drawback of a single membership value in *fuzzy set theory* by applying interval-based membership that captures three kinds of evidence with respect to an object in a universe of discourse: *support*, *hesitation and against* [8]. Using measures like *hesitation and attractiveness* of an item, which are derived from vague membership in vague sets, hesitation information has been successfully modeled [9].

Over the past years, several set theories have been proposed for the purpose of handling the uncertainty and imprecision that exists in databases. A brief introduction of these set theories is given in the section that follows.

2. SET THEORIES FOR MANAGING UNCERTAIN DATA

2.1 Fuzzy Set Theory

Fuzzy set theory proposed by Zadeh [10] is the most successful approach to handling vagueness. It makes use of partial membership of elements to a set which represents the degree to which the element belongs to the set.

Let I be a classical set of objects, called the universe of discourse, where an element of I is denoted by x. A fuzzy set $A = \{< x, \mu_A(x) > | x \in I\}$ in a universe of discourse I is characterized by a membership function, μ_A, as follows:

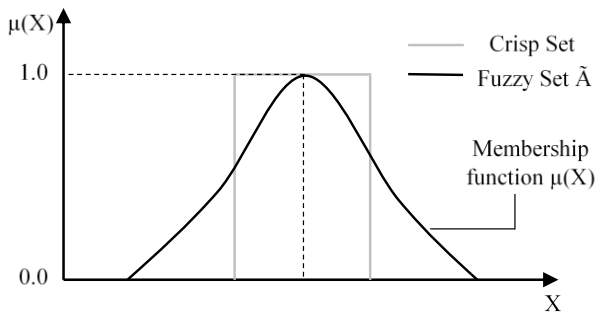$$\mu_A : I \rightarrow [0, 1].$$

Fig 1: Membership Function of a Fuzzy Set

2.2 Rough Set Theory

Rough set theory which was given by Z. Pawlak [11] is another powerful tool for dealing with the imprecision and uncertainty arising from inexact, noisy, or incomplete information.

In rough set theory, sets are defined by approximations instead of membership function. The notion of rough sets is based on the assumption that, unlike traditional set theory, there exists some additional information about elements of a universe of discourse. Elements that reveal the same information are indiscernible (similar) and form blocks that can be seen as elementary granules of knowledge about the universe. These granules, also known as elementary sets, can be considered as elementary building blocks of knowledge. Any union of elementary sets is referred to as a crisp set, and any other sets are called rough sets.

Let I be a universe for a finite set of objects and R be an indiscernibility relation or equivalence relation over the universe I, i.e. $R \subseteq I \times I$. The indiscernibility relation R represents our lack of knowledge about elements of I. Let X be a subset of I. The aim is to characterize the set X with respect to R. For this purpose, rough set theory incorporates the concept of approximations and boundary regions.
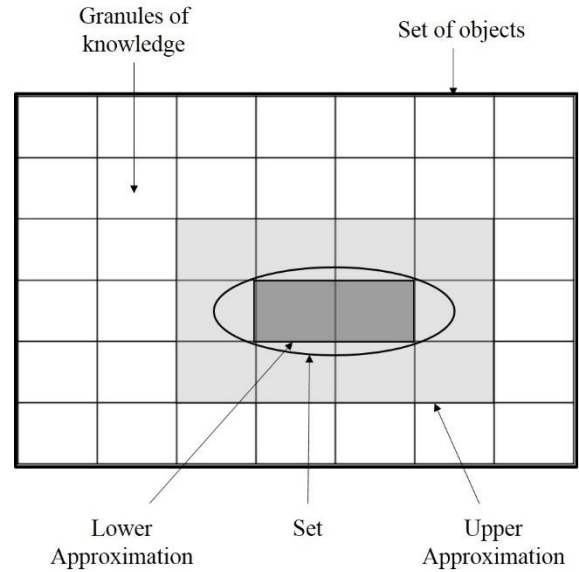

Fig 2: Rough Set [11]

- The lower approximation of a set X with respect to R is the set of all objects, which can be certainly classified as X with respect to R.

- The upper approximation of a set X with respect to R is the set of all objects which can be possibly classified as X with respect to R.

- The boundary region of a set X with respect to R is the set of all objects, which can be classified neither as X nor as not-X with respect to R.

Thus, we can say that a set is rough (imprecise) if it has nonempty boundary region; otherwise the set is crisp (precise).

2.3 Soft Set Theory

The idea of soft sets was initially proposed by Molodtsov [12] as a completely new mathematical approach for processing uncertain data. Soft set is defined as a parameterized family of subsets of a universe set where each element is considered as a set of approximate elements of the soft set.

Let I be an initial universe set and E a set of parameters or attributes with respect to *I*. Let $P(I)$ denote the power set of I and $A \subseteq E$. A pair *(F, A)* is called a soft set over I, where F is a mapping given by $F : A \rightarrow P(I)$. In other words, a soft set *(F, A)* over I is a parameterized family of subsets of I. For $e \in A$, *F(e)* may be considered as the set of e-elements or e-approximate elements of the soft sets *(F, A)*. Thus *(F, A)* is defined as

$$(F,A) = \{F(e) \in P(I) : e \in E, F(e) = \varnothing \text{ if } e \notin A\}.$$

2.4 Vague set Theory

Gau and Buehrer [7] introduced the notion of vague sets. A vague set V in a universe of discourse I is characterized by a true membership function α_V, and a false membership function β_V, where

$$\alpha_V : I \rightarrow [0, 1],$$

$$\beta_V : I \rightarrow [0, 1], \text{ and}$$

$$\alpha_V(x) + \beta_V(x) \leq 1.$$

Here $\alpha_V(x)$ 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 negation of grade of membership of x derived from the evidence against x. Suppose I = {x$_1$, x$_2$, …, x$_n$ }.

Let *I* be a set of objects, called the universe of discourse, where an element of *I* is denoted by *x*.
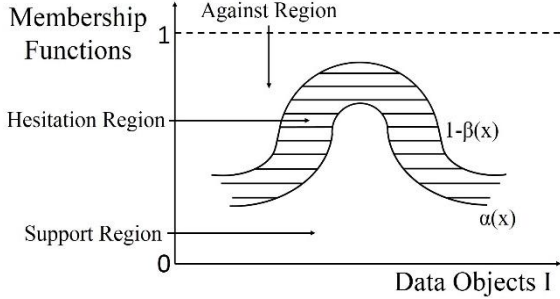

Fig 3: The true (α) and False (β) Membership Functions of a Vague set [8]

The grade of membership of x is represented by the interval $[\alpha_V(x); 1- \beta_V(x)]$, which is a subinterval of [0, 1] as depicted in the Fig. 3. 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*.

3. BASIC TERMINOLOGY RELATED TO VAGUE SETS AND HESITATION INFORMATION

3.1 Hesitation and Overall Hesitation

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

This is directly implied from the above definition that $H(x) = 1 - \alpha(x) - \beta(x)$.

3.2 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 $int(x, s_i)$ is a vague value $[\alpha_i(x), 1 - \beta_i(x)]$ which is sub interval of $[\alpha(x), 1 - \beta(x)]$. The overall intent of x denoted as $INT(x)$ is the interval $[\alpha(x), 1 - \beta(x)]$.

3.3 Attractiveness and Overall Attractiveness

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

Given the intent $[\alpha_V(x); 1- \beta_V(x)]$ of an item *x* for an HS s_i, we have a one-to-one corresponding pair of the attractiveness and hesitation, known as *AH*-pair, denoted as *[att(x; s$_i$); h(x)]*. Attractiveness and hesitation are two significant ideas, since one might be interested in gaining ARs with the items having high hesitation or high attractiveness.

3.4 Weighted Attributes

The variables selected to calculate weight are known as weighted attributes $A(a_1, a_2, \ldots, a_n)$. Depending on the domain, it could be any kind of variable such as the item weight in case of supermarket domain.

3.5 Item Weight

Item weight is the value attached to each item representing its significance. In case of supermarket transactions, it can represent the profit associated with the sale of a certain item. The item weight is a function of the weighted attribute. If item weight is represented as $w(i)$ then $w(i) = f(a)$.

3.6 Itemset Weight

Weight of an itemset is the weight of the enclosing items. It can be denoted as $w(is)$. The item weight is a special case of the itemset weight, when the itemset has only one item. The average value of the item weight is given by

$$w(is) = \frac{\sum_{k=1}^N w(i_k)}{N}$$

3.7 Transaction Weight

Transaction weight is a type of an itemset weight that is attached to all transactions. Higher the transaction weight, larger is the contribution of the transaction in mining results.

3.8 AH-pair Transaction and Database

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

4. VAGUE ASSOCIATION RULE MINING

In this section, the concept of Hesitation Statuses (*HSs*) is shown and modeling of hesitation information is discussed. Then the notion of VARs as well as four different types of confidence and support used to evaluate the quality of VARs have been discussed.

4.1 Hesitation Information Modeling

While purchasing an item during a transaction, a particular state between two different situations of "buying" and "not buying" is termed as a *Hesitation Status (HS)*.

In order to capture the hesitation evidence, a subinterval of $[\alpha_V(x); 1- \beta_V(x)]$ is used to represent the customer's *intent* for each item with respect to various HSs. To find the intent value, the linear extensions concept of a partial order is used.

4.2 Vague Association Rule

A Vague Association Rule (VAR), $r = (X \Rightarrow Y)$, is an association rule obtained from an AH-pair database.

Based on attractiveness and hesitation of an item with respect to a hesitation status (HS), various types of confidence and support of a VAR can be described. *Attractiveness-Hesitation* confidence and *Attractiveness-Hesitation* support of a VAR have been defined in order to evaluate the quality of the VAR. Accordingly, four kinds of support and confidence have been described to evaluate the VARs as follows.

4.3 Support

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

1. The A-support of Z,

$$Asupp(Z) = \frac{\sum_{T \in D} \prod_{z \in Z} M_A(z)}{|D|}$$

2. The H-support of Z,

$$Hsupp(Z) = \frac{\sum_{T \in D} \prod_{z \in Z} M_H(Z)}{|D|}$$

3. The AH-support of Z,

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

4. The HA-support of Z,

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

4.4 Confidence

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

1. If both X and Y are A frequent-itemsets, then the confidence of r, called the A-confidence of r is defined as

$$Aconf(r) = \frac{A_{supp}(Z)}{A_{supp}(X)}.$$

2. If both X and Y are H frequent-itemsets, then the confidence of r, called the H-confidence of r is defined as

$$Hconf(r) = \frac{H_{supp}(Z)}{H_{supp}(X)}.$$

3. If X is an A frequent-itemset and Y is an H frequent-itemset, then the confidence of r, called the AH-confidence of r is defined as

$$AHconf(r) = \frac{AH_{supp}(Z)}{A_{supp}(X)}.$$

4. If X is an H FI and Y is an A frequent-itemsets, then the confidence of r, called the HA-confidence of r is defined as

$$HAconf(r) = \frac{HA_{supp}(Z)}{H_{supp}(X)}.$$

5. LITERATURE REVIEW

An Lu, Yiping Ke et al. [8] applied the theory of vague sets in the context of AR mining as to include the hesitation information into ARs. Concepts like *attractiveness* and *hesitation* have been introduced, which characterize the overall information of a customer's intent on an item. Depending on these two concepts, the notion of vague association rules (VARs) has been proposed and an algorithm to mine the VARs has been designed. Experiments demonstrate that the algorithm is effective and the VARs capture more exact and better information as compared to conventional ARs.

Bernecker et al. [9], and Q. Zhang et al. [13] examined that employing predictable support may perhaps cause the loss of some important patterns. For computing the probability that a pattern is frequent, they presented the notion of Probabilistic Frequent Itemsets (PFI). Dynamic programming based solutions were previously being used to regain PFIs from vague databases. This algorithm computed probabilities, and validated that an itemset was a PFI in $O(n^2)$ time. However, the proposed algorithm avoids the use of dynamic programming, and is capable of validating a PFI much faster, i.e. in O(n) time.

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

An Lu et al. [16] devised an algorithm for the issue, given a vague relation *r* over a schema *R* and FDs *F* over *R* set, what is"best" approximation with respect to *F* when taking into median membership (*m*) and the imprecision membership (*i*) thresholds account. Employing these two vague set thresholds, defined mi-overlap among vague sets notion and a merge operation on *r*. FD satisfaction in *r* is defined in values terms being mi-overlapping. Presenting Lien's and Atzeni's axiom system is sound and complete for FDs being satisfied in the vague relations. Chase procedure is studied for a vague relation *r* over *R*, named VChase(*r*, *F*), as a means to maintain consistency of *r* with respect to *F*. The main outcomes is that the output of the procedure is the most object-precise approximation of *r* with respect to *F*. The complexity of VChase(*r*, *F*) is polynomial time in the sizes of *r* and *F*.

A. Pandey et al. in "A Model for Mining Course Information using Vague Association Rules" [17] have proposed an algorithm for mining vague association rule that discovers the hesitation information of items. The algorithm was devised to mine the courses and the hesitation of students to attend the courses. Using experiments on real datasets they have proved their algorithm to be effective in mining vague association rules. In contrast to the conventional association rules mined from transactional databases, vague association rules mined from the AH-pair databases have been found to be more detailed and accurate, and are capable of capturing better information. Later in their work, "A Model for Vague Association Rule Mining in Temporal Databases", [18] A. Pandey et al. have proposed an extension of their algorithm for mining temporal association rules and analyzing the obtained hesitation information which can be utilized in making the courses more efficient and attractive.

B. Y. Chilakalapudi et al. [19] presented an algorithm for extracting frequent itemsets from a huge vague database, interpreted under PWS. They have suggested that the mining procedure can be modeled as a Poisson binomial distribution, and have designed an algorithm is implemented which can effectively and accurately determine frequent itemsets in a large vague database. The devised mining algorithm facilitate PFI outcomes to be re-energized. This lessens the requirement of re-executing the entire mining algorithm over the fresh database, which is often more costly and redundant. The devised algorithm can maintain incremental mining and provides precise outcomes on mining the vague database. A broad estimation on real datasets has been performed in order to prove the effectiveness of the proposed algorithm.

Priya Bajaj et al. [20] proposed an algorithm that enhances the efficiency of web access by utilizing the next demanded web page prior to the user requests. It is an intelligent predictive mining that examines the user' web access history and predicts the next page. Here vague enhanced Markov model is devised to execute the prediction. Suggestions have been made for vague rules to carry out the pruning at distinct levels of Markov model. When the prediction table is formed, the association mining will be implemented to discover the efficient next page. This model enhances the correctness and efficiency of the prediction.

Vivek Badhe et al. [21] proposed a model for profit pattern mining. They have demonstrated the difficulty in making decisions, particularly in financial problems which is a critical job in industry. Although profit pattern mining serves the purpose, this scheme relies on the inaccurate and vague surroundings. Most of the approaches to association rule mining concentrate on inferring rules by means of associations between data and searching frequent patterns that exist in the data. The key procedure makes use of support and confidence measures for forming rules. But as the data has turned out to be more composite today, it is necessary to search solutions which aid in dealing with these issues.

The recent literature survey indicates that there is a requirement of optimizing the extracted results further, so that the final outcomes are more improved and accurate. Several optimization techniques have already been successfully applied in various fields of data mining such as classification and clustering. Some of these techniques have been discussed in the following section.

6. OPTIMIZATION ALGORITHMS
6.1 Genetic algorithm (GA)
GA is an algorithm of adaptive heuristic search which is based on the concepts of genetics and natural selection. These algorithms signify an intelligent random exploitation search used to resolve improvement issues. Although randomized, Genetic Algorithms are by no means random, instead they exploit historical data to direct the search into a region of improved execution inside the search space.

The GA is a technique for resolving both unconstrained and constrained optimization issues that which are similar to the process of natural selection. The genetic algorithm then simulates the survival of the fittest individual among individuals over consecutive generations for the purpose of problem solving [22, 23].

The genetic algorithm proceeds in the following manner:

Selection of initial population

Evaluation of the fitness of the individual

Defining the average fitness of the population

Repeat

    Best suited individuals are selected for reproduction

    Mating is performed at random

    Crossover is applied

    Mutation is performed

    Evaluation of the fitness of all individuals

    Determining average fitness of the population until terminating condition is reached

6.2 Artifcial Bee Colony (ABC) Algorithm
Artificial Bee Colony (ABC) algorithm is a problem solving method, developed based on the behaviors of honey bee colony, searching and sharing the information with other colony members in the hive, to be able to find out richest food sources in shortest possible time [24, 25, 26, 27].

A possible solution in the problem is represented by a position of a food source in nature. The amount of nectar in that food source represents the fitness value of that solution.

There are three types of bees in the ABC algorithm.

6.2.1 Employed Bees
Every employed bee works on a food source. Position of this food source symbolizes a possible solution. An employed bee calculates the fitness of this solution and saves the position information in its memory. Number of employed bees in the hive "N_e", represents the number of solutions in the algorithm in one step of iteration.

6.2.2 Onlooker Bees
The bees that are waiting in the hive receive information about the position of food sources from employed bees. Each onlooker bee selects a food source to exploit, depending on the nectar amount. This selection is done by modeling the nectar amount as a probability (P_q, probability of qth employed bee being selected by onlookers). Higher the amount of nectar, higher is the probability that the employed bee is selected. After this selection, each onlooker bee searches a new position near the selected food source (location of the employed bee).

6.2.3 Scout Bees
A bee which searches for new food sources without any information is called scout bee. They randomly search the whole environment. A scout bee becomes an employed bee when it starts to work on a food source.

6.3 Animal Migration Optimization (AMO)
In the behavior of animal ecology, migration is a widespread phenomenon that has been studied intensively. Animal migration is a comparatively long-distance movement of individuals, typically on a seasonal basis. It is a common phenomenon found in most of the animal groups, for example mammals, birds, fish, reptiles, amphibians, and insects. The trigger for the migration may be local climate, local food availability, the season of the year, and so on [28].

Animal migration optimization simulates the migration phenomenon of animals through variation of position, substitution of individuals with one another, and gradually arriving at the optimal solution. Studies have shown that every bird changes its position related to the six or seven animals around it, no matter how close or how far away those animals are. Interactions between flocks are thus based on a topological rule rather than a metric rule.

The AMO algorithm is divided into two parts.

6.3.1 Animal Migration Process
At the time of the animal migration process, an animal should follow three rules:

- Avoid collisions through your neighbors;
- Move in a similar direction as your neighbors; and
- Remain close to the neighbors.

The idea of local neighborhood of an individual is described through the use of topological ring. For simplicity, the neighborhood length is set to five for each individual dimension. In this algorithm the neighborhood topology is static and is described on the set of indices of vectors.

If the animal index is j, then its neighborhood contains of animal having indices j−2, j−1, j, j+1, j+2, if the animal index is 1, the neighborhood consists of animals having indices $NP−1$, NP, 1, 2, 3, and so forth. Once the topology of the neighborhood has been constructed, we choose one neighbor randomly and also update the position of the individual according to this neighbor, as can be seen in the following formula:

$$X_{i,G+1} = X_{j,G} + \delta.(X_{neibor\ hood,G} - X_{j,G})$$

Where, $X_{neibor\ hood,G}$ is the neighborhood's present position, δ is created by applying a random number generator controlled through a Gaussian distribution, $X_{j,G}$ is the current position of the ith individual, and $X_{j,G+1}$ is the new position of the ith individual.

6.3.2 Population Updating Process

At the time of population updating process, the algorithm simulates how some of the animals leave the group and some join the new population. Individuals will be changed through some novel animals with a probability P_a. The probability is used according to fitness quality. We sort the fitness in descending order, so the probability of the individual with best fitness is $1/NP$ and the individual with worst fitness, through contrast, is 1.

After creating a novel solution $X_{i,G+1}$, it will be compared and evaluated with $X_{i,G}$, and the individual with the higher objective fitness will be selected.

$$X_i = \begin{cases} X_{i,} & if\ f(X_{i,})is\ better\ tHan\ f(X_{i,G+1}), \\ X_{i,G+1} & \text{otherwise} \end{cases}$$

6.4 Particle Swarm Optimization (PSO)

PSO is an optimization technique based on the stochastic optimization method developed by Dr. Eberhart and Dr. Kennedy. PSO is based on the social behavior of a fish school or bird flock [29].

PSO shares numerous similarities with evolutionary computing methods such as genetic algorithm. An initial random population of possible solutions is taken and the system searches the population for an optimal solution by updating generations. However, in contrast with GA, PSO does not involve any evolution operators, for example mutation and crossover. In PSO, potential solutions, or particles, fly through the problem space following the present optimum particles.

Every particle updates itself by following two "best" values. The particle keeps track of its coordinates in the problem space which are associated with the best solution (fitness) it has attained until now. This value is stored, and is known as pbest. The second "best" value that PSO tracks is the best value that has been found out so far by any particle in the population, which is a global best and is denoted by gbest. When a particle considers only its topological neighbors, the best value is a local best and is known as lbest.

The basic idea of PSO consists of altering/ accelerating the velocity of all the particles toward the pbest and lbest locations (PSO local version). Acceleration is weighted through a random term, with separate random numbers being produced for acceleration toward lbest and pbest locations.

In the recent years, PSO has been effectively applied in numerous research and application areas. It has been demonstrated that PSO gets higher results in a faster, cheaper way compared to other approaches. PSO offers certain advantages over genetic algorithm as well. The implementation of PSO is easier as compared to GA. In addition, a relatively smaller number of parameters is required to adjust in PSO. Thus, the same version, with slight variations, works well in an extensive variety of applications.

6.5 Ant Colony Optimization (ACO)

ACO is a technique that simulates the behavior of ants as social insects that work together to accomplish a general aim with the use of crowd wisdom. ACO algorithms put swarm intelligence into particular action. Swarm intelligence, which is based on the concept of collective behavior, has occupied ACO in numerous fields and domains of problem solving. ACO has been applied successfully in several domains of data mining, such as classification and clustering, and has provided scalable solutions.

ACO is a collection of algorithms of Artificial Intelligence that rely on imitation of social insects' behavior, especially ants'. These algorithms use agents, that we call "ants", for investigation of the best solution to a problem, such as shortest path between few places that might be food for colony, just like the case with real world ant colonies. These agents are iteratively construct problem solutions. The probability for an ant to visit a town is depends upon the quantity of pheromone that all agents detect at the time of its exploration. Pheromone is a substance that ants create and deposit along paths that they have traversed, creating those paths more attractive for the next ones that might pass through the same point, while previously existent pheromones vapourize as time passes. At the time of algorithm improvement, artificial pheromone is placed after complete tour-solution construction on each and every town that was selected and visited for its construction. In this way, the pheromone amount is the heuristic data at a given point of time, reflecting colony experience about feasible solutions of the problem.
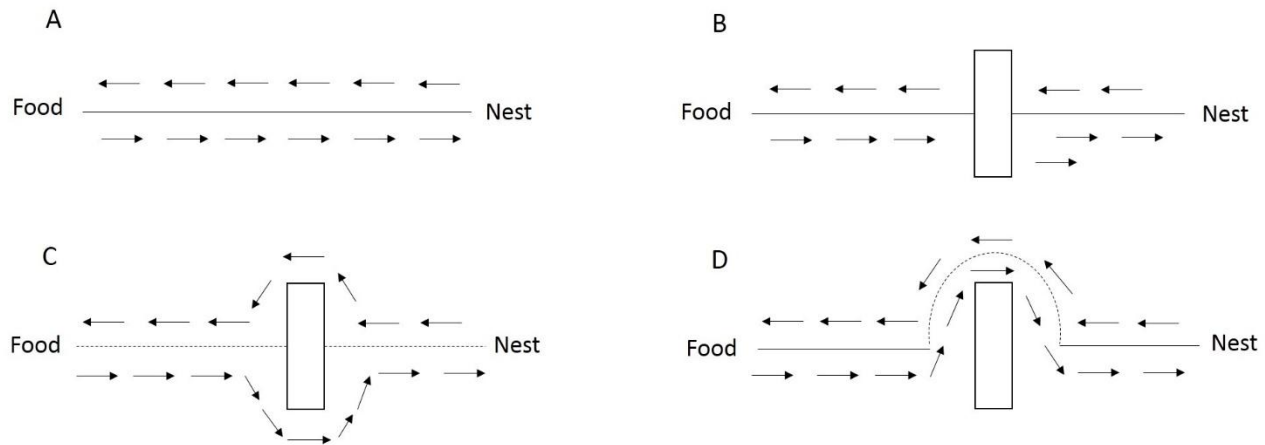
Fig 4: Ant Colony Optimization

7. CONCLUSION

Association rule mining is a crucial task of the data mining process. However, the conventional process of association rule mining has some drawbacks since it assumes that the available information is exact and precise, whereas most of the data in the real world scenario is much more complex. This makes the conventional association rule mining process insufficient to handle such data that is vague, uncertain, incomplete and imprecise.

Vague sets have been successfully applied to manage the uncertainty in datasets. Various implementations of vague set theory for mining different types of vague information have been discussed in this paper. Moreover, this paper introduces several techniques that can be applied to the rules obtained after mining the hesitation information which will further optimize the result.

8. REFERENCES

[1] U. M. Fayyad, G. P. Shapiro and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," 0738-4602-1996, AI Magazine (Fall 1996). pp: 37–53.

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

[3] R. Agrawal, R. Srikant, "Fast algorithms for mining association rules". In: Proceedings of the 20th International Conference on Very Large Databases, Santiago, Chile, 1994.

[4] An Lu and Wilfred Ng, "Mining Hesitation Information by Vague Association Rules". 2007© Springer.

[5] A. Lu, W. Ng, "Managing merged data by vague functional dependencies," ER. (2004) 259–272.

[6] A. Lu, W. Ng, "Vague sets or intuitionistic fuzzy sets for handling vague data: Which one is better," 2005 Springer.

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

[8] An Lu, Yiping Ke, James 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.

[9] T. Bernecker, H. Kriegel, M. Renz, F. Verhein, and A. Zuefle, "Probabilistic Frequent Itemset Mining in Uncertain Databases," Proc. 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2009.

[10] L. Zadeh, "Fuzzy Sets", © 1965 Elsevier Inc.

[11] Z. Pawlak, "Rough sets", 1982 International Journal Computer Information Science 11 (5) 341–356.

[12] D Molodtsov, "Soft Set Theory – First Results", © 1999 Elsevier Inc.

[13] Q. Zhang, F. Li, and K. Yi, "Finding Frequent Items in Probabilistic Data," Proc. ACM SIGMOD International Conference on Management of Data, 2008.

[14] An Lu and Wilfred Ng, "Vague Sets or Intuitionistic Fuzzy Sets for Handling Vague Data- Which One Is Better?" 2005 © Springer.

[15] An Lu, Yiping Ke, James Cheng, and Wilfred Ng, "Mining Vague Association Rules" 2007 © Springer.

[16] An Lu and Wilfred Ng, "Handling inconsistency of vague relations with functional dependencies," Proceedings of the 26th international conference on Conceptual modeling, Pages 229-244, Springer-Verlag Berlin, Heidelberg ©2007.

[17] A. Pandey and K.R. Pardasani, "A Model for Mining Course Information using Vague Association Rule", International Journal of Computer Applications (0975 – 8887) Volume 58– No.20, November 2012.

[18] A. Pandey and K.R. Pardasani, "A Model for Vague Association Rule Mining in Temporal Databases," Journal of Information and Computing Science Vol. 8, No. 1, 2013, pp. 063-074, ISSN 1746-7659, England, UK.

[19] 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 (0975 – 8887) Volume 73– No.12, July 2013.

[20] P. Bajaj and S. Raheja, "A Vague Improved Markov Model Approach for Web Page Prediction," International

Journal of Computer Science & Engineering Survey (IJCSES) Vol.5, No.2, April 2014.

[21] V. Badhe et al., "Vague Set Theory for Profit Pattern and Decision Making in Uncertain Data", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 6, 2015.

[22] H. De. Garis, "Genetic programming: building artificial nervous systems using genetically programmed neural network modules," Proc. of the 7th international conference on machine learning. 132-139. Morgan Kaufmann 1990.

[23] John J. Grefenstette, "Optimization of control parameters for genetic algorithms," IEEE Transaction on Systems, Man, Cybernetics, SMC-16(1):122-128, 1986.

[24] C. Zhang, J. Ning and D. Ouyang, "An artificial bee colony approach for clustering," Expert Systems with Applications, vol. 37, (2010), pp. 4761-4767.

[25] D. Karaboga and C. Ozturk, "A Novel clustering approach: Artificial bee colony (ABC) algorithm," Applied Soft Computing, vol. 11, (2011), pp. 652-657.

[26] X. Yan, Y. Zhu, W. Zou and L. Wang, "A new approach for data clustering using hybrid artificial bee colony algorithm", Neural computing, vol. 97, (2012), pp. 241-250.

[27] D. Karaboga, C. Ozturk, N. Karaboga, B. Gorkemli, "Artificial bee colony programming for symbolic regression". Information Science 209:1–15 (2012).1

[28] X. Li, J. Zhang, and M. Yin, "Animal migration optimization: an optimization algorithm inspired by animal migration behavior," Neural Computing and Applications, 2013.

[29] J. Kennedy and R. C. Eberhart, "Particle Swarm Optimization," Proc. IEEE int'l conf. on neural networks, Vol. 4, pp. 1942-1948. IEEE service center, Piscataway, NJ, 1995.