INTEGRATION OF SELF ORGANIZING FEATURE MAPS AND HONEY BEE MATING OPTIMIZATION ALGORITHM FOR MARKET SEGMENTATION

Babak Amiri and Mohammad Fathian
Department of Industrial Engineering, Iran University of Science and Technology
Email: Amiri_Babak@ind.iust.ac.ir

ABSTRACT

This study is dedicated to proposing a two-stage method, which first uses Self-Organizing Feature Maps (SOM) neural network to determine the number of clusters and cluster centroids, then uses honey bee mating optimization algorithm based on K-means algorithm to find the final solution. The results of simulated data via a Monte Carlo study show that the proposed method outperforms two other methods, SOM followed by K-means (Kuo, Ho & Hu, 2002a) and SOM followed by GAK (Kuo, An, Wang & Chung, 2006), based on both within-cluster variations (SSW) and the number of misclassification. In order to further demonstrate the proposed approach's capability, a real-world problem of an internet bookstore market segmentation based on customer loyalty is employed. The RFM model is used for comparison of customers' loyalty. Then the proposed method is used to cluster the customers. The results also indicate that the proposed method is better than the other two methods.

Keywords: *Clustering, Self organizing feature map, Honey bee mating, Market segmentation.*

1. INTRODUCTION

Market segmentation has been mentioned as an important and avenue of research in the field of electronic commerce (Chang, 1998). In this new and competitive commercial framework, market segmentation techniques can give marketing researchers a leading edge: because the identification of such segments can be the basis for effective targeting and predicting of potential customers (O'Connor and O'Keefe, 1997).
Among the segmentation methods, the post-hoc methods, especially clustering methods are relatively powerful and frequently used in practice (Dillon et al., 1993; Wedel and Kamakura, 1998). Among clustering methods, the K-means method is the most frequently used, since if can accommodate the large sample sizes associated with market segmentation studies (Anil et al., 1997). Due to increasing computer power and decreasing computer costs, artificial neural networks (ANNs) have been recently applied to a wide variety of business areas (Vellido et al., 1999), such as market segmentation (Balakrishnan, Cooper & Jacob, 1996; Kuo et al., 2002a, b), sales forecasting (Kuo & Xue 1998), and bankruptcy prediction (Cadden, 1991). One type of unsupervised neural networks, the Self-Organizing Feature Maps (SOM), can project high-dimensional input space on a low-dimensional topology, allowing one to visually determine out the number of clusters (Lee et al., 1977; Pykett, 1978).

Besides, over the last decade, modeling the behavior of social insects, such as ants and bees, for the purpose of search and problem solving has been the context of the emerging area of swarm intelligence. Honey-bees are among the most closely studied social insects. Honey-bee mating may also be considered as a typical swarm-based approach to optimization, in which the search algorithm is inspired by the process of marriage in real honey-bee.

Bozorg Haddad and Afshar (2004) presented an optimization algorithm based on honeybee mating that successfully applied to a single reservoir optimization problem with discrete decision variables. Later, Bozorg Haddad et al (2005) applied the same algorithm to three benchmark mathematical problems.
Honey-bee has been used to model agent-based systems Afshar et al (2006) developed an optimization algorithm based on the honey-bee marriage process.

In (Afshar et al, 2006) the honey bee mating optimization algorithm is presented and tested with a nonlinear, continues constrained problem with continues decision and state variables to demonstrate the efficiency of the algorithm in handling the single reservoir operation

optimization problems. They showed that the performance of the model is quite comparable with the result of the well-developed traditional linear programming solvers.

In this paper, we proposed application of honeybee mating optimization in clustering (HBMK-means). Therefore, this research proposes a two-stage method. It first determines the number of the cluster and cluster centroids via the Self-Organizing Feature Maps (SOM) neural network. Then honey bee mating optimization algorithm is presented and applied to find the final solution (defined as SOM+HBMK in this paper). It is compared with the other two methods, including SOM followed by K-means (Kuo et al., 2002a) and SOM followed by GAK (SOM+GAK) (Kuo, An, Wang & Chung, 2006), via a Monte Carlo simulation (Milligan & Cooper, 1985). The simulated results show that SOM+HBMK provides better performance than the other two methods, based on both within-cluster variations (SSW) and number of misclassifications.
In order to further demonstrate the feasibility of the proposed approach, a real-world problem, the market segmentation of an internet bookstore, is employed based on customer loyalty. In order to segment the customers based on loyalty, the RFM model variables used as the basis for market segmentation. Then, SOM+HBMK is utilized as clustering method for market segmentation. The computational results also indicate that SOM+HBMK has the smallest SSW compared with the other two methods.

The remainder of this paper is organized as follows. Section Two discusses the general idea of market segmentation, applications of ANNs in marketing segmentation, and meta- heuristic algorithms in clustering analysis. The proposed two stage method is described in Section Three, while Section Four presents the simulation algorithm and results. Finally the real world problem results are detailed in Section Five

2. BACKGROUNDS

This section introduces some necessary backgrounds, including market segmentation as well as applications of neural networks and meta-heuristic algorithms for clustering analysis. Detailed discussion of these topics is in the following subsections.

2.1. Market segmentation

Due to multiple and varied requests from different customers, companies must satisfy the needs of discriminating customers who can choose from a multitude of alternatives in the marketplace. To select its markets and serve them well, many companies are embracing target marketing (Kotler, 1997). Kotler (1997) suggested one common three-step approach to identifying the major segments in a market: (1) Survey Stage, (2) Analysis Stage, and (3) Profiling Stage. Wedel and Kamakura (1998) put forward the following six criteria for determining the effectiveness and profitability of market segmentation: (1) identifiability, (2) substantiality, (3) accessibility, (4) stability, (5) responsiveness, and (6) actionability. They also summarized those segmentation methods in the four categories.

2.2. Applications of Artificial Neural Networks for Market Segmentation

The learning algorithms of ANNs can be divided into two categories: supervised and unsupervised. In supervised learning, the network has its output compared with a known answer, and receives feedback about any errors. The most widely applied unsupervised learning scheme is Kohonen's feature maps. Venugopal and Baets (1994) presented the possible applications of ANNs in marketing management, using three examples, retail sales forecasting, direct marketing and target marketing, to demonstrate the capability of ANNs. Bigus (1996) suggested that ANNs can be employed as a tool for data mining and presented a network with three different dimensions of data: demographic information (sex, age, and marriage), economic information (salary and family income), and geographic information (states, cities, and level of civilization). Balakrishnan, Cooper and Jacob (1994) compared self-organizing feature maps with the K-means method, showing that the K-means method has a higher rate of classification through the Monte Carlo algorithm. Subsequently, Balakrishnan et al., (1996) employed the frequency-sensitive competitive learning algorithm (FSCL) and the K-means method for clustering the simulated data and real-world problem data, presenting a combination of these two methods. Although the neither simulated nor real-world problem data can determine which single method is better, the combination of these two methods seems to provide better managerial explanation for the brand choice data. A modified two-stage method, which first uses the self organizing feature maps to determine the number of clusters and the starting

points and then employs the K-means method to find the final solution, was proposed by Kuo et al (Kuo et al., 2002a) for market segmentation. The simulation results show that the modified two-stage method is slightly more accurate than the conventional two-stage method (Ward's minimum variance method followed by the K-means method) with respect to the rate of misclassification.

2.3. Applications of Meta Heuristic Algorithms on Clustering

One popular class of data clustering algorithms is the center based clustering algorithm. K-means used as a popular clustering method due to its simplicity and high speed in clustering large datasets (Fogy, 1996). However, k-means has two shortcomings: dependency on the initial state and convergence to local optima (Afshar et al, 2006) and global solutions of large problems cannot found with reasonable amount of computation effort (Spath, 1989). In order to overcome local optima problem lots of studies done in clustering.
Maulik and Bandyopadhyay (2000) proposed a genetic algorithm based method to solve the clustering problem and experiment on synthetic and real life datasets to evaluate the performance. The results showed that GA-based method might improve the final output of k-means.
Krishna and Murty (1996) proposed a novel approach called genetic k-means algorithm for clustering analysis. It defines a basic mutation operator specific to clustering called distance-based mutation. Using finite Markov chain theory, it proved that GKA converge to the best-known optimum.

Shokri et al (1991) discussed the solution of the clustering problem usually solved by the K-means algorithm. The problem known to have local minimum solutions, which are usually what the K-means algorithm obtains. The simulated annealing approach for solving optimization problems described and proposed for solving the clustering problem. The parameters of the algorithm discussed in detail and it shown that the algorithm converges to a global solution of the clustering problem.
Sung and Jin (2000) considered a clustering problem where a given data set partitioned into a certain number of natural and homogeneous subsets such that each subset is composed of elements similar to one another but deferent from those of any other subset. For the clustering problem, a heuristic algorithm exploited by combining the tabu search heuristic with two complementary functional procedures, called packing and releasing procedures. The algorithm numerically tested for its electiveness in comparison with reference works including the tabu search algorithm, the K-means algorithm and the simulated annealing algorithm.

Over the last decade, modeling the behavior of social insects, such as ants and bees, for the purpose of search and problem solving has been the context of the emerging area of swarm intelligence. Using ant colony is a typical successful swarm-based optimization approach, where the search algorithm inspired by the behavior of real ants.
KUO et al, (2005) proposed a novel clustering method, ant K-means (AK) algorithm. AK algorithm modifies the K-means as locating the objects in a cluster with the probability, which updated by the pheromone, while the rule of updating pheromone is according to total within cluster variance.

Shelkor et al (2004) present an ant colony optimization, methodology for optimally clustering N objects into K clusters. The algorithm employs distributed agents who mimic the way real ants find a shortest path from their nest to food source and back. They compared result with other heuristic algorithms in clustering, GA, Tabu search, SA. They showed that their algorithms are better than other algorithms in performance and time.

3. METHODOLOGY

After discussing the general backgrounds of neural networks and application of meta-heuristic algorithms on clustering, this section explains the proposed approach for clustering in two stages. The first stage employs the SOM network to determine the number of clusters and cluster centroids; and honey bee mating optimization algorithm is used in this study for finding the final solution in the second stage. For comparison of proposed method with K-means, SOM network followed by GAK-means (Kuo et al., 2006) and SOM network followed by K-means (Kuo et al., 2002a), using the data from Monte Carlo simulation demonstrates the effectiveness of the proposed approach. Finally, this method is utilized for market segmentation of an internet bookstore in Iran, based on customer loyalty. Fig. 1 illustrates the proposed method's structure.

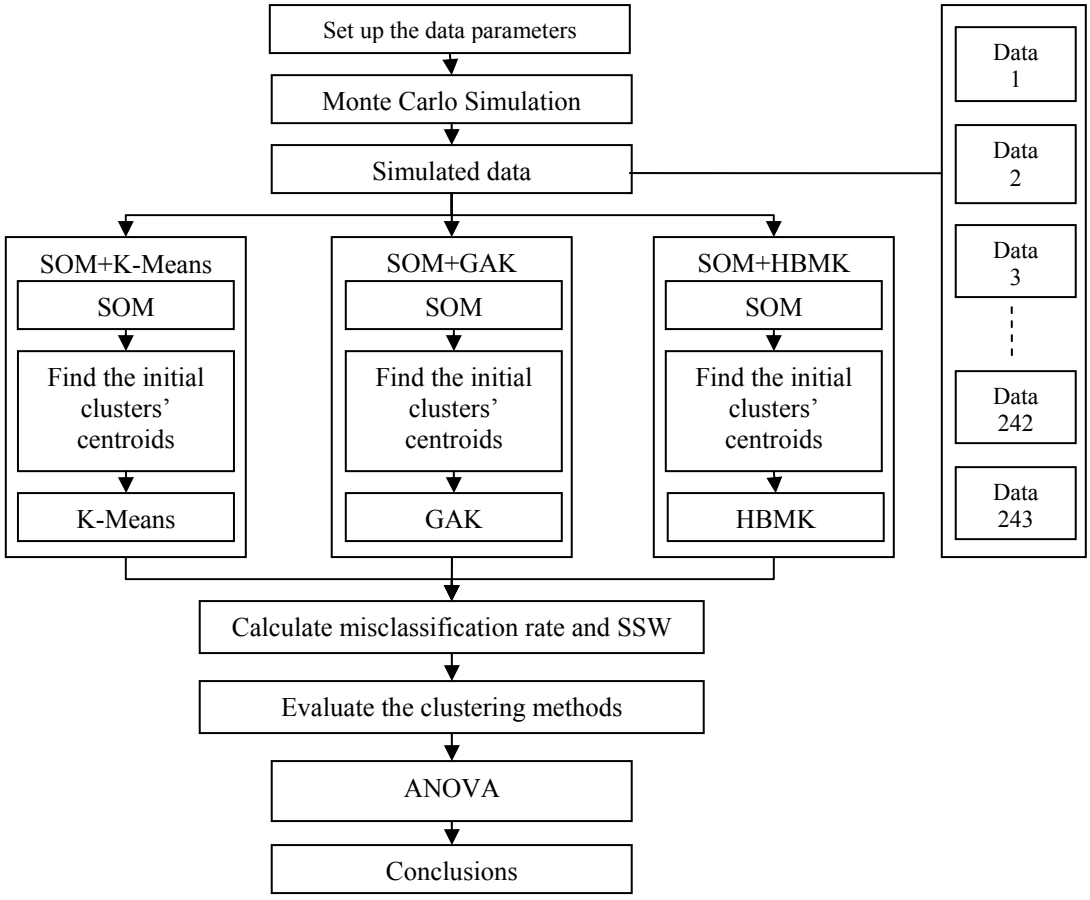

Figure.1. Performance evaluation of the proposed method

3.1. Monte carlo simulation

The data for this research were generated similar to Milligan's and Cooper, (1985) procedure, which has been used in several studies that examine the properties of clustering algorithm (Balakrishnan et al., 1994; 1996; Cowgill et al., 1999). Several factors can affect the quality of cluster recovery, e.g., the number of clusters in the data, the number of dimensions used to describe the data, the density level, and the level of errors in the data (Milligan & 1985). The three levels of density in this research are described as follows.

1. The equal condition: The number of points is as close to equal as possible in each cluster.
2. The 10% condition: One cluster must contain 10% of the data points.
3. The 60% condition: One cluster must contain 60% of the data points.

The simulated data sets for this study contain 3, 5, or 7 distinct non-overlapping clusters. The number of dimensions is varied so that all points in a data set are described by a 6, 8, or 10 dimensional spaces, while the three levels of error are error free, low error and high error, as in Table 1. Therefore, the full factorial design results in 3×3×3×3 combinations, and each experiment is conducted with three replications, so there is a totally 243 data sets, with each set containing 120 data points (Milligan & Cooper, 1985).

Table 1. The experimental design

Factor	I	II	III
Cluster	3	5	7
Dimension	6	8	10
Density	0.1	0.5	0.6
Error	No	Low	High

3.2. Proposed SOM+HBMK algorithm

The proposed method use SOM to find number of clusters and centroids, then use HBMK to get the final solution as shown in Fig. 2. Each of these two stages is explained in more detail in the following subsections.

3.2.1. Self-organizing feature maps (SOM)

The Self-Organizing Feature Map (SOM), which is typical of the unsupervised learning neural networks, can project a high-dimensional input space on a low-dimensional topology so as to allow one to visually determine the number of clusters directly (Lee et al., 1977; Pykett, 1978). The most widely used unsupervised learning scheme is the self-organizing feature maps developed by Kohonen (Kohonen, 1982). The learning rule of adjusted weights is as follows:

$$\Lambda(i,i^*) = \exp\left(-\frac{|r_i - r_{i*}|^2}{2\omega^2}\right) \quad (2)$$

where s is a width parameter that is gradually decreased with every learning cycle.

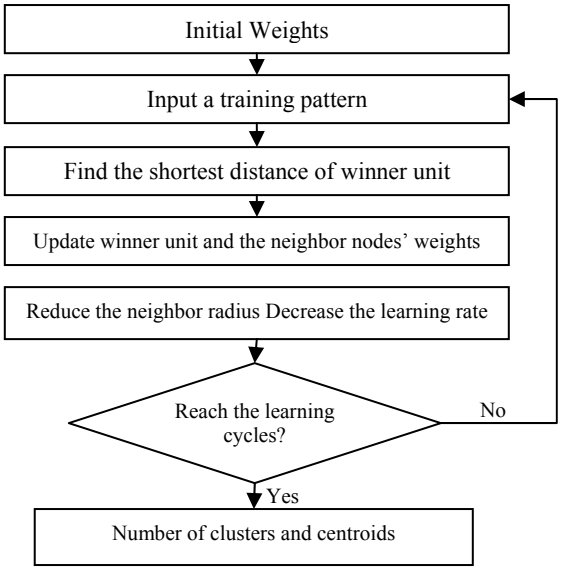

Figure2. SOM algorithm

3.2.2. Application of Honey-bee Mating Optimization Algorithm in Clustering

$$\Delta w_{ij} = \eta \wedge (i,i^*)(\xi_j - w_{ij}) \quad (1)$$

Where η h is a learning rate that is gradually decreased, ξ_j is the value of input node j, and i* is the winner node. The neighborhood function $\Lambda(i,i^*)$ is 1 for i=i* and it decrease with distance $|r_i-r_i^*|$ between unit i and i* in the output array. A typical choice for $\Lambda(i,i^*)$ is as follows:

3.2.2.1. Clustering

Data clustering, which is an NP-complete problem of finding groups in heterogeneous data by minimizing some measure of dissimilarity, is one of the fundamental tools in data mining, machine learning and pattern classification solutions (Garey, 1982). Clustering in N-dimensional Euclidean space R^N is the process of partitioning a given set of *n* points into a number, say *k*, of groups (or, clusters) based on some similarity (distance) metric in clustering procedure is Euclidean distance, which derived from the Minkowski metric (equations 2 and 3).

$$d(x,y) = \left(\sum_{i=1}^{m} |x_i - y_j|^r\right)^{1/r} \quad (2)$$

$$d(x,y)=\sqrt{\sum_{i=1}^{m}\left(x_i-y_j\right)^2} \quad (3)$$

In this study, we will also use Euclidian metric as a distance metric. The existing clustering algorithms can be simply classified into the following two categories: hierarchical clustering and partitional clustering. The most class of popular class of partitional clustering methods is the center based clustering algorithms (Zulal et al, 2006).

The k-means algorithms, is one of the most widely used center based clustering algorithms (Forgy, 1965).

To find *K* centers, the problem is defined as an optimization (minimization) of a performance function, Perf(X, C), defined on both the data items and the center locations. A popular performance function for measuring goodness of the k clustering is the total within-cluster variance or the total mean-square quantization error (MSE), equation 4 (Zulal et al, 2006).

$$Perf(X,C)=\sum_{i=1}^{N} Min\left\{\left\|x_i-c_l\right\|^2 \middle| l=1,...,K\right\} \quad (4)$$

The steps of the k-means algorithm are as follow (Mualik and Bandyopadhyay, 2000):

Step 1: Choose *K* cluster centers randomly from *n* points

Step 2: Assign each point to clusters

Step 3: Compute new cluster centers

Step 4: If termination criteria satisfied, stop otherwise continues from step 2

Note that in case the process close not terminates at step 4 normally, then it executed for a mutation fixed number of iterations.

3.2.2.2. Honey-bee Modeling

A honeybee colony typically consists of a single egg-laying long-lived queen, anywhere from zero to several thousands drones (depending on the season) and usually 10,000 to 60,000 workers (Afshar, 2001). Queens are specialized in egg laying. A colony may contain one queen or more during its life cycle, which named monogynous and/or polygynous colonies, respectively .Only the queen bee is fed "royal jelly", which is a milky-white colored, jelly-like substance "Nurse Bees" secrete this nourishing food from their glands, and feed it to their queen. The diet of royal jelly makes the queen bee bigger than any other bee in the hive. A queen bee may live up to 5 or 6 years, whereas worker bees and drones never live more than 6 months. There usually several hundred drones that live with the queen and worker bees. Mother Nature has given the drones' just one task, which is to provide the queen with some sperm. After the mating process, the drones die.

Drones are the fathers of the colony. They are haploid and act to amplify their mothers‘ genome without altering their genetic composition, except through mutation. Therefore, drones considered as agents that propagate one of their mother‘s gametes and function to enable females to act genetically as males. Workers specialized in brood care and sometimes lay eggs. Broods arise either from fertilized or unfertilized eggs. The former represent potential queens or workers, whereas the latter represent prospective drones (Afshar et al, 2006).

In the marriage process, the queen(s) mate during their mating flights far from the nest. A mating flight starts with a dance performed by the queen who then starts a mating flight during which the drones follow the queen and mate with her in the air. In each mating, sperm reaches the spermatheca and accumulates there to form the genetic pool of the colony. Each time a queen lays fertilized eggs, she randomly retrieves a mixture of the sperm accumulated in the spermatheca to fertilize the egg (Page, 1980).

The queen is pursued by a large swarm of drones (drone comets), when copulation occurs. Insemination ends with the eventual death of the drone, and the queen receiving the’’ mating sign ‘‘. The queen mates multiple times but the drone, inevitably, only once. These features make bee mating the most spectacular mating among insects (Afshar et al, 2006).

The mating flight may considered as a set of transitions in a state-space (the environment) where the queen moves between the different states in some speed and mates with the drone encountered at each state probabilistically. At the start of the flight, the queen initialized with some energy content and returns to her nest when the energy is within some threshold from zero to full spermatheca (Afshar et al, 2006).

In developing the algorithm, the functionality of workers is restricted to brood care and therefore, each worker may be represented as a heuristic which acts to improve and/or take care of a set of broods (i.e., as feeding the future queen with royal jelly). A drone mates with a queen probabilistically using an annealing function as follows (Afshar, 2001):

$$\Pr ob(Q, D) = \exp[-\Delta(f)/S(t)] \quad (5)$$

Where *Prob(Q ,D)* is the probability of adding the sperm of drone D to the spermatheca of queen Q (that is, the probability of a successful mating); $\Delta(f)$ is the absolute difference between the fitness of D (i.e., f (D)) and the fitness of Q (i.e., f (Q)); and S (t) the speed of the queen at time t .It is apparent that this function acts as an annealing function, where the probability of mating is high when the queen is still at the beginning of her mating flight, therefore her speed is high, or when the fitness of the drone is as good as the queen‘. After each transition in space, the queen‘s speed and energy decays according to the following equations:

$$S(t+1) = \alpha(t) \times S(t), \quad (6)$$

where α is a factor $\in [0,1]$ and is the amount of speed reduction after each transition.

Workers that used to improve the brood‘s genotype may represent a set of different heuristics. The rate of improvement in the brood‘s genotype, as result of a heuristic application to that brood, defines the heuristic fitness value.

The fitness of the resulting genotype is determined by evaluating the value of the objective function of the brood genotype and/or its normalized value. It is important to note that a brood has only one genotype.

Thus, an HBMO algorithm maybe constructed with the following five main stages:

- The algorithm starts with the mating flight, where a queen (best solution) selects drones probabilistically to form the spermatheca (list of drones). A drone then selected from the list randomly for the creation of broods.
- Creation of new broods (cluster centers) by crossover the drone‘s genotypes with the queens.
- Use of workers (heuristics) to conduct local search on broods (trial solutions).
- Adaptation of worker‘s fitness, based on the amount of improvement achieved on broods.
- Replacement of weaker queens by fitter broods.

The algorithm starts with three user-defined parameters and one predefined parameter.The predefined parameter is the number of workers)W ,(representing the number of heuristics encoded in the program. However, the predefined parameter may be used as a parameter to alter the number of active heuristics if required; that is, the user may choose the first heuristic, where W is less than or equal to the total number of heuristics encoded in the program. The three user-defined parameters are the number of queens, the queen‘s sperm theca size representing the maximum number of mating per queen in a single mating flight, and the number of broods that will be born by all queens. The speed of each queen at the start of each mating flight initialized at random. A set of queens then initialized at random. A randomly selected heuristic then used to improve the genotype of each queen, assuming that a queen is usually a good bee. A number of mating flights are undertaken. In each mating flight, all queens fly based on the speed of each, where speed generated at random for each queen before each mating flight commences. At the start of a mating flight, a drone generated randomly and the queen positioned over that drone. The transition made by the queen in space based on her speed that represents the probability of flipping each gene in the drone‘s genome. At the start of a mating flight, the speed may be higher and the queen may make very large steps in space. While the energy of the queen decreases, her speed decreases, and as a result, the neighborhood covered by the queen, decreases. At each step in the space, the queen mates with the drone encountered at that step using the probabilistic rule in Eq. (5). If the mating is successful (i.e., the drone passes the probabilistic decision rule), the drone‘s sperm is stored in the queen‘s spermatheca. To sum up, the algorithm starts with a mating flight where a queen selects a drone with a predefined probabilistic rule. By cross-overing the drone‘s genotypes with the queen‘s, a new brood (trial solution) is formed which later can be improved, employing workers to conduct local search (Afshar et al, 2006).

When all queens complete their mating flight, they start breeding. For a required number of broods, a queen selected in proportion to her fitness and mated with a randomly selected sperm from her spermatheca. A worker chosen in proportion to its fitness in order to improve the resultant brood. After all broods have been generated, they are sorted according to their fitness. The best brood replaces the worst queen until there is no brood that is better than any of the queens are. Remaining broods then killed and a new mating

flight begins until all assigned mating flights are completed or convergence criteria met (Afshar et al, 2006). The main steps of the HBMO algorithm presented in figure 3. In addition, a full-scale computational flowchart illustrated in figure 4.

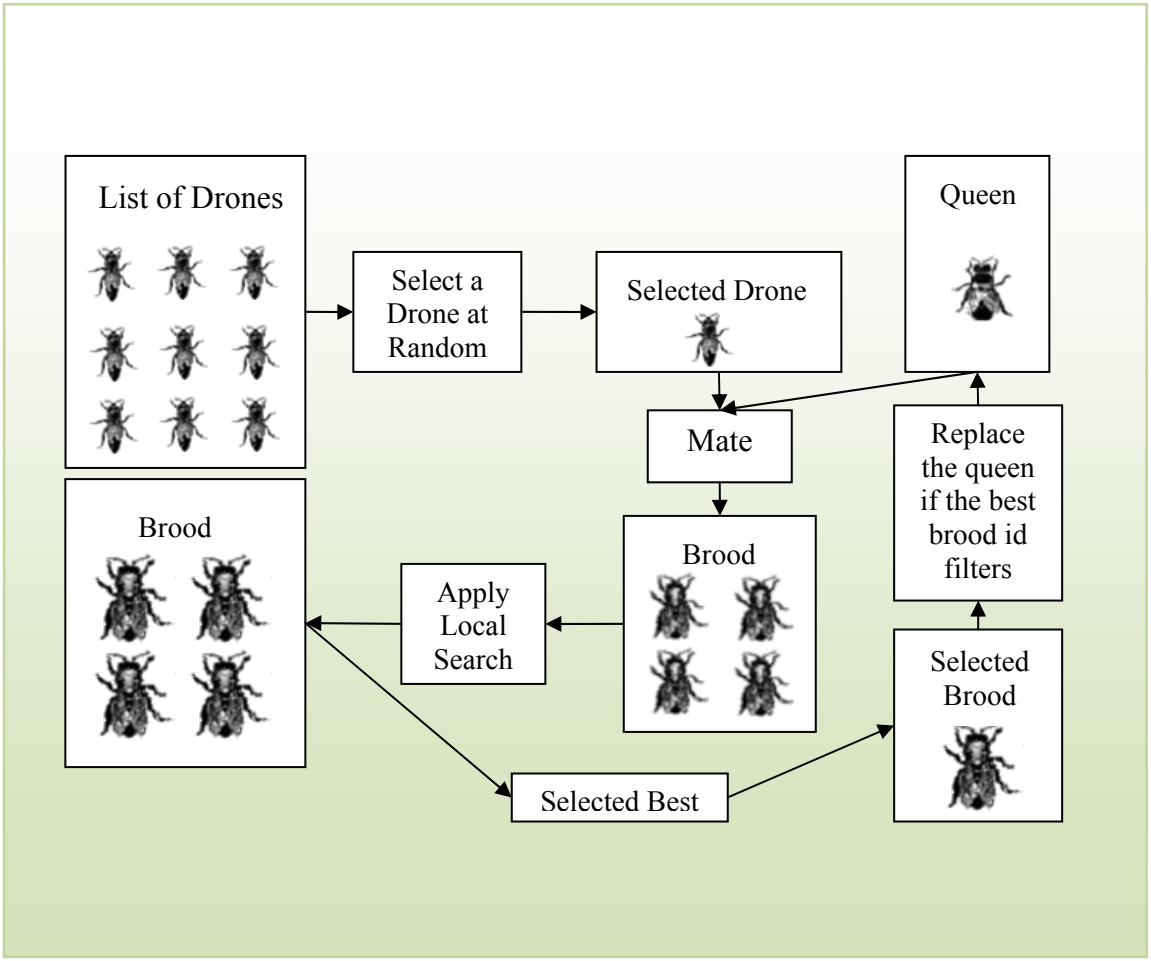

Figure 3.The HBMO algorithm

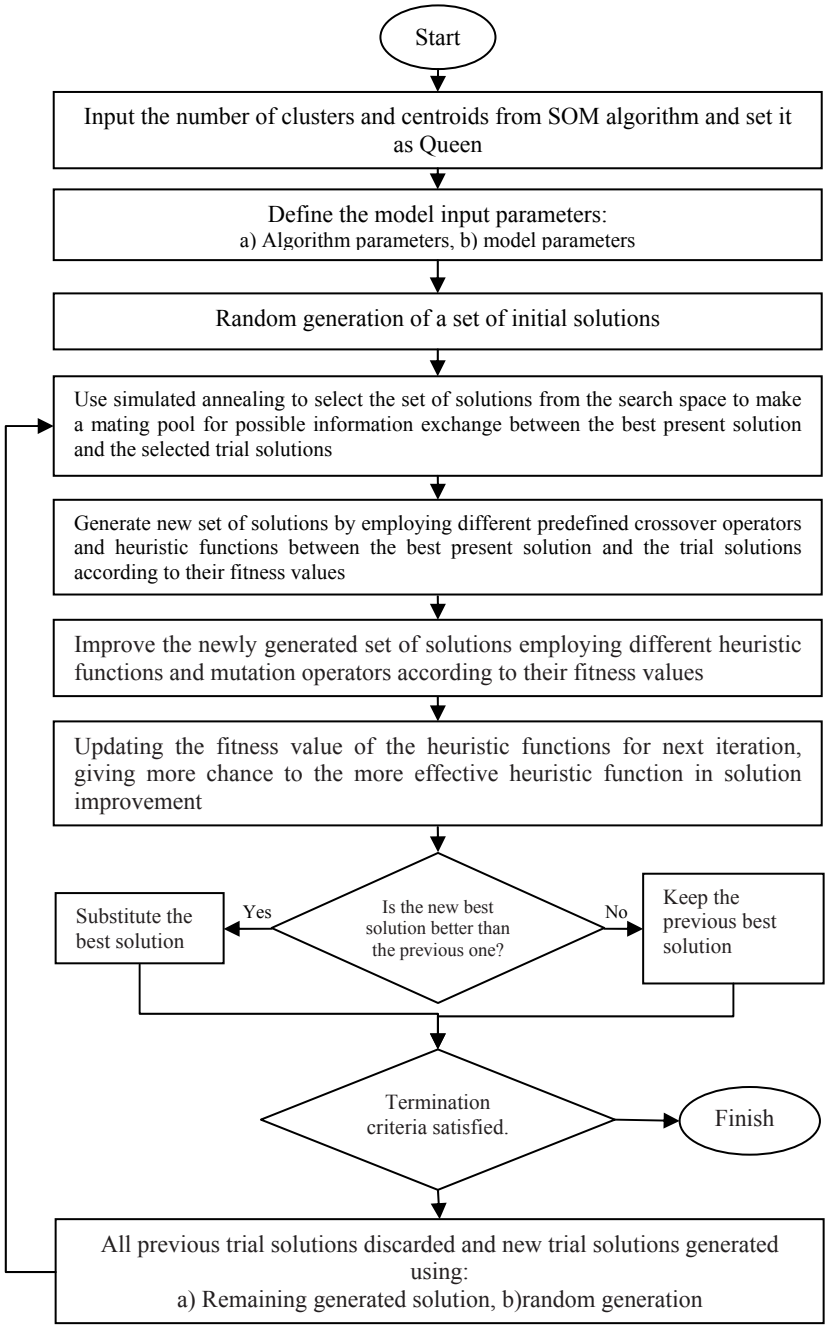

Figure 4. HBMO Algorithm representation

3.2.2.3. HBMK Clustering Algorithm

The search capability of HBM algorithm used in this article for the purpose of appropriately determining a fixed number of *K* cluster centers in R^N; thereby suitably clustering the set of *n* unlabelled points the clustering metric that has been adopted is the sum of the Euclidean distance of the points of the points from their respective cluster centers. The steps of the proposed algorithm are shown in figure 2, there are now described in detail.

Step 1: String representation

A chromosome has used to represent a candidate solution to a problem where each gene in the chromosome represents a parameter of the candidate solution. In this study, a chromosome regarded as a set of K initial cluster centers and each gene is a cluster center dimension. Specifically, a chromosome can be represented as

$C = [c_1 \ldots c_j \ldots c_K]$ where c_j is the jth gene and K is total number of genes. Figure 5, illustrate a chromosome encoding example. $C_1 = (2, 5, 1)$, $C_2 = (6, 3, 2)$, $C_3 = (5, 7, 4)$.

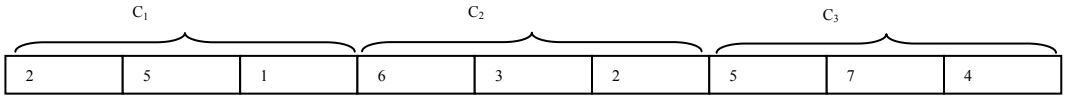

Figure 5. A chromosome-encoding example

Step 2: Input number of clusters and centroids
Input number of clusters and centrodis from SOM algorithm and set it as queen.

Step 3: Define the model inputs parameters
The algorithm starts with three user-defined parameters and one predefined parameter. The predefined parameter is the number of worker (W), representing the number of heuristics encoded in the program. However, the predefined parameter may be used as a parameter to alter the number of active heuristics if required; that is, the user may choose the first heuristic, where W is less than or equal to the total number of heuristics encoded in the program. The user-defined parameters are number of queens, the queen's spermatheca size representing the maximum number of broods that will be born by all queens. The speed of each queen at the start of each mating flight initialized at random,

Step 4: Random generation of a set of initial solutions
In this stage, a set of initial cluster centers generated randomly from the dataset points. Each solution represents *K* cluster centers as shown in figure1.

Step 5: Mating flight
Use simulated annealing to select the set of solutions from the search space to make a mating pool for possible information exchange between the best present solution (queen) and the selected trial solutions.

Step 6: Breeding process
Generate new set of solutions by employing predefined crossover operators and heuristic functions between the present solutions and the trial solution according to their fitness values. In this study, we adopt intermediate crossover. It creates children by taking a weighted average of parents. You can specify the weights by a single parameter, Ratio, which can be a scalar or a raw vector of length number of variables. The default is a vector of all 1's. The function creates the child from parent 1 and parent 2 using the following equation.

$$\text{Child} = \text{parent 1} + \text{ran} * \text{Ratio} * (\text{parent 2} - \text{parent 1}) \quad (7)$$

If all the entries of Ratio lie in the range [0, 1], the children produced are within the hypercube defined by placing the parents apposite vertices. If ratio is not in that range, the children might lie outside the hypercube. If ratio is a scalar, then all the children lie on the line between the parents.

Stage 7: feeding selected broods and queen with the royal jelly by workers
Improve the newly generated set of solutions employing different heuristic functions and mutation operators according to their fitness values. For binary representation of chromosomes a bit position mutated, by simply flipping its value. Since we are considering floating point representation in this article, we use the following mutation. A number δ in the range [0, 1] generated with uniform distribution. If the value at a gene position is *v*, after mutation it becomes:

$$v \pm 2 * \delta * v \qquad v \neq 0$$
$$v \pm 2 * \delta \qquad v = 0$$

The + or – sign occurs with equal probability. Note that we could have implemented mutation as:

$$v \pm \delta * v$$

However, one problem with this form is that if the values at a particular position in all the chromosomes of a population become positive (or negative), then we will never be able to generate a new chromosome having a negative (or positive) value at the position. In order to overcome this limitation, we have incorporated a factor of 2 while implementing mutation other form like

$$v \pm (\delta + \varepsilon) * v$$

Where $0 < \varepsilon < 1$ would also have satisfied our purpose.

Step 8: If the new best solution is, better than the queen replace it with queen.

Step 9: Check the termination criteria

If the termination criteria satisfied finish the algorithm, else discard all previous trial solutions and generate new trial solutions. Then go to step 5 until all assigned iteration (mating flights) completed or convergence criteria met.

The pseudo code for application of HBM algorithm in clustering illustrate in figure 6.

- *Set k* — *Number of clusters*
- *Set m* — *Number of Drones*
- *Set L* — *Capacity of Spermatheca*
- *Set T_{max}* — *Speed of Queen at the start of a mating flight*
- *Set T_{min}* — *Speed of Queen at the end of a mating flight*
- *Set t* — *Speed reduction schema*
- *Set P* — *Number of iteration*
- *Set $T= T_{max}$*
- *Set w* — *Number of workers*
- *Set b* — *Number of broods*

Begin
Set the centroids from SOM algorithm as Q (Queen)
Generate m Drones with k length randomly from X matrices, and set them as D matrices
Calculate their objective function
Repeat
Repeat
Select a D_i from D randomly
Calculate $\Delta(f)=|f(Q)-f(D_i|$
Generate r randomly,
If $exp(-\Delta(f)/T)>r$
Add D_i to spermatheca S
Else $T=t*T$
Until Capacity of spermatheca completed or $T=T_{min}$

Repeat
Select a crossover function from list W, according to its probability
Select a S_i randomly and generate new solution by crossover S_i and Q and set it as B_i
Calculate crossover fitness
Update probability matrices of crossover function selection
Until Number of broods equal to b

Begin
Select a mutation function from list W, according to its probability
Mutation B_i and set it as E_i
If $f(E_i)>f(Q)$ swap E_i and Q
Else keep the previous solution
Calculate mutation fitness
Update probability matrices of mutation function selection
End

Generate new drones list randomly

Until the termination criteria satisfied
End

Figure 6. Pseudo code for HBM clustering Algorithm

4. SIMULATION

Clustering methods have been presented and evaluated in numerous studies (Punj & Steward, 1983, Vriens et al., 1996, and Kuo et al., 2002a, b). Though these techniques are optimal for some specific distributional assumption or dimensionality, further study is still necessary for determining their robustness to data, which do not satisfy the assumed structure. However, in real-world problems, it is quite difficult to determine which clustering method is the best, since the true, or real, clustering solutions are unknown. Thus, most validation literature tries to solve this problem through Monte Carlo simulation. One of the main advantages of this is that the researcher can use analytical data with a known structure. This section presents the simulation results for three methods through Monte Carlo study proposed by (Milligan & Cooper, 1985).

4.1. Generating simulation data sets

According to the simulation algorithm described in the Section 3.1, Matlab 7 is used to program the algorithm. Discriminant analysis using SPSS 9.0 shows that the clusters can discriminate fairly well for the simulation algorithm.

4.2. Results of SOM

For SOM network, the number of learning epochs is set as 100 and the training rate is set to 0.5. The two-dimension output topology is 10×10. The input data for SOM are generated from the simulation algorithm. The error function (Kohonen, 1991) used to evaluate the convergence of SOM is:

$$E = \sum_{i=1}^{P} \sum_{j=1}^{M} \sqrt{\left(X_i - W_{ij}\right)^2} \qquad (8)$$

where P is the number of training patterns, M is the number of output units and W_{ij} is the weights of the winner unit for every training pattern.

4.3. The Results of SOM+HBMK

SOM+HBMK method is programmed in Matlab 7 on a Pentium IV, 2.8 CPU and 1 GB RAM. The average computational time of a generation for SOM+HBMK is 37.33 seconds. The SSW value of SOM+HBMK does not decrease significantly after 20 generations. Thus, this study sets up the number of generations to be 20.

4.4. The evaluation of three clustering methods

This study compares three methods using 243 simulated data sets by calculating misclassification rate and SSW. To examine the performance of these three clustering methods, SPSS 9.0 is utilized for the ANOVA test. The performance of the proposed SOM+HBMK with lowest misclassification rate, 1.54%, is the best among the three clustering methods, as listed in Tables 2 and 3. For further discussion, five hypothesizes are examined, as follows.

Hypothesis 1: The number of misclassifications does not differ across the levels of error.

The error factor affecting cluster recovery of three methods is as significant as the number of clusters at 0.05, significance level according to Table 3. The mean misclassifications of three methods increase for the error free, low-error, and high-error data sets.

Hypothesis 2: The number of misclassifications does not differ across the number of clusters in the data set.

According to Table 2, it is shown that the number of clusters affects the cluster recovery of three methods significantly at the 0.05 significance level. As the number of clusters increases, cluster recovery becomes better for each method. Table 3 illustrates this finding.

Hypothesis 3: The number of misclassifications does not differ across the number of dimensions of each observation.

The number of dimensions does not affect cluster recovery of the three methods according to Table 2. In Table 3, the cluster ability is the best for SOM+HBMK but the worst for SOM+HBMK while the number of dimensions is 6. The cluster ability is the best for SOM+K as the number of dimensions is 10. For SOM+GAK the cluster ability is the best as the number of dimensions is 8.

Hypothesis 4: The number of misclassifications does not differ across the levels of density.

The levels of density dose not affect the method according to Table 2. The cluster recovery is bad for SOM+K while the level of density is 10%. For SOM+GAK, the cluster recovery is bad while the level of density is 50%. The cluster ability is the best for SOM+HBMK as the level of density is 10% in Table 3.

Hypothesis 5: The number of misclassifications does not differ across the three clustering methods.

In Table 3, the misclassifications of SOM+K are significantly higher than other cluster methods. But it is shown that the number of misclassifications does not differ across the three clustering methods. So, ANOVA test is used, with the results as shown in Table 4.

Table 2. The computational results of multivariate analysis of variance for three clustering methods.

Factors	SOM+K-means	SOM+GAK	SOM+HBMK
Cluster Number	0.000*	0.000*	0.000*
Dimension	0.736	0.478	0.252
Density Level	0.297	0.759	0.825
Error Level	0.000*	0.000*	0.000*

* the mean difference is significant at the α=0.05 level.

Table 3. The number of misclassification rates across the factor level

	Level	SOM+K-means	SOM+GAK	SOM+HBMK
Average		0.0411	0.0211	0.0154
Cluster Number	3	0.0744	0.0593	0.0328
	5	0.0253	0.0033	0.0025
	7	0.0235	0.0036	0.0019
Dimension	6	0.0460	0.0285	0.0264
	8	0.0389	0.0175	0.0155
	10	0.0383	0.0202	0.0148
Density Level	0.1	0.0493	0.0186	0.0157
	0.5	0.0417	0.0256	0.0160
	0.6	0.0322	0.0219	0.0163
Error Level	No	0.0083	0.0024	0.0011
	Low	0.0215	0.0113	0.0098
	High	0.0933	0.0525	0.0318

Table 4. The ANOVA test of three clustering methods (α=0.05)

S.V.	S.S.	DF	MS	F Test	P-Value
SSB	2935,357	2	1467,679	18,61251	0.000
SSE	18925,058	240	78,854		
SST	21860,415	242			

The three clustering methods have significant differences according to Table 4. To examine their performance, SPSS 9.0 is utilized for Scheff's multiple comparison test. The performance of the proposed SOM+HBMK is the best among the three clustering methods. Use critical values from Scheffe's S procedure, derived from the F distribution. This procedure provides a simultaneous confidence level for comparisons of all linear combinations of the means, and it is conservative for comparisons of simple differences of pairs. According to Scheffe's multiple comparison tests, as shown in Table 5, the mean difference between clustering methods is not significant at 0.05, significance level. The reason may be that the simulated samples are not vague enough. But the MSW (mean of SSW) of SOM+HBMK is smaller than other two methods shown in Table 6.

Table 5. The Scheff's multiple comparison test (α=0.025)

Clustering Method I	Clustering Method II	P-Value
SOM+K	SOM+GAK	0.125
SOM+K	SOM+HBMK	0.083
SOM+GAK	SOM+HBMK	0.102

Table 6. The MSW of three clustering method

Clustering methods	Monte Carlo simulation	SOM+K-means	SOM+GAK	SOM+HBMK
SSW	46203	53318	47672	32528

5. MODEL EVALUATION RESULTS

SOM+HBMK is the best method for clustering analysis as shown in Section 4. To further demonstrate the proposed method, an advanced comparison of three methods was made, using real-world data of an internet bookstore for market segmentation based on customer loyalty.

5.1. RFM Model

Direct marketing professionals have been trying to gain such insight ever since the end of the nineteenth century, when the first catalogue of products that could be ordered by mail appeared in the USA (Raphael, 2002). However, it was only at the beginning of the 1960s that a simple and effective quantitative method to separate customers who are likely to make purchases from those who are not was devised: the recency-frequency-monetary value (RFV or RFM) analysis (Cullinan, 1977). Generally, shortened to RFV, it is sometimes known as "RFM" analysis. In this approach to market segmentation, customers are clustered together into an arbitrary number of segments according to their most recent day of purchase, the number of purchases they have made and the monetary value of their purchases. A random sample taken from the segmented customer database is then subjected to a direct marketing campaign. As a result, some customer segments may reveal themselves to be profitable, while others may do the reverse. Subsequently, the remaining customers in the database who belong to profitable segments are targeted by the same campaign (Armondo et al, 2006).

5.2.1. Clustering analysis

For the 243 simulated data sets, there is no significant difference for SOM+K and SOM+GKA as shown in Section 4. However, the SSW of SOM+GKA is smaller than that of SOM+K. Thus, this study will further compare their clustering capabilities.

(1) The determination of segmentation number

The number of clusters is determined by SOM network. The parameters for the SOM network are the same as those set in Section 4. There are 700 training samples with 20 input data for training. Fig. 9 displays the training result of SOM network, indicating that there are clearly five clusters.

(2) Evaluation of three clustering methods

After determining the number of clusters using SOM, three methods are used to cluster 700 samples. The within cluster variation (SSW) is also calculated since it is used to evaluate three methods, SOM+K, SOM+GKA and SOM+HBMK. Table 7 shows that SOM+HBMK has the smallest SSW, which is identical to the previously experimental result. Thus, SOM+HBMK is employed as the clustering tool for market segmentation.

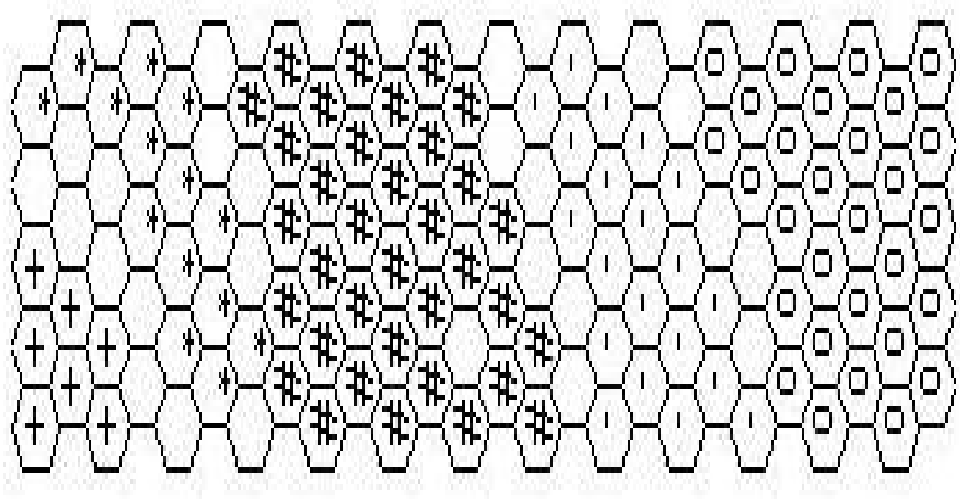

Figure.7. Clustering result of SOM.

Table 7. SSW of three methods

Clustering methods	SOM+K-means	SOM+GAK	SOM+HBMK
SSW	1.542×10^5	1.815×10^5	1.542×10^5

6. CONCLUSIONS

In a clustering problem, it is always difficult to determine the number of clusters. This study shows that the auto clustering feature of SOM is more effective and objective than the K-means method. Thus, SOM can be utilized as the initial stage for market segmentation to determine the number of clusters and starting points. According to Section 4, the proposed SOM+HBMK

clustering method,, using the 243 simulated data sets and the customer loyalty data of an internet bookstore in Iran, is better than both SOM followed by K-means and SOM followed by GAK-means. Thus the proposed two-stage method, which first uses the SOM to determine the number of clusters and then employs honey bee mating optimization algorithm based on K-means algorithm to find the final solution of clustering, is a robust clustering method. It can be applied as a new clustering method for market segmentation or other clustering problems.

6. REFERENCES

[1]. Afshar, O.Bozog Haddad, M.A. Marino, B.J.Adams, (2006), Honey-bee mating optimization (HBMO) algorithm for optimal reservoir operation, Journal of the Franklin Institute.

[2]. Anil, C., Carroll, D., Green, P. E., & Rotondo, J. A. (1997). A feature-based approach to market segmentation via overlapping K-centroids clustering. Journal of Marketing Research, Vol. XXXIV (August), 370–377.

[3]. Armando Leite Ferreira, Eber Assis Schmitz, Priscila M.V. Lima and Fernando Silva Pereira Manso, Optimized RFV analysis, Marketing Intelligence & Planning Vol. 24 No. 2, 2006, pp. 106-118.

[4]. Balakrishnan, P.V.(Sundar), Cooper, M. C., & Jacob, V. S. (1994). A study of classification capabilities of neural networks using unsupervised learning: A comparison with K-means clustering. Psychometrika, l.59(4), 509–525.

[5]. Balakrishnan, P.V.(Sundar), Cooper, M. C., Jacob, V. S., & Lewis, P. A. (1996). Comparative performance of the FSCL neural net and K-means algorithm for market segmentation. European Journal of Operational Research, 93, 346–357.

[6]. Bigus, J. P. (1996). Data mining with neural networks. New york: McGraw-Hill.

[7]. O. Bozorg Haddad, A. Afshar, MBO (Marriage Bees Optimization), A new heuristic approach in hydro systems design and operation, in: Proceedings of 1st International Conference On Managing Rivers In The 21st Century: Issues and Challenges, Penang, Malaysia, 21–23 September 2004, pp. 499–504.

[8]. O. Bozorg Haddad, A. Afshar, M.A. Marin˜ o, Honey bees mating optimization algorithm (HBMO); a new heuristic approach for engineering optimization, in: Proceeding of the First International Conference on Modeling, Simulation and Applied Optimization (ICMSA0/05), Sharjah, UAE, 1–3 February 2005.

[9]. Cadden, D. T. (1991). Neural networks and the mathematics of chaos-an investigation of these methodologies as accurate predictors of corporate bankruptcy First international conference on artificial intelligence applications on wall street (pp.582–589). IEEE Computer Society Press.

[10]. Chang, S. (1998). Internet segmentation:State-of-the-art marketing applications. Journal of Segmentation in Marketing, 2(1), 19–34.

[11]. Cowgill, M. C., Harvey, R. J., & Watson, L. T. (1999). A genetic algorithm approach to cluster analysis. Computers & Mathematics with Applications, 37(7), 99–108.

[12]. C.S. Sung, H.W. Jin, A tabu-search-based heuristic for clustering, Pattern Recognition, Volume 33, 2000, Pages 849-858.

[13]. Cullinan, G.J. (1977), Picking them by their Batting Averages: Recency-Frequency-Monetary Method of Controlling Circulation, Direct Mail/Marketing Association, New York, NY.

[14]. Dillon, W. R., Kumar, A., & Borrero, M. S. (1993). Capterin individual differences in paired comparisons: an extended BTL model incorporating descriptor variables. Journal of Marketing Research, 30, 42–51.

[15]. E.W. Forgy, Cluster analysis of multivariate data: Efficiency versus interpretability of classifications, Biometrics 21, 3, 1965, 768–769.

[16]. M.R. Garey, D.S. Johnson, H.S. Witsenhausen, The complexity of the generalized Lloyd–Max problem, IEEE Trans. Inform. Theory 28 ,2, 1982, 255–256.

[17]. H. Spath, Clustering Analysis Algorithms, Ellis Horwood, Chichester, UK, 1989.

[18]. H.A. Abbass, (2001), A monogamous MBO approach to satisfiability, in: Proceeding of the International Conference on Computational Intelligence for Modeling, Control and Automation, CIMCA'2001, Las Vegas, NV, USA,.

[19]. H.A. Abbass, (2001), Marriage in honey-bee optimization (MBO): a haplometrosis polygynous swarming approach, in: The Congress on Evolutionary Computation (CEC2001), Seoul, Korea, pp. 207–214.

[20]. Kohonen, T. (1982). A simple paradigm for the self-organized formation of structured feature maps. In S. Amari, & M. Berlin (Eds.), Competition and cooperation in neural nets, Lecture notes in biomathematics. Berlin: Springer.

[21]. Kohonen, T. (1991). Self-organizing maps: Optimization approaches. In T. Kohonen, K. Makisara, O. Simula, & J. Kangas (Eds.), Artificial neural networks (pp. 981–990). Amsterdam, The Netherlands: Elsevier.

[22]. Kotler, P. (1997). Marketing management: analysis, planning, implementation, and control (9th ed). Upper Saddle River, NJ: Prentice Hall.

[23]. Krishna, K., & Murty, M. (1999). Genetic K-means algorithm. IEEE Transactions on Systems, Man, and Cybernetics-Part: Cybernetics, 29(3), 433–439.

[24]. Kuo, R. J., & Xue, K. C. (1998). A decision support system for sales forecasting through fuzzy neural network with asymmetric fuzzy weights. Journal of Decision Support Systems, 24(2), 105–126.

[25]. Kuo, R. J., Ho, L. M., & Hu, C. M. (2002). Integration of self-organizing feature map and K-means algorithm for market segmentation. International Journal of Computers and Operations Research, 29, 1475–1493.

[26]. R. I. kuo, H. S. Wang, Tung-Lai Hu, S. H. Chou, Application of Ant K-Means on Clustering Analysis, Computers and Mathematics with Applications 50 (2005) 1709-1724.

[27]. Kuo, R.J., An, Y.L, Wang, H.S, Chung, W.J, Integration of self-organizing feature maps neural network and genetic K-means algorithm for market segmentation, Expert Systems with Applications, 30, pp. 313–324, (2006).

[28]. Lee, R. C. T., Slagle, J. R., & Blum, H. (1977). A triangulation method for the sequential mapping of points from N-space to two-space. IEEE Transactions on Computers, 26, 288–292.

[29]. Maulik, U., & Bandyopadhyay, S. (2000). Genetic algorithm-based clustering technique. Pattern Recognition, 33(9), 1455–1465.

[30]. Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. Psychometrika, 50, 159–179.

[31]. O'Connor, G. C., & O'Keefe, B. (1997). Viewing the web as a marketplace: the case of small companies. Decision Support Systems, 21(3), 171–183.

[32]. Punj, G., & Steward, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for applications. Journal of Marketing Research, 20, 134–148.

[33]. Pykett, C. E. (1978). Improving the efficiency of Sammon's nonlinear mapping by using clustering archetypes. Electronics Letters, 14, 799–800.

[34]. Raphael, M. (2002), "Where did you come from?", Direct Marketing, Vol. 62 No. 11, pp. 36-8.

[35]. P.S. Shelokar, V.K. Jayaraman, B.D. Kulkarni, An ant colony approach for clustering, Analytica Chimica Acta 509, 2004, 187–195.

[36]. R.E. Page, The evolution of multiple mating behaviors by honey-bee queens (Apis mellifera L.), J. Genet. 96,1980, 263–273.

[37]. Shokri Z. Selim and K. Al-Sultan, A Simulated Annealing Algorithm for the Clustering problem, Pattern Recognition, Volume 24, Issue 10, 1991, Pages 1003-1008.

[38]. Vellido, A., Lisboa, P. J. G., & Vaughan, J. (1999). Neural networks in business: a survey of applications (1992-1998). Expert Systems with Applications, 17(1), 51–70.

[39]. Venugopal, V., & Baets, W. (1994). Neural networks and their applications in marketing management. Journal of Systems Management, 45(9), 16–21.

[40]. Vriens, M., Wedel, M., & Wilms, T. (1996). Metric conjoint segmentation method: A Monte Carlo comparison. Journal of Marketing Research, 33, 73–85.

[41]. Wann, C-D. , & Thomopoulos, C.A. S. (1997). A comparative study of self organizing clustering algorithm dignet and ART2. Neural Networks, 10(4), 737–753.

[42]. Wedel, M., & Kamakura, W. A. (1998). Market segmentation: conceptual and methodological foundations. Boston: Kluwer Academic.

[43]. Zulal Gungor, Alper Unler, (2006). K-harmonic means data clustering with simulated annealing heuristic, Applied Mathematics and Computation, in press.