

**DEVELOPMENT OF A HYBRID SYSTEM OF ARTIFICIAL NEURAL NETWORKS
AND ARTIFICIAL BEE COLONY ALGORITHM FOR PREDICTION AND
MODELING OF CUSTOMER CHOICE IN THE MARKET**

J. Shahrabi*¹, S. Mottaghi Khameneh²

¹Academic Rank: Assistant Professor, Amirkabir University

²Master of Industrial Engineering, Amirkabir University

Published online: 15 February 2017

ABSTRACT

With the increasing growth of technology and the emergence of various industries, numerous manufacturers have entered this field. In today's world, sellers and manufacturers find themselves among a vast number of competitors. Therefore, they need to adopt a variety of policies and strategies for their own survival and profitability. Companies should identify their customers' needs and adopt their own policies based on customers' purchase behaviors. To this end, attempts have been made to identify the customer choice model since the past decades. These models aim at modeling and predicting customer choice among several brands. Traditional models were of interest for many years and these methods were frequently used with the advent of artificial intelligence and machine learning systems. They could demonstrate very good results. In this study, it has been attempted to present a new method for the modeling and prediction of customer choice in the market using the combination of artificial intelligence and data mining. Indeed, the new model is to be applied in helping managers with decision-making. Hence, probabilistic neural networks have been combined with artificial bee colony algorithm.

Author Correspondence, e-mail: jamalshahrabi@aut.ac.ir

doi: <http://dx.doi.org/10.4314/jfas.v9i1s.685>



The proposed model was tested in a real market and its efficiency and accuracy were finally compared with those of other models, including neural network trained with back-propagation, probabilistic neural networks, and the neural networks trained with genetic algorithm. The results reveal that the hybrid model shows better performance than the other models.

Keywords: Consumer Choice Model, Data Mining, Artificial Intelligence, modeling, predicting, probabilistic neural network, artificial bee colony algorithm.

1- INTRODUCTION

Quantitative analysis has become widespread in business and marketing, especially in the identification of customer behavior and decision making process (customer choice model) in commercial activities (1). Customer choices models are model the choice of different brands by customers in the market. These models can be applied as simulators of market reaction to the marketing strategies of companies in order to suggest appropriate alternative decisions on price promoting, policies of companies, and so on. Thus, development of a proper model with high accuracy on the prediction of customer behavior in the market can be of help to companies as a strategic tool to improve the adoption of decision-making policies in the domain of marketing. (2)

Research methodology and theoretical framework in customer choice was first introduced by McFadden in 1973 through Conditional Logit Analysis. (1)

Customer choice models can be considered as a classification problem wherein customers are placed in different categories based on the choices they make. Logit and Probit models constitute the traditional models used in this area. Due to the complexity of customer choice process, the above linear models cannot exercise a desired accuracy in this context. (3)

Today, artificial intelligence techniques, such as artificial neural networks, fuzzy logic, and genetic algorithms have become the common subject of research in various fields due to their high capabilities in modeling complex engineering issues. The integration of the main advantages of these relatively new techniques has been of great concern in recent years. Accordingly, new hybrid models, including neural-fuzzy networks, genetic-fuzzy, and so on have come out (3).

Hybrid models of artificial intelligence benefit from greater flexibility and enjoy the ability to accurately estimate nonlinear relationships without any requirement for the observation of assumptions and limitations of traditional methods such as Logit and the like. Therefore, many

researchers tend to use a hybrid model for prediction and classification within the scope of their work (4).

The employment of artificial intelligence in the field customer choice modeling has received unprecedented attention and a large number of studies have been carried out in this area. As an example, one may refer to Suh, et al.'s hybrid model, which resulted from the combination of artificial neural networks and Logit model (5).

The study of consumer choice has revealed artificial neural networks as suitable models for this purpose. Similarly, it has been found that probabilistic neural networks (henceforth referred to as PNN) outperform back propagation neural networks. Therefore, PNNs have been used for customer choice modeling in this study. The study showed that PNNs are so advantageous to neural networks that Gan, et al. used it for customer choice. They compared it with neural networks and concluded that PNNs may outweigh them (4). In addition, different algorithms have been used for neural network training; however, bee colony algorithm has not been used for neural network training in the scope of customer choice. Therefore, it was decided to use the combination of PNN and bee colony algorithm in customer choice domain.

The advantages of PNNs to neural networks are listed below.

- 1- PNNs are much faster than perceptron neural networks.
- 2- PNNs are more accurate than perceptron neural networks.
- 3- PNNs are relatively insensitive to outliers.
- 4- PNNs predict the probability of exact goals.

The only drawback of PNNs is that they need to have a very long memory. Indeed, one the main advantages of PNNs is that they act much faster than back propagation algorithm in problems in such a way that the back propagation algorithm time is a significant fraction of the total calculation time. In waves' Hull-to-Emitter Correlation, which has been conducted by Naval Ocean Systems Center (NOSC), PNN accurately identified the bodies. Marchette & Preibe explained the problem and classification results using back-propagation algorithm and traditional techniques. Maloney described the results of using PNN on the same dataset. The dataset contained 113 reports on emitting waves with three continuous input parameters for each. PNN output has increased classification accuracy to 89%. Since the author has used the same dataset, it took .7 second to run PNN with 85% accuracy. Compared to back-propagation algorithm with 82% accuracy, the results showed that the speed promotion is about 200,000 times in PNN. (6)

Bee colony algorithm is an optimization method that has the ability to find the optimal solution with relatively low computational requirements. (7)

Karaboga & Basturk in 2007 and Karaboga & Akay in 2009 used bee colony algorithm to optimize some multi-dimensional numerical problems. They compared the obtained results with those of famous evolutionary algorithms, such as genetic algorithms, particle swarm optimization, and innovative algorithm. Indeed, bee colony algorithm was found to have better performance than other algorithms.

The advantages of using bee colony algorithm to other methods include the enjoyment of memory, local searches, and solution mechanism improvement and, thereby, this algorithm can be effective in finding the optimal solution with high quality. Bee colony algorithm can be optimized to provide a trade-off between performance and complexity of optimization. (8, 9)

In this study, a new hybrid approach was developed using probabilistic neural networks and bee colony algorithm to find the optimal standard deviation in artificial neural networks and to predict customer choice in the market. In comparison with hybrid neural network models with genetic algorithms, etc., the proposed model has such advantages as higher accuracy, insensitivity to outliers, a smaller number of control parameters. In this regard, a case study is undertaken in order to test the proposed model and measure its modeling and predictive power.

In this study, the training dataset will be held constant at the first stage of the “choice model”, but the network design changes from the standard model of neural network with back propagation training into a hybrid of probabilistic neural network and bee colony algorithm. All the networks are optimized to predict customer choice.

In the second stage, price sensitivity analysis was performed to demonstrate the usability and applicability of the PNN-ABC model.

2- PNN application in marketing and customer choice modeling

Neural network usually classify models on the basis of samples. Various neural networks apply different learning rules, determine patterns through training samples, and then categorize new patterns based on the obtained results. The common methods, such as back-propagation training model use innovative approaches to explore statistical categories. Innovative methods usually involve many small corrections in system parameters that ultimately improve the system performance. In addition to the above calculations for training, it can be shown that the gradual

adaptation of back-propagation method is susceptible to wrong optimal acceptance. A classification method has come into existence based on statistical rules to improve this method.

Although probabilistic neural network entails a structure similar to that of back-propagation network, the main difference between the two networks lies in that PNN has an activation statistical function and other unique features. For instance, PNN asymptotically approaches the Bayes optimal level under simple and clear conditions of decision-making boundaries. (6)

The following is a brief description of PNN.

2-1- Structure of probabilistic neural networks

PNNs consist of three layers of nodes. Since their training occurs in a path instead of a few numerical values per training vector, they are quickly learned. Probabilistic neural networks estimate the probability density functions of each category based on training samples. Figure 2-1 shows the structure of PNNs, which identify K classes. Probabilistic neural network can be extended to any number of categories.

The input layer contains N nodes: there is one node for each input feature of the input vector. Each feature input node is connected to all nodes in the latent layer; therefore, all the latent nodes fully receive the eigenvector.

Hidden nodes are divided into groups, that is, one group for each of the K categories. These groups are associated with eigenvectors in k-th class (here, there is a Gaussian function for each eigenvector). All Gaussian functions in the same category send the category of their values to one node in the output layer. Thus, there will be the number of K output nodes. (6)

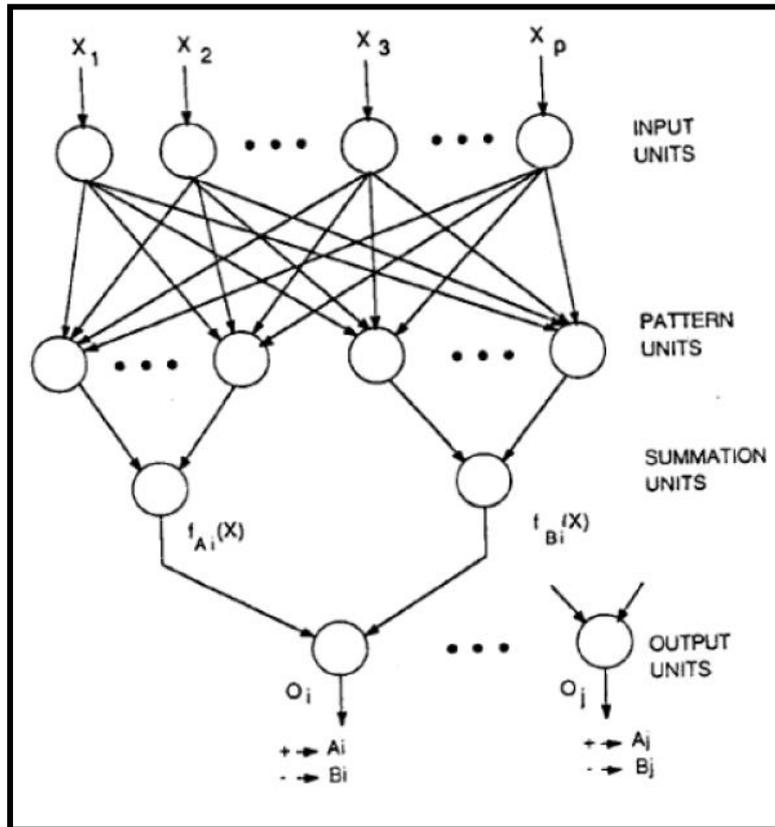


Fig.2.1. Probabilistic neural network structure (6)

2-1-1- Probabilistic neural network performance

In the output node for class K, all Gaussian values of category K gather together and the total is scaled. Since the magnitude of probability is unique under the total function, the total appears as the probability density function. Here special symbols are used for clarity.

P represents the eigenvectors of category 1, i.e. $\{x(p): p = 1, \dots, P\}$ and Q represents the eigenvectors of category 2, i.e. $\{y(q): q = 1, \dots, Q\}$. In the hidden layer, node P is placed in category 1 and Q lies in category 2.

Gaussian equations focused on the points of both categories 1 and 2, $x(p)$ and $y(q)$ are shown in vectors 2-1 and 2-2 for each input vector where N denotes the number of vectors:

Equation (2-1)
$$g_1(x) = \left[\mathbf{1} / \sqrt{(2\pi\sigma^2)^N} \right] \exp\{-||x - x^{(p)}||^2 / (2\sigma^2)\}$$

Equation (2-2)
$$g_2(y) = \left[\mathbf{1} / \sqrt{(2\pi\sigma^2)^N} \right] \exp\{-||y - y^{(q)}||^2 / (2\sigma^2)\}$$

values can be considered as half the size of the mean distances among the eigenvectors in the same group or they can be considered as half the distance of the sample with the nearest sample vector in each sample. The sum of the K-th node, which is obtained from the hidden nodes in the K-th group, is referred to as Parzen and/or Gaussian windows. The sums are defined through equations 2-3 and 2-4 as follows:

$$(Equation\ 2-3) \quad f_1(x) = \left[\frac{1}{\sqrt{(2\pi\sigma^2)^N}} \right] \frac{1}{P} \sum_{p=1,P} \exp\left\{-\frac{\|x - x^{(p)}\|^2}{2\sigma^2}\right\}$$

$$(Equation\ 2-4) \quad f_2(y) = \left[\frac{1}{\sqrt{(2\pi\sigma^2)^N}} \right] \frac{1}{Q} \sum_{q=1,Q} \exp\left\{-\frac{\|y - y^{(q)}\|^2}{2\sigma^2}\right\}$$

Here, x is the input eigenvector; σ represents extension parameters (standard deviation) for Gaussian function in classes 1 and 2. Subsequently, N represents the number of input vectors, P denotes the number of central vectors in class 1, Q shows the number of centers in class 2, and $x^{(p)}$ and $y^{(q)}$ represent centers of classes 1 and 2, respectively. $(\|x - x^{(p)}\|^2)$ is the Euclidean distance between x and $x^{(p)}$. Each input vector x is placed in both functions of $f_1(x)$ and $f_2(x)$ and the maximum values of the two functions determine the output class. For problems with more than two classes, the same process is at play. Here there are no repetition and weight calculations (10). For more information about probabilistic neural networks, readers are welcomed to refer to SPECT 1990 (10).

2-2- PNN Application in marketing and customer choice model

Probabilistic neural networks have been used in numerous studies in the field of marketing as mentioned in Table 2-1.

Table 2.1. PNN application in marketing

No.	Year	Author(s)	PNN application
1	1998	Kim	He used PNN to predict the stock market index because PNN enjoys more favorable circumstances for modeling and forecasting(11) .
2	1999	Yang	He used PNN to predict bank insolvency(12) .

3	2005	Gan, et al.	They compared PNN with artificial and regression neural networks to model customer choice. The results indicated that PNN enjoys higher speed and accuracy (4).
4	2009	Mostafa	He used PNN in the efficiency modelling of competitive market for Egyptian companies. PNN has a high potential to classify the relative performance of companies because of its high strength and flexibility(13) .
5	2009	Mostafa	He used the hybrid of data envelopment analysis and PNN to model the efficiency of superior Arab banks. The results showed that neural networks have high predictive power . (14)
6	2009	Ogut	He combined PNNs with support vector machines to predict the manipulation of financial information. This method was compared with logistic regression (logit), discriminant analysis, and Probit classifiers and the results indicated the superior performance of the proposed hybrid model(15) .
7	2010	Cheng, et al.	They combined PNNs with rough sets and C4.5 to predict marketing time model. It was revealed that the hybrid model is more accurate than any of the individual models . (16)
8	2010	Jandaghi & Hashemi	They compared PNNs with discriminant analysis to model product choice (cell phone) with respect to its properties. The results showed that PNNs require fewer variables than discriminant analysis does(17) .
9	2013	Kazemi, et al.	They integrated PNN and Data preprocessing for brand choice modeling and constructing a market response simulator. The evaluation results show that the offered approach outperforms all previous methods, so it can be considered as an effective tool for consumer behavior modeling and simulation. (18)

3- Bee colony algorithm and its application in neural network modeling

Bee colony algorithm is a relatively new method of particle intelligence algorithms and a population-based technique that was developed by Karaboga in 2005. This method is inspired by behavioral intelligence of bees' search. Since the bee colony algorithm has a simple concept and easy implementation, and contains fewer control parameters, it has been used in a variety of optimization problems. (19)

The following is a brief description of the bee colony algorithm:

3-1- Performance of bee colony algorithm

ABC algorithm creates an initial population i from the solutions ($i= 1,2,\dots,E_b$), which have been distributed randomly. Here, i denotes the population size and E_b is representative of the number of employed bees. Each x_i solution includes a D -dimensional vector where D represents the number of parameters that must be optimized. The location of a food source in bee colony algorithm represents a possible solution for the optimization problem. After initialization, the population situation is placed in iterative cycles of search process for employed bees, onlooker bees, and scouts (cycle = 1, 2, ..., MCN) where MCN is the maximum number of cycles for the search process. Then, employed bees correct the position in their memories based on local information (visual information) and embark on assessing the nectar rate of the new situation (modified solution). If the new amount of nectar were more than the previous nectar, bees will maintain the new position and will erase the previous position from their minds. Otherwise, they hold the previous position in their memory.

After all the employed bees carried out the search process, they provide information about the nectar rate and their position to the onlooker bee that are waiting on the dance place. The onlooker bee examines all the information taken from the employed bees and selects the food source due to the possible amount of the nectar. The onlooker bee and employed bees apply the same procedure regarding the change of criteria and situations. Greedy selection is appropriate for the optimization problems without constraints.

The probability of selecting food source P_i by scouts is calculated according to equation 3-1:

(Equation 3-1)

$$P_i = \frac{fit_i}{\sum_{n=1}^{E_b} fit_n}$$

It is noteworthy that fit_i shows the nectar rate of the i -th source (the fitness value of the i -th solution), which is calculated by employed bees. In addition, E_b represents the number of food sources, which is equal to the number of employed bees (in real problems, E_b represents the number of possible solutions). Clearly, a good food source attracts a higher number of onlooker bees. Consequently, onlookers select their preferred food source and create the food situation of the neighbor $i+1$ to make the i -th choice. The same choice feature of employed bees is used by onlookers. Furthermore, if the i -th solution is not improved in a specific number of iterations (limits), the employed bee discards it in accordance with the solution. It is then converted into a scout and searches for new food sources. When a new position is determined, another cycle of ABC algorithm begins. The same procedure is repeated until the stopping criterion occurs.

The ABC algorithm changes a parameter that has been selected randomly and leaves the other parameters unchanged in order to determine the neighborhood of the food position.

In other words, the sum of the current parameter and the numerical uniform parameter from the interval $[-1, 1]$ and also the difference between the selected parameter value and the random parameter value determine the neighboring position of the food source. Equation 3-2 clearly shows it as follows:

$$(Equation 3-2) \quad V_{ij} = X_{ij} + \Phi_{ij}(X_{ij} - X_{kj})$$

In equation 3-2, both parameters are the members of $\{1, 2, \dots, E_b\}$ and $k \neq i$.

Moreover, Φ factor represents a random value between 1 and -1 and $j \in \{1, 2, \dots, D\}$. In other words, X_{ij} is the j -th parameter of solution X_i , which has been selected to be changed. Here V_{ij} is the new candidate's position for food source, which is obtained according to the formula.

When food position is discarded, the employed bee becomes a scout. Then, the scout creates an entirely new food position using equation 3-3:

$$(Equation 3-3) \quad X_i^{j(\text{new})} = \min x_i^j + \Phi(\max x_i^j - \min x_i^j)$$

This equation is used for each parameter j and Φ is a random value within the range of 1 and -1. If the parameter value (obtained using equation 2-7 or 7-2) exceeds the predetermined limit, the parameter can be considered as an acceptable value. (20, 21)

Clearly, employed bees and onlookers select the new food source based on the neighborhood, visual information, and the comparison of food sources with each other. However, scouts search for food sources without and guidance and discover a completely new source. In the case of artificial bees, scouts can have high speed possible solution discoveries. (8)

Parameter setting in meta-heuristic optimization algorithms obviously affects the algorithm performance. Divergence, entrapment in local extremum, and time consumption are the consequences of inappropriate parameter setting. ABC algorithm has the advantage of having a low number of control parameters. ABC algorithm is not dependent on the initial population because its initialization is sometimes heavy with the feasible region at random. In return, its performance effectively directs the population to the feasible region. (22)

For more information on the neural network training by bee colony algorithm, readers are referred to Karaboga et al., 2007. (19)

3-2- Application of ABC algorithm in neural network training

Bee colony algorithm has been used in numerous articles on in neural network training, which is presented Table 3-1.

Table 3.1. Application of bee colony algorithm in neural network training

No.	Year	Author	Result
1	2007	Karaboga et al	They used the ability to explore and utilize bee colony algorithm in search for optimal weights (19).
2	2011	Ozturk, et al.	They combined bee colony algorithm with Levenberg-Marquardt method for training the neural network. Finally, they concluded that the use of this hybrid approach would be much better than traditional methods because traditional methods may be trapped in local optimum and are slow (23).
3	2011	Ozkan, et al.	They calculated the mean absolute error of the hybrid approach of neural network combined with bee colony algorithm and discovered that the hybrid method would be superior to artificial neural networks in terms of evapotranspiration (24).

4	2011	Hsieh	She combined recurrent neural network with bee colony algorithm to predict the stock price (25).
5	2011	Irani & Nasimi	They showed that this hybrid method acts better than the methods based on gradient descent. The combination of neural network with bee colony algorithm increases the predictive power of neural networks (26).
6	2012	Yeh & Hsieh	They compared the performance of neural network training with genetic algorithm, bee colony algorithm, and particle swarm optimization and concluded that ANN-ABC is superior (27).
7	2012	Kisi, et al.	They showed that this hybrid method acts better than fuzzy neural method, neural network, curve points, and differential evolution and enjoys higher accuracy (28).
8	2014	Uzlu, et al.	They used ANN model with the ABC algorithm to estimate annual hydraulic energy production of Turkey. They compared ANN-ABC model performance with results of classical ANN models trained with the BP (back propagation) algorithm. Results showed the ANN-ABC method predicted hydroelectric generation better than the classical ANN trained with the BP algorithm.(29)
9	2015	Wang, et al.	They developed system for fruit classification consists of wavelet entropy (WE), principal component analysis (PCA), feed forward neural network (FNN) trained by fitness-scaled chaotic artificial bee colony (FSCABC) and biogeography-based optimization (BBO). They results showed that their new methods outperform better than the previous ones.(30)
10	2014	Karaboga, et al.	They did A comprehensive survey about artificial bee colony (ABC) algorithm and applications and they mentioned some more ANN trained with Bee colony.(31)

For more information about training ANN with ABC readers referred to Karaboga et al. (2014).
(31)

4- Hybrid model of PNN and bee colony algorithm for customer choice modeling and prediction

In this model, probabilistic neural networks have been used to model and predict consumer choice in the market, but with the difference that it has been tried to find the best variance for classes. For this purpose, bee colony algorithm has been applied. The model steps are as follows.

First step: data entry

At this step, the required data of the problem is entered. It should be noted that the data must be suitable and free of any distorted data.

Second step: creation of a random population

First, a random population of 10 food positions (random possible solutions) is created for the standard deviation of each class (there are six classes in the present case study) within the range of .0001 to 2. These values are considered as the standard deviation of the PNN. In fact, a PNN is produced in each iteration and the model is implemented. In each iteration, the number of classification errors is obtained. (As it was mentioned earlier, bee colony algorithm parameters are determined using trial and error method and user skills. In the current case study, the number of 10 initial population and the range of .0001 to 2 have been selected after several times running the model with various values because they result in the best answers.)

Third step: Calculation of the fitness of each population member

As it was mentioned earlier, the number of the created positions equals the number of employed bees and each employed bee goes to a position. Value of the fitness function should be calculated for every member of the population. Here, the number of misclassifications is calculated, equation 4-1. Since this number should be minimized and the algorithm seeks a position with maximum fitness function, so the fitness function of each member of the population is obtained via equation 4-2:

(Equation 4-1)
$$fit_i = \text{number of misclassification}$$

The population members should maximize the following relation:

(Equation 4-2)
$$fit_i = \frac{1}{1+fit_i}$$

The maximum value of the above equation is considered as the superior answer. At this step, the initial is applied in probabilistic neural network.

Fourth step: creation of a new position in the neighborhood of employed bees

Employed bees change one of the parameters to determine the new position. Employed bees provide the onlookers with information about the positions. Onlookers seek the points with the highest amount of nectar and they may not find a better answer. As a result, scouts look for a better answer. New answers are produced using the equation 4-3 with regard to the available initial population. Here, there are the number of j dimensions (equal to the classes) and i and k range from 1 to 10. New s are calculated.

(Equation 4-3)

$$V_{ij} = X_{ij} + \Phi_{ij}(X_{ij} - X_{kj})$$

Fifth step: calculation of the fitness function for the changed position

The value of fitness function for new answers is calculated.

Sixth step: greedy selection

Greedy selection is conducted between X_{ij} and V_{ij} . If the fitness function V_{ij} is better than X_{ij} , it will be selected and will replace X_{ij} . The fitness function V_{ij} should be compared with the 10 initial positions; otherwise, a new food source is created similar to the method mentioned in the previous step.

If all the observer bees (onlookers) have not been distributed, the algorithm will be transferred to the seventh step and if they have been distributed, the algorithm will be transferred to the tenth step.

Seventh step: Calculation of P_i values for X_i positions

P_i , probability, is calculated according to the equation 4-2. This probability actually represents the possibility of a position selection by the onlooker bees. (In fact, P_i is the nectar of the food source in the main algorithm is the situation. The more the nectar, the greater the likelihood that the onlooker bees go to that position.)

The number of 10 food sources has been considered in this case study.

Eighth step: X_i positioning with the highest P_i value

As it was mentioned above, the position with a higher P_i contains more nectar. In other words, this is the best position in real problems and is closer to the optimal solution.

Ninth step: change of the determined position of X_i

The onlooker creates a new position in the neighborhood of point P_i .

Tenth step: reaching the value of limit

All the above steps are repeated to achieve the best answer or ultimate iterations. If the value of limit was not achieved and the best answer was obtained: In this case, X_i would be selected as the best food position when the number of cycles has been finished. In addition, the algorithm will return to the fourth step when the number of cycles has not been finished.

If it reached the value of limit, an answer higher the limit (it equals 100 in the present case study) was repeated, and no improvement occurred; then, the algorithm progresses to the eleventh step. (The parameter's value of limit has been obtained by the user based on trial and error method.)

Eleventh step: creation of a new food position using equation 4-4

In this step, a new answer is created by scouts according to the equation 4-4.

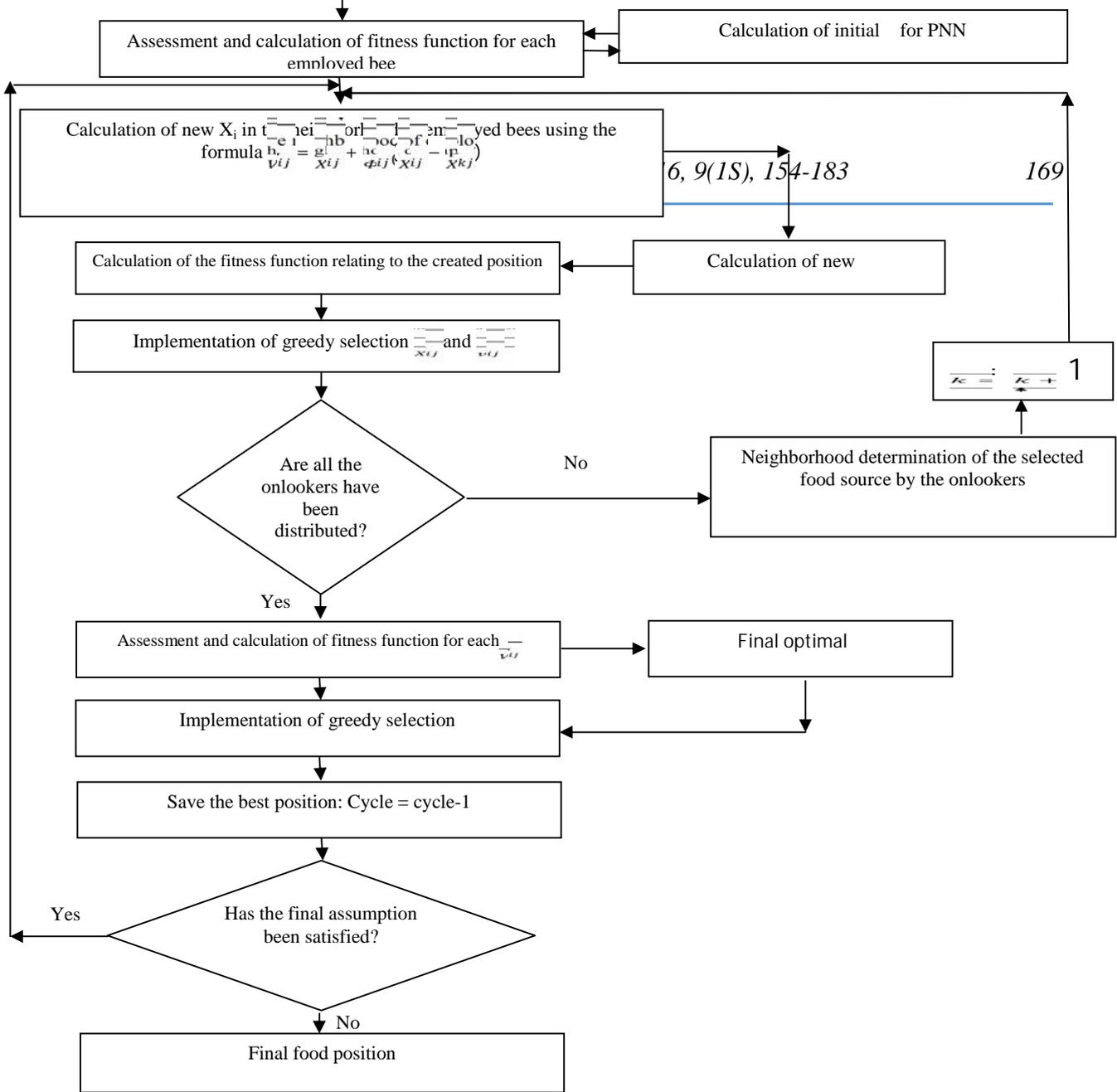
(Equation 4-4)
$$X_i^{j(\text{new})} = \min x_i^j + \Phi(\max x_i^j - \min x_i^j)$$

Twelfth step: calculation of fitness function for new food positions

In this step, the values of fitness function for new positions are assessed and examined. The best position is kept in memory and the above steps are repeated until the optimal solution is achieved.

4-1- Schematic view of the model

In Figure 4-1 below, the proposed model is presented in the form of a flowchart.



4-2- Comparison of the model with other models

In this section, the hybrid PNN-ABC and artificial bee colony algorithm is compared with other conventional models in predicting and modeling customer choice.

Logit models are traditionally used for the customer choice problem. Even, some scholars use these methods for the modeling and prediction of consumer behavior. As it was mentioned above, the most important constraint of the logit function a linear form of the profit function which is used to calculate the probability or odds ratio of make a particular decision. This method requires much more time to examine the relationship between different functions and different possible algorithms towards obtaining the appropriate model. Moreover, the relationships between the objective function and effective factors are non-linear and complex. In real classification

problems such as brand choice modeling, it is not possible to purely and incrementally produce an explicit assumption of the linear effects out of the variables.

Artificial neural network with back-propagation training algorithm is an exemplary model for the modeling of marketing problems. wherein have been

The neural networks used in almost all marketing studies are training neural networks with back-propagation algorithm. For more information on back-propagation algorithms, interested readers are referred to Rumelhart et al. (32)

Although the recursive back-propagation algorithm is a highly common neural network training technique, there are four primary reasons for the inadequacy of this algorithm for the training neural network:

- 1- Recursive algorithm search process is based on a point-to-point model which is inherently convergent to local extremum. Therefore, the trained neural network with recursive back-propagation algorithm usually leads to erroneous optimization in the convergence process.
- 2- Recursive back-propagation algorithm can lead to temporary instability. This is due to the network proneness to forgetting what has been learned in the past. Indeed, the node weights are updated after the beginning of each course.
- 3- The node weights are set proportional to the node outputs and the inputs are transformed using non-linear cyclic function. Accordingly, the neural network may undergo some inability when the nodes weight move to higher levels.
- 4- Recursive back-propagation algorithm requires a differentiable objective function that constrains the problem that can be solved by neural networks. (33)

4-3- The reasons for the selection of evolutionary algorithms to combine with probabilistic neural network

Here are the advantages and disadvantages of evolutionary algorithms (bee colony algorithm is also part of such algorithms):

- Gradient-based optimization algorithms and back-propagation error algorithm are very fast.
- Gradient-based optimization algorithms and back-propagation error algorithm suffer from the problem of falling into the trap of local minima.
- Gradient-based optimization algorithms and back-propagation error algorithm are only applicable to a specific and standard class of neural networks, such as multilayer

perceptron and RBS, which have a closed form solution for the derivative of the objective function relative to the weights.

- Evolutionary optimization algorithms are slow in comparison with gradient-based algorithms.
- Evolutionary optimization algorithms have *the ability to escape from local optimum trap*.
- Evolutionary optimization algorithms are not dependent on a specific structure of the network and, thereby, they are applicable to any defined structure.

Bee colony algorithm is an optimization tool with the ability to find the optimal solution with relatively low computational requirements. (7)

Bee colony algorithm is one of the evolutionary algorithms, which has a simple concept and easy implementation, and contains fewer control parameters.

ABC algorithm consists of three control parameters, including the number of colony members, the value of limit, and the maximum cycle number (MCN). Colony members include an equal number of employed bees and onlookers. In fact, the positions created in the beginning of the algorithm should be set to be equal to the number of employed bees. However, when the initial population parameter is determined by the user for the first time, the value of limit will be easily calculated. Therefore ABC algorithm contains two parameters that must be set: initial population values and MCN.

5- Case study of PNN-ABC hybrid model

5-1- Data

In this section, same data as those used by Kalwani , et al. in 1990 and Fish, et al. in 2004 have been used. These data have been used since they pertain to three various brands and two sizes; therefore, they suit the customer choice problem, which examines the customer choice among the available brands. In addition, the size of the data is suitable for the implementation of neural networks because neural networks need a large volume of data for training and experimental phases. The data include coffee purchases by customers on an individual basis from four supermarkets in Kansas City over a period of 65 weeks. It is noteworthy that the data of 25 weeks were used to calculate the size and brand loyalty and the remaining 40 weeks are used for fitness/training and holdout set. Data for the remaining 40 weeks are divided into two categories, 1564 data for training and 1236 data for the holdout set have been considered. These data were cleaned by previous studies. The decision variables and output variable are listed in the table

below. Since the retailers individually improve coffee brands and sizes (discount), brand-sizes have been modeled as choice options. For example, Floger trademark contains large-sized three-pound packages and small-sized one-pound packages in the dataset. Each of these compounds constitutes a different brand-size. The six compounds with the highest sale rates in the market (three brands, each containing two sizes) were maintained and the sum of these six compounds included 82.4% of the total number of purchases and 87.1% of market share. Each of the brand-sizes that was removed have taken up than 1% of the market share. (34)

Table 5.1. Input and output variables for customer choice problem

Input variable	Description
$R_{i k}^{(n)}$	Original price of option i at the time of the n-th purchase by the k-th customer
$PROM_{i k}^{(n)}$	Variable 0 and 1, which indicates the presence or absence of discounts in the price of option i at the time of the n-th purchase by the k-th customer
$PCUT_{i k}^{(n)}$	The discounted price of option i at the time of the n-th purchase for the k-th customer
$PRV_{i k}^{(n)}$	Variable 0 and 1, which indicates whether the previous purchase of the k-th customer has been followed by a discount on the option with the same brand i at the time n or not
$SPRV_{i k}^{(n)}$	Variable 0 and 1, which indicates whether the two previous purchases of the k-th customer has been followed by a discount on the option with the same brand i at the time n or not
$BL_{i k}^{(n)}$	Customer loyalty to the brand of the i-th choice at the time of the n-th purchase
$SL_{i k}^{(n)}$	Customer loyalty to the size of the i-th choice at the time of the n-th purchase
Output variable	
$BC_{i k}^{(n)}$	Selection of the i-th option by the k-th customer at the time of the n-th purchase (34)

For an in-depth discussion of model's independent variables including their theoretical foundations and other informations, the reader is referred to Guadagni and Little (35)

The data should be normal to implement the models. To this end, all the data should lie in the range [0, 1] so that all the data can be placed in the same range. Indeed, the effects of the data in a larger range may exceed the limit. For this purpose, the datum of each Eigen has been divided by the largest amount of data in that Eigen and the processing operation of the data is done.

5-2- First stage: customer choice

Since the problem consists of 7 independent variables and 6 choice modes for each customer, there is the number of 42 input nodes. Agrawal et al.; and Fish, et al. could obtain a hidden layer with 11 nodes through the optimization of the process. Furthermore, the number of hidden nodes is equal to the selected options, that is, 6. With the modeling and implementation of this network structure, the number of 626 errors was obtained after 1500 iterations among the 1236 test data.

In neural networks, the model accuracy and complexity increase and computing speed comes down in parallel with increasing the number of hidden layers. Thus, a neural network with two hidden layers was built so that accuracy would increase. In this way, the number of errors reached 539 in the 1500-th iteration.

Since the problem contains 7 independent variables and 6 selection states for each customer, there will be the number of 42 input nodes for PNNs. The next layer in the PNN is the one in which all the 42 input nodes are connected to the 1564 data of the problem and the 1564 data are divided into 6 classes (this problem has 6 output nodes). Then, there exist the total units of each group and, at the end, the group with the highest aggregate is chosen.

The PNN network has 522 errors after modeling. The PNN has been used for the hybrid model because the number of errors is lower than that of feed-forward multilayer neural network with recursive back-propagation training and the computing speed is also higher.

In the next stage, the neural network containing 42 input nodes, 11 hidden nodes, and 6 output nodes is trained by the genetic algorithm to determine the pertaining weights. Fish, et al. in 2004 did so in their study. Their model was implemented; and 3000 iterations and 440 errors were recorded where the presented answer is better than the neural network with back-propagation training.

Finally, PNN-ABC hybrid model is implemented for customer choice. Bee colony algorithm includes the following control parameters:

- 1- CS: It includes the employed bees (E_b) and onlooker bees (O_b)
- 2- The value of limit: It entails the number of attempts to stop the food source
- 3- The maximum cycle number (MCN)

Although ABC algorithm has three parameters that must be set, when the CS parameter is determined by the user for the first time, the value of limit is easily calculated. As a result, ABC algorithm comes out with two parameters that must be set: CS and MCN values.

Here, the number of CS has been considered equal to 20. Thus, an initial population equal to 10 is created (the initial population is equal to the number of the employed bees). In addition, MCN is considered equal to 1458.

For the conduct of this experiment, the data are subdivided into three categories of training data, validation data, and holdout data.

Validation data have been used since they direct the algorithm towards better responses when implemented. To this end, 20% of the training data has been dedicated to the validation set, which is almost equal to 313 samples. The number of 391 errors occurred in iteration 1458 after running the model.

This shows that the proposed model has the lowest number of errors among the 5 above-mentioned models and also enjoys higher speed and accuracy. In Table 5-2, the number of errors in customer choice prediction is presented in the holdout set of various models.

Table 5.2. The number of errors in customer choice prediction in the holdout set

<i>Model</i>	<i>Number of errors</i>	<i>Number of holdouts</i>
ANN with back-propagation training method containing one hidden layer	626	1236
ANN with back-propagation training method containing two hidden layers	539	1236
PNN	522	1236
ANN with training method of genetic algorithm	440	1236
PNN with training method of bee colony algorithm	391	1236

5-2-1- Comparison of classifiers

In this section, different methods of classification are compared with each other. Therefore, the various models included in this study have once been compared with each other based on the overall accuracy and, next time, their classification accuracy have been tested on each class one by one.

5-2-1- Comparison of classifiers based on the overall accuracy of each model

In this section, the models are assessed based on their prediction accuracy as listed in Table 5-3.

Table 5.3. Prediction accuracy of models

<i>Model</i>	<i>Number of correct predictions</i>	<i>Number of holdout samples</i>
ANN	697	1236
PNN	714	1236
ANN-GA	796	1236
PNN-ABC	845	1236

As it can be observed in the table, PNN-ABC model is more accurate in predicting customer choice. Therefore, it can be argued that this model can be used as a tool to account for the marketing policies.

5-2-1-2- Comparison of the models based on prediction accuracy of individual classes

For the comparison of the models based on prediction accuracy of the individual classes, the number of properly categorized samples is divided by the total number of samples pertaining to that class. The obtained answer is considered as the classification accuracy for each class. The results of accuracy calculation of classes are presented in Tables 5-4, 5-5, 5-6, and 5-7 for each of the models. By correctly predicted samples, it is meant that these data, for example, have belonged to category 1-1 (1) and they have then been correctly predicted in this category. On the other hand, the misclassified model samples refer to the situation wherein the data, for example, have belonged to category 1-1 (1), but they have then been predicted in other classes by mistake. Since the neural network model with one hidden layer has shown weaker results compared to the neural network model with two hidden layers, the analyses pertinent to this model are not presented in the following.

Table 5.4. Comparison of classifiers' accuracy-neural network model with two hidden layers

Option	The number of relevant samples	The number of samples correctly classified by the model	The number of samples misclassified by the model	Classifier's accuracy
1-1	340	187	153	0.55
2-1	63	15	48	0.238
1-2	492	324	168	0.6585
2-2	83	32	51	0.3855
1-3	215	129	86	0.6
2-3	43	10	33	0.2325

Table 5.5. Comparison of classifiers' accuracy-PNN model

Option	The number of relevant samples	The number of samples correctly classified by the model	The number of samples misclassified by the model	Classifier's accuracy
1-1	340	208	132	0.6117
2-1	63	13	50	0.2063
1-2	492	322	170	0.6544
2-2	83	32	51	0.3855
1-3	215	129	86	0.6
2-3	43	10	33	0.2325

Table 5.6. Comparison of classifiers' accuracy-neural network model trained by evolutionary algorithm (genetic algorithm)

Option	The number of relevant samples	The number of samples correctly classified by the	The number of samples misclassified by the model	Classifier's accuracy
---------------	---------------------------------------	--	---	------------------------------

	samples	model		
1-1	340	248	92	0.7294
2-1	63	19	44	0.3015
1-2	492	376	116	0.7642
2-2	83	21	62	0.2530
1-3	215	128	87	0.5953
2-3	43	4	39	0.0930

Table 5.7. Comparison of classifiers' accuracy-PNN-ABC hybrid model

Option	The number of relevant samples	The number of samples correctly classified by the model	The number of samples misclassified by the model	Classifier's accuracy
1-1	340	257	83	0.7758
2-1	63	30	33	0.4761
1-2	492	389	103	0.7906
2-2	83	21	62	0.253
1-3	215	136	79	0.6325
2-3	43	12	31	0.2790

In Figure 5-1, the models' ability to correctly place the samples in classes has been examined. As it can be observed, the ABC- PNN model has the highest efficiency in the correct classification of the samples compared to other models. This model has taken up the highest accuracy in all categories other than the fourth one. Therefore, it can be concluded that the proposed model is more accurate than the other models in classification and the model can be used for classification.

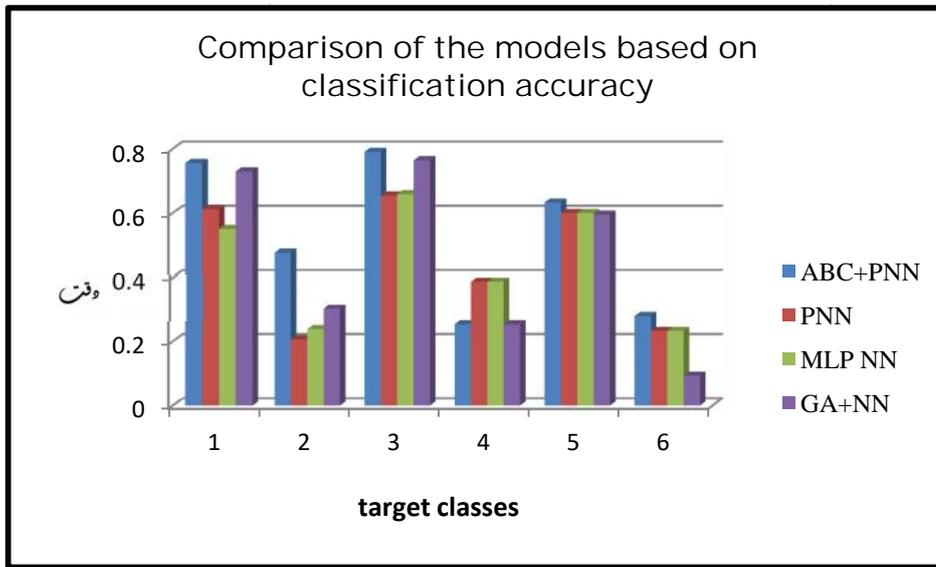


Fig.5.1. The ability of models in sample classification

5-3- Validation of PNN-ABC model using price sensitivity analysis

According to the results obtained in the previous section, PNN-ABC model took advantage of higher accuracy and speed in predicting customer choice; therefore, in this section, price sensitivity analysis has been conducted to demonstrate the usability and applicability of PNN-ABC model. For this purpose, 25 cents, 50 cents, and a dollar has been added to and then reduced from each of the six brand-size options in terms of price, respectively. In this way, the application of the proposed model is examined in terms of marketing and pricing. In Table 5-8, the price sensitivity analysis results obtained for each brand-size option are presented. In each of these columns, the number of customers who opt for this option due to the price change is given.

Table 5.8. Price sensitivity analysis using PNN-ABC model

<i>-\$1</i>	<i>-50 cents</i>	<i>-25 cents</i>	<i>+\$1</i>	<i>+50 cents</i>	<i>+25 cents</i>	<i>With the initial price</i>	<i>Brand-size</i>
451	436	402	238	256	269	349	1-1
97	81	72	10	13	17	58	2-1
712	691	636	433	448	464	544	1-2
263	232	181	81	87	93	156	2-2
455	431	396	222	243	249	348	1-3
161	146	127	73	78	82	109	2-3

The results are shown in Figure 5-2. As it is observed in the figure, PNN-ABC model displays a true predictability using price sensitivity analysis. In other words, there is an inverse relation between price and the number of customers, which means that the number of customers is reduced with rising prices and vice versa. Accordingly, it can be concluded that PNN-ABC model can be used as a strategic tool for the adoption of marketing policies and it can be offered to managers as the simulator of market reactions.

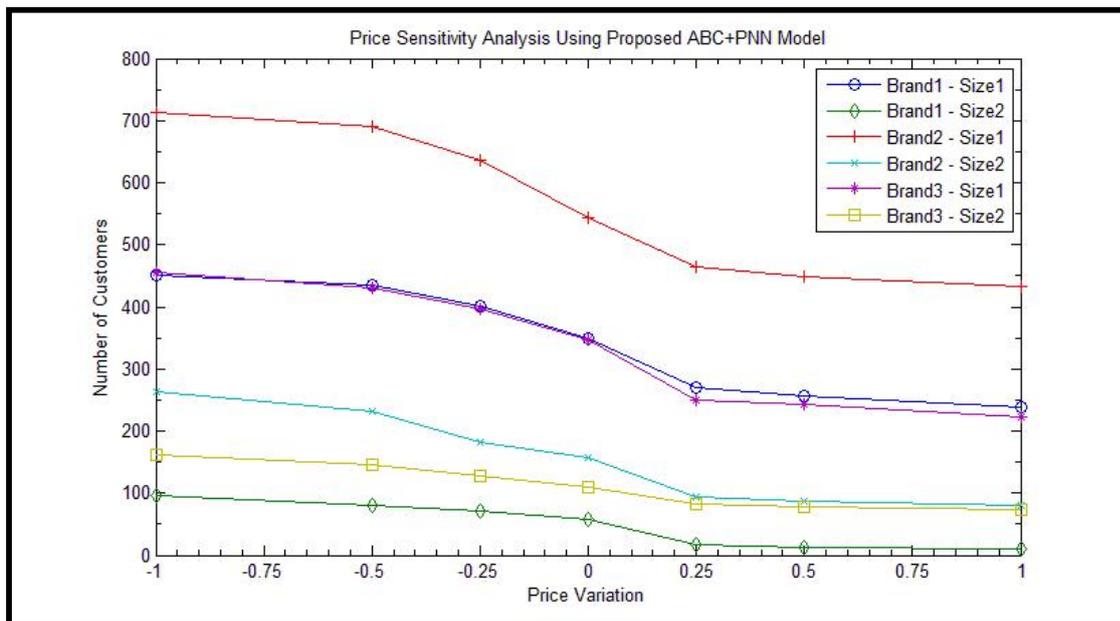


Fig.5.2. Results of price sensitivity analysis using PNN-ABC model

6- CONCLUSION

The typical structure of feed-forward neural network for this problem is very complex and it does not operate satisfactorily, as well. In this study, it was indicated that PNN outperforms the standard structure of feed-forward neural network both in speed and accuracy. Hence, probabilistic neural network was used in this study. Then, such models as standard feed-forward neural network, neural network with two hidden layers, probabilistic neural network, neural network trained with genetic algorithm, and probabilistic neural network trained with bee colony algorithm were implemented using the data pertaining to market coffee. The results showed that PNN-ABC is superior to all models in terms of accuracy and speed, which means that it shows fewer errors in shorter time. Those managers and planners dealing with a greater number and

volume of data are recommended to assign credit to the models with a simpler structure, high accuracy, and high speed. Therefore, PNN-ABC model is very suitable for this purpose.

It is noteworthy that all the models were compared with each other based on prediction accuracy of individual classes to ensure the superior performance of the proposed model. For this purpose, the correctly classified samples were divided by the total number of samples of that class and the resulting answer has been regarded as the classification accuracy for each class. As per the results, it was proved that PNN-ABC model is more accurate than other models and, thereby, this model can be used for classification problems.

Finally, price sensitivity analysis was performed on the proposed model to demonstrate its applicability. The results of the analysis showed that PNN-ABC model would demonstrate a correct predictability.

Based on the studies conducted in this area and the obtained results, it can be concluded that the new hybrid PNN-ABC model can be used as a strategic tool for managers and decision-makers to determine the exact marketing policies and it can act as the simulator of market reactions.

REFERENCES

1. Baltas, George. "A model for multiple brand choice." *European Journal of Operational Research* 154, no. 1 (2004): 144-149.
2. Banerjee, Arindam, Dheeraj Awasthy, and Vivek Gupta. "A choice modelling approach to evaluate effectiveness of brand development initiatives." *International Journal of Management and Decision Making* 6, no. 2 (2005): 180-198.
3. Chintagunta, Pradeep, Ekaterini Kyriazidou, and Josef Perktold. "Panel data analysis of household brand choices." *Journal of Econometrics* 103, no. 1 (2001): 111-153.
4. Gan, Christopher, V. Limsombunchao, Mike Clemes, and Amy Weng. "Consumer choice prediction: Artificial neural networks versus logistic models." (2005).
5. van Wezel, Michiel, and Rob Potharst. "Improved customer choice predictions using ensemble methods." *European Journal of Operational Research* 181, no. 1 (2007): 436-452.
6. Specht, Donald F. "Probabilistic neural networks." *Neural networks* 3, no. 1 (1990): 109-118.
7. Yeh, Wei-Chang, and Tsung-Jung Hsieh. "Artificial bee colony algorithm-neural networks for S-system models of biochemical networks approximation." *Neural Computing and Applications* 21, no. 2 (2012): 365-375.

8. Karaboga, Dervis, and Bahriye Basturk. "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm." *Journal of global optimization* 39, no. 3 (2007): 459-471.
9. Karaboga, Dervis, and Bahriye Akay. "A comparative study of artificial bee colony algorithm." *Applied mathematics and computation* 214, no. 1 (2009): 108-132.
10. Specht, Donald F. "Probabilistic neural networks and the polynomial adaline as complementary techniques for classification." *IEEE Transactions on Neural Networks* 1, no. 1 (1990): 111-121.
11. Kim, Steven H., and Se Hak Chun. "Graded forecasting using an array of bipolar predictions: application of probabilistic neural networks to a stock market index." *International Journal of Forecasting* 14, no. 3 (1998): 323-337.
12. Yang, Z. R., Marjorie B. Platt, and Harlan D. Platt. "Probabilistic neural networks in bankruptcy prediction." *Journal of Business Research* 44, no. 2 (1999): 67-74.
13. Mostafa, Mohamed M. "Modeling the competitive market efficiency of Egyptian companies: A probabilistic neural network analysis." *Expert Systems with Applications* 36, no. 5 (2009): 8839-8848.
14. Mostafa, Mohamed M. "Modeling the efficiency of top Arab banks: A DEA–neural network approach." *Expert Systems with Applications* 36, no. 1 (2009): 309-320.
15. Ö üt, Hulisi, Ramazan Akta , Ali Alp, and M. Mete Do any. "Prediction of financial information manipulation by using support vector machine and probabilistic neural network." *Expert Systems with Applications* 36, no. 3 (2009): 5419-5423.
16. Cheng, Jao-Hong, Huei-Ping Chen, and Yi-Min Lin. "A hybrid forecast marketing timing model based on probabilistic neural network, rough set and C4. 5." *Expert systems with Applications* 37, no. 3 (2010): 1814-1820.
17. Jandaghi, Gholamreza, and Ommolbanin Hashemi. "Modeling the Selection of a Product Through its Attributes by Using Probabilistic Neural Networks and Discriminant Analysis." *World Applied Sciences Journal* 8, no. 8 (2010): 959-964.
18. Kazemi, S. M. R., Esmaeil Hadavandi, Farhad Mehmanpazir, and Mohammad Masoud Nakhostin. "A hybrid intelligent approach for modeling brand choice and constructing a market response simulator." *Knowledge-Based Systems* 40 (2013): 101-110.

19. Karaboga, Dervis, Bahriye Akay, and Celal Ozturk. "Artificial bee colony (ABC) optimization algorithm for training feed-forward neural networks." In *International Conference on Modeling Decisions for Artificial Intelligence*, pp. 318-329. Springer Berlin Heidelberg, 2007.
20. Shukran, Mohd Afizi Mohd, Yuk Ying Chung, Wei-Chang Yeh, Noorhaniza Wahid, and Ahmad Mujahid Ahmad Zaidi. "Artificial bee colony based data mining algorithms for classification tasks." *Modern Applied Science* 5, no. 4 (2011): 217.
21. Karaboga, Dervis. *An idea based on honey bee swarm for numerical optimization*. Vol. 200. Technical report-tr06, Erciyes university, engineering faculty, computer engineering department, 2005.
22. Karaboga, Dervis, and Bahriye Basturk. "Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems." In *International Fuzzy Systems Association World Congress*, pp. 789-798. Springer Berlin Heidelberg, 2007.
23. Ozturk, Celal, and Dervis Karaboga. "Hybrid artificial bee colony algorithm for neural network training." In *2011 IEEE Congress of Evolutionary Computation (CEC)*, pp. 84-88. IEEE, 2011.
24. Ozkan, Coskun, Ozgur Kisi, and Bahriye Akay. "Neural networks with artificial bee colony algorithm for modeling daily reference evapotranspiration." *Irrigation Science* 29, no. 6 (2011): 431-441.
25. Hsieh, Tsung-Jung, Hsiao-Fen Hsiao, and Wei-Chang Yeh. "Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm." *Applied soft computing* 11, no. 2 (2011): 2510-2525.
26. Irani, Rasoul, and Reza Nasimi. "Application of artificial bee colony-based neural network in bottom hole pressure prediction in underbalanced drilling." *Journal of Petroleum Science and Engineering* 78, no. 1 (2011): 6-12.
27. Yeh, Wei-Chang, and Tsung-Jung Hsieh. "Artificial bee colony algorithm-neural networks for S-system models of biochemical networks approximation." *Neural Computing and Applications* 21, no. 2 (2012): 365-375.
28. Kisi, Ozgur, Coskun Ozkan, and Bahriye Akay. "Modeling discharge–sediment relationship using neural networks with artificial bee colony algorithm." *Journal of Hydrology* 428 (2012): 94-103.

29. Uzlu, Ergun, Adem Akpınar, Hasan Tahsin Öztürk, Sinan Nacar, and Murat Kankal. "Estimates of hydroelectric generation using neural networks with the artificial bee colony algorithm for Turkey." *Energy* 69 (2014): 638-647.
30. Wang, Shuihua, Yudong Zhang, Genlin Ji, Jiquan Yang, Jianguo Wu, and Ling Wei. "Fruit classification by wavelet-entropy and feedforward neural network trained by fitness-scaled chaotic ABC and biogeography-based optimization." *Entropy* 17, no. 8 (2015): 5711-5728.
31. Karaboga, Dervis, Beyza Gorkemli, Celal Ozturk, and Nurhan Karaboga. "A comprehensive survey: artificial bee colony (ABC) algorithm and applications." *Artificial Intelligence Review* 42, no. 1 (2014): 21-57.
32. DE Rumelhart, GE Hinton, RJ Williams; Learning representations by back-propagating errors. *Nature (London)*, 323 (1986), pp. 533–536.
33. Fish, Kelly E., John D. Johnson, Robert E. Dorsey, and Jeffery G. Blodgett. "Using an artificial neural network trained with a genetic algorithm to model brand share." *Journal of Business research* 57, no. 1 (2004): 79-85.
34. Kalwani, Manohar U., Chi Kin Yim, Heikki J. Rinne, and Yoshi Sugita. "A price expectations model of customer brand choice." *Journal of Marketing research* (1990): 251-262.
35. J.D.C. Guadagni, Little, A Logit Model of Brand Choice Calibrated on ScannerData, *Mark Sci.* 2 (1983) 203–238.

How to cite this article:

Shahrabi J, Mottaghi Khameneh S. Development of a hybrid system of artificial neural networks and artificial bee colony algorithm for prediction and modeling of customer choice in the market. *J. Fundam. Appl. Sci.*, 2016, 9(1S), 154-183.