Predicting and Optimizing the End Price of an Online Auction using Genetic-Fuzzy Approach

Kodhai E[1], RamPrasath C[2], Bubalan V[3], Pratheeban R[4]

[1]*Associate Professor, Sri Manakula Vinayagar Engineering College, Pondicherry-605104, India*

[2, 3, 4] *Dept. Of Information Technology, Sri Manakula Vinayagar Engineering College, Pondicherry-605104, India*

[1]kodhaiej@yahoo.co.in

[2]cramprasaath@gmail.com

[3]bubalan.art@gmail.com

[4]pratheebanr@gmail.com

***Abstract*— Online auction has now been an admired system in setting bid price by online users, which is being estimated to account for 30% electronic business. Auction price prediction, involving the modeling of uncertainty regarding the bidding process, is a challenging task due to the variety of factors changing the auction environment including human behavior. Thus, this paper focuses on predicting the end price in an online auction using a combination of Fuzzy Logic and Modified Genetic Algorithm (MGA). The prediction of final prices of the English auctions is by using real-world online auction data, which is collected from eBay Auction. The Fuzzy system is used to predict the imprecise value from the data set and then the predicted value along with the latest bid values is given to the hybrid genetic algorithm to result the optimized end-price value.**

***Keywords*—Modified Genetic Algorithm (MGA), Genetic Fuzzy, Fuzzy Sets, Fuzzy Logic System.**

I. INTRODUCTION

The internet started revolutionizing the commercial world in 90's, which paved the way for the online auctions. In mid 90's, the auctions on internet have grown at a tremendous rate. They have been considered as the best way of trading used items for scrap value and new items for profit. Online auction is being estimated to account for 30% electronic commerce by 2010 [1]. It has now been a popular process in setting bid price for online users. Auction price prediction involves in the structuring of improbability regarding the bidding process. Primarily it is a challenging task due to the variety of factors that happens to change in auction settings. Even if all the factors were considered, there is still the uncertainty in human behavior while bidding at auctions.

The prediction of the end price in an online auction is by using a combination of algorithms and logics. The consideration of the psychology determines the varying nature of the human mind with respect to the auction environment. It makes the patterns in the bidding process to be non-linear and so the linear prediction systems are not enough to predict the end price in an online auction. Hence, non-linear logic is used to bring in variables in auction that cannot be expressed in numerical terms otherwise. These factors have a profound effect on the auction bids and hence a suitable logic is used to represent these factors.

The system is constructed to predict the final prices of the auctions, using real-world online auction data, collected from eBay Auction. The observed results show that a non-linear system can catch the complicated relationship among the variables accurately much better than the others, which is of great help for the buyers to avoid overpricing and for the sellers to facilitate the auction.

II. RELATED WORKS

The economists have special interests on online auction and bidding process, several logics and techniques were proposed by researchers for predicting end bid prices. In such way, researchers proposed the prediction logic using the concepts of Data Mining, Fuzzy Logic, Genetic Algorithm, Neuro Fuzzy, Grey System Theory, etc.

A. Data Mining Concepts

Data Mining [3] technique are being approached using Neural Network and Bayesian Network. To forecast the winning bid prices, there includes four processes. They are *Data Collection, Variable Selection, Data Transformation* and *Data Mining [21]*. *Data Collection* is for identifying the available data from sources and to extract the data. The strategy of data collection process will be varying with respect to the objectives. In *Variable Selection [21]*, the variables selected for data mining are known as active variables. They are actively used to differentiate segments, make predictions, and perform some other data mining operations. To select the variables for input variables, regression analysis, stepwise regression, discrimination analysis and decision tree are used, based on selected variables and subjective judgment, input variables are selected. During *Data Transformation [21]*, the original value of data will be transformed for Data Mining. This is an important step for accuracy and validity of the result which also depends on how the data analyst decides to form and present the input variable. Finally, actual *Data Mining* [21] phase takes place to apply the selected data mining technique to transformed data. However, it has disadvantages at neural network process like it requires input

variables in the range from 0 to 1, it cannot explain the results and also it may converge prematurely to an inferior solution.

B. Fuzzy Logic

Fuzzy logic [1] utilizes fuzzy sets defined by membership functions in logical expressions to deal with the extent to which the object belongs to the set. Basically the construction of a fuzzy logic system consists of three major steps: *fuzzification*, *construction of knowledge base* and *defuzzification [22] [23] [24]*. *Fuzzification* is the process of converting crisp values to fuzzy values (e.g., low, medium, high) .Each linguistic variable after mapping can have difference membership function values for different linguistic term, it breaks traditional binary logic that a case can only belong to or not belong to a category. This process is what we call *Fuzzification [22] [23] [24]*. *Knowledge Base* is constructed by a series of "IF-THEN" rules. After *Fuzzification* and *fuzzy inference*, each input value will have a corresponding value for each linguistic term of the output variable. The process to convert fuzzy values to the corresponding crisp value is called *Defuzzification [22] [23] [24]*. Basically it consists of two main steps. In the first step, a representative value is determined for each term in the linguistic variable. In the second step, the best crisp value for the linguistic result is computed. However, it gives only the predicted values at random, thus it cannot be considered as optimized values.

C. Genetic Algorithm Concepts

Genetic Algorithm [4] is one of the search algorithm based on the technicalities of natural selection process and genetics and they combine the fittest survival among string structures to structure a search algorithm. Genetic Algorithm is particularly suitable for multi-parameter optimization problems with an objective function subject to numerous hard and soft constraints. The main idea of GA is to start with a population of solutions to a problem, and attempt to produce new generations of solutions which are better than the previous ones. Genetic Algorithm operates through a simple cycle consisting of the following four stages: initialization, selection, crossover, and mutation. In the *initialization* stage, a population of genetic structures (called chromosomes) that are randomly distributed in the solution space is selected as the starting point of the search. These chromosomes can be encoded using a variety of schemes including binary strings, real numbers or rules. After the initialization stage, each chromosome is evaluated using a user-defined fitness function. The goal of the fitness function is to numerically encode the performance of the chromosome. For real- world applications of optimization methods such as GA, the choice of the fitness function is the most critical step. The mating convention for reproduction is such that only the high scoring members will preserve and propagate their worthy characteristics from generation to generation and thereby help in continuing the search for an optimal solution. The chromosomes with high performance may be chosen for replication several times whereas poor-performing structures may not be chosen at all. Such a *selective* process causes the best-performing chromosomes in the population to occupy an increasingly larger proportion of the population over time. *Crossover* causes to form a new offspring between two randomly selected 'good parents'. *Crossover* operates by swapping corresponding segments of a string representation of the parents and extends the search for new solution in far-reaching direction. The crossover occurs only with some probability (the crossover rate). There are many different types of crossover that can be performed: the one-point, the two-point, and the uniform type. *Mutation* is a GA mechanism where we randomly a member of the population is chosen randomly and change one randomly chosen bit in its bit string representation. Although the reproduction and crossover produce many new strings, they do not introduce any new information into the population at the bit level. If the mutant member is feasible, it replaces the member which was mutated in the population. The presence of mutation ensures that the probability of reaching any point in the search space is never zero. But the value cannot be assured as perfect, since it just gives the selected value from the unpredicted value range.

D. Neuro Fuzzy Technique

Neural networks [1] impersonate biological information processing mechanisms which are designed to perform a nonlinear mapping from a set of inputs to a set of outputs. The mapping is carried out by the processing elements, called *artificial neurons*, which are interconnected to form a network divided into layers (usually three): the *input layer* receives inputs from outside, the *output layer* sends outputs to the outside and one or more *intermediate layers* (hidden layer) connect the input and output layers. Basically a neuro-fuzzy system is a fuzzy logic system with a learning algorithm derived from the neural network theory to determine its parameters, including the parameters of the membership function and the relative importance of each fuzzy rule. The most common approach used to combine the two techniques is so-called *Fuzzy Associative Memory*. It attempts to use neural networks to implement the desired mapping for fuzzy systems by applying fuzzy rules to a set of inputs, combining the consequents of each rule, and producing a value for the output variable. Each rule is associated with a weight that represents the importance of the rule in relevance to the other rules in the system. The errors between the results computed by the Fuzzy Associative Memory system and the desired output are used to modify the weights. The training process will stop until the error is less than a certain threshold value. A fuzzy logic system is constructed by using the complete knowledge base to describe the relationship among independent and dependent variables. Then the knowledge base is fine -tuned by using the learning ability of neural network based on the training data set. Finally we use the testing data set to validate the obtained model. However, the fuzzy prediction is not so accurate and the process involved in neural network to bring out the solutions is not transparent.

III. GENETIC- FUZZY APPROACH

A. Introduction

Online auctions are making its own popularity among the economists and consumers. At this trend, a system that predicts the end price of these online auctions in an unbiased manner will be a necessary tool. Prediction should be unbiased and should not have human intervention which may bias the results. Using the Genetic Algorithm and Fuzzy Logic capabilities we can design a system that can satisfy the unbiased criteria.

The proposed work is to predict the end bid value of an online auction using the previous bids and other factors that affect an auction. The genetic approach used here is well suited for auctions but they can also be used to predict similar environments like the stock market. The test results show that the Genetic Fuzzy works well in predicting and optimizing the final bid. The Fuzzy logic makes the prediction closer to the actual bid thus justifying its use here. Output of Fuzzy logic is given as input to the genetic algorithm and the predicted end price is optimized. Future work can be done on adjusting the system in order for it to work for other type of auctions as now it can work only on the English auction.

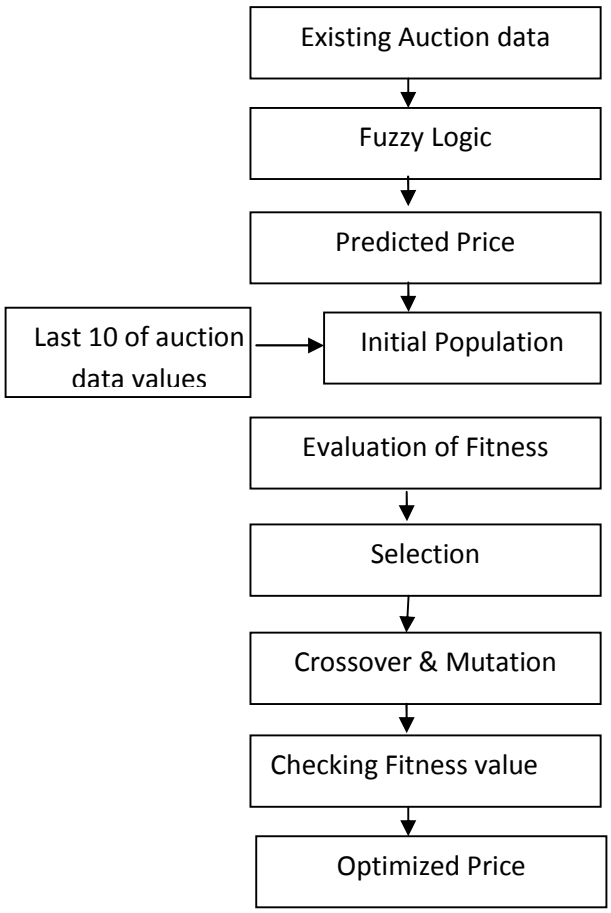

Fig.1. Flow chart of genetic fuzzy approach

The proposed system can also be adjusted to get inputs from databases with the bid details. Another way of getting the bids as input is to enter each one manually. The system does not need input other than the bids. Unlike in linear prediction systems which need inputs other than the bids, there is no need for such information to be provided and hence the system is independent of the user view and thus is unbiased and finally the predicted end price is optimized.

B. The Working of a Genetic Fuzzy System

Since genetic fuzzy is basically a fuzzy logic system combined with the learning ability of genetic algorithm, the fuzzy logic, genetic algorithm and finally the Genetic Fuzzy system has been introduced in the following.

C. Fuzzy Logic System

The fuzzy logic [1] is a process to predict the value from a set. The fuzzy process results a predicted value which will be always a imprecise one. The process involves in creation of member functions. When the member function is created, the fuzzification process can be applied. Certain fuzzy rules shall be applied on fuzzified sets. This application of fuzzy rules on fuzzification is the mechanism of Inference mechanism.

For understanding the fuzzy process, we shall work on an illustration. Let us consider 1 to 10 as bid values in price and .0 to .60 as time in seconds. By creating member function, the bid values from 1 to 3 below, 2 to 8 be medium, 7 to 10 be high, and for time values, from 0 to 25 below, 15 to 45 be medium, 40 to 60 be high.

Let the input variables be B_s, B_m, B_f, which is low, medium, and high respectively. Now the process of fuzzification is implemented on the fuzzy values. Fuzzification is the process of converting crisp values to fuzzy values (e.g., low, medium, high). For example, we use low, medium, and high to describe the extent of starting price (B_s) and external reference price (B_m). For every linguistic variable, each term is defined by its membership function. If the figures of one data are $\{B_s, B_m\} = \{6, 9\}$, for example, then the corresponding values of each term are described as follows.

B_s： $\mu low\ (6) = 0.6$，$\mu\ medium\ (6) = 0.4$，$\mu high\ (6) = 0$
B_m：$\mu low\ (9) = 0.34$，$\mu\ medium\ (9) = 0.66$，$\mu\ high\ (9) = 0$

In other words, the corresponding values can be written as follows.

B_s： $\{low, medium, high\} = \{0.60, 0.40, 0.00\}$
B_m： $\{low, medium, high\} = \{0.34, 0.66, 0.00\}$

As B_s equal to 6, its membership function values for low, medium, and high are 0.60, 0.40, and 0.00 respectively. Since each linguistic variable after mapping can have different membership function values for different linguistic term, it breaks traditional binary logic that a case can only belong to

or not belong to a category. This process is called as *Fuzzification*.

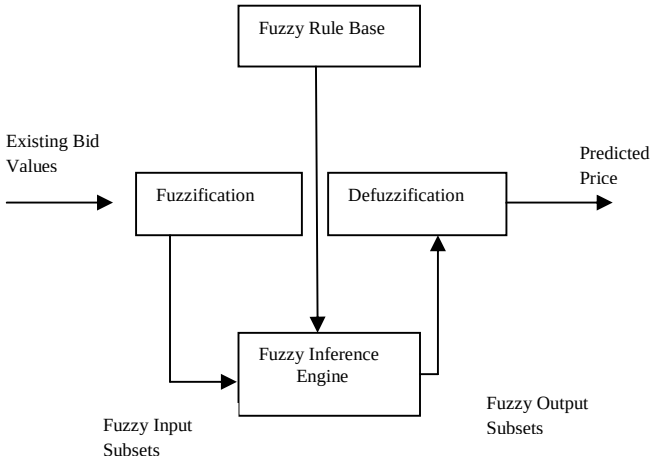

Fig.2. Fuzzy logic overview

After *fuzzificatio*n and fuzzy inference, each input value will have a corresponding value for each linguistic term of the output variable. For example the corresponding value of the linguistic term "the B_f is medium" is 0.6. Assume that the corresponding values for the other linguistic terms are "the B_f is very low" is 0, "the B_f is low" is 0.1, "the Pf is high" 0.35, and "the B_f is very high" is 0. The process to convert fuzzy values to the corresponding crisp value is called defuzzification. This process is carried out using centre of gravity method in which we are using weighted average method which is one of the most commonly used defuzzification method. The Fuzzy system has a characteristic to represent human knowledge or experiences using fuzzy rules.

D. Modified Genetic Algorithm

Genetic Algorithm [2] is processes with a set of solutions which is represented by chromosomes, called as *population*. Solutions from one population are taken and used to form a new population. This is carried out for the new population will be better than the old one. Solutions which are selected to form new solutions which are called as *offspring* are selected according to their fitness. According to the concept of fitness, the more suitable they are the more chances they have to reproduce. This is repeated until some condition is satisfied, like number of populations or improvement of the best solution.

The Modified Genetic Algorithm procedure consists of selecting a parent pair of chromosomes and then performing crossover and mutation operators (generating the offspring chromosomes – the new population) until the termination condition is reached; then the best individual in the population is selected as a solution to the problem. The crossover operator is used for exchanging information from two parents (vectors p1 and p2) obtained in the selection process by a roulette wheel approach.

Traditional approach of Genetic algorithm, where evaluation is performed after the initialization of population. The next step is the selection of child population on the basis of fitness and application of genetic operators. Genetic algorithm is very uncertain also. It may be possible that the initial parent population has the best result and after so many generations we could not find the better result as it has been found before. At the same time application of genetic operator means to generate new set of child population. In simple genetic algorithm approach evaluation is done after crossover and mutation but again it may be possible that the required result can be found after crossover only. After this observation we have modified the simple genetic algorithm with evaluation after each new set of population. At the same time we are keeping the best result found previously and it is replaced with the next best result if it is better than the previous result. The modified genetic algorithm,

MODIFIED GENETIC ALGORITHM

1. **Initiation of Parent population**
2. **Evaluation for Fittest**
3. **While Termination Criteria Not Satisfied**
   {
   1. **Selection Of child population**
   2. **Apply Crossover**
      1. **Evaluation**
      2. **Replace the result if it is better than previously stored**
   3. **Apply Mutation**
      1. **Evaluation**
      2. **Replace the result if it is better than previously stored**

   }
4. **Go to Step 3 until termination criteria satisfies**

IV. OBJECT ORIENTED (PROJECT) DESIGN

Architectural Design

The architectural design describes the layers used in this project. The Interface layer, Business layer and Database layer are integrated to perform the technical task for our problem.

The Fig 3 explains the overall design of the "Optimization and Bidding" application.

The Gene-Fuzzy system starts with bid values from the Auction site, the previous bids and previous time bids. These bid values are stored in the bid cluster, which is the database. These three values are been entered in the Fuzzy system. The fuzzy system predicts the values with the help of fuzzy logic process and stores in the imprecise predicted bid cluster.

Using the predicted value from the fuzzy system, it is given to Genetic system to optimize the value. The genetic

system optimizes the value using the algorithm of Hybrid Genetic. In genetic fuzzy, the population is determined their fitness value. Based on the fitness value, the selection process is made with the help of random selection method. The selected population is performed by the crossover operation, which is helped with the single point crossover. The mutation process is performed as normal. With these processes, a new offspring will be produced and replaced in the population. All these processes will be looped until the new values are obtained. The resulted value is the optimized value and it is stored in the bid cluster.

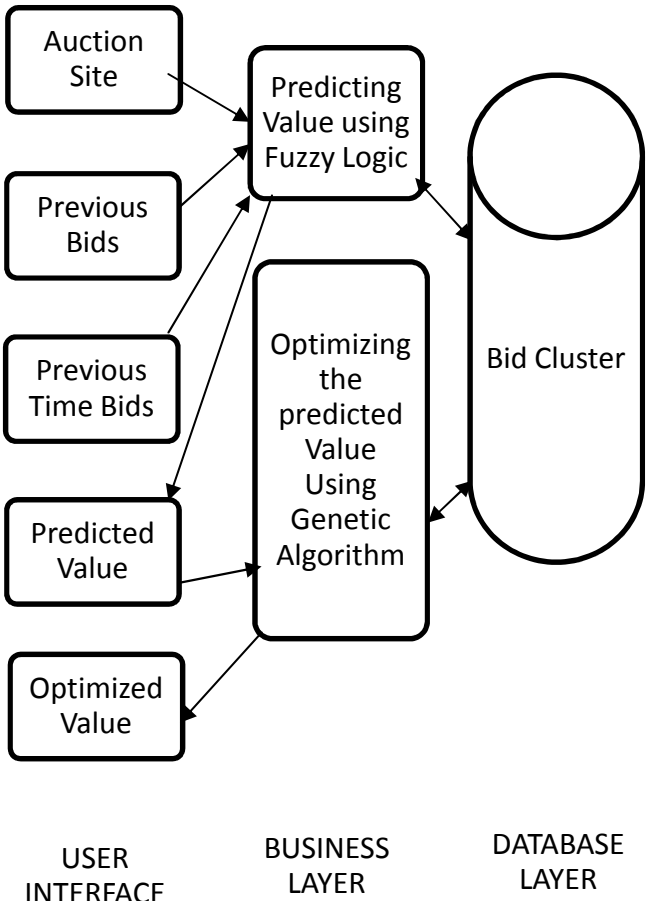

Fig.3. Architectural design

V. CONCLUSION

The proposed system is trying to propose a final price prediction model for online auction based on the real world data with the help of genetic fuzzy technique to catch the complicated relationship among the final price and key factors involved in an auction. The empirical results show that genetic fuzzy performs the best no matter in the training data sets or testing data sets. In addition to the better prediction accuracy, genetic fuzzy system also provides the knowledge base obtained from the data set which describes the delicate relationship among the variables. This knowledge can not only provide practitioners some insights in understanding the bidding process but also provide the academia the basis for constructing hypotheses for further investigation. This proposed genetic fuzzy system will show a promising perspective to explore the online auction.

Applications

The ability to predict the final price of online auction items can be of great help for both the buyers and sellers, the following applications are recommended for buyers and sellers.

Buyers

Being able to make predictions about the likely closing prices of the various auction items, the buyer can determine which auction is more close to his/her willingness-to-pay to place a bid. In doing so, the bidder can make bidding decision with the aid of the predicting model and diminish the uncertainty with respect to winning the item in the auction.

Sellers

The model of the final-price based on the attributes of the auction can also be used to help sellers to set the auction rules of their items to facilitate the completion of the auction. When the seller enters the information of the item they want to sell, our model would give suggestions for the auction attributes to maximize the final price and enhance the efficiency of the transaction.

REFERENCES

[1] Chin-Shien Lin, Shihyu Chou, Shih-Min Weng, Yu-Chen Hsieh, 2011, “A Final Price Prediction Model for English Auctions — A Neuro-Fuzzy Approach”.
[2] Ricardo de A. Araujo, 2011, “Translation Invariant Morphological Time-lag Added Evolutionary Forecasting”, Intelligent Computing Department, RiA Prediction Systems, Recife, PE, Brazil.
[3] KIM Hongil, BAEK Seung, 2004, “Forecasting Winning Bid Prices in an Online Auction Market - Data Mining Approaches”, KIM Hongil, BAEK Seung, College of Business Administration, Hanyang University Seoul 133-791 Korea.
[4] Darrell Whitley, “A Genetic Algorithm Tutorial”, Computer Science Department, Colorado State University, Fort Collins, CO 80523.
[5] Rayid Ghani, 2004, “Predicting the End-Price of Online Auctions”, Hillery Simmons.
[6] Srinivas. M and Patnaik. L, 1994, "Adaptive probabilities of crossover and mutation in genetic algorithms", IEEE Transactions on System, Man and Cybernetics, vol.24, no.4, pp.656–667.
[7] Kim, Yongseg, 2004, “An Optimal Auction Infrastructure Design: An Agent-based Simulation Approach”, Proceedings of the Tenth Americas Conference on Information System, New York, 1851-1859.
[8] Lucking-Reiley, D., 2000, “Auctions on the Internet: What’s Being Auctioned, and How?” Journal of Industrial Economics, 48(3), 227-252.
[9] Pinker, E.J., Seidmann, A. & Vakrat, Y., 2003, “Managing Online Auctions: Current Business and Research Issues,” Management Science, 49(11), 1457–1484.
[10] Karl Nygren, 2004, “Stock Prediction – A Neural Network Approach”.
[11] Fei Dong, Sol M. Shatz, Haiping Xu, 2011, “An Empirical Evaluation on the Relationship between Final Auction Price and Shilling Activity in Online Auctions”.
[12] Madhu Goyal, Jie Lu, Guang Zhang, 2007, “Decision Making in Multi-Issue e-Market Auction Using Fuzzy Techniques and Negotiable

Attitudes", University of Technology, Sydney, PO BOX 123 Broadway NSW Australia.
[13] Charles W. Richter, Jr., Gerald B. Sheblé, "Genetic Algorithm Evolution of Utility Bidding Strategies for the Competitive Marketplace", Electrical and Computer Engineering Iowa State University Ames, Iowa 50011.
[14] Wu Xiao-qin, 2012, "Research on Prediction Model of Time Series Based on Fuzzy Theory and Genetic Algorithm", Key Laboratory of Network and Intelligent Information Processing, Hefei University, Hefei 230601 , China.
[15] José Luis Aznarte, Jesús Alcalá-Fdez, Antonio Arauzo-Azofra, José Manuel Benítez, 2012, "Financial time series forecasting with a bio-inspired fuzzy model"
[16] Jin Ho Choi, Hyunchul Ahn, Ingoo Han, 2006, "Utility-based double auction mechanism using genetic algorithms", Graduate School of Management, Korea Advanced Institute of Science and Technology, 207-43 Cheongrangri-Dong, Dongdaemun-Gu, Seoul 130-722, Republic of Korea.
[17] Chi-Bin Cheng, Chao-Jung Cheng, 2011, "Available-to-promise based bidding decision by fuzzy mathematical programming and genetic algorithm".
[18] A. Azadeh, S.F. Ghaderi, B. Pourvalikhan Nokhandan, M. Sheikhalishahi, 2011, "A new genetic algorithm approach for optimizing bidding strategy viewpoint of profit maximization of a generation company", Department of Industrial Engineering and Center of Excellence for Intelligent Based Experimental Mechanics, College of Engineering, University of Tehran, Iran.
[19] Shanshan Wang, Wolfgang Jank, Galit Shmueli, 2004, "Forecasting eBay's Online Auction Prices using Functional Data Analysis", University of Maryland, College Park, MD 20742.
[20] Chin-Shien Lin, Shihyu Chou, Shih-Min Weng, 2006, "A Final Price Prediction Model for English Auctions— A Neuro-Fuzzy Approach", Department of Business Administration National Chung Hsign University, Taiwan.
[21] Arun K Punjari, 2001, "Data mining techniques" Universities Press (India) Pvt. Ltd.
[22] Zadeh, L.A., 1975, "The concept of a linguistic variable and its application to approximate reasoning I", Inf. Sci., 8: 199–249.
[23] Zadeh, L.A., 1975, "The concept of a linguistic variable and its application to approximate reasoning II", Inf. Sci., 8: 301–357.
[24] Zadeh, L.A., 1976, "The concept of a linguistic variable and its application to approximate reasoning III", Inf. Sci., 9: 43–80.