IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITP4593

An Exploratory Study: Forecasting Winning Bid Prices in Online Auction Markets

Hong-Il Kim(itlime@ihanyang.ac.kr)[1)], Bongjun Kim[2)], Sungbin Cho[3)], Seung Baek[1)]
[1)] College of Business Administration, Hanyang University 17 Haengdang-dong, , Seongdong-gu, Seoul 133-791, KOREA
Tel: +82-2-2290-1062 Fax: +82-2-2290-1169
[2)] Electronics and Telecommunications Research Institute, [3)] Konkuk University

ABSTRACT

To solve the information asymmetry problem in online auction markets, this study suggests and validates forecasting models of winning bid prices. Specially, it explores the usability of Neural network, Bayesian network, and Logistic regression in building the forecasting models. This research empirically shows that, in forecasting winning bid prices in online auction markets, data mining techniques such as Bayesian network and Neural network, have showed better performance that a traditional statistical model, Logistic regression. In addition, depending on the nature of data and the data transformation strategy, we have to select an appropriate data mining technique carefully.

1. INTRODUCTION

One of main problems in operating offline auction markets efficiently is the information asymmetry problem. The information asymmetry problem is caused by either lack of expertise (information regarding the product being auctioned) or lack of pricing information in the retail market[9]. If either buyers or sellers do not have a complete set of market information, they cannot guarantee that they get fair market prices for transaction items through auction. The online auction has solved the problem partially by providing powerful searching engines. However, online auction users have still suffered from the information asymmetry problem. Depending on what information online auction users have about transaction items, their perceived prices might have a huge gap. If the gap between the perceived prices of buyers and sellers is huge, it might take long time to negotiate the final price. The Revelation Principle of McAffee and McMillan[8] states that, by announcing the prices with which the seller wants to sell and the buyer to buy, auction markets can operate efficiently. In online auction markets, the online auction companies employ various methods to improve communication productivity between buyers and sellers. If sellers know fair market prices of their selling items before negotiating prices with buyers, the productivity of their price negotiation processes will be enhanced. In the case of buyers, if they know fair market prices of their buying items before negotiation, they can shortly reach to the final prices.

The objective of this research is to suggest and validate models to forecast the fair market prices of the products that are exchanged in online auctions. Since the transaction data in online auction markets normally has high noise, this study employs two AI-based forecasting techniques, Bayesian network and Neural network.

2. RESEARCH METHODS

2.1 Neural Network

Neural network emulates the human pattern recognition functions through a similar parallel processing structure of multiple inputs. It is designed to capture the causal relationships between dependent and independent variables in a given data set. Neural network has been applied widely in industries, from predicting financial series to diagnosing medical conditions, from identifying clustering of valuable customers to identifying fraudulent credit card transaction, from recognizing numbers written on checks to predicting the failure rates of engines[1].

Neural network generally consists at least a layer of input nodes, one or more layers of hidden nodes, and a layer of output nodes. Training the network is the process of setting the best weights among nodes in the network. Throughout the training process, Neural network produces weight where the actual output is as close to the forecasted outputs as possible for as many of the examples in the training set as possible[1]. One of the most popular training techniques is the back-propagation network, originally developed by John Hopfield. It is a supervised learning technique with the objective of training the network to map input vectors to a desired output vector[6]. Depending on Network structure, the performance of the Neural network is varied.

2.2 Bayesian Network

The Bayesian network consists of a set of probabilistic nodes and a set of directed arcs connecting the nodes. A central feature of the Bayesian network is to allow inference based on observed evidence on any of the nodes.

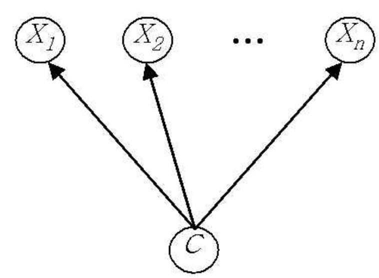


Figure 1. Naïve Bayesian Network

IF A and B are random variable, Bayes' rule states that

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Bayesian network, one of the most effective classifiers, is a rapidly growing field of research that has seen a great deal of activity in recent years[3]. It is a graphical representation of a probabilistic model that most people find easy to construct and interpret[5]. Over the last decade, the Bayesian network has become a popular representation for encoding uncertain expert knowledge in expert systems[4]. A key step in the Bayesian network is the computation of the marginal likelihood of a data set given a model[2]. In this study, the naïve Bayesian is used to predicting the winning bid prices.

3. RESEARCH METHODOLOGY

3.1 Data Collection

The data set used in this study consists of five hundred bid information of notebook computers from eBay which were collected in six weeks. The data collection has been restricted in four major notebook manufactures, *Compaq*, *Dell*, *IBM*, and *Toshiba*, which are popularly being auctioned on eBay. Drawing random samples from unbalanced populations is likely to yield biased samples, thereby adversely affecting the learning process of the models. As a result, the performance of the network and statistical model may be poor when tested in realistic situations[6, 7]. To solve the problem, each brand has been gathered equal population.

The data set consists of 12 variables. The product related information is notebook computer specification, for example, brand, TFT size, CPU type, RAM, HDD etc. The auction related information contains seller rating, bidding period, starting bid price. The input variables are summarized in Table 1.

Table 1. Input Variables

Category	**Variables**
Product related information (9)	Brand, TFT size, CPU Type, CPU Speed, RAM, HDD, Ethernet, Modem, CD drive
Auction related information (3)	Seller Rating, Bid Period, Starting Bid amount

3.2 Data Transformation

Before data transformation, data cleaning job has been performed firstly. The 71 data that have missing or incomplete information have been removed from the data set, as a result, total 429 data are left. To enhance the performance of forecasting models, two normalization strategies have been performed.

First, all the values of the data have been converted 0 to 1 ranges. Massaging the values to 0~1 range helps networks to recognize patterns in the input data. Second, by grouping the continuous values of data into several groups, the skewed distributed data is transformed into the symmetric distributed data [See Figure 1]. Although a major weakness of these data transformation strategies is losing information, many studies use these methods to enhance to model accuracy.

3.3 VARIABLE SELECTION

The choice of input variables used in the classification and prediction models is one of the important considerations in the design of Neural network and Bayesian network. In this research, to choose input variables for training networks, multiple regression analysis, stepwise method, discriminant analysis and decision tree are used. Through the analyses, nine variables (Brand, TFT size, CPU speed, RAM, HDD, Ethernet, CD drive, seller rating, and starting amount) are selected and three variables (CPU type, modem, and bidding period) are removed.

3.4 Data Mining

The objective of this research is to suggest and validate models to forecast the fair market prices of the items that are exchanged in online auctions using Neural network and Bayesian network. *SAS Enterprise Miner 4.0* is used as a tool for Neural network and the macro function of *Microsoft Excel 2002* is used for Bayesian network.

3.5 Analysis of Results

Through twenty times simulation, averages of classification rates of the networks are computed. The results is showed in Figure 2. Neural network, Logistic regression and Bayesian network have showed 51.05%, 53.96% and 73.32% of average classification rate, respectively. In forecasting prices based on highly noisy data, such as online auction market information, Bayesian network showed best performance among three forecasting models. Many studies empirically state that, in handling highly noisy data, data mining techniques, such as Neural network or Bayesian network, outperform traditional statistical models. Surprisingly, in this study, a traditional statistical method, Logistic regression, outperformed a popular data mining technique, Neural network. This result might be caused by the data transformation. The information loss throughout the data transformation process undermines the performance of Neural network.

In order to investigate the information loss effect on forecasting model's performance, based on before-massaging training data, Neural network and Logistic regression are re-evaluated [See Figure 3]. The average classification rates of Neural network and Logistic regression are 57.47% and 55.79% respectively. The Neural network slightly outperforms Logistic regression. This result states that there is information loss effect on forecasting model's performance. This research empirically proves that, in handling noisy data such as online auction markets, data mining techniques, such as Neural network and Bayesian network, show better performance than traditional statistical methods. However, depending on the nature of data and the data transformation strategy, we have to select an appropriate data mining technique carefully.

4. CONCLUSIONS AND FUTURE STUDY

To solve the information asymmetry problem in online auction markets, this study suggests and validates forecasting models of winning bid prices. Specially, it explores the usability of Neural network, Bayesian network, and Logistic regression in building the forecasting models. This research shows that, in forecasting winning bid prices in online auction markets, data mining techniques such as Bayesian network and Neural network, have showed better performance that a traditional statistical model, Logistic regression.

As a future study, we will empirically investigate how differently sellers and buyers behave in online auction markets, if they know fair market prices before the actual negotiation process from our forecasting model.

Figure 1. Data Transformation

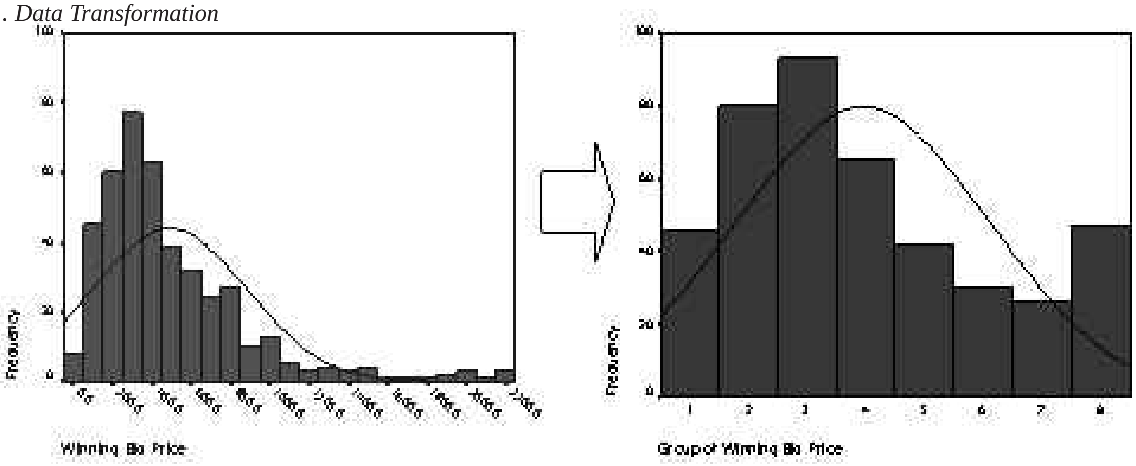


Figure 2. Classification Rate with Neural, Logistic, and Bayesian (After Data Massaging)

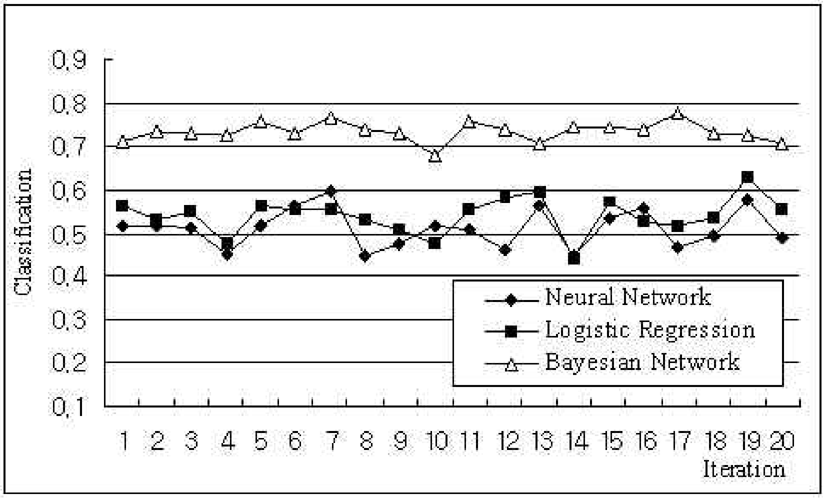
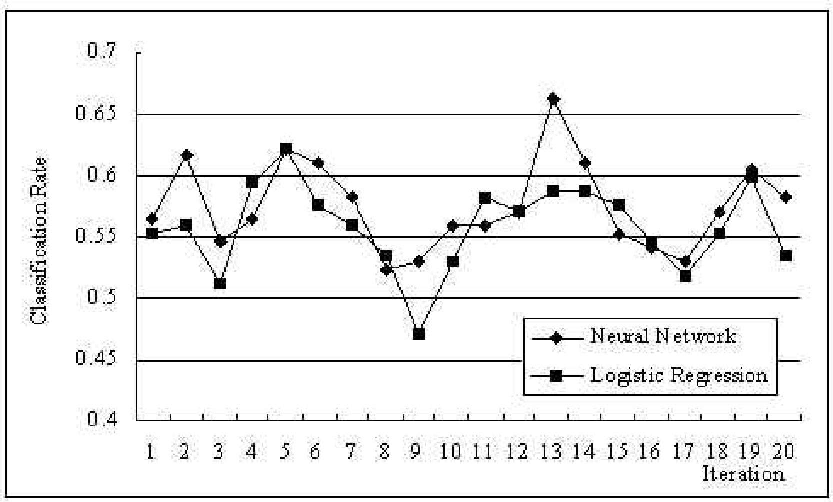


Figure 3. Classification Rate with Neural and Logistic (Before Data Massaging)



5. REFERENCES

1. Berry, M.J.A. and Linoff, G., *Data Mining Techniques : for Marketing, Sales, and Customer Support*, 1997, Wiley Computer Publishing.

2. Chickering, D.M. and Heckerman, D., "Efficient approximations for the marginal likelihood of bayesian networks with hidden variables", *Machine learning*, 1997, Vol. 29**,** Iss. 2, pp. 181-212.

3. Friedman, N., "Bayesian network classifiers", *Machine learning*, 1997, Vol. 29**,** Iss. 2, pp. 131-163.

4. Heckerman, D., Geiger, D., and Chickering, D., "Learning Bayesian networks: The combination of knowledge and statistical data", *Machine learning*, 1995, Vol. 20**,** Iss., pp. 197-243.

5. Howard, R. and Matheson, J., *Influence diagrams*. Readings on the Principles and Applications of Decision Analysis, ed. J. Matheson, Vol. II, 1981, Menlo Park, CA, Strategic Decision Group.

6. Jain, A.B. and Nag, N.B., "Performance Evaluation of Neural Network Decision Models", *Journal of Management Information Systems*, 1997, Vol. 14**,** Iss. 2, pp. 201-216.

7. Korobow, L. and Stuhr, D., "Performance measurement of early warning models", *Journal of Banking and Finance*, 1985, Vol. 9**,** Iss. 2, pp. 267-273.

8. McAffee, R.P. and McMillan, J., "Auctions and bidding", *Journal of Economic Literature*, 1987, Vol. 25**,** Iss. 2, pp. 699-738.

9. Metha, K. and Lee, B., *An empirical evidence of winner's curse in electronic auction*, Proceeding of the 20th international conference on Information Systems, 1999, Charlotte, North Carolina, pp. 465-471.

Related Content