International Journal of Applied Science and Engineering
2005. 3, 2: 101-109

Online Auction Customer Segmentation Using a Neural Network Model

Chu-Chai Henry Chan *

E-Business Research Lab
Department of Industrial Engineering and Management,
Chaoyang University of Technology,
Wufong, Taiwan, R.O.C.

Abstract: Online auction is a huge growing business. Most of online buyers face a big problem of predicting the seller's behavior to submit a reasonable price for wining a bid. Unfortunately, auction web sites e.g. eBay provide only a user ID or nickname to identify a consumer. This situation built up a wall between users for truly knowing each other. To overcome such a problem, this research proposes a neural network model called SOM to segment online auction customers into homogenous groups. Based on the segmented groups, the behavior of online bidders can be divided into three types: patient deals, impulsive deals and analytic deals. To demonstrate the feasibility of the proposed methodology, 1470 records retrieved from Taiwan eBay are used to conduct an empirical study. In conclusion, the percentages of each customer type are 39.3 % (impulsive deals), 27.8 % (analytic deals) and 32.2 % (patient deals). The analyzed result shows that more than sixty percent of bidder's behave rationally and patiently.

Keywords: customer segmentation; neural network; online auction.

1. Introduction

Online auction is a very important market for e-business. In year 2003, the trading amount is about 2.1 million U.S. dollars and the growing rate is 78 % per year [1, 3]. In such a huge growing business, customers never know and see each other face to face. Most of auction web sites such as eBay provide only a user ID or nickname for identifying a seller and a buyer [9, 10]. This situation built up a wall between consumers. In this way, customers have difficulty in making a right decision of conducting a bid. Even in some real cases, bidders place an irrational price for wining a bid. This phenomenon is called "Winner Curse" [13]. So far, no existing research provides information of the seller's behaviors on an online auction web site. "Winner Curse" still bothers most of bidders and happens daily. To improve such a situation, this study proposes a model to differentiate customer behavior at Taiwan eBay. Based on the differentiated customer behavior model, a bidder can know the basic types of deals and how to make a right decision.

In order to construct such a customer behavior model, this study applies customer segmentation as a key technology to find the general characteristics of most profitable customers [3]. By customer segmentation, transaction data are differentiated into several ho-

* Corresponding author: e-mail: ccchan@mail.cyut.edu.tw *Accepted for Publication: July 28, 2005*

mogenous categories [6]. Each category presents a type of customer behavior. Finally, a customer behavior model is generated. Based on the proposed customer behavior model, bidders can know the basic characteristics of e-auction and decide how to win a bid.

2. Consumer behavior model

To conduct customer segmentation for e-auction market, this study proposes three steps to generate a consumer behavior model (see Figure 1) [14]. The first step is to develop a spider program to retrieve a data from online auction site e.g. eBay. The second step is to apply the proposed neural network to segment customers. In this stage, the retrieved auction data are differentiated as several homogenous groups. Finally, he third step is to interpret segmented data as customer types. Each segmented data would be used to describe one type of customers.

The final goal of this study is to generate a customer behavior model. We have to know the basic definition of a customer behavior model. According to the model of EC (Electronic Commerce) consumer behavior defined by Turban, the purchasing decision making is basically a customer's reaction to stimuli [12].

The decision making process is influenced by the characteristics of sellers and buyers, the environment, the technology and the EC logistic (see Figure 2). In an e-auction process, a buyer can only find the id (identity) of the seller who he/she is contacting right now. This situation was making in understanding each other's behavior very difficultly. To cope with such a difficulty, this study chooses several important factors of stimuli (price, promotion, product and quality) and personal characteristics (Latest Login Time, Total bid time and number) to describe the behaviors of customers. The behavior model of online auction is described as follows:

$$B_{xy}= f_{seg}\,(C_{xy}, R_{xy}, T_{xy}, N_{xy}, P_{x,} F_{xy}) \qquad (1)$$

B_{xy}= The customer behavior model of e-auction by seller x and buyer y

f_{seg} = A Function of doing customer segmentation

C_{xy}= Credit by seller x and buyer y

R_{xy}= Price Change Rate by seller x and buyer y (Initial bid- Final Bid)

T_{xy}= The total bid period by seller x and buyer y

N_{xy}= The number of buyer's bid by seller x and buyer y

P_x= Product Description

F_{xy} = Final Price

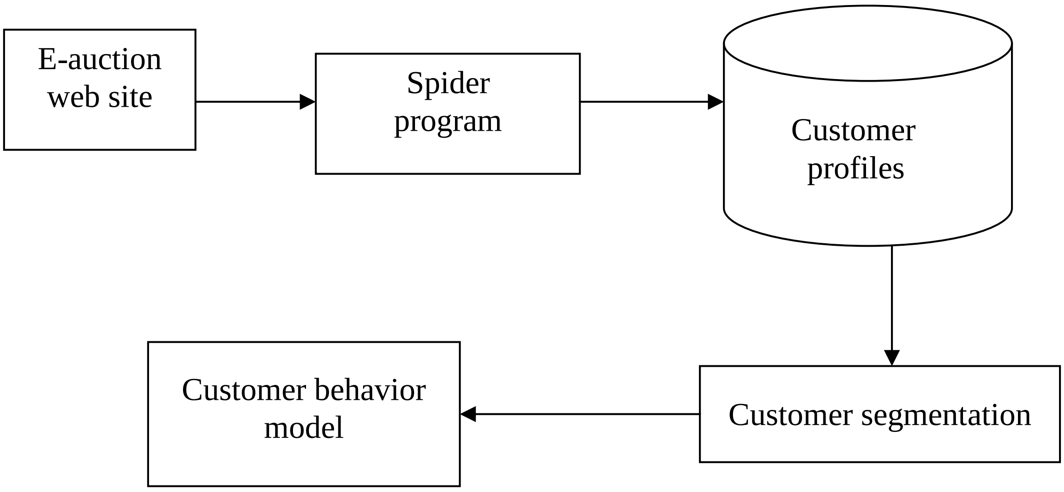


Figure 1. Customer segmentation processes

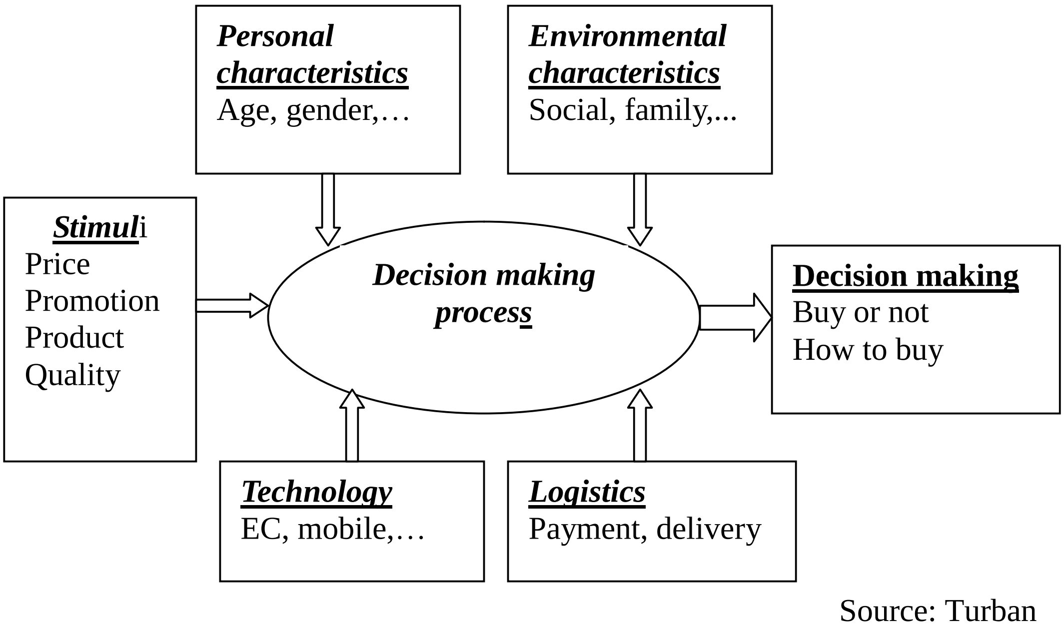


Figure 2. E-commerce consumer behavior model

3. Methodology

For this study, segmenting customers to generate a customer behavior model is the goal we are pursuing. Six steps are proposed to complete such a mission (See Figure 3). First, all of consumer transaction data have to be retrieved from an e-auction web site. The second step is to select significant variables as input data of the proposed customer segmentation algorithm. Before data analysis raw data must be normalized into a variable which is between zero and one and can be analyzed by a neural net, so the third step is to convert data into a normalized format. The fourth step is applying neural networks model to cluster data into several homogenous groups. The fifth step is to interpret ting those clustered customer data into several groups. The final step is to select and construct a customer behavior model based on customer segmentation by the previous step.

Both data collection and data cluster are very challengeable tasks for this study and need profound mechanisms to support. For data collection, the proposed mechanism had developed a spider program to crawl data from an auction web site. For data cluster, a neural network is applied for mining customer data.

In data collection stage, this study had developed a **spider program** with a Java language on a Linux system to collect data from eBay. This spider program helps to search an URL No. for each product. Based on the URL No., this program can request E-auction web site to provide detail information related to the searched product. The information is organized as a database called customer profiles [8].This spider program completes data set collection task by including a URL searching agent and an auction data agent. The functionalities for both intelligent agents are listed as follows [2]:

- URL searching agent:
  In eBay, there are three databases (transaction database, product database and customer database). This agent is developed to search URL address on eBay for each product from product da-

tabase. The identified URL addresses are recorded into one auction URL database.

- Auction data agent:
  After URLs are recorded, electronic auction data is extracted from an auction web site by an auction data agent and stored into an auction-based database for further analysis.

For data cluster, the research proposes SOM (Kohonen self-organizing network) to cluster associated customer data into homogenous groups. Several researches had proved that SOM is an effective algorithm to cluster customers [4, 17]. Kohonen self-organizing network (T. Kohonen, 1980) [5] is an unsupervised learning neural network, which applies neighborhood and topology to cluster associated data into one group. The detail for Kohonen self-organizing network is described as follows [5]:

STEP 1. Select the winning output unit as the one with the largest similarity measure (or smallest dissimilarity measure) between all weight vectors w_i and the input vector x. If the Euclidean distance is chosen as the dissimilarity measure, then the winning unit c satisfies the following equation.

$$\|x - w_c\| = \min \|x - w_i\| \tag{2}$$

where the index c refers to the winning unit.

STEP 2. Let NB_c denote a set of index corresponding to a neighborhood around winner c. The weights of the winner and its neighboring units are then updated by

$$\Delta w_i = \eta \gamma(\mathrm{i})\,(x - w_i), i \in NB_c \tag{3}$$

where η is a small positive learning rate. Instead of defining the neighborhood of a winning unit, we can use a neighborhood function γ *(i)* around a winning unit c. The Gaussian function can be used as the neighborhood function.

$$\gamma(\mathrm{i}) = \exp\left(\frac{-\|p_i - p_c\|^2}{2\sigma^2} \right) \tag{4}$$

The order of the weight updates on an individual layer is not important. Be sure to calculate the error term

$$E_p = \frac{1}{2} \sum_{k=1}^{M} \delta_{pk}^{2} \tag{5}$$

SOM will map input data into several output groups in Figure 4 [16]. During the process, the area of related data will gradually be narrowed down into a smaller group based on the Neighboring Concept in Figure 5 [16].

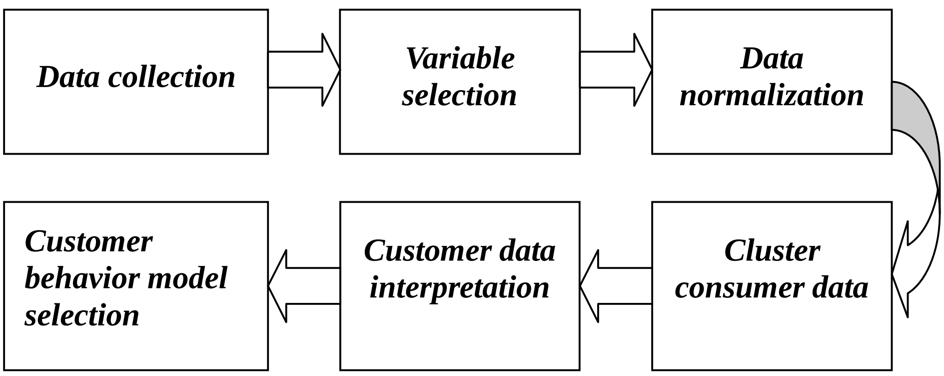
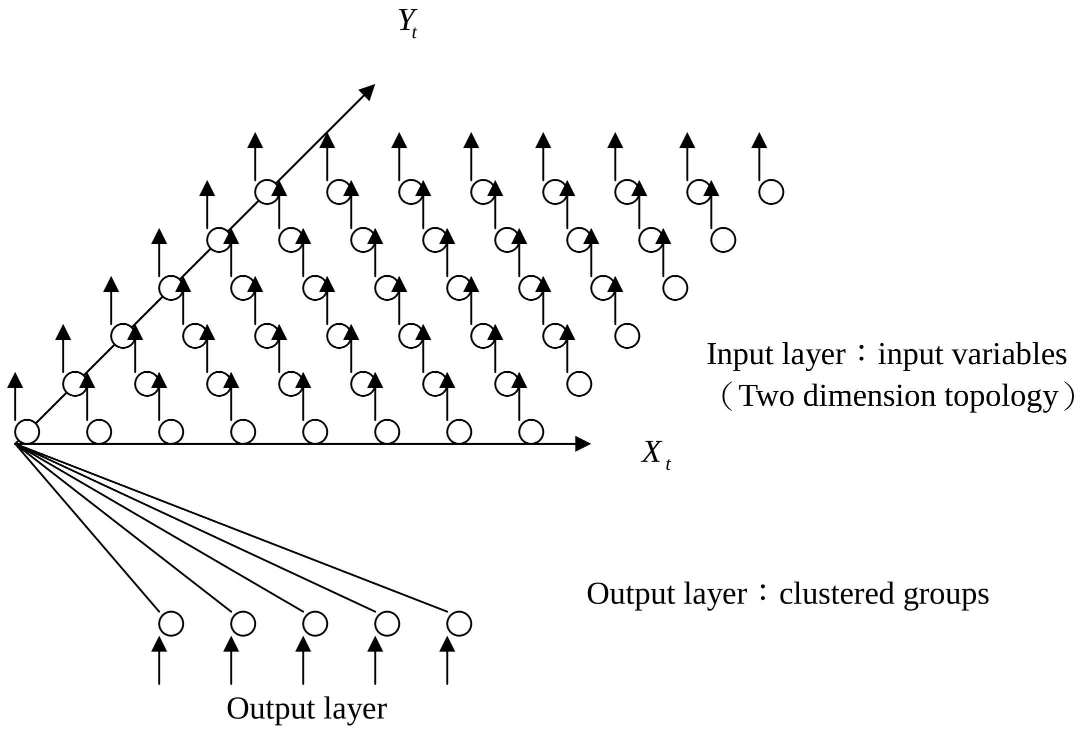
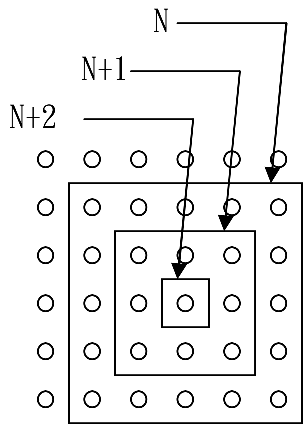


Figure 3. Procedure of data cluster



Source : Yeh (1993)

Figure 4. The architecture of SOM



Source : Yeh (1993)

Figure 5. The neighboring concept of SOM

4. Empirical study

To study feasibility of the proposed methodology, a prototype of web-based spider program had been developed. Through the developed spider program, a data set containing 1470 records related to "digital cameras" is retrieved for further analysis. The crawled data set was held between February11, 2004 and August 9, 2004 at Taiwan eBay (http://www.ebay.com.tw). The web pages of Taiwan eBay responds to the query contains details of the specific auction, including last bid (if any), opening and closing time and date, seller's ID and rating, minimum bid, number of bids, and a listing of bid history [1]. The bid history contains information on each bidder, including buyer's ID and rating, as well as the price, time and date of bids.

Too many factors possibly affect the result of electronic auction. By taking the concept of the black-box nature of neural network, the research proposed a SOM as a data mining tool for data cluster [15]. After data is clustered, auction customer behaviors of eBay for bidding digital camera are automatically clustered into nine neurons.

Based on literatures by Turban [11, 12], normally customers are divided into three major types: "impulsive deals", who want to sell or buy products quickly; "patient deals", who want to sell or buy products for a longer period; and "analytic deals", who do substantial research before making the decision.

Even there are a number of ways of identifying customer behavior. The most well-known method called RFM (Recency, Frequency and Monetary) model is used to represent the characteristics of each customer [4]. This model clusters customer behavior from three dimensions of customer's transactional data: Recency, Frequency and Monetary [15]. The first dimension is Recency, which show how long it has been since the bid began. The second one is Frequency, which indicate of how often the customer places a bid. Monetary Value is used to measure the amount of money that the customer has spent in this bid [6].

In this study case, the data values are derived from databases such as average of latest login day, auction cycle day and purchase monetary amount. The basic assumption of applying the RFM model is that future patterns of consumer trading are similar to past and existing patterns. The calculated RFM values are summarized to realize the behavior patterns of customers in the research. In this study, we propose to use the following RFM variables [6]:

1. Recency (R): the total bid period.
2. Frequency (F): the total number of bids.
3. Monetary (M): the final bid price

To cluster customer data, a neural network with six input nodes and nine output nodes had been constructed for this study and the details is listed in Table 1. The meanings of input variables are summarized in Table 2. In this study case, a 3x3 Map Neuron is used to cluster customers into three customer types in Figure 6 [4]. First, neuron 1 and 2 are clustered as impulsive deals. In this type of deals, buyers usually provide a higher price (M>= 0.98719) in a short period (R <= 0.26554) to win a bid. Second, neuron 4, 7 and 8 are clustered as patient deals. Patient buyers usually win a bid with a reasonable price (M>=0.954871) for a longer period (R >= 0.519772). Third, neuron 3, 5, 6 and 9 are clustered as analytic deals. Analytic buyers get used to compare the price with other deals and usually place a lower price (M<=0.738946) to win. The failure rate of analytic deals is very high. Normally sellers are not willing to sell their products in such a circumstance.

Summarily, the total number for each customer type are 587 (impulsive deals), 409 (analytic deals), and 474 (patient deals). From above observation, about 32.5 % (analytic deals) of customers like to spend some time in competing the price during bidding. 39.9 % (impulsive deals) of consumers want to buy theirs products in a shorter period, which could cause "Winner Curse". The distributions of customer type are summarized in Figure 7.

Table 1. Neural network model

Neural network model	SOM
Input layer (No. of neurons)	6
Output layer (No. of neurons)	9
Segment groups	3

5. Conclusion and future research

In a C2C e-commerce environment, sellers and buyers have difficulty in knowing each other's behaviors. Unpredictable results such as "Winner Curse" could happen to bring a lot of negative impacts to bidders [13]. To realize the behavior of e-customers, the study proposes a SOM neural network model to cluster customer data into homogenous groups. To conduct an empirical study, more than one thousand of data were retrieved from Taiwan e-Bay as analysis basis. A 3x3 Map Neuron is applied to cluster customers into patient deals, impulsive deals and analytic deals. From data analysis, two interesting findings are as follows:

1. More than sixty percent of online auction customer's behaviors are rational. Specially, more than 27.8 percent of bidders like to bid the price with other ones.

Table 2. Definitions of input variables

Variable names	Description	Types
R_x (Price change rate)	Initial bid- final bid	Promotion
T_{xy} (Total bid period)	Initial bid time- final bid time	Personal characteristics
F_{xy} (Final price)	Final bid	Price
C_{xy} (Credit of seller)	Credit of the current seller	Quality
P_x (Product description)	Web page of product description	Product
N_{xy} (Number of bid)	The total number of bids	Personal characteristics

Impulsive deals (98) R: 0.26554 F: 0.012806 M: 0.987194 1	Impulsive deals (489) R: 0.211633 F: 0.012806 M: 0.98719 2	Analytic deals (32) R: 0.28599 F: 0.424895 M: 0.575105 3
Patient deals (83) R: 0.519772 F: 0.030607 M: 0.969393 4	Analytic deals (40) R: 0.388273 F: 0.261054 M: 0.738946 5	Analytic deals (310) R: 0.61066 F: 0.992853 M: 0.007147 6
Patient deals (220) R: 0.995135 F: 0.023562 M: 0.976438 7	Patient deals (171) R: 0.728627 F: 0.045129 M: 0.954871 8	Analytic deals (27) R: 0.762979 F: 0.585842 M: 0.414158 9

Figure 6. Use a 3x3 map neuron for data cluster

2. Many researches worry that "winner curse" will reduce the interests of bidders [13]. Because "winner curse" makes buyers paying more money, they will get a bad impression on e-auction. This study found that impulsive deals are only 39.3 % of total e-auction transactions. The possibility of "winner curse" is lower than 39.3 %. It means that "Winner Curse" could be a smaller problem of e-auction.

There are still some limitations for this study.

1. The research chose e-auction data of "digital camera" as our empirical study sample. Because digital camera is a standard product, our observation can only represent the perspective of the behavior of e-auction consumers, who bid standard products. The customer behavior of bidding irregular products e.g. antique is still unknown. More related researches need to be explored in the future.
2. The behavior of new generation keeps changing rapidly. The results of the research only represent the behavior of current e-auction consumers. A database of customer profiles needs to be developed for keeping track of customer footsteps in the future.
3. This study proposes a SOM neural network to segment customers [5, 7]. There are a bunch of data cluster algorithms, e.g. K-means, Fuzzy C-means, Mountain Clustering, Subtractive Clustering. Further researches need to be explored to prove which method is the best for segmenting auction customers.
4. This study selects six significant variables to segment e-auction bids. Unfortunately, the RFM model can only use three variables to interpret the customer behavior. In the future, how to develop a new model which can combine RFM and the other three variables in presenting the phenomenon of e-auction is crucial.

The research develops a methodology to cluster the behavior of e-auction customers. It is a big step for knowing the styles of e-auction user's behavior. Our future research is going to develop an intelligent agent to assist bidders predict the important factors of e-auction such as price, length (time) and so on for assisting decision making. In this way, the possibility of "Winner Curse" will be reduced. The behaviors of sellers and buyers will be more rational and predictable in the long run.

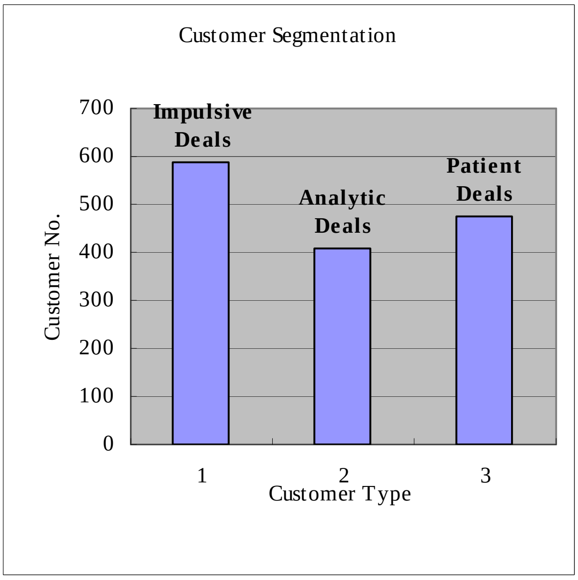


Figure 7. Customer distribution

Acknowledgements

The author acknowledges his research students, Zu-Chow Hwang and Ming-Shang Lin for collecting data in this study.

References

[1] Bryan, D., Reiley, L. D., Prasad, N., and Reeves. 2000. Pennies from eBay: the Determinants of Price in Online Auction. *Working Paper, Vanderbilt University:* 1-24.

[2] Chan, Chu-Chai Henry. 2005. Developing a Data mining-Based Spider Program for Price Prediction of Electronic

Auction. *Working Paper, Chaoyang University of Technology.*

[3] Deitel, H.M., P.J. Deitel, and K. Steinbuhler. 2001. *"E-business and e-commerce for managers"*. Prentice Hall Inc., New Jersey.

[4] Hsieh, Nan-Chen. 2004. An integrated data mining and behavioral scoring model for analyzing bank customers. *Expert Systems with Applications*, 27: 623–633.

[5] Jang, Jyh-Shing, Chuen-Tsai Sun, and Eiji Mizutani. 1997. *"Neuro-Fuzzy and soft computing"*. Prentice Hall, New York.

[6] Jonker Jedid-Jan, Nanda Piersma, and Dirk Van den Poel. 2004. Joint optimization of customer segmentation and marketing policy to maximize long-term profitability. *Expert Systems with Applications*, 27: 159-168.

[7] Kohzadi, Nowrouz, Milton S. Boyd, Badman Kermanshahi, and Iebling Kaastra. 1996. A Comparison of artificial neural network and time series models for forecasting commodities prices. *Neurocomputing*, 10: 169-181.

[8] Shaw, Michael J., Chandrasekar Subramaniam, Gek Woo Tan, and Michael E. Welge. 2001. Knowledge management and data mining for marketing. *Decision Support Systems*, 31, 127–137.

[9] Sherstyuk, Katerina. 2002. Collusion in private value ascending price auctions. *Journal of Economic Behavior and Organization*, 48: 177–195.

[10] Standifird, Stephen S. 2001. Reputation and e-commerce: eBay auctions and the asymmetrical impact of positive and negative ratings. *Journal of Management*, 27: 279–295.

[11] Turban, E. 1997. Auctions and Bidding on the Internet: An Assessment. *Electronic Markets*, 7, 4: 7-11.

[12] Turban, E, Jae Lee, David King, and H. Michael Chung. 2000. *"Electronic Commerce A Managerial Perspective"*. Prentice Hall International Inc., New Jersey.

[13] Oh, Wonseok. 2002. C2C Versus B2C: A comparison of the Winner's Curse in two types of electronic auctions. *International Journal of Electronic Commerce*, 6, 4: 115-138.

[14] Rayals, Lynette. 2002. Are your customer worth more than money ? *Journal of Retailing and Consumer Service*, 9: 241-251.

[15] Yao, Jingtao, Yili Li, and Chew Lim Tan, 2000. Option price forecasting using neural networks. *The International Journal of Management Science*, 28: 455-466.

[16] Yeh, I-Cheng. 1993. "*The applications and practices of neural networks*". Scholars Book, Taiwan, R.O.C.

[17] Yuan, Soe-Tsyr and Wei-Lun Chang. 2001. Mixed-initiative synthesized learning approach for web-based CRM. *Expert Systems with Applications*, 20: 187-200.