An Empirical Evaluation on the Relationship Between Final Auction Price and Shilling Activity in Online Auctions*

[1]Fei Dong, [2]Sol M. Shatz and [3]Haiping Xu
[1,2]Computer Science Department
University of Illinois at Chicago, Chicago, IL 60607, USA
{fdong, shatz}@cs.uic.edu
[3]Computer and Information Science Department
University of Massachusetts Dartmouth, North Dartmouth, MA 02747, USA
hxu@umassd.edu

Abstract. *In this paper, we are interested in the relationship between final prices of online auctions and possible shill activities during those auctions. We conduct experiments on real auction data from eBay to exam the hypotheses that (1) A lower-than-expected final auction price indicates that shill bidding was less likely to occur in that auction; and (2) A higher-than-expected auction final price indicates possible shill bidding. In the experiments, a neural network approach is used to learn the expected auction price. In particular, we trained the LArge Memory Storage and Retrieval (LAMSTAR) Neural Network based on features extracted from item descriptions, listings and other auction features. The likelihood of shill bidding is determined by a previously proposed Dempster-Shafer theory based shill certification technique. The experimental results imply that both a lower-than-expected final auction price and a higher-than-expected final auction price might be used as direct evidence to distinguish trustworthy auctions from likely shill-infected auctions, allowing for more focused evaluation of those shill-suspected auctions.*

1. Introduction

Online auction platforms, such as eBay, have become ideal places for people to purchase bargain-priced and hard-to-find items. Besides, they also allow people to become instant businessmen at a low cost. However, frauds develop in online auctions as online auction platforms expand in use. Shill bidding, which is one type of auction fraud, aims to inflate the final price of an auction. Such bidding activity has been found to have a severe impact on the fairness of auction markets, and in the worst scenario it can result in insufficient market or even market failure [1].

Many online consumers do not realize that shill bidding is a serious illegal behavior. In fact, if convicted, a shill bidder can serve several years in prison and pay a significant fine [2]. In 2004, an eBay store owner who conducted shill bidding pleaded guilty to Combination in Restraint of Trade, a violation of the New York antitrust law. It is a felony punishable by a maximum of four years in prison [2].

To protect online auction bidders from shill bidding, shills, who intentionally drive up an auction's final price, should be detected as they are placing the shill bids so that the auctions can be cancelled and bidders are protected from fraudulent activities. However, it is not easy to detect shills due to their concealment efforts and the lack of non-deniable evidence. Yet, if we can easily and quickly distinguish auctions that are suspected of involving shills from auctions without shills, the process of shill investigation may become somewhat less difficult since efforts can then be focused on just those auctions with potential shills. In this paper, we propose a method to classify auctions into two such categories – those that are likely to involve shills and those that are not likely to involve shills. The primary classification feature we consider is the actual final price of the auction (in comparison to an expected final price).

The rest of this paper is organized as follows. Section 2 discusses related work. Section 3 describes the hypotheses we evaluate and why we formed these hypotheses. Section 4 introduces how the experimental data is collected and how the desired features of the data can be defined. Section 5 presents our experimental procedures and reports the results. Finally, Section 6 provides conclusions and mentions some future work.

2. Related Work

Generally, a substantial amount of work has been done in the study of auction data. Heijst et al., combined text mining and boosting algorithms to predict auction final prices [3]. Ghani et al., compared a regression model,

* This material is based upon work supported by the U.S. National Science Foundation under grant numbers CNS-0715648 and CNS-0715657.

a neural network and a decision tree, and they achieved the best result using the neural network when treating the price prediction problem as a series of binary classification problems [4]. In [5], Lim et al., employed grey system theory to predict auction closing prices in a simulated auction environment. In this paper, we use a neural network approach, but for the special purpose of predicting shilling activities. In particular, the expected auction price is learned from the LArge Memory Storage and Retrieval (LAMSTAR) Neural Network, where price prediction is based on features extracted from item descriptions, listings and bids features.

The topics of shill detection and verification also have attracted significant attention from researchers. Patel et al., proposed a real-time shill monitor for agent based online auction systems using role-based access control mechanisms [6]. Xu, et al., introduced a formal model checking approach for detecting shilling behaviors, especially the competitive shilling behaviors [7]. Dong et al., proposed a decision support system based on Dempster-Shafer (D-S) theory [8] to compute the likelihood of shill bidding activities [9]. Kauffman and Wood discovered that the existence of shill bids in an online auction can drive up the final selling price of the auction [10]. In contrast, in this paper we are interested in some different questions: If an auction finishes at a price that exceeds the expected closing price range, can we consider it to be likely that the auction involves shill bidding? And similarly, if an auction finishes at a price that is much lower than the expected closing range, is it likely that the auction does not involve any shill bidding?

3. The Hypotheses

Since a shill bidder's primary goal is to drive up the final price, it seems reasonable that for a shill-infected auction, the final auction price should be significantly higher than it would have been if no shill bids were placed in the auction. Therefore, if the final price of an auction can be predicted accurately (assuming there are no shill bids), the actual final price of the auction would be very useful in deciding if this auction is likely to involve shills. This would provide direct evidence to both shill investigators and other bidders.

Kauffman and Wood found that shill bidding acts as a signal for other bidders to place higher bids and thus increase the auction's winning bid [10]. In other words, if shill bidding occurs in an auction, the auction will very likely have a higher final auction price. Equivalently, we suggest the following linguistic, fuzzy-logic type expression:

Shill bidding is highly likely to result in a higher-than-expected final auction price.

According to the *modus tollens* rule in classic logic, we obtain our first actual hypothesis.

□ **Hypothesis 1 (The non-shill hypothesis):** A lower-than-expected final auction price indicates unlikely shill bidding.

Auction prices can reflect the current market of the auctioned item. If a final auction price falls in the normal price range – for example within +/-20% of the average price – the price of the auction conforms to market discipline. In contrast, if the market for a type of item is depressed, the final auction prices for these items are not expected to be high. Under this circumstance, if the final prices of a particular seller's auctions are significantly and consistently higher than those of the same-item auctions, the seller is suspected of employing shill bidding or other types of fraudulent bidding activities.

Therefore, the final price of an auction is believed to be an indicator of trustworthiness.

An auction that does not involve shill bidding is likely to have a lower-than-expected, or as-expected, final auction price.

Again, according to the *modus tollens* rule, we obtain our second hypothesis:

□ **Hypothesis 2 (The shill hypothesis):** A higher-than-expected auction final price indicates possible shill bidding.

4. Data Collection

EBay provides listings for a broad range of auctioned items. It also makes available detailed information of each auction and some limited information on sellers and bidders. Because there is no available public auction database, we designed a software agent to collect auction data from eBay. Given some search criteria, the agent is able to retrieve specified data for completed auctions, and store the data on local disks.

We designed the data collecting agent as shown in Figure 1. A central server first obtains the completed auctions' URLs from auction listing pages according to the search criteria. Then the server establishes and maintains a global queue that can be accessed by crawlers to keep track of the URLs. Following the given URLs, the crawlers sequentially scan HTML tags to extract the desired data. The collected data are then stored in a database. An advantage of this design is that multiple crawlers work collaboratively to create an efficient data collecting agent. In addition, the task queue can be used to avoid duplicated data collection.

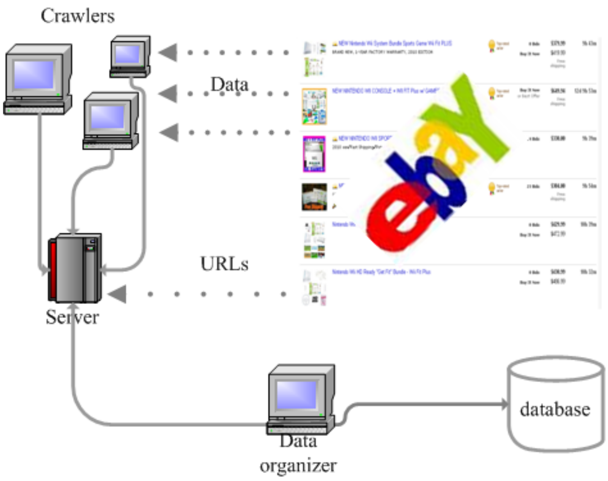

Figure 1. Data collecting agent

The data used in this project is under the category of Nintendo WII game console and systems. Note that although the broad categorization of the data is WII game console system, the items bundled with the game systems vary from auction to auction. For example, the items for sale in one of the auctions include a WII game console and a new WII FIT, which is an accessory of WII game system; while in another auction the auctioned item is a bundle of Nintendo WII System Console, Steering Wheel and 13 Games. Therefore, the category of WII console system is still a broad category that contains many types of items. This partially explains why the prices of this category cover a wide range. The final price distribution for the data used in this study is shown in Figure 2.

For each auction, we collected data that is filled in by the seller when listing an item for auction, including information about the seller, details of the item (name, specifications, description, photos, etc.) and attributes about the auction (length, starting bid, reserve price, shipping charges, etc.). The data is processed to extract attributes and create new attributes that are then used to predict the final price for that auction. The data features classified in 4 different groups are listed in Table 1.

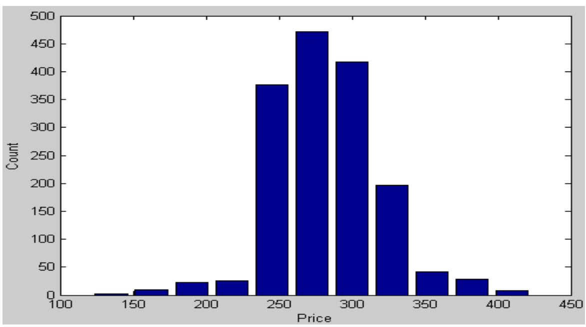

Figure 2. Final price distribution in the study

Table 1. Data features in 4 different groups

Group Name	Features	
Item	Condition (new, used, refurbished, unspecified)	Shipping cost ($)
	Number of pictures	Description
Seller	Reputation percentage (%)	Reputation score
	Has web-store (Yes/No)	TopRatedseller (Yes/No)
Bid Details	Number of bids	Number of bidders
	Duration	Starting time
	Starting bid ($)	End Time
	Month	Day
Category-Specific	Fit (Yes/No)	Game (number)
	Bundle (Yes/No)	Wheel (Yes/No)

5. Experiments

In this section, we report experiments we conducted to study the two hypotheses proposed in Section 3.

5.1 Overview

First, we built and trained a neural network to predict auction prices (the neural network based model is introduced in Section 5.2). Once the neural network based price predictor achieved good performance, we employed the price predictor to predict the final prices of new auctions that were not used in the training and testing phases. Since the price predictor can achieve a relatively high accuracy, we consider the predicted prices as "expected" prices. We selected 30 auctions whose predicted prices were higher than their actual prices and another 30 auctions whose predicted prices were lower than their actual final prices. The two groups of data were used to test Hypothesis 1 and Hypothesis 2, respectively. Next, we computed a shill score, as described in [9], for each of the 60 auctions. The shill score is the highest belief of shilling behavior among all bidders in an auction. The shill score is a number between 0 and 1 (inclusive) that indicates the likelihood of an auction involving shills. The shill score and belief of shill are defined in Section 5.3. In brief, the predicted price is compared with the actual final auction price. If the actual price is higher than the expected price, and the shill score for the auction is higher than or equal to 0.9, the shill hypothesis is verified. Otherwise, if the shill score is lower than 0.9, the shill hypothesis is not verified. Similarly, if the expected price is higher than the actual price, and if the shill score for the auction is less than 0.5, the non-shill hypothesis is verified. Otherwise if the shill score is higher than 0.5, the non-shill hypothesis is not verified. Note that the thresholds of 0.9 and 0.5 are subjective values defined in [9]; but may be adjusted later based on our further experience. In the following sections, we explain further the process for obtaining expected prices and the method for computing shill scores.

5.2 Price Prediction

We built a price predictor based on the Large Scale Memory Storage and Retrieval (LAMSTAR) network [11]. The LAMSTAR, which combines Self Organization Map (SOM) and statistical decision tools, has been successfully applied to diagnosis, prediction and detection type of applications [12]. The trained network for predicting auction final prices is shown in Figure 3.

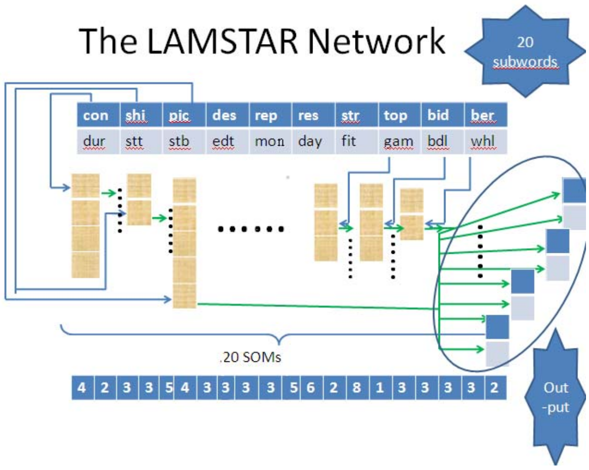

Figure 3. The network for predicting auction price

The grids on the top of the figure are the subwords, representing features from Table 1. Each of them is abbreviated by three letters. These features are preprocessed and then provided as inputs to the neural network. For every subword there is an associated SOM module (in the middle of the figure) that is used to store and retrieve information in the training process. For each subword, a winning neuron in the associated SOM module is determined based on the similarity between the input and a storage-weight vector (stored information).

In the middle of the figure, the arrows between SOM modules encode the correlations between them. The links-weights between different SOM modules and the link-weights from the SOM modules to the output decision layer are continuously trained during normal operational runs. They are adjusted on a reward/punishment principle. Specifically, for the weights of links to the output layer, if the output of the particular output neuron is desired, the link weight for that neuron is rewarded by a small non-zero increment, while if the output is not desired, the link weight is punished by a small non-zero decrement. The link-weights between SOMs can be trained in a similar way.

The grids on the rightmost part of the figure (pointed to by arrows and circled in an oval) are the output decision layer of the network. The network is designed with multiple output layers and each layer consists of two neurons, so each layer represents a binary classifier: whether the final auction price is within certain $*X* range or not. The value of *X* in this study is set as 50. In other words, the price predictor is trained to predict if an auction's final price is in a $50 range or not, such as ($185, $235], rather than a specific numeric value. Since the minimum price in the collected data set is $135 and the maximum price is around $410, there are 6 output layers in the neural network.

Because the actual final price is a specific number, while the expected price is defined using a range, to compare these prices, the actual price is compared to the average price of the expected range. For example, if a predicted price falls in the range ($185, $235], $210 is used as the comparator. In this study, we define "higher" as at least $50 more and "lower" as at least $50 less.

After training the neural network on 1000 auctions and testing it on another 600 auctions, the neural network achieved a precision as high as 95%. Thus, given an auction, the price predictor can determine with a small chance of error if the final auction price will fall in a range span of $50. Note that in this experimental study, we only predict the final price of one specific category of items. We leave the work of predicting the final price of general items as future work.

5.3 Shill Analysis

In [9], a shill certification method based on the mathematical theory of evidence, Dempster-Shafer (D-S) theory was introduced. Six bid-level properties and two auction-level properties are quantified to compute the *belief of shill* for every bidder in an auction. The bid-level properties include the time of a bidder's last bid in an auction, the bidder's concurrent bidding activities, the bidder's reputation score, the bidder's average bid increment, the bidder's winning ratio, and a bidder's affinity for the seller. The auction-level properties include the number of bids and the starting price of the auction.

The D-S theory considers a universe of discourse Θ (also called frame of discernment) that consists of a finite set of mutually exclusive atomic states in a problem domain [8]. For example, in the auction shill detection domain, the frame of discernment for a bidder is Θ = {shill, ~shill}. The power set 2^{Θ}, which is the set of all possible subsets of Θ including the empty set, can be denoted as $2^{\Theta} = \{\emptyset, \{shill\}, \{\sim shill\}, \Theta\}$. The D-S theory assigns a belief mass to each subset of the power set by function $m: 2^{\Theta} \rightarrow [0,1]$. The function is called basic mass assignment (BMA) if it satisfies the following two equations:

$$\sum_{A \in 2^{\Theta}} m(A) = 1 \quad (1)$$

$$m(\emptyset) = 0 \quad (2)$$

Given a certain piece of evidence, $m(A)$ represents one's belief exactly on state A, not any subset of A. The empty set Ø represents a contradiction, which cannot be true in any state. Therefore, the BMA for Ø is assigned 0. The basic mass assignment $m(\Theta)$ can be interpreted as the total ignorance of the problem domain, where one feels uncertain about the truth because every state is present. For the shill detection problem, Eq. (1) and Eq. (2) imply that $m(shill) + m(\sim shill) + m(\Theta) = 1$.

To obtain the overall belief on state A, one must take the sum of beliefs on all subsets of A. As defined in Eq. (3), a belief function is defined as the mass sum of all Bs, which are subsets of A. Thus, in D-S theory, a degree of belief is represented as a belief function rather than a Bayesian probability function, and mass values are assigned to sets of elements rather than singletons.

$$bel(A) = \sum_{B \subset A} m(B) \quad (3)$$

Based on Dempster's rule of combination, the formula for computing the belief of shill is as follows:

$$belief(shill_i) = m(shill_i) \quad (4)$$

$$m(shill_i) = m_1(shill_i) \oplus m_2(shill_i) \oplus \cdots \oplus m_n(shill_i) \quad (5)$$

The shill certification approach is demonstrated to be effective and accurate. In the experiments, we compute the shill score for every bidder in an auction. The shill score for an auction is defined as the highest shill score among the bidders. The auction is suspected to involve shill bidding when the shill score for the auction is higher than 0.9; while the auction is considered to be free of shills when the shills score for the auction is less than 0.5.

5.4 Results and Discussion

After building the price predictor, we ran the price predictor on a distinct collection of data. As mentioned in Section 5.1, we selected 30 auctions that have predicted prices higher than actual prices; we name this group of data as Higher Group. We also selected another 30 auctions that have predicted prices lower than actual prices, and name this as Lower Group. According to the hypotheses proposed in Section 3, the shill scores of most of the auctions in Higher Group should be under 0.5 and the shill scores of most of the auctions in Lower Group should be above 0.9.

The actual prices, predicted prices, and the shill scores for Higher Group and Lower Group are shown in Figure 4 and Figure 5, respectively. In order to make shill scores visible in the figures with a Y-axis scaled from [0, 400], the shill scores are multiplied by 400 in both figures. To enhance visibility, points are connected by lines.

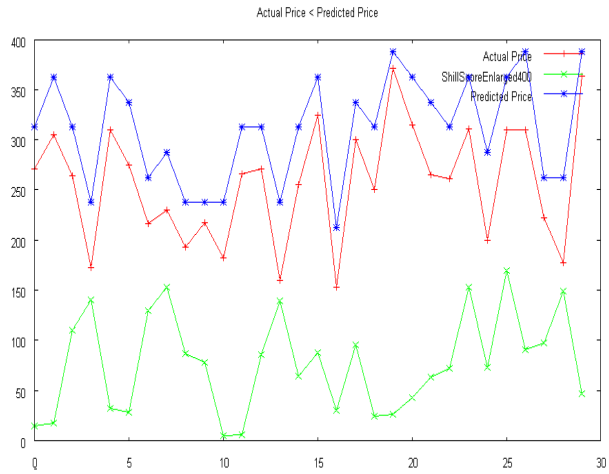

Figure 4. Analysis results of Higher Group auctions

From Figure 4, we can see that when the actual prices are less than the predicted prices, the shill scores for all 30 auctions are under 200, which means that each of them is smaller than 0.5 before being scaled by 400. Hence in accordance with the shill certification rules, no auctions in this group involved shill bidding activities. Therefore, the experimental results support the non-shill hypothesis (Hypothesis 1). .

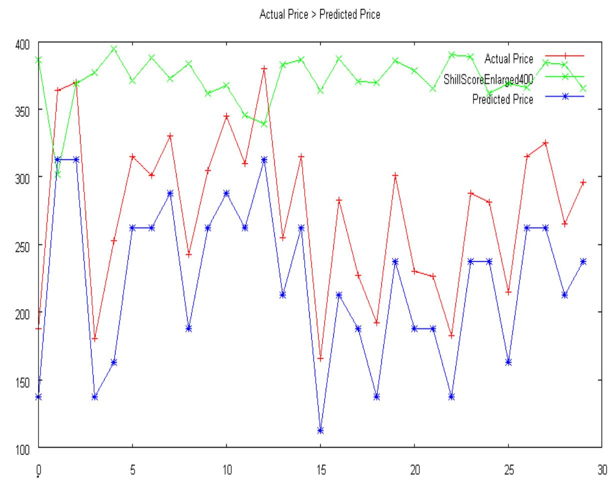

Figure 5. Analysis results of Lower Group auctions

From Figure 5 we can see that when the actual prices are higher than the predicted prices, most shill scores are beyond 360 (i.e., bigger than 0.9 before being scaled). However, there are several exceptions: 3 out of 30 shill scores are smaller than 0.9 but still much higher than the non-shill threshold, 0.5. As discussed in [9], a bidder with a shill score between 0.5 and 0.9 implies the bidder is suspicious rather than innocent. Thus, the shill hypothesis (Hypothesis 2) is also supported by the data.

The experimental results provide us two very important pieces of evidence for shill bidding investigation: 1) if an auction has a higher-than-expected final price, the auction is likely to involve shills, and 2) if an auction has a lower-than-expected final price, the auction is unlikely to involve shills. Note that we did not explicitly consider "normal" (or as-expected) priced auctions, but we believe shilling activities are not likely in such cases as in the auctions with low-than-expected final auction prices.

The auctions with lower-than-expected prices can be eliminated from the set of auctions that need to be investigated. Moreover, the two pieces of evidence can be easily obtained. Both experienced investigators and ordinary bidders can use the price-based evidence as signals for determining the trustworthiness of a seller.

Note that in order to make our approach more practical, we need to consider the following three issues: First, since the market of an item may change with time, the price predictor should be trained periodically in order to make the expected price as accurate as possible. Second, in this paper, we only analyzed a small number of auction data in one category to verify the hypotheses; using more auction data from other categories may strengthen our results. Third, we should determine if the price of a general category of items can be predicted as accurately as that of the WII gaming system studied in this paper.

6. Conclusions and Future Work

Shill bidding has been a serious issue faced by innocent bidders in online auctions for a long time. However, due to the characteristic of concealment of shill bidding activities, both bidders and investigators lack an easy yet effective way to evaluate the trustworthiness of a seller. Thus, there is a pressing need to explore a new simple method to investigate auction shills. In this paper, we first propose two hypotheses: (1) A lower-than-expected final auction price indicates that shill bidding is less likely to occur in the auction, and (2) A higher-than-expected auction final price indicates possible shill bidding. We then present experiments to test these hypotheses. The experimental results indicate that both hypotheses can be used to provide direct evidence in determining auctions with shill bids. The evidence can help people distinguish auctions with shills from auctions without shills, so effort can be focused on auctions with potential shills and thus save precious investigation time.

In our future work, we will look to improve the precision of the price predictor as well as the shill certification techniques. We will also strengthen the empirical study in terms of considering prediction-price intervals that reflect prediction errors, and performing analysis aimed at uncovering threats to the validity of the findings. In addition, we will also consider studying other types of auctions with unexpected prices in order to widen the scope of our results.

Acknowledgement. We thank D. Graupe for his advice on designing and improving the precision of the LAMSTAR neural network.

References

[1] F. Dong, S. Shatz, and H. Xu, "Combating Online in-auction Fraud: Clues, Techniques and Challenges," *Computer Science Review*, Vol. 3, No.4, 2009, pp. 245-258.

[2] Consumer Affairs, "Shill Bidding Exposed in Online Auctions," Retrieved on January 31, 2010 from http://www.consumeraffairs.com/news04/shills.html

[3] D. V. Heijst, R. Potharst, and M. V. Wezel, "A Support System for Predicting eBay End Prices," *Decision Support Systems,* Vol. 44, No. 4, 2007, pp. 970-982.

[4] R. Ghani and H. Simmons, "Predicting the End-price of Online Auctions," *Proc. of the International Workshop on Data Mining and Adaptive Modelling Methods for Economics and Management*, Pisa, Italy, 2004.

[5] D. Lim, P. Anthony, and C. Ho, "Agent for Predicting Online Auction Closing Price in a Simulated Auction Environment," *Proc. of Pacific Rim International Conference on Artificial Intelligence - PRICAI*, 2008, pp. 223-234.

[6] R. Patel, H. Xu, and A. Goel, "Real-Time Trust Management in Agent Based Online Auction Systems," *Proc. of the 19th International Conference on Software Engineering and Knowledge Engineering*, July 9-11, 2007, Boston, USA, pp. 244-250.

[7] H. Xu and Y.-T. Cheng, "Model Checking Bidding Behaviors in Internet Concurrent Auctions," *International Journal of Computer Systems Science & Engineering*, Vol. 22, No. 4, 2007, pp. 179-191.

[8] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[9] F. Dong, S. M. Shatz, and H. Xu, "Reasoning Under Uncertainty for Shill Detection in Online Auctions Using Dempster-Shafer Theory," *To appear in Int. Journal of Software Engineering and Knowledge Engineering (IJSEKE)*, Vol. 20, No. 7, Nov. 2010.

[10] R.J. Kauffman and C.A. Wood, "The Effects of Shilling on Final Bid Prices in Online Auctions," *Electronic Commerce Research and Applications*, Vol. 4, No. 1, 2005, pp. 21-34.

[11] D. Graupe and H. Kordylewski , "A Large-Memory Storage and Retrieval Neural Network for Adaptive Retrieval and Diagnosis," *International Journal of Software Engineering and Knowledge Engineering*, Vol. 6, No.1, March 1998, pp. 115-138.

[12] D. Graupe, *Principles of Artificial Neural Networks (2nd Edition)*, World Scientific Publishing Co. Pte. Ltd., Singapore, 2007.