ELSEVIER

Expert Systems with Applications 31 (2006) 542–550

Expert Systems with Applications

www.elsevier.com/locate/eswa

Predicting the final prices of online auction items

Li Xuefeng *, Liu Lu, Wu Lihua, Zhang Zhao

School of Economics and Management, BeiHang University, Mailbox 10-42 XueYuan Road No. 37, Beijing 100083, China

Abstract

With the prevalence of the Internet and e-commerce, the online exchange market, especially the online auction market develops very fast. The activities of online auction produce a large number of transaction data. If utilized properly, these data can be of great benefit to sellers, buyers and website administrator. Typically, the final price prediction results may help sellers optimize the selling price of their items and auction attributes. At the same time, part of the information asymmetry problems may be solved for buyers. Thus, transaction time can be shortened and cost can be saved. In this paper, we collect large amounts of historical exchange data from Eachnet, an online auction website most famous in China and use machine learning algorithms and traditional statistical methods to forecast the final prices of auction items. We propose an attribute construction method to overcome the problem that auction bid list changes dynamically. Some experiments are performed and the prediction results are discussed to verify the proposed solution.

Keywords: Online auctions; Price prediction; Machine learning; Neural network

1. Introduction

E-commerce, especially the online auction has experienced great development for the recent years. Everyday, there are thousands of users to do their businesses and exchanges on the Internet. The online auction market has already become an important business entity. According to estimation, online exchanges will take 25 percent of all business exchange amount by 2005. As the online auction website most successful in the world, eBay was in possession of 135 million register users and contained 1400 million auction items in its website by the end of 2004. Influenced by the flourish of global online auction market, the online auction business in China burgeons fast too. Eachnet (www.ebay.com.cn, Ebay–Eachnet) who once was the biggest online auction website in China and was purchased by eBay in 2004 witnessed the progress of Chinese online auction market. According to Iresearch China online auction research report 2004, While the market size will go up to 21 billion Yuan with an annual average growth rate of 84 percent, the users of Chinese online auction market will amount to 35 million in 2007 with an annual average growth rate of 43 percent during the next 3 years.

Different from the traditional auction mode, online auction can produce a large number of electronic data during the exchange process. These data contain sufficient information on products and economic behaviors of actors. If utilized properly, they will bring sellers, buyers and the website administrator great benefits.

Online auctions are exchange mechanism for determining prices of items in the Internet. They connect buyers and sellers together in ways that were impossible previously. Although online auctions provide many benefits over traditional exchange methods, they also have several problems. One of the main problems observed on online auctions is the winner's curse. The winner's curse is that bidder pays a price higher than the true value of the item. The cause for the occurrence of winner's curse is lack of the information on the part of the bidder regarding the true vale of the item (Metha & Lee, 1999). This shows that the information asymmetry, which is caused either by the lack of expertise or lack of pricing information of products, still exists in online auction.

Since a number of attributes, e.g. the feedback rating of a seller, freight and the auction start and end time can vary over time in the auction situation, it is a challenging task to use traditional machine learning technique to predict the final prices of the online auction items. This paper employs the attributes of the seller and the auction item, the intrinsic attributes of the auction itself, and the historical exchange data to predict the final prices of the auction items. By learning, a stable prediction model is acquired first. Then the prediction is performed and the prediction results are provided.

Compared with previous work, the contributions of our study can be summarized as follows: we provide a solution for

* Corresponding author. Tel.: +86 108 232 4787; fax: +86 108 231 5601.
E-mail address: leexuefeng@263.net (L. Xuefeng).

0957-4174/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2005.09.077

dealing with the time-dependent dynamics of bidding. By attribute selection and construction, traditional machine learning algorithms and statistical methods can be used to predict the final price of the online auction items. We also conduct a comparison between the prediction qualities of regressions with neural networks and show that clustering of data has a measurable effect on prices prediction result.

The rest of this paper is organized as follows. The Section 2 summarizes the related work to our research. In Section 3, we describe the Ebay–Eachnet's auction mechanism, and introduce the collection and preprocessing of exchange data. Construction and selection of attributes are also presented in detail. Section 4 reports our experimental work. It provides the details of data set, evaluation metrics, procedure and results of different experiments, as well as the discussion of the results. Section 5 provides some concluding remarks and directions for future research.

2. Related work

In this section, we briefly present some of research literature related to data analysis of online auctions and the final price or the winning price prediction.

Considerable work that applied traditional techniques to the online auction analysis has been made in the economics domain. In the data mining and information systems field, there also existed some researches on online auction. For example, Bapna found that significant heterogeneity exists among the users of electronic markets like Ebay and developed a stable taxonomy of biding behavior in online auctions (Bapna et al., 2004). Bajari and Hortacsu (2003) explored the determinants of bidder and seller behavior. The winner's curse is that bidder pays a price higher than the true value of the item. Lucking-Reiley analyzed the effect of various Ebay features on the final price of auctions (Lucking-Reiley, Bryan, Prasad, & Reeves, 2000). They found that seller's feedback ratings have an important effect on his auction prices, with negative comments having a much greater effect than positive comments. However, these researches mainly focused on the historical exchange data analysis. They only described the past auction and did not involve the price prediction (Shah, Joshi, Sureka & Wurman, 2003). As for the researches to use the machine learning techniques to predict the final price of the online auction items, fewer can be found.

There has been some work on price prediction of items in online markets, e.g. airlines fares (Oren Etzioni, Rattapoom Tuchinda, Craig A. Knoblock, & Alexander Yates, 2003). They explored how to predict ticket purchase price through mining airfare data to minimize ticket purchase price. In the AI community, literature on time-dependent product price predition was fewer. The trading agent competition (TAC) (Wellman, Reeves, Lochner, & Vorobeychik, 2002; 2004) was the only work that involved implicitly prices prediction in auctions. TAC relied on a simulator of airline, hotel, and ticket prices and the competitors built agents to bid on these. TAC simulated flight prices using a stochastic process that followed a random walk with an increasingly upward bias. In addition, the TAC auction of airline tickets assumed that the supply of airline tickets is unlimited. Several TAC competitors have explored a range of methods for price prediction including historical averaging, neural nets, and boosting. In the TAC, all of work was performed with artificially generated data and did not use any real auction data.

In this paper, we focus on solving this problem on how to overcome the dynamics of bidding which changes drastically over time. On the basis of data collection, construction and selection of attributes, prices prediction is done.

3. Ebay–Eachnet's auction mechanism, data collection and preprocessing

3.1. Ebay–Eachnet's auction mechanism

Ebay–Eachnet is the largest and most successful consumer–consumer auction site in China. It has possessed of 11.6 million register users at the end of the first quarter in 2005. The exchange amounts exceeded 2500 million Yuan in 2004.

Similar to Ebay, Ebay–Eachnet runs a second price English auction for a variety of consumer goods. The auction is 'second price' because the winner pays the next highest bid, and they are 'English auctions' because bids are ascending. The seller can decide to have an auction last for 1, 3, 5, 7 or 10 days. When a potential bidder views a current auction on EBay–Eachnet, they see a description of the item for sale. This description is up to the seller's discretion, but is usually fairly detailed and may include a picture of the item. The bidder also can review the recent history of the seller and ratings by previous customers. Most importantly, the bidder observes the current high bid for the item. They also observe information on other bidders for that auction including when they bid, but not what their bid was. Lastly, all bidders know the exact time remained before the close of the auction.

After observing this information, the bidder may decide to bid on the item. The amount entered by the bidder is actually a 'proxy' bid. That is, EBay–Eachnet will take that bid and automatically bid slightly above the next highest bidder up to the amount entered. This system allows the bidder to leave the auction, without having to worry about being outbidden, as the proxy system will keep on bidding automatically. However, if a bidder is outbidding, then Ebay–Eachnet can notify them by email. The winner of the auction is the highest bidder at the end of the auction. The winner pays an amount slightly above the next highest bid.

In addition, Ebay–Eachnet also runs 'buy it now' and 'English auction + buy it now'. In this paper, only English auction is studied.

3.2. Data collection

A lot of information on Ebay–Eachnet auctions is publicly available. Ebay–Eachnet posts on its website the complete bid histories of closed auctions for at least one month after the ending date. We used a specially designed data collection program to gather the exchange data from Eachnet. The collection program first constructs URL dynamically according to product ID. Then it