Available online at www.sciencedirect.com
ScienceDirect
Decision Support Systems 42 (2006) 1521–1538
Decision Support Systems
www.elsevier.com/locate/dss

Exploring auction databases through interactive visualization

Galit Shmueli [a,*], Wolfgang Jank [a], Aleks Aris [b,c], Catherine Plaisant [b], Ben Shneiderman [b,c,d]

[a] Department of Decision and Information Technologies, Robert H. Smith School of Business, University of Maryland, College Park, MD 20742, USA
[b] Human–Computer Interaction Laboratory, Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA
[c] Department of Computer Science, University of Maryland, College Park, MD 20742, USA
[d] Institute for Systems Research, University of Maryland, College Park, MD 20742, USA

Received 11 May 2005; received in revised form 28 December 2005; accepted 18 January 2006
Available online 9 March 2006

Abstract

We introduce *AuctionExplorer*, a suite of tools for exploring databases of online auctions. The suite combines tools for collecting, processing, and interactively exploring auction attributes (e.g., seller rating), and the bid history (price evolution represented as a time series). Part of *AuctionExplorer*'s power comes from its coupling of the two information structures, thereby allowing exploration of relationships between them. Exploration can be directed by hypothesis testing or exploratory data analysis. We propose a process for visual data analysis and illustrate *AuctionExplorer*'s operations with a dataset of eBay auctions. Insights may improve seller, bidder, auction house, and other vendors' understanding of the market, thereby assisting their decision making process.

Keywords: Online auctions; Bid history; Auction dynamics; User interface; Time series

1. Introduction and motivation

Empirical research of online auctions has been flourishing in recent years due to the important role that these auctions play in the marketplace, and the availability of large amounts of high-quality bid data from websites such as eBay, Yahoo!, OnSale, and uBid. Academic researchers as well as practitioners in the eCommerce world have been trying to unveil relationships between selling, bidding and winning. In particular, a popular question has been the influence of auction design, auction dynamics, and market characteristics on the auction outcome. This information is then used to come up with recommendations for sellers, buyers, and auction sites.

In practice, there exist a variety of online services that are aimed at providing recommendations for bidders or sellers in online auctions (e.g., www.coolebaytools.com). Quite often, these recommendations are based on historical auction data and are used for strategic decision making. The online service hammertap.com, for instance, provides historical data from eBay's auctions to learn about the best-selling products, the most profitable eBay categories, and the most successful sellers and

* Corresponding author. Tel.: +1 301 405 9679.
E-mail addresses: gshmueli@rhsmith.umd.edu (G. Shmueli), wjank@rhsmith.umd.edu (W. Jank), aris@cs.umd.edu (A. Aris), plaisant@cs.umd.edu (C. Plaisant), ben@cs.umd.edu (B. Shneiderman).

0167-9236/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2006.01.001

their strategies. A similar service, Whizanalysis.com, provides complete sales analysis of the entire eBay market and identifies products and sellers that "make money." Many of these services are geared towards eBay.com, but tools for other online auction sites like Yahoo! or uBid.com also exist. Services based on historical data typically provide statistics about auctions, products, categories and sellers. These statistics arrive in the form of summary tables or simple charts. One problem is that such services do not clean the data, and therefore the historic data, and their statistics, are often based on irrelevant information (e.g., a search for a certain model of an iPod device will combine prices of both iPod devices and accessories for that model). Even if the data are manually cleaned, such summarizations are limited. While information about past transactions can be of potential strategic value to bidders and sellers, their summary in static tables and charts does not reveal their full power. Rather than summarizing the information in a static way, users would benefit from the ability to interactively explore the information. An interactive exploration would allow users to obtain an overview of all the auctions of interest and, if desired, to drill down on selected auctions and their attributes for investigating specific patterns and anomalies. Current approaches do not allow the user to explore auctions in such an interactive way.

The limited information that is provided by commercial services has led most academic researchers to collect their own raw data from online auction websites, and use these data for modeling purposes. The range of research questions that have been tested empirically is wide: multiple researchers [3,20] used eBay data to investigate the determinants of price. Ref. [4] used uBid.com data to detect types of bidding strategies. Ref. [22] compare eBay and Amazon data in terms of auction dynamics towards the auction end ("last minute bidding"). Ref. [27] used OnSale.com and SurplusAuction.com data to model the arrival times of bidders to online auctions. Ref. [10] examined eBay's reputation mechanism to learn about trust between buyers and sellers. Refs. [3] and [7] used eBay data to compare bidding strategies in auctions for common versus private value items. Ref. [16] used eBay data to study the price evolution and its dynamics, showing that dynamics differ markedly even within comparable auctions for the same good. Finally, Ref. [18] use eBay data to detect the fraudulent act of bid shilling (a seller placing dummy bids on their auction to raise the price). These are just a few of the many empirical studies that have appeared in the online auction literature.

Empirical research in online auctions has focused almost exclusively on collecting data and/or fitting statistical or other analytical models to the data, while exploratory data analysis (EDA) has been rarely addressed. EDA is a crucial first step in any data analysis and can aid in detecting data-anomalies or interesting patterns in the data. Typical tools for EDA are of static nature (summary statistics, tables and charts) and require the user to specify in advance which relationships are to be scrutinized (e.g., the user has to specify to the software first that a table of X vs. Y is desired which is then produced). This approach requires the user to have some advance knowledge of interesting patterns and will therefore not be as powerful in detecting surprising patterns. An interactive approach can be more powerful and can lead to a better understanding of the database. This is especially so for auction data which are typically of complicated nature. Current static tools will not fully display the information in auction databases for the following reasons:

(1) the need to accommodate two forms of data: typical auction data include *auction attributes* (such as seller's design choices, item attributes, seller information) and the *bid history* that describes the bids as they arrive during the auction. Users often need to study relationships across these differing data forms. In classical time series analysis, there is usually a single or a small number of time series that are of interest. In contrast, here we have a large number of auction attributes that are coupled with bid histories. Existing statistical summaries and graphical methods tend to cater either to cross-sectional data for showing auction attributes (histograms, summary statistics), or to time series (such as time plots) for showing bid histories. Combined visualizations are not standard. Our goal is to contribute in this aspect.
(2) the unique structure of bid histories makes it difficult to apply existing time series analysis tools. The bid history is a non-standard time series: it consists of bids placed at unevenly spaced time points with sparse and dense time periods. Furthermore, if we look at many auctions simultaneously, these time series differ markedly: auctions have different durations, bids in different auctions are placed at unrelated time points, and the total number of bids can vary vastly.

In a recent paper, Ref. [23] introduced several general graphical techniques for visualizing online auction data