BUILDING BUSINESS HEURISTICS WITH DATA-MINING INTERNET AGENTS

Steven Walczak
University of Colorado, Denver
swalczak@carbon.cudenver.edu

Dawn G. Gregg
University of Coloraod, Denver
dawn.gregg@cudenver.edu

Abstract

The information age is in full swing with new information being deposited on the world-wide-web (web) at exponentially increasing rates. Likewise, the ever more global economy is increasing the complexity of business decision-making. Business heuristics help simplify complex business problems and enable near optimal decision making with partial information. This article examines the use of web data-mining through internet agents for the development of business heuristics. The case of developing heuristics for the e-commerce setting of online auctions is examined.

Introduction

The world-wide-web (web) has promulgated a virtual revolution in information availability and services, with an estimated 4 billion pages of information content (Lawrence & Giles 1998). There are over half a billion Internet users in the United States and Canada alone and the market for Internet access and value-added services was estimated at over $154 billion (Datamonitor 2000). Furthermore, online content subscriptions ended 2002 with over $1.3 billion in revenue, which is a 95 percent increase in just one year for revenues from individuals or organizations willing to pay for online content (OPA 2003), and it is estimated that online content doubles every two years (Lawrence & Giles 1998).

With more information more readily available now than ever and the introduction of data warehousing as a means for storing extremely large quantities of information and data-mining as a means for extrapolating knowledge from the stored information, the late 1990's and early 2000's have been dubbed "the knowledge age" (Wiig, 1993). Companies have mastered collecting volumes of data from data warehouses, ERP and CRM systems but have yet to master the complex task of making decisions based on this information (Walczak 2001). In addition, few companies even attempt to automate collection and use of the vast amount of web based data to support their internal decision-making. Data sleuthing or "business intelligence" (BI) systems are one way organizations are beginning to utilize their internal information resources (Schlosser 2003). As these systems evolve they will need to support the automated mining and sleuthing of both internal and external data resources.

Solutions to business problems, especially in complex domains, are often approximated through the use of heuristics (Barr & Feigenbaum 1981). Developing business heuristic based on available information (inference) is one way business can create business intelligence from raw data. Acquiring the most relevant information available is a problematic issue now due to the rapid and global access to information permitted by the web. Business heuristics prior to the PC revolution were developed using intuition, induction, and deduction. While these same processes apply to current business heuristic development, utilizing information for inductive and deductive reasoning to develop new business heuristics requires higher performance due to the exponential increase in information availability.

Software agents can be used to automatically perform information retrieval and processing (mining) from the web as well as traditional databases and data warehouses. Utilizing agents to automatically acquire and synthesize information from various information resources facilitates the development and testing of new business heuristics. This article examines how agents and other services currently retrieve information from the web and then describes how information revealed through agent information acquisition may be used to develop e-commerce business heuristics in the domain of online auctions. Once developed, new business heuristics may be tested empirically with similar Internet agents to determine the validity and reliability of the proposed heuristic solution.

Agents in Web Business Intelligence Systems

An essential part of any BI system is the acquisition of the information necessary for identifying relevant trends, problems or opportunities facing an organization. This is traditionally accomplished using data-mining. When the data source is a traditional database or data warehouse, information retrieval is done through query engines. Information retrieval from the web is problematic in that all of the possible data sources may not be identified and the data content is highly dynamic, changing from minute to minute (Ackermann and Hartman 1998).

Traditional web-search engines use an index of catalogued web sites that may be several days or even weeks old to track and identify keywords on web sites for retrieval (Ackermann and Hartman 1998). Keyword searches on unconstrained domains like the web will return a large quantity of matching web sites, many of which are completely unrelated to the desired search criteria (but simply contain the keyword). Meta-search engines such as Dogpile and MetaCrawler attempt to improve the relevancy of search finds, by utilizing the indexed searches of multiple standard search engines and then aggregating them based on the number of different search engines that returned hits for a specific page and other relevancy criteria.

Agents provide a means for conducting information searches on the web that are both current and more relevant. A software agent is an autonomous system with varying levels of mobility and intelligence (Maes 1994, Papazoglou 2001). Autonomy implies that the software agent performs its tasks with little or no human direction. Necessarily, agents for heuristic development and verification require mobility, the ability to traverse an intranet or Internet, to acquire information that then may be used for induction, deduction, or empirical testing of heuristic rules.

Agents have been previously developed to serve as personal assistant/reminders, perform Internet purchasing (for a specified product), and as information retrieval specialists (Maes 1994). Examples of agents that perform information retrieval and data-mining include: an agent that data-mines a very large CRM database to plan for future customer needs (Durbin et al. 2002). an agent that autonomously tracks browsing paths of users to identify new information sources (Fong & Wong 2002), and agents that extract information from multiple heterogeneous information sources including the web and internal databases (Chalupsky et al. 2002).

One advantage of agent-based systems is that new information retrieval and data-mining techniques may easily be incorporated into agents as they become available (Bigus & Bigus 2001) without affecting other functionality of the agent. These up-to-date agents can then be used to support evolving organizational BI systems.

Development of E-Commerce Business Heuristics Through Agent Data-Mining

Hagel and Singer (1999) state that consumer efficiency is dependent on having knowledge of their own information aggregated with that of other consumers and the marketplace. The web has enabled e-commerce retailers to provide goods and services to consum related to ers anyplace in the world 24 hours a day, thus significantly increasing the need for and availability of information other consumer preferences and behaviors. Agents provide a tool for acquiring vast quantities of information and data-mining the acquired information which can then be applied to develop heuristic methods to make sense of all the latest information.

Agents are a reliable technology for gaining access to the most current information available on the web (Papazoglou 2001). However agents must still be told what information to collect, which corresponds to the intuitive part of business heuristic development. As information is being collected from the web and other information resources, such as a corporate database of current customer accounts, the intelligence part of agents is utilized to perform pattern matching and other current data mining techniques to acquire inductive or deductive evidence for business heuristic development. Papazoglou (2001) claims that consumer behavior cannot be predicted with traditional analytic methods, supporting the need for development of business heuristics in online markets.

This research uses the "Auction Advisor" system (Gregg & Walczak 2003) to develop and test new business heuristics. Auction Advisor is a multi-agent system with agents that collect information from online auction sites and additional agents that use business heuristics to advise buyers and sellers on optimal strategies for participating in online auctions. This ongoing research is developing additional agents that will automatically develop new business heuristics to further optimize online auction participation.

The Auction Advisor collects information on quantities of open and closed auctions, current status of open auctions, and results of closed auctions for specific items specified by the user across multiple online auction sites. The information acquired includes the auction bid history. The bid history of closed auctions permits identification of many different values and behaviors that are encapsulated into online auction bidding and selling heuristics. E-commerce in general and online auctions specifically, because of their immediate and global nature, are changing market dynamics and consumer behaviors.

One value that can be tracked using bid history data is the frequency of proxy bidding. Proxy bidding allows a bidder to specify their maximum bid and a proxy bidding agent, supplied by the online auction service, will then incrementally increase the user's bid to maintain a winning bid until the maximum is reached. While proxy bidding can help simulate active bidding at the online auctions, proxy bidding can also artificially drive up the closing price of an auction item (Gregg & Walczak 2003). This knowledge can be used to develop a new e-business auction heuristic:

> *Proposed Heuristic 1: Whenever two or more proxy bids are detected for an open auction, wait to place another proxy bid until closer to the close of the auction.*

Another behavior identified by the auction agents is that different classes of auction products have different percentages of proxy bidding (proxy bidding appears to be more prevalent in the more technological product categories). Knowledge of the lack of proxy bidding would indicate a more specialized type of market and could lead to the following online auction seller heuristic:

> *Proposed Heuristic 2: If proxy-bidding activity in this category is less than the average, utilize special advertising features offered by online auction services to attract more bidders*

These heuristics will help to minimize the buyer cost and maximize the seller return.

Another behavior that is unique to online auctions is a practice called "slamming", which occurs when a bidder waits until the closing minutes of an auction to place a bid (usually within the last two minutes). In traditional English-style offline auctions, bidding is kept open as long as active bidding continues. Most online auctions, e.g. eBay, currently have a hard close time after which no new bids are accepted. Slamming is potentially detrimental to online auction sellers because the practice when widely followed tends to limit competitiveness and lowers the resulting closing price. Additionally, the practice of slamming is detrimental to uniformed bidders who may use a proxy bid early in the auction, just to be outbid in the last minute by a slam bidder.

The information on closed auctions collected by the Auction Advisor agents is analyzed to determine the presence of slamming and the frequency of slamming for products in specific categories. The practice occurs at vastly different levels (percentages of auctions) for different products categories and across different online auction services. Analysis of the slamming data can lead to potential heuristics for both buyers and sellers. A possible seller heuristic is:

> *Potential Heuristic 3: If a product's sales category has a high percentage of slam wins* (where the actual value of high is determined heuristically and for Heuristic 3 is currently 33.3 percent)*, the seller should consider auctioning the item at an online auction service that enables automatic auction extensions, thus permitting continued competition for the auction item and realizing a potentially higher closing price.*

A possible buyer heuristic relating to slamming prevalence is:

> *Potential Heuristic 4: If a product's sales category has a high percentage of slam wins, the buyer should wait until the closing minutes of the auction to place a bid, or should monitor the auction closely to maximize the probability that the buyer can successfully respond to a last minute slam bid by another bidder.*

The data-mining performed by the Auction Advisor agents has permitted the development of several new heuristics that are unique to online auctions. Agents operating in different electronic commerce markets should also result in discovery of new e-consumer behaviors based on acquiring and mining the vast collections of commerce and consumer information available on the web. Consequent heuristics to predict subsequent behaviors in the e-marketplace may then be developed either by individuals or automatically by the agents themselves.

Evaluating Business Heuristics with Agents

When new business heuristics are proposed, the utility of the new heuristic must be evaluated to determine its applicability in current and future business problems (Walczak 2001). Means for evaluating new heuristics include the use of artificial neural networks, simulation, and statistical analysis. Every heuristic needs to be validated with multiple data sets to confirm the reliability of the heuristic. This research uses an empirical analysis to evaluate Potential Heuristic 4 as it would be used in the "real-world". The other three potential heuristics have undergone a preliminary evaluation and validation using simulated auctions modeled after historic auction data retrieved by the agents. Future research will examine the practical application and analysis of the results for the other three potential heuristics.

An example of the validation process is for the heuristic that recommends that a bidder be aware of the potential for slam-style bidding. The product category selected historically had 50 percent slam bidding for closed auctions that received any bids (Gregg & Walczak 2003). Bidding on twenty independent auctions for similar products within the specific auction category is initiated. Proxy bids are placed no closer than one hour before the hard auction close time for six of these auctions and the remaining six all have bids placed within the last two minutes. The other eight auctions exceeded the maximum bid amount recommended by the Auction Advisor agents prior to the opportunity to place a bid.

Results for the 12 auctions on which bids were placed are displayed in Table 1. While the result of twelve auctions is certainly not exhaustive, the consistency of the results implies that future empirical evidence will be similar. The results from these twelve auctions indicate that although only 50 percent slamming was previously identified from the mined agent acquired data, the practice of slam bidding may be much more common for this specific auction category.

Table 1. Evaluation of "Wach Out for Slamming" Heuristic

Auction type	Average/(latest) time of last bid	Percentage of wins	Average time of winning bid
6 auctions without Slam bids	97.5/(63) minutes before close	0	84 seconds before close
6 auctions with Slam bids	81/(47) seconds before close	83.33	69 seconds before close

The six test cases that placed early bids and thus ignored the "watch out for slam bidding in this category" heuristic resulted in zero wins and each was outbid in the final two minutes of the auction. The six test cased that followed a slam bidding strategy won 5 out of 6 auctions. The auction slam bid that defeated the agent-recommended slam bid was placed within 20 seconds of the auction close and the researcher did not have sufficient time to verify a new bid placement. Timing of the winning bids for these auctions was unexpected since the agents had identified that slam bidding occurred on average 50 percent of the time, but for the 12 cases of the heuristic evaluation study 100 percent slam bidding occurred (of course for four of these cases, the evaluation study is responsible for the presence of the slam bid, two of the other slam bid tests received outside slam bids, one before and one after the agent).

The application of the perform slam bidding or "be aware of slam bidding potential" heuristic is conditional on other economic factors. Before placing a bid, the current value of the highest bid must still be lower than the perceived value of the item for the bidder, which was the reason that bids were never placed in the other eight auctions for the auction category in which the heuristic was evaluated.

Summary

Agents are a well known solution for performing information retrieval from the web and other data sources (Chalupsky et al. 2002, Durbin et al. 2002, Fong & Wong 2002, Gregg & Walczak 2003, Maes 1994, Papazoglou 2001). In addition to performing information retrieval, agents are now being used to perform data-mining and inference. This article has demonstrated how knowledge mined from the web for online auctions may be used to develop new e-commerce/online auction heuristics. Furthermore, these same agents with little or no modification may be used to induce the applicability of the newly formed e-commerce business heuristics.

An example of a new consumer behavior enabled by most online auction providers, slam (last minute) bidding, is verified through agent information collection and yielded several new e-commerce business heuristics. The new heuristic, "watch out for slam

bidding" and "perform slam bidding" for appropriate online auction product categories was tested and resulted in zero percent auction wins when the heuristic was ignored and over 80 percent auction wins when the heuristic was followed. Future research is needed to evaluate the development of business heuristics in other e-economy domains through the use of agent technology.

References

Ackermann, E., and Hartman, K. *Searching and Researching on the Internet and World Wide Web*, Wilsonville, OR: Franklin, Beedle, and Associates, 1998.

Barr, A., and Feigenbaum, E. A. *The Handbook of Artificial Intelligence* Volume 1, Reading, MA: Addison-Wesley, 1981.

Bigus, J. P., and Bigus, J. *Constructing Intelligent Agents Using Java*, 2nd Edition. New York: Wiley, 2001.

Chalupsky, H., Gil, Y., Knoblock, C. A., Lerman, K., Oh, J., Pynadath, D. V., Russ, T. A., and Tambe, M. Electric elves: agent technology for supporting human organizations. *AI Magazine* (23:2), 2002, pp. 11-24.

DataMonitor. Powering the Internet Revolution: 1999-2003. Available as of Jan. 2001 at **http://www.datamonitor.com**, 2000.

Durbin, S. D., Warner, D., Richter, J. N., and Gedeon, Z. Information self-service with a knowledge base that learns. *AI Magazine* (23:4), 2002, pp. 41-59.

Fong, J., and Wong, H. K. Online analytical mining of path traversal patterns for Web measurement. *Journal of Database Management* (13:4), 2002, pp. 39-61.

Gregg, D. G., and Walczak, S. E-commerce Auction Agents and Online-Auction Dynamics. *Electronic Markets* (13:2), 2003, in press.

Hagel, J., and Singer, M. *Net worth*. Boston: Harvard Business School Press, 1999.

Lawrence, S., and Giles, C. L. Searching the World Wide Web. *Science* (280:5360), 1998, pp. 98-100.

Maes, P. 1994. Agents that Reduce Work and Information Overload. *Communications of the ACM* (37:7), 1994, pp. 31-40,146.

Online Publishers Association (OPA). U.S. Consumer Spending on Online Content Totals $1.3 Billion in 2002, According to Online Publishers Association Report. Available as of Mar. 2003 at **http://www.comscore.com/news/**, 2003.

Papazoglou, M. P. Agent-Oriented Technology in Support of E-Business. *Communications of the ACM* (44:4), 2001, pp. 71-77.

Schlosser, J. Looking for Intelligence in Ice Cream. *Fortune* (147:5), 2003, pp. 114-117.

Walczak, S. Neural Networks as a Tool for Developing and Validating Business Heuristics. *Expert Systems with Applications* (21:1), 2001, pp. 31-36.

Wiig, K. M. *Knowledge Management Foundations*, Arlington, TX: Schema Press, 1993.