2000

Intelligent Agent-Based Data Mining in Electronic Markets

Merrill Warkentin
Northeastern University, mwarkentin@cba.neu.edu

Ravi Bapna
Northeastern University, rbapna@cba.neu.edu

Intelligent Agent-Based Data Mining in Electronic Markets

Merrill Warkentin, Northeastern University, Boston, MA 02115, mwarkentin@cba.neu.edu
Ravi Bapna, Northeastern University, Boston, MA 02115, rbapna@cba.neu.edu

Abstract

The advent of web-based electronic commerce has brought a tremendous increase in the volume of "collectable data" that can be mined for valuable managerial knowledge. Utilizing intelligent agents can enhance the data mining procedures that are employed in this process. We focus on the role of data mining and intelligent agent technology in the B2C and B2B e-commerce models. By identifying the complex nature of information flows between the vast numbers of economic entities, we identify opportunities for applying data mining that can lead ultimately to knowledge discovery.

Introduction

Millions of people surf the web every day and interact with electronic commerce web sites around the world. While most companies capture user interactions within their web site, they do not utilize its maximum potential. According to Forrester Research, only 18% of the companies it surveyed use their web data for marketing purposes and only 16% use it for customer support. It also indicates that 72% of the companies that collect web data admit that they do not analyze this data, or use it in any application. However, organizations are beginning to realize the value of this web data and are allocating vast resources for creating the necessary infrastructure to analyze this data, which would enable them to learn more about their customers and gain competitive advantage.

In order to get the most mileage out of this "digital footprint" data, each company must decide: 1) what data to collect and how to organize it, 2) what kind of analysis to perform on the data, 3) how frequently to perform data analysis, and 4) how to validate and integrate the results into decision making and planning. Through data mining, a company can synthesize Web site patterns into meaningful information, enabling it to understand and engage customers and prospects over the internet. The mining of Web-based data is the key to creating a lasting relationship with online customers and establishing a productive online storefront.

There is a myriad of data mining tools available in the market that employ a variety of data mining algorithms and techniques. For a novice user, it is often difficult to determine which tools or techniques are appropriate for a particular data analysis and data mining scenario. To reduce some of this cognitive load, "intelligent agents" are beginning to be employed (Chan et al. 1999; Sugumaran, Bose 1999). These agents can automate some of the mundane activities such as data cleansing, data transformation, etc., and also help the user in the selection of appropriate tools and data mining methodology. Typically, intelligent agents act on behalf of the human user in problem solving activities and decision-making.

The objective of this research is to: (a) study the information flow between the various entities in different electronic markets, and (b) investigate how intelligent agent technology can be used in data mining for discovering new relationships and nuggets of knowledge that could be incorporated into managerial decision making.

Generic Architecture for Agent-Based Data Mining

While data mining applications are coming to the forefront of business data analysis and decision making, in order to successfully execute these applications, a significant amount of *a priori* knowledge about the various data mining techniques, their applicability to different scenarios, relevant data selection and transformation, etc. is required. Hence, for a casual user interested in applying data mining techniques to decipher trends and buying behaviors from customer "digital footprint" data, shielding some of the complexities involved in normal data mining operations would be a welcome change. To this end, we adapt the generic agent-based data mining environment proposed by Sugumaran et al. (2000), which would assist the user in appropriate data selection, data cleansing, data mining method selection and execution, and interpreting the results. This generic architecture can be tailored to meet the needs of a particular application domain. It consists of the following agents (Sugumaran et al. 2000): (1) User Interface Agent, (2) Control Agent, (3) Data-Centric Agent, (4) Data Mining Agent, and (5) Visualization Agent.

The above mentioned agents in the generic architecture are designed using object-oriented principles and are modular in nature. Each agent is structured using a standard agent-template, which consists of the following three high level modules: (a) interface module, (b) process module, and (c) knowledge module. The interface module deals with the public interface of the agent, which is visible to other agents and users. The process module contains methods that implement a variety of functions. The agent responds to requests from

other agents and users by executing one or more methods. The knowledge module contains domain specific and domain independent knowledge relevant to problem solving.

Data Mining in B2C E-Commerce

In this section we specify the role of data mining in the B2C e-commerce models. We focus on the flow of information between the various stakeholders and identifies opportunities for applying data mining.

Data Mining Agents and Mass-Customization

One of the promises of e-tailing is that businesses will be able to provide a personalized shopping experience to buyers based on some prior determination of their preferences. This involves having a non-invasive scheme for obtaining such preferences, and subsequently using data mining techniques to determine patterns that can be exploited to enhance the consumers shopping experience. Firms that have been successful in mass-customization have these two tasks inherently built into their core business functionality, rather than as an afterthought. This often provides the elusive stickiness and loyalty attributes that e-tailers yearn for.

Typically, the front end of a web-based B2C storefront consists of modules for product searching, order management, payment processing, and customer relationship management. At the back end there are modules that integrate with the enterprise systems such as inventory control, shipment processing, and order fulfillment. We propose that this architecture be enhanced to include a data acquisition agent and a data mining agent. The former is responsible for connecting to the enterprise data storage and collect micro-level "clickdata" that can subsequently be mined in a variety of ways. The latter connects to the managerial decision support module and supports the knowledge discovery process for managers.

Data Mining Agents in Online Auctions

Data mining agents have a significant role to play in the emerging web-based dynamic pricing schemes such as online auctions. Such auctions and exchanges are ideal pricing mechanisms for managing uncertainty about the price of a good or service. Until recently, the high transaction costs associated with dynamic pricing mechanisms limited their application to specific sectors such as finance, commodities and art. The role of data mining in general and intelligent agent based data mining in particular in such environments becomes important because of the tremendous volume of market data that can be collected and potentially analyzed by businesses.

One such application is the micro-level analysis of consumer bidding patterns in consumer related auctions. These could be B2C auctions of rapidly aging commodities like consumer electronics and computer hardware or C2C auctions such as those on eBay. Such an analysis can reveal significant insights into the heterogeneity of bidding strategies of consumers and can help auctioneers design the auction parameters to exploit such strategies. Bapna, Goes, Gupta (2000) demonstrate how auctioneers can benefit from data mining utilizing an intelligent agent that tracked over 100 B2C online auctions round-the-clock. The intelligent agent was able to keep track of each and every bid that was submitted, including bids that did not figure in the final list of winners. Subsequently the data was mined for bidding patterns adopted by the consumers and the following three types of bidders were identified: a) evaluators (early one time high bidders), b) participators (monitoring the progress of auction and never bidding higher than the minimum required), and c) opportunists (bargain hunters).

While comparing the relative performance of these three groups with respect to loss of surplus, Bapna, Goes and Gupta (2000) found that the evaluators as a group fared worst, the participators were best off, and the opportunists lay in between. In addition, data mining technology allowed the researchers to establish relationships between the auction parameters, such as the bid-increment, chosen by the auctioneer and the mix of consumer bidding strategies.

Data Mining for Adserver Targeting of Banner Ads and Website Content

There are numerous data flows between the individual surfing the web, the adserver, the content-provider website, and other websites in the adserver's network. There is also a very sophisticated revenue sharing model linked to this complex network relationship. Leading adservers intercept the user's click patterns, along with metadata about the content requests users make (click on "Boston weather"), in order to evaluate his surfing pattern. The adserver has also evaluated the user's click patterns (and content requests) at many other sites in its network. To improve the "stickiness" of the content provider site, an adserving company like DoubleClick may share knowledge with the provider about a user's preferences in order to ensure that the provider's server dynamically delivers interesting content to him. Content manipulation agents can be utilized to dynamically serve appropriate content to each user based on agent-based data mining analysis.

Agent-based data mining technology can play a role in the real-time analysis of this clickdata and related information. Data is shared by businesses within the B2B component of this relationship, and between the user and

each of the businesses in this network. Managers in these new marketspaces who effectively implement proper agent-based data mining technologies and mechanisms can gain competitive advantage in the race to build brand equity and lock in consumer loyalty. The knowledge resulting from data mining will assist managers in creating intimate customer experiences that will ultimately lead to improved revenues.

Data Mining in B2B E-Commerce

The role of data mining and intelligent agents becomes important in this sector as one examines the integration of typical back-end functions like inventory control with issues of customer service, marketing and production planning. Because of the vast numbers of buyers and sellers, data-mining agents can play a role in the procurement process by looking for substitutes, prices and supply levels in real-time. Inventory controllers can benefit from better demand forecasting.

In electronic vertical markets, agents have been developed to support the information exchange function. For instance, Application Service Providers (ASP) like www.CommerceOne.com, that build and host vertically integrated trading communities on the Internet, utilize intelligent agents to aggregate horizontal information such as shipping patterns. Subsequently, the agent can use this information to negotiate competitive advertising rates with adserver companies such as www.DoubleClick.com and www.adforce.com. Ultimately, the B2B e-commerce space will far exceed the volume of transactions in the B2C space, given the volume of corporate procurement (about half of all Fortune 1000 spending) and given its mission critical status.

We envisage that the current B2B exchanges will graduate towards being more than simple market makers that match buyers and sellers. The real benefits will come when they support a collaborative environment that will include project management, forecasting and knowledge sharing. Each of these tasks adds layers of complexity to the exchange, leading us to believe that technologies such as agent-based data mining will be in great demand.

Conclusion

Agent-based data mining technologies are being implemented by many companies in both B2C and B2B markets. But for firms using data mining processes within these electronic commerce marketspaces, there are some key managerial implications to consider. Success will be determined by the way the tools are used, not by the tools themselves.

The emergence of data mining and intelligent agents at the same time that millions of individuals have gone online to purchase from thousands of new websites has created an exciting opportunity for practical technological convergence. As e-commerce moves many business processes and activities online, new data streams are born and a chance for greater efficiencies is generated for those willing to carefully perform the correct data analysis procedures. Agent-based data mining will enable firms to capture the full potential of this technological convergence.

References

Bapna R, Goes P, Gupta A. 2000. A Theoretical and Empirical Investigation of Multi-item On-line Auctions. *Information Technology and Management* **1**(1): 1-23.

Chan P. K. , Fan W., Prodromidis A.L., Stolfo S.J.. 1999. Distributed Data Mining in Credit Card Fraud Detection. *IEEE Intelligent Systems and their Applications* **14**(6): 67-74.

Sugumaran V, Bose R. 1999. Data Analysis and Mining Environment: A Distributed Intelligent Agent Technology pplication. *Industrial Management and Data Systems* **99**(2): 71-80.

Sugumaran V., Warkentin M., Bapna R. 2000, The Role of Intelligent Agents and Data Mining in Emerging Electronic Marketspace Relationships, Under review.