Mining for Web-Enabled E-Business Applications

Richi Nayak
Queensland University of Technology, Australia

INTRODUCTION

A small shop owner builds a relationship with its customers by observing their needs, preferences and buying behaviour. A Web-enabled e-business will like to accomplish something similar. It is an easy job for the small shop owner to serve his customers better in future by learning from past interactions. But, this may not be easy for Web-enabled e-businesses when most customers may never interact personally, and the number of customers is much higher than of the small shop owner.

Data mining techniques can be applied to understand and analyse e-business data, and turn into actionable information, that can support a Web enabled e-business to improve its marketing, sales and customer support operations. This seems to be more appealing, when data is produced and stored with advance electronic data interchange methods, the computing power is affordable, the competitive pressure among businesses is strong, and the efficient and commercial data mining tools are available for data analysis.

BACKGROUND

Data mining is the process of searching the trends, clusters, valuable links and anomalies in the entire data. The process benefits from the availability of large amount of data with rich description. The rich descriptions of data such as wide customer records with many potentially useful fields allow data mining algorithms to search beyond obvious correlations. Examples of data mining in Web-enabled e-business applications are generation of user profiles, enabling customer relationship management, and targeting Web advertising based on user access patterns extracted from the Web data. With the use of data mining techniques, e-business companies can improve the sales and quality of the products by anticipating problems before they occur.

When dealing with Web-enabled e-business data, a data mining task is decomposed into many sub tasks (figure 1). The discovered knowledge is presented to user in an understandable and useable form. The analysis may reveal how a Web site is useful in making decision for a user, resulting in improving the Web site. The analysis may also lead into business strategies for acquiring new customers and retaining the existing ones.

DATA MINING OPPORTUNITIES

Data obtained from the Web-enabled e-business transactions can be categorised into (1) primary data that includes actual Web contents, and (2) secondary data that includes Web server access logs, proxy server logs, browser logs, registration data if any, user sessions and queries, cookies, etc (Cooley, 2003; Kosala & Blockeel, 2000).

The goal of *mining the primary Web data* is to effectively interpret the searched Web documents. Web search engines discover resources on the Web but have many problems such as (1) the abundance problem, where hundreds of irrelevant data are returned in response to a search query, (2) limited coverage problem, where only a few sites are searched for the query instead of searching the entire Web, (3) limited query interface, where user can only interact by providing few keywords, (4) limited customization to individual users, etc (Garofalakis, Rastogi, Seshadri, & Hyuseok, 1999). Mining of Web contents can assist e-businesses in improving the orga-

Figure 1. A mining process for Web-enabled e-business data
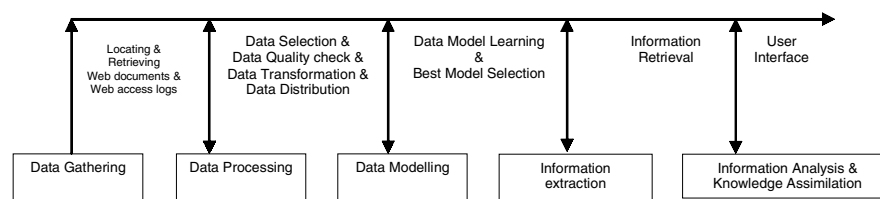
nization of retrieved result and increasing the precision of information retrieval. Some of the data mining applications appropriate for such type of data are:

- Trend prediction within the retrieved information to indicate future values. For example, an e-auction company provides information about items to auction, previous auction details, etc. Predictive modelling can analyse the existing information, and as a result estimate the values for auctioneer items or number of people participating in future auctions.
- Text clustering within the retrieved information. For example structured relations can be extracted from unstructured text collections by finding the structure of Web documents, and present a hierarchical structure to represent the relation among text data in Web documents (Wong & Fu, 2000).
- Monitoring a competitor's Web site to find unexpected information e.g. offering unexpected services and products. Because of the large number of competitor's Web sites and huge information in them, automatic discovery is required. For instance, association rule mining can discover frequent word combination in a page that will lead a company to learn about competitors (Liu, Ma, & Yu, 2001).
- Categorization of Web pages by discovering similarity and relationships among various Web sites using clustering or classification techniques. This will lead into effectively searching the Web for the requested Web documents within the categories rather than the entire Web. Cluster hierarchies of hypertext documents can be created by analysing semantic information embedded in link structures and document contents (Kosala & Blockeel, 2000). Documents can also be given classification codes according to keywords present in them.
- Providing a higher level of organization for semistructured or unstructured data available on the Web. Users do not scan the entire Web site to find the required information, instead they use Web query languages to search within the document or to obtain structural information about Web documents. A Web query language restructures extracted information from Web information sources that are heterogenous and semi-structured (Abiteboul, Buneman, & Suciu, 2000). An agent based approach involving artificial intelligent systems can also organize Web based information (Dignum & Cortes, 2001).

The goal of *mining the secondary Web data* is to capture the buying and traversing habits of customers in an e-business environment. Secondary Web data includes Web transaction data extracted from Web logs.

Some of the data mining applications appropriate for such type of data are:

- Promoting cross-marketing strategies across products. Data mining techniques can analyse logs of different sales indicating customer's buying patterns (Cooley, 2003). Classification and clustering of Web access log can help a company to target their marketing (advertising) strategies to a certain group of customers. For example, classification rule mining is able to discover that a certain age group of people from a certain locality are likely to buy a certain group of products. Web enabled e-business can also be benefited with link analysis for repeat buying recommendations. Schulz, Hahsler, & Jahn (1999) applied link analysis in traditional retail chains, and found that 70% cross-selling potential exists. Associative rule mining can find frequent products bought together. For example, association rule mining can discover rules such as "75% customers who place an order for product1 from the /company/product1/ page also place the order for product2 from the /company/product2/ page".
· Maintaining or restructuring Web sites to better serve the needs of customers. Data mining techniques can assist in Web navigation by discovering authority sites of a user's interest, and overview sites for those authority sites. For instance, association rule mining can discover correlation between documents in a Web site and thus estimate the probability of documents being requested together (Lan, Bressan, & Ooi, 1999). An example association rule resulting from the analysis of a travelling e-business company Web data is: "79% of visitors who browsed pages about *Hotel* also browsed pages on *visitor information: places to visit*". This rule can be used in redesigning the Web site by directly linking the authority and overview Web sites.
- Personalization of Web sites according to each individual's taste. Data mining techniques can assist in facilitating the development and execution of marketing strategies such as dynamically changing a particular Web site for a visitor (Mobasher, Cooley, & Srivastave, 1999). This is achieved by building a model representing correlation of Web pages and users. The goal is to find groups of users performing similar activities. The built model is capable of categorizing Web pages and users, and matching between and across Web pages and/or users (Mobasher, et al, 1999). According to the clusters of user profiles, recommendations can be made to a visitor on return visit or to new visitors (Spiliopoulou,

Related Content

A Methodology for Building XML Data Warehouses
Laura Irina Rusu, J. Wenny Rahayuand David Taniar (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 530-555).*
www.irma-international.org/chapter/methodology-building-xml-data-warehouses/7663

From User Requirements to Conceptual Design in Warehouse Design: A Survey
Matteo Golfarelli (2010). *Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction (pp. 1-16).*
www.irma-international.org/chapter/user-requirements-conceptual-design-warehouse/36605

A TOPSIS Data Mining Demonstration and Application to Credit Scoring
Desheng Wuand David L. Olson (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 1877-1887).*
www.irma-international.org/chapter/topsis-data-mining-demonstration-application/7738

On the Rim Between Business Processes and Software Systems
Maria Estrela Ferreira da Cruz, Ricardo J. Machadoand Maribel Yasmina Santos (2019). *New Perspectives on Information Systems Modeling and Design (pp. 170-196).*
www.irma-international.org/chapter/on-the-rim-between-business-processes-and-software-systems/216338

Comparative Genome Annotation Systems
Kwangmin Choiand Sun Kim (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 1784-1798).*
www.irma-international.org/chapter/comparative-genome-annotation-systems/7731