Chapter 7

Summary and Conclusions

The expansion of World Wide Web has enabled businesses operations to be executed fast and easier. E-commerce web sites easily reach to customers and due to good speed and low cost, many businesses in retail, education, communication, banking etc are setting up on web. Web Data Mining is the application of data mining techniques to extract useful knowledge from web data like contents of web documents, hyperlinks structure of documents and web usage logs. It is the area converge to Machine Learning, Information retrieval, Databases, Statistics and Natural Language Processing. Web data mining enables improving website design, customer behavior analysis, improving customer services and relationship, provide personalized services, business decision making in e-commerce. To discover the novel, potentially useful and interesting information, several pattern discovery techniques can be applied such as Path Analysis, Association Rules, Sequential Patterns, Clustering, and Classification etc. Web mining may involve personal and confidential data. Thus, privacy is the issue concern with individual and rising.

We focus on finding two types of traversal patterns the Maximal Forward Path and Maximal Backward Path using web mining. User traversal path patterns can help to improve web site structure and navigability of web site. For traversal pattern mining, the sequence of activities needs to be performed is: parsing and cleaning web server log, dividing the extracted and cleaned data into user sessions and finding traversal patterns. A maximal forward path is defined as the longest consecutive sequence of forward references before the first backward reference is made to visit some previously visited page in the same session. Thus, the last reference in a maximal forward sequence indicates a content page that is desired by the user. The Maximal Backward Path demonstrates groups of nodes in the backward sequence combination. This presents a good indication of how well the infrastructure of a site is constructed.
and arranged. The longer the combination of nodes MBP holds the less organized a site appears to be. The emphasis can also be given to the conditional restrictions which can be applied with specific patterns to find the more knowledge about users’ patterns. In e-commerce, traversal path patterns represent the navigation behavior of customers. The information about purchasing behavior of customer can be used to find association between purchasing items and this can help in improvement of cross-selling. Thus, considering both traversal pattern and purchasing behavior of customer can add value to association rule finding.

The technique of automatic advertising personalization has been presented which is based on combination of web content and web usage mining. It uses knowledge extracted from web site content and user online ad behavior for advertising personalization. In this technique, some important factors like most appropriate web site content, click through probability, advertising policy are considered. In e-commerce, personalized advertising provides advantages like higher click-through rates and purchase conversion. Also customer can be guided to relevant products and flawless personalized shopping experience can be achieved. Using web data mining, return on advertising investment can be measured by computing ad click through rates and purchase conversations. In automatic advertising personalization, for every user request, user visiting the web site is assigned to patterns discovered and all the information related to user behavior and advertising policy are integrated to provide appropriate personalized advertisement. For, particular user request, user is assigned to nearby publisher content group. In this case, the nearby content group indicates that current user is visiting which category like news, sports etc. and the selected usage pattern specifies the group of users with same type of behavior like sports, buyers etc. Related to each publisher content group, there is advertising content group and related to each usage pattern, there is ad visiting pattern. By processing different vectors, the list of appropriate advertisements is created for recommendation. Other benefits are also achieved like eliminating repetition of advertisements for particular user, tracking maximum number of advertisements per user, presenting advertisements already clicked by users visited similar pages etc. Advertising policies can be specified to filter list of advertisements that may include emission time, shape etc. and according
to advertising policy, periodically changing advertisements can be presented. Return on advertising investment can be measured by applying web data mining on computing ad click through rates and purchase conversations.

Web data mining in e-commerce can answer the important business questions such as recognizing profitable customers, identifying top and bottom selling products, identifying pages customers frequently buy from, conversion rate for particular banner ad, knowing products purchased by customers from a particular banner ad purchase, identifying banner ads carrying more traffic etc. Intelligent web mining techniques can be combined with traditional web mining approaches to improve the quality of mining. Intelligent web mining in e-commerce is useful for increasing the sale, fetching customers to site, and measuring the return on investment. In E-commerce Intelligence system, data is collected from various sources such as web servers, commerce servers and enterprise data system and integrated into system to enable business intelligence and analysis. The approach to build data warehouse enables storing data from variety of sources to support decision making. Integration of data warehouse, data mining, OLAP into e-commerce system reduces the attempt for web mining and provides effective business intelligence. The data needed to perform the analysis should include the sources such as: web server logs, product information, content information of the web site, customer purchase data, customer demographics information. Some of challenging issues in data collection task is sessionizing clickstream data, filtering crawlers sessions and collecting customer information. A data warehouse provides the data source for online analytical processing and data mining. Designing a proper data warehouse schema and populate the data from the OLTP system to the warehouse is very time consuming and complex. A well-designed data warehouse would feed business with the right information at the right time in order to make the right decisions in e-commerce system.

For web mining purposes, it needs to have a database schema for data warehouse designed to support decision-making and data analysis. The data in a data warehouse system are normally organized in multidimensional modeling with star schema using
fact tables plus the surrounding dimension tables. The requirement of clickstream data in the data warehouse makes the schema design more complex. The data warehouse is required to make the customer clickstream available for analysis. The warehouse is the source data of data mining and business intelligence it contains the fundamental business content of what a web store sells with web services and capabilities. A warehouse allows analyzing all hits on a web site, all the products sold in the web store from many viewpoints. It is necessary to include additional information such as order information from the web store, product information about the product, user browsing sequence from clickstream, and customer information from user table. Dimensions are qualifiers that give meaning to measures. Dimensions are stored in dimension tables made up of dimensional elements and attributes. Some of the important dimensions in the warehouse are: session, page, time, and user. For data mining and business intelligence, the data in the data warehouse need to be transformed or aggregated to different level of granularity such as session level, order-header level or customer level depending on the mining and analysis goals. The last step is to evaluate the mining results and then adopt actionable results. After the mining algorithms are applied for pattern discovery, many patterns may be identified but not all of them are interesting or actionable. Unlike most of the pattern evaluation approaches, which rely on an SQL statement to query the database and evaluate the results, the data cube is an important component in the mining procedure and one can dice and roll up the data cube to easily verify mining results. Based on actionable mining results, analyst can take action useful to improve the business.