Exploiting OLAP and Data Mining for augmenting E-business

Monika Mangla

Computer Department

Terna Engineering College, Nerul, Navi Mumbai, INDIA

manglamona@gmail.com

Abstract—**Mushrooming of e-business has been at a swift rate during past decade. In this scenario every day a new e-retailer is entering the world of e-business thus increasing competition in the area [1]. Any e-retailer would be able to sustain in this neck to neck competition if he is capable of attracting new customers along with retaining persistent customers. The data mining techniques, when implemented on the data available with the store is capable of understanding interesting patterns in the transactions made by customers [2] [3]. The paper presented discusses the usage of OLAP and data mining in detecting buying patterns from the data lying with the log files.**

Keywords-Data mining, Online Analytic Processing (OLAP), Association rules, e-business, transactions.

I. INTRODUCTION

The evolution of e-business comes with handling various challenges as retaining the persistent customer base along with attracting new customers. To achieve the said objective some policies should be evolved to attract the customers. This paper discusses the importance of maintaining records of visitors, buyers and transactions made by buyers [4]; techniques that can be utilized in doing so and consequent benefits of the same. In this paper, the methodology of Data mining is discussed [5]. Section 2 discusses the storage of web log files for mining purpose and OLAP. Section 3 discusses techniques for data mining in the web log files.

II. DATA WAREHOUSING AND OLAP

While new e-retailers are entering into the business they are necessitated to have data security and integrity to survive into this neck to neck competition. Mushrooming into the area demands for analyzing the data available that is maintained in data warehouses, a repository of an organization's electronically stored data that is designed to facilitate mining the data [6]. Apart from providing a common data model Data warehouses also facilitate decision support system applications such as frequent item sets (e.g., items with maximum sales), sequential patterns (e.g., if an item is purchased then another item is likely to be purchased in near future) and association rules (e.g., if a n item is purchased then probability for purchase of another related item is very high).

Web history log files can be maintained in form of multidimensional data where each axis represents a dimension of the information. Figure 1 represents data mining architecture consisting of OLAP server and data base server. However, accessing data alone is not sufficient rather data stored in multiple dimensions need to be mined to answer these queries. Thus OLAP and data mining are accepted to be integral parts of any decision support system reinforcing each other thus providing correct results.
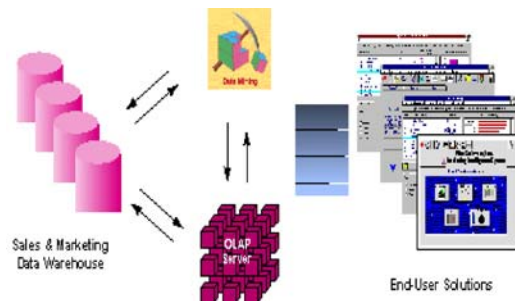

Figure 1.   Integrated Data Mining Architecture

For Data Mining some machine assistance is required as if done manually, users may not have enough time to analyze data in every dimension and consider all relevant scenario that may result in generating ambiguous or incorrect results. OLAP data mining systems are capable of mining the data along multiple dimensions, finding patterns along dimensions and considering all relevant scenarios.

III. DATA MINING

Data Mining is a key tool to help make sense of all of the data and better understand and facilitate electronic commerce, as it provides hidden information about customers buying patterns, browsing patterns, usage times and preferences. Web mining techniques can be applied to understand and analyze such data, and turning into information that can support a web enabled electronic business to improve its sales and customer support [7].

Information regarding browsing and buying patterns of existing customers can be used to generate customized offers [8]. Better revelation about navigation patterns of the web pages by using web mining techniques can lead to

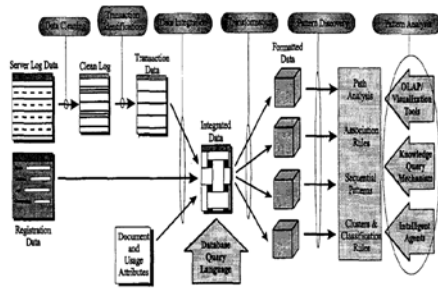customization for existing customers. Steps in Data Mining are

Figure 2.   Web Usage Mining

A. Mining of web usage

Web Usage Mining, commonly termed as Web Log Mining, is to discover patterns in the browsing and navigation of Web users. It discovers interesting patterns in Web Logs. Process of Web Usage Mining is depicted in Figure 2.

B. Frequent Item sets

Web Mining can help in finding Frequent Item Sets from the web log. This helps in finding items with fewer sales, which needs to be promoted to increase sale.

C. Pattern Detection from Web Transaction

Patterns can be associative in nature or sequential patterns. Associations refer to the pattern that if an item A is purchased then item B is likely to be purchased. Example for association is vehicle and petrol. Sequential patterns are the patterns in which if an item is purchased then item B is likely to be purchased in near future. Contact lens and lens solution are examples of sequential patterns. All such patterns are to be generated from the data in web log history available with web log history. Such association rules help e-business in developing cross-selling and up-selling strategies.

D. Clustering of Customers and Products

Number of users visit e-store site on regular basis out of which some are mere visitors to the site while others are promising customers of the e-store. Client browsers hold some of the very important information in the history files, cookie files that can further facilitate to analyze and classify data. These customers can be classified as persistent customers and just visitors to the e-store. Clustering techniques can be used to classify the customers among different clusters. Further persistent customers can be classified on the basis of web server which they access most of the times, which is going to help identify customers of particular products or brand [9] [10]. The process is going to help in finding the product that gives maximum returns to the e-retailer along with finding product having least sale. Some measures need to be taken to promote sale of such products either by giving luring schemes or marketing the product. Sometimes even relocation of such products on the shopping site or highlighting the product may help to boom the sale.

Some corrective measures should also be taken to convert mere visitors into persistent customers of the store. Persistent customers can be retained by giving them some extra benefits; where benefits can be monetary or even non-monetary.

E. Applications of analyzed data to e- business

Classification of customers into infrequent and persistent customers helps to customize offers to retain persistent customers while measures are also taken to convert infrequent customers into persistent customers. Understanding the navigational patterns of the customers on the web store may be used to customize web site so that associated items are put on same web page. Mining the data may also help in forecasting sales in the future to estimate the stock.

IV.CONCLUSION

It has been extensively shown in the paper that Web mining can serve as the backbone of any e-business as it can be appropriately utilized for the mushrooming of e-business. To survive this neck to neck competition in e-retailing; it has become demand of time to take correct action at very right time. Manual analysis of data may result in producing some incorrect results therefore discovering and extracting information from Web documents services need to be automated. With the inherit nature of Internet, e-businesses are bound to tackle massive, random and sometimes incomplete data, so integration of e-business with web mining is inevitable and this combination is bound to progress.

V. REFERENCES

[1] R. Armstrong, D. Fkeitag, T. Joachims, and T. Mitchell. Webwatcher: A learning apprentice for the world wide web. In Proc. AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments. 1995.

[2] Han, J., Kamber, M. (2000). Data mining: Concepts and Techniques. New York: Morgan-Kaufman.

[3] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In Proc. of the 20th VLDB Conference,pages 487-499, Santiago, Chile, 1994. Berry, M., J., A., & Linoff, G., S., (2000). Mastering data mining. New York: Wiley.

[4] Edelstein, H., A. (1999). Introduction to data mining and knowledge discovery (3rd ed). Potomac, MD: Two Crows Corp.

[5] Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). Advances in knowledge discovery & data mining. Cambridge, MA: MIT Press.

[6] Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). The elements of statistical learning : Data mining, inference, and prediction. New York: Springer.

[7] C. Wan, M. Liu, and L. Wang, "Content-based sound retrieval for web application," inWeb Intelligence: Research and Development,N. Zhong, Y. Yao, J. Liu, and S. Oshuga, Eds, Singapore: Springer-Verlag, 2001, vol. LNCS 2198, pp. 389–393.

[8] C.-H. Lee and H.-C. Yang, "Developing an adaptive search engine for e-commerce using a web mining approach," in Proc. Int. Conf. Inform.Technol.: Coding and Computing, 2001, pp. 604–608.

[9] A. Z. Broder, S. C. Glassman, M. S. Manasse, and G Zweig. Syntactic clustering of the web. In Proc. Of 6th International World Wide Web Conference, 1997.

[10] K. Yanai, M. Shindo, and K. Noshita, "A fast image-gathering system on the world wide web using a PC cluster," in Web Intelligence: Research and Development, N. Zhong, Y. Yao, J. Liu, and S. Oshuga, Eds, Singapore: Springer-Verlag, 2001, vol. LNCS 2198, pp. 324–334.