A Study of Knowledge Discovery Requisites for Business Intelligence

Pruthvi R[1], Panduranga Rao M.V[2]

[1]Department of Computer Science and Engineering
B.T.L Institute of Technology, Bangalore, India
pruthvishantharam@gmail.com

[2] Department of Computer Science and Engineering
B.T.L Institute of Technology, Bangalore, India

*Abstract***— Business Intelligence (BI) refers to the set of tools and systems that transform raw data into meaningful information and then convert meaningful information into the form that is beneficial for business. In this paper, we discuss the concept of web usage mining to achieve the conversion of raw data into meaningful information. When the necessity of web usage becomes more, there will be increase in the number of users using it. If the web starts predicting the web site or the web page required while the user is using the web, the access time can be decreased. To address this issue, we discuss the concept of web prediction. We discuss various models used in the process of web prediction. Our studies found that the combination of all-Kth Association Rule Mining model and all-Kth Markov model along with two-tier prediction framework helps in achieving prediction accuracy. We aim at using the concept of prediction accuracy in the big business organizations and discuss how it can be beneficial for efficient business intelligence.**

Keywords **— Business Intelligence, Web usage mining, Web prediction, all-Kth ARM model, all-Kth Markov model, Two-tier framework.**

I. Introduction

BUSINESS INTELLIGENCE (BI) represents the tools and systems that take part in the key role of strategic planning process in the company. Companies must acquire knowledge of their competitors to define business strategies and establish their business network in an efficient manner [1]. This required knowledge can be gained with the help of Web. Sources of data for data mining are provided by World Wide Web. Discovering and analyzing the usage patterns from web logs can be achieved by one of the data mining areas i.e. web usage mining. The problems associated with web data collection can be overcome by the three phases of web usage mining namely- Pre-processing, Pattern discovery and Pattern analysis. Once the data is as per user's requirement, user can use the data for specific needs. Once the user finds the advantage of web usage and once he is comfortable with web usage, the user starts using the web more. As a result of this, there is need to help the user to navigate to the desired websites efficiently. Web prediction helps to lessen the access time by predicting the next set of web pages. It does so, by acquiring the knowledge of users' previous website visits. In this paper, with respect to web prediction, we discuss the combination of all-Kth Association Rule Mining model and all-Kth Markov model along with two-tier prediction framework.

II. Literature Survey

In this section, we discuss business intelligence history, web usage mining, predictive caching, web prediction and the necessity of web prediction in the field of business intelligence

A. History of the emergence of Business Intelligence

Information provide answers and the answers help the people to make decisions. Business people get answers from business data. Business data includes the information about the people i.e. the customers, the products manufactured in the business organization and the places i.e. where actually the purchase and sales are happening for the business organization. Business people have the questions regarding the best products to be manufactured, loyal customers and proper places for purchase and sales transactions with the organization.

Initially, the data was stored manually, then in computers, then in storage devices and later in the databases. The databases provide a way to store the business data about the people, products and places. Storing data in the database requires expertise. Business applications were created to enter business data. Business applications provided better way to collect data. Thus, data collection did not have any problem but the data access was not easy since they come from multiple locations. Therefore, the concept of data warehouse was introduced. The data was moved to data warehouse, as a result of which the data coming from multiple locations were managed and accessed efficiently. Then the business intelligence concept emerged.

Business Intelligence 1.0 tools were capable of generating and analyzing the reports. Business Intelligence vendors promised more access across multiple locations to report and

© 2013 IJAIR. ALL RIGHTS RESERVED

analyze data. As a result, demand for data increased. Next, the business wanted data faster. The World Wide Web helped fetching data faster and things moved quickly. Since business needed more answers fast, lots of websites and business applications were created. Business Intelligence tools were everywhere and they provided more access. Now, the challenge of data management was, cost. Business Intelligence vendors needed a way to offer more functionality at low cost. Business Intelligence platforms emerged.

Business Intelligence 2.0 came out with more Business Intelligence tools and more Business Intelligence functionalities. There was intuitive access and Business Intelligence was about usability i.e. to turn data to right format. The concept of turning data to right format is discussed in the next sub-section. There was unstructured data and it came from multiple locations. 80 percent of business is conducted on unstructured information. Since the Business Intelligence should be people centric, unstructured data must be transformed to right format.

In 90's the business needed to do more to collaborate, search and communicate to drive innovation. Business Intelligence can work the way you do, to provide better insight, better decision for more people and to drive more innovation.

B. *Web Usage Mining*

Web usage mining deals with discovering and analyzing usage patterns from web logs and aim at improving the web based applications. The data collection in web usage mining can be made from server logs, browser logs, etc. From the server's perspective, mining uncovers information about sites where server resides. This is used to improve design of sites. From the client's perspective i.e. from client's sequence of clicks, information about users or groups of users is detected. This is used to perform pre-fetching and caching of pages. Web data is huge and unstructured. Therefore, it has to be pre-processed and parsed before extracting required information. This can be achieved with the help of three phases of web usage mining namely- Pre-processing, Pattern discovery and Pattern analysis.

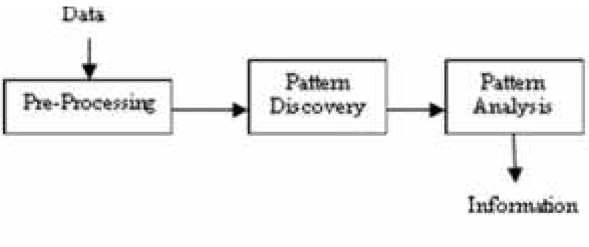

Fig. 1 Three phases of web usage mining

1) Pre-processing [2]: This phase helps to transform the data into the format that will be more easily and effectively processed for the purpose of the user.

2) Pattern discovery [2]: This phase is carried out on samples of data. The patterns are discovered using various pattern discovery methods such as association rules, clustering, classification, etc.

3) *Pattern analysis* [2]: This phase is done to filter the uninteresting information and to extract useful information.

Web usage mining is a part of Business Intelligence. It is crucial for Customer Relationship Management (CRM) in order to ensure customer satisfaction.

C. *Predictive Caching*

Caching is the well-known approach for improving the performance of web-based systems. Mining web-logs for caching web objects enhance the performance of web-caching systems. Web-caching reduces the network traffic, decreases access latency and lowers the server load. Powerful caching techniques are obtained when the web-caching concept is combined with data mining. Predictive caching [3] helps to improve the overall performance of the system. In this paper, predictive caching involves the discussion on cache replacement policies.

Cache replacement algorithm uses the past information to predict the future. The method of ranking the objects in the cache, based on their chances of being reused in future, is called as zero-order prediction. For example, if a web page A has been used 100 times then according to zero-order prediction, it can be concluded that A might be accessed in next 5 minutes also. Sequential access rules results in the method called first-order prediction. For example, if A ⟶ B, then according to the first-order prediction, it can be concluded that B will also be accessed in future, even though B has not been accessed more in the past.

Absolute count of web pages and predicted accesses according to association rules, enhance the priority of cached web pages that are not accessed frequently but will be in the future. It also provides objects that may be popular in future, even though they are not yet popular in the past.

D. *Web Prediction*

Web prediction helps to lessen the browsing access time by predicting the next set of web pages. It does so, by acquiring the knowledge of users' previous website visits. In this paper, with respect to web prediction, we discuss all-Kth Association Rule Mining model, all-Kth Markov model and two-tier prediction framework.

1) Association Rule Mining [4]: ARM is a technique that has been applied successfully to discover related transactions. ARM focuses on association of frequent item sets. For example, in a store, ARM helps to find out the items purchased together which can be utilized for shelving and ordering processes. In web prediction, the prediction is conducted according to the association rules that satisfy certain support and confidence. ARM addresses efficiency and scalability problems.

© 2013 IJAIR. ALL RIGHTS RESERVED

2) All-K^{th} Markov model [4]: Markov model is the model that predicts the next step based on current step. All-*K*th Markov model is the model in which all orders of Markov models are generated and utilized collectively in prediction. For example, when a user session $x = <P1, P5, P6>$ is given, prediction of all-*K*th model is performed by consulting third-order Markov model. If the prediction using third-order Markov model fails, then the second-order Markov model is consulted. This process repeats until the first-order Markov model is reached. Therefore, the all-*K*th-order Markov model achieves better prediction.

For all-*K*th-order Markov model, the running time of building it is linear because building each order of Markov model takes linear time with the training set size.

3) Two-Tier Prediction Framework [4]: The Two-Tier Prediction Framework is a novel framework for Web navigation prediction. Here, a unique prediction model, namely, *EC*, will be generated and later consulted to assign examples to the most appropriate classifier. In this framework, all classifiers are trained on training set, each example is mapped to one or more classifiers and then, each example is mapped to one classifier. The two-tier framework improves the prediction time without compromising prediction accuracy.

Combining more than one model improves prediction accuracy. Therefore, it can be concluded that the combination of all-K^{th} Association Rule Mining model and all-K^{th} Markov model along with two-tier prediction framework helps in achieving greater prediction accuracy.

E. Web Prediction in the Field of Business Intelligence

In business organization, most users can be classified into –

1) *Viewers*: Executives and Managers.

2) *Casual users*: Managers and Supervisors.

3) *Functional users*: Managers, Supervisors and Analysts.

4) *Super users*: Analysts.

Getting the right Business Intelligence tools into the right hands require IT to know the user population and the different needs to be served. *Executives and managers* want BI dashboard or scorecard. Some managers prefer performance indicators. *Business Intelligence analysts* need metrics, report and analysis. They also need time to understand the process within the given role. They spend time with users and understand key drivers of their roles. IT report developers require Microsoft reporting services and the crystal reports. Some users want Business Intelligence to be delivered in PDF files and data in excel spreadsheets. Some other users want the freedom to keep updated on what's going on. The information must be relevant and modified based on the person accessing it.

Therefore, based on the accesses being made, if web predicts the next set of web pages or web sites required by the user, the access time will be decreased. Instead of thinking upon the navigation of the web pages or web sites, the business user can search for the required information quickly. The prediction models and framework discussed in this paper, can be used to achieve the prediction accuracy thus facilitating easy navigation to the required web sites. This helps to ensure efficient business intelligence.

III. CONCLUSIONS

In this paper, we have looked into the concept of web usage mining to get the required data in the right format. The concept of predictive caching is discussed by which we have come to know that the power of web-caching is high in predicting the user's next actions in selecting the web site or web page. The all-K^{th} Association Rule Mining model, all-K^{th} Markov model and two-tier prediction framework that are discussed under web prediction, have revealed that all three of them put together helps in achieving greater prediction accuracy.

Based on the history of business intelligence emergence, we had come to know that there is the requirement for converting raw data into meaningful information. As we have already discussed, this issue can be addressed by the concept of web usage mining. Once the required information is available, users expect the access to the information at faster rate. At this point, the concept of web prediction can be applied and best models of web prediction can be used to achieve better prediction accuracy, thereby ensuring efficient business intelligence.

.

REFERENCES

[1] Xin Chen and Yi-fang Brook Wu "Web Mining for Business Intelligence: Discovering Novel Association Rules from Competitors' Websites", *IRMA International Conference, pp. 679-682, 2005*.

[2] Yogish H K, Dr. G T Raju and Manjunath T N, "The Descriptive Study of Knowledge Discovery from Web Usage Mining", *IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, pp. 225-230, September 2011*.

[3] Qiang Yang and Haining Henry Zhang, "Web-Log Mining for Predictive Web Caching", *IEEE Transactions on Knowledge and Data Engineering, Vol. 15, No. 4, pp.1050-1053, July/August 2003.*

[4] Mamoun A. Awad and Issa Khalil, "Prediction of User's Web-Browsing Behavior: Application of Markov Model", *IEEE Transactions on Systems, Man, And Cybernetics—Part B: Cybernetics, Vol. 42, No. 4, pp.1131 -1142, August 2012.*

© 2013 IJAIR. ALL RIGHTS RESERVED