Analysis of Web Mining and Evolving of User Profiles

M. Venkata Krishna, L. Raghavendar Raju, D.Jamuna, M. Gayathri

Abstract— A good and effective Customer Relationship Management (CRM) needs clear understandings of customer requirements. The management should have up to date needs of the customers, and accordingly they should act. Gathering information about the web user (customer) profiles and analysis of the profiles is not so easy for the large data. For example, Nokia site is developed to search exact cell phone models based on the customer interests. If they analyze all the user profiles they can know the most wanted cell phone model. Analyzing web user profiles and making decisions and predictions are possible with data mining techniques. We aim to analyze and track the web user profiles.

Index Terms: **Customer Relation Manager, Web mining, Preprocessing.**

I. INTRODUCTION

Customer Relationship Management (CRM) can use data from within and outside an organization to allow an understanding of its customers on an individual basis or on a group basis such as by forming customer profiles. An approach is presented for discovering and tracking evolving user profile using web mining techniques [15]. An improved understanding of the customer's habits, needs, and interests can allow the business to profit by, for instance, "cross selling" or selling items related to the ones that the customer wants to purchase [2]. Hence, reliable knowledge about the customers' preferences and needs forms the basis for effective CRM. As businesses move online, the competition between businesses to keep the loyalty of their old customers and to lure new customers is even more important. The fast pace and large amounts of data available in these online settings have recently made it imperative to use automated data mining or knowledge discovery techniques to discover Web user profiles. Web usage mining can use various data mining or machine learning techniques to model and understand Web user activity. In [4] clustering was used to segment user sessions into clusters or profiles that can later form the basis for personalization. In [3], notion of an adaptive Web site was proposed, where the user's access pattern can be used to automatically synthesize index pages. The work in [1] is based on using association rule discovery as the basis for modeling Web user activity; whereas the approach proposed in [5] are used probabilistic grammars to model Web navigation patterns for the purpose of prediction.

II. EXISTING SYSTEM

A good and effective Customer Relationship Management (CRM) needs clear understandings of customer requirements. The management should have up to date needs of the customers [14], and accordingly they should act. Gathering information about the web user (customer) profiles and analysis of the profiles is not so easy for the large data. For example, Nokia site is developed to search exact cell phone models based on the customer interests [15]. If they analyze all the user profiles they can know the most wanted cell phone model. Analyzing web user profiles and making decisions and predictions are difficult. The accessed data may not be stored at the Server side [1]. So many Web Sites provides web user interface to their Customers/Users, but they may not collect all the accessed information, because they feel that is not so important.

III. PROPOSED SYSTEM

The Proposed System provides collecting and mining of the collected data, that which can improve the company's growth and mining techniques also provides the quality of the User-Interface. The Proposed Model constitutes the following activities.

- Collecting all the Web accessed data on a web server
- Preparing the collected information as a data set. [Pre pruning Process].
- Creating per-user profiles
- Creating decisions based on the user profiles.
- Showing the decisions to the company in the Website.
- Collection of Web data such as activities/Click Streams recorded in Web Server logs
- Preprocessing of Web data such as filtering request and identifying unique regions
- Interpretation/Evaluation of the discovered profiles.

IV. IMPLEMENTATION

The user enters into a website and searches [11] for the required features by submitting a request to the web server. Like way, distinct users send different requests to acquire the information about their interested features [7]. The web server collects all the requests from distinct users and generates a access log file. This access *log file* is considered as input to the proposed system. Access log file undergoes different [13] web mining techniques such as preprocessing, pattern discovery, clustering and finally new requests will be generated. These new requests represent [6] new features which will be incorporated in the website that increases the commercial view of the concerned site.
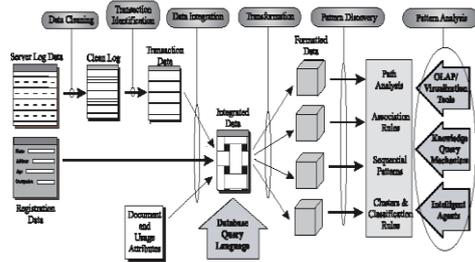
Fig1: General Architecture for Web Usage Mining

a) Preprocessing

It formats and extracts the useful data from access logs [10] received from web server.

b) Distinct Users

After preprocessing the data, pattern discovery is applied to it in order to recognize distinct users those who sent the requests [9].

c) Distinct Requests

After identifying Distinct Users, their distinct requests are recognized.

d) Clustering

Clustering is performed to group the similar features from distinct requests [12].

e) Generating New Request

Similar features are used to generate new request [15].

g) Access

The new request is incorporated as new feature which can be accessed by customer in [8] the web site.

V. RESULTS

The implementation of my application is shown in a stand alone system (OS : Windows XP) where VMware is used to create virtual environment with two systems. The two systems are Linux installed, created with the help of VMware.The stand alone system and other two Linux installed systems are differentiated by their IP addresses. System 1(Linux installed) acts as web server which generates access log file. System 2 (Linux installed) is used to send requests by distinct users, acts as a client. Various web mining techniques are applied to generate interesting patterns. The Results can be viewed in stand alone system (system 3) by retrieving the data form Post Gre SQL 8.3 database by using queries.

VI.CONCLUSION

This application has attempted to provide an up-to-date survey of the rapidly growing area of Web Usage mining. With the growth of Web-based applications, specifically electronic commerce, there is significant interest in analyzing Web usage data to better understand Web usage, and apply the knowledge to better serve users. This has led to a number of commercial offerings for doing such analysis. However, Web Usage mining raises some hard scientific questions that must be answered before robust tools can be developed. For Web usage mining, the session dissimilarity measure is not a distance metric, and dealing with relational data is impractical given the huge size of the data sets. Therefore, evolutionary techniques which can deal with ill-defined features and non-differentiable similarity measures are suitable. Evolutionary techniques can handle a vast array of subjective, even non-metric dissimilarities, making them suitable for many applications in data and Web mining. Moreover, they are meaningful only within well defined distinct profiles/contexts (context-sensitive) as opposed to all or none of the data (context-blind). Today's web sites are a source of an exploding amount of click stream data that can put the scalability of any data mining technique into question.

Moreover, the Web access patterns on a web site are very dynamic, due not only to the dynamics of Web site content and structure, but also to changes in the user's interests, and thus their navigation patterns. The access patterns can be observed to change depending on the time of day, day of week, and according to seasonal patterns or other events in the world.

ACKNOWLEDGEMENT

REFERENCES

[1] R. Cooley, B. Mobster, and J. Srivastava, "Web Mining: Information and Pattern Discovery on the World Wide Web,"Proc. Ninth IEEE Int'l Conf. Tools with AI (ICTAI '97), pp. 558-567, 1997.

[2] O. Nasraoui, R. Krishnapuram, and A. Joshi, "Mining Web AccessLogs Using a Relational Clustering Algorithm Based on a Robust Estimator," Proc. Eighth Int'l World Wide Web Conf. (WWW '99),pp. 40-41, 1999.

[3] O. Nasraoui, R. Krishnapuram, H. Frigui, and A. Joshi, "Extracting Web User Profiles Using Relational Competitive FuzzyClustering," Int'l J. Artificial Intelligence Tools, vol. 9, no. 4,pp. 509-526, 2000.

[4] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, "WebUsage Mining: Discovery and Applications of Usage Patternsfrom Web Data," SIGKDD Explorations, vol. 1, no. 2, pp. 1-12, Jan.2000.

[5] M. Spiliopoulou and L.C. Faulstich, "WUM: A Web UtilizationMiner," Proc. First Int'l Workshop Web and Databases (WebDB '98),1998.

[6] T. Yan, M. Jacobsen, H. Garcia-Molina, and U. Dayal, "From userAccess Patterns to Dynamic Hypertext Linking," Proc. Fifth Int'lWorld Wide Web Conf. (WWW '96), 1996.

[7] M. Perkowitz and O. Etzioni, "Adaptive Web Sites: Automatically Learning for User Access Pattern," Proc. Sixth Int'l WWW Conf. (WWW '97), 1997.

[8] J. Borges and M. Levene, "Data Mining of User Navigation Patterns," Web Usage Analysis and User Profiling, LNCS, H.A. Abbass, R.A. Sarker, and C.S. Newton, eds. pp. 92-111, Springer-Verlag, 1999.

[9] O. Zaiane, M. Xin, and J. Han, "Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs," Proc. Advances in Digital Libraries (ADL '98), pp. 19-29, 1998.

[10] O. Nasraoui and R. Krishnapuram, "A New Evolutionary Approach to Web Usage and Context Sensitive Associations Mining," Int'l J. Computational Intelligence and Applications, special issue on Internet intelligent systems, vol. 2, no. 3, pp. 339-348, Sept. 2002.

[11] O. Nasraoui, C. Cardona, C. Rojas, and F. Gonzalez, "Mining Evolving User Profiles in Noisy Web Clickstream Data with a Scalable Immune System Clustering Algorithm," Proc. Workshop Web Mining as a Premise to Effective and Intelligent Web Applications (WebKDD '03), pp. 71-81, Aug. 2003.

[12] P. Desikan and J. Srivastava, "Mining Temporally Evolving Graphs," Proc. Workshop Web Mining and Web Usage Analysis (WebKDD' 04), 2004.

[13] O. Nasraoui, C. Rojas, and C. Cardona, "A Framework for Mining Evolving Trends in Web Data Streams Using Dynamic Learning and Retrospective Validation," Computer Networks, special issue on Web dynamics, vol. 50, no. 14, Oct. 2006.

[14] M.A. Maloof and R.S. Michalski, "Learning Evolving Concepts Using Partial Memory Approach," Working Notes AAAI Fall Symp. Active Learning 1995, pp. 70-73, 1995.

[15] Olfa Nasraoui, Member, IEEE, Maha Soliman, Member, IEEE, Esin Saka, member, IEEE, Antonio Badia, Member, IEEE, and Richard Germain "A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites "Vol. 20,no. 2, February 2008.

AUTHOR BIOGRAPHY

M. Venkata Krishna Reddy, Working as Assoc. Professor in CSE Dept. Jayaprakash Narayan College of Engineering, Mahabubnagar, M.Tech(CSE) from Vidya Vikas Institute of Technology, Hyderabad. B.Tech(CSE) from Sri Kottam Tulasi Reddy Memorial College of Engineering, Gadwal. His areas of Interest are in Mobile Adhoc Networks, Data Mining, Networking and guided M. Tech and B. Tech Students IEEE Projects.

L. Raghavendar Raju Working as Assoc. Professor in CSE Dept. Jayaprakash Narayan College of Engineering, Mahabubnagar, M.Tech(CSE) from Sree Datta Institute of Engineering, Hyderabad B.Tech(CSE) from Jayaprakash Narayan College of Engineering, Mahabubnagar His areas of Interest are in Networking, Data Mining, Wireless Sensor Networks and guided M. Tech and B. Tech Students IEEE Projects.

Prof.D.Jamuna, Working as Professor & Head of CSE Dept. Jayaprakash Narayan College of Engineering, Mahabubnagar, M.Tech(SE) from School of Information Technology, JNTUH, Hyderabad. BE(CSE) from Vijayanagara Engineering College, Bellary. Experience 15 Years in Teaching Profession. Her areas of Interest are in Wireless Sensor Networks, Data Mining, Networking and guided M. Tech and B. Tech Students IEEE Projects. She is a Member of CSI.

M. Gayathri, Working as Asst. Professor in CSE Dept. Jayaprakash Narayan Group of Educational Institutions, Mahabubnagar, M.Tech(CSE) from Vidya Vikas Institute of Technology, Hyderabad. B.Tech(ECE) from Sri Kottam Tulasi Reddy Memorial College of Engineering, Gadwal. Her Areas of Interest are in Computer Networks, Data Mining.