Performance Evaluation of Clustering in Web-Log Analysis Based on Agent

Himani[1], Gaurav Arora[2], Sunita Sangwan[3]

[1]M.Tech. Scholar, S.B.M.N College of Engg., Rohtak (India)
Himanivirmani3@gmail.com

[2]Associate Professor, S.B.M.N College of Engg., Rohtak (India)

[3]Associate Professor, S.B.M.N College of Engg., Rohtak (India)

Abstract

Web mining is the use of data mining Technique to automatically discover & extract information from web documents. When user searches for goods the management agent receives order from graphical user interface .Management agent receives information, update agent information store house and feedback the mining result to user. Intelligent agent can help making computer system easier to use, enable finding & filtering information. The mining agent is the analytical center of whole agent system.It mainly adopts two kind of analytical method: related rule mining and cluster analysis. Cluster of objects are formed so that objects with in a cluster have high similarity. The aim of this paper is to analyze the web log data .To achieve this clustering tool is used. It performs in two phases. First it captures the web-log data. Then it analyzes the data& discovers the hidden pattern. Agent requires an agent communication language to describe & process agent request. The future internet will use PERL to encode information with meaningful structure & semantics.

Keywords: Web mining, Clustering, Diagnostic and Verification tools.

I. INTRODUCTION

The widespread usage of internet under World Wide Web platform in business practice, websites brings in a huge mound of data to organizations. Web Mining based on Agents has become a highly focused issue. Web Mining has matured as a field of basic and applied research in computer science in general and websites in particular. Websites has changed the face of most business functions in competitive enterprises.

1.1 WEB MINING

Web Mining is the use of data mining techniques to automatically discover and extract information from web documents and services [1]. Web mining technology has enabled e-commerce to do personalized marketing, which eventually results in

higher trade volumes. The companies can establish better customer relationship by giving them exactly what they need. Companies can understand the needs of the customer better and they can react to customer needs faster.

Web Mining refers to the overall process of discovering potentially useful and previously unknown information or knowledge from the web data. It implicitly covers the standard process of knowledge discovery in database (KDD) [3]. We could simply view web mining as an extension of KDD that is applied on the web data. From the KDD point of view, the information and knowledge terms are interchangeable [4].

2. LITERATURE REVIEW

2.1 The Work Process of System

When users search for goods on the e-commerce website, the Management Agent will carry on pretreatment for data and establish user Agent, then mine the data and give information feedback to users finally, offer the individualized service.

The general steps are:

- Management Agent (MA) receives order submitted from graphical user interface (GUI).
- Management Agent (MA) looks over related information in the Agent information storehouse activates or creates user agent and transfers task to User agent (UA).
- User Agent (UA) transfers information to Mining Agent (MiA), then Mining Agent carries on data mining for user's task with

IJCSMS International Journal of Computer Science & Management Studies, Special Issue of Vol. 12, June 2012
ISSN (Online): 2231 –5268
www.ijcsms.com

105

various kinds of mining methods and provides feedback foe User Agent.

- User Agent communicates with other User Agent according to provided systematic serial number and transfer information to Management Agent.

- User Agent assembles mining information to shape rule, update user model and returns to Management Agent.

- Management Agent receives information, updates Agent Information Storehouse (AIS) and feedbacks the mining result to user [45].

Intelligent Agents can help making computer systems easier to use, enable finding and filtering information, customizing views of information and automation work. An Intelligent Agent is software that assists people and acts on their behalf. It works by allowing people to delegate work that they could have done to the agent software [11].

If the Data Mining system produces large number of meaningful information by using a specialized Data Mining algorithm (association, clustering, decision trees etc.), it will take more time for the end-users to choose the appropriate knowledge for the problem discussed. A solution for this problem could be an intelligent system based on Agents. Data Mining and Intelligent Agents can make a common front to help people   in the decision making process, to elaborate decision models and take decision in   real time [20].

2.2 CLUSTERING

Clustering analyzes data objects without consulting a known class label. In general, the class labels are not presented in the training data simply because they are not known how to begin with. Clustering can be used to generate such labels. The objects are clustered or grouped based on the principle of maximizing the intraclass similarity and minimizing the interclass similarity. That is, clusters of objects are formed so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters. Each cluster that is formed can be viewed as a class of objects, from which rules can be derived [33]. E-commerce website can predict the customer's purchase behavior according to other customers' records or past purchase behaviors of this customer, and simulate sales-person to help the customer to finish buying course.

3. RESEARCH METHODOLOGY

3.1 TOOL AND METHOD USED

- In websites, Logs Cluster Analysis is one of the mining methods.

- The cluster objects include user groups and web pages. The cluster of user groups plays an important role in providing personalized services in websites. In order to cluster user groups, we must describe their browser action.

- Tool used for the web log file analysis:
Cluster Diagnostics and Verification Tool, Agents require a common Agent Communication Language (ACL) to describe and process agent's requests. The future internet will use PERL to encode information and services with meaningful structure and semantics that computers and people can readily understand. With PERL and Web-Mining model, it becomes easier to implement Agents.

3.2 THE CLUSTER DIAGNOSTICS AND VERIFICATION TOOL

The Cluster Diagnostics and Verification Tool (ClusDiag.exe) is a tool that performs basic verification and configuration analysis checks and creates log files to help system administrators identify configuration issues prior to deployment in a production environment. ClusDiag.exe can capture all relevant log files and event logs from each node of a server cluster and merge them into a single file for easy analysis and troubleshooting. Administrators can analyze these log files with built-in filtering, merging, and bookmarking functionality and generate various diagnostics reports.
Where to Download the Tool You can obtain ClusDiag.exe from the Microsoft Windows Server 2003 Resource Kit, or you can download it from the Cluster Diagnostics and Verification Tool (ClusDiag.exe) Web site.

You can use this dialog box to go to the folder that contains the cluster log file.
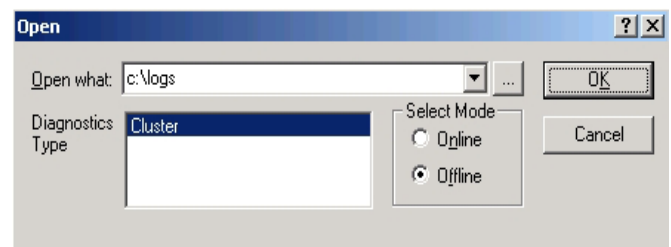
Adding a Bookmark

Cluster logs can be quite large, and one of the features that is very useful in ClusDiag.exe is the ability to place a bookmark in to the cluster log file. You create a bookmark by clicking the far left side of a line in the log file as shown in the following figure:
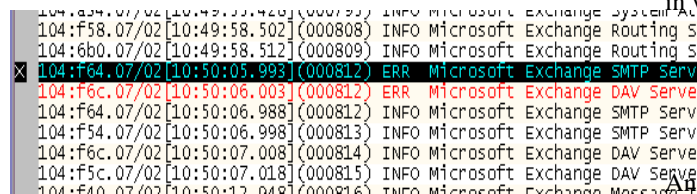

Figure : Adding a Bookmark

4. AGENT COMMUNICATION LANGUAGE

An Agent Communication Language (ACL) is a collection of speech-act-like message types, with agreed-upon semantics, which facilitate the knowledge and information exchange between software agents.

Agents require a common agent communication language (ACL) to describe and process agent's requests. The future internet will use PERL to encode information and services with meaningful structure and semantics that computers and people can readily understand. With PERL and Web-Mining model, it becomes easier to implement Agents.

4.1 WEB MINING WITH PERL

Why Perl? Why not?

Perl is an excellent tool for a web mining project. Perl's basic but powerful built-in data structures, easily accessible regular expressions and large selection of CPAN (Comprehensive Perl Archive - the repository of Perl module/libraries) modules show that Perl easily meets the application's requirements.

Perl programs are easy to learn and write. Perl has great text processing capabilities. As comparatively to other programming languages (such as C, C++), Perl is having fewer lines of code. Perl has become the most widely used option for custom CGI script. Perl is a scripting language which means it does not have to be compiled. Instead, an interpreter executes the Perl script, this makes it easy to write and test Perl scripts, because they do not have to go through the typical edit-compile-link cycles.

5. SCOPE OF DISSERTATION

Web Mining based on Agents may help and benefit societal areas (such as E-learning, E-government, Recommendation systems, Security, E-mining etc.) in ways of:

- extracting new knowledge
- providing support for decision making
- empowering valuable management of societal issues

An obvious societal benefit is that Web Mining research efforts lead to user (or group of users) satisfaction:

- by providing accurate and relevant information retrieval
- by providing customized information
- by learning about user's demands so that services can target specific groups or even individual users
- by providing personalized service

6. CONCLUSION

The Cluster Diagnostics and Verification tool capture log files as well as a analyze log files which found to be quite useful for the study of web log analysis.
In the process of Agent communication, Perl language is used as an Agent Communication Language (ACL). Perl as an ACL and Agent-based model for e-commerce website can effectively improve the intelligence of the system. But there are many related technologies that need to be further research.

7. FUTURE WORK

The future work is to make the web mining system not only efficient but also more intelligent. The future internet will use PERL to encode information and services with meaningful structure and semantics that computers and people can readily understand. With PERL and Web-Mining model, it becomes easier to implement Agents.

REFERENCES

[1] Agrawal & Srikant, "Popular algorithm for mining frequent itemsets is Apriori", in: Proceedings of 20th International Conference on Very Large Data Bases, September 12-15, Santiago de Chile, Chile, 1994.

[2] Alex G. Buchner, Maurice D. Mulveena, "Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining", in: ACM SIGMOD Record, Volume 27, Issue 4 (December 1998), Pages: 54 – 6, ISSN:0163-5808

[3] Ansari S., Kohavi R., Mason L., Zijian Zheng, "Integrating E-commerce and data mining: architecture and challenges", in: Proceedings of IEEE International Conference on Data Mining, 2001:27 – 34.

[4] Auguste D M, "Customer service in E-business", in: Proceedings of IEEE Internet Comput. 5(5): 90–91, 2001.

[5] Common Log File Format, Available: http://www.w3.org/Daemon/User/Config/Logging.html#common-logfile-format

[6] D.Konopnicki and O.Shmueli, "W3qs: A Query System for the World Wide Web", in: proceedings of the 21th VLDB Conference, pages 54-65, Zurich, 1995.

[7] Danish Khan, "CAKE – Classifying, Associating & Knowledge DiscovEry: An Approach for Distributed Data Mining (DDM) using PArallel Data Mining Agents (PADMAs)", in: Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008.

[8] Elhadi S, Sadd Abu-Draz, "Agent-mediated E-commerce System [A]", in: proceedings of the 19th International Conference on Advanced Information Networking and Applications, 2005.

[9] Extended Log File Format, Available: http://www.w3.org/TR/WD-logfile.html

[10] Gazis D. PASHAs, "Advanced intelligent agents in the service of electronic commerce", Derek Leebaert, The Future of the Electronic Marketplace[C], Cambridge : MIT Press ,1998, 145—177.

[11] Gilbert, D, "Intelligent agents: The Right Information at the Right Time", IBM corporation, Research Triangle park, NC USA.

[13] Guttman, R. H. and Maes, P., "Cooperative vs. Competitive Multi-Agent Negotiations in Retail Electronic Commerce", in Proceedings of Second Int. Workshop on Cooperative Information Agents II, Learning, Mobility

and Electronic Commerce for Information Discovery on the Internet, Jul. 1998, pp. 135-147.

[14] Guttman, R.H. and Maes, P., "Agent-mediated negotiation for retail electronic commerce", in: Agent-Mediated Electronic Commerce: First International Workshop on Agent Mediated Electronic Trading, (Noriega, P., and Sierra, C. ed.), Berlin: Springer 1999, 70-90.

[15] H.Baazaoui Zghal, S.Faiz, H.Ben Ghezala, "A Framework for Data Mining based Multi-Agent: An Application to Spatial Data", World Academy of Science, Engineering and Technology, 2005.

[16] Hollfelder S, Oria V, Ozsu M T 2000, "Mining user behavior for resource prediction in interactive electronic malls", in: Proceedings of IEEE Int. Conf. on Multimedia and Expo (New York: IEEE Comput. Soc.) pp 863–866.

[17] Hongwei Ke, Liming Cao, and Xiaoping Wang, "Study on Data Mining Based on Mobile Agent", Fujian Computer, Fujian, 2003, pp. 10-12, 20

[18] I.Khosla, B.Khun, and N.Soparkar, "Database Search using Information Mining", in: proceedings of 1996 ACM-SIGMOID International Conference on Management of Data, 1996.

[19] IIS Standard Log File Format, Available: http://www.iis.net/

[20] Irina Tudor, Liviu Ionita, "Intelligent Agents as Data Mining techniques used in Academic environment", Petroleum-Gas University of Ploiesti, Department of Informatics, Bd. Bucuresti, No. 39, 100680, Romania.

[21] J. Bertot, C.McClure, W.Moen, J.Rubin, "Web Usage Statistics: Measurement Issues and Analytical Techniques", Government Information Quarterly, 14(4), 375-395, 1997.