

Web Data Mining through Software Agents

Shakti Kundu

Dept. of Computer Science & Engineering
DIT University, Dehradun, India

M. L. Garg, PhD

Dept. of Computer Science & Engineering
DIT University, Dehradun, India

ABSTRACT

Web data mining research efforts lead to user or group of user's satisfaction by providing accurate and relevant information retrieval; by learning about user's demands so that services can target specific groups or even individual users and by providing personalized service. Due to the explosive growth of web content in the internet world in the past few years, there is tremendous requirement and scope of doing research in the category of e-commerce agents. Also user requirements such as getting accurate information in a prompt manner are not getting fulfilled as per their expectations. This has motivated the researcher to pursue his research in the field of web data mining through software agents. This paper has highlighted the implementations of software agents for e-commerce applications.

Keywords

Agents, Negotiation, Software, System, Web.

1. INTRODUCTION

A software agent is a computer system, situated in some environment that is capable of flexible autonomous action in order to meet its design objectives. It is well suited for developing e-commerce systems. The relationship between agent and e-commerce can be found easily from agent's characteristics. From the end-user perspective, an agent is a program that assists people and acts on their behalf [1].

Agents often need to interact in order to achieve their objectives or improve their performance. To negotiate automatically, agents must share a cooperative mechanism [2].

Software agents play the role of intelligent agents that perform some group of tasks autonomously or with some level of liberty. The tasks employed by the agents are description of user's goal or expectations. All the related activities are pursued on behalf of a user or another program [3]. Intelligent agents may be defined as computer program that helps a user with routine computer tasks. Other names of intelligent agent are Software Agents, Intelligent Software Robots and Soft bots.

Various software agents have been developed that search for interesting information using the domain characteristics and user profiles. Agents such as Harvest [4], FAQ-Finder [5], Information Manifold [6] and Parasite [7] depends either on pre-requisite information of domain or on coded design to extract and interpret documents. Other agent such as ShopBot [8] regains product knowledge from range of vendor sites using only overall information about the scope of the product.

Information categorization is another agent based approach. Several software agents via information extraction techniques and hierarchical clustering techniques provide filter and categorical information [9, 10, 11].

An important advantage of e-commerce data mining or web data mining is the possibility of conducting controlled

experiments and measuring their effects on business goals. Direct marketing, for example, can target particular groups, and the results can be data mined for predictive behavior modeling of consumers [12].

Though the existing data mining systems have advantages of their own, yet they are not so perfect enough concerning the intelligence aspect, and still need to be perfected. The agents play high quality intelligence role in the field of web data mining. The combination of the agent and data mining systems can realize the whole intellectuality of mining process [13].

Multi-Agent System (MAS) is a system composed of many agents, and generally these agents exchange information through network infrastructure. In MAS, the ability of an individual agent is limited, but multi-agent system can finish a lot of complicated tasks through cooperation [14]. MAS can improve enterprises capability to mine customer's information effectively, which can basically helps in saving time and energy.

Negotiation can be defined as "mechanisms that allow a recursive interaction between a principal and a respondent in the resolution of a good deal" [15]. Automated negotiation is a key issue for e-commerce because it provides the de-facto means of interaction between stakeholders with different aims and objectives. To improve the efficiency of the software system, an application should be developed with latest tools and techniques along with standard parameters.

The remainder of the paper is organized as follows: In the subsequent Section 2, the Problem Description has been highlighted. In Section 3, Negotiation System for E-commerce has been discussed and Conclusions are given in Section 4.

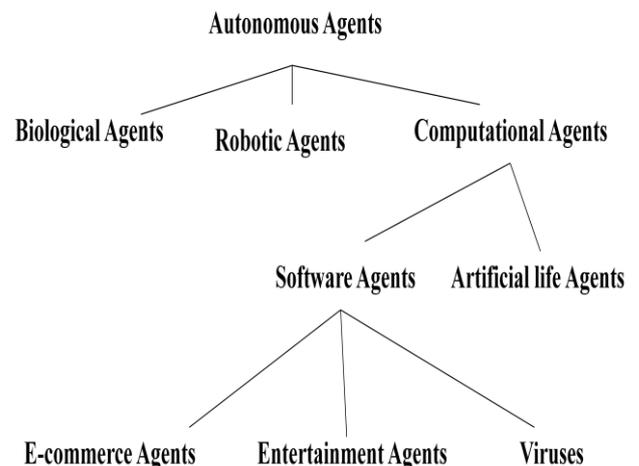


Figure 1: Taxonomy Tree to classify Autonomous Agents

2. PROBLEM DESCRIPTION

Many researchers have provided different techniques for web data mining, such as:

- R. Cooley et al. highlighted WEBMINER system [16] which works on SQL-like query mechanism for extracting the discovered knowledge. However, the expressiveness of this language is restricted by the input parameters acceptable by the miner (Current miners do not support generic specifications on the structure of patterns to be discovered e.g. page revisit, cycles etc.)
- Sandeep Neli et al. [17] proposed automated data mining using Perl script from web servers. The proposed search tool prepared on the platform of Perl / CGI script.
- Liren Chen et al. [18] proposed WebMate, a personal software agent for browsing and searching.
- Robert Armstrong et al. [19] presented the design and implementation of an agent called WebWatcher that is intended to assist users both by interactively advising them as they traverse web links in search of information, and by searching autonomously on their behalf.
- Henry Liberman [20, 21] introduced an agent Letizia that tracks the user's browsing behavior and tries to anticipate what items may be of interest to the user.
- Henry Liberman et al. [22] proposed reconnaissance agents with two examples: Letizia and Powerscout. The main difference is Letizia uses local reconnaissance (searching the neighborhood of current page), while Powerscout uses global reconnaissance (making

use of traditional search engine to search the Web in general).

- WebSIFT (Website Information Filter) [23] is designed to perform web usage mining from server logs. The preprocessing algorithms include identifying users, server sessions, and identifying cached page references through the use of the referrer field. It identifies interesting information and frequent item sets from mining usage data.
- Most of the web log analyzer tools are used for pattern discovery and pattern analysis. Using various web analyzer tools, it is possible to detect the number of accesses to the server and URL of users. Such tools provide knowledge about number of hits, files, page visited. However such tools are having certain criticism. For instance, Webalizer [24] which is one of the popular web log analyzer tool but it is having one major limitation. It cannot differentiate between the human user visit and robot visit.

3. NEGOTIATION SYSTEM FOR E-COMMERCE

WEBMINTEL system is a collection of various agents. Multi-Agent system (MAS) is a powerful paradigm in now a day's distributed systems. In multi-agent environments, agents often need to interact in order to achieve their objectives or improve their performance. If agents are to negotiate automatically with one another they must share a negotiation mechanism [25]. The involvement of various software agents collaboratively achieves the goal of negotiation for e-commerce activities.

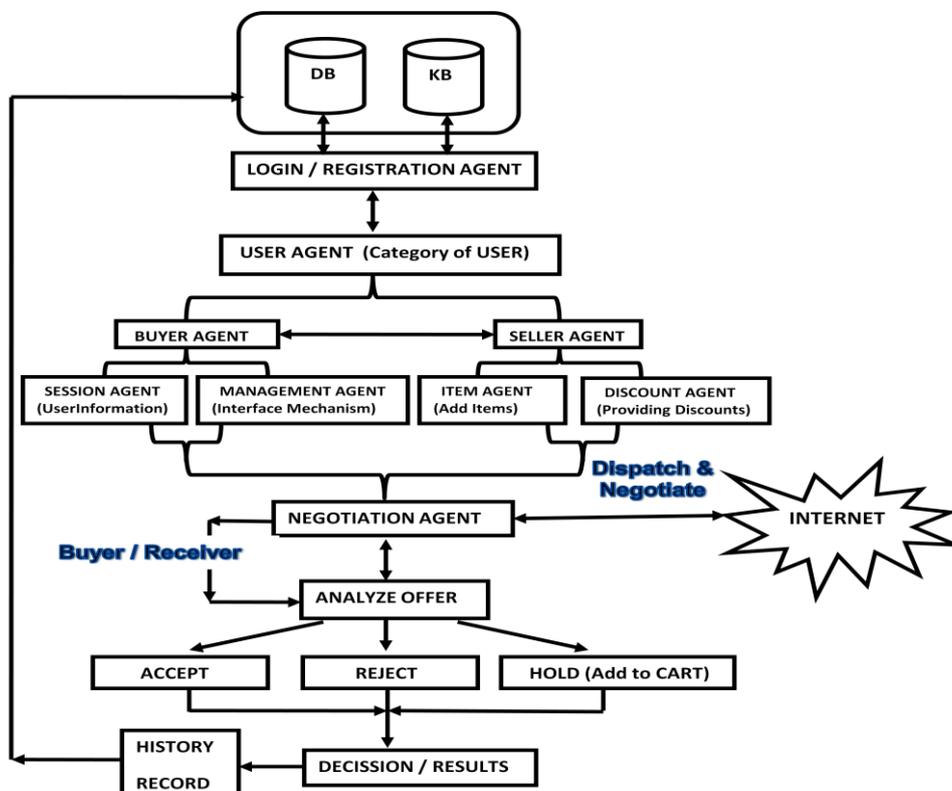


Figure 2: Web Mining Intelligent Agents (WEBMINTEL) System

negotiation profiles have been developed through JSP and JDBC for e-commerce. This approach seems to be intelligent due to the usage of latest tools and methodologies. The outcome of proposed system is positive and efficient in comparison to the outdated directions which had been followed by many researchers in the past such as Perl and SQL like query mechanism. The prospective work is to make the web data mining system more efficient and intelligent. A future research direction is to investigate an alternative approach to update the proposed system. This may be done using genetic algorithms.

5. REFERENCES

- [1] Cen Q., Zhao J., Zhu X., "The data mining system based on multi-agent under the circumstances of e-commerce", Proceedings of 3rd International Conference on Natural Computing (ICNC 2007) IEEE, Vol. 7, pp. 34-38, 24-27 August, 2007.
- [2] Tang Q., Xie F., "Agent based negotiation model for e-commerce", Wireless Communications, Networking and Mobile Computing, Proceedings of WiCom 2007, International Conference, pp. 3466-3469, 21-25 September 2007.
- [3] "The IBM Agent", Available at: <http://www-03.ibm.com/systems/director/downloads/agents.html>. Accessed 09-October-2016.
- [4] Brown C. M., Danzig B. B., Hardy D., Manber U., Schwartz M. F., "The harvest information discovery and access system", Proceedings of 2nd International World Wide Web Conference, Chicago, 1994.
- [5] Hammond K., Burke R., Martin C., Lytinen S., "FAQ-Finder: A case based approach to knowledge navigation", Proceedings of 11th Conference on Artificial Intelligence for Applications, IEEE Xplore, pp.80-86, 20-23 Feb, 1995.
- [6] Kirk T., Levy A. Y., Sagiv Y., Srivastava D., "The information manifold", Proceedings of the AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments, pp.85-91, 1995.
- [7] Malinowski, Wilamowski B. M., "Compiling computer programs through internet", Proceedings of International Conference on Information Technology Based Higher Education and Training, ITHET-2000, Istanbul, Turkey, pp. 343-348, July 3-5, 2000.
- [8] Doorenbos R. B., Etzioni O., Weild D. S., "A scalable comparison shopping agent for the world wide web", Available at: Technical Report 96-01-03, University of Washington, Deptt. of Computer Science and Engineering, 1996.
- [9] Broder A. Z., Glassman S. C., Manasse M. S., Zweig G., "Syntactic clustering of the web", Proceedings of 6th International World Wide Web Conference, pp. 1157-1166, 1997.
- [10] Chang C., Hsu C., "Customizable multi-engine search tool with clustering", Proceedings of 6th International World Wide Web Conference, 1997.
- [11] Wulfekuhler M. R., Punch W. F., "Finding salient features for personal web page categorization", Proceedings of 6th International World Wide Web Conference, 1997.
- [12] Chajri M., Fakir M., "Application of data mining in e-commerce", Journal of Information Technology Research, IGI Publications, US, Vol. 7(4), pp. 79-91, 2014.
- [13] Qingtian H., Xiaoan G., Wenguo W., "Study on web mining algorithm based on usage mining", Proceedings of 9th International Conference on Computer Aided Industrial Design and Conceptual Design, IEEE Xplore, pp. 1121-1124, 22-25 November, 2008.
- [14] Tang Y., Liu X., Huang Y., Lu G., "The design of data mining system", Theory and application of System Project, pp. 56-63, September 2000.
- [15] Wang T. H., Guan S. U., Ong S.H., "Migration control for mobile agents based on passport and visa", Proceedings of Future Generation Computer Systems, Vol. 19, pp. 173-186, 2003.
- [16] Cooley R., Mobasher B., Srivastava J., "Web mining: Information and pattern discovery on the world wide web", Proceedings of 9th International Conference on Tools with Artificial Intelligence, IEEE Xplore, pp. 558-567, 3-8 November, 1997.
- [17] Neli S., Govindasamy K., Wilamowski B. M., Malinowski A., "Automated data mining from web servers using perl script", Proceedings of 12th International Conference on Intelligent Engineering Systems, Miami, Florida, pp. 191-196, February 25-29, 2008.
- [18] Chen L., Sycara K., "WebMate : A personal agent for browsing and searching", Proceedings of the second International Conference on Autonomous Agents, pp. 132-139, 1998.
- [19] Armstrong R., Freitag D., Joachims T., Mitchell T., "WebWatcher: A learning apprentice for the world wide web", AAAI Spring Symposium on Information Gathering, Stanford, CA, March, 1995. Available at: <http://www.cs.cmu.edu/afs/cs/project/theo-6/web-agent/www/project-home.html>
- [20] Liberman H., Letizia, "An agent that assists web browsing", Proceedings of the 14th International Joint Conference on Artificial intelligence, Vol. 1, pp. 924-929, 1995.
- [21] Liberman H., "Autonomous interface agents", Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 67-74, 1997.
- [22] Liberman H., Fry C., Weitzman L., "Exploring the web with reconnaissance agents", Communication of the ACM, Vol. 44(8), pp. 69-75, August, 2001.
- [23] Cooley R., Tan P., Srivastava J., "Discovery of interesting usage patterns from web data", Proceedings of WEBKDD '99 Revised Papers from the International Workshop on Web Usage Analysis and User Profiling, pp.163-182, 1999.
- [24] Top 10 Web Log Analyzers, Available at: <http://www.valencynetworks.com/articles/top-10-web-log-analyzers.html>. Accessed 19-November-2016.
- [25] Tang Q., Xie F., "Agent based negotiation model for e-commerce", Proceedings of International Conference of Wireless Communications, Networking and Mobile Computing, pp. 3466-3469, 2007.

6. AUTHOR PROFILE

Shakti Kundu received his M.Tech. in Computer Science & Engineering from Guru Jambheshwar University of Science & Technology, Hisar, Haryana, India in 2010, M.Phil. in Computer Science from Chaudhary Devi Lal University, Sirsa, Haryana, India in 2008, M.C.A. from Kurukshetra University Kurukshetra, Haryana, India in 2006. Presently he is pursuing his Ph.D. in CSE from DIT University, Dehradun, India. The author current research interests are Web Mining, Knowledge Management and Web Testing. He is life member of CSI, ISTE, IAENG, AIRCC and IAEME.

Dr. M L Garg, presently Professor, Department of Computer Science & Engineering at DIT University, Dehradun, India, has obtained his Ph.D. degree in Computer Science & Engineering from Thapar Institute of Engineering & Technology (Deemed University), Patiala, India with collaborative research work at IIT Delhi, in the year 1992. His area of research includes Fuzzy Logic, Genetic Algorithms & Knowledge Representation and Reasoning. He has published about thirty research articles in peer reviewed International and National Journals and Conferences.