A Survey : An Approach for semantic-Synaptic Web Entropy

Ramkaran Lovewanshi[1] Pankaj Kumar Sahu[2] Mrs. Pranjal Tivari[3]
Dept. of computer science,RGPV University
Ramkaran.bpl@gmail.com,
pankajkumarsahu99@gmail.com
career.gargi@gmail.com

Abstract: **Semantic web mining and Synaptic web entropy mining is an important and recent research area today where number of technique presented in order to mine the web crawl efficiently and to find the web page rank of various data available in the web, in the present paper which is taken by us for the further research is the hybrid approach where the entropy is calculated based on the semantic-synaptic based approach and the important role of the entropy required today to monitor the todays web fluctuation and various stages of multiple portals and web data available today , here as the authors of the paper mentioned about the scope of the entropy monitoring ,we assume it to be a great way to make it better to experience and query search for the extracted data and entropy monitoring, we have monitored a paper Line – up approach which is efficient and determined to visualized the ranking data and optimize according to the user requirement and monitored the data efficiently, here we would like to further enhance research work on analyzing and using the entropy data as input and to use them in Line up technique to visualize and to optimize according to the user requirement in the web entropy visualization.**

INTRODUCTION

Web mining is the integration of information gathered by traditional data mining methodologies and techniques with information gathered over the World Wide Web. It is used to understand customer behaviour, evaluate the effectiveness of a particular Web site, and help quantify the success of a marketing campaign. It also allows looking for patterns in data through content mining, structure mining, and usage mining. Content mining is used to examine data collected by search engines and web spiders. Structure mining is used to examine data related to the structure of a particular Web site and Web Usage Mining is applied to many real world problems to discover interesting user navigation patterns for Improvement of web site design by making additional topic or recommendations observing user or customer behaviour.Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data, in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behaviour at a Web site. Web usage mining itself can be classified further depending on the kind of usage data considered. They are web server data, application server data and application level data. Web server data correspond to the user logs that are collected at Web server. Some of the typical data collected at a Web server include IP addresses, page references, and access time of the users and is the main input to the present Research. This work concentrates on web usage mining and in particular focuses on discovering the web usage patterns of websites from the server log files.This Web Usage Mining is the process in which user access patterns are discovered and analyzed by mining the log files and related data associated with a certain website. It is a kind of web mining which automatically discovers user usage patterns and is helpful in studying and analyzing user interests. Web usage mining consists of mainly three stages, namely data preprocessing, pattern discovery and pattern analysis.

A. Data Pre-processing

The data should be pre-processed to improve the efficiency and ease of the mining process. The main task of data pre-processing is to prune noisy and irrelevant data, and to reduce data volume for the pattern discovery phase. Field Extraction and data cleaning algorithms parse the web log records separating the fields and purging. Covering

B. Pattern discovery

Few techniques to discover patterns from pre-processed data are listed like converting IP addresses to domain names, filtering, dynamic site analysis, cookies, path analysis, association rules, sequential patterns, clustering, decision trees etc.

C. Pattern Analysis

Following statistics are a few listed ones which are the end products of analysis such as the frequency of visits per document, most recent visit per document, who is visiting which documents, frequency of use of each hyperlink, and most recent use of each hyperlink. The common techniques used for pattern analysis are visualization techniques, OLAP techniques, Data & Knowledge Querying, Usability Analysis.

II. LITERATURE REVIEW

In the literature there are authors and paper such as [1][3] they have studied and provided detailed review on web mining focusing on different dimensions of this field, the web mining and its approach for the database users are immense where the work is depends on the web structure and further the syntax based approach for the data extraction from the web resources is provided. The data accessing system and arranging them in the required order and formatting in users way is provided by the authors.

In this paper authod [1] highlighted use of cloud computing in web mining , where nowadays a number of usage are migrating on cloud storage and further the web data is keeping over the cloud with the security point of view and also on the view of remote accessibility with large number of options on web. The web structure mining over the cloud using their approach works well with the cloud and further the execution performed by them on amazon cloud with web structure mining.

In this paper [3] author have focused on scope of agent technology in it whereas [4] provided details on web personalization through web mining, where a personal data storage and mining the data such as profile data, provide media content as image, video and other document related entity were extracted and performed using this technique. The technique work well with dom structure in web availability.
In this paper author [2] provided semantic web mining and suggested an ontology learning mechanism for the extraction of semantics through grammatical rule extraction technique. Where the onotology and semantic bag keyword is used to manipulate and data usage. The semantic data web technique is used in the scenario where the matching probability is increased while comparing with the existing syntax based approach working with data mining. The semantic web mining gives the large number of data output as it match with related works to query driven by the user.

In this paper author [5] proposed an agent based web mining model for e-buisness. The theme for the research demonstrated that the web agent based technique can be used for the business e-commerce purpose for the data mining. Such as OMG and other group may apply the mining level technique with the existing e-commerce architecture. While dealing with the large data on web there is a high probability of extracting product and hence providing the search product solution to the user using agent based web mining model in web data mining approach.

In this [8] provided a multi-agent module working as knowledge crawler. [6] employed web mining for on-line social network analysis, however strategy for selecting appropriate sample size to reflect exact real social networks andactual implementation is left as future research. In [7] proposed an agent based web text mining system for mining HTML based documents on the web, however it still lacks efficient algorithm for very large document collections and use of XML specifications.

Critical review of literature highlights this fact that agent technology has widely been employed in semantic web applications at various fronts and researchers have agreed on its applicability for mining semantic web contents. Although some efforts had already been made to propose application specific agent based solution in diverse areas like e-business[5] or for social networking[6], but there is no standard framework for semantic web content mining. Thus, there is scope of research in this direction. Upcoming section elaborates our proposed framework. [15] proposed the next agent based web mining but there is a scope to research in content file in contrast of unstructured data mining with concept of web. Multimedia mining already included in agent based web mining al[15] but the user timing log mining and file size mining can provide a better way to meet the requirements.

CHALLENGES WITH WEB MINING

Today the World Wide Web is popular and interactive medium to distribute information. The web is huge, diverse, dynamic and unstructured nature of web data , web data research encountered lot of challenges for web mining. Information user could encounter following challenges when interacting with web.

1. **Finding Relevant Information-**

People either browse or use the search service when they want to find specific information on the web. Today's search tools have problems like low precision which is due to irrelevance of many of the search results. This results in a difficulty in finding the relevant information. Another problem is low recall which is due to inability to index all the information available on the web.

2. Creating new knowledge out of the information available on the web-

This problem is basically sub problem of the above problem. Above problem is query triggered process (retrieval oriented) but this problem is data triggered process that presumes that already has collection of web data and extract potentially useful knowledge out of it.

3. Personalization of information-
When people interact with the web they differ in the contents and presentations they prefer.

4. Learning about Consumers or individual users-
This problem is about what the customer do and want. Inside this problem there are sub problem such as customizing the information to the intended consumers or even to personalize it to individual user, problem related to web site design and management and marketing.

5. Finding or Analysing the Large Data –
Large Amount of the data is unable to monitor and optimize according to the user requirement, so here the requirement is to find the best way to analyse it efficiently.

REFERENCES

[1]. Huiping Peng"Discovery of Interesting Association Rules Based on Web Usage Mining" 2010 International Conference.

[2]. Sanjay Kumar Malik, Nupur Prakash, S.A.M. Rizvi" Ontology and Web Usage Mining towards an Intelligent Web focusing web logs" 2010 International Conference.

[3]. Hao Yan, Bo Zhang, Yibo Zhang, Fang Liu, Zhenming Lei "Web usage mining based on WAN users' behaviours" 2010 International Conference.

[4]. Han J., Pei J., Yin Y. and Mao R., "Mining frequent patterns without candidate generation: A frequent-pattern tree approach" Data Mining and Knowledge Discovery, 2004.

[5]. Agrawal R, Srikant R., "Fast Algorithms for Mining Association Rules", VLDB. Sep 12-15 1994, Chile, 487-99, ISBN 1-55860-153-8.

[6]. Ting I.H., _Web Mining Techniques for On-line Social Networks Analysis'. In Proceedings of the 5th International Conference on Service Systems and Service Management, Melbourne, Australia, 30 June-2 July 2008, pp. 696-700.

[7]. Jicheng W., Yuan H., Gangshan W. &Fuyan Z., _Web Mining: Knowledge Discovery on the Web'. In Proceedings of IEEE International Conference on System, Man and Cybernetics 1999 (IEEE SMC'99), Vol. 2 , pp. 137-141.

[8]. Zhan L. &Zhijing L., _Web Mining based on Multi-Agents'. Published in proceedings of Fifth International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'03), 2003.

[9]. C.Dimou, A.Batzios, A.L.Symeonidis and P.A.Mitkas, _A Multi-agent framework for Spiders Traversing the Semantic Web'. In Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence.

[10]. F. Buccafurri, G. Lax, D. Rosaci and D. Ursino, _Dealing with Semantic Heterogeneity for Improving Web Usage'. Data Knowledge Eng. Vol. 58, Issue 3, pp. 436–465,2006.

[11]. Singh A., Juneja D. and Sharma A.K., _Design of Ontology-Driven Agent based Focused Crawlers'. In proceedings of 3rd International Conference on Intelligent Systems & Networks (IISN-2009),Organized by Institute of Science and Technology, Klawad, 14 -16 Feb 2009, pp. 178-181. Available online in ECONOMICS OF NETWORKS ABSTRACTS, Volume 2, No. 8: Jan 25, 2010.

[12]. Singh A., Juneja D., Sharma A.K., 'Design of An Intelligent And Adaptive Mapping Mechanism For MultiagentInterface'.In Proceedings of International Conference on High Performance Architecture and Grid Computing Communications in Computer and Information Science (HPAGC'11), 2011, Volume 169, Part 2, 373-384, DOI: 10.1007/978-3-642-22577-2_51.

[13]. Singh A., Juneja D., Sharma A.K., _General Design Structure of Ontological Databases in Semantic Web'. Published in International Journal of Engineering, Science & Technology, Vol. 2, Issue 5, pp. 1227-1232, 2010.

[14]. Karayannidis N. &Sellis T., _Hierarchical Clustering for OLAP: The CUBE File Approach'. Published in The VLDB Journal — The International Journal on Very Large Data Bases, Vol. 17, Issue 4, July 2008.

[15]. Aarti Singh, _Agent Based Framework for Semantic Web Content Mining'. Published in International Journal of Advancements in Technology,Vol. 3 No.2 (April 2012), ISSN 0976-4860.

