Volume 2, Issue 3, July 2015 ISSN: 2348-8565 (Online)
International Journal of Modern Engineering and Research Technology
Website: http://www.ijmert.org Email: editor.ijmert@gmail.com

Line Up: A Technique for Semantic-Synaptic Web Entropy Visualization

Anchal Singh Thakur
M.Tech Scholar
Department of Computer Science and Engineering
Lakshmi Narain College of Technology
Bhopal, (M.P.), [INDIA]
Email : thakuranchal02@gmail.com

Dr. Vineet Richhariya
Head of the Department
Department of Computer Science and Engineering
Lakshmi Narain College of Technology
Bhopal, (M.P.), [INDIA]
Email : vineetrich@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.

Keywords:—*Web mining, Semantic mining, Web structure.*

I. 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 website 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. SYNAPTIC WEB

Here we are observing the various synaptic observation and the various reasons to find synaptic analysis of web data:

- You are creating new *synaptic* connections with every new activity you engage in.
- The *synaptic* network that finally emerges is only partly determined by smaller information also.

It is widely accepted that the synapse plays a role in the formation of memor. As neurotransmitters activate receptors across the synaptic cleft, the connection between the two neurons is strengthened when both neurons are active at the same time, as a result of the receptor's signaling mechanisms. The strength of two connected neural pathways is thought to result in the storage of information, resulting in memory.

III. LITERATURE REVIEW

Huiping Peng in 2010 stated the use of FP-growth algorithm for processing the web log records, obtaining a set of frequent access patterns, then using the combination of browse interestingness and site topology interestingness of association rules for web mining [1]. Analysis of web usage mining by using Web Log analyzer tool, “Web Log Expert” was carried out by Sanjay Kumar Malik et al. in 2010. They focused on the

development of Ontology for an intelligent or efficient web and it's relation with web usage mining. Finally, they also summarize some other research challenges towards an intelligent machine and web environment.

Hao Yan proposed a two-step K-means clustering algorithm to search user groups in realistic data collected from WAN. They gave some useful practical conclusions to facilitate design of targeting and recommending applications, K-mean is efficient algorithm which is already proven for the best clustering approach, further more K-mean is extended to K-Medoit and other scheme to make further more efficient clustering from the available data and document from the user.

Jiawei Han in 2004 proposed a novel frequent-pattern tree structure, which is an extended prefix-tree structure for storing compressed, crucial information about frequent patterns, and develop an efficient FP-tree based mining method, FP-growth, for mining the complete set of frequent patterns by pattern fragment growth.

Rakesh Agrawal and Ramakrishan Srikant in 1994 consider the problem of discovering association rules between items in a large database of sales transactions. They present two new algorithms for solving this problem that are fundamentally different from the known algorithms. They also show how the best features of the two proposed algorithms can be combined into a hybrid algorithm, called AprioriHybrid

Mohd helmy Abd Wahab in 2008 describes the pre-processing techniques on IIS Web Server Logs ranging from the raw log file until before mining process can be performed.

C.P Sumathi in 2011 presented an overview of the various steps involved in the preprocessing stage.

Renata Ivancsy in 2006 investigated three pattern mining approaches from the web usage mining point of view.

Vaibhav Kant singh in 2008 shows how the different approaches achieve the objective of frequent mining. They also look for hardware approach of cache coherence to improve efficiency of the above process.

Dr. R. Krishnamoorthi and K. R Suneetha in 2009 has done the in-depth analysis of Web Log Data of NASA website to find information about a web site, top errors, potential visitors of the site etc. which help system administrator and Web designer to improve their system by determining occurred systems errors, corrupted and broken links by using web using mining.

IV. Problem Formulation

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

analyse the texual, semantic and synaptic results.

The complete flow is shown in below figure:

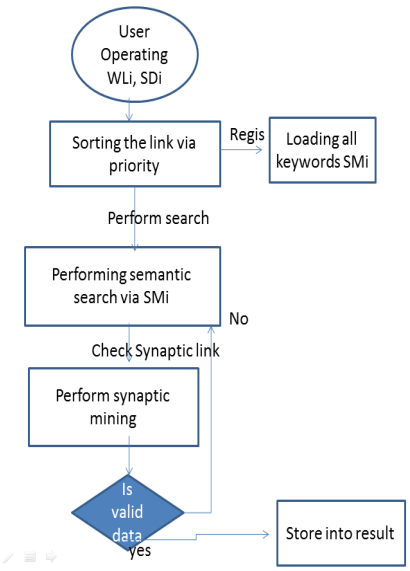

Figure 1- Flow of proposed work

In figure above we presented the steps which followed in order to perform the proposed work.

VII. CONCLUSION

We have discussed data mining and web mining categories have been discussed. Semantic –Synaptic mining also has been analyzed along their advantages and Entropy having their usage. An algorithm has been proposed for Mining and Visualizing of mining in web usage mining which is efficient than traditional Tabular scheme. The first part of algorithm, i.e. backward scan. firstly scan the web log database and obtain the longest candidate level length. After that count the occurrence of each candidate. Each candidate count satisfy the minimum threshold value and then obtain the maximum forward reference from candidate count length. The new approach requires minimum repeated database scan for mining in web usage mining. It will reduce the time and space execution.

Web usage mining is the application of data mining techniques to discover usage patterns from Web data, in order to understand and better serve the needs of Web-based applications. Web usage mining consists of three phases, namely pre-processing, pattern discovery, and pattern analysis. One of the algorithms which is very simple to use and easy to implement is the Hybrid algorithm. In this paper a new technique is proposed to discover the web usage patterns of websites from the server log files with the foundation of clustering and improved Hybrid algorithm. The effective algorithm will be proposed with the improvements as well as the implementation of Hybrid Algorithm. The forthcoming step in the research work shall be to design the improved version of the Hybrid Algorithm that shall be implemented on the Server Log Files for Association Rule Mining.

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

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.

V. PROPOSED WORK

Here we briefly describe a technique to discovered frequent item pattern.

A. Symantec Mining

A Web mining from the crawl is done first, we are extracting the information from the web based on the similar type of object and their availability in semantic manner, the data is been extracted and use to create Entropy.

B. Synaptic Mining

In this algorithm, the patterns are categorized according to the length executed on lattice model. Patterns will form a lattice based on the pattern-length and pattern-frequency. And using this lattice, frequent patterns are searched depth first.

Lattice Construction: The basic element of the lattice is an atom i.e. single page. Each atom or page stands for length-1 prefix equivalence class. Beginning from bottom elements the frequency of upper elements with length n can be calculated by using two n-1 length patterns belonging to the same class.

C. Applying Line Up on Entropy and Mined Data

We are applying now the Line-up technique where the things can be substitute in various dataset and the result observed from the various semantic data and user can optimize according to the visualize and observation required.

VI. PROPOSED ALGORITHM :

Input:

Input web links WLi, Semantic keywords dataset SDi.

Output : input query result, semantic count, synaptic count.

Steps:

- Read all the inputs links WLi and store into the map according to the priority. TreeMap.
- Load all the semantic keywords from the dataset based on the input query.
- Perform semantic web search using web crawl structure from web dataset SDi.
- Performing Synaptic search on performing search on branches of semantic WL web links.
- Finding synaptic results and add to recursive semantic search on it.
- Store into output array count and

Rules", VLDB. Sep 12-15 1994, Chile, 487-99, ISBN 1-55860-153-8.

[6] Mohd Helmy Abd Wahab, Mohd Norzali Haji Mohd,Mohamad Mohsin "Data Pre-processing on Web Server Logs for Generalized Association Rules Mining Algorithm" 2008.

[7] C.P. Sumathi, R. Padmaja valli,"An overview of pre-processing of web log files for web usage mining" 2011.

[8] Renata Ivancsy, Istvan Vajk "Frequent Pattern Mining in Web Log Data"2006.

[9] Vaibhav Kant Singh, Vijay Shah, Yogendra Kumar Jain, "Proposing an Efficient Method for Frequent Pattern Mining" 2008

[10] K. R. Suneetha, Dr. R. Krishnamoorthi, "Identifying User Behaviour by Analyzing Web Server Access Log File" 2009.

[11] R. Cooly, B. Mobasher and J. Srivastava, "Web Mining: Information and Pattern Discovery on the World Wide Web", IEEE, August 1997.Pp.558 – 566.

[12] Charu C. Aggarwal and Philip S. Yu, "An Automated Sys - tem for Web Portal Personalization", Proceedings of the 28th VLDB Conference, Hong Kong, China, 2002

[13] Renáta Iváncsy, István Vajk, "Frequent Pattern Mining in Web Log Data", Acta Polytechnica Hungarica, 2006

[14] B.Santhosh Kumar, K.V. Rukmani, "Implementation of Web Usage mining Using Apriori and FP Growth Algorithm", Int. J. of Advanced networking and Applications, (2010), Pages: 400-404

[15] Shui Wang, Le Wang, "An implementation of FP growth algorithm based on high level data structure of Weka-JUNG framework", Journal of Convergence Information Technology, Number 9, 2010.

[16] Kotsiantis S, Kanellopoulos D., "Association Rules Mining: A Recent Overview", GESTS International Transactions on Computer Science and Engineering, 2006, pp.71- 82

[17] J. Han, J. Pei, and Y. Yin. "Mining frequent patterns with out candidate generation". IEEE, Sept.1998 pp-365-378.

* * * * *