

A Fast and Efficient Hybrid Approach for Web Caching and Pre-Fetching To Enhance End User Experience over Web Environment

Monti Babulal Pal

Department of Computer Science & Engineering
Shri Vaishnav Institute of Technology & Science,
Indore, India

Dinesh Chandra Jain

Department of Computer Science & Engineering
Shri Vaishnav Institute of Technology & Science,
Indore, India

ABSTRACT

With Rapid growth of web environment end user want a fast and effective response over internet. Effective way to serve fast response is web caching is a well-known strategy for improving the performance of Web based system by keeping Web objects that are likely to be used in the near future in location closer to user. In this paper an effective web caching and pre fetching technique has been proposed that use FP growth algorithms for frequent page generation , relative weighted rule for determining relative weight of each page with respect to other in order to enhance response and Markov model use to store relative weight of page to their relative position for fast and efficient web pre-fetching in order to has improved user response of web page and expedites users visiting speed.

Keywords

Web Services, log preprocessing, FP growth, relative weight, and Markov model

1. INTRODUCTION

Over last decade World Wide Web has grow to be an imperative medium for communication namely for storing, sharing and distribution of information globally. WWW is treated as global information service centre for widely distributed information like news ,e-commerce, advertisement. Thrombus enhancement in employment of WWW with time leads to make difficulty in information retrieval process via WWW which tends to make user bore and impatience while waiting for information due to network delay. This situation would be resolve efficiently by Web mining.

The web page pre-fetching plays an important role in order to enhance the response time and make the application well-organized. The web pre-fetching is a technique in order to preprocess the user requests, before they are actually demanded. Therefore, the time that the user must wait for the requested documents can be reduced by hiding the request latencies. Pre-fetching is the method for reducing Latencies. The user always expects an interactive response, better satisfaction and quality of output. There are various approaches and algorithms have been proposed for improving the web performance.

The proposed work has improved the hit ratio of the web page and expedites users visiting speed. Predictive Web pre-fetching refers to the mechanism of deducing the forth coming page accesses of a client based on its past accesses. In this work we demonstrate the frequent mining pattern which is obtain on the basis of input and on the basis of that caching

and prefetching ratio is calculated. Thus we present a new idea for the interpretation of Web pre-fetching and web caching from the given usage items. The approach works on the basis is web mining with the combination of markov model.

This paper is divided into seven sections. First one is introduction in which give the brief description of work. The second section discusses the previous work related to the topic. The third section describes the approach used in the presented work. The next section describes the proposed architecture of the presented work. After this the simulation result has discussed. Finally paper concludes in the section eight's.

2. PREVIOUS WORK

Research over log file , log preprocessing, web mining and web pre-fetching is going very fast in last decades. A server log is a log file (or several files) automatically created and maintained by a server of activity performed by it. But unfortunately log pre processing received very less attention [1,2]. Ravindra Gupta et al [3] present an Methods for user identification, session identification, and path completion based on F-P tree. In another work T. Revathi et al present and describe different pre processing phases [4]. In recent year research towards web personalization is done by web uses mining [5,6,7] and give an idea of web pre-fetching. It also discuss about the web object pre-fetching. Some authors describe web mining techniques. They explain how the web mining works. What is the taxonomy of web mining? As the web mining uses in order to pre-fetch the web page. Kavita Sharma etal [8] give an idea to explore the usefull information from the web with the help of data mining. Wang Yong-gui etal [1] throws some light on semantic-based Web mining techniques. These techniques are mentioned by number of people in order to improve Web service levels and deal with the existing Web services which are supported by the sort of semantic problem. P. Sampath etal [9] present an weight estimation process with span time, request count and access sequence details.

3. WEB PRE-FETCHING

Web pre-fetching is a technique for reducing web latency based on predicting the next future web objects to be accessed by the user and pre-fetching them during times. If finally the user requests any of these objects, it will be already on the client cache. This technique takes advantage of the spatial locality shown by the web objects [1,2].

The pre-fetching technique has two main components: The prediction engine and the pre-fetching engine. The prediction

engine runs a prediction algorithm to predict the next user's request. The pre-fetching engines decide to prefetch them or not depending on some conditions like available bandwidth. Each engine can work at any element of the web architecture [2].

The predicted page (PP) is the number of objects. It predicted by the prediction engine. Pre-fetch page (PFP) represents the number of objects pre-fetched. The number of objects pre-fetched that are requested later by the user is the pre-fetch hit (HPFP). The opposite of the pre-fetch hit is the pre-fetch miss (MPFP), which represents the number of pre-fetched objects that were never demanded by the user (i.e., extra traffic). Finally, request page (RP) refers to the total amount of objects requested by the user (pre-fetched or not), and the user request not pre-fetched (URnP) represents the number of objects demanded by the user that were not pre-fetched [1].

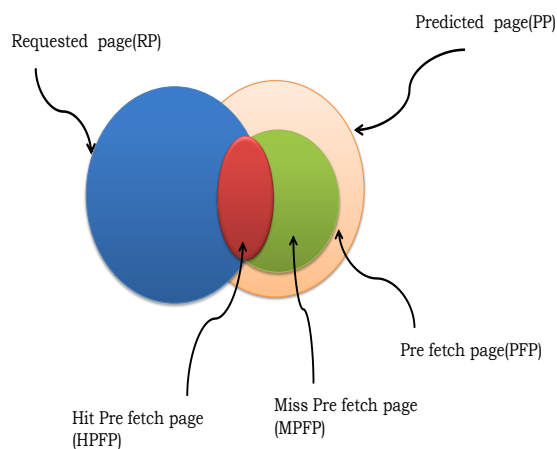


Figure1 Web pre-fetching type of request

As shown in fig. 1, the set of pre-fetch page (PFP) is a subset of the predicted page (PP). The result of the intersection between the requested page (RP) and Pre-fetch page is the pre-fetch hit subset (HPFP). This subset is the main factor to reduce the perceived latency. In Fig. 1, A represents a user request not pre-fetched (URnP), which is a user request neither predicted nor pre-fetched. Red circle is a pre-fetch request made by the pre-fetching engine that is requested later by the user, thus becoming a pre-fetch hit. Green circle is a pre-fetch miss (MPFP) resulting from an unsuccessful prediction that was pre-fetched but never demanded by the user. This request becomes extra traffic and extra server load.

4. PROPOSED ALGORITHMS

The proposed algorithms in order to improve hit ratio of the web pre-fetch page and expedite users visiting speed. Prediction of upcoming Web page depend upon forth coming page accesses of a client based on its past accesses which is gather from web log file. Web log file is used to capture client request at server side, whether web log file is an rough set having higher degree of noise so it is need to pre process log file before taking any decision over web pre fetching. Proposed algorithms apply preprocessing process over raw log file.

Log Pre-Processing- Along with that important information there is also some inconsistent data like noise, null value and other error information [12] which is not so important for web personalization so in order improve web mining result its need to refine web log file before mining. Data cleaning, user and session identification, data integration and so on are main important part of log pre processing.

Data Cleaning:- In proposed methodology data cleaning process use to recognized useful token and remove unwanted and redundant token then store in data base after normalization [10].

User & Session identification:- User and session identification is very important step towards web personalization generally IP address is used to distinguish but when there is an proxy server then number of user having same IP address then some more attribute like browses information ,operating system and Refer URI field is used as per concern[11].

F-P tree for frequent page generation:- Proposed algorithms used F-P growth [4] concept to generate frequent web page. F-P growth algorithms use to generate frequent item set without need to generating candidate item set so it is very time efficient to use F-P growth .

Assignment of relative weight - Proposed algorithms use Weight rule concept to assign relative weight to each page with respect to each other page. This relative weight is use to represent probability of page p request just after page q .

$$R_w = \frac{\text{Number of occurrence of page q and page p together}}{\text{Number of occurrence of page q}} \dots 1$$

Weighted Markov matrix:- proposed algorithms use M*M two dimensional table to store relative weight. M is total number of unique transaction /page that had be identify in above step. for example relative weight of p wrt q is store at qth row and pth column then at time pre fetching if q page is called then at qth row the page having highest relative weight is to be pre fetched with page q.

Algorithms shows in figure 2 demonstrate each and every step of whole work very clearly like token identification, session and transaction identification. The frequent patterns are extracted with the weight values. The weighted support is estimated and used for the pages. As suggested in Algorithm below.

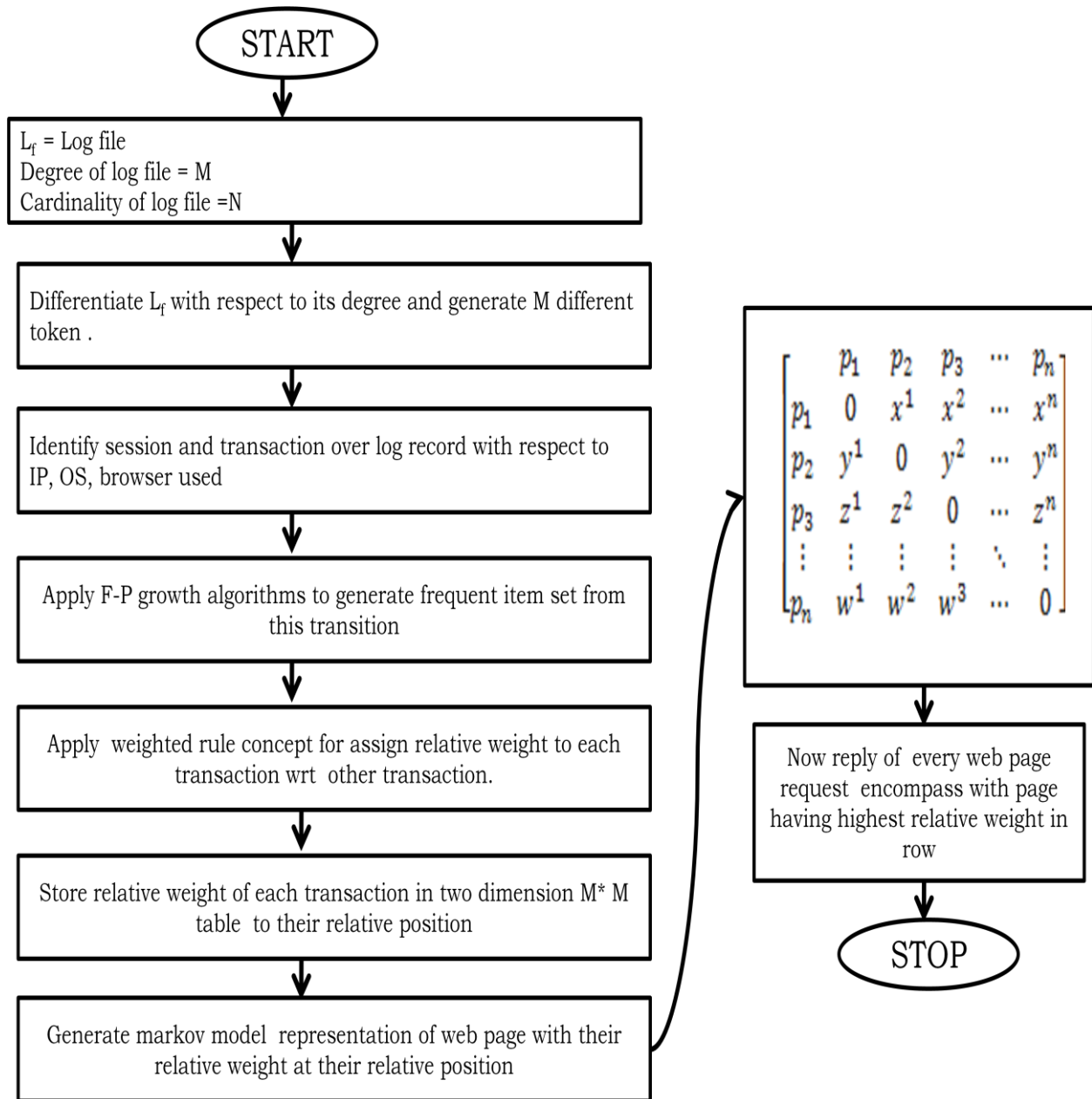


Figure: 2 Proposed Algorithms

5. SIMULATION AND RESULTS

For simulation and result analysis a real time scenario of client server architecture having 30 clients and one server is taken as a scenario for verification of proposed work whole verification is done over MATLAB 10 and used My Sql for data base support.

Time Complexity – Existing weighted systolic tree [2]concept use a balance tree to crate link between pages with pre-fetch page. Whether time complexity to evaluate pre-fetch page is $O(\log N)$ where N is total number of web page /transaction over the web log file. Basically $\log(n)$ time is related to height of systolic tree.

Where time required taking decision about pre-fetch page in proposed model $O(1)$ same as simple markov model concept [3].

Methods	Time Complexity	Space Complexity
Weighted Systolic Tree	$O(\log N)$	$2*N$
Markov Model	$O(1)$	$[(2^n-1)*N]$
Proposed	$O(1)$	$N*N$

Space Complexity - Proposed methodology required higher space than systolic tree but required nominal as compare to simple Markov model present in [3].

6. CONCLUSION

Proposed methodology present an efficient web fetching and web caching methodology in order to improve end user experience with fast and effective response. Propose methodology having optimal space and time complexity as compare to earlier. This paper present an hybrid approach that encompasses FP growth Tree for generations of frequent item set without candidate set generation then use relative weight rule concept to find approximation of pre fetch page and then Markov Model to store web page in order fast retrieval.

7. ACKNOWLEDGEMENT

I would like to thank to my guide “Dr. Dinesh Chandra Jain” who has given his knowledge and time in order to complete this paper. This paper will never complete without the support faculty member of CSE department of SVITS, Indore.

8. REFERENCES

- [1] WANG Yong-gui and JIA Zhen, “Research on Semantic Web Mining” IEEE 2010, pp 67-70.
- [2] P. Sampath, C. Ramesh, T. Kalaiyarasi, S. Sumaiya Banu and G. Arul Selvan, “An Efficient Weighted Rule Mining for Web Logs Using Systolic Tree”, IEEE 2012, pp 432-436.
- [3] Nizar R. Mabroukeh and C. I. Ezeife, “Semantic-rich Markov Models for Web Prefetching”, IEEE 2009, pp 465-470.
- [4] A.B.M.Rezbaul Islam and Tae-Sun Chung, “An Improved Frequent Pattern Tree Based Association Rule Mining Technique”, IEEE 2011
- [5] Toufiq Hossain Kazi, Wenying Feng and Gongzhu Hu, “Web Object Prefetching: Approaches and a New Algorithm”, IEEE 2010, pp 115-120.
- [6] Brijendra Singh and Hemant Kumar Singh, “Web Data Mining Research: A Survey”, IEEE 2010.
- [7] Kavita Sharma, Gulshan Shrivastava and Vikas Kumar, “Web Mining: Today and Tomorrow”, IEEE 2011, pp 399-403.
- [8] Kavita Sharma, Gulshan Shrivastava and Vikas Kumar, “Web Mining: Today and Tomorrow”, IEEE 2011, pp 399-403.
- [9] P. Sampath, C. Ramesh, T. Kalaiyarasi, S. Sumaiya Banu and G. Arul Selvan, “An Efficient Weighted Rule Mining for Web Logs Using Systolic Tree”, IEEE 2012, pp 432-436.
- [10] L. Ramaswamy, A. Iyengar, L. Liu, F. Douglass, “Automatic Fragment Detection in Dynamic Web Pages and Its Impact on Caching”, IEEE Transactions On Knowledge And Data Engineering, Vol. 17, No. 6, June 2005.
- [11] Nacim Fateh Chikhi, Bernard Rothenburger, Nathalie Aussenac-Gilles “A Comparison of Dimensionality Reduction Techniques for Web Structure Mining”, Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence 2007, pp 116-119.
- [12] Nikhil Kumar Singh, Deepak Singh Tomar, Bhola Nath Roy, “An approach to understand the end user behavior through log analysis” International Journal of Computer Applications (0975 – 8887), August 2010.