



www.ijarcsse.com

Volume 6, Issue 5, May 2016

ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Improving the Performance of Proxy Server Using Modified Apriori Algorithm for Pre-fetching

Devendra Parmar
Student of CSE, PIET
India

Yask Patel
Asst. Prof. CSE, PIET
India

Abstract— *Web mining is a large area for finding patterns from the web documents. Pre-fetching is used improving the performance of browser for response of any request from user, that makes copy of document on browser cache memory that user hits in future. Modified Apriori improves the hit ratio of pre-fetched pages for browser.*

Keywords— *Web usage mining, Pre-fetching, Caching*

I. INTRODUCTION

Web mining is a process of mining the data from the large amount of data that is on web pages. It helps users to understand the nature and behavior of the users that are accessing the web.

There are three sub types of web mining.

- 1) Web Usage Mining,
- 2) Web Content Mining,
- 3) Web Structure Mining.

Web Mining is the extraction of interesting and potentially useful patterns and implicit information from artifacts or activity related to the World Wide Web. Web usage mining provides the support for the web site design, providing personalization server and other business making decision, etc.

Web caching is used to reduce the network traffic by caching web pages at the proxy server level but nowadays caching alone is not sufficient because World Wide Web has evolved rapidly from a simple information-sharing mechanism to dynamic and multimedia data. Pre-fetching is next to give better performance because web caching caches the pages that are recently accessed by the user so to get better performance pre-fetching is used that fetch the pages that are being accessed by the user in future. By using Pre-fetching time to requested page reduces. Pre-fetching keeps copy of the pages that are being accessed by user in future so it also increases availability of the page.

In pre-fetching there is a little problem that if a single page from the pre-fetch is not hit by user then overall pre-fetch fails. For this we are using modified apriori that improves over this problem.

II. LITERATURE SURVEY

“Web Usage Mining: A Survey on Pattern Extraction from Web Logs” paper gives brief detail for the prefetching. Also discuss web catching that is discussed. Proxy server and prediction at proxy server using proxy server log. This also gives discuss of user latency and provides framework. Steps in that model are finding similar pages, clustering pages, calculate frequency of each pages, selecting a page from the cluster in sequence.

For prediction in this paper author used sequential ranking algorithm for predicting the page. After that the results are given compared using graph. Conclusion of the paper is this that major advantage of sequential rank selection algorithm is it selects only one web page of a website for pre-fetching purpose of user, hence consumes less space of users. Apriori and FP Growth are used to find most frequent items. The Apriori algorithm generates candidate set during each pass. It reduces the dataset by discarding the infrequent itemsets that do not meet the minimum threshold from the candidate sets[2]. To avoid the generation of candidate set which is expensive the FP Growth algorithm is used to mine the database[2].

III. PROBLEM DEFINITION

Web caching is used to reduce the network traffic by caching web pages at the proxy server level but nowadays caching alone is not sufficient because World Wide Web has evolved rapidly from a simple information-sharing mechanism to dynamic and multimedia data. Pre-fetching is next to give better performance because web caching caches the pages that are recently accessed by the user so to get better performance pre-fetching is used that fetch the pages that are being accessed by the user in future. By using Pre-fetching time to requested page reduces. Pre-fetching keeps copy of the pages that are being accessed by user in future so it also increases availability of the page.

In pre-fetching there is a little problem that if a single page from the pre-fetch is not hit by user then overall pre-fetch fails. For this we are using modified apriori that improves over this problem.

Table I Comparison Based on Parameters

Algorithm	Hit ratio	Time taken to pre-fetch page	No. of patterns Found
Apriori	Less than other	More compare to FP Growth	Less than Others
FP Growth	More than apriori	Less than Apriori	Very much than apriori
Modified Apriori	high than both	High	More than others

After checking the literature on web usage mining, we have found a number of advantages of improved web mining algorithm. In this paper improved version of algorithm (apriori) for the web usage mining. Apriori algorithm has been modified to perform pre-fetching for better hit ratio. For the pre-fetching here URL is used and by using the URL pages pre-fetched according to pattern found by apriori.

IV. RELATED WORK

Modified Algorithm:

- Step 1: Candidate itemset of size k, itemset L_k
 $k = 1$ 2. Generate frequent itemsets of length l
- Step 2: Generate length (k+1) candidate itemsets from length k frequent itemsets
- Step 3: Prune candidate itemsets containing subsets of length k that are infrequent
- Step 4: How many k-itemsets contained in a (k+1)-itemset?
- Step 5: Count the support of each candidate by scanning the DB
- Step 3: Eliminate candidates that are infrequent, leaving only those that are frequent

Pseudo Code for proposed work:

```

Pseudo-code:
Ck: Candidate itemset of size k
Lk : frequent itemset of size k
L1 = {frequent items};
for ( k = 1; L k != ∅; k++) do begin
Ck+1 = candidates generated from L k;
Lk+1 = candidates in Ck+1 with min_support
End
return U k Lk
    
```

Here Modified Apriori algorithm works as same to the apriori but there is little modification that makes it different is this that is does not provides long patterns by repeating joining and pruning steps but joining steps is done at once and pruning is done twice in overall algorithm so we are getting more patterns but that are short patterns . This short patterns are used for pre-fetching the pages that are being accessed by user in future.

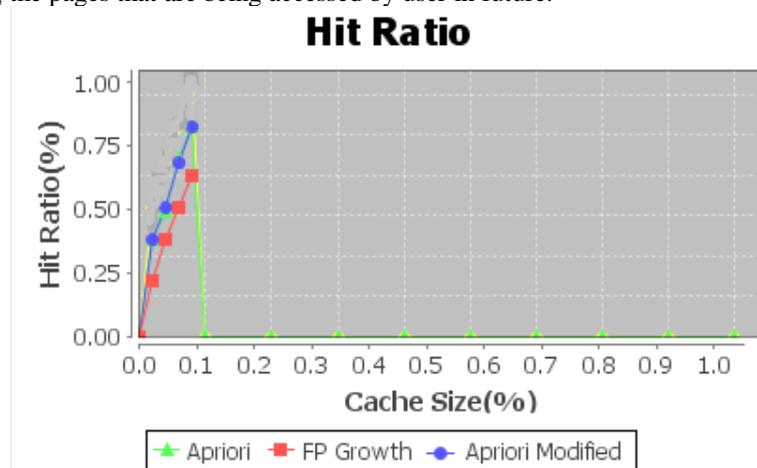


Fig. 1 Hit Ratio Comparison of Apriori, FP Growth and Modified Apriori

V. CONCLUSION AND FUTURE WORK

Modified Apriori is proposed for prediction of web requests of users and accordingly, pre-fetching the content from the server. The modified algorithm improves performance of browser using web usage mining and pre-fetching scheme which is clear in the result section. Later cluster the user according to their access pattern and usage behavior with the help of K-Means algorithm and then modified Apriori algorithm is applied to generate rules for pre-fetching pages that works more better than apriori.

Future work is extending algorithm for Clustering based on location with both Apriori and FP-Growth algorithms. Implementation of the remaining proposed work will using Clustering with both IP based and location based is for as future work. In this Modified algorithm Time required for generation of pattern is high than apriori and FP Growth algorithm so the future work is to reduce that time for generation of pattern.

REFERENCES

- [1] Ashika Gupta, Rakhi Arora, Ranjana Sikarwar, neha Saxena “Web Usage Mining Using Improved Frequent Pattern Tree Algorithms” IEEE 2014
- [2] Mr.Rajesh K. Ahir, Ms. Mital B. Ahir “Algorithms For Mining Frequent Patterns: A Comparative Study” International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 12, December 2013
- [3] Ankita Kusmakar, Sadhna Mishra “Web Usage Mining: A Survey On Pattern Extraction From Web Logs” International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 9, September 2013
- [4] Ramya C. Shreedhara K. S., And Kavitha G “Preprocessing: A Prerequisite For Discovering Patterns In Web Usage Mining Process” International Journal of Information and Electronics Engineering, Vol. 3, No. 2, March 2013
- [5] Greeshma G. Vijayan and Jayasudha J. S. “A survey on web pre-fetching and web caching techniques in a mobile environment” 2012
- [6] Chintan R. Varnagar, Nirali N. Madhak, Trupti M. Kodinariya, Jayesh N. Rathod “Web Usage Mining: A Review on Process, Methods and Techniques” IEEE, 2012.
- [7] Shahnaz Parvin Nina, Md. Mahamudur Rahaman, Md. Khairul Islam Bhuiyan, Khandakar Entenam Unayes Ahmed “Pattern Discovery of Web Usage Mining” IEEE, 2009.
- [8] Rajni Pamnani, Pramila Chawan “Web Usage Mining: A Research Area in Web Mining”, Department of computer technology, VJTI University, Mumbai, 2009
- [9] Suneeha, K. R. and D. R. Krishnamoorthi, "Identifying User Behavior by Analyzing Web Server Access Log File", (IJCSNS) International Journal of Computer Science and Network Security, VOL.9, No.4, April 2009.
- [10] Yan Li, Boqin Feng, Qinjiao Mao, “Research on Path Completion Technique in Web Usage Mining”, Computer Science and Computational Technology, 2008. ISCSCT '08. International Symposium on Volume 1, 20-22 Dec. 2008
- [11] Zidrina Pabarskaite, Aistis Raudys, “A process of knowledge discovery from web log data: Systemization and critical review”, Journal of Intelligent Information System, Springer, 2007.
- [12] J. Srivasta, R.Cooley, M.Deshpande, P.Tan, "Web usage mining: discovery and applications of usage patterns from Web data", ACM SIGKDD Vol.7, No.2, Jan-2000.
- [13] Kosala and Blockeel, “Web Mining Research: A Survey”, SIGKDD Exploration, Newsletter of SIG on Knowledge Discovery and Data Mining, ACM, Vol.2, 2000.
- [14] Seung Won Shin, Byeong Hag Seong & Daeyeon Park, (2000)"Improving World-Wide- Web Performance Using Domain-Top Approach to Prefetching”, Fourth International Conference on High-Performance Computing in the Asia-Pacific Region vol. 2, pp. 738-746.
- [15] Evangelos P. Markatos & Catherine E. Chronaki, (1998) “A Top-10 “Approach to Prefetching on the Web”, Proceedings of INET’ 98 (The Internet Summit).
- [16] R. Cooley, B. Mobasher, J. Srivastava, "Web mining: information and pattern discovery on World Wide web”, tools with artificial intelligence, Ninth IEEE International November 1997.