IJRCCT International Journal of Research in Computer and Communication Technology ISSN (O) 2278-5841 ISSN (P) 2320-5156

Survey of Recent Web Prefetching Techniques

Sonia Setia
Echelon Institute of Technology
Faridabad
setiasonia53@gmail.com

Dr. Jyoti
YMCAUST
Faridabad
justjyoti.verma@gmail.com

Dr. Neelam Duhan
YMCAUST
Faridabad
neelam_duhan@rediffmail.com

***Abstract*—Web caching and web prefetching are the two major areas of research focused at reducing the user perceived latency. Both if used well can greatly help in reducing this latency as web caching helps in exploiting temporal latency while web prefetching helps in exploiting spatial latency. However if prefetched pages are not visited by the users in their future accesses, they can increase the network traffic and overload the web server. This paper aims at surveying various research papers who have worked in this direction.**

***Index Terms*— Web Prefetching, Predictions, Proactive Caching.**

I. INTRODUCTION

With the advancement in information technology, the World Wide Web (WWW) has become a huge information repository. Today, it covers almost every topic including text, image, audio, video, and metadata, in which a human user could be interested. The enormous spread of Internet and consequently, the growth of the World Wide Web have implied a dramatic increase in the number of connected users and therefore, a tremendous increase in the global traffic. Thus it can damage the quality of service (availability, reliability, and security) and specially the latency perceived by the users.

People use the WWW to access information from remote sites. But they do not like to wait long for their results. The latency of retrieving a Web document depends on several factors.

- *Speed of Servers*
- *Speed of clients*
- *Network Bandwidth and Propagation Delay*

Some of these delays, such as client or server slowness or transmission time, can be reduced by using faster computers or higher bandwidth links. However, propagation delay which is basically determined by the physical distance traversed cannot be reduced beyond a point.

One possible remedy to the latency problem is to use caching. Caching is used to reduce the response time by storing copies of popular web documents in a local cache or a proxy server cache close enough to the end user. However, WWW documents are becoming increasingly dynamic (i.e. have short lifetimes) which limits the potential benefits of caching. The performance of a WWW caching system can be increased by integrating document prefetching into its design. Prefetching attempts to overcome these limitations by pro-actively fetching content before users actually request it.

The rest of the paper is organized as follows: Section 2 discusses basic concepts of prefetching. Section 3 deals with various web prefetching techniques. Section 4 provides some conclusion remarks.

II. Basic Concept

Web prefetching is a technique which reduces the web latency by predicting the web objects that are to be accessed by the user in future and prefetching those objects before user actually demands for them. This technique takes advantage of the spatial locality shown by the web objects.

Prefetching systems are usually based on the basic web architecture. The basic web architecture [1] is characterized by web clients, that is, the software employed by users to access the web and web servers, which contain the information that users demand. This technique is implemented by means of two extra elements:

A. *Prediction Engine:* It is the part of the prefetching system aimed at guessing the user's future accesses. The output of the prediction engine is the hint list, which is composed of a set of URLs which are likely to be requested by the user in subsequent requests.

B. *Prefetching Engine:* Depending on the parameters like available bandwidth, server load, prefetching engine decides which objects from the predicted hints to be prefetched. By prefetching the requests in advance, the user's waiting time is reduced when the

ISSN (Online) 2278- 5841
ISSN (Print) 2320- 5156

object is actually demanded. But without careful usage, Prefetching can increase the network traffic as well as server's load if its predictions are not accurate enough.

III. Web Prefetching Techniques

In literature, various web prefetching techniques have been proposed for improving the web performance. This section surveys a large variety of recent techniques in this area. These techniques can be categorized as follows:

A. Markov Model Based Approach

In [2] access time and frequency based page rank like algorithm (TFPR) is proposed for conducting web page predictions which works on Markov chain model. These two factors, time length and frequency, are used to bias the page rank algorithm to give a higher ranking to the pages that are visited for a longer time and more frequently than others. Firstly this approach extracts the navigational path followed by the user and makes a graph G. Then it expands this graph G by including all those pages that point to the pages already in G and finds the weighted links between them. The length of the path of the sub graph depends on the order of the Markov model used. Until it reaches this predefined depth, it follows the same process for the newly included pages in the same graph. Then TFPR algorithm computes the local ranking for the pages in the sub graph and provide prediction list to the current user in decreasing order based on the ranking values. In this paper, authors used 1st order Markov chain model to expand the sub graph which is memory less to declare the weights in the directed graphs and to expand the sub graph.

Another Page Rank like algorithm [3] based on higher order Markov chain model is proposed to improve the 1st order Markov chain model [2]. Authors showed the experimental results that the proposed approach yields better prediction accuracy as compare to existing approach [2].

B. Prediction by Partial Match based approach

The standard prediction by partial match model [4] uses multiple Markov models to store historical URL paths. Therefore Std PPM model occupies more space and hence another model named Longest Repeating Sequence (LRS) [5] was derived from the standard PPM model that keeps the longest repeating subsequences of URL's that are frequently accessed. In both cases URL is recorded multiple times. Therefore it consumes more memory.

In [6] two models named Memory efficient PPM (MePPM) and memory efficient LRS (MeLRS) are proposed which are based on the access patterns of the users. The MePPM is constructed with the first accessed URL as the child of the root. Subsequently the next URL's are the child node of the previous URL in addition to a link to it from the root. Thus it uses a single n-order Markov tree to store URL Paths where n is the number of URL's accessed in a session. Using MePPM, the proposed MeLRS model corresponding to LRS model is reconstructed in which less number of nodes are required compared to LRS. The advantage of two proposed models is that these models utilize memory efficiently in addition to supporting prefetching.

C. Double Dependency Graph based Approach

In [7] the intelligent web prefetching mechanism is proposed which works on DDG algorithm [8]. In this paper author argues that despite the benefits of prefetching it can increase the network traffic, if not implemented carefully. An aggressive prefetching can overload the web server by generating extra prefetched requests. Thus it can cause extra traffic on the overall network and hence can decrease the system performance. In this paper authors presented a new mechanism which adjusts dynamically the aggressiveness of the prediction algorithm according to the server load. Instead of using a threshold or a fixed number of prefetching requests, this mechanism generates more or less prefetching requests according to the system performance and server load. Thus it controls the traffic and server load caused by prefetching.

D. Compression algorithms based approach

An online prediction model based on LZ78 and LZW algorithms is proposed in [24]. These algorithms are lossless compression algorithms and also used in sequence mining. This online prediction model does not need to preprocess the historical data to make a prediction model which is time consuming also. Thus it reduces the computational complexities. LZ78 and LZW construct the dictionary using the sequences, which is used to create prediction tree. Using LZ78 the node counts in prediction tree are less but some sequences can be lost which are not inserted in tree. Because of this, LZ78 hit rate is less but its memory efficiency is better. In LZW algorithm, dictionary contains all alphabets that are in the sequence. Therefore LZW has best precision and hit rate and memory consumption is also less than other models. Author argues that LZW algorithm is the best if precision, memory efficiency and running speed together are considered.

E. Sequential mining based approach

Sequential mining approach is used to extract frequent web access patterns from the log of server data. In [9] author argues that only those access sequences are meaningful in which user spends his most of the time. Session of thirty minutes is taken for finding the access

ISSN (Online) 2278- 5841
ISSN (Print) 2320- 5156

sequence of each user in longer sessions. Thus the web access found in all sessions for each user makes the web access sequence data (WASD). Then it filters out the less frequent access sequences. Now when a user requests for a page, it checks whether request is frequent or not by analyzing the frequent sequences found in WASD. If match is found then all frequent pages of that request are prefetched otherwise no prefetching is done. As the cache size is limited, the authors used various cache replacement algorithms such as Least Recently used (LRU) and Least frequently used(LFU) whenever required.

F. Clustering Based Approach

In this paper [10] author proposed an approach based on rough set clustering which is used to form the clusters of sessions. It classifies uncertain, imprecise or incomplete information in terms of data acquired from past experience. Using rough set clustering only meaningful sessions are obtained in which user spends his quality time. In this paper, author presented an algorithm called RST which uses the concept of rough sets to calculate equivalence between objects and then finds lower approximation and upper approximation. Lower approximation is the union of all equivalence objects which are contained in the target set, which is generally supposed by the user. The upper approximation is the union of all equivalence objects which have non-empty intersection with target set. Authors proposed the concept of PPE (Prediction Prefetching Engine) which resides at proxy server. When user requests for a page it matches that request with existing rough set clusters and then decides whether to prefetch the page or not. By clustering using RST, only the meaningful sessions of Web log are fed to rule generator phase of PPE and thus the complexity of PPE is also reduced.

An another approach [11] which integrates the clustering with distance measure technique, SAM[13] and SABDM[14] , which is based on sequence alignment and is used to compute the similarities between two clusters. In the proposed model, firstly user sessions are created based on IP address, date and time using the web access logs. Then clusters of these sessions are formed by using modified k-means algorithm [12]. Whenever the user requests for a page, the nearest cluster to the requested page is determined by measuring the distance with all cluster centers. Then it retrieves the next page in the cluster and counts the number of sessions in which this page is present followed by the requested page in the cluster. Based on frequency, top n pages are selected for the prediction list.

Another approach [22] presented an optimized predictive prefetching technique based on clustering. Using web log data file, it makes clusters of similar pages. Prediction algorithm works on these clusters. Author has compared the results of proposed technique with existing technique [23] for enhancing his previous technique. Overall performance of this technique is calculated by the summation of percentage of each web object. In previous technique[23] this percentage of each web object is equal to number of request sessions accessed by the user to the total number of web objects in the web logs .But proposed technique optimized this prediction by considering frequency of each predicted cluster to calculate the percentage of each web object. The association rules defined for proposed technique showed the frequency for each page's usage by user in each cluster. The proposed technique provided the results with higher probability of prediction accuracy.

G. Integrated Approach

In [15] author used an integrated approach for the prediction of web page based upon web logging and sequential rank based selection technique. This research focused on how to enhance the overall performance of web prefetching mechanism and how to improve the efficiency of web resources' usage. In the proposed approach, web logging is used to improve the accuracy of accessing the web document and sequential rank based selection techniques used to optimize the prediction for clustered accesses. In this approach, when a user requests for a web page, first the prediction engine finds the similar pages from the web logs of a server requested by user and makes a cluster of these similar pages. Different pages are clustered in different groups according to their groups and URL addresses. Then it finds out the frequency of each user's request from clusters. Then sequential rank based selection technique [16] finds the cluster which has maximum probability for prefetching a page for the user.

H. Content based Approach

In [21], authors suggest that the prediction of user's accesses can be improved if the algorithm considers the organization of the web space. Therefore, this subset of prefetching techniques is that where predictions are made from the analysis of the content of recently visited web pages to find links clicked with high probability instead of using only the sequence of accesses. This category of prefetching techniques is known as semantic prefetching. This semantic prefetching makes efficient predictions of web objects for satisfying user's subsequent requests.

A keyword-based semantic prefetching approach was designed by Cheng and Ibrahim [19, 20] that could predict future requests based on semantic preferences of past retrieved Web documents. The scheme was evaluated by considering the Internet news services. Neural network

was applied over the keyword set to predict future requests.

P. Venketesh [17] proposed semantic prefetching scheme that uses anchor texts presented in the web page to make effective predictions. Naïve Bayes Classifier was applied to compute the probability values of anchor texts, based on which the predictions were generated.

In [18], author's proposed fuzzy logic based web prefetching scheme that generates effective predictions for satisfying user's requests. Predictions are generated based on the anchor text information associated with hyperlinks in a web page. Based on the user's browsing pattern in each session, prediction engine dynamically computes the value and generates the list of predictions.

IV. CONCLUSION

Prefetching techniques are applied to reduce the latency, control the network traffic and server load and improve the user satisfaction. Lot of research has been done in the area of web prefetching in various directions. In this paper various web prefetching approaches have been discussed. Markov model and prediction by partial match based approaches utilize memory efficiently. Double dependency graph based approach mainly focuses on the network traffic and server load. Sequential mining, clustering and content based approaches mainly focus on the precision. Compression algorithm based approach considers both the precision and memory efficiency.

REFERENCES

[1] G. H. Kuenning and G. J. Popek, "Automated hoarding for mobile computers," In: Proceedings of the 16th ACM symposium on operating system principles, Malo France, pp. 264-275, October 1997.

[2] Y. Z. Guo, K. Ramamohanarao, and A. F. Park, "Personalized PageRank for Web Page Prediction Based on AccessTime-Length and Frequency," International Conference on Web Intelligence, July 2007.

[3] R.Khanchana and Dr. M. Punithavalli, "An Efficient Web Page Prediction Based on Access Time-Length and Frequency," IEEE Computer, August 2011

[4] T. Palpanas and A. Mendelzon, "Web prefetching using partial match prediction," Proceedings of web caching workshop, San Diego California, March 1999.

[5] X. Chen and X. Zhang, "A Popularity-Based Prediction Model for Web Prefetching," IEEE Computer, Vol. 36 No. 3, pp. 63-70, 2003.

[6] C. D. Gracia and S. Sudha, "MePPM- Memory Efficient Prediction by Partial Match Model for Web Prefetching," IEEE Computer, 2012.

[7] J. Marquez, J. Domenech, J. A. Gil, and A. Pont, "An intelligent technique for controlling web prefetching costs at the server side," International Conference on Web Intelligence and Intelligent Agent Technology, IEEE Computer, 2008.

[8] J. Domenech, J. A. Gil, J. Sahuquillo, and A. Pont, "DDG: An efficient prefetching algorithm for current web generation," In Proc. of the 1st IEEE Workshop on Hot Topics in Web Systems and Technologies, 2006.

[9] A. Singh and A. K. Singh, "Web Pre-fetching at Proxy Server Using Sequential Data Mining," Third International Conference on Computer and Communication Technology, 2012.

[10] Ms. Jyoti, Dr. A. K. Sharma, Dr. Amit Goel, and Ms. Payal gulati "A Novel Approach for clustering web user sessions using RST" International Conference on Advances in Computing, Control, and Telecommunication Technologies, IEEE Computer, 2009.

[11] G. Poornalatha and P. S. Raghavendra, "Web Page Prediction by Clustering and Integrated Distance Measure" International Conference on Advances in Social Networks Analysis and Mining, IEEE Computer, 2012

[12] G. Poornalatha and P. S. Raghavendra, " Web User Session Clustering Using Modified K-means Algorithm," First International Conference on Advances in Computing and Communications (ACC – 2011), CCIS(191),Springer-Verlag, pp.243-252, 2011

[13] B. Hay, G. Wets, and K. Vanhoof, "Mining Navigation Patterns Using a Sequence Alignment Method," Journal of Knowledge and Information Systems, Springer-Verlag, pp. 150-163, 2004

[14] G. Poornalatha, P. S. Raghavendra, "Alignment Based Similarity Distance Measure for Better Web Sessions Clustering," Journal of Procedia CS, vol.5, pp. 450-457, 2011 .

[15] N. Ahmad, O. MaliIk, M. Hassan, M. S. Qureshi, and A. Munir, "Reducing User Latency in Web Prefetching Using Integrated Techniques" IEEEComputer, 2011.

[16] B. Parhami "Introduction to Parallel Processing Algorithms and Architectures", Kluwer Academic Publishers New York, Boston, pp.111-112, 2002.

[17] P.Venketesh, R Venkatesan, and L. Arunprakash, "Semantic Web Prefetching Scheme Using Naïve Bayes Classifier," International Journal of Computer Science and Applications, Vol. 7, No. 1, pp. 66 – 78, 2010.

[18] P.Venketesh and R. Venkatesan "Adaptive Web Prefetching Scheme Using Link Anchor Information," International Journal of Applied Information Systems, Vol. 2, No. 1, May 2012.

[19] T. I. Ibrahim and C. Xu, "Neural net based predictive pre-fetching to tolerate WWW latency," Proceedings of the 20th International Conference on Distributed Computing Systems, 2000.

[20] C. Z. Xu and T. I. Ibrahim, "A Keyword-Based Semantic Prefetching Approach in Internet News Services," IEEE Transactions on Knowledge and Data Engineering, Vol. 16, No. 5, pp.601 -611, 2004.

[21] J. I. Khan and Q. Tao, "Exploiting Webspace Organization for Accelerating Web Prefetching," In Proceedings of the

ISSN (Online) 2278- 5841
ISSN (Print) 2320- 5156

IEEE/WIC International Conference on Web Intelligence, Halifax, Canada, 2003.

[22] N. Ahmad, A. Khan, and F. Bibi "Optimizing Predictive Prefetching in Multi-Client Single-Server Environment," Frontiers of Information Technology, IEEE Computer, 2011

[23] Q. Yang, H. H. Zhang, and T. Li, "Mining Web Logs for Prediction Models in WWW Caching and Prefetching," International Conference on Knowledge Discovery and Data mining in proceedings of the seventh ACM SIGKDD, San Francisco, California, USA, pp. 473-478, 2001.

[24] A. Moghaddam and E. kabir, "Dynamic and memory efficient web pageprediction model using LZ78 and LZW algorithms" Proceedings of the 14th International CSI Computer Conference, IEEE Computer, 2009.