International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 10, ISSN No. 2455-2143, Pages 76-80
Published Online August - September 2016 in IJEAST (http://www.ijeast.com)

PERFORMANCE IMPROVEMENT BY COMBINING INTELLIGENT WEB CACHING TECHNIQUES WITH SPATIAL WEIGHT MATRIX TECHNIQUE FOR WEB PRE-FETCHING

V.Vijilesh
Associate Professor,
Department of IT,
Kumaraguru College of Technology, India

Dr. R. Nedunchezhian
Professor,
Department of CSE,
Kalaignar Karunanidhi Institute of Technology, India

Dr. K.R. Baskaran
Associate Professor,
Department of IT,
Kumaraguru College of Technology, India

Abstract: Web caching and web pre-fetching techniques are the commonly used techniques for solving the internet related problems. By caching the frequently used data, web caching technique helps in efficient usage of network band width and reduces the network latency. Web pre-fetching technique preloads the web objects that are expected to be requested in the future and caches them. The above two techniques help in increasing Hit Ratio (HR) and Byte Hit Ratio (BHR). In this work, it is demonstrated that integration of spatial weight matrix technique for web pre-fetching with intelligent web caching techniques will further increase the HR and BHR. It is also demonstrated that SVM-LRU combination with pre-fetching technique shows good performance in terms of HR and BHR in comparison with other caching techniques.

Keywords: Machine Learning • Spatial Weight Matrix • SVM • Caching • Pre-fetching.

I. INTRODUCTION

With rapid advances in the technology, Web users are able to make the best use of the enormous volume of information obtainable over the internet. Still the web users are experiencing greater access delay due to the exponential increase of web users and web traffic [1]. By using the web caching and web pre-fetching techniques, individually, it is possible to minimize the access delay that is experienced by the web users. To further minimize this access delay, it is recommended to combine intelligent web caching and techniques based on web pre-fetching.

II. LITERATURE SURVEY

A. Web caching and Web pre-fetching Techniques

Temporal locality and spatial locality are the properties made use by Web caching and Web pre-fetching techniques respectively in order to reduce access latency and predict future requests of web users. HR and BHR are the two common metrics used in accessing web caching techniques. Cache pollution may occur when non-intelligent page replacement techniques are used. Therefore intelligent web caching techniques need to be used to avoid this situation.

B. Intelligent Machine Learning Techniques

There is a need for substantial training data, which can be collected from users log files when intelligent caching techniques have to be deployed. The source for training data is the contents of users log files. The literature survey on intelligent web caching techniques is listed as below:

Datta et al [1] proposed methods those results in lesser access delays. Mookherjee et al [2] proposed a structure for evaluating LRU cache replacement method for different sizes of cache. Back-propagation neural networks for effective cache replacement were analyzed by Jake Cobb and Hala ElAarag [3, 4]. Calzarossa [5] suggested that cache replacement can be done by fuzzy logic. Farhan [6] has experimented usage of BPNN (back-propagation neural network) for LRU cache replacement.

Yang et al [7], experimented the integration of web caching and web pre-fetching techniques using model based predictive pre-fetching technique. Craswell et al [8], analyzed the effective ranking method based on content, in finding the future requests of web users. Cheng-Zhong Xu et al [9], proposed Keyword-based semantic pre-fetching technique where future prediction of user requests are based on the usage of past web objects. Georgakis and Li [10], highlighted the importance of textual information in the pages that are browsed to know about the user's preference.

III. PROPOSED METHODOLOGY

Log file contents are used in identifying access patterns of web users. Log file contents are reprocessed and then the web objects are classified into class 0 and class 1 using the factors: object size, recency, retrieval time and frequency. Clustering of web objects accessed by a web user is done for every 30 minutes time interval. Spatial weight matrix method is used for clustering technique. The chosen 30 minute time interval neither overloads the server nor increases the pre-fetching time. Intra-site clustering technique, in which clustering of all the browsed pages of a URL in the specified time interval is used in this research. Frequently browsed web objects for every user are identified using the support and confidence parameters.

Cache memory is split into two partitions. The first partition is called as short-term cache memory and the second partition is called as long-term cache memory. Sixty seven percentage of the total space is assigned to short-term cache memory and thirty three percentage of the total space is assigned to long-term cache memory [11, 12, 13]. Web objects that are pre-fetched are loaded into the short-term cache. Access count of the web object found in the short-term cache is incremented by one, when the user requests it. All other web objects that are present in that user's cluster are pre-fetched and put into the short-term cache. When this count reaches the threshold value the web object is classified using SVM/Bayesian classier/Neuro-fuzzy techniques. Those objects which are classified as class 1, are moved to the top of the long-term cache and other objects are shifted to the rear of the long-term cache.

When the long-term cache does not have the required space, adequate class 0 web objects are removed and shifted to short-term cache. When the short-term cache does not contain required space the web objects are deleted using LRU method until adequate space is created.

A. Dataset

The dataset for testing is obtained from the URL: http:\\ita.ee.lbl.gov/html/contrib/NASA-HTTP.html. A total of about 600452 items, approximately 46 GB of dataset, about 387,543 items were used for clustering, approximately 34 GB of bytes were used for clustering, about 167,801 items were used for testing, and approximately 11 GB for testing.

IV. EXPERIMENTAL RESULTS

Computed HR and BHR for different values of support and confidence combinations are plotted in the graphs shown below in Figures 1(a) to 2(d):

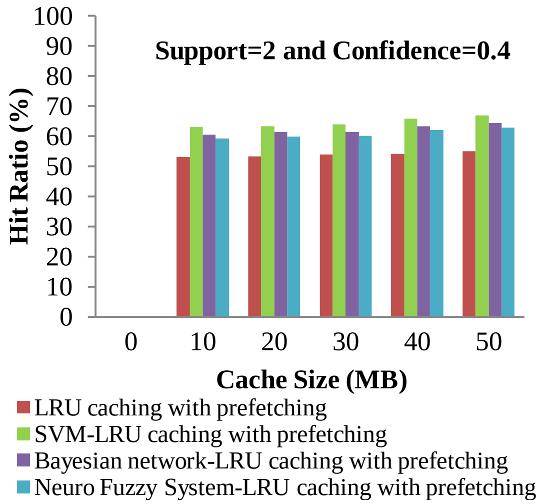

Figure 1(a). Significance of HR considering various intelligent web caching techniques combined with spatial weight matrix web pre-fetching for Support=2 and Confidence= 0.4

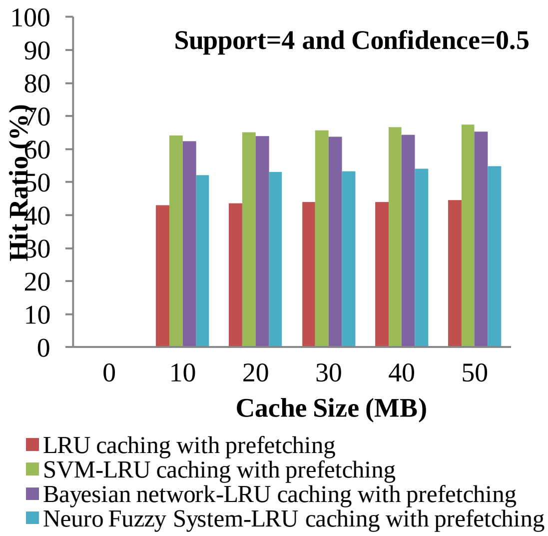

Figure 1(b). Significance of HR considering various intelligent web caching techniques combined with spatial weight matrix web pre-fetching for Support=4 and Confidence= 0.5

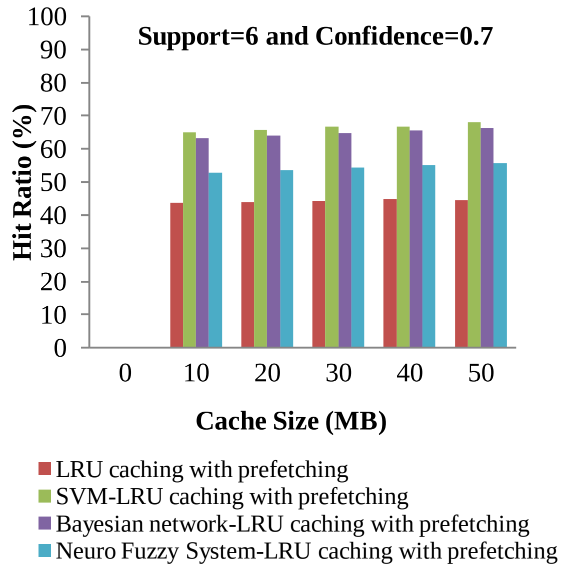

Figure 1(c). Significance of HR considering various intelligent web caching techniques combined with spatial weight matrix web pre-fetching for Support=6 and Confidence= 0.7

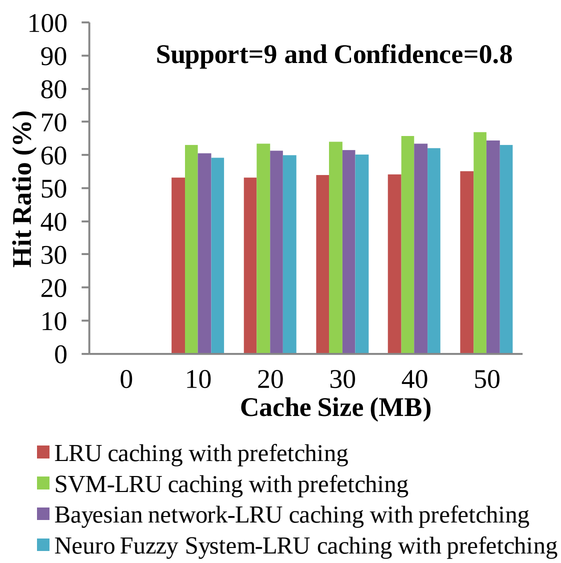

Figure 1(d). Significance of HR considering various intelligent web caching techniques combined with spatial weight matrix web pre-fetching for Support=9 and Confidence= 0.8

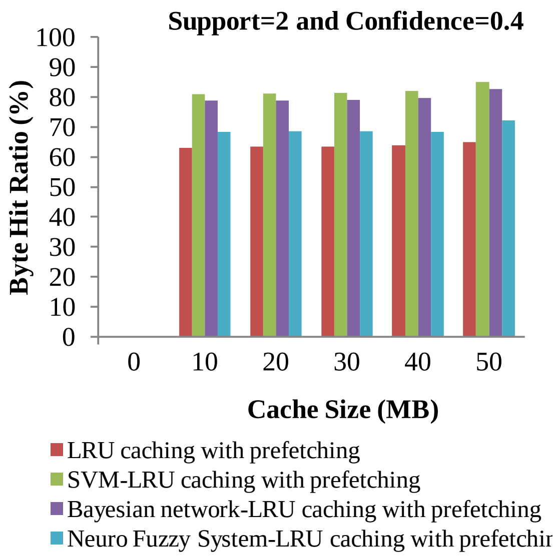

Figure 2(a). Significance of BHR considering various intelligent web caching techniques combined with spatial weight matrix web pre-fetching for Support=2 and Confidence= 0.4

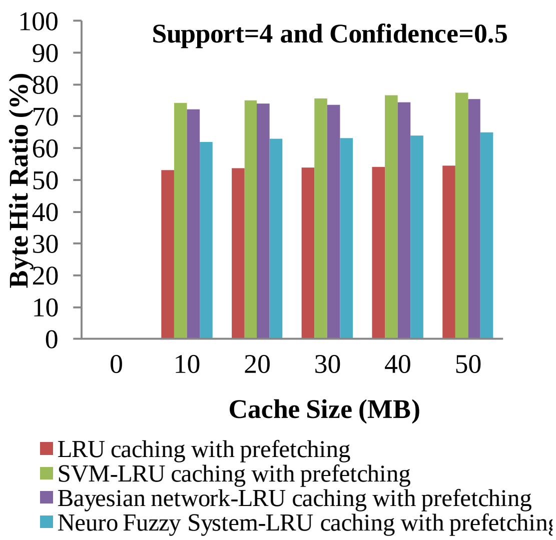

Figure 2(b). Significance of BHR considering various intelligent web caching techniques combined with spatial weight matrix web pre-fetching for Support=4 and Confidence= 0.5

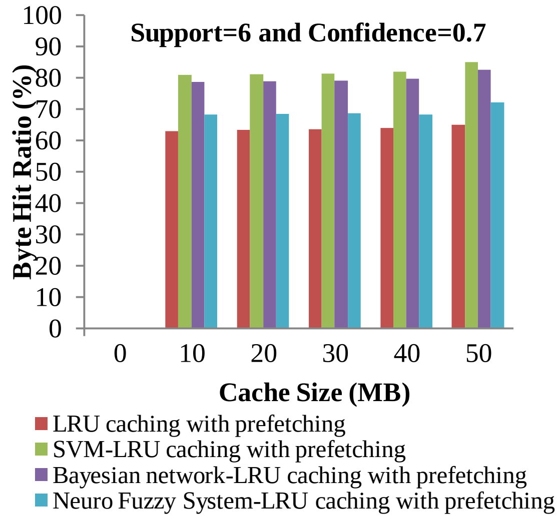

Figure 2(c). Significance of BHR considering various intelligent web caching techniques combined with spatial weight matrix web pre-fetching for Support=6 and Confidence= 0.7

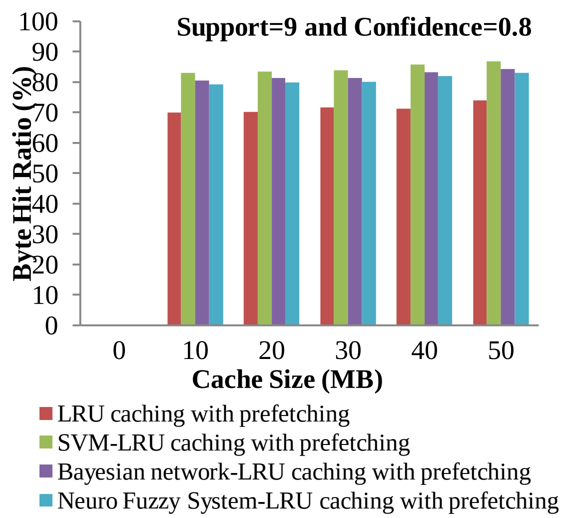

Figure 2(d). Significance of BHR considering various intelligent web caching techniques combined with spatial weight matrix web pre- fetching for Support=9 and Confidence= 0.8

The result obtained for HR and BHR by considering different set of values for Support and Confidence is given in Table 1.

Table 1. Comparison of HR and BHR achieved for various values for Support and Confidence parameters using combined Intelligent web caching and Spatial weight matrix web pre-fetching

Machine Learning Techniques			LRU	Neuro-fuzzy-LRU	Bayesian network-LRU	SVM-LRU
Support (S) and Confidence values (C)	S=2, C=0.4	HR (%)	43.8	49.4	59.8	62.1
		BHR (%)	44.3	66.4	64.8	63.4
	S=4, C=0.5	HR (%)	43.9	53.4	63.9	65.8
		BHR (%)	53.8	75.8	74.0	64.3
	S=6, C=0.7	HR (%)	44.3	54.3	64.8	66.4
		BHR (%)	63.8	82.1	79.8	69.2
	S=9, C=0.8	HR (%)	53.9	60.8	65.2	67.6
		BHR (%)	71.4	84.6	82.2	80.8

From the Table 1, it is inferred that SVM-LRU web caching technique combined with spatial weight matrix based web pre-fetching technique leads to increased HR and BHR in comparison with LRU and other intelligent web caching techniques combined with web pre-fetching technique.

V. CONCLUSION

In this research work, for the chosen dataset, it is evident that by integrating spatial weight matrix pre-fetching technique with SVM-LRU caching technique leads to increased HR and BHR compared to other techniques.

VI. REFERENCES

1. A. Datta, K. Dutta, H. Thomas, D. VanderMeer, 2003. World wide wait: a study of Internet scalability and cache-based approaches to alleviate it. Management Science, 49 (10), 1425–1444.
2. V.S. Mookherjee, Y. Tan, 2002. Analysis of a least recently used cache management policy for web browsers. Operations Research 50(2), pp: 345–357.
3. J. Cobb, H. ElAarag, 2008. Web proxy cache replacement scheme based on back-propagation neural network. Journal of System and Software, 81(9), pp: 1539-1558.
4. J. Cobb, H. ElAarag, 2006. Training and simulation of neural networks for web proxy cache replacement. Proceedings of the International Symposium on Performance Evaluation of Computer and Telecommunications Systems (SPECTS2006), Summer Simulation Multi-conference, Calgary, Canada, pp: 279-298.
5. Calzarossa, V.G., 2003. A Fuzzy Algorithm for Web Caching. Simulation Series Journal 35(4), pp: 630–636.
6. Farhan, 2007. Intelligent Web Caching Architecture. Master thesis, Faculty of Computer Science and Information System, UTM University, Johor, Malaysia.
7. Qiang Yang, Zhen Zhang, 2001. Model Based Predictive Pre-fetching. Proceedings of IEEE International Workshop on Database and Expert Systems Applications.
8. N. Craswell, D. Hawking, S.E. Robertson, 2001. Effective Site Finding Using Link Anchor Information. Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval.
9. Cheng-Zhong Xu, Tamer I. Ibrahim, 2004. A Keyword Based Semantic Pre-fetching Approach in Internet News Services. IEEE Transactions on Knowledge and Data Engineering, 16(5), pp: 601 -611.
10. A. Georgakis, H. Li, 2006. User behavior modeling and content based speculative web page pre-fetching. Data and Knowledge Engineering - Elsevier, 59, pp: 770 - 788.
11. Baskaran, KR, Kalaiarasan, C, 2014. Pre-eminence of Combined Web Pre-fetching and Web Caching Based on Machine Learning Technique. Arabian Journal for Science and Engineering, 39(11), pp: 7895-7906.
12. Baskaran, KR, Kalaiarasan, C, 2016. Improved Performance by Combining Web Pre-fetching Using Clustering with Web Caching Based on SVM Learning Method. International Journal of Computers Communications & Control, 11(2), pp: 167-178
13. Baskaran, KR, Kalaiarasan, C 2014, Improving Hit Ratio and Byte Hit Ratio Using Combined Pre-fetching and Web Caching, International Review on Computers and Software, vol. 9, no. 8, pp. 1426 -1433.