IRJSE *Int. Res. J. of Science & Engineering,* 2020; Special Issue A7: 829-833 SJIF Impact Factor 6.68
ISSN: 2322-0015
RESEARCH ARTICLE OPEN ACCESS

SVM Based Implementation of Intelligent Web Caching

Patelpaik HB

Lokmanya Tilak Mahavidyalaya, Wani, Yavatmal, MS, India
Email: hbpatelpaik@rediffmail.com

Manuscript Details
Available online on http://www.irjse.in
ISSN: 2322-0015

Cite this article as:
Patelpaik HB. SVM Based Implementation of Intelligent Web Caching, *Int. Res. Journal of Science & Engineering,* February, 2020, Special Issue A7 : 829-833.

ABSTRACT
The World Wide Web contains huge amount of information almost on every subjectsuch as education, environment, defense, business, sports, medical science, banking, shopping etc. Increasing popularity of World Wide Web has introduced new issues such as Internet traffic, bandwidth consumption thereby leading to latency in service being provided by the application servers. Despite technological advances this huge traffic can lead to considerable delays in accessing objects on the web. Caching of objects in the World Wide Web is widely used technique to reduce network traffic, latency and server load. Caching involves storing copies of objects at locations that are relatively closer to the user. Caching plays a vital role to increase the performance of web sites. Various caching policies are available such as Least-Recently-Used (LRU), Least frequently used (LFU), Greedy-Dual-Size (GDS) and Greedy-Dual-Size-Frequency (GDSF).etc. Traditional caching policies do not perform well in every situation of caching the web object.Performances of conventional Web proxy caching policies are increased dramatically using Machine learning techniques. Support Vector Machine or SVM algorithm is a simple and powerful Supervised Machine Learning technique used for building both regression and classification models. Intelligent caching approaches are designed by combining conventional caching policies LRU and GDSF with support vector machine (SVM) and termed as SVM–LRU, SVM–GDSF. These models are tested through trace-driven simulation. The results are compared with Web proxy caching polices based on traditional techniques. The observations revealed that the performance is significantly improved due to the proposedSVM–GDSF, SVM–LRUtechniques.

Key words: - WWW, Band-Width, Latency, Web caching, LRU, LFU, Machine learning, SVM

INTRODUCTION

The WWW has become an essential tool for accessing information and interaction among people on the globe.It provides a wide range of Internet-based services including shopping, banking, entertainment, education, governance, etc. WWW traffic has been increasing rapidly due to web applications on every web site. The fast growth of online applications is making network traffic congested and increase serve load which delays response, thereby annoying the web user. Caching the popular web objects, is an efficient solution to the latency problem which bringdocuments closer to the web users. The web objects can be cached at various placesat WWW:at the client browserat or near the server (reverse proxy) to reduce the server load, or at a proxy server. The proxy cachingserver's double purposes: first, it reduces the access latency for a document; second, it reduces the amount of "external" traffic that is transported over the wide-area network, which also reduces the user's perceived latency. Generally, a proxy cache have limited storage in which it stores "popular" documents. Whenever the cache is full and theproxy needs to cache a new document, it has to decide which document to evict from the cache to accommodate the new document. The policy used for the eviction decision is referred to as the *replacement policy*. Traditional caching policies, such as LFU, LRU, SIZE, GDSF, etc. for memory systems do not necessarily perform well when applied to WWW traffic for the following reasons:

- A large amount of caches deal with fixed-size pages in memory system, so the size of the page does not play any role in the replacement policy. In contrast, Web objects are of variable size, and web object size can affect the performance of the caching policy.
- The cost of missed Web objects from original servers depends on several factors, such as distance, bandwidth between the proxy and the original servers and the size of the web object. Such dependence does not exist in traditional memory systems.
- Web objects areregularly updated, which means that it is very important to consider the web object expiration period at replacement instances. In memory systems, pages are not generally associated with expiration period.
- Web object popularity needs to be considered in any Web caching policy to optimize a desired performance metric but it has not been considered in memory system.

Web proxy caching policies have attempted to combine some factors which can affect the performance of Web proxy caching for making decisions about caching. It is not an easy task, because one factor in a particular environment may bemore important in other environments [4, 5]. WhichIdealWeb objectswill be re-visited is still a major challenge faced by the existing Web proxy caching techniques. In other words, there is the question of which Web objects should be cached and which Web objects should be replaced to make the best use of available cache space, improve hit rates, reduce network traffic, and alleviate loads on the original server [4, 7, 8]

In a Web proxy server, Web proxy log files record the activities of the users and can be considered to contain complete and prior knowledge of future accesses. Using log records machine learning algorithm decides which web objects will be revisited or not.

Recent studies have proposed machine learning techniques to cope with the above problem [9, 10].Most of these studies utilize an artificial neural network (ANN) in Web proxy caching, although ANN training may consume considerable amounts of time and require extra computational overheads. More importantly, the integration of intelligent techniques in Web cache replacement is still being researched.

Machine learning algorithm which can be used for both classification and regression challenges. Machine learning algorithms have a wide range of applications such as text classification, finding, extracting and summarizing relevant data, making predication based on the analysis data, Web page classification, and bioinformatics application.

1. Web Proxy Caching:

Web caching is one of the most successful solutions for improving the performance of Web-based systems.

In Web caching, the popular Web objects that are likely to be used in the near future are stored on devices closer to the Web user such as client's machine or proxy server. Thus, Web caching has three attractive advantages to Web users. Web caching decreases user perceived latency, reduces network bandwidth usage and reduces load on the origin servers. Typically, a Web cache is located in a browser, proxy server and/or origin server as shown in Fig. 1. The browser cache is located in the client machine. At the origin server, Web pages can be stored in a server-side cache for reducing the redundant computations and the server load.

The proxy cache is found in the proxy server, which is located between the client machines and origin server. It works on the same principle as the browser cache, but on a much larger scale. Unlike the browser cache which deals with only a single user, the proxy server serves hundreds or thousands of users in the same way. As shown in Fig. 1, when a request is received, the proxy server checks its cache. If the object is available, the proxy server sends the object to the client. If the object is not available, or it has expired, the proxy server will request the object fromthe origin server and send it to the client. The requested object will be stored in the proxy's local cache for future requests.

Web proxy caching is widely utilized by computer network administrators, technology providers, and businesses to reduce both user delays and Internet congestion [11, 8].

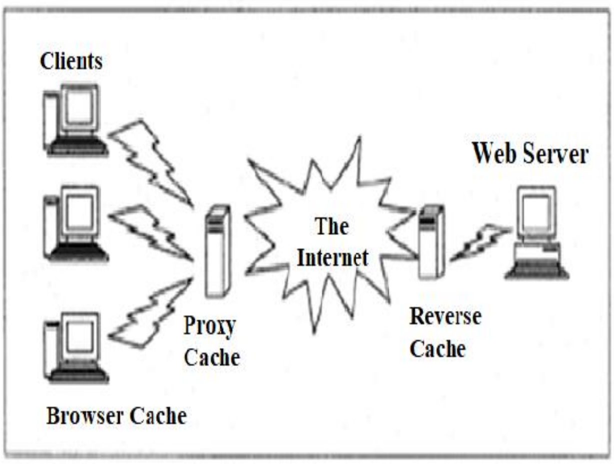

Fig 1: Web object caching positions

In this study, much emphasis will be placed on Web proxy caching, due to the fact that it is the most common strategy used for caching Web pages.

2. Machine learning

Machine Learning is a sub-area of artificial intelligence, whereby the term refers to the ability of IT systems to independently find solutions to problems by recognizing patterns in databases. Machine Learning enables to recognize patterns on the basis of existing algorithms and data sets and to develop adequate solution. In general, the machine learning algorithm takes the past examples as inputs, analyzes them, and outputs abstract patterns or rules. Thus, the machine learning mechanisms form the basis for adaptive systems [1].

1. Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. It performs classification more accurately and faster than other algorithms. These machine learning algorithms have a wide range of applications such as text classification, Web page classification and bioinformatics application. Hence SVM can be used to produce promising solution for web proxy caching.

In SVM, each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper plane that differentiate the two classes very well. Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line) as shown in Fig.2 below.

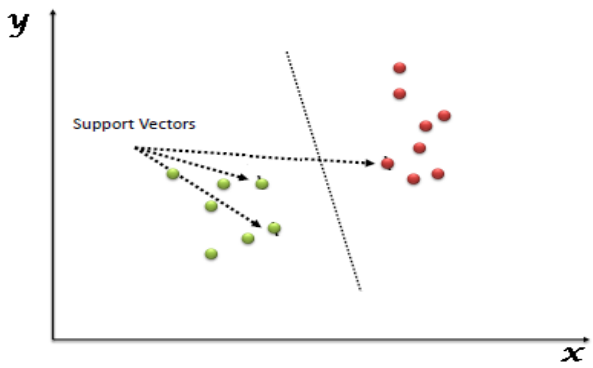

Fig. 2 – SVM-Segregated Hyperplane

Support vector machine or SVM algorithm is based on the concept of 'decision planes', where hyper planes are used to classify a set of given objects. Pictorial examples of support vector machine algorithm in Fig.3 shows that two sets of data. These datasets can be separated easily with the help of a line, called a **decision boundary**. There can be several decision boundaries that can divide the data points without any errors. The nearest points from the optimal decision boundary that maximize the distance are called **support vectors.**

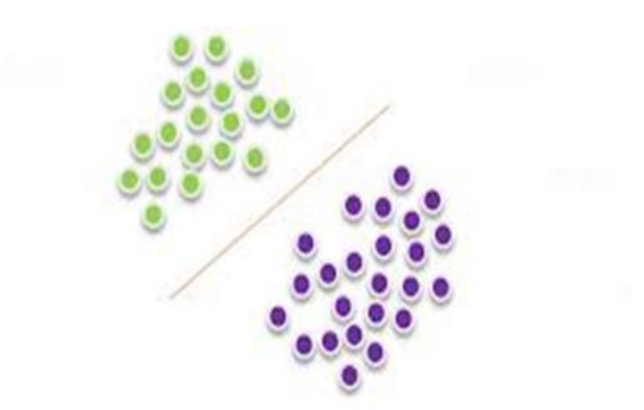

Fig 3. SVM Decision Boundary

2. The proposed model of Intelligent Web Caching

We are presenting here an Intelligent web caching model that is based on machine learning technique "support Vector Machine". The model uses SVM in collaboration with the traditional caching algorithms to identify the potential cacheable and non-cacheable web objects. The model consists of two major functional components termed as'Online Component' and 'Offline Component'. In the online component, when the user requests Web page, the user communicates with proxy directly for retrieving that page from proxy cache or from server as shown in Fig. 4. When the cache is full and a new Web object is fetched from the server, the proposed intelligent caching approach is used to identify unwanted Web objects for replacement. The intelligent caching approaches are executed in the online component.

Whereas, the Offline Component is responsible for training the machine to learn from past data when the proxy server is not busy. It does not interact with the clients directly to handle the users' requests. The trained classifiers is then used in Online Component to evict or cache proper web object.

In this paper,we present two intelligent Web proxy caching approaches known as SVM–GDSF and SVM–LRU.

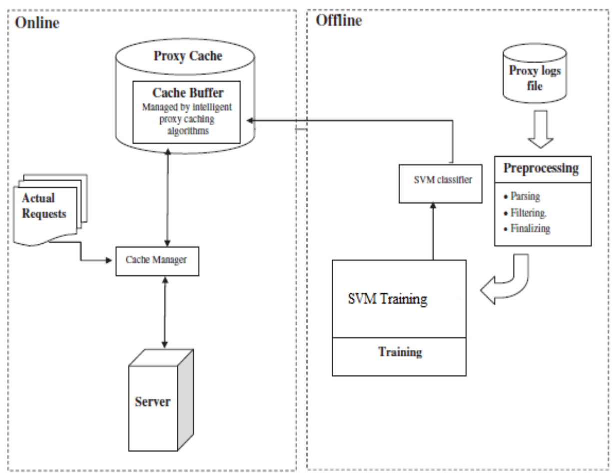

Fig 4.A Proposed intelligent Web proxy caching model

3. SVM-GDSF

Greedy dual-size frequency (GDSF) is a function based cache replacement policy.With this policy, cache performance can be optimized by selecting the appropriate weighting parameters. This policy is applicable to a variety of usage scenarios, as it can account for multiple influencing factors. However, the selection of appropriate weighting parameters is extremely challenging and new problems may arise from the computation of function values. The calculation cost is high and the parameter adjustment is complex.The byte hit ratio of GDSF policy is too low. Therefore, the SVM classifier is integrated with GDSF for improving the performance in terms of the byte hit ratio of GDSF. The proposed intelligent proxy caching approach is called SVM–GDSF.

In SVM–GDSF, a trained SVM classifier is used to predict the classes of Web objects either it may be re-visited later or not [6]. Thereafter, the classification decision is integrated with traditional cache replacement policy (GDSF) to give a key value for each object in the cache buffer, as shown in following equation. Consequently, the objects with the lowest values are removed first.

$K(g) = L + F(g) * C(g) / (S(g) + W(g))$

Where, W(g) represents the value of predicted class of object g, based on SVM classifier. W(g) will be assigned to 1 if object g is classified by SVM as an object to be re-visited, otherwise W(g) will be assigned to 0. The key value of object g is determined not by its past occurrence frequency, but also by the class predicted depending on the six factors such as URL ID, timestamp, elapsed time, size and type of Web object. The type of web object can be from HTML, image, audio, video application and other. The motivation behind the proposed SVM–GDSF approach is to enhance the priority of those cached objects that may be revisited in the near future, according to the SVM classifier, even if they are large in size or not accessed regularly..

7. SVM-LRU

Least Recently Used caching technique is one of the most common techniques in Web caching.Least Recently Used (LRU) Page Replacement Algorithm replaces the page in memory that has not been used for the longest period of time. The problem with this algorithm is the complexity in implementation. It suffers from caching pollution due to unpopular object will remain in a cache for a long time.

SVM classifier is combined with LRU to form a new algorithm called SVM–LRU. When the Web object g is requested by the user, SVM predicts whether the class of that object will be revisited again or not. If the object g is classified by SVM as an object to be re-visited again, the object g will be placed at the top of the cache stack. Otherwise, the object g will be placed in the middle of the cache stack. Hence, SVM–LRU can efficiently remove the unwanted objects at an early stage to make space for the new Web objects. By using this mechanism, cache pollution can be reduced and the available cache space can be utilized more effectively.

CONCLUSION

This work proposed two Intelligent Web proxy caching approaches, namely SVM–LRU and SVM–GDSF for improving the performance of the conventional Web proxy caching algorithms. SVM learn from Web proxy logs file to predict the classes of objects to be re-visited or not. Other intelligent classifiers can be utilized to improve the performance of traditional Web caching policies. Several intelligent Web proxy caching approaches canbe proposed to help in improving both the hit ratio and the byte hit ratio.

Conflicts of interest: The authors stated that no conflicts of interest.

REFERENCES

1. Intelligent Web Caching Using Machine Learning Methods, Sarina Suleman, Sitimariyan Shamsuddin, Ajith Abraham, Shahida Sulaiman
2. An Overview of Web Caching Replacement Algorithms Abdullah Balamash ,Marwan Krunz
3. Tutorial on Support Vector Machine (SVM) Vikramaditya Jakkula,
4. Pre-Fetching and Re-Fetching in Web Caching Systems: Algorithms and Simulation, H.T. Chen, Trent University, Peterborough, Ontario, Canada, 2008.
5. Web cache replacement policies: a pragmatic approach, IEEE, W. Kin-Yeung, Network 20 (2006) 28–34.
6. Intelligent Web proxy caching approaches based on machine learning techniquesWaleed Ali, Siti Mariyam Shamsuddin, Abdul Samad Ismail
7. Web proxy cache content classification based on support vector machine, W. Ali, S.M. Shamsuddin, A.S. Ismail, Journal of Artificial Intelligence 4 (2011)100–109.
8. Web cache optimization with nonlinear model using object features, T. Koskela, J. Heikkonen, K. Kaski, Computer Networks 43 (2003) 805–817.
9. Web proxy cache replacement scheme based on back-propagation neural network, J. Cobb, H. ElAarag, Journal of Systems and Software 81 (2008) 1539–1558.
10. Intelligent Web Caching Architecture, Farhan, Malaysia, 2007
11. An admission-control technique for delay reduction in proxy caching, C.C. Kaya, G. Zhang, Y. Tan, V.S. Mookerjee, Decision Support Systems 46 (2009) 594–603.