ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 1 (2014) No. 1, pp. 33-40

Mobile Web Cache Prefetching Technology Based on the Combination of Context and User Interest Degree

Cong Du[1], SuoZhu Wang[2*]
[1] College of Information Engineering, Capital Normal University, Beijing, China, 100048.
[2] School of Management, Capital Normal University, Beijing, China, 100089.
ducong101@gmail.com; wsz_wsz@163.com

(*Received Jan 2014, accepted Apr 2014*)

Abstract: Web prefetching technology can reduce user access latency and play an important role in improving the network quality and performance. In this paper, in order to improve the web performance, we analyze several common prefetching strategies and propose a prefetching model. By studying user interest degree model, the character of the mobile user's behavior and proxy server mode, we propose a prefetching algorithm which based on user interest degree and context information. Finally we design the experiment and the results show that the algorithm in improving prefetching performance has certain advantages.

Keywords: Web cache, prefetching technology, user interest degree, Context Information

1. Introduction

Nowadays, the rapid development of wireless networks makes it more convenient to access Internet. However, due to the rapid growth of network applications and mobile users, it causes serious network latency. Traditional web uses caching and prefetching technology to solve this problem. According to the principle of locality, it saves the resources, which user frequently access, into proxy server or client. When the mobile users access the same data again, it can avoid sending a new request to a remote server to reduce access latency.

According to the traditional web, mobile web can also use this way to dispose access latency. But mobile web also has many differences from traditional web. For example, mobile host frequently broken link, the limited resource of host (battery, CPU, etc).So the prefetching algorithm which mobile web used should not be complexity, cockamamie, and the obtained pages should not be too large.

For web prefetching mechanism, experts and scholars have made lots of research. For example, Padmanabhan V N proposed a hyperlink-based prefetching mechanism [1]. An algorithm based on access sequence model is proposed by Han and Tian [2] to predict access objects. By analyzing the access behavior of SNS users to predict the access objects is proposed by Lin and Zhang [3].Song and Shen use the DG (Dependency graph) to describe different web pages access mode [4]. Each node in the diagram represents a visited Web page. There are some others to use data mining techniques [5, 6]. According to the web access log records, establish association rules to predict access object sets. Data mining method is more suitable for the proxy server and server mode, the data which user focused on will be analyzed by the server and prefetch in advance. Venkataramani et al proposed a method which uses the Markov model [7]. Others prefetching mechanism are based on cost function. It takes the lifetime of web objects and bandwidth resources into account. For example, Good fetch prefetching algorithm [8] and APL prefetching algorithm [9].In addition, there are other methods which based on semantics and popularity to build prefetching model.

The above methods are commonly methods which used in traditional Web. These algorithms have their own advantages and shortcomings. But this paper studies the prefetching mechanism in mobile web. According to previous results on traditional web, this paper proposes a prefetching algorithm which based on the combination of context information and user interest degree.

2. Prefetching Model and Algorithm

2.1. Prefetching Model

The mobile web prefetching strategies can be classified as client prefetching, proxy prefetching and server prefetching according to the location. But this paper uses proxy server mode as a means to dispose the prefetching mechanism. Based on the traditional research results, this paper proposes a framework of a mobile web prefetching mechanism in Figure 1.The historical prediction model is established by mobile user access records and predict user access data

according to certain criteria. Then the prefetching control achieves the real prefetching (To update or keep the data sets).Finally, the prefetching control transfers the prefetching data sets to mobile user and feedback the prefetching parameter(hit rate)to prefetching model in order to predict more accurately.

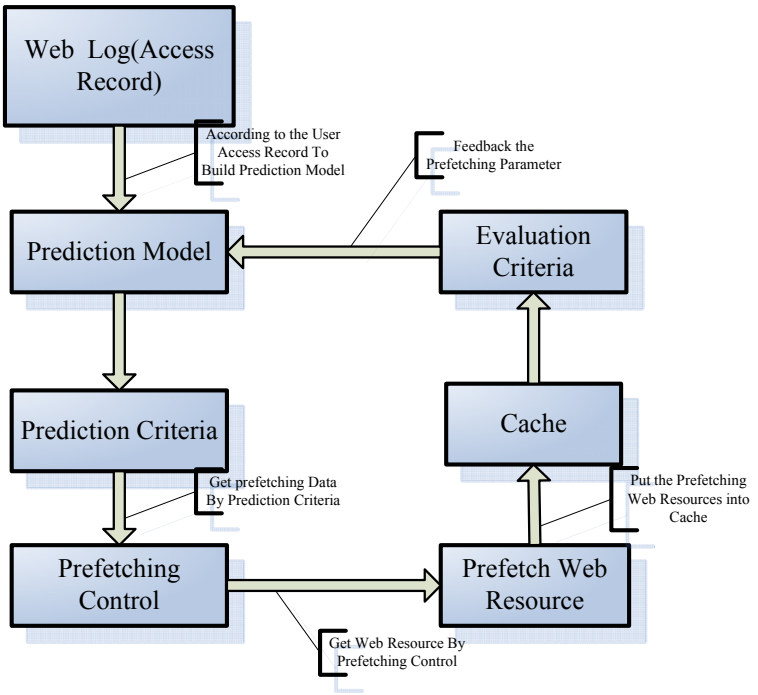

Fig. 1: Prefetching Model

Then how to evaluate different prefetching technologies by different evaluation criteria is also very important. This paper uses the following two methods.

2.1.1. The size of hit rate $MR:PR$ means the request hits number; R means the request number by user. The size of the prefetching hit rate is expressed as: $MR = PR / R$.

2.1.2. Delay ratio $Delay^R : Delay^P$ means the prefetching latency; $Delay$ means the latency which doesn't use prefetching technology.

The delay ratio is measured from user's point of view. So it will take benefits for users.

2.2. Related Definition

Compared with the traditional web users, mobile web users have their own characteristics. (1)The certainty of the mobile user's behavior, one device is usually a fixed number corresponding to a fixed user. By the fixed number, it can get the information which users get from the web. Therefore, the user interest degree can be determined according to the characteristics. (2)Context information, due to the characteristics of the mobile network, it can predict the behavior of the current user according to the user's staff, the location, time and other factors. (3) Instability, due to the instability of the wireless environment, it can cause the host to frequently break link .So the data which users get should not comprehensive[10,11].According to the above characteristics, the definition is given as follows:

Definition 2.2.1.Context is the combination of vary kinds of information which is related to users. The information describes user behavior and surrounding circumstances. The context is expressed as follows:

$$Context = \{Context(1), Context(2)...Context(i)...Context(n)\} \quad (i = 1...n)$$

For mobile user i, the context information is defined as follows:

$$Context(i) = \{UserId(i), Location(i), CurrentTime(i)\}$$

And, $UserId(i)$ means the mobile user's unique identification; $Location(i)$ means the mobile user's current location; $CurrentTime(i)$ means the mobile user's access time.

Definition 2.2.2.Context Similarity: When a mobile user logs into the network via a mobile device, it can get the user context information based on the user ID, GIS technology and system time. Then the context similarity is calculated as follows:

$$\rho = \frac{Location(i) * Location(j) + CurrentTime(i) * CurrentTime(j)}{\sqrt{Location(i)^2 + Location(j)^2}\sqrt{CurrentTime(i)^2 + CurrentTime(j)^2}}$$

And, $Location(i)$ means the current location; $CurrentTime(i)$ means the current access time.

Definition 2.2.3.The mobile user interest degree is defined as:

$$Interest = \{Interest(1), Interest(2)...Interest(i)...Interest(n)\}; \quad (i = 1...n)$$

The mobile user interest degree in certain context information is calculated as:

$$Interest(i)^{(Context=Context(i))} = F(i,j) * T(i,j) * S(i,j)$$

And, $F(i,j)$ means the number of visits from user i to page j in $Context(i)$; $T(i,j)$ means the access time from user i to page j in $Context(i)$; $S(i,j)$

means the scroll times from user i to page j in $Context(i)$.

2.3. The Prefetching Algorithm

In this paper, we proposed an algorithm based on the combination of context information and user interest degree. The basic idea of this algorithm is as follows:

As the mobile users' behavior have certainty and related to the context information. So this paper prefetches the Top-N pages for users by analyzing user interest degree in different contexts and stores them into the proxy server. When the page which mobile user visits exists in the proxy server, get the data from proxy server. On the contrary, get the page from the server.

The prefetching algorithm is described as follows:

Input:

(1): The historical context information

$$Context = \{Context(1), Context(2)...Context(i)...Context(n)\}$$

(2): The user interest degree information in different context: $F(i, j)$, $T(i, j)$, $S(i, j)$.

(3): The current context information:

$$Context(i) = \{UserId(i), Location(i), CurrentTime(i)\}$$

Output: The Top-N page which the user is interested in.

Algorithm description:

1). Initialization, N represents the number of prefetching pages;

2).Calculate the context similarity $\rho(C, H)$ between current context and historical context by definition 2, then get the highest context similarity;

3).Calculate the size of user interest degree in $Context(H)$ by definition 3 and sort the pages by the size of interest degree in $Context(H)$;

4).Output the Top-N pages and put them into the proxy server in advance.

By the above algorithm, we can get the Top-N pages which the user most interested in, then put the N pages into proxy server in advance. When the mobile user access the web, access the proxy server first. If the web resource is not in proxy server, request the resource from real server. By this way, it can reduce the user's access latency.

3. Experimental Design and Analysis

We used the web log in database as the simulation data. Firstly, we prefetch data from database according to prefetching algorithm and put prefetching data into proxy server. Then we simulate users to access the proxy server (Redis Server) and database (Mysql). We observed cache hit ratios and the delay ratio for different cache size. The results are shown in Figure 2.

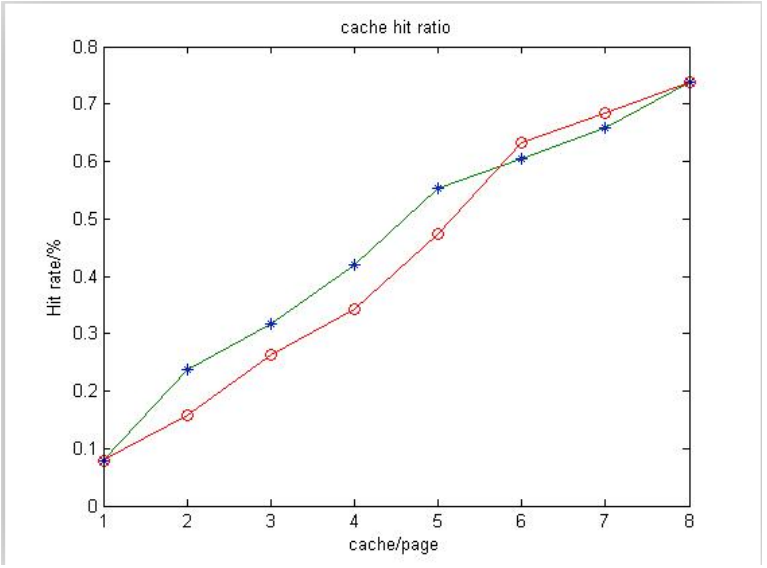

Fig. 2: Comparison of the size of the cache hit rate

From the Figure 2, We can find that the algorithm's cache hit rate which based on context and user interest degree is higher than the algorithm which based on user interest degree (○represents the algorithm which based on user interest degree, * represents the algorithm which based on context and user interest degree).Therefore, the algorithm which based on context and user interest degree has a certain effect in improving cache hit ratio. From the Figure 3, we can also find that when the cache size is large, the algorithm's cache hit rate which based on user interest degree is higher than the algorithm which based on context and user interest degree. It because the effect of cache size on cache hit ratio is not greatly.

In conclusion, this paper studies the delay ratio about two algorithms. When the user access the Redis proxy server, the access time is 3ms.When the user access the Mysql database, the access time is 452ms.So the delay ratio is $Delay^R$ =3/452<1.So it found that use the proxy server mode has a significant effect in prefetching performance.

4. Conclusion

Web prefetching mechanism as a means to improve the cache hit rate has played a more and more important role. This paper analyzes the characteristic of traditional web prefetching mechanism and combines with the character of mobile users, proposed an algorithm which based on context and user interest degree. By studying the model of user interest degree and analyzing the mobile user access preference in different context, it predicts data and stores it in the proxy server in advance. It improves the cache hit rate successfully and has an effect in reducing access latency. And the algorithm has no rigid requirements for hardware. It can apply to different operating system platforms. However, with the growth of mobile users, the amount of data about users will become more and more. So, how to get the data accurately which user concerned about from the mass data by using data mining techniques, or by adjusting the structure of network to accelerate the user's access speed is still a problem which needs further research and discussion.

References：

Padmanabhan V N, Mogul J C. (1996). Using predictive prefetching to improve World Wide Web latency. *ACM SIGCOMM Computer Communication Review*, 26(3), 22-36.

Han X C, Tian Y G. (2010). Web cache replacement algorithm based on prediction. Computer Engineering and Design, 31(1), 110-113.

Rongheng Lin, Hui Zhang, Hua Zou. (2012). Web Replacement Policy Based on User Requests for SNS. *Journal of Beijing University of Posts and Telecommunications*, 35(1), 111-114.

JiangChun Song, JunYi Shen. (2007). Research on a new clustering algorithm of Web user communities and Web sites URLS. *Control and Decision*, 22(3), 284-288.

Nanopoulos A, Katsaros D, Manolopoulos Y. (2003). A data mining algorithm for generalized web prefetching. *Knowledge and Data Engineering, IEEE Transactions,* 15(5), 1155-1169.

Zukerman I, Albrecht D W, Nicholson A E. (1999). Predicting users' requests on the WWW. *Courses and Lectures-International Centre for Mechanical Sciences*, 275-284.

Venkataramani A, Yalagandula P, Kokku R, et al. (2002). The potential costs and benefits of long-term prefetching for content distribution. *Computer Communications*, 25(4), 367-375.

Jiang Y, Wu M Y, Shu W. (2002). Web prefetching: Costs, benefits and performance. *Proceedings of the 7th international workshop on web content caching and distribution (WCW2002*). Boulder, Colorado.

Davison B D. (2002). Predicting web actions from html content. *Proceedings of the thirteenth ACM conference on Hypertext and hypermedia. ACM*, 159-168.

MuHai Hu, ShuQin Cai. (2011). The Model of Mobile Customer Segmentation Based on Hyper graph of Context Preference Knowledge. *Chinese Journal of Management*, 8(10), 1509-1516.

Beihong JIN, Sihua TIAN, and Chen Lin, et al. (2007). Incorporating caching with prefetching for wireless Web access. *Journal of Frontiers of Computer Science and Technology*, 1(2), 180- 190.

Shi Y C, Meng X W, Zhang Y J, et al. (2012). Adaptive learning approach of contextual mobile user preferences. *Ruanjian Xuebao/Journal of Software*, 23(10), 2533-2549.