

Data Mining Technology as the Core of the Intelligent Prefetching Web Application Analysis

Hui Yuan^{1,a} and Lei Zhao^{2,b*}

¹Xianning Vocational Technical College, China

²China Light Industry Design & Engineering Co., Ltd, China

^azhaoleies@163.com, ^b437209812@qq.com

Keywords: Data mining; Prefetching technology; Cache technology; Intelligent construction

Abstract. With the further development of in Internet, the rapid growth of users led to network congestion and server overload. Caching technology is considered to be one of the effective ways to reduce server load, reduce network congestion, and reduce customer access latency, but has limited role. To further improve the performance of the cache system, the introduction of prefetching technology. The data mining technology combines the knowledge of databases, artificial intelligence and statistics. This paper takes the data mining technology as the core to study the intelligent Web prefetching technology, and plays an important role in the future data mining in the accurate retrieval of information, personalized service, improve the design and service of the portal site, and build the intelligent Web site.

Data Mining Technology

Data mining is a process that uses various analysis tools to discover the relationship between models and data in massive amounts of data that can be used to predict. Data mining involves a variety of technologies such as databases, artificial intelligence, machine learning, and statistical analysis. Data mining technology can automatically analyze data from large databases or data warehouses, conduct inductive reasoning, discover potential patterns from them, or create associations, build new business models, help decision makers adjust their marketing strategies, and make the right decisions [1]. Data mining shows that knowledge is hidden in the vast amount of data accumulated in the day, and only by complex algorithms and reasoning can not find knowledge. At the same time, the idea of data mining also points out a new research direction for the development of artificial intelligence.

Data mining refers to a complete process that draws previously unknown, effective, and useful information from large databases and uses them to make decisions or enrich knowledge. A typical data mining system structure shown in Fig. 1.

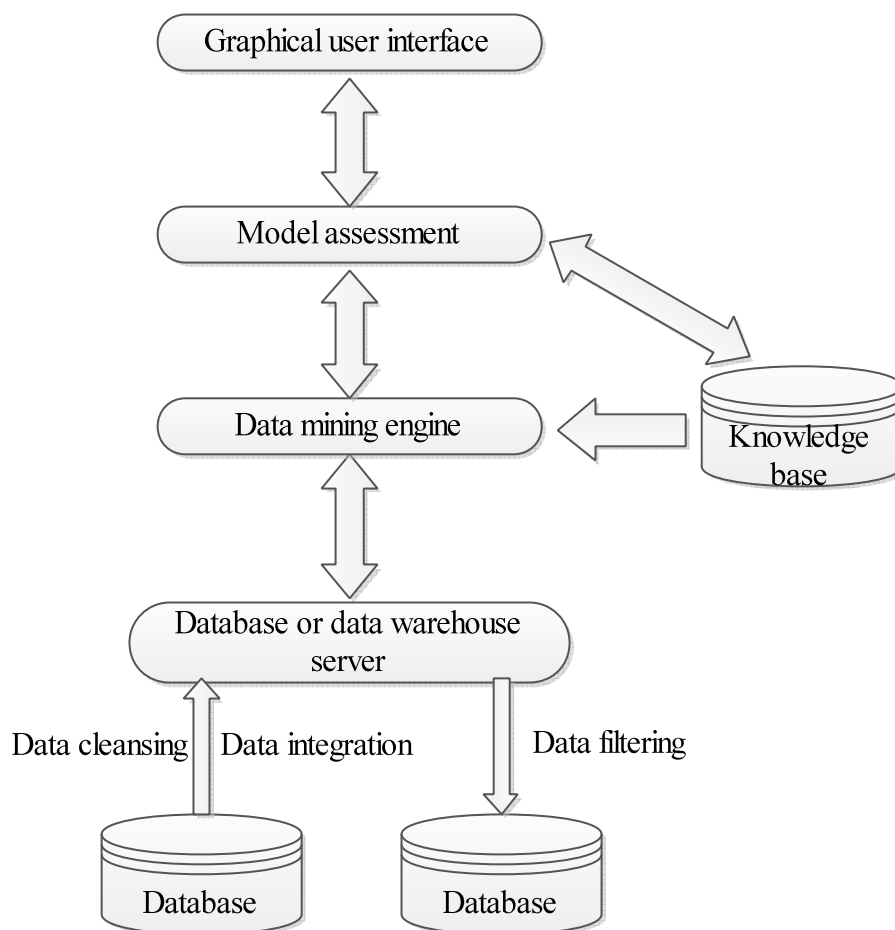


Figure 1. Data mining system structure diagram

Database, data warehouse, or other repository: It indicates that the object of data mining is made up of a (or group) database, data warehouse, spreadsheet, or other type of repository. It is often necessary to use data cleansing and data integration operations to perform initial processing of these data objects.

Database or data warehouse server: According to the user's mining request, the database or data warehouse server is responsible for extracting the relevant data. Knowledge Base: This is a domain knowledge that will be used to guide the search process of data mining or to help evaluate the mining results [2]. The user-defined threshold used in the mining algorithm is the simplest domain knowledge.

Data Mining Engine: This is the most basic part of a data mining system and consists of a set of functional modules for characterization, association, classification, clustering, and evolution and deviation analysis.

Mode Evaluation Module: The module can assist the data mining module to focus on mining more meaningful pattern knowledge based on interest in ness measures. The pattern evaluation module can also be integrated with the mining module, depending on the implementation of the data mining method used.

Graphical user interface: The module communicates between the user and the data mining system, allowing the user to interact with the system. On the one hand the user through the module to their own mining requirements or tasks submitted to the excavation system, and provide the necessary knowledge of mining search; the other hand, the system through the module to show or explain the results of data mining or intermediate results [3]. In addition, the module can also help users browse the data object content and data definition mode, assess the mining of the model knowledge, and in a variety of forms to show the mining of the model knowledge.

Web Prefetching Technology

The rapid development of Internet technology makes the sharing and distribution of information across time and space constraints, the network has become an important source of information for people to obtain. However, by the limitations of network bandwidth, users have to wait for a long wait time before accessing the page. In recent years, with the gradual upgrading of network bandwidth, the delay problem has been alleviated to a certain extent, but the impact on the quality of network services still exist, especially in some specific areas such as low-speed Modem Internet, wireless network. In order to solve the user access delay, people put forward the cache and prefetching technology [4]. Caching technology is based on the time of WWW access locality, in the application of the client's browser and the local proxy server, has access to the Web document saved, so as to avoid once again access the file has been accessed, but also to the remote server request. When the Web resources frequently change (such as dynamically generated Web files), cache technology to solve the problem of delay is no longer obvious [5]. Prefetching technology is based on the user's access to the spatial locality, according to the user's history access records, the use of data mining technology to dig out the user's access mode, and then based on the user's current situation to predict the upcoming page, it is an active cache technology.

According to the prefetching technology in the implementation of the location of www, can be divided into client prefetching, proxy server prefetching and server prefetching [6]. The client prefetches are prefaced by the client according to the customer's historical access information and the current request and the prefetching content is cached in the client cache. The client prefetching is relatively accurate when analyzing customer history access information, but not with other customers Share the prefaced content, which may cause a large amount of network traffic; proxy server prefetching by the proxy server to analyze the customer's historical access information, the customer's current request prefetching and prefetching content cache in the proxy server cache or send to the client cache, the proxy server prefetching can play a group of customers to visit the same site when the content of the contents of the request to achieve the prefaced content of multi-customer sharing, the lack of proxy server is facing multi-client, multi-site Access to information, the complexity of prefetching is relatively high; server-side prefetching by the server to visit the customer's information, the customer's current request prefetching and prefaced content into the cache or sent to the proxy server and customers [7]. Server-side prefetching than proxy server-side prefetching can play a better part of the content of customer access to the content of the group, and achieved better results, and the server-side to save the customer access content is only the content of this site, you can better analyze the customer Content preferences, preferences, and thus better content prefetching.

Model Established

Development Environment. Operating system: Windows 7 and above.

Memory: 1G.

Development language: JAVA, XML, XSL.

Development environment: JBuilder 7.0, XML SPY.

System Design. The system is based on data mining, the purpose is to web prefetching technology research. The workflow is shown in Fig. 2:

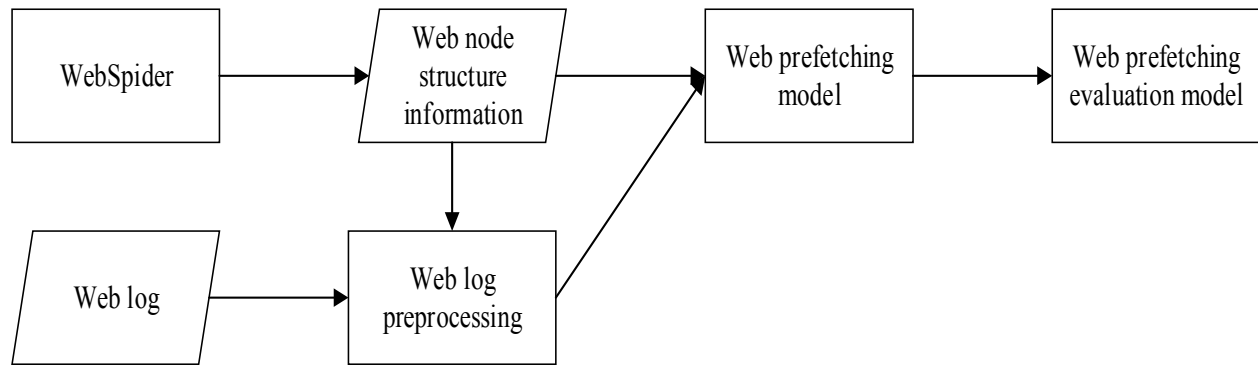


Figure 2. Web prefetching system

(1) Web Spider module access site to obtain the site structure information and Web page document size, used to calculate the LCR value and other purposes.

(2) Web preprocessing module combined with Web site structure information to analyze the Web port to be processed after the user session set.

(3) Web prefetching model module to generate a prefetching model by analyzing the user session set.

(4) Web prefetching evaluation model to evaluate the performance of Web prefetching model.

Prefetching Model Study. In this paper, we use the prefetching model based on sequence pattern. Sequential pattern mining is a further generalization of association rule mining [6]. It excavates the temporal relations among item sets in sequence database and is an important part of data mining. The sequence pattern can be used for prefetching, that is, by mining the log on the server to get the sequence pattern of the user access sequence. According to these modes and the user's current visit to the web page, the user's most likely access to the page is sent to the user.

Although the granularity of the user session is still too large for the sequence pattern mining after the previous data preprocessing process, it is necessary to further decompose the session to make it a certain semantic transaction. Often transaction identification uses the Maximal Forward References method, and a maximum forward reference path is equivalent to a meaningful transaction. A backward reference means that the same user has requested a page that was viewed before. When a backward reference occurs, the forward reference is aborted, and a maximum forward reference path can be obtained, and a maximum forward reference path can be obtained when the user session is aborted. $\{x_1, x_2, \dots, x\}$ represents a user session, $\{y_1, y_2, \dots, y\}$ represents a string containing a potential MFP, The initial value is empty. The Flag indicates whether the current traversal direction is forward or backward, Flag 1 indicates forward, and 0 indicates backward.

Begin:

for each session $sk \in \text{Ses}$;

$y_1 = x_1; j = 2; i = 2;$

Flag = 1;

while($i \leq n$) {

for $1 \leq k < j$

if($x_i == y_k$) {

if(Flag == 1) output $\{y_1, y_2, \dots, y_n\}$ as MFP to DF;

$j = k+1; i++;$

Flag = 0;

```
    }  
    else{  
        yj = xi; j++; i++;  
        Flag = 1;  
    }  
} //end while  
if(Flag == 1) output {y1, y2, ....., yn} as MFP to DF;  
end.
```

After the transaction identification is carried out, the sequence pattern mining can be performed on the result set after the data preprocessing.

Results Analysis

The Web prefetching evaluation model is used to evaluate the Web prefetching model. The core idea is to determine the prefetching model through statistical data. In the experiment, we maintain two variables hit Ratio and byte Hit Ratio in the program, according to the value of these two variables to determine the performance of the Web prefetching model is good or bad. Prefetching technology reduces the client response time by increasing the load on the server and increasing the network traffic. Therefore, a good prefetching model must take into account the access characteristics of Web clients, server load and network traffic status.

References

- [1] M. Zhu. Data Mining, Hefei: China University of Science and Technology Press, 2002.
- [2] J.C. Wang, J.G. Pan and F.Y. Zhang. Study on Web Text Mining Technology. [J]. Computer Research and Development, 2000: 37 (5)
- [3] B. Mobasher, R. Cooley and J. Srivastara. Automatic personalization based on web usage minim. Communications of the ACM, 2000, 43(8):142-151
- [4] X. Su, X.L. Jiao and P. Wu. Web Mining Application Research [J]. Information Systems, 2005, 28 (36): 651-655
- [5] W. Wang, W. Gao and J.T. Li. Web Data Mining [J] .Computer Science, 2000: 27 (4)
- [6] H.Q. Xu and Y.C. Wang. Web Prefetching Model Based on User Access Path Analysis [J]. Journal of Software 2003 14 (6): 1142-1147
- [7] M. Yu. Study on the Development and Application of Computer Data Mining Technology [J] .Journal of Urban Construction Theory (Electronic Edition) 2016 (22).