

The Application and Research of Web Log Mining in Network Resources of Ceramic Information

Wan Fang

Jingdezhen Ceramic Institute, Jingdezhen City, Jiangxi Province, China 333403

Keywords: Web Log Mining Ceramics Application of resources

Abstract: this article does research on the application web log mining technology in network resources of ceramic information. Through the analysis and pretreatment of the raw log data, and then relevance analysis of the information in WebPages, the structure of ceramic enterprise website is optimized and the information of ceramic product is improved. The research results can provide some useful data for the development of ceramic enterprises.

1. Introduction

With the development of computer technology and the popularity of the world wide web, many users browsing and ordering ceramic products on the internet. More and more ceramic enterprises build their own webs of ceramic resources information on the internet for the convenience of users. With the network of ceramic information resources to provide abundant ceramic resources to users, its structure and function also becomes more and more complex, and then the problem of excessive resource information also arise. Users tend to set lost in the face of huge amount of ceramic resources information and then unable to successfully find the resources which needed.

The managers of ceramic products and designers of website construction urgently need the advanced technology and management concepts, to help the users quickly and efficiently find the ceramic resources that meet the personalized needs of users. How to find the need of the users from the use of the site's users, and to optimize the website structure and content to, provide personalized service has become the urgent problem of ceramic information website, and web log mining provides the most effective way to solve this problem.

2. Web log mining technology

Web mining is the application of data mining in web. It uses data mining technology to extract interesting, useful and hidden information from the network related resources and behaviors. It is also a comprehensive technology which involves many fields including web technology, data mining, computer linguistics, information science.

Web log mining belongs to the web application of mining areas. It uses data mining technology, through deep analysis of a large number of users access records in web server log, to such as the user access pattern and their interests, find some interesting novel, potentially useful, and understandable unknown information and knowledge. It is used for the analysis of the use of site, so as to assist management and support decisions. The Web log mining can help enterprises to better analyze and understand the behavioral characteristics of the users, and find the cause and regulation which deeply hides behind the user behavior, so as to get the greatest economic potential of exploration of ceramic enterprise Web.

3. Technology application

The main research contents of this article is to extract interesting information original web log access records from the user's, record information about the user's access and exchange, and through the analysis of these data, it can help understand user's behavior in order to optimize the structure of the site, or to provide personalized service to users. The Web log mining process as shown in Figure 1:

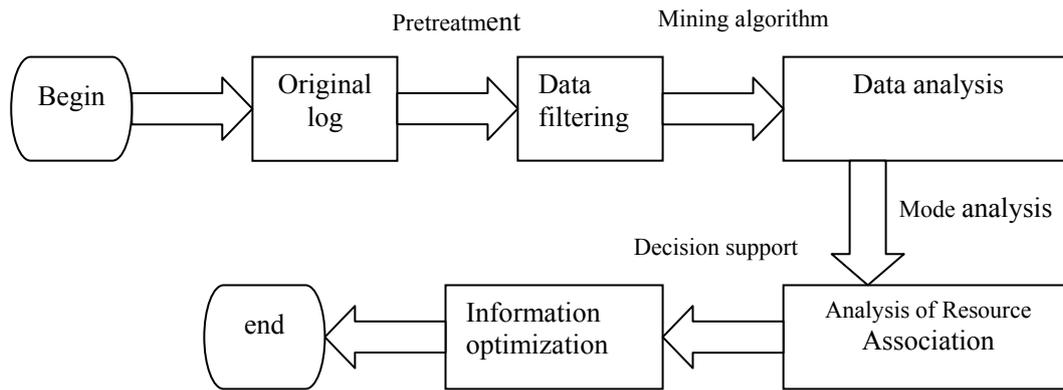


Figure 1 Process of web log mining application

In order to optimize the resources of ceramic information network, improve the utilization rate of cyber resources, and to help the users quickly find target resources, this research is based on the Webpage browsing records on the server, according to the user's browsing preferences and searching keywords, to analyze the correlation between the WebPages. In the ceramic resources information network which applies Web log mining technology, the website can provide a lot of useful information resources. In access to product information, visitors get access to the system with the user login, then the system records the user's access behavior, information including access products, the time that stay in the Webpage, record key words used in searching information, calculate the amount of visits and access time, and analyze the correlation between Webpage, and the users' most interesting products and so on, to provide a basis for ceramic enterprise decisions.

3.1 Data filtering

Users' setting access to the web site will produce a lot of log information, we are impossible to analyze the data to all the log information, so we have to filter the data, to filter out the useful information, filtering the effective data information, according to the constraints which defines data access.

3.2 Data analysis

Statistical analysis of data is the most commonly used way to extract knowledge from Web user behavior. Statistical analysis of filtered web journal can obtain the relevant network using information. Although it does not seem to make deeper data analysis, the results tend to improve system performance, enhance the security of the system. It is of great help to optimize the site structure and provide marketing decision etc.

3.3 Analysis of Resources Association

Association analysis mainly depends on the technology of association rules, which is mainly used to find user session by users often visited pages, there is no order relationship between these pages. Mining association rules usually use the Priority algorithm and its deformation algorithm, a maximal frequent item sets mining access from transaction database, this set is the user access patterns from association rules mining. Association can not only be used as a reference for designers to optimize web sites, but also a basis for commercial and shopping decision-making in Web.

4. Case analysis

In the increasingly rich of resources network ceramic information, we take a daily ceramic enterprise resource website as an example, according to the data content characteristics of users browsing records from the server, or according to the constraints which define data winning, we can filter out the effective data (Table 1). In this research, we define the following three conditions according to the demand.

4.1 User limit

In this research the user must set the account and password to login the website to browse the website, considering each user has a different purpose and interest, users with higher interest degree

tend to have a higher frequency of Webpage access records. In order to have a high reference to the analysis results, this research defines a minimum access value, so as to filter out the last 15 days of browsing records. During this time, the number of access users is less than 5 times, the browsing history will be ignored.

4.2 Attention Limited

When a user browsing Webpage data over a longer period of time, we can say that the user has a higher Webpage attention, so this research defines a minimum residence time threshold, to filter out the user browsing data that residence time is higher than the threshold. If the user in the target Webpage retention time is less than 8 seconds, then his order browsing record will be ignored.

4.3 Key word records

When users browse a website, and can not timely and quickly find the information resources which they need, they often find relevant information by key word search. By filtering and recording the user access key words, we can get to know the information that the user is interested in, help enterprise optimize network information resources (for example: the tings set of products classification, key words of product information).

Table 1 users browsing the resources record

User ID	Access time	Objective Webpage	Residence time (S)	Search key word
User01	2013/10/01 8:10:20	A1	12
User01	2013/10/02 18:09:50	C3	10	Chinese red
User01	2013/10/05 15:25:45	D4	32	Antique
User01	2013/10/05 16:05:10	C1	14
.....
User03	2013/10/01 12:11:21	D4	18	Blue and white
User03	2013/10/01 12:18:42	C2	12	Folk style
User03	2013/10/05 09:54:30	A2	10	antique
.....
User12	2013/10/02 18:08:21	A1	11	European style
User12	2013/10/02 18:09:50	C2	15	Folk style
User12	2013/10/04 15:40:15	D4	20	antique
User12	2013/10/11 22:05:20	B2	18
.....

We set up a formula, "att" represents the degree of attention. "f" represents a single user in a particular period to a single target Webpage visits, "fmin" represents the minimum number of users access the target Webpage experimental data, "fmax" represents maximum user access to target Webpage number experimental data. "T" represents the time that user stays on the Webpage, "tmin" represents the minimum residence time, "tmax" represents the longest residence time. The formula can be a single user to target Webpage attention degree, the maximum of "att" from each user goals Webpage list. According to the related column recommended level will Webpage index recommended to Webpage side, improve the efficiency of user browsing. "kword" represents the frequency that experimental data of key words appear, through sorting from the list before the extraction 3 most keywords, then the Webpage information optimization. In order to verify the correctness and effectiveness of the experimental results, we recommend before and after the user browsing sequence length and flow rate were compared, to achieve the expected goal, through recommendation and optimization of personalized information resources, improve the convenience of user browsing, and reduces the search time related resources.

5. Summary

This research is about the Web log data mining technology used in ceramic information web log analysis. Through filtering and processing the raw log data, through correlation analysis, optimize the network structure and ceramic products resources information according to the user's requirements, achieved the network personalized recommended target, and it provide a theoretical basis to the development and decisions of ceramic enterprises. In a word, the application of Web log data mining technology in ceramic cyber resources plays a certain role in promoting the development.

References

- [1] Zhang Jiaai, Research on Data Mining Technology, Journal of Ji Lin Agriculture Science and Technology College,2008.02
- [2] Zhang Xiaofeng, Research and Application of Personalized Recommendation Method of Web Oriented Data Mining Technology in Website Optimization, Manufacturing Automation,2012.01
- [3] Jia Lv, Study on Application of the Technology of Web Log Mining, Journal of Chongqing Normal University,2006.12
- [4] Chen Xiaohong, Research on Efficient Association Rules of Data Mining Based on Web, Computer Engineering and Science, 2005.11
- [5] Guo Yan, Analysis of Log Scale Network and User Interest Mining, Chinese Journal of Computers,2005.09