Research on Web User Behavior Compliance Detection Method Based on Clustering Data Analysis Technology

Cheng Qin1,*

1Kunming Power Supply Bureau of Yunnan Power Grid Co., Ltd,Yunan, China

*Corresponding author e-mail: qincheng@kunming.psb.org

Abstract. With the rise of network technology, the Internet has been integrated into all aspects of people's life, which brings convenience to our daily life. However, the rise of the network has led to the emergence of a large number of web pages, which also makes it difficult for people to accurately need the information. At the same time, the proliferation of Internet users leads to a large number of user behavior data, which has been applied to analyze user behavior. By mining Web log information, we can extract the behavior rules and interests of users, which will put forward more decision direction. By improving the system platform, we can improve better user service, which will improve the user experience. Facing the marine user behavior generated by web, we can analyze it by clustering algorithm, which will better analyze the compliance detection of user behavior. Through user behavior clustering, we can extract the user's browsing behavior in the web server log, which will better mine the user's interest.

Keywords: Clustering Analysis Technology, Web, User Behavior, Compliance Testing

1. Introduction
With the development of the Internet, the amount of interactive data in web is increasing, which will make it more and more difficult for users to collect interest. Website operators need to mine the information that users are interested in in big data [1]. By analyzing user behavior, operators can identify the events that users are interested in, which will better recommend useful information to users. In the big data environment, the artificial filtering mode has been unable to meet the analysis of user behavior, which needs to be solved by clustering analysis [2-4]. Through cluster analysis, web operators can make clear the user's behavior preferences, which will better predict the user's preferences. Through clustering algorithm, we can find the user groups with the same interest as the target users. According to the user access log, we can recommend appropriate network resources to the target users. User access log data is extremely easy to obtain, which only needs to browse the web page according to their interests to generate recommended data sources [5, 6]. Users' interests will not change much in a period of time, which can be done by clustering analysis. Through clustering and recommender, the web system can filter many users with different interests, which can improve the accuracy of recommendation [7].
2. Cluster analysis

2.1. The definition of clustering
Cluster analysis is the process of dividing data objects into multiple classes or clusters according to certain algorithm rules. Try to ensure that the objects in the same cluster or class have high similarity, which will lead to large object differences. Because of the clustering characteristics, in order to reduce the calculation amount and accuracy of the overall data, we often treat the data objects in a cluster as a whole. If some tasks only need some characteristic data, we can eliminate the interference of other irrelevant data after clustering. In addition, clustering analysis can be used as a preprocessing step of other algorithms, which can simplify the calculation and analysis efficiency.

2.2. Introduction of FCM clustering algorithm
FCM is a fuzzy clustering method, which is the minimum variance sum to calculate the membership degree of samples for clustering criteria. By iteratively optimizing the objective function, we can get the local minimum, which will get the optimal result.

Given Data Set \( X = \{ x_j, j = 1, 2, \ldots, N \mid X_j \in \mathbb{R}^q \} \), the objective function of FCM is minimized as Formula 1.

\[
J_m = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^m D_{ij}
\]  

(1)

\( D_{ij} = d^2(x_j, v_i) \) is the similarity measure of cluster centers \( v_i \) and \( x_j \) of class \( i \). According to Euclidean principle, we can get formula 2.

\[
D_{ij} = d^2(x_j, v_i) = \| x_j - v_i \|^2
\]  

(2)

Therefore, we can obtain the constraint conditions as shown in Formula 3.

\[
u_{ij} \in [0,1], \sum_{i=1}^{C} u_{ij} = 1 \forall j
\]  

(3)

\[
0 \leq \sum_{j=1}^{C} u_{ij} < N \forall i
\]

Pixels far from the cluster center are given low membership values, while those near the cluster center are given high membership values.

By making \( u_{ij} \) and \( v_i \) zero, we obtain the condition of minimizing JM, as shown in formula 4.

\[
u_{ij} = \left[ \sum_{k=1}^{C} \left( \frac{D_{ij}}{D_{kj}} \right)^{\frac{1}{m-1}} \right]^{-1}
\]  

(4)

\[
v_i = \frac{\sum_{j=1}^{N} u_{ij}^m x_j}{\sum_{j=1}^{N} u_{ij}^m}, (i = 1, 2, \ldots, C)
\]

By iterating formula 4, we can get formula 5.

\[
max x_i \in [1, C] \quad \| v_i^{(l)} - v_i^{(l+1)} \|_{\infty} < \epsilon
\]  

(5)
3. Characteristics of cluster analysis

3.1. Process target set without classification basis
Clustering analysis does not need a classification standard as input in advance, which can select the cluster eigenvalues from the sample data. The biggest difference between clustering and classification is that the target of classification has been known in advance, while clustering can only automatically classify from sample data. The result is the same as classification, but the category has no prior definition.

3.2. Deal with classification determined by multiple characteristic variables
In a loan institution, it is necessary to assess the credit level of its users. When mining the data provided by users, we can use multiple data of users as characteristic variables for clustering analysis, such as the number of real estate, family income and expenditure amount, user age, etc. However, other general analysis methods may not be able to solve such complex and multi-dimensional classification problems.

3.3. Exploration of classification methods
We should be based on the external characteristics of things, which will analyze the internal characteristics and laws of things. By combining the similarity distance, we can calculate the clustering groups that will be generated by grouping things. Through clustering analysis, we can analyze the undiscovered grouping mode, which is often used in data mining.

4. Web user behavior detection

4.1. Collection of web user information
Every time you click on the website, the server will record the corresponding time and date information. For this kind of service application system, the length of time we stay on a web page can reflect the user's interest in the content of the page. For example, in distance education, if a user is learning a distance mathematics subject, his stay time in the mathematics section will increase compared with other subjects, which will spend most of his time learning. Therefore, in addition to recording the user's access path, the server records the browsing time of the page is also the data that must be collected.

4.2. Calculate user browsing time
Each time the user clicks the corresponding connection, the corresponding time data will be generated. For users who are interested in the website system, they may click on the same page and link several times in a row. Therefore, we can record a time parameter for each browsing time, which can partially show his interest in behavior each time. However, this way can not accurately reflect his real interests. Therefore, we must record the number of times users click on certain page links on the website and the total time spent browsing information on the system within a month. By quoting the total time and the number of hits, we can calculate the average time spent browsing pages of interest during this period, as shown in formula 6.

\[
Browsing\ time = \frac{The\ total\ time\ to\ browse\ the\ URL\ page}{The\ number\ of\ clicks\ by\ the\ user}
\]  

(6)

Among them, the browsing time of each URL is the difference between the time of the next link page and the time of the current link click.

4.3. Design flow of algorithm
Web log clustering is mainly divided into three processes: data preprocessing, feature analysis and clustering, which can obtain web server log data. By preprocessing the data, we can extract the user
behavior characteristics. Finally, for the analysis of user behavior characteristics, we can cluster according to the similarity calculation results. The flow chart is shown in Figure 1.

![Flow chart](image)

**Figure 1.** The design process of the algorithm.

4.4. **Cluster feature analysis**

This paper introduces the web page type. By combining with clustering, the user behavior is clustered in this paper. Through the graphic options of SPSS modeler, this paper draws the corresponding bar chart, as shown in Figure 2.

![Bar chart](image)

**Figure 2.** Clustering bar graph.

5. **Conclusion**

This paper analyzes the web user behavior by clustering. The traditional website has the disadvantages of complex structure and many page sites, which will be difficult to analyze the user's hobbies. Therefore, we can do Web Clustering Analysis Based on user path, which can mine user behavior. In this paper, the main process of cluster analysis is established. In addition, an example is given in this
paper. Through fuzzy clustering, this paper analyzes the clustering algorithm of the system log file, which can better get the needs of customers.

Acknowledgment
Project Name: research and application of business dynamic sensing system based on full flow monitoring technology, Project Number: KJKM2018302.

References
[1] Huang Zhenhua, Xiang Yang, Zhang Bo, Wang Dong, Liu Xiaoling. An effective method for k-means clustering [J]. Pattern recognition and artificial intelligence, 2010, (5): 516-521.
[2] Jia Rongfei, Jin Maozhong, Wang Xiaobo. Query clustering based on user query log [J]. Journal of Beijing University of Aeronautics and Astronautics, 2010 (4): 500-503.
[3] Rong Huigui, Huo Shengxu, Hu Chunhua, Mo Jinxia. Collaborative filtering recommendation algorithm based on user similarity [J]. Acta communication Sinica, 2014, (02):16-24.
[4] Wang Jicheng, pan Jingui. Research on Web text mining technology [J]. Computer research and development, 2010, 37 (5): 513-520.
[5] Wang Xinjun. An improved recommendation algorithm based on collaborative filtering and partition clustering [J]. Computer research and development, 2011, (3):205-212.
[6] Xiao Qiang, Qian Xiaodong. Association Rules Mining Algorithm Based on K-means algorithm for fteb access users [J]. Library and information work, 2011, (16):136-139.
[7] Zhao Wei, he pilian, Chen Xia, et al. Research on data preprocessing technology in Web log mining [J]. Computer applications, 2008, 23 (5): 62-64.