ISSN 2616-5775 Vol. 5, Issue 14: 40-43, DOI: 10.25236/AJCIS.2022.051406

Design of Network Personalized Information Recommendation System Based on Big Data

Hu Xiaoqin, Pan Jinfeng

Quanzhou University of Information Engineering, Quanzhou, Fujian, China

Abstract: *At present, some network personalized information recommendation systems have the problem of high CPU usage, so this paper designs the network personalized information recommendation system based on big data. The network personalized information tags are obtained, and the keywords are selected to annotate the recommended resources. Based on big data, the potential factors of user behavior are extracted and the data association features are identified. Under different recommended strategy, the retention rate and click-through rate of each traffic group ID are counted, in order to improve the traffic allocation function of the recommendation system. Test results: the average CPU utilization rate of the network personalized information recommendation system in this paper is 40.43%, which is 7.02% and 7.26% lower than the other two network personalized information recommendation systems, respectively. This shows that the designed network personalized information recommendation system has better compression resistance after combining big data technology.*

Keywords: *Flow distribution; Personalized information; Information bias; Hidden features*

1. Introduction

Influenced by the development of information technology, network personalized information recommendation service has become a new product. At the same time, the impact of the rise of the Internet is not only reflected in the commercial value, but also penetrated into the daily life of most people [1-2]. In particular, the online network covers the corresponding user behavior and information transmission trajectory, which also leads to the trace of all data and information on the network, providing a data basis for the design and application of the network personalized information recommendation system [3]. At the same time, with the rapid growth of information, a series of problems may occur. Therefore, it is necessary to introduce big data technology to preprocess relevant data. In essence, the network personalized information recommendation system is to screen out the information that meets the needs of users from the numerous data information, which has high practical significance.

2. Design of network personalized information recommendation system

2.1. Access to network personalized information tags

Network personalized information label is a kind of integration of resources such as user interest and information characteristics, including corresponding labeling times, frequencies and other information [4]. In the above label information, big data technology is used to mine the information and rules required by users, and select keywords to annotate the recommended resources. Generally, data and information contained in labels can reflect user interests and preferences [5]. Under the influence of specific resources, the probabilistic expression formula of recommendation labels is described as follows:

$$G = \frac{(\eta, \varepsilon | l)}{\delta(\eta, \varepsilon)} \tag{1}$$

In formula (1), η represents user, ε represents resource, and δ represents label set. Formula (1) works on the premise that ε and η have the same weight, but considering the actual scenario,

ISSN 2616-5775 Vol. 5, Issue 14: 40-43, DOI: 10.25236/AJCIS.2022.051406

η contains more labels. Therefore, the formula (1) is smoothed as follows:

$$G' = \log_2(\eta, \varepsilon) \times \frac{|\varepsilon|}{\log_2(|\eta|+1)^T} \tag{2}$$

In Formula (2), T represents label dimension. In addition, when invalid labels exist in the recommendation system, the negative impact on the recommendation system should be reduced by data cleaning [6]. Although the popularity of a resource can be judged by the network personalized information label in most networks, when the amount of information is too large, it still needs to be classified according to the degree of association between the labels [7]. In order to mine the hidden features of user behavior from different angles, as the basis for subsequent system recommendation. Based on the above description, the steps of obtaining the network personalized information label are completed.

2.2. Extracting potential factors of user behavior based on big data

Use big data technology to extract potential factors of user behavior from massive network data according to historical data and information, and guide the subsequent system design and application direction. Improve system functions according to user preferences and other behavioral information. In the face of problems such as missing data in the data set, it is necessary to identify data association features and predict unknown information through big data technology. The specific expression formula is as follows:

$$\hat{\mu} = \frac{L}{\phi_\mu} \tag{3}$$

In Formula (3), L represents the data set, ϕ represents the total number of vectors, and μ represents the inner product of vectors. At the same time, because each system user has different scoring standards for different resource types in the scoring process, information bias needs to be considered in the process of extracting potential factors of user behavior [8-9]. Under the condition that the average value of user related items is known, the potential factor is taken as the optimization coefficient of the loss function, and the mathematical expression formula of the offset information is obtained as follows:

$$\sigma = \sum \frac{k + \phi_\mu}{\|\lambda\|} + \overline{h} \tag{4}$$

In Formula (4), λ represents the potential factor, and k represents the optimization coefficient of the loss function. In addition, since the data information reflected by labels is more multi-dimensional, it is necessary to further combine semantic analysis methods to compare the differences of data information after extracting the potential factors of user behavior [10]. Based on this, the steps of extracting potential factors of user behavior are completed.

2.3. Improve the flow distribution function of the recommendation system

The traffic distribution function of network personalized information recommendation system is to select different traffic distribution schemes under different network modes according to the established recommendation strategy and user behavior preferences [11]. When users browse information on the web page, a series of behavior logs will be generated accordingly, and the system's traffic allocation function will automatically and spontaneously send the traffic group ID to users, and enter the traffic group allocation phase [12]. After continuous operation, the system will count the retention rate and click through rate of each traffic group ID under different recommendation strategies to determine whether the recommendation strategy needs to be adjusted. Based on the above, the basic structure of the flow allocation function use case is obtained, as shown in Figure 1:

ISSN 2616-5775 Vol. 5, Issue 14: 40-43, DOI: 10.25236/AJCIS.2022.051406

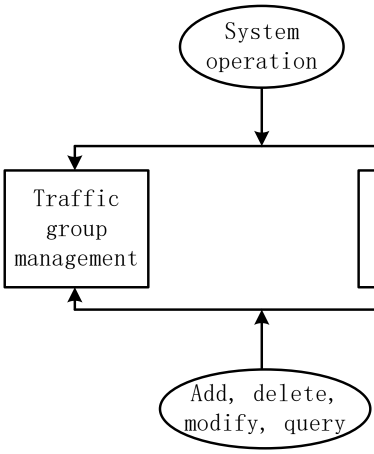

Figure 1: Basic Structure of Flow Allocation Function Use Case

As can be seen from Figure 1, the traffic group management and traffic allocation management of the system have the same permissions for user recommendation policies. If a user's traffic group ID has different attributes, the system administrator can query and modify the attributes. In addition, in the face of multi-dimensional and heterogeneous data sources, the recommendation policy grouping function of the system is triggered, and the data dimension is ignored to a certain extent, so as to ensure that the traffic allocation function is not disturbed [13]. Based on this, the steps of improving the traffic distribution function of the recommendation system are completed.

3. System pressure test

3.1. Setting up the Test Environment

According to the actual application environment of the network personalized information recommendation system designed this time, build the system test environment before the test. Server configuration: CPU: Intel Xeon E3-1220 v6, memory: DDR4 16G, hard disk: SATA 512GB; In the cluster environment, two master nodes are allocated as the primary node and the standby node. In addition to resource scheduling management, three or more slave slave nodes are allocated for data storage and parallel resource processing. Java and Scala are selected as the development language, Tomcat is used as the Web server, the development framework is SpringBoot, and the MysQL tool is Navicat Premium. Before testing, after installing MysQL, create Hive users, download files to the $HIVE HOME/lib directory, and conduct system testing.

3.2. Stress test results

Table 1: Stress test results of network personalized information recommendation system

Number of concurrent users	CPU utilization (%)		
	Network personalized information recommendation system based on improved neural network	Network personalized information recommendation system based on association rules	Web personalized information recommendation system in this paper
50	5.16	6.22	3.49
100	22.15	23.69	18.94
150	26.99	26.04	23.17
200	43.61	45.48	38.55
250	47.93	49.36	41.02
300	52.61	53.59	43.31
350	56.94	57.24	53.19
400	62.15	63.48	58.47
450	75.61	72.48	60.94
500	81.33	79.33	63.26

ISSN 2616-5775 Vol. 5, Issue 14: 40-43, DOI: 10.25236/AJCIS.2022.051406

In order to get accurate test results, this test is conducted by experimental comparison. The network personalized information recommendation system based on improved neural network and the network personalized information recommendation system based on association rules are used as references. The cpu utilization of the proposed network personalized information recommendation system and the other two network personalized information recommendation systems in different user concurrency scenarios are respectively tested, as shown in Table 1:

As can be seen from Table 1, the average cpu utilization of the network personalized information recommendation system in this paper is 40.43%. The average cpu utilization of the network personalized information recommendation system based on improved neural network is 47.45%. The average cpu utilization of the network personalized information recommendation system based on association rules is 47.69%. The network personalized information recommendation system in exposit-text is superior to the other two network personalized information recommendation systems in terms of pressure performance.

4. Conclusion

Based on big data technology, this paper improves the flow distribution function of the system. The effect of user behavior on recommendation strategy is fully considered and the performance of the existing network personalized information recommendation system is improved. In the subsequent research, we need to focus on the difference between long-term interests and short-term interests of users, so as to obtain more accurate recommendation results.

References

[1] LIU Chunmao, ZHANG Xuejia, ZHOU Yue. An Empirical Study on Information Encounter Traits of Personalized Information Recommendation Systems Based on Eye Movement Experiment——In Case of University Students[J]. Journal of Modern Information, 2022, 42(05):26-37.

[2] WANG Xinyue, WANG Hong. Design of personalized recommendation system for network associated data based on hybrid filtering[J]. Electronic Design Engineering,2021,29(01):110-114.

[3] LIU Junliang, LI Xiaoguang. Techniques for Recommendation System: A Survey[J]. Computer Science,2020,47(07):47-55.

[4] FU Chenbo, ZHENG Yongli, ZHOU Mingming, et al. Personalized restaurant recommendation combining user behavior network information[J]. Journal of Zhejiang University of Technology, 2020, 48(5):574-580.

[5] LI Nanxing, SHENG Yiqiang, NI Hong. Conditional Convolution Implicit Factor Model for Personalized Recommendation[J]. Computer Engineering, 2020, 46(4):85-90+96.

[6] HUANG Dong, CHEN Guang, LI Haibin, et al. Spark personalized place recommendation system[J]. Journal of Liaoning Technical University(Natural Science Edition), 2020(6):533-540.

[7] GE Suhui, WAN Quan, BAI Chengjie. Personalized movie recommendation system based on Spark and alternate least squares[J]. Journal of Nanjing University of Science and Technology, 2020, 44(5):583-589.

[8] LIU Min, LEI Jue. Design of personalized recommendation system for shipping lines based on social network location[J]. Ship Science and Technology, 2020, 42(22):190-192.

[9] LI Kunlun, SU Huading, RONG Jingyue. Personalized Recommendation System Incorporated with the Implicit Feedback of Projects and Users[J]. Journal of Chinese Computer Systems, 2020, 41(3):519-525.

[10] KUANG Rong, YANG Zhenguo, LIU Wenyin. Multiple factors based personalized learning recommendation system[J]. Application Research of Computers, 2020(1):183-187.

[11] DAI Jie, LI Qingshan, CHU Hua, et al. Breakthrough in Smart Education: Course Recommendation System Based on Graph Learning[J]. Journal of Software, 2022, 33(10):3656-3672.

[12] XU Bianyun. Design of university library bibliography personalization recommendation system based on association rules[J]. Modern Electronics Technique, 2021, 44(08):68-72.

[13] JIANG Wuxuan, YI Ming, WANG Ling. Research on Personalized Recommendation Combining User Interest Drift[J]. Information Studies: Theory & Application, 2022, 45(1):38-45+37.