



Intelligent Push Method of News and Information for Network Users Based on Big Data

Ting Chen¹(✉) and Zihui Jin²

¹ Chengdu College of University of Electronic Science and Technology of China,
Chengdu 611731, China
chenting08282022@163.com

² Tianfu College of SWUFE, Chengdu 621050, China

Abstract. In view of the poor management effect of massive news and information, this paper proposes an intelligent push method of news and information for network users based on big data. Based on big data technology to identify network users news interest preferences, build network user news information classification recommendation algorithm, simplify the network user news information intelligent push steps, the experimental results prove that this paper design method in network user news information push has high practicability, can fully meet the research requirements, has certain application value.

Keywords: Big data · Network users · News and information · Intelligent push

1 Introduction

Online users ‘news and information facilitate people’s daily life to the greatest extent. Since the intelligent push of news and information has been put forward, it has attracted wide attention. Users want to get the information from their urgent needs in real time, and different users also have different needs for different products, so the needs for information becomes more personalized. However, the current network users’ news and information is characterized by complex structure, dynamic change and scattered distribution, which makes information overload and information fan become the key problems that hinder the efficiency of digital Internet [1]. How to quickly and comprehensively get the information needed by users from a large amount of data, improve the ability of news active information service and meet the personalized needs of users has always been a hot topic of information resource experts. The main purpose of information push is to solve the intelligent information push technology proposed by information overload. By reflecting users’ interest and preference information, and through personalized recommendation calculation, it can provide different users with different recommendation information. With the deepening of the research and application of personalized information push, this technology has been applied to various industries. For example,

Hu Yue and others designed an intelligent push technology for continuing education information based on deep neural network [2]; Huang Weihua and others designed an accurate marketing push system based on intelligent analysis of user characteristic information [3], but the above push method only evaluates user push information by testing user browsing history, which has obvious drawbacks, because when users browse computers, it will cause a large number of browsing history to be biased. If browsing history is used to evaluate push content, it will lead to inaccurate push information, Or the push information data is too complicated to be pushed, resulting in a delay in pushing. In order to solve the above problems, this paper designs an intelligent push method of news information for network users based on big data. A user interest model is constructed to accurately determine the user preference weight and label it. At the same time, depending on the final push mathematical model, the attenuation of user interest can also be observed, so as to better realize personalized push work and give play to the advantages of high push efficiency.

2 Intelligent Push Method of News and Information for Network Users

2.1 Screening of News and Interest Information of Network Users

Traditional media news client relying on the maternal information dissemination, in its specific interest screening process is relatively simplified, “innate genes” and “early development” are good, late interest screening is relatively easy its interest screening process can not well reflect the mobile news client interest screening, so the following research for aggregate news client. And both belong to the commercial website news client, in some interest screening ideas are much the same. In this study, for research convenience, it is divided into internal interest screening and external interest screening; the internal interest screening of mobile news client focuses on the interest screening of the editorial department in the production process of news client, from the detection of news sources, to the later evaluation. According to the mobile news client news interest screening process of time order, the internal interest screening is divided into early interest screening and late interest screening, early interest screening is before the news information push news interest screening, late interest screening refers to the news after the audience interest screening process, late interest screening is obviously different from the traditional media news interest screening process [4]. The flow chart of the internal interest screening and integration of China’s mobile news client, as shown in Fig. 1:

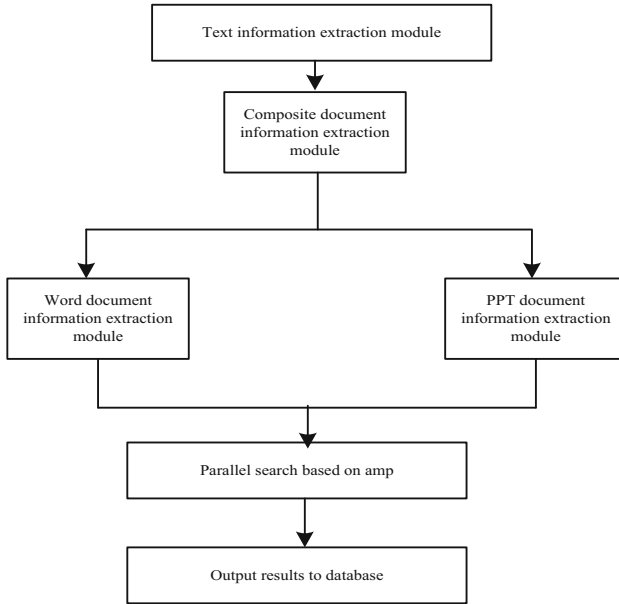


Fig. 1. News dissemination information screening step

In an information system, if the user's evaluation of the information is obtained directly from the user, the feedback will be described as explicit feedback. Although explicit feedback is easy to achieve, in most applications, the user is required to explicitly evaluate the relevance of all documents, which may lead to user disgust, limited learning of the user interest model, and reduced the availability of the whole system [5]. Although it will lead to indirect correlation between feedback and users' usability evaluation of all documents, it has more potential than explicit feedback in supporting user interest model processing, because it is easy to collect and will not affect users' normal lives. The collected temporal behavior preference characteristics are sent to the server and analyzed simply through the data parallel structure. When data is used as algorithm input, the collected data needs to be preprocessed first. The main purpose of preprocessing is to extract algorithm related data from a large number of data and convert it into the required format [6]. The main flow of data preprocessing is shown in Fig. 2.

- (1) Extract the relevant site data, and analyze the user behavior data of a certain site;
- (2) Filter useless data items, collect the data sent to the server in a specific log format, and be separated by “/ h” characters;
- (3) Confirm the recommended range, and filter according to the specific site URL naming rules;
- (4) Extract the content page.

After data preprocessing, the parallel characteristics of diversity key data are further analyzed, and big data is used to study the recommendation algorithm.

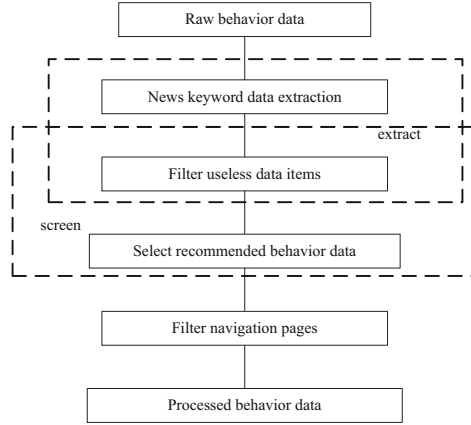


Fig. 2. Main process of news and information data processing of network users

2.2 Intelligent Push Evaluation Algorithm for News and Information

User interest is obtained by using a series of operations such as observing the news of network users, so the user interest expression should be consistent with the expression of network user news, that is, it can express users' interest in certain fields through the vector. When a user has multiple fields of interest, then the user's interest model should be a set of vectors composed of multiple vectors [7]. Set users have N areas of interest, each field of interest with m keywords, then you can use a m dimensional vector V to describe a field of interest, so, the user's interest model can use N vector life to describe t with interest with h network user news with vector, both can use the vector estimation for unified processing. So without the original information about the user's interest, the user's interest can be gradually learned by observing the user's action, so the original set S is empty, when observing the user is interested in a network user news application d , describe vector V by trid and the user's interest in the application R update vector set S . The specific process of user interest model construction is as follows, the original value of vector set S of user interest model is empty for all observed user interested news d for the following processing: application preprocessing, the corresponding analysis of the network user news documents, then with the specific language of document correlation processing, to give them large weights. Estimate the tid description vector V of application d . Estimate the user's interest R in application d ; Update the tid description vector V of application d according to the user's interest R in application d :

$$V_i = Vd - N - htid * R_i \tag{1}$$

A new vector set is formed by the vector in S and the new document vector V , and the similarity between the two random vectors in the new vector set is estimated:

$$\text{sim}(V_j, V_k) = \frac{V_i}{|V_j| \times |V_k|}jk \tag{2}$$

Combining the two vectors V with maximum similarity, The time start and completion identification is described by the timestamp, The time interval is mainly used

in describing the user’s reading time, Then the time ontology weight T (abel) can be described as the ratio of the user’s Internet access time and ontology browsing time, Data from the data in the preference analysis were obtained from the data ports in the mobile communication network, In addition to the data in the communication network, User-to-content access logs are also collected in the product operation platform, Taking the access data of the operating platform as the input of two-layer association rule data mining, Therefore, the data mining method, Get the network data of user interest.Let A and B be content item sets, and AT and BT be the types of A and B respectively, then the double-layer association rule set is g:

$$Z = \text{sim}(V_j, V_k)\{A \rightarrow B \rightarrow A \rightarrow B \text{ and } AT \rightarrow BT\} \tag{3}$$

where $A \rightarrow B$ is a basic content layer association rule, which means that when users access content set A , users will also access content set B with a high probability; $AT \rightarrow BT$ is a content type layer association rule, which means that when users access content type set AT , users will also access content type BT with a high probability. Basic content layer association rules are based on basic content fact table data extraction, and content type layer association rules are based on user access type facts. The behavior collection layer is the most important part of the overall design of the method hardware structure. It is mainly responsible for collecting relevant information between customers and projects and feeding back the recommendation results. For the incentive scoring mechanism designed by the method, it is necessary to actively score the user experience value. The data is relatively sparse, but the corresponding weight is large. Therefore, the following processing should be done for data collection: the corresponding final calculation should be made according to the actual situation, and the evaluation scores should be uniformly processed. Only in this way can the comparison be made in the final calculation, and the recommendation reliability is high. Table 1 shows the main behavior analysis of general users.

Table 1. User behavior analysis

User behavior	Scoring mode	Value size	Remarks
Direct scoring	Active evaluation	0–5	The preference is reflected by the value of the score, and 5 is the favorite
Direct voting	Active evaluation	0–1	No need for excessive user involvement
Direct forwarding	Active evaluation	0–1	Unable to draw a conclusion, the result is inaccurate
Direct comment	Passive evaluation	0–1	Unable to draw a conclusion, the result is inaccurate
Page dwell time	Passive evaluation	–	Not easy to use

The studies on the corresponding implicit scoring of news and information are shown in Table 2.

Table 2. Recommends the information recessive score

Execution name	Score (5-point system)
Search accuracy	3
Simple browsing	1
Browse carefully	3
Favorite page	4

Table 3. Experimental parameter setting

Parameter	Numerical value
node	10
CPU	2
Core frequency	1.9 GHz
Memory	8 GB

After the above behavioral and implicit scoring research, it can be dynamically adjusted through various pages, and jump into the relevant business areas. In the whole dynamic processing stage, large user behavior information will be carried. In order to realize all the above dynamic requests, interceptors should be set up to meet different business needs. The setting of interceptors needs to meet the characteristics of easy to expand. The unified own abstract interceptors mainly have the following three types:

See if the user has their own ID, if not, read the session control ID in the threshold related text file or create the session control ID to the text file and take the user ID as the key to recording all behavioral data.

Since there is a time difference when users query the information, it is necessary to record the dynamic request information and complete the page jump interception.

There are various scoring modes, and the special business logic relationship needs to be intercepted. The timing diagram design for news recommendation is shown in Fig. 3.

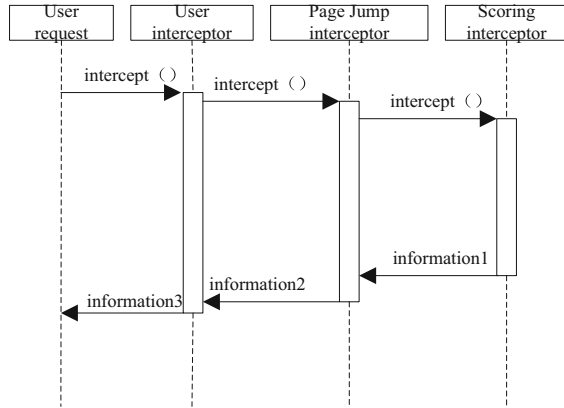


Fig. 3. News recommendation timing diagram

Data parallelism divides the training data into different Windows, and each window has a complete network model, using different data for training. In order to obtain a network model containing all the training information, the different network information needs to be synchronized, as shown in Fig. 4.

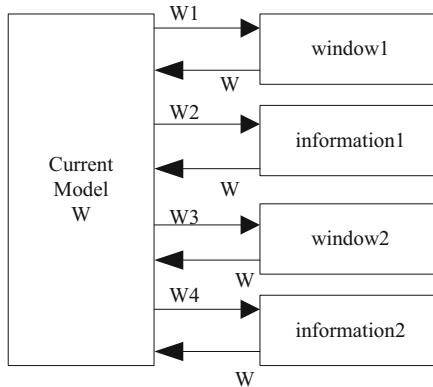


Fig. 4. Data Parallel Structure

After one iteration begins, each window obtains a new network model from the server and is trained. W is transmitted back to the server. The iteration can not stop until all nodes W are updated. If $1/n$ of size for a single window is used under n windows, then multiple windows are fully equivalent to a single window for training. After iterative processing, if a window data is abnormal, the whole training speed will be slowed down. In order to further improve the data use rate, it needs to be preprocessed.

2.3 Implementation of Intelligent Push of User News

Based on the user interest model, multimedia applications are pushed from two different directions. On the one hand, after the user operation record and the user interest model based on the multimedia application scenario are established, the content that the user may be interested in the multimedia application is searched out, and the content is predicted and pushed vertically by relying on certain branch prediction information score; On the other hand, through the user interest model, similar user interest news applications, application score prediction, and complete horizontal push can be found. The implementation of hierarchical recommendation algorithm is divided into three steps: establishing users, finding nearest neighbors, and calculating recommendation data. The user's personal preference information is counted, and the user preference matrix is obtained by analyzing the similarity according to the multi-level decision-making; After obtaining the user preference matrix, find the user's nearest neighbor; According to the user's preference for the project, set the nearest neighbor set of target user Z as S_u , and calculate the user's evaluation result of project B . the specific calculation formula is as follows:

$$P_{z,B} = P'_z \frac{\sum_{j \in S_u} sim(z, j) \times (P'''_{j,B} - P''_j)}{\sum_{j \in S_u} sim(z, j)} \quad (4)$$

In the formula: P'_z and P''_j represent the average score of user z and j on item B respectively; $P'''_{j,B}$ is the score of user j on item B ; $sim(z, j)$ indicates the similarity between users z and j . Combined with the project evaluation results of users, design the hierarchical recommendation scheme. Before project matching, the project ontology should be established first, so as to describe the relationship between different projects. The multi-level decision analysis algorithm is adopted to match users with the project ontology, and the projects with the highest matching degree are recommended to users. Constructing the project ontology in the hierarchical recommendation of digital news information can realize the correlation recommendation between different digital information. Defines an ontology as a quintuple:

$$Q = P_{z,B}(a'b', c', d', e') \quad (5)$$

In the formula: a' represents the concept set; b' represents the set of concept instances; c' represents the binary relationship between concepts; d' represents the constraints between examples and concepts; e' represents the inclusion relationship between any two concepts. After building digital news information projects according to the form of five ples, ranking according to similarity, and the projects with high similarity are recommended to the user's intelligent push system into three levels: application layer, processing layer and data layer. The application layer mainly has home page, knowledge scenario, push process interface; the processing layer is the core part of the system, mainly by hybrid push method based on knowledge scenario, is the key formation part of the algorithm (Fig. 5).

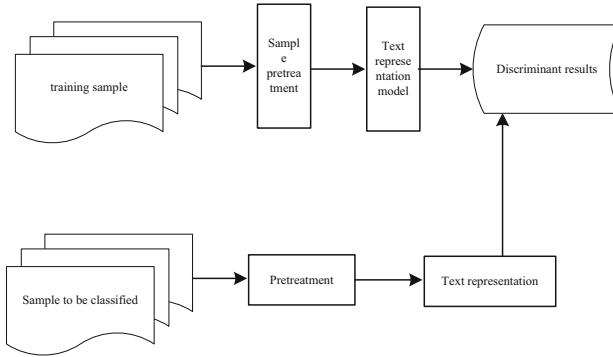


Fig. 5. Mixed push step of news and information

Intelligent recommendation method of network user news and information based on user interest model. From the application level, the intelligent recommendation system here needs to include home page, knowledge situation, push process, resource center, result analysis and other modules, so that users can see the push information when they log in. After the user logs in, the home page shows the most concerned information, push information and user history browsing information. The push information module is the key module of the system. The results are obtained by using the estimation based on the user interest model proposed in this paper. According to the encrypted content, the hierarchical implementation scheme is recommended, as shown in Fig. 6.

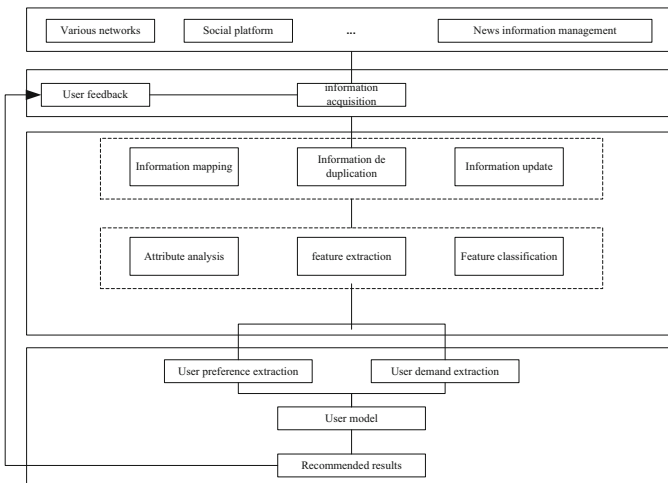


Fig. 6. Recommended implementation scheme of news and information classification

Based on the above preprocessing, the recommendation algorithm is studied under large-scale data distribution conditions, and the recommendation accuracy is low. In order to improve this problem, the recommendation algorithm based on big data is proposed. The implementation process is shown in Fig. 7.

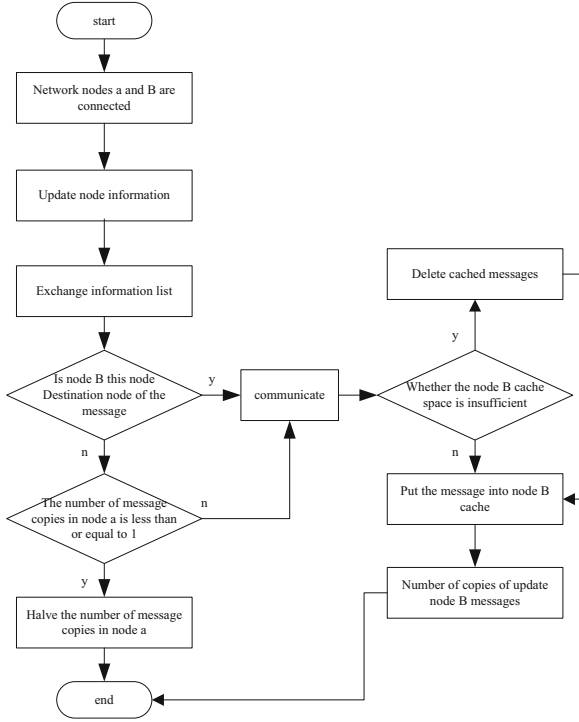


Fig. 7. Implementation Process

The specific implementation process of the algorithm is as follows: assuming that there are K information layer and N kinds of output layers, the weight parameter θ of the output layer is a $A \times B$ matrix, which can be expressed as $\theta \in C^{A \times B}$. The feature obtained after pooling the sample X is a K -dimensional vector, that is $f \in C^K$. The probability that sample X is divided into the Y -th category is:

$$P = (Y|X, C) = \frac{e^{(c_y \cdot f + q_y)}}{\sum_{h=1}^N e^{(c_y \cdot f + q_h)}} \quad (6)$$

In the formula: q_h represents the h -th offset term of the full connection layer, and the loss function can be obtained by maximizing the likelihood probability:

$$W = - \sum_y^R \log(p(g_y | x_y, \theta)) \quad (7)$$

In the formula: R is the training data set, and g_y represents the real data type of the y -th sample. In order to prevent overfitting, the convolutional layer neurons structure should be simplified with a certain probability to ensure that the weights do not work. After feature compression processing, based on the stability of the database storage space, the internal state and behavior control of the data can operate freely. Through the above process, the diversity of key data parallel recommendation schemes based on multi-level decision analysis is realized. The specific implementation process of the digital news information classification and recommendation scheme is: users multi item source into the feature model, store the collected information in the information base through various collection methods, and form a small user database. The core of the user model is established, and the information in the information base is integrated by using the information fusion method to extract the user's needs and preferences. The established project ontology is matched with the demand information, and the recommendation results are obtained, so as to realize the hierarchical recommendation of digital news information.

3 Analysis of the Experimental Results

The experiment mainly evaluated the influence of parameters and dimensions on the analysis of personalized recommendation model, randomly divided the data set with 80% as training set and the remaining 20% as test set, 256 GN memory CPU and 5 NVIDIA Tesla K40 as GPU, with a single precision peak of 4.25 Tflops, display memory of 12 GB and bandwidth of 280 Gbytes/s. Based on the past experimental experience, the value range of the parameter values is selected first, and these values may make the algorithm achieve good results, and then an optimal value is determined for the experiment. The experimental environment settings are shown in the table (Table 3).

Different from calculating the fixed number of nodes, the algorithm of this paper needs to calculate the maximum value of adjacent nodes. All nodes within a certain range are considered as current neighbors. Although the number of neighbors is uncertain, the similarity will not vary greatly, and it is appropriate to handle isolated nodes. The traditional method adopts the display scoring method to make news recommendation for users, while the dual recommendation method adopts the implicit scoring method to make news recommendation for users. The two recommended methods compare the accuracy of the recommended results under the influence of noise interference and human factors, respectively. The specific comparison is as follows. Comparing the traditional recommendation method with the recommended method in this paper, we find a more suitable calculation method for the similarity of news projects in scenic spots, and divide the data set for 20%, 40%, 60%, 80% and 90%, respectively. The data MAE error calculation results using different algorithms are shown in Table 4.

Table 4. Information recommendation error calculation results

Proportion of information	Traditional recommendation method	Recommended methods in this paper
20%	0.12	0.01
40%	0.11	0.02
60%	0.16	0.04
80%	0.16	0.05
90%	0.13	0.06

Based on the algorithm errors collected in the table, the calculation results for different similar algorithms largely depend on the sparsity of the data, and the error results of different methods under 5 experiments are shown in Fig. 8.

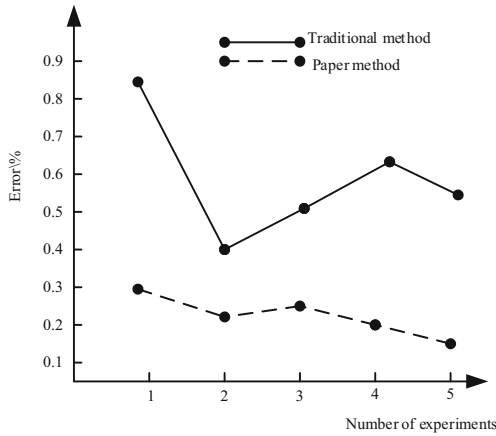


Fig. 8. The two methods compare the results using different algorithms

When traditional news intelligent recommendation algorithms extract news data, they are easy to be affected by noise interference. The final extracted data can not be counted completely, and the error is large, up to 0.9%. It is difficult to recommend effective information. The news intelligent recommendation algorithm designed in this paper is not affected by noise interference when extracting news data, and the final extracted data error is small, up to 0.32%, which proves that the recommended information is more accurate. To further prove the effectiveness of the two methods, the news recommendation accuracy of the two methods under noise interference is compared again, and the results are shown in Fig. 9.

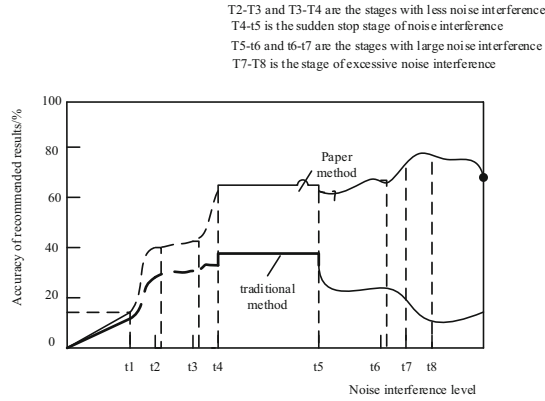


Fig. 9. The two methods recommend accuracy comparison under the influence of noise interference

It can be seen from Fig. 9 that the traditional recommendation method is affected by human factors, and obtains accurate recommendation information, resulting in vague types of news recommendation and low accuracy, with a maximum of only 38%. However, the method designed in this paper can obtain accurate recommendation information through implicit scoring, and uses a dual recommendation mechanism to improve the accuracy of recommendation results, so as to recommend recommendation results under the same environmental background. The recommendation accuracy is as high as 80%. The validity of the design method is proved again.

To sum up, the big data based intelligent push method of news information for network users designed in this paper has high push accuracy and high application value in the field of intelligent push of news information for network users.

4 Conclusion

In order to help users find valuable information more effectively and improve the accuracy of news information recommendation, this paper proposes an intelligent push method of news information for network users based on big data. This method builds user interest model according to different emphases of users, and provides users with the news information they need. Experiments show that the design method is influenced by noise interference, so that all the data are counted, the extraction effect is good, and the recommended results are more accurate. However, the recommendation time is not analyzed this time, which can not guarantee the recommendation efficiency of the design method. In the next study, further analysis is needed.

References

1. Peng-fei, C., Jie, L., Jun, Y.: Research on construction of real-time push service in cloud environment. *Comput. Technol. Dev.* **30**(03), 204–208 (2020)

2. Yue, H., Xiaonan, L., Bin, W., et al.: Research on intelligent push technology of continuing education information based on deep neural network. *Electron. Des. Eng.* **29**(14), 42–46 (2021)
3. Weihua, H., Xin, F.: Precise marketing pushing system based on intelligent analysis of user feature information. *Modern Electron. Tech.* **44**(06), 43–46 (2021)
4. Jing, L., Si-yu, F., Ya-fu, Z.: Model predictive control of the fuel cell cathode system based on state quantity estimation. *Comput. Simul.* **37**(07), 119–122 (2020)
5. Yingchun, Y.U., Yanjie, W.A.N.G., Xincheng, W.A.N.: Realization of accurate intelligent early-warning push technology based on WeChat. *Meteorol. Sci. Technol.* **48**(02), 195–199 (2020)
6. Xinnan, H., Bindong, S., Tinglin, Z.: The influence of geographical distance on the dissemination of internet information in the internet society. *Acta Geogr. Sinica* **75**(04), 722–735 (2020)
7. Wenhong, Z., Yiwen, S., Linxu, D., et al.: A study of the progress and development of web archiving for digital memory preservation. *Library Tribune* **40**(09), 42–52 (2020)