Characteristics Modeling and Interest Calculation of Learners’ Characteristics in Active Push Service of National Education Information Resources Under Big Data

Yu Yan, Bin Wen, Jianhou Gan and Jun Wang

ABSTRACT

In the traditional way of education information resources service, it is often learners learn the resources which are only provided to them by builders’. The meaning of information resources is simply to put into Internet to solve the problem of time and space, that is to realize long-distance learning and remote downloads. The sheer volume of data creates the dilemma of getting lost in the information world. In this article, we analyse the learners’ characteristics of education information resource in the national area and establish the characteristic model and feature calculation model of learners in big data environment.

Yu Yan, Master degree students. He received a Bachelor of engineering degree in software engineering at the end of 2015 in Henan University. In 2015, he was admitted to the National education informatization & Key Laboratory of Ministry of Education. His main research direction are the personalized recommendation of national education resources and semantic network.

Bin Wen, born in Yunnan province in 1978. He received Ph.D. degree in computer application technology from China University of Mining & Technology, Beijing, China, in 2013. In 2005, he was a faculty member at Yunnan Normal University, China. Currently, he is associate professor in Yunnan Normal University, China. Now, his research interest covers intelligent information processing and emergency management.

Jianhou Gan, born in Yunnan province in 1976. He received Ph.D. degree in Metallurgical Physical Chemistry from Kunming University of Science and Technology, China, in 2016. In 1998, he was a faculty member at Yunnan Normal University, China. Currently, he is professor in Yunnan Normal University, China. He has published over 40 refereed journal and conference papers. His research interest covers education informalization for nationalities, semantic Web, database, intelligent information processing.

Jun Wang, born in Yunnan province in 1988. He received master's degree in modern education technology from Yunnan Normal University, China, in 2012. In 2013 he was a faculty member at Yunnan Normal University, China. Currently, he is assistant research fellow in Yunnan Normal University, China. Now his research interest covers education informalization for nationalities and knowledge engineering.
1. INTRODUCTION

In recent years, with the support of a series of national policies, the national education resources gradually formed the trend of mass digitization, and the national education resources showed the characteristics of the quantitative multi-level of the sea. In 2014, Xiao-Hong Yang who is the professor of Northwest Normal University investigated the situation of national education informatization, it display that only 13.3% researchers join in education resources construction research and only 5.8% researchers join in the information network infrastructure in all national education informatization. In this year, the key laboratory of the national education information education ministry has formed a good environment in the integration of industry and research in Yunnan ethnic areas through the practice and construction of recent years. For example, in the project of heritage of wa, we have built a variety of digital resources such as the wa dictionary, Wengding Village roaming 3D scene. The laboratory constructed the national education information resource ontology library and the minority costume element library and other national culture and education digital resource library. After the massive national education resources are digitized, the learners accurately find the target information is more and more difficult, when the learners' personalized learning needs are not met, the massive data resources will lose their own significance. In addition, the characteristics of learners' interest are diverse and dynamic. This drives the national education information resource to use a new service method to alleviate the "information lost" in the big data environment.

Through sorting out the resource service model of big data in recent years, we can find that personalized push model has been the focus of researchers in recent years. Li bao[3] (2017) proposed the framework of personalized resource push service in education data environment. Using the analysis framework of Hadoop platform, education resource personality service framework is designed, from the dimension of learner characteristics analysis to push implementation approach analysis, it is more comprehensive model research. Geguili[4] (2017) proposed the modeling of user personalized interest based on situational awareness, M-C-W user interest model is put forward and update mechanism, Considering the user behavior, scene and other factors in order to better construct the dynamic user interest model. The educational information resource service model based on information push is a good bridge between learners and ethnic education information resources. The similarity of the learners and the characteristics of resources are matched to realize the active push of the national education information resources. From what kind of resources can learners accept and what kind of service they need, System analyzes learners' individual characteristics, learning tendencies, media preferences, learning objectives, learning readiness, cognitive styles, information processing methods, learning persistence and learning situations to determine the different learning needs of different learners. [5] From the point of view of resources, in the resource can provide what services and provide services, the system by analyzing the type of resources, knowledge presentation, learning objectives, learning time and knowledge of the degree of difficulty and other characteristics of information, to determine the appropriate user resources. [6] Lifei[7] (2016) put forward association rule to optimize personalized recommendation system, the data analysis is divided into offline and online processing two parts, and put forward a kind of user interest model based on
adjacent clustering processing optimization algorithm. In this article, the learners' characteristic dimensions are analyzed and the learners' feature models are gradually established.

2. THE ANALYSIS OF LEARNERS' CHARACTERISTICS OF EDUCATION INFORMATION RESOURCES

The people provided by the service of ethnic education resources not only have the native people who need to learn their own ethnic cultures, but also are the researchers who are concerned with the culture of ethnic minorities. Such researchers are part of the nation's researchers, as well as some other ethnic groups. Therefore, the national education resources of learners with diversity, the purpose of learning is different, in the analysis of the characteristics of learners, learners should be considered in the process of learning data, comprehensive analysis of learners' personality characteristics.

The extraction of the characteristics of learners, the most important is to get all data in the learning process, learners' generated data, including static data and dynamic data, the subjective intention interest feature static data representative of learners, like the chosen concerns and the selected resource items when learners enter the learning process. And dynamic data is representative of learners in the learning process of a series of data, such as learners' interaction information, operation information, etc. Objective performance data information of learners' interest characteristics. Remove some information that has nothing to do with education, From the basic information, media preferences, learning objectives, learning preparation, cognitive style, learning persistence, information processing methods to pay attention to and access, and ultimately build a model of learners' characteristics. Specific acquisition methods and data storage patterns are not the same in different feature classifications, such as, learners in the learning process, will produce a series of data, such as retrieving data record, learning and interaction between the message log files, log information is dynamic, after mining, the data information will get the learner's dynamic characteristics of the data. This study summarizes the characteristics of learners into five categories, namely learning schedule, input strategy, motivation characteristics, learning style, and ability foundation. Specific acquisition methods and data description are shown in TABLE 1:

| Characteristic class | Characteristic item | Way       | source                  | describe                                                                 |
|---------------------|--------------------|-----------|-------------------------|--------------------------------------------------------------------------|
| learning progress   | learning progress  | data mining| computer                | Observe the learning progress and obtain the data.                        |
| Investment in strategy | learning progress | data mining| computer                | Fuzzy sets                                                                 |
|                     | Degree of engagement | questionnaire | Learners, computer   | 0-40°                                                                     |
|                     |                     |            |                         | 40-80°                                                                    |
|                     |                     |            |                         | More than 80°                                                            |
| motive              | Interest in learning | Questionnaire and statistical analysis | Learners, computer | Observations are made through selected course resources.                |
|                     | Learning attitude   | Pre-school test | Learners, computer      | Fuzzy sets                                                                 |
### Learning expectation

| Learning expectation | Method                | Objectives                                      |
|---------------------|----------------------|-------------------------------------------------|
| Self efficacy       | Self-efficacy scale  | High, medium, low                               |
| Emotional state     | Data mining          | Fuzzy sets                                      |
| Learning motivation | Questionnaire and interview | Self-improvement, outside drive, cognitive style |

### Learning style

| Learning style | Method                | Learners                                       |
|----------------|----------------------|------------------------------------------------|
| Felder-Silverman learning style scale | Learners | Sensation / intuition, visual / verbal perception, inductive / deductive, impulsive / reflective, sequence / whole |

### Ability foundation

| Cognitive structure | Method            | Learners, computer | Complex data structures |
|---------------------|-------------------|--------------------|-------------------------|
| Cognitive ability   | Pre-school test   | Learners, computer | Observation, imagination, attention, etc. |

| Information literacy | Method            | Learners | Fuzzy sets |
|----------------------|-------------------|----------|------------|

## 3. THE NATIONAL EDUCATION INFORMATION RESOURCES DATA UNDER THE ACTIVE PUSH LEARNER DELIVERY SERVICE MODEL IN FEATURE MODELING

Facing massive educational information resources of ethnic minorities, the traditional manual methods of obtaining and processing data have obviously been unable to use data effectively. The use of new techniques and tools makes data acquisition efficient, and new methods of processing data make data analysis more comprehensive, which is a hot topic that researchers have been discussing in recent years. The data acquisition and processing mode of Hadoop open source distributed framework platform in big data acquisition and analysis has gained a lot of researchers' approval. Hadoop is mainly distributed processing of large-scale data, so even cheaper computers can build a distributed cluster environment. The distributed computing framework (Map Reduce) and distributed file system (HDFS) together form the core content of Hadoop. Among them, the latter is mainly responsible for the distributed storage of large-scale data, while the former is distributed parallel computation of massive data based on the latter. For the massive data generated by the interaction between learners and resource platforms, the distributed file system of open source Hadoop can be used for distributed processing of data. As shown in FIG 1:
In this mode, the data mining module of Hadoop can excavate the daily data of learners, use the underlying HDFS to store all data files on the node, and then analyze the data using the MapReduce data processing framework. Among them, HDFS has a complete data storage mechanism, generally consisting of Client, NameNode, DataNode, SecondaryNameNode, Client provides the invocation interface to the learner's data file, NameNode stores reference relationship between each data file and data block in memory, and DataNode stores small blocks of data, which is the smallest addressing unit in HDFS, and the size of a block is 64M, the data blocks are stored in parallel according to the different types of learners, and the data is passed to the MapReduce module, the MapReduce module will analyze the dynamic learning data and static data of the learners that are stored in the HDFS, dynamic data here is any data of the learner in the learning process. MapReduce will divide the data processing of the learning process into two stages: Map and Reduce. In the Map phase, the program processes a lot of data to get a series of keyword key for the learning behavior data and the value that corresponds to it, a keyword name corresponds to a value; Reduce phase, will get all the key words and the corresponding values for aggregation operation, all values corresponding to the same keyword Key are aggregated to together. This method can use keyword Key to bring together the data attribute values of the same type of learner, the same kind of knowledge unit corresponding to the learning process, When the dynamic data gathered in the data analysis layer is cleaned up, transformed, migrated and analyzed, the learners' characteristics can be constructed or the data of learners' learning behavior can be generated.

Data analysis and acquisition method of ethnic education resources learners based on Hadoop platform, This paper puts forward a learner characteristic analysis model under the active push service model of national education information resources service under big data. The learner characteristics model is divided into data acquisition layer, data storage layer, data analysis layer, business layer and application layer. As shown in FIG2 below:
The data acquisition layer uses the existing data mining technology to mine the learning data of the minority education information resources learners, including learners' retrieval information, operation information, learning record information in the learning process, etc. The data storage layer is the HDFS storage engine, which stores data from each node, including learners' static data, learners' retrieval data, learners' operation logs and other logs. Data analysis layer, using MapReduce data processing framework for data analysis, data source is stored in the HDFS nodes of each learner's characteristic data. In the business layer, the data of the upper layer are modeled by learners' characteristics, which are divided into static model of learners and learners' dynamic feature model. The application layer is mainly responsible for the calculation and feature matching of the learner's characteristic data. This part of data is stored in the database cluster, which is the data base for the final recommendation model.

4. CALCULATION OF LEARNERS’ INTEREST DEGREE IN ETHNIC EDUCATION INFORMATION RESOURCES BASED ON HADOOP PLATFORM

This paper puts forward a model of learners' characteristics of ethnic education resources based on Hadoop platform. The main function of the business layer is to construct the national education information resource service model of the information active push service in the application layer. In order to realize the national education information resources active push, One way is to analyze the learner's interestingness through the analysis of learners' characteristics, then match the national education information resources; another way is to offer proactive push services based on the characteristics of similar learners. Whichever way you choose to push it, the interest degree of the learner's characteristic data should be calculated.
first. This chapter calculates the interest degree of learners in the active push model of national education information resources by constructing the three-tuple model.

In the second part of the article has introduced five dimensions of learners' characteristics in the resource service model of national education, which is the learning progress, investment strategy, motivation characteristics, learning style and ability foundation. In addition to the learners' subjective characteristics data can be through the form of questionnaire and scale are analyzed, such as learning expectation, learning, learning attitude, the learner's interest in feature information is divided into explicit information and implicit information, such as retrieving information, operation information, learning process data, and so on. Among them, explicit information is the interest classification and learning evaluation that learners actively label in the process of learning, which is static data. By analyzing the interaction, attention and evaluation information of learners and systems, the explicit interest information of learners can be obtained. Hidden information is the learners in the learning process showed learning preferences and learning motivation, which belongs to the dynamic data, need means the use of systems analysis to obtain the learning in the process of learning resources, browse operation records, records of the three information retrieval were collected and analyzed, and then calculate the comprehensive weight. Eventually incorporating explicit information and implicit information analysis, access to the learners' learning preferences and the diversity and dynamics to calculate the similarity of each attribute, so it is concluded that the learner's interest model. In the form of data, the above data can be decomposed into the acquisition of static information data and the acquisition of dynamic information data.

4.1 The Calculation of Interest Characteristic of Learners based on Static Data

When learners interact with the system, they will actively choose some information they are interested in and leave their own evaluation information, such as paying attention to different interest labels, curriculum resources, curriculum evaluation and other information, The information is the explicit information displayed by the learner himself. After the data is acquired, the data can be processed, and all the resources with the same attention and related concerns can be presented. To construct the relationship system of learners, concerns and resources, using the triples \(<i,t,w>\) to express. That is, the learner \(i\) focuses on resource \(t\) for a resource classification \(w\). In resources, concerns are usually set. Therefore, the relationship between learners and resources can be obtained from the relationship between learners and concerns. Among them, the collection of interest of each learner shows the learner's interest preference, and the more repetition is concerned, the higher the learner's interest in related resources. The specific calculation methods are as follows:

1) In order to show the annotation of the item in the resource, the resource concern matrix is firstly established. As shown in formula 1:

\[
    f_{ijw} = \begin{cases} 
    1 & \text{if } p_{ijw} \geq 0.5 \\
    0 & \text{other} 
    \end{cases}
\]

(1)
2) The number of items that have been associated with the resource $w$ in the learner's learning process has shown the popularity of the concern, as shown in the formula (2):

$$f_{ij} = \frac{e_{ij}}{\sum_{j=1}^{n} e_{wj}}$$

(2)

Among them, $f_{ij}$ is the concern item, $e_{ij}$ is concerned about the number of times on the resource $w$; $\sum_{j=1}^{n} e_{wj}$ is all the concerns in the resource classification $w$.

3) Establishing learner resource scoring matrix $r(i,w)$. As shown in formula (3):

$$r_{iw} = \begin{cases} \text{User's evaluation over resource } w & \text{if } w \text{ is item } \text{rating} \\ \text{otherwise} & \end{cases}$$

(3)

Throughout the process, the need for learners to collect, grade, and evaluate data cleaning and transformation, and finally form a learner-resource score matrix, learners score matrix resources the size of the weight calculation, get the learner resources appetite. Among them, $r(i,w) = r_{iw}$ is the evaluation of learner $i$ to resource $w$, $m_{iw}$ is the learner's preference for resources. When the model of learners' preference concerns is established, the sequence of learners' attention is sorted by weight. According to the item and the resource, the database of the server is analyzed, and the learner's interest resource is gathered.

4.2 Calculation of Learners' Interest Characteristics based on Dynamic Data

The implicit information of the learner is usually obtained from the relationship between the learning operation behavior information, resource browsing information, learning duration, and so on. Among them, the operation behavior of learning information is the key to the direct reaction of user interest, browse information resources is the performance of the learner's interest, and the content of information is a direct response to the learner's preferences and interests. Compared with the explicit information of learners, learners' implicit information is more dynamic and immediate. Combining with explicit information, learners can construct complete information of learners' interest characteristics.

1) THE LEARNER OPERATES THE BEHAVIOURAL INFORMATION

In the process of learners' learning, a series of information about the operation behavior will be produced, which shows the learners' interest to a great extent, such as learners encounter interest resources will choose to download or copy operation; read or view a resource information in a certain period of time; learners will deliberately search for information that interests them, etc. Here, the learner's behavior information is transformed into a set of vector values. Set the $M$ collections for the learner behavior characteristic value, among them, the behavior of learners is divided into elements such as copy, download, click content, mouse drag, and so on, that is $M = (M1, M2, M3, \ldots, Mn)$. And use $(m1, m2, m3, \ldots, mj)$ to represent the behavior of a learner $Mi$. This paper mainly discusses the operation behavior of learners from three dimensions, namely, resource browsing object, visit frequency and duration. Combining the information of the three, the learners' interest in
behavior can be inferred. In the three-dimensional model of operation behavior of learners, assuming visit frequency $fq$ said at the time $t$ learners $i$ on resource $u$ browsing times, browsing time $t$ and search times reflects the level of interest in learner $i$ resources to a certain extent. At this point, $Mi(u,f,t)$ is used to represent the interest of the learner $i$ in browsing the resource $u$ in the time $t$. The specific formula of learner behavior is shown in formula (4):

$$M_{i}(u,f,t) = M_{i}(u,f,t) = M_{i}(u,f,t)$$

Among them, $M_{i}(u,f,t)$ stated in when $t$, Learners' interest degree in resource $u$, the learner's interest degree in accessing frequency $f$. The specific formula is shown in formula (4):

$$M_{i}(u,f,t) = M_{i}(u,f,t) = M_{i}(u,f,t)$$

2). RESOURCE BROWSING INFORMATION

The resource information that learners browse is a reflection of learners' interest directly, and a relatively important feature item will be given a higher value. If the frequency of the feature items of two resources is the same in a resource, then the less frequent items in other documents will assign higher values. Suppose that $tf_{ik}$ is the frequency of the occurrence of the characteristic item $D_i$ in the resource $t_k$, and $id_{jk}$ represents the resource frequency of the feature item $t_k$, which is the index of the frequent occurrence of $t_k$ in the resource set according to the resource. The improved interest expression is selected here, and the expression is calculated as shown in formula (5):

$$R_{i}(resource) = tf_{ik} \times \log(N/df_{ik} + 0.1)$$

Where $N$ is the number of the training set contains all text, text frequency $tf_{ik}$ said the training of $t_k$, in order to avoid the value of the training of $t_k$, to maintain a balance with the logarithmic function is 0.01, in order to avoid when $N=0$, log is equal to 0.

3). LENGTH OF STUDY

In the learners' learning process, the length of learning is the main interest of learners to a resource. In this paper, the time factor is introduced to judge the validity of learners' learning. Suppose that the resource item that the learner accesses is $m$, the concern item that the resource item accesses is $[t_1, t_2, ..., t_n]$, the length of time spent on the concerns is $[t_1, t_2, ..., t_n]$, the time factor is introduced here to calculate the value of the time effect factor as the learner's interestingness, The concrete formula is shown in formula (6):

$$\lambda(t) = \begin{cases} 0 & \text{if } \Delta t < 1 \\ 1 - \frac{\Delta t}{t_{-1}} & 1 \leq \Delta t \leq 30 \\ 1 & \Delta t > 30 \end{cases}$$

Among them, the timeliness factor, $\lambda(t)$ said learners’ attention in resources in time, $\lambda(t) \in [0,1]$, here $t_0$ for learners to start learning time, set $t$ to the present moment, can come to learning when the length of the difference between the two, namely $\Delta t = t_k - t_{k-1}$. The learner's timeliness factor fully considers the decline of learners' interest with time, which effectively improves the accuracy of interestingness calculation.
In summary, the change and intensity of interest are reflected in the calculation of the interest degree of the dynamic information to obtain the learner's information. Then the recursive formula for the calculation of the dynamic interest of the learner is:

\[ m_{xw} = \lambda(t) \times R(\text{Operation}) \times R(\text{resource}) \]  

(7)

4.3 The Calculation of the Comprehensive Interest of the Learner

Through the analysis of learners' explicit information and implicit information calculation can draw the corresponding static and dynamic characteristics of the learner model feature model, feature model learning interest is the key content calculated on the basis of learning interest, interest determines model learners interested in information selection. In the comprehensive analysis of learners' interest degree, we need to consider the subjective weight and objective weight of learners. The subjective weight of learners is mainly expressed by learners' preference for subjective ideas. For example, in the previous analysis, learners are interested in the part of their preferences, which is the subjective choice of learners. Therefore, the study of subjective weight can be based on the study of explicit information. Accordingly, the objective weight is mainly manifested in the learners' unconscious preference in the learning process, that is, learners' implicit information.

When calculating the learners' comprehensive interest, we should first consider the proportion of the similarity vector set of explicit interest and implicit interest. Using the constraint ratio to adjust the subjective explicit weight and objective implicit weight of the learner, the results of the first two sections are integrated to update the learners' comprehensive interestingness in the constraint ratio change.

**Definition:** The constraints of learners' comprehensive interestingness are expressed as follows:

\[ p = \mu m_{xw} + (1 - \mu m_{xw}) \]  

(8)

Among them, the learners' comprehensive weight \( p \) satisfies the condition \( 0 \leq p \leq 1 \). \( \mu \) is the weight scale factor, it is used to adjust the subjective explicit weight and objective implicit weight of the index. The weight proportionality coefficient satisfies \( 0 \leq \mu \leq 1 \). In general, when the learners' comprehensive interest and their concerns are consistent, this shows that the interest degree is relatively stable, and the value of \( \mu \) is relatively large. Correspondingly, when the learner's interest degree is larger than that of his own concerns, the value of the \( \mu \) is relatively small. It should be noted that the value of \( \mu \) can be 0, which means that learners do not have any explicit information tagging, and it is necessary to dynamically mine the specific interest characteristics of learners.

5. SUMMARY

Under the environment of big data, education has undergone tremendous changes. Established in national education and Resource Recommendation
Model of resource matching is necessary and urgent, this paper through the analysis of large data model, constructs the calculation model of minority learners characteristic model, interest, do the basic work for the future of personalized service. Of course, there are also shortcomings, Examples are not introduced for analysis, to match the resource analysis, it will be necessary to study the next step in the research direction, resource matching and learning from the perspective of similarity matching and gradually build a national education resource service personalized push model.

ACKNOWLEDGEMENTS

The research is supported by a National Nature Science Fund Project with Nos.61562093, Nos.61661051, Key Project of Applied Basic Research Program of Yunnan Province Nos. 2016FA024 and Project of Applied Basic Research Program of Yunnan Province Nos. 2016FD022.

REFERENCES

[1] F Q Yang and X H Yang 2014 The research status and development countermeasures of national education informationization in China. Modern remote education research, 04:71-79+88.

[2] J J Chen, Y Sun, L Tian and X Sun 2013 Analysis of the status quo of the digitalization of ethnic education information resources and Research on the attribution and Countermeasures . China Educational Technology 07:74-79+85

[3] B Li and D H Zhang 2017 Design of personalized resource push service framework under education big data [J]. Distance Education in China 09:62-69+80.

[4] G L Ge, L Y Yuan, X C Wang. 2017 User personalized interest modeling based on context awareness .Application Research Of Computers; 34(04):995-999.

[5] Grubert J, Langlotz T, Zollmann S, et al 2017 Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. IEEE Transactions on Visualization & Computer Graphics PP(99):1-1.

[6] Champiri Z D, Shahamiri S R, Salim S S B. 2015 A systematic review of scholar context-aware recommender systems. Pergamon Press, Inc..

[7] F Li 2016 Personalized recommendation system based on association rules optimization [J]. Journal of Inner Mongolia Normal University (NATURAL SCIENCE CHINESE EDITION), 45(04):515-520.

[8] Q Jiang, Y Zhao, P J Wang and L P Wang 2015 Personalized adaptive online learning analysis model and Implementation Based on big data. China Educational Technology,(01):85-92.

[9] S He, X L Feng, Q H Wu, T CXiong and R P Li 2017 Research on library personalized service based on user behavior modeling and big data mining. Library and Information Service;61(01):40-46.

[10] X M Yang, S C Chen and S S Tang 2017 Regional education data network construction and key issues in the era of big data. E-education Research; 38(01):37-46.

[11] T H Sun, A N Li, M Li and Q S Zhu. Research on Hadoop based distributed improved clustering collaborative filtering recommendation algorithm. Computer Engineering and Applications 51(15):124-128.

[12] X Q Gu, L W Zheng, J Jian 2014 To educational big data: the acquisition and sharing of learning experience data based on xAPI specification. Modern Distance Education Research (05):13-23.