Research Article

Network Education Resource Information Sharing System Based on Data Mining

Xinlei Yuan

School of Education Science, Mudanjiang Normal University, Mudanjiang 157011, China

Correspondence should be addressed to Xinlei Yuan; yuanxinlei@mdjnu.edu.cn

Received 17 February 2022; Revised 11 March 2022; Accepted 16 March 2022; Published 7 April 2022

Copyright © 2022 Xinlei Yuan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To enter a new stage of development of education informatization, network education has become a global trend of education training, the development of lifelong education, and resourcesharingplatformforhigher vocational colleges to some extent, with network technology and multimedia technology as the core of the rapid development of modern information technology. The sharing of network education resource sharing system and the application in the whole advancement provide the conditions, it promotes the construction and development of higher vocational education resource system. This study examines the problems that exist in the construction of higher vocational college education resources, such as unreasonable management of education resources, unclear goals of higher vocational colleges, and unclear school-enterprise cooperation, based on the research status and achievements of higher vocational college education resources at home and abroad. On the basis of combing and analyzing the related concepts, theoretical basis, and construction principles of higher vocational education resources, the paper designs the network sharing system of higher vocational education resources in detail from its design objectives, design ideas, user needs, overall framework system, hierarchical structure, and functional modules. The core part is the realization of system function modules, including the resource management function module, system management function module, and user management function module. At present, domestic scholars have carried out a lot of research on the construction of basic education network resources. Although the research focuses on resource allocation, coconstruction and sharing mechanisms, and the current state of building and development methods, it also serves as a guide for the development of basic education network resources. However, most of these research results analyze problems from the qualitative perspective, and most of the research results are speculative research at the macro level, lacking practicality. With the rapid development of network education, the construction of network resource libraries is developing in the direction of intelligence and individuation to meet the different learning needs of learners. Based on the analysis of the problems encountered by learners in using the resource library for learning, this paper puts forward the application of data mining technology (DM) to guide the construction of the resource library and discusses how to improve the organizational structure of the resource library, improve the utilization rate of the resource library, and adapt to personalized learning. At present, the system is still in the system testing and experience stage, through the continuous testing of relevant functional modules, further improves the function of the sharing system, and constantly increases the scalability of the system function, in order to establish a perfect, comprehensive, and nationwide higher vocational education resource network sharing system.

1. Introduction

With the advent of the information age, more and more people put the learning process on the network, using the massive information provided by the educational resource library for learning. Different learning interests and learning objectives have different requirements on educational resources [1]. What users need is personalized information with the highest reference value and high quality, rather than related information with mixed advantages and disadvantages and low quality. Users’ needs have changed, and higher requirements have been put forward for the quality of educational resources [2]. Resource library builders have also realized that people no longer only care about the “quantity” of information resources as
before, but pay attention to the time cost of using network information. Payment for the use of the cost is only one aspect, and user's waste time to pay a larger price. In order to better adapt to users, resource library builders must respect users, study users' behavior habits and interests, provide rich and high-quality information, and provide users with more needed, personalized education resources [3].

At present, there is still no unified concept of what an education resource bank is. It is generally considered that an educational resource bank is a collection of various educational resources. The construction of the education resource database mentioned here mainly refers to the construction of an online teaching resource database. Online teaching resources mainly include the following contents [4, 5]: Material resources: multimedia material database, curriculum database, learning document database, case database, electronic documents, a question bank, and so on.

Data mining technology can effectively solve the following three problems in the construction of resource databases.

1.1. Improving Resource Utilization. Learners choose learning resources randomly or blindly at first. To improve the utilization of resources, one is to let more learners know about the resource library, that is, the mining of potential users; at the same time, users who have been using the resource library should not be lost, that is, users reside [6]. How to make learners stay longer on their own resource library website is a challenge for resource library designers. In order to prolong the resident time of learners, we should understand the browsing behavior of learners, know the interests and needs of learners, and dynamically adjust the web page to meet the special needs of learners. Through the mining of learners' access information, learners' browsing behavior can be known, so as to understand their interests and needs [7]. For example, the sequential pattern discovery technology in data mining can be used to discover learners' navigation behavior in the form of pages on the site and dynamically adjust the site structure for learners. Make the connection between the relevant files accessed by learners more direct, so that users can easily access the page to be visited, increase the convenience of access, leave users a better impression, and increase the probability of the next visit [8].

For example, through the analysis of the results of 15 volunteers web mining, found that there are some learners who have spent some time browsing “multimedia material,” “multimedia courseware” page after analyzing cluster as a set of them, know that learning is interested in multimedia content, can adjust the page and page content, at present this kind of learners in the content of the page [9]. Add multimedia content and links to meet the needs of learners. Other characteristics of learners, such as their education level, location, and age, can also be clustered to present resources that meet their requirements for learners [10].

1.2. Improving the Organizational Structure of Resources. Improve resources organization structure, realize the personalized service, is the highest goal of the construction of the repository, as much as possible is to make each student feel when browsing repository site is the site the only user, as much as possible to meet the needs and interests of each student, and constantly adjust the resource organization structure to adapt demand change of learners. Make the learning interface required by learners easier.

1.3. Improving Download Speed. If learners personalized original database each access transaction, there are many links to browse the web all at the same time to browse web pages and browsing B or learners often at A web page to browse the web after B, are surfing the web can be thought of as A and B have a certain correlation between if a user's visit to ask for browsing the web. A. Then, browsing web page B is likely to be valuable to the user and can be added to the list of recommendations. By using mathematical statistics and association rules and other methods, the results of mining and analysis are put into the personalized database. When learners enter the system next time, the system can provide them with pages that meet their learning needs according to the database. In addition, since the learning needs of learners are constantly changing, it is necessary to collect the original data at a fixed time and combine the historical information in the personalized database of learners to conduct personalized analysis again, generate results, and update personalized data.

The following is a summary of the research: Section 2 discusses the research status; Section 3 discusses the functional module design of network sharing system; Section 4 shows experimental results of the proposed concepts; finally, the conclusion is given in Section 5.

2. Research Status

2.1. Problems Existing in the Current Online Education Resource Library

(1) Low resource utilization rate: learners cannot find resources to solve problems or resources suitable for their learning ability when they use the resource library for learning, both of which affect learners' utilization of resources. The resource that cannot be found to solve the problem can be divided into two situations [11]. One is that the resource does not exist, and the other is that the resource exists but has not been discovered. Learners search for a clear purpose, if the search is exactly the gap of resources, then it is bound to be fruitless. Even if resources exist, it involves the process of reselection or comprehensive analysis of resources. This process includes the intelligence and credibility of the information agent and the reliability and validity of the resource itself. The quality of solving each problem directly affects learners' access to resources [12]. The following reasons are the main reasons why learners cannot find resources suitable for their learning ability: first, there is a difference between the existing knowledge structure of learners and the difficulty of resources. The smaller the difference, the higher the utilization rate. Once the differences are too great,
the best resources are of no use to the learner. Second, learners’ cultural background affects the utilization rate of resources. Language barriers and different modes of thinking are related factors that affect the utilization of resources. Third, the relationship between the age of learners and their ability to accept knowledge, and the relationship between learning subjects and the learning effect based on network resources also have an impact on the learning process.

(2) Inappropriate learning process for personalized learning is the fundamental problem existing in the current network resource library, which is mainly reflected in the lack of diversity in the organization of resources. The existing resource library is still unified in the presentation of content, although students can find what they need according to their needs. However, it cannot meet the needs of analytically oriented personalized learning services. Dynamic or static educational resources are difficult to adapt to the needs of individual learning. Different learning objectives and interests have different requirements for learning resources, and the richer the resources, the more important the accuracy and efficiency of choice. Therefore, it is necessary to build a personalized learning service system to assist learners to screen information, discard the information that has no value or little impact on their learning, accurately and quickly obtain information resources that meet the needs of learners, and meet the needs of personalized learning.

(3) Difficult resource retrieval and long download time: the current network environment is one of the reasons leading to difficult information retrieval and slow download, while the organizational structure of the resource library and the setting of search and query functions is another reason leading to this phenomenon [13]. At present, most resource libraries are built according to the storage type of resources, which is easy to build but difficult to use. Learners need to constantly jump from one page to another to query related knowledge points, and the search function is single and it is difficult for learners to immediately obtain the required content. There are significant differences in personality among different learners, and most of the current resource banks are in the resource sharing stage, which cannot meet the personalized and efficient learning requirements of learners. The application of data mining technology to the construction of network resource library can effectively change this situation.

2.2. Methods of Data Mining. Common data mining methods are as follows.

2.2.1. Association Analysis. Association analysis is a practical data mining technology, which refers to finding useful

dependence or association knowledge from a large number of data sets. Association rules can mine interesting knowledge and patterns from a large amount of transaction data or relational data.

2.2.2. Decision Tree. Decision tree is mainly used for induction and classification based on attribute values of data. Its algorithms include ID3, C4.5, CART, and CHAID. Currently, SLIQ and SPRINT, two new algorithms, can be used for decision tree induction based on very large training sets and can deal with classification belonging and continuity attributes.

2.2.3. It Consists of Three Basic Operators. They are Reproduction (selection), crossover (recombination), and mutation (mutation).

2.2.4. Bayesian Network. Based on Bayesian theorem of posterior probability, Bayesian network is a method based on statistical processing of data. It has the functions of classification, clustering, prediction, and causality analysis.

2.2.5. Rough Set Method. It is a new mathematical tool, which is often used to deal with ambiguity and uncertainty, finds the inherent structural relationship between inaccurate data or noisy data, and also can be used for feature reduction and correlation analysis.

2.2.6. Neural Network. Neural network is one of the most commonly used data mining techniques, which seeks to develop and test neural computational simulations. Neural network can be subdivided into feed-forward, feedback, and self-organizing neural network, which has the functions of optimal calculation, clustering, and prediction.

2.3. The Construction Status of Online Education Resource Database in China. Many schools in our country attach great importance to the construction of website and network education resource database. It provides a wealth of educational resources. Students also enjoy access to learning materials and distance education through the Internet, but take a look at the construction of educational resources. The following problems are common:

Educational resources provided by the school. Most of them are teaching courseware, electronic teaching plans, and open class videos. Its teaching content is basically similar, for example, the distance teaching of primary and secondary schools is basically based on Chinese, mathematics, foreign language, physics, politics, biology, and other main learning subjects, and what university opens is current popular class major, such as computer, economy, etc.; when learners search for a topic, they will get a large amount of similar or even the same information, which requires a considerable amount of time for manual secondary retrieval. Filter useful information.
There are differences in knowledge structure, expression, and understanding between resource developers and learners, even though learners can find a large number of educational resources on the Internet. However, if the difference between the two is too large, the resources obtained cannot be well used. It may even have the opposite effect. Progress in knowledge cannot be made even through long hours of hard work. The utilization of resources will be reduced.

The basic problem of network education is that network education resources are not suitable for personalized learning. At present, the form of online education resources in China is single. The content is not systematic and targeted. Many resource media have simply changed form. The content is still a copy of the traditional face-to-face teaching materials. Lack of teaching design and pertinence to students’ learning. It is not suitable for students to learn individually and independently.

3. Functional Module Design of Network Sharing System

According to the functional requirements of network sharing system, in order to meet the various needs of user objects, the design of the system platform should be divided into three modules: resource management module, including resource audit, upload, download, preview, search, statistics, collection, evaluation, delete, data backup, and other functions; as indicated in Figure 1, the user management function module includes general user and administrator user rights allocation, as well as the system management module, which includes basic data management, system configuration, and security management.

Kohonen neural network is a neural network trained by unsupervised learning method. When the vector enters Kohonen neural network, the neuron with the minimum Euclidean norm value as the winning neuron. If \( I(x) \) is used to specify the winning neuron, \( I(x) = k \), when \( ||w_k - X|| < ||w_j - X||, j = 1, 2, 3, \ldots, n \), the learning process of the winning neuron can be expressed as

\[
w_j(n + 1) = \begin{cases} 
    w_j(n) + \eta(n)\left[ X - w_j(n) \right], & \text{if } I(x) = k, \\
    w_j(n), & \text{otherwise}
\end{cases}
\]

where \( f(x)(n) \) represents the neighborhood function of the other iteration.

Suppose that the number of applications sent by learners to page A is \( N_i \), the number of applications sent to page B is \( N_j \), and the number of simultaneous applications sent between any two pairs of pages \( (A, B) \) is \( N \). So, calculate

\[
P(A | B) = \frac{N_{ij}}{N_j}, \\
P(B | A) = \frac{N_{ij}}{N_i}
\]

Construct the distance vector matrix between pages, where the distance between pages is defined as

\[
D(A, B) = \left[ \frac{1}{P(A | B) * P(B | A)} \right]^{1/2}.
\]

The similarity between the calculated pages is expressed as

\[
SIM(A, B) = \frac{P(A | B) + P(B | A)}{2}
\]

According to the characteristic vectors \( T \) and \( T_1 \) (student number) of learners’ personalized information, the characteristic matrix \( T \) of learners’ personalized information is constructed.

Using the Euclidean distance formula of fuzzy concept,

\[
d(x, y) = \left( \sum_{i=1}^{n} |x_i - y_i|^2 \right)^{1/2}.
\]

Then, the Euclidean distance of the feature coding of a learner and learners similar to the learner is

\[
D(x, y) = \sum_{k=1}^{n} |u_x(x_k) - u_y(y_k)|.
\]

If \( D(x, y) \) is small, it indicates that the characteristics of this learner are very similar to that of another learner, and then the conversation mode of another learner can be recommended to this learner.

In addition, FCCRD algorithm is used to cluster user sessions to obtain user session clustering. The access patterns among users in the same session cluster should be as similar as possible, while the access patterns among users in different session clusters should be as different as possible. Given user session cluster \( C \) and significance threshold \( P \), the
calculation method of overall usage feature PRC of user session cluster C is as follows:

\[ pr_c = \{ < p, \text{weight}(p, pr_c) > | p \in P, \text{weight}(p, pr_c) > \} \]  \hspace{1cm} (10)

Assume that the current session S can be represented as \( S = \{s_1, s_2, \ldots, s_n\} \). Overall usage characteristic C can be represented as \( C = \{w_1^c, w_2^c, \ldots, w_k^c\} \). Cosine similarity function is used to calculate the matching coefficient between C and S:

\[ \text{Match}(s, c) = \left[ \frac{\sum_k (w_k^c \cdot s_k)}{\sqrt{\sum_k (s_k^2)^2 \sum_k (w_k^c)^2 \cdot 1/2}} \right] \]  \hspace{1cm} (11)

Assuming that Recolnmend (P, S) is a metric to determine whether a page is recommended for learners, you can set a threshold or set the maximum number of recommended pages. Select the largest pages of Recolnmend (P, S) and recommend them to learners. This way of adding links varies depending on how users recommend pages at different times of visit. Then, the recommendation coefficient of judging whether a web page is recommended to learners by using web structure mining technology and web usage mining technology can be expressed as

\[ \text{Recommend}(p, s) = [\text{weight}(p, c) \cdot \text{match}(s, c)]^{1/2} \]  \hspace{1cm} (12)

The main function of the model is to learn and track users’ personalized interests in the network teaching resource database system and filter teaching resources according to the characteristics of users’ personalized interests, so as to help users quickly and accurately search and recommend the teaching resources they are interested in in the massive network teaching resource database system. Figure 2 shows the structure.

4. Experimental Results and Analysis

To validate the effectiveness of the suggested approach, VC++ is utilized to create simulation tools. First is the simulator generated models (BA model and Waxman model) based on the network topology to generate network node distribution. Then, when the number of nodes was 200 to 1000, the performance of L_flooding and flooding algorithm in this paper was compared according to the number of redundant information and system transmission cost.

In order to improve the accuracy of the experimental number of cockles, the average value of the 10 nine cap VIII experiment was used for the number matching of each nine clam. Figures 3 and 4 show the outcomes of the experiments.

Experimental results show that in the same system scale, the number of redundant messages generated by this policy is reduced by nearly 1/3 compared with flooding and the transmission cost is better than flooding. In this way, a large amount of network bandwidth is saved and the quality of service of distributed application system is improved.

The author creates a discipline-media type two-dimen-sional table based on the statistical results in order to better comprehend the amount of resources of various media types in the courses of nine disciplines, as shown in Table 1 and Figure 5. From the perspective of horizontal view, all the subjects show the lack of text resources and comprehensive multimedia resources. The courses of art, physical education and health, comprehensive practical activities, and cross-field learning activities are generally lacking in audio resources, and the resources of pictures, animation, and application are also significantly insufficient. Vertically, comprehensive multimedia resources only exist in Chinese and literature and mathematics courses, but not in other
The user interface

Feature extraction module

Personalized teaching resource filtering

Personalized retrieval module

Personalized user interface adjustments

Teaching Resource Management

User model building blocks

**Figure 2**: Structure of personalized retrieval model.

---

**Figure 3**: Number of redundant messages and system size.

**Figure 4**: Transport overhead versus system size.
Table 1: Discipline-media type distribution of the number of resources available.

<table>
<thead>
<tr>
<th>Subject/media type</th>
<th>Text</th>
<th>Picture</th>
<th>Video</th>
<th>Audio</th>
<th>Animation</th>
<th>Application</th>
<th>Integrated multimedia</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>001 language and literature</td>
<td>1301</td>
<td>16617</td>
<td>7776</td>
<td>716</td>
<td>3167</td>
<td>6066</td>
<td>317</td>
<td>31703</td>
</tr>
<tr>
<td>002 math</td>
<td>6391</td>
<td>7377</td>
<td>7676</td>
<td>113</td>
<td>6330</td>
<td>3016</td>
<td>10</td>
<td>39316</td>
</tr>
<tr>
<td>003 human and social health</td>
<td>307</td>
<td>9163</td>
<td>3363</td>
<td>1363</td>
<td>3177</td>
<td>6116</td>
<td>0</td>
<td>33133</td>
</tr>
<tr>
<td>004 natural science</td>
<td>13</td>
<td>3067</td>
<td>736</td>
<td>71</td>
<td>10</td>
<td>3766</td>
<td>0</td>
<td>6716</td>
</tr>
<tr>
<td>005 technology</td>
<td>363</td>
<td>1167</td>
<td>1301</td>
<td>1313</td>
<td>717</td>
<td>661</td>
<td>0</td>
<td>6193</td>
</tr>
<tr>
<td>006 art</td>
<td>313</td>
<td>33</td>
<td>3710</td>
<td>0</td>
<td>1137</td>
<td>1113</td>
<td>0</td>
<td>6373</td>
</tr>
<tr>
<td>007 sport and health</td>
<td>0</td>
<td>70</td>
<td>177</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>339</td>
</tr>
<tr>
<td>008 comprehensive practical activities</td>
<td>1</td>
<td>0</td>
<td>337</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>339</td>
</tr>
<tr>
<td>009 cross-disciplinary learning activities</td>
<td>3</td>
<td>1</td>
<td>91</td>
<td>1</td>
<td>1</td>
<td>19</td>
<td>0</td>
<td>131</td>
</tr>
<tr>
<td>Total</td>
<td>7367</td>
<td>31136</td>
<td>30699</td>
<td>3170</td>
<td>13163</td>
<td>11136</td>
<td>337</td>
<td>101131</td>
</tr>
</tbody>
</table>

Figure 5: Distribution of subject-media species of available resources.

Figure 6: Download frequency ratio of resources in the same discipline.
courses. On the whole, there is a lack of discipline-media structure in the existing resources of canal.

The ratio of download frequency of resources in each discipline with four grades of A, B, C, and D was calculated, and the radar chart was selected to display the statistical results, as shown in Figure 6.

With a total of 4.46 million downloads, language and literature resources are the most popular, as shown in Figure 6. Among them, the download frequency grade is D, that is, frequently downloaded resources account for about 40% of the total resources of this discipline. Mathematics resources were downloaded 1.98 million times, accounting for 21.6% of the total number of resources downloaded, and mathematics resources frequently downloaded accounted for more than 50% of the total number of resources downloaded. In comparison, sports and health resources and technology resources are less downloaded, and resources rarely downloaded account for a large proportion of such resources. This shows that the downloads of traditional “main subject” resources in China still account for a large proportion of the total number of downloads, while the resources of sports and technology, which are in a weak position in traditional school education, are rarely downloaded, and most resources cannot be fully utilized.

5. Conclusion

A multidimensional, multiuser, and multilevel peer system is used to build basic education network resources. The system’s internal elements are interconnected and mutually constrained, and the system is connected to and influenced by external systems. Only by achieving the balance between the internal and external of the system can the BIAN realize the balanced and sustainable development of the basic Church network resources. Against the ecological education letter as a kind of new thinking pattern, used to guide basic education network resource construction, to make our work in the resources construction of bow needs, hardware environment, software tools, unifies, the resource service and the standard, policy system, the investment, the combination of network based teaching storehouse resources construction implement internal unity and external coordination. In research methods, as a result of the author of Master Data Mining Co., Ltd., also cannot do it for the various data mining algorithms and models become adept, flexible use, combined with Guan Ding data mining in network resources face to draw lessons from the experience of the project owner is very little, and national basic education resource network of large amount of data and higher requirements to the server hardware configuration, in the study. The author did not use multidimensional mining algorithm and did not conduct in-depth research on key technologies. The majority of existing online education resource service platforms offer a keyword search resource push mode, however this passive resource search mode simply looks for the degree of convergence between the target resource and the key provided, which represents the quality of the target resource. Further research can be done to see if there is a link between the times it takes to search for resources and the time it takes to download them.

Data Availability

The data used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares that he has no conflicts of interest.

Acknowledgments

This study was supported by the Education Department of Heilongjiang Province under Grant 1451ZD013.

References