Research Article

Optimizing Information Dissemination Model for Improvement of College Students’ Education Based on Learning Community

Jing Chang 1, Jian Huang 2, and Ying Hu 1,3

1 School of Mathematics and Economics, Hubei University of Education, Wuhan, Hubei 430000, China
2 School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan, Hubei 430000, China
3 Big-Data Modeling and Intelligent Computing Research Institute, Hubei University of Education, Wuhan, Hubei 430000, China

Correspondence should be addressed to Ying Hu; yinghu@hust.edu.cn

Received 10 June 2021; Accepted 21 July 2021; Published 19 August 2021

Abstract

In the context of the popularization and diversified application of information technology in higher education, efficient information dissemination has a significant impact on the learning effect of the learning community. Improving the efficiency of information dissemination and driving the force of learning to enhance the learning effect are the hot issues in the field of higher education data analysis. This paper proposes a new method of feature fusion using information entropy and ReliefF algorithm, applies the improved PageRank algorithm and K-means algorithm to optimize the information transfer mode, and finally develops a new and efficient network information model. The comparative test results show that the new model can complete the dissemination of the same amount of information with a smaller delivery ratio. This research result can play an advantageous role in information interaction feedback, curriculum quality analysis, and teaching information transmission.

1. Introduction

Learning activities will ultimately be attributed to learning output. High-quality learning output is affected by many factors, such as students' learning drive, learning methods, and learning environment. How to improve these factors to a certain extent is a problem worth studying [1]. A new round of worldwide information technology revolution is reconstructing people's way of life, learning, and thinking. Higher education in China is also facing unprecedented challenges and innovative opportunities. New information technology breaks through the time and space limitations of traditional teaching methods, and at the same time, a new kind of educational productivity is born. It is this new product that will affect students' learning drive and other factors, thereby determining the learning level of output quality. Combining the fact that information technology is used in the general trend of education and teaching, this article intends to analyze the improvement path of the effectiveness of information dissemination based on the characteristics of the learning community from the technical level [2].

About the learning community, Siemens et al. [3] introduced the theory of connectivity and stated that learning is no longer an internalized personal activity. The use of information technology has greatly changed the way of learning, and the process of learning is seen as a process of connecting specialized nodes and information sources. Regardless of the way of learning, learning itself is an information exchange and information receiving behavior, and the carrier and manifestation of its information dissemination is the learning community. As early as 1989, Brown [4] and other scholars introduced the concept of a learning community, believing that learning is full communication between different individuals and groups such as peer learners, teachers, experts, and families and is a social construction. Hunter [5] described that in the learning community, driven by common interests and goals,
individuals will continue to contribute meaningful information, resources, and knowledge to the group. Group members can not only interact but also learn from each other, and the learners have a sense of belonging. Kowch et al. [6] added that the group members will have a sense of mutual dependence and mutual beliefs resulting in a stronger learning drive. Johnson [7] and Seinkuchler [8] believed that learning communities are not necessarily the result of spontaneous formation but can also be the result of teaching design, and they can be designed to transform and improve operational efficiency. The activities of all human beings are affected by an internal driving force, which is an internal stimulus. When the organism has a need, the stimulus-response caused by the internal driving force will actively pursue the satisfaction of the need. Ausubel declared [8] that internal drive comes from three aspects: one is a cognitive drive, which refers to the satisfaction of acquiring knowledge and solving problems; the other is the self-improving drive, thinking that academic achievement is equivalent to a kind of the corresponding status; the third is the attachment drive, which believes that hard work and study are for the recognition of the elders and peers. In terms of learning activities, internal driving forces include strong learning interests, clear learning goals, and satisfactory learning effectiveness. According to Hwang et al. [9], the learning community provides learners with space and tools for communication, promotes the establishment of social relationships, and encourages learners to generate learning motivation.

How to use the communication network to spread information more efficiently, improve the efficiency of the learning community, and enhance the individual learning capability is a hot issue in the field of higher education big data analysis. In the current popular interactive communication network, the influence of individuals and the relationship between individuals directly determine the ability of individuals to disseminate information. Within the network, individuals with high influence play a key role in the dissemination of information. A large number of studies have shown that the dissemination of information within the same characteristic groups is more efficient. By establishing an information dissemination model, using real information data to verify the pros and cons of model assumptions is a very effective way to achieve this goal [10–12]. In this context, this article conducted a simulation study on the characteristics of university information network groups and information dissemination methods. This article can deepen people’s understanding of the information dissemination methods of college learning communities and make better use of information dissemination based on information networks. Furthermore, this study will enhance the internal drive of the learning community and ultimately serve the educational practice of higher education.

The rest of the paper is structured as follows. Section 2 describes the proposed model framework and data statistics. In Section 3, the simulation model and future fusion methods are explained. Section 4 is about results, and the conclusion is given in Section 5.

2. Model Framework and Data Statistics

2.1. Model Framework. This paper studies the information dissemination process based on the characteristics of the learning community and finally establishes an information network information dissemination model based on the feature fusion group preclustering. The overall structure of the proposed model is shown in Figure 1. Initially, the data were preprocessed to avoid unwanted and data points from the collected dataset. Next, the information entropy and ReliefF algorithm were employed to calculate the feature weights of each attribute as follows. The improved PageRank and K-means algorithms were applied to perform the preclustering of the feature group, and then the delivery rate and user placement rate were computed.

2.2. Data Statistics. The data used in the research were collected from desensitized college students. The collected data were preprocessed, retaining active communication variables, passive communication variables, and communication duration variables, accumulating active communication. The communication duration of users with the same passive communication attributes was also taken into account, and the outliers and missing values were excluded. The distribution of the communication duration attribute data after processing is shown in Figure 2.

It can be seen from the figure that the Q-Q points are more evenly scattered around a straight line. It is also evident that the attribute data follow the normal distribution and the data are of good quality.

3. Information Dissemination Simulation Model

In this section, we give a mathematical overview of the information entropy and PageRank algorithm. We describe the main notions, definitions, and theoretical results of the PageRank method that we used.

3.1. Feature Fusion Based on Information Entropy and ReliefF Algorithm. In the past, when performing cluster analysis on multidimensional data, it was generally assumed that the weight of each value of data attribute should be the same, that is, it was considered that different attributes had the same effect on clustering. However, this was an unreasonable assumption [13, 14]. The effect of attribute features in the data on cluster analysis must be greater in some dimensions and smaller in other dimensions. The present study shows that based on this assumption, users can be clustered more accurately. By combining information entropy and the ReliefF algorithm [15], we calculated the feature weights of each attribute as follows.

Use information entropy to describe the amount of data information. Suppose there is a dataset $A = \{a_1, a_2, \ldots, a_n\}$ and a probability measure $p_i = P(A = a_i)$, then for the given dataset, the information entropy can be expressed as follows:
For any two features $A_j$ and $A_i$, if $H(A_i) > H(A_j)$, it means that the feature $A_j$ plays a larger role than the feature $A_i$ in the clustering process.

The basic idea of the ReliefF algorithm is to select a sample $x_i$ randomly selected from the sample set, select $k$ neighboring samples of the same class $H_i$, select $k$ samples of different classes at the same time $M_i$, and calculate the feature weight of the attribute according to the weight using the following equation:

$$W(j) = W(j) + \sum_{c \neq \text{class}(i)} \frac{p(c)/1 - p(\text{class}(x_i))}{mk} \sum_{j=1}^{k} d(x_i(j), M_i(j)) - \sum_{j=1}^{k} d(x_i(j), H_i(j)),$$

where $\text{class}(x_i)$ represents the category to which the sample $x_i$ belongs, $p(c)$ is the prior probability of the user category $c$, $c$ represents categories other than $x_i$, and $x_i(j)$ represents the $j^{th}$ eigenvalue of the sample $x_i$, and its distance function can be computed as

$$d(x_i(j), M_i(j)) = \frac{x_i(j) - M_i(j)}{\max(j) - \min(j)}.$$

The results of the attribute feature weight values listed in Table 1 show that the feature weights of calling duration and called duration are significantly higher than the weights of calling times and called times.

The pseudocode for calculating the attribute weight value using the ReliefF algorithm for the first time is as follows (Algorithm 1).

3.2. Feature Group Preclustering Based on Improved PageRank Algorithm and K-Means Algorithm. Before computing the clustering using PageRank and K-means algorithms, we define the intimacy between users $A$ and $B$ using the following equation:

$$D(A, B) = \sum_{j=1}^{n} w(j) \cdot d(A(j), B(j)),$$

where $w$ is the weight of the attribute feature $j$.

In the traditional information dissemination model based on the PageRank algorithm [16], it is generally believed that the information spread between related users has the same probability. However, in the information network, the probability of different users contacting other users is
different. The degree of intimacy directly affects the probability of contact between them and can be calculated as

\[ G(i, j) = \frac{D(i, j)}{\sum_{k=\text{all}} D(i, k)}, \]  

(5)

where \( G(i, j) \) represents the probability of user \( i \) contacting user \( j \) and \( i \) and \( j \) represent the directed intimacy between user \( i \) and user \( j \). It means traversing all users who are in contact with the user. For users who do not have a direct contact record in the information data, we use \( E/n \) to represent the random probability of their mutual contact, and the probability transition matrix can be expressed as

\[ G = \alpha G^T + \frac{(1 - \alpha)E}{n}, \]  

(6)

where \( \alpha \) is the damping coefficient.

In the information network, user \( A \) contacting user \( B \) can be regarded as a random event, and the probability of this random event can be expressed as a chain structure using the following equation:

\[ P(X_{n+1} = j | X_n = i_0, X_1 = i_1, \ldots, X_n = i_n) = P(X_{n+1} = j | X_n = i_n), \]  

(7)

where \( P(X_{n+1} = j | X_n = i) \) represents the one-step transition probability of the user calling user \( X_n \) in the state.

According to the stable distribution condition of the Markov chain, the PageRank algorithm can be used to find the limit distribution of the corresponding random event:

\[ \text{PR}_{n+1}(A) = \sum_{B \in \text{all}(A)} \text{PR}_n(B) \ast G(B, A), \]  

(8)

where \( \text{PR}_{n+1}(A) \) represents the probability of user \( A \) being accessed for the \((n+1)\)th time, \( \text{PR}_n(B) \) represents the probability of user \( B \) being accessed for the \(n\)th time, and \( \text{PR}_n \) represents traversing all users who are connected to user \( A \). The uniform initial importance is set to \( \text{PR} = [1, 1, \ldots, 1] \); then, the user’s influence \( \text{PR} \) can be finally determined by using equation (9). Table 2 lists the top 10 influential users in the dataset.

The pseudocode of the algorithm for calculating the user’s PR value is as follows (Algorithm 2).

When using the K-means method for clustering, the randomly selected initial cluster center directly affects the clustering speed and clustering results [17]. According to the characteristics of the information network, this paper directly selects high-influence users as the cluster center, and determines the turning point and divides the 100 users of the cluster into 8 categories at the same time, and the number of users in each category is shown in Table 3.

We clustered the information users in the dataset based on intimacy. Figure 3 shows the projection results of the clustering results in a two-dimensional space. We chose the best projection angle, and we can see that the users are clustered into 8 categories.

The pseudocode of the improved K-means algorithm is as follows (Algorithm 3).

4. Comparative Test of the Information Dissemination Efficiency of the New Model

When verifying the effectiveness of the information dissemination model proposed in this article, we carried out initial delivery of information based on the idea of “birds of a feather flock together,” that is, information spreads more quickly among users of the same type but relatively slow between classes, so we target different types of users. Each category of users separately carries out information delivery. To save computational overhead and realize the real-time analysis of the information dissemination process, we implemented an information delivery scheme based on trial backtracking in the model: first, we compared the total contribution of several users with larger PR values to network information dissemination and selected the total contribution. Large users are used for delivery. For scenarios where university information is 100% known to students, the

| FA  | NAC | NPC | ACD | PCD |
|-----|-----|-----|-----|-----|
| 0.196 | 0.199 | 0.299 | 0.306 |

FA, NAC, NPC, ACD, and PCD represent the feature attribute, number of active communication, number of passive communication, active communication duration, and passive communication duration, respectively.
subsequent model verification must reach the end of the full coverage of the information after all the information is received by such users. Under this premise, the model with a small amount of information has more good information dissemination efficiency.

The total contribution of users to network information dissemination is defined as

\[ \text{able inf} (i) = \sum_{j \in \text{all}} \text{over inf} (j), \]  

(9)

where \( \text{able inf} (i) \) represents the effective cumulative influence of user \( i \) on other users and it represents the cumulative influence of user \( i \) on user \( j \) that is less than 100%. This indicates to traverse all users who are in contact with user \( i \).
4.1 Delivery Rate. The results about the information delivery of the proposed model are shown in Table 4. In the fifth group with the largest number of users, the delivery rate is as low as 35.0%, whereas the fourth group has the highest delivery rate of 60%. The average delivery rate that satisfies the full coverage of the information is 48.5%.

| Category | Serving user ID (in order of delivery) | Delivery rate (%) |
|----------|----------------------------------------|-------------------|
| First    | 2, 84, 23, 1, 100                      | 50.0              |
| Second   | 22, 85, 49, 28, 89, 86, 44             | 54.5              |
| Third    | 38, 16, 33, 43, 57, 87                 | 42.9              |
| Fourth   | 42, 52, 68, 69, 17, 83                 | 60.0              |
| Fifth    | 47, 59, 13, 62, 35, 95, 54             | 35.0              |
| Sixth    | 53, 40, 9, 26, 78, 20                  | 50.0              |
| Seventh  | 65, 99, 46, 80, 27, 20                 | 54.5              |
| Eight    | 74, 11, 41, 64, 20                     | 41.7              |
| Average  |                                        | **48.5**          |

4.2 User Placement Rate. In the traditional information network information dissemination model, to achieve the same effect of full coverage of information dissemination, the user placement results are shown in Table 5. The average placement rate exceeds 60%. In the fifth group with the largest number of users, the placement rate is high as 55%, and the placement rate in the first type of user group even reached 70%, which validates the efficiency of the proposed model.

4.3 Performance Comparison. Figure 4 shows the comparative test results of the information dissemination efficiency between the proposed model and the traditional information dissemination model. It can be seen that under the same information coverage target, the information placement rate in the proposed model has been significantly reduced compared with the traditional model. Its average delivery rate dropped from 61.71% to 48.58%. Especially in the fifth user category with the largest number of users, the delivery rate dropped from 55.0% to 35.0%. The above
results seem to be based on the proposed model, which can achieve the same information dissemination effect with less information input.

5. Conclusion

This paper studied the characteristics of information dissemination in college information networks using information entropy and ReliefF algorithm to fuse individual attribute characteristics, calculate different attribute weights, and correct the problem of feature equality in traditional clustering algorithms. The influence and intimacy of individuals greatly improved the accuracy of learning community clustering. Based on learning community preclustering and optimizing information delivery methods, a new type of information network information dissemination model was finally developed. The comparative test results showed that the proposed model can achieve the same degree of information dissemination coverage with a smaller information delivery rate than the traditional model.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the Teaching Research Project of Hubei Province of China under project no. 2017452 (The multidimensional incentive mechanism of college students’ learning ability and the design of talent training program).

References

[1] A. Mo and K. Cai, Analysis of the Characteristics and Lack of Learning Motivation of Local Undergraduate College Students, Vol. 3, University Education, London, UK, 2017.
[2] D. Zhu, Study on the Learning Motivation of College Students—Taking a College of Education as an Example, Hunan University of Science and Technology, Hunan, China, 2017.
[3] G. C. Siemens, “A learning theory for the digital age,” 2004, http://www.elearnspace.org/Articles/connectivism.htm.
[4] J. S. Brown, A. Collins, and P. Duguid, “Situated cognition and the culture of learning,” Educational Researcher, vol. 18, no. 1, pp. 32–42, 1989.
[5] M. Hunter, “Teaching is decision making,” Educational Leadership, vol. 37, no. 1, pp. 62–65, 1979.
[6] G. Kowch, A New Paradigm for Teaching, Leading and Learning in Participatory Learning Environments, Sense publishers, Rotterdam, Netherlands, 2016.
[7] P. Johnson, Essential Learning: Authentic Teaching Situations, Rowman & Littlefield Publishers, Lanham, Maryland, US, 2019.
[8] D. P. Ausubel, J. D. Novak, and H. Hanesian, Educational Psychology: A Cognitive View, Holt Rinehart & Winston, New York, NY, US, 1978.
[9] G.-J. Hwang, C.-C. Tsai, and S. J. H. Yang, “Criteria, strategies, and research issues of context-aware ubiquitous learning,” Educational Technology & Society, vol. 11, no. 2, pp. 81–91, 2008.
[10] Si. Cheng, Research on the Endogenous Motivation of College Students for Application-Oriented Talent Cultivation in Colleges and Universities—Taking Guangxi University as an Example, Guangxi University, Nanning, China, 2018.
[11] F. Wang, “Cultivation of applied talents in colleges and universities based on supply-side reform,” Jiangsu Higher Education, vol. 3, no. 5, 2016.
[12] K. Niu, Research on the Evaluation of College Students’ Learning Process, Shanghai Normal University, Shanghai, China, 2015.
[13] G. Kuh, J. Kinzie, J. A. Buckley, B. K. Bridges, and J. C. Hayek, “What matters to student success: a review of the Literature,” National Postsecondary Education Cooperative, vol. 4, 2006.
[14] L. A. Heidi and G. J. Cizek, “Handbook of formative assessment,” Routledge: Taylor&Francis Group, vol. 5, 2010.
[15] M. Robnik-Šikonja and I. Kononenko, “Theoretical and empirical analysis of ReliefF and RReliefF,” Machine Learning Volume, vol. 53, pp. 23–69, 2003.
[16] L. András, N. E. Tamás, and C. Tibor, “A local PageRank algorithm for evaluating the importance of scientific articles,” Annales Mathematicae et Informaticae, vol. 44, pp. 131–140, 2015.
[17] L. Young and H. Wu, “A clustering method based on K-means algorithm,” in Proceedings of the 2012 International Conference on Solid State Devices and Materials Science, Kyoto, Japan, September 2012.