Journal of Information Technology Review

DLINE JOURNALS

Print ISSN: 0976-3511
Online ISSN: 0976-2922
JITR 2025: 16 (2)
https://doi.org/10.6025/jitr/2025/16/2/63-70

Personalized Recommendation of Educational Resources Based on K-Means Clustering

Jiang Jing
Southeast University Chengxian College, 210088
Nanjing, China
logovictory@163.com

ABSTRACT

In today's society, the moral development of college students is crucial for maintaining social stability and growth, making it an urgent issue to address. To provide more practical information on moral development, we utilize K-Means clustering technology to group users based on their preferences, enabling the provision of more precise and valuable information. Firstly, the data is classified through a collaborative filtering algorithm to ensure data comparability and determine user preferences. Based on this, an effective model for recommending educational resources is created. Subsequently, the K-Means clustering algorithm is employed to develop a targeted recommendation process based on the specified objective function, yielding an effective recommendation of educational resources. Through experiments, we have found that this approach not only aligns well with the content of moral courses but also ensures that students are more engaged and achieve good results in reality.

Keywords: Educational Resources, Personalized, Recommendation Method, K-Means Clustering

Received: 10 October 2024, Revised 11 January 2025, Accepted 27 January 2025

1. Introduction

General Secretary Xi Jinping emphasized that the country's development relies on the development of talents, as well as national prosperity and stability, with higher education institutions being the crucial carriers of this foundation [1]. Therefore, we should strive to cultivate talents who support the party and national leaders, guiding their development with their personality and values. Currently, scholars are investigating the needs of moral development among college students and exploring practical ways to implement them. Currently, education in colleges and universities consists of three major courses: Marxism and Socialism with Chinese Characteristics. The main difference lies in that the first two focus on cultivating students' moral character to achieve the goal of cultivating virtues, while the latter focuses on cultivating students' moral standards to achieve a higher level of morality. To enhance the moral cultivation of college students and their sense of social responsibility, we must integrate the concept and practical application of moral cultivation into the curriculum. However, currently, the lack of practical moral cultivation courses and activities hinders the work of moral cultivation [2]. In today's dynamic society, we should seize opportunities, actively explore more effective moral education models, strengthen the practice and activities of moral courses, fully tap and cultivate college students' moral qualities, actively explore and practice effective moral courses, and adopt effective measures, such as the K-Means clustering algorithm, to provide personalized recommendations for college moral courses, thereby better meeting the needs of college students' moral cultivation [3].

2. Related Analysis

Personalized recommendations refer to providing relevant information tailored to users' behavioral patterns and browsing habits. Standard methods in personalized recommendation include collaborative filtering, association rule-based recommendation, clustering algorithm-based recommendation, and hybrid recommendation. In 1992, Goldberg et al. proposed the concept of collaborative filtering, which was first applied to the Tapestry system for filtering emails and news for users, showcasing a new recommendation idea. The appearance of the collaborative filtering algorithm has led to in-depth research by many scholars, making it the most widely used recommendation algorithm in the field. Abroad, scholars have researched personalized recommendation models for educational resources, mainly proposing four types: (1) Information Processing Type, emphasizing guided teaching and students' problem-solving through memorizing and understanding theoretical knowledge [4]. (2) Personalized Education Type, emphasizing active teaching patterns and providing personalized, multidisciplinary content tailored to students' developmental status [5]. (3) Cooperative Communication Type, emphasizing group cooperation in teaching activities, further developing students' autonomous exploration ability during cooperation and communication [6]. (4) Behavioral Control Type, proposing program teaching patterns by constructing a reasonable disciplinary knowledge system, decomposing knowledge into small units, and further developing "control teaching methods" [7]. In China, there are three types of personalized recommendation models for education: (1) the Indoctrination Type, where teachers actively impart textbook theoretical knowledge to students, thereby limiting the improvement of teaching quality and the enhancement of student knowledge levels [8]. (2) Preaching Type, where teachers cannot accurately understand students' learning needs, possibly leading to misjudgments in students' learning interests and preferences, resulting in reduced teaching efficiency [9]. This study aims to address the current issues of poor timeliness, single format, and limited effectiveness of education in colleges, actively explore college students' new needs and changes based on mobile internet technology, and construct a personalized recommendation model for college education with the support of the internet and big data technology [10].

3. K-means Clustering Algorithm

3.1 Principles of the K-means Clustering Algorithm

K-means Algorithm is considered an effective clustering algorithm, which originated in 1967 when mathematician MacQueen used a k value to divide a set into k groups and adjusted the relationships between the groups to improve the overall objective function of the groups. The K-means algorithm is characterized by its ease of use and efficiency [11,12]. Its working principle is as follows: Firstly, the entire dataset is divided into k clusters, and k feature values are used as starting points. Then, the remaining feature values are gradually adjusted upward based on a certain Euclidean distance until the endpoint of the objective function is reached.

This completes one iteration of the algorithm. In a Q-dimensional space SQ with a finite set X={*x1, x2, ..., xn*}, it is initialized and divided into k clusters, each containing *C1, C2, ..., Ck*. When each cluster contains n objects, the clustering centers *Z1, Z2, ..., Zk* of the *i*-th cluster are determined. According to the expression of the objective function, it is represented as formula (1).

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n_j} D^2_{x_j\,,\,z_i} \tag{1}$$

Here, the Euclidean distance refers to the relative distance between the *j*-th feature and the *i*-th cluster. The steps of the K-means algorithm are as follows: First, select k features as the core of clustering from the finite set *x*; based on the Euclidean distance, the distance between two objects can be determined, and they are sorted according to distance for better comparison. According to formula (1), the clustering centers of the clusters can be further defined, and then formula (2) is executed again to achieve the perfect result of the algorithm. The *K*-means algorithm extracts *K* features from a large dataset, groups them, and then iteratively adjusts these features to achieve the optimal distribution state [13,14]. Through the K-means algorithm, we can cluster data with strong correlations into the same category while isolating data with less correlation. Figure 1 illustrates an example of the *K*-means clustering algorithm, which divides the data into three classes, with each class having a clustering value *K* equal to 3. The 'o' represents the clustering core.

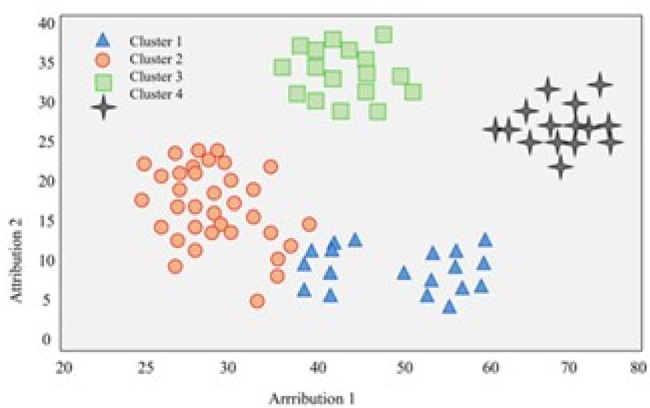

Figure 1. K-means Clustering Algorithm

3.2 Recommending Educational Resources Based on K-Means Clustering

By adopting personalized ways to specify learning content, we can not only provide students with abundant learning materials but also give them opportunities for active learning. However, due to the vast amount of learning materials, students find it challenging to select the most suitable learning content based on their needs, which makes their learning process less efficient [15]. By leveraging users' favorite features, we can rank content from courses according to their specific values, providing more accurate recommendations. At the same time, we utilize the collaborative filtering algorithm to ensure optimal results. Selecting the collaborative filtering algorithm to establish the recommendation model for education resources involves two steps: data behavior collection and similarity calculation [16]. In the data behavior collection of users, data diversity needs to be ensured by collecting user preference data in various forms, constraining the data set within [0, 1], and normalizing it. The similarity calculation is performed using formulas (2) and (3).

$$a(s,d) = \sqrt{\sum\left(s_f - d_f\right)^2} \tag{2}$$

$$sim(s,d) = \frac{1}{a(s,d)} \tag{3}$$

In this formula, *s* and *d* refer to two points in the dimensional space, and *f* refers to the number of these two points. $a(s, d)$ represents their Euclidean distance. Through this distance, we can quantify the preference level of users for specific educational resources and utilize the K-Means clustering algorithm to provide personalized recommendations. By adopting the K-Means clustering technique, personalized recommendations can be effectively realized by randomly extracting *g* groups from the user data and calculating the distances between each group and the initial clustering centers. The objective function is set as:

$$\boldsymbol{h}_j = \sum_{k=1}^{g} \sum_{l \in j_k} \left| l - c_k \right|^2 \tag{4}$$

Where *jk* is a special category, and its squared difference is determined by the square root of ck, the square root of which is determined by the square root of *jk*. Additionally, *hj* is another special category, and the square root of the square root of *hj* affects its square root and, consequently, the square root of *hj*. If l is a special category, then the square root of *hj* will affect its square root and, consequently, the square root of l. By using the clustering algorithm, we can randomly extract a group of special categories from a set of special categories as needed and make their square roots similar [17,18]. In this way, we can better provide the content of courses.

4. Experimental Design and Analysis

4.1 Experimental Design

Through comparative testing, we found that both the hybrid recommendation method and the big data-based recommendation method can effectively provide personalized educational resources. We used students from the first to fourth year of a university as the test subjects. We extracted thousands of sets of educational resources from the resource database to evaluate the recommendation effects of these two methods. To ensure the reliability of the data, we randomly selected 1,000 students from universities in various provinces across the country. We collected educational materials from the national resource database [19,20]. We carefully selected thousands of data sets based on the teaching courses of different grades for reference. By improving 80% of the samples, we can better measure the efficiency of the collaborative algorithm and test its performance on these samples. This way, we can more accurately evaluate the efficiency of the collaborative algorithm and predict its advantages and disadvantages. Through systematic simulation, we can accurately estimate the satisfaction of target consumers when purchasing a specific item and compare it with the model used to evaluate the model's performance.

4.2 Experimental Results Analysis

Through clustering technology, we can effectively evaluate the performance of the algorithm to understand its characteristics better, including clustering centers, clustering time, intra-cluster similarity, inter-cluster similarity, and clustering accuracy. “High intra-cluster object similarity and low inter-cluster object similarity” are excellent clustering results, which can effectively divide objects into more reasonable categories. Therefore, using cluster similarity as an effective method in indicator evaluation is very effective. Using the MATLAB testing platform, we can recommend various educational resources for different stages of teaching at multiple universities to achieve effective dissemination. To verify the effectiveness of each group of recommendation methods, we connected them to three groups separately and tested their timeliness and complexity, as shown in Figure 2.

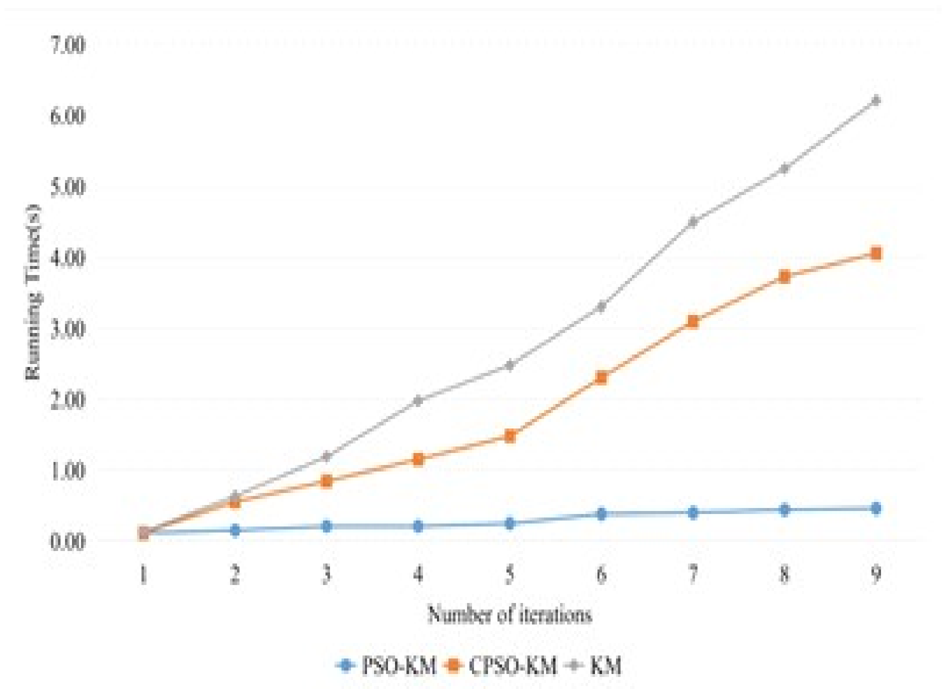

Figure 2. Efficiency Differences of Three Algorithms over Time

Through the analysis of Figure 2, we can see that the efficiency and effectiveness of the KM algorithm are not satisfactory, as both are inferior to those of PSO-KM. The efficiency and effectiveness of the latter are both superior to the former. This can be attributed to the insufficient global search capability of the KM algorithm, which tends to enter partial optimization, resulting in its efficiency and effectiveness being inferior to PSO-KM. However, when the PSO-KM algorithm introduces particle swarm, its global search capability is enhanced, and the movement of its particles becomes more random, resulting in significant improvements in search efficiency and effectiveness, thereby significantly improving its overall performance. By incorporating chaotic sequences into the calculation, we significantly enhance the overall search capability and cause the particles to move in a more orderly manner. Compared to the PSO-KM algorithm, we completed the task more quickly. Although this method is not optimal, we still believe it has high value in improving overall efficiency. Through the K-Means clustering algorithm, we generated personalized recommendations of education resources for 1,000 students from the first to fourth year, and the results are presented in Figure 3. From the figure, we can see that this method effectively integrates different types of resources and provides accurate recommendations, meeting the requirements of various education courses at different grade levels and possessing specific practical value.

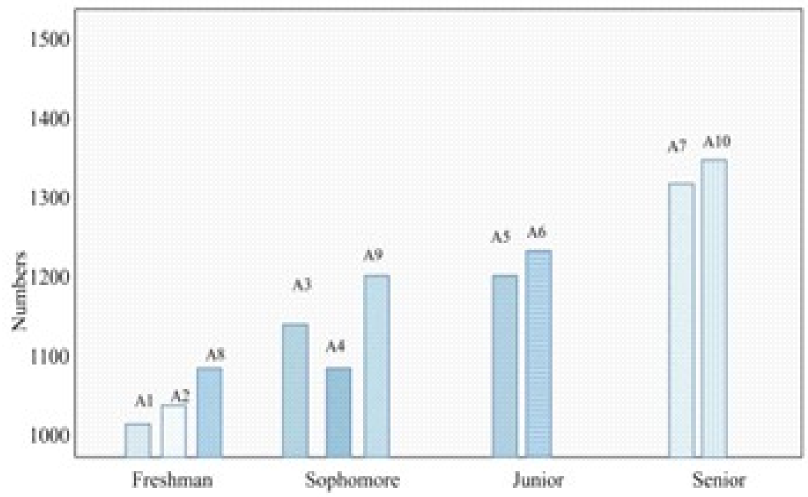

Figure 3. K-means Algorithm Recommendation Results

Through systematic experiments, we found that utilizing collaborative filtering, clustering, and logistic regression classification algorithms can significantly enhance the system's accuracy, recall rate, and F1 score. This finding provides strong support for the reliability and effectiveness of our algorithm. Through a 45-minute actual investigation, we can more accurately evaluate the effectiveness of various recommendation schemes and divide the respondents into multiple stages to better understand their preferences for these recommended materials.

5. Conclusions

In recent years, personalized recommendation technology has developed rapidly, not only quickly and accurately collecting a large number of personal needs but also becoming a mainstream technology in web data analysis. Therefore, this paper focuses on exploring the collection, analysis, and personalized recommendation technology for addressing personal needs, thereby effectively overcoming the shortcomings of traditional web search technology. At the same time, it has been verified by relevant practices and demonstrated good reliability and operability. By adopting clustering algorithms, this paper proposes a novel personalized recommendation method for educational resources, ensuring the orderly progression of education for college students and enhancing their ability to engage actively in education. The experimental results demonstrate that this method can meet the needs of students across various university grades and accurately reflect their educational interests, thereby achieving effective personalized recommendations. Although the time for this study is limited, we still strive to explore various academic resources and assign specific names to them, enabling us to better screen and recommend suitable courses, thereby providing more theoretical support for future research.

Acknowledgements

The study was supported by the Project of Philosophy and Social Science Research of Jiangsu Universities in 2021: A Research on Ideological and Political Education of Art Students in Application-Oriented Undergraduate Colleges and Universities (No. 2021SJB1170).

References

[1] Vankayalapati, R., Ghutugade, K. B., Vannapuram, R., et al. (2021). K-Means algorithm for clustering of learners' performance levels using machine learning techniques. *Journal of Physics: Conference Series, 11*(2), 11–18.

[2] Ghaderi, F., Moradhasel, B. (2021). Resources and expenditures management in the field of medical supplies using K-means clustering algorithm: A case study. *Tehran University of Medical Sciences, Centre for Electronic Resources Provision and Journal Improvement, 9*(4), 30–36.

[3] Wati, M., Rahmah, W. H., Novirasari, N., et al. (2021). Analysis K-means clustering to predicting student graduation. *Journal of Physics: Conference Series, 1844*(1), 012028. https://doi.org/10.1088/1742-6596/1844/1/012028

[4] Yusniyanti, A. L., Virgantari, F., Faridhan, Y. E. (2021). Comparison of average linkage and K-means methods in clustering Indonesia's provinces based on welfare indicators. *Journal of Physics: Conference Series, 1863*(1), 012071. https://doi.org/10.1088/1742-6596/1863/1/012071

[5] Sabri, M. Z. M., Majid, N. A. A., Hanawi, S. A., et al. (2021). Model development in predicting academic performance of students based on self-efficacy using K-means clustering. *Journal of Physics: Conference Series, 2129*(1), 012047. https://doi.org/10.1088/1742-6596/2129/1/012047

[6] Qi, F., Chang, Y., Ramesh, K., et al. (2021). Online and offline teaching connection system of college ideological and political education based on deep learning. *Progress in Artificial Intelligence, 12*(2), 163–174.

[7] Yanhong, D. (2021). Study on the ideological and political construction of college courses from the content of tobacco control education. *Tobacco Regulatory Science, 9*(5), 40–50.

[8] Ye, W., Qin, J., Mu, W., et al. (2021). Analysis on the cohesion of red cultural resources into ideological and political education in primary and secondary schools. *Francis Academic Press, 9*(3), 22–28.

[9] Xia, T., Ahmad, M. T. (2022). Method of ideological and political teaching resources in universities based on school-enterprise cooperation mode. *Mathematical Problems in Engineering, 11*(5), 55–60.

[10] Ma, J. (2021). University ideological and political education management based on K-means mean value algorithm. *Journal of Physics: Conference Series, 1852*(4), 042023.

[11] Sathyamoorthy, M., Kuppusamy, S., Dhanaraj, R. K., et al. (2022). Improved K-means based Q-learning algorithm for optimal clustering and node balancing in WSN. *Wireless Personal Communications, 122*(3), 2745–2766.

[12] Moubayed, A., Injadat, M., Shami, A., et al. (2020). Student engagement level in an e-learning environment: Clustering using K-means. *American Journal of Distance Education, 34*(2), 137–156. https://doi.org/10.1080/08923647.2020.1719201

[13] Trivedi, S., Patel, N. (2020). Clustering students based on virtual learning engagement, digital skills, and e-learning infrastructure: Applications of K-means, DBSCAN, hierarchical, and affinity propagation clustering. *Sage Science Review of Educational Technology, 3*(1), 1–13.

[14] Irawan, Y. (2019). Implementation of data mining for determining majors using K-means algorithm in students of SMA Negeri 1 Pangkalan Kerinci. *Journal of Applied Engineering and Technological Science (JAETS), 1*(1), 17–29.

[15] Feng, Z., Niu, W., Zhang, R., et al. (2019). Operation rule derivation of hydropower reservoir by K-means clustering method and extreme learning machine based on particle swarm optimization. *Journal of Hydrology, 576*, 229–238.

[16] Li, Y., Chu, X., Tian, D., et al. (2021). Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm. *Applied Soft Computing, 113*, 107924. https://doi.org/10.1016/j.asoc.2021.107924

[17] Umargono, E., Suseno, J. E., Gunawan, S. K. V., et al. (2020). K-means clustering optimization using the elbow method and early centroid determination based on mean and median formula. *The 2nd International Seminar on Science and Technology (ISSTEC 2019)*, Atlantis Press, 121–129.

[18] Pitchai, R., Supraja, P., Victoria, A. H., et al. (2021). Brain tumor segmentation using deep learning and fuzzy K-means clustering for magnetic resonance images. *Neural Processing Letters, 53*, 2519–2532.

[19] Shamrat, F. M. J. M., Tasnim, Z., Mahmud, I., et al. (2020). Application of K-means clustering algorithm to determine the density of demand of different kinds of jobs. *International Journal of Scientific & Technology Research, 9*(02), 2550–2557.

[20] Li, D., Wong, W. E., Wang, W., et al. (2021). Detection and mitigation of label-flipping attacks in federated learning systems with KPCA and K-means. *2021 8th International Conference on Dependable Systems and Their Applications (DSA)*, IEEE, 551–559.