Research on intelligent recommendation system of course resources based on central enterprise training platform

Kaili Guo[a*], Qianhui Yu[b], Yixing Qi[c], Jie Zhan[d], Zhuoyue Li[e], Guihua Lin[f]
Kai Zhang[g], Ran Fang[h]

State Grid of China Technology College, Jinan, China

[a*]guokaili2014@163.com.cn,[b]liuguangpeng@sgtc.sgcc.com.cn
[c]qiyixing@sgtc.sgcc.com.cn,[d]zhanjie@sgtc.sgcc.com.cn
[e]lizhuoyue@sgtc.sgcc.com.cn,[f]linguihua.sgtc.sgcc.com.cn
[g]zhangkai@sgtc.sgcc.com.cn,[h]fangran@sgtc.sgcc.com.cn

Abstract. State Grid E-learning is an online training service platform for the unified construction and promotion and application of the State Grid Corporation of China. This paper carries out research on personalized recommendation technology of courseware resources of the State Grid School, adopts hybrid Collaborative filtering algorithm and Knowledge graph classification system, establishes an intelligent recommendation system based on user's stereoscopic portrait on the State Grid School, connects user attributes and behaviors with the digital training resources of the State Grid School, greatly improves the efficiency of resource utilization in various application scenarios, and realizes intelligent push of resources at the same time, Enhance user learning experience.

Keywords: State Grid E-learning; personalized recommendation; Collaborative filtering algorithm; Knowledge graph.

1 Introduction

With the rapid development of technology today, the amount of content on the Internet is increasing day by day, and people increasingly feel helpless in the face of massive data. Under the development of content being king, how to effectively distribute reasonable content to users in different dimensions has become a new challenge for operations. The recommendation method has evolved from manual filtering by traditional editors to algorithmic recommendation, achieving the effect of personalized content customization with thousands of people and faces[1]. The recommendation system optimizes algorithms based on users' historical behavior, interest preferences, and statistical characteristics to generate a list of items that users may be interested in, achieving personalized customization and improving content delivery efficiency and accuracy. State Grid E-learning undertakes online training services for 1.5 million employees of the company, providing rich skills and knowledge training resources.

M. Yu et al. (eds.), *Proceedings of the 2024 5th International Conference on Big Data and Informatization Education (ICBDIE 2024)*, Advances in Intelligent Systems Research 182,
https://doi.org/10.2991/978-94-6463-417-4_48

How to accurately push these rich learning resources to the student end is an urgent technical problem that needs to be solved.

2 Principle of recommendation algorithm structure

The content recommendation algorithm model based on user profiles mainly includes five modules, namely user data module, course data module, calculation and analysis module, statistical ranking module, and result generation module[2].

2.1 User Data Module

The user data module is mainly responsible for collecting basic user information and behavior data, and generating corresponding feature values for subsequent calculations based on the system defined feature vectors.

2.2 Course Data Module

The course data module is mainly responsible for collecting basic information related to the course and user learning and evaluation information about the course. At the same time, based on the system defined feature vectors, corresponding feature values are generated for subsequent calculations.

2.3 Calculation and analysis module

The calculation and analysis module is mainly responsible for reading user and course data, calculating the similarity between users and courses based on intelligent recommendation algorithms, and matching the similarity to the target user's preferred training resource set[3].

2.4 Statistical module

The statistics module is responsible for compiling course evaluations and cumulative learning frequency rankings within the platform, generating an ordered set with ranking attributes as an alternative dataset for recommendation results.

2.5 Result generation module

The recommendation result generation module is responsible for aggregating the candidate datasets generated from the above paths, including operations such as deduplication, filtering, and sorting.

Finally, based on the system defined weight of each data path, extract and generate the final recommendation result list.

3 Implementation of an intelligent recommendation system based on State Grid School

3.1 Establish a three-dimensional portrait of users in the State Grid School

The three-dimensional user profile constructed in this article has static characteristics of the user, such as their job level, educational level, and working years. It has dynamic characteristics based on behavior, such as the cumulative number of learning courses and cumulative learning duration on the platform.[4] A feature vector is constructed to form the three-dimensional user profile. A three-dimensional user profile that combines static and dynamic feature information can more accurately depict the uniqueness of users.

Static Properties. Static attributes refer to attributes that are fixed or have extremely low frequency of change in user attributes. The static attributes selected in this paper include information such as job level, educational level, and working years[5].

Dynamic Properties. Dynamic attributes refer to the dynamic behavior attributes of users in the State Grid School. The dynamic attributes selected in this paper include cumulative learning duration and cumulative number of learning courses.

Definition of eigenvectors and eigenvalues. According to the selected feature vectors and business rules, corresponding feature values can be obtained, which are: educational level (1 point for high school, 2 points for junior college, 3 points for undergraduate, 4 points for master's, and 5 points for doctoral degree), job level (1 point for junior high school, 2 points for intermediate level, and 3 points for advanced level), work experience (1 point for 1-3 years, 2 points for 4-5 years, 3 points for 6-10 years, 4 points for 11-20 years, and 5 points for over 20 years) The feature vectors and eigenvalues of five dimensions, including cumulative learning duration (1-3 hours and 1 minute, 4-10 hours and 2 minutes, 11-30 hours and 3 minutes, 31-60 hours and 4 minutes, and 60 hours and above 5 minutes), number of learning courses (1-5 hours and 1 minute, 6-10 hours and 2 minutes, 11-20 hours and 3 minutes, 21-40 hours and 4 minutes, and 40 hours and above 5 minutes). Select the characteristic vectors of 10 usersas shown in Table 1. as shown in Table 1.

Table 1. Statistical Table of Partial User Feature Vectors

User	**Zhang***	**Wang***	**Li***
Job level	senior	senior	intermediate
Education	master	undergraduate college	undergraduate college
Years of service	3years	6 years	5 years
Learning duration	3 hours	6 hours	10 hours
Number of learning courses	3	10	10
Feature vector	(3,4,1,1,1)	(3,3,3,2,2)	(2,3,2,2,2)

3.2 Build a Knowledge graph of the training resources of the State Grid School

Research and construct the Knowledge graph, classification, clustering system and professional labels of training courses of the training resources of the State Grid School. Sort out the curriculum resources of the State Grid School, classify 3300 courses, establish a Knowledge graph system of 111 items in 11 categories, establish curriculum specialty labels, and realize the labeling of courses[6].

Table 2. Knowledge graph of 11 Specialties in Technical Standards Area

First level label	Secondary label
Plan and design	foundation, system, transformer substation, Converter station, line, distribution network, communication
Engineering construction	foundation, transformer substation, Converter station, aerial conductor, cable, Thermal power, hydropower, communication, economy, other
Equipment and materials	foundation, Ultra-high voltage, high pressure, medium voltage, low pressure, aerial conductor, cable, protect, Power electronics, monitor, machinery, material, Thermal power, hydropower, communication, check and accept, other
Scheduling and Trading	foundation, stabilize, Scheduling plan, Reactive power, Network source coordination, hydropower, dispatching operation, protect, dispatching automation, Distribution network operation, power trading, communication, other
Operation and maintenance	foundation, transformer substation, Converter station, aerial conductor, cable, Thermal power, hydropower, Tools and instruments, other
Experiments and Metrology	foundation, high pressure, medium voltage, aerial conductor, cable, protect, Power electronics, monitor, metering, detection, Thermal power, hydropower, communication, other
Safety and environmental protection	foundation, secure, labor protection, health, environmental protection, meet an emergency, fire fighting, other
Technical supervision	foundation, power quality, equipment performance, Electrical measurement, thermotechnical, metal, chemistry, environmental protection, protect, communication, other
Internet	foundation, network, computer, communication security, Communication resources, Communication Applications, Communication operation, other
Electricity sales and marketing	foundation, metering, trade, energy conservation, Selling electricity, demand-side, Electricity consumption, other
New energy	on-grid, construct, device, run, test, other

3.3 Hybrid Collaborative filtering algorithm based on user profile

In this paper, the content recommendation algorithm uses a hybrid Collaborative filtering algorithm based on users and goods. The hybrid Collaborative filtering algorithm based on user profile in this paper adopts the user based Collaborative filtering algorithm, the item based Collaborative filtering algorithm, the course tag similarity algorithm based on association rules and the hybrid algorithm based on course evaluation results. Based on the characteristics of different analysis objectives, selecting different algorithms is more in line with practical laws and improves the accuracy of recommendation algorithms[7]. The specific instructions are as follows:

User based similarity algorithm for Collaborative filtering. The training target of State Grid E-learning is 1.5 million, and the number of service targets is relatively large. Generally, after the number of users reaches a certain level, the user based Collaborative filtering is no longer recommended from the perspective of computing efficiency. However, the State Grid E-learning is a specific business training for the State Grid Corporation and belongs to an enterprise with distinctive job business characteristics. The user similarity has a high reference value for the selection of recommended content. At the same time, the trainees in the company are relatively stable, Therefore, this project selects the user based Collaborative filtering algorithm[8].

The cosine similarity algorithm is used for user similarity calculation, and the user feature vector formula is selected.

$$\cos\theta = \frac{\sum_{i=1}^{n}(A_i * B_i)}{\sqrt{\sum_{i=1}^{n} A_i^2} * \sqrt{\sum_{i=1}^{n} B_i^2}} \tag{1}$$

Among them, represents the i-th vector value of user A and represents the i-th vector value of user B. By calculating the cosine of the angle between the n-dimensional vectors (n>0) of two users, the similarity between users is calculated. The range of cosine similarity values is [0,1], and the magnitude of the included angle is inversely proportional to the cosine similarity value. The smaller the angle between two vectors, the greater the cosine similarity value. The closer the calculated cosine value is to 1, the higher the similarity between these two users. Similarity is the basis of user based Collaborative filtering. Based on the above calculation formula, 10 sample data were selected for algorithm validation, and the results are shown in Table III.

Table 3. Similarity Graph of Some Users

User	**Zhang***	**Wang***	**Li***
Job level	senior	senior	intermediate
Education	master	undergraduate college	undergraduate college
Years of service	3 years	6 years	5 years
Learning duration	3 hours	6 hours	10 hours
Number of learning courses	3	10	10
Feature vector	(3,4,1,1,1)	(3,3,3,2,2)	(2,3,2,2,2)
Similarity	1.0	0.8944	0.8625

From the calculation results, it can be observed that the user feature vectors selected by the model generate feature values and apply formulas to calculate the similarity results, which comply with objective logical laws[9].

Item based similarity algorithm for Collaborative filtering. In this project, items are understood as training courses and the formula for calculating course similarity.

$$W_{\mathrm{ab}} = \frac{N(a) \cap N(b)}{\sqrt{N(a) * N(b)}} \tag{2}$$

Among them, represents the similarity between course a and course b, represents the number of users who have studied course A, represents the number of users who have studied course B, and represents the number of users who have studied both course a and course b. The similarity between courses is the basis of Collaborative filtering based on items. According to the above formula, select 10 sample data for algorithm validation, and the results are shown in Table IV.

Table 4. Similarity Map of Part 1 Courses

Serial Number	**Course Name**	**Accumulated number of learners**	**Number of students studying Course A together**	**Similarity**
1	CourseA	139	-	-
2	CourseB	243	90	0.3264
3	CourseC	187	120	0.7443
4	CourseD	69	30	0.3063
5	CourseE	55	20	0.2287
6	CourseF	843	80	0.2337
7	CourseG	670	68	0.2228
8	CourseH	530	95	0.3500
9	CourseI	440	112	0.4528
10	CourseJ	70	55	0.5575

Course Label Similarity Algorithm Based on Association Rules. Association rules are a common form of connection between things, and they can provide feedback on the dependencies between transactions to a certain extent. Whether it is a recommendation algorithm based on Collaborative filtering or a content based recommendation algorithm, the core of the algorithm focuses more on analyzing the relationship between users or projects, while the association rule based recommendation algorithm takes into account the relationship between recommendation options and recommenders. Objectively speaking, because curriculum association rules are extracted and analyzed from users' Big data, It to some extent eliminates the influence of subjective factors on course association rules. To understand recommendation algorithms based on association rules, it is necessary to first understand the following concepts. Here, we will directly use the course labels in the technical standards section of the State Grid E-learning as an example to illustrate. Implement labeling of courses in the technical standards section. Support refers to the frequency of the occurrence of a combination of two course labels in all user data. Recorded as:

$$Support(X,Y) = \frac{Freq(X,Y)}{N} \tag{3}$$

Due to the large volume of courses included in the State Grid E-learning and the fact that most of them are professional courses such as electrical engineering, the individual classification of courses is not clear. Therefore, in considering course recommendation based on association rules, this article uses course labels as items to participate in the frequency calculation of the itemset, where X and Y represent any two course labels, and Freq (X, Y) represents the probability of X and Y appearing together. Confidence represents the degree of dependence of one course label on another among all course label combinations. Recorded as:

$$Confidence(X,Y) = \frac{Freq(X,Y)}{Freq(X)} \tag{4}$$

In general, the higher the confidence level, the stronger the correlation between the two course labels X and Y. It is worth noting that the confidence level of X to Y and Y to X can be different values. Therefore, when calculating the correlation level of course labels for course recommendation, it is important to pay attention to whether the pre course labels of the confidence level are accurate.

This article selects effective user data from 100 technical standards zones, and calculates the support and confidence levels of pairwise combinations of course labels. By ranking the confidence levels from high to low, a confidence matrix can be obtained (the matrix has as many dimensions as there are course labels), which is the association rule of course labels. Subsequently, based on the user's latest browsing courses and the course labels to which they belong, a confidence matrix is used to search for predicted labels, and a random algorithm is used to enumerate the courses in that label, ultimately achieving course recommendation.

Similarity Algorithm Based on Course Evaluation Results. The State Grid School has launched a course evaluation system, where each course is evaluated and scored from one to five points, with one being dissatisfied and five being very satisfied. Based on the evaluation results of learners, courses are recommended, and the results are shown in Table V.

Table 5. Score Table of Course Evaluation Results

Course evaluation	Score
Very satisfied	5
Quite satisfied	4
Satisfied	3
Dissatisfied	2
Not satisfied	1

3.4 Multiple recall

The recommendation results of this paper adopt a multi-channel recall method, which recalls the candidate result sets generated by different sources and algorithms to meet diverse content recommendation needs. At the same time, the system can flexibly configure according to the recall results of different periods and ranges, achieving the goal of manual intervention in recommendation results. The recall channels include: recall based on user similarity, recall based on course similarity, recall based on course label similarity, recall based on course rating ranking, and recall based on specific content set by the system. There are a total of 5 recall channels. The recall results defined in this project include: user similarity alternative result set (priority 1), course similarity alternative result set (priority 2), course label similarity alternative result set (priority 3), course scoring alternative set (priority 4), and system specified course alternative result set (priority 5), totaling 5 result alternative sets. The 5 alternative result sets are all ordered data lists. Each recall can return a maximum of 5 alternative result records. If there are no records that meet the recall criteria, an empty alternative result set will be returned. The system ultimately provides recommended content results for each user[10].

The system can configure the weight of the default 5-way alternative result set in the final recommendation result list according to business needs. The system defaults to a consistent weight of 20% for all 5 alternative sets. If there is a missing alternative result set for a certain path, other alternative result sets will be used in sequence to fill in. The priority of the 5-way alternative result set can be configured by the business administrator.

Dynamic configuration and weight adjustment.

In addition to multi-channel recall, the system can also configure the weight of alternative result sets recalled by different ways in the final recommendation results according to business needs, to meet the requirements of manageable and controllable recommendation information. The default weight for all 5 recall paths is 20%. For empty alternative result sets returned by the recall path, other recall path result sets will be used in sequence to supplement.

4 Summary

This article studies the intelligent recommendation system for courseware resources on central enterprise training platforms. By analyzing user characteristics and behaviors, users' potential interests are explored, and personalized data services are provided. The intelligent distribution of over 200 courses in the technical standards section of the State Grid School is achieved, achieving a recommendation effect of thousands of people and faces, meeting the personalized training needs of students, and improving experience and satisfaction. This paper belongs to the technical research of machine learning in the field of artificial intelligence. In the future, the recommendation system will explore the application process in more than 80 specialized areas of the State Grid School platform, and continuously improve algorithms based on actual application situations, so that intelligent distribution technology can be widely applied to the entire platform.

References

1. Shadi A, Mohammad D, Amjad G, et al. An Intelligent Recommendation System for Automating Academic Advising Based on Curriculum Analysis and Performance Modeling[J]. Mathematics,2023,11(5).
2. Jing Y. An intelligent recommendation method of personalised tour route based on association rules[J]. International Journal of Reasoning-based Intelligent Systems,2023,15(1).
3. Hong W, Zhiyong L. Intelligent Recommendation Method of Mobile Wireless Communication Information Based on Speech Recognition Technology Under Strong Multipath Interference[J]. International Journal of Information Security and Privacy (IJISP),2022,16(2).
4. Luo J, Duan X. Intelligent recommendation of personalised tourist routes based on improved discrete particle swarm[J]. International Journal of Computational Science and Engineering,2022,25(6).
5. E. A O, Ram M, Eduard P, et al. Uses of the digital twins concept for energy services, intelligent recommendation systems, and demand side management: A review[J]. Energy Reports,2021,7.
6. Xiajun S, Yunting M. Research on the impact of artificial intelligence recommendation on academic procrastination under the background of big data——The mediating role of mobile phone addiction[J]. Journal of Physics: Conference Series,2021,1948(1).
7. Yuanyi C. Research and Application of Personalized Intelligent Recommendation Algorithm for Computer Network Search Engine[J]. Journal of Physics: Conference Series,2021,1915(3):
8. Daijian Q W. Research on Intelligent Recommendation Business Model of Tourism Enterprise Value Platform from the Perspective of Value Cocreation[J]. COMPLEXITY, 2021,2021.

9. Qin F Z. Intelligent Recommendation Algorithm Combining RNN and Knowledge Graph[J].Journal of Applied Mathematics,2022,2022.
10. Huang J. Influences and Solutions of the Cocoon Room Caused by the Intelligent Recommendation Algorithm of the E-commerce Platform[J].American Journal of Applied Psychology,2022,11(6).