International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-21

https://doi.org/10.70917/ijcisim-2026-0143
Article

User Image Modeling and Interaction Optimization Strategies in Interaction Design of Short Video Parenting Platform Construction in Colleges and Universities

Yan Yang *

Art and Design School, Luoyang Institute of Science and Technology, Luoyang, Henan, 471000, China; yang11202024@163.com

Abstract: With the rapid development of science and technology, students' behavior of using short video parenting platform to watch videos in their daily life is becoming more and more popular. Therefore, the article proposes a user classification method based on the improved FRM model for college short video parenting platform, determines the weights of the four indicators in the RFMS model through the combined assignment method, and then applies the K-means clustering algorithm to analyze the users by clustering. And on this basis, a short video recommendation algorithm based on association rules is proposed aiming to improve the accuracy of short video recommendation, and its effect is verified through experiments. Finally, an interaction optimization strategy based on college short video parenting platform is proposed in combination with the basic user profile. The article analyzes the changes in the distribution of topics and the number of topic types in the recommendation list obtained from the model of this paper and the collaborative filtering recommendation model. The results show that the user-based collaborative filtering recommendation model declines relatively sharply in the three indicators. The degree of decline becomes significantly smoother at [30,60], while the overall degree of decline of the three indicators of this paper's model is relatively smooth. From this, it can be concluded that the model in this paper has obvious advantages in improving the diversity of recommendation results.

Keywords: RFMS model; K-means; user profile; association rules; university short video parenting platform

1. Introduction

University short videos are 5-minute videos created by university educators for college students with the aim of inspiring education. They mainly focus on value guidance, teaching and research, and campus culture [1-2]. In March 2018, the Communist Youth League of China released its first short video, “Hello, I am Tuantuan,” on the Douyin platform, pioneering the use of short video accounts by government agencies [3]. Since then, many leading universities have joined short video platforms such as Douyin. In September 2022, Douyin collaborated with Tsinghua University and other renowned institutions to launch the “Back-to-School Open Class” initiative, rolling out 100 open classes to disseminate scientific knowledge to the public. The series of short video courses has been widely embraced by college students online [4]. Meanwhile, a review of short video content across Chinese universities reveals common issues, such as unclear positioning of short video platforms for educational purposes, significant content homogenization, and a lack of brand distinctiveness [5-7]. The selection of topics and presentation formats for short videos are too narrow, failing to incorporate contemporary youth language and trending topics, making it difficult to sustain students' interest. Additionally, short video platform teams lack systematic management, and their political and digital literacy require improvement [8-11]. How to leverage short videos to integrate educational content into every aspect of college students' learning and daily lives, thereby supporting their growth and development, is an urgent and important issue for higher education professionals to explore.

The concept of user profiling was first proposed by Alan Cooper, the father of information interaction technology [12]. Through data mining and research on various types of users in the University of California Digital Library, he screened hidden data from the abundant library information, discovered the interrelationships among seemingly complex and disordered data, and identified the patterns of connection between the dwell times of different types of users [13-15]. Subsequently, research on user profiling modeling began. Sang, J et al. [16] proposed a user modeling method across open social networks, which quantifies similar behaviors and integrates social data using local social rules. This method was validated through personalized video recommendation experiments conducted on Google+ and YouTube datasets. Shi, R et al. [17] explored how the characteristics of short video platforms influence users' willingness to share marketing information, finding that information quality has the most significant impact on this willingness. This study provides guidance for platforms to optimize their marketing effectiveness. Yang, C et al. [18] utilized a new model called Multi-Site Probabilistic Factorization (MPF) to model user profiles across six popular short video websites, resulting in significant improvements in short video recommendations. Liu, Q and Deng [19] proposed a user profiling method based on the RFM model under big data conditions. This method involves collecting and processing user behavior data to create comprehensive user profiles and has been proven to more accurately identify users. Xu, Y et al. [20] proposed a multi-dimensional user profiling model that incorporates psychological and emotional factors, providing personalized services by analyzing users' emotions and preferences in a Web3.0 personalized environment.

The article first proposes the improved RFMS model, prepares and preprocesses the data according to the specific meanings of the indicators in the model, and standardizes the processed data. The combination assignment method is used to determine the four indicator weights of the improved RFMS model, and at the same time, the value of the users of the short video parenting platform of a university is analyzed by clustering, and the characteristics of the users are analyzed according to the results obtained from the classification. Subsequently out of the IGA-ARU algorithm for association rule mining to achieve short video recommendation for users with more behavioral data, the genetic algorithm and the downward closure characteristics of the association rules are fused, and the selection, crossover, mutation and other parts of the genetic algorithm are improved. The final recommendation list is generated based on the target user's predicted score for the short videos mined by association rules. Then, the accuracy and diversity of the traditional collaborative filtering recommendation model and this paper's model are compared and analyzed based on the accuracy and diversity indicators to verify the effectiveness of this paper's model. Finally, based on the experimental results, the interaction optimization strategy based on college short video parenting platform is proposed from three levels: recommendation technology, legal system and platform management.

2. Modeling of Platform User Profiles Based on Improved RFM Models

2.1. Description of the Methodology

In order to better classify the value of users of college short video parenting platform, this paper proposes a user value classification method for college short video parenting platform based on the improved RFM model, which improves the corresponding indexes so that they are more applicable to the college short video parenting platform, and proposes to use the combinatorial assignment method to determine the weight of the indexes in the model, and use the k-means clustering algorithm to classify the users.

2.2. Establishment of the RFMS Model

In view of the fact that the operation mode of short video parenting platform in colleges and universities is different from traditional platforms. This paper proposes the RFMS improvement model, and the meaning of each index in the model is explained as follows:

R : the average browsing time interval of users in a period of time. For college short video parenting platforms, the smaller the average browsing time interval is, the higher the user's activity is. The formula for the R indicator is:

$$R = \frac{T_L - T_F}{F} \tag{1}$$

Where T_L denotes the last browsing time of the user in a period of time, T_F denotes the first browsing time of the user in a period of time, and F is the number of browsing times in the corresponding time.

F : the number of times a user has browsed in a period of time.

F metrics have the same meaning as F metrics in traditional RFM models.

M : The amount of traffic generated by the user for the school over a period of time.

S : the number of comments posted by the user over a period of time.

2.3. Data Preparation and Pre-Processing

The research data is selected from the user browsing data of the short video parenting platform of a university. Before the data analysis, we need to pre-process the data, the original data data have vacant values, duplicate records, incorrect data, these data will have a bad impact on our analysis, so we need to clean these data. Delete the data in the interest table whose browsing time is "0", these data are dirty data. Extract and integrate the required fields in the RFMS model using SQL statements.

2.4. Normalization of RFMS Indicators

Due to the improved RFMS model, the average browsing time interval, the number of user browsing, user traffic, the number of reviews of the four indicators of the value range of data differences are relatively large, the quantitative outline is not uniform, in order to eliminate the impact of the value of the classification of the data needs to be feature scaling, this paper adopts the standardization of the deviation standardization method to standardize the data, mapping the data between [0,1], standardization processing The formula is:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

Where x is the data sample, x' is the processed sample data, $x_{\min}$ is the minimum value in the data sample, and $x_{\max}$ is the maximum value in the sample data.

2.5. Determination of RFMS Indicator Weights Using Combined Assignment Methods

After comprehensively considering the advantages and disadvantages of subjective and objective assignment methods, this paper decides to use the combined assignment method to determine the indicator weights of the improved RFMS model [21]. The steps are as follows:

1) The subjective assignment method uses the hierarchical analysis method (AHP) to determine the RFMS indicator weights $w1$.

2) The objective assignment method uses entropy weighting to determine RFMS indicator weights $w2$.

3) The indicator weights obtained by the two methods are weighted and averaged using the linear weighted combination method. The final indicator weights are:

$$w = \alpha w1 + (1-\alpha) w2 \tag{3}$$

Where α is the importance degree of hierarchical analysis method in the whole process of determining indicator weights, $1-\alpha$ is the importance degree of entropy weight method, in this study, α is taken as 0.5, and it is considered that hierarchical analysis and entropy weight method have the same importance degree.

The steps of using hierarchical analysis method to determine the index weight $w1$ are as follows:

(1) Establish the hierarchical model of the RFMS model of the short video parenting platform for colleges and universities. According to the requirements, the single hierarchy model constructed in this study is shown in Figure 1.

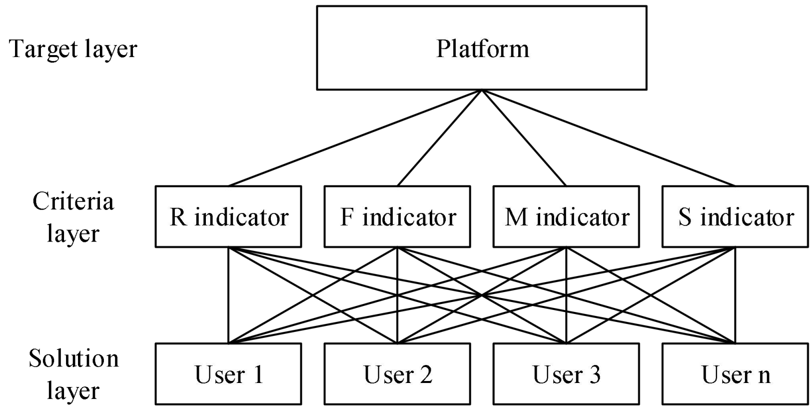


Figure 1. Single hierarchy model of user value of shopping guide platform.

(2) Construct the judgment matrix in each level, and use the pairwise comparison method and the 1~9 scale to construct the pairwise comparison array of each level for each factor in the previous level.

Entropy weight method is an objective assignment method, its basic idea is to determine the objective weights based on the information entropy of the indicator, the smaller the information entropy E of a certain indicator is, indicating that the indicator has a strong variability, the greater the information that the indicator can provide, and the greater the role it plays in the results of the comprehensive evaluation, and the greater the weight of the indicator is [22]. The specific process of determining the weights of the four indicators of the RFMS model using the entropy weight method is entered as follows:

(a) Normalize the four indicators of RFMS model.

The raw data of RFMS indicators are written as $x=(x_{ij})_{mn}$, where m represents the number of samples, n represents the number of indicators, and the j th indicator value of the i th sample is written as x_{ij}:

$$x=\begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & x_{m4} \end{bmatrix} \tag{4}$$

Then the data for the standardized indicator is written as $y_{ij}=\dfrac{x_{ij}-\min(x_j)}{\max(x_j)-\min(x_j)}$.

(b) Calculate the weight f_{ij} of the i th sample under the j th metric for that metric, using the formula:

$$f_{ij}=\frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}} \tag{5}$$

(c) Calculate the information entropy of the j th indicator, denoted as H_j:

$$H_j=-k\sum_{i=1}^{m} f_{ij}\ln f_{ij} \text{ Which } k=\frac{1}{\ln m} \tag{6}$$

(d) Calculate the weight of the j th indicator, denoted as w_j:

$$w_j=\frac{1-H_j}{\sum_{j=1}^{n}(1-H_j)}=\frac{1-H_j}{n-\sum_{j=1}^{n} H_j} \tag{7}$$

2.6. Modeling K-Means Clustering

Based on the improved RFMS model, the four indicators R, F, M and S of the users of the university short video parenting platform are clustered, and according to the experience and the actual operation of the university short video parenting platform, here the user value is categorized into four categories, i.e., the value of K takes the value of 4. The specific steps are as follows:

Input: four-dimensional dataset composed of four indicators R, F, M, S. The number of clusters K=4.

Output: clustering results.

(1) Randomly select 4 points in the dataset as clustering centers.

(2) Calculate the distance from other data points to the 4 clustering centers according to a specific distance calculation formula, and classify the data point into the cluster where the clustering center with the smallest distance is located. The 4 initial classifications are obtained.

(3) Calculate the mean of the 4 initial classifications and use them as new cluster centers. Recalculate the distance of other data points to the new clustering center and divide the data points into the cluster where the clustering center with the smallest distance is located.

(4) Repeat the steps (2) and (3) until there is no significant change in the mean value of the initial classification or the maximum number of iterations has been reached. Output clustering results.

In steps (2) and (3), the selection of different distance calculation formulas will produce different clustering results, and in this study, the Euclidean distance is selected as the distance calculation formula. The Euclidean distance is a more commonly used distance in practice, and since the Euclidean distance is related to the magnitude of the variables, it is necessary to standardize the data to eliminate the effect of the magnitude on the results before performing the clustering analysis.

3. Modeling User Profiles of Short Video Parenting Platforms in Higher Education Institutions

3.1. Introduction of User Data of Short Video Parenting Platforms in Higher Education Institutions

3.1.1. Introduction to Data

The data in this paper is quoted from the user data of a university short video parenting platform. There are four tables of raw data: user_info, login_info, visit_info, result. Now introduce these four forms respectively: user_info is mainly for the user's basic information, including user ID, browsing time, grade, specialty and other 8 aspects of the user's basic information. login_info is the user's login information, in addition to the user ID to record the user's login days, login interval, the last login days from the end of the period and other aspects of information. visit_info is the user's visit information, including the number of visits to the home page, the number of scheduled visits and other information. Finally, result is the user evaluation result.

3.1.2. Data Pre-Processing

Missing values and duplicate values: There are four aspects of user data information, namely, basic user information, user login information, user access information and user browsing information, which are combined according to the user ID for a total of 136,592 users with 50-dimensional feature data. For user purchase information, the number “1” indicates browsing, and the number “0” indicates not browsing. At the same time, it can be noticed that there are duplicate values in user IDs, and duplicate values should be filtered to ensure the uniqueness of user IDs. The above deletion of missing user information and duplicate user information leaves 135,629 pieces of user data.

Outliers processing: In addition to the basic user information in the part of the information belongs to the non-numerical variables or single variables, the rest of the information is required to carry out the test of basic statistics, the data distribution of the general test and the initial identification of outliers. Generally speaking, for the online education platform user age should be between 0-100 years old, and for the number of login days, login interval, the last login distance from the end of the period days and other information should be non-negative indicators. On this basis, an outlier test is performed for each feature

Regarding the age information, which is expressed in terms of the number of months, the scatter plot of the age characteristics is shown in Figure 2. The normal range of age is 0-100 years old corresponding to 0-1200 months, and the basic statistical information about age in the dataset & age months interval is located in [0, 25634], according to the normal situation, we need to exclude [00, 1200] other than the data, a total of 28 articles. For these data this paper chooses to use the average value of 72 instead.

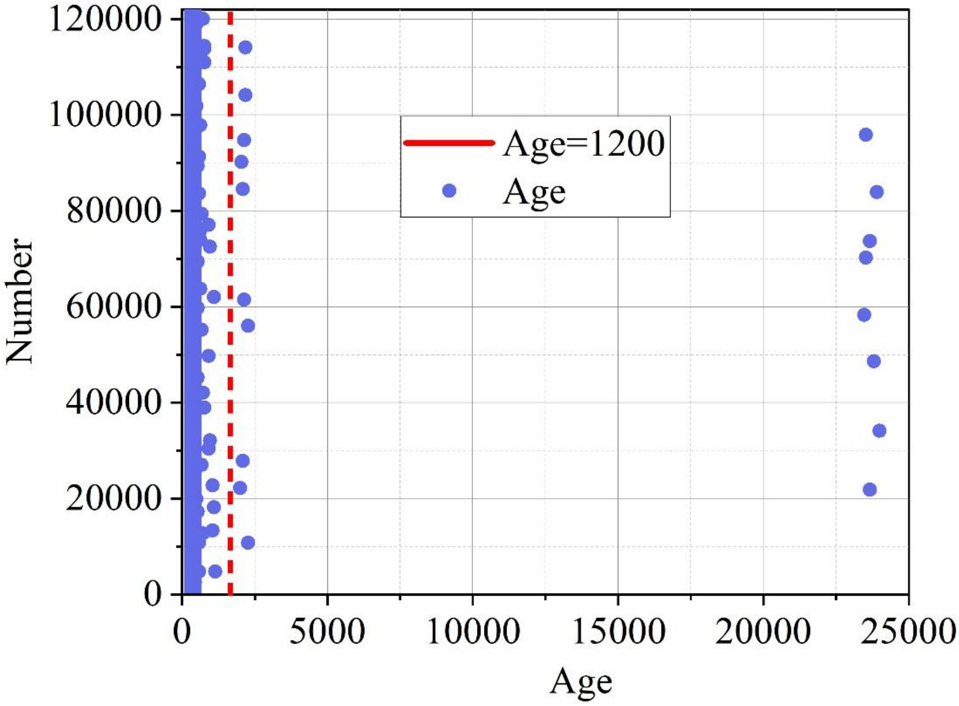
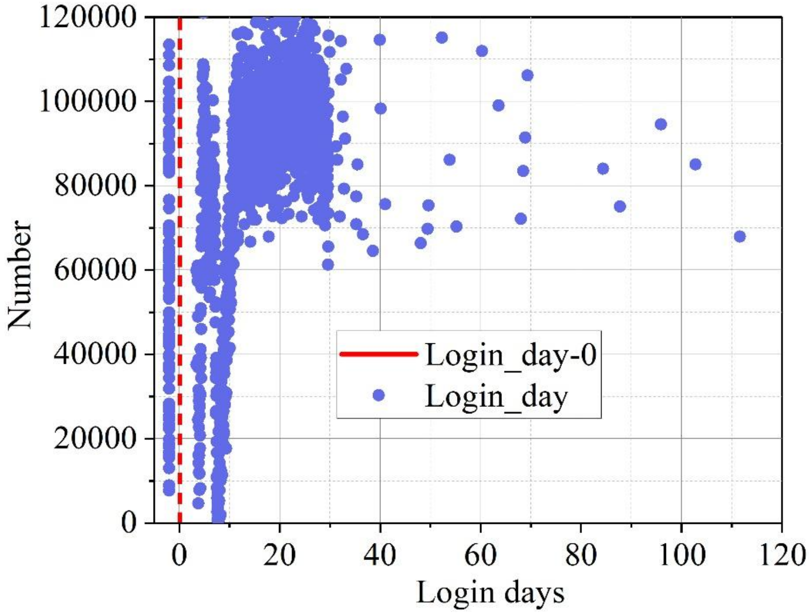
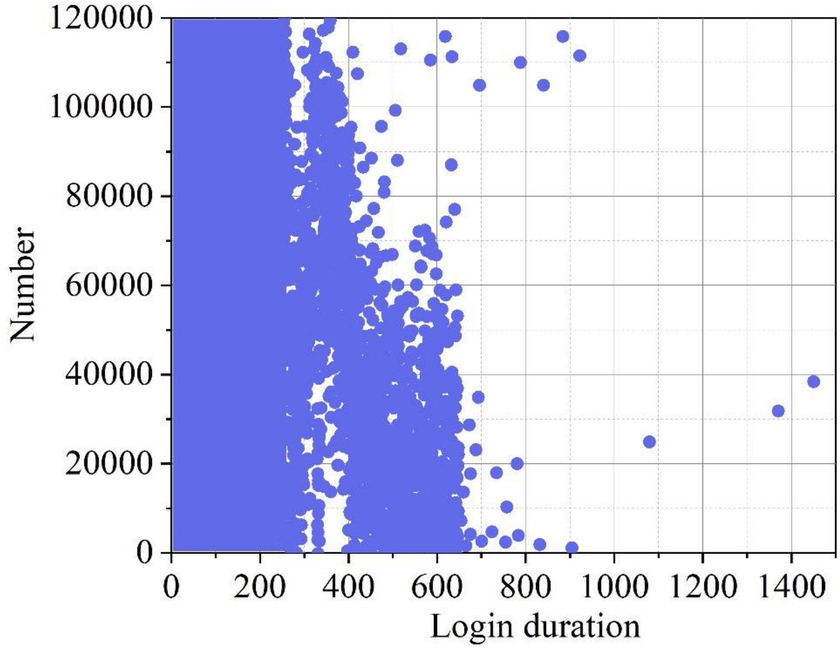


Figure 2. Age characteristics.

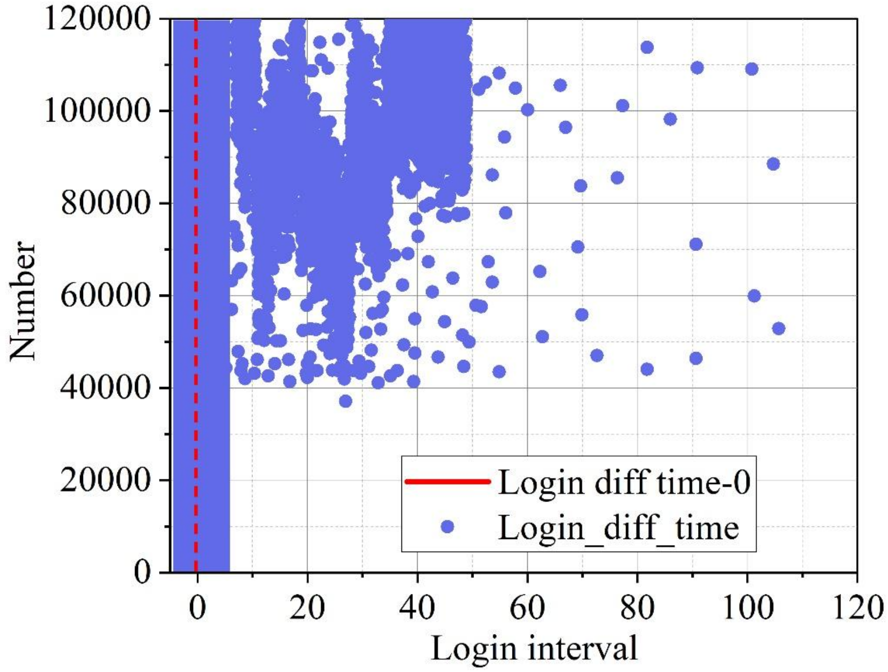
Faced with the three metrics of days logged in, logging interval and last logged in days from the end of the period, it is not possible to be negative considering the actual situation. By screening these three features there are 3715, 3895 and 4051 negative values respectively. The scatter plot of the distribution of the five features is shown in Fig. 3 (Fig. a~e are the number of login days, login duration, login interval, last login distance from the end of the period and the number of landing pages for two visits, respectively). According to the figure, it is observed that the other indicators corresponding to the read 3715 data are also mostly 0, which cannot provide much favorable information, so they are eliminated, leaving 12165 sets of data.



(a) Login days

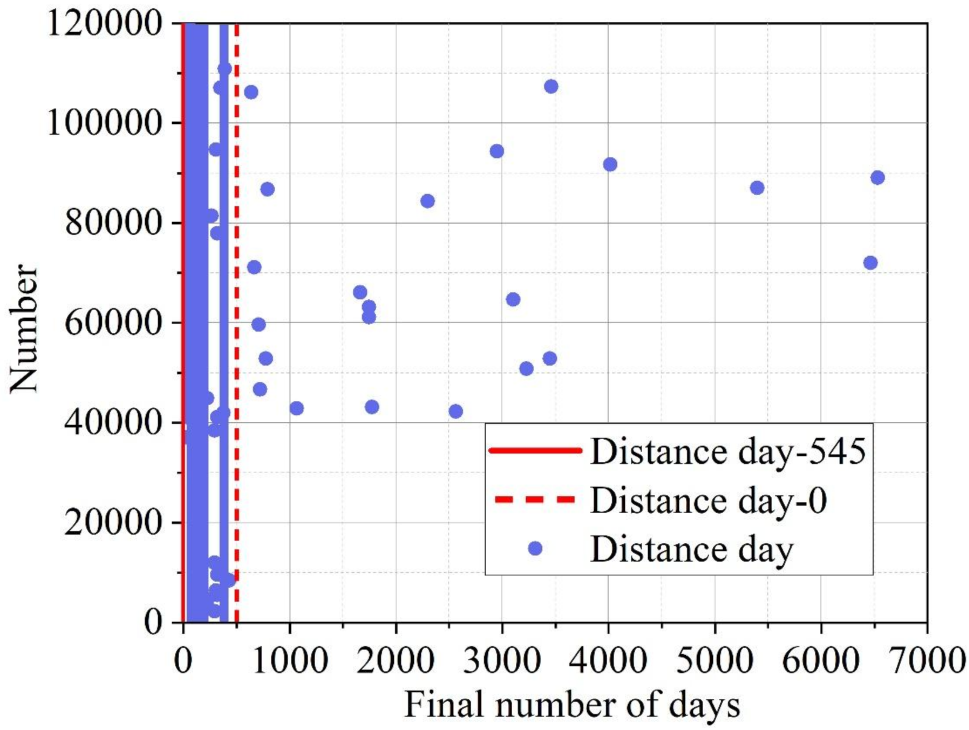


(b) Login duration

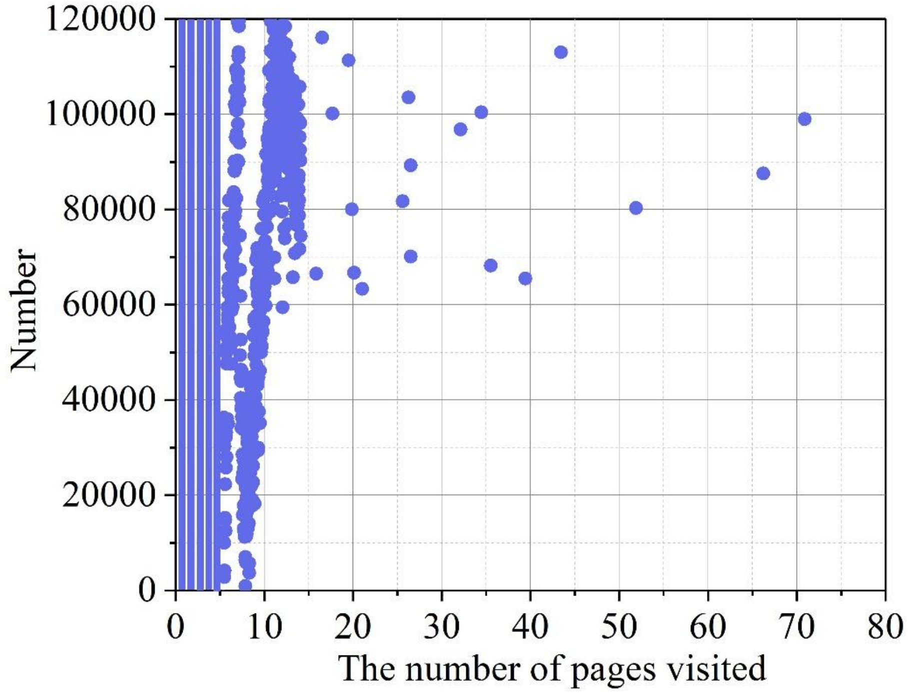


(c) Login

interval



(d) Final login



(e) The number of pages visited

Figure 3. Five characteristics distributed scatter points.

3.2. Indicator Assignment

In this paper, the weights of the features are considered in terms of both the comparative strength and conflictiveness of the data.The RFMS model metrics weights are shown in Table 1. The obtained weights are multiplied with the normalized RFMS metric counterparts to obtain the dataset used for clustering.

Table 1. RFMS model index weight.

	W_R	W_F	W_M	W_S
Right planting	0.049052	0.091365	0.268541	0.573261

3.3. User Segmentation Based on RFMS and Clustering Algorithm

The K-means algorithm was used to cluster the RFMS metrics after assigning weights. In this paper, the elbow method was chosen to determine the optimal number of clusters, and the optimal number of clusters is shown in Table 2. It can be seen that the slope decline of the line segment after K=4 slows down significantly, so the recommended number of clusters is obtained as 4.

Table 2. Optimum number of clustering.

Number of clusters	2	3	4	5	6	7	8	9
VESS	590	196	87	61	49	45	33	26

The results of user clustering are shown in Table 3.

Table 3 User clustering results

Categories	$\overline{R}$	$\overline{F}$	$\overline{M}$	$\overline{S}$	Quantity
Cluster 1	0.03808	0.00383	0.04635	0.01244	5126
Cluster 2	0.0408	0.00252	0.00413	0.00439	76265
Cluster 3	0.01419	0.00484	0.00744	0.01338	21362
Cluster 4	0.03954	0.00516	0.00983	0.0393	19255

3.4. Analysis of User Segmentation Results

The RFMS metrics here correspond to the values after normalization, and the user score rankings are shown in Table 4.

Table 4. User score.

Ranking	$\overline{R}$	$\overline{F}$	$\overline{M}$	$\overline{S}$	Score	Quantity	Cluster class
1	0.03794	0.00386	0.04639	0.01254	0.10389	5126	Cluster 1
2	0.04083	0.00256	0.00419	0.0042	0.09378	19255	Cluster 4
3	0.0141	0.00476	0.00735	0.01333	0.05183	76265	Cluster 2
4	0.03944	0.00522	0.00965	0.03921	0.03806	21362	Cluster 3

The higher the score indicates the greater the value of the user, reflecting the user's loyalty and satisfaction with the short-term video parenting platform. According to the scores of the above four types of users, they can be categorized into: important development users, loyal promotion users, short-term active users, and ordinary loss users.

(1) Important development users: This type of user has the highest score, reflecting the value of this type of user is also the largest. Their M index is much larger than the other three categories, indicating that this type of user learns the most courses on the platform, reflecting a stronger reliance on the platform. In addition this category has the second highest S-indicator, which also reflects the high level of satisfaction of this category of users. R index and F index also have good performance, but there is still room for development. Combining the above four indicators, it can be analyzed that this type of user has the largest number of learning courses on the platform, the number of logins is relatively balanced, and the average length of each use is more reasonable has a continuous and consistent demand for the platform, and needs to be provided with focused attention to achieve its conversion.

(2) Loyal enhancement users: This type of user ranks first in F and S indicators, which means that this type of user has the highest frequency of use and the greatest satisfaction, and this type of user is a stable group of loyal users of the platform. The M index is higher than the other two categories, but much lower than the important promotion users, which indicates that although this type of user logs into the platform more often, but each time he logs in to watch the course, the time is not much.

(3) Short-term active users: this type of users is characterized by large R indicators and high activity, but their performance in other indicators is very average.

(4) Ordinary churning customers: the activity of this type of users is far lower than the other three

types, but the performance in F indicators is not bad, M, S indicators are also lower than the development of users and enhancement of the user, proving that this type of users have a certain habit of using the platform before, but with the decline in activity and satisfaction gradually reduced to churning customers.

3.5. Validation of results

In order to verify the difference between the unimproved RFM model and the improved RFMS model, this paper performs the same normalization, assigns weights, clusters, and calculates scores to get the user segmentation results for the comparison of the unimproved RFM model.The RFM model indicator weights are shown in Table 5.

Table 5. RFM model index weight.

	W_R'	W_F'	W_M'
Weight	0.07865	0.147652	0.776516

After the normalization of the three indicators were multiplied by the weight elbow method under the still recommended to be divided into four classes, take K - meas algorithm for clustering, RFM model user score ranking as shown in Table 6.

Table 6. RFM model user scores.

Ranking	$\overline{R}$	$\overline{F}$	$\overline{M}$	Score	Quantity	Cluster class
1	0.05842	0.00636	0.17569	0.24072	2031	Cluster 4
2	0.06299	0.0064	0.06868	0.13796	12598	Cluster 3
3	0.06738	0.00599	0.0101	0.08334	85639	Cluster 1
4	0.02289	0.00804	0.01783	0.02322	21659	Cluster 2

The model identified 2,031 important development users, which still have a much larger M metric than the other three categories, spend the most total time on the platform, and also have good performance in the other two metrics, and are overall important users of the platform. The second-ranked user with all three metrics in second place is a loyal boost user. The third-ranked 85,639 users have the largest R metrics, indicating the most active and are short-term active users of the platform. The last place scoring user group is the least active, but the total login frequency is high, which can be reflected as normal type churn users.

The comparison of RFMS model and RFM model user identification is shown in Table 7. It can be seen that the login rate for important development users under the RFMS index is higher than that of the RFM index, while the login rate for ordinary churn users is lower than that of the RFM index, reflecting that the results of the RFMS model in the identification of user value are more in line with the results of the conversion of actual users.

Table 7. RFMS model and RFM model user identification.

	Important development user	Ordinary drain users
RFMS	9.526%	0.274%
RFM	9.206%	0.295%

4. Personalized User Interaction Model Based on IGA-ARU

4.1. Association Rule Mining Based on Improved Genetic Algorithm

The idea of applying the improved genetic algorithm to mining association rules proposed in this section is to achieve mining to higher quality association rules by designing operations such as crossover, mutation, selection, etc. for the characteristics of the association rules to meet their characteristics. The association rule mining process based on improved genetic algorithm is shown in Fig. 4.

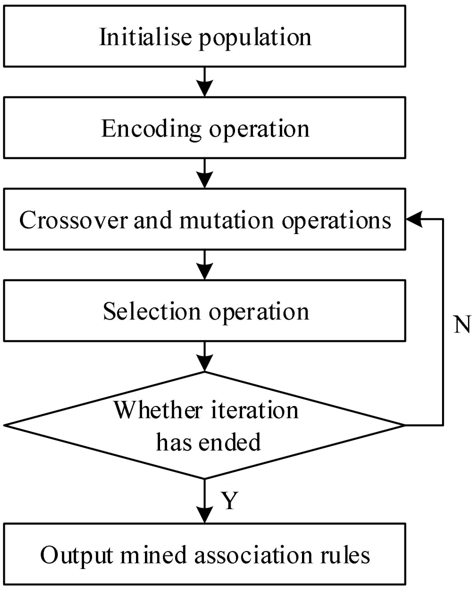


Figure 4. Based on the association rule mining process of improved genetic algorithm.

(1) Initialization

This part of the operation is mainly used to initialize the parameters of the association rules and the genetic algorithm, which include the settings of the set of frequent itemsets $FreItemset = \varnothing$ and the set of valid association rules $AssociationRules = \varnothing$, the settings of the thresholds of the minimum support $\min\_\sup$ and the minimum confidence threshold $\min\_conf$ settings, etc.

(2) Coding operation

Coding in genetic algorithm is actually mapping the feasible solutions in the problem space into the chromosomes in the genetic space, which enables the problem solving to be transformed into the operation on the chromosomes [23]. When the genetic algorithm finds the global optimal solution, it needs to decode its corresponding encoding string in order to obtain the final association rule.

This algorithm uses binary encoding because the binary encoding method is simple and efficient in genetic algorithms, making operations such as crossover and mutation easier to implement. Each individual in the current population $Candidate\ Group$ consists of n binary bits $(n = GSize)$, where each bit represents a video in the transaction set. If a video appears in an individual, the corresponding binary bit is "1", otherwise it is "0".

(3) Crossover operation

Crossover operation in genetic algorithms involves exchanging a portion of the genes in the chromosomes of two individuals to produce a new individual. The role of crossover operation is to increase the diversity of the population and avoid falling into local optimal solutions. Crossover operations can somehow make the new generation of individuals more favorable in terms of traits and may produce more new species.

In order to incorporate the downward closure property of association rules, this algorithm is adapted by using certain strategies during the crossover operation. In the crossover operation in individual $individual_i$ of $Candidate\ Group$ in the current population, a frequent itemset is randomly selected from two parent individuals $individual_i$ and merged with the frequent itemset in another parent individual $individual_j$. After merging, all the subsets in the offspring individuals are checked to see if they satisfy the minimum support threshold, and if not, the subset is removed from the offspring, i.e., the binary position where it is located is 0, so as to generate a new generation of individual $individual_i'$, which is added to the candidate population $Candidate\ Group$.

(4) Mutation operation

The common way of mutation in genetic algorithm is single-point mutation, i.e., mutation operation on one of the individual chromosomes to increase the diversity of the population, so as to avoid the genetic algorithm from falling into local optimization. In order to improve the efficiency of the algorithm and optimize the results, the algorithm uses a control coefficient $\zeta\left(2\le\zeta\le10\right)$ to adjust the size of the candidate populations, and at the same time in the mutation operation, the use of "directed mutation" strategy, that is, to select the individual of a binary position 0 to 1, rather than the individual of a position 0 to 1, and to choose the individual of a position 0 to 1, and to choose the individual of a position 0 to 1. At the same time, when performing the mutation operation, the "directed mutation" strategy is used, i.e., selecting an individual whose binary position is 0 to be changed to 1, and not allowing 1 to be changed to 0, so as to generate a new individual $individual_i''$, and adding it to the candidate population $Candidate\ Group$. The process is performed continuously until the candidate population size $Candidate\ Group$ reaches or exceeds the preset limit $\zeta GSize$, or reaches the upper limit of the number of mutations. In practical applications, the results of the mutation operation need to be de-duplicated to avoid duplicate individuals from being added to the candidate population $Candidate\ Group$.

(5) Selection operation

The purpose of the selection operation is to choose the best adapted $GSize$ individuals to form the frequent k itemsets, which will be used as the starting populations for the next round of operations in order to discover better quality association rules. At the time of selection, there may be itemsets in the candidate population $Candidate\ Group$ with unsatisfied support or confidence that should not be included in the selection operation. These itemsets can be filtered out before selection by using for the fitness function based approach. At this time, some selection operation algorithms can be used to increase the variability between individuals, making the selection operation more diverse and thus increasing the evolutionary speed of the population. Tournament selection method is a common selection strategy in genetic algorithms. This algorithm uses an improved tournament selection method:

Step 1: Select the individual with the highest fitness in the candidate population $Candidate\ Group$ as the first individual in the new candidate population $Candidate\ Group$.

Step 2: Randomly select k individuals in the current candidate population, where k is specified in advance. Select an individual from the k individuals that has the highest fitness and is not duplicated in the new population. Add the selected individual to the new candidate population $Candidate\ Group$.

Step 3: Repeat step 2 until enough $GSize$ individuals have been selected.

Before performing the selection operation, individual merit needs to be evaluated using fitness to determine whether to include in the next generation population. The fitness function affects the quality of association rule mining and the accuracy of recommendation results, because the goal of this method is to mine frequent item sets, the evaluation criterion is the support degree, and the fitness function is designed:

$$Fitness(X)=Support(X)-\min\_\sup \tag{8}$$

Where $Support(X)$ is the current individual support value and $\min\_\sup$ is the set minimum support threshold.

When the value of an individual's fitness is too low, it will result in the individual's contribution to the next generation of the population being small or even ineffective. Therefore, when the value of the individual's fitness is less than 0 it will be eliminated, otherwise its confidence and boost are calculated and the formula is shown below:

$$Confidence(X\rightarrow Y)=\frac{Support(X\cup Y)}{Support(X)} \tag{9}$$

$$Lift(X\rightarrow Y)=\frac{Support(X\cup Y)}{Support(X)\times Support(Y)} \tag{10}$$

When an individual satisfies the minimum confidence threshold and the lift is greater than 1, it is decoded and added to the set of valid association rules $Association\ Rules$ for storage.

(6) Iterative process

Through iterative operation, the genetic algorithm continuously selects the excellent individuals from the current population and generates new individuals through crossover and mutation operations to gradually optimize the quality of the population. In this process, the number of individuals will gradually decrease. When the size of the current population $Current\ Group$ is less than or equal to 1, no further operation can be performed. At this point, the set of valid association rules $Association\ Rules$ is output.

4.2. Short Video Recommendation Based on Association Rules

In this algorithm, association rule mining is applied to form a short video recommendation list. Firstly, the set of short videos that the target user has already watched a is clarified, and then those rules with a as antecedent are searched from all the association rules, and the confidence level of the rules is used as a scoring criterion, and the confidence level is accumulated and calculated as shown in the formula below:

$$Confidence\_sum = \sum Confidence(b) \tag{11}$$

Where $Confidence(b)$ denotes the confidence level of the association rule containing the short video b.

Finally, the top N videos with the highest scores are selected and constructed into a recommendation list to be returned to the user.

5. Testing the Effectiveness of Model Applications

In order to verify the validity of the model proposed in this paper, this paper will utilize the empirical analysis of the model on the real dataset. In this paper, the user data of the short video parenting platform in colleges and universities are acquired. The dataset includes 632 users, as well as 78,582 likes data, 48,241 works data and 53,343 concerns data of these users.

5.1. Model Recommendation Accuracy Effect

In order to clearly show the recommendation effect of the recommendation lists generated by the traditional user-based collaborative filtering recommendation model and the diversified recommendation model proposed in this paper on the same experimental data, the accuracy and recall metrics computed by the two are compared with each other separately, using accuracy and recall as the criteria. The topic diversity model mainly selects the top N items with better diversity from a larger recommendation list to recommend, so as to improve the diversity of the recommendation list. Therefore, in this paper, the number of recommended items in the final recommendation list is set to 10, i.e., N=10, the number of similar users is 15, and the length of the initial recommendation list is set to 20, 30, 40, 50, 60, and 70, and the values of the recommendation result indicators of the two models are calculated. The comparison results of accuracy and recall are shown in Fig. 5 and Fig. 6.Comparison of F-value between UserCF model and this paper's model is shown in Fig. 7.

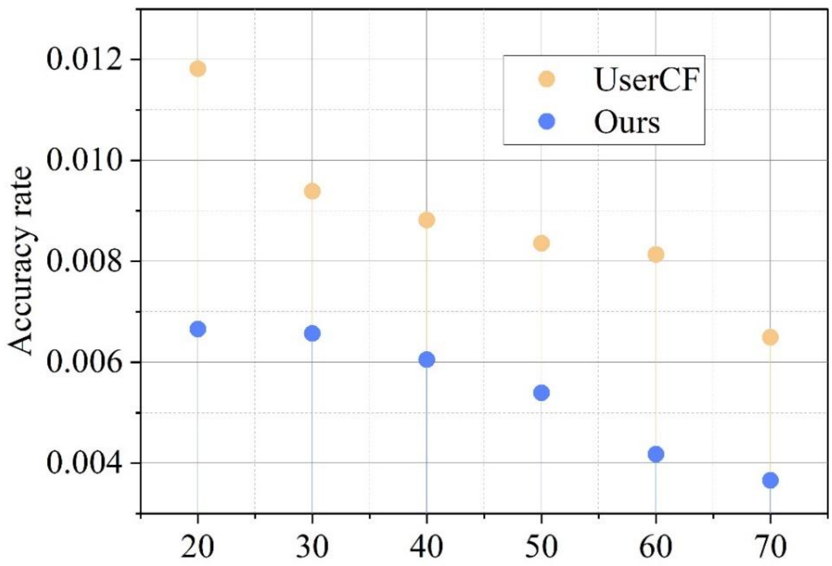


Figure 5. UserCF model is compared with the accuracy of the model in this paper.

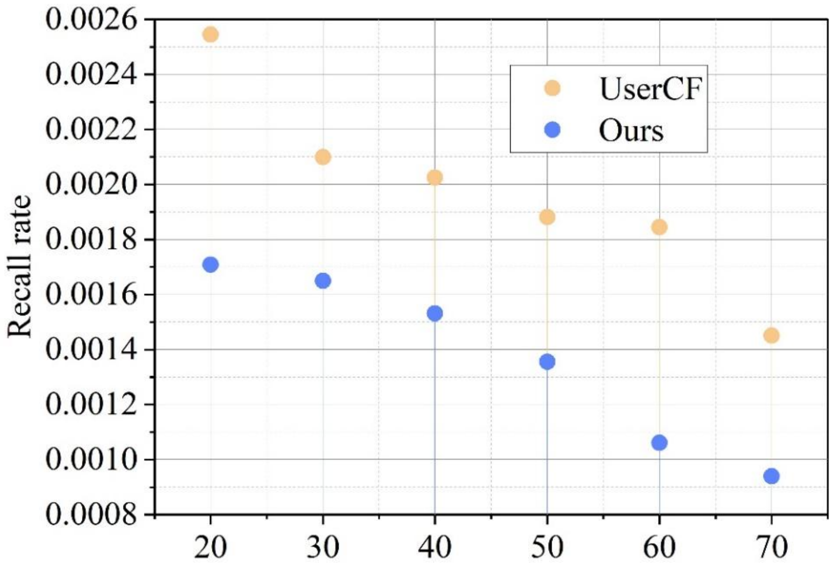


Figure 6. IGA-ARU model is compared with the recall rate of this article.

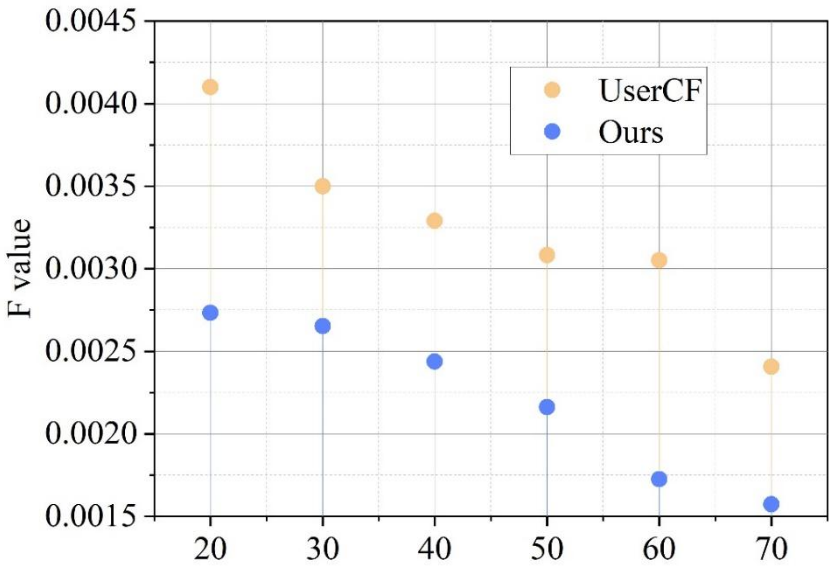


Figure 7. IGA-ARU model is compared to the value of the model F in this article.

(1) Separately, the accuracy, recall and F-value of the user-based collaborative filtering recommendation model show a decreasing trend as the length of the initial recommendation list increases. The accuracy, recall and F-value of this paper's model show a decreasing trend. Moreover, the decreasing trends of all three metrics are approximately the same in both models.

(2) The user-based collaborative filtering recommendation model has a relatively sharp decline in the three indicators when the initial recommendation list length is [20,30]. At [30,60], the degree of decline becomes significantly smoother. After the number of recommended items in the initial recommendation list is more than 60, the degree of decline increases again. The overall degree of decline of the three indicators in this paper's model is relatively gentle.

(3) The user-based collaborative filtering recommendation model is always higher than this paper's model, which indicates that this paper's model loses accuracy compared with the user-based collaborative filtering recommendation model, but the decrease in accuracy means that the degree of diversification increases, which means that the degree of diversification of this paper's model is better than that of the user-based collaborative filtering recommendation model.

5.2. Model Recommendation Diversity Effect

In order to be able to more clearly show the change in the diversity of recommended items in the recommendation list formed by the traditional user-based collaborative filtering recommendation model and the model proposed in this paper, a comparative analysis of the degree of diversity of the two models is carried out to quantitatively represent the degree of diversity of the recommended content by using the index of intra-list similarity, the Hemming distance, and the coverage rate as the criteria, so that a more intuitive and objective conclusion can be drawn. The intra-list similarity of the two models is shown in Figure 8.

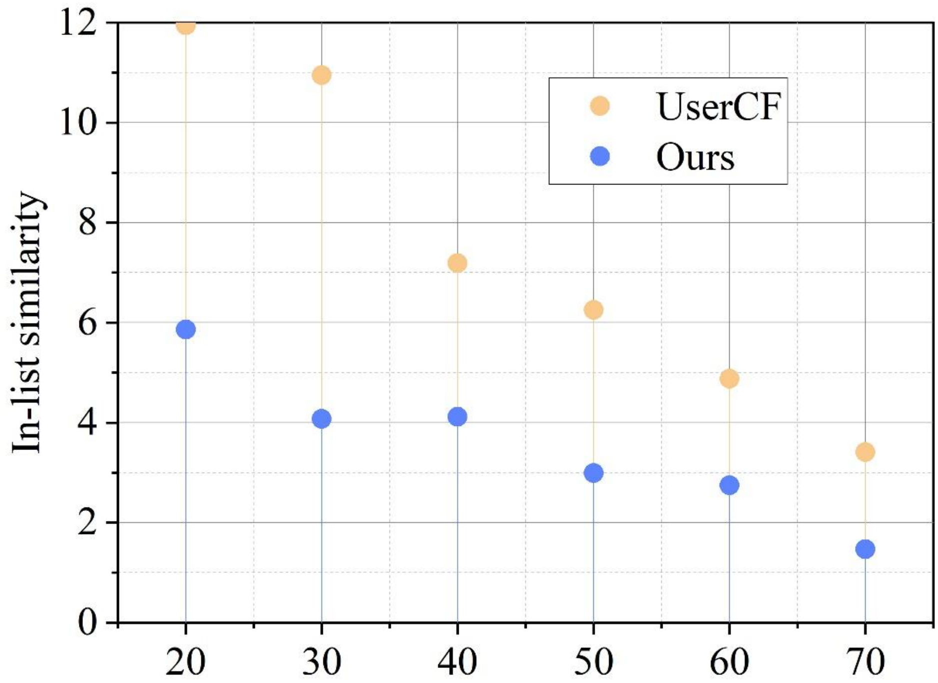


Figure 8. The similarity of the list of two models.

(1) Separately, in the user-based collaborative filtering recommendation model, with the increase of the initial list length, the similarity within the list gradually decreases, indicating that the degree of diversity of the recommendation results gradually increases. In the model proposed in this paper, also with the increase in the number of recommended items in the initial recommendation list, the similarity within the list gradually decreases, indicating that the degree of diversity of the recommendation results gradually increases.

(2) In the recommendation model based on collaborative user filtering, when the length of the initial recommendation list is [20,40], the degree of similarity within the list decreases relatively significantly, and the degree of decline slows down significantly after it is larger than 40. In the model of this paper, when the length of the initial recommendation list is [20,30], the degree of similarity within the list decreases relatively significantly, and after it is greater than 30, the degree of decline slows down significantly. At the same time, the gap between the two is shrinking.

(3) On the whole, under the same initial recommendation list length, the intra-list similarity value of the model proposed in this paper is smaller than that of the user-based collaborative filtering recommendation model, indicating that the degree of individual diversity of the recommendation results obtained in this paper is significantly better than that of the user-based collaborative filtering recommendation model, and that this paper's model has a better performance in terms of improving the diversity of the individuals in the recommendation results. The comparison graph of the Hemming distance calculated by the two models is shown in Fig. 9. The coverage comparison graph is shown in Figure 10.

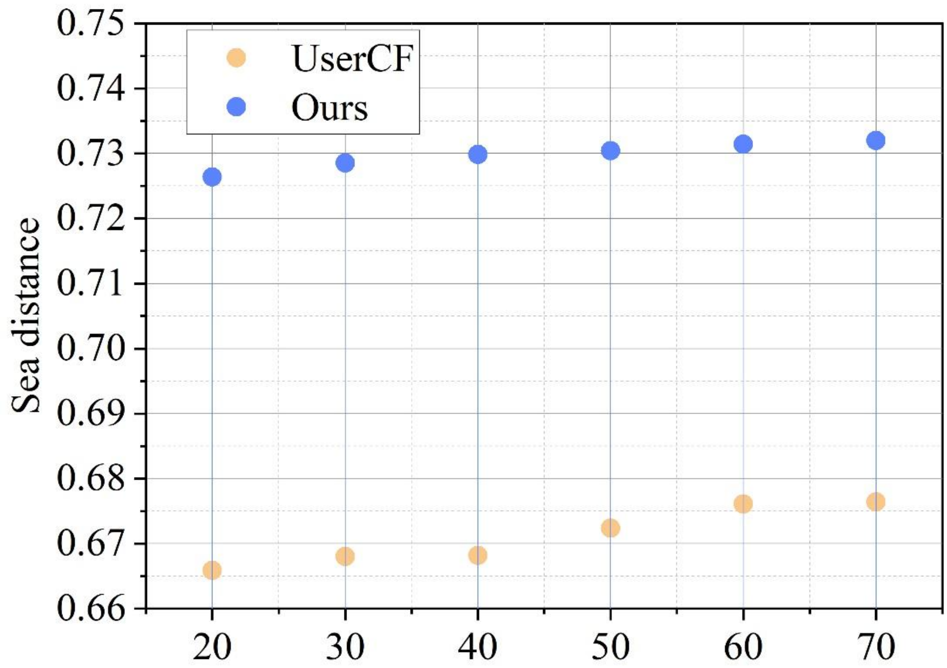


Figure 9. IGA-ARU model is compared with the distance of the model.

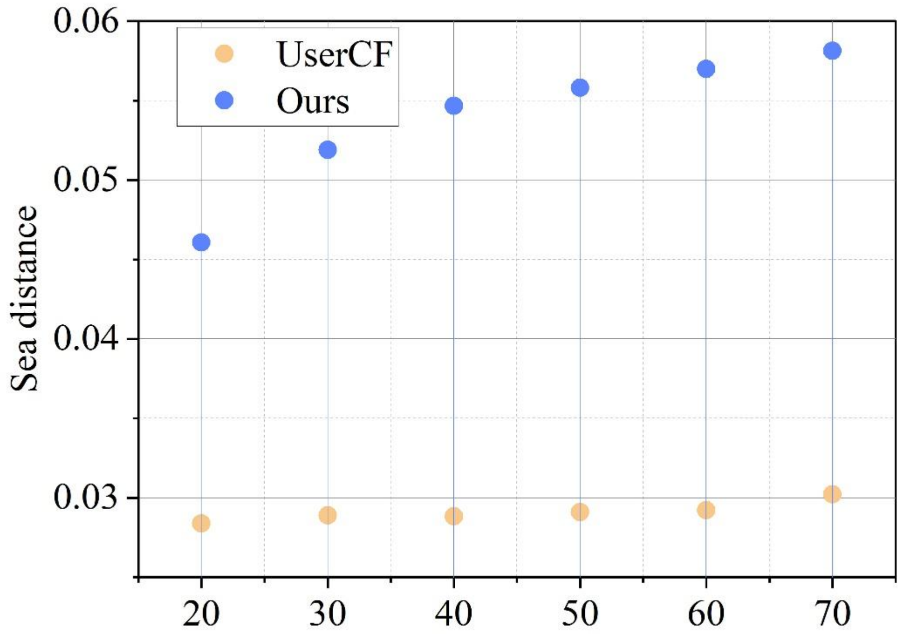


Figure 10. IGA-ARU model is compared with the coverage of this model.

(1) Separately, the Hemming distance and coverage of both this paper's model and the user-based collaborative filtering model increase with the increase in the length of the initial recommendation list, which, according to the relationship between the Hemming distance, coverage, and the degree of diversification, suggests that the overall degree of diversification of both models increases with the increase in the length of the initial list.

(2) The degree of growth of the Hemming distance and coverage of both models is relatively flat, which indicates that both models are relatively flat in terms of the improvement of the overall degree of diversification.

(3) As a whole, under the same initial recommendation list length, the values of the Hemming distance and coverage of this paper's model are larger than those of the user-based collaborative filtering model, which indicates that this paper's model is better than the user-based collaborative filtering

recommendation model in improving the overall diversity effect of the recommendation results.

5.3 Models Improve Information Cocooning Effects

In this paper, by comparing the collaborative filtering recommendation model with the recommendation list generated by this model for the same user, the length of the recommendation list generated by the two models is 10. Observe the changes in the type and number of topics in each item of the two recommendation lists to analyze the improvement of diversified recommendations on the phenomenon of "information cocoon". The distribution of topic types in the recommendation list of the collaborative filtering model is shown in Figure 11, and the distribution of topic types in the recommendation list of this model is shown in Figure 12.

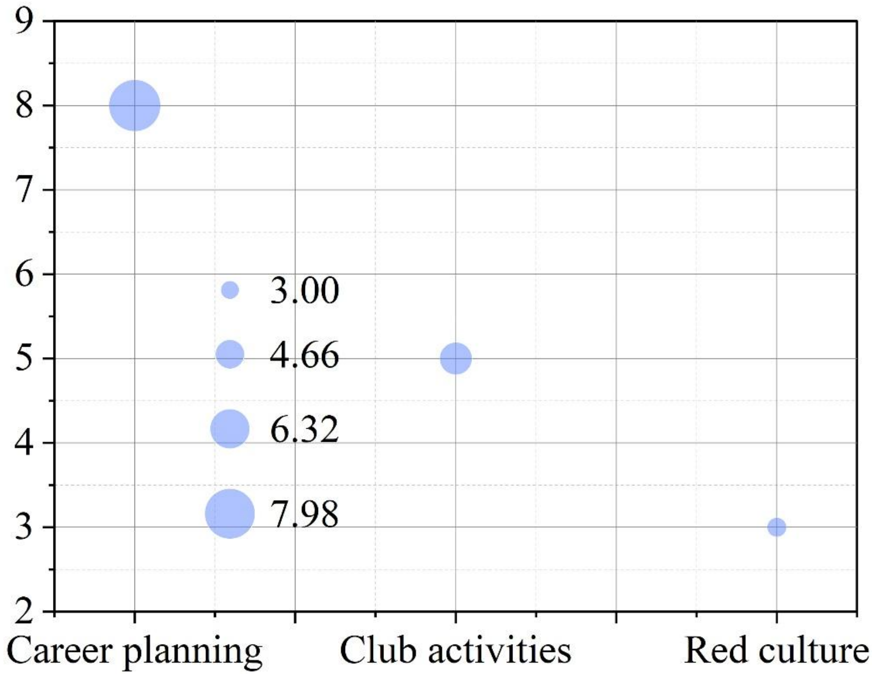


Figure 11. The topic type distribution of the collaborative filtering model recommendation list.

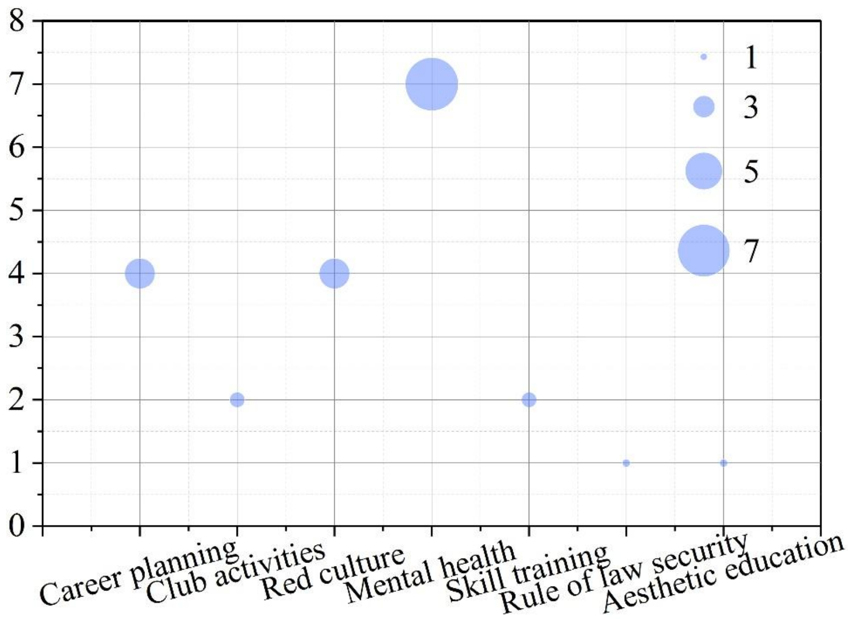


Figure 12. The topic type distribution of the recommended list of this article.

The recommendation lists obtained by five users were also selected, and the number of topic types in the recommendation lists was analyzed by comparing the number of topic types in them. The topic type variety is shown in Figure 13. From multiple users, the collaborative filtering recommendation model recommends videos with less number of topic type kinds, while the number of topic type kinds of recommended videos by this paper's method are increased.

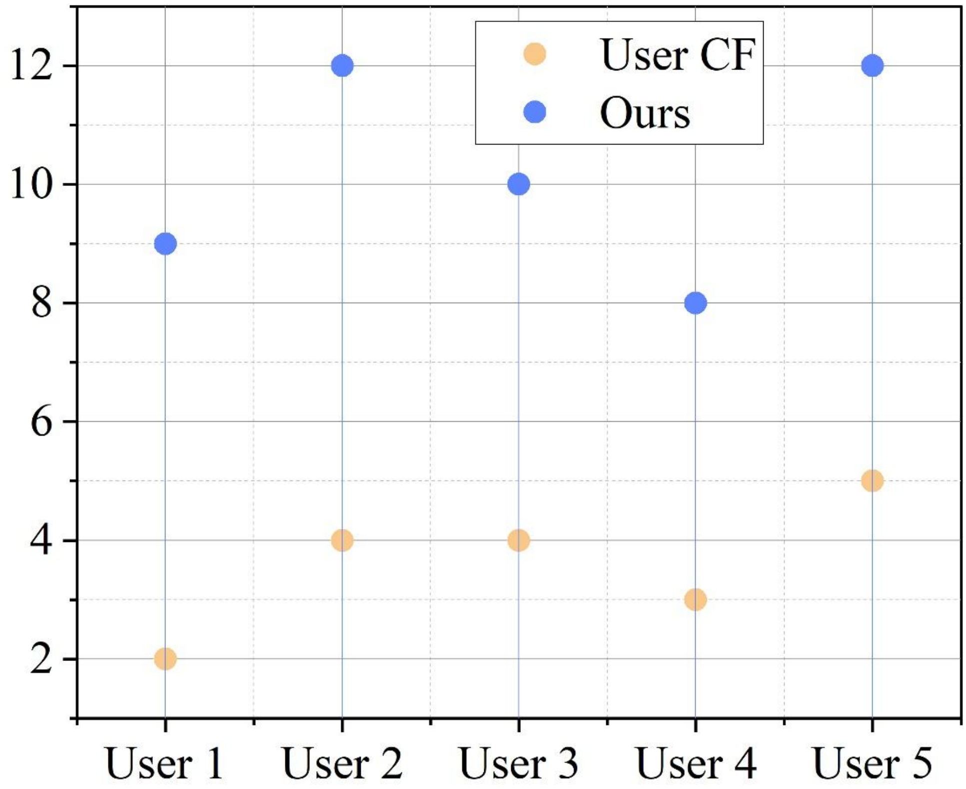


Figure 13. Topic type.

6. Interaction Optimization Strategies Based on Short Video Parenting Platforms in Colleges and Universities

6.1. Improve Personalized Information Recommendation Technology

6.1.1. Ensuring the Timeliness of Personalized Information Recommendation Techniques

In the process of research and development of technical means, to keep abreast of the times. Meet the needs of users at the same time, need to keep pace with the pace of technological development of the times. In the personalized recommendation system, the technology is continuously adjusted and updated, and it is constantly being developed and improved.

The short video parenting platform can be flexibly adjusted to the personalized information recommendation technology through the actual situation of the situation in the later application. Avoid defective recommendation technology, and make real-time improvements and innovations to the existing technology.

6.1.2. Ensuring the Fit of Personalized Information Recommendation Techniques

Personalized information recommendation technology needs to be matched with the short video parenting platform. The author believes that the weighted hybrid technology can be used to apply all the performance of the recommender system in the most direct way in the process of recommendation, to realize the weight distribution of the recommender method in the system in the process of recommendation of the short video parenting platform, and to give full play to the advantages of information recommendation as much as possible.

Technology is an important part of the personalized information recommendation process to achieve the goal. Technical means determine the accuracy of personalized information recommendation. In the personalized information recommendation of the short video parenting platform, while selecting and adopting the recommendation method, the technical means also need to make progress together with the recommendation method.

6.1.3. Use of encryption to protect the security of user information

Short video parenting platforms often extract user behavioral information in the process of personalized recommendation. In order to be able to more accurately understand the user's information as well as behavioral preferences and interest preferences, the short video parenting platform will collect all kinds of private information about the user in the process of the user's use. User behavior includes users' browsing of short videos, liking of short videos, commenting on videos, and so on.

The short video parenting platform needs to set up encryption technology to realize the protection of users' personal privacy. While providing users with recommended videos in the background, user behavior information is strongly protected and encrypted. Prevent access to user information from other areas besides the short-video parenting platform.

6.2. Sound Legal System for Short Video Parenting Platforms

6.2.1. Sound Legal System for Short Video Parenting Platforms

Short video parenting platforms need to collect users' private information during the recommendation process. User privacy protection is particularly important, especially the basic information of users. While improving encryption technology, it is also necessary to improve the legal system of user privacy protection to regulate and supervise the short video parenting platform.

In personalized recommendation, improve and sound the legal system of user privacy protection. Prevent the leakage of personal information of users of short-video parenting platforms to platform merchants for the purpose of profit. To strengthen the protection of personal privacy at the legal level, short video parenting platforms need to strictly comply with the laws and regulations, and do not cross the red line of leaking user privacy.

6.2.2. Improve the Construction of the Copyright Legal System for Short Videos

Strengthen the improvement of online copyright and intellectual property rights, control the publication and broadcasting of short-video parenting platforms, and strictly comply with laws and regulations. If videos published by users of short-video parenting platforms are found to contain similar content, piracy or plagiarism, they should be promptly intercepted and penalized. At the same time as regulating short-video users, it is necessary to formulate penalty regulations for short-video parenting platforms.

6.2.3. Strengthening Legal Awareness of Short Video Parenting Platforms and Users

(c) Strengthening the supervision of short-video parenting platforms in terms of reviewing videos by short-video parenting platforms. First, train short video parenting platforms by means of legal education, so that short video parenting platforms are fully aware of the seriousness of violating the law. Secondly, realize the real name system of short video parenting platform. The real-name system is strictly implemented when users of short-video parenting platforms register and post video content, so that short-video parenting platforms and users are always regulated.

6.3. Strengthening the Management of Short Video Parenting Platforms

6.3.1. Establishment of an Evaluation and Feedback Mechanism for Short-Video Parenting Platforms

In the process of continuously improving and improving the personalized recommendation system of the short video parenting platform, the concept of always meeting the needs of users should be carried out throughout.

First, through the establishment of short video education platform evaluation feedback mechanism. When users in the process of using, found that the platform has unreasonable recommendations, can be timely evaluation and feedback to the short video parenting platform.

Second, when the short-video parenting platform receives the user's evaluation and feedback, it can detect the short-video parenting platform in a timely manner, and make timely adjustments and optimization for unreasonable recommendations.

Finally, the evaluation and feedback mechanism of the short-video parenting platform can be utilized. It can also realize mutual supervision and mutual constraints through mutual supervision between users and users.

6.3.2. Creating a User Guidance Mechanism for Short Video Parenting Platforms

This paper finds that the personalized information recommendation of short video parenting platform relies on the user's usage information, and the subsequent video recommendation is based on the user's behavioral preference information. It is necessary to establish a short video parenting platform user guidance mechanism to actively guide users to actively use the short video parenting platform. Points can be taken to attract customers. The video played on the short video parenting platform has a profit-oriented advertising video, etc., to create revenue for the short video parenting platform.

6.3.3. Establishment of an Audit and Management Mechanism for Short-Video Parenting Platforms

With regard to the broadcasting and forwarding of videos on short-video parenting platforms, it is necessary to establish an audit management mechanism for short-video parenting platforms. It is clear that short-video parenting platforms should fulfill the audit management mechanism when disseminating short videos, and strictly control the disseminated videos. The short videos disseminated should have positive energy and set up correct values. Short-video parenting platforms should increase their auditing and control of short videos.

7. Conclusion

In this paper, based on the short video parenting platform in colleges and universities, we studied the value recognition of users and user portrait modeling, and proposed interaction optimization strategies for the experimental results. The main conclusions are as follows:

(1) Based on the improved RFM model to model the user portrait in the college short video platform, the platform users can be divided into: important development users, loyal promotion users, short-term active users, and ordinary churn users.

(2) The degree of decline of the user-based collaborative filtering recommendation model becomes obviously flat at [30,60]. When the number of recommended items is more than 60, the degree of decline increases, while the overall degree of decline of the model in this paper is relatively smooth. This concludes that the model in this paper has good operational performance.

(3) In the development of short video education platform in colleges and universities, it is necessary to continuously improve the personalized information recommendation technology, improve the legal system of user information protection, and also strengthen the management of short video platform. The establishment of evaluation feedback mechanism, the creation of short video platform user guidance mechanism and the establishment of short video platform audit management mechanism, to achieve the continuous optimization of short video platform personalized information recommendation system.

Funding

This research was supported by the Special Task Project of Humanities and Social Science Research of the Ministry of Education, "Research on the Construction of Short video education platform in colleges and universities under the perspective of big data" (22JDSZ3174).

References

1. Yang, S. (2021). Construction of video courses of physical education and health education in colleges and universities under the MOOC platform. Mobile Information Systems, 2021(1), 9925838.
2. Yuan, Y. (2021, September). Research on the influence of short video platform on college students' values in based on information technology. In 2021 4th International Conference on Information Systems and Computer Aided Education (pp. 908-911).
3. Guo, X. (2025). Short Video Platforms for Learning: What Drives Undergraduates' Satisfaction and Continued Use in Chengdu. Journal of Interdisciplinary Research (ISSN: 2408-1906), 10(1), 144-156.
4. Xiong, S., Chen, J., & Yao, N. (2024). A multidimensional framework for understanding problematic use of short video platforms: the role of individual, social-environmental, and platform factors. Frontiers in Psychiatry, 15, 1361497.

5. Roy, P. (2024). An Investigation into How Global Convergence of Media is Influencing Cultural Exchange, Diversity and Homogenization of Media Content in Different Regions. Global Media and Cultural Exchange, 1.
6. Zhang, N., Hazarika, B., Chen, K., & Shi, Y. (2023). A cross-national study on the excessive use of short-video applications among college students. Computers in Human Behavior, 145, 107752.
7. Alley, A., & Hanshew, J. (2022). A long article about short videos: A content analysis of US academic libraries' use of TikTok. The Journal of Academic Librarianship, 48(6), 102611.
8. Fritz, A. M., & Smith, A. M. (2024). Marketing higher education on YouTube: a content analysis of college promotional videos. Journal of Marketing for Higher Education, 1-22.
9. Pham, H. H., Farrell, K., Vu, H. M., & Vuong, Q. H. (2017). Using YouTube videos to promote universities: a content analysis. Technics Technologies Education Management (TTEM), 12(1), 58-72.
10. Chen, X., Nie, W., & Zhou, M. (2020, March). Short Video: A New Way to Spread Campus Culture. In 4th International Conference on Culture, Education and Economic Development of Modern Society (ICCESE 2020) (pp. 561-563). Atlantis Press.
11. Liu, Y., Chiu, D. K., & Ho, K. K. (2023). Short-form videos for public library marketing: performance analytics of Douyin in China. Applied Sciences, 13(6), 3386.
12. Zhang, J., & Zhang, X. (2024, December). Exploration of Precise Knowledge Service Mode of Video Websites Based on Group User Portrait. In Proceedings of the 2024 8th International Conference on Natural Language Processing and Information Retrieval (pp. 130-135).
13. Li, G., Chen, W., Yan, X., & Wang, L. (2022). Modeling and analysis of group user portrait through WeChat mini program. Wireless Communications and Mobile Computing, 2022(1), 2515962.
14. Chen, Y., He, J., Wei, W., Zhu, N., & Yu, C. (2021). A multi-model approach for user portrait. Future Internet, 13(6), 147.
15. Xinlin, J., & Wenting, L. (2023, September). Precise Positioning and Personalized Recommendation Algorithm of Smart Library Based on User Portrait. In 2023 International Conference on Power, Electrical Engineering, Electronics and Control (PEEEC) (pp. 531-535). IEEE.
16. Sang, J., Deng, Z., Lu, D., & Xu, C. (2015). Cross-OSN user modeling by homogeneous behavior quantification and local social regularization. IEEE Transactions on Multimedia, 17(12), 2259-2270.
17. Shi, R., Wang, M., Liu, C., & Gull, N. (2023). The influence of short video platform characteristics on users' willingness to share marketing information: based on the SOR model. Sustainability, 15(3), 2448.
18. Yang, C., Yan, H., Yu, D., Li, Y., & Chiu, D. M. (2017, August). Multi-site user behavior modeling and its application in video recommendation. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 175-184).
19. Liu, Q., & Deng, W. (2022). Animation user value portrait based on RFM model under big data. Mathematical Problems in Engineering, 2022(1), 8246540.
20. Xu, Y., Wang, Q., Roxas, R. E. O., Sun, Y., Li, X., & Lv, X. (2022). ONLINE SHOPPING USER PORTRAIT MODEL CONSIDERING PSYCHOLOGICAL AND EMOTIONAL FACTORS. International Journal of Neuropsychopharmacology, 25(Supplement_1), A21-A21.
21. Yiliang Xie,Chen Chen & Rui Zhang. (2025). Research on User Profiling of “Internet + Nursing Service” Platform Based on Improved RFM Model. Open Journal of Social Sciences,13(04),184-195.
22. Victor Hugo Antonius & Devi Fitrianah. (2024). Enhancing Customer Segmentation Insights by using RFM + Discount Proportion Model with Clustering Algorithms. International Journal of Advanced Computer Science and Applications (IJACSA),15(3).
23. Sinem Guney,Serhat Peker & Cigdem Turhan. (2020). A Combined Approach for Customer Profiling in VIDeo on Demand Services Using Clustering and Association Rule Mining. IEEE Access,8,84326-84335.