Study on New Media Micro-Video Personalized Recommendation System Based on Neural Network

Lyu Jiaqi

School of Media and Communication, Taylor's University, Malaysia, 47500

lyujiaqi@sd.taylors.edu.my

Keywords: Micro-Video; Personalized Recommendation; Neural Network

Abstract: In recent years, the rapid growth of cloud computing, big data, mobile Internet and various video services has led to the explosive growth of micro-video data. How to quickly and effectively find videos that users are interested in and provide personalized recommendation services has become an inevitable trend of micro-video business development. In this article, a three-dimensional DNN is proposed to extract the depth features of micro-videos, and a user behavior preference model with joint depth feature labels is constructed to better describe the user's behavior preferences. The test results show that, compared with FCA, this method has obvious advantages in the later stage of operation, and the error is reduced by 33.84%. The accuracy of this algorithm for micro-video image feature recognition is higher, which is 21.82% higher than that of the comparison algorithm. The model abstract, by introducing the relevant weight update feedback, updates the user's behavior preference in real time, and significantly improves the accuracy of personalized recommendation results of micro-videos.

1. Introduction

The popularity of micro-video platform has brought creativity and entertainment to people's lives [1]. On the micro-video platform, every user who watches micro-video is both a content consumer and a content producer [2]. A great quantity micro-video materials are produced in the Internet every day, which leads to serious overload of user information, and it is difficult for users to independently choose their interesting content [3]. With the vigorous growth of micro-videos, the amount of data around micro-videos will also expand rapidly in a short time [4]. For example, there will be more and more information about micro-videos, video introductions, video reviews, sharing and other related information, and the relationship between information will become more and more complicated [5-7]. Faced with such a huge amount of new media micro-videos, it is difficult for traditional collaborative filtering algorithms to take conventional means to extract the content features of new media micro-videos, and it is impossible to accurately recommend new media micro-videos [8-9].

The detection of micro-video content requires video frame segmentation, that is, random sampling. Video frame extraction refers to using key frames of video to represent the complete meaning of the whole video, aiming at different types of video coding formats, frame rates, bit rates, video resolutions and different kinds of videos [10]. The input layer of the DNN model is a set of high-dimensional space vectors, that is, feature vectors of users and videos, which are textual descriptive information of videos and users [11]. At the same time, deep learning has made technological breakthroughs in the fields of natural language processing, image processing and speech processing, and new media micro-video recommendation technology has emerged as the times require [12]. In micro-video software, every user who watches micro-video is both a consumer and a producer of micro-video content [13]. There are a great quantity micro-video materials produced in the Internet every day, which causes serious information overload problems for users, and it is difficult for users to independently select interesting content from them [14]. This article constructs a personalized recommendation model of micro-video based on neural network driven by big data. The model uses neural network model to model historical data, optimizes the parameters of ANN prediction model, and outputs the optimal solution, thus realizing personalized recommendation of new media micro-

Copyright © (2023) Francis Academic Press, UK DOI: 10.25236/icmmct.2023.007

videos.

In recent years, with the rapid growth of the Internet, especially the sudden emergence of e-business, personalized recommendation has become an essential service of major mainstream websites [15]. However, compared with today's booming e-business websites, there is still a big gap in the personalized recommendation service level of micro-videos. Micro-videos are short, rich in content and interactive. At the cognitive level, users identify the external rich world through a great quantity micro-videos [16]. The input layer of the DNN model is a set of high-dimensional space vectors, that is, feature vectors of users and videos, which are textual descriptive information of videos and users. The output layer of the DNN model is a set of low-dimensional dense semantic vector spaces, which is the abstract description result of the low-level text features extracted by the neural network. In order to ensure the integrity and reliability of video data, before formal modeling and analysis, text description data needs to be preprocessed, which mainly includes video filtering and video fusion. Video filtering is mainly to remove some new media micro-videos with missing important text information. Video fusion is mainly to merge new media micro-videos with similar important text information. This article mainly analyzes the personalized recommendation mechanism of micro-video driven by big data and the specific application of data mining in micro-video platform. The research contains the following innovations:

(1) In this article, a three-dimensional DNN is proposed to extract the depth features of micro-videos, and a user behavior preference model with joint depth feature labels is constructed to better describe the user's behavior preferences.

(2) The model carries out modeling and in-depth training of users and new media micro-videos, grasps the internal relationship between users' interest preferences and new media micro-videos, and realizes personalized recommendation of new media micro-videos.

(3) Through the experimental analysis of the recommendation system, it is found that through the learning and training of the deep neural network (DNN) model, the recommendation system has high recommendation performance and can fully grasp the users' interest preferences in new media micro-videos.

2. Methodology

2.1 Feature Extraction of Micro-Video Based on Neural Network

Micro-videos on the Internet have the characteristics of wide diversity, single content and short duration. If traditional visual features such as color, texture, shape and other low-level visual features are directly used, it is difficult to form an effective video content representation. In this study, an optimized 3D convolution neural network algorithm is proposed to extract the depth features of micro-videos. In the short video recommendation application scenario, you can first analyze the feature information of the short videos that users watched before, then compare these feature information with the features of other short videos, and finally recommend TopN short videos with the highest similarity to users [17]. It can solve the problem of cold start of new objects well, but it is difficult for users to obtain novel recommendation results. Because the features of the recommended objects match the user's interest model, the recommendation results obtained are similar to those before.

The framework of micro-video depth feature extraction based on 3D DNN is shown in Figure 1. For a micro-video, a three-dimensional DNN model is first established, and then the micro-video data set is used for network training and optimization. Finally, the trained network is used to extract depth features, and the classification of micro-video is realized through Softmax classifier.

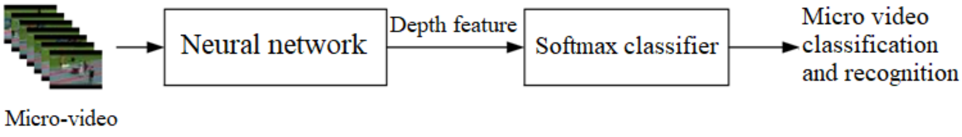

Figure 1. Framework of micro-video depth feature extraction based on 3D neural network

In order to extract the depth features of micro-video, a three-dimensional DNN based on modified linear units is proposed in this article, as shown in Figure 2, which includes three basic components: the activation function ELU (yellow block), the convolution layer conv (blue block) and the pool layer pool (brown block). The three components gradually improve the abstraction ability of the proposed features by continuously stacking the convolution layer and pool layer, so as to obtain deeper features with more representation ability.

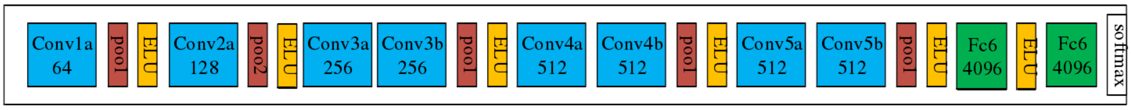

Figure 2. Three-dimensional DNN

In order to improve the ability of feature representation, the DNN must be trained on large data sets to obtain effective feature representation [18]. Therefore, after designing the three-dimensional DNN, this section will use data preprocessing to enhance the data, and then realize the debugging and optimization of network parameters based on the gradient descent algorithm. Finally, the depth features of micro-video will be extracted based on the trained network.

2.1.1 Data Preprocessing

These text description data are mainly used for feature extraction of new media micro-video content, which lays the foundation for the subsequent new media micro-video modeling. Remove the keywords with less importance, keep those with more importance, and use them as the tag information of the video, so as to describe the whole video features and complete the video modeling. Data sets usually need to be preprocessed before training DNNs, including data enhancement and improving the randomness of data. Considering the characteristics of short time and single content of micro-video, the environment between the beginning and end of micro-video is not different, so it can be divided into multiple segments to enrich the database. The micro-video data preprocessing process is shown in Figure 3. First, each video in the training set is divided into three parts: head, middle and tail, and then the video clips are randomized on the database to complete the data preprocessing. Finally, the preprocessed data set is input into three-dimensional DNN to realize the network training.

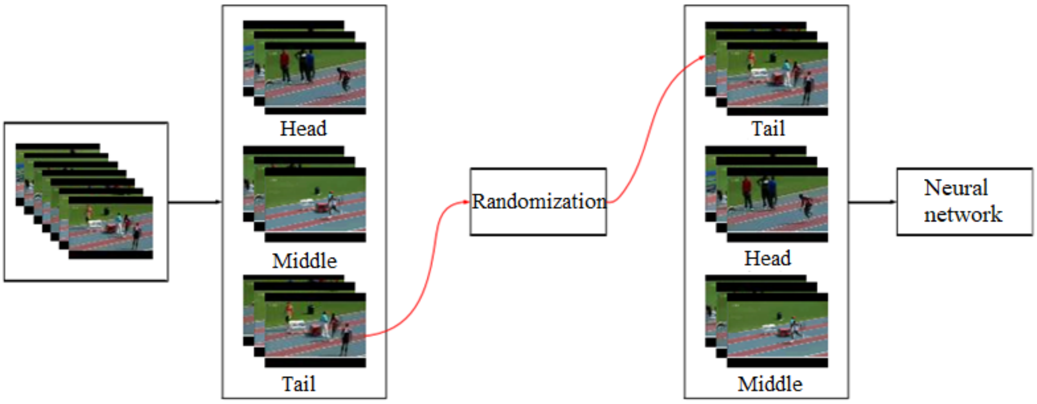

Figure 3. Micro-video data preprocessing process

2.1.2 Network Training and Parameter Adjustment

Network training and parameter adjustment is an important step in the application of deep network to practical applications. The quality of network training directly affects the performance of network in practical applications [19]. Network parameter adjustment refers to selecting appropriate initial parameters for gradient descent method, while optimization refers to optimizing network parameters based on gradient descent algorithm. In this study, momentum optimization algorithm is used to optimize network parameters. The momentum optimization algorithm is an improved version of the traditional gradient descent algorithm, which alleviates the ill conditioned problem of the traditional optimization algorithm, thus improving the efficiency of network training. The flow of momentum random gradient descent algorithm is shown in Table 1.

Table 1. Flow of momentum random gradient descent algorithm

Input: learning rate ε, momentum parameter α, initial weight θ, initial velocity v
While stop criteria not reached do
Take m samples from the training set $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$, corresponding target as $y^{(i)}$
Compute gradient estimation: $g \leftarrow \frac{1}{m}\nabla_\theta \sum_i L\left(f\left(x^{(i)};\theta\right), y^{(i)}\right)$
Calculation speed update: $v \leftarrow \alpha v - \varepsilon g$
Apply updates: $\theta \leftarrow \theta + v$
End while
Output: neural network after training

For the learning rate, the value of $0.1 \sim 10^{-5}$ is generally taken. In this study, the learning rate is set to 0.0001 to prevent gradient oscillation and ensure training convergence [20]. The momentum parameter is set as 0.9 to ensure the maximum step size when the continuous gradient points to the same direction. The batch size is set to 50 to ensure the accuracy of gradient estimation, while the number of iterations is set to 50000 to ensure that all data is fully utilized.

2.1.3 Depth Feature Extraction of Micro-Video

According to the feature extraction results of video and users, the initial feature information of video and users is abstracted layer by layer, and finally mapped into two feature vectors with low-dimensional semantic correlation. The distance between these two feature vectors corresponds to the correlation between users and videos, and the correlation results are used as the scoring basis for new media micro-video recommendation. If the work doesn't meet the standards of each level, it won't enter the next traffic pool. That is to say, if all data indicators of the first traffic push are qualified, it will enter the second traffic pool. If all data of the first traffic pool are qualified, it will be pushed for the third time. If the data doesn't meet the standards during the third push, the video playback volume will stay at the third level.

After completing the training, the three-dimensional DNN is used to extract the depth features of micro-video. The depth feature extraction process of micro-video based on three-dimensional DNN includes video cutting, feature extraction, feature averaging and regularization. For a micro-video, it is first divided into 16 frames of video clips, and the overlap of 8 frames between adjacent clips is maintained, as shown in Figure 4.

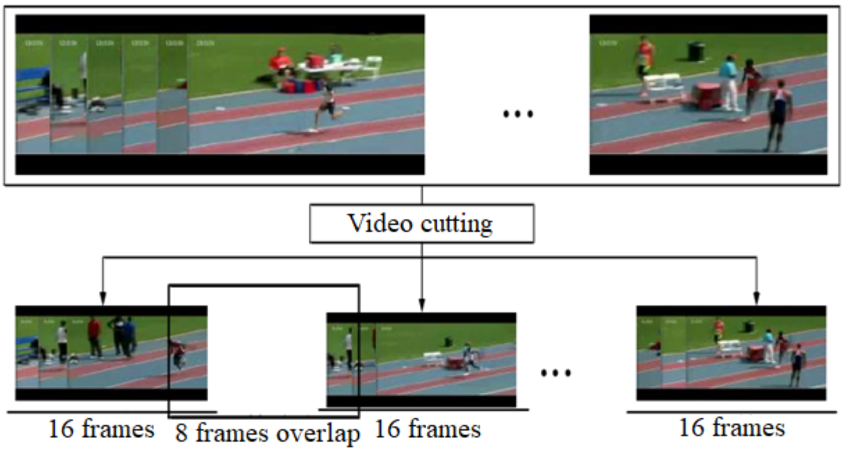

Figure 4. Micro-video cutting scheme

The features extracted by the semantic-based image classification and retrieval system are actually the bottom-level visual features. However, because they are extracted from the image database with specific known semantic categories, these features already have semantic information. Under this premise, they can be called semantic features of images, thus realizing the mapping from the bottom-level features to the high-level semantics. The structure of micro-video image mosaic detection model based on deep learning is shown in Figure 5.

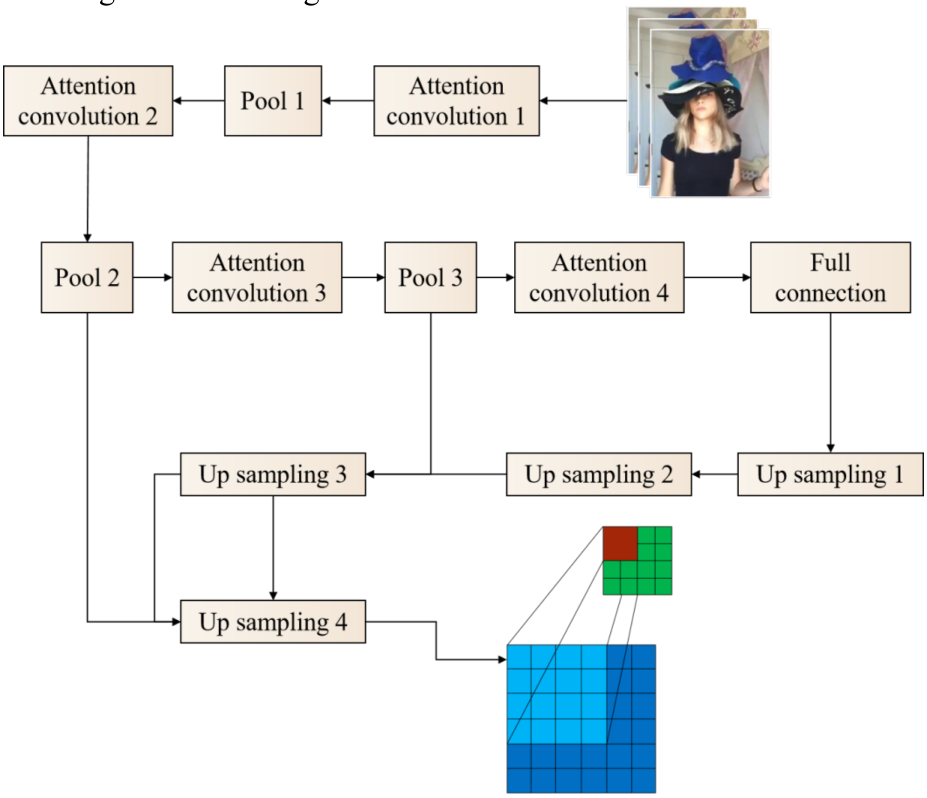

Figure 5. Detection model of micro-video image stitching

In the task of video natural language description generation, the representation and extraction of video features is the first step and the key link, which plays a vital role in the output of the subsequent natural language model. The input micro-video signal $I(X,t)$ is compared with N distribution models, and then the matching model is updated. If:

$$\left|I_j(X,t)-\mu_{ij}(X,t)\right|<\tau D_{ij}(X,t) \tag{1}$$

The $I(X,t)$ matches the model p_i . Where τ is a global threshold, i represents the i -th distribution model, and j represents the component in space (s,r,g).

If $I(X,t)$ matches more than one p_i at the same time, the distribution model with high probability, small variance and small difference from $I(X,t)$ is selected for updating. That is, the distribution model satisfying the minimum similarity distance $d_i(X,t)$ is updated. $d_i(X,t)$ is defined as:

$$d_i(X,t)=\sum_{j=s,r,g}\frac{\left|I_j(X,t)-\mu_{ij}(X,t)\right|D_{ij}(X,t)}{h_{ij}(X,t)} \tag{2}$$

Update the matching p_i according to the following formula:

$$\mu_{ij}(X,t+1)=(1-\alpha)\mu_{ij}(X,t)\alpha I(X,t) \tag{3}$$

$$D_{ij}(X,t+1)=\min\left\{\left[(1-\beta)D_{ij}^2(X,t)+\beta\left(I(X,t)-\mu_{ij}(X,t)\right)_2\right]^{1/2},D_{\max}\right\} \tag{4}$$

$\alpha\in(0,1)$ is the update factor of mean, which determines the update rate of mean, $\beta\in(0,1)$ is the update factor of variance, which determines the update rate of variance, and $D_{\max}$ is the estimated value of the largest variance in all models, which is the global upper limit of variance.

If the color distribution in a picture conforms to a certain probability distribution, the moment of color can be used as a feature to distinguish different color distributions. The three lower moments of color are mathematically expressed as:

$$\mu = \frac{1}{M \times N} \sum_{i=1,j=1}^{i=M,j=N} p_{i,j} \tag{5}$$

$$\sigma = \left(\frac{1}{M \times N} \sum_{i=1,j=1}^{i=M,j=N} \left(p_{i,j} - \mu\right)^2 \right)^{\frac{1}{2}} \tag{6}$$

$$S = \left(\frac{1}{M \times N} \sum_{i=1,j=1}^{i=M,j=N} \left(p_{i,j} - \mu\right)^3 \right)^{\frac{1}{3}} \tag{7}$$

Where $p_{i,j}$ is the color component value of the pixel located at the coordinate (i,j) in the image, and M,N is the length, width and pixel number of the image, respectively. Key frame extraction is the key to establish video shot data classification and index. It not only solves how to extract key frames that can represent the content of the shot from the sequence of shots, but also analyzes their content and category according to the key frames.

2.2 Construction of Micro-Video User Behavior Preference Model

In the process of user behavior preference modeling, if we only rely on a single tag or micro-video visual features to build a model, it is difficult to effectively describe user behavior preferences because of the failure to consider the complementarity between the data. For tags, they can provide semantic clues of video, but it is difficult to achieve a fine representation of user behavior preferences. Although visual features can represent video content, they lack semantic information. Compared with traditional low-level visual features, depth features have higher semantic abstraction and stronger representation ability. Therefore, this article combines the depth features and tags. First, it uses the tag sorting technology to build the user's tag behavior preference model, then uses the micro-video depth features to build the user's depth feature preference model, and finally combines the tag depth features to build the user's behavior preference model, so as to realize the effective description of the user's behavior preference.

The overall framework of the user behavior preference model constructed in this article is shown in Figure 6. Firstly, based on the extracted depth features of micro-video, the micro-video representation is established, and the maximum likelihood estimation method is used to construct the user depth feature behavior preference model; then, according to the universality of tags, the tags are sorted to realize the representation of Video Tags, and the moving average method is used to build the user tag behavior preference model; finally, a user behavior preference model based on the joint tag depth feature is constructed by the method of weighted voting.

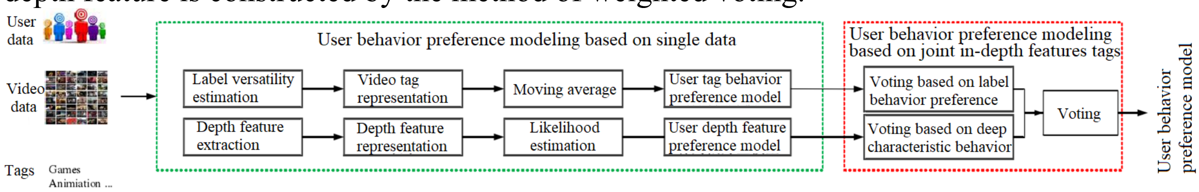

Figure 6. The overall framework of the user behavior preference model

As can be seen from Figure 6, the overall architecture of the user behavior preference model mainly includes two modules: user behavior preference modeling based on single data and user behavior preference modeling based on joint in-depth features tags [21].

In the process of user behavior preference modeling, if we only rely on a single tag or micro-video visual features to build a model, because we fail to consider the complementarity between the data, the model can effectively describe user behavior preferences.

Therefore, this article combines the depth feature tag feature to build a user behavior preference model to realize the description of user behavior preference. U_T is used to represent the user label behavior preference model, and U_D represents the user depth feature behavior preference model, then the user behavior preference model U of the joint depth feature tag is expressed as follows:

$$U=\lambda U_T+(1-\lambda)U_D \tag{8}$$

Where λ is the weight of user label behavior preference model in the joint model.

The micro-video user behavior preference model in this study is built based on the depth features and labels used to plan. For new micro-video, because there is no label and rating data temporarily, the paper will use the user depth features in the proposed model for personalized initial recommendation, which reduces the data sparsity and cold start problems to a certain extent.

2.3 Design of Micro-Video Personalized Recommendation System

After building the user behavior preference model, we need to comprehensively consider the impact of video tags and key frames on users' viewing decisions according to the characteristics of micro-video, so as to update the user behavior preference model in real time and realize personalized recommendation. In this study, a micro-video personalized recommendation system based on user behavior preference model is designed and implemented, which can detect the key frames of micro-video through neural network algorithm, and then use the updated user behavior preference model to recommend the optimized micro-video results to users.

The overall architecture of the micro-video personalized recommendation system is shown in Figure 7, including four modules: data set collection module, data preprocessing module, user behavior preference model update module and personalized recommendation module.

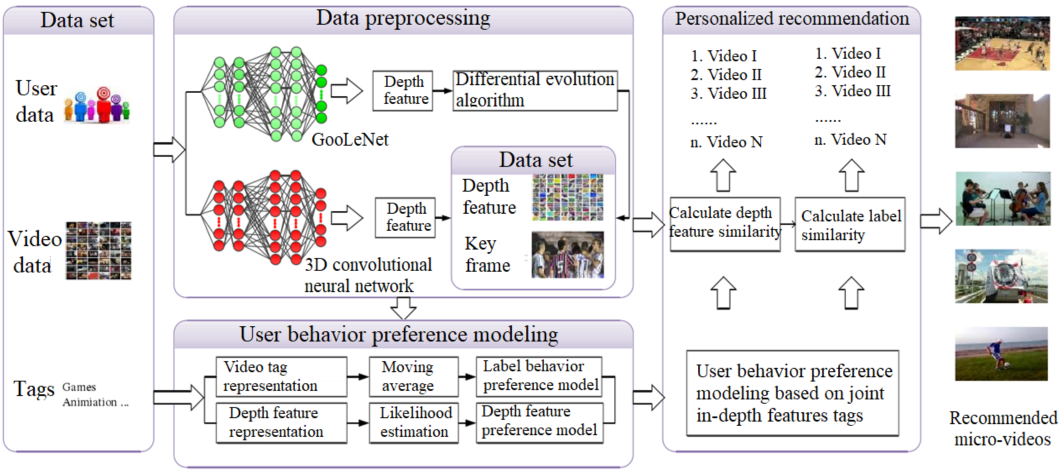

Figure 7. The overall framework of micro-video personalized recommendation system

2.3.1 Data Set Acquisition Module

The data set module collects data related to personalized recommendation, including user data, micro-video data, tags, etc. User data is the user's historical micro-video viewing record, which can reflect the user's behavior preferences; micro-video is the video data to be recommended to users in the personalized recommendation system; tags are data added by users and represent the semantic information of micro-video.

(1) User data

User data mainly refers to the historical operation records of users on new media micro-video. Users' historical operation data of micro-video usually includes two situations: first, the user moves the video playback progress bar and browses the new media micro-video roughly, and the actual viewing time of the user is less than the total video time, which probably indicates the user is not really interested in this new media micro-video; second, the user watch the new media micro-video completely without fast forward browsing, and the actual viewing time of the user is equal to the total length of the video, which can be considered that the user is interested in this video. Therefore, in order to distinguish the two operations of users on new media micro-video, the index of new media micro-video playback duration ratio is added when judging whether users' operations on new media micro-video are effective. The specific calculation formula of micro-video playback duration ratio is as follows:

$$\tau_{ia} = \frac{t_{ia}}{T_i} \tag{9}$$

Where τ_{ia} is the length of time that user a watches the new media short visual frequency i; T_i is the total duration of short visual frequency i. Only when the ratio of τ_{ia} to T_i is greater than a given threshold can the user be considered to be effective in new media micro-video operations and record them. Generally speaking, the actual threshold of new media micro-video is about 0.3. So far, the user's historical operation record of new media micro-video is obtained, and the user feature vector is obtained, which lays the foundation for subsequent user modeling.

(2) Micro-video data

Like other conventional videos, the text description data of micro-video are nothing more than: video title, basic video content, video publishing region, video planning, screenwriter and starring, video publishing time and publishing author, etc. These text description data are mainly used for feature extraction of new media micro-video content, which lays the foundation for subsequent new media micro-video modeling.

2.3.2 Data Preprocessing Module

In order to ensure the integrity and reliability of video data, text description data needs to be preprocessed before formal modeling and analysis. The preprocessing module includes two parts: depth feature extraction and key frame detection of micro-video. As mentioned above, the depth features of micro-video are extracted by three-dimensional DNN to obtain better feature representation, and the extracted depth features participate in the construction of user behavior preference model; for the key frame detection part, firstly, GoogLeNet network, which is recognized to have the best classification effect, is selected to extract the depth characteristics of the current micro-video frame, and then the key frame of micro-video is detected by differential evolution method, which is used as a summary description of micro-video and acts on the personalized recommendation mechanism. The specific process is shown in Figure 8.

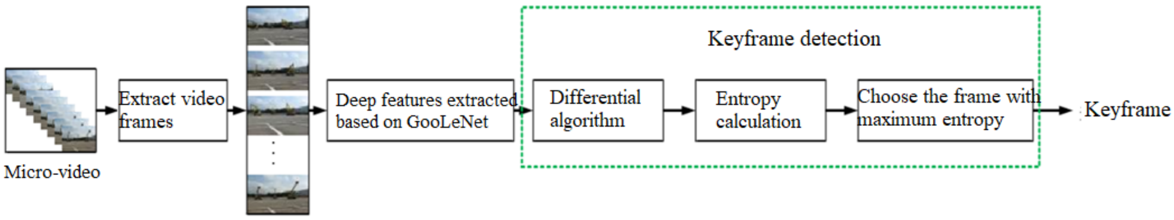

Figure 8 The flowchart of key frame extraction form micro-video

2.3.3 User Behavior Preference Model Update Module

The internal characteristics or attributes of the areas separated by the boundary are the same, while the internal characteristics or attributes of different areas are different. The duration of micro-video is generally defined within five minutes, but now it is more within one minute. Micro-videos on the media platform are mostly people and things in life that users share independently. Their logic is simple, and they are more like graphic content. These micro-videos are different from the previous video research objects, because they do not have the exact behavior range or character target, so they can't only analyze human behavior, and the general feature extraction algorithms can't fully represent all the information of this kind of micro-videos. There are differences among different modal feature data, and direct addition will eliminate the uniqueness of various features, but will have a negative impact on the results. Under the background of multi-core mapping, the high-dimensional space is decomposed into the combination of several low-dimensional feature spaces.

Since user behavior preferences will change with time, geographical location and other factors, updating user behavior preference model based on user real-time feedback is a necessary means to improve user modeling ability. In this study, the relevance feedback of weight update is used to update the user label behavior preference model and the parameters of the user depth feature behavior preference model. First, collect the user's real-time viewing data, and then use the data as the basis for updating the user's behavior preference model. That is, if the user clicks and watches a video, it means that the user is interested in the video. At this time, it is necessary to update the user's behavior preference model to ensure the consistency between the model and the real behavior preference. The parameter update process of the user behavior preference model is shown in Figure 9. By tracking the user's feedback on the push results, and based on the feedback, the dynamic update of the user behavior preference model is realized.

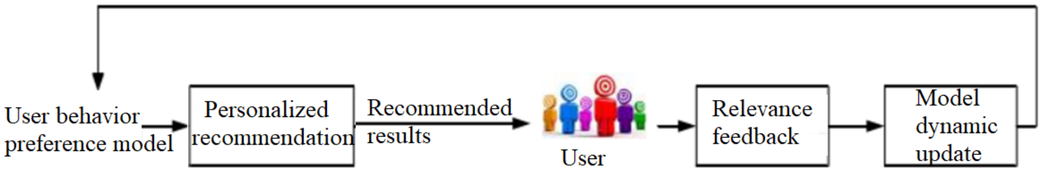

Figure 9 Parameter update process of the user behavior preference model

For users, the personalized recommendation system first recommends micro-video that may be of interest to them based on their behavioral preference model of joint tag depth characteristics. Then, if the user clicks and watches a micro-video, it is considered that the micro-video meets the user's behavior preference, so it is necessary to update the user's behavior preference model according to the current micro-video. The weight updates of user depth feature behavior preference model and user label behavior preference model are shown in formula (10) and formula (11) respectively:

$$U_D^{t+1} = \frac{n \times U_D^t + v_D}{n+1} \tag{10}$$

$$U_D^{t+1} = \alpha U_T^t + (1-\alpha) v_T \tag{11}$$

Where U_D^t and U_D^{t+1} represent the user behavior preference model before and after the update respectively; U_T^t and U_T^{t+1} represent the pre update and post update user tag behavior preference models respectively. The updated user behavior preference model is shown as follows:

$$U^{t+1} = \lambda U_T^{t+1} + (1-\lambda) U_D^{t+1} \tag{12}$$

2.3.4 Personalized Recommendation Module

Based on the feature depth extraction of micro-video and user behavior preference model, the personalized recommendation module is designed. Firstly, the similarity between video and user tag behavior preference model and user depth feature behavior preference model is calculated respectively; then, combine the similarity results of the two to construct the joint similarity results and sort them; finally, recommend the top n micro-video to users. The similarity matching module mainly determines whether the micro-video meets the user's behavior preference, that is, when the content of the micro-video is highly similar to the user's behavior preference model, it means that the user may be more interested. The similarity matching criteria between micro-video and user tag behavior preference model and user depth feature behavior preference model are shown in equations (13) and (14):

$$S_T = \frac{1}{N}\sqrt{(U_T)^T v_T} \tag{13}$$

$$S_D = \frac{1}{M}\sqrt{(U_D)^T v_D} \tag{14}$$

Where S_T and S_D are the similarity between video v and user tag behavior preference model U_T and user depth feature behavior preference model U_T; v_T and v_D represent micro-video under depth feature representation and micro-video under label representation respectively.

After getting the similarity between micro-video and user depth feature behavior preference model and user label behavior preference model, we need to calculate the final similarity based on the results of the two, and then recommend the micro-video with the top ranking to users. The calculation formula of joint similarity S is as follows:

$$S = \lambda S_T + (1-\lambda)S_D \tag{15}$$

3. Result analysis and discussion

Select the features that need to remove the irrelevant information, establish an evaluation standard to distinguish which features or combinations of features are helpful for classification, judge which features or combinations of features are redundant or complementary, and select complementary features to remove redundant features, thus avoiding the problem of over-training in a labeled data set. It can be measured by data mining with the corresponding model. By capturing the effect data and continuously optimizing the delivery model, optimizing the resource recommendation model based on data analysis, and optimizing the delivery combination strategy, the best marketing effect can be achieved. The convergence index is used to compare the mining results of FCA and the recommended method in this article. The convergence of training loss is shown in Figure 10.

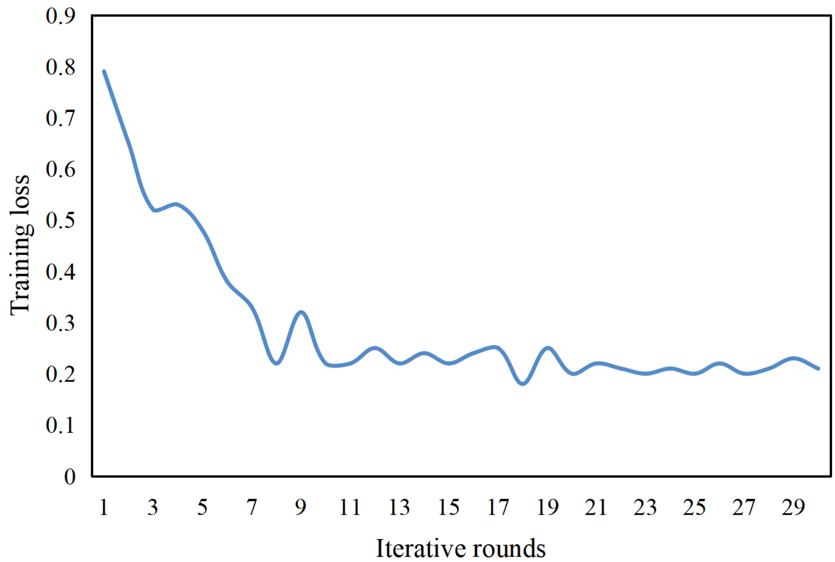

Figure 10. Convergence comparison results

This method can obtain reasonable and scientific image feature analysis results. The method in this article is applied to video classification and retrieval, and the optimization characteristics are good and the convergence speed is fast. The performance of the classification system largely depends on the discrimination ability of visual features. If features have good discrimination, the accuracy of the classifier will be improved.

The output feature dimension reflects the intrinsic feature relationship between users and videos. With the increase of neural network layers, the standard average error value and recall rate of recommendation system are increasing, but the root mean square error value is decreasing, which indicates that the higher the neural network layers, the better the performance of recommendation system. Because with a certain amount of data, the larger the number of layers of neural network, the better the keyword extraction and feature analysis of video and users will be, which can further improve the performance of recommendation system. The comparison of the average absolute errors of the algorithms is shown in Figure 11.

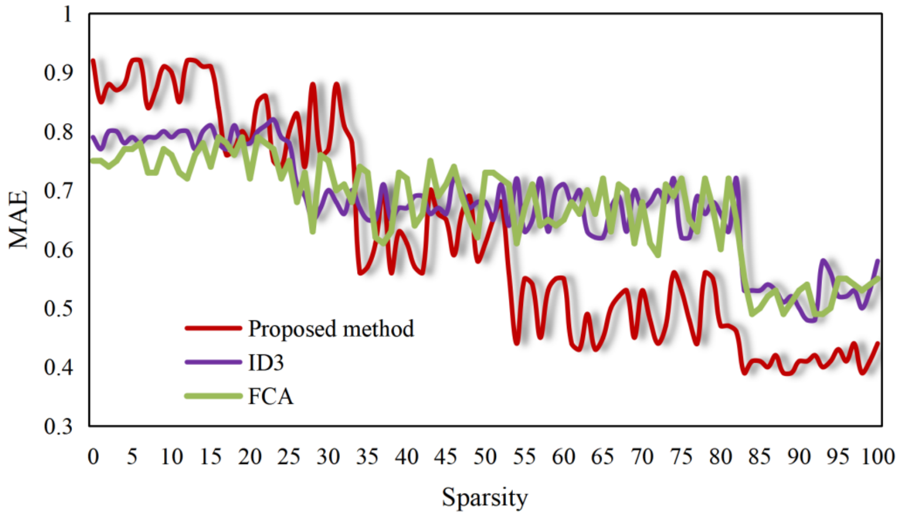

Figure 11. Comparison of Average Absolute Errors of Algorithms

It can be seen that, compared with FCA, this method has obvious advantages in the later stage of operation, and the error is reduced by 33.84%. For video sequences, the audio signal is suitable to express descriptive semantics, and the visual signal is suitable to express mandatory semantics. Only by integrating vision and hearing can a complete and rich semantic information be expressed.

For the micro-video platform, on the basis of big data and artificial intelligence, the user data is continuously collected and analyzed, and classified and marked according to the types of videos most frequently visited by each user. The platform makes accurate and personalized distribution according to the content that users are interested in, so that users can immerse themselves in the micro-video environment, increase their curiosity about the next video, and finally enhance their stickiness. The accuracy of micro-video image feature recognition is taken as the test index, and ID3 and FCA are selected as the comparison objects. The experimental results are shown in Table 2, Table 3 and Table 4.

Table 2 ID3 method

Sample size	Accuracy of micro-video image feature recognition (%)
15	93.84
30	93.39
45	92.54
60	91.88
75	91.62
90	91.41
105	90.88

Table 3 FCA

Sample size	Accuracy of micro-video image feature recognition (%)
15	97.55
30	95.81
45	95.55
60	94.38
75	93.84
90	93.35
105	93.22

Table 4 Proposed method

Sample size	Accuracy of micro-video image feature recognition (%)
15	98.99
30	98.41
45	97.99
60	97.69
75	96.95
90	96.52
105	96.47

The detection of micro-video content requires video frame segmentation, that is, random sampling. Video frame extraction refers to using key frames of video to represent the complete meaning of the whole video, aiming at different types of video coding formats, frame rates, bit rates, video resolutions and different kinds of videos. The input layer of the DNN model is a set of high-dimensional space vectors, that is, feature vectors of users and videos, which are textual descriptive information of videos and users. The output layer of the DNN model is a set of low-dimensional dense semantic vector spaces, which is the abstract description result of the low-level text features extracted by the neural network. In order to ensure the integrity and reliability of video data, before formal modeling and analysis, text description data needs to be preprocessed, which mainly includes video filtering and video fusion. Video filtering is mainly to remove some new media micro-videos with missing important text information. Video fusion is mainly to merge new media micro-videos with similar important text information.

Video content analysis mainly studies the relationship between feature description and high-level semantic concepts, and its ultimate goal is to automatically extract video semantic concepts from various features and related original video data. Figure 12 shows the scatter plot of the predicted value and the actual value of the test sample tested by ID3. Scatter chart of predicted value and actual value of test sample tested by FCA is shown in Figure 13. The scatter diagram of the predicted value and the actual value of the test sample tested by the micro-video personalized recommendation model using the micro-video description method in this article is shown in Figure 14.

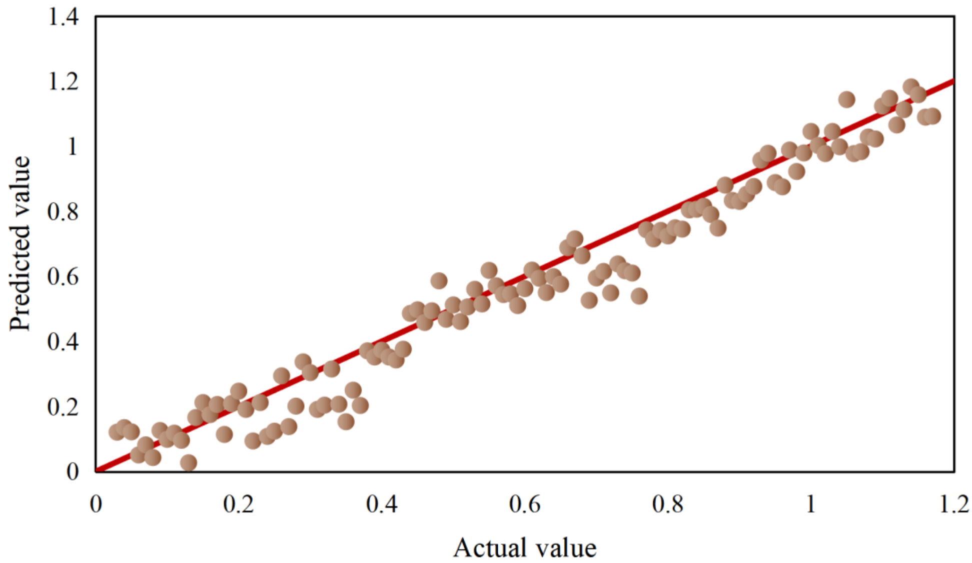

Figure 12. Scatter diagram of ID3 actual value and predicted value

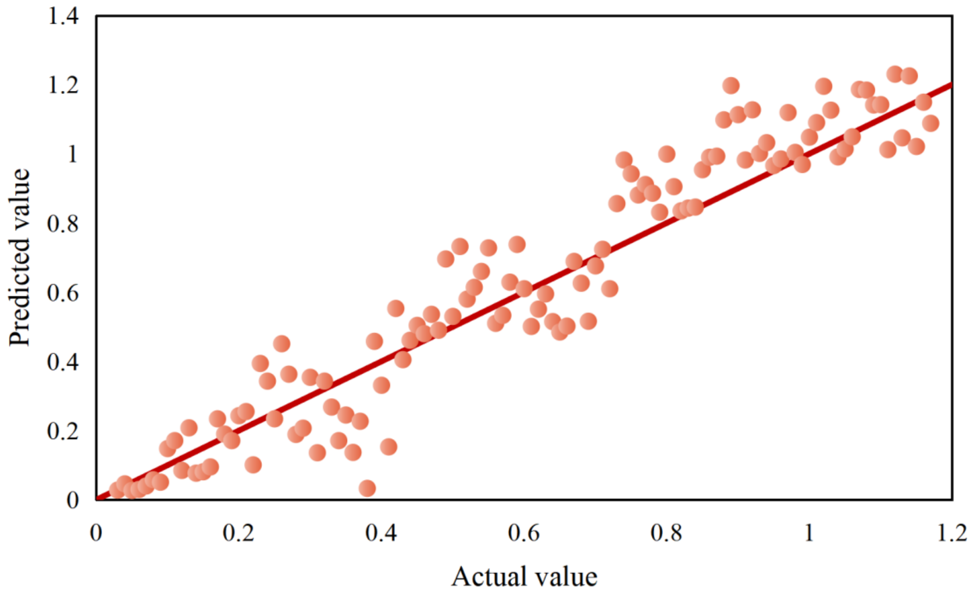

Figure 13. scatter plot of FCA actual value and predicted value

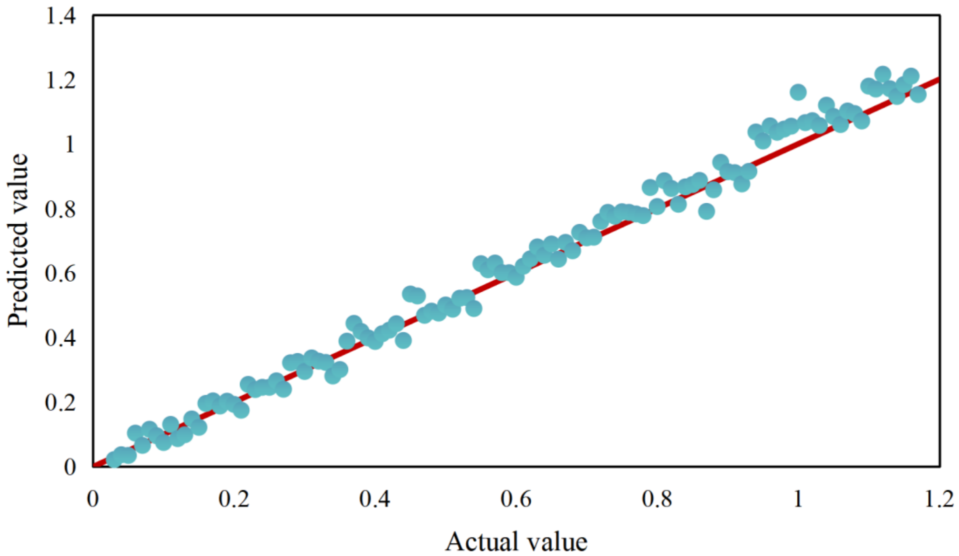

Figure 14. Scatter chart of actual value and predicted value of the micro-video description method in this article

It can be analyzed that the micro-video personalized recommendation model based on the micro-video description method in this article is better than the comparison algorithm in both accuracy and efficiency. In video processing applications, a large amount of video data in network resources have great complexity and diversity, so it is difficult to achieve ideal results if only a single mode of video processing is used. Therefore, it is necessary to use multimodal methods to process the appearance features and motion features in video data. Compare the accuracy of the algorithm in feature recognition of micro-video images, as shown in Table 5 and Figure 15.

The detection results show that the accuracy of this algorithm for micro-video image feature recognition is higher, which is 21.82% higher than that of the comparison algorithm. After completing the overall framework design of video description system, as well as the design of each sub-module and its corresponding neural network structure, it is necessary to build a framework on the system platform and write programs to realize the system. Through experimental analysis, it is found that the personalized recommendation system of new media micro-videos based on DNN model proposed in this article has a good recommendation effect, and can meet the recommendation needs of users for new media micro-videos. After the implementation, we also need to use the training data set to train the model, so that the model can continue to learn and update the network parameters, so as to obtain a mature description system that can directly input video and output correct video description. For content creators, data mining will help them understand the public's preferences more quickly, better meet the needs of the public, promote the accuracy of micro-video content production and improve the output efficiency. For the monitoring of micro-video marketing effect, data mining will help brand owners understand the effect of micro-video marketing, and finally form and optimize the marketing model to better serve brand building.

Table 5 Proposed method

Sample size	Proposed method	ID3	FCA
10	0.597	0.573	0.621
20	0.573	0.559	0.61
30	0.616	0.562	0.616
40	0.612	0.57	0.602
50	0.592	0.633	0.595
60	0.681	0.702	0.627
70	0.683	0.659	0.649
80	0.682	0.639	0.643
90	0.666	0.644	0.624
100	0.672	0.67	0.632

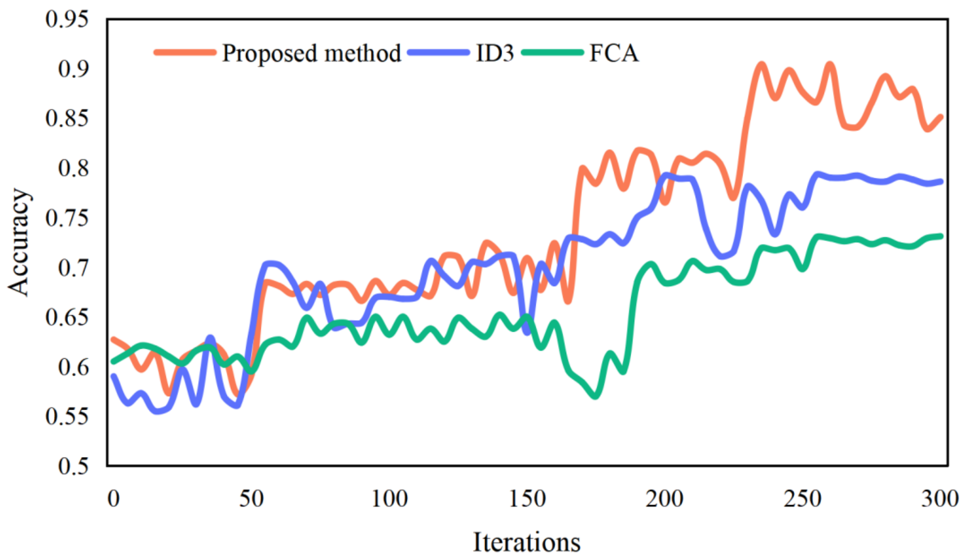

Figure 15. Comparison of feature recognition accuracy of micro-video images

5. Conclusion

Driven by big data, all walks of life clearly recognize the convenience brought by data mining, and use data mining in fast and accurate decision-making of enterprises, which not only promotes the rapid economic development, but also causes a new wave of competition. Because of its wide diversity, single content, short duration and low traffic consumption, micro-video has become the main entertainment way for people to spend their leisure time in modern life. However, the rapid growth of micro-video not only brings convenience to people, but also brings the problem of information overload. Therefore, it is of great practical engineering application value to find out possible interesting videos from a great quantity videos and recommend them to users. In this article, a three-dimensional DNN is proposed to extract the depth features of micro-videos, thus effectively representing the content of micro-videos. In addition, considering the complementarity of micro-video user data, taking tags as semantic information of video and depth features as visual information of video depth, a user behavior preference model with joint depth feature tags is constructed to better describe user behavior preferences. Finally, a personalized recommendation mechanism is designed.

The test results show that, compared with FCA, this method has obvious advantages in the later stage of operation, and the error is reduced by 33.84%. The accuracy of this algorithm for micro-video image feature recognition is higher, which is 21.82% higher than that of the comparison algorithm. The video data contains rich semantic scenes, and the high-level semantic information is closer to people's thinking and understanding, so more semantic content can be extracted to make the retrieval results more in line with users' needs.

References

[1] Zheng Y. Decision Tree Algorithm for Precision Marketing via Network Channel[J]. International journal of computer systems science & engineering, 2020(4):35.

[2] Yang, CJ, Wei YZ, et al. Research on Precision Marketing of University Teaching Materials in Big Data Era[J]. ADV SOC SCI EDUC HUM, 2017,80(10):262-266.

[3] Hongyong Y U, Qiu C, Wen J. IPTV video personalized recommendation system[J]. Telecommunications Science, 2017, 003: 102-114.

[4] Du X, Yin H, Chen L, et al. Personalized Video Recommendation Using Rich Contents from Videos[J]. 2016, 12(001): 98-129.

[5] Tripathi A, Ashwin T S, Guddeti R. EmoWare: A Context-Aware Framework for Personalized Video Recommendation Using Affective Video Sequences[J]. IEEE Access, 2019, 7:51185-51200.

[6] Li H, Murata T. Improving Personalized Recommendation in VSS Based on User's Preference[J]. Lecture Notes in Engineering & Computer Science, 2012, 2195(1): 26-59.

[7] Deng Z, Yan M, Sang J, et al. Twitter is Faster: Personalized Time-Aware Video Recommendation from Twitter to YouTube[J]. Acm Transactions on Multimedia Computing Communications & Applications, 2015, 11(2):1-23.

[8] Ma D, Luo L, Fang Y. Short video recommendations based on analytic hierarchy process and collaborative filtering algorithm[J]. Journal of Physics: Conference Series, 2021, 1774(1):012014 (7pp).

[9] Yang K, Duan Y. Personalized movie recommendation method based on ensemble learning[J]. High tech communication, 2022, 28(1):7.

[10] Tian Z, Liu Y, Sun J, et al. Exploiting Group Information for Personalized Recommendation with Graph Neural Networks[J]. ACM Transactions on Information Systems (TOIS), 2021,4(2):11-29.

[11] Chen H, Lian Q. Poverty/investment slow distribution effect analysis based on Hopfield neural network[J]. Future Generation Computer Systems, 2021, 122(11).

[12] Cao L, Zhu C. Personalized next-best action recommendation with multi-party interaction learning for automated decision-making[J]. PLoS ONE, 2022, 17.

[13] Jie L, Shambour Q, Xu Y, et al. BizSeeker: A hybrid semantic recommendation system for personalized government-to-business e-services[J]. Internet Research, 2010, 20(3):342-365.

[14] Kim Y, Jung S, Ji S, et al. IoT-based personalized NIE content recommendation system[J]. Multimedia Tools and Applications, 2018,11(3):58-79.

[15] Noorian A, Harounabadi A, Ravanmehr R. A Novel Sequence-Aware Personalized Recommendation System Based on Multidimensional Information[J]. Expert Systems with Applications, 2022, 4(1): 128-133.

[16] Song H Y, Zhang H, Xing Z H. Research on Personalized Recommendation System Based on Association Rules[J]. Journal of Physics: Conference Series, 2021, 1961(1):012027 (10pp).

[17] Bg A, Lj A, Mg C, et al. A two-step personalized location recommendation based on multi-

objective immune algorithm - ScienceDirect[J]. Information Sciences, 2019, 475:161-181.

[18] Feng R, Li T, Dong J. New media fine art education platform based on internet of things technology[J]. International journal of internet protocol technology, 2021(3):14.

[19] Sharma A, Singh A K. The Expansion of New Media Art in Indian Visual Art Practices: An Analysis[J]. ECS transactions, 2022(1):107.

[20] Liu H. Big data precision marketing and consumer behavior analysis based on fuzzy clustering and PCA model[J]. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 2021(4):40.

[21] Kenny L. Precision-Guided Marketing Frames the Future of Information Operations[J]. Signals, 2021(11):75.