

Applied Mathematics and Nonlinear Sciences

<https://www.sciendo.com>

Feasibility of the development of short videos in vocational education based on the context of deep learning

Dong Wang^{1,†}, Huaizhi Xu², Zhuo Wang¹

1. College of Art and Design, Yangzhou Polytechnic Institute, Yangzhou, Jiangsu, 225127, China.

2. College of Intelligent Manufacturing, Yangzhou Polytechnic Institute, Yangzhou, Jiangsu, 225127, China.

Submission Info

Communicated by Z. Sabir

Received May 2, 2024

Accepted August 25, 2024

Available online October 4, 2024

Abstract

The development of information technology has brought new opportunities for the development of vocational education teaching, and short video, as a product of information technology, breaks through the limitations brought by the traditional teaching mode and provides a new thinking direction for the transformation of the teaching mode. This paper proposes a recommendation system for short videos based on deep learning, where the scene and behavioral features of the target short video are first extracted to obtain the joint features. Then, user and video modeling is realized using deep learning technology. Finally, the similarity and preference are calculated by combining the joint features of the target short video and the user short video preference features. The TOP-N with the highest user preference is obtained according to the actual demand. The recommendation list is generated to complete the intelligent recommendation of short videos based on deep learning. The experimental analysis found that the recommendation system has high performance, with a maximum recommendation accuracy of 98.2%, and is capable of fully comprehending the students' demand preferences for short videos. The use of short videos in course teaching can make a great positive change in students' learning attitudes, and the short video recommendation algorithm based on deep learning allows students to more accurately find short videos suitable for their learning conditions, which further improves their learning efficiency.

Keywords: Deep learning; Short video teaching; Preference features; Intelligent recommendation; Similarity.

AMS 2010 codes: 00A35

[†]Corresponding author.

Email address: wd201553@163.com

1 Introduction

In recent years, the fountain-like rapid development of short video platforms represented by Shutterbugs, TikTok, Second Shot, etc., has won the enthusiasm of the majority of young people and even social groups of all ages. It has become the norm for young people to have multiple short video platforms on their cell phones at the same time, and brushing TikTok, watching seconds, and grabbing red packets on Shutterbugs have become part of many people's lives [1-2]. The content presented behind these short videos has become the main channel and way for audience groups, including vocational education students, to receive information, which is a new environment with both challenges and opportunities for vocational education [3-4].

"Internet + education" is a product of the integration of emerging electronic information technology and the education industry. With the diversified development of the Internet in recent years, "Internet + education" continues to innovate models, development, and evolution [5-6]. As a cutting-edge example, "short video + education" caters to the trend of fragmentation of demand, takes a short video platform as a carrier, drives resource allocation by market mechanism, and builds a large-scale ecosystem [7-9]. The document of the education ecology report of Shuttle released by Shuttle Big Data Research Institute shows that the cumulative generation volume of education short videos is more than 200 million, the average daily play volume is more than 2.2 billion, and the average daily like volume is more than 60 million [10-12]. On the basis of making full use of the high penetration rate of the short video platform, "short video + education" reaches more users, drives a wider range of groups to participate in the interaction, and shows many characteristics in line with the platform model [13-14].

Literature [15] points out that we are entering a fragmented information age. Short videos have become the focus of people's attention. The low-cost feature of short videos meets the needs of people in a fast-paced society, and the wide popularity of short video activities related to education is also reflected in short videos. Literature [16] argues that the quality of education embodied in audiovisual media is unsatisfactory compared to distance education. It is pointed out that the potential of video in education has not been fully utilized as far as the current situation is concerned. Literature [17] found through research that video lectures for vocational and technical education and training can be categorized into six different designs, with dubbed slides being the most commonly used one. Three of these six designs were found in video lectures. Literature [18] stated that the advancement of technology has drawn the attention of education to hypervideo. Hyper-video presents a great opportunity for education because it is interactive and can support learning more effectively than traditional courses. Literature [19] aims to improve student learning outcomes by implementing a video-supported learning model. The results of the study showed that students' performance improved compared to the previous one. Literature [20] specified that the current theoretical and practical learning in vocational education has become flexible and is no longer limited to physical space. It was concluded from the study that it is very important to understand the position of teachers and students in vocational education, and only then can online education be effective. Literature [21] emphasized that short videos are very popular among students, but the immersive nature of short video platforms tends to be addictive, which requires teachers and parents to be able to pay more attention to the overuse of short videos by students. Literature [22] illustrates that multimedia technology enables teachers to realize simulated real-life scenarios, which is of great value to vocational education, which requires the practice and application of skills.

In this paper, a deep learning-based algorithm for short video content recommendation is proposed to extract the features of the target short video and match them with user preference features. Then, the inter-frame matching method is introduced to preprocess the short video to improve the efficiency of subsequent feature extraction. Deep learning uses a graph convolutional neural network video

recommendation model to predict the user's preference for the target video. The directed edges from the previous short video to the next short video are used as the propagation path of vectors, and the user's long-term and short-term preferences are decoupled and modeled separately. After obtaining the user preferences, we calculate the comprehensive vector of user preferences and the feature vector of the target short video, select the short video with the highest preference degree to form the recommendation list, and complete the intelligent recommendation of short videos. The experimental results show that the method presented in this paper has a higher recommendation accuracy, and the combination of short video lectures in vocational education can have a better teaching effect.

2 Development of short videos in vocational education

2.1 Characteristics of Short Video Teaching

2.1.1 The Innovative Nature of Short Video Instruction

Innovation is the driving force behind the sustainable development of all things, and teaching is no exception. Both young and old users find short videos, with their unique and innovative nature, very popular. Short video teaching is also of great interest to the "user" as a student. Through short videos and voice narration, students can experience multiple auditory and visual experiences, deepening their understanding and memorization. Short videos allow users to learn anywhere, anytime, for free. Students can start the learning mode with only a cell phone. This not only allows students to make full use of fractions of time but also greatly increases their motivation to learn. In the case of students do not understand once, you can listen again and again, watch again and again, through continuous learning to deepen the impression of the teaching video for the teaching of doubtful places, but also through the video below the message to put forward the point of view. Often, a valid point of view will lead to many students discussing with each other. In a one-minute video for a single point of knowledge to explain, through manual operation and short language explanation, a point of knowledge is quickly explained to the students. If the students still have doubts, they can click on the comments section to view the discussion of others. The comments section not only gives an understanding of the video, but even professionals will be the expansion of the point of knowledge so that readers have a deeper understanding.

2.1.2 Interactivity of short video instruction

Communication is the most significant ability that distinguishes humans from lower animals, and interaction is a way to achieve it. The quality of teaching and learning can be improved by good interaction between teachers and students, and the strong interactivity of short videos is one of its most advantageous features. Through the interaction between teachers and students, students can understand and improve their performance, thus realizing the improvement of their learning ability. In short video teaching, there are no geographical boundaries, no age boundaries. The same video teaching has multi-age students, students from different regions, different races, and even students from different countries, who are formed into a learning community due to the same needs and desire for knowledge. They can communicate and explore just like traditional teaching. It is not only the interaction between learners and video producers but also the interaction between learners and video producers. Suppose learners have any opinions about the teaching of the video. In that case, they can let the video producer know the first time through the comment section so that the video producer can improve his teaching video through the learners' opinions or suggestions and be closer to the learners' needs.

Most importantly, the learners' messages in the comment section can be seen by the teacher the first time, and the teacher can answer the questions the first time. And if the students still don't understand something, they can continue to ask questions, and the teacher can answer them again in a short time. Such teacher-student interaction is two-way, can provide efficient feedback, save both sides a lot of time, and can help learners achieve higher learning efficiency in a short period. This interactive and personalized learning approach enables each learner to better develop themselves and also promotes interpersonal relationships of mutual respect and solidarity, which helps to cultivate a sense of teamwork and collaborative learning among learners.

2.1.3 The Interestingness of Short Video Instruction

The short videos are able to attract the students' attention based on creative short videos and present the most interesting videos in the shortest possible time. Interesting teaching features from photography, painting, English, language, office software, science, and technology are presented differently in the video. Through short videos, it integrates four modes of presentation: graphic, text, sound, and visual, allowing learners to combine visual and auditory senses in an all-around way. Short teaching videos like situational, visual, and interesting ones are not only conducive to students' personalized learning but also serve the purpose of stimulating their enthusiasm and interest.

2.2 Rationale

2.2.1 Microlearning Theory

Microlearning is a method of learning that relies on short videos. Microlearning is a new type of learning that utilizes micro-content and micro-media through a new media system. Microlearning is a kind of informal learning for smaller learning units anytime, anywhere, with short videos as the carrier of learning activities, which is characterized by: short learning time, less learning content, and various forms of learning. Micro-learning mainly reflects its "micro" characteristics, from the content point of view, for the content of a small unit module. In terms of time, it takes a short amount of time. From the platform perspective, micro-learning for micro-sustainable devices is intended to realize the fragmentation of learning across time and space. Micro-learning caters to the requirements of people's knowledge economy in the modern information network era, and micro-courses are precisely the teaching resources under the guidance of micro-learning theory, which are presented in the form of short videos and last about 10 minutes, which is in line with the characteristics of the long-lasting poor attention of students in vocational schools. Its content focuses on a specific knowledge point or a specific link, is short and clear, and is famous for its miniature content. At the same time, microclasses are connected to modern network information technology, which presents teaching content in the form of short videos to facilitate informal learning, fragmented learning, and independent learning.

2.2.2 Theories of Educational Communication

From the point of view of communication science, education is known as an information dissemination activity in which educators select the appropriate information content according to the requirements of the curriculum and, with the help of an effective medium, transfer knowledge and information to the specific educational target of the process of dissemination is a complex process of information transfer and exchange, and a variety of modes of dissemination have been developed in the course of its evolution. Educational communication has a clear purpose, specificity, and diversity of media communication. According to the theory of educational communication and its

characteristics, the teaching process should be designed with students as the main body of the course content to enhance the relevance and clarity of teaching and to choose the appropriate communication media to deliver the learning content with the principle of minimum cost and optimal learning effect. In the teaching process based on short videos, teachers design the corresponding teaching design and teaching links according to the teaching objectives and syllabus, balance online and offline teaching resources, and optimize the combination of microteaching resources and offline teaching through the integration of teaching resources, teaching links, media channels, and other elements, so as to achieve the most efficient teaching at the lowest cost.

2.3 Short Video Intelligent Recommendation Algorithm

2.3.1 Short video personalized recommendation system design

The overall process of the new media short video personalized recommendation system based on the deep neural network model is shown in Figure 1 [23], from which the basic structure of the system can be seen.

- 1) Acquire the user's short video history operation data and the text description data of the short video through the Hadoop big data platform and complete the preprocessing of the two parts of data [24].
- 2) Adopt a traditional text mining algorithm to data mine the text description data of short videos to obtain the key features of short video content, combined with the user's playback records of short videos, to realize user and video modeling.
- 3) Taking the user and short video modeling situation as the training samples of the deep neural network, the short video and user feature vectors are used as the input data of the deep neural network, and the deep neural network is trained in a supervised manner. Make full use of the feature extraction advantage of deep learning technology to grasp the semantic features of short videos and users, mine the intrinsic relationship between users' interests and preferences and short videos, and construct a personalized recommendation system.

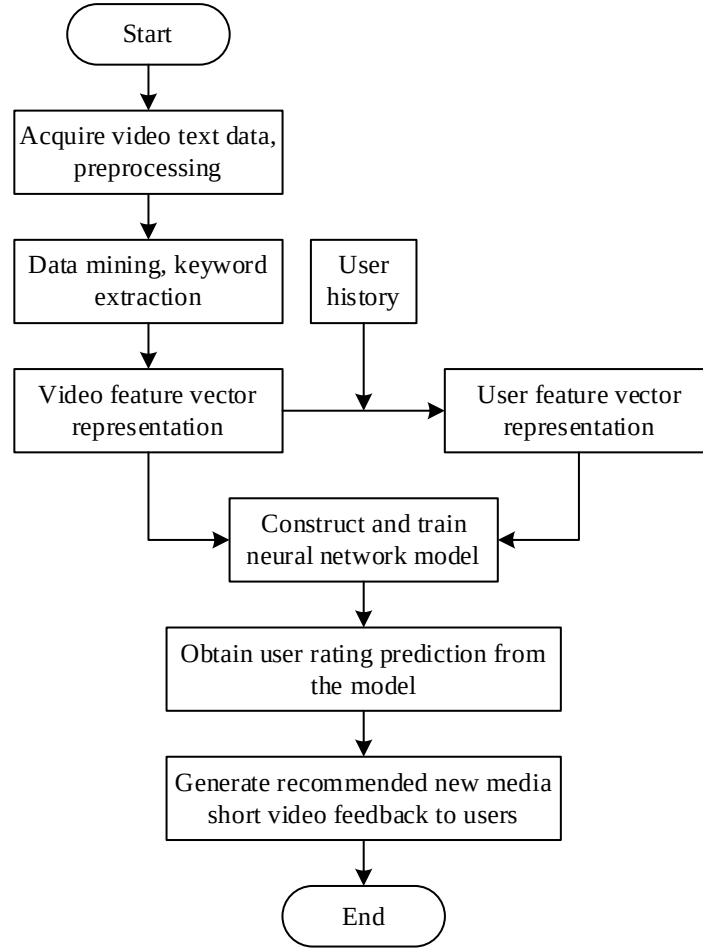


Figure 1. Short frequency recommendation system process

2.3.2 Deep Learning-Based Intelligent Recommendation Algorithm for Short Video

Before using deep learning algorithms to recommend short videos for users, it is necessary to extract the features of the target short video and match them with user preferences to achieve more accurate results. Deep learning-based short video intelligent recommendation algorithm technology introduces the inter-frame matching method to preprocess the short video and improve the efficiency of subsequent feature extraction [25].

The image feature point is extracted by the FAST corner point detection method with a sampling threshold $N=9$. $G(p)$ is used to represent the gray value of the pixel point in the center of the feature point, $G(q)$ is used to represent the gray value of the pixel point around the point, δ is the threshold set in advance, and $F(G(p), G(q))$ is used to represent the result of comparing the gray values, and the judgment condition is obtained as shown below:

$$\begin{cases} N = \sum_{q \in \text{cirl}(p)} F(G(p), G(q)) \\ F(G(p), G(q)) = \begin{cases} 0, & |G(q) - G(p)| \leq \delta \\ 1, & |G(q) - G(p)| > \delta \end{cases} \end{cases} \quad (1)$$

If there are more than 9 points in the pixel points surrounding the target pixel point whose difference in gray value from the target point is within the threshold range, the target point is determined to be a feature point.

BRIEF is used to describe the feature points extracted by the FAST algorithm, using (a_i, b_i) to denote the i th pixel pair and $K(a_i, b_i)$ to denote the result of comparison of pixel values a_i and b_i . The BRIEF eigenvalue $f_B(P)$ of feature point P is calculated using the formula shown below:

$$\begin{cases} f_B(P) = \sum_{i=1}^{138} 2^{i-1} K(a_i, b_i) \\ K(a_i, b_i) = \begin{cases} 0, a_i \geq b_i \\ 1, a_i < b_i \end{cases} \end{cases} \quad (2)$$

Let A_i and B_j denote the i th and j th feature points in the feature point set A of the c th point and the feature point set B of the $c-1$ th frame image, respectively, and $f_n(A_i)$ and $f_n(B_j)$ denote the BRIEF values of A_i and B_j , respectively, to compute the fuzzy similarity $\varphi(A_i, B_j)$ of A_i and B_j , and the computational formula is shown below:

$$\varphi(A_i, B_j) = \frac{\text{Sum}(f_B(A_i) \cdot f_B(B_j))}{128} \quad (3)$$

Where Sum is the binary conversion and summing operation by bit; $\cdot$ is the same-or operation, $\varphi(A_i, B_j)$ is proportional to the matching degree of A_i and B_j points, and m feature points $A_1, A_2, \dots, A_w$ and $B_1, B_2, \dots, B_w$ in A and B are established as the similarity set $S = \{ \langle A_i, B_j, \varphi(A_i, B_j) \rangle \}$.

Let the total number of extracted feature points in the images of frame c and frame $c-1$ be X and Y , respectively, $X \geq Y$, obtain a set of feature points $u_1, u_2, \dots, u_n$ with one and only one feature point in the image of frame c , construct a point set U containing n feature points, obtain a set of feature points $v_1, v_2, \dots, v_n$ with the same number of feature points as that in U in the image of frame $c-1$, construct a point set V , and in U and V , the fuzzy matching degree of the two likenesses at the same position is greater than 0.5, U and V are the matching schemes for the images of frame c and frame $c-1$, and the total matching degree $\bar{S}$ for U and V is calculated as follows:

$$\bar{S} = \sum_{i=1}^n \varphi(A_i, B_j) \quad (4)$$

Through feature matching, the feature points that cannot be matched in the two images are obtained, and the corresponding screen is cut, and the target screen is gridded and processed to be divided into a rectangle with a side length of L . The width and height of the image are set to be W and H , respectively, and the formula for determining L is shown below:

$$L = \left\lfloor \sqrt[4]{W \times H} \right\rfloor \times 2 + 1 \quad (5)$$

The image will be divided into a number of $L \times L$ grid, after which the content of the changes occurring in a close comparison set a threshold ζ , when the proportion of feature points in the grid not matched successfully ρ is higher than ζ , it is considered that the content of the grid in the content of the larger changes need to be carried out to further feature recognition, set the target grid there are a total of k feature points, of which the matching of the i th feature point is γ_i , matching successful feature points $\gamma_i = 0$, unsuccessful feature points $\gamma_i = 1$ and ρ are calculated as follows:

$$\rho = \frac{\sum_{i=1}^k \gamma_i}{k} \quad (6)$$

After obtaining all the regions with large changes by the above method, they are grouped and inputted into the feature recognition extraction algorithm for a new round of feature extraction.

Feature extraction is divided into two parts: scene feature extraction and behavioral feature extraction. Let there exist a short video collection $S = \{s_1, s_2, \dots, s_m\}$, which contains a total of m short videos, corresponding to a scene classification collection $T = \{t_1, t_2, \dots, t_n\}$, which contains a total of n scenes, and a deep fusion network is used to extract the global and local features in the scenes and to fuse the extracted results. The scene feature extraction is used to obtain the set of probability values of the i th short video s_i of S in the total scenes $\eta_{s_i}^T$, with $\eta_{s_i}^{t_j}$ denoting the probability that s_i corresponds to the j th scene t_j , to obtain $\eta_{s_i}^T$ as follows:

$$\eta_{s_i}^T = \{\eta_{s_i}^{t_j}\} (i \in (1, m), j \in (1, n)) \quad (7)$$

The behavioral features of the short video are extracted by the variable convolutional neural network to obtain the behavioral classification set $W = \{w_1, w_2, \dots, w_z\}$, which contains a total of z behaviors, similar to the scene feature extraction, using $\eta_{s_i}^{w_j}$ to denote the probability that s_i corresponds to the j th scene w_j , to obtain the set of probability values of s_i in all the behaviors $\eta_{s_i}^W$ is shown below:

$$\eta_{s_i}^W = \{\eta_{s_i}^{w_j}\} (i \in (1, m), j \in (1, z)) \quad (8)$$

Based on the above calculation, the joint features of the target short video s_i are obtained as shown below:

$$\eta_{s_i} = \left(\eta_{s_i}^T \right)^T \eta_{s_i}^W \quad (9)$$

A graph convolutional neural network video recommendation model in deep learning is used to predict the user's preference for the target video. The switching behavior b_{ij} of a user switching from

short video i to short video j constitutes another type of edge set B , and the weight θ of this edge is determined by the number of switching times of the user in the dataset, if $\theta = 0$, no switching behavior is generated, i.e., the edge does not exist.

Design the embedding layer based on unique thermal encoding by combining the user's identity, construct the embedding layer matrix C for m user, and construct the embedding layer matrix E for n short videos, then the dimension of C is $m \times D$, and the dimension of E is $n \times E$, where D denotes the adjustable low-dimensional spatial dimensions, and set the unique thermal encoding ϕ_y and ϕ_x for the user and the short videos, respectively, combining the matrices C and E , the low-dimensional representation vectors of the user and the short videos can be obtained c_y and e_x are shown below:

$$\begin{cases} c_y = C^T \phi_y \\ e_x = E^T \phi_x \end{cases} \quad (10)$$

In the short video intelligent recommendation algorithm, it is necessary to model the user's long-term and short-term preferences separately, and the user's long-term preferences are modeled by the node representation vectors of all short videos that the user has watched, with B denoting the set of all short videos that the user y has interacted with, Y_x denoting the set of all users that have watched the short video x , $\alpha_l^{(k+1)}$ and $\beta_l^{(k+1)}$ denoting the network parameters and the bias parameter when propagating from layer 1 to layer $k+1$, respectively. $\mathcal{G}$ denotes the nonlinear activation function, $aggregate(\cdot)$ denotes the vector aggregation operation, $g_{y-l}^{(k)}$ and $g_{y-l}^{(k+1)}$ denote the user y layer k and layer $k+1$ representation vectors, respectively, and $h_{x-l}^{(k)}$ and $h_{x-l}^{(k+1)}$ denote the video x layer k and layer $k+1$ representation vectors, respectively, and the propagation formulas of $g_{y-l}^{(k+1)}$ and $h_{x-l}^{(k+1)}$ are obtained as shown below:

$$\begin{cases} g_{y-l}^{(k+1)} = \mathcal{G} \left(\alpha_l^{(k+1)} \left(g_{y-l}^{(k)} + aggregate(h_{x-l}^{(k)} | x \in Y_x) \right) \right. \\ \quad \left. + \beta_l^{(k+1)} \right) h_{x-l}^{(k+1)} \\ = \mathcal{G} \left(\alpha_l^{(k+1)} \left(h_{y-l}^{(k)} + aggregate(g_{x-l}^{(k)} | y \in B_y) \right) \right. \\ \quad \left. + \beta_l^{(k+1)} \right) \end{cases} \quad (11)$$

The user's short-term preference is modeled by the user's last-moment interaction behavior representation vector, i.e., the directed edge from the previous short video to the next short video is used as the propagation path of the vector. Thus, the long-term preference and the short-term preference can be decoupled and modeled to avoid the mutual influence of the two.

Using the set R_x to denote all videos with x directed edges, $\alpha_s^{(k+1)}$ and $\beta_s^{(k+1)}$ to denote the network parameters and bias parameters when propagating from layer 1 to layer $k+1$, respectively, $aggregate(\cdot)$ to denote the vector aggregation operation combining the weights, $h_{x-s}^{(k)}$ and $h_{x-s}^{(k+1)}$ to denote the video x k th and $k+1$ th layer characterization vectors, and $h_{i-s}^{(k)}$ to denote the next short video i k th layer characterization vectors, the propagation formula for obtaining $h_{x-s}^{(k+1)}$ is shown below:

$$h_{x-l}^{(k+1)} = \mathcal{G}\left(\alpha_s^{(k+1)}\left(h_{y-s}^{(k)} + \text{aggregate}\left(h_{i-s}^{(k)} \mid i \in R_x\right)\right) + \beta_s^{(k+1)}\right) \quad (12)$$

According to the long and short-term preference vector propagation to obtain the two components of the short video, the two components are merged and processed to obtain the comprehensive vector of user preferences $h^{(k+1)}$ is shown below:

$$h^{(k+1)} = h_{x-1}^{(k+1)} + h_{x-s}^{(k+1)} \quad (13)$$

After acquiring the user preferences, the cosine similarity ω between the user preference composite vector $h^{(k+1)}$ and the target short video feature vector η_{s_i} is calculated as:

$$\omega = \frac{\sum_{k=0, i=1}^N (h^{(k+1)} \times \eta_{s_i})}{\sqrt{\sum_{k=0}^N (h^{(k+1)})^2} \sqrt{\sum_{i=1}^N (\eta_{s_s})^2}} \quad (14)$$

Where N is the vector dimension [26].

Combining the user-preferred short videos with ω to calculate the user's preference f_{pq} for the target short videos, let the similarity of short videos i and j be ω_{ij} , the user's y preference for short videos of category v be $T(y, v)$, the set of user y preferred short videos be B_y , and the set of L short videos with the highest degree of similarity to short video i be $K(i, L)$, to get f_{pq} as follows:

$$f_{pq} = T(y, v) \sum_{i \in B_y} \omega_{ij} \quad (15)$$

According to the actual demand to, obtain the TOP-N with the highest user preference, generate the recommendation list, and complete the short video intelligent recommendation based on deep learning.

3 Performance and effectiveness of the system

3.1 System implementation and testing

In this paper, deep learning-based video recognition and intelligent recommendation algorithms are used. This system uses a cloud preprocessing module to provide static file distribution and load-balancing services. The network interface module serves as a channel for transmitting data between the client and the server. The video display module is used to decode and render video, thus further enabling video playback and push display. This paper utilizes the system hardware and software platform to conduct key frame sampling test experiments and user preference threshold setting experiments.

3.1.1 Keyframe Sampling Test Experiments

To study the effect of sampling step size and similarity threshold on the key frame extraction algorithm, a short video with a duration of 100s is used for simulation testing. In this paper, we tested the number of keyframes extracted under different combinations of sampling step size and similarity threshold, and the results are shown in Fig. 2. The group of videos contains 20 keyframes, as shown by manual testing conducted in multiple groups. Comparing the results in Fig. 2, it can be seen that the number of key frames extracted when the sampling step length is 2-4s and the similarity threshold is 0.6-0.8 is in line with the real situation when increasing the similarity threshold or decreasing the sampling step length, the number of keyframes proposed will increase. When the similarity threshold is decreased or the sampling step is increased, the number of proposed keyframes will decrease. The sampling step is set to 3s, and the similarity threshold is set at 0.7, given the amount of computation and the number of keyframes extracted.

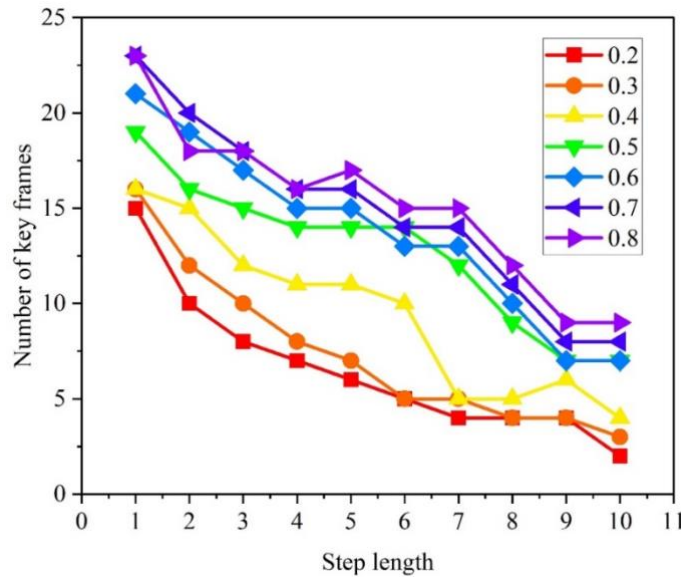


Figure 2. Number of extracted keyframes

3.1.2 User Preference Threshold Settings

The algorithm generates a user preference table that records the user's preference scores for different short videos based on their interaction behavior with the videos in this paper. The user preference score for each category of short videos is set as an integer between 0 and 100. During the system's operation, this paper utilizes histograms to determine the distribution of user preference scores and sets up a set of experiments that involve 200 different short videos and 200 users for simulation testing. The histogram distribution of user preference scores is shown in Figure 3. If the preference threshold is too high, only a small number of short videos can be pushed to users. And when the threshold is set too low, the requirement of reducing object categories cannot be achieved. The preference threshold for this paper is determined by weighing the user's preference and the number of short videos, and it is set to 60.

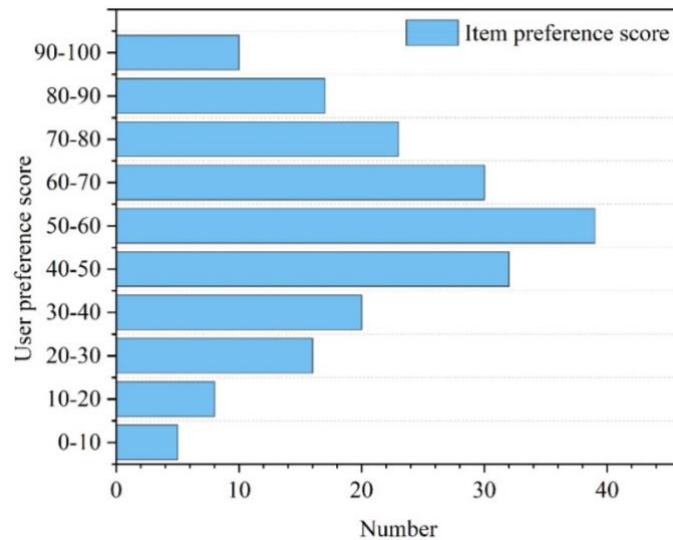


Figure 3. Distribution histogram of user preference scores

3.1.3 Recommendation Algorithm Testing Experiments

Based on the above analysis and parameter settings, this paper uses 20 short videos of 350s duration to test the effectiveness of the recommendation algorithm proposed in the paper. The 20 short videos include 15 different object categories, and the algorithm in this paper can match these objects sequentially. The 20 videos are categorized into 5 groups of videos with different durations, and Table 1 shows the recommendation results for these short videos using different methods. Table 1 shows that this paper's algorithm can achieve a maximum recommendation accuracy of 98.2%. Compared to the recommendation algorithm based on causal inference, the recommendation accuracy has been improved. It shows that the short video recommendation algorithm based on deep learning can more accurately recommend more suitable short video teaching courses for each student according to their different personalities, learning abilities, and learning states.

Table 1. Recommendation accuracy of videos with different lengths

Duration/s	Effective frame number	Based on causal inference	Accuracy / %	The algorithm is recommended for frame number	The accuracy of the algorithm / %
0-70	13, 11, 8, 12	11, 8, 7, 11	84.1	12, 10, 7, 10	88.6
70-140	28, 29, 20, 36	27, 28, 17, 34	93.8	27, 28, 18, 35	95.6
140-210	45, 56, 59, 60	42, 52, 57, 57	94.5	44, 55, 58, 58	97.7
210-280	65, 69, 70, 68	64, 65, 68, 62	95.2	64, 68, 68, 67	98.2
280-350	78, 76, 82, 93	75, 72, 77, 91	95.7	77, 75, 80, 91	98.2

3.2 Analysis of practical effects

3.2.1 Analysis of Questionnaires

In order to understand the effect of the application of short videos in the teaching of vocational school courses in the context of deep learning and the impact on student's learning outcomes, this survey was conducted by using a questionnaire star to collect the actual learning situation of the students in the classes where the lectures were given. The students of a secondary school's 22nd-grade E-commerce 1 class. The questionnaire set includes multiple-choice questions centered on the students'

learning attitude, learning habits, learning efficiency, and learning ability in four aspects of the content, with specific questions set in Table 2.

Table 2. Questionnaire Content Settings

Dimension	Issue number
Learning attitude	1-4
Learning habits	5-9
Learning efficiency	10-13
Learning ability	14-18

3.2.2 Analysis of Learning Attitudes

Table 3 shows the statistics of the results of the questions related to students' learning attitudes. According to Table 3, 89.87% of the students believed that short video teaching in the context of a deep learning classroom increased their motivation to learn logistics, and 86.25% and 90.88% of the students were able to complete the learning requirements set by the teacher and actively participate in the group discussion under the implementation of this teaching mode. It can be seen that short video teaching has enhanced students' motivation to learn, and their learning attitudes have greatly transformed compared to before.

Table 3. Student Learning Attitude Analysis

Topic	Very fit/%	Coincidence/%	Indeterminate/%	Discrepancy/%	Very inconsistent/%
The short visual frequency teaching has improved my motivation to learn	54.1	35.77	10.13	0	0
Short visual frequency teaching can complete the teacher's study requirement	31.56	54.69	10.96	2.09	0.7
Short visual teaching makes me feel happy about learning	48.69	43.56	7.75	0	0
I will actively participate in the group discussion	36.58	54.3	6.09	3.03	0

3.2.3 Analysis of study habits

Table 4 shows the statistics of the responses to the questions related to students' study habits. According to Table 4, students' study habits have improved greatly, and many students pay more and more attention to pre-study before class and review after class, of which 89.99% and 92.47% of the students said that they would take the initiative to do pre-study before class and complete the homework after class, and 89.96% of the students would deal with the study problems they encountered in the class in a timely manner. Short video teaching provides students with a means of communication that is effective and increases the frequency of their interactions with their fellow teachers. By analyzing the data results, it has been found that short video teaching in the context of a deep learning classroom is conducive to the formation of good learning habits and the enhancement of students' learning confidence.

Table 4. Student Study Habit Analysis

Topic	Very fit/%	Coincidence/%	Indeterminate/%	Discrepancy/%	Very inconsistent/%
I will actively prepare for the content.	30.53	59.46	7.89	2.12	0
I will actively solve the problem of learning.	30.52	59.44	7.91	2.13	0
I will do my homework.	32.69	59.78	6.13	1.4	0
Short visual frequency teaching improved the interaction frequency of my classmates and my classmates.	42.78	51.16	6.06	0	0
Short-frequency teaching provides me with effective communication channels.	36.49	53.09	6.15	1.79	2.57

3.2.4 Learning Efficiency Analysis

Table 5 shows the statistics of the results of the answers to the questions related to students' learning efficiency. According to Table 5, 89.37% of the students think that their learning efficiency has been significantly improved, and 93.25% of the students think that their attention in the classroom has been improved. The short video teaching in the context of the deep learning classroom enabled 86.26% of the students to have a deeper understanding of the knowledge taught, and under the guidance of the teacher, 91.26% of the students had a clearer understanding of the teaching content and a clearer teaching schedule. In the survey, there are still a small number of students who do not have a deep enough understanding of the knowledge, which indicates that the teachers still need to strengthen the creation of video content and strive to enable students to better grasp the knowledge.

Table 5. Student Learning Efficiency Analysis

Topic	Very fit/%	Coincidence/%	Indeterminate/%	Discrepancy/%	Very inconsistent/%
The short visual frequency teaching makes my learning efficiency significantly improved	42.78	46.59	8.12	2.51	0
Short-frequency teaching improves my classroom attention	38.12	55.13	6.75	0	0
Short-frequency teaching makes my understanding of knowledge deeper	34.13	52.13	8.46	5.28	0
The short visual frequency teaching makes me clear about learning content and learning progress	42.11	49.15	8.74	0	0

3.2.5 Learning Capacity Analysis

The short video teaching has greatly enhanced the students' learning ability in the context of a deep-learning classroom. Table 6 shows the statistics of the responses to the questions related to the student's learning ability. 85.59% of the students said that their knowledge level has improved and been broadened. 85.54% of the students felt that their practical abilities were improved. The ability to analyze and solve problems was effectively improved by this teaching method, according to 84.67% of the students. 86.58% of the students believed that their teamwork ability had been improved. Improved their teamwork skills. Additionally, 94.51% of the students reported improvement in their communication and expression skills, which indicates that the short video feedback provided after class had a significant impact. A comprehensive analysis of the above shows that short video teaching

in the context of a deep learning classroom is conducive to improving students' learning ability and personal comprehensive level.

Table 6. Student Learning Ability Analysis

Topic	Very fit/%	Coincidence/%	Indeterminate/%	Discrepancy/%	Very inconsistent/%
The myopia has promoted my knowledge level	34.56	51.03	12.11	2.3	0
The short frequency has improved my practice ability	38.76	46.78	8.17	6.29	0
The short time frequency allows me to analyze the ability to resolve the problem	30.55	54.12	10.23	1.29	3.81
The short time band has promoted my team collaboration ability	37.89	48.69	7.16	6.26	0
Myopia has improved my ability to communicate and express	42.38	52.13	5.49	0	0

4 Conclusion

As an emerging media form in the era of mobile Internet, short videos bring new ideas and new opportunities for vocational education with their unique dissemination method and broad coverage. Based on this, this paper proposes a short video-based intelligent recommendation algorithm based on deep learning.

- 1) The algorithm in this paper can achieve a maximum recommendation accuracy of 98.2%, which is improved compared with the recommendation algorithm based on causal inference, which can only achieve a maximum recommendation accuracy of 95.7%.
- 2) From the teaching practice, it is found that 90.88% of the students can actively participate in the group discussion, 92.47% of the students said that they would take the initiative to complete the homework after the class, 94.51% of the students said that their communication and expression skills have been improved, and 91.26% of the students have a clearer understanding of the teaching content. The use of short videos in course teaching can make a great change in students' learning attitudes. The desire for knowledge is maximized, making students more actively involved in classroom learning and greatly improving students' learning motivation and learning effect.

References

- [1] Jiang, C. (2023). Film and Television Communication in the New Media Era: Content Innovation in the Age of Short Videos. *Highlights in Art and Design*, 4(3), 164-167.
- [2] Sun, H. (2021). TikTok: The era of short videos (Doctoral dissertation, Master's thesis). San Francisco State University, San Francisco, CA. Retrieved from <https://scholarworks.calstate.edu/downloads/44558m19f>.
- [3] Lin, J., Swart, J., & Zeng, G. (2023). Theorising TikTok cultures: Neuro-images in the era of short videos. *Media, Culture & Society*, 45(8), 1550-1567.
- [4] Jiang, J., Hong, Y., Li, W., & Li, D. (2022). A study on the impact of official promotion short videos on tourists' destination decision-making in the post-epidemic era. *Frontiers in Psychology*, 13, 1015869.
- [5] Wang, X. (2024). The application of new media communication technology in vocational education teaching under the background of big data. *Applied Mathematics and Nonlinear Sciences*.

- [6] Schwendimann, B. A., De Wever, B., Hämäläinen, R., & Cattaneo, A. A. (2018). The state-of-the-art of collaborative technologies for initial vocational education: A systematic literature review. *International Journal for Research in Vocational Education and Training*, 5(1), 19-41.
- [7] Han, W. (2022, May). Research on short video marketing model in the New Media Era. In 2022 International Conference on Comprehensive Art and Cultural Communication (CACC 2022) (pp. 195-198). Atlantis Press.
- [8] Ensi, T. A. N. G. (2020). Analysis of mass communication of short video clips in era of convergence media. *Sino-US English Teaching*, 17(3), 93-98.
- [9] Banica, L., Burtescu, E., & Enescu, F. (2017). The impact of internet-of-things in higher education. *Scientific Bulletin-Economic Sciences*, 16(1), 53-59.
- [10] Finkelhor, D., Walsh, K., Jones, L., Mitchell, K., & Collier, A. (2021). Youth internet safety education: Aligning programs with the evidence base. *Trauma, violence, & abuse*, 22(5), 1233-1247.
- [11] Abbas, N., & Ojha, U. (2019). Not just a medical student: delivering medical education through a short video series on social media. *JMIR Medical Education*, 5(1), e11971.
- [12] Si, X. (2020). Short video—A new approach to language international education. *Communication trends in the post-literacy era: polylingualism, multimodality and multiculturalism as prerequisites for new creativity.—Ekaterinburg, 2020*, 121-129.
- [13] Bulca, Y., Bilgin, E., Altay, F., & Demirhan, G. (2022). Effects of a short video physical activity program on physical fitness among physical education students. *Perceptual and Motor Skills*, 129(3), 932-945.
- [14] Fyfield, M., Henderson, M., Heinrich, E., & Redmond, P. (2019). Videos in higher education: Making the most of a good thing. *Australasian Journal of Educational Technology*, 35(5), 1-7.
- [15] Zhang, T. (2020, December). A brief study on short video platform and education. In 2nd International Conference on Literature, Art and Human Development (ICLAHD 2020) (pp. 543-547). Atlantis Press.
- [16] Laaser, W., & Toloza, E. A. (2017). The changing role of the educational video in higher distance education. *The International Review of Research in Open and Distributed Learning*, 18(2).
- [17] Rahim, M. I., & Shamsudin, S. (2019). Categorisation of video lecture designs in MOOC for technical and vocational education and training educators. *Journal of Technical Education and Training*, 11(4).
- [18] Cattaneo, A. A., Van der Meij, H., & Sauli, F. (2018). An empirical test of three instructional scenarios for hypervideo use in a vocational education lesson. *Computers in the Schools*, 35(4), 249-267.
- [19] Ariyanto, S. R., Hidayatullah, R. S., & Nurtanto, M. (2020). Practical Learning Innovation: Real Condition Video-Based Direct Instruction Model in Vocational Education. *Online Submission*, 6(1), 79-91.
- [20] Roslin, A. R., Rahmatullah, B., Zain, N. Z. M., Purnama, S., & Yas, Q. M. (2022). Online learning for vocational education: Uncovering emerging themes on perceptions and experiences. *Journal of Vocational Education Studies*, 5(1), 1-15.
- [21] Ye, J. H., Wu, Y. F., Nong, W., Wu, Y. T., Ye, J. N., & Sun, Y. (2023, January). The association of short-video problematic use, learning engagement, and perceived learning ineffectiveness among Chinese vocational students. In *Healthcare* (Vol. 11, No. 2, p. 161). MDPI.
- [22] Wu, S. (2024). Application of multimedia technology to innovative vocational education on learning satisfaction in China. *PLoS One*, 19(2), e0298861.
- [23] Sania Fatima & Shameem Akther. (2024). Melanoma detection using Egret search golden optimization – Deep convolutional neural network model. *Biomedical Signal Processing and Control(PB)*,106653-106653.
- [24] Ranjeet Kumar Singh. (2024). Developing a big data analytics platform using Apache Hadoop Ecosystem for delivering big data services in libraries. *Digital Library Perspectives*(2),160-186.

- [25] Wen-Hsiao Peng & Chun-Chi Chen. (2013). An Interframe Prediction Technique Combining Template Matching Prediction and Block-Motion Compensation for High-Efficiency Video Coding. *IEEE Trans. Circuits Syst. Video Techn.*(8),1432-1446.
- [26] Sean Eli & Michael Schweinberger. (2024). Non-asymptotic model selection for models of network data with parameter vectors of increasing dimension. *Journal of Statistical Planning and Inference*106173-.