Efficient Barrage Video Recommendation Algorithm Based on Convolutional and Recursive Neural Network

Ping He, Siyuan Ma, Weidong Li

School of Information and Technology, Hebei University of Economics and Business, China
pinghe@heuet.edu.cn, masy_19@163.com, lwdbox@126.com

Abstract

Convolutional neural network has been widely used in recommendation methods. Recursive convolutional neural network not only has the advantages of automatic feature extraction, but also effectively captures the features before and after the time series. However, the traditional video recommendation process has the problems of low time efficiency and low accuracy. To solve the above problems, this paper proposes a method that is based on convolutional recurrent neural network in the barrage video recommendation. Firstly, according to the number of barrage videos, the method selects the preferable video fragments of users, and adopts the k-means clustering method to extract key frames from video. Secondly, we use a convolution and recursive neural network model (RCNN) to classify the similar video fragments. Finally, the recommendation can be achieved by the similarity between video fragments. Experimental results demonstrate the effectiveness and better performance of the proposed method.

Keywords: Barrage, Convolutional neural network, Recursive neural network, Video recommendation

1 Introduction

With the advances in digital process technology and mobile communication, the domestic and foreign appear many big video websites. After more than ten years of exploration and practice, major video sites (domestic video sites such as iQIYI, Tencent, Youku, foreign video sites such as YouTube and Netflix, etc.) have already occupied a majority position in the field of video website. And watching the video in the website has been integrated into people's daily life. Nevertheless, how to timely and accurately recommend the suitable videos that meet the users' needs, tastes and preferences has become a positive solution for promoting the development of some video websites.

Nowadays, deep learning has been widely used in image processing, speech recognition and natural language processing. Furthermore, the applications of deep learning in recommendation system have become a research hot spot of researchers [1-2]. Convolutional neural network as a kind of multi-layer perception is an efficient solution to increase deep learning applied in the application fields. The reasonable utilization of convolutional neural network tends to highlight the effect of applications. Recall that convolutional neural network owns the ability of learning characterization. It can achieve the translation invariant classification of input information by hierarchical structure. In addition, convolutional neural network still has the other advantages, such as strong robustness, better fault tolerance, easier to train and optimize. Therefore, it has obtained some research results in the aspect of application fields, such as visual recognition and image classification [3]. Recursive neural network is an artificial neural network that is similar to the tree hierarchical structure composed by the network nodes. It can recurse the input information on the basis of the connection order of network node. Moreover, it is a special kind of recurrent neural network, which has been widely paid attention to the field of natural language processing. Consequently, because of the advantages of deep learning technology, such as strong learning ability, wide coverage, strong adaptability and good portability, the researchers applied the deep learning technology to the field of video recommendation for the reason to improve the recall rate of users in a certain extent and achieve a better recommendation effect in general.

Currently, the barrage video has been extensively accepted by a large number of people and become a mainstream watching style. It results that the researchers pay more and more attention to study the related problems of barrage video recommendation for obtaining the best experiences of users when watching video. The barrage text sent by users in the video can reflect the users' feelings about the current video clip, and the number of barrages can reflect the preference degree of the current video clip by users. Therefore, under this circumstance, this paper explores the deep potential information by studying the number of barrages and combines with deep learning technology in order to improve the accuracy of video

*Corresponding Author: Weidong Li; E-mail: lwdbox@126.com
DOI: 10.53106/160792642021112206004

recommendation. Most studies of the barrage video are analyzed on the basis of the barrage contents. For example, Deng [4] proposes a video recommendation model based on barrage emotions, which uses Dirichlet function to classify barrage words, and calculates emotional similarity based on Jaccard distance. Hong [5] use dynamic time warping algorithm (DTW) to calculate the distance between emotional distributions. But the literature only gives priority to calculate the positive and negative emotions of the users who send a large amount of barrage data, and deletes the related information of the users who sent a small amount of barrage data. Therefore, it results that the users who send a small amount of barrage data and also prefer this video cannot be classified. In order to improve the monitoring intensity of the living anchors' behaviors, the literature [6] proposes an evaluation system of anchors' behaviors based on barrage texts, which predicts the emotions of audiences according to barrage texts and further evaluates the behaviors of anchors.

In the above literatures, the analysis of emotional value is seriously depended on the barrage contents and ignores the number of barrages. At the same time, the traditional video recommendation process has low time efficiency and low accuracy. To solve the above problems, this paper proposes a video recommendation model that is based on the data of barrages and combines with the recursive convolutional neural network. Specifically, the main contributions of this paper are summarized as follows:

- First, we propose a preprocessing model of the barrage data to analyze the user's preferred video clips in the video.
- Second, we propose a recursive convolutional neural network to achieve image classification and learning tasks. The proposed model can find some videos that the video clips contain the similar human behaviors, and recommend the discovered videos to users.
- Third, experimental results on several benchmark datasets demonstrate the effectiveness of the proposed based on the data of barrages.

The rest of this paper is organized as follows: In Section II, we describe some related works about the traditional video hybrid recommendation schemes and the video hybrid recommendation schemes based on deep learning. In Section III, we define the barrage video recommendation problem based convolutional recursive neural network and prove it NP-hard. In Section IV, we describe the work flow of proposed video recommendation scheme in details, including the data processing process and RCNN model. In Section V, we show our experimental results to evaluate the proposed recommendation model. In Section VI, we conclude the paper and describe the further researching direction.

2 Related Works

In this section, we describe the research results of the traditional video hybrid recommendation method and the video recommendation method based on deep learning.

2.1 Traditional Video Hybrid Recommendation Method

The traditional video hybrid recommendations are generally divided into the content-based video hybrid recommendation method and the collaborative filtering based video hybrid recommendation method. The descriptions of two recommendations are followed in details.

2.1.1 Video Hybrid Recommendation Method Based on the Content

The advantage of content-based recommendation algorithm is that it does not need the data of other users when recommending a user, so it has better interpretability. The disadvantage is that it is difficult to automatically extract the features of images, music and videos, and there is a cold start problem for new users. For example, Kong [7] proposes a recommendation method based on label weight score that each label is set the corresponding score, where the score simply denotes the weight of the item or the user on the label, in order to reduce the influence of objective factors on user scoring and improve the accuracy and authenticity of the score. On the basis of context, Lin [8] proposes a recommendation model based on user decision that selects the required features or the combinations of features according to the factors affecting user decision, and takes the preference of users' interests as the direct influencing factor of decision. Tzamousis [9] applies various machine learning algorithms to learn the efficient combination method of various recommendation algorithms, and can rapidly select the best hybrid method according to the given input information. Moreover, this method can be easily extended to the other recommendation methods.

2.1.2 Video Hybrid Recommendation Method Based on the Collaborative Filtering

In the aspect of video hybrid recommendation method based on collaborative filtering, Nguyen [10] proposes a probability model that combines the displayed and hidden feedback to a unified model. In order to solve the problem that a single user is only interested in some special fields, Zhang [11] proposes a collaborative clustering recommendation method that adopts clustering algorithm to group the users and items according to their interests or characteristics, and then makes the corresponding recommendations on the basis of each group. Zhang [12] improves the original frequency based and ranking-based information kernel extracting methods, and proposes a method that considers the similarity between users and fully utilizes the scoring in formation of users and goods when seeking the list of neighbors for optimizing the process of finding the most similar neighbors.

2.2 Video Recommendation Method Based on Deep Learning

2.2.1 Video Recommendation Method Based on Convolutional Neural Network

For solving the cold start problem resulted by no related score when new videos are introduced, Li [13] proposes a method that computes the video correlation directly from the content, the method uses the deep convolution neural network to process the video information. In addition, in order to improve the applicability of the convolutional neural network, Li [14] uses the condition convolution to extract the behavior features of users. The method introduces the feature vector of goods as the convolution kernel and does not set the training parameters. Cai [15] proposes an method that combines the matrix decomposition and cross channel convolution neural network. Wang [16] applies the dynamic convolution probability matrix decomposition model to the grouping recommendation, integrates the text representation method of convolutional neural network into the potential factor model, and meanwhile integrates the state space model into the potential factor model.

2.2.2 Video Recommendation Method Based on Auto-encoder

To alleviate the data sparseness problem, Feng [17] proposes a method that introduces the stack autoencoders into the recommendation system. The method inputs the noisy data to the automatic encoder for generate a new scoring matrix, and it uses the collaborative filtering model in the probability framework for deriving the Bayesian processing algorithm of collaborative depth learning based on sampling. Zeng [18] constructs a context-aware recommendation model that fuses an unsupervised variational auto-encoder into a probabilistic matrix decomposition model.

2.2.3 Video Recommendation Method Based on Boltzmann Machine

Zhang [19] integrates the user labels into Boltzmann machines with restricted real conditions, and uses the ability that the Boltzmann machines with real conditions can fit the arbitrary discrete distributions in order to predict the users' scoring values for new goods. In addition, this method adopts the TF-IDF algorithm in the classification of texts to predict the user's preference for the applied labels. Liu [20] introduces the deep network structure model and combines the model with the clustering algorithm based on time weighting. Moreover, the multi-layer restricted Boltzmann is used to form the deep Boltzmann machine. Shen [21] proposes a hybrid recommendation algorithm that combines the restricted Boltzmann machine (RBM) with the weighted Slope One. The algorithm uses the real value RBM based on the project to fill in the score matrix and introduces the project attribute information in the calculation process of similarity.

2.2.4 Video Recommendation Method Based on Deep Neural Networks

To alleviate the data sparseness problem, Hong [22] propose Cross-domain Deep Neural Network (CD-DNN) for the cross-domain recommendation. Wen [23] apply matrix factorization and Multi-Layer Perception (MLP) to capture both linearity and nonlinearity of user-item interactions. Then it fuse the effects of multiple implicit feedback through neural networks to boost the quality of recommendation.

3 The RCNN Model of BVR

In this section, we firstly define the video recommendation problem based on barrage and then prove it NP-hard.

3.1 Problem Definition

In this subsection, we describe the characteristics and working mechanism of barrage video and the related definitions.

When a user sees a scene in one video, he or she may write some texts and send those texts to the video in the period of time. The other users may send the texts to this video too when they watch the same scene of video later. When the number of barrages in this time is more than a certain number during a period of time, we can conclude that those users who send the video barrages are more interested in this scene. By analyzing these video segments interested by users, we can find the video with similar scene in order to recommend those similar videos to the users. The above questions can be summarized as definition 1.

Definition 1. (The barrage video recommendation problem based on convolutional recursive neural network, BVR) Given the set of videos is $V = \{v_1, v_2, \cdots, v_n\}$, where n denotes the number of videos in V and the playing time of the video v_i is T_{v_i}, the set of users is $C = \{c_1, c_2, \cdots, c_m\}$, the number of barrages in time t_j for video v_i is denoted by d_{i,t_j} and the number of viewing video v_i in time t_j is g_{ij} et al. Therefore, the barrage video recommendation problem based on convolutional recursive neural network (BVR) refers that use a RCNN model for the accuracy ε of the recommendation system is significantly improved.

Definition 2. (RCNN model) The model firstly takes the time point t_j in the video v_i when the number of barrages $d_{i,t_j} > \lambda$ (λ is obtained after testing), and secondly extracts the video fragments $f_i = \{q_{i,1}, q_{i,2}, \cdots, q_{i,|f_i|}\}$ on the basis of the time point t_j, where $q_{i,k}$ denotes the $k-th$ video segment in the $i-th$ video. Next, $K-means$ clustering scheme is used to extract the key frame in the set of extracted fragments f_i to generate the structured data set U. Finally, the problem can be obtain the set of classified predicting results S that is classified on the basis of the structured data set U by introducing the recursive convolutional neural network model, such that the BVR is solved.

The mathematical model of this problem is formally shown in equation (1).

$$\begin{array}{ll} Input & : V = \{v_1, v_2, \cdots, v_n\} \\ & C = \{c_1, c_2, \cdots, c_m\} \\ & v_i = \{(d_{i,t_j}, g_{ij})\} \\ & T_i = \{t_1, t_2, \cdots, t_{T_{v_i}}\} \\ & f_i = \{q_{i,1}, q_{i,2}, \cdots, q_{i,|f_i|}\} \\ Output & : U, S \\ Goal & : max\ \varepsilon \\ s.t. & (1)\ 1 \le i \le n \\ & (2)\ 1 \le j \le T_{v_i} \\ & (3)\ n, m \in N \\ & (4)\ t_j \in T_i \\ & (5)\ d_{i,t_j} > \lambda \end{array} \qquad \textbf{(1)}$$

3.2 NP-hard

In this subsection, we reduce BVR to the clique cover problem, and give the proving process of its $NP-hard$. We prove that the clique cover problem of graphs with K vertices can be accurately reduced to the problem that a set of videos f_i is classified into K subsets of videos. The NP hardness of clique cover problem is the $NP-hard$ problem, which can clearly show the problem that how exactly the recommended videos can be classified to some subsets of videos.

Theorem 1. *The barrage video recommendation problem (BVR) based on convolutional recursive neural network is $NP-hard$ problem.*

Proof: The clique cover problem can be accurately reduced to a video classification problem in polynomial time.

The clique cover problem (CCP) refers that all the vertices in a graph whether can be exactly divided into K cliques. The clique cover problem is Non-deterministic Polynomial complete problem (NPC), this problem can be reduced from the k-graph coloring problem. To derive this result, we transform the input graph G of the k-graph coloring problem to its complement graph G', so the problem that how the vertices in G' are divided into K cliques is same as the problem that how the vertices in G are divided into K independent sets. We can set one color for each independent set to get K colors, and make sure that all the vertices with the same color in the graph G are not connected to each other. Therefore, finding the solution the clique cover problem of graph G' is equal to solve the coloring problem of graph G. Because we know that the k-graph coloring problem is $NP-hard$, the clique cover problem can be proved by using the reduction of k-graph coloring problem. Therefore, the problem of clique cover is $NP-hard$.

From the above descriptions, BVR can be proved by using reduction of clique cover problem. In the clique cover problem, given an undirected simple graph $G = (V_i, E_i)$, a positive integer J, it finds an optimal grouping method such that the vertices of graph G can be divided to K independent sets and there are independent sets $\mathring{V}_i^c = \mathring{V}_1 \cup \mathring{V}_2 \cup \mathring{V}_3 \cup \cdots \cup \mathring{V}_K$ such that $K < J$, and a complete subgraph of G can be derived by $\mathring{V}_1$, $\mathring{V}_2$, $\mathring{V}_3$, $\cdots$, $\mathring{V}_K$.

First, we defined an example of BVR through the clique cover problem.

(1) Assume that the set of videos in BVR is $V = \{v_1, v_2, \cdots, v_n\}$, the set of processed extracted clips in video v_i is $f_i = \{q_{i,1}, q_{i,2}, \cdots, q_{i,|f_i|}\}$, the set of users is $C = \{c_1, c_2, \cdots, c_m\}$, an undirected graph is $G_i = (V_i, E_i)$, where the set of vertices $V_i = \{\overline{V}_{i,1}, \overline{V}_{i,2}, \overline{V}_{i,3}, L, \overline{V}_{i,|f_i|}\}$ in graph G_i represents the set of $|f_i|$ videos to be classified $f_i = \{q_{i,1}, q_{i,2}, \cdots, q_{i,|f_i|}\}$. The vertices $V_i = \{\overline{V}_{i,1}, \overline{V}_{i,2}, \overline{V}_{i,3}, L, \overline{V}_{i,|f_i|}\}$ of graph G can be divided into K cliques $\mathring{V}_i^c = \mathring{V}_{i,1} \cup \mathring{V}_{i,2} \cup \mathring{V}_{i,3} \cup \cdots \cup \mathring{V}_{i,K}$, where at most $K \le |f_i|$.

(2) $\exists \overline{V}_{i,j} \in \mathring{V}_{i,t}$, and $\mathring{V}_{i,t} \subseteq \mathring{V}_i^c$, where $1 \le j \le |f_i|$ and $1 \le t \le K$.

(3) $\mathring{V}_i^c = \mathring{V}_{i,1} \cup \mathring{V}_{i,2} \cup \mathring{V}_{i,3} \cup \cdots \cup \mathring{V}_{i,K}$ is constructed by the subsets in $\mathring{V}_i$, and meet the following conditions:

- $\mathring{V}_{i,j} \ne \varnothing, \forall j$.
- $\bigcup_{j=1}^{K} \mathring{V}_{i,j} = \mathring{V}_i^c$.

- $\mathring{V}_{i,p} \cap \mathring{V}_{i,q} = \varnothing (p \neq q)$, $\mathring{V}_{i,p} \subset \mathring{V}_i^c$, $\mathring{V}_{i,q} \subset \mathring{V}_i^c$.

Then, we prove that there is a minimum number of cliques K for the graph $G = (V_i, E_i)$ if and only if there is an optimal video classification method such that the set of videos f_i can be divide into K cliques and $K \leq |f_i|$.

$\Rightarrow$: Assume that there is an undirected graph $G_i = (V_i, E_i)$, where $V_i = \{\overline{V}_{i,1}, \overline{V}_{i,2}, \overline{V}_{i,3}, L, \overline{V}_{i,|f_i|}\}$ is the set of vertices in graph G, and the vertices have different degrees of $0 < \delta_1 < \delta_2 < \delta_3 < \cdots < \delta_a < |V_i|$. The set of vertices V_i can be divided into K cliques denoted by $\mathring{V}_i^c$, and each each clique in $\mathring{V}_i^c = \mathring{V}_{i,1} \cup \mathring{V}_{i,2} \cup \mathring{V}_{i,3} \cup \cdots \cup \mathring{V}_{i,K}$ denotes an independent set of vertices, where $\mathring{V}_{i,p} \cap \mathring{V}_{i,q} = \varnothing (p \neq q)$, $\mathring{V}_{i,p} \subset \mathring{V}_i^c$, $\mathring{V}_{i,q} \subset \mathring{V}_i^c$. Because $\exists G_i = (V_i, E_i)$ is true and the vertex in $\overline{V}_i$ represents the video segment $q_{i,k}$, so the problem that the vertices in V_i of graph G_i can be divided into K cliques in $\mathring{V}_i^c$ can be transformed into the problem that the set of video segments f_i is reasonably classified into K subsets of video segments, where $K \leq |f_i|$.

$\Leftarrow$: Assume that the set of videos is V, for each video v_i, according to the number of barrages d_{i,t_j} we can get the time point t_j, and can extract the set of video segments f_i to be classified based on t_j. For the video segment $q_{i,k}$ in the video segment set f_i, it needs to be divided into K video clusters, and all video segments meet $q_{i,k} \in f_i$. Since $\exists V$ is true, the problem of dividing all video segments $q_{i,k}$ into K video clusters in the set of video segments f_i can be reasonably transformed into the problem of dividing the vertices V_i in graph G_i into K cliques in $\mathring{V}_i^c$, where $K \leq |f_i|$.

As long as the above reduction can be easily completed in polynomial time, the set of video segments can be accurately classified and pushed to the user. Therefore the BVR is $NP-hard$. □

4 Proposed Video Recommendation Model Based on RCNN

In this section, we describe the video recommended workflow that consists of two sub-modules: (1) data preprocessing; (2) RCNN model, as shown in Figure 1.

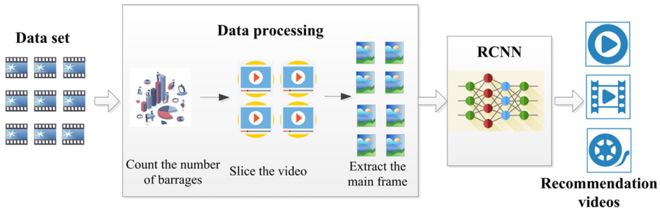

Figure 1. The structure of video recommendation

4.1 Barrage Data Preprocessing

The stage of data preprocessing mainly includes three steps: sorting out the number of barrages, slicing the video, and extracting the main frame. The barrage is shown in the forms of dynamic text, scroll text or static text, so the number of barrages in the same playing video is defined as the number of barrages on the same screen at the same time. The rolling time and stationary time of each barrage in Bilibili website are 7 seconds, and it is found by testing that there are few cases when the number of barrages on the same screen is more than 50 [24]. Guo [25] drew a "time-barrage" polyline graph with coordinates (t_j, d_{i,t_j}) at an interval of 5 seconds. On this basis, this paper firstly counts the number of barrages d_{i,t_j} of each video v_i, finds the time point t_j of the number of barrages $d_{i,j} > \lambda$ in video v_i, and then slices the video to form the video segments f_i according to the time point t_j. λ is simply computed in equation (2), where T_{v_i} denotes the total playing time of video v_i.

$$\lambda = \frac{1}{T_{v_i}} \sum_{t_j=1}^{T_{v_i}} d_{i,t_j} \qquad \textbf{(2)}$$

Considering the interference at the beginning and ending of video, it is necessary to remove the number of barrages in the video music beginning stage and the music ending stage, and remove the influence of herd effect on the number of barrages when dealing with the number of barrages. In this paper, the key frame extraction algorithm based on $K-means$ clustering is used to extract the main frames of video segments f_i and selects the video frames that can fully represent the video scenes.

4.2 RCNN Model

The RCNN model is reasonably composed of convolutional neural network (CNN) and recursive neural network (Elman recursive neural network). The network structure is shown in Figure 2. After the convolutional network passes through the full connection layer, it does not go directly to the function layer (classification layer) of the convolution layer, but directly to the added recursive neural network layer. The $Relu$ function is used as nonlinear mapping function in the network.

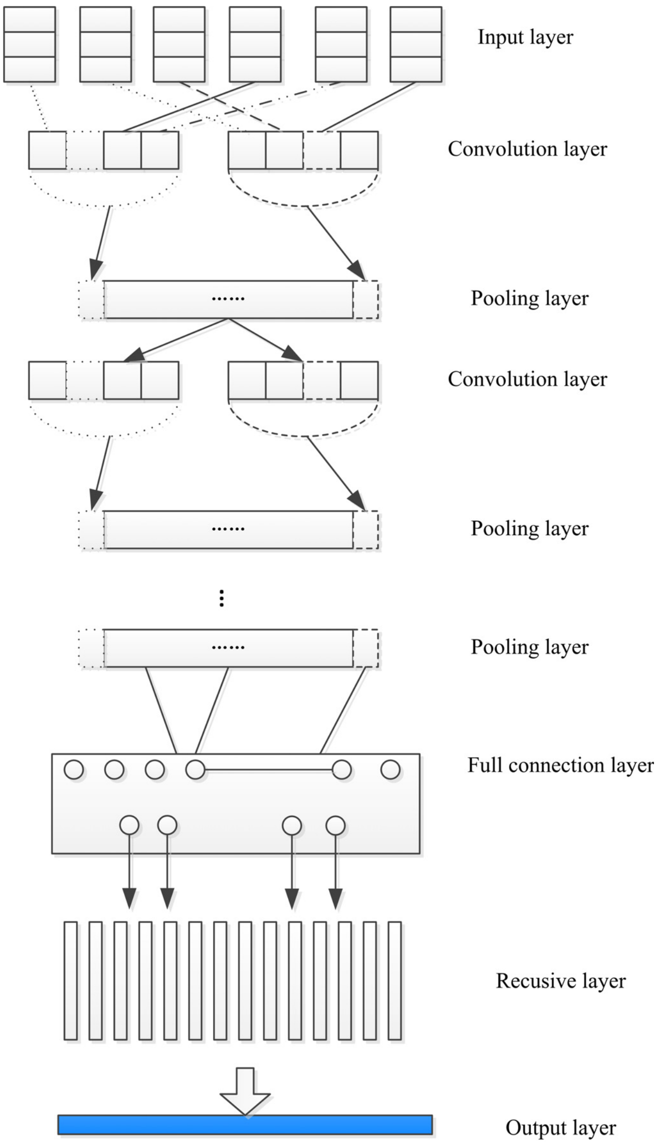

Figure 2. RCNN model

The construction steps of video recommendation model based on recursive convolutional neural network proposed in this paper are described as follows:

Step 1: according to the barrages, the video frames $\tilde{X}$ that can typically represent the video scenes are selected. Processes the video frame $\tilde{X}$ to the original image X.

Step 2: the convolutional neural network in RCNN model is used to extract convolutional information from the spatial features of original image X obtained by Step 1, then the feature graph is transformed to the information of low dimension features H_a through the pooling layer, and finally the information of low dimension features H_a is output through the full connection layer. Note that the adopted convolutional model is the effective mode, that all the pixels of convolutional kernel are selected from the input image.

Assume that the input of convolutional neural network is the original image X, and the feature of $a-th$ layer in convolutional neural network is denoted by H_a (where $H_0 = X$), so the generating process of H_a is exactly computed by equation (3).

$$H_a = f(H_{a-1} * W_a + b_a) \quad \textbf{(3)}$$

Where W_a is the weight vector of convolutional kernel in $a-th$ layer, the operator $*$ represents the convolutional operation between the convolutional kernel and the image or feature graph of $(a-1)-th$ layer H_{a-1}. Then, the convoltional output sums up the bias b_a of $a-th$ layer. Finally, the feature graph H_a of $a-th$ layer is contained by the non-linear active function $f(^\circ)$.

In pooling layer we adopt the Max pooling scheme that the maximum value is selected from local acceptance domain. Let H_a denote the pooling layer, so H_a is exactly shown in equation (4).

$$H_a = \max pooling(H_{a-1}) \quad \textbf{(4)}$$

In full connection layer we use *Relu* function as activation function. Assume that H_a is the full connection layer, and the output of full connection layer H_a is contained by the sum of weighted value in input layer H_{a-1} and the activation function $f(^\circ)$, where w_a is the weighted value of full connection layer, as shown in equation (5).

$$H_a = f(w_a H_{a-1} + b_a) \quad \textbf{(5)}$$

Step 3: take the information with low dimensional feature $u(k-1)$ as the input ($u(0) = H_a$), and then input it into the Elman model of RCNN model. It is shown in Figure 3.

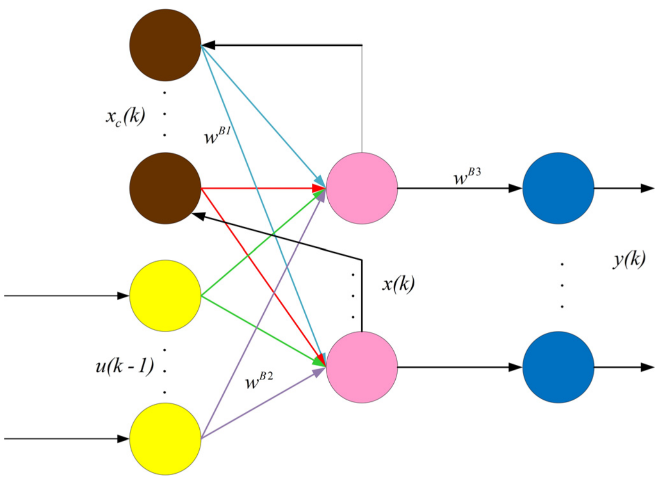

Figure 3. Elman network structure diagram

We can see from Figure 3 that the activation value of receiving layer in the current time (k) and the input value in the previous time ($k-1$) are used as the input value of network. The output relationships equation among input layer, hidden layer, receiving layer and output layer of Elman network are shown in the equation (6-8).

$$x(k) = f\left(w^{B1}x_c(k) + w^{B2}u(k-1)\right) \quad \textbf{(6)}$$

$$x_c(k) = \alpha \cdot x_c(k-1) + x_c(k-1) \quad \textbf{(7)}$$

$$y(k) = w_i^{B3}x(k) \quad \textbf{(8)}$$

Where $x_c(k)$ denotes the output of receiving layer in k time, $u(k-1)$ represents the output of input layer in $k-1$ time, $x(k)$ denotes the output of hidden layer unit in k time, $f(\circ)$ is the activated function, w^{B2} is the weight of the input layer to the hidden layer, w^{B1} is the weight of the receiving layer to the hidden layer, w^{B3} is the weight of the hidden layer to the output layer, the range of self-connection feedback gain factor is $0 \le \alpha < 1$. When the input $u(k-1)$ is delayed before introducing the input layer, $x_c(k)$ will be replaced by $x_c(k-1)$. $y(k)$ is the output of output layer.

Step 4: the $y(k)$ obtained by the RNN network will be input to the classification layer of the network, and we choose the *softmax* layer as the classification layer. In this layer, assume the output of $i-th$ neuron in previous layer is defined as $y_i(k)$, w_{ij} denotes the connected weight between the $i-th$ neuron of previous layer and the $j-th$ neuron of this layer, b_j is the bias. So the input x_j^l of $j-th$ neuron in the *softmax* layer is accurately shown in equation (9).

$$x_i^l = \sum w_{ij} \times y_i(k) + b_i \quad \textbf{(9)}$$

The output of neuron j in *softmax* layer is denoted by y_j^l and C is the total number of classifications. Each output y_j^l represents the probability that the corresponding input x_j^l will be classified into the category l, ($0 \le l \le C$). The relationship is shown in equation (10), where the initial value of c is 0.

$$y_j^l = \frac{e^{x_j^l}}{\sum_{c=0}^{C} e^{x_c^l}} \quad \textbf{(10)}$$

Step 5: for the loss function model in this paper, the loss function is shown in equation (11). According to the parameters set, we firstly use the Gradient Descent (SGD) algorithm to take the derivative about the weights and bias of each layer in convolutional network. And then, we update the parameters and simultaneously use the back propagation algorithm to adjust the feedback weights of Elman model. If the model reaches the maximum number of iterations or the loss function within a reasonable range, the model will stop the training process, otherwise the model will return to Step 2. In equation (11), where $y(k)$ is the predictive value, $y_d(k)$ is the real value.

$$Cost\ J = \frac{1}{m}\sum_{i=0}^{m}(y_d(k) - y(k))^2 \quad \textbf{(11)}$$

Step 6: when the train of model is end, we will obtain the specific video segments with user preferences that have been classified. Finally, the related videos are recommended to the users according to the similarity among video segments.

Algorithm 1. barrage recommendation algorithm

Input: video training set X_{train}, video test set X_{test}

Output: video recommendation set D

1. for each video X_{train} do
2. while ($a > 0$)
3. according to equation (3-4) calculate H_a;
4. according to GD calculate W_a, w_a, b_a;
5. end while
6. while ($k \ge 1$)
7. according to equation (6) calculate $x(k)$;
8. according to equation (7) calculate $x_c(k)$;
9. according to equation (8) calculate $y(k)$;
10. according to BP update w^{B1}, w^{B2}, w^{B3};
11. end while
12. according to equation (9) calculate x^l;
13. according to equation (10) calculate y^l;
14. end for
15. return D;

5 Experimental Results

5.1 Dataset

In this section, we briefly introduce the sources of data and experimental environment used in the experiment. The videos of this experiment are downloaded from the domestic barrage website "Bilibili" by "Jiji Down" software. From the set of classic videos, we randomly choose multiple types of videos, such as the modern comedy-Home with Kids, ancient costume comedy-Bronze Teeth Ji Xiaolan, humanities war drama-Bright Sword, historical costume drama-Youth Bao Zheng, action comedy-World for the Monkey King, martial arts love drama-The Heaven Sword and Dragon Saber By Jin Yong respectively. We choose the first 5 episodes of each drama, so the total number of episodes is 30. Since Bilibili website periodically maintains the barrage pool and limits the amount of barrage data, so the video with more than 20 minutes has about 3,000 barrages in per episode, and the video with more than 40 minutes has 6,000-8,000 barrages in per episode. The example of barrage is < p = "the time when the barrage appears at video (second), the mode of barrage (such as 1-3 denotes rolling barrage, 4 denotes bottom barrage, 5 denotes top barrage, 6 denotes the reverse barrage, 7 denotes the specified bits, 8 denotes the advanced barrage), font size (px), font color (HTML color (decimal)), the generating time of barrage (Unix format), barrage pool, the sender id of barrage, barrage id in barrage database" > barrage content, as shown in Figure 4.


```
<chatid>3473629</chatid>
<mission>0</mission>
<maxlimit>6000</maxlimit>
<state>0</state>
<real_name>0</real_name>
<source>e-r</source>
<d p="466.52100,1,25,16777215,1537589892,0,eed5d550,5477613164298244">真是年轻啊</d>
<d p="695.31600,1,25,16777215,1537590121,0,eed5d550,5477733165432832">故意说不会水</d>
<d p="758.83500,1,25,16777215,1537590185,0,eed5d550,5477766523256832">本来就是少女</d>
<d p="786.32700,1,25,16777215,1537590212,0,eed5d550,5477780895039488">和珅大头</d>
<d p="953.97900,1,25,16777215,1537590380,0,eed5d550,5477868729008128">真会拍马屁</d>
<d p="26.35200,1,25,16777215,1537590839,0,51671930,5478109172727808">真的停不下来 本来只打算看一集</d>
<d p="863.71700,1,25,16777215,1537593776,0,51671930,5479649300185090">和大人你是不是回去复习了哈哈哈哈</d>
<d p="930.55300,1,25,16777215,1537593861,0,51671930,5479693672775680">爱批注爱盖章嘛</d>
<d p="1026.55900,1,25,16777215,1537593971,0,51671930,5479751557840896">和珅的武功还不赖呢</d>
<d p="2385.80300,1,25,16777215,1537608030,0,373b0528,5487122667012100">扇子戏很多</d>
<d p="1517.79900,1,25,16777215,1537613903,0,7186a5fe,5490201565593604">ooo</d>
<d p="1279.42500,1,25,16777215,1537614966,0,bf745471,5490758836551680">四姑娘，我看你是想做皇后啊</d>
<d p="1811.69900,1,25,16776960,1537616845,0,40e8d4fc,5491744182370308">住手！</d>
<d p="1576.88700,1,25,16777215,1537619719,0,3cb9249f,5493250977169408">烟才是本体</d>
<d p="1832.32400,1,25,16777215,1537620005,0,3cb9249f,5493400705433600">怎么感觉文豪撩文盲都是从名字开始的</d>
<d p="2003.00300,1,25,16777215,1537620192,0,3cb9249f,5493499065532416">老板好淡定</d>
<d p="2598.63900,1,25,16777215,1537620809,0,3cb9249f,5493822558044160">教你写字的时候你就写倒了</d>
<d p="1855.50100,1,25,15138834,1537624402,0,b3e2015b,5495706352615424">囍</d>
```


Figure 4. Barrage list

By Python program, it crawls the sending time, content and sender id of the barrage. Then, the barrage data is simply processed, and is figuratively visualized, as shown in Figure 5.

Barrage transmission time	Barrage sender id	Barrage content
00:00:50	9.51E+07	oh亲爱的上帝 快看看我发现了什莫
00:00:54	1699353	下冰雹
00:00:54	85df764b	我叫夏雨，我叫夏雪，我叫下冰雹
00:00:54	5bfd1ca2	小紫猴！
00:00:55	b7f1c21a	b站竟然有这个
00:00:56	849f240f	夏东海秒变刘新建
00:00:56	9acff6d	从组家庭
00:00:57	e86c7df3	合影！
00:00:57	2f1cb5d8	合影
00:00:58	552d84c2	合影
00:00:59	1eb025be	合影
00:00:59	d6fd0afb	好暖啊
00:00:59	b8bda1cd	小雪美的勒
00:00:59	b749017e	居然已经15年了
00:00:59	d6b55f0c	合影
00:00:59	66de1888	合影！
00:01:00	407b7aca	合影！
00:01:00	70bb9da8	合影
00:01:00	15819157	合影
00:01:00	14550831	合影
00:01:01	46ed6115	合影
00:01:01	2a01391f	合影!
00:01:01	b693a27d	合影！！
00:01:01	74238a05	希望所有二婚家庭都这么美满鸭!
00:01:01	2b0e5b12	合影
00:01:01	e535b27c	合影
00:01:01	d6992e63	合影
00:01:01	2ab2ef26	合影
00:01:01	b3fbcc72	合影

Figure 5. The list of processed barrage

In this paper, we remove the number of barrages at the beginning and ending of video in the final number of barrages for reducing the noise of barrages, as shown in Figure 6.

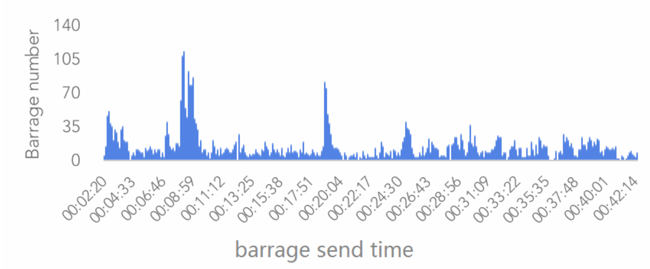

Figure 6. The changing number of barrages

Besides data preprocessing in this paper, we set the training model and some parameters, as follows:

(1) The operating system is Windows 7, Intel core i5 processor and 16GB RAM.

(2) The programming environment is Python 3.0, and the keras library environment is built. In order to train the weight of RCNN, the random the Gradient Descent (SGD + momentum) algorithm is used. We adopt Dropout operation with a dropout rate of 0.5 and use leaky ReLU with a negative slope of 0.2 in each layer. we use SGD with a learning rate of 0.0001 and momentum of 0.9 to fine-tune.

5.2 Baseline

For evaluating the whole performance of the proposed model, we randomly divide the data set into a training set (80%) and a testing set (20%). At present, in the study of the recommendation system, the mean absolute error (MAE) [26] and the root mean square error (RMSE) are commonly used as the evaluation indexes in the prediction task. Therefore, the mean absolute error and root mean square error (RMSE) are used as evaluation indexes in this paper.

The mean absolute error represents the average value of summing the absolute differences between the predicted value and the real value, and the corresponding calculation formula is shown in equation (12).

$$M_{MAE} = \frac{\sum_{i=1}^{N} | y_i - \hat{y}_i |}{N} \quad \textbf{(12)}$$

The root mean square error represents the square root value of average value after the square sum of the differences between the predicted value and the real value. The corresponding calculation formula is shown in equation (13). The smaller the RMSE value is, the more accurate the prediction model is to describe the experimental data.

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad \textbf{(13)}$$

In equation (12-13), where n is the number of samples, y_i is the actual result and $\hat{y}_i$ is the predicted result of experiment.

5.3 Evaluation Results

In order to verify the effectiveness of this model, we design two groups of experiments and the proposed method was compared with the collaborative filtering subject regression algorithm (CTR) [27], the collaborative filtering deep learning algorithm (CDL) [28], Convolutional Neural Networks (CNN) [29] and the convolutional matrix decomposition algorithm (ConvMF) [30] on the data set of this paper. In first group of experiments, we verify the influence of size to training set on the performance of these algorithms. We set that the size of training set is respectively 20%, 40%, 60%, and 80%. The experimental results are shown in Figure 7. From Figure 7, we can learn that the performance of four algorithms has little difference, but when the size of training set is more bigger the performance of RMSE in the proposed model is better than that of other models in different proportions, and the performance is best when the training set accounts for 80% of the data set.

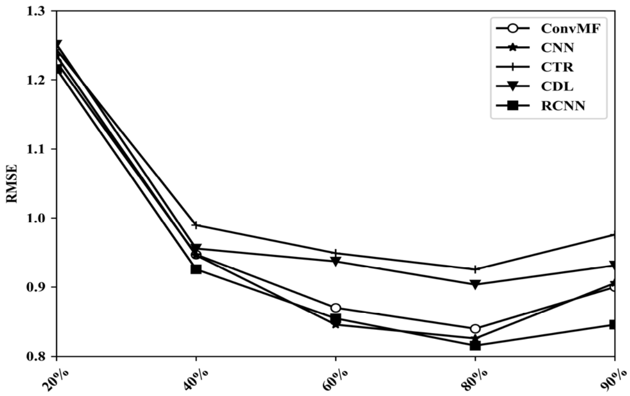

Figure 7. The experimental results of RMSE when the size of training data set is 20%, 40%, 60%, 80% of total data set

In the second group of experiments, we verify the predicting results of proposed recommendation system based on barrage recursive convolutional neural network. The experimental results are shown in Figure 8 and Figure 9. We observe that the performance of proposed recommendation method is better than those of CTR, CDL,CNN and ConvMF. From Figure 8, the average absolute error of proposed model is 0.6492, and is respectively increased by 11.56%, 7.64%,10.66%, and 5.64% compared with those of the other models in CTR, CDL, ConvMF and CNN. From Figure 9, the results show that the root mean square error of the proposed model is 0.8155 and is respectively risen by 10.96%, 8.78%, 2.43% and 0.90% compared with those of the other models in CTR, CDL, ConvMF and CNN. This indicates that the recursive convolutional neural model in this paper has a better recognition rate for images than those of the traditional convolutional neural network models, which means that the recognition and evaluation of users' preference for video segments based on barrage data plays a positive role.

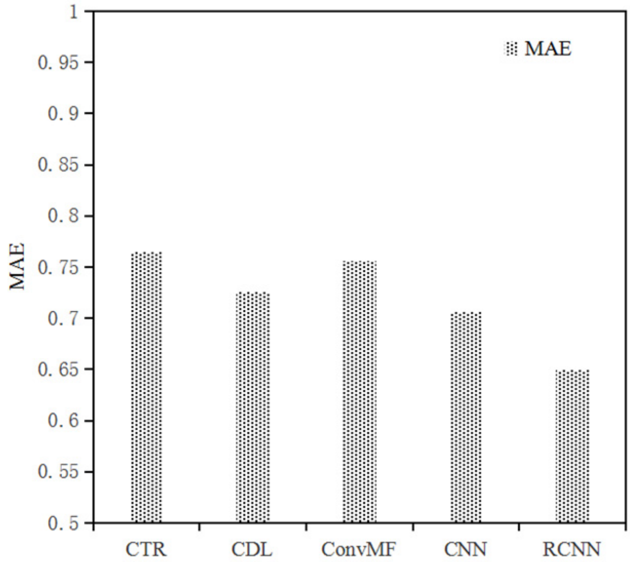

Figure 8. The experimental results of MAE for CTR, CDL, ConvMF and RCNN

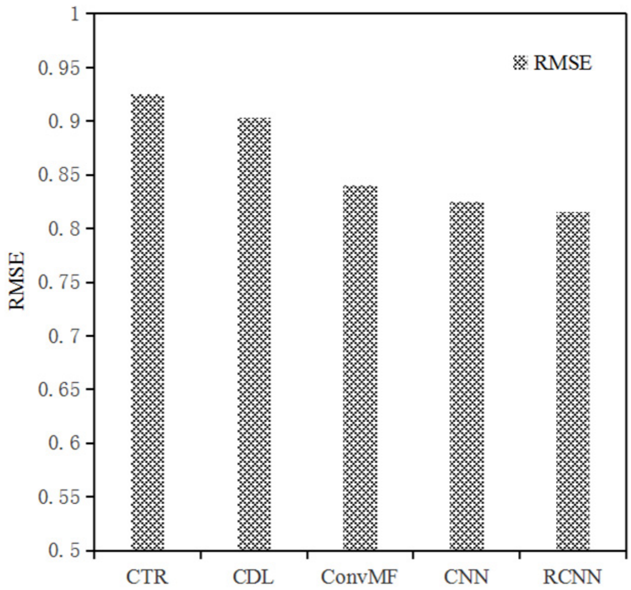

Figure 9. The experimental results of RMSE for CTR, CDL, ConvMF and RCNN

6 Conclusion

In order to improve the accuracy rate of the recommendation model, this paper proposes a video recommendation method that adopts the convolutional recursive neural network based on barrage. The barrage video can quickly capture the preference features of users by barrages, and improve the overall recommendation performance of prediction model with the help of convolutional recursive neural network.

The experimental results show that the mean absolute error of the proposed model is respectively increased by 11.56%, 7.64% 10.66%, and 5.64% compared with those of CTR, CDL, ConvMF and CNN model; The root mean square error of the model in this paper is respectively increased by 10.96%, 8.78%, 2.43% and 0.90% compared with those of CTR, CDL, ConvMF and CNN model. Therefore, the proposed model improves the accuracy of recommendation in a certain extent. However, the model has the lower recognition ability for the video without motion features. Therefore, we will add the content-related descriptions on the basis of researching the video key frame and further improve the recursive neural network [31-33] for further improving the accuracy of the prediction model in the next research. In addition, Convolutional transform learning (CTL) [34] can combine the benefits of unsupervised learning and convolutional neural network, so we attempt to introduce it to minimize the data fidelity loss function for improve the performance.

Acknowledgments

This work is supported by Scientific Research Projects of Hebei Education Department (Key Projects) #ZD2019017. The corresponding author is Weidong Li.

References

[1] L. Huang, B. Jiang, S. Lv, Y. Liu, D. Li, Survey on Deep Learning Based Recommender Systems, *Chinese Journal of Computers*, Vol. 41, No. 7, pp. 1619-1647, July, 2018.

[2] S. Ma, P. He, A Review of the Research on Recommendation Methods for Application Fields, *Computer Science and Application*, Vol. 9, No. 7, pp. 1317-1327, July, 2019.

[3] F. Zhou, L. Jin, J. Dong, Review of Convolutional Neural Network, *Chinese Journal of Computers*, Vol. 40, No. 6, pp. 1229-1251, June, 2017.

[4] Y. Deng, C. Zhang, J. Li, Video Shot Recommendation Model Based on Emotion Analysis Using Time-sync Comments, *Journal of Computer Applications*, Vol. 37, No. 4, pp. 1065-1070+1134, April, 2017.

[5] Q. Hong, S. Wang, Q. Zhao, J. Li, W. Rao, Video User Group Classification Based on Barrage Comments Sentiment Analysis and Clustering Algorithms, *Computer Engineering & Science*, Vol. 40, No. 6, pp.1125-1139, June, 2018.

[6] C. Zhao, Y. Li, R. Hong, J. Zhang, L. Gong, Supervision of Webcasting-Anchor Behavior Evaluation Based on Barrage Emotion Analysis, *In 2018 4th International Conference on Big Data Computing and Communications (BIGCOM)*, Chicago, IL, USA, 2018, pp. 66-71.

[7] X. Kong, B. Su, H. Wang, H. Gao, J. Li, Research on the Modeling and Related Algorithms of Label-weight Rating based Recommendation System, *Chinese Journal of Computers*, Vol. 40, No. 6, pp. 1440-1452, June, 2017.

[8] X. Lin, Y. Sang, C. Long. Personalized Recommendation Based on User Decision-making Mechanism, *Library and Information Service*, Vol. 63, No. 2, pp. 99-106, February, 2019.

[9] E. Tzamousis, M. Papadopouli, On Hybrid Modular Recommendation Systems for Video Streaming, *Information Retrieval*, pp. 1-23, January, 2019.

[10] T. Nguyen, A. Takasu, A Probabilistic Model for the Cold-start Problem in Rating Prediction Using Click Data, *International Conference on Neural Information Processing*, Guangzhou, China, 2017, pp. 196-205.

[11] F. Zhang, L. Zhang, T. Luo, Y. Wu, A Feature-Based Co-Clustering Model, *Journal of Computer Research and Development*, Vol. 55, No. 7, pp. 1508-1524, July, 2018.

[12] W. Zhang, J. Li, J. Yang, Improved extraction Method of Information Core in Collaborative Filtering Recommendation, *Application Research of Computers*, Vol. 37, No. 1, pp. 140-143, January, 2020.

[13] Y. Li, H. Wang, H. Liu, B. Chen, A Study on Content-based Video Recommendation, *2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, China, 2017, pp. 4581-4585.

[14] N. Li, Y. Sheng, H. Ni, Conditional Convolution Implicit Factor Model for Personalized Recommendation, *Computer Engineering*, Vol. 46, No. 4, pp. 85-90+96, April, 2020.

[15] N. Cai, G. Liu, H. Cai, Improved Model Combining Improved Matrix Decomposition and Convolutional Neural Networks, *Computer Engineering and Applications*, Vol. 55, No. 19, pp. 178-184, October, 2019.

[16] H. Wang, M. Dong, Latent Group Recommendation Based on Dynamic Probabilistic Matrix Factorization Model Integrated with CNN, *Journal of Computer Research and Development*, Vol. 54, No. 8, pp. 1853-1863, August, 2017.

[17] C. Feng, G. Si, W. Ni, Research on Collaborative Deep Learning Recommendation Algorithm, *Computer Systems & Applications*, Vol. 28, No. 1, pp. 169-175, January, 2019.

[18] X. Zeng, Y. Yang, S. Wang, T. He, J. Chen, Hybrid Recommendation Algorithm Based on Deep Learning, *Computer Science*, Vol. 46, No. 1, pp. 126-130, January, 2019.

[19] G. Zhang, B. Wang, Y. Hou, Real-Valued Conditional Restricted Boltzmann Machines with Tag for Recommendation Algorithm, *Journal of Frontiers of Computer Science and Technology*, Vol. 13, No. 1, pp. 138-146, January, 2019.

[20] K. Liu, *Research and Application of Movie Recommendation System Based on Deep Learning*, Master's thesis, Beijing University of Technology, Beijing, China, 2018.

[21] X. Shen, C. He, X. Meng, Research on Hybrid Recommendation Algorithm of Restricted Boltzmann Machine and Weighted Slope One, *Application Research of Computers*, Vol. 37, No. 3, pp. 684-687, March, 2020.

[22] W. Hong, N. Zheng, Z. Xiong, Z. Hu, A Parallel Deep Neural Network Using Reviews and Item Metadata for Cross-Domain Recommendation, *IEEE Access*, Vol. 8, pp. 41774-41783, February, 2020.

[23] H. Wen, X. Liu, C. Yan, L. Jiang, Y. Sun, J. Zhang, H. Yin, Leveraging Multiple Implicit Feedback for Personalized Recommendation with Neural Network, *Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM 2019)*, Dublin, Ireland, 2019, pp. 1-6.
[24] S. Liu, *Research on Human-computer Interaction Efficiency of Modern Web Barrage*, Master's thesis, Sun Yat-sen University, Guangdong, China, 2015.
[25] K. Guo, *Danmaku Data Mining Algorithm Study Based on Quantity Characteristics*, Master's Thesis, Huazhong University of Science and Technology, Hubei, China, 2017.
[26] J. Zhou, H. Huo, Multi-dimensional Space Video Recommendation Algorithm in Social Network Service, *Computer Engineering*, Vol. 41, No. 1, pp. 245-250, January, 2015.
[27] M. Zhang, B. Ding, W. Ma, Y. Tan, Y. Liu, S. Ma, Hybrid Recommendation Approach Enhanced by Deep Learning, *Journal of Tsinghua University (Science and Technology)*, Vol. 57, No. 10, pp. 1014-1021, October, 2017.
[28] C. Wang, D. M. Blei, Collaborative Topic Modeling for Recommending Scientific Articles, *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, CA, USA, 2011, pp. 448-456.
[29] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, *International Conference on Neural Information Processing Systems*, Lake Tahoe, Nevada, USA, 2012, pp. 1097-1105.
[30] D. Kim, C. Park, J. Oh, S. Lee, H. Yu, Convolutional Matrix Factorization for Document Context-Aware Recommendation, *the 10th ACM Conference on Recommender Systems*, Boston, Massachusetts, USA, 2016, pp. 233-240.
[31] C. Hsieh, M. Campo, A. Taliyan, M. Nickens, M. Pandya, J. J. Espinoza, Convolutional Collaborative Filter Network for Video Based Recommendation Systems, ArXiv:1810.08189, October, 2018.
[32] E. Liu, Y. Chu, L. Luan, G. Li, Z. Wang, Mixing-RNN: A Recommendation Algorithm Based on Recurrent Neural Network, *International Conference on Knowledge Science, Engineering and Management*, Athens, Greece, 2019, pp. 109-117.
[33] M. Xiong, Y. Feng, T. Wu, J. Shang, B. Qiang, Y. Wang, A Novel Recommendation Framework Fusing Temporal Dynamics, CNN-Based Text Features and Item Correlation, *IEICE Transactions on Information and Systems*, Vol. E102.D, No. 8, pp. 1517-1525, August, 2019.
[34] Z. Li, H. Zhao, Y. Guo, Z. Yang, S. Xie, Accelerated Log-regularized Convolutional Transform Learning and Its Convergence Guarantee, *IEEE Transactions on Cybernetics*, pp. 1-15, April, 2021. Doi: 10.1109/TCYB.2021.3067352

Biographies

Ping He received the Ph.D. degree in computer science and technology from Beijing Jiaotong University in July 2015. She is associate professor and currently working at Hebei University of Economics and Business. Her research interests include design and analysis of algorithms for optimization problems in wireless networks and data dissemination.

Siyuan Ma is a graduate student in the School of Computer Science of Hebei University of Economics and Business. Her research interests include recommendation systems, algorithm optimization, and multi-modal learning analysis.

Weidong Li received his Ph.D. in Control Theory and Engineering from Beijing University of Science and Technology in April 2006. He is working at Hebei University of Economics and Business. His research interests include design and analysis of algorithms for data mining and machine learning.