# Research on College Chinese Online Learning System Based on Content Recommendation Algorithm

Liang Chen(✉)

Gongqing College of Nanchang University, Jiangxi 332020, China
yuerushui@163.com

**Abstract.** With the development of Internet scale, recommendation algorithm is widely used in e-commerce, advertising, community and so on, and collaborative filtering algorithm is also widely used because of its simple implementation and clear process. However, the traditional recommendation algorithm has the problems of long cold start time, insufficient suppression of Matthew effect and insufficient utilization of item information. In the existing methods for cold start problems, the recommended contents are not fully utilized. If we want to use the recommended content as the information source of the recommendation system, we need an appropriate modeling method to analyze the content. Therefore, the focus of this paper is how to analyze and model the College Chinese online learning system, and finally use this information for recommendation.

**Keywords:** Content recommendation algorithm · College Chinese · Online learning system

## 1 Introduction

The emergence of recommendation system has a great impact on the traditional information retrieval business and traditional Internet services. Due to the penetration of recommendation algorithm, it is an active mining of content and information in the Internet. In other words, the recommendation algorithm is a more accurate automatic search, because it does not require user intervention. In the traditional e-commerce model, there are only two ways for users to retrieve the goods they want to buy: classified catalog and search. When user needs are not clear, search cannot work. On the other hand, the huge directory tree will make users lose patience. While increasing the time cost, users are unlikely to buy their favorite items. After the rise of recommendation system, e-commerce websites can establish user preference model for each user by using recommendation system. When the user's operation and data are more detailed, the accuracy of the model will be better. When the recommendation system generates a recommendation list for users, users can browse and buy goods that meet their interests in the list. These goods often meet the needs of users [1]. There are many information sources in the recommendation system, among which the most common is user history


M. A. Jan and F. Khan (Eds.): BigIoT-EDU 2022, LNICST 465, pp. 580–585, 2023.
https://doi.org/10.1007/978-3-031-23950-2_63

data, including user browsing, clicking, downloading, comments and so on. At the same time, the information source also has historical data of items, such as click through rate, sales volume, etc. After many optimization and changes, more and more data are also used to recommend algorithms, such as labels, time, geographical location and other additional information.

In order to better improve contemporary college students' Chinese learning ability, this paper takes the popular wechat applet for college students' Chinese online learning as the research carrier, takes the frequent and compulsory Chinese test types at the university stage as the direction, pursues objective reality, and constructs a self-defined and improved recommendation system to make personalized recommendation for college students in Chinese learning. The recommendation system mainly focuses on content-based recommendation algorithm, supplemented by user based collaborative filtering algorithm And make appropriate amendments and improvements according to the type framework of College Students' Chinese test questions, so that it can make corresponding personalized recommendations to college students more effectively, so as to effectively improve college students' Chinese learning ability.

## 2 Content Based Recommendation Algorithm

Content based recommendation (CB) is a recommendation algorithm proposed and used earlier, even earlier than user based collaborative filtering. This recommendation method is to recommend corresponding items or contents for users according to the items or contents users use or click, which was first applied to information retrieval system.

There are three main steps of CB algorithm:

(1) Item representation: extract some features (that is, the content of the item) for each item to represent the item. For an item, the data contained in it is divided into two categories: structured data and unstructured data. Taking text as an example, its structured data includes such as title, label, length, publishing time, etc., while the text itself is unstructured data. Because of its content and ideas, we can't store it directly through a simple data structure. In the object representation, we need to extract these two parts of the object at the same time and make a unified structured representation. In text analysis, the familiar vector space model (VSM) and term frequency inverse document frequency (TF IDF) are a kind of modeling after extracting features from text documents in the field of information retrieval. Music can be recommended based on metadata, such as common writers, albums, release dates, etc., which is the earliest content-based recommendation method involving music. However, this has great limitations [2]. The reason is that the styles of these composers and singers are relatively changeable, which are affected by their own conditions such as their experience, times and age. If you make recommendations based solely on this information,

The results obtained are not entirely accurate.

(2) User learning: use the feature data of a user's favorite (and disliked) content in the past to learn the user's preference profile. There are many different methods to learn

user preferences, such as k-nearest neighbor (KNN), Rocchio algorithm and so on. The former is a clustering method. For a new item, the algorithm finds K items that user u has scored and similar to this item, and judges the preference for the new item according to U's preference for these K items. The difference from the collaborative filtering algorithm is that the KNN method here is the preference calculated according to the item attribute vector, not the user score. Rocchio algorithm is an algorithm for dealing with relevance feedback in information retrieval. The feedback data generated by the user's behavior in the operation process constantly modifies the vector representing the user's preference to approximate the real result. The formula is as follows:

$$u_{k+1}(t) = u_k(t) + \Gamma_{l1}\dot{e}_k(t) + \Gamma_{l2}\dot{e}_{k+1}(t) + \Gamma_{p1}\Delta\dot{e}_k(t) + \Gamma_{p2}\Delta\dot{e}_{k+1}(t) \tag{1}$$

(3) Recommendation generation: by comparing the user profile obtained in the previous step with the characteristics of candidate items, recommend a group of items with the greatest relevance for this user. The method used in this step depends on the type of model used in step 2). If the learning method is used, the algorithm will return K items most related to the user's attributes to the user as recommendations; If the traditional classification method is adopted, the algorithm will directly predict K users' favorite items through the model and recommend them to users. The specific similarity measurement can be obtained by calculating cosine similarity or Euclidean distance, and the formula is as follows:

$$\begin{aligned}(E(t) - M_2C)\Delta\dot{x}_{k+1}(t) = & f(t, x_d(t)) - f(t, x_{k+1}(t)) + B\Delta u_k(t) \\ & - \left(\Gamma_{p1}C + M_1C + M_2C\right)\dot{x}_d(t) + \Gamma_{p1}C\dot{x}_{k-1}(t)\end{aligned} \tag{2}$$

CB algorithm has advantages, such as independence between users, generating a user's recommendation list does not depend on the behavior of other users. For example, it is better interpretive because it models items and users separately and quantifies some specific attributes, which can well explain the reasons for the recommendation results.

# 3 College Chinese Online Learning System Based on Content Recommendation Algorithm

## 3.1 Overall Structure

The composition of convolutional neural network crnn is divided into two main parts: one is feature learning layer and the other is feature mapping layer. The crnn structure designed in this paper has four hidden layers, plus the input layer and output layer, there are six layers. The first layer is the input layer. As mentioned above, we process the text into 96 × Mel spectrum of 1366, which will be directly used as the input of the network.

The next 2 to 5 layers are hidden layers. The hidden layer is the main component of convolutional neural network, which is also the difference between it and general depth neural network. The hidden layer is composed of neurons in accordance with the number of pixel points. The convolution kernel scans and learns the features on the plane of each layer. After pooling, the features are mapped to the next layer, which is a layer by layer mapping process from low-level features to high-level features. Finally, we access an output layer in the form of full connection, and the output layer is also a classifier. The results we get can be classified and recommended [3].

Each H × W × The convolution layer of D will learn D features from the image with width of W and height of H. each convolution layer is followed by a pool layer, which reduces the number of feature maps of each layer. If it is the maximum pool, it means that the internal maximization function is adopted to screen the most significant features.

The model structure diagram of crnn is shown in Fig. 1:

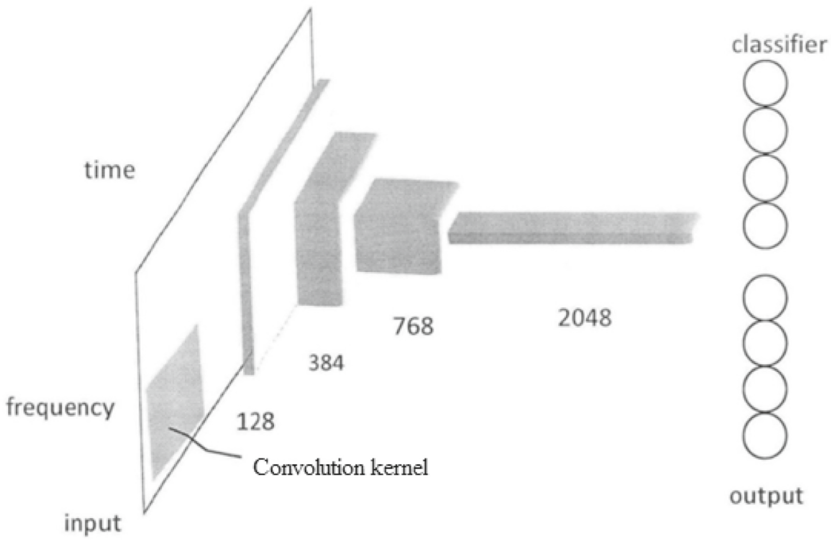


**Fig. 1.** Model structure diagram of crnn

### 3.2 Dropout Optimization and Regularization

As mentioned above, over fitting and gradient disappearance are easy to occur after the model is complicated. According to statistics, if the content of the article is too long, it will be divided into two parts in the Chinese text processing stage and processed into a Mel spectrum. Therefore, for a music, in the tensor diagram of n × 1 × 15 output in the CNN substructure of crnn, the maximum value of n is 2, so the number of layers of RNN is finally two. The RNN substructure adopts Gru structure (as shown in Fig. 2), because its required number of iterations and convergence speed are better than the LSTM model with the same function.

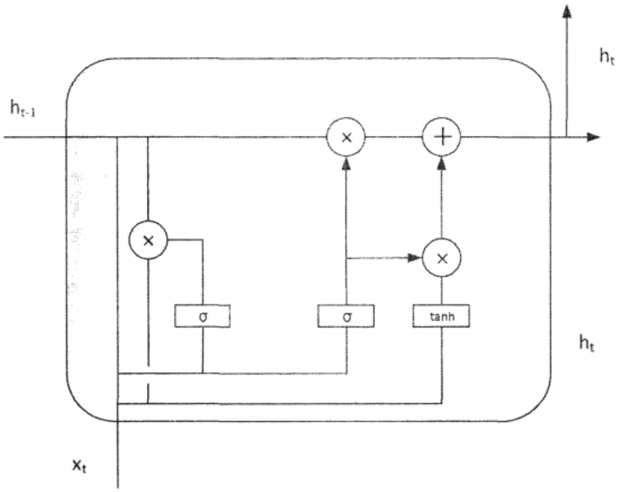


**Fig. 2.** Gru structure

### 3.3 Output and Recommendation

As mentioned above, the model needs to take the output of CNN substructure as the input of RNN, so adjust the output of CNN substructure to n × one × Tensor of 15. RNN is divided into two layers. The output of the first layer will be the whole sequence. For the second layer, we only select the last value of the sequence as the output, representing the most significant feature quantity in this group of feature graphs. This is because we only need to output 1 at a time × 1 to the classifier [4].

We combine the output layer of RNN with the classifier activated by sigmoid function, and each neuron is constrained by 1 × The tensor of 1 corresponds to the predicted offset of 50 labels, and the value of this 50 offset is used as the vector describing the category of Chinese articles.

## 4 Conclusion

In the analysis and comparison of the existing methods to optimize the recommendation algorithm, it is found that the traditional methods generally do not pay enough attention to the information contained in the recommended content itself. This part of information can not be used as the information source of the recommendation algorithm effectively and completely, which is actually one of the reasons for the cold start. When a new item is added to the database, there is no user data, so it is difficult to be scanned by the traditional collaborative filtering algorithm. If the recommendation algorithm can use the information of the content itself to analyze the "semantics" of the content, it can define the nature of a content and directly make basic recommendations when new content is added. Based on this, taking the College Chinese online learning system as an example, this paper puts forward a recommendation method to analyze the content semantics.

## References

1. Liang, X.: Practice of recommendation system. People's Posts and Telecommunications Press (2012)
2. Shuying, Y., Hua, Z.: Pattern recognition and intelligent computing. Electronic Industry Press (2015)
3. Yuhang, Z., Wenjuan, Y., Shan, J.: Overview of personalized recommendation system. Value Eng. **39**(02), 287–292 (2020)
4. Bing, S., Pengfei, D.: Speech emotion recognition method based on convolutional neural network. Sci. Technol. Innov. Guide **13**(6), 87–90 (2016)