Academic literature recommendation technology based on two-layer attention network

Hanyang Deng\textsuperscript{1,a} and Shiping Tang\textsuperscript{1,b}\textsuperscript{*}

\textsuperscript{1}Computer Science, Beijing Institute of Technology, Beijing, 100081, China
\textsuperscript{a}email: 3120180986@bit.edu.cn
\textsuperscript{b}email: simontangbit@bit.edu.cn

Abstract. Recently, academic literature recommendation for learners has become an important topic. Recently deep learning based models are used in literature recommendations, which follow a similar Embedding&MLP paradigm. However, in the base model and its follow-up model Deep Interest Network (DIN), users only have to click or not click on items, limiting the expression of users' specific interests. Therefore, we propose a two-layer attention model (BIH) based on DIN. BIH adds specific behaviors to the user behavior sequence and adds a behavior attention layer, which can learn the expression of user interests more accurately. Experiments on a public dataset and users' behavior logs of academic literature demonstrate the effectiveness of proposed approaches, which achieve superior performance on both Macro-F1 and Micro-F1 compared with the base model and DIN.

1. Introduction

In recent years, more and more people are studying online. However, how to choose the one that suits them from the numerous academic literature is an important problem for learners. Recommending appropriate academic literature for learners based on their historical behavior has become an important topic.

The structure of the recommendation system model has evolved from shallow to deep. In applying these models to academic literature recommendations, the unique characteristics of academic literature are more and more. More models are proposed for better academic literature recommendation. Zhang\textsuperscript{[6]} put learning styles into user-based and item-based collaborative filtering recommendations to help learners quickly find learning resources. Ma\textsuperscript{[5]} proposed a literature recommendation method that combines authors and literature influence. This method integrates information such as authors and literature influence based on matrix decomposition to recommend academic literature. Deep Edu\textsuperscript{[3]} improved based on NeuralCF\textsuperscript{[1]}, using the network of Embedding&MLP to recommend books, which significantly improved the recommendation effect. Fan\textsuperscript{[4]} uses the network of embedding + CNN + MLP to recommend academic literature. CNN can obtain the text features from the text information of the literature.

Inspired by the success of the attention mechanism in computer vision and natural language processing, the Deep Interest Network\textsuperscript{[2]} (DIN) was proposed to predict user behaviors and make corresponding recommendations to users. DIN is the first work in the recommendation system that uses the attention mechanism, which indicates the diversity of users' interests and uses the attention mechanism to activate historical behaviors about candidate items. However, the DIN model only considers the correlation between a given candidate item and the user's historical behavior sequence.
but ignores the user's different behaviors on different items in the historical behavior sequence. For example, in the users' behavior logs of academic literature, a user may search, browse, or download specific literature. These different behaviors may have different effects on the user's interest. In DIN, the user's behavior is only clicking or non-click, limiting the expression of the user's specific interest. Therefore, we propose a two-layer attention model (BIH) based on DIN. BIH considers the correlation between items and specific behaviors in the user behavior sequence by a behavioral attention layer (HA) and an item attention layer (IA), and adaptively calculates the representation vector of user interest. Finally, verifies the effectiveness of the BIH through experiments.

2. The two-layer attention network (BIH)
This section will introduce the BIH model in detail.

![Network architecture](image)

Figure 1. Network architecture. The Base Model is the model without the item attention layer (IA) and behavior attention layer (HA). DIN is the Base Model + IA. The BIH is the DIN + HA. The user's interest can adaptively change according to the user's different behaviors on the literature in the historical behavior.
2.1. Behavioral attention layer (HA)

Among all the features, the users’ behavioral features are important, and in the literature recommendation, they play a crucial role in modeling user interests. A user’s behavior on items only has to click or not click in DIN, which means that the user has only two attitudes towards items. In reality, users’ behaviors on items are diverse, which means that users’ interests in items are complex. A behavioral attention layer is added to the model based on DIN to better express users’ interests.

The structure of the BIH model is shown in Figure 1. Compared with the DIN model, the BIH adds the user’s specific behavior to the items in the user’s historical behavior part of the input layer and adds a behavior attention layer to obtain the relationship between the literature and the user’s historical behavior. The other structure remains unchanged. Specifically, the activation unit is applied to user behavior features, and it is used as a weighted sum pool to adaptively calculate the user interest representation $v_U$ given the candidate literature $A$ and the user’s specific behavior:

$$v_U(A) = f(b_1, b_2, ..., b_H; e'_1, e'_2, ..., e'_H) = \sum_{j=1}^{H} a(e'_j, b_j) e'_j = \sum_{j=1}^{H} w_j e'_j$$  \hspace{1cm} (1)

$$e'_j = f(v_A, e_j) = a(e_j, v_A)e_j = w_j e_j$$  \hspace{1cm} (2)

Where $E = \{e_1, e_2, ..., e_H\}$ is a list of embedding vectors of items in the historical behavior of user $U$ of length $H$, $B = \{b_1, b_2, ..., b_H\}$ is a list of embedding vectors of specific behaviors in the user’s historical behavior, $v_A$ is the embedding vector of candidate literature $A$, $E' = \{e'_1, e'_2, ..., e'_H\}$ is the item-level representation of the user’s historical behavior. In this way, $v_U(A)$ changes with the specific behaviors in candidate literature and user historical behaviors.

2.2. Loss function

The loss function is the cross-entropy loss function:

$$L_{target} = -\frac{1}{N} \sum_{(x,y) \in S} \sum_{t=1}^{T} y_t \log o_t$$  \hspace{1cm} (3)

Where $S$ is the training set of size $N$, and $T$ is the number of classifications, $x$ is the network’s input, $y = \{y_1, y_2, ..., y_T\}$ is the label, and $y_t \in \{0, 1\}$. $o = \{o_1, o_2, ..., o_T\}$ is the network’s output, and $o_t$ represents the probability that the predicted category is $t$.

3. Experiments

This section will introduce the experiment in detail, including datasets, evaluation indicators, experimental setup, model comparison, and corresponding analysis. Experiments on public datasets with user behaviors and users’ behavior logs of academic literature prove that the effectiveness of this method is better than the latest methods.

3.1. Datasets and Experimental Setup

MovieLens Dataset. The MovieLens dataset contains 138,493 users, 27,278 movies, 20 categories and 20,000,263 samples. In the dataset, each user’s rating of the movie is on a 5-point scale, increasing on a half-point scale (0.5 to 5 points). It is transformed into three-category data to make it suitable for prediction tasks. Mark the samples with a score of 4.5 to 5 as 2, which means they like it very much; mark the sample with 3.5 to 4 as 1, which means they like it; the rest are marked as 0, which means they do not like it.

Users’ behavior logs of academic literature. Collect user behavior logs from the literature website, and obtain the user’s browsing and download records of the literature from July 2020 to December 2020. The data contains 6697 users, 434093 articles, 10 categories, and 2532180 samples. Each user's behavior on the literature has two kinds of browsing and downloading. It is transformed into four-category data to make it suitable for the prediction task. Mark the sample that the user browses and downloads as 3; the sample that only performs downloading is marked as 2; the only browsing is marked as 1; for negative sampling, literature that does not appear in the user’s historical behavior is randomly selected and marked as 0.
The statistical information of the data set is shown in Table 1.

| dataset     | user  | item  | category | sample       | behavior |
|-------------|-------|-------|----------|--------------|----------|
| MovieLens   | 138493| 27278 | 20       | 20000263     | 3        |
| literature  | 6697  | 434093| 10       | 2532180      | 4        |

3.2. Model comparison

Base Model. The base model follows the Embedding&MLP architecture and is the basis of most of the predictive model deep networks developed subsequently. In the input, the specific behavior of each item in the user's historical behavior is added as a feature of the behavior, and the item id and item category are combined and then summed, fed into the subsequent network, and finally output the level of user's interest in the candidate item.

DIN (Base Model + IA). The DIN model adds a layer of attention to the items in the user's historical behavior and candidate items based on the base model. The specific behavior of each item is added to the user's historical behavior. After the items in the behavior and the candidate item passes the attention layer, the weighted vector of each item in the behavior is obtained, and the specific behavior is combined and then summed and fed into the subsequent network, and finally output the level of user's interest in the candidate item.

BIH(DIN + HA). A layer of HA is added based on the DIN model. The specific behavior of each item in the user's historical behavior is added to the input. After the items in the behavior and the candidate item passes the IA, the weighted vector of each item is obtained, and then the specific behavior is passed through the HA, and the weighted sum is fed into the subsequent network, and finally output the level of user's interest in the candidate item.

3.3. Metrics

In multi-class prediction, Macro-F1 and Micro-F1 are widely used.

\[ Micro_F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \]  \hspace{1cm} (4)
\[ \text{precision} = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N} TP_i + \sum_{i=1}^{N} FP_i} \]  \hspace{1cm} (5)
\[ \text{recall} = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N} TP_i + \sum_{i=1}^{N} FN_i} \]  \hspace{1cm} (6)
\[ F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \]  \hspace{1cm} (7)
\[ Macro_F1 = \frac{\sum_{i=1}^{N} F_1_i}{N} \]  \hspace{1cm} (8)

Where TP stands for predicting positive classes as positive classes, FP stands for predicting negative classes as positive classes, and FN stands for predicting positive classes as negative classes. N represents the number of species.

3.4. Results

The results of each model on the two datasets are shown in Table 2. The results of each model on the literature dataset are shown in Figure 2. All experiments were repeated three times, and the average results were taken. The effect of all networks with attention layers is better than Base Model, which proves that the attention mechanism is effective. The effect of BIH is better than the DIN model, which proves that adding a behavioral attention layer based on DIN is effective. The HA can obtain the influence of the user's specific behavior on the user's specific interest. The BIH model obtains the self-adaptive changes of user's interests along with their behavior and candidate items through two
attention layers. Compared with other networks, it dramatically improves the expressive ability of the model.

On the MovieLens dataset, the Macro-F1 and Micro-F1 of each model are not much different, while on the literature dataset, the two metrics of each model are quite different. Analyzing the reason, it may be that in the public dataset, the number of each category is not very different, while in the literature dataset, the number of downloads is much more than the number of browsing, and the unevenness of the dataset leads to such a result.

Table 2 Macro-F1 and Micro-F1 values of each model on two datasets

| model                  | MovieLens | Literature |
|------------------------|-----------|------------|
|                        | Macro-F1  | Micro-F1   | Macro-F1  | Micro-F1   |
| BaseModel              | 0.6934    | 0.6916     | 0.7281    | 0.7416     |
| DIN(BaseModel+IA)      | 0.7046    | 0.7028     | 0.7512    | 0.7641     |
| DIN+HA                 | 0.7179    | 0.7162     | 0.7662    | 0.7804     |

Figure 2. The results of each model on the literature dataset

4. Conclusions

This paper focuses on predicting the behavior of users with rich historical behavior data and then recommending literature to users. In the literature recommendation, users may have multiple behaviors such as browsing and downloading instead of simply clicking and not clicking, and these different behaviors also represent users' different interests in different literature. To better express the relationship between users' interests and behaviors, we proposed the BIH model based on DIN. In the input layer, the specific behavior of each literature in the user's historical behavior is added, and a behavioral attention layer (HA) is added to make the user's interest adaptively change with his behavior on literature. The experimental results show that, on the literature dataset, the BIH improves the Macro-F1 and Micro-F1 by 0.015 and 0.0163, respectively, compared to DIN.

At present, we can only recommend some of the most interesting literature to users. In future, we will focus our work on recommending a complete learning route for each user.
Acknowledgments
The authors would like to thank the editor and the anonymous reviewers for their suggestions that have helped us improve the work.
This work was supported in part by the National Key R&D Program of China under Grant 2019YFB1406303.

References
[1] He, Kaiming, et al. Neural collaborative filtering. Proceedings of the 26th international conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2017.
[2] Zhou, Guorui, et al. Deep interest evolution network for click-through rate prediction. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2018.
[3] F.Ullah et al. “Deep Edu: A Deep Neural Collaborative Filtering for Educational Services Recommendation” in IEEE Access, vol. 8, pp. 110915-110928, 2020, doi: 10.1109/ACCESS.2020.3002544.
[4] Fan, Haiwei et al. Based on MLP improved deep neural network learning resources promotion Recommended Algorithm [J/OL]. Computer Application Research. https://doi.org/10.19734/j.issn.1001-3695.2019.04.0109.
[5] Ma, Huifang et al. Scientific papers combining the influence of authors and literature Recommend Method [J]. Journal of Northwest Normal University (Natural Science Edition), 2021, 57(02): 55-62.
[6] Zhang, Junge. Design and implementation of an online learning platform supporting personalized recommendation [D]. Shaanxi: Shaanxi Normal University, 2014.