*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*March-April 2022 Page No. 686-697*
*Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022*

# Self-focus Sequence Recommendation Model for Fusion Context Information

**Wanjun Yu, Yu Tian***
School of computer science and Information Engineering, Shanghai Institute of Technology, Shanghai 201418, China
*Corresponding Author.

***Abstract:***

Sequence recommendation is a hot research content in the system. The core idea of sequence is to dig all the relationships between users and projects from the sequence of users and projects, and personalize the user's next item may interact. Existing Most of the research methods are modeling the user ID and the project ID interactive sequence, and the impact of the project and user context information is ignored. For this issue, this paper proposes a converged project and user context information. Self-focus sequence recommendation model. This model consists of two parallel modules of the embedded layer module, convolutional neural network and self-focus mechanism network, where the convolutional neural network module is modeled for dynamic preferences of the user and the project interaction. Self-focus mechanism network module capture user and user contextual information characteristics. Finally, the user's dynamic preferences and projects learned from the two modules are combined with the user context information feature to enhance the recommended performance. The experimental results show that the model of this paper has increased by at least 23.2%, the accuracy rate and the recall rate are also significantly improved compared to the current baseline method.

***Keywords****: Context information, Self-focused mechanism, Sequence recommendation, Convolutional neural network*

## I. INTRODUCTION

Now, both e-commerce applications closely related to our lives, or under the short video application of the current hot, they are popular behind the recommendation system. At present, users have a large number of hidden feedback behavior when using these applications (such as browsing, click, collection, add to cart, sharing, etc.), but traditional collaboration algorithm uses these data in defects, so sequence recommendation becomes the value of these data Maximize key technologies.

The sequence is recommended to capture the user's dynamic preferences, and the usual practice is to model the sequence of users interacting with the project in the past period. If the traditional sequence recommended Markov chain model is the typical application [1]. The Markov chain will be used from the information captured in the sequence modeling for the next interaction forecast, but the Markov chain model has an obvious drawback, that is, it thinks that the current interaction is only one or a few or a few the interaction behavior is related. One of the first-order Markov chain models believes that the current preference is only related to the last sequence behavior of the user, and the high-end Markov chain model

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 686-697
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

assumes the current interaction preference of the user and the user's recent sequence behavior [2]. Related. Can only capture users' recent preferences, and ignore the long-term preference of users is a common problem with low-end Markov chains and high-end Markov chain models. Another commonly used sequence recommendation model is based on a method of decomposition machine. It is mainly encoding the interaction between users - projects into two corresponding low-dimensional matrices. In addition, Liu [3]. Also uses the probability matrix decomposition recommendation algorithm to solve the traditional matrix decomposition of data sparse problems. Bayes Personalized Sort (BPR-MF) model method optimizes the PAIRWise target function by random gradient, which is well captured to the user's preference [4]. The most representative baseline algorithm FPMC can achieve recommendations for the next item, because It has the advantages of the Markov chain model and the matrix decomposition model [5]. The drawbacks of the algorithm model based on matrix decomposition are that they are often modeling for low-order users - project interaction, and did not consider capturing high levels. Interaction conversion mode.

Subsequently, in the field of recommendation system, deep capture technology has been striving for the researchers, whether in the industry or academic community, it has an extremely important position. Circulating Neural Network (RNN) is the first depth capture technology applied to sequence recommendation systems one[6]. Compared to traditional sequence recommendations, RNN-based sequences are recommended to deal with sequence data in a session, and the work-dependent relationship between the project is accurately interacting, and He[7]. Work is of the typical. Due to tradition Sequence recommendation model and sequence recommendation technology based on circular neural networks There is only point-dependent relationships in the user-project interaction sequence, and the Xiao[8]. Takes the first to collect convolutional neural networks in sequence recommendations. The jump relationship between the interactive sequence and the joint relationship between the projects. Another kind of deep capture technology in the sequence is recommended is a focus mechanism, which can solve the traditional sequence recommendation, and the CNN-based sequence recommendation exists. The case where there is noise in the sequence. Currently, there is a two-point question for the traditional sequence recommendation method and the depth capture-based sequence push method: 1) Most of the prior methods only consider the impact of the project ID, and other characteristics are ignored. Influence, such as user and project context information, that is, the user behavior context information or project context information. There is no relationship between the neighboring items, there is no relationship between the non-adjacent projects, and the above problem is solved. This paper proposes a mixed neural network model of parallel structure, and its parallel structure consists of convolutional neural network Composition with self-focus mechanism.

The main contributions are as follows: 1) This paper proposes a self-focus sequence recommendation model of fusion project and user context information, which includes the following modules, which are embedded in encoding layers, convolutional neural network modules, and self-focus mechanism modules. This paper implements simultaneous modeling user dynamic preferences and context information of user and project interaction, solving problems that currently have a current research method ignore the impact of context information characteristics. 2) Convolutional neural network modules combined with PaIRSE

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*March-April 2022 Page No. 686-697*
*Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022*

encoding methods in this method, This method pairs the user with any two items in the project interaction sequence to encode into sheets. Finally, the work is completed by the convolutional neural network. So this paper can learn between the projects of the interactive sequence, this Solving the existing research method does not build a problem of non-adjacent projects in mutual sequence. 3) This paper proposes that the SSRF model performs parameter analysis and ablation test on Alibaba and Retailrocket data set, SSRF model effectiveness and robustness is better than current Popular method.

## II. SELF-FOCUS SEQUENCE RECOMMENDATION MODEL FOR FUSION CONTEXT INFORMATION

### 2.1 Overall Architecture

The frame structure of this model method is shown in Figure 1. The item1, item2, ..., item represents the item embedded sequences for all of the user all history interactions, and context1, context2, ..., contextn is user interaction history. Project corresponds to projects and user context information. The model mainly includes the following components, namely embedded layer modules, convolutional neural network processing modules, self-focus mechanism modules, and all-connection layer modules. The flow of the entire model is as follows: 1) in the embedding layer the module encodes the user data into a vector, the user-project interaction history project data encodes the project sequence vector, encodes the interactive project type data into matrix. 2) Convolution Neural Network (CNN) is responsible for processing user embedding, user-project interaction Information, solving the short-term preference problem of the user. 3) The self-focus mechanism is responsible for handling the type sequence data of the context information in the user's interactive project, and the aggregation is obtained by the context information on preferences. 4) Combined with all preferences and context information rights Implement the user's next possible interaction item recommended in the interaction sequence.

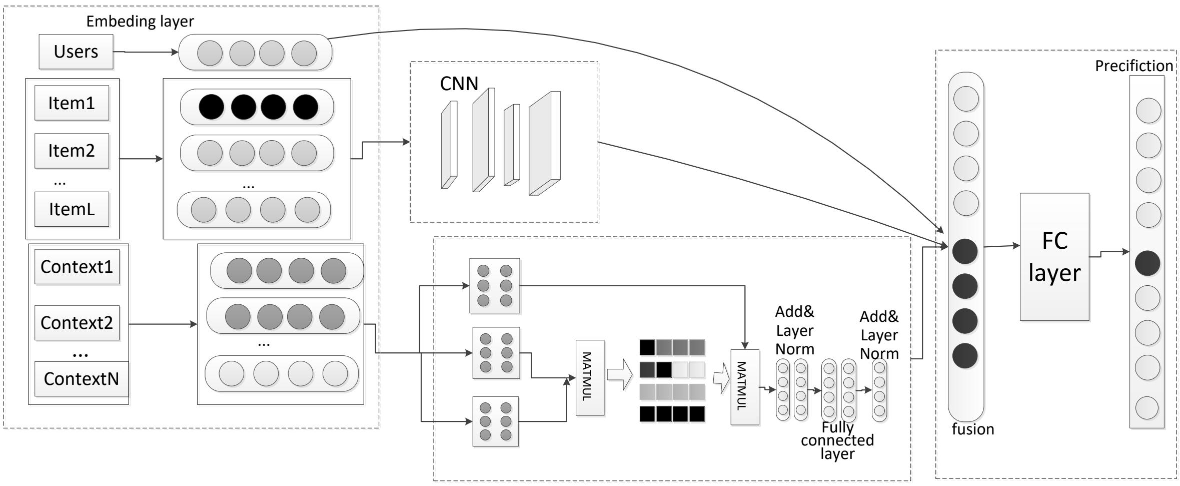


Fig 1: structure of our model

This article expresses $U$ to all user sets, $U = \{u_1, u_2, \ldots, u_{|U|}\}$, $|U|$ is the number of users. The collection

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 686-697
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

of all items uses $I$ represents, $I = \{I_1, I_2, \ldots, I_{|I|}\}$, the total number of items is $|I|$. The collection of users and project interactions $S$, $S = \{S_1, S_2, \ldots, S_{|S|}\}$, $S_i \in I$. This article The next interactive item will be predicted by training the user's interaction, predict the next interaction item of the user $U$. In this part, the coded layer, convolutional neural network module, self-focus network module, full connect layer, and predictive layer pair The method proposed in this paper is described in detail.

### 2.2 Embedding Encoding Layer

The field of computer vision and natural language processing has made great progress in convolutional neural network applications, so this article has also taken CNN to capture high-order sequence relationships that can be interactive with the project. First, the user $U$ interacts at time $t$. The embedded matrix $E_{(u,t)}^{\mathrm{L}} \in R^{L \times d}$ input to the CNN module. The CNN module is composed of four-layer convolution layer having a volume core of $1 \times 1$ and $3 \times 3$, and the convolution of $1 \times 1$ is to obtain a rich interaction characteristic representation, $3 \times 3$ Cemented nucleus is to learn to interact with complex user projects in order to learn high-order sequence interactions. After each layer is completed, it will be normalized, and then activated by activation functions, the activation function of the application is Relu. After all convolution layer convolution, through a full connection with a certain dropout. The final sequence feature sheet $T_{(u,t)}^{L} \in R^{d}$ is obtained.

### 2.3 Self Attention Network Module

Since the self-focus network module needs to process a sequence input matrix $E_c \in R^{n \times d}$ with time-to-time order relationship, the self-focus network is not like a circular neural network and a convolutional neural network to acquire location information, which is required. The self-focus network module plus the position encoding matrix $P \in R^{n \times d}$. The final self-focused network input matrix is as follows:

$$F_c = \begin{bmatrix} E_{cl} + P_I \\ E_{c2} + P_2 \\ \cdots \\ E_{cn} + P_n \end{bmatrix} \tag{1}$$

The formula for scaled dot-product attention is defined as follows:

$$SA(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{2}$$

In Formula (2), the query is represented by $Q$, the key represents $K$, the value is $V$, and $d$ represents the dimension of each feature. The function of scaling factor $\sqrt{d}$ is to make the dimension size of the

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 686-697
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

result obtained by the dot product moderate with the gradient value of softmax function, which can enhance the back propagation to a certain extent. $Q$, $K$, and $V$ in this paper are all equal to $Q$. After linear transformation, three mapping matrices are obtained, which are input into $SA$.

$$F_h = SA\left(F_c * W^Q, F_c * W^K, F_c * W^V\right) \tag{3}$$

This is the mapping matrix for $W^Q, W^K, W^V \in R^{d \times d}$. Because multiplex attention has the advantage of representing subspace information with different positions, it is more flexible and better than scaling dot product attention. Therefore, this paper applies multiple attention from the attention network module, and multiple attention is defined as follows:

$$M_h\left(F_c\right) = \text{Concat}\left(F_1^h, F_2^h, \ldots, F_{nh}^h\right) W^O \tag{4}$$

$$F_i^h = SA\left(F_c * W_i^Q, F_c * W_i^K, F_c * W_i^V\right) \tag{5}$$

In Formula (4) and (5), $W^O$, , $W_i^Q$ ,$W_i^K$ , $W_i^V$ is the parameter that the model needs to learn, and the hyperparameter $nh$ is the number of heads in the multi-attentional network. In addition, the self-attention network module uses Residual connection, Layer normalization, and full connection Layer with ReLu activation functions. The purpose of the method in this paper is to add residual connection to effectively prevent the problem of gradient disappearance, while the purpose of layer normalization is to make the data still normalized after the effect of the neural network layer, so that the model can avoid the problem of gradient disappearance and gradient explosion to a certain extent. The output definition of the self-attention network is as follows:

$$M_h = \text{LayerNorm}\left(M_h + F_c\right) \tag{6}$$

$$OT_f = \text{ReL}U\left(\left(M_h W_I + b_1\right) W_2 + b_2\right) \tag{7}$$

$$OT_f = \text{LayerNorm}\left(M_h + OT_f\right) \tag{8}$$

$W_*$, $b_*$ is the parameter of the model. For simplicity, self-attention network blocks are defined as follows:

$$OT_f = \text{SAB}\left(F_c\right) \tag{9}$$

After passing through the first self-attention network block, $OT_f$ aggregates all the characteristic information that user $U$ has previously interacted with. It will be more helpful to learn complex sequence transformation after another self-attention network block based on $F_c$, which is defined as follows:

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 686-697
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

$$OT_f^p = SAB\left(OT_f^{p-I}\right) \tag{10}$$

When $p=0$, $OT_f = F_c$.

2.4 Fully Connected Layer

The output of the convolutional neural network module and the output of the attention network module are spliced together at the full connection layer in order to simultaneously capture the item features, brand information, user interaction behavior and other contextual information features in the user interaction sequence.

$$O_{cf} = \left[T_{(u,t)}^L; O_f^p\right] W + b \tag{11}$$

The $W \in R^{3d \times d}$, $b \in R^d$ we embed user to $P_u$ with $O_{cf}$ is pieced together to $u$ for a long time user preferences, and then mapped to $|I|$ I a node in the output layer.

$$y^{(u,t)} = W' \begin{bmatrix} O_{cf} \\ P_u \end{bmatrix} + b' \tag{12}$$

Where $W' \in R^{|I| \times 2d}$ and $b' \in R^{|I|}$ are the weight matrix and bias term of the output layer respectively. Output layer $y^{(u,t)}$ represents the probability that user $u$ interacts with item $i$ at timestamp $t$.

The binary cross entropy loss function is adopted as the objective function:

$$-\sum_u \sum_t \log\left(s\left(y_{S_t'}^{(u,t)}\right)\right) + \sum_{j \notin S_\mu} \log\left(1 - s\left(y_j^{(u,t)}\right)\right) \tag{13}$$

In this paper, the Adam optimizer, a variant of adaptive stochastic gradient Descent (SGD), is used to optimize the network model. In each iteration, N negative samples (j) are randomly selected for the target term $S_t^u$.

## III. EXPERIMENT

3.1 Data Set and Data Preprocessing

The data set is alibaba user behavior data set and Retailroket e-commerce data set. Each experimental data set is composed of user ID, commodity ID, commodity category ID, behavior type, time stamp and other information. In order to verify the valid lines of the model, 85,000 and 31,000 user behavior records were extracted from the two data sets respectively, and the data were statistically and preprocessed. Python was used to clean the data to remove dirty data. The first 70% of each data set after processing was

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
March-April 2022 Page No. 686-697
Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022

selected for model training, and the last 30% was used for test and evaluation. The data statistics after processing are shown in Table I.

**TABLE I. Characteristic of dataset**

| Dataset | users | items | actions per user | actions per item |
|---|---|---|---|---|
| Alibaba | 21960 | 19897 | 15.1 | 16.6 |
| Retailrocket | 6040 | 3417 | 165.5 | 296.6 |

3.2 Evaluation Indices

When evaluating the performance of this model, the accuracy Pre@K, recall Rec@K, AUC and RelaImpr indexes are used to evaluate the performance. This is because the widely used indicator AUC (area under the ROC curve) is used for evaluation in the CTR prediction task, as shown in Formula (14). In addition, RelaImpr index was further introduced to measure the relative improvement compared with the basic model (FM) [9]. Since the AUC from the random strategy is 05, the RelaImpr in this task is formalized into formula (15). In addition, the estimated click-through rate can also be used to generate top-K recommendation list, so the accuracy rate Pre@K and recall rate Rec@K are used to evaluate the quality of recommendation list. The accuracy rate represents the proportion of correctly recommended items to all recommended items, while the recall rate represents the proportion of correctly recommended items to actual purchased items. The definition is shown in Formula (16) and Formula (17).

$$AUC = \frac{1}{|U|}\sum_{u}\frac{1}{|E(u)|}\sum_{(i,j)\in E(u)}\delta\left(\hat{x}_{ui} > \hat{x}_{uj}\right) \tag{14}$$

$E(u)=\{(i,j)|(u,i)\in S_{test} \wedge (u,j)\notin (S_{test}\cup S_{train})\}$, the higher THE AUC, the higher the ranking quality, the AUC of the random normal distribution AUC =0.5, and the upper limit of AUC is 1.

$$\mathrm{Re}la\,\mathrm{Im}\,pr = \frac{AUC(measured\ \mathrm{mod}\,el) - 0.5}{AUC(base\ \mathrm{mod}\,el) - 0.5} - 1 \tag{15}$$

$$\mathrm{Pr}\,e@N = \frac{|L_t \cap L_r|}{L_r} \tag{16}$$

$$\mathrm{Re}\,c@N = \frac{|L\_t \cap L\_r|}{L_t} \tag{17}$$

Where, $L_t$ is the item set actually browsed by users, and $L_r$ is the recommended result set. In general, the higher the accuracy rate and recall rate, the better the performance of the recommendation algorithm.

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*March-April 2022 Page No. 686-697*
*Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022*

### 3.3 Compare the Model

In order to verify the validity of this model, this paper adopts comparative experiment to compare the model with the following model.

(1) The decomposing machine model proposed by FM, Song et al. [10] can model second-order feature interaction for CTR, which is the basic model for evaluation.

(2)Wide&DEEP, proposed by Zhou et al. [11]. It is composed of the Wide part that processes the original features and the Deep part that extracts the interaction of higher-order features.

(3)NFM, Feng et al. [12] used a dual interaction layer to carry out feature interaction before DNN layer.

(4)DEEP FM, Sun et al. [13]. DeepFM replaces the Wide component in Wide&Deep with an FM layer.

(5)DCN, Amir et al. [14]: DCN captures the interaction of bounded degree features with their crossover networks.

(6)DIN, Zhang et al [15]. DIN is the classic model for session-based recommendations. It takes into account the weight of items in the user's historical behavior.

(7)AutoInt, Santiago et al. [16]. AutoInt introduces a self-attentional neural network for feature interaction.

### 3.4 Experimental Environment and Setup

Experimental environment: the processor is AMDRyzen5 4600HwithRaden Graphics 3.00ghz, the memory is 16GB, and the system is windows10 X64 bit. Deep capture frameworks: python3.0+Anacanda, Tensor Flow, Numpy. The maximum length of all three behavior sequences in the SPRIF model is 30, and the number of feature fields is 47. The dimension $n_h$=64 of each feature is embedded, and the dimensions of the 2-layer MLP in Deep are 32 and 16, respectively. Weight of click, unclick and delete shopping cart loss $\lambda_c$: $\lambda_d$: $\lambda_u$ = 1:1:10.

### 3.5 Click Rate Forecast

SSRF was first evaluated on classical CTR predictions to verify its ability to model positive user preferences. CTR prediction results on the two datasets are shown in Table II.

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*March-April 2022 Page No. 686-697*
*Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022*

**TABLE II. Experimental results on Alibaba and Retailrocket datasets**

| model | Alibaba | | Retailrocket | |
|---|---|---|---|---|
| | AUC | Relal-mpr | AUC | Relal-mpr |
| FM | 0.698 | 38.91% | 0.728 | 41.35% |
| Wide& DEEP | 0.680 | 27.89% | 0.715 | 30.73% |
| NFM | 0.678 | 52.34% | 0.703 | 56.84% |
| DEEP FM | 0.704 | 69.72% | 0.732 | 70.08% |
| DCN | 0.697 | 74.78% | 0.726 | 76.23% |
| DIN | 0.676 | 85.23% | 0.692 | 86.19% |
| AutoInt | 0.714 | 87.52% | 0.758 | 88.57% |
| SSRF | 0.880 | 92.22% | 0.912 | 94.51% |

The experimental results of this model and the benchmark model to be compared have been recorded in Table 2. According to the data in Table 2, it can be clearly found that the AUC and Relalmpr of AutoInt model are better than that of DIN model in Alibaba data set. The model we proposed achieves the best performance in all the above evaluation indexes on Alibaba. The AUC of the model in this paper was improved by at least 23.249%, and Relalmpr was improved by at least 5.37%. The data of N in Retailrocket data set showed that THE SSRF model was also improved compared with the baseline model in various evaluation indicators.

3.6 Accuracy and Recall Tests

The accuracy and recall experiments were carried out on two datasets to compare the performance of various recommendation algorithms with different loss weights (used to measure the influence of dislike). The horizontal axis represents the value of loss weight, and the vertical axis represents the accuracy rate and recall rate respectively.

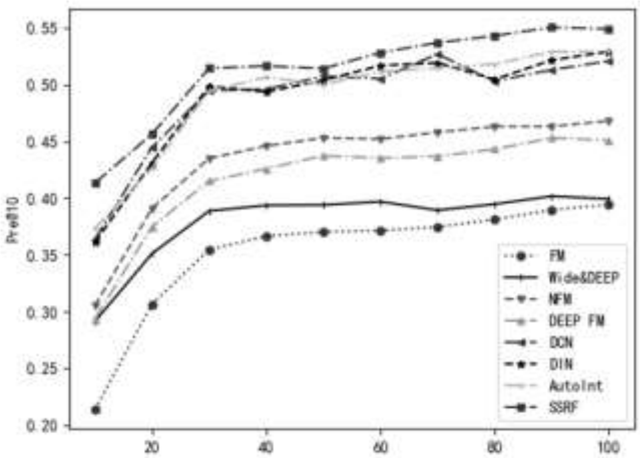


Fig 2: comparison of accuracy of various algorithms with different loss weights

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*March-April 2022 Page No. 686-697*
*Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022*

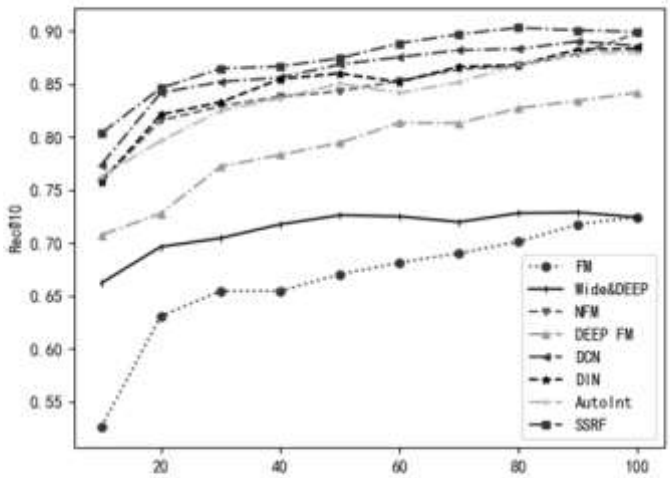


Fig 3: comparison of accuracy of various algorithms with different loss weights

It can be seen from FIG. 2 and 3 that, compared with other baseline recommendation algorithms, the accuracy and recall rate of the algorithm proposed in this paper are significantly improved after calculation combined with user interest feedback, and the changes of algorithm accuracy and recall rate tend to be stable as the value increases, indicating the effectiveness of the proposed method. Among them, the performance of FM and Wide&DEEP models is low, which is because the models tend to extract low-order or high-order combination features and cannot extract these two types of features at the same time. The proposed method uses the hybrid neural network structure of convolutional neural network and self-attention mechanism to simultaneously model the dynamic preferences of users and the context information of user interaction with the project, and divides user interests into more fine-grained ones, which is helpful to improve the recommendation performance.

3.7 Model Variation Experiment

In order to prove the effectiveness and necessity of different components in the user interest feedback module, two variant models were set up for experiment: SSRF-UC and SSRF-O. The SSRF-UC model does not consider the influence of the characteristics of context information, and the SSRF-O model does not adopt the pairse coding method and adopts the traditional ONE-HOT vector coding method. Compared with SSRF, the prediction results of CTR are shown in Figure 4 and the prediction results of rec@10 are shown in Figure 5.

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*March-April 2022 Page No. 686-697*
*Article History: Received: 08 February 2022, Revised: 10 March 2022, Accepted: 02 April 2022, Publication: 30 April 2022*

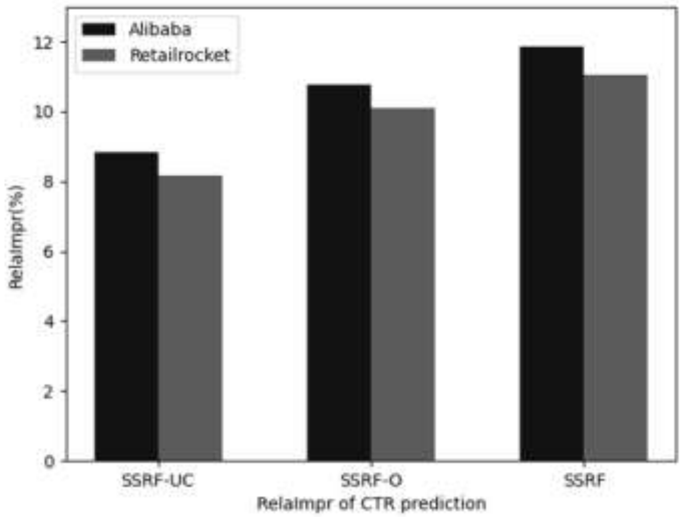


Fig 4: Comparison of Relalmpr index in CTR prediction experiment of model variation

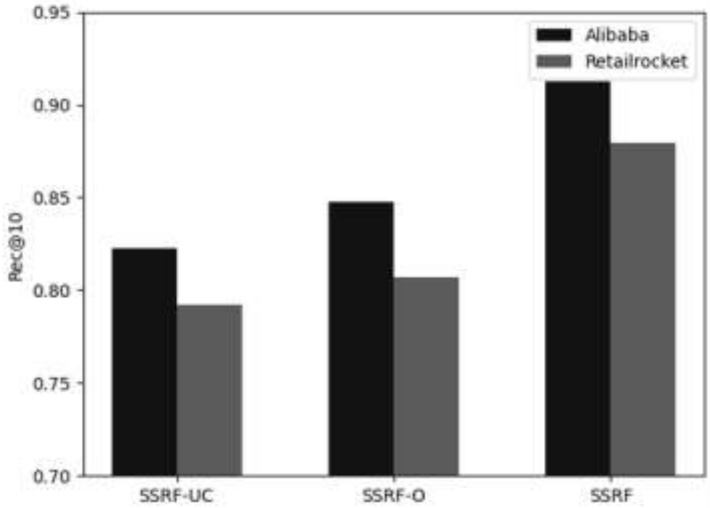


Fig 5: Experimental comparison of model variation accuracy

The experimental results show that: (1) SSRF performs better than ssrf-uc, which confirms the importance of context information features. (2) The performance from SSRF-O to SSRF was significantly improved, indicating that the use of PAIRse to encode more effective capture interaction sequence between items is helpful to understand the user's justice preference and improve the model's recommendation performance.

## V. CONCLUSION

In this paper, a hybrid neural network recommendation model combining item and user context information is proposed. In this model, convolutional neural network module is used to extract the sequence features of user-item interaction, and the context information features of user sequence are modeled by self-attention mechanism. Experiments on two data sets closely related to real life prove that the proposed model method can further improve the recommendation performance than the benchmark model and achieve the best recommendation performance. In the following work, we will further study the influence of the combination of item and user context information on sequence recommendation and the long-term preference of users.

## REFERENCES

[1] Qian Zhao, F Maxwell Harper, Gedimin as Adomavicius, and Joseph A Konstan. (2018) Explicit or implicit feedback? engagement or satisfaction? : a field experiment on machine-learning-based recommender systems. In Proceedings of SAC.

[2] Ruoxi Wang, Bin Fu, Gang Fu, and Mingliang Wang. Deep&cross network for ad click predictions. In Proceedings of ADKDD, 2017.

[3] Jian Liu, Chuan Shi, Binbin Hu, Shenghua Liu, and S Yu Philip. (2017) Personalized ranking recommendation via integrating multiple feedbacks. In PAKDD.

[4] Badrul Munir Sarwar, George Karypis, Joseph A Konstan, John Riedl. (2001) Item-based collaborative filtering recommendation algorithms. In WWW.

[5] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. Wide&deep learning for recommender systems.

[6] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. (2017) Deepfm: a factorization-machine based neural network for ctr prediction. In Proceedings of IJCAI.

[7] Xiangnan He and Tat-Seng Chua. (2017) Neural factorization machines for sparse predictive analytics.In Proceedings of SIGIR.

[8] Jun Xiao, Hao Ye, Xiangnan He, Han wang Zhang, Fei Wu, and Tat-Seng Chua. (2017) Attentional factorization machines: Learning the weight of feature interactions via attention networks.In IJCAI.

[9] Menghan Wang, Mingming Gong, Xiaolin Zheng, and Kun Zhang. (2018) Modeling dynamic miss ingness of implicit feedback for recommendation. In Proceedings of NIPS.

[10] Weiping Song, Chence Shi, Zhiping Xiao, Zhijian Duan, Yewen Xu, Ming Zhang, and Jian Tang. (2019) Autoint: Automatic feature interaction learning via selfattentive neural networks. In Proceedings of CIKM.

[11] Chang Zhou, Jinze Bai, Junshuai Song, Xiaofei Liu, Zhengchao Zhao, Xiusi Chen, and Jun Gao. (2018) Atrank: An attention-based user behavior modeling framework for recommendation.In Proceedings of AAAI.

[12] Yufei Feng, Fuyu Lv, Weichen Shen, Menghan Wang, Fei Sun, Yu Zhu, and Keping Yang. (2019) Deep session interest network for click-through rate prediction. In Proceedings of IJCAI.

[13] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. (2019) Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer.In Proceedings of CIKM.

[14] Amir H Jadidinejad, Craig Macdonald, and IadhOunis. (2019) Unifying explicit and implicit feedback for rating prediction and ranking recommendation tasks. In Proceedings of ICTIR.

[15] Quangui Zhang, Longbing Cao, Chengzhang Zhu, Zhiqiang Li, and Jinguang Sun. (2018) Cou pledcf:Learning explicit and implicit user-item couplings in recommendation for deep collaborative filtering. In Proceedings of IJCAI.

[16] Santiago Hors-Fraile et al. (2018) Analyzing recommender systems for health promotion using a multidisciplinary taxonomy: A scoping review. International Journal of Medical Informatics, 114: 143-155.