Knowledge and Information Systems (2023) 65:3337–3352
https://doi.org/10.1007/s10115-023-01867-w

REGULAR PAPER

A new interest extraction method based on multi-head attention mechanism for CTR prediction

Haifeng Yang[1] · Linjing Yao[1] · Jianghui Cai[1,2] · Yupeng Wang[1] · Xujun Zhao[1]

Received: 25 November 2022 / Revised: 21 January 2023 / Accepted: 14 March 2023 /
Published online: 5 April 2023

Abstract
Click-through rate (CTR) prediction plays a vital role in recommendation systems. Most models pay little attention to the relationship between target items in the user behavior sequence. The attention units used in these models cannot fully capture the context information, which can be used to reflect the variations of user interests. To address these problems, we propose a new model named interest extraction method based on multi-head attention mechanism (IEN) for CTR prediction. Specifically, we design an interest extraction module, which consists of two sub-modules: the item representation module (IRM) and the context–item interaction module (CIM). In IRM, we learn the relationship between target items in the user behavior sequence by a multi-head attention mechanism. Then, the user representation is gained by integrating the refined item representation and position information. At last, the correlation between the user and the target item is used to reflect user interests. In CIM, the context information has valuable temporal features which can reflect the variations of user interests. Therefore, user interests can be further acquired through the feature interaction between the context and the target item. After that, the learned relevance and the feature interaction are fed to the multi-layer perceptron (MLP) for prediction. Besides, experiments on four Amazon datasets were conducted to evaluate the effectiveness of our method in capturing user interests. The experimental results show that our proposed method outperforms state-of-the-art methods in terms of AUC and RI in the CTR prediction task.

✉ Jianghui Cai
Jianghui@tyust.edu.cn

Haifeng Yang
hfyang@tyust.edu.cn

Linjing Yao
s20202011059@stu.tyust.edu.cn

Yupeng Wang
yupengwang@tyust.edu.cn

Xujun Zhao
zxj0226@126.com

1 School of Computer Science and Technology, Taiyuan University of Science and Technology, Waliu road, Taiyuan 030024, China

2 School of Computer Science and Technology, North University of China, Xueyuan road, Taiyuan 030051, China

Keywords Recommendation system · Multi-head attention · Feature interaction · Click-through rate prediction

1 Introduction

In the past decades, with the rapid growth of information, recommendation systems [1] have played a key role in numerous domains, such as news [2], e-commerce [3–5], and online advertising [6, 7]. Inspired by the successful applications of deep learning in computer vision [8] and natural language processing [9, 10], deep neural network-based approaches have also been extended to the field of recommendation systems.

Deep neural network-based models have been proposed to learn feature interactions and the representative models include Wide & Deep, PNN, DIN, and other models. Wide & Deep [11] combines a linear model and a nonlinear model to learn both low-order and high-order feature interactions, but the linear part still relies on manual learning which leads to poor model performance. PNN [12] is proposed to better capture feature interactions by the product layer. Since user behavior sequence is important for mining user interests, models such as DIN [13], DIEN [14], and DMR [15] obtain user representation from the user's historical behavior to reflect user interests. On the one hand, they pay little attention to the relationship between target items in the user behavior sequence when these models learn user representation. On the other hand, when using context information to reflect the variations of user interests, the attention units used in these models are hard to express the diversity of user preferences. As a result, these models fail to obtain the real interests of the user, which in turn makes the CTR prediction results inaccurate.

In order to extract user's real interests more accurately, a new model named interest extraction method based on multi-head attention mechanism (IEN) is proposed in this paper. Specifically, we design an interest extraction module which consists of two sub-modules: the item representation module (IRM) and the context–item interaction module (CIM). In the IRM, the relationship between target items in the user behavior sequence is learned by using the multi-head attention mechanism which helps to obtain refined item representations. Then, by integrating refined item representations and position information, user representation is gained. At last, the correlation between the user and the target item is calculated by the inner product. In the CIM, a multi-head attention mechanism is used to learn feature interaction between the context and the target item to further get the user interest.

In summary, the main contributions of this paper are summarized as: We propose a new model named interest extraction method based on multi-head attention mechanism (IEN) to capture user interests in this paper. On the one hand, item representation module (IRM) is introduced to learn the relationship between target items in the user behavior sequence and the refined item representation is acquired. On the other hand, the context–item interaction module (CIM) is designed to capture the feature interaction between the context and the target item by utilizing the multi-head attention mechanism.

The rest of the paper is organized as follows. Section 2 discusses related work. Section 3 details our model structure. We perform experimental validation and analysis in Sect. 4. Finally, in Sect. 5, we summarize our work and point out the direction of future work.

2 Related work

The previous CTR prediction models are based on LR and its variants [16]. LR is a linear model and lacks the ability to learn complex feature interactions. This makes its feature representation generally weak. To overcome this drawback, the FM [17] model is proposed, which can learn second-order feature interactions. Meanwhile, FFM [18] and FWFM [19] are proposed based on FM. FFM introduces the concept of "field" into feature interactions, while FWFM learns different feature interactions by the inner product of embedding vectors and field weights. However, these methods cannot fully apply the data features which can be grouped by certain rules in real scenarios. Since clustering [20–23] and classification [24] can divide the features, they are applied to solve the above problem in CTR prediction. In addition, NFM [25] combines FM and neural networks to improve the model's performance. Learning the interactions of second-order features can improve the performance of the model, but some redundant feature interactions may lead to noises. So, AFM [26] is proposed to learn feature interactions by attention mechanism.

Recently, recommendation models based on deep neural networks (DNN) have received much attention and achieved remarkable results. Among them, Wide & Deep [11] combines linear and nonlinear deep models to learn low-order and high-order feature interactions, but the linear part still relies on manual learning and results in inferior model performance. To address this problem, DeepFM [27] combines the power of DNN and FM for feature representation. In addition, DCN [28] applies a novel cross-network to automatically learn high-order features. PNN [12] is proposed to better capture high-order feature interactions by the product layer. Besides, xDeepFM [29] is proposed to learn feature interactions by compressed interaction network. In general, the above deep models improve recommendation performance by capturing low-order or high-order features.

Because the user's historical behavior contains items viewed by the user, it is crucial to capture user interests. DIN [13] is proposed to learn user interests by activation unit, but DIN rarely considers the changing trend of interest. Thus, DIEN [14] learns the dependencies between sequential behaviors by designing an auxiliary loss and AUGRU. And the model not only extracts user interests but also captures the temporal evolution of user interests. Meanwhile, BST [30] is used to capture the neglected sequential nature in DIN by utilizing a transformer. DSIN [31] learns user interests by using session information from the user behavior sequence. DMR [15] is proposed to obtain the relevance between the user and the target item through user-to-item network and item-to-item network. DMIN [32] learns potential multiple user interests from the user behavior sequence. MIAN [33] extracts feature interactions between multi-field features by utilizing a multi-interaction layer and a global interaction module.

Referring to the literature [34], we know that the above methods focus on learning user representations in the user behavior sequence but they cannot fully learning the contextual information. In addition, we agree with the literature [35] that learning refined item representations is important for learning user representations. We propose a model called interest extraction method based on multi-head attention mechanism (IEN) in this paper. It can better learn context information and the refined representation of items. User interests can be captured more precisely, which is helpful to improve the accuracy of CTR prediction.

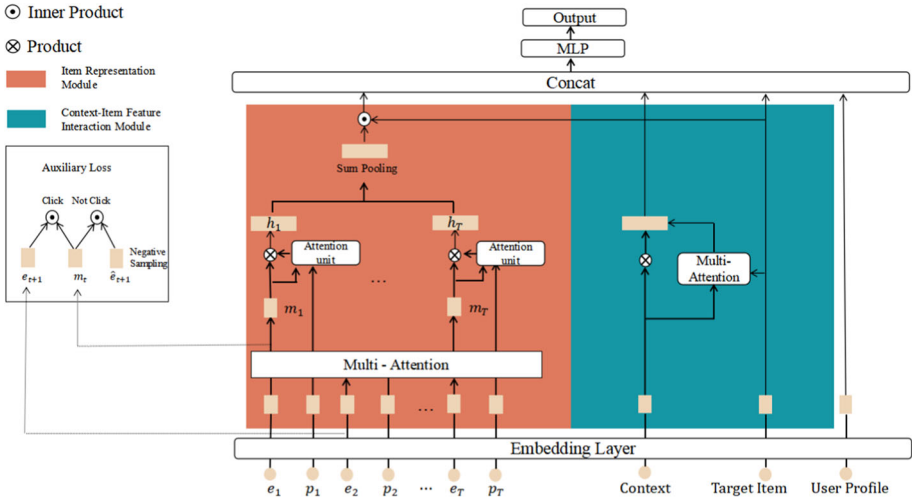


Fig. 1 Interest extraction method model

3 Interest extraction method based on multi-head attention mechanism

Since the user behavior sequence is a list of items visited by the user, the user behavior can reflect the user's interests, and extracting good user representation from the user behavior sequence is beneficial to acquire user interests. In addition, relying only on the user behavior sequence may lead to outdated recommendations. Moreover, valuable features in context information (e.g., review time, ratings) are meaningful for deriving the user's current interests. So, learning the feature interaction between the context and the target item is advantageous for obtaining user interests.

The multi-head attention mechanism [36] can learn the relationship between different features. Thus, the relationship between target items in the user behavior sequence and feature interaction between the context and the target item can be learned by the multi-head attention mechanism. Therefore, a new model called interest extraction method based on multi-head attention mechanism (IEN) is designed in this section. The framework of IEN is shown in Fig. 1, and we design an interest extraction module, which is composed of two sub-modules: the item representation module (IRM) and the context–item interaction module (CIM).

In the IRM, to capture user representation, we integrate position information with refined item representations learned by a multi-head attention mechanism. After that, the relevance between the user and the target item is derived by the inner product. In the CIM, the feature interaction between the context and the target item is learned via the multi-head attention mechanism.

3.1 Item representation module

Four types of features are used in our model processing: *User Profile*, *Target Item*, *User Behavior Sequence*, and *Context*. User Profile (x_p) contains features related to the user, such as user ID, consumption level, and gender. The Target Item (x_t) contains item ID, category ID, etc. And the item ID can be expressed as an embedding matrix $V = [v_1; v_2; \ldots; v_K] \in R^{K \times d_v}$,

where K is the total number of items and d_v is the embedding dimension of the jth item v_j. The User Behavior Sequence contains multiple items and can be denoted as $x_b = [e_1; e_2; \ldots; e_T] \in R^{T \times d_e}$, where d_e is the embedding dimension of the tth item e_t and T is the length of user behavior sequence. Context (x_c) includes the time, the method of matching, the corresponding match score, and other valuable information. As there is sequential information in the user behavior sequence, the position information $[p_1; p_2; \ldots; p_T] \in R^{T \times d_p}$ is represented to capture the sequential information in the user behavior sequence, where d_p is the dimension of the tth position.

It is known that user's historical behavior contains some items visited by the user, so user interests denoted by the user representation can be extracted from user behaviors. Thus, to better extract user interests, we require a good method to capture the user representation from the user behavior sequence.

The item representation module (IRM) is proposed to learn the user interests implied the user behavior sequence. In the IRM, the refined representation of each item is acquired via a multi-head attention mechanism. After that, the user representation is obtained by integrating refined item representations and position information. Lastly, the inner product is used to get the correlation between the user and the target item to denote the user's interest in the target item. The input of the multi-head attention mechanism used in this paper consists of the query (Q), key (K), and value (V). The specific calculation equations are as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^{\mathrm{T}}}{\sqrt{d}})V \tag{1}$$

$$\begin{aligned} \text{head}_h &= \text{Attention}(x_b W_h^Q, x_b W_h^K, x_b W_h^V) \\ &= \text{softmax}(\frac{x_b W_h^Q \cdot (x_b W_h^K)^{\mathrm{T}}}{\sqrt{d_k}}) \cdot x_b W_h^V \end{aligned} \tag{2}$$

$$M = \text{MultiHead}(x_b) = \text{Concat}(\text{head}_1, \text{head}_2, \ldots, \text{head}_H) W^O \tag{3}$$

where $W_h^Q, W_h^K, W_h^V \in R^{d \times d}$ is the weight matrix, d and d_k are the scale factors, W^O is the linear matrix, and H is the number of the head. The output of each head is concatenated to get the high-order item representations, denoted as $M = (m_1; m_2; \ldots; m_t; \ldots; m_T)^{\mathrm{T}}$, where m_t is the tth item.

In addition, this paper will use auxiliary loss [14] to supervise the learning of the refined representation of each item. The auxiliary loss uses the $(t+1)$th item e_{t+1}^i to supervise the learning of the tth item representation. The real next action of the user is used as the positive sample, while the negative sample is sampled from the set of items that were not clicked. The formulation of auxiliary loss is expressed as,

$$L_{\text{aux}} = -\frac{1}{N}(\sum_{i=1}^{N}\sum_{t}(\log\sigma(m_t^i, e_{t+1}^i) + \log(1 - \sigma(m_t^i, e\hat{Z}_{t+1}^i)))) \tag{4}$$

where N is the number of training samples.

The user representation can be learned by integrating the refined item representation m_t and position information p_t. The calculation formulas are expressed as,

$$\alpha_t = \frac{\exp(\tanh(W_p p_t + W_m m_t + b))}{\sum_{t=1}^{T}\exp((\tanh(W_p p_t + W_m m_t + b))} \tag{5}$$

$$u = \sum_{t=1}^{T}(\alpha_t m_t) = \sum_{t=1}^{T}(h_t) \tag{6}$$

where $p_t \in R^{d_p}$ is the tth position embedding and $W_p \in R^{d_h \times d_p}$, $W_m \in R^{d_h \times d_m}$, $b \in R^{d_h}$ are learning parameters. And α_t is the normalized weight of the tth item. The feature vector of the user behavior sequence can be mapped into a fixed-length feature vector $u \in R^{d_v}$ by weighted sum pooling.

Finally, the item representation $v \in R^{d_v}$ of the target item is queried by the embedding matrix. After that, the relevance between the user and the target item is gained by the inner product: $r = u^{\mathrm{T}} v$.

3.2 Context–item interaction module

The valuable temporal features in the context information are critical to derive the user's current interests. However, the IRM focuses on learning the relationship between the user behavior sequence and the target item. It does not sufficiently learn the context information and thus lacks learning of the user's current interests. Thus, IRM may lead to outdated recommendations. Therefore, we propose the context–item interaction module (CIM). In the CIM, we learn the feature interaction between the context and the target item by applying the multi-head attention mechanism. In this way, the interests of the user are gained. The specific steps of this module are as follows. Firstly, the context representation x_c is concatenated with the target item representation x_t and $Z = Concat(x_t, x_c)$ is gained. Secondly, the feature interaction (R) between the context and the target item is learned using Eqs. 7–8.

$$\begin{aligned} \mathrm{head}'_h &= \mathrm{Attention}(ZW'^{Q}_h, ZW'^{K}_h, ZW'^{V}_h) \\ &= \mathrm{softmax}\left(\frac{ZW'^{Q}_h \cdot (ZW'^{K}_h)^{\mathrm{T}}}{\sqrt{d_k}}\right) \cdot ZW'^{V}_h \end{aligned} \tag{7}$$

$$R = \mathrm{MultiHead}(Z) = \mathrm{Concat}(\mathrm{head}'_1, \mathrm{head}'_2, \ldots, \mathrm{head}'_H) W'^{O} \tag{8}$$

3.3 The overall structure of interest extraction method

The interest extraction method based on multi-head attention mechanism (IEN) mainly consists of the embedding layer, the interest extraction module, and the multi-layer perceptron. The specific structure is shown in Fig. 1. Among them, the interest extraction module mainly includes IRM and CIM.

Most features can be encoded as high-dimensional one-hot vectors. To begin with, in the embedding layer, the one-hot vectors are transformed into low-dimensional dense features. After that, in the interest extraction module, IRM and CIM can sufficiently capture the user interests in the target item. Finally, the relevance between the user and the target item, feature interaction, the user profile, context information, user behavior sequence, and target item are concatenated together and fed into the MLP for the final CTR prediction. Since CTR prediction is a binary classification task, the widely used cross-entropy loss function is chosen for the loss function. It uses the label of the target item to supervise the whole prediction. So, the cross-entropy loss function is defined as,

$$L_{\mathrm{target}} = -\frac{1}{N} \sum_{(x,y)\in D}^{N} (y \log f(x) + (1-y)\log(1-f(x))) \tag{9}$$

Table 1 Datasets used in this paper

Dataset	Users	Items	Categories
Electronics	192,403	63,001	801
Beauty	290,703	199,829	248
CDs & Vinyl	418,950	436,599	658
Book	603,668	367,982	1600

where $x = [x_p, x_t, x_b, x_c]$, D is the training set with the total number of N, $y = \{0, 1\}$ indicates whether the user clicked the target item, and $f(x)$ is the prediction result of the MLP output.

In this paper, we uses cross-entropy loss and auxiliary loss to supervise the overall prediction and the extraction of refined item representations, respectively. So, the final prediction loss function is defined as,

$$L_{\text{final}} = L_{\text{target}} + \beta L_{\text{aux}} \tag{10}$$

where β is a hyperparameter that balances the losses of the two parts.

4 Experiments

In this section, we conduct comparison experiments on four Amazon datasets with the proposed method IEN and seven existing popular algorithms, and the experimental results are analyzed and evaluated. In addition, we validate the effectiveness of each part in the IEN model with the ablation experiments.

4.1 Datasets

The four real datasets derived from the Amazon datasets[1]. are used to evaluate the model's performance. All Amazon datasets contain abundant user behavior, user profile, and context information. We select *Electronics*, *Beauty*, *CDS&Vinyl*, and *Book* datasets which are already extensively used in CTR prediction task in the Amazon datasets. We get training set and testing set by random sampling from the original dataset with split rate of 80% and 20%, respectively. Table 1lists the statistics of the four datasets.

4.2 The comparison models

To evaluate the performance of the IEN model, we compare it with the most popular methods based on deep neural network frameworks. The methods include DNN, Wide & Deep, PNN, DIN, DIEN, DMR, and DMIN. Among them, DIN, DIEN, DMR, and DMIN acquire user interests from the user behavior sequence.

DNN [37]: DNN is a standard deep neural network, which consists of the embedding layer and MLP. It is also a prototype for other DNN-based models for CTR prediction task.

Wide&Deep [11]: This method combines linear and nonlinear models to learn feature interactions and further capture user interests.

[1] http://jmcauley.ucsd.edu/data/amazon/.

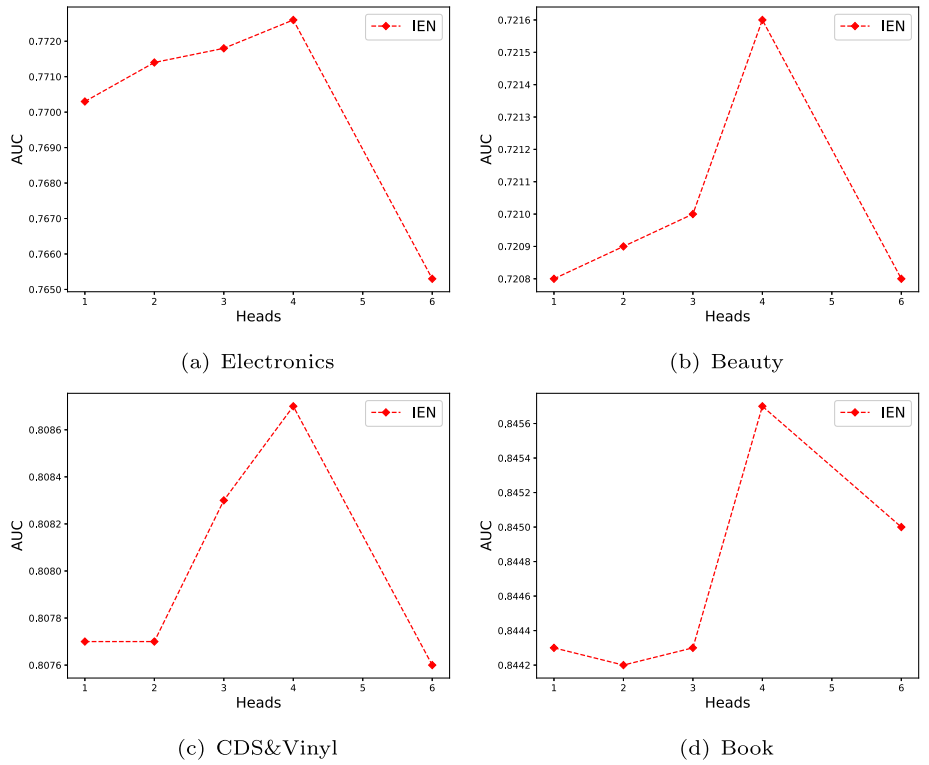


Fig. 2 Experimental results for different numbers of heads in the attention mechanism

PNN [12]: This method introduces the product layer to learn high-order feature interactions.

DIN [13]: DIN uses attention units to learn user interests by combining attention mechanisms with DNN.

DIEN [14]: In this method, the disadvantages of DIN are solved by applying the auxiliary loss and AUGRU to get the evolving process of user interests.

DMR [15]: This method learns the relevance between the user and the target item by designing a user-to-item network and an item-to-item network.

DMIN [32]: DMIN utilizes the behavior refiner layer and the multiple interest extraction layer to learn the multiple interests of the user.

4.3 Experimental setups

In this subsection, the parameters in our model are set the same as the ones in the references [14, 15]. The learning rate is set to 0.001, the batch size is set to 256, and the weight of the auxiliary loss is set to 1.

To check the effects of heads number in the multi-head attention mechanism on the model's performance, an experiment was conducted. The results are displayed in Fig. 2. It is interesting to notice that the model performs best when H is 4 on all datasets. So, H in IEN is set to 4, and the setting of H in the comparison algorithms refers to the literature.

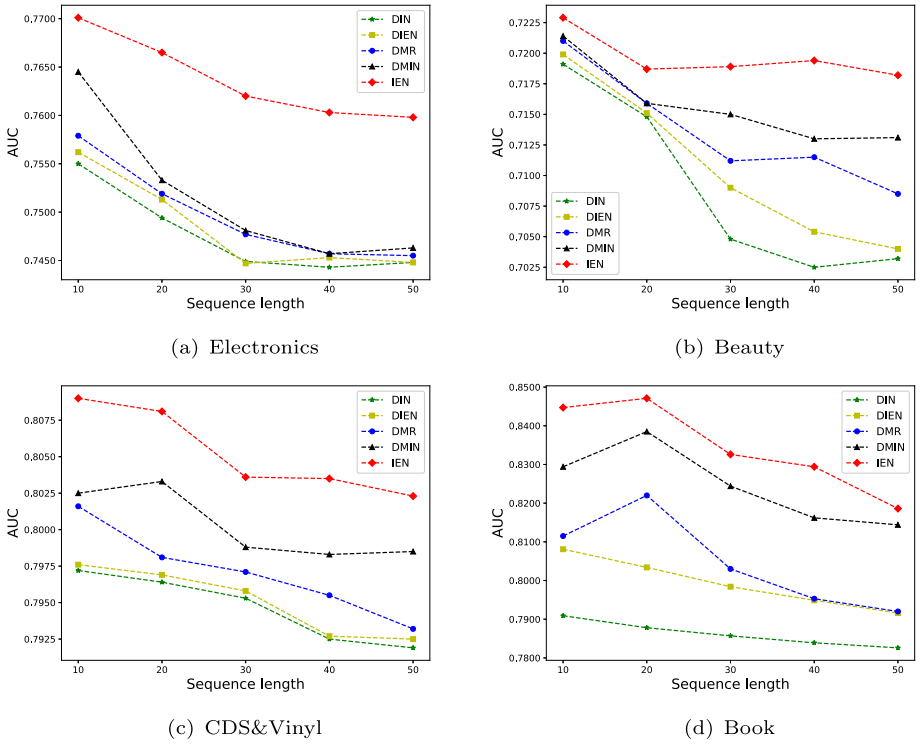


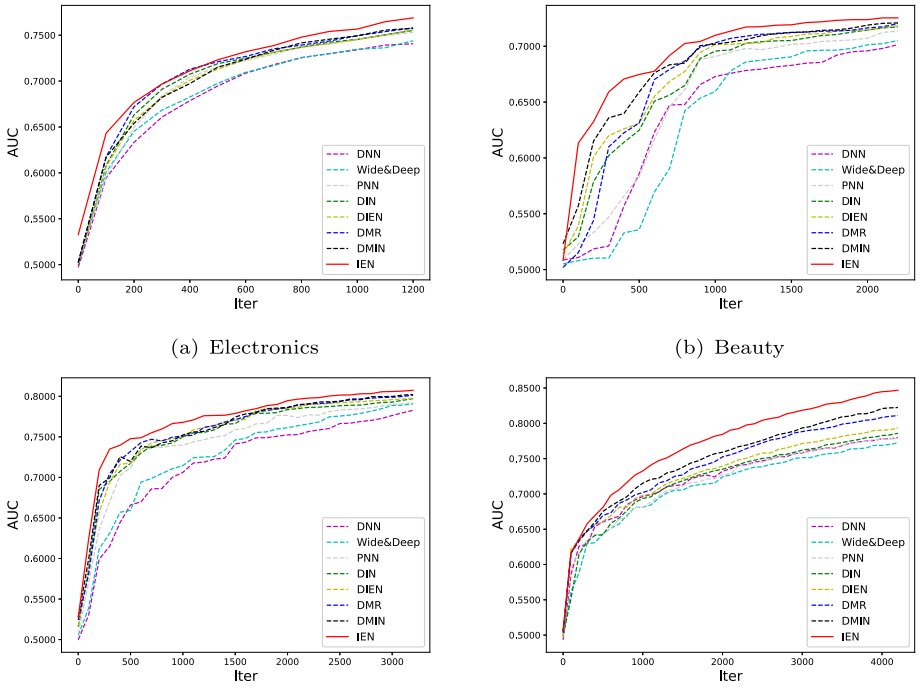
Fig. 3 Results of the effect of user behavior sequence length on model's performance

Furthermore, since the user behavior sequence contains valuable information, choosing the appropriate length of the behavior sequence is critical to the model's performance. We compare the IEN with models that make predictions based on the user behavior sequence, such as DIN, DIEN, DMR, and DMIN. A quick look from Fig. 3, we can see the impact of different user behavior sequence lengths on the model's performance. It is clear that IEN shows the best effect when the sequence length is set to 10 on the *Electronic*, *Beauty*, *CDS&Vinyl* datasets. And on the *Book* dataset, IEN performs the best when the sequence length is set to 20. In addition, DIN and DIEN show the best performance on the four datasets with sequence length set to 10. Additionally, DMR fares the best when the sequence length is 10 on the *Electronics*, *Beauty*, *CDS&Vinyl* datasets and 20 on the *Book* dataset. The model performs well when DMIN is set to 10 on the top two datasets and 20 on the bottom two datasets.

Finally, we use Adam [38] as the optimizer of IEN. Besides, we use the area under the ROC curve (AUC) [39] and DNN-based RelaImpr (RI) [40] as evaluation indexes. The experiments are repeated five times, and the average results are recorded.

4.4 Experimental results

In this subsection, to demonstrate the validity of IEN, we analyze the experimental results of IEN and comparison algorithms (DNN [37], Wide & Deep [11], PNN [12], DIN [13], DIEN [14], DMR [15], DMIN [32]) on four datasets *Electronics*, *Beauty*, *CDS&Vinyl*, *Book*, where



(a) Electronics (b) Beauty

(c) CDS&Vinyl (d) Book

Fig. 4 Training curves on the four datasets

Table 2 AUC experimental results on four datasets

Method	Electronics	Beauty	CDS & Vinyl	Book
DNN	0.7393 ± 0.0029	0.7047 ± 0.0048	0.7853 ± 0.0025	0.7818 ± 0.0023
Wide & Deep	0.7439 ± 0.0011	0.7071 ± 0.0059	0.7917 ± 0.0010	0.7847 ± 0.0028
PNN	0.7526 ± 0.0016	0.7149 ± 0.0012	0.7939 ± 0.0009	0.7848 ± 0.0028
DIN	0.7550 ± 0.0005	0.7191 ± 0.0011	0.7972 ± 0.0016	0.7909 ± 0.0018
DIEN	0.7562 ± 0.0013	0.7199 ± 0.0011	0.7976 ± 0.0011	0.8081 ± 0.0035
DMR	0.7579 ± 0.0012	0.7210 ± 0.0009	0.8016 ± 0.0012	0.8220 ± 0.0009
DMIN	0.7645 ± 0.0015	0.7214 ± 0.0017	0.8025 ± 0.0025	0.8335 ± 0.0017
IEN	0.7701 ± 0.0005	0.7241 ± 0.0010	0.7701 ± 0.0010	0.8471 ± 0.0011

DNN is used as the baseline. Figure 4 presents the training process of all algorithms on four datasets, where the red line represents IEN. Taking a close look at the results, the IEN model far exceeds other algorithms on four datasets. Besides, our method performs best on the *Book* dataset.

The comparison experimental results on the four datasets are given in Table 2, where the best experimental results are indicated in blue. It is observed that IEN beats the other models on all metrics on the four datasets. Concretely, the AUC values of IEN proposed in this paper are 0.7701, 0.7241, 0.8090, and 0.8471 on the four datasets. Observing Table 3, the AUC

Table 3 RI based on DNN for all models on each dataset

Method	Electronics (%)	Beauty (%)	CDS & Vinyl (%)	Book (%)
DNN	0.00	0.00	0.00	0.00
Wide & Deep	1.92	1.17	2.24	1.03
PNN	5.56	4.98	3.01	1.06
DIN	6.56	7.03	4.17	3.23
DIEN	7.06	7.43	4.31	9.33
DMR	7.77	7.96	5.71	14.27
DMIN	10.53	8.16	6.02	18.35
IEN	12.87	9.48	8.31	23.17

of IEN increased by 12.87%, 9.48%, 8.31%, and 23.17% on the four datasets, respectively, compared to DNN. Meanwhile, compared with the suboptimal model, the AUC value of the IEN model is improved by 2.12%, 1.22%, 2.15%, and 4.08%, respectively. All in all, the effectiveness of the IEN is much improved compared to the other algorithms.

The reasons for the good performance of our approach can be attributed to the following aspects. On the one hand, the refined representation of each item is obtained by learning the relationship between target items in the user behavior sequence. It is beneficial to gain a better user representation to reflect user interests. In addition, learning the feature interaction between the context and the target item by multi-head attention mechanism is beneficial to capture the current interests of the user.

On the other hand, even though models such as DIN, DIEN, DMR, and DMIN all acquire user interests from the user behavior sequence. DIN fails to capture the evolving process of user interests due to ignoring position information. DIEN pays little attention to the relevance between the user and the target item. Finally, DMR and DMIN ignore the effect of context information on the user interests; thus, these models are inferior to IEN.

Besides, DNN, Wide & Deep, and PNN both improve on the feature interactions. But they focus more on the feature interactions without modeling user behavior sequence features. It results in ignoring the user interests in the user behavior sequence. In particular, DNN does not learn feature interactions sufficiently. Therefore, it performs worse in the comparison algorithms. In addition, Wide & Deep performs not well because the "wide" part relies on manually designed feature interactions. Additionally, PNN focuses on the learning of high-order features and ignores the valuable information in the original features.

4.5 Ablation experiments

In this subsection, to check the effectiveness of each component in the IEN, we conducted ablation experiments. The ablation results are summarized in Tables 4 and 5. As expected, the results of the ablation experiments illustrate the validity of each part. Moreover, the experimental results demonstrate that each module improves the performance of CTR prediction to a certain degree.

Firstly, to verify the significance of the relevance, we compare the IEN with the IEN w/o UI after removing the correlation between the user and the target item. The AUC value of IEN w/o UI on the *Electronics* dataset is 0.7591, which is a 4.07% decrease compared to IEN. Its performance is also worse on all other datasets. It is worth mentioning that it drops

Table 4 AUC results of the ablation experiment on four datasets

Method	Electronics	Beauty	CDS & Vinyl	Book
IEN w/o UI	0.7591 ± 0.0016	0.7089 ± 0.0017	0.7912 ± 0.0019	0.7855 ± 0.0017
IEN w/o BI	0.7598 ± 0.0009	0.7209 ± 0.0011	0.7984 ± 0.0015	0.7893 ± 0.0015
IEN w/o CI	0.7616 ± 0.0014	0.7235 ± 0.0009	0.8062 ± 0.0007	0.8348 ± 0.0012
IEN	0.7701 ± 0.0005	0.7241 ± 0.0010	0.8090 ± 0.0010	0.8471 ± 0.0011

Table 5 IEN-based RI in ablation experiments

Method	Electronics (%)	Beauty (%)	CDS & Vinyl (%)	Book (%)
IEN w/o UI	−4.07	−6.78	−5.59	−17.75
IEN w/o BI	−3.81	−1.42	−3.43	−16.65
IEN w/o CI	−3.15	−0.27	−0.91	−3.54
IEN	0.00	0.00	0.00	0.00

mostly on the *Book* dataset. This is because the correlation between the user and the item reflects the user's preference for the item.

Secondly, to confirm the refined representation of each item is essential in IEN, we compare the IEN with the IEN w/o BI after deleting the refined representation of each item. On the *Electronics* dataset, the AUC value of IEN w/o BI is 0.7598, which is 3.81% lower than the IEN. In a straightforward view, it validates the necessity of learning the high-order item representation by applying a multi-head attention mechanism.

Finally, to check the validity of the feature interaction module between the context and the target item, the IEN is compared against the IEN w/o IC with the feature interaction of the context and the target item removed. The IEN w/o IC on the *Electronics* dataset has an AUC value of 0.7616. It decreases by 3.15% compared to IEN and the performance is inferior to IEN on the other datasets. Therefore, learning the feature interaction between the context and the target item is proven to be reasonable.

In summary, we found that removing the correlation between the user and the target item had the largest impact on the performance of CTR prediction. Removing the learning of the refined item representations has a close second impact on the performance of CTR prediction. Deleting the feature interaction between the context and the target item has the least impact on the performance of CTR prediction. Nonetheless, each part contributes to the performance of CTR prediction.

5 Conclusion

In this paper, we propose a new model called interest extraction method based on multi-head attention mechanism (IEN). We design an interest extraction module, which consists of the item representation module (IRM) and the context–item interaction module (CIM). In the IRM, the relationship between the items in the user behavior sequence is learned to obtain the refined representation of each item. This operation assists to acquire a good user representation which is used to reflect user interests. Besides, in the CIM, the feature interaction between the context and the target item is learned by a multi-head attention

mechanism. Furthermore, experimental results on four public datasets demonstrate that our proposed method IEN helps to improve the performance of CTR prediction. The ablation experiments further illustrate the effectiveness and necessity of each part in the IEN. As far as we know, the time interval information is useful to reflect the variations of user interests, and this model does not take time interval information into account. Therefore, how to better explore the time interval information will be the task in the next stage.

Acknowledgements The work was supported by the National Natural Science Foundation of China (Grant No. U1931209), the Central Government Guides Local Science and Technology Development Funds (Grant No. 20201070), and the Fundamental Research Program of Shanxi Province (Grant Nos. 20210302123223, 202103021224275).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval The study is original and has not been submitted to any other journal/conference.

References

1. Wang J, Huang P, Zhao H, Zhang Z, Zhao B, Lee DL (2018) Billion-scale commodity embedding for e-commerce recommendation in alibaba. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, pp 839–848
2. An M, Wu F, Wu C, Zhang K, Liu Z, Xie X (2019) Neural news recommendation with long- and short-term user representations. In: Proceedings of the 57th conference of the association for computational linguistics, pp 336–345
3. Chen W, Huang P, Xu J, Guo, X, Guo C, Sun F, Li C, Pfadler A, Zhao H, Zhao B (2019) POG: personalized outfit generation for fashion recommendation at alibaba ifashion. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp 2662–2670
4. Ni Y, Ou D, Liu S, Li X, Ou W, Zeng A, Si L (2018) Perceive your users in depth: Learning universal user representations from multiple e-commerce tasks. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, pp 596–605
5. Pei C, Zhang Y, Zhang Y, Sun F, Pei D (2019) Personalized context-aware re-ranking for e-commerce recommender systems
6. He X, Pan J, Jin O, Xu T, Liu B, Xu T, Shi Y, Atallah A, Herbrich R, Bowers S, Candela JQ (2014) Practical lessons from predicting clicks on ads at facebook. In: Proceedings of the eighth international workshop on data mining for online advertising, pp 5–159
7. Huang Z, Pan Z, Liu Q, Long B, Ma H, Chen E (2017) An ad CTR prediction method based on feature learning of deep and shallow layers. In: Proceedings of the 2017 ACM on conference on information and knowledge management, pp 2119–2122
8. Huang G, Liu Z, van der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: 2017 IEEE conference on computer vision and pattern recognition, pp 2261–2269
9. Lauriola I, Lavelli A, Aiolli F (2022) An introduction to deep learning in natural language processing: models, techniques, and tools. Neurocomputing, pp 443–456
10. Devlin J, Chang M, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the north American chapter of the association for computational linguistics: human language technologies, pp 4171–4186
11. Cheng H, Koc L, Harmsen J, Shaked T, Chandra T, Aradhye H, Anderson G, Corrado G, Chai W, Ispir M, Anil R, Haque Z, Hong L, Jain V, Liu X, Shah H (2016) Wide & deep learning for recommender systems. In: Proceedings of the 1st workshop on deep learning for recommender systems, pp 7–10
12. Qu Y, Cai H, Ren K, Zhang W, Yu Y, Wen Y, Wang J (2016) Product-based neural networks for user response prediction. In: IEEE 16th international conference on data mining, pp 1149–1154
13. Zhou G, Zhu X, Song C, Fan Y, Zhu H, Ma X, Yan Y, Jin J, Li H, Gai K (2018) Deep interest network for click-through rate prediction. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, pp 1059–1068

14. Zhou G, Mou N, Fan Y, Pi Q, Bian W, Zhou C, Zhu X, Gai K (2019) Deep interest evolution network for click-through rate prediction. In: The thirty-third AAAI conference on artificial intelligence, pp 5941–5948
15. Lyu Z, Dong Y, Huo C, Ren W Deep match to rank model for personalized click-through rate prediction. In: The thirty-fourth AAAI conference on artificial intelligence, pp 156–163
16. McMahan HB, Hol G, Sculley D, Young M, Ebner D, Grady J, Nie L, Phillips T, Davydov E, Golovin D, Chikkerur S, Liu D, Wattenberg M, Hrafnkelsson AM, Boulos T, Kubica J (2013) Ad click prediction: a view from the trenches. In: The 19th ACM SIGKDD international conference on knowledge discovery and data mining, pp 1222–1230
17. Rendle S (2010) Factorization machines. In: Webb GI, Liu B, Zhang C, Gunopulos D, Wu X (eds) ICDM 2010, The 10th IEEE international conference on data mining, Sydney, pp 995–1000
18. Juan Y, Zhuang Y, Chin W, Lin C (2016) Field-aware factorization machines for CTR prediction. In: Proceedings of the 10th ACM conference on recommender systems, pp 43–50
19. Pan J, Xu J, Ruiz AL, Zhao W, Pan S, Sun Y, Lu Q (2018) Field-weighted factorization machines for click-through rate prediction in display advertising. In: Proceedings of the 2018 World Wide Web Conference on World Wide Web, pp 1349–1357
20. Yang Y, Cai J, Yang H, Zhang J, Zhao X (2020) TAD: a trajectory clustering algorithm based on spatial-temporal density analysis. Expert Syst Appl 139:112846
21. Yang Y, Cai J, Yang H, Li Y, Zhao X (2022) Isbfk-means: a new clustering algorithm based on influence space. Expert Syst Appl 201:117018
22. Yang Y, Cai J, Yang H, Zhao X (2022) Density clustering with divergence distance and automatic center selection. Inf Sci 596:414–438
23. Yang H, Shi C, Cai J, Zhou L, Yang Y, Zhao X, He Y, Hao J (2022) Data mining techniques on astronomical spectra data-i. clustering analysis. Monthly Notices Astron Soc 517(4):5496–5523
24. Yang H, Zhou L, Cai J, Shi C, Yang Y, Zhao X, Duan J, Yin X (2022) Data mining techniques on astronomical spectra data-ii. classification analysis. Monthly Notices R. Astron Soc 518(4):5904–5928
25. He X, Chua T (2017) Neural factorization machines for sparse predictive analytics. In: Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval, Shinjuku, pp 355–364
26. Xiao J, Ye H, He X, Zhang H, Wu F, Chua T (2017) Attentional factorization machines: learning the weight of feature interactions via attention networks. In: Proceedings of the twenty-sixth international joint conference on artificial intelligence, pp 3119–3125
27. Guo H, Tang R, Ye Y. Li Z, He X (2017) Deepfm: a factorization-machine based neural network for CTR prediction. In: Sierra, C. (ed.) Proceedings of the twenty-sixth international joint conference on artificial intelligence, pp. 1725–1731
28. Wang R, Fu B, Fu G, Wang M (2017) Deep & cross network for ad click predictions. In: Proceedings of the ADKDD'17, pp 12–1127
29. Lian J, Zhou X, Zhang F, Chen Z, Xie X, Sun G (2018) xdeepfm: Combining explicit and implicit feature interactions for recommender systems. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, pp 1754–1763
30. Chen Q, Zhao H, Li W, Huang P, Ou W (2019) Behavior sequence transformer for e-commerce recommendation in alibaba. In: Proceedings of the 1st international workshop on deep learning practice for high-dimensional sparse data, pp 1–4
31. Feng Y, Lv F, Shen W, Wang M, Sun F, Zhu Y, Yang K (2019) Deep session interest network for click-through rate prediction. In: Proceedings of the twenty-eighth international joint conference on artificial intelligence, pp 2301–2307
32. Wu M, Xing J, Chen S (2022) Deep user multi-interest network for click-through rate prediction. In: knowledge science, engineering and management—15th international conference. lecture notes in computer science, vol 13369, pp 57–69
33. Zhang K, Qian H, Cui Q, Liu Q, Li L, Zhou J, Ma J, Chen E (2021) Multi-interactive attention network for fine-grained feature learning in CTR prediction. In: WSDM '21, The fourteenth ACM international conference on web search and data mining, pp 984–992
34. Yan C, Li X, Chen Y, Zhang Y (2022) JointCTR: a joint CTR prediction framework combining feature interaction and sequential behavior learning. Appl Intell 52, 4701–4714 (2022). https://doi.org/10.1007/s10489-021-02678-8
35. Jiang W, Jiao Y, Wang Q, Liang C, Guo L, Zhang Y, Sun Z, Xiong Y, Zhu Y (2022) Triangle graph interest network for click-through rate prediction. In: Proceedings of the fifteenth ACM international conference on web search and data mining

36. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. In: Advances in neural information processing systems 30: Annual conference on neural information processing systems 2017, pp 5998–6008
37. LeCun Y, Bengio Y, Hinton GE (2015) Deep learning. Nature 521(7553):436–444
38. Kingma DP, Ba J (2015) Adam: A method for stochastic optimization. In: 3rd International conference on learning representations
39. Fawcett T (2006) An introduction to ROC analysis. Pattern Recognit. Lett. 27(8):861–874
40. Yan L, Li W, Xue G, Han D (2014) Coupled group lasso for web-scale CTR prediction in display advertising. In: Proceedings of the 31th international conference on machine learning. JMLR workshop and conference Proceedings, vol 32. pp 802–810

Haifeng Yang is a professor Computer Application Technology, Taiyuan University of Science and Technology, Taiyuan, China. He is the longterm member of the institute for intelligent information and data mining. Pro. Yang is a member of China Computer Federation (CCF) and Chinese Astronomical Society (CAS). His research interests concern the data mining and machine learning methods in the specific backgrounds especially for the astronomical big data.

Linjing Yao is a graduate student at Taiyuan University of Science and Technology. Her main research interests are data mining and recommendation systems.

Jianghui Cai is the chief professor of Computer Application Technology, Taiyuan University of Science and Technology, Taiyuan, China. He is the long-term member of the institute for intelligent information and data mining. Pro. Cai is a senior member of China Computer Federation (CCF). His research interests concern the data mining and machine learning methods in specific backgrounds of astronomical informatics, seismology, and mechanical engineering.

Yupeng Wang received the MS degree in School of Mathmatical Sciences from Beihang University, Beijing, China, in 2014. She is currently working toward the PhD degree in the College of Mechanical Engineering from Taiyuan University of science and technology, Taiyuan, China. Her main research interests include clustering, time series analysis and data mining.

Xujun Zhao received the BS in the School of Computer Science and Technology from Taiyuan University of Technology (TYUT), and the MS and Ph.D. degrees from Taiyuan University of Science and Technology. He is currently a associate professor in the School of Computer Science and Technology at TYUST. He is a member of China Computer Federation (CCF). His research interests include data mining and parallel computing.