



Article

Deep Filter Context Network for Click-Through Rate Prediction

Mingting Yu, Tingting Liu and Jian Yin *

School of Mechanical and Information Engineering, Shandong University, Weihai 264209, China;
202000800470@mail.sdu.edu.cn (M.Y.); ltting@mail.sdu.edu.cn (T.L.)

* Correspondence: yinjian@sdu.edu.cn

Abstract: The growth of e-commerce has led to the widespread use of DeepCTR technology. Among the various types, the deep interest network (DIN), deep interest evolution network (DIEN), and deep session interest network (DSIN) developed by Alibaba have achieved good results in practice. However, the above models' use of filtering for the user's own historical behavior sequences and the insufficient use of context features lead to reduced recommendation effectiveness. To address these issues, this paper proposes a novel article model: the deep filter context network (DFCN). This improves the efficiency of the attention mechanism by adding a filter to filter out data in the user's historical behavior sequence that differs greatly from the target advertisement. The DFCN pays attention to the context features through two local activation units. This model greatly improves the expressiveness of the model, offering strong environment-related attributes and the adaptive capability of the model, with a significant improvement of up to 0.0652 in the AUC metric when compared with our previously proposed DICN under different datasets.

Keywords: DeepCTR; context features; filter; local activation unit; users' historical behavior features; deep filter context network (DFCN)



Citation: Yu, M.; Liu, T.; Yin, J. Deep Filter Context Network for Click-Through Rate Prediction. *J. Theor. Appl. Electron. Commer. Res.* **2023**, *18*, 1446–1462. <https://doi.org/10.3390/jtaer18030073>

Academic Editor: Jorge Bernardino

Received: 18 April 2023

Revised: 11 August 2023

Accepted: 20 August 2023

Published: 22 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the increasing popularity of the Internet and the continuous development of computer science and technology, Internet finance has become an important part of the country's economic and financial life. Among the various effects, the rapid development of the e-commerce industry has provided new opportunities for the e-commerce operation of Internet finance [1]. In the information age, although people have more options for shopping or browsing information, the sheer volume of information makes it impossible for people to select products that meet their needs or preferences, and they can only shop by searching precisely. This greatly reduces the efficiency of shopping and the user's shopping experience. As an information filtering system, recommender systems have been introduced to the e-commerce industry, learning from users' personal preferences and historical behavior to predict users' preferences and make effective filtering recommendations. This not only saves advertising costs for e-commerce platforms but also improves the shopping experience for users [2,3]. Initially, recommender systems were divided into three categories, based on recommendation mechanisms: content-based recommender systems [4], collaborative filtering-based recommender systems [5], and hybrid recommender systems that combine both these systems [6]. With the advent and development of deep learning, recommender systems were upgraded to incorporate deep learning into the recommender system. Deep learning can further mine and analyze data to find hidden relationships and patterns between the data, helping recommender systems to make recommendations more accurately and efficiently [7].

Click-through rate (CTR) prediction models are a crucial group of recommendation systems. Click-through rate prediction analyzes the probability of clicking on a recommended advertisement or target by analyzing the users' historical clicking behavior and known context features, enabling more accurate targeting of advertisements and, thus,

saving costs [8,9]. Common CTR models include the factorization machine (FM) [10] and logistic regression (LR) [8] as base models. Such a model is mainly a manual or automatic cross-construction of feature vectors and weighted summation employed to obtain click-through rate predictions. However, the problem is that the basic CTR prediction model can only predict the relationships on the surface of features, i.e., it can only complete the intersection of lower-order features and cannot make effective judgments regarding the deep hidden relationships and regular characteristics among higher-order features. With a combination of deep learning and recommender systems, the CTR prediction model has been upgraded to DeepCTR, which mainly adds a deep learning component to help address the shortcomings of traditional CTR prediction models that cannot complete higher-order feature intersections and extract hidden vectors. Initially, DeepCTR models were basically improvements on traditional CTR models with a deep learning component, such as DeepFM [11], NFM [12], xDeepFM [13], and PNN [14]. These methods incorporate deep learning components; however, they mostly follow the approach of first compressing and embedding high-dimensional user features into a fixed-length representation vector, and then feeding them into a multilayer perceptron (MLP). Due to the high dimensionality of user feature behavior, such as the wide variety of interests exhibited by each individual, compressing the embedding stage of the model can result in a loss of information and it may not fully utilize the user feature behavior [15]. As neuroscience research continues to advance, the idea of attentional mechanisms has been proposed [16] and applied to recommender systems. Influenced by the combination of the attention mechanism and DeepCTR, the Alibaba Group introduced the deep interest network (DIN) [15]. Subsequently, the deep interest evolution network (DIEN) [17] and deep session interest network (DSIN) [18] were successfully presented, which are improvements on DIN. Although these models have performed well in reality and have advanced the DeepCTR model, they still have problems. Specifically, they ignore the influence of context feature vectors on the historical behavioral characteristics of users and users' click-through rates. Moreover, the models only treat the target items with strong context-related attributes as ordinary vectors for compressed embedding, meaning that they are not fully utilized and, thus, limit the expressive ability of the model. To address this problem, we propose the deep interest context network (DICN) [19], thereby adequately solving the problem of ignoring contextual vectors. However, since DICN is based on the improved DIN model, filtering of the users' historical behavior features is still at the primary stage, and is still based on DIN. That is, the historical behavior features are directly embedded and compressed and are then fed into the attention mechanism without much processing, resulting in a great deal of redundant information. Although the attention mechanism network can be assigned weights, this still limits the expressive ability of the model.

In response to this issue, this paper makes the following contributions:

- This paper presents a new and simple filtering machine for users' historical behavior features. This filtering machine makes full use of the characteristics of the targeted advertisements to filter the users' historical behavior, helping the attention mechanism process the input feature vector more efficiently and expressively.
- In this paper, a new algorithmic model is proposed: the deep filter context network (DFCN). The DFCN introduces a filter to the original DICN model. The filter enhances the model's ability to capture those of the users' historical behavior features that align with target advertisements while preserving user interest diversity. The model's self-adaptability and expressiveness are also improved by processing the user history feature vector in a prior step.
- In this paper, experiments are conducted on the open Taobao user dataset and the Amazon user dataset. The experimental results demonstrate the effectiveness and superiority of the DFCN model.
- This paper designs two sets of comparative experiments so as to verify that the filtering layer can effectively enhance the ability of the attention mechanism in capturing and helping the model to improve its predictive ability. In addition, the importance of the

newly added local activation unit for context features is demonstrated. At the same time, this paper highlights the fact that the newly developed filtering layer is more suitable for the pre-processing of users' historical behavior feature data, which means it cannot replace attention mechanism empowerment.

The rest of the paper is organized as follows: Section 2 discusses related works in the literature. Section 3 describes the filters, attention mechanisms, and structure of the components of the overall DFCN model. Section 4 describes the setup and analysis of the results of related experiments. Section 5 demonstrates some details of the models and makes comparisons with some inadequate models that were used during the experiment. Finally, Section 6 concludes the paper and presents ideas for future work.

2. Related Works

2.1. Attention Mechanism and DICN

The attention mechanism presents a model that simulates the attention paid by the human brain. The core principle is to use the probability distribution of attention to capture the effect of a key input on the output [20]. With the emergence and continuous development of deep learning, people have added the attention mechanism to deep learning and have proposed many deep learning algorithms about the attention-based mechanism [21]. This is a good thing for the recommendation system when used as a commercial data mining algorithm; it can push the recommendation system to continue to progress, improve the model's expressiveness and self-adaptive ability, and improve data mining. In our previous paper, we proposed a new model called the deep interest context network (DICN) which makes full use of attention mechanisms and deep learning. The model takes the DIN proposed by the Alibaba Group as the base model, optimizes the attention mechanism of the DIN, and uses those context features that are not valued by the DIN model to operate the attention mechanism to empower the users' historical behavior features. The DICN operates on the attention mechanism by adding a new local activation unit that takes the timestamped feature vector from the users' historical behavior features and the context features of the target advertisement, resulting in an attention weight matrix called the contextual attention weight matrix. At the same time, the original local activation unit operates the attention mechanism, using the feature of the users' historical behavior and the target advertisement to obtain the users' attention weight matrix. After multiplying these two weight matrices to obtain the total attention weight matrix, it is multiplied with the original users' historical behavior features matrix to achieve weighting of the users' historical behavior features matrix, helping the model to focus on those features with high relevance to the target advertising attributes and context, and improving the expressive and predictive power of the model.

2.2. Bandpass Filter

In the beginning, the term bandpass filter was used in radio communication systems. In such systems, the superimposition of other frequencies of noise in the channel with the modulated signal produces distortion, which can convey incorrect information and affect the quality of communication [22]. As can be seen, signal filtering is a very important part of the process, ensuring the reliability and accuracy of the signal [23]. As one of the most common types of filters, bandpass filters are used to select signals within a certain frequency range and suppress signals at other frequencies [24]. With the rise and continuous development of digital image-processing techniques, filters are used in the field of digital image processing. An image can be represented as a discrete function of pixel values versus the plane coordinates and can be viewed as two-dimensional signal data [25]. When filters are used for image processing, they allow the image to be enhanced or restored to avoid the distortion caused by interference from other noisy signals [26]. The effect is shown in Figure 1.

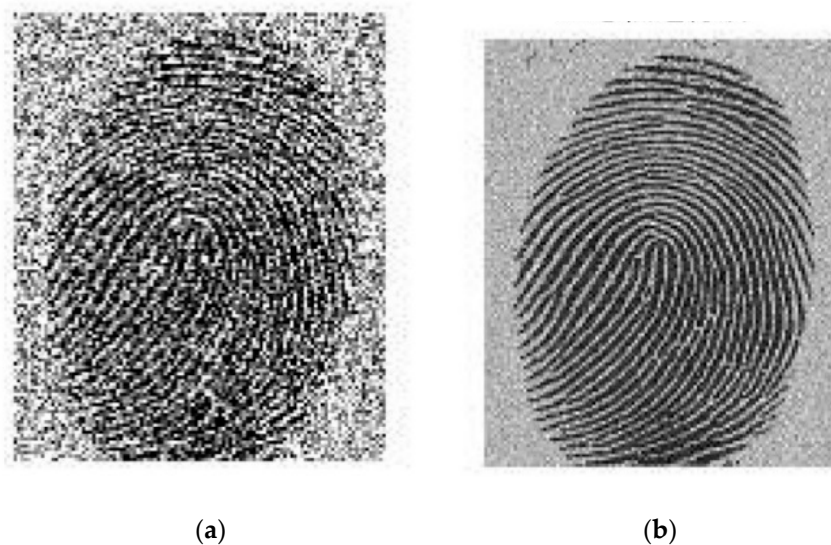


Figure 1. A demonstration of the usefulness of filters in digital image processing: (a) is the blurred resultant image, with Gaussian noise superimposed on the image; (b) is the image with Gaussian noise, processed with a low-pass filter to produce a clear image.

In the DeepCTR model, the users' historical behavior data can be seen as multi-dimensional vector signals. In a large database of users' historical behavior, there must be fluctuations and deviations in users' interests at different times or variations in users' interests in the different types of items, which can be treated as noise signals that have a negative effect on the prediction model. In the DIN and DICN models, users' historical behavior data is manipulated directly, without much processing, by the attention mechanism. Although the attention mechanism itself can be understood as a filtering operation, the complexity and size of the input data lead to a decrease in the effectiveness of the attention mechanism filtering process, which makes it necessary to pre-process the users' historical behavior data. In this paper, we introduce the bandpass filter principle used in electronic communication systems into the DeepCTR model and formulate the passband and blocking band of the bandpass filter algorithm according to the target advertising vector, so as to achieve the initial screening process of the users' historical behavior data and help the attention mechanism to further complete the weight allocation.

3. Model Structure

The structural flow of the deep filtered contextual network model is as follows: the input layer pre-processes the data in the dataset according to its characteristics. The embedding layer then transforms the user features, users' historical behavior features, payment activity, and context features in the dataset into sparse vectors, variable length sparse vectors, and dense vectors, classified according to the characteristics of data length. Specifically, user features and context features are converted to sparse vectors since they are fixed-length sparse features. Sparse features with variable lengths of users' historical behavior are then converted to variable-length sparse vectors. Payment activity is converted to a dense vector. After conversion into a vector, the users' historical behavior features are entered into the filtering layer. In the filtering layer, the target advertisement vector is expanded into a tensor of the same shape as the tensor of the users' historical behavior features, after which the historical behavior tensor is subtracted from the target advertisement tensor to obtain a bandstop filter with the target advertisement as the blocking band. After this, the target advertisement tensor is again subtracted from the bandstop filter tensor to obtain a bandpass filter, with the target advertisement tensor as the passband. The bandpass filter is multiplied by the Hadamard product of the original users' historical behavior sequence tensor to obtain the filtered bandpass historical behavior sequence tensor. In the attention layer, the bandpass historical behavior sequence tensor and the context feature tensor will

be fed into the two local activation units, resulting in the users' historical weight matrix and the context weight matrix, respectively. They will first be multiplied to obtain the total weight matrix and then multiplied with the bandpass users' historical behavior tensor to perform the additive pooling operation together. Ultimately, the result of the sum pooling operation and the user feature sparse vector, the target advertisement sparse vector, and the context feature sparse vector are passed through the MLP layer to obtain the final result. The overall block diagram is shown in Figure 2.

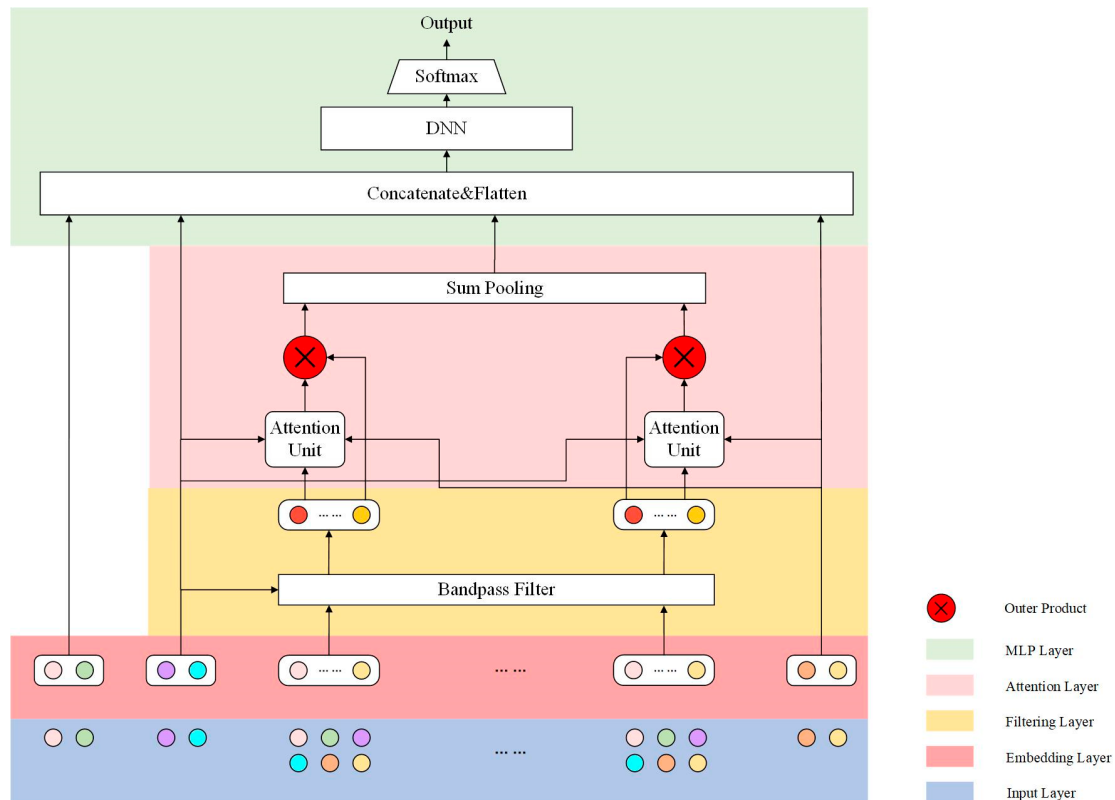


Figure 2. The structure of the DFCN model.

3.1. Input Layer

In this layer, the raw data that are fed into the model are pre-processed. As the input data are sparse, high-latitude data and not directly usable by the model, this layer pre-processes the input data to encode them in preparation for the input embedding layer.

For sparse features of low dimensionality and fixed length, such as user features, target advertisements, and contextual features, we preprocessed the data using one-hot encoding [27]:

$$e_i \in D^{K_i} \quad (1)$$

$$e_i[j] \in \{0, 1\} \quad (2)$$

$$\sum_{j=1}^{K_i} e_i[j] = 1 \quad (3)$$

where e_i is the i -th feature group in the dataset D , and K_i denotes the dimensionality of this feature group, i . The equation indicates that only one element in a feature group is coded as 1, while the rest of the elements are all coded as 0.

However, there is an obvious problem with the unique thermal encoding, which is very detrimental to the model's embedding compression operation if the feature length is not fixed and the unique thermal encoding is of varying lengths. Therefore, for users' historical behavior features with data of variable lengths, we use label encoding to encode the discrete text and numbers. In these equations, N_i denotes that there are N_i different categories in feature group i . The equation indicates that the element $e_i[j]$ is encoded using

consecutive integers in the interval, effectively solving the problem of using unique heat to encode too great a dimensionality:

$$e_i \in D^{K_i} \quad (4)$$

$$\alpha \in [0, N_i - 1] \quad (5)$$

$$e_i[j] = \alpha \quad (6)$$

3.2. Embedding Layer

After the input layer has pre-processed the data, the embedding layer takes the pre-processed data and compresses them for embedding. Since the high dimensionality of the vectors after pre-processing is not conducive to model learning, the sparse vectors are mapped from the high-latitude to the low-latitude vector space in the embedding layer and are then converted into fixed-length embedding vectors to facilitate the learning of non-linear relationships between features in the fully connected layer. The embedding layer formula is as follows:

$$G^i = [g_1^i, g_2^i, \dots, g_j^i, \dots, g_{K_i}^i] \in \mathbb{R}^{K_i \times K_a} \quad (7)$$

The embedding matrix G^i of the i -th feature group stitches together the embedding vector g_j^i from dimension 1 to K^i , while the embedding vector g_j^i takes the values in the set of real numbers in dimension K^i :

$$g_j^i \in \mathbb{R}^{K_a} \quad (8)$$

If the feature group is encoded using one-hot encoding, the embedding vector of the feature group is represented as a single embedding vector:

$$t_i = g_j^i \quad (9)$$

If the feature group is encoded using label encoding, the embedding vector of feature group i is represented as a tensor, i.e., a list of embedding vectors:

$$\{t_{i_1}, t_{i_2}, \dots, t_{i_j}\} = \{g_{i_1}^i, g_{i_2}^i, \dots, g_{i_j}^i\} \quad (10)$$

3.3. Filtering Layer

After the data have been compressed by the embedding layer, they move to the filtering layer, where the users' historical behavior features are processed again. The core idea of this layer is based on the bandpass filter found in radio communication systems; here, we construct a bandpass filter with the target advertisement as the passband to re-filter the users' historical behavior features data, helping the attention layer to better implement the attention mechanism and assign higher weights to those features of the users' historical behavior that are similar to the target advertisement.

Since the users' historical behavior feature is a tensor and the target advertisement is a vector, the target advertisement vector first needs to be expanded into a tensor, so that the target advertisement tensor is shaped into the form of the users' historical behavior tensor:

$$T_a = \{G_1, G_2, \dots, G_n\} \quad (11)$$

Here, n represents the length of the second dimension of the users' historical behavior profile tensor.

The users' historical behavior features tensor, T_h , is subtracted from the target advertising tensor, T_a , to obtain a bandstop filter, H_s , with the target advertising tensor as the stop band. The target advertising tensor, T_a , is then subtracted from the bandstop filter, H_s , to obtain a bandpass filter, H_p , with the target advertising tensor as the gain.

$$H_s = T_h - T_a \quad (12)$$

$$H_p = T_a - H_s \quad (13)$$

Multiplying the bandpass filter H_p with the original users' historical behavior feature T_h yields the final users' historical behavior feature tensor T_p , which passes through the bandpass filter:

$$T_p = H_p \times T_h \quad (14)$$

After the band-pass filter, the values of those elements in the original users' historical behavior profile tensor with low relevance to the target advertisement will be reduced, while the values of elements with high relevance to the target advertisement will be retained. This layer effectively filters the users' historical behavior features, helping the attention layer to give greater weight to users' historical behavior features that are highly relevant to the target advertisement and to reduce the weight of other non-relevant features. The structure of the bandpass filter is shown in Figure 3.

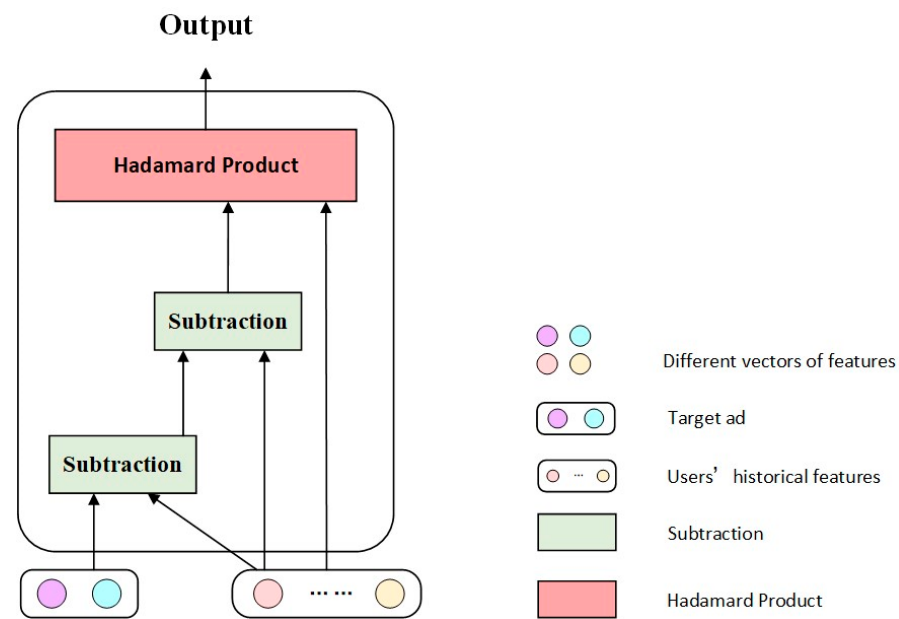


Figure 3. Structure of the bandpass filter in the filtering layer.

3.4. Attention Layer

This layer mainly provides the attention mechanism empowerment operations. This layer is centered on the attention unit, which contains two local activation units, to learn the attention weight matrix for the band-pass filtered users' historical behavior and context features, respectively. The formula is as follows:

$$\begin{cases} T_p = \{G_g, G_p, G_c, G_s, G_w\} \\ G_{ad} = \{t_g, t_p, t_c\} \\ G_{env} = \{t_s, t_w\} \end{cases} \quad (15)$$

G_g, G_p, G_c represent the item, behavior, and item-type embedding matrices in the users' historical behavior features, while G_s and G_w represent the month and date embedding matrices in the users' historical behavior features, respectively, while t_g, t_p and t_c correspond to the target advertisement item, behavior, and type of embedding vectors, respectively. t_s and t_w denote the current context feature embedding the vectors.

Taking the bandpass users' historical behavior feature for the attention weight matrix as an example, first, the target advertisement matrix is expanded into a tensor with the same shape as the bandpass users' historical behavior feature tensor. Then, the bandpass historical behavior feature tensor and the target advertisement tensor are used to perform the Hadamard product operation and subtraction operation.

$$T_a = \{G_{ad_1}, G_{ad_2}, \dots, G_{ad_n}\} \quad (16)$$

$$T_p * T_a = \begin{bmatrix} h_{11}a_{11} & \cdots & h_{1j}a_{1j} \\ \vdots & \ddots & \vdots \\ h_{i1}a_{i1} & \cdots & h_{ij}a_{ij} \end{bmatrix} \quad (17)$$

$$T_p - T_a = \begin{bmatrix} h_{11} - a_{11} & \cdots & h_{1j} - a_{1j} \\ \vdots & \ddots & \vdots \\ h_{i1} - a_{i1} & \cdots & h_{ij} - a_{ij} \end{bmatrix} \quad (18)$$

The result is then spliced with the bandpass historical behavior feature matrix and the target advertisement matrix to feed into the Dice activation function [15] and the linearized part of the DNN model, which is used to obtain the weight matrix ω_h of the bandpass historical behavior features.

$$\omega_h = DNN(T_p, T_h, T_p * T_h, T_p - T_h) \quad (19)$$

The structure of the attention unit is shown in Figure 4.

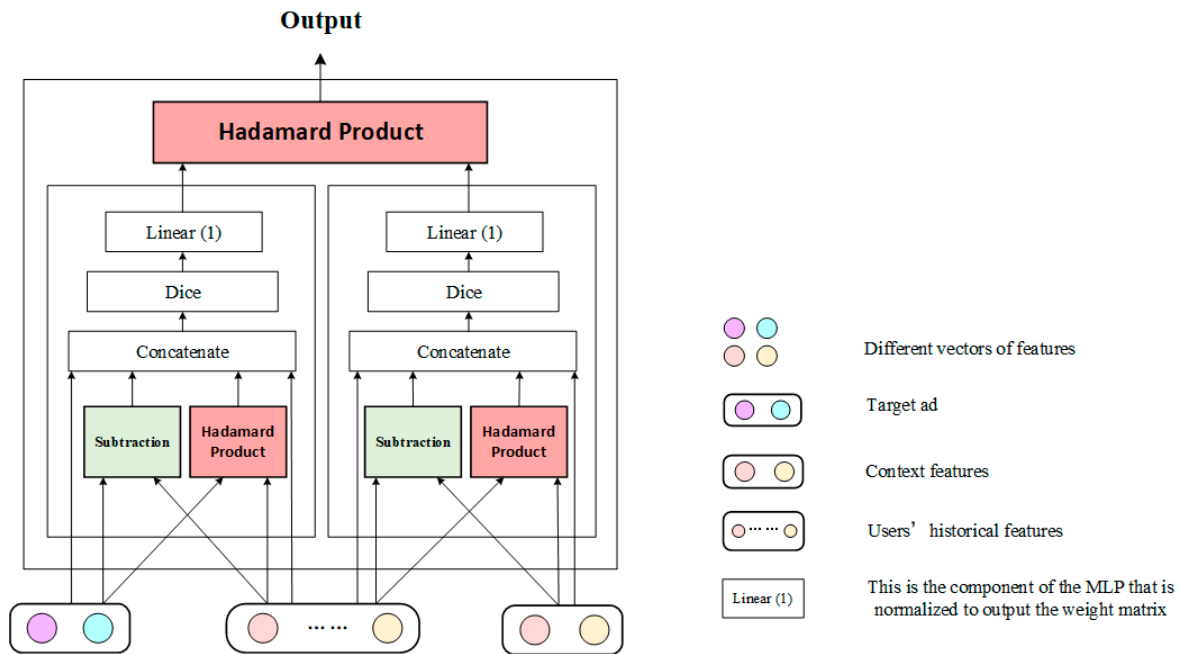


Figure 4. Structure of the attention unit.

The attention weight matrix of the context features is identified using the same steps as above, resulting in an attention weight matrix of context features, ω_e .

Ultimately, the Hadamard product of ω_h and ω_e gives the total attention weight matrix Ω ; then, the outer product with the bandpass users' historical behavior tensor gives the bandpass weight of the users' historical behavior tensor, T_{att} .

$$\Omega = \omega_h * \omega_e \quad (20)$$

$$T_{att} = T_p \times \Omega \quad (21)$$

After the local activation unit yields the users' historical behavior tensor with bandpass weights, the addition pooling operation is performed. This effectively solves the problem that the fixed length of user interests makes the model's learning efficiency decrease.

3.5. MLP Layer

This layer is the deep learning part of the model, which learns the non-linear relationships between features by using a fully connected neural network DNN model. The multi-layer perceptron layer first concatenates and then flattens the user feature embed-

ding matrix, the target advertisement embedding matrix, the context feature embedding matrix, and the additive pooled users' historical behavior feature matrix with pass weights, using them to one-dimensionalize the multi-dimensional input, avoiding high-dimensional vectors and facilitating fully connected neural network learning. After flattening, first, the results are fed into the DNN model, then ReLU [28] is selected as the activation function of the DNN. Finally, the normalization operation is completed using the SoftMax function [29] to output the click-through rate prediction results.

4. Experiments and Analysis

In order to validate the performance and learning ability of the new model that is proposed in this paper and prove the superiority of the new model, this experiment uses TensorFlow-2.1.0 as the learning framework and Python 3.7 as the running environment, using them to compare the various classical DeepCTR models.

4.1. Datasets

To prevent the occurrence of coincidences, two datasets were chosen for this paper: the Alibaba Taobao user history behavior dataset and the Amazon clothing, shoes, and jewelry dataset. At the same time, the items in both datasets have a high degree of contextual relevance to the environment, i.e., clothing, shoes, etc., have a strong seasonal relevance, with different clothing choices for different seasons. The Alibaba Taobao user history behavior dataset contains information on the users, items, item types, users' historical behavior (click, favorite, add to cart, or purchase), and behavior timestamps [30–32]. The Amazon clothing, shoes, and jewelry dataset contains users, items, users' ratings, and behavior timestamps [33–35]. Details of the datasets are given in Table 1.

Table 1. Basic statistics for the datasets.

Datasets	Features	Numbers	Total Samples
Taobao	Users	376	11,198
	Items	9066	
	Categories	1248	
	Behavior Types	4	
	Timestamps	11,198	
Amazon	Users	88,462	91,206
	Items	8510	
	Scores	5	
	Timestamps	91,206	

4.2. Evaluation Indicators

In order to accurately and effectively evaluate the learning and prediction performance of the DFCN model, the experiment in this paper divides the datasets into a training group and a test group, according to a certain ratio. Meanwhile, the AUC (area under the curve), the log loss function, and RelaImpr-DIN [15,19] are used as the evaluation indicators in this paper, where the loss function formula refers to the loss function formula of the DIN model:

$$L = -\frac{1}{N} \sum_{(i,b) \in T} (b \log p(i) + (1-b) \log(q - p(i))) \quad (22)$$

RelaImpr-DIN is based on an improved version of the RelaImpr formula. Originally, the RelaImpr formula was designed to be able to reflect the gap between the DIN model and BaseModel [15] more intuitively. In this paper, the AUC parameters of the embedding and MLP paradigms were replaced with the AUC parameters of the DIN model in order to further reflect the performance gap between the DFCN model and the DIN and DICN models. It is expressed as:

$$RelaImpr - DIN = \left(\frac{AUC(\text{measured model}) - 0.5}{AUC(\text{DIN model}) - 0.5} - 1 \right) \times 100\% \quad (23)$$

4.3. Comparison Models

To verify that the DFCN model in this paper performs well in all the above metrics, we use the following, widely used, CTR prediction model for comparison to make the results more intuitive.

- FNN [36]: FNN is a combination of FM and DNN. The FNN model is one of the more classical embedding and MLP paradigms, which uses the hidden vectors obtained from FM training as initial values to feed into the DNN, i.e., a combination of embedding and the multilayer perceptron.
- AFM [37]: This model employs the attention mechanism, which is an evolutionary update of the NFM [12] model, by introducing the attention mechanism into the FM model and assigning weights to the vectors after the embedding and interaction layers, through the attention pooling layer.
- DeepFM [11]: This model is an evolutionary upgrade of the Wide and Deep model. DeepFM uses the FM model algorithm in the wide part and deep learning in the deep part to extract the non-linear relationships between the features.
- DIN [15]: A CTR prediction model with significant advances was proposed by Zhou et al. DIN introduces local activation units into the embedding and MLP paradigm and uses an attention mechanism to assign weights to users' historical behavior features as a way to explore the similarity between historical features and target advertisements.
- DICN [19]: An evolutionary update of DIN that adds an additional local activation unit to the DIN model to explore the similarity of environmental and contextual features in the historical features.
- DFCN: The new model that is proposed in this paper and that is described in Section 3 introduces a filtering layer to process the users' historical behavior features of the compressed embeddings, reducing the parameters of those elements with little similarity to the target advertisement and helping the local activation unit to perform the assignment operation more accurately and efficiently.

4.4. Parameter Settings

This paper compares experiments that use different models with the same parameters as a way to verify the superiority of the new model DFCN. When using the Taobao user history behavior dataset, the number of iterations for each epoch is set to 10, while the model batch size is set to 256. The number of training sets is 8958 data units, of which 7166 data units are for training, 1792 data units are for validation after training, and the number of test sets is 2240. The final ratio of the data training set:validation set:test set was 14:6:5. The number of DNN hidden layers in the MLP was 256, 128, and 64, and the number of DNN hidden layers in the local activation unit was 80, 40. When using the Amazon clothing, shoes, and jewelry dataset, the number of iterations of the epoch was increased to 15, and the remaining parameters were kept constant to verify that overfitting would not occur in the case of larger datasets. However, since the total amounts of data in the two datasets are not the same, the number of training and test sets is also different. In the Amazon clothing, shoes, and jewelry dataset, the total amount of data in the training set comprised 72,965 units, of which 58,372 data units were used to train the model, representing 56% of the total dataset, while 14,197 data units were used to validate the model, representing 24% of the total dataset. The final 18,241 data units were used to test the model in the test set, representing 20% of the total dataset.

4.5. Analysis of Results

This section will show the experimental results visually, including tables and images, to verify the superiority of the model.

4.5.1. AUC and the RelaImpr-DIN

This subsection uses tables and bar charts to present the indicator data for the above comparison model and the DFCN model, as shown in Table 2.

Table 2. AUC and RelaImpr-DIN, as predicted using CTR.

Model	Taobao		Amazon	
	AUC	RelaImpr-DIN	AUC	RelaImpr-DIN
FNN	0.5165	−89.73%	0.5180	−98.66%
AFM	0.5270	−83.20%	0.5248	−81.53%
DeepFM	0.5222	−86.19%	0.5188	−86.00%
DIN	0.6607	0.00%	0.6343	0.00%
DICN	0.7661	65.59%	0.6350	0.52%
DFCN	0.8313	106.16%	0.6355	0.89%

From Table 2, it can be seen that the DFCN model proposed in this paper outperforms DIN and DICN and far outperforms the rest of the mainstream CTR models, both in the test with the Taobao users' historical behavior dataset and in the test with the Amazon clothes, shoes, and jewelry dataset. When we carefully analyze the indicators and the related data, we can see that DFCN has improved by 0.1706 in the AUC indicator and 106.16% in the RelaImpr-DIN indicator compared to the DIN model for the Taobao user history behavior dataset test. The AUC metric improved by 0.0012 and the RelaImpr-DIN metric improved by 0.89% compared to the DIN model in the Amazon dataset test, which is quite significant. In contrast, when comparing DICN, the results for the AUC indicator derived under the two datasets improved by 0.0652 and 0.0005, respectively, and the results for the RelaImpr-DIN indicator improved by 40.57% and 0.37%, respectively. This is a good example of the superiority of the new DFCN model presented in this paper. A visual comparison of the other DFCN models and the comparison model for the above two evaluation metrics is shown in Figure 5.

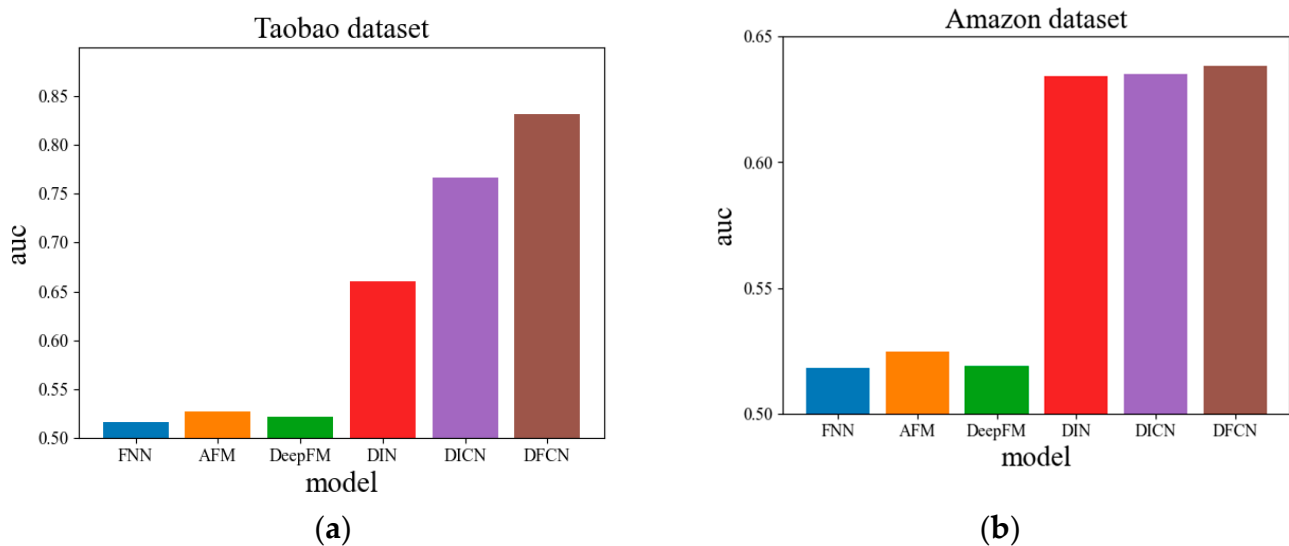


Figure 5. Histogram comparing the AUC metrics of each model, using the Taobao dataset and the Amazon dataset: (a) comparison of the AUC of various models using the Taobao dataset; (b) comparison of the AUC with the Amazon dataset.

4.5.2. Test and Log Loss

In this subsection, we plot the log loss of the different CTR prediction models under the two datasets, tested as the vertical axis, with the number of experimental iterations

shown as the horizontal axis in a line graph. This is intended to explore the loss rate of each model that is tested. See Figure 6 for a folded-line diagram.

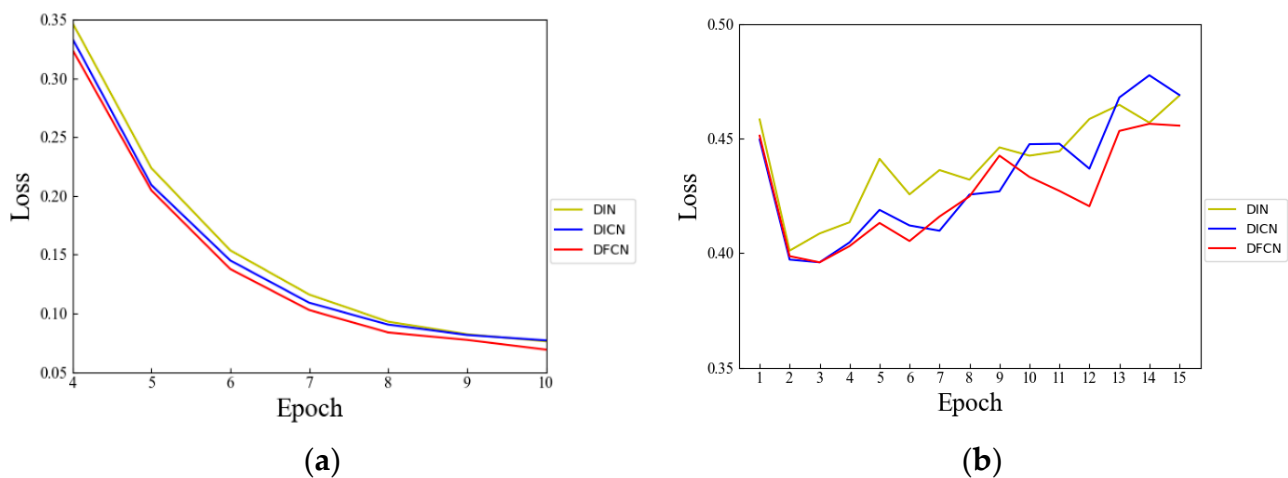


Figure 6. Log loss comparison line chart for each model using different datasets: (a) a comparison of the log loss of each model under the Taobao dataset; (b) a comparison of the log loss of each model under the Amazon dataset.

From the line graph, we can clearly see that the test log loss of DFCN is mostly at its lowest for different numbers of iterations. The test log loss of the Taobao dataset decreases as the number of epoch iterations increases, while the test log loss of the Amazon dataset fluctuates somewhat, but not significantly, and basically tends to be stable and remains at its lowest value. The test log-loss comparison of these two datasets visually illustrates that the DFCN model can effectively avoid over-fitting or under-fitting, ensuring the accuracy of the model's recommendations.

5. Comparisons and Contributions

The method proposed in this paper is a refinement and development of both the classical model and the model previously proposed by our team.

5.1. Comparison to the Classical Models and DICN Models

5.1.1. Comparison to the Classical Models

Compared with the classical model, firstly, the DFCN model proposed in this paper retains the local activation unit introduced by the Alibaba team in the DIN model, as well as the variable length sparse vector, which means that the user's historical behavior features are no longer embedded as a fixed-length vector, but instead interact with the target advertisement in the local activation unit; finally, the weight matrix is obtained according to relevance. Secondly, an obvious shortcoming in the DIN model is that although Zhou et al. introduced a context feature variable, they did not make good use of it. For this reason, we continue to adopt the strengths of our previous model, namely, the introduction of another new local activation unit, to investigate the relevance of the contextual features in the user's historical behavioral patterns to the contextual features of the target advertisement. This improvement is extremely suitable for those cases where the target advertisement itself is closely related to the context of the environment and greatly improves the accuracy of the recommendation system.

5.1.2. Comparison to the DICN Model

Compared to the DICN model, the DFCN model proposed in this paper incorporates an additional filtering layer. The core role of this layer is to filter the sequence of users' historical behavior features, according to the target advertisement. The filtering layer converges with the attention mechanism in terms of its main purpose, which is to ignore

data that are less relevant to the target advertisement and value data that are more relevant to the target advertisement. However, the filtering layer is simpler than the local activation unit structure and does not require a separate deep learning network to learn the non-linear relationships between vectors, which means that those features of the users' historical behavior that are highly relevant to the target advertisement can be filtered out in less time and at a lower cost. However, the filtering layer also has certain drawbacks. Due to the simplicity of its structure, the non-linear relationship between the user's historical behavior features vector and the target advertising vector cannot be mined; therefore, the filtering layer cannot replace the local activation unit to obtain the weight matrix. The filtering layer can pre-process those data with simple linear relationships to help the local activation unit to exclude some variables with very low correlation and help the attention mechanism of the local activation unit to learn non-linear relations, thus improving the reliability of the weight matrix output by the local activation unit. This experiment also designed a DFCN model without the need for any local activation unit, and verified by the use of comparison experiments that the filtering layer cannot replace the local activation unit. However, the expressiveness of the model, with the filtering layer and the local activation unit together, is higher than that of the model with only the local activation unit. The DFCN model without any local activation units was compared with a DIN model with only local activation units and no filtering layer, then the complete DFCN model was input into the Amazon dataset to derive the AUC values for each of the three models. The conclusions were verified by comparing the AUC values of the three models. A visual comparison is presented in Figure 7.

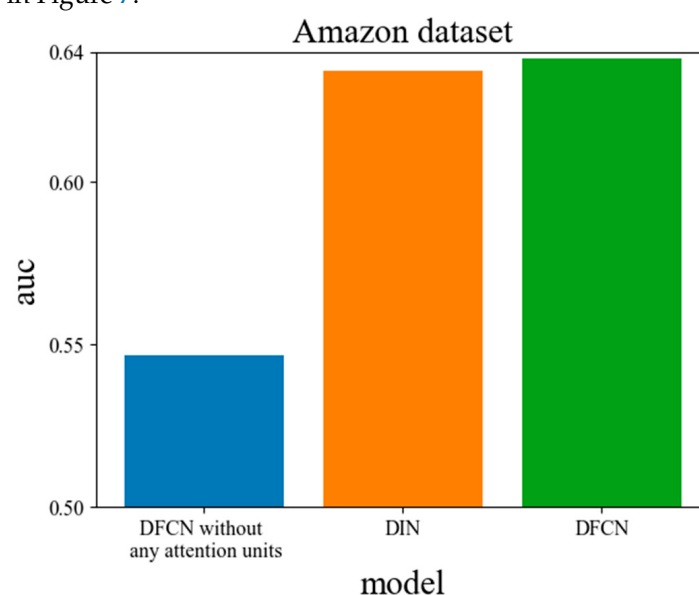


Figure 7. Visual comparison of the AUC of the three models.

It is clear from Figure 7 that the expressive ability of the model would be greatly reduced if only the filtering layer were added without the local activation unit implementing the attention mechanism. The filtering layer is, therefore, the proverbial icing on the cake for the overall model but cannot directly replace it for expressing the non-linear relationship between the target advertisement and the user's historical behavior.

5.1.3. Importance of the Context Feature Attention Unit

In the course of our experiments, we also designed a DFCN model without the introduction of local activation units for context features, i.e., we directly used the original DIN model to add the new filtering layer proposed in this paper, to investigate whether the addition of more local activation units would have an overfitting side effect on the model and, thus, reduce the predictive power of the model. The results demonstrate that the complete DFCN model has better expressive abilities compared to the DFCN model,

which lack context features and local activation units, with better values of AUC for both the Taobao user history behavior dataset and the Amazon dataset, as well as a smaller test log-loss value. A visual comparison is presented in Figure 8.

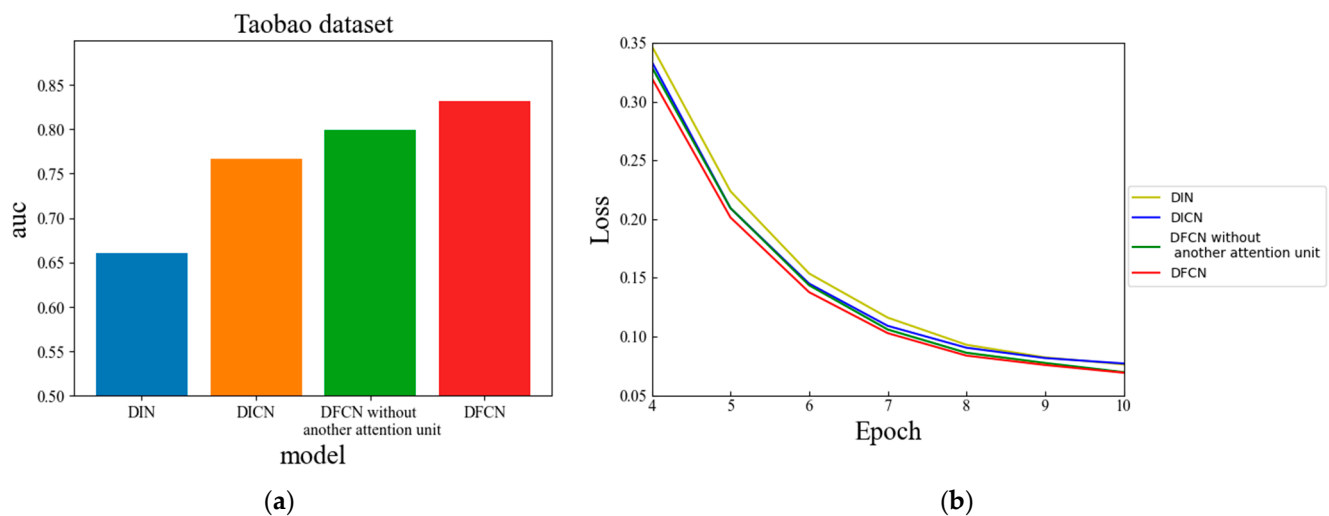


Figure 8. Comparison of the no-context feature attention unit model with the full DFCN model: (a) a comparative histogram of AUC evaluation metrics for the four models, using the Taobao dataset; (b) a line chart comparing the log-loss evaluation metrics of the four models, using the Taobao dataset.

From Figure 8, we can clearly see that the AUC results are also significantly better than the original model DIN, as well as the previously proposed DICN model when the filtering layer proposed in this paper was added. However, the lack of a local activation unit made it impossible to generate a weight matrix based on the correlation between contextual features and the user's historical behavior features, which reduces the expressive and predictive power of the model, to some extent.

5.2. Contributions

The main contribution of this paper is to propose a method for processing the user's historical behavior sequence feature data, i.e., a filtering layer. By means of a linear operation between the target advertising data and the user history behavior sequence data, the user history behavior sequence is primed and filtered for the next step regarding data entry into the local activation unit and, thus, the relevance weights are obtained. The filtering layer improves the representation of the user's historical behavioural features, first, by eliminating the complex and time-consuming deep learning module by means of linear operations in the middle, and second, by filtering the vector of the user's historical behavioral features with high relevance to the target advertisement by means of simple operations. At the same time, this paper also demonstrates that the filtering layer cannot completely replace the local activation unit because of its simplicity, which prevents the filtering layer from learning the non-linear relationships between vectors. This also illustrates the importance of local activation units from another perspective.

6. Conclusions

In an era of diversified market economies, the Internet economy accounts for an increasing share of the overall economy [38], and the rise and continuous development of e-commerce promote the progress of recommendation systems. As a data mining model, recommendation systems can analyze data in detail to help e-commerce companies to improve their decision-making, increase operational efficiency, and provide a better service to their customers [39]. In this paper, the DFCN model proposed in the context of e-commerce display advertising filters the huge volume of users' historical behavioral feature data, effectively suppressing the interference of non-relevant user history features

with the relevant features and helping the attention mechanism to assign weights to each element more precisely and effectively. At the same time, this paper continues the design advantages of DICN, focusing on the contextual variables in the users' historical behavior features and making full use of the context features. After conducting several comparison tests, it is proved that the DFCN model proposed in this paper has significantly improved recommendation accuracy and the loss rate, and can achieve greater learning ability and more accurate and efficient recommendations.

The limitations of this study mainly stem from the newly proposed filtering layer and the local activation unit. For the filtering layer, this study simply performs a linear addition and subtraction operation between the users' historical behavior features and the target advertising sequence. The resulting filter is used directly to filter the data regarding the users' historical behavior features that have a high correlation with the target advertisement. This saves time and money but, to some extent, it ignores the non-linear relationship between the user's historical behavior features and the targeted advertisements, which causes the local activation unit to experience some limitations. In future work, we intend to continue to follow up on the filter layer, balancing the cost of time against the filter layer's ability to filter the data. For the attention local activation unit, more meaningful local activation units can be added, thereby mining other vectors for correlation in the user's historical behavioral features and further improving the model.

Author Contributions: Conceptualization, M.Y.; methodology, M.Y.; software, M.Y.; validation, M.Y.; formal analysis, M.Y.; data curation, M.Y.; writing—original draft preparation, M.Y.; writing—review and editing, M.Y., T.L. and J.Y.; visualization, M.Y. and T.L.; project administration, J.Y.; funding acquisition, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under grant 61971268.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank the National Natural Science Foundation of China for funding our work, grant number 61971268.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yuan, L. Exploring the operation models and development trends of internet finance. *Trade Fair Econ.* **2023**, *3*, 110–112.
2. Zhao, Y.; Liu, H.W. A survey on recommender systems. *Intell. Comput. Appl.* **2021**, *11*, 228–233.
3. Zhang, S.J.; Wang, Y.H. Personalized recommender system based on matrix factorization. *J. Chin. Inf. Process.* **2017**, *31*, 134–139+169.
4. Lops, P.; De Gemmis, M.; Semeraro, G. Content-based recommender systems: State of the art and trends. In *Recommender Systems Handbook*; Springer: Boston, MA, USA, 2011; pp. 73–105.
5. Xue, H.; Dai, X.; Zhang, J. Deep matrix factorization models for recommender system. In Proceedings of the International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 3203–3209.
6. Chen, W.; Niu, Z.; Zhao, X.; Li, Y. A hybrid recommendation algorithm adapted in e-learning environments. *World Wide Web* **2014**, *17*, 271–284. [[CrossRef](#)]
7. Huang, L.; Jiang, B.; Lü, S.; Liu, Y.; Li, D. Survey on deep learning based recommender systems. *Chin. J. Comput.* **2018**, *41*, 1619–1647.
8. McMahan, H.B.; Holt, G.; Sculley, D.; Young, M.; Kubica, J. Ad click prediction: A view from the trenches. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; pp. 1222–1230.
9. Deng, L.; Liu, P. Research on ad click-through rate prediction based on gmm-fms. *Comput. Eng.* **2019**, *45*, 122–126.
10. Steffen, R. Factorization machines. In Proceedings of the 10th International Conference on Data Mining, Sydney, Australia, 13–17 December 2010; pp. 995–1000.
11. Guo, H.; Tang, R.; Ye, Y.; Li, Z.; He, X. Deepfm: A factorization-machine based neural network for ctr prediction. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 1725–1731.

12. He, X.N.; Chua, T.S. Neural factorization machines for sparse predictive analytics. In Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval, Tokyo, Japan, 7–11 August 2017; pp. 355–364.
13. Lian, J.X.; Zhou, X.H.; Zhang, F.Z.; Chen, Z.X.; Xie, X.; Sun, G.Z. xDeepFM: Combining explicit and implicit feature interactions for recommender systems. In Proceedings of the 24th ACM SIGKDD International Conference, London, UK, 19–23 August 2018; pp. 1754–1764.
14. Qu, Y.; Han, C.; Kan, R.; Zhang, W.; Wang, J. Product-based neural networks for user response prediction. In Proceedings of the 2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain, 12–15 December 2016; pp. 1149–1154.
15. Zhou, G.R.; Song, C.R.; Zhu, X.Q.; Fan, Y.; Zhu, H.; Ma, X.; Yan, Y.H.; Jin, J.Q.; Li, H.; Gai, K. Deep interest network for click-through rate prediction. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 1059–1068.
16. Yang, L.; Wang, S.; Zhu, B. Point-of-interest recommendation algorithm combining dynamic and static preferences. *J. Comput. Appl.* **2021**, *41*, 398–406.
17. Zhou, G.R.; Mou, N.; Fan, Y.; Pi, Q.; Bian, W.J.; Zhou, C.; Zhu, X.Q.; Gai, K. Deep interest evolution network for click-through rate prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 5941–5948.
18. Feng, Y.F.; Lv, F.Y.; Shen, W.C.; Wang, M.H.; Sun, F.; Zhu, Y.; Yang, K.P. Deep session interest network for click-through rate prediction. In Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019; pp. 2301–2307.
19. Yu, M.T.; Liu, T.T.; Yin, J.; Chai, P.L. Deep interest context network for click-through rate. *Appl. Sci.* **2022**, *12*, 9531. [[CrossRef](#)]
20. Treisman, A.M.; Gelade, G. Feature-integration theory of attention. *Cogn. Psychol.* **1980**, *12*, 97–136. [[CrossRef](#)]
21. Chen, H.H.; Wu, G.D.; Li, J.X.; Wang, Y.J.; Tao, H. Research advances on deep learning recommendation based on attention mechanism. *Comput. Eng. Sci.* **2021**, *43*, 370–380.
22. Tang, J.Y.; Wang, C.Z.; Ren, X.F.; Yang, X.X. Research on the design and simulation of band stop filter. *Instrum. Technol.* **2023**, *1*, 64–68.
23. He, Y.; An, X. Design of an infinite impulse response chebyshev digital bandpass filter. *Pract. Electron.* **2023**, *31*, 96–99.
24. Wang, P.B.; Yan, Y.E. Design of bandpass and low-pass cascade filter. *Chin. J. Electron Devices* **2018**, *41*, 1473–1476.
25. Zhang, Y.Y. Application of filtering algorithm in digital image denoising. *Autom. Appl.* **2020**, *12*, 49–51.
26. Zhou, Z.M.; Song, L.X. Research on the application of frequency domain filter in digital image processing. *China Comput. Commun.* **2021**, *33*, 198–200.
27. Liu, H.L.; Tao, J.; Qiu, L. Implementation of one-hot encoding based on python. *J. Wuhan Inst. Shipbuild. Technol.* **2021**, *20*, 136–139.
28. Jiang, A.B.; Wang, W.W. Research on optimization of relu activation function. *Transducer Microsyst. Technol.* **2018**, *37*, 50–52.
29. Huang, G.H.; Lin, G.D.; Wu, E.J.; Zhao, X.D.; Song, L.L. Design of fixed-point algorithm for softmax of dnn. *China Integr. Circuit* **2022**, *31*, 60–64.
30. Zhu, H.; Li, X.; Zhang, P.Y.; Li, G.Z.; He, J.; Li, H.; Gai, K. Learning tree-based deep model for recommender systems. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 1079–1089.
31. Zhu, H.; Chang, D.Q.; Xu, Z.R.; Zhang, P.Y.; Li, X.; He, J.; Li, H.; Xu, J.; Gai, K. Joint optimization of tree-based index and deep model for recommender systems. In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; pp. 3971–3980.
32. Zhuo, J.W.; Xu, Z.R.; Dai, W.; Zhu, H.; Li, H.; Xu, J.; Gai, K. Learning optimal tree models under beam search. In Proceedings of the International Conference on Machine Learning, Vienna, Austria, 26–28 August 2020.
33. McAuley, J.; Targett, C.; Shi, Q.F.; Hengel, A.V.D. Image-based recommendations on styles and substitutes. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago de Chile, Chile, 9–13 August 2015; pp. 43–52.
34. He, R.; McAuley, J. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In Proceedings of the 25th International Conference on World Wide Web, Montréal Québec, QC, Canada, 11–15 April 2016; pp. 507–517.
35. Veit, A.; Kovacs, B.; Bell, S.; McAuley, J.; Bala, K.; Belongie, S. Learning visual clothing style with heterogeneous dyadic co-occurrences. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 4642–4650.
36. Zhang, W.N.; Du, T.M.; Wang, J. Deep learning over multi-field categorical data: A case study on user response prediction. In Proceedings of the European Conference on Information Retrieval, Padua, Italy, 20–23 March 2016; pp. 45–57.
37. Xiao, J.; Ye, H.; He, X.N.; Zhang, H.W.; Wu, F.; Chua, T.S. Attentional factorization machines: Learning the weight of feature interactions via attention networks. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 3119–3125.

38. Li, T.R. Research on the marketing models of e-commerce enterprises under the background of the internet economy. *Trade Fair Econ.* **2023**, *4*, 55–57.
39. Mydyti, H.; Kadriu, A.; Bach, M.P. Using Data Mining to Improve Decision-Making: Case Study of A Recommendation System Development. *Organizacija* **2023**, *56*, 138–154. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.