

Spatiotemporal-Enhanced Network for Click-Through Rate Prediction in Location-based Services

Shaochuan Lin¹, Yicong Yu², Xiyu Ji², Taotao Zhou², Hengxu He^{2,†}, Zisen Sang², Jia Jia^{2,†}, Guodong Cao³ and Ning Hu¹

¹Alibaba Group, Hangzhou, China

²Alibaba Group, Shanghai, China

³Alibaba Group, Beijing, China

Abstract

In Location-Based Services(LBS), user behavior naturally has a strong dependence on the spatiotemporal information, *i.e.*, in different geographical locations and at different times, user click behavior will change significantly. Appropriate spatiotemporal enhancement modeling of user click behavior and large-scale sparse attributes is key to building an LBS model. Although most of existing methods have been proved to be effective, they are difficult to apply to takeaway scenarios due to insufficient modeling of spatiotemporal information. In this paper, we address this challenge by seeking to explicitly model the timing and locations of interactions and proposing a Spatiotemporal-Enhanced Network, namely StEN. In particular, StEN applies a Spatiotemporal Profile Activation module to capture common spatiotemporal preference through attribute features. A Spatiotemporal Preference Activation is further applied to model the personalized spatiotemporal preference embodied by behaviors in detail. Moreover, a Spatiotemporal-aware Target Attention mechanism is adopted to generate different parameters for target attention at different locations and times, thereby improving the personalized spatiotemporal awareness of the model. Comprehensive experiments are conducted on three large-scale industrial datasets, and the results demonstrate the state-of-the-art performance of our methods. In addition, we have also released an industrial dataset for takeaway industry to make up for the lack of public datasets in this community.

Keywords

spatiotemporal systems, click-through rate prediction, location-based services

1. Introduction

Location-Based Services (LBS) are mobile services that provide the user with current location-relevant content on smartphones or other services. Among them, takeaway service is the most popular and convenient commercial service. Like other LBS, it also requires timely delivery, which results in a strong dependence on time and geographical location for users. In this way, recommending products suitable for the user's temporal and spatial demands in LBS is a pretty challenging problem.

Recently, some methods[1, 2, 3] have been proved effective in e-commerce through the user's historical behavior, but it is not easy to adapt them into the LBS scenario. The main reason is that most of them do not pay attention to users' strong spatial and temporal demands. For instance,

a user prefers fast food in the work area on weekdays and may choose fried chicken in his or her residential area on weekends. This changes in user behavioral interests are bonded with the changes of location and time. Although there are some initial efforts[4, 5] to integrate spatiotemporal information into sequential recommendation, most of them consider partial spatiotemporal information, and efforts to fully and thoroughly model such integrated spatiotemporal patterns are still lacking. Different from the above scenarios, there are some common attributes in the takeaway scenario which have a weak correlation with the user's historical behavior. For example, milk tea is naturally suitable to be recommended at afternoon tea. On the other hand, the historical behaviors of users imply their personal dietary preferences.

To tackle above problems, we propose a Spatiotemporal-Enhanced Network(StEN), to better meet users' temporal and spatial demands. Specially, StEN applies Spatiotemporal Profile Activation (StPro) module to model user's common spatiotemporal preference by activating attribute features (user and item). For the personalized spatiotemporal preference of users, a novel Spatiotemporal Preference Activation (StPre) and a Spatiotemporal-aware Target Attention (StTA) module are proposed. StPre disassembles the spatiotemporal preference embodied by the user's historical behavior in detail, which including Temporal Evolving

DL4SR'22: Workshop on Deep Learning for Search and Recommendation, co-located with the 31st ACM International Conference on Information and Knowledge Management (CIKM), October 17-21, 2022, Atlanta, USA

*Corresponding author.

✉ lin.lsc@alibaba-inc.com (S. Lin); yicongyu.yyc@alibaba-inc.com (Y. Yu); jixiyu.jxy@alibaba-inc.com (X. Ji); taotao.zhou@lazada.com (T. Zhou); hengxu.hhx@alibaba-inc.com (H. He); zisen.szs@koubai.com (Z. Sang); jj229618@alibaba-inc.com (J. Jia); guodong.cao@alibaba-inc.com (G. Cao); huning.hu@alibaba-inc.com (N. Hu)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Activation(TEA), Temporal periodic Fusion(TPF) and Spatial Preference Activation(SPA). While StTA employs different spatiotemporal information to generate different parameters and feed them into target attention to improve the personalized spatiotemporal awareness of the model. In addition, we have released an industrial dataset for takeaway industry to make up for the lack of public datasets in this community.

All our contributions can be summarized as follows:

- StEN applies **Spatiotemporal Profile Activation (StPro)** module to model user’s common spatiotemporal preference by activating attribute features (user and item).
- For the personalized spatiotemporal preference of users, a novel **Spatiotemporal Preference Activation (StPre)** is proposed, which disassembles the spatiotemporal preference embodied by the user’s historical behavior in detail, and extracts preferences from three small modules: Temporal Evolving Activation (TEA), Temporal Periodic Fusion (TPF) and Spatial Preference Activation (SPA).
- We also propose a **Spatiotemporal-aware Target Attention (StTA)** module, which employs different spatiotemporal information to generate different parameters and feed them into target attention to improve the personalized spatiotemporal awareness of the model
- In addition, we have also released an industrial dataset for takeaway industry to make up for the lack of public datasets in this community. Experimental results demonstrate that our method has achieved the state-of-the-art on three large-scale industrial datasets and the online A/B testing results further show its practical value.

2. Related Work

2.1. Sequence-based Model

Earlier deep CTR approaches hope to eliminate the complicated work of feature engineering jobs and focus more on automatically mining the correlations between features[6, 7, 8, 9, 10]. Later on, researchers[1, 2, 3] found that the users’ historical behavior sequence contains richer and more direct information, which brought breakthroughs to the entire recommendation community. Many researches focus on exploring potential interests in the user’s historical behavior sequence. They extract sequence features by incorporating structures such as Pooling, RNN, and Attention into the model. YoutubeDNN[11] proposes a feature embedding on items method and then takes the average value to extract historical sequence features. DIN[1] believes that interests

in the user’s historical behavior sequence are diverse. Faced with a particular product, only part of the interests associated with that product will influence user’s behavior. Based on this, DIN designs a local activation module to extract different user interests from the sequence for various target commodities. DIEN[2] further explores the interrelationships between users’ historical behaviors and proposes the concept of user interest evolution. It designs an auxiliary loss and a structure based on GRU. Inspired by the success of the self-attention mechanism in sequence-to-sequence tasks, BST[12] leverages a transformer layer instead of GRU to mine information about the user’s interest. DSIN[3] observes that the user’s interests in a short period are concentrated, while long-term interests are scattered. It splits the sequence into different sessions and explores the information through the self-attention mechanism and Bi-LSTM module. SIM[13] proposes an interest mining method for life-long user sequences. However, all historical behavior sequences of users are very long, which may lead to time-consuming and noise problems. To overcome this, SIM provides a search-based long sequence extraction method to extract top-k behavior sequences from life-long sequences through soft and hard search technology.

2.2. Time Aware Attention Model

The above deep CTR models do not explicitly make use of the click time information in the user’s historical behavior, where the click time information has an impact on the user’s evolutionary behavior and the user’s periodic behavior. The user’s evolutionary behavior denotes that the user’s interest changes over time, and the user’s periodic behavior indicates the user’s periodic actions. Specially, TIEN[14] pays more attention to the user’s evolutionary behavior, and believes that the closer the historical behavior is to the current time, the greater the weight should be. TLSAN[15] leverages the absolute value of the time difference and then uses its reciprocal as the time position embedding. TiSASRec[16] models items’ relative time intervals by sine and cosine function to explore the evolutionary behavior of users and then utilizes items’ absolute temporal signals, such as month(M), weekday(W), date(D) and hour(H), to detect periodic behavior of users. TimelyRec[17] captures potential irregularity information in user’s periodic patterns, and then integrates the information to compute the similarity between target time and users interactions with an attention mechanism.

2.3. Spatiotemporal Model

Spatial location is also important for some location-aware platforms, such as Facebook Places[18] and Airbnb[19]. Thus, it is a natural way to integrate temporal in-

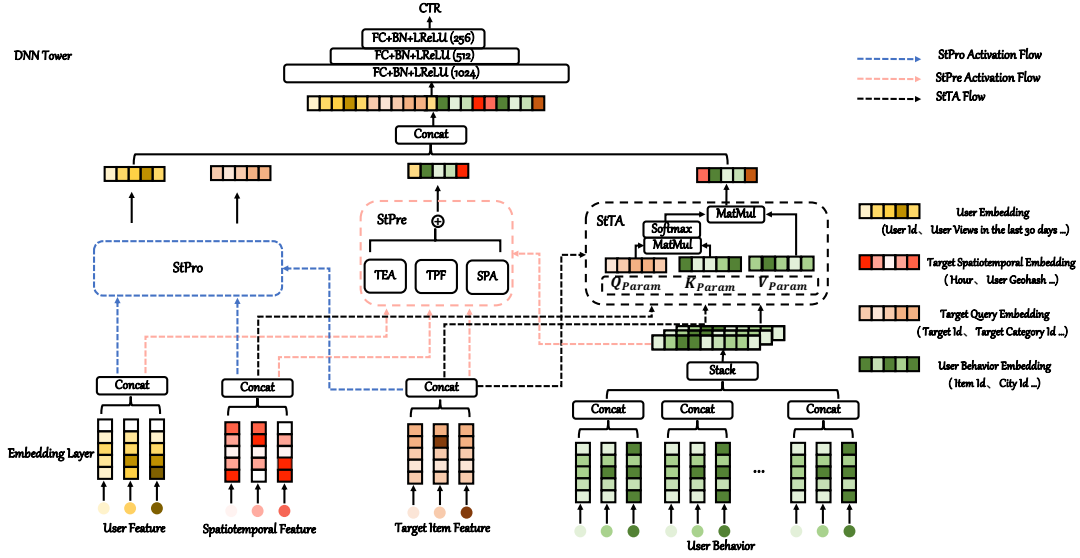


Figure 1: Our StEN consists of three modules: Spatiotemporal Profile Activation(StPro), Spatiotemporal Preference Activation(StPre) and Spatiotemporal-aware Target Attention(StTA).

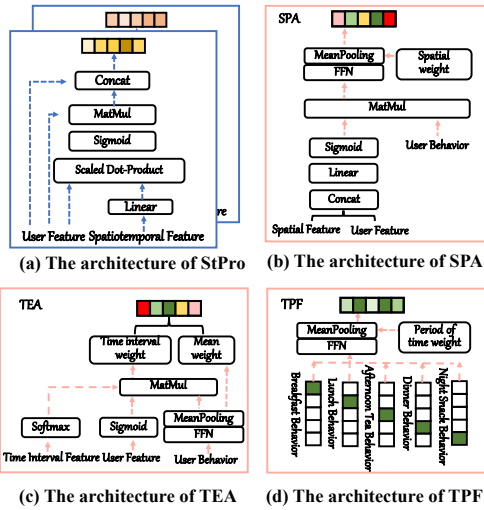


Figure 2: The architecture of Spatiotemporal Profile Activation(StPro) and Spatiotemporal Preference Activation(StPre). StPre includes three models: Temporal Evolving Activation(TEA), Temporal periodic Fusion(TPF) and Spatial Preference Activation(SPA).

formation and spatial location to optimize recommendation models. However, due to the complexity of model design, publicly available existing work is limited. CalendarGNN[20] utilizes GNN and GRU to extract

the segmented time and geographic information in the user’s historical behavior sequence. While effective, it is applied to article browsing of web pages without regard to the geographic location of the item. So it is not suitable for our takeaway industry. TRISAN[21] extracts the spatiotemporal information from the user’s historical behavior sequence by employing two spatial activation and one temporal similarity activation modules in the model. However, it does not detail the information contained in the user’s spatiotemporal behavior, which leads to insufficient spatiotemporal information exploring. While TRISAN is of great relevance for our purposes, unfortunately, the method has not been open-sourced and the dataset used in this paper is not publicly available. So we cannot perform method comparisons with it in the Section 4.

3. Spatiotemporal-Enhanced Network

3.1. Preliminary

In this paper, we denote $x = (m, u, st, b) \in \mathcal{X}$ as input data, where m is the target item feature, u is the user, b is the user click behavior and st is the spatiotemporal feature.

In particular, we geocode¹ the user’s latitude and longitude and convert them to hexadecimal numbers to obtain

¹<https://en.wikipedia.org/wiki/Geohash>

geohash-6, which is then combined with the user’s Area-of-Interest(AOI)[22] and serve as the spatial feature g in this paper. While the temporal feature is represented by hour of day, time period of day(breakfast, lunch, afternoon tea, dinner and night snack) and day of the week. User features u include user id, user gender and other features, while item features i include item id, item category and other features. Before all features enter the model, we will perform a vectorized representation of them. For the convenience of description, in the latter part of this article, m, u, st, b all represent the embedding vectors of the corresponding features. Denoting $y \in \mathcal{Y}$ as the click label, and our CTR prediction task can be defined as:

$$\mathcal{P}(y = 1|x) = f(x; \theta)(x \in \mathcal{X}) \quad (1)$$

where $f(x; \theta)$ is a probability value obtained after we forward the input data x into any CTR network, and then activate by a sigmoid function. θ represents the parameters of the network. Typically, each of our user history behaviors includes the item v , the item’s location l , the click time t and the click period of time p . The CTR task of Equation 1 above is then mainly achieved by minimizing the following cross-entropy loss function during training,

$$\mathcal{L}(f, x_i, y_i) = \frac{1}{N} \sum_{i=1}^N -y_i \log f(x_i; \theta) - (1 - y_i) \log(1 - f(x_i; \theta)) \quad (2)$$

where $y_i \in \{0, 1\}$ is the ground-truth label, N is the mini-batch size and i is the index of the input data. We set N to 1024 in this paper.

3.2. Spatiotemporal Profile Activation

This module is mainly used to capture common spatiotemporal preferences that are less correlated with user behavior. E-commerce scenarios only need to consider the personalized behavior of user, but in the takeaway scenario, we need to consider the impact of time and location on users and items. For instance, there is a natural difference between the user’s order in the workplace and the residential area. Therefore, we use spatiotemporal features st to extract common spatiotemporal preference for the static item and user features. Below we will take the user feature as an example,

$$att_u = \text{sigmoid}\left(\frac{FC_u(st) \cdot u^T}{\sqrt{d_u}}\right)u \quad (3)$$

where $FC_u(st) \in \mathbb{R}^{d_{st} \times d_u}$, is the linear transformation of the st , d_u is the last dimension of u , d_{st} is the last dimension of st . Inspired by [1], we then concatenate u and att_u and add their differences, their common values, to get the final activation value $h_u = \text{concat}(u, att_u, u - att_u, u * att_u)$.

Through the above same activation method, we can obtain the final activation value of the item and is denoted as h_m . Finally, we concat the above activation values to obtain the spatiotemporal profile activation value h_{stpro} . Fig. 2(a) shows the structure.

3.3. Spatiotemporal Preference Activation

We further propose a Spatiotemporal Preference Activation(Stpre) to model the personalized spatiotemporal preference embodied by user behaviors in detail.

3.3.1. Temporal Evolving Activation(TEA)

The time sequence of user clicks will have a certain impact on the current behavior. For example, a user who frequently clicks on milk tea in a short period of time will cause him to be more willing to click on dessert in the next time slot. To model this temporal evolving pattern, we first calculate the time interval t_i between request time t_r and each historical behavior click time t^j . Then we eliminate the noise by applying a nonlinear transformation to the time interval, thus obtaining the temporal evolution factor f_{te} ,

$$f_{te} = FC_2(\text{LeakyReLU}(FC_1(e^{-t_i}))) + e^{-t_i} \quad (4)$$

where $FC_1 \in \mathbb{R}^{N \times N_h}$ and $FC_2 \in \mathbb{R}^{N_h \times N_i}$ denotes two fully connected layers, $t_i \in \mathbb{R}^{N \times N_i}$, N_h is the hidden size, and N_i is the sequence length we set. In this paper, we abbreviate the structure of Equation 4 as FFN. Then we normalize the above temporal evolution factor f_{te} through a softmax function to get the weight of temporal evolution w_{te} . After that, w_{te} can help to obtain temporal activation features related to the behavior order,

$$att_{tea} = w_{te} \cdot \text{FFN}(\text{sigmoid}(FC_t(u)) \cdot b) \quad (5)$$

where $FC_t(u) \in \mathbb{R}^{N_u \times 1}$, N_u is the last dimension of the feature u . Finally our robust temporal evolution fusion feature be obtained by $h_{tea} = w_m * \text{MeanPooling}(\text{FFN}(b)) + w_{tea} * att_{tea}$. Mean weight w_m and time interval weight w_{tea} are two trainable weight parameters used to balance the output. The module is depicted in Fig. 2(c).

3.3.2. Temporal periodic Fusion(TPF)

User historical behavior contains rich but scattered behavioral interests. However, when we explore user behavior from the perspective of time period, we are pleased to find that users’ behavioral interests are more concentrated and periodic. Model would be messy if we directly learn mixed user behavior without any behavioral slices.

In this case, we propose a Temporal periodic Fusion module to learn the user periodic preference in takeaway industry.

Based on the period of time p , we first divide the user historical behavior b into five time slices $b = \{b_{pb}, b_{pl}, b_{pt}, b_{pd}, b_{ps}\}$. Then we feed each period of time sequence into the FFN and mean pooling in turn to get the characteristics of breakfast behaviors $mean_{pb}$, lunch behaviors $mean_{pl}$, afternoon tea behaviors $mean_{pt}$, dinner behaviors $mean_{pd}$, and night snack behaviors $mean_{ps}$. Take the breakfast behavior as an example,

$$mean_{pb} = MeanPooling(FFN(b_{pb})) \quad (6)$$

Further, to obtain a more general periodic representation h_{tpf} , we fuse the above periodic characteristics through an average operation. Fig. 2(d) illustrates an outline of this architecture.

3.3.3. Spatial Preference Activation(SPA)

User’s geographic location affects his personalized dietary choices. For example, when the user works in company, he may choose rice, and when the user is at home, he may prefer fried chicken. We call this the user’s spatial preference. To capture this spatial preference, we utilize the spatial features g and combine them with the user’s feature u . We then feed the above-combined values into a fully connected layer and activate through a sigmoid function to get the geolocation activation value of q_{spa} .

$$q_{spa} = sigmoid(FC_q(concat(g, u))) \quad (7)$$

where $FC_q \in \mathbb{R}^{N_{gu} \times 1}$, N_{gu} is the dimension of the combine value g and u . Further, we use q_{spa} to activate all of the user history behavior to explore the user’s spatial preferences h_{spa} through FFN and mean pooling. The architecture can be observed in Fig. 2(b).

Finally, we fuse the output of the above three small modules together to obtain our final spatiotemporal preference activation value $h_{stpre} = h_{tef} + w_{tpf} * h_{tpf} + w_{spa} * h_{spa}$. Period of time weight w_{tpf} and spatial weight w_{spa} are also two trainable weight parameters used to balance the output.

3.4. Spatiotemporal-aware Target Attention

To more effectively explore the spatiotemporal relationships between historical user behavior and target item, we propose a Spatio-temporal-aware Target Attention(StTA) mechanism. Drawing on the ideas of CAN[23] and AdaptPGM[24], we generate different parameters through spatiotemporal information for target attention, thereby improving the personalized spatiotemporal

awareness of the model. Taking W_Q, b_Q as an example, we can get that,

$$Q_{Param} = W_q \cdot st + b_q \rightarrow W_Q, b_Q \quad (8)$$

where $W_q \in \mathbb{R}^{D \times (d_i * d_o + d_o)}$ and $b_q \in \mathbb{R}^{d_i * d_o + d_o}$ are the parameters of a fully-connected layer. D is the dimension of st , d_i is the dimension of input embedding (such as target item embedding m or user behavior embedding b) and d_o is the dimension of final output embedding. Then we can split Q_{Param} into two parts(W_Q, b_Q) as parameters of the subsequent target attention fully connected layer. Specially, we take the first $d_i * d_o$ parameters as W_Q and the last d_o parameters as b_Q . In the same way, we can obtain $K_{Param}(W_K, b_K)$ and $V_{Param}(W_V, b_V)$ through the spatiotemporal feature st . After that, we utilize the primitive target attention mechanism to obtain the final module output h_{ta} ,

$$\begin{aligned} Q &= W_Q \cdot m + b_Q, \\ K &= W_K \cdot b + b_K, \\ V &= W_V \cdot b + b_V \\ h_{ta} &= softmax\left(\frac{QK^T}{\sqrt{d_K}}\right)V \end{aligned} \quad (9)$$

Where d_k is the dimension of K . Fig. 1(a) illustrates the structure.

3.5. Dense Tower for StEN

Once we have all the feature vector representations, we can fuse all the above module outputs to get the final prediction $dense_0 = concat(h_{stpro}, h_{stpre}, h_{ta})$. A three-layer perceptron structure is then applied,

$$dense_{i+1} = LeakyReLU(BN(FC_{fi}(dense_i))) \quad (10)$$

where $i = 0, 1, 2$. We then get the prediction of click via a sigmoid activation $\mathcal{P}(y = 1|x) = sigmoid(FC_{sigmoid}(dense_3))$. $FC_{sigmoid} \in \mathbb{R}^{N \times 1}$. Finally, we optimize the parameters of our whole model by Equation 2 defined above. The detail is illustrated in Fig. 1(a).

4. EXPERIMENTS

4.1. Datasets

Due to the lack of public spatiotemporal datasets in the takeaway industry, we conducted experimental comparisons on three industrial datasets (D_1, D_2 and D_3) collected from Ele.me, a major LBS platform in China. The dataset D_1 mainly recommend stores to users, which consists of over 5 billion samples. Dataset D_2 and D_3 mainly recommend meals to users and contain more than

Table 1

Statistics of the dataset used in this paper. ML indicates median length.

Datasets	\mathcal{D}_1	\mathcal{D}_2	\mathcal{D}_3
Total Size	5541799773	575941170	177114244
# Feature	388	218	38
# Users	49249999	28706270	14427689
# Items	2750505	12302502	7446116
# Clicks	343277081	5626279	3140831
ML of User Behaviors	39.66	41.59	41.19

Table 2

Overall performance on \mathcal{D}_1 , \mathcal{D}_2 and \mathcal{D}_3 . StPro: Spatiotemporal Profile Activation. StPre: Spatiotemporal Preference Activation. DIN+StPro+StPre, DHAN+StPro+StPre, DIEN+StPro+StPre are three variation models to investigate the generalization of our module.

Model	\mathcal{D}_1	\mathcal{D}_2	\mathcal{D}_3
DIN	0.7209	0.7294	0.6403
DHAN	0.7265	0.7312	0.6419
DIEN	0.7346	0.7452	0.6531
DIN+StPro+StPre	0.7236	0.7324	0.6434
DHAN+StPro+StPre	0.7271	0.7336	0.6445
DIEN+StPro+StPre	0.7348	0.7458	0.6571
StEN	0.7353	0.7535	0.6627

500 million and 100 million samples, respectively. For \mathcal{D}_3 , we collected one week’s data from the server logs as training set and one day’s data as the test set. We have publicly released the dataset \mathcal{D}_3 ² to further advance the exploration of spatiotemporal patterns in the LBS community. The details of our datasets can be seen in Table 1.

4.2. Experimental Settings

All models in this paper are implemented with Python 2.7 and Tensorflow 1.4. AdagradDecay[25] is chosen as our optimizer to train the model. To avoid overfitting in the early stage of model training and maintain the training stability, we adopt a warm-up[26] strategy for all methods. We set the learning rate to 0.001 and gradually increased it to 0.015 within 1M steps. We set the batchsize N to 1024. We repeated all the experiments five times and averaged the metrics to obtain more reliable results. In our experiments, We adapt Area Under Curve (AUC) and RelAImpr[27] as our evaluation metric.

To show the effectiveness of our method, we select three well-known and industry-proven CTR prediction models as our baselines.

DIN: Deep Interest Network (DIN) designs a local activation module to capture the information in the user behavior sequence that will affect the user behavior when facing the target item. At the same time, DIN does not model the interrelationships among items in a sequence of actions.

DHAN: Deep Hierarchical Attention Networks(DHAN) designs a set of attention networks with multi-dimensional and multi-level structures, which can capture the interest expression of users in various dimensions. At the same time, the attention network can extract features that are similar to the knowledge expression of the tree structure.

DIEN: Deep Interest Evolution Network (DIEN) adapts the interest evolution factors in user behavior. It designs an AUGRU-based module to model the evolution process and trend of user interests.

4.3. Overall Performance Comparison

Table 2 compares StEN with three well-known CTR prediction models on \mathcal{D}_1 , \mathcal{D}_2 and \mathcal{D}_3 . We find that DHAN[28] performs better than DIN[1] on both datasets due to the addition of a multi-dimensional and multi-level attention mechanism. For example, DHAN surpasses DIN on by margins of 0.56% on dataset \mathcal{D}_1 . Notably, 0.1% improvement of AUC is significant for online model deployment to improve the actual CTR in production. Due to the excellent performance of LSTM module in exploring user behavior sequence, DIEN[2] outperforms DHAN in both datasets. However, it is worth noting that recurrent neural networks such as LSTM have slow training and prediction problems and are prone to high response time problems when serving online. By comparison, our StEN advantages all of them to a new level. We have achieved **AUC=0.7353**, **AUC=0.7525** and **AUC=0.6627** on \mathcal{D}_1 , \mathcal{D}_2 and \mathcal{D}_3 , respectively. Our method is 0.96% higher than current best results (DIEN) on dataset \mathcal{D}_3 .

At the same time, to investigate the generalization of our module, we have conducted variation experiments by adding StPre and StPro to the above baseline models. Note that the main difference among the above three methods is the attention module, so our StTA will not

²<https://tianchi.aliyun.com/dataset/dataDetail?dataId=131047>

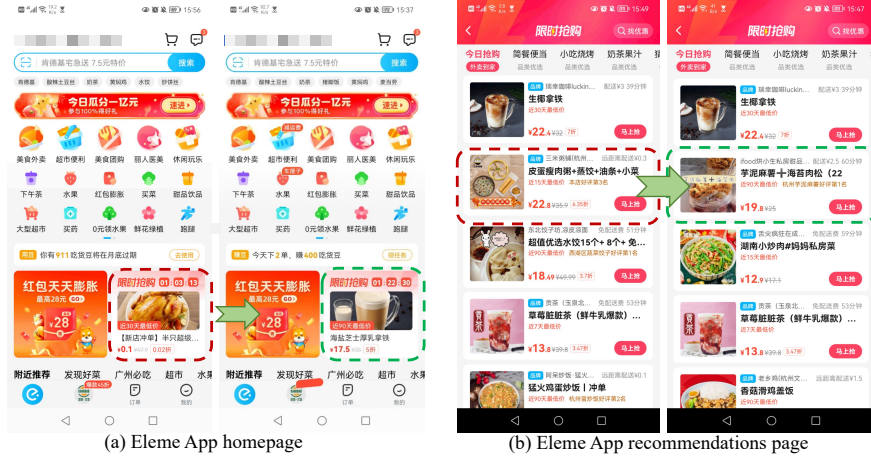


Figure 3: Screenshots of the Eleme mobile App. (a) and (b) are the recommendation results of the online-serving model (red box) and StEN (green box) during afternoon tea, where the right of (a) and (b) (green box) are more suitable for afternoon tea.

Table 3

Ablation study on \mathcal{D}_1 and \mathcal{D}_2 . StPro: Spatiotemporal Profile Activation. TEA: Temporal Evolving Activation. TPF: Temporal Periodic Fusion. SPA: Spatial Preference Activation. StPre: Spatiotemporal Preference Activation. StTA: Spatiotemporal-aware Target Attention.

Methods	\mathcal{D}_1		\mathcal{D}_2	
	AUC	Relalmp	AUC	Relalmp
BaseModel	0.7332	0.00%	0.7414	0.00%
w/ StPro	0.7345	0.56%	0.7474	2.49%
w/ TEA	0.7345	0.56%	0.7500	3.56%
w/ TPF	0.7342	0.43%	0.7479	2.69%
w/ SPA	0.7348	0.69%	0.7476	2.57%
w/ StPre	0.7349	0.73%	0.7521	4.43%
w/ StTA	0.7350	0.77%	0.7499	3.52%
StEN	0.7353	0.90%	0.7535	5.01%

be added to interfere. It can be observed from Table 2 that when we directly adapt our two proposed activation modules to the three baselines mentioned above, there is a certain stable improvement in performance. For example, DIN obtains a significant improvement of 0.27% on \mathcal{D}_1 , 0.30% on the \mathcal{D}_2 and 0.31% on the \mathcal{D}_3 , while DIEN has the weaker improvement of 0.02% on \mathcal{D}_1 , 0.06% on \mathcal{D}_2 and 0.4% on the \mathcal{D}_3 . All these variation models further demonstrate that our proposed modules have good generalizability and can be added to other existing models as a plug-and-play module.

4.4. Ablation Study

To investigate the effectiveness of our proposed method, we conduct ablation studies in Table 3. Our BaseModel

in this paper consists of a primitive Target Attention module mentioned in Section 3.4. Observed from Table 3, each module has played a different positive role after being added.

We then show the effect of Spatiotemporal Profile Activation (StPro) by adding it to the BaseModel. Observed From Table 3, we can see that our "w/ StPro" has brought a relatively stable improvement in effect. In particular, compared to BaseModel, the offline AUC rises from 0.7332 to 0.7345 (+0.13%) and 0.7414 to 0.7474 (+0.6%) when tested on \mathcal{D}_1 and \mathcal{D}_2 , respectively. The results demonstrate that Spatiotemporal Profile Activation is an effective way to model user's common spatiotemporal preference.

Next, we validate the effectiveness of Spatiotemporal Preference Activation (StPre) over the model. As reported in table 3, "w/ StPre" increases the results of "BaseModel" by 0.17% and by 1.07% on dataset of \mathcal{D}_1 and \mathcal{D}_2 , respectively. In order to see the effect of the three small modules (TEA, TPF and SPA) in StPre, we also performed some ablation experiments in Table 3. We can observe that module SPA shows the best performance when tested on dataset \mathcal{D}_1 , while module TEA achieves better performance when tested on dataset \mathcal{D}_2 . This illustrates that in different scenarios, the user's spatiotemporal preferences will focus on different emphasis, specific focus needs to be specifically determined.

We also evaluate the effect of Spatiotemporal-aware Target Attention (StTA) mechanism. In Table 3, we observe a significant improvement after adding Spatiotemporal-aware Target Attention into the system. For example, "w/ StTA" achieves an offline AUC of 0.7350 when tested on the dataset of \mathcal{D}_1 . This is higher than "BaseModel" by 0.18%. The improvement demonstrates that our proposed Target Attention mechanism can meet

the user's spatiotemporal demands compared to the primitive target attention module. Injecting our StTA into the model could improve the effectiveness of system in LBS. Furthermore, our "StEN(StPre+StPro+StTA)" consistently improves the results of "w/ StPre", "w/ StPro" and "w/ StTA". This is because more appropriate spatiotemporal enhancement has been conducted by integrating the three module we proposed in this paper.

4.5. Online A/B Testing

We have deployed our method on the Ele.me platform and conducted an online A/B test for one month in November 2021, which is under the bucket test. One bucket is the BaseModel we have defined in Section 4.4 and the other bucket is our model StEN. Compared with the online-serving BaseModel, our method has increased the CTR of one-hop by 1.6%, the CTR of the second-hop by 2.4%, the order volume by 2.1%, and the order UV by 2.4%. These online benefits from our method are crucial for the recommendation systems of Ele.me. On the one hand, an efficient model can improve user click efficiency. On the other hand, the emphasis on spatiotemporal characteristics can also improve user experience and increase the user stickiness of the platform.

For better understanding, we also compare the recommendation results of the online-serving model with our StEN on the Ele.me platform, as shown in Figure 3. The items (red box) on the left of Figure 3(a) and Figure 3(b) are not suitable for afternoon tea, but are more appropriate for breakfast and staple food, respectively. While our StEN (green box) recommends the sweetmeats and milk tea that are suitable for afternoon tea. Therefore, StEN does a better job of capturing users' strong spatial and temporal demands and can improve the user experience.

5. Conclusions

In this paper, we propose a novel spatiotemporal-enhanced network StEN. In particular, StEN applies a StPro module to capture common spatiotemporal preference by activating attribute features. A StPre module is further applied to model the personalized spatiotemporal preference embodied by the behaviors in detail. Moreover, a StTA mechanism is adopted to generate different parameters for target attention at different locations and times, thereby improving the personalized spatiotemporal awareness of the model. Comprehensive experiments are conducted on three large-scale industrial datasets, and the results demonstrate the state-of-the-art performance of our methods.

References

- [1] G. Zhou, X. Zhu, C. Song, Y. Fan, H. Zhu, X. Ma, Y. Yan, J. Jin, H. Li, K. Gai, Deep interest network for click-through rate prediction, in: Y. Guo, F. Farooq (Eds.), Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018, ACM, 2018, pp. 1059–1068.
- [2] G. Zhou, N. Mou, Y. Fan, Q. Pi, W. Bian, C. Zhou, X. Zhu, K. Gai, Deep interest evolution network for click-through rate prediction, in: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019, AAAI Press, 2019, pp. 5941–5948.
- [3] Y. Feng, F. Lv, W. Shen, M. Wang, F. Sun, Y. Zhu, K. Yang, Deep session interest network for click-through rate prediction, in: S. Kraus (Ed.), Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019, ijcai.org, 2019, pp. 2301–2307.
- [4] Q. Cui, C. Zhang, Y. Zhang, J. Wang, M. Cai, ST-PIL: spatial-temporal periodic interest learning for next point-of-interest recommendation, in: G. Demartini, G. Zuccon, J. S. Culpepper, Z. Huang, H. Tong (Eds.), CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, November 1 - 5, 2021, ACM, 2021, pp. 2960–2964.
- [5] W. Zhang, J. Wang, Location and time aware social collaborative retrieval for new successive point-of-interest recommendation, in: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM '15, Association for Computing Machinery, New York, NY, USA, 2015, p. 1221–1230.
- [6] H. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir, R. Anil, Z. Haque, L. Hong, V. Jain, X. Liu, H. Shah, Wide & deep learning for recommender systems, in: Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, DLRSRecSys 2016, Boston, MA, USA, September 15, 2016, ACM, 2016, pp. 7–10.
- [7] H. Guo, R. Tang, Y. Ye, Z. Li, X. He, Deepfm: A factorization-machine based neural network for CTR prediction, in: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017, ijcai.org, 2017, pp. 1725–1731.

- [8] R. Wang, B. Fu, G. Fu, M. Wang, Deep & cross network for ad click predictions, in: Proceedings of the ADKDD'17, Halifax, NS, Canada, August 13-17, 2017, ACM, 2017, pp. 12:1-12:7.
- [9] J. Lian, X. Zhou, F. Zhang, Z. Chen, X. Xie, G. Sun, xdeepfm: Combining explicit and implicit feature interactions for recommender systems, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018, ACM, 2018, pp. 1754-1763.
- [10] W. Song, C. Shi, Z. Xiao, Z. Duan, Y. Xu, M. Zhang, J. Tang, AutoInt: Automatic feature interaction learning via self-attentive neural networks, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019, ACM, 2019, pp. 1161-1170.
- [11] P. Covington, J. Adams, E. Sargin, Deep neural networks for youtube recommendations, in: S. Sen, W. Geyer, J. Freyne, P. Castells (Eds.), Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, USA, September 15-19, 2016, ACM, 2016, pp. 191-198.
- [12] Q. Chen, H. Zhao, W. Li, P. Huang, W. Ou, 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, 2019, pp. 1-4.
- [13] Q. Pi, G. Zhou, Y. Zhang, Z. Wang, L. Ren, Y. Fan, X. Zhu, K. Gai, Search-based user interest modeling with lifelong sequential behavior data for click-through rate prediction, in: CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020, ACM, 2020, pp. 2685-2692.
- [14] L. Xiang, W. Chao, T. Bin, T. Jiwei, Z. Tao, Deep time-aware item evolution network for click-through rate prediction, in: CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020, ACM, 2020, pp. 785-794.
- [15] J. Zhang, D. Wang, D. Yu, Tlsan: Time-aware long-and short-term attention network for next-item recommendation, volume 441, 2021, pp. 179-191.
- [16] Y. Wang, L. Zhang, Q. Dai, F. Sun, Y. Bao, Regularized adversarial sampling and deep time-aware attention for click-through rate prediction, in: CIKM '19: The 28th ACM International Conference on Information and Knowledge Management, Beijing, China, November 3-7, 2019, ACM, 2019, p. 349-358.
- [17] C. Junsu, H. Dongmin, K. Seongku, Y. Hwanjo, Learning heterogeneous temporal patterns of user preference for timely recommendation, in: WWW '21: Proceedings of the Web Conference 2021, Ljubljana, Slovenia, April 19-23, 2021, ACM, 2021, p. 1274-1283.
- [18] C. Ting, S. Yizhou, Task-guided and path-augmented heterogeneous network embedding for author identification, in: WSDM '17: The 10th ACM International Conference on Web Search and Data Mining, Cambridge, United Kingdom, February 2-6, 2017, ACM, 2017, pp. 295-304.
- [19] Grbovic, Mihajlo, H. Cheng, Real-time personalization using embeddings for search ranking at airbnb, in: KDD '18: The 24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, London, United Kingdom, August 19-23, 2018, ACM, 2018, pp. 311-320.
- [20] D. Wang, M. Jiang, M. Syed, O. Conway, V. Juneja, S. Subramanian, N. V. Chawla, Calendar graph neural networks for modeling time structures in spatiotemporal user behaviors, in: KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020, ACM, 2020, pp. 2581-2589.
- [21] Y. Qi, K. Hu, B. Zhang, J. Cheng, J. Lei, Trilateral spatiotemporal attention network for user behavior modeling in location-based search, in: CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Gold Coast, Australia, November 1-5, 2021, ACM, 2021, pp. 3373-3377.
- [22] Y. Hu, S. Gao, K. Janowicz, B. Yu, W. Li, S. Prasad, Extracting and understanding urban areas of interest using geotagged photos, *Comput. Environ. Urban Syst.* (2015) 240-254.
- [23] G. Zhou, W. Bian, K. Wu, L. Ren, Q. Pi, Y. Zhang, C. Xiao, X. Sheng, N. Mou, X. Luo, C. Zhang, X. Qiao, S. Xiang, K. Gai, X. Zhu, J. Xu, CAN: revisiting feature co-action for click-through rate prediction, *CoRR abs/2011.05625* (2020).
- [24] Shishi, Thinking and practice of alimama's search advertising prediction model 2021, 2022. URL: <https://zhuanlan.zhihu.com/p/446993392>.
- [25] J. C. Duchi, E. Hazan, Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization, *J. Mach. Learn. Res.* 12 (2011) 2121-2159.
- [26] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, IEEE Computer Society, 2016, pp. 770-778.
- [27] L. Yan, W. Li, G. Xue, D. Han, Coupled group lasso for web-scale CTR prediction in display advertising, in: Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014, volume 32 of *JMLR Workshop and Conference Proceedings*, JMLR.org, 2014,

- pp. 802–810.
- [28] W. Xu, H. He, M. Tan, Y. Li, J. Lang, D. Guo, Deep interest with hierarchical attention network for click-through rate prediction, in: Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020, 2020, pp. 1905–1908.