where $\mathbf{W}^I$ has the corresponding shape. Similarly, the adaptive representation of session interests mixed with contextual information w.r.t. the target item is calculated as follows:

$$\begin{aligned} a_k^H &= \frac{\exp(\mathbf{H}_k \mathbf{W}^H \mathbf{X}^I))}{\sum_k^K \exp(\mathbf{H}_k \mathbf{W}^H \mathbf{X}^I)} \\ \mathbf{U}^H &= \sum_k^K a_k^H \mathbf{H}_k \end{aligned} \tag{10}$$

where $\mathbf{W}^H$ has the corresponding shape. Embedding vectors of $User\ Profile$ and $Item\ Profile$, $\mathbf{U}^I$ and $\mathbf{U}^H$ are concatenated, flattened and then fed into the MLP layer.

4 Experiments

In this section, we first introduce experiment datasets, competitors and evaluation metric. Then we compare our proposed DSIN with competitors and analyse the results. We further discuss the effectiveness of critical technical designs in DSIN empirically.

4.1 Datasets

Advertising Dataset

Advertising Dataset[2] is a public dataset released by Alimama, an online advertising platform in China. It contains 26 million records from ad display/click logs of 1 million users and 800 thousand ads in 8 days. Logs from 2017-05-06 to 2017-05-12 are for training and logs from 2017-05-13 are for testing. Users' recent 200 behaviors are also recorded in logs.

Recommender Dataset

To verify the effectiveness of DSIN in the real-world industrial applications, we conduct experiments on the recommender dataset of Alibaba. This dataset contains 6 billion display/click logs of 100 million users and 70 million items in 8 days. Logs from 2018-12-13 to 2018-12-19 are for training and logs from 2018-12-20 are for testing. Users' recent 200 behaviors are also recorded in logs.

4.2 Competitors

- **YoutubetNet**. YoutubeNet [Covington *et al.*, 2016] is a technically designed model which uses users' watching video sequence for video recommendation in Youtube. It treats users' historical behaviors equally and utilizes average pooling operation. We also experiment with YoutubeNet without $User\ Behavior$ to verify the effectiveness of historical behaviors.
- **Wide&Deep**. Wide&Deep [Cheng *et al.*, 2016] is a CTR model with both memorization and generalization. It contains two parts: wide model of memory and deep model of generalization.
- **DIN**. Deep Interest Network [Zhou *et al.*, 2018c] fully exploits the relationship between users' historical behaviors and the target item. It uses attention mechanism to learn the representation of users' historical behaviors w.r.t. the target item.

[2]https://tianchi.aliyun.com/dataset/dataDetail?dataId=56

Model	Advertising	Recommender
YoutubeNet-NO-UB[a]	0.6239	0.6419
YoutubeNet	0.6313	0.6425
DIN-RNN	0.6319	0.6435
Wide&Deep	0.6326	0.6432
DIN	0.6330	0.6459
DIEN	0.6343	0.6473
DSIN-PE[b]	0.6357	0.6494
DSIN-BE-NO-SIIL[c]	0.6365	0.6499
DSIN-BE[d]	**0.6375**	**0.6515**

[a] YoutubeNet without *User Behavior*.
[b] DSIN with positional encoding.
[c] DSIN with bias encoding and without session interest interacting layer and the corresponding activation unit.
[d] DSIN with bias encoding.

Table 1: Results (AUC) on the advertising and recommender dataset

- **DIN-RNN**. DIN-RNN has a similar structure as DIN, except that we use the hidden states of Bi-LSTM, which models users' historical behaviors and learns the contextual relationship.
- **DIEN**. DIEN [Zhou *et al.*, 2018b] extracts latent temporal interests from user behaviors and models interests evolving process. Auxiliary loss makes hidden states more expressive to represent latent interests and AUGRU models the specific interest evolving processes for different target items.

4.3 Metrics

AUC (Area Under ROC Curve) reflects the ranking ability of the model. It is defined as follows:

$$\text{AUC} = \frac{1}{m^+ m^-} \sum_{x^+ \in D^+} \sum_{x^- \in D^-} (I(f(x^+) > f(x^-)))) \tag{11}$$

where D^+ is the collection of all positive examples, D^- is the collection of all negative examples, $f(\cdot)$ is the result of the model's prediction of the sample x and $I(\cdot)$ is the indicator function.

4.4 Results on the advertising and recommender Datasets

Results on the advertising dataset and recommender dataset are shown in Table 1. YoutubeNet performs better than YoutubeNet-No-User-Behavior owing to *User Behavior*, while Wide&Deep gets the betters result due to combining the memorization of wide part. DIN improves AUC obviously by activating *User Behavior* w.r.t. the target item. Especially, the results of DIN-RNN in both datasets are worse than those of DIN due to the discontinuity of users' behavior sequences. DIEN obtains better results while auxiliary loss and specially designed AUGRU lead to deviating from the original expression of behaviors. DSIN gets the best results on both datasets. It extracts users' historical behaviors into session interests and models the dynamic evolving procedure