RECOMMENDATION SYSTEMS

FINAL PROJECT - SPRING 2023

SASRec4SV: Self-Attention Sequential Recommendation for Short Video

Anran Wang, Jiacheng Shen, Krabs Siew

Supervised by

Dr. Hongyi Wen

Contents

1 Introduction

Over the past 5 years, the rapid surge in popularity of short videos has significantly transformed the social media landscape[2]. Figure 1 illustrates the interface of TikTok, a widely-used short video platform. Unlike the long-video watching website YouTube, short video platforms operate under a unique mechanism. Firstly, these platforms present content directly to users without the need for searching. Users either watch the video or scroll down if it does not pique their interest. This sequential interaction between users and the platform gives rise to a session-based recommendation problem. Secondly, the user interface design of short video platforms allows for capturing both explicit and implicit user feedback, including the watch time, likes or dislikes, comments, and sharing behaviors. How to leverage this wealth of feedback to generate accurate and effective short video recommendations within a session-based context poses a meaningful and intriguing research question.

In previous studies, Self-Attention Sequential Recommendation (SASRec) has been widely used to address time series problems[1]. Utilizing the attention mechanism, SASRec can model the entire user sequence, and adaptively considers only recent items for prediction, which could potentially be applied in the short video recommendation scenario. However, it remains unclear how this method can be adapted to incorporate explicit and implicit feedback. Therefore, in this project, our objective is to propose a novel model architecture, called SASRec4SV, specifically tailored for short video recommendation. Subsequently, we aim to evaluate the performance of our model using a real-world dataset, providing insights into its effectiveness.

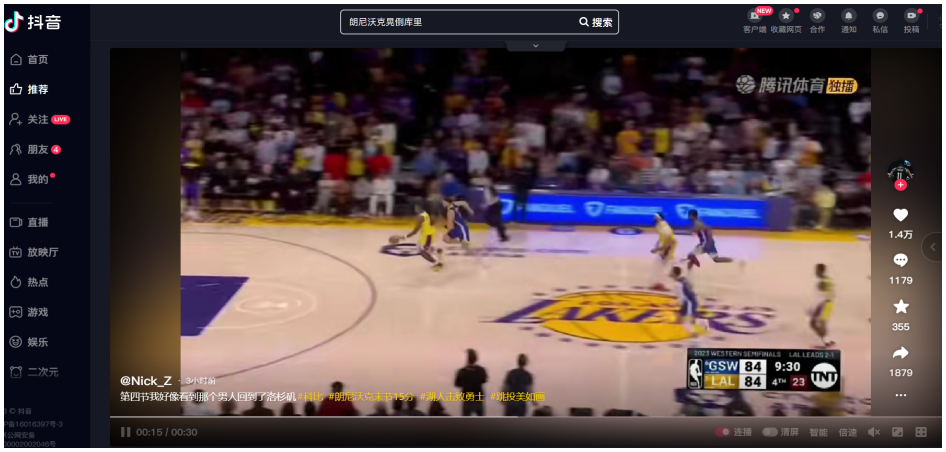

Figure 1: A screenshot of the short video platform TikTok

2 Methods

SASRec4SV SASRec4SV, our new model, is built upon SASRec with several improvements. To handle numerous implicit feedback, such as user likes, comments, and shares, we aggregate the number of interactions into a single variable called auxiliary feedback. We then combine the user ID, video ID, author ID, the auxiliary feedback embeddings, and then add the position embeddings as input for the self-attention layer. We restrict a user's interaction log to the 500 most recent ones, considering the session-based nature of short video recommendation, $S^u = \{S_1^u, S_2^u, \ldots S_{500}^u\}$. For each interaction log, we begin by establishing a timeline mask to prevent the subsequent information from influencing the current output prediction. To accommodate for different video durations, we choose the watch time ratio (WTR) as the ground truth label. We calculate the WTR by dividing the watching time of a video by its total duration, which reflects the relative proportion of the video a user has watched. As the WTR ranges between 0 and 1, we propose the following two approaches to train the model:

1) Given a user and video pair, the model directly outputs the prediction for the WTR. We use ReLU as the activation function and the Huber loss as the loss function. This linear implementation of the model is shown in Figure 2.
2) We transform the WTR into a binary label based on the threshold of 0.5. If the WTR is larger than 0.5, we set the label to 1, meaning the user likes the video. Otherwise, we set the label to 0, indicating that the user does not like the video. In the feed-forward layer, we use the Sigmoid functions as the activation function, as shown in Figure 3. The goal of the model is to understand the user's preferences for each video, and then output the prediction for the most likely video.

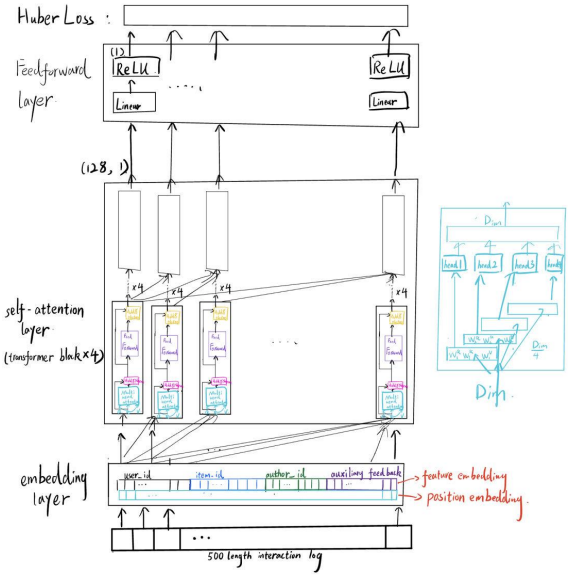

Figure 2 shows the linear type model

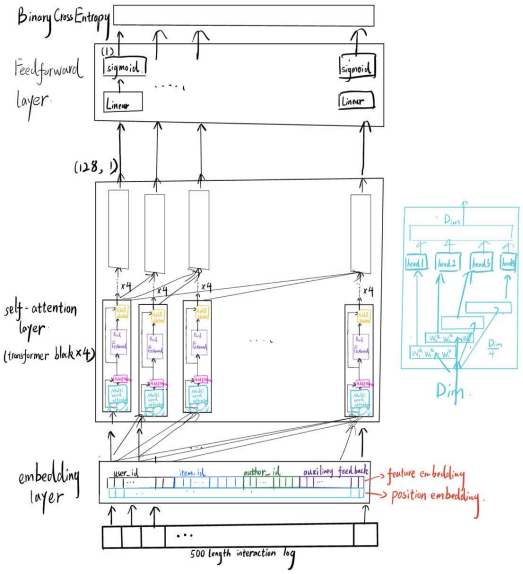

Figure 3 shows the binary type model

Watch Time Gain (WTG) WTG, different from WTR, is another metric that helps us to determine the relevance between a user and a video based on the implicit feedback of watch time. Zheng proposed Watch Time Gain[3], which evaluates the relative watch time among videos of similar durations, enabling a fairer assessment of relevance between a user and a video. By

comparing the watch time of a particular video to the average watch time of other videos with similar durations, WTG provides a more accurate measure of the user's engagement and interest in a specific video than other metrics directly derived from watch time. We apply the WTG metric to process the implicit feedback for the user and derive the relevance between a user and a video. We first divide videos into groups based on the duration of a video. Then within each group, we normalize the watch time and use z-score to reflect the relative watch time. Finally, we set a threshold for the z-score and make a binary classification based on it.

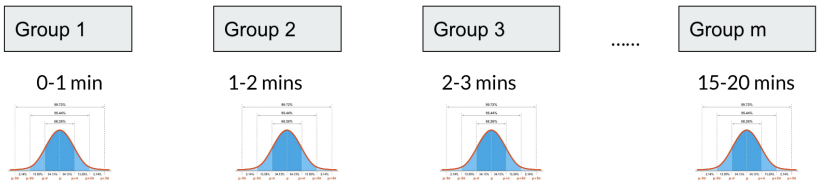

Figure 4 Group videos by duration and normalize the watch time within each group

3 Experiments

In this project, we evaluate our model on the KuaiRand-1k dataset collected from Kuaishou[4]. We split the interaction logs for training, validating, and testing. Specifically, the logs between April 9th and April 20th are the training dataset, the logs on April 21st are the validation set, and the logs on April 22nd are the test set. Table 1 shows the basic statistics of the dataset.

	Users	Items	Interactions	Item features
KuaiRand-1k	1000	4,369,953	11,713,045	62

Table 1: the statistics of KuaiRand-1k

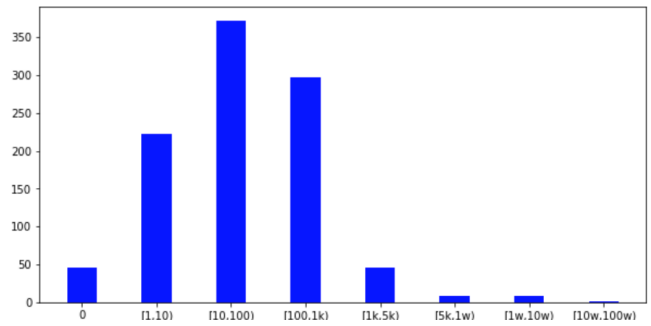

Fig 5: Distribution of fan number of each user

The dataset also includes some features about the users. For example *fans_user_num* is the attribute representing how many fans a user has. Figure 3 shows the distribution of *fans_user_num.* We find that this variable follows a long-tail distribution, in which over 85% of the users have less than 1000 fans. We set 1000 as the threshold for the binary classification and split all the users into two groups, those having less than 1000 fans (normal users) and those having more than 1000 fans (influencers). The skewed distribution indicates that the recommendation utility may favor the normal users and against the influencers primarily due to the larger number of normal users in the dataset. This phenomenon introduces a potential popularity bias against influencers. However, it is crucial to acknowledge that platform influencers play a vital role in the short video platforms as they are highly active in content creation and hold significant influence in shaping the community culture. Therefore, it is important for us to assess whether our model has the popularity bias or not.

4 Results

We first train and perform hyperparameter tuning of our model by F1 score for binary type model and RMSE for linear type model. We set the epoch 75, learning rate 1e-3, weight decay 1e-3, and batch size 200. Figure 6 and Figure 7 shows the training loss and validation loss for the binary model and the linear model. We observe that validation loss is always lower than training loss. One possible reason could be that we are training on all the previous data to predict

the last day's watch behavior so that the model can potentially learn better. The F1 score of the binary type model is *0.5765* and RMSE for linear type model is *2.9390*.

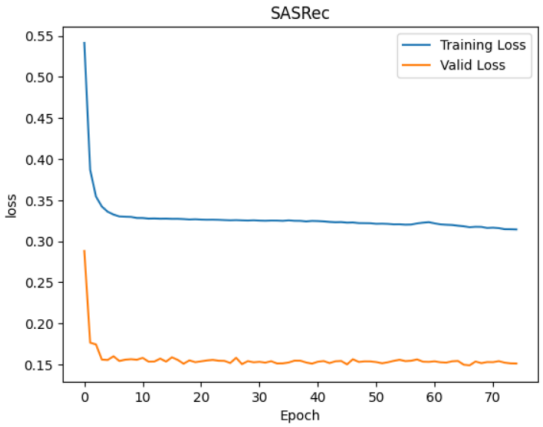

Figure 6: the training/valid loss with binary label

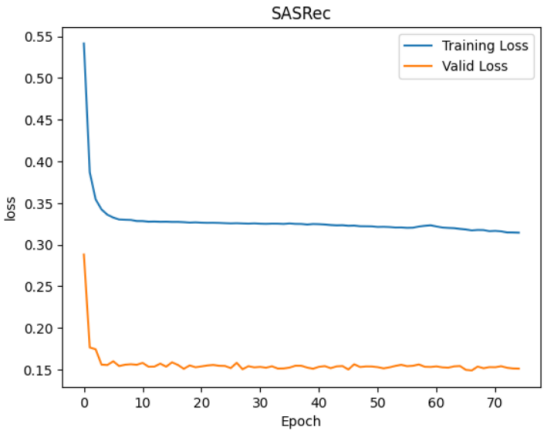

Figure 7: the training/validation loss with linear label

We then evaluate the popularity bias matrix for two user groups: normal users and influencers following our definition in the above section. Table 2 shows the Accuracy and F1 score results. We observe that for both Accuracy and F1, influencers score lower than normal users, meaning that their recommendation utility is worse than others. This indicates that the popularity bias does exist in our model, which could be a potential direction for future improvement.

	Normal users	Influencers
Accuracy	0.9099	0.8323
F1	0.5829	0.5743

Table 2: the evaluation on different user groups

For each interaction log, we also calculated the WTG and set the threshold of 1. If z >1, then the user and the video is relevant and we label relevance as 1. Otherwise, the user and the video is not relevant and we label relevance as 0. The following table shows a summary of the transformation of WTG. However, we don't have enough time to integrate this with the model.

	user_id	video_id	relevance	z_score	duration_ms	duration_bucket	play_time_ms
869	0	591142	0	-0.463669	342661	5	0
870	0	1496664	0	-0.616837	53600	0	0
871	0	1879529	0	-0.616837	23833	0	0
872	0	1823185	0	-0.616837	10533	0	0
873	0	2524252	0	-0.616837	28750	0	0
...	...	...	...	...	...	...	...
4645314	999	2293210	1	1.324074	70000	1	64285
4645315	999	501790	0	-0.499249	11500	0	1658
4645316	999	4300309	1	3.424308	187520	3	231097
4645317	999	2680433	1	1.920917	78083	1	84103
4645318	999	1134659	0	-0.117833	7233	0	7036

Table 3: a summary of the WTG transformation for the validation set

5 Discussions and Future Work

The main contributions of this project are summarized as follows:

- We propose a new model architecture SASRec4SV adapting SASRec for the short video recommendation tasks with two implementations, one with linear prediction and the other with the binary label.
- Secondly, we evaluate the performance of SASRec4SV on the practical dataset KuaiRand for two user groups, namely what we define as normal users and influencers and show that our model does exhibit popularity bias.
- Lastly, we intend to reproduce Watch Time Gain (WTG) [3] to provide a fair evaluation that alleviates video duration bias in short videos with different length. As the better standard to reflect user's preference and incorporate implicit feedback into our model, WTG could be very helpful. Regrettably, due to time constraints, we were unable to apply the WTG fully.

For some future directions, as we show that indeed popularity bias propagates in our model and influencers with fans more than 1000 gets lower recommendation utility, how to tackle this popularity bias becomes a problem. We might future integrate some bias intervention techniques in our model structure. At the same time, we also hope to fully operationalize WTG

as an unbiased metric for our dataset. This can also be used to examine other biases in different user groups, from user history or age groups. These explorations will contribute to a more comprehensive and accurate short video recommendation system.

References

[1]W. -C. Kang and J. McAuley, "Self-Attentive Sequential Recommendation," *2018 IEEE International Conference on Data Mining (ICDM)*, Singapore, 2018, pp. 197-206, doi: 10.1109/ICDM.2018.00035.

[2]Q. Cai "Constrained Reinforcement Learning for Short Video Recommendation." *arXiv preprint arXiv:2205.13248*, 2022.

[3]Y. Zheng "DVR: Micro-Video Recommendation Optimizing Watch-Time-Gain under Duration Bias." *Proceedings of the 30th ACM International Conference on Multimedia*, 2022.

[4]C. Gao. "KuaiRand: An Unbiased Sequential Recommendation Dataset with Randomly Exposed Videos." *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022.

[5]D. Kaye, *TikTok: Creativity and culture in short video*. John Wiley & Sons, 2022.