Towarads fully personalized attention for Sequential Recommendation - An experimental Study

Kangrui Yu
ky2390@nyu.edu

ABSTRACT

Timestamp information is crucial for recommendation problems because users naturally have long and short term preferences over certain items. Recent progress in deep learning, notably the exploration of different attention mechanisms and newer architectures alongside well-established techniques like RNNs and CNNs in natural language processing, has enhanced the utilization of the chronological sequence of user interactions. One standout example is the TiSASRec model[1], drawing inspiration from the widely-used Transformer model in natural language processing, which has attained cutting-edge performance. However, TiSASRec, akin to the original Transformer model, while it tries to use relative time information to capture the importance of the recently interacted items of the user, it yet neither utilize the absolute timestamp information, which may be helpful in modeling user's inherent constant interest , nor does it utilize the personal information of each user. To overcome this limitation, we proposed the SAS-FPT model, outperforming the original TiSASRec by around 2 percent in terms of NDCG@10 and Recall@10.

KEYWORDS

Recommendation Systems, Self Attention, Personalized Embedding, Sequential Recommendation

1 INTRODUCTION

Traditionally, recommender systems have primarily focused on conventional collaborative filtering and ranking techniques. The issue of sequential recommendation is relatively recent but significant in research, with leveraging temporal data to enhance recommendations proving to be challenging. TiSASRec, introduced by [1] for sequential recommendation tasks, has achieved remarkable results and significantly accelerated processing compared to earlier CNN/RNN-based methods. However, TiSASRec employs a standard Transformer model with relative information about timestamp, which inherently lacks personalization. In practical scenarios, incorporating personalized elements into TiSASRec, especially for recommender systems, is crucial. In our model (SAS-FPT), we proposed to utilize personalized embeddings for attention network along with combining relative and absolute time information.

2 RELATED WORK

2.1 SSE-PT

SSE-PT[2] adopts a similar embedding technique with personalized user embedding concated into the attention mechanism. In addition, the model utilizes a novel stochasic shared embeddings to overcome overfitting. In comparison, our model utilizes a different group-embedding-loss method to overcome overfittings. In addition, we futher customize the embedding input of the attention architecture to make it fully personalized for each user and capable of capturing long-term and short-term shifts of the user preference.

3 METHODOLOGY

3.1 Sequential Recommendation

With a pool of n users, each interacting with a selection of m items over time, sequential recommendation aims to develop a tailored hierarchy of the top K items from the entire set for any user, at any moment. Our approach is based on sequences of interactions between users and items. Sequences of duration T include the indexes of the most recent T interactions by each user, ordered chronologically from oldest to newest. While the lengths of these sequences may differ for each user, we can standardize them by padding shorter sequences to length T. Due to the temporal order of the data, a simple random split into train/validation/test sets isn't feasible. Instead, we must ensure that our training data precedes the validation data, which precedes the test data in temporal sequence. We designate the last items in the sequences as the test sets, the second-to-last items as the validation sets, and the remainder as the training sets. For assessment, we utilize ranking metrics like NDCG@K and Recall@K.

3.2 Absolute & Relative Time Embedding

We interpret time intervals within an interaction sequence as the relationship between two items. Certain users engage in interactions more frequently than others, and our focus is solely on the relative duration of time intervals within a single user sequence. Therefore, for all time intervals, we normalize by dividing them by the shortest time interval (excluding 0) within a user sequence to obtain personalized intervals.

In addition, we further embed the absolute timestamp of user-item interaction. Specifically, we propose that the absolute timestamp is embedded according to the week it falls in within a year, i.e. ($t_{u,i} \in 1, 2, 3...52$)

3.3 Personalized User Embedding

We define the user embedding lookup table as $U \in R^{d_u x d_{emb}}$, further more, we utilize a personalized group-wise embedding loss:
$L = \sum_{i \in I^-} max(0, cos(emb_{i_0}, emb_{i_1} - margin)) + \sum_{i \in I^+}(1 - cos(emb_{i_0}, emb_{i_1}))$
where n user embedding positive and negative pair I is randomly sampled from user embedding controlled by hyperparameter n, which is the sample size per batch.

3.4 Fully Personalized Attention Architecture

To incorporate both the absolute and relative time embedding and the personalized user embedding, we define the input of our SAS-FPT model as follows:
$r_{input} = concat(item_{emb}, week_{emb}, user_{emb})$

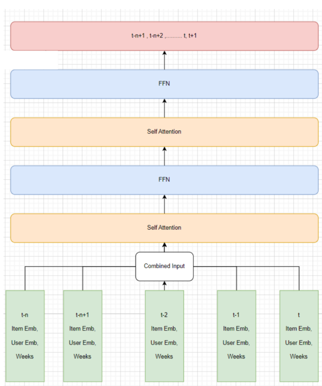

Figure 1: SAS-FPT Architecture

4 EXPERIMENTS

In this section, we compare our proposed algorithms, SAS-FPT with another well performing model leveraging relative time data - TiSASRec. We implement our code in Pytorch and conduct our research with a 4060-Laptop GPU.
Datasets: We use MovieLens 1M dataset, to avoid cold-start problem, we remove users that have less than 5 interactions and items that receive less than 5 interactions. In addition, we also test the cold-start case, where we conserve the users and items with few interactions

Table 1: Hyperparameters

Parameter	Value
Item embedding dimension	50
User embedding dimension	25
User Group Loss Regularization	0.01
Absolute Time Bucket	52

Table 2: Performance

Model	Recall@10	NDCG@10
TiSASRec	0.7881	0.5505
TiSASRec+	0.7894	0.5550
TiSASRec++	0.8002	0.5630
SAS-FPT	**0.8056**	**0.5702**

Table 3: Performance with cold start

Model	Recall@10	NDCG@10
TiSASRec	0.7998	0.5645
SAS-FPT	**0.8030**	**0.5778**

5 CONCLUSION

From the performance we can conclude that our fully personalized attention model outperforms its ancestor TiSASRec by around 2 percent in the Movielens1M Dataset. We also did the ablation study to further investigate how different components of our work contribute to the increase in performance. In TiSASRec+, we refine the relative time interation matrix by adding additional dimensions in the embedding lookup table to decrease the skewness of data, in TiSASRec++, we further implement absolute time embedding into the attention architecture. Finally, in SAS-FPT, which is our proposed full model, we implement the user embedding and groupwise embedding loss and achieve the best performance.

6 FUTURE WORKS

We would like to explore the following fields of study for expanding our proposed model. We will further investigate on how these hyperparameters works and come up with a novel method to determine the relative portion of the embedding dimensions of User, Item and Time to overcome overparameterization, (which is shown in some cases where larger embedding dimensions don't achieve higher performance).

REFERENCES

[1] Jiacheng Li, Yujie Wang, and Julian McAuley. 2020. Time Interval Aware Self-Attention for Sequential Recommendation. In The Thirteenth ACM International Conference on Web Search and Data Mining (WSDM '20), February 3–7, 2020, Houston, TX, USA. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3336191.3371786

[2] Liwei Wu, Shuqing Li, Cho-Jui Hsieh, and James Sharpnack. 2020. SSE-PT: Sequential Recommendation Via Personalized Transformer. In Fourteenth ACM Conference on Recommender Systems (RecSys '20), September 22–26, 2020, Virtual Event, Brazil. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3383313.3412258