TGSRec

Tags	
Column	
Column 1	

▼ Abstract

- sequential parttern & temporal collaborative signals를 unify 하여 recommandation quality를 향상 > 어려움
- 이유

1) sequential patterns 과 collaborative signals을 simultaneously encode 하기 어려움

2) collaborative signals의 temporal efects를 express 하는 것은 사소하지 않음.

- Temporal Graph Sequential Recommender (TGSRec) 제안 : continuous-time bi-partite graph 를 정의
    - novel Temporal Collaborative Transformer (TCT) layer 제안 > users 와 items에서, collaborative signals을 동시에 capture. 또한, sequential patterns안의 temporal dynamics을 고려.
    - TCT 계층에서 학습한 정보를 temporal graph를 통해 propagate하여 sequential patterns과 temporal collaborative signals를 통합.
    - novel collaborative attention을 채택함으로써 self-attention mechanism 향상

▼ 1. introduction

- users에 대한 historical time-ordered item purchasing sequences to predict future items. > sequential recommendation (SR) problem
- self-attention model은 infers sequence embeddings at position t by assigning an attention weight to each historical item and aggregating these items. The attention weights reveal impacts of previous items to the current state at time point t.
- 이유 : crucial temporal collaborative signals을 무시해서.

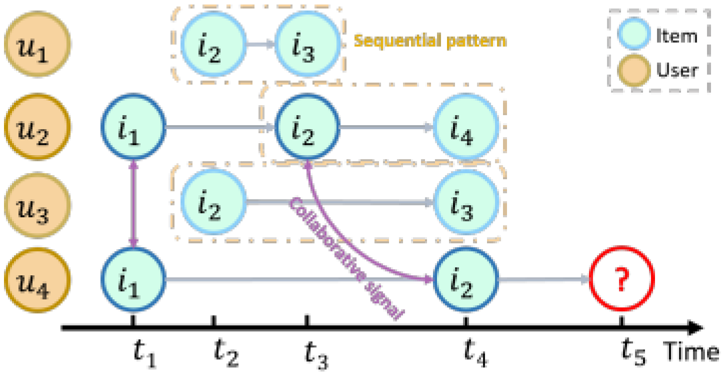


Figure 1: A toy example of temporal collaborative signals. Given the items that users u_1, u_2, u_3 and u_4 like in the past timestamps t_1, t_2, t_3 and t_4, the target is to recommend an item to u_4 at t_5 as the next item after i_2.

1) sequence의 collaborative signal을 model하기 부족. 동시에 encode하기 어려움

- SASRec : Self-attention > users에서만 효과
- SSE-PT : collaborative signals, same embedding(sequential encoding) > user-item interaction fail. & 명시적 collaborative signal을 express 불가

2) time gap이 있는데 이걸 같은 contribution 으로 볼건가

- contribution

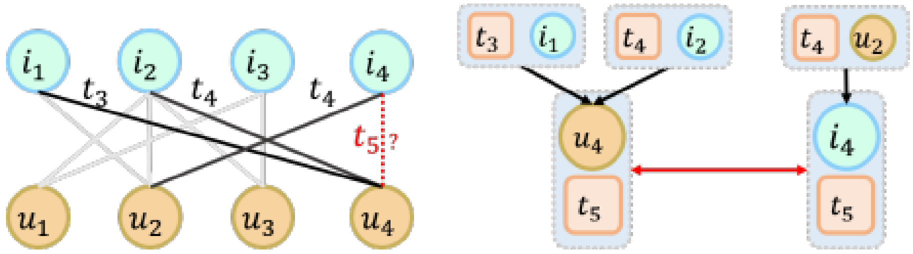
1) graph sequential recommendation : sequential patterns 과 temporal collaborative signals 를 통합하는 graph embedding method를 학습

2) temporal collaborative transformer : collaborative signals와 temporal effect를 공동으로 모델링하는 temporal node embeddings을 추론하는 새로운 temporal collaborative attention mechanism.

3) extensive experiments : five real-world datasets으로 실험-비교

▼ 3. Definitions and Preliminaries

- 일반적인 SR과 다른 user interaction sequence 사용 > Continuous-Time Bipartite Graph (CTBG)



(a) Example of CTBG (b) Temporal Inference

Figure 2: The associated CTBG of Figure 1 and the inference of temporal embeddings of u_4 and i_4 at t_5.

- edge : timestamp as attribute
- every user/item node in this graph preserve the sequential order via the timestamps at edges

- Definition 3.1 (Continuous-Time Bipartite Graph). N nodes 와 E edges 연속적인 시간 bipartite graph 는$\mathcal{B} = \mathcal{U}, \mathcal{I}, \mathcal{E}_\mathcal{T}$,
  - 여기서 $\mathcal{U}, \mathcal{I}$는 two disjoint node sets of users and items.
  - 모든 edge $e \in \mathcal{E}_\mathcal{T}$ 는 **tuple *e* = (*u*,*i*, *t*)** 로 정의. where $u \in \mathcal{U}, i \in \mathcal{I}, and t \in \mathcal{R}^+$
    as the edge attribute.
  - Each triplet (u,i, t) denotes the interaction of a user u with item i at timestamp t.
- Definition 3.2 (Continuous-Time Recommendation). At a specific timestamp t, given user set U, item set I, and the associated CTBG, the **continuous-time recommendation** of u is to generate a ranking list of items from $\mathcal{I} \setminus \mathcal{I}_u(t)$, where the items that u is interested will be ranked top in the list.
- Definition 3.3 (Continuous-Time Seqential Recommendation). For a specific user u, given a set of future timestamps$\mathcal{T}_u > T$, the continuous-time sequential recommendation for this user is to make a continuous-time recommendation for every timestamp $t \in \mathcal{T}_u$

▼ 4. proposed model

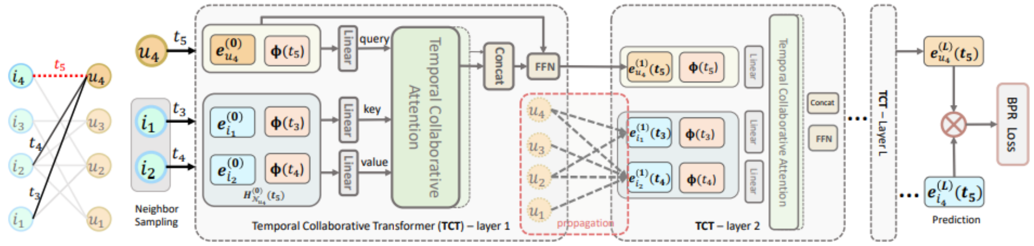


Figure 3: The framework of TGSRec. The query node is u_4, whose final temporal embedding at time t_5 is $h_{u_4}^{(2)}(t_5)$. The TCT layer samples its neighbor nodes and edges. Timestamps on edges are encoded as vectors by using mapping function Φ. Node embeddings for the first TCT layer are long-term embeddings. Node embeddings for other TCT layers (e.g. layer 2) are propagated from the previous TCT layer, thus being temporal node embeddings.

4.1 embedding layer

- **two types of embeddings** 를 encode
  - 1) one being the long-term embeddings of nodes
  - 2) the other being the continuous- time embeddings of timestamps on edges.

4.1.1 Long-Term User/Item Embeddings

- long-term collaborative signals representation에 대한 Long-term embeddings 은 중요.
- user(item) node 는 vector $e_u(e_i) \in \mathbb{R}^d$로 parameterized.
- embeddings은temporal user/item embeddings inference을 위한 시작 상태 역할을 한다는 점에 유의.
- During the training process, embeddings will be optimized.

4.1.2 Continuous-Time Embedding

- The time encoding function embeds **timestamps(scalar)** into vectors so as to
  represent the time span as the dot product of corresponding encoded time embeddings.
- Define the temporal effects as a function of time span in continuous time space

$$\psi(t_1 - t_2) = \mathcal{K}(t_1, t_2) = \Phi(t_1) \cdot \Phi(t_2), \qquad (1)$$

- temporal correlation between two timestamps

$$\Phi(t) \mapsto \sqrt{\frac{1}{d_T}}\left[\cos(\omega_1 t), \sin(\omega_1 t), \dots, \cos(\omega_{d_T} t), \sin(\omega_{d_T} t)\right]^\top, \tag{2}$$

where $\boldsymbol{\omega} = \left[\omega_1, \dots, \omega_{d_T}\right]^\top$ are learnable and d_T is the dimension.

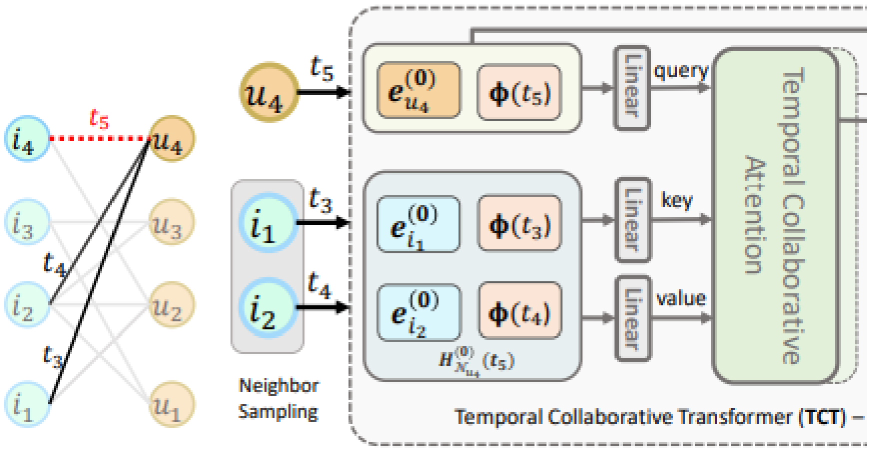
positional encoding

4.2 Temporal Collaborative Transformer

- 2가지 장점
    - 1) correlations의 temporal effects을 explicitly하게 characterizes하는 user/item embeddings 및 temporal embedding 모두에서 information 구성.
    - 2) TCT layer는 user-item interactions간 collaborative attention을 채택
        - the **query input** to the **collaborative attention** is from the **target node (user/item),**
          while the **key and value inputs** are from **connected neighbors**

4.2.1 Information Construction.

- long term node embedding + time embedding > TCT layer의 input으로 쓰기 위해 구성



- The query input information at the l-th layer for user u at time t (eq3) temporal embedding of u

$$\boldsymbol{h}_u^{(l-1)}(t) = \boldsymbol{e}_u^{(l-1)}(t) \| \Phi(t),$$

the information for u at t time vector of t

- $h_u^{(l-1)}(t) \in \mathbb{R}^{d+d_T}$: user u 의 time t에서의 information
- $e_u^{(l-1)}(t) \in \mathbb{R}^d$: temporal embedding of u
- $\Phi(t) \in \mathbb{R}^{d_T}$: time vector of t
- $||$: concat operation

- Note that when *l* = 1, it is the first TCT layer. > The temporal embedding $e_u^{(0)}(t) = E_u$
- When *l* > 1, the temporal embedding is generated from the previous TCT layer.

- In addition to the query node **itself**, also propagate temporal collaborative information from its **neighbors**.
- Randomly sample S different interactions of u before time t

$$\mathcal{N}_u(t) = \{(i, t_s) | (u, i, t_s) \in \mathcal{E}_t \text{ and } t_s < t\}$$

t_s < t sampling (eq4) 각 (i, t_s) pair l-th layer

$$\boldsymbol{h}_i^{(l-1)}(t_s) = \boldsymbol{e}_i^{(l-1)}(t_s) \| \Phi(t_s), \tag{4}$$

4.2.2 Information Propagation.

- Propagate the information of **sampled neighbors**$N_u(t)$ to infer the temporal
  embeddings (Eq5) historical interaction $\pi_t^u(i, t_s)$: (u, i, t_s) 의 영향 반영

$$\mathbf{e}_{\mathcal{N}_u}^{(l)}(t) = \sum_{(i,t_s)\in\mathcal{N}_u(t)} \overbrace{\pi_t^u(i,t_s)}^{\text{the importance of an interaction }(u,i,t_s)} \underbrace{\mathbf{W}_v^{(l)}}_{\text{the linear transformation matrix}} \boldsymbol{h}_i^{(l-1)}(t_s),$$

- $W_v \in \mathbb{R}^{(d\times(d+d_T))}$: linear transformation matrix

4.2.3 Temporal Collaborative Attention

- Adopt the novel temporal collaborative attention mechanism to measure the **attention weight** $\pi_t^u(i, t_s)$,

$$\pi_t^u(i,t_s) = \frac{1}{\sqrt{d+d_T}}\left(\mathbf{W}_k^{(l)}\boldsymbol{h}_i^{(l-1)}(t_s)\right)^\top \mathbf{W}_q^{(l)}\boldsymbol{h}_u^{(l-1)}(t), \qquad (6)$$

$W_k^{(l)}$와 $W_q^{(l)}$: linear transformation matrices, factor $\frac{1}{\sqrt{d+d_T}}$: dot-product from growing large with high dimensions

eq3, eq4를 eq6의 오른쪽 side는 재표현 가능

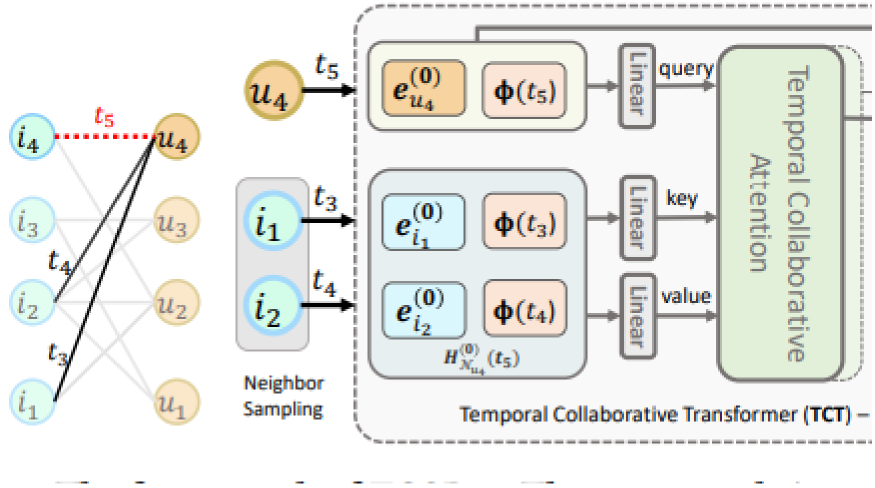
$$\underbrace{\boldsymbol{e}_u^{(l-1)}(t)\cdot\boldsymbol{e}_i^{(l-1)}(t_s)}_{\text{the user-item collaborative signal}} + \underbrace{\Phi(t)\cdot\Phi(t_s)}_{\text{temporal effect}},$$

temporal effect : eq1

- more stacked layers, collaborative signals and temporal effects are entangled and tightly connected됨
- dot-product attention can characterize impacts of temporal collaborative signals
- **normalize** the attention weights across **all sampled** interactions by employing a softmax function

$$\pi_t^u(i, t_s) = \frac{\exp\left(\pi_t^u(i, t_s)\right)}{\sum_{(i',t_s')\in N_u(t)} \exp\left(\pi_t^u(i', t_s')\right)}. \tag{8}$$

- $H_{\mathcal{N}_u}^{(l-1)}(t) \in \mathbb{R}^{(d+d_T)\times S}$ > fig3 (추정하고자 하는 user와 time과 연관된 sample로 만듬)



- $K_u^{(l-1)}(t) = W_k^{(l)} H_{\mathcal{N}_u}^{(l-1)}(t)$, $V_u^{(l-1)}(t) = W_v^{(l)} H_{\mathcal{N}_u}^{(l-1)}(t)$, $q_u^{(l-1)}(t) = W_q^{(l)} H_{\mathcal{N}_u}^{(l-1)}(t)$ > key, query, value
- For **simplicity** and without **ambiguity**, **superscripts** and **time *t*** 무시하고 combine Eq. (6) and Eq. (8).

$$e_{\mathcal{N}_u} = V_u \cdot \text{Softmax}\left(\frac{K_u^\top q_u}{\sqrt{d + d_T}}\right), \tag{9}$$

- dot-product attention in Transformer

1) multi-head attention operation

2) concatenate the output from each head as the information for aggregation

- self-attention X > temporal collaborative attention, user-item interactions 과 temporal information 를 결합해 modeling

4.2.4 Information Aggregation

- temporal node embedding의 결과물로 , TCT layer의 final step으로 query information을 (eq3), neighbor information in Eq. (5) aggregate

$$\boldsymbol{e}_u^{(l)}(t) = \mathrm{FFN}\left(\boldsymbol{e}_{\mathcal{N}_u}^{(l)}(t) \| \boldsymbol{h}_u^{(l-1)}(t)\right), \tag{10}$$

- concat > FFN two linear transformation layers with ReLU

4.2.5 Generalization to items.

- user query perspective에서만 TCT 레이어를 제시하지만 query 가 특정 시간의 항목인 경우에도 유사.
- user query information과 item query information를 번갈아 사용하고, user time pair으로 Eq(4)와 Eq(5)의 neighbor information를 변경하면 됨.
- 다음 layer로 보내지는 $e_i^{(l)}(t)$로서 시간 t, item i의 temporal embedding에 대한 추론 가능

4.3 Model Prediction

- TGSRec model은 L TCT layers 로 구성. 각 test triplet (u,i,t) 에서, last TCT layer 의 시점 t에서 u와 i 의 temporal embedding > $e_u^{(L)}(t)$, $e_i^{(L)}(t)$
- prediction score $r(u,i,t) = e_u^{(L)}(t) \cdot e_i^{(L)}(t)$ (eq11)
    - $r(u,i,t)$: score to recommend i for u at time t
    - generalized continuous-time embeddings과 제안된 TCT 계층을 통해, 어떤 타임스탬프에서도 사용자/항목 임베딩을 일반화 및 추론할 수 있으므로 기존 작업이 다음 항목만 예측하는 동안 여러 단계 권장 사항을 실현할 수 있음.
    - Definition 3.3을 기반으로 각 사용자에게 주어진 타임스탬프의 항목 순위 목록을 권장함. 따라서 Eq. (11)를 사용하여 모든 후보 항목의 점수를 계산하고 점수별로 정렬가능.

4.4 Model Optimization

- pairwise BPR loss > top-N recommendation

$$\mathcal{L}_{bpr} = \sum_{(u,i,j,t) \in O_T} -\log\sigma\left(r(u,i,t) - r(u,j,t)\right) + \lambda||\Theta||_2^2, \quad (12)$$

- mini-batch Adam optimization > Binary Cross Entropy (BCE) loss

$$\mathcal{L}_{bce} = \sum_{(u,i,j,t) \in O_T} \log\sigma\left(r(u,i,t)\right) + \log\sigma\left(1 - r(u,j,t)\right) + \lambda||\Theta||_2^2, \quad (13)$$

5. Experiments

5.1 datasets

Table 1: Statistics of datasets.

Dataset	Toys	Baby	Tools	Music	ML100K
#Users	17,946	17,739	15,920	4,652	943
#Items	11,639	6,876	10,043	3,051	1,682
#Edges	154,793	146,775	127,784	54,932	48,569
#Train	134,632	128,833	107,684	51,765	80,003
#Valid	11,283	10,191	10,847	2,183	1,516
#Test	8,878	7,751	9,253	984	1,344
Density	0.07%	0.12%	0.08%	0.38%	6.30%
Avg. Int.	85 days	61 days	123 days	104 days	4.8 hours

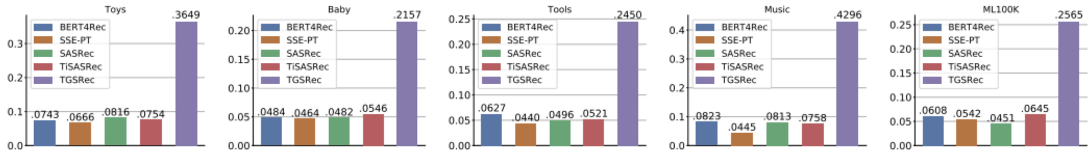
"Av. Int." denotes average time interval.

5.2 Experimental Settings

5.3 Performance Comparison (RQ1)

Table 2: Overall Performance w.r.t. Recall@{10,20} and MRR.

Datasets	Metric	BPR	LightGCN	SR-GNN	GRU4Rec	Caser	SSE-PT	BERT4Rec	SASRec	TiSASRec	CDTNE	TGSRec	Improv.
Toys	Recall@10	0.0021	0.0016	0.0020	0.0274	0.0138	0.1213	0.1273	<u>0.1452</u>	0.1361	0.0016	**0.3650**	0.2198
	Recall@20	0.0036	0.0026	0.0033	0.0288	0.0238	0.1719	0.1865	<u>0.2044</u>	0.1931	0.0045	**0.3714**	0.1670
	MRR	0.0024	0.0018	0.0018	0.0277	0.0082	0.0595	0.0643	<u>0.0732</u>	0.0671	0.0025	**0.3661**	0.2929
Baby	Recall@10	0.0028	0.0036	0.0030	0.0036	0.0077	0.0911	0.0884	0.0975	<u>0.1040</u>	0.0218	**0.2235**	0.1195
	Recall@20	0.0039	0.0045	0.0062	0.0048	0.0193	0.1418	0.1634	0.1610	<u>0.1662</u>	0.0292	**0.2295**	0.0663
	MRR	0.0019	0.0024	0.0024	0.0028	0.0071	0.0434	0.0511	0.0455	<u>0.0521</u>	0.0157	**0.2147**	0.1626
Tools	Recall@10	0.0023	0.0021	0.0051	0.0048	0.0077	0.0775	<u>0.1296</u>	0.0913	0.0946	0.0186	**0.2457**	0.1161
	Recall@20	0.0036	0.0035	0.0092	0.0059	0.0161	0.1155	<u>0.1784</u>	0.1337	0.1356	0.0271	**0.2559**	0.0775
	MRR	0.0026	0.0023	0.0028	0.0051	0.0068	0.0419	<u>0.0628</u>	0.0460	0.0480	0.0203	**0.2468**	0.1840
Music	Recall@10	0.0122	0.0142	0.0051	0.0549	0.0183	0.0915	0.1352	<u>0.1372</u>	<u>0.1372</u>	0.0071	**0.5935**	0.4563
	Recall@20	0.0152	0.0183	0.0092	0.0589	0.0346	0.1494	0.2093	<u>0.2094</u>	0.1951	0.0163	**0.5986**	0.3892
	MRR	0.0057	0.0064	0.0028	0.0540	0.0106	0.0423	<u>0.0824</u>	0.0768	0.0681	0.0037	**0.3820**	0.2996
ML100k	Recall@10	0.0461	0.0565	0.0045	0.0996	0.0246	0.1079	0.1116	0.09450	<u>0.1332</u>	0.0350	**0.3118**	0.1786
	Recall@20	0.0766	0.0960	0.0060	0.1168	0.0417	0.1801	0.1786	0.1808	<u>0.2232</u>	0.0536	**0.3252**	0.1020
	MRR	0.0213	0.0252	0.0012	<u>0.0938</u>	0.0147	0.0519	0.0600	0.0448	0.0605	0.0162	**0.2416**	0.1478



(a) NDCG@10 in Toys (b) NDCG@10 in Baby (c) NDCG@10 in Tools (d) NDCG@10 in Music (e) NDCG@10 in ML100K

Figure 4: NDCG@10 Performance in all Datasets. We ignore other methods because of their low values.

5.4 Parameter Sensitivity (RQ2)

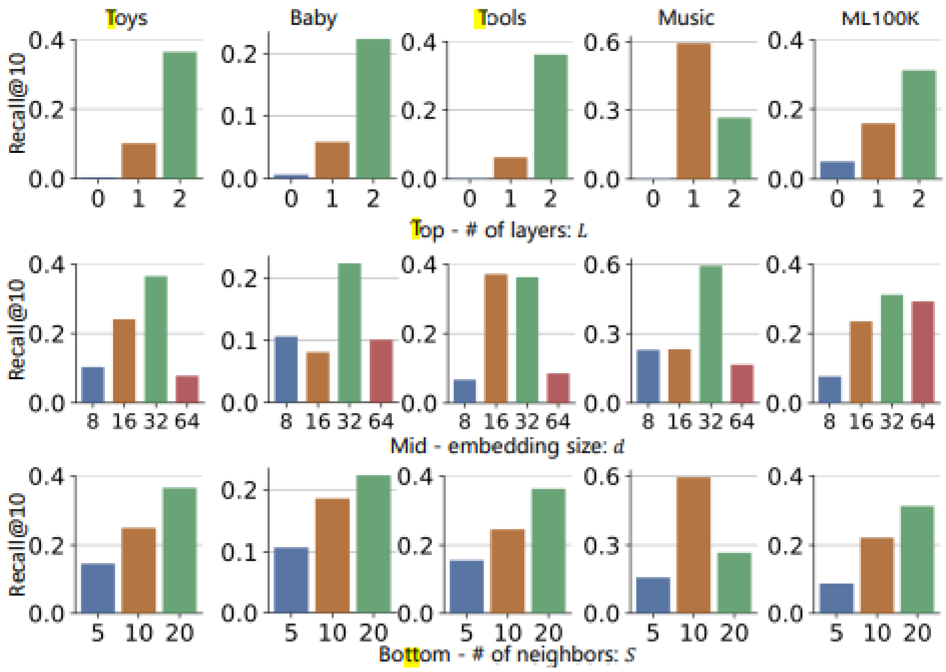


Figure 5: Recall@10 w.r.t. L, d and S on 5 datasets.

5.5 Ablation Study (RQ3 & RQ4)

Table 3: Ablation analysis (Recall@10) on five datasets. Bold score indicates performance better than the default version, while $\downarrow$ indicates a performance drop more than 50%.

Architecture	Toys	Baby	Tools	Music	ML100K
(0) Default	**0.3649**	**0.2235**	**0.3623**	**0.5935**	0.3118
(1) Mean	0.0027$\downarrow$	0.0210$\downarrow$	0.0055$\downarrow$	0.0051$\downarrow$	0.0647$\downarrow$
(2) LSTM	0.0991$\downarrow$	0.1237	0.1266$\downarrow$	0.3740	0.3088
(3) Fixed ω	0.0854$\downarrow$	0.0944$\downarrow$	0.0910$\downarrow$	0.3679	0.2789
(4) Position	0.0380$\downarrow$	0.0243$\downarrow$	0.0209$\downarrow$	0.0742$\downarrow$	0.0878$\downarrow$
(5) Empty	0.0139$\downarrow$	0.0240$\downarrow$	0.0018$\downarrow$	0.0346$\downarrow$	0.0603$\downarrow$
(6) BCELoss	0.2200	0.1916	0.1763$\downarrow$	0.4624	**0.3542**

5.6 Temporal Correlations (RQ4)

Table 4: Variants of Temporal Information Construction

Variant	Toys	Baby	Tools	Music	ML100K
TGSRec	**0.3649**	**0.2235**	**0.3623**	**0.5935**	**0.3118**
$\mathcal{U}$ w/o T	0.0103	0.0138	0.0106	0.0112	0.1555
$\mathcal{I}$ w/o T	$\underline{0.1013}$	$\underline{0.0961}$	$\underline{0.0836}$	$\underline{0.2785}$	$\underline{0.2336}$

Table 5: Recommendation w.r.t. time increments after the last interaction at timestamp $T = 1159142400$. 'next' is the timestamp of the test interaction. The ground truth item is in red color. Items also predicted by SASRec and TiSASRec are in blue color.

Time	Rank-1	Rank-2	Rank-3	Rank-4
T+5d	Letoya	H. of Blue L.	Ult. Prince	Veneer
T+30d	J. of A Gemini	Living Lgds.	Killing Joke	Crane Wife
next	Buf. S.F.	Killing Joke	Empire	Stadium Arc.
T+60d	D. of Future P.	Even Now	L. Mks. Wd.	Przts. Author
SAS.	Crane Wife	Empire	H. Fna. Are	You in Rev.
TiSAS.	Crane Wife	Empire	WTE. P. S.	Stadium Arc.