Efficiently Leveraging Multi-level User Intent for Session-based Recommendation via Atten-Mixer Network

Peiyan Zhang*
The Hong Kong University of Science and Technology
Hong Kong
pzhangao@cse.ust.hk

Jiayan Guo*
School of Intelligence Science and Technology, Peking University
Beijing, China
guojiayan@pku.edu.cn

Chaozhuo Li*
Microsoft Research Asia
Beijing, China
cli@microsoft.com

Yueqi Xie
The Hong Kong University of Science and Technology
Hong Kong
yxieay@cse.ust.hk

Jae Boum Kim
The Hong Kong University of Science and Technology
Hong Kong
Upstage
Republic of Korea
jbkim@cse.ust.hk

Yan Zhang
School of Intelligence Science and Technology, Peking University
Beijing, China
zhyzhy001@pku.edu.cn

Xing Xie
Microsoft Research Asia
Beijing, China
xing.xie@microsoft.com

Haohan Wang
University of Illinois at Urbana-Champaign
IL, USA
haohanw@illinois.edu

Sunghun Kim
The Hong Kong University of Science and Technology
Hong Kong
hunkim@cse.ust.hk

WSDM 2023

Reported by liang li

Motivation

Details:

- However, GNNs achieve relatively marginal improvements with exponential growth in model complexity.
- We dissect the classical GNN-based SBR models and empirically find that some sophisticated GNN propagations are redundant, given the readout module plays a significant role in GNN-based models.

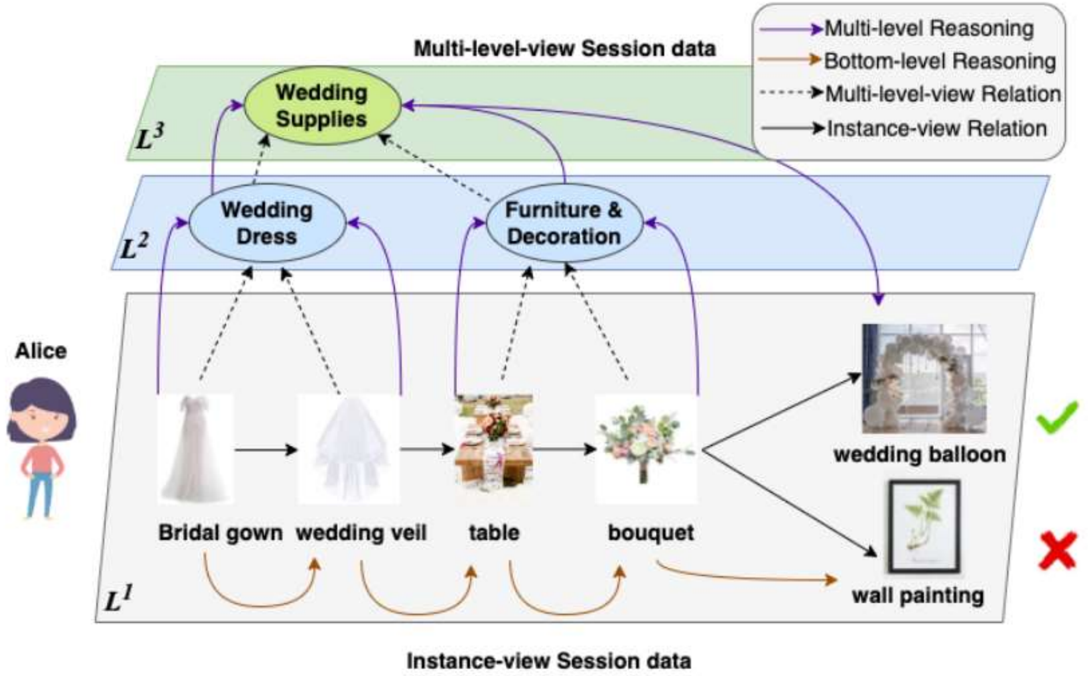


Figure 1: An example of multi-level reasoning over session data.

Motivation

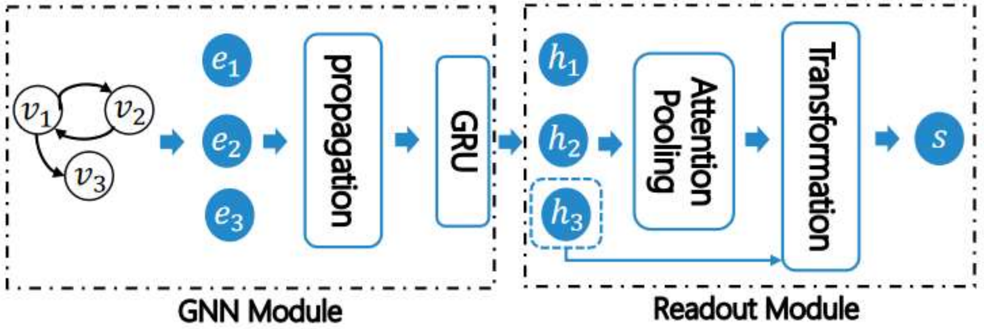


Figure 2: Decomposition of typical GNN-based SBR models

$$w_{mj} = \mathcal{N}(\gamma_{mj}, \delta_{mj})$$

$$\gamma_{mj} = \sum_{i=1}^{I} a_{mi}\theta_{ij}, \quad \delta_{mj} = \sum_{i=1}^{I} a_{mi}^2 \sigma_{ij}^2 \tag{1}$$

$$\mathcal{L}_{reg} = -D_{KL}\left(q(w_{ij}|\theta_{ij}, \alpha_{ij})||p(w_{ij})\right) \tag{2}$$

$$\rho_{\text{density}} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} I(w_{ij} > \alpha) \tag{3}$$

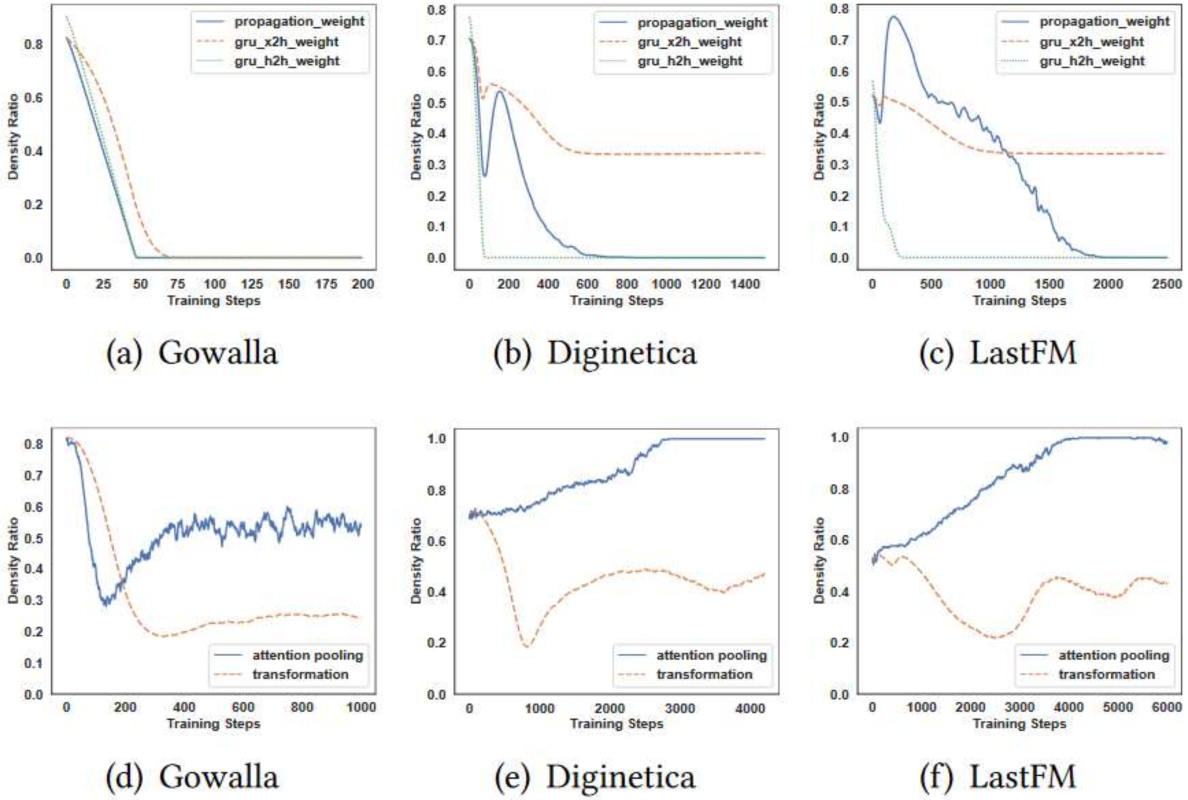


Figure 3: Sparsification result by applying SparseVD on GNN Module.

Problem Statement

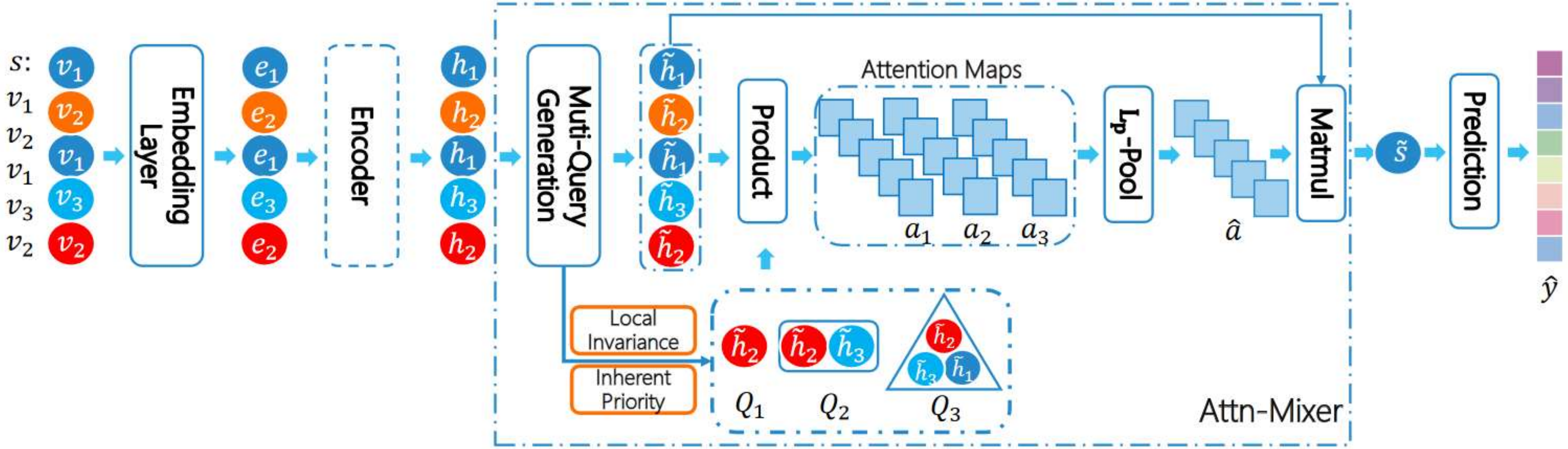


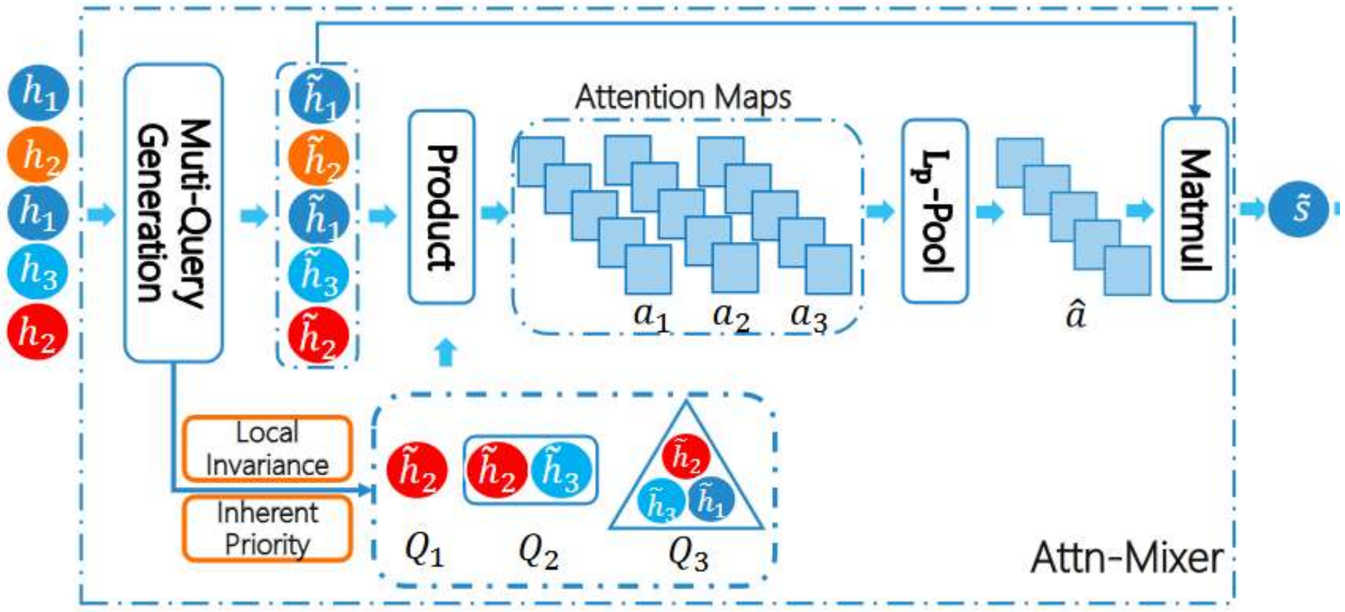
Figure 4: Overview of the Atten-Mixer. Given a session, we first gather normalized item embeddings from the embedding layer and apply Multi-head Level-L Attention (L = 3) by using deep sets operation on the last l normalized hidden states. Then the L_p pooling is used on the generated attention maps to get session representation and make recommendation. Our framework supports employing various item encoders (doted square) before Attn-Mixer.

$$\tilde{h}_j = \frac{h_j}{\|h_j\|_2}$$

$$\hat{z}_j = (\mathrm{W}_m(\tilde{s}||\tilde{h}_n))^T \tilde{h}_j, \tag{4}$$

$$\hat{y}_j = \mathrm{softmax}(\sigma \hat{z}_j), \tag{5}$$

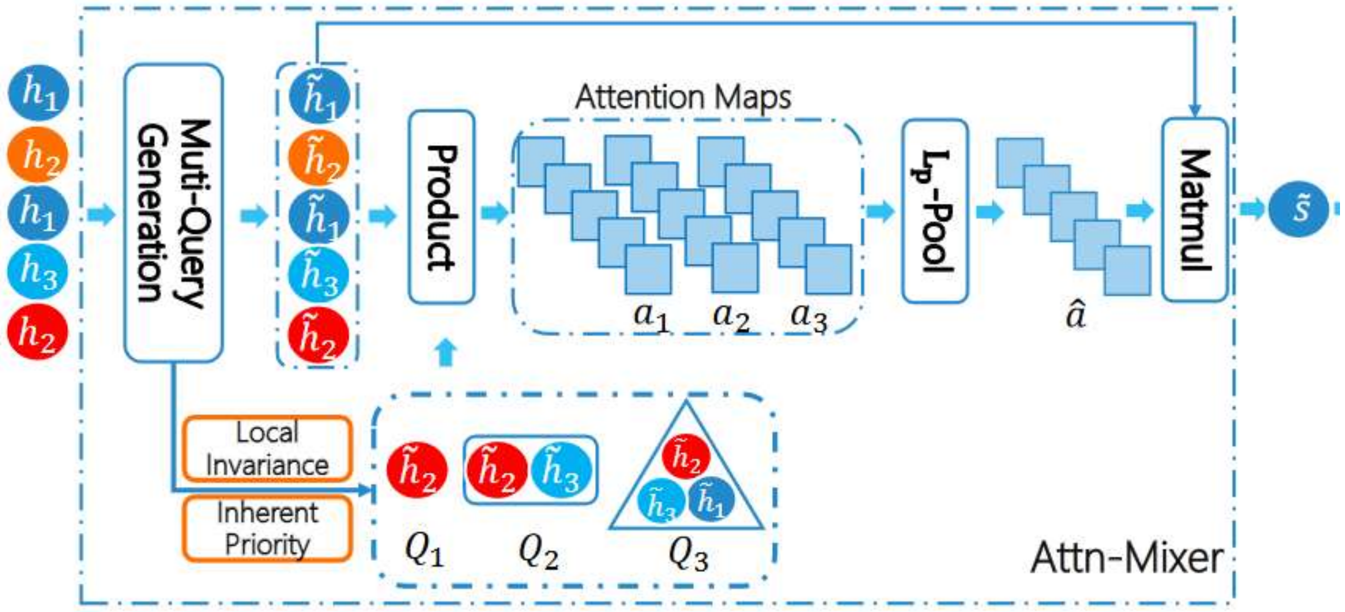
Method



$$
\begin{aligned}
Q_1 &= W_{q1}(\tilde{h}_n),\\
Q_2 &= W_{q2}(\Sigma(\{\tilde{h}_j\}_{j=n,n-1})),\\
&\ldots,\\
Q_L &= W_{qL}(\Sigma(\{\tilde{h}_j\}_{j=n,\ldots,n-L+1})),
\end{aligned} \tag{6}
$$

$$
\alpha_h = \mathrm{softmax}\left(\frac{QW_h^Q(KW_h^K)^T}{\sqrt{d}}\right), \tag{7}
$$

Method



$$\begin{aligned}
\hat{\alpha}_{j,h} &= [\Sigma_{m=0}^{l-1}(\alpha_{j,mH+h})^p]^{\frac{1}{p}}, \\
s^h &= \Sigma_{j=1}^{n}\hat{\alpha}_{j,h}\tilde{h}_j, \\
s &= \oplus\{s^h\}_{h=1,\ldots,H} \\
\tilde{s} &= \frac{s}{||s||_2},
\end{aligned} \tag{8}$$

Experiments

Table 1: Statistics of datasets used in the experiments.

Statistic	*Diginetica*	*Gowalla*	*Last.fm*
No. of Clicks	981,620	1,122,788	3,835,706
No. of Sessions	777,029	830,893	3,510,163
No. of Items	42,596	29,510	38,615
Average Length	4.80	3.85	11.78

Experiments

Table 2: Results of main experiments. The results of the best performing baseline are underlined. The numbers in bold indicate statistically significant improvement (p < .01) by the pairwise t-test comparisons over the other baselines.

Model	*Diginetica*			*Gowalla*			*Last.fm*		
	HR@20	MRR@20	Time (s)	HR@20	MRR@20	Time (s)	HR@20	MRR@20	Time (s)
NextItNet	35.60	9.66	91.06	38.69	16.48	67.94	21.02	6.46	413.27
NARM	48.27	16.43	107.61	49.67	22.14	80.52	21.73	6.87	427.14
SR-GNN	51.16	17.67	341.68	50.16	24.58	338.62	22.49	8.30	1626.94
GC-SAN	50.63	17.37	437.27	50.35	24.65	398.04	22.63	8.40	1814.78
SGNN-HN	51.57	17.54	365.38	50.72	24.97	326.91	23.66	8.34	1595.59
LESSR	51.71	18.15	440.84	51.34	25.49	511.68	23.37	<u>8.84</u>	1927.20
NISER+	<u>54.18</u>	18.36	292.15	53.89	25.73	278.65	23.82	8.36	279.80
DHCN	53.85	18.50	2169.87	53.77	24.13	2452.76	22.86	7.78	21059.94
DSAN	54.02	<u>18.62</u>	273.48	<u>54.09</u>	<u>26.64</u>	279.17	<u>24.17</u>	8.42	1203.81
Atten-Mixer	**55.66**	**18.96**	288.12	**55.12**	**27.01**	267.37	**24.50**	**9.05**	1140.09

Experiments

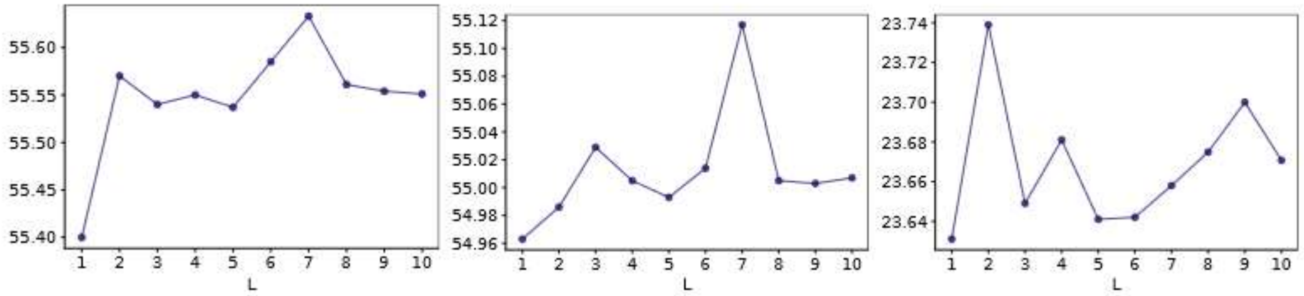
Table 3: Performance comparison of different session-based recommendation methods with their Atten-Mixer incorporated version. All the improvements are statistically significant at the level of p < .01.

Dataset	Metric	SRGNN w/o	SRGNN w	SGNN-HN w/o	SGNN-HN w	Improv. SRGNN	Improv. SGNN-HN
Diginetica	HR@5	26.90	**29.67**	24.88	**27.68**	10.3%	11.25%
	HR@10	38.24	**41.97**	36.48	**39.45**	9.75%	8.14%
	HR@20	51.16	**55.73**	51.57	**52.76**	8.93%	2.31%
	MRR@5	15.28	**16.58**	13.64	**15.48**	8.51%	13.49%
	MRR@10	16.78	**18.20**	15.17	**17.04**	8.46%	12.33%
	MRR@20	17.67	**19.16**	17.54	**17.97**	8.43%	2.45%
Gowalla	HR@5	34.27	**37.62**	33.10	**36.37**	9.78%	9.88%
	HR@10	42.18	**46.07**	42.98	**44.92**	9.22%	4.51%
	HR@20	50.16	**54.46**	50.72	**53.28**	8.57%	5.05%
	MRR@5	22.97	**24.46**	21.45	**23.27**	6.49%	8.45%
	MRR@10	24.03	**25.59**	23.06	**24.42**	6.49%	5.90%
	MRR@20	24.58	**26.17**	24.97	**25.00**	6.47%	0.12%
Last.fm	HR@5	12.08	**12.53**	12.04	**12.64**	3.73%	4.98%
	HR@10	16.57	**17.17**	16.72	**17.51**	3.62%	4.72%
	HR@20	22.49	**23.59**	23.66	**23.89**	4.89%	0.97%
	MRR@5	7.81	**8.22**	7.61	**8.06**	5.25%	5.91%
	MRR@10	8.40	**8.51**	8.23	**8.72**	1.31%	5.95%
	MRR@20	8.80	**8.96**	8.84	**9.15**	1.82%	3.51%

Experiments

Table 4: Ablation studies on different components.

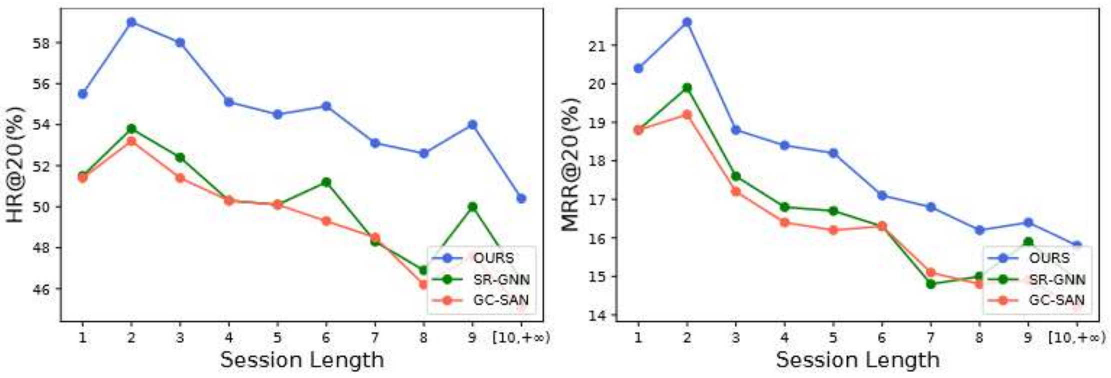
Model	*Diginetica* HR@20	*Diginetica* MRR@20	*Gowalla* HR@20	*Gowalla* MRR@20	*Last.fm* HR@20	*Last.fm* MRR@20
Atten-Mixer-M	52.27	17.52	50.20	25.06	22.14	8.30
Atten-Mixer-IP	53.76	17.99	52.62	25.03	22.56	8.83
Atten-Mixer-LI	53.81	17.78	52.27	24.91	22.28	8.81
Atten-Mixer-LP	53.48	17.89	53.29	25.54	23.44	8.90
Atten-Mixer	**55.66**	**18.96**	**55.12**	**27.01**	**24.50**	**9.05**



(a) *Diginetica* (b) *Gowalla* (c) *Last.fm*

Figure 5: HR@20 w.r.t. the value of L.

Experiments



(a) HR@20 on *Diginetica* (b) MRR@20 on *Diginetica*

Figure 7: Performance w.r.t. different session length.

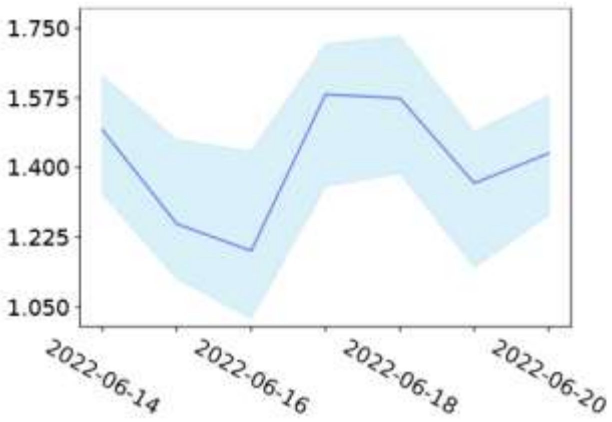


Figure 8: Top business metric improvement percentage (y-axis) over days (x-axis) in online experiments.

Thanks