A Self-Correcting Sequential Recommender

Yujie Lin
Shandong University
Qingdao, China
yu.jie.lin@outlook.com

Chenyang Wang
Shandong University
Qingdao, China
201900122032@mail.sdu.edu.cn

Zhumin Chen
Shandong University
Qingdao, China
chenzhumin@sdu.edu.cn

Zhaochun Ren
Shandong University
Qingdao, China
zhaochun.ren@sdu.edu.cn

Xin Xin
Shandong University
Qingdao, China
xinxin@sdu.edu.cn

Qiang Yan
WeChat, Tencent
Guangzhou, China
rolanyan@tencent.com

Maarten de Rijke
University of Amsterdam
Amsterdam, The Netherlands
m.derijke@uva.nl

Xiuzhen Cheng
Shandong University
Qingdao, China
xzcheng@sdu.edu.cn

Pengjie Ren*
Shandong University
Qingdao, China
renpengjie@sdu.edu.cn

code: https://github.com/TempSDU/STEAM.

WWW 2023

Reported by Minqin Li

Introduction

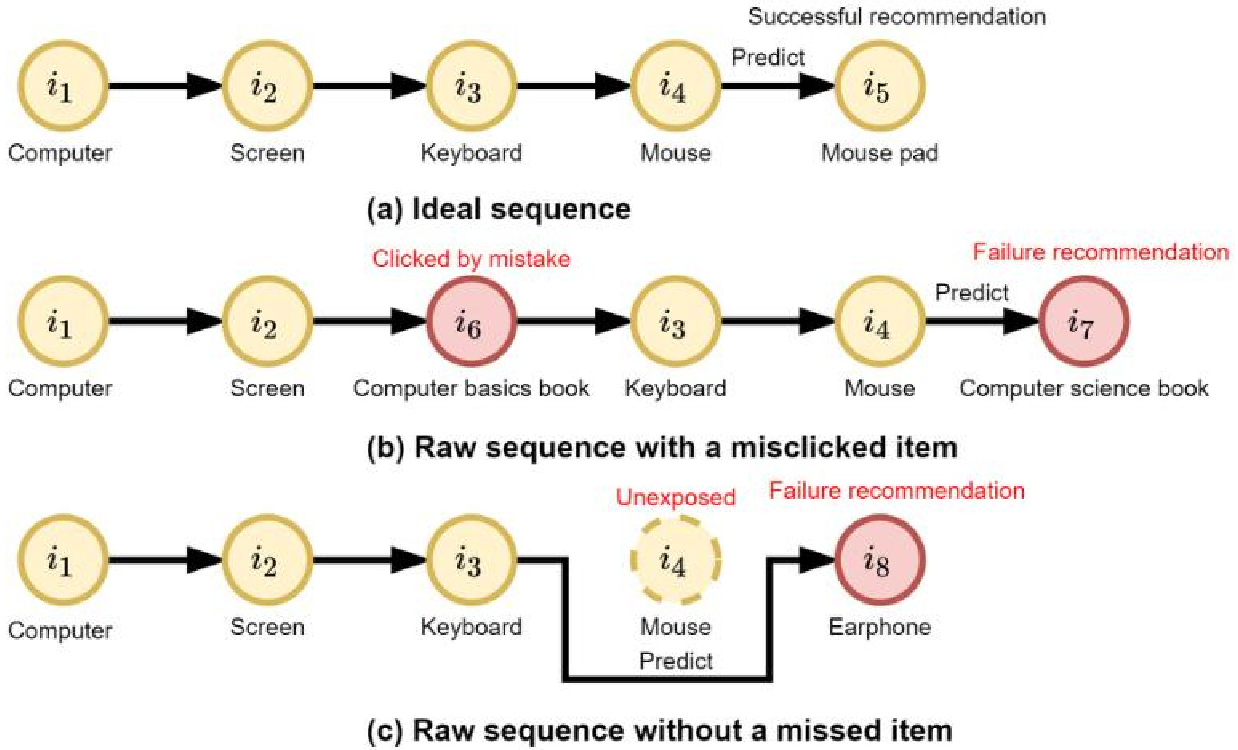


Figure 1: Examples for two kinds of imperfect item sequences. Sub-figure (a) is an ideal item sequence without any imperfection. Sub-figure (b) is an imperfect item sequence that contains a misclicked item (i.e., i_6). Sub-figure (c) is an imperfect sequence that lacks a missed item (i.e., i_4).

Method

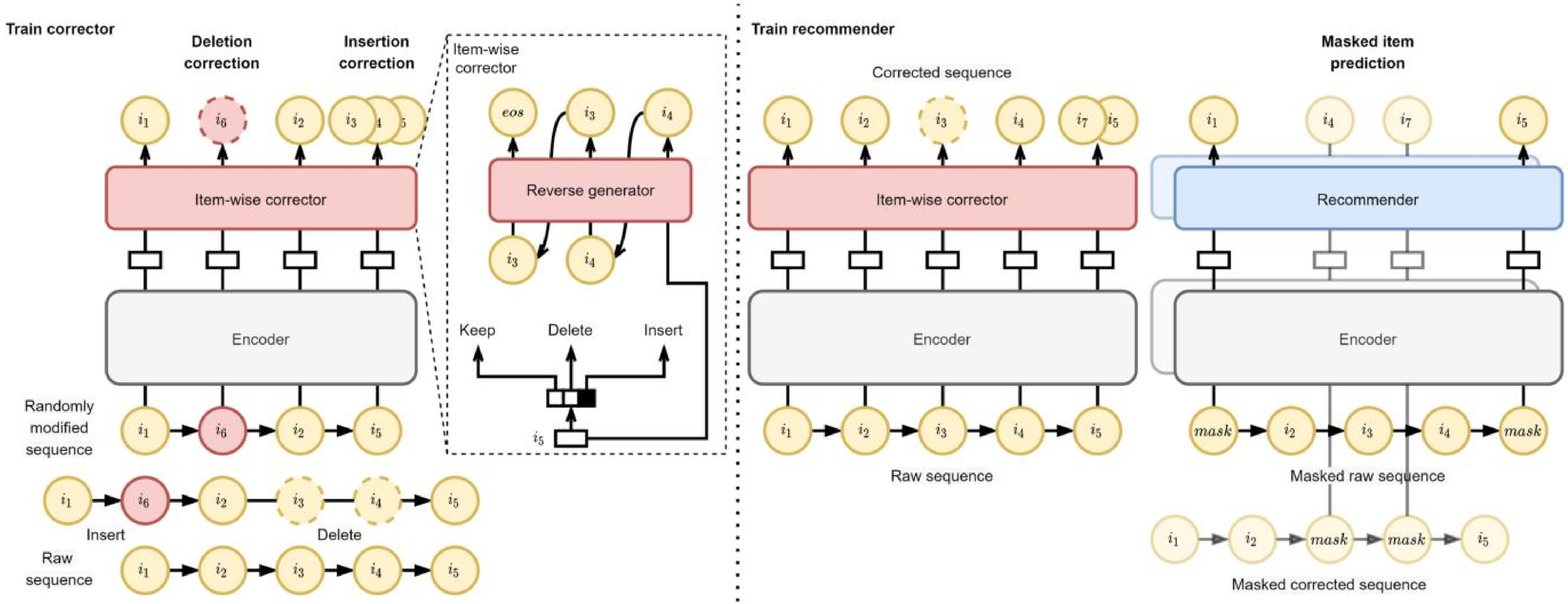
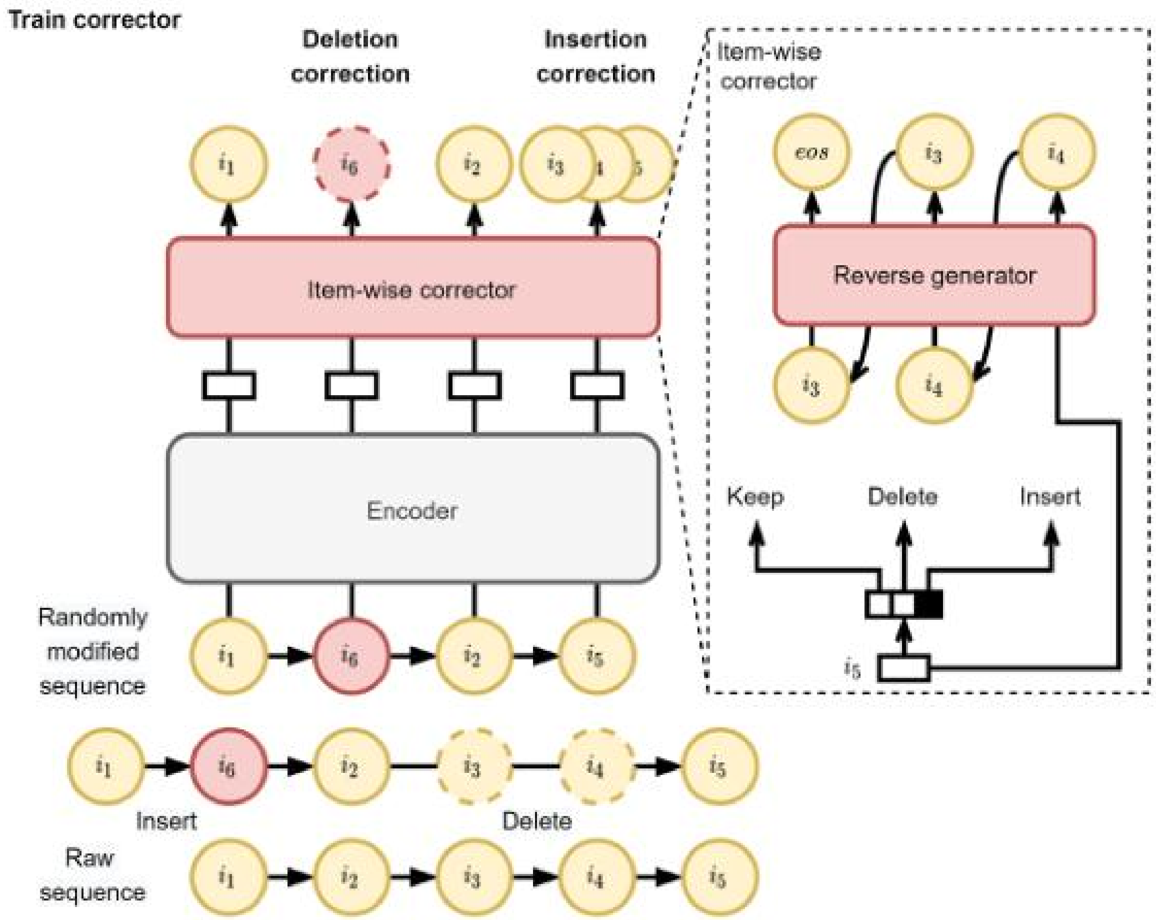


Figure 2: An overview of STEAM. For training the corrector, the item-wise corrector is asked to perform deletion correction and insertion correction on items to recover the raw item sequence that has been randomly modified. The raw item sequence with its corrected version are both used to train the recommender using the masked item prediction task. Finally, STEAM is optimized by the joint loss from the corrector and the recommender.

Method



$$\mathbf{e}_t = \mathbf{E}\mathbf{i}_t \tag{1}$$

$$\mathbf{h}_t^0 = \mathbf{e}_t + \mathbf{p}_t \tag{2}$$

$$\mathbf{H}_e^l = \mathrm{Trm}_{\mathrm{bi}}(\mathbf{H}_e^{l-1}) \tag{3}$$

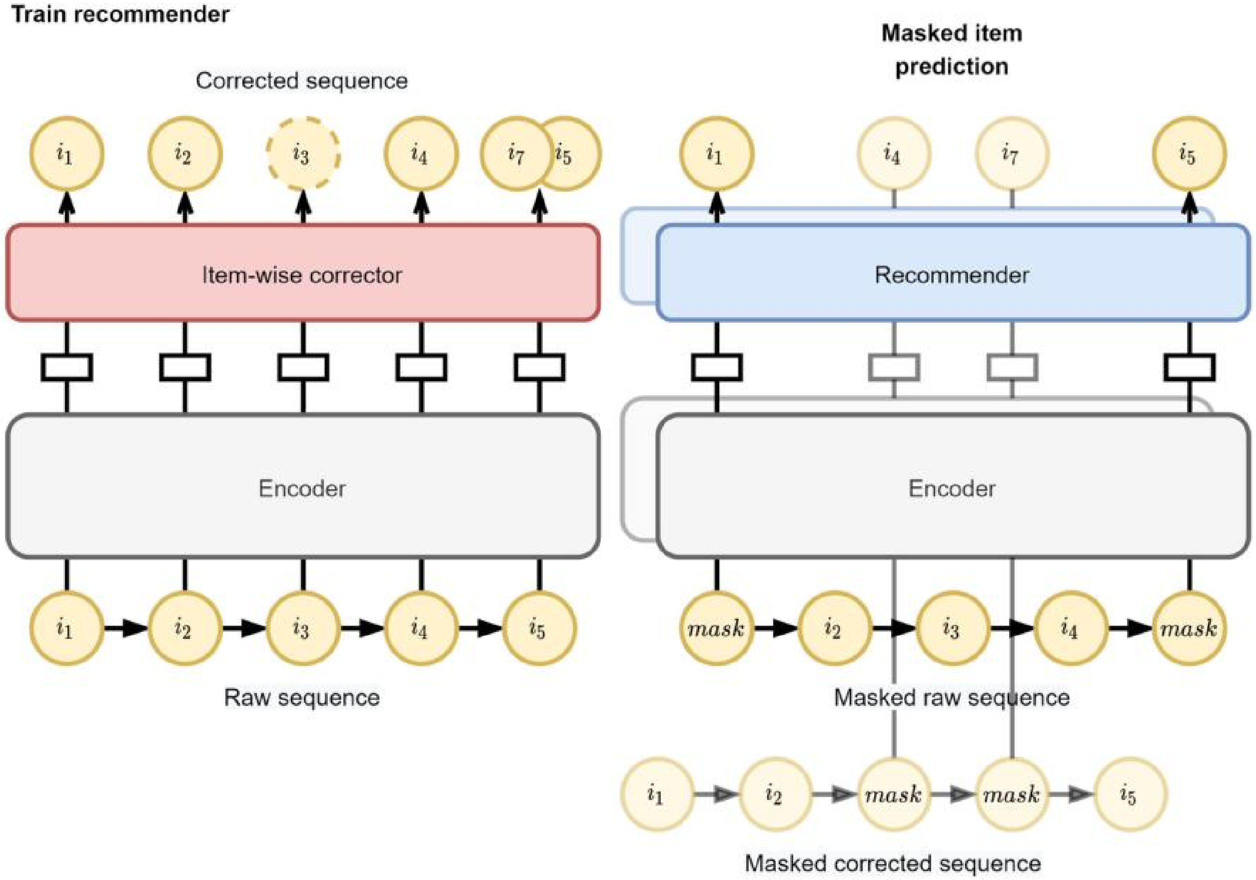
$$P(\hat{o}_t \mid S) = \mathrm{softmax}(\mathbf{W}\mathbf{h}_t) \tag{4}$$

$$\mathbf{H}_c^0 = \begin{bmatrix} \mathbf{h}_t + \mathbf{p}_1 \\ \mathbf{e}_1 + \mathbf{p}_2 \\ \vdots \\ \mathbf{e}_{n-1} + \mathbf{p}_n \end{bmatrix} \tag{5}$$

$$\mathbf{H}_c^l = \mathrm{Trm}_{\mathrm{uni}}(\mathbf{H}_c^{l-1}) \tag{6}$$

$$P(\hat{i}_n \mid S_{1:n-1}^{<i_t}, S) = \mathrm{softmax}(\mathbf{E}^\top \mathbf{h}_n) \tag{7}$$

Method



$$\mathbf{H}_r^l = \mathrm{Trm}_{\mathrm{bi}}(\mathbf{H}_r^{l-1}) \tag{8}$$

$$P(\hat{i}_t \mid S) = \mathrm{softmax}(\mathbf{E}^\top \mathbf{h}_t) \tag{9}$$

$$\begin{aligned} L_1 &= -\log P(S^r|S^m) \\ &= -\left(\log P(O|S^m) + \sum_{i \in I^{ins}} \log P(S^{<i}|S^m)\right) \end{aligned} \tag{10}$$

$$\begin{aligned} L_2 &= -\left(\log P(\widetilde{I^r} \mid \widetilde{S^r}) + \log P(\widetilde{I^c} \mid \widetilde{S^c})\right) \\ &= -\left(\sum_{i \in \widetilde{I^r}} \log P(\hat{i} = i \mid \widetilde{S^r}) + \sum_{i \in \widetilde{I^c}} \log P(\hat{i} = i \mid \widetilde{S^c})\right) \end{aligned} \tag{11}$$

$$L = L_1 + L_2 \tag{12}$$

Experiments

Table 1: Statistics of the datasets after preprocessing.

Dataset	#Users	#Items	#Actions	Avg. length	Sparsity
Beauty	22,362	12,101	194,682	8.7	99.93%
Sports	35,597	18,357	294,483	8.3	99.95%
Yelp	22,844	16,552	236,999	10.4	99.94%

Experiments

Table 2: Performance comparison of different methods on the real test sets. The best performance and the second best performance are denoted in bold and underlined fonts respectively. $*$ indicates that the performance gain of STEAM against the best baseline is statistically significant based on a two-sided paired t-test with $p < 0.05$.

	Real Beauty				Real Sports				Real Yelp			
Model	HR@5	HR@10	MRR@5	MRR@10	HR@5	HR@10	MRR@5	MRR@10	HR@5	HR@10	MRR@5	MRR@10
GRU4Rec	32.95	42.59	21.63	22.90	30.58	42.85	18.35	19.97	55.40	76.57	32.23	35.05
SASRec	36.58	45.57	25.43	26.62	34.51	46.20	21.91	23.46	58.24	77.96	35.07	37.72
BERT4Rec	36.67	47.28	23.38	24.79	35.16	47.91	21.54	23.24	61.18	79.72	37.64	40.13
SRGNN	37.33	47.65	25.15	26.52	35.92	48.32	22.44	24.08	59.86	78.96	36.74	39.30
CL4SRec	39.29	48.75	27.59	28.84	37.91	49.83	24.53	26.11	62.15	80.16	39.29	41.70
DuoRec	<u>40.95</u>	<u>50.78</u>	**28.84**	**30.15**	<u>39.80</u>	<u>51.93</u>	<u>25.97</u>	<u>27.58</u>	<u>64.01</u>	<u>82.63</u>	<u>40.85</u>	<u>43.34</u>
FMLP-Rec	39.69	48.72	28.01	29.20	37.67	49.32	24.66	26.21	61.85	80.76	38.38	40.92
Recommender	35.73	46.47	22.84	24.27	35.02	47.78	21.34	23.03	61.41	80.57	37.67	40.22
STEAM	**42.57***	**52.89***	<u>28.75</u>	<u>30.14</u>	**42.14***	**55.16***	**26.87***	**28.61***	**67.22***	**84.49***	**43.45***	**45.77***

Experiments

Table 3: Performance analysis of STEAM on different groups of the real test sets. Overall-R (Overall-C) is the performance on all raw (corrected) test item sequences. Changed-R (Changed-C) is the performance on the raw (corrected) test item sequences of the changed sequence group. Unchanged is the performance on the test item sequences of the unchanged sequence group.

	Real Beauty				Real Sports				Real Yelp			
STEAM	HR@5	HR@10	MRR@5	MRR@10	HR@5	HR@10	MRR@5	MRR@10	HR@5	HR@10	MRR@5	MRR@10
Overall-R	42.21	52.75	28.27	29.68	42.03	55.04	26.75	28.48	67.19	84.49	43.42	45.75
Overall-C	42.57	52.89	28.75	30.14	42.14	55.16	26.87	28.61	67.22	84.49	43.45	45.77
Changed-R	41.35	51.59	27.04	28.40	35.04	47.64	21.56	23.23	56.05	74.19	34.36	36.80
Changed-C	42.56	52.06	28.66	29.94	35.54	48.12	22.08	23.76	57.46	74.40	35.45	37.73
Unchanged	42.58	53.25	28.79	30.22	44.21	57.36	28.37	30.12	67.44	84.71	43.62	45.95

Experiments

Table 4: Statistics of correction operations by STEAM on the real test sets. #Changed is the percentage of the changed test item sequences after correction. #Keep, #Delete and #Insert are the percentages of different types of correction operations during correction.

Dataset	#Changed	#Keep	#Delete	#Insert
Real Beauty	29.91	88.60	4.03	7.37
Real Sports	23.82	95.72	4.21	0.07
Real Yelp	2.17	99.63	0.15	0.22

Experiments

Table 5: Performance comparison of different methods on the simulated test sets.

	Simulated Beauty				Simulated Sports				Simulated Yelp			
Model	HR@5	HR@10	MRR@5	MRR@10	HR@5	HR@10	MRR@5	MRR@10	HR@5	HR@10	MRR@5	MRR@10
GRU4Rec	32.22	42.13	21.28	22.59	29.96	42.26	17.99	19.61	54.64	75.87	31.66	34.49
SASRec	35.97	45.26	24.97	26.20	33.63	45.23	21.47	23.01	57.71	77.12	34.64	37.23
BERT4Rec	35.83	46.79	22.79	24.25	34.10	46.49	20.62	22.26	59.46	78.07	36.36	38.85
SRGNN	36.64	46.81	24.50	25.85	35.39	47.55	22.00	23.60	57.55	76.82	35.09	37.68
CL4SRec	38.66	48.22	26.96	28.23	37.10	48.93	23.95	25.52	61.08	78.99	38.48	40.88
DuoRec	<u>40.26</u>	<u>50.13</u>	<u>28.39</u>	<u>29.71</u>	<u>38.87</u>	<u>50.95</u>	<u>25.36</u>	<u>26.96</u>	<u>63.06</u>	<u>82.07</u>	<u>40.24</u>	<u>42.78</u>
FMLP-Rec	39.38	48.47	27.85	29.06	37.23	48.86	24.33	25.87	61.17	80.37	37.97	40.56
Recommender	35.14	45.96	22.22	23.66	33.70	46.40	20.38	22.06	60.33	79.08	36.52	39.03
STEAM	**42.09***	**52.21***	**28.45**	**29.81**	**41.72***	**54.82***	**26.43***	**28.17***	**66.46***	**84.05***	**42.83***	**45.19***

Experiments

Table 6: Robustness analysis of different models. Each value is a performance disturbance.

	Beauty				Sports				Yelp			
Model	HR@5	HR@10	MRR@5	MRR@10	HR@5	HR@10	MRR@5	MRR@10	HR@5	HR@10	MRR@5	MRR@10
GRU4Rec	-2.21%	-1.08%	-1.62%	-1.35%	-2.03%	-1.38%	-1.96%	-1.80%	-1.37%	-0.91%	-1.77%	-1.60%
SASRec	-1.67%	<u>-0.68%</u>	-1.81%	-1.58%	-2.55%	-2.10%	-2.01%	-1.92%	**-0.91%**	-1.08%	<u>-1.23%</u>	-1.30%
BERT4Rec	-2.29%	-1.04%	-2.52%	-2.18%	-3.01%	-2.96%	-4.27%	-4.22%	-2.81%	-2.07%	-3.40%	-3.19%
SRGNN	-1.85%	-1.76%	-2.58%	-2.53%	-1.48%	-1.59%	-1.96%	-1.99%	-3.86%	-2.71%	-4.49%	-4.12%
CL4SRec	-1.60%	-1.09%	-2.28%	-2.12%	-2.14%	-1.81%	-2.36%	-2.26%	-1.72%	-1.46%	-2.06%	-1.97%
DuoRec	-1.68%	-1.28%	-1.56%	-1.46%	-2.34%	-1.89%	-2.35%	-2.25%	-1.48%	-0.68%	-1.49%	-1.29%
FMLP-Rec	<u>-0.78%</u>	**-0.51%**	<u>-0.57%</u>	<u>-0.48%</u>	<u>-1.17%</u>	<u>-0.93%</u>	<u>-1.34%</u>	<u>-1.30%</u>	-1.10%	**-0.48%**	**-1.07%**	**-0.88%**
Recommender	-1.65%	-1.10%	-2.71%	-2.51%	-3.77%	-2.89%	-4.50%	-4.21%	-1.76%	-1.85%	-3.05%	-2.96%
STEAM	**-0.28%**	-1.02%	**+0.64%**	**+0.44%**	**-0.74%**	**-0.40%**	**-1.20%**	**-1.09%**	<u>-1.09%</u>	<u>-0.52%</u>	-1.36%	<u>-1.22%</u>

Thanks