APPENDIX

TABLE 6
Representative datasets in various domains for recommendation, where 'SNs' and 'LBSNs' are short for social networks and the location-based social networks, respectively.

Domain	Representative Datasets
Movie	MovieLens (100K/1M/10M/20M/25M/Latest), Netflix, Amazon, FilmTrust, Douban, Yahoo!Movie, Flixster, CiaoDVD, EachMovie, IMDB, Watcha [113]
Consumable	Amazon (Clothing, Home, Office, Sports, Electronic, Cell Phones, Beauty, Food, Health, Toy, Baby), Taobao, Beibei, Tmall, Alibaba, Retailrocket, Flipkart [37]
Music	Last.fm, Yahoo!Music, Douban, Amazon-CDs, KKBox, Kollect.fm, EchoNest, MSD [191], TasteProfile [35]
Book	Amazon-Book/Kindle, Douban, Dianping, IntentBooks, Book-Crossing, LibraryThing, GoodReads
News	Bing-News, Baidu-News, Medium, Adressa-News [68], Microsoft News [93]
SNs	Epinions, Delicious, Douban, Last.fm, Yelp, Ciao, Xing, FilmTrust, Twitter, Wechat, Weibo
LBSNs	Foursquare, Yelp, Gowalla, BrightKite
Paper	Aminer, CiteULike, PRSDataset [24]
Image	Pinterest, Flickr, Aesthetic Visual Analysis [135]

TABLE 7
The abbreviation and full names of the eight conferences.

Abbre.	Full Names
AAAI	AAAI Conference on Artificial Intelligence
CIKM	Conference on Information and Knowledge Management
IJCAI	International Joint Conference on Artificial Intelligence
KDD	SIGKDD Conference on Knowledge Discovery and Data Mining
RecSys	ACM Conference on Recommender System
SIGIR	SIGIR Conference on Research and Development in Information Retrieval
WSDM	International Conference on Web Search and Data Mining
WWW	The ACM Web Conference

TABLE 8
The equations for the six evaluation metrics.

$$Precision@N = \frac{1}{|\mathcal{U}|}\sum_u \frac{1}{N}\sum_{j=1}^{N} rel_j$$

$$Recall@N = \frac{1}{|\mathcal{U}|}\sum_u \frac{1}{|T(u)|}\sum_{j=1}^{N} rel_j$$

$$HR@N = \frac{1}{|\mathcal{U}|}\sum_u \delta(R(u)\cap T(u) \neq \emptyset)$$

$$MAP@N = \frac{1}{|\mathcal{U}|}\sum_u \frac{1}{N}\sum_{k=1}^{N} Precision@k$$

$$MRR@N = \frac{1}{|\mathcal{U}|}\sum_u \sum_{j=1}^{N} j^{-1} rel_j$$

$$NDCG@N = \frac{1}{|\mathcal{U}|}\sum_u \frac{DCG@N}{IDCG@N},\ DCG@N = \sum_{j=1}^{N} \frac{2^{rel_j}-1}{log_2(j+1)}$$

TABLE 9
The update rules of commonly-used optimizers.

Optimizer	Update Rule
GD	$\boldsymbol{\theta}_{t+1} \leftarrow \boldsymbol{\theta}_t - \eta\nabla_\theta\mathcal{L}(\boldsymbol{\theta}_t)$
SGD	$\boldsymbol{\theta}_{t+1} \leftarrow \boldsymbol{\theta}_t - \eta\nabla_\theta\mathcal{L}(\boldsymbol{\theta}_t;(u,i))$
MB-SGD	$\boldsymbol{\theta}_{t+1} \leftarrow \boldsymbol{\theta}_t - \eta\nabla_\theta\mathcal{L}(\boldsymbol{\theta}_t;B(u,i))$
AdaGrad	$\boldsymbol{\theta}_{t+1} \leftarrow \boldsymbol{\theta}_t - \frac{\eta}{\sqrt{\boldsymbol{G}_t+\epsilon}}\cdot \boldsymbol{g}_t;\ \boldsymbol{g}_t = \nabla_\theta\mathcal{L}(\boldsymbol{\theta}_t)$
RMSProp	$\boldsymbol{\theta}_{t+1} \leftarrow \boldsymbol{\theta}_t - \frac{\eta}{\sqrt{\mathbb{E}[g^2]_t+\epsilon}}\cdot \boldsymbol{g}_t;\ \mathbb{E}(\boldsymbol{g}^2)_t = \gamma\mathbb{E}[\boldsymbol{g}^2]_{t-1} + (1-\gamma)\boldsymbol{g}_t^2$
Adam	$\boldsymbol{\theta}_{t+1} \leftarrow \boldsymbol{\theta}_t - \frac{\eta}{\sqrt{\hat{\boldsymbol{v}}_t+\epsilon}}\cdot \hat{\boldsymbol{m}}_t;\ \hat{\boldsymbol{m}}_t = \frac{\boldsymbol{m}_t}{1-\beta_1^t}\ \hat{\boldsymbol{v}}_t = \frac{\boldsymbol{v}_t}{1-\beta_2^t}$ $\boldsymbol{m}_t = \beta_1\boldsymbol{m}_{t-1} + (1-\beta_1)\boldsymbol{g}_t;\ \boldsymbol{v}_t = \beta_2\boldsymbol{v}_{t-1} + (1-\beta_2)\boldsymbol{g}_t^2;$
Remark	$\boldsymbol{\theta}$ is the learnable parameter; η is the learning rate; (u,i) is one training sample; $B(u,i)$ is a batch of training samples; $\boldsymbol{G}_t \in \mathbb{R}^{d\times d}$ is a diagonal matrix where each diagonal element is the sum of the squares of the gradients w.r.t. $\boldsymbol{\theta}$ up to time step t; ϵ is a smoothing term to avoid division by zero (usually on the order of $1e-8$); $\gamma = 0.9$ is the momentum term; $\boldsymbol{m}_t$ is the decaying average of past gradients; $\boldsymbol{v}_t$ is the decaying average of past squared gradients;

TABLE 10
The performance of ten baselines on ML-1M ($N = 10$) across the six metrics, where the best performance is highlighted in bold.

Origin	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.2554	0.0530	0.6997	0.1734	0.8290	0.4952
ItemKNN	0.2218	0.0676	0.6521	0.1428	0.6804	0.4261
PureSVD	0.2389	0.0760	0.6840	0.1558	0.7430	0.4562
SLIM	–	–	–	–	–	–
BPRMF	0.2854	0.0819	0.7585	0.1924	0.8932	0.5275
BPRFM	**0.4379**	**0.1156**	**0.8667**	**0.3360**	**1.3718**	**0.6745**
NeuMF	0.4210	0.1023	0.8515	0.3215	1.3105	0.6528
NFM	0.4162	0.0999	0.8426	0.3171	1.2995	0.6473
NGCF	0.3684	0.0971	0.8426	0.2646	1.1560	0.6230
Multi-VAE	0.4201	0.1023	0.8532	0.3218	1.3131	0.6538
5-filter	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.2558	0.0531	0.7001	0.1737	0.8303	0.4957
ItemKNN	0.2284	0.0666	0.6502	0.1482	0.6987	0.4319
PureSVD	0.2288	0.0626	0.6541	0.1534	0.7062	0.4334
SLIM	–	–	–	–	–	–
BPRMF	0.2858	0.0712	0.6665	0.2068	0.8888	0.4814
BPRFM	**0.4191**	**0.1077**	**0.8627**	**0.3200**	**1.3213**	**0.6651**
NeuMF	0.4184	0.1073	0.8571	0.3157	1.3029	0.6568
NFM	0.4042	0.0972	0.8397	0.3066	1.2672	0.6381
NGCF	0.2151	0.0618	0.6491	0.1328	0.6487	0.4186
Multi-VAE	0.4056	0.1017	0.8487	0.3062	1.2673	0.6415
10-filter	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.2569	0.0532	0.7024	0.1745	0.8339	0.4975
ItemKNN	0.2208	0.0619	0.6381	0.1443	0.6821	0.4233
PureSVD	0.2270	0.0676	0.6900	0.1416	0.6984	0.4479
SLIM	0.2182	0.0642	0.6832	0.1377	0.6879	0.4450
BPRMF	0.2760	0.0670	0.6421	0.2001	0.8505	0.4601
BPRFM	**0.4032**	**0.0992**	**0.8540**	**0.3034**	**1.2637**	**0.6466**
NeuMF	0.2123	0.0638	0.6471	0.1378	0.6803	0.4322
NFM	0.3898	0.0916	0.8303	0.2923	1.2141	0.6232
NGCF	0.2865	0.0793	0.7909	0.1903	0.9134	0.5494
Multi-VAE	0.3901	0.0913	0.8298	0.2952	1.2335	0.6287

TABLE 11
The performance of ten baselines on LastFM ($N = 10$) across the six metrics, where the best performance is highlighted in bold.

Origin	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.0712	0.0716	0.3978	0.0338	0.2509	0.2311
ItemKNN	**0.4219**	**0.4332**	0.9580	**0.3324**	**1.5841**	**0.8026**
PureSVD	0.4173	0.4304	0.9480	0.3321	1.5807	0.7989
SLIM	–	–	–	–	–	–
BPRMF	0.3796	0.3906	0.9411	0.2867	1.4303	0.7624
BPRFM	0.3367	0.3465	0.8970	0.2415	1.2322	0.6897
NeuMF	0.3561	0.3663	0.9320	0.2552	1.3112	0.7279
NFM	0.2656	0.2723	0.8449	0.1772	0.9723	0.6058
NGCF	0.4059	0.4175	**0.9596**	0.3004	1.4575	0.7688
Multi-VAE	0.2662	0.2724	0.8396	0.1772	0.9659	0.5998
5-filter	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.0743	0.0875	0.4067	0.0388	0.2873	0.2450
ItemKNN	0.2410	0.3065	0.8551	**0.1647**	**0.9835**	0.6341
PureSVD	0.2374	0.2959	0.8507	0.1567	0.9453	0.6197
SLIM	–	–	–	–	–	–
BPRMF	0.2368	0.2975	0.8643	0.1551	0.9349	0.6222
BPRFM	0.2429	0.3042	0.8507	0.1586	0.9399	0.6135
NeuMF	0.2452	0.3074	0.8540	0.1627	0.9731	0.6284
NFM	0.1198	0.1428	0.5717	0.0675	0.4543	0.3625
NGCF	**0.2532**	**0.3252**	**0.8880**	0.1631	0.9698	**0.6381**
Multi-VAE	0.1176	0.1405	0.5695	0.0676	0.4515	0.3577
10-filter	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.0677	0.0880	0.4097	0.0349	0.2712	0.2476
ItemKNN	0.1851	0.2621	0.7840	0.1188	0.7708	0.5577
PureSVD	0.1862	0.2522	0.7556	0.1201	0.7673	0.5418
SLIM	**0.2120**	**0.3064**	**0.8369**	**0.1380**	**0.8713**	**0.6018**
BPRMF	0.1915	0.2657	0.8009	0.1193	0.7696	0.5592
BPRFM	0.1891	0.2631	0.7949	0.1157	0.7551	0.5513
NeuMF	0.1690	0.2382	0.7529	0.1001	0.6621	0.5044
NFM	0.0745	0.0967	0.4495	0.0336	0.2574	0.2532
NGCF	0.1821	0.2659	0.8020	0.1060	0.7019	0.5349
Multi-VAE	0.0855	0.1113	0.4768	0.0447	0.3309	0.2901

TABLE 12
The performance of ten baselines on Book-X ($N = 10$) across the six metrics, where the best performance is highlighted in bold.

Origin	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.0048	0.0150	0.0411	0.0023	0.0214	0.0249
ItemKNN	0.0214	0.0283	0.1027	0.0159	0.1010	0.0782
PureSVD	0.0776	0.1761	0.3247	0.0560	0.3178	0.2344
SLIM	–	–	–	–	–	–
BPRMF	0.0868	0.2309	0.3924	0.0586	0.3450	0.2687
BPRFM	0.0954	**0.2970**	**0.4603**	0.0636	**0.3836**	**0.3115**
NeuMF	0.0693	0.2067	0.3578	0.0421	0.2655	0.2289
NFM	0.0858	0.2757	0.4278	0.0576	0.3548	0.2919
NGCF	**0.0957**	0.2343	0.3947	**0.0673**	0.3787	0.2778
Multi-VAE	0.0858	0.2762	0.4282	0.0576	0.3550	0.2921
5-filter	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.0097	0.0177	0.0789	0.0047	0.0420	0.0476
ItemKNN	0.0419	0.0661	0.2186	0.0283	0.1937	0.1580
PureSVD	0.1030	0.1833	0.4480	**0.0697**	**0.4115**	0.3105
SLIM	–	–	–	–	–	–
BPRMF	0.1057	0.2127	0.4945	0.0637	0.3911	0.3159
BPRFM	0.1021	0.1994	0.4766	0.0609	0.3747	0.3033
NeuMF	0.0839	0.1619	0.4148	0.0480	0.3041	0.2571
NFM	0.0802	0.1549	0.3971	0.0480	0.3030	0.2529
NGCF	**0.1099**	**0.2311**	**0.5083**	0.0652	0.4021	**0.3249**
Multi-VAE	0.0810	0.1568	0.4001	0.0483	0.3046	0.2542
10-filter	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.0134	0.0192	0.1081	0.0062	0.0547	0.0630
ItemKNN	0.0516	0.0786	0.2780	0.0335	0.2314	0.1921
PureSVD	**0.1079**	**0.1622**	0.4740	**0.0695**	**0.4223**	**0.3216**
SLIM	0.0753	0.1014	0.3475	0.0510	0.3264	0.2504
BPRMF	0.0998	0.1619	**0.4801**	0.0565	0.3594	0.2983
BPRFM	0.0931	0.1401	0.4474	0.0524	0.3332	0.2779
NeuMF	0.0985	0.1490	0.4582	0.0564	0.3476	0.2852
NFM	0.0741	0.1119	0.3841	0.0415	0.2748	0.2380
NGCF	0.0818	0.1415	0.4345	0.0417	0.2788	0.2533
Multi-VAE	0.0751	0.1136	0.3867	0.0416	0.2736	0.2378

TABLE 13
The performance of ten baselines on Epinions ($N = 10$) across the six metrics, where the best performance is highlighted in bold.

Origin	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.0016	0.0009	0.0157	0.0005	0.0052	0.0076
ItemKNN	0.0480	0.0193	0.1599	0.0365	0.1753	0.1176
PureSVD	0.0681	0.0413	0.2462	0.0490	0.2424	0.1675
SLIM	–	–	–	–	–	–
BPRMF	0.0721	0.0467	0.2861	0.0483	0.2523	0.1844
BPRFM	0.0774	0.0448	0.3018	0.0494	0.2576	0.1884
NeuMF	**0.0882**	**0.0573**	**0.3519**	**0.0588**	**0.3176**	**0.2329**
NFM	0.0693	0.0416	0.2822	0.0431	0.2322	0.1745
NGCF	0.0819	0.0558	0.3234	0.0553	0.2852	0.2069
Multi-VAE	0.0696	0.0420	0.2822	0.0432	0.2317	0.1743
5-filter	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.0033	0.0039	0.0299	0.0012	0.0111	0.0152
ItemKNN	0.0354	0.0309	0.1668	0.0222	0.1271	0.1037
PureSVD	0.0335	0.0311	0.1651	0.0188	0.1117	0.0962
SLIM	–	–	–	–	–	–
BPRMF	0.0369	0.0378	0.1825	0.0219	0.1317	0.1115
BPRFM	**0.0391**	0.0354	0.2002	0.0220	0.1368	0.1192
NeuMF	0.0382	0.0383	**0.2095**	0.0215	**0.1423**	**0.1279**
NFM	0.0289	0.0252	0.1638	0.0150	0.1008	0.0946
NGCF	0.0385	**0.0406**	0.1777	**0.0230**	0.1335	0.1086
Multi-VAE	0.0283	0.0245	0.1611	0.0148	0.0996	0.0932
10-filter	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.0040	0.0058	0.0370	0.0015	0.0139	0.0189
ItemKNN	0.0261	0.0315	0.1388	0.0150	0.0916	0.0813
PureSVD	0.0238	0.0285	0.1332	0.0124	0.0782	0.0736
SLIM	0.0260	0.0327	0.1423	0.0147	0.0936	0.0840
BPRMF	0.0258	0.0318	0.1407	0.0139	0.0893	0.0820
BPRFM	0.0295	0.0335	0.1636	0.0158	0.1061	0.0979
NeuMF	0.0302	0.0375	0.1772	0.0164	**0.1160**	**0.1084**
NFM	0.0200	0.0219	0.1278	0.0101	0.0756	0.0756
NGCF	**0.0326**	**0.0402**	**0.1876**	**0.0165**	0.1092	0.1052
Multi-VAE	0.0199	0.0216	0.1265	0.0100	0.0748	0.0749

TABLE 14
The performance of ten baselines on Yelp ($N = 10$) across the six metrics, where the best performance is highlighted in bold.

Origin	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.0012	0.0058	0.0113	0.0005	0.0048	0.0061
ItemKNN	0.0128	0.0302	0.0699	0.0094	0.0628	0.0524
PureSVD	0.0365	0.1168	0.1938	0.0218	0.1417	0.1222
SLIM	–	–	–	–	–	–
BPRMF	0.0425	0.1573	0.2447	0.0234	0.1594	0.1471
BPRFM	**0.0508**	**0.2144**	**0.3141**	0.0271	**0.1925**	**0.1859**
NeuMF	0.0500	0.2092	0.3048	**0.0272**	0.1916	0.1827
NFM	0.0349	0.1598	0.2431	0.0175	0.1369	0.1425
NGCF	–	–	–	–	–	–
Multi-VAE	0.0352	0.1612	0.2445	0.0176	0.1379	0.1433
5-filter	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.0024	0.0044	0.0199	0.0010	0.0090	0.0109
ItemKNN	0.0646	0.1227	0.3061	0.0452	0.2815	0.2208
PureSVD	0.1066	0.2259	0.4823	0.0671	0.4053	0.3162
SLIM	–	–	–	–	–	–
BPRMF	0.1061	0.2260	0.4900	0.0623	0.3809	0.3060
BPRFM	0.1099	0.2262	0.5079	0.0629	0.3811	0.3088
NeuMF	0.1106	0.2370	0.5259	0.0636	0.3975	0.3264
NFM	0.0612	0.1149	0.3322	0.0317	0.2142	0.1931
NGCF	**0.1274**	**0.2832**	**0.5599**	**0.0764**	**0.4574**	**0.3573**
Multi-VAE	0.0621	0.1168	0.3348	0.0322	0.2165	0.1944
10-filter	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.0031	0.0040	0.0262	0.0012	0.0111	0.0137
ItemKNN	0.1174	0.1765	0.4882	0.0792	0.4600	0.3380
PureSVD	0.1508	0.2239	0.5837	0.0975	0.5493	0.3918
SLIM	0.1128	0.1655	0.4819	0.0778	0.4625	0.3432
BPRMF	0.1455	0.2171	0.5849	0.0886	0.5104	0.3770
BPRFM	0.1448	0.2117	0.5847	0.0866	0.4955	0.3701
NeuMF	0.1562	0.2354	0.6187	0.0961	0.5515	0.4032
NFM	0.0697	0.0894	0.3510	0.0360	0.2322	0.2028
NGCF	**0.1685**	**0.2583**	**0.6347**	**0.1051**	**0.5878**	**0.4187**
Multi-VAE	0.0702	0.0899	0.3539	0.0361	0.2331	0.2041

TABLE 15
The performance of ten baselines on AMZe ($N = 10$) across the six metrics, where the best performance is highlighted in bold.

Origin	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.0025	0.0169	0.0245	0.0006	0.0063	0.0104
ItemKNN	0.0018	0.0114	0.0176	0.0008	0.0074	0.0096
PureSVD	0.0202	0.1095	0.1474	0.0117	0.0971	0.0969
SLIM	–	–	–	–	–	–
BPRMF	0.0323	0.2060	0.2642	0.0170	0.1510	0.1655
BPRFM	0.0518	0.3530	0.4249	0.0286	0.2549	0.2744
NeuMF	**0.0518**	**0.3531**	**0.4250**	**0.0286**	0.2550	**0.2745**
NFM	0.0517	0.3524	0.4243	0.0285	0.2546	0.2741
NGCF	–	–	–	–	–	–
Multi-VAE	0.0518	0.3531	0.4249	0.0286	**0.2552**	0.2745
5-filter	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.0064	0.0193	0.0612	0.0026	0.0257	0.0333
ItemKNN	0.0072	0.0224	0.0640	0.0041	0.0378	0.0419
PureSVD	0.0614	0.2104	0.4070	0.0348	0.2791	0.2654
SLIM	–	–	–	–	–	–
BPRMF	0.0657	0.2178	0.4335	0.0375	0.3012	0.2852
BPRFM	0.0759	0.2350	0.4776	**0.0442**	**0.3466**	**0.3177**
NeuMF	**0.0761**	0.2354	0.4789	0.0440	0.3453	0.3172
NFM	0.0760	**0.2356**	**0.4792**	0.0441	0.3458	0.3176
NGCF	0.0492	0.1711	0.3449	0.0239	0.1960	0.2038
Multi-VAE	0.0759	0.2349	0.4775	0.0441	0.3461	0.3174
10-filter	Pre	Rec	HR	MAP	MRR	NDCG
MostPop	0.0085	0.0179	0.0793	0.0040	0.0387	0.0470
ItemKNN	0.0136	0.0312	0.1136	0.0067	0.0603	0.0679
PureSVD	0.0758	0.1781	0.4530	0.0429	0.3248	0.2935
SLIM	0.0342	0.0802	0.2426	0.0213	0.1769	0.1702
BPRMF	0.0806	0.1875	0.4738	0.0463	0.3500	0.3117
BPRFM	0.0864	0.1901	0.4919	0.0504	0.3746	0.3264
NeuMF	0.0839	0.1852	0.4861	0.0493	0.3694	0.3244
NFM	0.0866	0.1906	0.4919	0.0509	0.3780	0.3276
NGCF	0.0663	0.1561	0.4127	0.0370	0.2880	0.2664
Multi-VAE	**0.0869**	**0.1909**	**0.4925**	**0.0511**	**0.3790**	**0.3285**

TABLE 16
The opitmal hyper-parameter settings found by Bayesian HyperOpt for different baselines with time-aware split-by-ratio under origin view.

Origin	Parameter	ML-1M	LastFM	Book-X	Epinions	Yelp	AMZe	Searching space	Description
ItemKNN	–makx	8	49	83	14	31	9	$[1, 100]$	the number of neighbors
PureSVD	–factors	39	24	10	99	96	6	$[1, 100]$	the number of singular values
BPRMF	–num_ng	4	5	5	5	5	4	$[1, 5]$	the number of negative items
	–factors	83	92	48	97	72	74	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0056	0.0080	0.0098	0.0095	0.0089	0.0047	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0002	0.0012	0.0008	0.0002	0.0088	0.0016	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–batch_size	128	64	64	128	1024	512	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
BPRFM	–num_ng	4	5	5	5	5	2	$[1, 5]$	the number of negative items
	–factors	30	24	83	33	45	66	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0030	0.0100	0.0093	0.0091	0.0100	0.0004	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0007	0.0056	0.0006	0.0095	0.0012	0.0037	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–batch_size	256	512	64	256	256	1024	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
NeuMF	–num_ng	4	2	3	3	3	4	$[1, 5]$	the number of negative items
	–factors	15	49	100	78	63	60	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0002	0.0016	0.0005	0.0003	0.0012	0.0006	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0010	0.0016	0.0064	0.0086	0.0061	0.0003	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–num_layers	3	3	2	3	3	3	$[1, 3]$	the number of layers for MLP
	–dropout	0.9531	0.7890	0.0156	0.0001	0.0010	0.7948	$[0, 1]$	dropout ratio
	–batch_size	64	512	64	64	256	256	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
NFM	–num_ng	1	2	3	5	3	3	$[1, 5]$	the number of negative items
	–factors	11	63	42	5	3	100	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0006	0.0030	0.0005	0.0019	0.0100	0.0007	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0017	0.0006	0.0005	0.0002	0.0067	0.0030	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–num_layers	3	3	3	2	3	1	$[1, 3]$	the number of layers for MLP
	–dropout	0.7108	0.9386	0.9338	0.7830	0.9957	0.9694	$[0, 1]$	dropout ratio
	–batch_size	256	64	64	256	512	1024	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
NGCF	–num_ng	4	4	4	5	4	3	$[1, 5]$	the number of negative items
	–factors	35	53	94	91	84	45	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0007	0.0034	0.0007	0.0008	0.0013	0.0013	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0065	0.0084	0.0001	0.0002	0.0071	0.0088	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–prop_layer	3	3	3	3	3	3	–	the number of propagation layers
	–node_dropout	0.0106	0.4160	0.9431	0.0019	0.0009	0.0015	$[0, 1]$	node dropout ratio
	–mess_dropout	0.0573	0.0023	0.0001	0.0382	0.0048	0.0080	$[0, 1]$	message dropout ratio
	–batch_size	128	128	128	128	512	256	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
Multi-VAE	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0003	0.0006	0.0001	0.0001	0.0007	0.0004	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0003	0.0062	0.0037	0.0001	0.0010	0.0048	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–kl_reg	0.0012	0.5975	0.0570	0.0009	0.0005	0.4945	$[0, 1]$	vae kl regularization
	–dropout	0.0415	0.6912	0.0001	0.0240	0.0263	0.9603	$[0, 1]$	dropout ratio
	–batch_size	128	64	512	128	512	512	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
Remarks	1. The searching space of batch size on ML-1M, LastFM, Book-X, and Epinions is $[2^6, 2^7, 2^8, 2^9]$; while on Yelp and AMZe is $[2^8, 2^9, 2^{10}]$ to speed up the training. 2. Early-stop mechanism is applied to all approaches. 3. The results of SLIM with origin and 5-filter views are not available due to either lack of computational memory or unreasonable amount of time in searching the best hyper-parameter settings. 4. The following results with origin view are not available due to unreasonable amount of time in searching the best hyper-parameter settings. In particular, we use the optimal parameter settings on 5-filter for NGCF and Multi-VAE on Book-X; Multi-VAE on Epinions; NeuMF and NFM on Yelp; BPRMF, BPRFM, NeuMF, NFM and NGCF on AMZe. We use the optimal parameter settings on 10-filter for NGCF and Multi-VAE on Yelp; and Multi-VAE on AMZe.								

TABLE 17
The opitmal hyper-parameter settings found by Bayesian HyperOpt for different baselines with time-aware split-by-ratio under 5-filter view.

5-filter	Parameter	ML-1M	LastFM	Book-X	Epinions	Yelp	AMZe	Searching space	Description
ItemKNN	–makx	73	55	65	30	75	41	$[1, 100]$	the number of neighbors
PureSVD	–factors	2	33	80	70	100	2	$[1, 100]$	the number of singular values
BPRMF	–num_ng	1	5	4	5	5	4	$[1, 5]$	the number of negative items
	–factors	100	56	99	70	92	74	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0018	0.0060	0.0093	0.0098	0.0051	0.0047	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0007	0.0003	0.0014	0.0029	0.0001	0.0016	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–batch_size	64	128	128	128	512	512	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
BPRFM	–num_ng	3	5	5	4	5	2	$[1, 5]$	the number of negative items
	–factors	96	66	98	40	66	66	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0023	0.0099	0.0080	0.0095	0.0097	0.0004	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0001	0.0001	0.0010	0.0026	0.0003	0.0037	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–batch_size	64	256	512	256	512	1024	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
NeuMF	–num_ng	5	4	4	3	3	4	$[1, 5]$	the number of negative items
	–factors	47	67	16	48	63	60	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0007	0.0002	0.0021	0.0001	0.0012	0.0006	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0039	0.0009	0.0056	0.0001	0.0061	0.0003	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–num_layers	1	3	3	3	3	3	$[1, 3]$	the number of layers for MLP
	–dropout	0.5919	0.9973	0.8592	0.0932	0.0010	0.7948	$[0, 1]$	dropout ratio
	–batch_size	128	512	256	128	256	256	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
NFM	–num_ng	5	5	5	3	3	3	$[1, 5]$	the number of negative items
	–factors	29	45	2	2	3	100	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0012	0.0031	0.0041	0.0027	0.0100	0.0007	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0059	0.0062	0.0004	0.0047	0.0067	0.0030	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–num_layers	2	1	1	3	3	1	$[1, 3]$	the number of layers for MLP
	–dropout	0.9815	0.9977	0.8360	0.6564	0.9957	0.9694	$[0, 1]$	dropout ratio
	–batch_size	512	512	256	256	512	1024	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
NGCF	–num_ng	2	4	4	5	4	3	$[1, 5]$	the number of negative items
	–factors	96	97	94	96	84	45	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0005	0.0018	0.0007	0.0010	0.0013	0.0013	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0003	0.0054	0.0001	0.0035	0.0071	0.0088	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–prop_layer	3	3	3	3	3	3	–	the number of propagation layers
	–node_dropout	0.0192	0.0029	0.9431	0.5553	0.0009	0.0015	$[0, 1]$	node dropout ratio
	–mess_dropout	0.0178	0.3631	0.0001	0.0013	0.0048	0.0080	$[0, 1]$	message dropout ratio
	–batch_size	512	256	128	128	512	256	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
Multi-VAE	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0007	0.0002	0.0001	0.0001	0.0007	0.0004	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0003	0.0001	0.0037	0.0001	0.0011	0.0048	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–kl_reg	0.0004	0.6766	0.0570	0.0009	0.0005	0.4945	$[0, 1]$	vae kl regularization
	–dropout	0.0222	0.0004	0.0001	0.0240	0.0263	0.9603	$[0, 1]$	dropout ratio
	–batch_size	256	128	512	128	512	512	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
Remarks	The following results with 5-filter view are not available due to unreasonable amount of time in searching the best hyper-parameter settings. In particular, we use the optimal parameter settings on 10-filter for NGCF and Multi-VAE on Yelp; and Multi-VAE on AMZe.								

TABLE 18
The opitmal hyper-parameter settings found by Bayesian HyperOpt for different baselines with time-aware split-by-ratio under 10-filter view.

10-filter	Parameter	ML-1M	LastFM	Book-X	Epinions	Yelp	AMZe	Searching space	Description
ItemKNN	–makx	100	28	71	14	73	68	$[1, 100]$	the number of neighbors
PureSVD	–factors	39	12	98	98	93	9	$[1, 100]$	the number of singular values
BPRMF	–num_ng	3	5	3	5	5	3	$[1, 5]$	the number of negative items
	–factors	22	92	88	49	98	97	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0004	0.0057	0.0098	0.0097	0.0099	0.0078	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0002	0.0088	0.0001	0.0050	0.0007	0.0004	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–batch_size	64	128	256	128	1024	1024	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
BPRFM	–num_ng	4	5	3	5	5	3	$[1, 5]$	the number of negative items
	–factors	37	61	94	72	100	37	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0014	0.0096	0.0067	0.0078	0.0096	0.0013	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0004	0.0010	0.0001	0.0019	0.0004	0.0049	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–batch_size	128	512	64	64	256	256	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
SLIM	–l1_ratio	0.4462	0.8330	0.8188	0.3951	0.0459	0.3299	$(0, 1]$	the ElasticNet mixing parameter
	–lambda	0.0001	0.0013	0.0072	0.0022	0.0040	0.0020	$[10^{-4}, 10^{-2}]$	constant to multiply penalty terms
NeuMF	–num_ng	5	2	5	5	5	5	$[1, 5]$	the number of negative items
	–factors	31	93	79	41	68	83	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0008	0.0036	0.0003	0.0012	0.0001	0.0004	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0008	0.0011	0.0096	0.0057	0.0009	0.0015	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–num_layers	2	2	3	3	2	2	$[1, 3]$	the number of layers for MLP
	–dropout	0.9592	0.5608	0.4719	0.4771	0.7371	0.7162	$[0, 1]$	dropout ratio
	–batch_size	64	512	64	128	1024	256	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
NFM	–num_ng	5	4	1	2	4	3	$[1, 5]$	the number of negative items
	–factors	19	69	74	88	12	99	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0045	0.0002	0.0019	0.0005	0.0005	0.0016	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0026	0.0006	0.0003	0.0071	0.0009	0.0004	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–num_layers	3	3	2	2	1	3	$[1, 3]$	the number of layers for MLP
	–dropout	0.9818	0.8898	0.9979	0.8666	0.8630	0.9862	$[0, 1]$	dropout ratio
	–batch_size	256	64	128	64	1024	512	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
NGCF	–num_ng	5	5	4	5	4	2	$[1, 5]$	the number of negative items
	–factors	18	74	36	85	84	68	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0004	0.0010	0.0004	0.0017	0.0013	0.0001	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0006	0.0006	0.0005	0.0066	0.0071	0.0042	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–prop_layer	3	3	3	3	3	3	–	the number of propagation layers
	–node_dropout	0.0005	0.0021	0.0002	0.0169	0.0009	0.3658	$[0, 1]$	node dropout ratio
	–mess_dropout	0.0175	0.0015	0.0016	0.0011	0.0048	0.0115	$[0, 1]$	message dropout ratio
	–batch_size	512	256	128	128	512	256	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
Multi-VAE	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0001	0.0005	0.0003	0.0003	0.0007	0.0004	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0010	0.0004	0.0050	0.0004	0.0011	0.0048	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–kl_reg	0.0003	0.0408	0.0009	0.0003	0.0005	0.4945	$[0, 1]$	vae kl regularization
	–dropout	0.0018	0.5885	0.0083	0.2841	0.0263	0.9603	$[0, 1]$	dropout ratio
	–batch_size	64	512	256	64	512	512	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
Remarks	The detailed explanation for the parameters of SLIM is available at https://lijiancheng0614.github.io/scikit-learn/modules/generated /sklearn.linear_model.ElasticNet.html								

TABLE 19
The opitmal hyper-parameter settings found by Bayesian HyperOpt for different baselines with time-aware leave-one-out under 10-filter view.

10-filter	Parameter	ML-1M	LastFM	Book-X	Epinions	Yelp	AMZe	Searching space	Description
ItemKNN	–makx	22	70	100	51	52	40	$[1, 100]$	the number of neighbors
PureSVD	–factors	45	26	89	96	97	11	$[1, 100]$	the number of singular values
BPRMF	–num_ng	5	4	5	5	5	5	$[1, 5]$	the number of negative items
	–factors	49	66	76	56	61	100	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0040	0.0082	0.0098	0.0098	0.0047	0.0051	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0014	0.0007	0.0013	0.0016	0.0012	0.0006	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–batch_size	128	256	256	64	1024	256	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
BPRFM	–num_ng	5	5	5	5	5	1	$[1, 5]$	the number of negative items
	–factors	99	71	99	87	72	71	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0074	0.0075	0.0096	0.0091	0.0098	0.0002	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0004	0.0079	0.0011	0.0001	0.0017	0.0007	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–batch_size	256	256	128	256	512	1024	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
SLIM	–l1_ratio	0.6215	0.0126	0.0023	0.2328	0.0754	0.1056	$(0, 1]$	the ElasticNet mixing parameter
	–lambda	0.0084	0.0055	0.0098	0.0072	0.0090	0.0066	$[10^{-4}, 10^{-2}]$	constant to multiply penalty terms
NeuMF	–num_ng	5	5	3	3	3	2	$[1, 5]$	the number of negative items
	–factors	29	65	68	61	84	78	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0010	0.0001	0.0091	0.0002	0.0003	0.0013	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0049	0.0004	0.0093	0.0023	0.0005	0.0050	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–num_layers	2	1	3	3	3	2	$[1, 3]$	the number of layers for MLP
	–dropout	0.1664	0.0933	0.0008	0.2623	0.5016	0.7352	$[0, 1]$	dropout ratio
	–batch_size	512	512	64	512	256	512	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
NFM	–num_ng	4	1	5	2	1	5	$[1, 5]$	the number of negative items
	–factors	40	2	73	37	1	54	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0026	0.0044	0.0024	0.0019	0.0026	0.0004	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0004	0.0002	0.0003	0.0025	0.0005	0.0003	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–num_layers	1	3	3	3	3		$[1, 3]$	the number of layers for MLP
	–dropout	0.7825	0.9421	0.8430	0.9916	0.1267	0.9864	$[0, 1]$	dropout ratio
	–batch_size	128	512	64	64	256	256	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
NGCF	–num_ng	5	5	5	5	5	4	$[1, 5]$	the number of negative items
	–factors	75	85	63	74	85	2	$[1, 100]$	the dimension of latent factors
	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0007	0.0010	0.0004	0.0004	0.0008	0.0002	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0061	0.0093	0.0100	0.0096	0.0030	0.0006	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–prop_layer	3	3	3	3	3	3	–	the number of propagation layers
	–node_dropout	0.0067	0.0007	0.9751	0.0001	0.0288	0.0001	$[0, 1]$	node dropout ratio
	–mess_dropout	0.0048	0.9363	0.0032	0.0007	0.0040	0.9876	$[0, 1]$	message dropout ratio
	–batch_size	64	64	256	512	512	256	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
Multi-VAE	–epochs	50	50	50	50	50	50	–	the number of epochs
	–lr	0.0006	0.0019	0.0003	0.0007	0.0012	0.0023	$[10^{-4}, 10^{-2}]$	learning rate
	–lambda	0.0004	0.0017	0.0030	0.0062	0.0002	0.0030	$[10^{-4}, 10^{-2}]$	L2 regularization coefficient
	–kl_reg	0.0517	0.0003	0.0007	0.0014	0.0036	0.0069	$[0, 1]$	vae kl regularization
	–dropout	0.0166	0.6342	0.0111	0.8804	0.0242	0.0165	$[0, 1]$	dropout ratio
	–batch_size	256	256	512	512	512	512	$[2^6, 2^7, 2^8, 2^9, 2^{10}]$	batch size
Remarks	The detailed explanation for the parameters of SLIM is available at https://lijiancheng0614.github.io/scikit-learn/modules/generated/sklearn.linear_model.ElasticNet.html								