

Are We Really Making Much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches

Трус Владлена, БПМИ172

Рассматриваемые вопросы

Задача: Top-n recommendation

- Воспроизводимость: реально ли воспроизвести результаты статьи?
- Прогресс: насколько новейшие алгоритмы лучше, чем базовые методы?

Возможные проблемы

- Слабые бейзлайны
- Установка слабых методов в качестве новых бейзлайнов
- Трудности при сравнении или воспроизведении результатов из разных статей

Воспроизводимые статьи

Conference	Rep. ratio	Reproducible
KDD	3/4 (75%)	[17], [23], [48]
RecSys	1/7 (14%)	[53]
SIGIR	1/3 (30%)	[10]
WWW	2/4 (50%)	[14], [24]
Total	7/18 (39%)	
<i>Non-reproducible:</i> KDD: [43], RecSys: [41], [6], [38], [44], [21], [45], SIGIR: [32], [7], WWW: [42], [11]		

Базовые методы

- TopPopular
- ItemKNN
- UserKNN
- ItemKNN-CBF
- $P3\alpha$
- $RP3\beta$

Collaborative Memory Networks (CMN)

	CiteULike-a			
	HR@5	NDCG@5	HR@10	NDCG@10
TopPopular	0.1803	0.1220	0.2783	0.1535
UserKNN	0.8213	0.7033	0.8935	0.7268
ItemKNN	0.8116	0.6939	0.8878	0.7187
$P^3\alpha$	0.8202	0.7061	0.8901	0.7289
$RP^3\beta$	0.8226	0.7114	0.8941	0.7347
CMN	0.8069	0.6666	0.8910	0.6942

	Pinterest			
	HR@5	NDCG@5	HR@10	NDCG@10
TopPopular	0.1668	0.1066	0.2745	0.1411
UserKNN	0.6886	0.4936	0.8527	0.5470
ItemKNN	0.6966	0.4994	0.8647	0.5542
$P^3\alpha$	0.6871	0.4935	0.8449	0.5450
$RP^3\beta$	0.7018	0.5041	0.8644	0.5571
CMN	0.6872	0.4883	0.8549	0.5430

	Epinions			
	HR@5	NDCG@5	HR@10	NDCG@10
TopPopular	0.5429	0.4153	0.6644	0.4547
UserKNN	0.3506	0.2983	0.3922	0.3117
ItemKNN	0.3821	0.3165	0.4372	0.3343
$P^3\alpha$	0.3510	0.2989	0.3891	0.3112
$RP^3\beta$	0.3511	0.2980	0.3892	0.3103
CMN	0.4195	0.3346	0.4953	0.3592

Neural Collaborative Filtering (NCF)

	Pinterest			
	HR@5	NDCG@5	HR@10	NDCG@10
TopPopular	0.1663	0.1065	0.2744	0.1412
UserKNN	0.7001	0.5033	0.8610	0.5557
ItemKNN	0.7100	0.5092	0.8744	0.5629
$P^3\alpha$	0.7008	0.5018	0.8667	0.5559
$RP^3\beta$	0.7105	0.5116	0.8740	0.5650
NeuMF	0.7024	0.4983	0.8719	0.5536
	Movielens 1M			
	HR@5	NDCG@5	HR@10	NDCG@10
TopPopular	0.3043	0.2062	0.4531	0.2542
UserKNN	0.4916	0.3328	0.6705	0.3908
ItemKNN	0.4829	0.3328	0.6596	0.3900
$P^3\alpha$	0.4811	0.3331	0.6464	0.3867
$RP^3\beta$	0.4922	0.3409	0.6715	0.3991
NeuMF	0.5486	0.3840	0.7120	0.4369
SLIM	0.5589	0.3961	0.7161	0.4470

Spectral Collaborative Filtering (SpectralCF)

	Cutoff 20		Cutoff 60		Cutoff 100	
	REC	MAP	REC	MAP	REC	MAP
TopPopular	0.1853	0.0576	0.3335	0.0659	0.4244	0.0696
UserKNN CF	0.2881	0.1106	0.4780	0.1238	0.5790	0.1290
ItemKNN CF	0.2819	0.1059	0.4712	0.1190	0.5737	0.1243
$P^3\alpha$	0.2853	0.1051	0.4808	0.1195	0.5760	0.1248
$RP^3\beta$	0.2910	0.1088	0.4882	0.1233	0.5884	0.1288
SpectralCF	0.1843	0.0539	0.3274	0.0618	0.4254	0.0656

Variational Autoencoders for Collaborative Filtering (Mult-VAE)

	REC@20	REC@50	NDCG@100
TopPop	0.0782	0.1643	0.1570
ItemKNN CF	0.2088	0.3386	0.3086
$P^3\alpha$	0.1977	0.3346	0.2967
$RP^3\beta$	0.2196	0.3560	0.3246
SLIM	0.2551	0.3995	0.3745
Mult-VAE	0.2626	0.4138	0.3756

Выводы

- Не всегда понятно, как делать предобработку данных, настроить гиперпараметры и тд
- Новую нейросеть сравнивают с предыдущими, но не с базовыми методами
- Не используются общеизвестные датасеты
- Метрика должна зависеть от контекста
- Не всегда высокая точность приводит к более качественным рекомендациям

Вопросы

- 1. Назовите базовые методы, с которыми сравнивались рассматриваемые работы.
- 2. Опишите методы CMN, NCF, SpectralCF, Mult-VAE. Какие вышли результаты сравнения с базовыми методами?
- 3. Какие основные выводы привели авторы статьи?