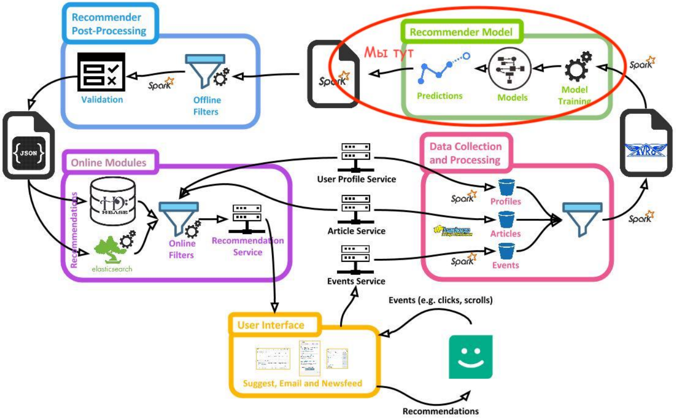
Нейросетевые рекомендеры

Николай Анохин

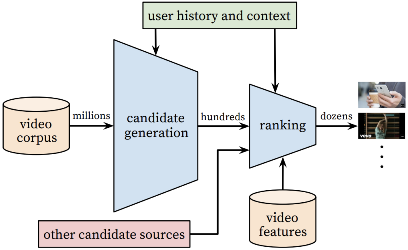
12 октября 2022 г.

Контекст

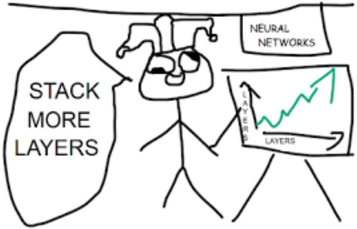

Recommender Post-Processing
Validation
Offline Filters
Recommender Model
Мы тут
Predictions
Models
Model Training
JSON
Online Modules
Recommendations
elasticsearch
Online Filters
Recommendation Service
User Profile Service
Article Service
Events Service
Data Collection and Processing
Profiles
Articles
Events
User Interface
Suggest, Email and Newsfeed
Events (e.g. clicks, scrolls)
Recommendations

От классики к нейросетям

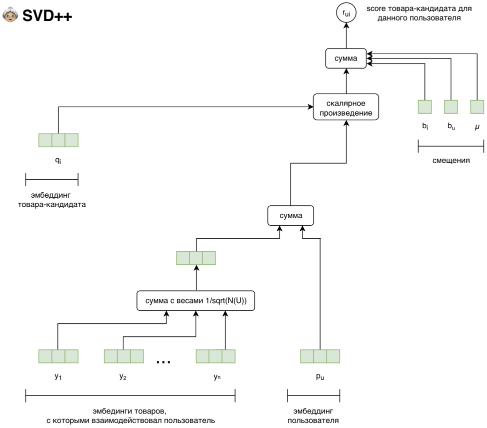
От классики к нейросетям



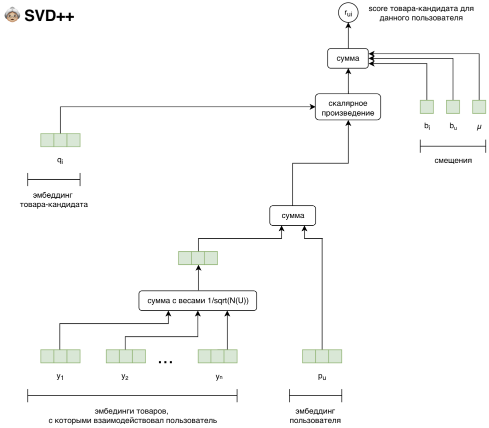
	Классика	Нейросетевые
Отбор кандидатов	MF	NN
Ранжирование	GBM	NN

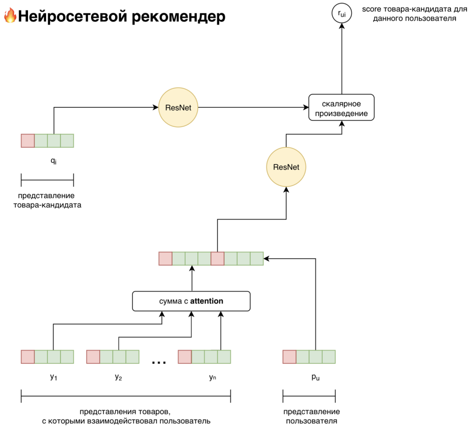


SVD++



$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T \left(p_u + \frac{1}{\sqrt{|N(u)|}} \sum_j y_j \right)$$


SVD++
r_{ui}
score товара-кандидата для данного пользователя
сумма
скалярное произведение
b_i
b_u
μ
смещения
q_i
эмбеддинг товара-кандидата
сумма
сумма с весами 1/sqrt(N(U))
y_1
y_2
...
y_n
p_u
эмбединги товаров, с которыми взаимодействовал пользователь
эмбеддинг пользователя


Нейросетевой рекомендер
r_{ui}
score товара-кандидата для данного пользователя
ResNet
скалярное произведение
q_i
представление товара-кандидата
ResNet
сумма с **attention**
y_1
y_2
...
y_n
p_u
представления товаров, с которыми взаимодействовал пользователь
представление пользователя

Истории успеха: отбор кандидатов

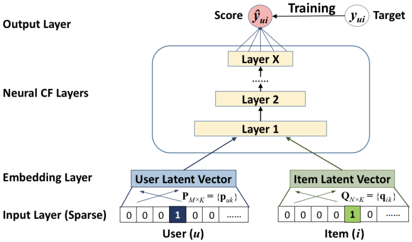
Истории успеха: отбор кандидатов

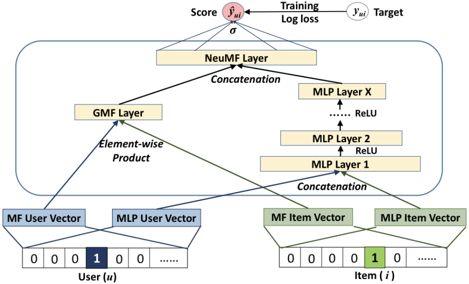
ML = модель + лосс + алгоритм оптимизации + данные

Как оставить след в науке

- Заменить скалярное произведение чем-нибудь покруче
- Заменить эмбединги чем-нибудь покруче
- Изобрести или прикрутить хитрый лосс
- Изобрести новый метод семплирования данных

Neural Collaborative Filtering [HLZ+17]

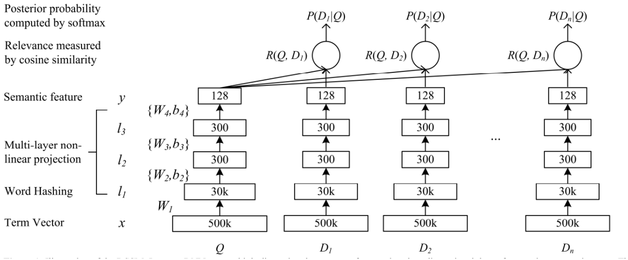




Интересность ⋆⋆⋆

Полезность ⋆⋆⋆

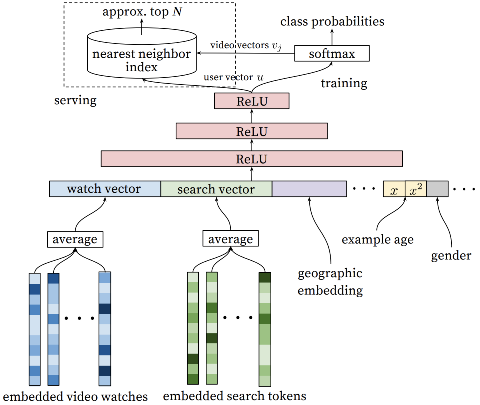
Learning Deep Structured Semantic Models for Web Search using Clickthrough Data [HHG+13]



Интересность ⋆⋆⋆

Полезность ⋆⋆⋆⋆

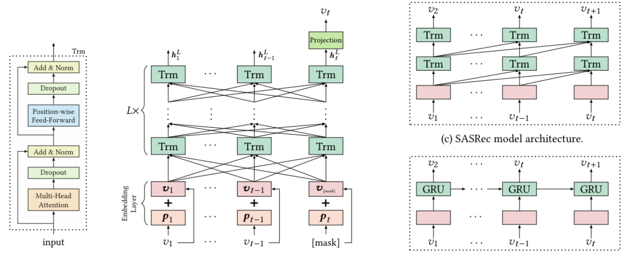
Deep Neural Networks for YouTube Recommendations [CAS16]



Интересность ⋆⋆⋆⋆⋆

Полезность ⋆⋆⋆⋆⋆

BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer [SLW+19]



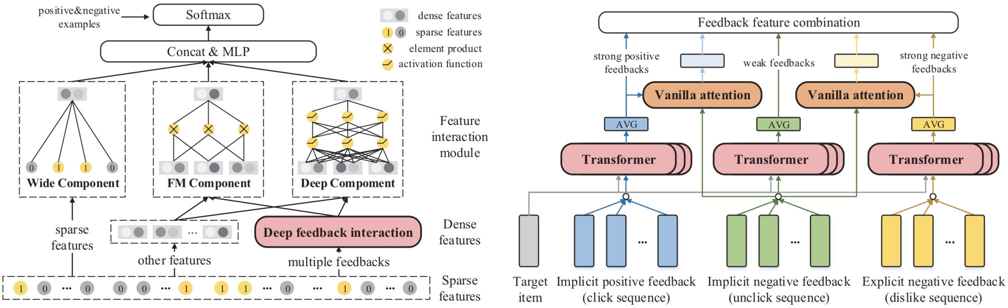
(a) Transformer Layer. (b) BERT4Rec model architecture. (c) SASRec model architecture. (d) RNN based sequential recommendation methods.

Интересность	⋆⋆
Полезность	⋆

BERT4Rec: эксперименты

Datasets	Metric	POP	BPR-MF	NCF	FPMC	GRU4Rec	GRU4Rec^{+}	Caser	SASRec	BERT4Rec	Improv.
Beauty	HR@1	0.0077	0.0415	0.0407	0.0435	0.0402	0.0551	0.0475	0.0906	**0.0953**	5.19%
	HR@5	0.0392	0.1209	0.1305	0.1387	0.1315	0.1781	0.1625	0.1934	**0.2207**	14.12%
	HR@10	0.0762	0.1992	0.2142	0.2401	0.2343	0.2654	0.2590	0.2653	**0.3025**	14.02%
	NDCG@5	0.0230	0.0814	0.0855	0.0902	0.0812	0.1172	0.1050	0.1436	**0.1599**	11.35%
	NDCG@10	0.0349	0.1064	0.1124	0.1211	0.1074	0.1453	0.1360	0.1633	**0.1862**	14.02%
	MRR	0.0437	0.1006	0.1043	0.1056	0.1023	0.1299	0.1205	0.1536	**0.1701**	10.74%
Steam	HR@1	0.0159	0.0314	0.0246	0.0358	0.0574	0.0812	0.0495	0.0885	**0.0957**	8.14%
	HR@5	0.0805	0.1177	0.1203	0.1517	0.2171	0.2391	0.1766	0.2559	**0.2710**	5.90%
	HR@10	0.1389	0.1993	0.2169	0.2551	0.3313	0.3594	0.2870	0.3783	**0.4013**	6.08%
	NDCG@5	0.0477	0.0744	0.0717	0.0945	0.1370	0.1613	0.1131	0.1727	**0.1842**	6.66%
	NDCG@10	0.0665	0.1005	0.1026	0.1283	0.1802	0.2053	0.1484	0.2147	**0.2261**	5.31%
	MRR	0.0669	0.0942	0.0932	0.1139	0.1420	0.1757	0.1305	0.1874	**0.1949**	4.00%
ML-1m	HR@1	0.0141	0.0914	0.0397	0.1386	0.1583	0.2092	0.2194	0.2351	**0.2863**	21.78%
	HR@5	0.0715	0.2866	0.1932	0.4297	0.4673	0.5103	0.5353	0.5434	**0.5876**	8.13%
	HR@10	0.1358	0.4301	0.3477	0.5946	0.6207	0.6351	0.6692	0.6629	**0.6970**	4.15%
	NDCG@5	0.0416	0.1903	0.1146	0.2885	0.3196	0.3705	0.3832	0.3980	**0.4454**	11.91%
	NDCG@10	0.0621	0.2365	0.1640	0.3439	0.3627	0.4064	0.4268	0.4368	**0.4818**	10.32%
	MRR	0.0627	0.2009	0.1358	0.2891	0.3041	0.3462	0.3648	0.3790	**0.4254**	12.24%
ML-20m	HR@1	0.0221	0.0553	0.0231	0.1079	0.1459	0.2021	0.1232	0.2544	**0.3440**	35.22%
	HR@5	0.0805	0.2128	0.1358	0.3601	0.4657	0.5118	0.3804	0.5727	**0.6323**	10.41%
	HR@10	0.1378	0.3538	0.2922	0.5201	0.5844	0.6524	0.5427	0.7136	**0.7473**	4.72%
	NDCG@5	0.0511	0.1332	0.0771	0.2239	0.3090	0.3630	0.2538	0.4208	**0.4967**	18.04%
	NDCG@10	0.0695	0.1786	0.1271	0.2895	0.3637	0.4087	0.3062	0.4665	**0.5340**	14.47%
	MRR	0.0709	0.1503	0.1072	0.2273	0.2967	0.3476	0.2529	0.4026	**0.4785**	18.85%

Deep Feedback Network for Recommendation [XLW+20]



Интересность ⋆⋆⋆

Полезность ⋆

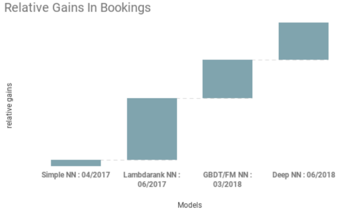
Истории успеха: ранжирование

Истории успеха: ранжирование

Как оставить след в науке

- Победить xgboost
- Пофиксить смещения

Applying Deep Learning To Airbnb Search [HAR+19]



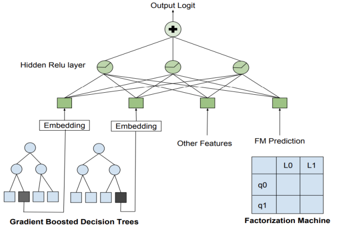


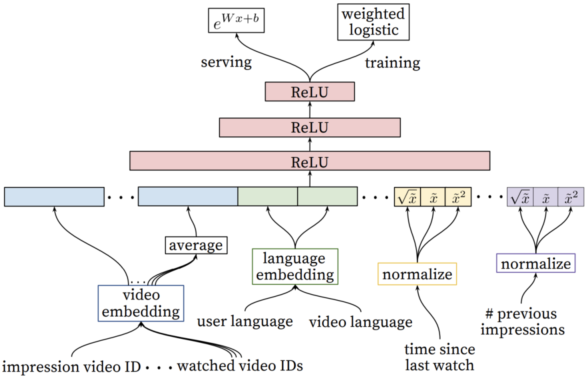
Figure 3: NN with GBDT tree nodes and FM prediction as features

> ...we were able to deprecate all that complexity by simply scaling the training data 10x and moving to a DNN with 2 hidden layers...

Интересность ⋆⋆⋆⋆

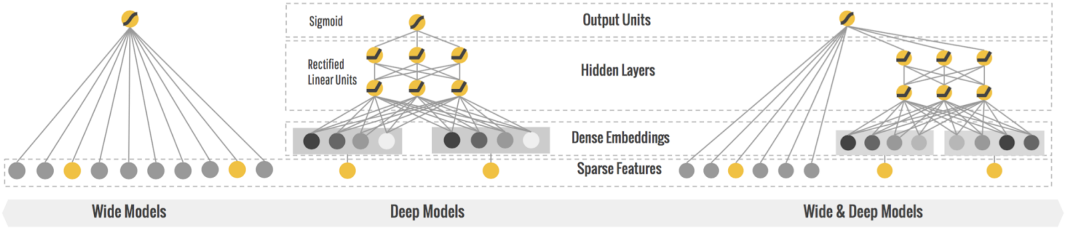
Полезность ⋆⋆⋆⋆⋆

YouTube: ранжирование



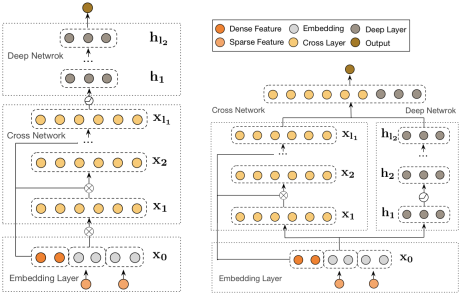
Интересность ⋆⋆⋆⋆⋆
Полезность ⋆⋆⋆⋆⋆

Wide & Deep Learning: Better Together with TensorFlow [Che16]



Интересность ⋆⋆
Полезность ⋆ ⋆ ⋆

DCN V2: Improved Deep & Cross Network and Practical Lessons for Web-scale Learning to Rank Systems [WSC+21]



(a) Stacked (b) Parallel

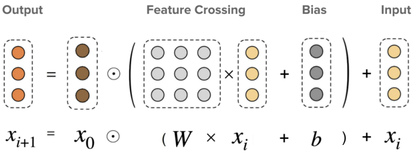
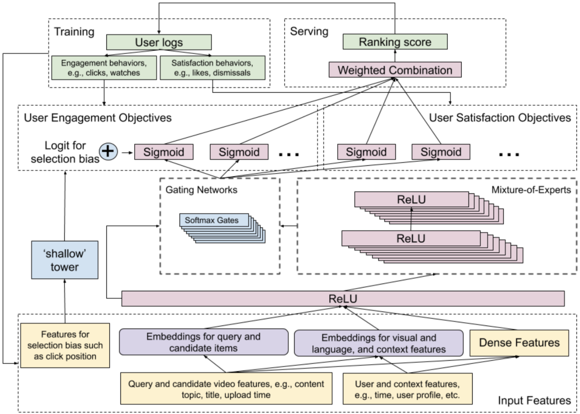


Figure 2: Visualization of a cross layer.

Интересность ⋆⋆⋆

Полезность ⋆⋆⋆⋆

Recommending What Video to Watch Next: A Multitask Ranking System [ZHW+19]



Интересность ⋆⋆⋆⋆⋆

Полезность ⋆⋆⋆⋆

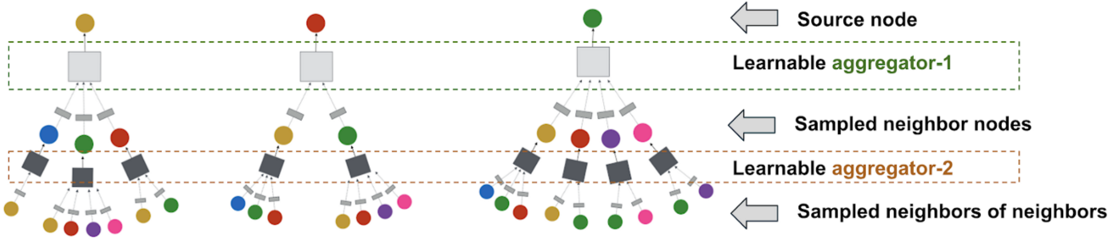
Истории успеха: контент

Истории успеха: контент

Как оставить след в науке

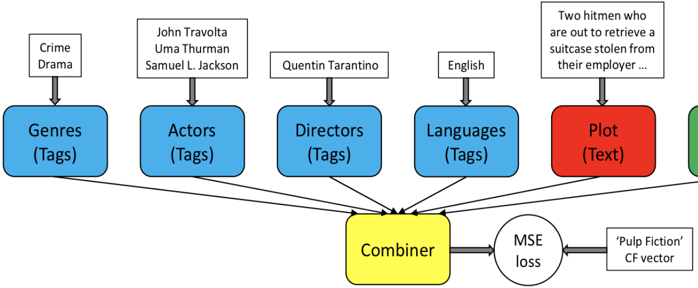
- Решить проблему холодного старта, хитро обучив эмбединги

PinSage: A new graph convolutional neural network for web-scale recommender systems [YHC+18]



Интересность ⋆⋆⋆⋆⋆
Полезность ⋆⋆⋆

CB2CF: A Neural Multiview Content-to-Collaborative Filtering Model for Completely Cold Item Recommendations [BKYK19]



Интересность ⋆⋆⋆⋆

Полезность ⋆⋆⋆⋆⋆

Проблемы нейрорекомендеров

Проблема воспроизводимости [DCJ19]

Многие результаты из статей невозможно воспроизвести

Некоторые новые алгоритмы работают хуже, чем затюненные бейзлайны

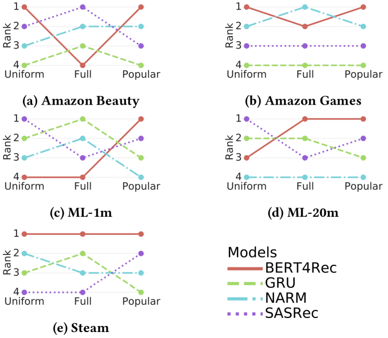
The CMN method was presented at SIGIR 18 and combines memory networks and neural attention mechanisms with latent factor and neighborhood models

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

Проблема сравнений [DZH21]

Результат сравнения может поменяться на обратный в зависимости от того, по какой метрике сравниваютъ



Итоги

Итоги

Нейросетевые модели могут заменить любой компонент рекомендательной системы: отборщик кандидатов, ранкер, item2item.

Нейросети помогают добавить inductive bias в рекомендательные модели.

Нейросетевой подход не гарантирует выигрыша – к выбору модели нужно подходить прагматично.


UNTIL WE MEET AGAIN.

Литература I

- Oren Barkan, Noam Koenigstein, Eylon Yogev, and Ori Katz, *Cb2cf: A neural multiview content-to-collaborative filtering model for completely cold item recommendations*, Proceedings of the 13th ACM Conference on Recommender Systems (New York, NY, USA), RecSys '19, Association for Computing Machinery, 2019, p. 228–236.
- Paul Covington, Jay Adams, and Emre Sargin, *Deep neural networks for youtube recommendations*, Proceedings of the 10th ACM Conference on Recommender Systems (New York, NY, USA), RecSys '16, Association for Computing Machinery, 2016, p. 191–198.
- Heng-Tze Cheng, *Wide and deep learning: Better together with tensorflow*, Jun 2016.

Литература II

Maurizio Ferrari Dacrema, Paolo Cremonesi, and Dietmar Jannach, *Are we really making much progress? a worrying analysis of recent neural recommendation approaches*, Proceedings of the 13th ACM Conference on Recommender Systems (New York, NY, USA), RecSys '19, Association for Computing Machinery, 2019, p. 101–109.

Alexander Dallmann, Daniel Zoller, and Andreas Hotho, *A case study on sampling strategies for evaluating neural sequential item recommendation models*, Fifteenth ACM Conference on Recommender Systems (New York, NY, USA), RecSys '21, Association for Computing Machinery, 2021, p. 505–514.

Литература III

- Malay Haldar, Mustafa Abdool, Prashant Ramanathan, Tao Xu, Shulin Yang, Huizhong Duan, Qing Zhang, Nick Barrow-Williams, Bradley C. Turnbull, Brendan M. Collins, and Thomas Legrand, *Applying deep learning to airbnb search*, Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (New York, NY, USA), KDD '19, Association for Computing Machinery, 2019, p. 1927–1935.
- Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry Heck, *Learning deep structured semantic models for web search using clickthrough data*, Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (New York, NY, USA), CIKM '13, Association for Computing Machinery, 2013, p. 2333–2338.

Литература IV

- Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua, *Neural collaborative filtering*, Proceedings of the 26th International Conference on World Wide Web (Republic and Canton of Geneva, CHE), WWW '17, International World Wide Web Conferences Steering Committee, 2017, p. 173–182.
- Xia Ning and George Karypis, *Slim: Sparse linear methods for top-n recommender systems.*, ICDM (Diane J. Cook, Jian Pei, Wei Wang, Osmar R. Zaïane, and Xindong Wu, eds.), IEEE Computer Society, 2011, pp. 497–506.
- Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang, *Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer*, Proceedings of the 28th ACM International Conference on Information and Knowledge Management (New York, NY, USA), CIKM '19, ACM, 2019, pp. 1441–1450.

Литература V

- Harald Steck, *Embarrassingly shallow autoencoders for sparse data*, The World Wide Web Conference (New York, NY, USA), WWW '19, Association for Computing Machinery, 2019, p. 3251–3257.
- Ruoxi Wang, Rakesh Shivanna, Derek Cheng, Sagar Jain, Dong Lin, Lichan Hong, and Ed Chi, *Dcn v2: Improved deep and cross network and practical lessons for web-scale learning to rank systems*, Proceedings of the Web Conference 2021 (New York, NY, USA), WWW '21, Association for Computing Machinery, 2021, p. 1785–1797.
- Ruobing Xie, Chen Ling, Yalong Wang, Rui Wang, Feng Xia, and Leyu Lin, *Deep feedback network for recommendation*, IJCAI, 2020.

Литература VI

- Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L. Hamilton, and Jure Leskovec, *Graph convolutional neural networks for web-scale recommender systems*, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (New York, NY, USA), KDD '18, Association for Computing Machinery, 2018, p. 974–983.
- Zhe Zhao, Lichan Hong, Li Wei, Jilin Chen, Aniruddh Nath, Shawn Andrews, Aditee Kumthekar, Maheswaran Sathiamoorthy, Xinyang Yi, and Ed Chi, *Recommending what video to watch next: A multitask ranking system*, Proceedings of the 13th ACM Conference on Recommender Systems (New York, NY, USA), RecSys '19, Association for Computing Machinery, 2019, p. 43–51.