

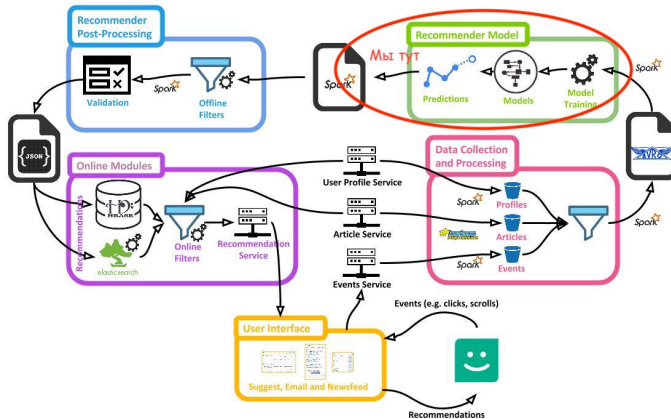
Нейросетевые рекоммендеры 2: ранкеры и контент

Николай Анохин

18 марта 2024 г.



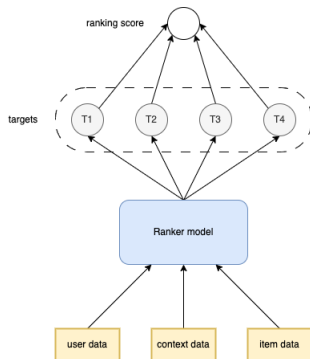
Контекст



Истории успеха: ранжирование



Постановка задачи ранжирования



Задача

Выбрать порядок кандидатов так, чтобы самые релевантные стояли в голове списка.



Истории успеха: ранжирование

Как оставить след в науке

- Победить xgboost
- Пофиксить смещения



Applying Deep Learning To Airbnb Search [HAR⁺19]

Relative Gains In Bookings

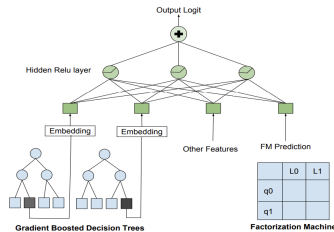
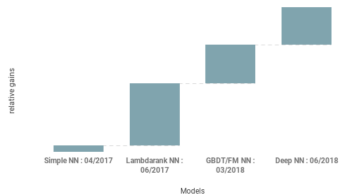


Figure 3: NN with GBDT tree nodes and FM prediction as features

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

Идея

Специфика применения NN ранкера на практике

Интересность

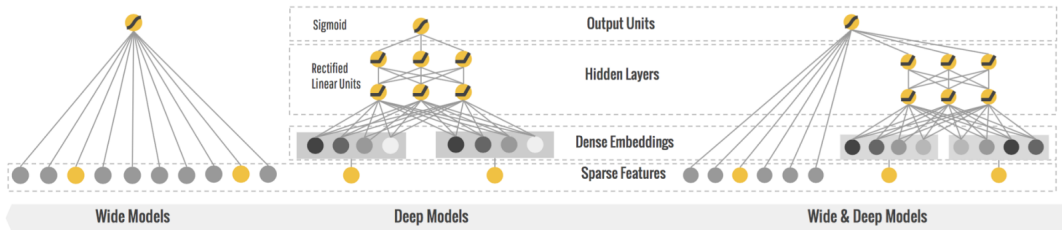
★★★★

Полезность

★★★★★



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



Идея

Скомбинировать “ручные” перемножения признаков с нелинейностью NN

Интересность

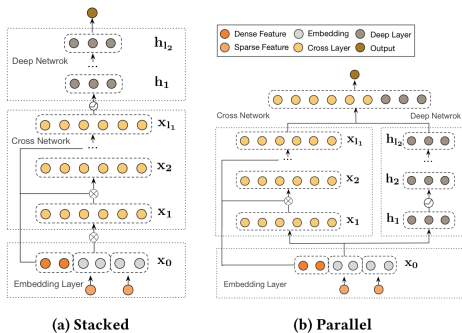
★★

Полезность

★★★



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



$$\text{Output} = \text{Feature Crossing} + \text{Bias} + \text{Input}$$

$$x_{i+1} = x_0 \odot (W \times x_i + b) + x_i$$

Figure 2: Visualization of a cross layer.

Идея

Интегрировать перемножения признаков в архитектуру NN

Интересность

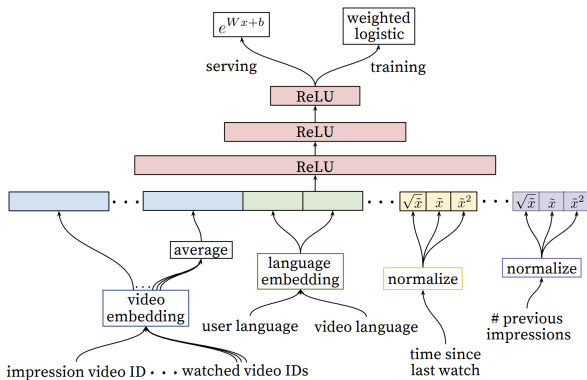
★ ★ ★

Полезность

★ ★ ★ ★



YouTube: ранжирование



Идея

Использовать последовательность айтемов как признак

Интересность

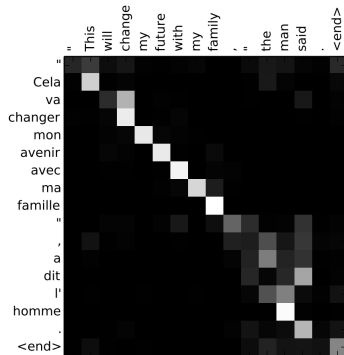
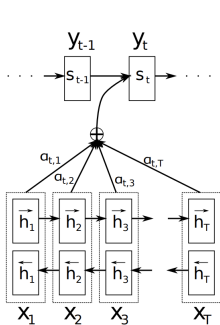
★★★★★

Полезность

★★★★★

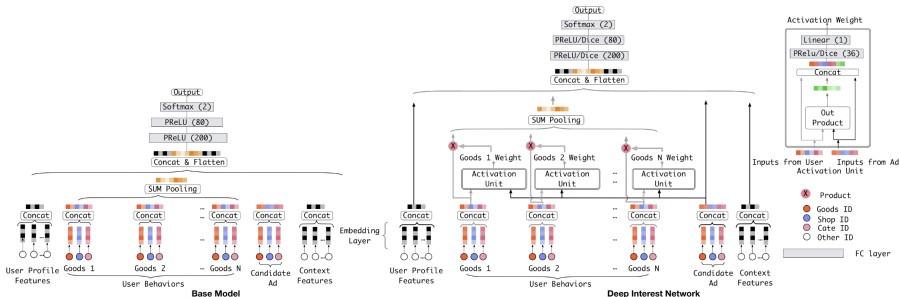


Механизм attention [BCB16]



$$\alpha_{ij} = \frac{\exp e_{ij}}{\sum_k \exp e_{ik}}, \quad e_{ij} = \text{ffn}(s_{i-1}, h_j)$$

Deep Interest Network for Click-Through Rate Prediction [ZSZ⁺18]



Идея
Интересность
Полезность

Attention (почти) для агрегации последовательности айтемов
★ ★ ★ ★
★ ★ ★ ★



MaskNet: Introducing Feature-Wise Multiplication to CTR Ranking Models by Instance-Guided Mask [WSZ21]

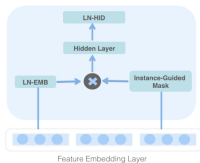


Figure 2: MaskBlock on Feature Embedding

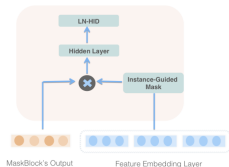


Figure 3: MaskBlock on MaskBlock

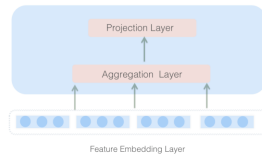


Figure 1: Neural Structure of Instance-Guided Mask



MaskNet: Introducing Feature-Wise Multiplication to CTR Ranking Models by Instance-Guided Mask [WSZ21]

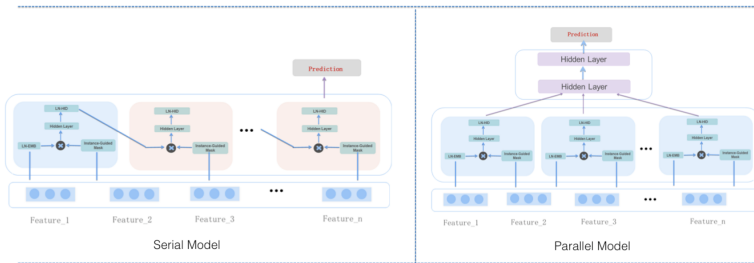


Figure 4: Structure of Serial Model and Parallel Model

Идея

Обучаемая маска выбирает “полезные” части выходов слоя

Интересность

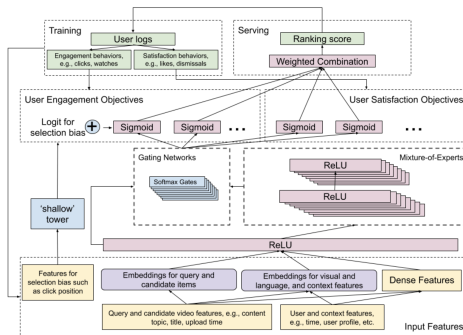
☆☆☆

Полезность

☆☆☆



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



Идея

Интересность

Полезность

Attention выбирает экспертов под разные таргеты

★★★★★

★★★★



Истории успеха: ранжирование
oooooooooooo

Истории успеха: контент
●ooo

Что работает на практике
oooo

Проблемы нейрорекомендеров
ooo

Итоги
oooo

Истории успеха: контент



Истории успеха: контент

Как оставить след в науке

- Решить проблему холодного старта, хитро обучив эмбединги



Learning Transferable Visual Models From Natural Language Supervision [RKH⁺21]

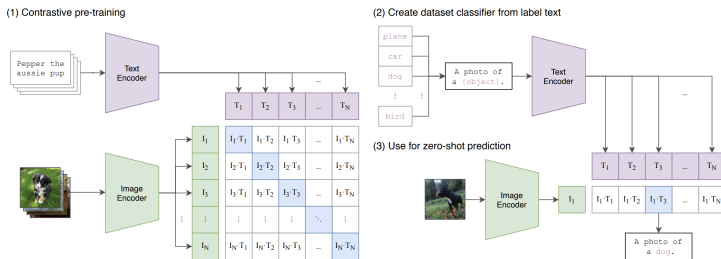


Figure 1. Summary of our approach. While standard image models jointly train an image feature extractor and a linear classifier to predict some label, CLIP jointly trains an image encoder and a text encoder to predict the correct pairings of a batch of (image, text) training examples. At test time the learned text encoder synthesizes a zero-shot linear classifier by embedding the names or descriptions of the target dataset's classes.

Идея

Мульти-модальные эмбединги решают вопросы

Интересность

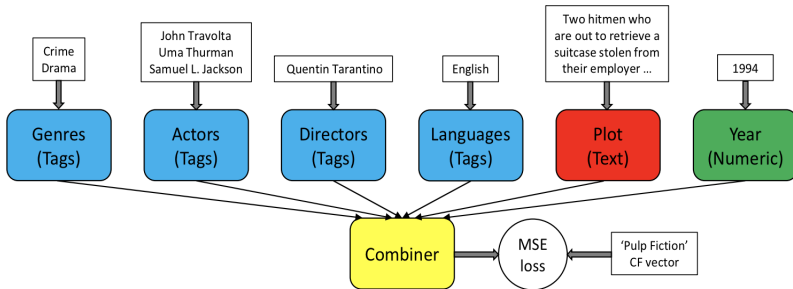
★ ★ ★

Полезность

★ ★ ★



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



Идея

Интересность

Полезность

Используем результат коллаборативной модели для обучения контентной

★ ★ ★ ★

★ ★ ★ ★ ★



Истории успеха: ранжирование
oooooooooooo

Истории успеха: контент
oooo

Что работает на практике
●ooo

Проблемы нейрорекомендеров
ooo

Итоги
oooo

Что работает на практике



Архитектуры [Wan23]

- Мульти-таргет решает.
- Авто пересечение признаков (DCN, MaskNet).
- МоЕ не дает профита при том же количестве параметров, как у базовых архитектур.
- DIN дает эффект для длинных последовательностей.



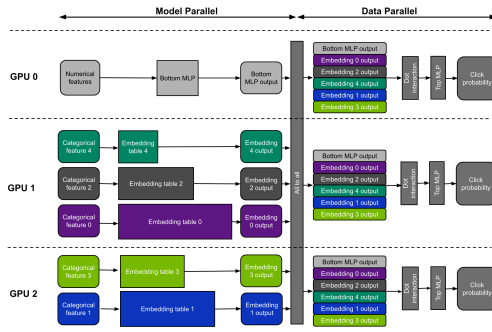
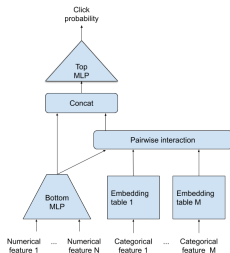
Improving Deep Learning For Airbnb Search [HAR⁺20]

If deeper nets were not the right architecture for us, we hypothesized, more specialized architectures might be. So we tried architectures [...] like **deep and wide** [...] followed by variants of **attention based networks**. [...] The short summary of those efforts is that they **failed to move the needle**.

Идея	Архитектура должна следовать из проблемы пользователя
Интересность	★★★★★
Полезность	★★★★★



Deep Learning Recommendation Model for Personalization and Recommendation Systems [NMS⁺19]



Идея

Простая архитектура, но много параметров в эмбедингах¹

Интересность

★★★

Полезность

★★★★★

¹https://catalog.ngc.nvidia.com/orgs/nvidia/resources/dlrm_for_pytorch



Истории успеха: ранжирование
oooooooooooo

Истории успеха: контент
oooo

Что работает на практике
oooo

Проблемы нейрорекомендеров
●oo

Итоги
oooo

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

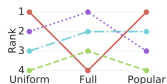
	Pinterest			
	HR@5	NDCG@5	HR@10	NDCG@10
TopPopular	0.1668	0.1066	0.2745	0.1411
UserKNN	0.6886	0.4936	0.8527	0.5470
ItemKNN	0.6966	0.4994	0.8647	0.5542
P ³ _α	0.6871	0.4935	0.8449	0.5450
RP ³ _β	0.7018	0.5041	0.8644	0.5571
CMN	0.6872	0.4883	0.8549	0.5430

	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 _α	0.3510	0.2989	0.3891	0.3112
RP ³ _β	0.3511	0.2980	0.3892	0.3103
CMN	0.4195	0.3346	0.4953	0.3592

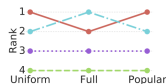


Проблема сравнений [DZH21]

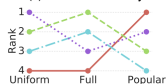
Результат сравнения может поменяться на обратный в зависимости от того, по какой метрике сравнивают



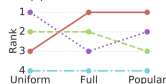
(a) Amazon Beauty



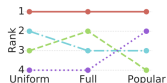
(b) Amazon Games



(c) ML-1m



(d) ML-20m



(e) Steam

Models

— BERT4Rec

- - GRU

- · - NARM

· · · SASRec



Истории успеха: ранжирование
○○○○○○○○○○○○

Истории успеха: контент
○○○○

Что работает на практике
○○○○

Проблемы нейрорекомендеров
○○○

Итоги
●○○○

Итоги



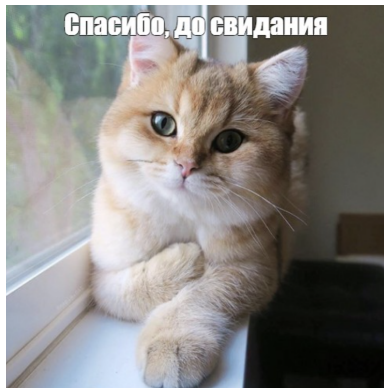
Итоги

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

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






Феерически завершающий лекцию слайд



<https://t.me/mlvok>





Литература I

-  Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio, *Neural machine translation by jointly learning to align and translate*, 2016.
-  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.
-  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.
-  Malay Haldar, Mustafa Abdool, Prashant Ramanathan, Tyler Sax, Lanbo Zhang, Aamir Mansawala, Shulin Yang, Bradley Turnbull, and Junshuo Liao, *Improving deep learning for airbnb search*, 2020.



Литература IV

-  Maxim Naumov, Dheevatsa Mudigere, Hao-Jun Michael Shi, Jianyu Huang, Narayanan Sundaraman, Jongsoo Park, Xiaodong Wang, Udit Gupta, Carole-Jean Wu, Alisson G. Azzolini, Dmytro Dzhulgakov, Andrey Mallevich, Ilia Cherniavskii, Yinghai Lu, Raghuraman Krishnamoorthi, Ansha Yu, Volodymyr Kondratenko, Stephanie Pereira, Xianjie Chen, Wenlin Chen, Vijay Rao, Bill Jia, Liang Xiong, and Misha Smelyanskiy, *Deep learning recommendation model for personalization and recommendation systems*, 2019.
-  Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever, *Learning transferable visual models from natural language supervision*, 2021.
-  Shusen Wang, *Methodologies for improving modern industrial recommender systems*, 2023.




Литература V

-  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.
-  Zhiqiang Wang, Qingyun She, and Junlin Zhang, *Masknet: Introducing feature-wise multiplication to ctr ranking models by instance-guided mask*, 2021.
-  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.



Литература VI

 Guorui Zhou, Chengru Song, Xiaoqiang Zhu, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai, *Deep interest network for click-through rate prediction*, 2018.

