

Recommender Systems

lecture 6: Deep Learning recommenders

Alexey Grishanov

Moscow Institute of Physics and Technology

Spring 2023

Recap: taxonomy

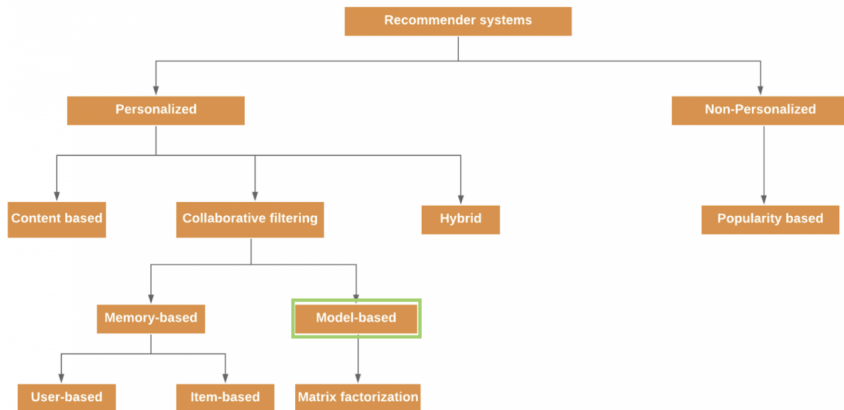


image source

Plan for today

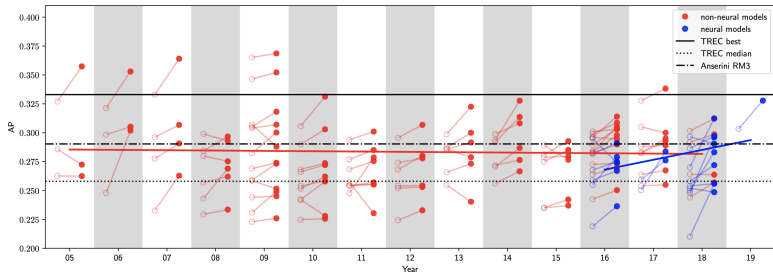
- 1 NeuroMF
- 2 MultVAE
- 3 Sequential models

«Neural hype», 2019

M. Dacrema et al. Are We Really Making Much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches (RecSys '19) - best paper

- considered 18 algorithms from top conferences in the last years
- only 7 of them could be reproduced with reasonable effort
- 6 can often be outperformed with comparably simple heuristic methods (nearest-neighbor or graph-based techniques)

W. Yang et al. Critically Examining the "Neural Hype": Weak Baselines and the Additivity of Effectiveness Gains from Neural Ranking Models. (SIGIR'19)



Neural Collaborative Filtering (NCF)

Motivation: generalization of MF (dot product)

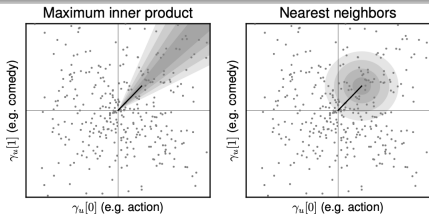
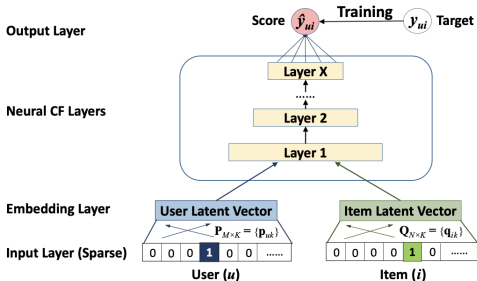
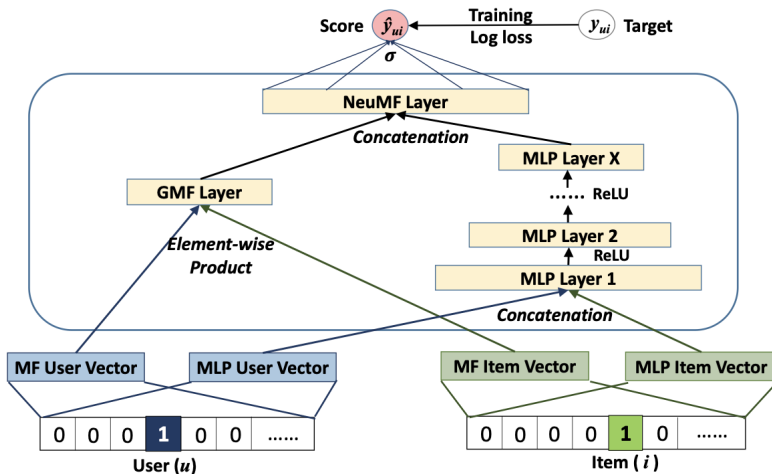


image source

Architecture:





X. He et al. (WWW 2017). Neural Collaborative Filtering

NeuroMF vs Matrix Factorization

Method	Movielens		Pinterest		Result from
	HR@10	NDCG@10	HR@10	NDCG@10	
Popularity	0.4535	0.2543	0.2740	0.1409	[8]
SLIM [25, 30]	<u>0.7162</u>	<u>0.4468</u>	0.8679	<u>0.5601</u>	[8]
iALS [20]	0.7111	0.4383	0.8762	0.5590	[8]
NeuMF (MLP+GMF) [17]	0.7093	0.4349	<u>0.8777</u>	0.5576	[8]
Matrix Factorization	0.7294	0.4523	0.8895	0.5794	Fig. 2

Learning dot product with MLP is hard:

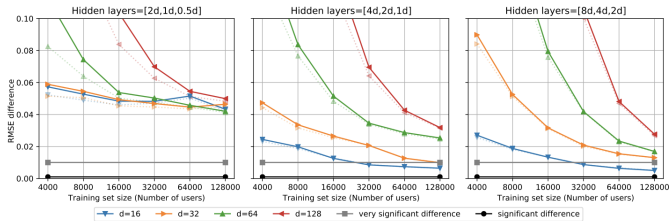


Figure 3: How well a MLP can learn a dot product over embeddings of dimension d . The ground truth is generated from a dot product of Gaussian embeddings plus Gaussian label noise. The graphs show the difference between the RMSE of the dot product and the RMSE of the learned similarity measure; the solid line measures the difference on the fresh set, the dotted on the test set. Noise and scale have been chosen such that 0.01 could indicate a very significant difference and 0.001 a significant difference.

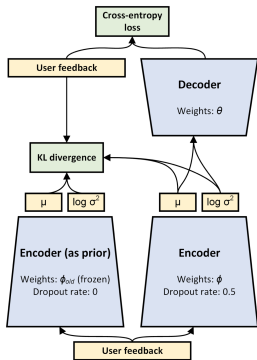
Steffen Rendle, Walid Krichene, Li Zhang, and John Anderson. (RecSys '20). Neural Collaborative Filtering vs. Matrix Factorization Revisited

Pros

- Easy refitting on new interactions
- Used in production by Bytedance (tiktok owner)
Z. Liu et al. (RecSys '22). Monolith: Real Time Recommendation System With Collisionless Embedding Table

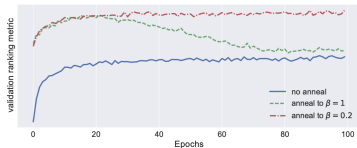
Cons

- Has issues with outperforming matrix factorization
- Overall quality far from SOTA



- model user history with multinomial distribution

$$\mathcal{L}_{\beta}(\mathbf{x}_u; \theta, \phi) \equiv \mathbb{E}_{q_{\phi}(\mathbf{z}_u | \mathbf{x}_u)} [\log p_{\theta}(\mathbf{x}_u | \mathbf{z}_u)] - \beta \cdot \text{KL}(q_{\phi}(\mathbf{z}_u | \mathbf{x}_u) \| p(\mathbf{z}_u)).$$



- 1 novel composite prior distribution
- 2 new approach to setting the hyperparameter
 $\beta = \beta(\mathbf{x}_u)$
- 3 new architecture (deep encoder, linear decoder)
- 4 novel training:
alternating updates for the encoder and decoder

	Recall@20	Recall@50	NDCG@100
MovieLens-20M Dataset			
WARP [38]	0.314	0.466	0.341
LambdaNet[6]	0.395	0.534	0.427
WMF [14]	0.360	0.498	0.386
SLIM [25]	0.370	0.495	0.401
CDAE [39]	0.391	0.523	0.418
Mult-DAE [22]	0.387	0.524	0.419
Mult-VAE [22]	0.395	0.537	0.426
RaCT [23]	<u>0.403</u>	<u>0.543</u>	<u>0.434</u>
EASE [33]	0.391	0.521	0.420
RecVAE (ours)	0.414 ± 0.0027	0.553 ± 0.0028	0.442 ± 0.0021
Netflix Prize Dataset			
WARP [38]	0.270	0.365	0.306
LambdaNet[6]	0.352	0.441	0.386
WMF [14]	0.316	0.404	0.351
SLIM [25]	0.347	0.428	0.379
CDAE [39]	0.343	0.428	0.376
Mult-DAE [22]	0.344	0.438	0.380
Mult-VAE [22]	0.351	0.444	0.386
RaCT [23]	0.357	<u>0.450</u>	0.392
EASE [33]	0.362	0.445	<u>0.393</u>
RecVAE (ours)	<u>0.361</u> ± 0.0013	0.452 ± 0.0013	0.394 ± 0.0010

Pros

- Successfully reproduced in «*Are we really making much progress*»
- Closer to SOTA
- Predictions are from user's click history. No need to obtain latent factor for user.

Cons

- Handling new items

Sequential recommendation

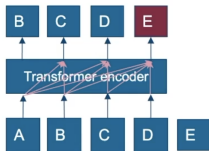


Task: predict next interaction in a set of user's interaction

Transformers for sequential recommendation

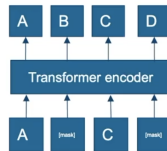
SASRec (Kang et al.):

- Predicts Input Sequence, shifted by 1 element
- Causal Self-Attention
- Use last prediction for recommendation



BERT4Rec (Sun et al.):

- Based on BERT language model
- Recover corrupted input for training
- Add [Mask] token to the end of the sequence for prediction



Sequential models comparison

Bert4Rec paper (CIKM'19):

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	<u>0.0906</u>	0.0953	5.19%
	HR@5	0.0392	0.1209	0.1305	0.1387	0.1315	0.1781	0.1625	<u>0.1934</u>	0.2207	14.12%
	HR@10	0.0762	0.1992	0.2142	0.2401	0.2343	0.2654	0.2590	<u>0.2653</u>	0.3025	14.02%
	NDCG@5	0.0230	0.0814	0.0855	0.0902	0.0812	0.1172	0.1050	<u>0.1436</u>	0.1599	11.35%
	NDCG@10	0.0349	0.1064	0.1124	0.1211	0.1074	0.1453	0.1360	<u>0.1633</u>	0.1862	14.02%
	MRR	0.0437	0.1006	0.1043	0.1056	0.1023	0.1299	0.1205	<u>0.1536</u>	0.1701	10.74%
Steam	HR@1	0.0159	0.0314	0.0246	0.0358	0.0574	0.0812	0.0495	<u>0.0885</u>	0.0957	8.14%
	HR@5	0.0805	0.1177	0.1203	0.1517	0.2171	0.2391	0.1766	<u>0.2559</u>	0.2710	5.90%
	HR@10	0.1389	0.1993	0.2169	0.2551	0.3313	0.3594	0.2870	<u>0.3783</u>	0.4013	6.08%
	NDCG@5	0.0477	0.0744	0.0717	0.0945	0.1370	0.1613	0.1131	<u>0.1727</u>	0.1842	6.66%
	NDCG@10	0.0665	0.1005	0.1026	0.1283	0.1802	0.2053	0.1484	<u>0.2147</u>	0.2261	5.31%
	MRR	0.0669	0.0942	0.0932	0.1139	0.1420	0.1757	0.1305	<u>0.1874</u>	0.1949	4.00%



A.Petrov. A Systematic Review and Replicability Study of BERT4Rec for Sequential Recommendation (RecSys'22):

Dataset	Number of papers	BERT4Rec wins	SASRec wins	Ties
Amazon Beauty*	19	12 (63%)	5 (26%)	2 (11%)
ML-1M*	18	13 (72%)	3 (17%)	2 (11%)
Yelp	10	6 (60%)	4 (40%)	0 (0%)
Steam*	8	7 (88%)	1 (12%)	0 (0%)
ML-20M*	8	7 (88%)	0 (0%)	1 (12%)
Total (including all other datasets)	134	86 (64%)	32 (23%)	16 (12%)



I want you to act as a song recommender. I will provide you with a song and you will create a playlist of 10 songs that are similar to the given song. You should also provide a name for the playlist. Do not choose songs that are same name or artist. Do not write any explanations or other words, just reply with the playlist name, description the songs.



Sure, I can act as a song recommender for you! What is the song you would like me to base the playlist off of?  



Gruppa krovi by Viktor Tsoi



Great choice! Based on "Gruppa krovi" by Viktor Tsoi, I have created a playlist with 10 similar songs. The playlist name is "Soviet Rock Anthems."  

Here are the songs in the playlist:

1. "Kino" by Akvarium
2. "Zvezda po imeni Solntse" by Nautilus Pompilius
3. "Nenavizhu" by DDT
4. "Kukushka" by Viktor Tsoi (from the "Assa" movie soundtrack)
5. "Vozmi menya za ruku" by DDT
6. "Deti gornogo Altaya" by Mashina Vremeni
7. "Pachka sigaret" by Brigada 5
8. "Kogda my byli na voine" by Alisa
9. "Kvartira" by Mashina Vremeni
10. "Gospodi, spasí, ohrani" by Akvarium

Enjoy the playlist!



- ① *M. Dacrema, P. Cremonesi, D. Jannach.* (RecSys '19). Are We Really Making Much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches.
- ② *X. He et al.* (WWW 2017). Neural Collaborative Filtering
- ③ *D. Liang et al.* (2020). Variational Autoencoders for Collaborative Filtering WWW '18.
- ④ *W. Kang, J. McAuley* (ICDM'19). BERT4Rec: Self-Attentive Sequential Recommendation
- ⑤ *F. Sun et al.* (CIKM'19). BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer
- ⑥ *A. Petrov, C. Macdonald* (RecSys '22). A Systematic Review and Replicability Study of BERT4Rec for Sequential Recommendation