Recommender Systems for Movielens-100k

Bayesian Factorization Machine vs Neural Collaborative Filtering

24.03.2025

Nam H. Le

Contents

1 Introduction

Domain of application

Movie recommendations for online movie streaming platform.

Aims

Recommending movies which users are most likely to watch… mainly for user retention!

Data description [1]

Around 100,000 ratings and 3,600 tag applications applied to 9,000 movies by 600 users.

```
sparsity: 98.3000%
```

	userId	movieId	rating	timestamp
0	1	1	4.0	964982703
1	1	3	4.0	964981247
2	1	6	4.0	964982224
3	1	47	5.0	964983815
4	1	50	5.0	964982931
...	...	...	...	...
100831	610	166534	4.0	1493848402
100832	610	168248	5.0	1493850091
100833	610	168250	5.0	1494273047
100834	610	168252	5.0	1493846352
100835	610	170875	3.0	1493846415

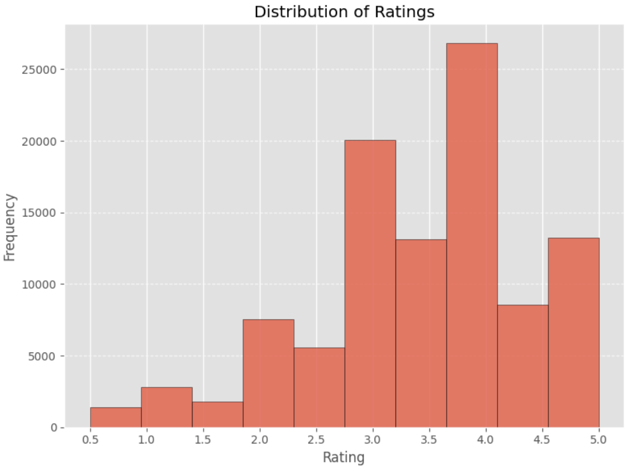
100836 rows × 4 columns

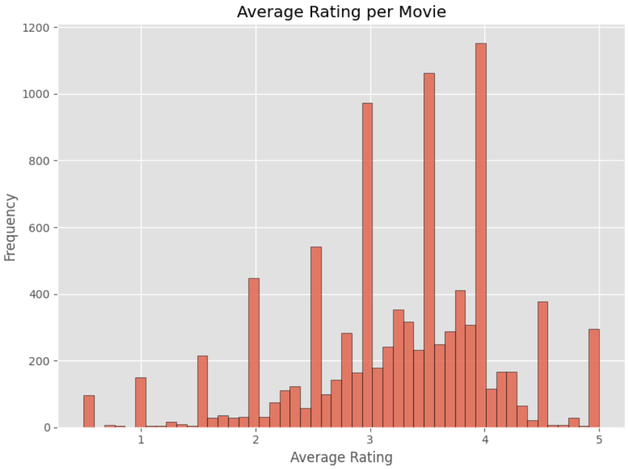
	movieId	title	genres
0	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
1	2	Jumanji (1995)	Adventure|Children|Fantasy
2	3	Grumpier Old Men (1995)	Comedy|Romance
3	4	Waiting to Exhale (1995)	Comedy|Drama|Romance
4	5	Father of the Bride Part II (1995)	Comedy
...	...	...	...
9737	193581	Black Butler: Book of the Atlantic (2017)	Action|Animation|Comedy|Fantasy
9738	193583	No Game No Life: Zero (2017)	Animation|Comedy|Fantasy
9739	193585	Flint (2017)	Drama
9740	193587	Bungo Stray Dogs: Dead Apple (2018)	Action|Animation
9741	193609	Andrew Dice Clay: Dice Rules (1991)	Comedy

9742 rows × 3 columns

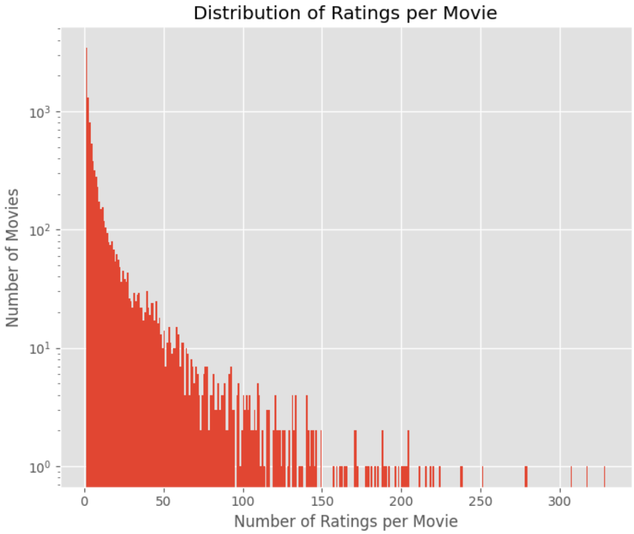
Only ratings and movies data was used. Tags was too complicated to work with.

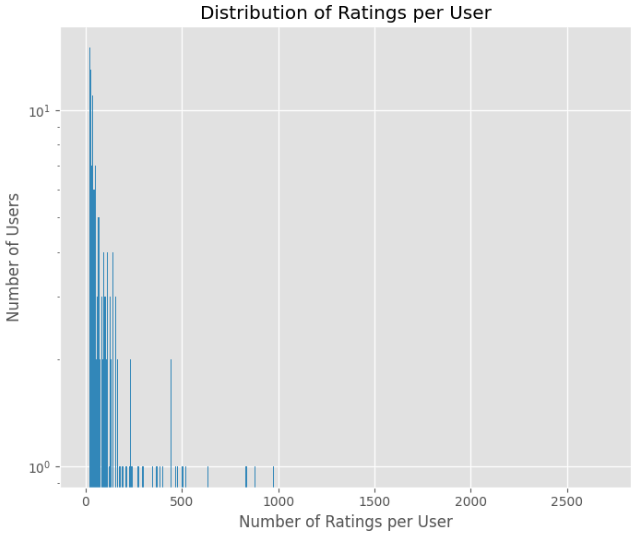
Data description [1] (ii)



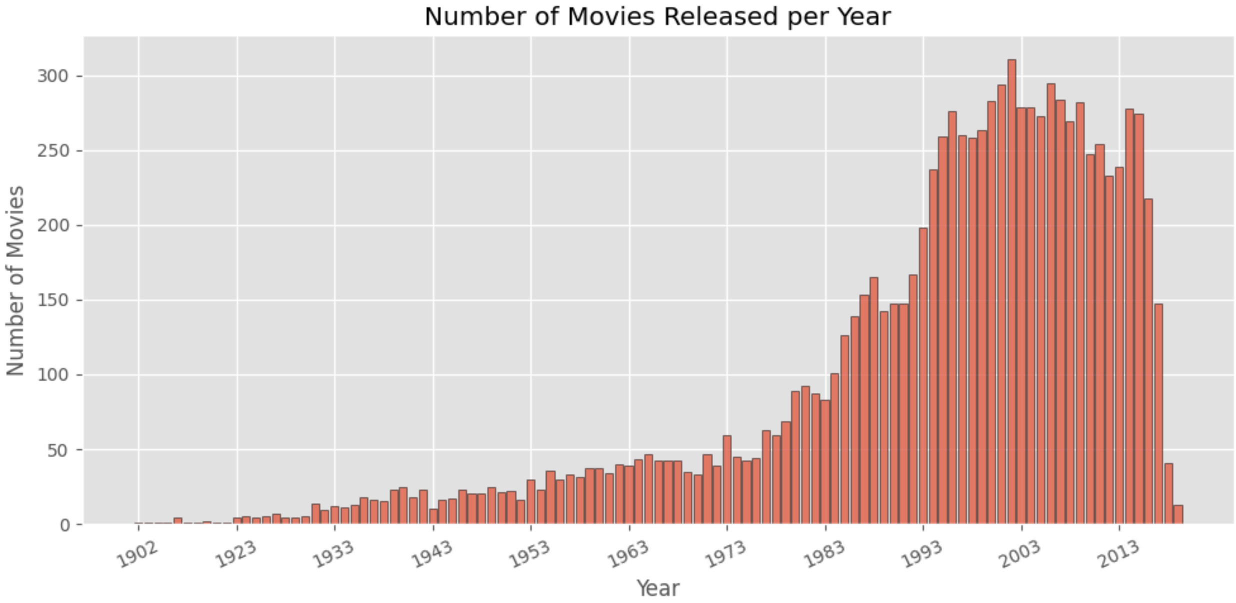


Data description [1] (iii)

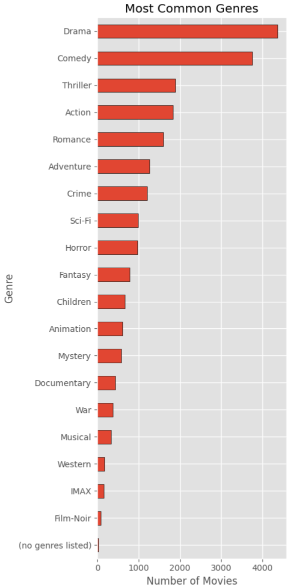


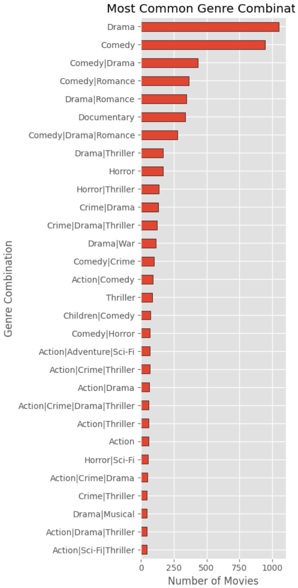


Data description [1] (iv)



Data description [1] (v)





Data preperation

Splitting dates from title and storing in its own column.

	movieId	title	genres	release_year
0	1	Toy Story	Adventure|Animation|Children|Comedy|Fantasy	1995
1	2	Jumanji	Adventure|Children|Fantasy	1995
2	3	Grumpier Old Men	Comedy|Romance	1995
3	4	Waiting to Exhale	Comedy|Drama|Romance	1995
4	5	Father of the Bride Part II	Comedy	1995
...	...	...	...	...
9737	193581	Black Butler: Book of the Atlantic	Action|Animation|Comedy|Fantasy	2017
9738	193583	No Game No Life: Zero	Animation|Comedy|Fantasy	2017
9739	193585	Flint	Drama	2017
9740	193587	Bungo Stray Dogs: Dead Apple	Action|Animation	2018
9741	193609	Andrew Dice Clay: Dice Rules	Comedy	1991

9742 rows × 4 columns

Data preperation (ii)

Splitting and one-hot encoded the genres.

	user	item	label	title	release_year	(no genres listed)	Action	Adventure	Animation	Children	...	Film-Noir	Horror	IMAX	Musical	Mystery	Romance	Sci-Fi	Thriller	War	Western
0	1	1	4.0	Toy Story	1995	0	0	1	1	1	...	0	0	0	0	0	0	0	0	0	0
1	1	3	4.0	Grumpier Old Men	1995	0	0	0	0	0	...	0	0	0	0	0	1	0	0	0	0
2	1	6	4.0	Heat	1995	0	1	0	0	0	...	0	0	0	0	0	0	0	1	0	0
3	1	47	5.0	Seven (a.k.a. Se7en)	1995	0	0	0	0	0	...	0	0	0	0	1	0	0	1	0	0
4	1	50	5.0	Usual Suspects, The	1995	0	0	0	0	0	...	0	0	0	0	1	0	0	1	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
100831	610	166534	4.0	Split	2017	0	0	0	0	0	...	0	1	0	0	0	0	0	1	0	0
100832	610	168248	5.0	John Wick: Chapter Two	2017	0	1	0	0	0	...	0	0	0	0	0	0	0	1	0	0
100833	610	168250	5.0	Get Out	2017	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	0
100834	610	168252	5.0	Logan	2017	0	1	0	0	0	...	0	0	0	0	0	0	1	0	0	0
100835	610	170875	3.0	The Fate of the Furious	2017	0	1	0	0	0	...	0	0	0	0	0	0	0	1	0	0

100836 rows × 25 columns

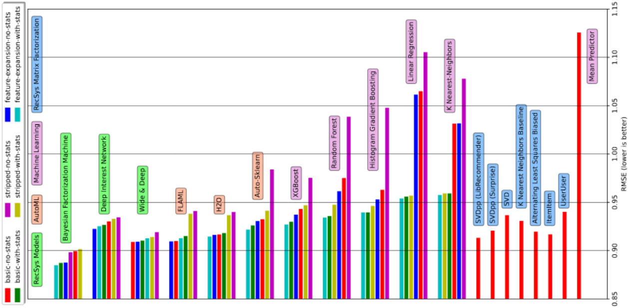
Mostly just trying to get the data in the right format for the different libraries since they all expect different formats:

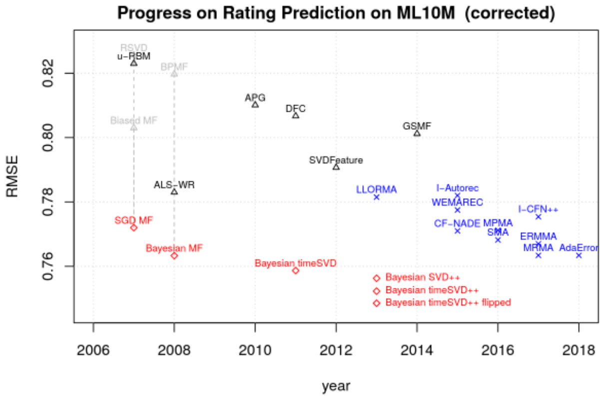
- Surprise
- MyFM
- LibRecommender

2 Methods and Justification

RS1: BFM [2]

BFM chosen as it was overlooked in literature despite it’s great performance given proper setup [8], [9].





RS1: BFM [2] (ii)

Simple version of equation

$$\hat{y} = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle v_i v_j \rangle x_i x_j$$

We have: Linear features and latent vectors. Trying to get a predicted ratings from these values.

Bayesian extension

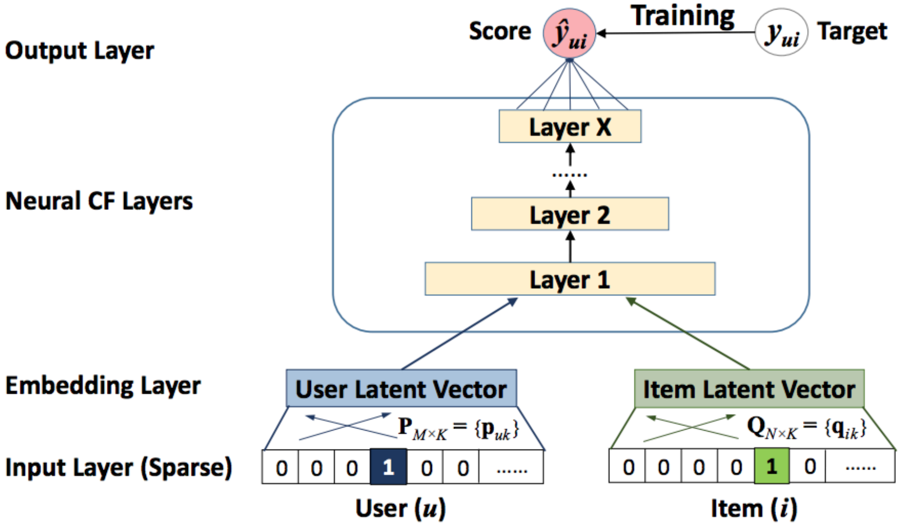
$$v_i \sim \mathcal{N}(\mu_u \Sigma_u)$$

$$v_j \sim \mathcal{N}(\mu_i \Sigma_i)$$

The latent vectors are drawn from a normal distribution. Why?

- Adds uncertainty
- Allows for confidence intervals

RS2: NCF [3]



Output Layer
Score
$\hat{y}_{ui}$
Training
y_{ui}
Target
Neural CF Layers
Layer X
......
Layer 2
Layer 1
Embedding Layer
User Latent Vector
Item Latent Vector
$\mathbf{P}_{M \times K} = \{\mathbf{p}_{uk}\}$
$\mathbf{Q}_{N \times K} = \{\mathbf{q}_{ik}\}$
Input Layer (Sparse)
0 0 0 1 0 0 ......
0 0 0 0 1 0 ......
User (u)
Item (i)

3 Evaluation

Evaluation metrics [4], [5]

Root Mean Squared Error

$$\text{RMSE} = \sqrt{\left(\frac{1}{T}\right) \sum_{u,i \in T} \left(\hat{r}_{ui} - r_{ui}\right)^2}$$

Most popular metric in literature. Shows up in almost all traditional model as THE main metric.

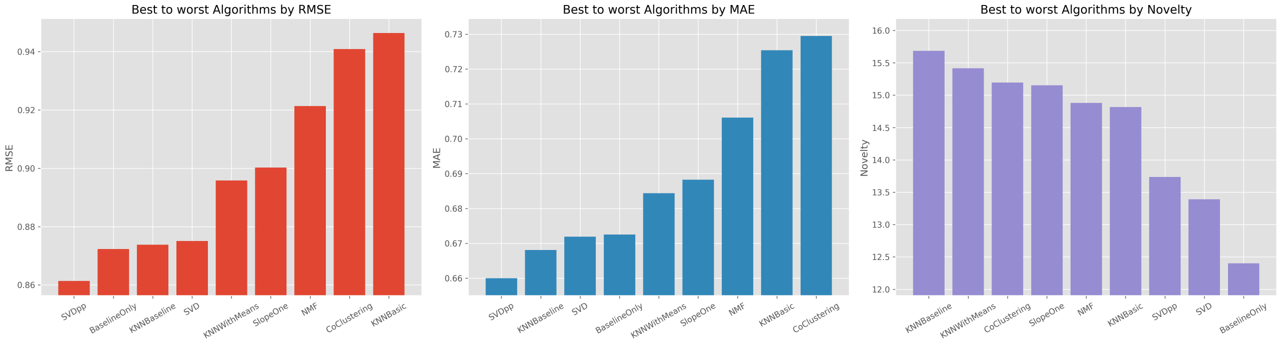
Novelty

$$\text{Novelty} = \frac{1}{|R|} \sum_{i \in R} -\log_2(p(i))$$

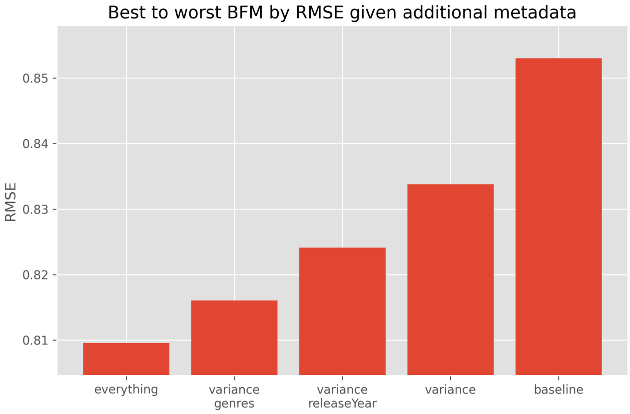
$$\text{where } p(i) = \frac{\text{number of interactions with i}}{\text{total interactions}}$$

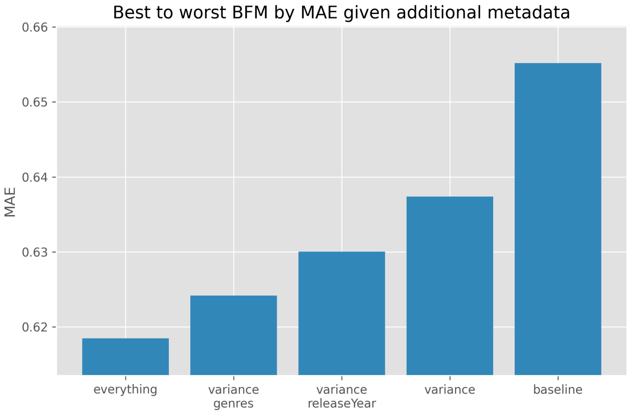
Gives a value for how well the model is at recommending items which few users have interacted with.

Evalution results

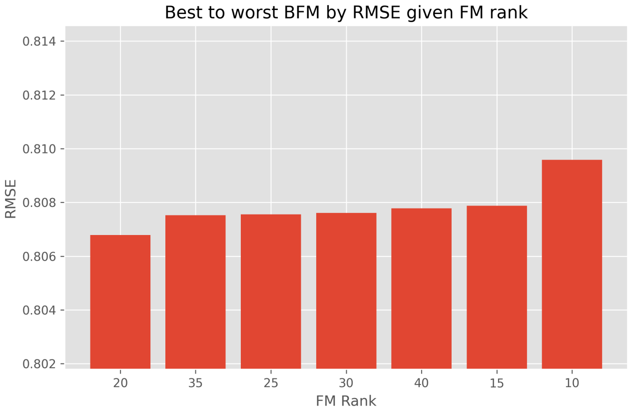


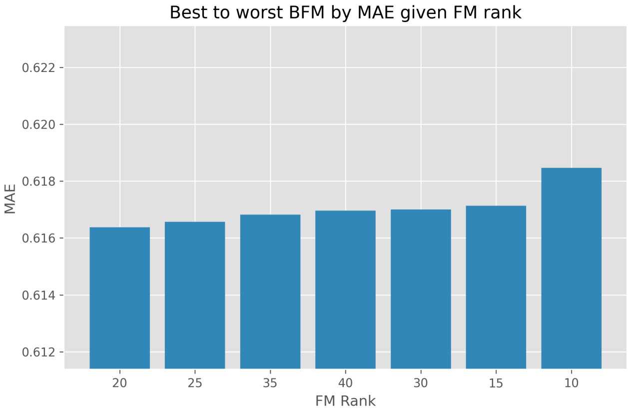
Evalution results (ii)



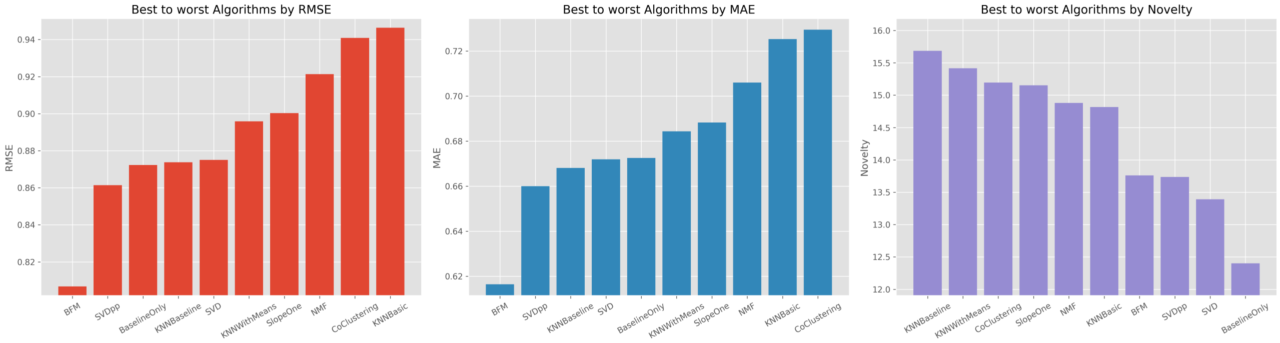


Evalution results (iii)

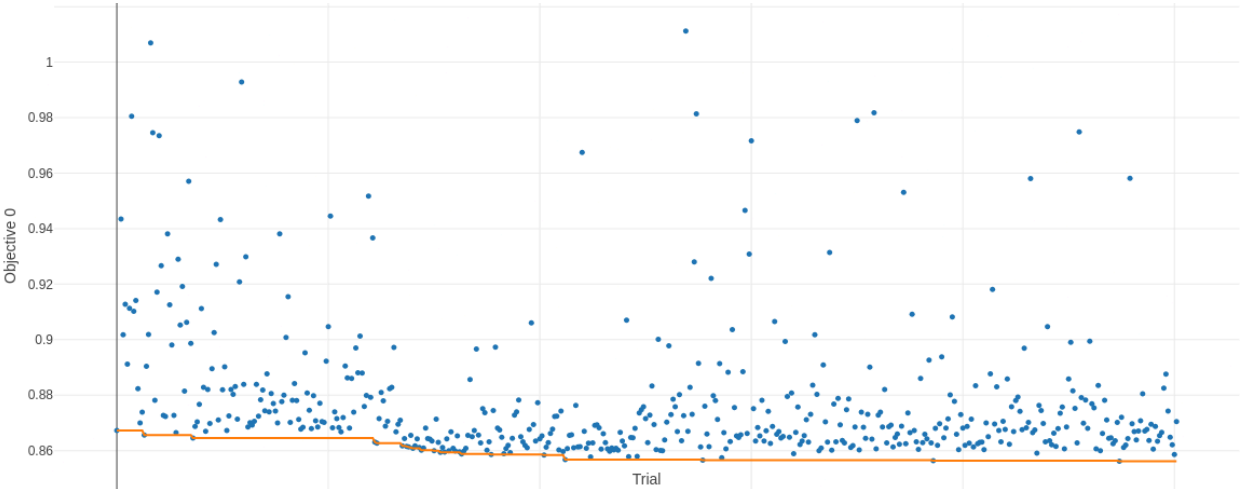




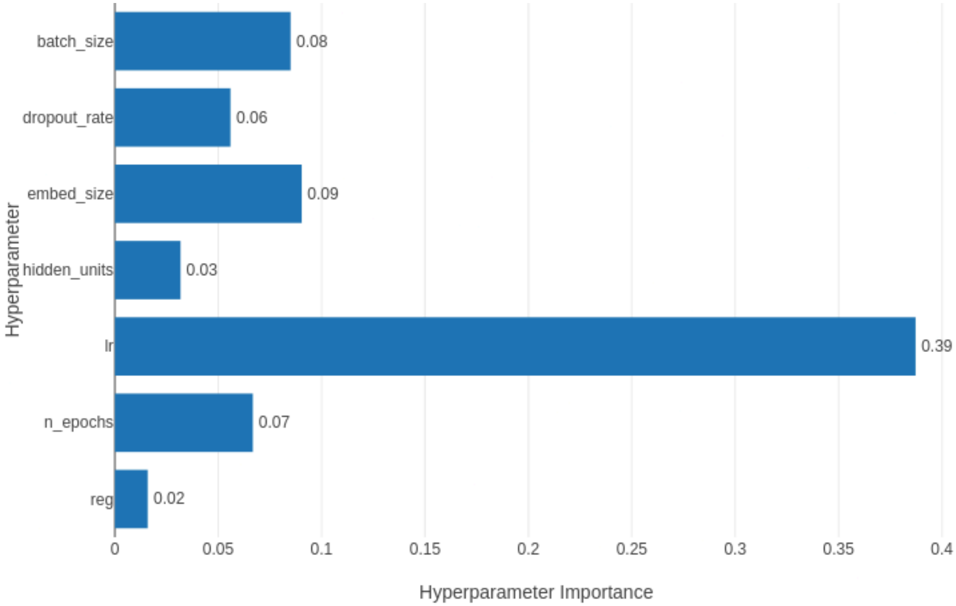
Evalution results (iv)



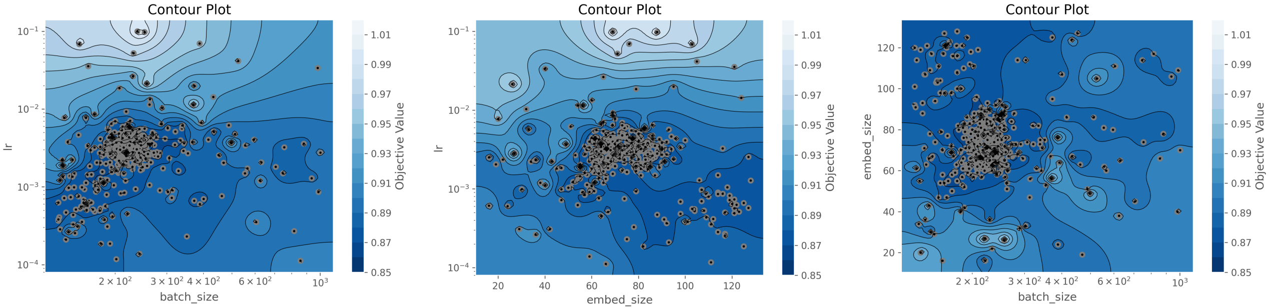
Evalution results (v)



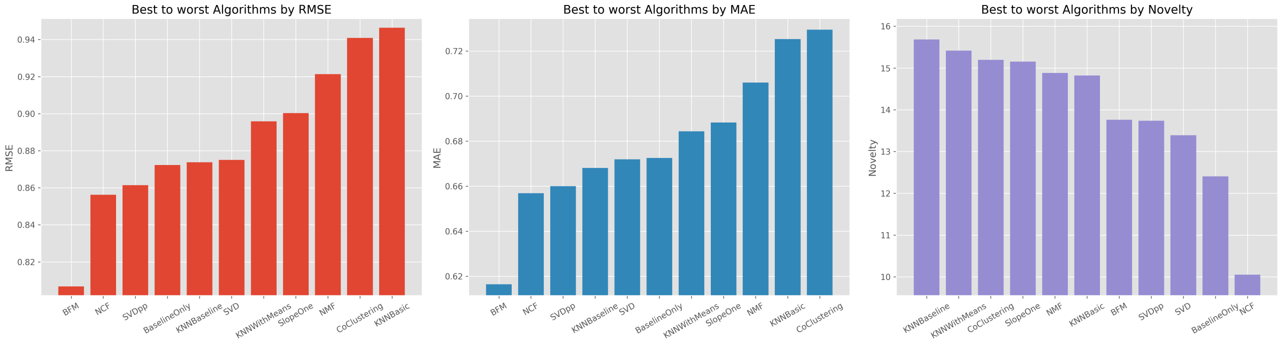
Evalution results (vi)



Evalution results (vii)



Evalution results (viii)



4 Ethical issues

Data leak [6]

Original movielens-100k dataset has all sorts of user data, such as gender, age, zipcode, …

Behaviour manipulation [7]

Is it worth trying to maximize viewer retention for shareholder value?

5 Bibliography

Bibliography

[1] F. M. Harper and J. A. Konstan, “The MovieLens Datasets: History and Context,” *ACM Trans. Interact. Intell. Syst.*, vol. 5, no. 4, Dec. 2015, doi: 10.1145/2827872.

[2] S. Rendle, “Factorization Machines,” in *Proceedings of the 2010 IEEE International Conference on Data Mining*, in ICDM '10. USA: IEEE Computer Society, 2010, pp. 995–1000. doi: 10.1109/ICDM.2010.127.

[3] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, “Neural Collaborative Filtering,” *CoRR*, 2017, [Online]. Available: http://arxiv.org/abs/1708.05031

[4] G. Shani and A. Gunawardana, “Evaluating Recommendation Systems,” vol. 12, 2011, pp. 257–297. doi: 10.1007/978-0-387-85820-3_8.

[5] M. Kaminskas and D. Bridge, “Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems,” *ACM Trans. Interact. Intell. Syst.*, vol. 7, no. 1, Dec. 2016, doi: 10.1145/2926720.

[6] A. Narayanan and V. Shmatikov, “How To Break Anonymity of the Netflix Prize Dataset,” *CoRR*, 2006, [Online]. Available: http://arxiv.org/abs/cs/0610105

Bibliography (ii)

[7] S. Milano, M. Taddeo, and L. Floridi, “Recommender systems and their ethical challenges,” *AI & SOCIETY*, vol. 35, p. , 2020, doi: 10.1007/s00146-020-00950-y.

[8] S. Rendle, L. Zhang, and Y. Koren, “On the Difficulty of Evaluating Baselines: A Study on Recommender Systems,” *CoRR*, 2019, [Online]. Available: http://arxiv.org/abs/1905.01395

[9] L. Wegmeth, “The Impact of Feature Quantity on Recommendation Algorithm Performance: A Movielens-100K Case Study.” [Online]. Available: https://arxiv.org/abs/2207.08713