AML 3204 SOCIAL MEDIA ANALYTICS

Comparing Collaborative filtering based recommender and Hybrid (collaborative plus content) recommender system

PRESENTED BY:

AJAY DAHIYA

AKSHITA KHATRI

ANGELICA SERRANO

JONATAS AGUIAR

FACULTY SUPERVISOR:

DEBASHISH ROY

RECOMMENDER SYSTEM

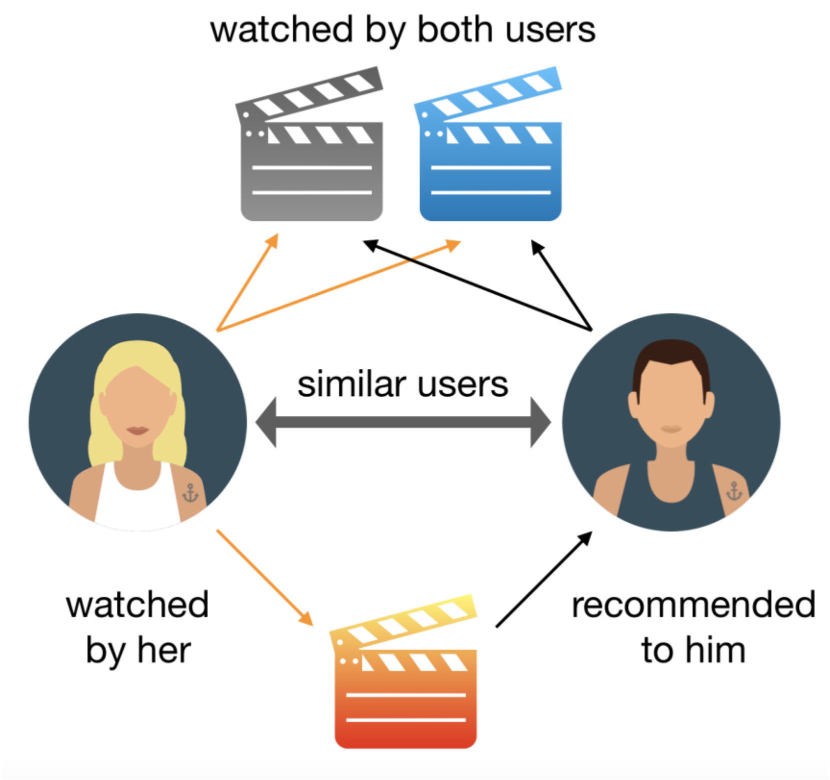


Image source: *https://towardsdatascience.com/recommendation-system-series-part-1-an-executive-guide-to-building-recommendation-system-608f83e2630a*

DATASET

01 Data was taken from MovieLens dataset collected by GroupLens research project

02 Dataset consist of 100,000 ratings from 943 users on 1682 movies.

03 Each user has rated at least 20 movies.

	movieId	movie title	Action	Adventure	Animation	Children's	Comedy	Crime	Documentary	Drama	Fantasy	Film-Noir	Horror	Musical	Mystery	Romance
0	1	Toy Story (1995)	0	0	1	1	1	0	0	0	0	0	0	0	0	0
1	2	GoldenEye (1995)	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	3	Four Rooms (1995)	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	4	Get Shorty (1995)	1	0	0	0	1	0	0	1	0	0	0	0	0	0
4	5	Copycat (1995)	0	0	0	0	0	1	0	1	0	0	0	0	0	0

- **u.data** consist of user, rating and item (movies) column.
- **u.Item** has information about the items.
- **u.genre** consist list of movie genre

TWEETS FROM TWITTER

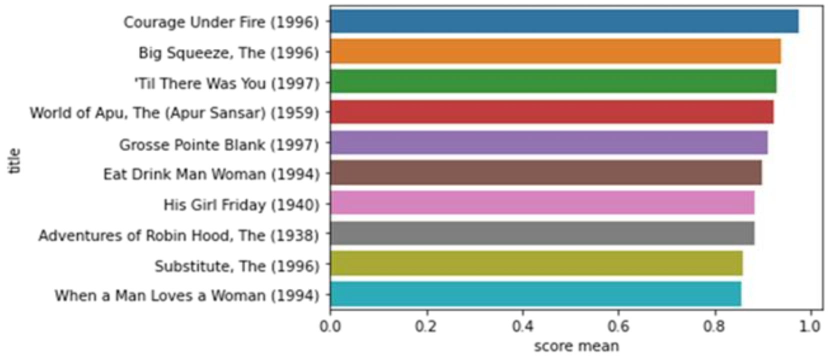
	title	tag
1160	Reckless (1995)	Reckless
1150	Last Summer in the Hamptons (1995)	LastSummerintheHamptons
117	Monty Python's Life of Brian (1979)	MontyPython'sLifeofBrian
1455	Grifters, The (1990)	TheGrifters
424	Species (1995)	Species
844	Carried Away (1996)	CarriedAway
328	William Shakespeare's Romeo and Juliet (1996)	WilliamShakespeare'sRomeoandJuliet
667	Amityville: Dollhouse (1996)	Amityville:Dollhouse
1357	Princess Bride, The (1987)	ThePrincessBride
1027	Carmen Miranda: Bananas Is My Business (1994)	CarmenMiranda:BananasIsMyBusiness

Tweepy was used to extract tweets with tag words as search value.

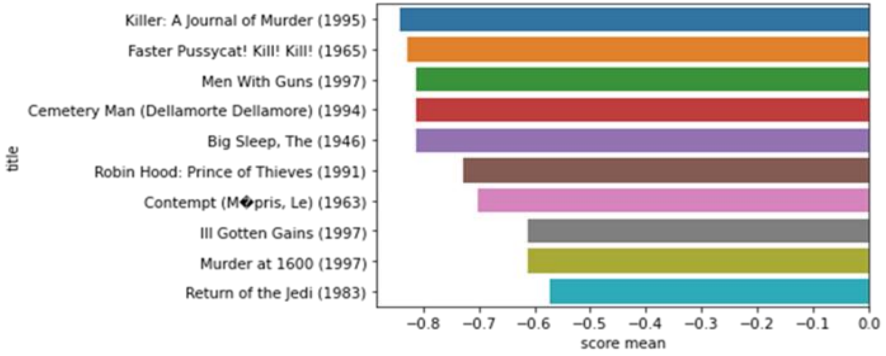
Extracted tweets were preprocessed and cleaned to prepare it for sentiment analysis.

	tag	tweets	clean_txt
3916	Hush	10:43am Hush by The Marias from Cinema https://t.co/07DqEpHPqB	hush maria cinema
1866	BadMoon	@OrksOf @YAlpharius As it zoggin' should 'ave went.	orksof yalpharius zoggin ave went
2352	VegasVacation	Watch the #video \n\nhttps://t.co/uT5vefbViR \nLeave a comment, LIKE, SHARE, TAG a friend.\n\n#WarGames: #Tank riding / shooting an #M9 #Flamethrower #LasVegas\n\n#guns #firearms #vegas #lasvegasstrip #sincity #love #vegasstrip #travel #vegasvacation #weapon #military #pewpew #outside https://t.co/PBwc8vlakW	watch video
2631	Fear	We all have anticipatory nerves before social events, but those with social #anxiety have persistent fear before, during and after - no matter how big or small the event is. I chatted to @drivetimerte about #socialanxiety on Monday. Catch up here ➡️ https://t.co/OxniD9t4Gz	anticipatory nerve social event social anxiety persistent fear matter big small event chatted drivetimerte socialanxiety monday catch
1368	WingsofDesire	✨ Opera Atelier 2021/2022: Wings of Desire ✨\n.\nIn case you missed our thrilling 21/22 Season Announcement, you can catch up on all of the exciting OA news in today's print edition of The Globe and Mail and online here: https://t.co/s8Z0oKaP1t\n.\n#OperaAtelier #WingsOfDesire #Art	opera atelier wing desire case missed thrilling season announcement catch exciting oa news today print edition globe mail online

SENTIMENT ANALYSIS



	title	score mean
268	Courage Under Fire (1996)	0.97500
138	Big Squeeze, The (1996)	0.93710
0	'Til There Was You (1997)	0.93000
1174	World of Apu, The (Apur Sansar) (1959)	0.92310
471	Grosse Pointe Blank (1997)	0.91180
340	Eat Drink Man Woman (1994)	0.89790
507	His Girl Friday (1940)	0.88340
16	Adventures of Robin Hood, The (1938)	0.88165
1016	Substitute, The (1996)	0.85740
1151	When a Man Loves a Woman (1994)	0.85550



	title	score mean
589	Killer: A Journal of Murder (1995)	-0.84020
372	Faster Pussycat! Kill! Kill! (1965)	-0.82710
680	Men With Guns (1997)	-0.81260
221	Cemetery Man (Dellamorte Dellamore) (1994)	-0.81260
137	Big Sleep, The (1946)	-0.81098
887	Robin Hood: Prince of Thieves (1991)	-0.72690
260	Contempt (M�pris, Le) (1963)	-0.70138
534	Ill Gotten Gains (1997)	-0.61240
720	Murder at 1600 (1997)	-0.61240
877	Return of the Jedi (1983)	-0.57268

COLLABORATIVE FILTERING

USING MATRIX FACTORIZATION

	DRAMA	COMEDY
MOVIE 1	1	1
MOVIE 2	3	2
MOVIE 3	1	4
MOVIE 4	1	1
MOVIE 5	3	3

	DRAMA	COMEDY
	✓	✗
	✗	✓
	✓	✗
	✓	✓

	MOVIE 1	MOVIE 2	MOVIE 3	MOVIE 4	MOVIE 5
	1	3	1	1	3
	1	2	4	1	3
	3	1	1	3	1
	4	3	5	4	4

Algorithm : Matrix Factorization based Recommendation

input : Set of users U, items Y,
Matrix of users-items rankings R,
Number of latent features l,
Number of items to be recommended $N \in \mathbb{N}$,

output: Top-N recommendations $R(U) \in \mathcal{I}^N$

for $i \leftarrow 1$ **to** numIterations **do**
  foreach *user* $\in U$ *and item* $\in Y$ *with rating* $R[u, i]$ **do**
    $predicted\_rating = U[u,] * M[i,]^t$
    $err = R[u, i] - predicted\_rating$
    $U[u,] = adjust\_by\_gradient\_descent(U[u,])$
    $M[i,] = adjust\_by\_gradient\_descent(M[i,])$

return N items with highest predicted ranking

Matrix Factorization is one of the most widely used technique for dimensionality reduction.

It works by compressing the initial sparse [user][item] and present them as separate matrix.

HYBRID RECOMMENDER SYSTEM

HYBRID RECOMMENDER SYSTEM USING PYTORCH

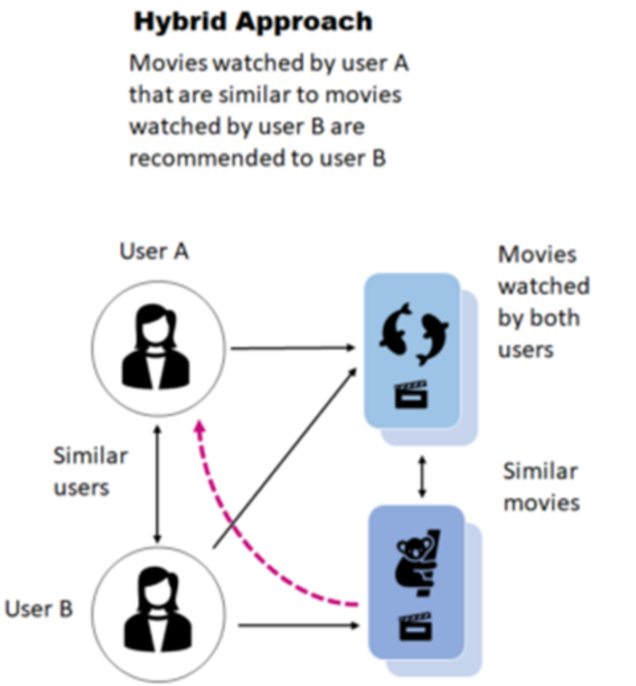


Image source: *https://www.relataly.com/building-a-movie-recommender-using-collaborative-filtering/4376/*

01 Hybrid recommender system has two branches: item-based and user-based.

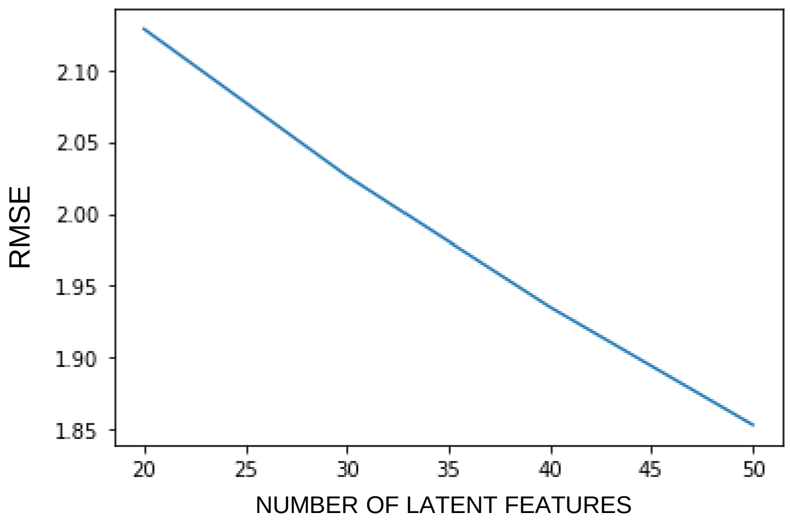
02 It works by combining two or more filtering techniques in different ways to increase the performance and accuracy of the recommender system

03 Pytorch embedding layers were used to create a hybrid recommender system.

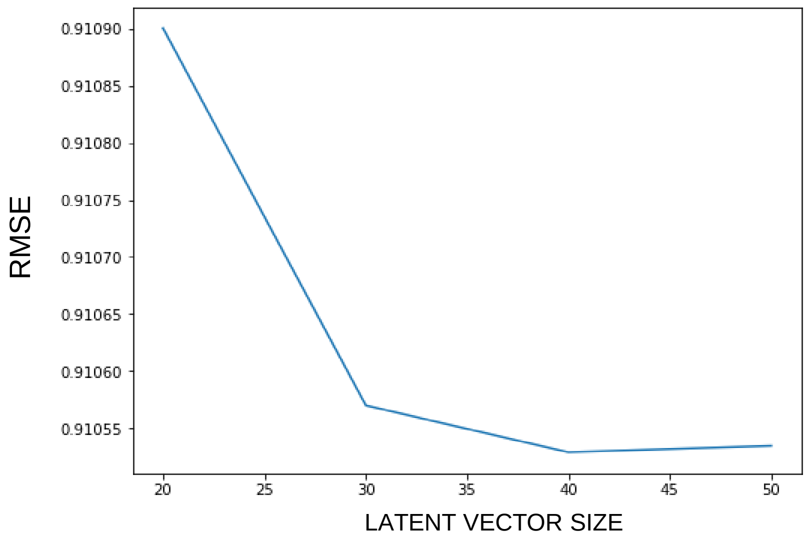
04 The Neural Embedding Layer is a tensor, a vector of vectors composed of random numbers like a conventional matrix.

EVALUATION METRICS

RMSE vs NUMBER OF LATENT FEATURES



RMSE vs NUMBER OF LATENT VECTOR SIZES



MATRIX FACTORIZATION

RMSE decreases as the number of latent features increases. The lowest RMSE value was around 1.85.

NEURAL NETWORK

RMSE decreases as the embedding layers size increase until it reaches the size of 40, when RMSE starts to remain constant at value 0.91 with lower fluctuation.

RESULT AND CONCLUSION

- RMSE decreases as the number of latent features increases from 20 to 50.
- Top 10 recommendation list was generated from each model. Based on the comparison result, all metrics, except recall, were better in the list generated by Hybrid Recommender.
- Top 20 recommendation list was generated from each model. Based on the comparison result, metrics of the Hybrid Recommender were better than the ones of Matrix Factorization.

Hybrid Recommender exhibited better performance than the other method. Thus, it can be a potential method to improve the traditional recommender systems that drive industry today.

Metrics of Top-10 recommendation

	Matrix Factorization	Hybrid Recommender
RMSE	1.289	0.879
MAE	1.099	0.826
Precision	0.143	0.220
Recall	0.862	0.755
F1-score	0.229	0.315

Metrics of Top-20 recommendation

	Matrix Factorization	Hybrid Recommender
RMSE	1.435	0.925
MAE	1.238	0.834
Precision	0.209	0.259
Recall	0.739	0.903
F1-score	0.295	0.367

REFERENCES

1. Ricci F., Rokach, L., Shapira, B. (2015) Introduction and Challenges. In: Ricci F., Rokach L., Shapira (Eds). *Recommender Systems Handbook*. (1-34). https://doi.org/10.1007/978-1-4899-7637-6_1

2. Ricci, F., Rokach, L., Shapira, B. (2011) Introduction to Recommender Systems Handbook. In: Ricci F., Rokach L., Shapira B., Kantor P. (Eds). *Recommender Systems Handbook*. (1-35). https://doi.org/10.1007/978-0-387-85820-3_1

3. Joel Azzopardi, Department of Artificial Intelligence Faculty of ICT University of Malta, Item-based vs User-based Collaborative Recommendation Predictions

4. Melville, P., Sindhwani, V. (2010). Recommender System. Encyclopedia of Machine Learning. https://doi.org/10.1007/978-0-387-30164-8_705

5. Resnick, P. J., Varian, H. R., (1997). Recommender Systems. *Communications of ACM, 519(3),* 56-58. https://doi.org/10.1145/245108.245121

6. G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," in IEEE Transactions on Knowledge & Data Engineering, Apr. 25, 2005, pp. 734-749.

7. B. Mobasher, “Recommender Systems,” Kunstliche Intelligenz, Special Issue on Web Mining, BottcherIT Verlag, Bremen, Germany, 2007, pp. 41-43.

8. M. Y. H. Al-Shamri and K. K. Bharadwaj, "Fuzzy-genetic approach to recommender systems based on a novel hybrid user model," Expert systems with applications, vol. 35, no. 3, pp.1386-1399., Oct. 2008.

9. Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction 12*, 331–370. https://doi.org/10.1023/A:1021240730564

10. Jiawei Han, Jian Pei, in Data Mining (Third Edition), 2012