HYBRID RECOMMENDATION SYSTEMS

DEVELOP AND EVALUATE HYBRID RECOMMENDATION SYSTEMS THAT COMBINE FILTERING AND AI TECHNIQUES.

Zufar Idoyatov
Zhashkey Erkanat
Kair Daniyal

RESEARCH GOAL

The research goal is to develop a sophisticated recommendation system that combines collaborative and content-based filtering techniques, addresses potential issues like hallucinations, and provides users with personalized, diverse, and relevant movie recommendations.

Keywords: Candidate Generation, Ranking, Post-processing, Feature Engineering

DATASETS

MovieLens 20M Dataset

The datasets describe ratings and free-text tagging activities from MovieLens, a movie recommendation service.

It contains 20M ratings and 465k tag applications across 27k movies. These data were created by 138k users between January 09, 1995 and March 31, 2015. This dataset was generated on October 17, 2016.

Users were selected at random for inclusion. All selected users had rated at least 20 movies.

	movieId	title	genres	userId	rating	timestamp
0	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy	3.0	4.0	1999-12-11 13:36:47
1	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy	6.0	5.0	1997-03-13 17:50:52
2	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy	8.0	4.0	1996-06-05 13:37:51
3	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy	10.0	4.0	1999-11-25 02:44:47
4	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy	11.0	4.5	2009-01-02 01:13:41

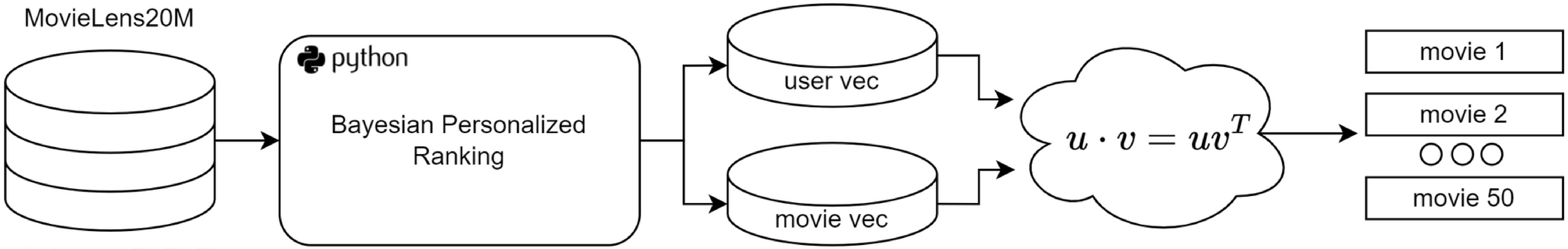
METHODS

BPR Model:
Bayesian Personalized Ranking is collaborative filtering approach that leverages user-item interactions. Recommends items by contrasting good (seen) item vectors with bad (unseen) item vectors based on user vectors. Recalculates all vectors using stochastic gradient descent.

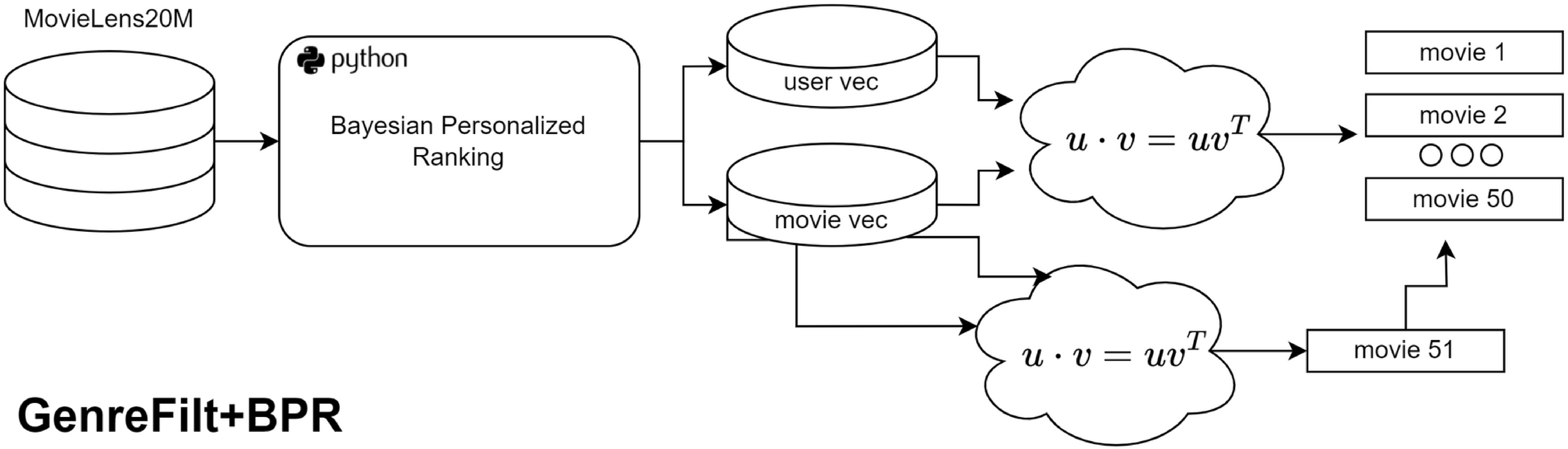
Similar Item:
Utilizes item vectors to find the most similar movies to the last seen movie.
Enhances recommendations by suggesting items that share similarities with the user's previously viewed content.

Genre Filtering:
Helps address hallucinations in the BPR model by filtering recommended items based on relevant genres.
Improves the relevance of recommendations by considering the user's preferences within specific genres.

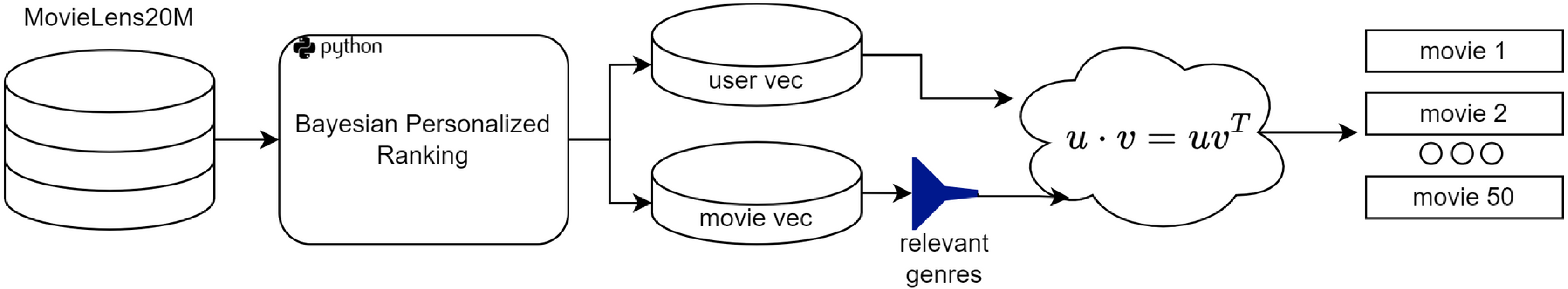
Plain BPR



Sim+BPR



GenreFilt+BPR



NEW METHOD

TF-IDF Ranking

For the following method we used 5 last movies that each user watched, extracted descriptions of this movies, then we vectorized all of the descriptions via tf-idf model. Next, we used average cosine similarity of each of the 5 last watched movies with recommended movie to get it's pseudo-rank. We then changed the order of the recommended movies according to the cosine similarity (descending).

This method allowed us to get incredible results within metrics at N

$$TF(t,d) = \frac{number\ of\ times\ t\ appears\ in\ d}{total\ number\ of\ terms\ in\ d}$$

$$IDF(t) = log\frac{N}{1+df}$$

$$TF-IDF(t,d) = TF(t,d) * IDF(t)$$

RESULTS

model	N	f1@N	recall@N	precision@N
genreFilt+BPR	50	0,0482	0,0784	0,0571
genreFilt+sim+BPR	50	0,0481	0,0784	0,0570
BPR+tfidfRank	50	0,0475	0,0788	0,0562
BPR	50	0,0475	0,0788	0,0562
Sim+BPR	50	0,0474	0,0785	0,0562
BPR+tfidfRank	20	0,0385	0,0514	0,0662
genreFilt+BPR	20	0,0370	0,0409	0,0702
genreFilt+sim+BPR	20	0,0365	0,0403	0,0693
BPR	20	0,0360	0,0396	0,0687
Sim+BPR	20	0,0356	0,0392	0,0679
BPR+tfidfRank	10	0,0370	0,0422	0,0807
genreFilt+BPR	10	0,0273	0,0242	0,0807
BPR	10	0,0266	0,0236	0,0787
genreFilt+sim+BPR	10	0,0265	0,0233	0,0784
Sim+BPR	10	0,0259	0,0229	0,0765
BPR+tfidfRank	5	0,0399	0,0378	0,1099
genreFilt+BPR	5	0,0191	0,0144	0,0917
BPR	5	0,0182	0,0140	0,0873
genreFilt+sim+BPR	5	0,0175	0,0133	0,0836
Sim+BPR	5	0,0169	0,0131	0,0821

CONCLUSION

- Similar item recommendations contribute by considering item features and enhancing diversity.
- Genre filtering helps mitigate potential issues with the BPR model and ensures recommendations align with user preferences within specific genres.
- The combination of these approaches in the hybrid model provides a robust and comprehensive recommendation system that considers both user-item interactions and content characteristics.

THANK YOU FOR YOUR ATTENTION