Comparing Hybrid Collaborative Filters to User/Item-Based Content Filtering

Javier Arteaga Puell, Rodrigo Castañón Martínez, Dakota Mellish

Explanation of Both Methods

User Based Collaborative Filtering + Item-Based Collaborative Filtering (UBCF + IBCF)

UBCF mentioned in 1998 in a research paper and 2001 for IBCF. [1] & [2]

Involves use of metrics cosine similarity, pearson correlation and Jaccard coefficient, either between users' or items' attributes. [3]

Many pairwise comparisons are required, and often the data matrix is sparse, which can compromise the reliability of results. Top similarities are then "recommended"

Hybrid Collaborative Filtering

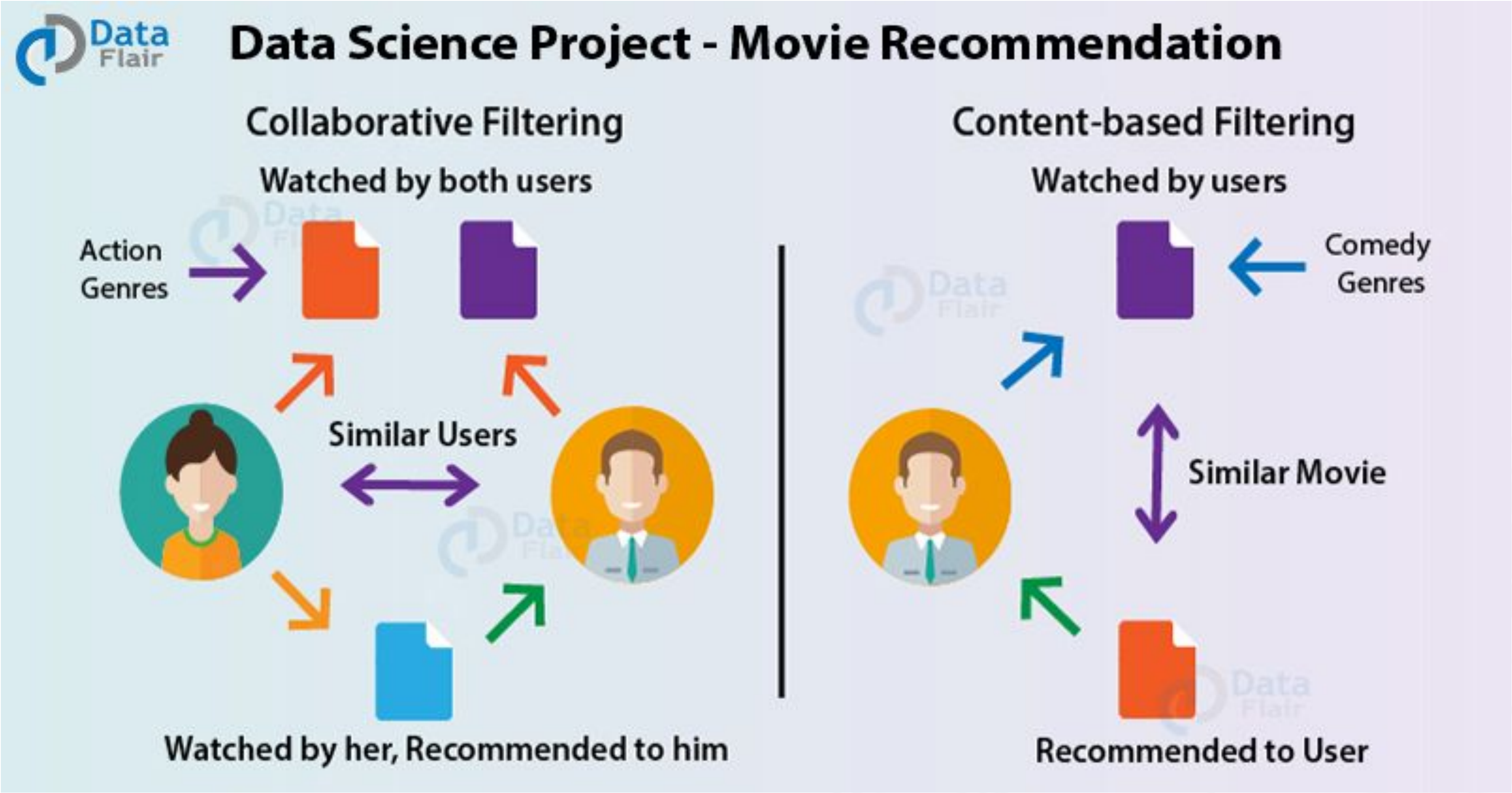
Paper considered was from 2016 [4]

Sequential approach of User based filtering >Clustering -> item-based filtering

Involves the use of Case Based Reasoning (CBR) for the which draws from past experiences and then uses average filling to "densify" the matrix.

This dense matrix is then fed into a genetic algorithm (self organizing map) for identifying clusters of users with their given attributes and then top results are fed into an item-based filter

Example


Data Flair
Data Science Project - Movie Recommendation
Collaborative Filtering
Watched by both users
Action Genres
Similar Users
Watched by her, Recommended to him
Content-based Filtering
Watched by users
Comedy Genres
Similar Movie
Recommended to User

Comparing the Two Methods

Aspect	UBCF + IBCF	Hybrid Collaborative Filtering
Data Handling	Sparse data can compromise results.	Uses data densification (CBR) to mitigate sparsity issues.
Accuracy	Sensitive to data sparsity, leading to inconsistent recommendations	More stable and accurate due to the preprocessing
Scalability	Less scalable; due to many pairwise comparisons.	More scalable; clustering reduces computations.
Cold Start Problem	Affected by lack of historical interactions.	Reduced impact due to clustering and densification.
Best Use Case	Works well with dense datasets and fewer users/items	Suitable for large-scale, sparse recommendation systems
Computational Cost	High due to continuous similarity calculations.	Lower, as clustering optimizes data processing.
Diversity of Recommendations	Can be more repetitive as it relies on user or item similarities alone	More diverse by integrating multiple approaches

Conclusions

- **When to Use UBCF + IBCF?**
  - Ideal for systems with **dense datasets**, smaller user bases, and where simplicity and quick implementation are priorities.
  - Works best when historical user-item interaction data is rich and consistent otherwise the pairwise comparison will struggle

- **When to Use Hybrid Collaborative Filtering?**
  - More effective for **large-scale**, **sparse** recommendation systems where data sparsity and scalability are significant challenges.
  - Preferred for **modern, dynamic applications** (e.g. streaming services like Netflix) that require real-time and accurate recommendations.

Improvement proposals

UBCF + IBCF

- Deep Learning integration
- Feedback profile
- Knowledge graphs
- Side information for cold start

Hybrid

- Clustering optimization
- Data Densification
- Knowledge graphs

References

[1] J. S. Breese, D. Heckerman, and C. Kadie, *"Empirical Analysis of Predictive Algorithms for Collaborative Filtering,"* in Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI'98), Madison, WI, USA, 1998, pp. 43-52.

[2] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, *"Item-Based Collaborative Filtering Recommendation Algorithms,"* in Proceedings of the 10th International Conference on World Wide Web (WWW'01), Hong Kong, 2001, pp. 285–295.

[3] H. Wang, Z. Shen, S. Jiang, G. Sun, R. Zhang, "User-based Collaborative Filtering Algorithm Design and Implementation" *Journal of Physics,* 2020. [Online]/ Available: doi:10.1088/1742-6596/1757/1/012168

[4] N. Kumar, Z. Fan, "Hybrid User-Item Based Collaborative Filtering," *Procedia Computer Science*, vol. 60, pp. 1453-1461, 2016. [Online]. Available:https://www.sciencedirect.com/science/article/pii/S1877050915023492.