International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Enhancing Recommendation Systems with Fuzzy Logic-Based Collaborative Filtering

Yernar Seitay

Kazakh-British Technical University, School of Information Technology and Engineering, Almaty, Kazakhstan
Email: *y_seitay[at]kbtu.kz*

Abstract: ***This research addresses the challenges of sparsity and uncertainty in user-product interaction data by integrating fuzzy logic with collaborative filtering. The proposed fuzzy CF framework utilizes cosine similarity metrics and triangular membership functions to refine similarity scores and predict ratings for unrated items. Experimental results demonstrate that while the fuzzy CF system slightly increases RMSE, it significantly enhances recommendation coverage, making it robust in sparse data scenarios. The findings suggest that fuzzy logic effectively complements traditional CF methods to improve recommendation quality and coverage.***

Keywords: collaborative filtering, recommendation system, user-product interaction, fuzzy inference system, sparse data handling

1. Introduction

Personalized recommendation systems improve user satisfaction through the offering of relevant product suggestions in e-commerce environments. Among the approaches to generating recommendations, collaborative filtering (CF) remains one of the most widely adopted methods. However, CF faces significant challenges such as sparsity, cold-start problems, and an inability to account for uncertainty in user preferences. Fuzzy logic has been a very promising way of dealing with uncertainty, demonstrating the potential to model user-item relationships and improve recommendation systems.

This, when integrated with FIS, leads to the fuzzy CF approach, combining user-based and item-based CF to enhance the quality of recommendations. As an extension of previous work on fuzzy-enhanced CF by Amatriain et al. [10] and others [6]-[8], the work addresses both sparsity and uncertainty challenges. The triangular membership functions, discussed in [9], will permit the fuzzy inference system to handle the uncertainty inherent in user preferences efficiently. This choice is a compromise between computational efficiency and the need for nuanced modeling of input variables.

This study uses Python libraries such as Pandas for data manipulation, scikit-learn for similarity computation, and scikit-fuzzy for fuzzification. In this work, the experiment uses the e-commerce dataset from Kaggle [4], which contains user reviews and product ratings. This has been used to build the user-item interaction matrix for evaluating the proposed recommendation framework.

2. Literature Survey

Collaborative filtering (CF) has been a cornerstone of recommendation systems, though the limitations, such as sparsity and cold-start issues, have motivated people toward hybrid approaches. In [1][2], Breese et al. [7] analyzed predictive algorithms for CF, emphasizing scalability and robustness.

Hybrid recommender systems emerged as a response to these challenges, combining multiple techniques for a broader predictive accuracy and adaptiveness [9]. Fuzzy logic, introduced by Zadeh [3], has proven particularly effective in handling uncertainty, making it a promising addition to recommendation models. Mendel [6] showed that a fuzzy system is robust when the data is noisy, while Gower et al. [5] applied fuzzy clustering to reduce the sparsity in user-item interaction.

Shambour and Lu [8] further extended the concepts by incorporating semantic information into multi-criteria CF and tried to handle more complicated user preferences. These techniques suffer from the problems of scalability and high computational loads very frequently. Based on these developments, the present work incorporates fuzzy logic into CF by using cosine similarity along with triangular membership functions to extend the system's predictability and coverage.

3. Methodology

The hybrid recommendation system in this paper integrates collaborative filtering with fuzzy logic. The methodology consists of the following sections: data preprocessing, similarity computation, fuzzy logic modeling, and evaluation.

3.1 Dataset

The dataset used for this study was sourced from Kaggle and contains user reviews and product ratings from an e-commerce platform [4]. Used attributes are user IDs, product IDs, and their ratings. The dataset provides a diverse representation of user-product interactions, making it suitable for building and evaluating recommendation systems. Due to its sparsity where many users rate only a small subset of products the dataset poses challenges that the hybrid approach aims to address [2].

	product_id	rating	user_id
0	B07JW9H4J1	4.2	AG3D6O4STAQKAY2UVGEUV46KN35Q,AHMY5CWJMMK5BJRBB...
1	B098NS6PVG	4.0	AECPFYFQVRUWC3KGNLJIOREFP5LQ,AGYYVPDD7YG7FYNBX...
2	B096MSW6CT	3.9	AGU3BBQ2V2DDAMOAKGFAWDDQ6QHA,AESFLDV2PT363T2AQ...
3	B08HDJ86NZ	4.2	AEWAZDZZJLQUYVOVGBEUKSLXHQ5A,AG5HTSFRRE6NL3M5S...
4	B08CF3B7N1	4.2	AE3Q6KSUK5P75D5HFYHCRAOLODSA,AFUGIFH5ZAFXRDSZH...
...	...	...	...
1460	B08L7J3T31	4	AHITFY6AHALOFOHOZEOC6XBP4FEA,AFRABBODZJZQB6Z4U...
1461	B01M6453MB	4.1	AFG5FM3NEMOL6BNFRV2NK5FNJCHQ,AGEINTRN6Z563RMLH...
1462	B009P2LIL4	3.6	AGVPWCMAHYQWJOQKMUJN4DW3KM5Q,AF4Q3E66MY4SR7YQZ...
1463	B00J5DYCCA	4	AF2JQCLSCY3QJATWUNNHUSVUPNQQ,AFDMLUXC5LS5RXDJS...
1464	B01486F4G6	4.3	AFGW5PT3R6ZAVQR4Y5MWVAKBZAYA,AG7QNJ2SCS5VS5VYY...

1465 rows × 3 columns

Figure 1: Dataset

3.2 Data Preprocessing

- User and Rating Processing: The user_id field is split into separate rows to handle concatenated entries.
- Rating Normalization: Ratings are converted to numeric values, with non-numeric entries removed.
- Aggregation: Duplicate (user_id, product_id) pairs are aggregated using the mean rating.
- User Filtering: Only users with at least two ratings are retained to ensure meaningful similarity computations.

3.3 Similarity Computation

Cosine Similarity is widely used in recommender systems due to its computational efficiency and the capability of operating with sparse data sets [1]. It calculates user-user and item-item similarities. Compared to other metrics such as Euclidean or Manhattan distances, Cosine similarity has proven to perform consistently well on sparse data [9]. In the system that will be described, it is used to compute the similarity matrices forming the basis for CF:

- **User Similarity**: The similarity of users is a measure that depends on how closely they are related through the rating given to the common items. User-user similarity is computed by treating each user as a vector in multi-dimensional space, where each item corresponds to a dimension. Missing ratings are considered as zeros to maintain computational consistency. The cosine similarity metric determines the angle between these vectors, giving a measure normalized between 0 and 1.
- **Similarity of Items:** Item similarity considers the similarities of products from the ratings given by users. Much like in the case of user-user similarity, the item-item similarity matrix is calculated by transposing the user-item interaction matrix and considering every item a vector. In this way, the method can ensure consistency regarding sparse data.

$$similarity(u,v)=1-\frac{\sum_{i=1}^{n} r_{ui}\cdot r_{vi}}{\sqrt{\sum_{i=1}^{n} r_{ui}^2}\cdot\sqrt{\sum_{i=1}^{n} r_{vi}^2}}$$

Figure 2: Cosine similarity

- **Construction of Matrices:** Both these similarity matrices are stored and then utilized to find the most similar users or items for any target user or item.

$$R=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}$$

Figure 3: User-Item Matrix

3.4 Fuzzy Inference System (FIS)

The fuzzy inference system builds on previous works [6][8] that takes a triangular membership function to represent input and output variables. The fuzzy inference system refines similarity scores and predicts ratings for unrated items. The FIS includes:

- Fuzzy Variables: Antecedent variables: aru (user-user similarity) and aru (item-item similarity).
- Membership Functions: Triangular membership functions model the input and output variables: aru, ari (Low, Medium, High) and rating (Very Low, Low, Medium, High, Very High)

$$Low(x)=trimf(x;0,0,2.5)$$
$$Mediumn(x)=trimf(x;1,2.5,4)$$
$$High(x)=trimf(x;2.5,5,5)$$

Figure 1: Antecedents (ARU, ARI)

$$Very\,Low(x)=trimf(x;0,0,1.25)$$
$$Low(x)=trimf(x;0.5,1.25,2.5)$$
$$Mediumn(x)=trimf(x;1.25,2.5,3.75)$$
$$High(x)=trimf(x;1.25,3.75,5)$$
$$Low(x)=trimf(x;3.75,5,5)$$

Figure 1: Consequent (Rating)

- Fuzzy Rules: Examples include: If aru is Low and ari is Low, then rating is Very Low. If aru is Medium and ari is Medium, then rating is Medium. If aru is High and ari is High, then rating is Very High.
- FIS Implementation: The scikit-fuzzy library is used to define and simulate the fuzzy control system.

3.5 Recommendation Generation

Predictions for unrated items are done based on the K most similar users or items, depending on pre-computed cosine similarities. The prediction involves:

- Similar Users or Items Identification: For a given target user, the K most similar users are identified from the user similarity matrix, then aggregate the ratings of those similar users for the target item to predict the target user rating.
- The calculation of ARU and ARI: Computes the average rating of similar users for the target item (aru) and the average rating of the target user for similar items (ari). These two values are fed as input into the FIS.

$$ARU_{u_t,i_t}=\frac{\sum_{u\in U_k} similarity(u_t,u)\cdot r_{ui_t}}{\sum_{u\in U_k} similarity(u_t,u)}$$

Figure 4: Aggregation of Similarities

- FIS-Based Prediction: The computed aru and ari values are input into the fuzzy inference system. The system evaluates the fuzzy rules and makes a prediction of the rating for the target item.

$$r\hat{}_{u_t, i_t} = FIS\left(ARU_{u_t, i_t}, ARI_{u_t, i_t}\right)$$

Figure 5: Predicted Rating Using Fuzzy Inference System

- Fallback Mechanism: The system falls back to using the average rating of the target item for robustness if similarity scores are inadequate, or if the FIS cannot make a prediction.
- Error Handling: The preprocessing handled such computational errors as missing similarity scores and division by zero using either global or item-level averages.

4. Results

4.1 Comparison: Fuzzy CF vs. Simple CF

While the Classic CF recommended less than 1%, the Fuzzy CF system was able to provide recommendations for 42% of the unrated items of a given user. The broader coverage hereby demonstrates the ability of the fuzzy CF system to provide comprehensive and personalized suggestions, even in sparse data environments.

Thus, though slightly higher than in traditional CF, the RMSE values remain within reasonable bounds, which assured that a reasonable prediction accuracy was achieved with the fuzzy CF. Furthermore, the results confirm broader recommendation coverage, pointing to the practical utility of the fuzzy CF approach as well.

4.2 Evaluation

The system's performance is measured using Root Mean Squared Error (RMSE) between actual and predicted ratings. Table~\ref{tab:metrics} summarizes the results for different values of K:

Table 1: RMSE comparison

	Fuzzy CF	CF
k=5	0.264069	0.0369336
k=10	0.264175	0.0372622
k=20	0.264175	0.0372622

These results indicate that while the fuzzy CF system has a slightly higher RMSE, it effectively provides broader and more comprehensive recommendations, which is critical for real-world e-commerce applications.

```
Total recommendations generated: 564
Top recommendations for User AHYXZVXUY3QTBP7IBFIUBSZVH2XQ
Product ID: B09WN3SRC7, Predicted Rating: 4.7
Product ID: B00K57MR22, Predicted Rating: 4.6
Product ID: B01J1CF05I, Predicted Rating: 4.6
Product ID: B07X2L5Z8C, Predicted Rating: 4.6
Product ID: B08TGG316Z, Predicted Rating: 4.6
Product ID: B0B9BXKBC7, Predicted Rating: 4.6
Product ID: B00LZLQ624, Predicted Rating: 4.5
Product ID: B00NH11KIK, Predicted Rating: 4.5
Product ID: B00NH11PEY, Predicted Rating: 4.5
Product ID: B00NH13Q8W, Predicted Rating: 4.5
Product ID: B00NNQMYNE, Predicted Rating: 4.5
Product ID: B00P93X6EK, Predicted Rating: 4.5
Product ID: B00V4BGDKU, Predicted Rating: 4.5
Product ID: B0148NPH9I, Predicted Rating: 4.5
Product ID: B014SZ090Y, Predicted Rating: 4.5
Product ID: B01DJJVFPC, Predicted Rating: 4.5
Product ID: B01HJI0FS2, Predicted Rating: 4.5
Product ID: B071VNHMX2, Predicted Rating: 4.5
Product ID: B0765B3TH7, Predicted Rating: 4.5
Product ID: B07G3YNLJB, Predicted Rating: 4.5
```

Figure 6: Predicted products for random user Id

5. Conclusion

This study demonstrates that integrating fuzzy logic with collaborative filtering can address significant challenges in recommendation systems, such as sparsity and uncertainty. By leveraging fuzzy inference systems, the approach enhances recommendation coverage, making it highly practical for real-world applications. Future studies could refine this framework further by incorporating external data sources and optimizing membership function parameters.

References

[1] 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, ser. WWW ’01.New York, NY, USA: Association for Computing Machinery, 2001, p.285–295. [Online]. Available: https://doi.org/10.1145/371920.372071

[2] J. Bobadilla, F. Ortega, A. Hernando, and A. Guti´eRrez, “Recommender systems survey,” Know.-Based Syst., vol. 46, p. 109–132, Jul. 2013. [Online]. Available: https://doi.org/10.1016/j.knosys.2013.03.012

[3] L. Zadeh, “Fuzzy sets,” Information and Control, vol. 8, no. 3, pp. 338–353, 1965. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S001999586590241X

[4] “Amazon dataset from kaggle,” https://www.kaggle.com/datasets/karkavelrajaj/amazon-sales-dataset.

[5] J. C. Gower, “A general coefficient of similarity and some of its properties,” Biometrics, vol. 27, no. 4, pp. 857–871, 1971. [Online]. Available: http://www.jstor.org/stable/2528823

[6] J. Mendel, “Fuzzy logic systems for engineering: a tutorial,” Proceedings of the IEEE, vol. 83, no. 3, pp. 345–377, 1995.

[7] J. S. Breese, D. Heckerman, and C. Kadie, “Empirical analysis of predictive algorithms for collaborative filtering,” in Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, ser. UAI’98. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998, p. 43–52.

[8] Q. Shambour and J. Lu, “A hybrid multi-criteria semantic-enhanced collaborative filtering approach for personalized recommendations,” pp.71 – 78, 09 2011.
[9] G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” IEEE Trans. on Knowl. and Data Eng., vol. 17, no. 6, p. 734–749, Jun. 2005. [Online]. Available: https://doi.org/10.1109/TKDE.2005.99
[10] Amatriain, X., Jaimes*, A., Oliver, N., Pujol, J.M. (2011). Data Mining Methods for Recommender Systems. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P. (eds)
[11] Recommender Systems Handbook. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-85820-3_2