ISSN Online: 2821-1693

Transactions on Machine Intelligence

Journal Homepage: https://tmachineintelligence.ir/index.php/TMI

Performance Improvement for Multi-Criteria Decision Making Using Collaborative Filtering-Based Recommender Systems

M. Fazlazad [1,*]

[1] Master's student in the Department of Computer Engineering, Faculty of Engineering, Yasouj Branch, Islamic Azad University, Yasouj, Iran

ARTICLE INFO

Article History:
Received 2 January 2024
Received in revised form 20 February 2024
Accepted 23 March 2024
Available online 24 March 2024

Keywords:
Recommender Systems, Collaborative Filtering, Ranking Matrix, Implicit Method

ABSTRACT

In the past, making decisions or recommendations, as well as processing data, did not pose significant challenges due to the limited data related to a small number of users. However, with the continuous growth of the population and the exponential increase in data and user profiles across global databases, the task of generating improved decisions or recommendations in terms of time, location, cost, and other characteristics has become more complex. Recommender systems, data mining techniques, and algorithms play a crucial role in addressing these challenges. The escalating attention of both researchers and practitioners towards recommender and data mining systems reflects the increasing difficulty in handling vast datasets efficiently. This article aims to analyze a relatively extensive dataset with diverse characteristics, seeking to achieve optimal clustering or categorization and regression in the shortest possible time, considering economic efficiency and other key features inherent to the dataset. The dataset under consideration exhibits data oscillation in four features. Initially, clustering and regression are performed using the implicit method. Additionally, employing the collaborative filtering approach based on recommender systems, specifically the collaborative filtering method using the ranking matrix (user-item collaboration), is employed. This method yields highly effective recommendations for new users based on a variety of essential criteria.

1. INTRODUCTION

Neighborhood-based filtering, also referred to as participatory filtering, is a method used to assess the degree of similarity (or closeness) among data points within a target dataset. This technique aims to group data items that share one or more common characteristics. However, this method faces several challenges, which have been discussed both in this study [1] and in previous literature.

Notable limitations include the "cold start" problem, which refers to the difficulty of making recommendations for new users or items, as well as the "first item" and "first user-item" issues. The concept of a "neighborhood and ranking matrix" is also introduced to address scenarios involving new users who enter the system or a dataset without

* Corresponding Author: mahnazffazlazad@gmail.com
Master's student in the Department of Computer Engineering, Faculty of Engineering, Yasouj Branch, Islamic Azad University, Yasouj, Iran

http://dx.doi.org/10.47176/TMI.2024.51

any prior interaction history. These users lack previous choices or preferences that the system can use to generate recommendations [2].

To mitigate these limitations, various solutions have been proposed. One common approach is the use of implicit feedback methods in place of explicit user ratings at the initial stage. In this method, users are presented with a series of questions that relate to attributes stored within a cluster head or embedded via regression models. Upon completion, the system identifies the most similar data points or clusters and associates the user accordingly [3].

Moreover, neighborhood-based recommender systems can be enhanced by integrating them with other types of filtering methods, such as collaborative filtering and content-based filtering. These hybrid systems aim to overcome the inherent limitations of individual approaches [4]. For example, when a user does not engage with the implicit method or fails to complete the initial form, the system may rely on alternative strategies such as identifying the most populated or relevant cluster, using algorithms like Top-N to retrieve similar data points, and applying other supplementary techniques [5].

Furthermore, due to the extensive use of data mining in processing large-scale datasets, there is considerable interest in the effectiveness of such recommendation methods. Many users seek meaningful and actionable insights from these systems to navigate the overwhelming volume of available information [6].

2. RECOMMENDER SYSTEMS BASED ON COLLABORATIVE FILTERING

Collaborative filtering, also known as neighborhood-based or memory-based filtering, represents one of the foundational algorithms in the development of recommender systems. These models were among the earliest approaches to adopt various methodologies such as implicit feedback processing and ranking mechanisms. Their core objective is to identify user preferences and behavioral patterns to generate personalized recommendations.

Collaborative filtering methods can be broadly categorized into two major types:

- **User-Based Collaborative Filtering**:

This method identifies a set of users with similar preferences and behaviors. Based on their historical ratings—typically on a fixed scale (e.g., 1 to 5)—the system predicts how a target user might rate a specific item. This prediction helps in recommending new items while mitigating issues like item "cold start." The method relies heavily on **explicit user input**, and its efficacy has been widely discussed in the literature [7].

- **Item-Based Collaborative Filtering**:

In contrast to user-based filtering, this approach focuses on the similarity between items. The recommendation is generated based on items that share similar rating patterns across users. Items that consistently receive high ratings from similar user profiles are prioritized, contributing to more accurate and personalized suggestions [8].

These collaborative filtering approaches both user- and item-based can utilize explicit (e.g., user ratings) or implicit (e.g., clickstream or browsing behavior) data. While such methods have successfully addressed many recommendation challenges, certain data instances—often referred to as extraneous or rare data can present unique difficulties. Depending on the algorithm used, these data points may be excluded or specifically handled to ensure system robustness [9].

Recommender systems strive to maintain fairness and data balance, which are crucial in providing reliable and equitable suggestions across diverse datasets. This balance reinforces the credibility of recommender systems, particularly when deployed at scale [10].

In pursuit of greater fairness and effectiveness, contemporary research explores the integration of artificial intelligence (AI) into collaborative filtering methods. When combined with content-based filtering, these hybrid models can enhance the system's capability to manage data complexity and user diversity [11].

An emerging development in this space is the session-based recommender system, which processes user interactions within a single session to deliver real-time, short-term personalized recommendations. These systems also rely on collaborative filtering principles to adjust dynamically to user preferences [12].

Modern recommender systems have extended beyond traditional domains. They are now utilized in diverse settings, such as food recommendation in restaurants, personalized hotel suggestions, and other commercial applications. By leveraging AI, these systems adapt in real-time to deliver context-aware suggestions that cater to user preferences [13].

In today's digital economy, almost every business—whether product-based or service-oriented—employs an internal e-commerce platform. These platforms are typically powered by recommender engines that guide users through vast amounts of available content. By presenting tailored recommendations, these systems ensure that users can quickly and efficiently discover relevant products or solutions, making recommender systems an essential tool for navigating large datasets and addressing user needs [14].

3. THE PROCESS OF DATA MINING AND PROVIDING RECOMMENDATIONS TO USERS

3.1. The First Stage (Initial Data and Values Refinement)

In this section, we utilize MATLAB as a computational and visualization tool to conduct an initial analysis of the dataset associated with user-item ranking. The purpose is to gain a preliminary understanding of the structure and distribution of the data, particularly in relation to item similarity and user preferences.

To begin, we explore the dataset at a superficial level to assess the initial degree of similarity among data points. Each entry in the dataset represents a user's interaction or rating of an item and is characterized by multiple features. These features reflect different dimensions of similarity and allow us to construct a similarity matrix that highlights how closely users align with various items based on shared attributes. For example, a subset of the data—comprising approximately 3,520 entries—represents individual data points, each considered as a distinct entity with its own coordinates in a two-dimensional space.

As illustrated in the accompanying figure, each point in the dataset occupies a unique position within a normalized coordinate system (with values ranging between 0 and 1 along both the X and Y axes). These coordinates are derived based on the feature values associated with each item or user profile. Each point in this two-dimensional projection serves as a representation of an entity within the system, and this spatial distribution allows us to visually interpret clusters of similarity. Such clustering can be instrumental in generating effective recommendations, as it helps the system identify the most relevant items for each user based on proximity in the feature space.

```
>> load(´data.mat´)
>> points
Points =
```

0.1790	**0.0250**
0.0079	0.1859
0.2085	0.0427
0.2322	0.1916
0.0095	0.1700
0.1112	0.0546
0.2026	0.1338
0.0181	0.1532
0.0836	0.2156
0.0519	0.0069
0.2206	0.1414
0.1895	0.0230
0.0708	0.0861

0.1198	0.1916
0.1870	0.1392
0.0818	0.0367
0.0742	0.2209
0.2130	0.0170
0.1408	0.0828
0.0941	0.0067
0.0231	0.0702
0.1183	0.0517
0.1680	0.1928
0.1014	0.0234
0.1625	0.0539
0.2421	0.0219
0.0427	0.1962
0.0316	0.1478
0.2288	0.0802
0.1808	0.1567
0.1662	0.2258
0.2077	0.0647
0.1529	0.0981
0.0715	0.0492
0.0123	0.2049

Fig.1. Dataset Values of Users-Items

In the next stage, the refined values and numerical data after being processed and converted into the appropriate format—are visualized in the form of a diagram illustrating the initial data refinement. This visualization maps each data point onto a two-dimensional coordinate space, allowing for a clearer interpretation of the dataset's structure. By plotting these values in a geographical-style layout, the diagram provides an intuitive representation of the distribution and relationships among the data points, thereby enhancing our understanding of the underlying patterns and potential item-user similarities.

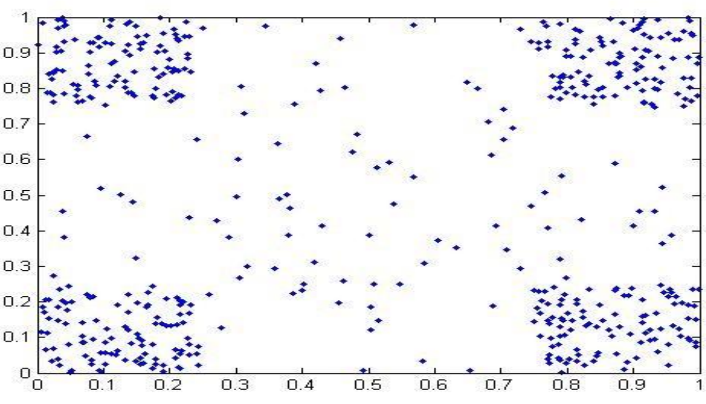

Fig. 2. Refined Points of Numbers and Values of Users-Items

3.2. 2nd Stage (Identification and Similarity of Data of the Same Type To Each Other)

At this stage, all data points—now represented within a two-dimensional space—are characterized by several distinct features. After transitioning from a matrix-based format to a diagrammatic representation, it becomes essential to assess and visualize the degree of similarity between these data points. This enables the identification and separation of similar and dissimilar entities, which are subsequently organized into distinct clusters. The clustering process aims to maximize intra-cluster similarity while minimizing inter-cluster similarity, thereby enhancing the structure and interpretability of the dataset. The following steps illustrate how this transformation and clustering are systematically carried out.

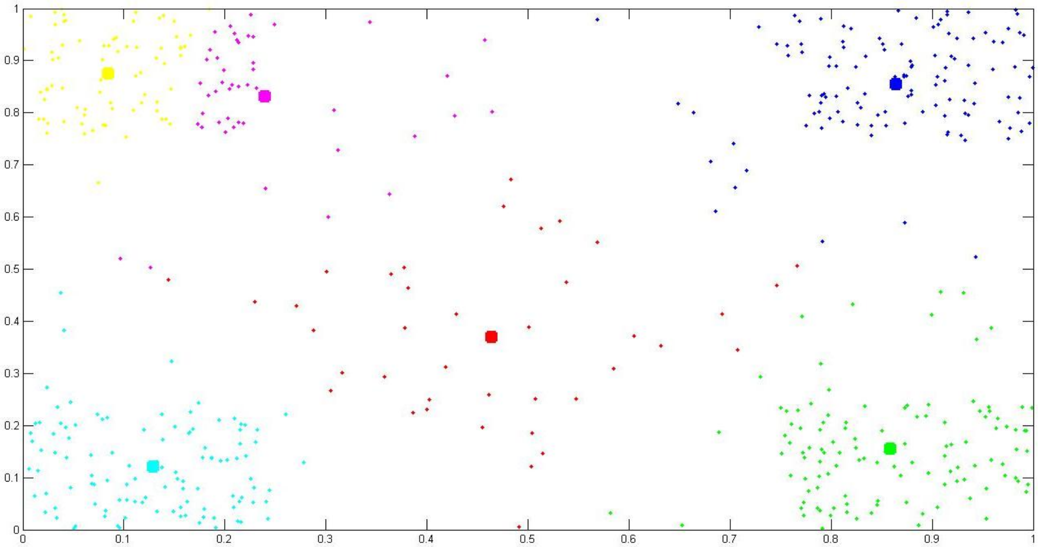

Fig. 3. The Process of The Degree of Similarity in the First Step of The Data

As illustrated in the figure above, the first step involves transforming the initial filtered data using a collaborative filtering-based recommender system algorithm. This algorithm identifies data with the highest user ratings within the dataset and the similarity matrix, while also considering the frequency of item selection by users. At this stage, the algorithm prioritizes top-rated items using the Top-K or Top-N approach, which ranks items based on their relevance and user interest. Initially, this filtering is applied implicitly, and in subsequent stages, it is refined into a matrix format to identify the most suitable and relevant recommendations aligned with user preferences. In the next figure, the second stage of this process is shown, where the similarity between items is further increased, resulting in enhanced clustering and more accurate recommendations.

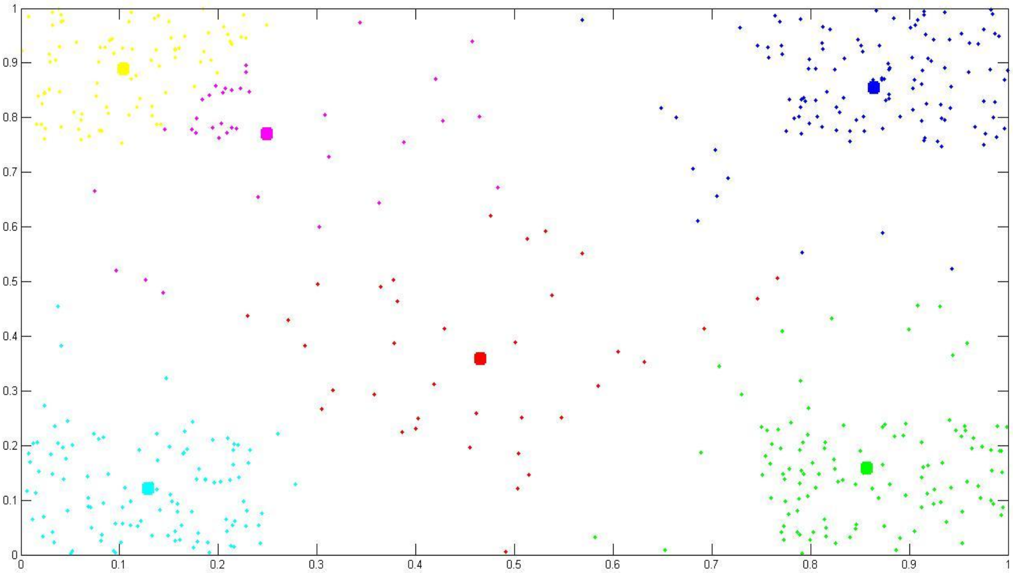

Fig. 4. Increasing the Similarity of Data by The Recommender System Algorithm Based on Collaborative Filtering

In the figure above, the degree of similarity among the target data within the dataset is enhanced, enabling the identification of similar data points within each cluster through collaborative filtering or neighborhood-based methods. This process facilitates the recommendation of relevant content for new users with diverse interests and preferences. Each data point is associated with a set of seven key features, which are reflected in the profiling of each cluster centroid. As the refinement process continues, the data becomes increasingly aligned with user profiles. Initially, implicit methods are employed to interpret user preferences, and in later stages, this information is transformed into a matrix-based structure to generate optimal recommendations that reflect high-priority items aligned with user tastes. This refinement and clustering process is illustrated in the next figure, representing the second stage, where similarity between data points is further increased.

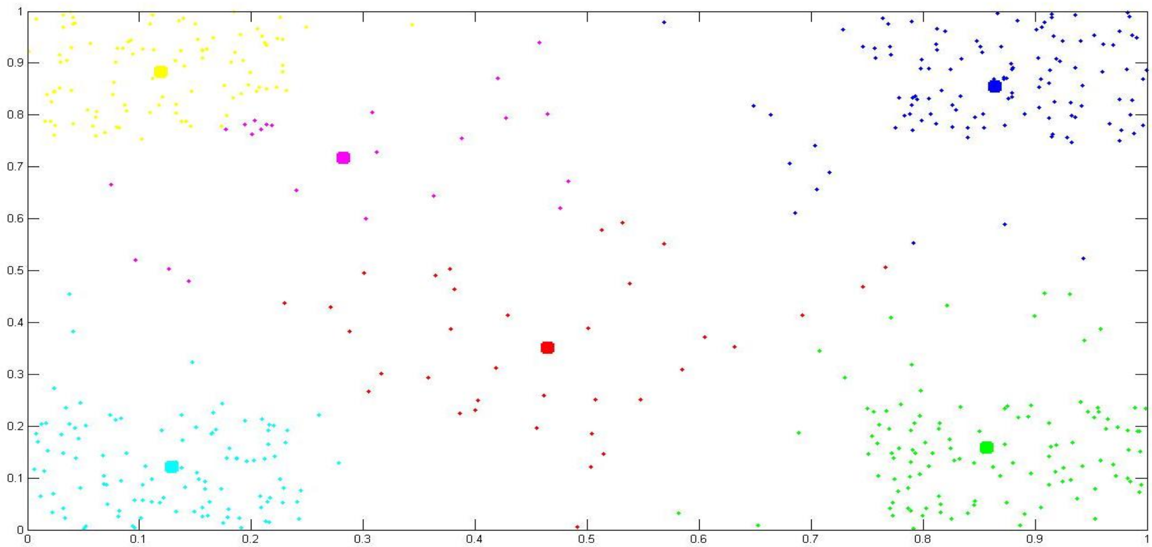

Fig. 5. Identification of Similar Data in The Cluster

3.3. The Third Stage (The Degree of Similarity of Related Data and Providing Recommendations)

In this stage of the process, the objective is to measure the distance between data points (the distance representing the degree of differentiation between each data point or cluster). This measurement is relative to the error rate associated with irrelevant data within each cluster (i.e., the percentage of data points that are unrelated or dissimilar to the other data points within the cluster). The resulting analysis, as depicted in the figure below, provides a clear understanding of how these error rates and distances are calculated.

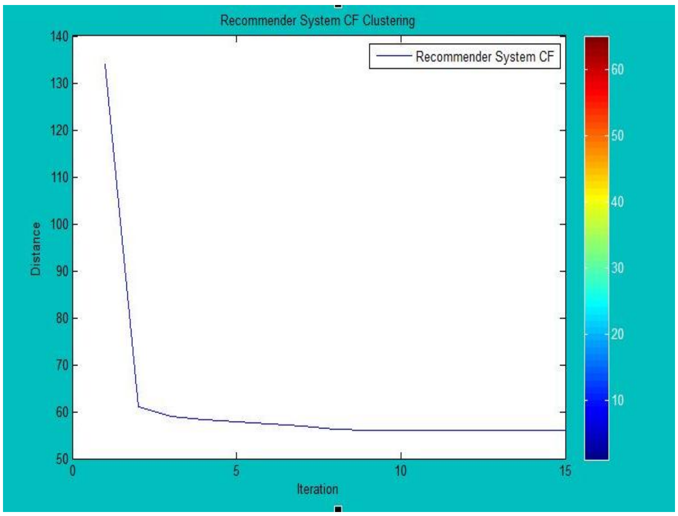

Fig. 6. The Degree of Similarity of Related and Unrelated Data in The Dataset and Clustering

At this stage, we aim to assess the degree of similarity of the desired data in a multi-dimensional environment. This is necessary due to the high number of characteristics associated with each data point. In this multi-dimensional space, each point or data is mapped across various geographic coordinates of the diagram, as illustrated in the figure above. The figure shows that, out of the 500 points, those with the smallest desired distances (ranging from an average distance of 0.10919 to 0.26494) were identified as having the highest degree of similarity. The best matching time for this process was recorded as 56.6998, with an optimal execution time of 17.6216.

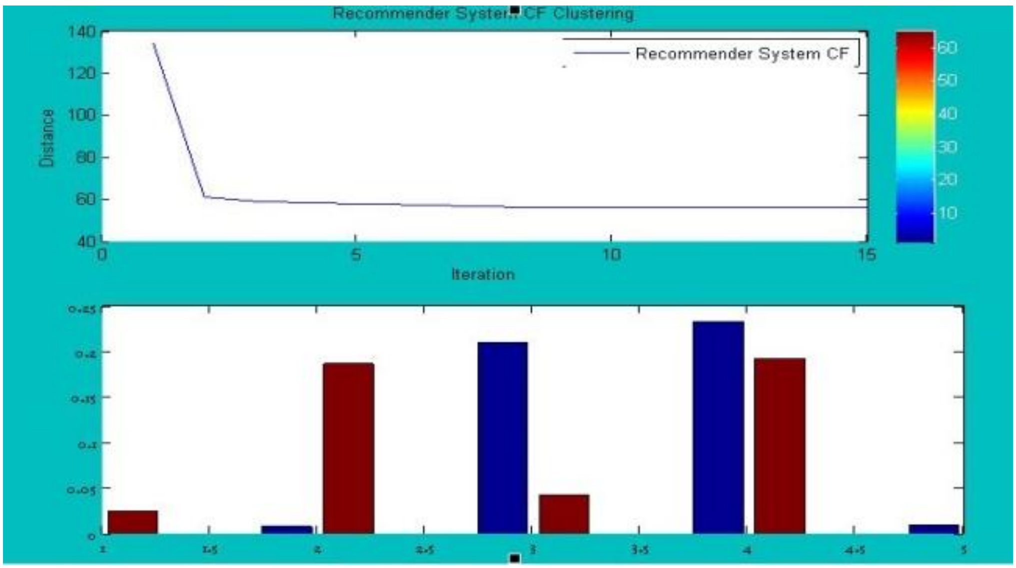

Fig. 7. Two-Dimensional Environment of Similar and Dissimilar Data of The Target Dataset

As shown in the figure above, the values and data associated with each cluster or category in recommender systems based on similarity can be visualized and specified in higher dimensions. By deepening the feature set within the target dataset, it becomes possible to more effectively distinguish between similar and non-similar data. This process can be achieved as follows.

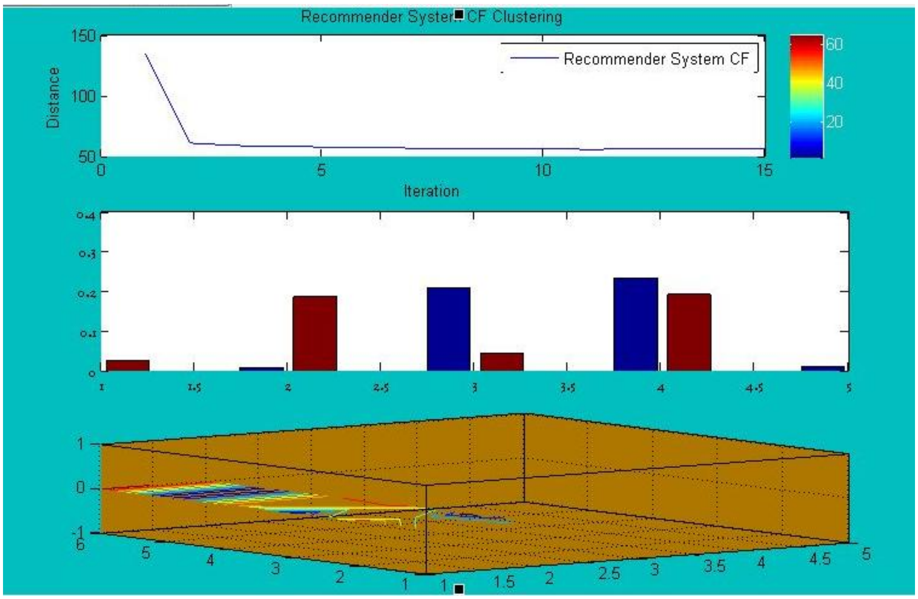

Fig. 8. Multidimensional Environment for Deep Differentiation Between Each of The Features

4. COMPARISON OF THE PROPOSED METHOD WITH PREVIOUS METHODS:

In the improved method outlined above, which surpasses previous techniques, it was observed that Adar Ibn Roush effectively identifies similar data within the target dataset, focusing on specific characteristics such as "Best Matching Time," "Minimum Distance Between Similar Data," and a number of desired attributes. During its implementation in the laboratory tool, the dataset was segmented into several distinct parts. Based on these improvements, a more refined method was developed, which is compared to the earlier methods in the table below.

Table 1. Data Preprocessing Specifications and Points in the Target Dataset:

Plot(Points)	Range-Location (Data)	Type Dataset
500 Points	Lower-Data: 0.0028 Upper-Data: 0.9312	**Dataset-Matrix**

After preprocessing the data in the target dataset and the initial processing of the points (specifying their geographical locations in various dimensions), we proceed with the main processing of the data. The results of this process are shown in the graphs above. The details of this process are summarized in the table below:

Table 2. Data Processing of the Target Dataset Using the Desired Characteristics

Time	Best-Fitness	Range-Distance	CF	Rmse
17.6216	56.6998%	Lower-Dist: 0.023517 Upper-Dist: 0.26494	0.75<=x<=1.5245	**0.23**

In the table above, the number of features in the target dataset is provided, taking into account the scaling of the desired dataset. The following table presents the associated time complexity and cost considerations:

Table 3. The Costs of Different Algorithms Compared to the Characteristics of the Target Dataset

Algorithm	Training	Prediction
C3m	O(m.n+k)	O(m)^2
User-based	-	O(mn)
Slope one	O(mn)^2	O(n)
Tendencies-Based	O(m+n)	O(1)
Rsvd	O(mnk)	O(1)
Rsvd2	O(mnk)	O(1)

Currently, the comparison of the proposed method with respect to the desired outputs, as compared to the previous methods, is presented as follows:

Table 4. Comparison of the Proposed Method with Previous Methods

Svd++	Rsvd	Rsvd2	So	Ub	C3m	Suggested-Method
0.73	**0.75**	**0.68**	**0.63**	**0.78**	**0.43**	0.37

5. CONCLUSION:

Recommender systems are a distinguishing feature of intelligence compared to other platforms. In this article, this was achieved through primary clustering after the data refinement stage. Each step involved calculating the degree of similarity with the desired points across various scales, along with continuous adjustments. Subsequent stages focused on measuring the distance between similar data points (each with different profiling) and the distance between similar and dissimilar data, which played a significant role in clustering. Additionally, the accuracy of finding similar data in collaborative filtering-based recommender systems was improved. As a result, the proposed method outperforms other clustering algorithms, such as K-Medoids or KNN, in terms of effectiveness and accuracy.

Declaration

We acknowledge that we used ChatGPT to enhance the academic writing of our manuscript while ensuring the originality and integrity of our work.

Transparency Statement

The data supporting this study are available upon reasonable request to the corresponding author, subject to ethical and confidentiality considerations.

Acknowledgments

We would like to express our gratitude to all individuals who contributed to this project.

Declaration of Interest

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

REFERENCES

[1] Kohavi, R., Longbotham, R., Sommerfield, D., & Henne, R. (2009). Controlled experiments on the Web: Survey and practical guide. Data Mining and Knowledge Discovery, 18(1), 140–181. https://doi.org/10.1007/s10618-008-0114-1

[2] Mahmood, T., & Ricci, F. (2007). Learning and adaptivity in interactive recommender systems. In Proceedings of the International Conference on Electronic Commerce (pp. 75–84). https://doi.org/10.1145/1282100.1282114

[3] Mahmood, T., & Ricci, F. (2009). Improving recommender systems with adaptive conversational strategies. In Proceedings of the ACM Conference on Hypertext and Hypermedia (pp. 73–82). https://doi.org/10.1145/1557914.1557930

[4] Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: An introduction. MIT Press. https://doi.org/10.1109/TNN.1998.712192

[5] Koren, Y. (2009). The BellKor solution to the Netflix grand prize. Netflix Prize Documentation, 81. http://www.netflixprize.com/assets/GrandPrize2009_BPC_BellKor.pdf

[6] Wang, Y., Ma, W., Zhang, M., Liu, Y., & Ma, S. (2022). A survey on the fairness of recommender systems. ACM Transactions on Information Systems. https://doi.org/10.1145/3547333

[7] McNee, S. M., Riedl, J., & Konstan, J. A. (2006). Being accurate is not enough: How accuracy metrics have hurt recommender systems. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1097–1101). https://doi.org/10.1145/1125451.1125659

[8] Cramer, H., Evers, V., Ramlal, S., van Someren, M., Rutledge, L., Stash, N., Aroyo, L., & Wielinga, B. (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. User Modeling and User-Adapted Interaction, 18(5), 455–496. https://doi.org/10.1007/s11257-008-9051-3

[9] Das, A. S., Datar, M., Garg, A., & Rajaram, S. (2007). Google news personalization: Scalable online collaborative filtering. In Proceedings of the World Wide Web Conference (pp. 271–280). https://doi.org/10.1145/1242572.1242610

[10] Zhang, H., Wu, B., Yuan, X., Pan, S., Tong, H., & Pei, J. (2022). Trustworthy graph neural networks: Aspects, methods and trends. arXiv preprint arXiv:2205.07424.

[11] Zhang, H., Yuan, X., Nguyen, Q. V. H., & Pan, S. (2023). On the interaction between node fairness and edge privacy in graph neural networks. arXiv preprint arXiv:2301.12951.

[12] Jin, D., Wang, L., Zheng, Y., Song, G., Jiang, F., Li, X., Lin, W., & Pan, S. (2023). Dual intent enhanced graph neural network for session-based new item recommendation. In Proceedings of the ACM Web Conference 2023 (pp. 684–693). https://doi.org/10.1145/3543507.3583526

[13] Yera, R., Alzahrani, A. A., Martínez, L., & Rodríguez, R. M. (2023). A systematic review on food recommender systems for diabetic patients. International Journal of Environmental Research and Public Health, 20, 4248. https://doi.org/10.3390/ijerph20054248

[14] Avini, H., Mirzaei ZavardJani, Z., & Avini, A. (2022). Improved recommender systems using data mining. Transactions on Machine Intelligence, 5(2), 97–114. https://doi.org/10.47176/TMI.2022.97