Available online at http://ijdea.srbiau.ac.ir

Int. J. Data Envelopment Analysis (ISSN 2345-458X)

Vol. 12, No. 4, Year 2024 Article ID IJDEA-00422, Pages 61-70
Research Article

International Journal of Data Envelopment Analysis

Science and Research Branch (IAU)

A New Hybrid Recommender System Integrating Density-Based Clustering and Neural Networks to Addressing the Cold Start Problem

L. Hanie Saleh Tabari[1], M. Nazari[1*], M. AlyanNezhadi[1], H. Pourrostami[1], S. M. R. Hashemi[2]

[1] Department of Computer Science, University of Science and Technology of Mazandaran, Iran
[2] Faculty of Mathematics, Statistics Computer Science, Semnan University, Semnan, Iran

Received 23 April 2024, Accepted 9 August 2024

Abstract

Recommender systems have become essential for filtering the vast amounts of online content available today, yet they often falter when confronted with the cold start problem: new users or items lack the historical data required for accurate suggestions. To overcome this, our study introduces a hybrid framework that fuses density-based clustering with a Multi-Layer Perceptron (MLP) neural network. Initially, density-based clustering groups users and items according to demographic attributes and content features, revealing latent structure even in sparse settings. Those cluster assignments, together with any available ratings, are then fed into the MLP, which learns complex, non-linear relationships to predict user preferences.

We evaluated our approach on the MovieLens 1M dataset, comprising one million ratings from 6,040 users on 3,952 films, and compared it against classic collaborative filtering and content-based baselines. Across metrics such as RMSE, our method consistently produced lower error rates and higher top-k recommendation accuracy, with the most pronounced gains observed for users with fewer than five prior ratings or newly introduced movies. By leveraging both side information and rating patterns, this hybrid model delivers more personalized recommendations in data-sparse scenarios, ultimately improving user satisfaction and engagement on digital platforms.

Keywords: Recommender Systems; Cold Start Problem; Density Clustering; Neural Network; Hybrid Filtering Techniques.

[*] Corresponding author: Email: Nazari.mousa@mazust.ac.ir

1. Introduction

The availability and management of the vast online content available is greatly enhanced by the implementation of recommender systems, especially in environments like emails, social networks, and e-commerce. The vast majority of these systems are based on matrices of user ratings to predict user preferences, either using collaborative filtering, content-based approaches, or both approaches as a combination. In content-based systems, the focus is on the user and the items based on the characteristics of the user-item relation. A collaborative approach uses similar users' likes to recommend items. Recommender systems design approaches help to worsen the systems, especially when using only one method [1, 2].

The cold start problem presents significant difficulties for traditional recommendation algorithms, especially in scenarios with sparse data. Existing methods, such as collaborative filtering and content-based filtering, often struggle to deliver accurate recommendations under these conditions.

One of these challenges is solving the cold start issue, which is described as when a new user is unable to receive effective recommendations since they have not interacted enough with the system. To aggravate this problem, issues like data sparsity and scalability are also obvious [3, 4]. To effectively tackle the cold start problem, there have been several approaches to fixing it, such as interrelating the traditional approach with clustering to devise a hybrid model. The approaches include evolutionary algorithms, artificial neural networks, and hybrid systems that have recently been examined, which effectively tackle the cold start issue within recommender systems [5-8]. Table 1 shows the challenges and their solutions.

Rahman et al. [9] have proposed a hybrid recommendation system that combines content-based and collaborative filtering techniques with adaptive clustering to tackle the cold start problem, particularly for new users. To improve collaborative filtering, they employ fuzzy user-based clustering, which leverages multidimensional user profiles based on ratings, age, and geographical data. This clustering method enables the grouping of users with similar preferences, addressing the sparsity of data for new users. Tanwar et al. [10] investigated a novel deep neural network-based hybrid recommender system that integrates user-user networks to improve recommendation accuracy and address cold start issues. This approach is the integration of user-user networks with the DNN model, enhancing collaboration and synergy in recommendations. The experiment results were achieved on publicly available datasets such as Flixster and MovieLens, which demonstrate significant performance improvements. Specifically, the model achieves a 19% reduction in RMSE, a 9.2% improvement in MAE, and a 4.1% increase in F1 Score compared to existing premier methods.

According to Alizadeh et al. [11], they proposed a hybrid recommendation system based on collaborative filtering and a content-based system. In the proposed method, a hybrid predictor system with a regression tree and a perceptron neural network is used for the cold start problem. Tahmasebi et al. [12] have proposed a method similar to user ranking, where the similarity of user demographics is used to select a stronger set of neighbors for the target user. An evolutionary neural network-based system and genetic algorithm provided by Şeref [13] are used to improve this challenge. The proposed system was evaluated by using the Movie Lens dataset, and the results indicated that the performance of the proposed recommended system improved compared to other methods.

A hybrid recommender system based on fuzzy clustering (FCM) and supervised learning has been proposed by Duan et al. [14]. The proposed method is defined in two phases: a fuzzy clustering technique to improve the cold start problem, and a recommended system based on a collaborative filtering system. Chen et al. [15] have proposed a hybrid solution based on evolutionary clustering for the cold start challenge. The emphasis of this method is on increasing system accuracy and preventing data congestion.

Panteli and colleagues [6] introduced a methodology that clusters existing users and discovers discriminant frequent patterns to predict the behavior of new users, thereby mitigating the cold start problem. The methodology begins by clustering old users based on their purchase behavior, allowing the identification of discriminant frequent patterns for each cluster. These patterns are then used to infer or "hallucinate" the purchase behavior of new users, enabling the system to provide personalized recommendations despite the absence of historical data. By focusing on frequent patterns within user clusters, the methodology offers a unique perspective on addressing the inherent sparsity challenges of traditional CF models.

The informed advisor system for travel tour planning by using an artificial neural network has been proposed by Bahramian [16]. The proposed method displays that it works better in the accuracy and satisfaction of the user and improves the cold start problem by using artificial neural network methods and a combination of case-based reasoning with the nearest neighbor algorithm.

Camacho et al. [17] used social network data and the K-Means clustering technique to improve the cold start challenge. Sedhain et al. [18] developed an efficient and accurate learning-based approach called LoCo to improve the problem of cold start by using social metadata such as friends of the user's group and his favorite pages.

In this paper, we present a Density-Based Adaptive Neural Collaborative Network (DANCN), which is effective in solving the cold start issue in recommender systems. It utilizes adaptive density-based clustering with artificial neural networks to recommend newly introduced users and items. It has innovatively selected the nearest neighbors set of most similar users as the nearest neighbors set for the target user without presuming the number of clusters. This method is more flexible and adaptive.

The key contributions of this paper are:

Table 1. Challenges and solutions of recommender systems

Challenges	Definition	Solution
Cold start problem	Problem offering suggestions to users who have not previously voted for items (cold start users) Problem presenting a new user suggestion for an item that has not been rated before (cold start of items)	Use of clustering techniques
Sparsity	Do not rate most items by users	Reduce dimensions for users or trivial items
Scalability	Lots of users and items	Use of clustering techniques and systems based on group refinement
Shilling attacks	Entering malicious users or competitors to apply incorrect points to reduce popularity	Apply systems based on group refinement to attacks

- A hybrid approach that integrates content-based and collaborative filtering techniques,
- A flexible framework allowing integration with various estimation methods.

The remainder of this paper is structured as follows: Section 2 details the DANCN methodology. Section 3 presents our experimental setup and results, providing a comparative analysis with existing methods. Finally, Section 4 concludes the paper with a summary of our findings and suggestions for future research directions.

2. Density-based Adaptive Neural Collaborative Network

The Density-based Adaptive Neural Collaborative Network (DBSCAN) groups users/items by similarity without requiring predefined cluster counts. DBSCAN was chosen for its ability to handle arbitrary cluster shapes and identify noise. The algorithm identifies clusters based on a maximum neighborhood radius (Eps) and a minimum number of points (MinPts), which are tuned empirically for optimal performance. DBSCAN operates through two interconnected phases: I) Adaptive User Clustering, and II) Neural Network Ranking.

The first phase utilizes density-based clustering without predetermining cluster numbers. This approach identifies users with similar preferences dynamically, incorporates demographic and rating information, and filters out users with random or unrealistic preferences. In the proposed method, the users' ontology information is used to measure their scores in addition to the information about the rankings given to various items by users. In the second phase, neural network ranking based on Multi-layer Perceptron (MLP) is used to learn complex user-item interaction patterns and generate personalized recommendations (see Fig. 1).

The clustering method is used to determine the degree of similarity between users. The density-based clustering view is employed, which is based on the information in the User-Item matrix, because density-based clustering techniques do not require knowing the number of clusters. The clustering process starts with randomly selecting a user from the User-Item matrix and considers it as the first member of the first cluster. Then all Eps_neighborhood (users with maximum Eps distance) are retrieved from the selected user and are considered as members of the cluster.

This process continues until Eps_neighborhood is restored to all cluster users. Users who are designated as members of the first cluster are users with similar tastes and interests. The mentioned procedure continues to cluster the rest of the users. At the end of the clustering process, users are placed in a group whose tastes are similar. Noise detection (outlier) is one of the most important features of density-based clustering algorithms, and due to the use of this view in the proposed method, users with random and unrealistic views are determined and removed from subsequent processes of the proposed method at the end of the clustering process. To determine the relationship between age, gender, user occupation, and movie genre, the following procedure is performed based on a combined approach after user clustering:

- Calculating the matrix of user information based on the movie,
- Determining users' interests in genres based on their personality traits,
- Select one or more genres for user characteristics.

Selecting the favorite movies of the users.

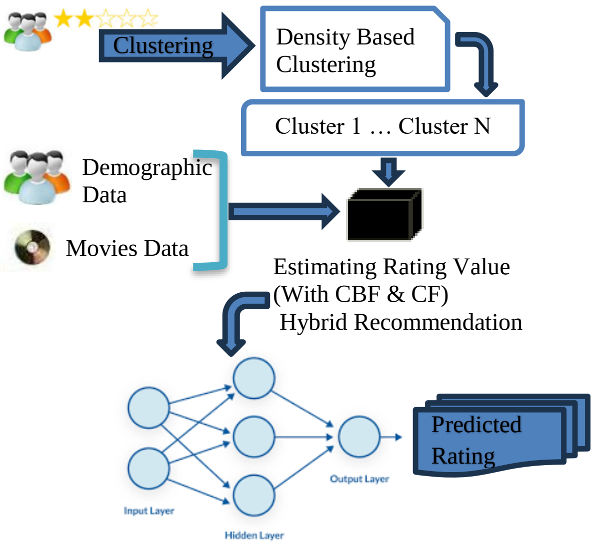

Figure 1. Density-based Adaptive Neural Collaborative Network Framework.

Since the user may have assigned the highest score to an item randomly, their performance scores have been re-evaluated to measure them accurately. Then, an artificial neural network is used to estimate the proposals in phase two of the proposed method after clustering and ranking estimation. To achieve the aim, a multilayer perceptron neural network with an acronym with the sigmoid activator function Equation 1) has been used. The number of hidden layer neurons is determined by performing many tests and examinations, and the weights are also adjusted based on the post-diffusion learning method.

$$y(v_i)=\frac{1}{(1+e^{-v_i})} \qquad (1)$$

This relationship is a logistic function, which varies from 0 to 1. y is the output of the i_{th} neuron and v_i is the sum of the weights of the input connections.

The number of input layer neurons is assumed to be equal to the number of inputs. Each node is connected in a layer with weights $w_{(i,j)}$, which are specified in each node in the next layer, and the weights are corrected by the error propagation learning method. After correcting the weights, the obtained signal will act as the input of the hidden layer, and in the hidden layer, the same operation will be repeated, and the desired output will be obtained, which has the lowest error rate.

3. Experiment and results

In this section, first, the used dataset is introduced. Then, the evaluation metrics which are used to measure the performance of the proposed method are investigated. Finally, the achieved results on the introduced dataset are discussed, and a comparison between the proposed method and trend methods has been made [11, 14, 15, 19].

3.1.Dataset

The Movie Lens 1M dataset investigated in this study can be retrieved from the Group lens website (https://grouplens.org/datasets/movielens). The collection contains 1,000,000 ratings, which have been given to 3952 videos by 6040 users. Available rankings include a 5-point numerical scale ranging from 1st to 5th rank, indicating very low interest to very high user interest, respectively. This dataset also includes the users' ontology information and includes features such as the user's ID, age, gender, and occupation. In this data set, each user has rated at least 20 items. The dataset is randomly divided into training and testing data, with 5,000 users which are selected for training and 1,040 users selected for testing.

3.2.Evaluation metrics

The proposed method was evaluated by using the MAE and RMSE criteria. The mean absolute difference between each prediction and score for all scores is acquired from users, and the test is scored by using Mean Absolute Error. “Eq. 2” and “Eq. 3” are explained as follows:

$$\text{MAE}=\frac{\sum u \sum i \left|p_{u,i}-r_{u,i}\right|}{n} \tag{2}$$

$$\text{RMSE}=\frac{1}{n}\sum_{i=1}^{n}\left|p_i-r_i\right| \tag{3}$$

Where p_(u,i) and p_i are the predicted rating of user u to item i, r_(u,i) and r_i are the real user rating, and n is the total number of test ratings.

3.3.Results

The experimental results validate the effectiveness of the proposed hybrid model in addressing the cold start problem. A detailed analysis of the impact of various parameters, such as epsilon (ε) in the density-based clustering, and the overall performance in cold start scenarios is presented. Additionally, we compare the model’s results against state-of-the-art methods.

As mentioned in the previous sections, the proposed method has two steps: user clustering and target rank estimating. Density-based clustering perspective is also used to cluster users. For clustering users, two parameters are needed, Eps and MinPts, which are the maximum neighborhood radius and the minimum number of users in a cluster, respectively

Due to the significance of setting and determining the values of Eps and MinPts, Fig. 2 illustrates the effect of varying epsilon on the MAE and RMSE of the recommendations. The results obtained in both forms indicate that Eps = 30 and MinPts = 20 determine the best performance of the proposed method based on both evaluation parameters.

Comparing these two parameters, Eps plays a more vital role than MinPts in the outcome and the quality of the clustering process. The general rule for setting Eps values depends on the level of data density; so, in dense and dense environments, the value of this parameter is considered less than in sparse environments, and different clustering results are obtained with very small changes. It can be obtained from the performance of the proposed method based on different values of this parameter. MovieLens data does not fall into the category of very dense or very thin data, and seems to be a bit thin in the dataset.

This analysis emphasizes the importance of parameter tuning in density-based clustering for achieving robust recommendations as Selecting an optimal epsilon ensures that the clustering phase captures meaningful relationships among users/items, enhancing the neural network’s input quality.

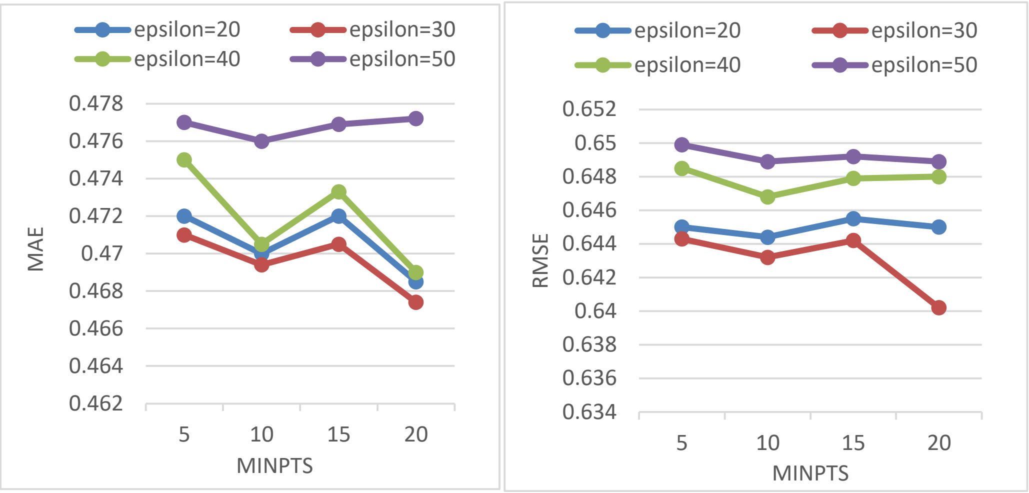

Figure 2. The impact of different values of epsilon on the MAE and RMSE of the recommender error

Table 2. The comparison of MAE and RMSE with other methods

Method	MSE	RMSE
Hernando et al. [1]	0.8	0.8
Chen et al. [2]	0.8	0.9
Duan et al. [3]	0.8	0.9
Proposed method	**0.6**	**0.7**

Building on the importance of parameter tuning demonstrated above, the performance of the proposed method is evaluated against state-of-the-art methods using the optimal parameter settings of Eps=30 and MinPts=20. Table 2 summarizes the comparison of the proposed method's performance with benchmark methods based on the MAE and RMSE metrics. The results based on the MAE and RMSE criteria indicate the appropriate and high performance of the proposed method in comparison with the compared methods.

Also, the method proposed by Chen et al. [15] has the highest RMSE error rate among comparative methods. In addition to the error rate, this method is due to the use of evolutionary algorithm-based clustering and its time overhead, compared to other algorithms with higher computational load. Among the methods that are compared, the best performance for the method is proposed by Alizadeh et al. [11]. This method uses various features of regression trees and perceptron neural networks to solve the cold start challenge and reach this efficiency level. However, the results of our approach have an approximate improvement of 25% compared to this method.

The combination of density-based clustering and neural network predictions in the proposed method demonstrates clear advantages over state-of-the-art approaches. By reducing MAE and RMSE significantly, the proposed hybrid model sets a benchmark for handling sparse and cold start scenarios in recommender systems. By comparing the proposed method to other methods, which will be examined in the following:

1) The maximum value of output estimating error from the actual output will not be more than one in this proposed method which means the proposed method has a very good performance and a small number of data have a significant error and if the number of them is large, it will have a significant impact on error results naturally. Fortunately, the number of points in the graph is small and has no impact on the estimation error.

2) The proposed method by Duan et al. [14], is defined in two phrases. The first phase uses the Fuzzy C-Means Clustering (FCM) to improve the cold start problem and in the next phase, they integrate by using a recommendation system based on collaborative filtering. The absence of fuzzy constraints (FCM) and collaborative filtering is considered an advantage of this proposed method in comparison to the Duan method [14]. Some of the advantages of using the DBSCAN algorithm are the ability to detect clusters with arbitrary shapes, the ability to cluster data with noise, and having a low time complexity, which shows better performance than the fuzzy method.

3) In the same vein, using a hybrid recommender system is much more effective than the other two systems solely because collaborative filtering and content-based systems provide more accurate suggestions. It will also overcome the sparsity problem and solve the cold start problem.

4. Conclusions

This study proposes an innovative hybrid recommendation model that incorporates density-based clustering and a neural network-based collaborative filtering model to solve the cold start problem. A combination of content-based filtering (CBF) and collaborative filtering (CF) based on demographic, movie, and user ratings is made to increase prediction accuracy. Cbf aims to predict the context of the actual user who gets on the system whereas the neural networks serve to process the offer by recognizing intricate patterns.

It was also found through experimentation that the hybrid system is a far more effective system than the conventional systems in diverse circumstances, especially when there is little information on the domain to make predictions on, thus leading to more user satisfaction.

Future works will enhance the clustering approaches, additional machine learning approaches will be tested and the system will be implemented in other domains besides the movie recommendation. In recommendation technologies, this work takes a step forward proposing an efficient solution for the cold start problem.

References

[1] Khusro, S., Ali, Z., & Ullah, I. (2016). Recommender systems: Issues, challenges, and research opportunities. In Information Science and Applications (ICISA) 2016 (pp. 1179–1189). Springer.

[2] Eirinaki, M., Gao, J., Varlamis, I., & Tserpes, K. (2018). Recommender systems for large-scale social networks: A review of challenges and solutions. Information Sciences, 78, 413–418. Elsevier.

[3] Peška, L., Tashu, T. M., & Horváth, T. (2019). Swarm intelligence techniques in recommender systems—A review of recent research. Swarm and Evolutionary Computation, 48, 201–219.

[4] Nilashi, M., Ibrahim, O., & Bagherifard, K. (2018). A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques. Expert Systems with Applications, 92, 507–520.

[5] Pancholi, D., & Selvi, C. (2023). Study of cold-start product recommendations and its solutions. In International Conference on Data Management, Analytics & Innovation (pp. 45–58). Springer.

[6] Panteli, A., & Boutsinas, B. (2023). Addressing the cold-start problem in recommender systems based on frequent patterns. Algorithms, 16(4), 182.

[7] Tsikrika, T., Symeonidis, S., Gialampoukidis, I., Satsiou, A., Vrochidis, S., & Kompatsiaris, I. (2018). A hybrid recommendation system based on density-based clustering. In Internet Science: INSCI 2017 International Workshops, IFIN, DATA ECONOMY, DSI, and CONVERSATIONS (pp. 49–57). Springer.

[8] Hao, B., Yin, H., Zhang, J., Li, C., & Chen, H. (2023). A multi-strategy-based pre-training method for cold-start recommendation. ACM Transactions on Information Systems, 42, 1–24. https://doi.org/10.1002/int.23011

[9] Rahman, M. M., Sifat, F. F., Islam, R., Molla, S., & Khan, M. R. K. (2024). Hybrid recommendation systems using adaptive clustering to address cold-start problems. In 2024 International Conference on Electrical, Computer and Energy Technologies (ICECET) (pp. 1–6). IEEE.

[10] Tanwar, A., & Vishwakarma, D. K. (2023). A deep neural network-based hybrid recommender system with user-user networks. Multimedia Tools and Applications, 82(10), 15613–15633.

[11] Alizadeh, S. H. (2018). A hybrid recommender system using multi-layer perceptron neural network. In 2018 8th Conference of AI & Robotics and 10th RoboCup IranOpen International Symposium (IRANOPEN) (pp. 7–13). IEEE.

[12] Tahmasebi, F., Meghdadi, M., Ahmadian, S., & Valiallahi, K. (2021). A hybrid recommendation system based on profile expansion technique to alleviate cold start problem. Multimedia Tools and Applications, 80, 2339–2354.

[13] Şeref, B., Bostanci, G. E., & Güzel, M. S. (2021). Evolutionary neural networks for improving the prediction performance of recommender systems. Turkish Journal of Electrical Engineering and Computer Sciences, 29(1), 62–77.

[14] Duan, L., Wang, W., & Han, B. (2021). A hybrid recommendation system based on fuzzy c-means clustering and supervised learning. KSII Transactions on Internet and Information Systems (TIIS), 15(7), 2399–2413.

[15] Chen, J., Zhao, C., Uliji, & Chen, L. (2020). Collaborative filtering recommendation algorithm based on user correlation and evolutionary clustering. Complex & Intelligent Systems, 6, 147–156.

[16] Bahramian, Z., Abbaspour, R. A., & Claramunt, C. (2017). A cold start context-aware recommender system for tour planning using artificial neural

network and case-based reasoning. Mobile Information Systems, 2017, Article 2017.

[17] Camacho, L. A. G., & Alves-Souza, S. N. (2018). Social network data to alleviate cold-start in recommender system: A systematic review. Information Processing & Management, 54(4), 529–544.

[18] Sedhain, S., Menon, A., Sanner, S., Xie, L., & Braziunas, D. (2017). Low-rank linear cold-start recommendation from social data. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 31, No. 1).

[19] Hernando, A., Bobadilla, J., Ortega, F., & Gutiérrez, A. (2017). A probabilistic model for recommending to new cold-start non-registered users. Information Sciences, 376, 216–232.