International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 17 • Issue 2 pp. 545-551 March 2026-Sept 2026 www.csjournals.com

Movie Recommendation System: Collaborative vs Content-Based

Anjali, Dr. Vishal Verma, Sandeep Rathi
MCA Student, Professor, Asst. Professor
Deptt. of Comp. Sc. and Application, Ch. Ranbir Singh University, Jind, Haryana, India

Abstract

The recommender system is a key part of a digital service, delivering personalized content that matches the customer's preferences. These systems help to search efficiently and avoid a large number of details in today's age of abundance of information. Within the many recommended approaches discussed in this research, the two of particular concern are Collaborative Filtering (CF) and Content-Based Filtering (CBF), and the one used for an experiment is MovieLens. Three main parameters are used in the comparative study: recommendation accuracy, diversity of recommended items, and handling of cold start problems. Content-Based Filtering predicts film recommendations using the attributes of the films and textual information, while Collaborative Filtering predicts film recommendations by analyzing the likes/dislikes of users and their interactions over time. The feature representation and cosine similarity between items are done using TF-IDF vectorization and cosine similarity, respectively. The Content-Based Filtering model can enhance the prediction accuracy in the selected experimental configuration, and the Collaborative Filtering model generates a large number of recommendations. The study also explores the pros and cons of both approaches and proposes a hybrid recommendation system that combines them to enhance the performance of the overall system and user satisfaction.

Keywords— Recommender Systems, Collaborative Filtering, Content-Based Filtering, MovieLens Dataset, Machine Learning, Recommendation Engine

1. Introduction

The recommender system has since become a crucial part of the digital landscape, helping users find content that resonates with their personal tastes and desires. These systems can analyze user activity, past interactions, and available item information to provide personalized recommendations to improve the overall user experience. These are widely adopted in platforms such as Netflix, Amazon, YouTube, Spotify, and various other streaming, e-commerce, and social media platforms, contributing to enhanced customer satisfaction, engagement, and content accessibility.

As data continues to grow, users have a hard time identifying the relevant data within large data collections. This is known as information overload and is a problem with the efficiency of searching and user satisfaction. The challenge is addressed by recommender systems using intelligent algorithms that anticipate user interests and display the most appropriate items according to the available data.

The various recommendation techniques that are used are Collaborative Filtering (CF) and Content-Based Filtering (CBF), which are still the most popular ones. Unlike content-based filtering, which is based on the features and metadata of items, Collaborative Filtering uses item ratings and interaction patterns of users. In this work, the two approaches are compared with the MovieLens dataset by analyzing various recommendation accuracy measures, diversity of recommendations, and the ability to tackle cold-start problems, and the practical advantages and disadvantages of these two approaches are explored.

2. Literature Review

There has been a lot done in recommender systems in the past 30 years, and many techniques have been proposed to enhance and optimise personalised information retrieval. One of the previous pioneering efforts in demonstrating personal recommendations based on collective user ratings was the GroupLens project being carried out by Resnick et al. (1994). Based on this, Sarwar et al. (2001) proposed an item-based collaborative filtering that has been shown capable of achieving greater scalability and efficiency given a large-scale data set. Considering the limitations of the single recommender techniques, Burke (2002) suggested hybrid recommender systems that integrate several of these algorithms to overcome the common problems associated with them, such as data sparsity and the cold-start phenomenon, and to improve the quality of the recommendations made.

Recently, efforts have been made to compare the effectiveness of the recommended techniques. According to Kurniawan et al. (2024), Content-Based Filtering works with detailed object attributes, particularly in small-sized data sets. They, however, had little discussion of recommendation diversity and scalability, though. Likewise, Raghuwanshi and Pateriya (2025) have discovered that an IB-CF system yields more stable recommendations than user-based systems; however, it has also presented some constraints in setting up a new user's profile and working with limited ratings. Analytically, the available studies have found that feature properties, feature quality, and user interaction properties have a strong impact on the quality of the recommendations. In recent years, the focus of research has shifted towards exploring the potential of combining the advantages of multiple algorithms in order to achieve better accuracy, diversity, and robustness in various areas of application of such hybrid recommendation systems.

3. Problem Statement and Research Gap

Although popular, current recommender systems have several drawbacks such as cold-start; sparsity of user/recommender and item/target matrices; limited diversity of recommendations; and scalability issues.

Numerous studies just present specific recommendations, whereas others do not consider in detail a comparison between such different techniques applying various performance measures. Furthermore, several current systems are not implemented in the real world, for example, with datasets like MovieLens.

Hence, this research is designed to develop and evaluate Collaborative Filtering and Content-Based Filtering techniques to study the effectiveness of Collaborative Filtering and Content-Based Filtering in movie recommendation systems.

4. Research Objectives

To investigate the effectiveness of movie recommendation techniques, the research pursues the following objectives:

- To implement Collaborative Filtering and Content-Based Filtering recommendation models.
- To preprocess and analyze the MovieLens dataset.
- To evaluate recommendation performance using accuracy and diversity metrics.
- To compare the strengths and limitations of both recommendation techniques.

5. Research Methodology

5.1 Dataset Description

This study will employ the MovieLens dataset that was created by the GroupLens Research at the University of Minnesota. The dataset includes movie information, genres, user ratings, and interactions. It is very popular for recommender systems research due to realistic patterns of user behavior.

5.2 Data Preprocessing

The data was preprocessed and machine-processable before training the recommendation models. This stage proved effective for data cleansing to improve data quality, remove inconsistencies, and prepare data for proper model training use. The data was preprocessed in the following manner:

- Removing duplicate records
- Handling missing values
- Cleaning movie metadata and textual features
- Converting textual information into numerical vectors
- Creating user-item interaction matrices

Python 3.11 was used for implementation along with Pandas, NumPy, and Scikit-learn libraries.

5.3 Content-Based Filtering

Content-Based Filtering (CBF) makes an individual movie recommendation based on the similarity between new items to be recommended and those items already liked by an individual user. This method does not depend on the ratings or comments of other users; rather, it uses the intrinsic characteristics of the movie itself, including genres, tags, and the textual description.

Text data was converted to numerical feature vectors for computational analysis using the Term Frequency-Inverse Document Frequency (TF-IDF) method, which gives more weight to informative terms that appear less frequently and less weight to words that appear more frequently.

TF-IDF formula:

$$\text{TF-IDF}(t,d) = \text{TF}(t,d) \times \log(N / \text{df}(t))$$

After feature extraction, the cosine similarity measure was used to verify the similarity between the movie vectors and to measure the similarity of two vector representations of features.

$$\text{Cosine Similarity}(A,B) = (A \cdot B) / (\|A\| \, \|B\|)$$

The model has been implemented in Python with the TfidfVectorizer class and the scikit-learn library. Before the vectorisation, information about a movie's genre, descriptive tags, and plot was packed together to form a full-length text description for each movie. The resulting feature vectors were then compared, and the highest similarity features were recommended to users; those with the highest scores were identified as movies with the greatest similarity.

An important advantage of the Content-Based Filtering approach is that it does not rely on community ratings information. Since the recommendation method is not based on user interactions, it still works in the event of limited historical rating information, thus improving its capabilities when facing cold-start scenarios with new users or datasets with few ratings.

5.4 Collaborative Filtering

Collaborative Filtering (CF) suggests movies by recognizing patterns in the way that users have rated movies in the past. This method does not look at movie features – if a user liked one movie, they will most likely be fond of the other one in the future. This is done by studying users' history of interacting with already rated movies and with each other, so that recommendations can be made.

To develop the model, a user–movie interaction matrix was created, with one row for every user and one column for every movie outcome. The electronic matrix was used as a basis for determining preference patterns and included the information available for ratings.

The methods used to estimate the similarity between users and movies were cosine similarity and Pearson correlation. These similarity measures were used to determine users with similar interests and/or movies that had similar rating patterns. From these relationships, the system

International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 17 • Issue 2 pp. 545-551 March 2026-Sept 2026 www.csjournals.com

was used to predict ratings for movies that the users hadn't yet seen and to generate personal recommendations by choosing those predicted to have the highest preference score.

By uncovering such relationships that are not apparent in movie metadata, Collaborative Filtering can generate a wide range of personalized recommendations. But it greatly relies on the number of user interaction data available and can lose performance when working with little user interaction data or with new user registrations.

5.5 Evaluation Metrics

The recommendation models were evaluated using RMSE (Root Mean Square Error), diversity score, and cold-start handling capability.

RMSE Formula:

$$RMSE = \sqrt{[(1/N)\ \Sigma(y_i - \hat{y}_i)^2]}$$

Lower RMSE values indicate better prediction accuracy.

RMSE analysis was also performed, and recommendation diversity was analyzed. Diversity reflects the variety of the selections of recommended movies by genres and categories. The greater the variety, the more engaged users will be, because they won't have to keep clicking on the same products.

Additionally, the performance in the Cold Start scenario was evaluated to see the performance of the recommendation system for new users and newly added movies having a small amount of interaction history.

6. Results and Discussion

The experimental results obtained from both recommendation techniques are summarized in Table 1.

Technique	RMSE Score	Diversity	Cold-Start Handling
Content-Based Filtering	3.07	Moderate	Good
Collaborative Filtering	3.74	High	Poor

From the results, it can be observed that the end-to-end system of CBF had lower RMSE values in the chosen experimental setup, which shows that the recommendation result is more accurate. Furthermore, CBF did not use the user interaction history; it was also able to deal with cold-start situations better.

However, Collaborative Filtering was able to give a more diverse recommendation, as it found the "relationship" between users and movies that might not have been obvious before. The method was limited by the fact that the data of user-item interaction was rare, and suffered from the cold-start problem.

The results indicate that both techniques have their own pro and cons for different application scenarios.

The results of this experimental analysis also showed that CF is more effective with the provision of a bigger data set of user interactions. The more users and ratings, the better the performance of CF depends on historical ratings. It has poor recommendation accuracy, though, due to its sparse dataset.

Content-based filtering, on the other hand, is based primarily on the content attributes like the genres, tags, and movie description. It means good performance even if there is little history of user interactions. However, the novelty of recommended content may decay over time with repeated recommendations of the same content.

The comparative evaluation provides an idea of the choice of the right recommendation approach in terms of the application's requirements, the dataset size, and users' behaviors.

7. Conclusion and Future Scope

For the evaluation of the performance of Collaborative Filtering (CF) and Content-Based Filtering (CBF) approaches for movie recommendation, a comparative study is carried out with the MovieLens dataset. The experiments demonstrated that the Content-Based Filtering model is more accurate at predicting products, while the Collaborative Filtering model is more effective at finding similarity in user preferences to provide a diverse set of products, under the experiment setup chosen. The result of this experiment leaves us with the conclusion that they both have their merits, and the choice of the proper recommendation method should rely upon the data characteristics and the purpose of the applications.

Both methods are successful, but each has some drawbacks. Cold start and sparsity are issues in collaborative filtering, and content-based filtering can lead to repetitive recommendations.

Future work could be based on a hybrid recommender system, which integrates both approaches, to augment the benefits of both systems. Incorporating hybrid recommendation models can overcome shortcomings of cold start problems and enhance the accuracy and diversity of the recommendation.

Deep learning recommender systems, reinforcement learning, and explainable artificial intelligence (AI) techniques can further enhance the quality of recommendations using intelligent methods. Using Real-time recommendation systems with streaming data and large-scale cloud-based systems might also be considered in industry applications.

Future studies could incorporate different sets of benchmarks and additional measures that are larger and more diverse to further evaluate RRS assessment. Precision, Recall, F1 Score, and NDCG would help to evaluate the quality of the recommendations, efficiency of ranking, and the recommendation system's effectiveness in various application areas.

References

[1] F. Ricci, L. Rokach, and B. Shapira, “Recommender Systems Handbook,” Springer, 2015.

[2] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Transactions on Knowledge and Data Engineering*, 2005.

[3] R. Burke, "Hybrid Recommender Systems: Survey and Experiments," *User Modeling and User-Adapted Interaction*, 2002.

[4] GroupLens Research, "MovieLens Dataset," University of Minnesota.

[5] GroupLens Research, "MovieLens: Movie Recommendation System," University of Minnesota, [Online].

[6] H. H. Kurniawan, W. S. Lukman, R. Fredyan, and M. A. Ibrahim, "Movie Recommendation System: A Comparison of Content-Based and Collaborative Filtering," ResearchGate, 2024, [Online].

[7] S. K. Raghuwanshi and R. K. Pateriya, "Movie Recommendation System: Content-Based and Collaborative Filtering," *International Journal of Computer Science*, 2025, [Online].

[8] G. Parthasarathya and S. Sathiya Devi, "Hybrid Recommendation System Based on Collaborative and Content-Based Filtering," *Journal of Information Technology Research*, vol. 15, no. 2, pp. 45–58, 2022, [Online].

[9] P. Seekhwal, M. Gumber, and S. Tyagi, "Movie Recommendation System Using Content-Based Filtering," SSRN Papers, 2025, [Online].

[10] N. Chaurasia and M. Goyani, "A Review of Movie Recommendation System: Limitations, Survey and Challenges," ResearchGate, 2020, [Online].