Hybrid Approaches for Enhanced Accuracy and Speed in Movie Recommendation Optimization

Levent Ural
Computer Engineering Student
Bahcesehir University
Istanbul, Türkiye
levent.ural@bahcesehir.edu.tr
https://github.com/uralevent/Artificial-Intelligence-Term-Project

***Abstract*—This article discusses hybrid approaches used for increased accuracy and speed in optimizing movie recommendation systems. While traditional recommendation systems often achieve limited success using a single methodology, this study examines a combination of different approaches. Hybrid modeling, which combines intrinsic and extrinsic factors to improve recommendation accuracy, can help predict user preferences more effectively.**

Ultimately, combining these hybrid approaches allows making movie recommendation systems more effective in terms of both accuracy and speed. This study may be useful for researchers looking for new ways to develop recommendation systems.

***Index Terms*—Movie, Hybrid Recommendation Systems, Collaborative Filtering, Content Based Filtering, Matrix factorization**

I. Introduction

Nowadays, with the rapid developments in the digital entertainment industry, movie recommendation systems have gained popularity and played an important role in enriching the user experience. However, traditional recommendation systems often achieve limited success by focusing on a single algorithm or methodology. At this point, adopting hybrid approaches to increase the accuracy and speed of recommendation systems has aroused great interest among researchers and industry professionals.

This article addresses current challenges in the development of movie recommendation systems and examines the role of hybrid models in overcoming these challenges. Where traditional methods have difficulty adequately understanding user preferences, hybrid modeling can provide more effective recommendations by combining internal and external factors. This offers the potential to increase user satisfaction as well as improve the overall performance of the platform.

Additionally, another important factor affecting the performance of movie recommendation systems is speed. Large data sets and complex algorithms can slow down recommendation processes, which can negatively affect user experience. This article will discuss various techniques that can be used to increase the speed of recommendation systems and content based, matrix factorization, collaborative filtering methods will be used.

In conclusion, this study provides a valuable resource for researchers and professionals in the digital entertainment industry, focusing on optimizing movie recommendation systems in terms of both accuracy and speed. The use of hybrid approaches is poised to play an important role in the future evolution of movie recommendation systems and has the potential to guide advances in this field.

II. DATASET

A 5-star movie scoring system was used in the dataset used in the project. In addition, the movies in the dataset are labeled in a way that best expresses their content. This dataset contains 100836 ratings and 3683 tags applied to 9742 movies. While determining the ratings, the ratings of 610 users were taken between March 29, 1996 and September 24, 2018. This dataset was created on September 26, 2018. All this data was obtained from ratings on MovieLens. While scoring, the users selected were selected from random users who had rated at least 20 movies.

III. LITERATURE REVIEW

Hybrid recommendation systems have come to the fore with the interactive integration of various recommendation techniques to address the accuracy and speed challenges directly encountered in movie recommendation optimization. The essence of hybrid approaches lies in their ability to leverage the strengths of multiple recommendation strategies and aim to reduce the weaknesses inherent in individual methods. These approaches often combine traditional content-based and collaborative filtering methods and aim to take advantage of simultaneously evaluating item properties and user-item interactions. Additionally, advances in areas such as model-based collaborative filtering, time-based modeling, and context-aware recommendations have further enriched hybrid frameworks, which promise to more precisely capture evolving user preferences.

As the landscape of movie recommendation systems evolves, this literature review focuses on examining innovative hybrid approaches that aim to strike a delicate balance between increased accuracy and fast recommendation processes.

A. Collaborative Filtering

Collaborative filtering is based on offering a recommendation to users through a model created by the interactions of users with the same tastes or interests.

While performing these operations, the ratings given by users to a movie or product are collected in a database and a user-item matrix is created. With the help of this matrix, similarities in the ratings given by users to a product are used. This process is often called ”neighborhood”.

There are two different methods for collaborative filtering. One of them is memory-based and the other is model-based:

- **Memory-Based Systems:** The basic concept in memory-based systems revolves around simulating and modeling user and product interactions within the system. This involves including factors that represent latent characteristics of users and products, such as user preference categories and product category classes.
  User-Based Method involves comparing users' rating patterns for an item and making predictions for a user based on the evaluations of another user within the neighborhood who has rated the same item. The prediction is calculated by assessing user similarity through a comparison of their ratings for the same item.
  On the other hand, **Item-Based Method** calculates predictions based on the similarity between items. The similarity between user-based and item-based recommendations can be determined using metrics such as the Pearson correlation coefficient and cosine similarity. These methods aim to provide accurate predictions by analyzing either user or item similarities, contributing to the effectiveness of recommendation systems without resorting to direct copying or imitation.

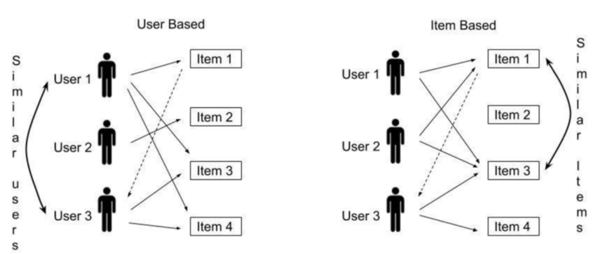

Fig. 1. Example of differences between item-based and user-based.

- **Model-Based recommendation systems** utilize machine learning to understand user ratings and preferences, allowing them to build a model that captures user behavior. By considering hidden characteristics and item preferences, these systems train the model to generate new predictions for users. Moreover, they incorporate users' implicit information, such as the music they listen to, books they read, or websites they visit, to enhance recommendations. Model-Based Collaborative Filtering encompasses various methods, including Cluster Models, Bayesian Networks, Association Rules, and Neural Networks. Ensemble techniques often combine multiple approaches for improved performance. Hybrid approaches integrate collaborative filtering with other techniques, offering a more comprehensive and personalized recommendation system for users. This approach goes beyond direct imitation, leveraging machine learning to provide tailored suggestions based on user behavior and preferences.

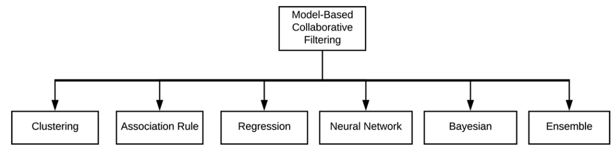

Fig. 2. Techniques of Model-Based Systems

Clustering algorithms, a subset of unsupervised machine learning techniques, organize data according to a predefined model. Widely employed in image processing, pattern recognition, and statistical parameter models, these algorithms segment data into clusters. Once stable clusters are formed, parameters are utilized to calculate averaged recommendations. In this method, users with similar interests are grouped together, facilitating neighborhood-based recommendations. A effective clustering method exhibits high intra-cluster similarity and low inter-cluster similarity. Among the most frequently utilized clustering algorithms is K-means. Notably straightforward to implement, this algorithm consistently demonstrates superior accuracy in recommendations compared to other algorithms.

Bayesian networks rely on conditional probability and Bayes' theorem to derive ratings from nodes that represent individual items.

Association Rule Mining is a technique employed to uncover relationships between items that are frequently purchased together. This mining algorithm predicts the likelihood of an item being chosen based on the historical or simultaneous purchase patterns of other items in transactions.

Neural Networks emulate the structure of the human brain, with numerous interconnected neurons. In a neural network, neurons form layers, encompassing input and output nodes.Collaborative filtering recommendation systems commonly use two types of neural networks: user-based and item-based.

Regression is a statistical technique employed to identify relationships within datasets, particularly aiming to understand the connection between independent variables and a dependent variable. Its primary objective is to establish the correlation between users' ratings and their respective neighborhoods.

An **ensemble model** is created by combining two or more algorithms and methods to enhance recommendation outcomes, surpassing the capabilities of a single method. This model comprises four essential components: ”Boosting,” ”Bagging,” ”Fusion,” and ”Randomness.”

B. Disadvantages of Collaborative Filtering

- Cold Start status is a disadvantage for this filtering. It is very difficult to make predictions for a new user group or just one user. Because the user did not make enough evaluations. There are no similarities yet, so you may receive incorrect suggestions at the beginning. When a new item is added, it is very difficult to recommend this item because it has not received any reviews yet. Someone needs to discover and evaluate that item.
- The algorithm has difficulty in terms of scalability because it has to calculate the evaluations and groupings of many users and items.
- Recommendation systems rely on a comprehensive catalog of product rankings. However, only a subset of this data is used for individual items, resulting in some items having a limited number of ratings. The sparseness of the rankings makes it difficult for the system to generate accurate recommendations.

C. Content-Based Filtering

Content-based filtering approaches attempt to recommend items to the active user that have similar characteristics to items he or she has rated positively in the past. This method is based on the principle that items with similar characteristics will be evaluated similarly. Content-based filtering systems often use information sources, such as text documents, to identify attributes that describe items and align with user preferences.

To better understand how it works, we can examine these steps:

- **Determining Content Properties:** The first step is to determine the content properties of each element. For example, attributes for a movie could be director, actors, genre, release year, etc. For a book, features such as author, genre, main theme can be considered.
- **Creating the User Profile:** A profile is created based on the user's past interactions. These interactions can be, for example, movies watched, books read or music listened to. These interactions are associated with features that reflect the user's interests and preferences.
- **Recommendation Structure:** Using the user's profile and the content properties of the items, items with similar content are recommended to the user. For example, it might focus on actors who starred in a movie the user previously liked or other movies in the same genre.
- **Calculation of Recommendation Points:** Scores of items that match the user's interests are calculated. These scores are used to determine how well they match the user's past preferences and interests.
- **Presenting Suggestions:** Based on the calculated scores, items with the highest scores are recommended to the user. These recommendations include content that is similar to content the user has previously liked.

Content-based filtering can provide personalized recommendations by focusing on the user's interests. However, it may result in the user only receiving recommendations based on their current preferences and may prevent the user from exploring a wider range of content. Therefore, more effective results can be achieved by combining collaborative filtering with a recommendation system.

IV. METHODOLOGY

It is very important to first determine the purpose for this experiment. The aim to be achieved in this experiment is that a hybrid method is expected to give a better result than collaborative filtering or content-based filtering alone. The most appropriate metric to use when making these performance evaluations is the RMSE value.

A. RMSE (Root Mean Square Error)

RMSE (Root Mean Square Error) is a performance metric that measures the amount of error between actual and predicted values. It is especially widely used in regression problems. The lower the RMSE value, the better the model's predictions.

RMSE is calculated by the following formula:

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\left(\frac{d_i - f_i}{\sigma_i}\right)^2} \tag{1}$$

RMSE measures the magnitude of errors between actual values and predicted values, and a lower RMSE value indicates better model performance. Therefore, it is frequently used in evaluating regression models.

B. Item-Based and User-Based Collaborative Filtering

First of all, the RMSE value can be calculated to measure the success of collaborative filtering. In this collaborative method, item-based and user-based collaborative filtering methods can be applied separately at the same time, aiming to provide the best results for collaborative filtering.

Below is a screenshot of the values obtained as a result of applying these two filtering methods on the dataset.

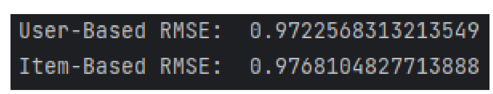

Fig. 3. Item-Based and User-Based RMSE Values

C. Content-Based Filtering

The Surprise library is a library specifically designed for recommender systems and focuses mainly on error metrics such as RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error). These metrics measure the differences between actual and predicted scores and are used to evaluate the model's predictive success.

Metrics used in classification problems, such as precision and recall, cannot be directly applied to recommender systems because recommendation systems often predict the user's actual interest levels, and classification labels, such as whether the user will like or dislike an item, cannot generally be used.

To use classification metrics such as precision and recall, we may first need to set a threshold value on the scores obtained from the recommendation system and classify the recommendations as ”liked” and ”disliked” using this threshold value. However, this approach is a form of converting recommendation systems into classification models and may require a carefully chosen threshold value.

The values obtained when the content-based filtering method is applied to the dataset used are shown in the screenshot below.

```
RMSE: 0.8718026976485529
```

Fig. 4. Content Based RMSE Value

D. *HYBRID Method*

A hybrid model was created by combining the two methods.

- The first step is to create TF-IDF vectors representing movies. These vectors are a feature vector for each movie and express the importance of a particular word in the document.
- Then, a matrix is created by calculating the cosine similarity between TF-IDF vectors. Each element of this matrix expresses the cosine similarity of two documents.
- Then, the cosine similarity matrix (content similarity) is calculated on these vectors.
- A function is defined that creates hybrid recommendations by combining collaborative filtering and content-based filtering scores.
- To measure the performance of the hybrid recommendation system, Root Mean Squared Error (RMSE) is calculated.

V. EXPERIMENTAL RESULTS

When examining whether the hybrid method is successful or not, it would be more accurate to compare it with the performance values of the Collaborative Filtering method and content-based filtering methods used in the hybrid method.

```
RMSE for Hybrid Recommendation: 0.5717793695737589
Accuracy: 0.4746317512274959
Precision: 0.00625
Recall: 0.4
    movieId                                 title
2         3               Grumpier Old Men (1995)
46       50           Usual Suspects, The (1995)
88      100                      City Hall (1996)
133     160                          Congo (1995)
141     169  Free Willy 2: The Adventure Home (1995)
158     187                     Party Girl (1995)
177     209            White Man's Burden (1995)
186     218              Boys on the Side (1995)
212     248                     Houseguest (1994)
231     269     My Crazy Life (Mi vida loca) (1993)
```

Fig. 5. Hybrid Method Results

TABLE I
RMSE VALUES OF METHODS

Method	RMSE Value	Increase Rate
Collaborative Method	0.9723	41.2%
Content-Based Method	0.8718	34.47%
Hybrid Method	0.5718	-

A. *Comparison With Collaborative Filtering*

First of all, it is necessary to look at the RMSE values compared to the Collaborative Filtering Method. In order to compare the performance of the Hybrid model, the value of the User-Based Method, which is slightly more successful in the experiments, was taken into account.

In this context, the User-Based RMSE value was 0.9723, while the RMSE value in the hybrid model was 0.5718. Lower RMSE value indicates better performance. When calculated, there was a 41.2 percent improvement.

B. *Comparison With Content Based Filtering*

When compared with the content-based filtering method, the RMSE value obtained thanks to the ”surprise” library can be used. The RMSE value obtained when only content-based filtering was used was 0.8718. In the hybrid method, RMSE values were calculated as 0.5718, as mentioned before. Considering these values, the development rate was 34.47 percent.

VI. DISCUSSION

Today, recommendation systems are used in a wide range of applications to provide users with easier access to personalized content. The two main types of recommendation systems, Collaborative Filtering and Content-Based Filtering, offer different advantages and limitations.

A. *Collaborative Filtering*

Collaborative Filtering allows users to provide recommendations using the preferences of other users with similar interests. This approach has the potential to offer users undiscovered content. However, it may face difficulties such as cold start issue. This issue occurs when new users or products that the recommendation algorithm has not yet interacted with are recommended.

B. *Content Based Filtering*

Content-Based Filtering recommends similar content based on users’ past preferences or profile information. This method tends to overcome the cold start problem and is effective in providing personalized recommendations to users. However, accuracy issues may arise if the user profile is incomplete or misleading.

C. *Hybrid Recommendation Systems*

Hybrid recommender systems combine collaborative filtering and content-based filtering methods, bringing together the advantages of both approaches. This can be more resilient to cold start issues and has the potential to provide more precise and personalized recommendations. For example, content-based filtering can take into account the user's personal preferences, while collaborative filtering recommends popular content.

Hybrid recommendation systems can improve recommendation accuracy by more comprehensively evaluating user behavior and content characteristics. However, given the complexity and difficulty of implementation of this approach, it requires a careful modeling and validation process.

As a result, which type of recommendation system to use varies depending on the application context and user behavior. While collaborative filtering and content-based filtering each offer unique advantages, hybrid recommender systems can combine these advantages to deliver a more powerful and flexible recommendation framework.

REFERENCES

[1] Yibo Wang, Mingming Wang, Wei Xu, "A Sentiment-Enhanced Hybrid Recommender System for Movie Recommendation: A Big Data Analytics Framework", Wireless Communications and Mobile Computing, vol. 2018, Article ID 8263704, 9 pages, 2018. https://doi.org/10.1155/2018/8263704

[2] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4: 19:1–19:19. ¡https://doi.org/10.1145/2827872¿

[3] Amin, S.A., Philips, J., Tabrizi, N. (2019). Current Trends in Collaborative Filtering Recommendation Systems. In: Xia, Y., Zhang, LJ. (eds) Services – SERVICES 2019. SERVICES 2019. Lecture Notes in Computer Science(), vol 11517. Springer, Cham.

[4] Upadhyaya, Prakash. (2023). Study of Mathematical Model for User-based Collaborative Filtering and Item-based Collaborative Filtering.

[5] K. Shah, A. Salunke, S. Dongare and K. Antala, "Recommender systems: An overview of different approaches to recommendations," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 2017, pp. 1-4, doi: 10.1109/ICIIECS.2017.8276172.

[6] Sun, X., Kong, F., Chen, H. (2005). Using Quantitative Association Rules in Collaborative Filtering. In: Fan, W., Wu, Z., Yang, J. (eds) Advances in Web-Age Information Management. WAIM 2005. Lecture Notes in Computer Science, vol 3739. Springer, Berlin, Heidelberg.

[7] Donghui Wang, Yanchun Liang, Dong Xu, Xiaoyue Feng, Renchu Guan, A content-based recommender system for computer science publications, Knowledge-Based Systems, Volume 157, 2018, Pages 1-9, ISSN 0950-7051,

[8] Burke, R. Hybrid Recommender Systems: Survey and Experiments. User Model User-Adap Inter 12, 331–370 (2002).

[9] E. Aslanian, M. Radmanesh and M. Jalili, "Hybrid Recommender Systems based on Content Feature Relationship," in IEEE Transactions on Industrial Informatics, doi: 10.1109/TII.2016.2631138.

[10] Lu, Z., Dou, Z., Lian, J., Xie, X., Yang, Q. (2015). Content-Based Collaborative Filtering for News Topic Recommendation. Proceedings of the AAAI Conference on Artificial Intelligence, 29(1).

[11] Koren, Y., Rendle, S., Bell, R. (2022). Advances in Collaborative Filtering. In: Ricci, F., Rokach, L., Shapira, B. (eds) Recommender Systems Handbook. Springer, New York, NY.

[12] Thakker, U., Patel, R., Shah, M. A comprehensive analysis on movie recommendation system employing collaborative filtering. Multimed Tools Appl 80, 28647–28672 (2021). https://doi.org/10.1007/s11042-021-10965-2