Recommender Systems Summative

Thomas Pettit   
Computer Science  
Durham UniversityDurham  
thomas.w.pettit@durham.ac.uk

# Introduction

## Domain of the Application

The domain of this application is movie/film recommendation. In particular, providing personalised recommendations to an active user of the system.

## Purpose and Aim of the Application

Movie and film streaming is a market that is growing very quickly, with an estimated market share of 60 billion dollars worldwide, and is predicted to reach 220 billion dollars by 2028 [2]. Therefore, it is in streaming companies’ interest to improve their service in order to maximise their portion of this market share. With the huge amount of data that companies such as Netflix and Amazon Prime Video now have on user ratings, genres and tags for each film on their platform, their ability to provide recommendations has increased vastly, and so recommender systems that harness this data are very valuable. The aim of this application is therefore to demonstrate how, using large amounts of data, recommendations can be provided to the active user, using content-based and collaborative-filtering approaches, that maximise both accuracy and diversity.

# Methods

## Data Description

The dataset used is the MovieLens 25 Million Dataset [1]. This dataset includes 5-star rating and free-text tagging activity from MovieLens, a movie recommendation service.

The dataset includes 25000095 ratings, and 1093360 tags across 62423 movies, from 162541 different users. In order to ensure that there is enough ratings data for training the recommender systems for each user, each user in the dataset has rated at least 20 different movies. In order to hide the identity of the users, only an ID can identify each user, which has no reference to their actual identity elsewhere. The dataset includes a Tag Genome. This is a data structure that contains the relevance of each tag to each movie, with each movie having a value for every tag in the genome. The tags indicate a range of features for a film, including: atmospheric, thought-provoking, realistic and more. This Tag Genome is represented by the files: genome-scores.csv and genome-tags.csv. The genome-scores.csv file contains the calculated relevance of each tag to each movie, and the genome-tags.csv file provides the description of each tag from each tag ID. The movies file contains the title and genres associated with each movie ID. The ratings file contains all the ratings given by users on the movies, as well as a timestamp indicating when the rating was left on the given film, represented in seconds since midnight Coordinated Universal Time (UTC) of January 1, 1970. The tags file contains all the tags given by users on films, as well as the timestamp that the user gave the tag for the corresponding film, also formatted in seconds since midnight Coordinated Universal Time (UTC) of January 1, 1970.

## Data Preparation and Feature Selection

The Genome itself was prepared by using a machine learning algorithm, trained upon user-contributed content such as tags, ratings and textual reviews, as described in [3]. The user reviews were web scraped from IMDB [11]. In the Genome, there were 1128 features (tags) selected for each movie, each with their own relevance score for every movie. These features were selected by looking at all 30,000 tags applied by users in the MovieLens dataset, then selecting only the tags that had been applied by 10 or more users, because otherwise the features would not be very useful and not lead to increases in performance for the systems. Then, as some of these remaining tags were misspellings of correct tags or too niche, these were filtered out using a tag quality metric developed by Sen et al [10]. This left 1570 features, including 442 of these being names of directors, which were removed, leaving 1128 features remaining. This whole feature selection process is described in [3]. All of these tags were used to train both the content-based and collaborative-filtering recommender systems. This is because the more tags, the more accurate the recommender systems can be, and so could lead to companies employing them earning a larger user base, as both the content-based and collaborative-filter systems can learn from a greater volume of data to learn user preferences from.

## Recommendation Algorithms

**Content-Based Filtering Approach**

Content-based approach works well in this application, as companies such as Netflix have access to a lot of data about each film they offer. This data, such as the Tag Genome with 1128 features used in this paper, provides lots of data that a content-based approach can exploit to make as accurate and diverse predictions as possible, by looking at all the different aspects of each movie in making its prediction for the active user. All the ratings and respective movie IDs for the active user are then collected, as well as the tag genome scores for each of these movies. The average rating for the active user is then calculated. In order to determine whether the active user liked or disliked the film given their rating of it, the average rating of the user is then subtracted from the rating given by the user for a particular film, as shown below:

Where rui represents the rating given by user u on item i, and represents the average rating of user u. However, this is only applied if the user has more than 5 ratings, as if the user only has 1 rating, the average will be the one rating they have given, and so this would cause their rating to be 0, not providing any information to the system, and with less than 5 ratings, there isn’t enough data available to accurately determine an average rating for the given user.

Then, the user profile is created by calculating the dot product between every tag’s relevance on every rated movie and the rating the user has given on every movie. The user profile therefore then represents the active users’ preferences for every one of the tags in the Tag Genome. Then, in order to make recommendation predictions, the dot product of the user profile and the movie Tag Genome scores for each of the possible movies is calculated. The sum of all the columns for every movie is then calculated, and this represents the prediction score for each movie. These values are then scaled using Min-Max Feature Scaling, using the Sklearn MinMaxScaler [5], to ensure all the prediction values are between 0 and 1. Then, the films that the active user has seen previously are removed. The dataframe is then sorted by the predicted score column in descending order, and the first 10 movies are returned. The prediction score is converted into a percentage, and then limited to only 1 decimal place. This makes for a more readable output for the active user. In order to make the predictions more explainable, the 10 tags with the highest relevancy values in the user profile are displayed, as well as the 5 tags with the highest relevancy scores for each of the top 10 predictions for the active user. An example of this can be shown below in Figure 1 on the next page.

**Collaborative-Filtering Approach**

Diagram, schematic

Description automatically generatedCollaborative-filter approach works well for this application. This is because companies such as Netflix have an ever-increasing userbase, which means they have access to lots of different people’s viewing habits, such as which movies they have given ratings. This increasing data source can be exploited by a collaborative-filtering model to learn which users are similar, and so make predictions for another user who is similar to a particular group. The dataset used in this paper has 162541 different users, a large number, and so also allows a collaborative-filtering model to make accurate and diverse predictions. In this approach, a model-based collaborative-filtering recommender system is created, based upon a Matrix-Factorisation algorithm, specifically SVD [4]. This approach works in the following way. The predicted rating for a movie i by user u, is given by the following formula:

Text

Description automatically generated with medium confidencewhere 𝜇 represents the average rating of all the movies in the dataset, bu represents the average rating given by user u, bi represents the average rating of item i, qi  represents the latent item vector and pu represents the latent user vector. Both the latent user and item vectors were set to 100 latent factors. During training, the SVD minimises the following regularized squared error:

Text, letter

Description automatically generatedThis is performed via stochastic gradient descent:

Where eui represents the difference between the actual rating given by user u on item i, and the predicted rating by the SVD. The model in this paper was trained using the learning rate, 𝛾, of 0.005, and the regularisation term, 𝜆, of 0.02.

As with the content-based approach, the top 5 tags of the recommended films are displayed in the table shown to the user, as shown below in Figure 2 on the next page. The user knows their own preferences, and so by displaying these tags, it explains to the user why they have been recommended the movies shown, as it aligns with their interests.

## Evaluation Methods

**Accuracy: RMSE**

RMSE was selected as the system needs to predict ratings close to what the actual user would rate the movie, as then it would be a good prediction. If the system predicts the user will really enjoy a film that they in reality dislike greatly, then the user will be put off the platform using this system. Thus, it is very important that errors between the actual and predicted ratings are low. As RMSE penalizes large errors highly, it was therefore selected to evaluate the accuracy of these models.In order to evaluate the accuracy of the models, Root-Mean Squared Error was used, using the sklearn metrics function mean\_squared error [6]. This follows the following formula:

**Accuracy: Content-Based Filtering Approach RMSE**

In this approach, 5 users are selected at random from the entire ratings dataset. All the ratings given by each of these 5 users are then collected. 75% of these ratings are then used to create the user profile, and 25% are used to test the algorithm. The user profile is created using the 75% split, and then the user profiles are then used to create the predictions for each of the films in the 25% split, in the same method as described in section II.C. However, unlike producing a match score measured in percentages, as shown in Figure 1, the predictions that were Min-Max Feature Scaled, were now multiplied by 5, to produce a rating between 0 and 5. The values were then rounded to the nearest 0.5, to be in the correct format to compare with the actual user ratings. Then, RMSE was calculated for each of these 5 users, using the predicted ratings for the films in the test set, and the actual ratings given by the users on the films in the test set. The 5 calculated RMSE values is then averaged, and the overall RMSE value for the content-based recommender system is created.

**Accuracy: Collaborative-Filtering Approach RMSE**

In order to measure the RMSE of this approach, the sklearn surprise accuracy library [7] was used. In this approach, in order to speed up the time taken to evaluate the model, only 7.5% of the ratings.csv data was used for testing. The model that was trained on the entire dataset was used to make predictions on this test set. The model’s predictions on this test set were then used, alongside the actual user ratings, using the RMSE formula, to calculate the model’s RMSE score.

**Diversity**

Diversity was the other metric used to evaluate the two approaches. Diversity determines how wide the spectrum of recommendations is. This was used in this paper as a recommender system that always recommends similar movies, which could quickly lead the user to become bored of this film genre, and so the recommender system would become obsolete. There are different methods of measuring diversity of recommendations. In this paper, it was measured via the proportion of common users between any 2 of the recommended films. If 2 items in the recommended set have a high proportion of common users, the recommended items are most likely very similar, and so the diversity of the recommender system is low, as described in [9]. This can be calculated via Cosine Similarity. The cosine similarity between every pair of predictions from the recommender system can be calculated, and the average value created. The Cosine Similarity was calculated using the following formula, where i and j are two different films:

This shows the average similarity between any 2 predictions in the top 10 recommendations provided by the recommender system.

Then, diversity can be calculated via the following formula:

where Sim10 is the average cosine similarity score between any 2 predictions in the top 10 provided by the recommender system. To use this metric on the 2 recommender systems in this paper, a user was first selected at random from the ratings dataset. Then, the top 10 recommendations were created for this user, and then the diversity metric was calculated following the method outlined above.

# implementation

## Input Interface

When the user first logs into the system, they are presented with the following menu:

The user can either login as an existing user in the dataset, or create a new user. However, if the user decides to create a new user, the collaborative filter approach for recommendations cannot be used, as it would require the whole SVD model to be retrained, which would be too time consuming. The user is warned of this issue, as shown in the screenshot above. If the user decides to login as existing user, they enter their user ID, and if they decide to create a new user, they are told their new user ID, and are logged in. Upon logging in, the user is presented with the following menu:

Text

Description automatically generatedText

Description automatically generatedHere, the user can view any ratings they have already made, add a new rating, edit an existing rating, or view their recommendations. Viewing ratings prints a dataframe to the command line showing the movie title and their rating for every film they have rated. Add a new rating allows the user to enter a movie ID, and then the rating for that film, as shown in the screenshot below:

However, this new rating is not taken into account for the collaborative-filtering recommendations, due to the same reason mentioned previously, whereby the model would need to be retrained, which is too time consuming.

Text

Description automatically generatedThe user can also choose to edit a rating they have previously made, which can be shown below in the screenshot:

The user can also select option 4 to view their recommended films. Upon pressing 4, they are then presented with a menu that allows the user to select either content-based or collaborative filter to use to create recommendations.

The user can also choose to logout of the system via option 5 of the menu.

## Output Interface

Both the recommender system approaches present the table of top 10 recommendations to the active user in the same way, with the movie ID, movie title, match score, and the top tags associated with the films. This can be seen below:

Figure 1 shows the output of the Content-Based recommendations. This approach states the top tags for the active user, which helps the user understand why they have been recommended the films they have been.

Figure 2 shows the output of the Collaborative-Filtering recommendations.

The tables are easy to read, and also have the advantage of displaying the top tags of each film, to help the active user determine which film out of these recommendations they may start watching.

# Evaluation

## Evaluation Metrics’ Results

**Accuracy: RMSE**

|  |  |
| --- | --- |
| **Content-Based Approach** | **Collaborative-Filter Approach** |
| 1.977 | 0.698 |

**Diversity**

|  |  |
| --- | --- |
| **Content-Based Approach** | **Collaborative-Filter Approach** |
| 0.703 | 0.955 |

## RS Techniques’ Comparison

The collaborative-filter approach has both lower Root Mean Squared Error, and a higher diversity score. This means that, when tested upon the testing data, the model predicts a rating by users on a particular film more closely to the actual user rating than the content-based approach. The collaborative-filter approach also has a higher diversity score. This means that it provides a wider spectrum of recommended films. This is particularly beneficial in this case, as it introduces the active user to a wider range of potential films to watch, outside of their usual watching habits. In industry, this would encourage the user to stick to the platform with this recommender system, as it recommends new and different films, thus making this an important metric. Therefore, due to the fact that Collaborative-Filter approach has both a lower RMSE score, and a higher diversity score, it is the better technique in this case, despite the time taken to train the model. However, the Collaborative-Filtering approach is not as explainable as the Content-Based approach. Thus, if the movie recommender system had to be more transparent, sacrificing accuracy and a small loss in diversity score may be worth it. The Collaborative-Filtering approach also suffers from the fact that new users or ratings or edited ratings would require the model to be retrained, which is a very time consuming process. Thus, if these weaknesses needed to be solved, then the small drawbacks in RMSE and diversity associated with the Content-Based approach may be bearable.

# conclusion

In conclusion, both content-based and collaborative-filtering approaches can be used to produce recommendations for the active user of a recommendation system. The Collaborative-Filtering approach performs best on both the evaluation metrics, however it’s practical use for when new users or ratings are created or users’ ratings are edited, is limited, as the SVD model would need to be retrained on the updated dataset, which is a very time-consuming process. Thus, if small sacrifices in RMSE and Diversity are allowable, then the Content-Based filter would potentially be best.

##### References

1. 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>
2. Research ltd and Markets. Global video streaming market size, share amp; trends analysis.
3. Jesse Vig, Shilad Sen, and John Riedl. 2012. The Tag Genome: Encoding Community Knowledge to Support Novel Interaction. ACM Trans. Interact. Intell. Syst. 2, 3: 13:1–13:44. <https://doi.org/10.1145/2362394.2362395>
4. <https://surprise.readthedocs.io/en/stable/matrix_factorization.html#unbiased-note>
5. <https://scikitlearn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>
6. https://scikitlearn.org/stable/modules/generated/sklearn.metrics.mean\_squared\_error.html
7. <https://surprise.readthedocs.io/en/stable/accuracy.html>
8. <https://medium.com/fnplus/deep-learning-22d971e3ad43>
9. <https://eugeneyan.com/writing/serendipity-and-accuracy-in-recommender-systems/#:~:text=Serendipity%20is%20measured%20as%20unexpectedness,the%20user%20interacted%20with%20i%20>.
10. SEN, S., HARPER, F. M., LAPITZ, A., AND RIEDL, J. 2007. The quest for quality tags. In Proceedings of the International ACM Conference on Supporting Group Work (GROUP’07). ACM, New York, NY, 361–370.
11. https://www.imdb.com/