

Movie Recommendation System Using Collaborative Filtering

Hitarthi Gandhi
University of Regina
Regina, Canada
hgw608@uregina.ca

Dr. Abdul Bias
University of Regina
Regina, Canada
Abdul.Bias@uregina.ca

Abstract— Recommendation system is a predicting system and which is used for filtering data where it tries to assumes the preferences of the individual and gives a recommendation based on these preferences. These recommendations are users to items and items to users, and sometimes one user to another user too. There are ample of information available on the internet for hotel, movies, books article, songs, video and many more. One of the simple methods is content based filtering because it suggest items based on the user's history, so the model provides a very limited degree of novelty. In this project, I used a collaborative filtering technique to recommend movies and a K-NN algorithm to prove the effectiveness of the model. By using this method, our recommendation system suggests the top 5 rated movies. In this proposed solution, we used a threshold value for the similarity between movies, and based on this threshold value results are changing.

Keyword: Machine learning, Movie Recommendation System, Collaborative Filtering, K-nearest neighbour Algorithm, Data analysis

I. INTRODUCTION

In this modern world, it's hard to think of our life without the internet because nowadays people are using the internet for shopping, to watch movies, to listen to music or to interact with people. The activities of a large number of individuals - Internet clients - comes with lots of possibility of data from thousands, millions, and billions of people. It is an opportunity for individuals who wants to research this information. But this data is very big so people are not able to handle this complicated and tedious computation on it. To solve this difficulty machine learning algorithms and methods are used. Machine learning algorithms are statistical methods and it is mostly based on regression and classification. Moreover, in advance developers creates a recom-

mendation algorithm to work with a given set of databases. Such systems give great opportunities.[5]

Recommender Systems (RSs) are the engine that knows about the preference of the individuals and gives suggestion which is most suitable to the user [1]. Nowadays, recommendation engine becoming more and more popular in serval areas such as hotels, movies, music, books article and many more. As there is a huge number of data, to find your relevant information is difficult and also it takes lots of time. Recommendation Systems filters the data which is a match to the user's requirements and It assists individuals to figure out suitable categorize about an item at the proper time. Mostly, Collaborative based filtering technique, Content based filtering technique , Hybrid filtering method, and Knowledge-based filtering methods are used to build a recommendation system [6].

Recommendation systems are depends on user ratings and textual reviews of the items. Non-personalized recommendation system displays the item that are popular among the people in general. Content-based recommendation system uses the information given by the individual such as text-based review, ratings and based on item system will generate the user profile and then the item is suggested by recommender system using this item description and user profile. Another method is a hybrid recommendation system. This method is a mixture of collaborative and content-based filtering algorithm. In this, a prediction is performed individually and then results of both methods combine to suggest recommendation [3]. This technique is use to get better result and overcome the limitation of single approach. Knowledge-based recommen-

dition system gives a recommendation based on users' specific queries. It might give an example of how results should look like. This framework searches through its information of things and gives similar results [8].

In the personalized recommendation engine, collaborative filtering method is the most commonly used recommendation technology [9]. In a simple word, a collaborative filtering method filters the user's information by collecting other similar user's information. Collaborative filtering suggests the item which is most rated by the user. CF is one of the recommendation techniques which gathers and analyses large data sets including user's activities, opinions or preferences.[10] The traditional collaborative filtering technique mainly has two types Item-Item based recommendation and User to User based recommendation and item to User based collaborative filtering technique is depending on the hypothesis, two individuals who have the same interests and may be interested in the same item. Item-based collaborative filtering method, a user may be choosing the item which is similar to the item they choose before [9].

To improve the result of a recommendation system, I used K-Nearest Neighbour (KNN) algorithm to build a movie recommendation system based on a collaborative filtering method. KNN is a well known and very simple machine learning algorithm. The main purpose of KNN algorithm relies on item feature similarity. KNN uses the Euclidean equation formula to measure the closest distance. KNN algorithm computes the neighboring distance between target film and all other movies in the database, and determined the movies based on the number of K, and suggest top K similar movies.

This paper has five sections. First section, a brief introduction to this work. Second section, a few relevant and important, recent research done in the area of recommender systems will be discussed. Third section gives details about this proposed system and its methodology. Forth section explains the experiment and result from analysis. Finally, fifth section gives the conclusion of this work.

II. RELATED WORK

In recent years, there have been many types of recommendation systems that are developed by a developer. Different types of systems, different types of approaches, different types of machine learning algorithms are used to build a recommendation system.

In one of the research papers, I read that, Tapestry was the first human being who developed a collaborative filtering technique. For this system, he used implicit and explicit data of the user's rating. And overall, using this collaborative filtering method he suggested rating to an intended user.[12] Authors in [3], developed a movie recommendation system for this they used a collaborative filtering method for collecting user ratings and use K-means algorithm to rank the movies based on ratings. On the other hand, authors Parvatikar and Joshi [2] proposed a book recommendation system, and to built this system various techniques are used. Mainly Item - based collaborative filtering technique is used.

Moreover, in the authors[8]. implemented a system which is combination of both User-based and Item-based collaborative filtering techniques. In this framework, they used the nearest neighbor machine learning algorithm and at the same time, they also developed a new algorithm that unifies the user-based and Item-based recommendation. Wang and Shao in [10] proposed a personalized system of web pages, in this clustering time-framed navigation sessions, are used to develop a model. Hierarchical Bisecting Medoids algorithm is used for these navigation sessions and analyzed those navigation sessions with associate-mining for the prediction in later. Through, their system improves the accuracy of recommendation.

III. PROPOSED FRAMEWORK

A. Collaborative filtering Technique

In this proposed, a movie recommendation engine used the collaborative filtering method which is more efficient to use and gives more accurate results. There are ample of movie recommendation website available in the market which recommends

content to a user based on their interest and it also shows the rating of that particular movie. Such as Netflix, Amazon prime, and so on.

Whenever people are planning things, then it's a human inclination that we first go through the web to search for relevant information and try to get a most suitable option from the various available option for that thing. Collaborative filtering technique makes automatic suggestions about the interest of users by collecting other users' preferences. A major appeal of collaborative filtering is that it is domain free. Collaborative based movie recommendation system mainly focuses on ratings given by the user to suggest another movie. In this section, we described both methods of collaborative filtering.

1) User-Based Collaborative filtering:

In the field of personalized system, user-based preferences are very general. In this method, system is estimate that the individual will choose the movie that are already chosen by individual with whom they have similar interest [4]. This method is memory based collaborative filtering and which is works on user-item similarity matrix. In this approach system checks the past user history and assumes that rating is not randomly given. Below figure demonstrate how user-based collaborative filtering works.

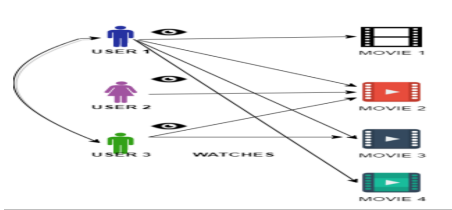


Fig. 1: Demonstration of User-based collaborative filtering[3]

2) Item-Based Collaborative filtering:

Item-based collaborative filtering predict human being will choose item which are related to the item that are already rated by the user. Item-based CF technique is model based ap-

proach for recommendation. In this Item-Item based collaborative technique we have used movie rating that are mostly same as rating of another movie, given by individuals. For this, we used item similarity weight matrix for items similarity weights. And after that, K most suitable items are selecting by the system, which is most rated and suggest these movies to the individuals. Below figure demonstrate how Item-based collaborative filtering works.

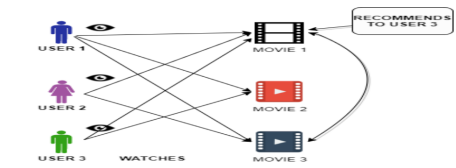


Fig. 2: Demonstration of Item-based collaborative filtering[3]

B. K Nearest Neighbour Algorithm

K nearest neighbour machine learning algorithm is useful algorithm in recommendation system for its low calculation time. K nearest neighbour algorithm is significantly important for user-item similarity matrix. In this movie recommendation system, to measure the distance between target movie from all other movies we use K- nearest machine learning algorithm.

After that it ranks the top K closet movies based on the distance matrix and which suggest people the top K-nearest neighbour. Finally, movie recommendation system recommends list of most similar movies. KNN analyses the pivot table and set the threshold value for similarity and applying threshold value in algorithm.

Above equation is used to measure the distance in K nearest neighbour algorithm.

$$Euclidiandistance = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

$$Manhattandistance = \sum_{i=1}^n |x_i - y_i| \quad (2)$$

$$Mankiwoskidistance = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p} \quad (3)$$

IV. EXPERIMENTAL EVALUATION

To build this movie recommendation system, here I used the IMDB movie rating database from Kaggle. There are more than 50,000 different movies with their rating from the year 2006. Dataset has two CSV files. One for movies and another one is for ratings. Movie dataset containing columns movie-id, genres, and title. While rating dataset containing user-id, movie-id, rating, count, and timestamps. From this dataset, I have used 10000 movie ratings. Ratings are given on the scales of 1 to 5.

After gathering the data from the internet, we need to clean that data because online data are noisy. The dataset containing lots of blank values and duplicate values. We removed this row from the dataset for easy processing. The below figure[4] shows the ratings of the movie dataset and The figure[5] shows the top 10 most rated movies from the movie's dataset.

In the data analysis task, various python libraries are introduced to building the module. Each movie is in the scale of 1 to 5 and the below figure shows the rating of movies. For each movie, we calculate the total count of ratings and also count that total number of movie ratings. As collaborative filtering methods works on user-item similarity matrix we create a matrix for movie names and movie ratings to get the proper user. There are many techniques available in market to find similarities or create a similarity matrix but here I used the popularity

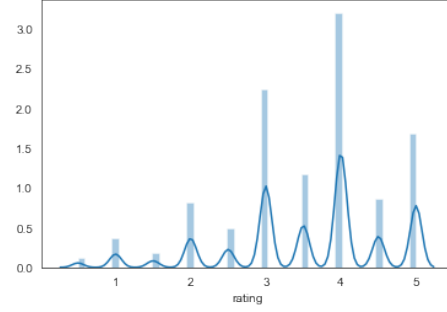


Fig. 3: Ratings

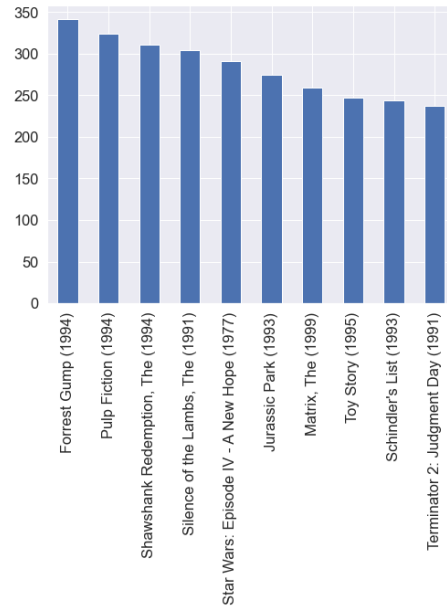


Fig. 4: Top 10 Most Rated Movies

threshold value. That means we set one threshold value for the recommendation, then if the total movie rating count is more than that threshold value then KNN considers it as a most rated movie.

We created a pivot table to create user-item similarity matrix. A pivot table is used to summarise, sort or to store average data in a table. In our case, the pivot table creates a matrix of each movie title and movie ratings and we took user-id as a column. As we discussed, K nearest neighbor algorithm

is one of the laziest machine learning methods. KNN algorithm measures the distance between two movies. Also, it creates a similarity matrix for a total number of ratings between count total number of ratings and based on popularity threshold value gives top K similar movie recommendation.

V. RESULT AND ANALYSIS

As explained, K-nearest neighbour algorithm examine the pivot table and evaluate the similarity using distance parameter. KNN examine the pivot table and the distance parameter measure distance or similarity between the target movie from all other movie. The pivot table containing many zero values for movies*users and this data is known as sparse data. We used scikit learn scipy.sparse library to overcome the sparse data size.

Below table is total rating count of our movie recommendation system.

TABLE I: Total rating counts for Particular movie

	Movie Title	TotalRatingCount
0	Round Midnight (1986)	38
1	Mitchell (1975)	18
2	Dangerous Minds (1995)	42
3	Dragonheart (1996)	48
3	Nutty Professor, The (1996)	54

To predict the movie, we need movie which is at least rated 30 times by the user. Because we select 30 as our popularity threshold value. Moreover, K nearest neighbour is used to give top 5 K-nearest neighbour.

Recommendations for 101 Dalmatians (1996):

- 1: 101 Dalmatians (One Hundred and One Dalmatians) (1961), with distance of 0.6122321188029379:
- 2: Dragonheart (1996), with distance of 0.6357671086605631:
- 3: Space Jam (1996), with distance of 0.6498805808437947:
- 4: James and the Giant Peach (1996), with distance of 0.6512329095368182:
- 5: Nutty Professor, The (1996), with distance of 0.6709869176898213:

Fig. 5: Top 5 Recommended movies

After implementation, we analyzed some common issues of KNN based collaborative filtering. In

analyzed a pivot matrix for the data and we found so many zero values in pivot matrix and it occupied lots of space. In this over main aim to give movie recommendation based on threshold value. So, in this system whenever user will add new movie into the system and it doesn't have any rating given by individuals then it cannot be recommended. This problem is known as item cold start. Whenever the users and items in the metrics increases, it is hard to manage the dataset, because there is more sparse data almost 90 percentage so its wastage of such a big space so database recommend mostly numeric value and it creates a scalability problem

VI. CONCLUSION

In this proposed paper, we developed a movie recommendation engine. This system is implemented based on collaborative filtering technique and we used K-Nearest Neighbour machine learning algorithm. Then we analyze the distance between the neighbours and proposed an algorithm which gives recommendation of movies. Threshold value counts the ratings of movies. However, this system uses very small amount of data so results are not accurate. So, it required more work to improve the accuracy of this system.

ACKNOWLEDGMENT

Professor Abdul Bias gives as knowledge from his experience and gives as an opportunity to improve our knowledge. He greatly assisted us to prepare this research paper and I would like to show how much I am thankful to him for their help without him it can't be possible.

REFERENCES

- [1] S. Kaur, R. K. Challa, N. Kumar, S. Solanki, S. Sharma and K. Kaur, "Recommendation generation using typicality based collaborative filtering," 20177th International Conference on Cloud Computing, Data Science Engineering - Confluence, Noida, 2017, pp. 210-215.ng".
- [2] S. Parvatikar and B. Joshi, "Online book recommendation system by using collaborative filtering and association mining", in 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2015, pp. 1-4.

- [3] C. M. Wu, D. Garg and U. Bhandary, "Movie Recommendation System Using Collaborative Filtering," 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2018, pp. 11-15, doi: 10.1109/ICSESS.2018.8663822.
- [4] M. Gupta, A. Thakkar, Aashish, V. Gupta and D. P. S. Rathore, "Movie Recommender System Using Collaborative Filtering," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 415-420, doi: 10.1109/ICESC48915.2020.9155879.
- [5] A. Nawrocka, A. Kot and M. Nawrocki, "Application of machine learning in recommendation systems," 2018 19th International Carpathian Control Conference (ICCC), Szilvasvarad, 2018, pp. 328-331, doi: 10.1109/CarpathianCC.2018.8399650.
- [6] Ruihai Dong, Barry Smyth, "User-based Opinion-based Recommendation", Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17), 2017.
- [7] M. T. Himel, M. N. Uddin, M. A. Hossain and Y. M. Jang, "Weight based movie recommendation system using K-means algorithm," 2017 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, 2017, pp. 1302-1306. doi: 10.1109/ICTC.2017.8190928
- [8] D. Pathak, S. Matharia, and C. N. S. Murthy, "ORBIT: Hybrid movie recommendation engine," In 2013 IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology (ICECCN), pp. 19-24, IEEE, Mar, 2013.
- [9] S. Zheng, Y. Shen, G. Zhang and Y. Gao, "A collaborative filtering recommendation algorithm based on dynamic and reliable neighbors," 2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2015, pp. 690-693, doi: 10.1109/ICSESS.2015.7339151.
- [10] R. Obeidat, R. Duwairi and A. Al-Aiad, "A Collaborative Recommendation System for Online Courses Recommendations," 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), Istanbul, Turkey, 2019, pp. 49-54. doi: 10.1109/Deep-ML.2019.00018
- [11] Okfalisa, I. Gazalba, Mustakim and N. G. I. Reza, "Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification," 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, 2017, pp. 294-298, doi: 10.1109/ICITISEE.2017.8285514.
- [12] M. K. Kharita, A. Kumar and P. Singh, "Item-Based Collaborative Filtering in Movie Recommendation in Real time," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 340-342. doi: 10.1109/ICSCCC.2018.8703362