Content-Based Movie Recommendation System Using Genre Correlation

SRS Reddy, Sravani Nalluri, Subramanyam Kunisetti, S. Ashok and B. Venkatesh

Abstract A recommendation system is a system that provides suggestions to users for certain resources like books, movies, songs, etc., based on some data set. Movie recommendation systems usually predict what movies a user will like based on the attributes present in previously liked movies. Such recommendation systems are beneficial for organizations that collect data from large amounts of customers, and wish to effectively provide the best suggestions possible. A lot of factors can be considered while designing a movie recommendation system like the genre of the movie, actors present in it or even the director of the movie. The systems can recommend movies based on one or a combination of two or more attributes. In this paper, the recommendation system has been built on the type of genres that the user might prefer to watch. The approach adopted to do so is content-based filtering using genre correlation. The dataset used for the system is Movie Lens dataset. The data analysis tool used is R.

1 Introduction

In this age of the Internet, the quantity of data transactions that happen every minute has increased exponentially. The huge amount of data has dramatically increased with the number of users on the Internet. However, not all the data available on the Internet is of use or provides satisfactory results to the users. Data in such huge volumes often turns out to be inconsistent and without proper processing of this information, it gets wasted. In such cases, users have to run their search multiple times before they finally obtain what they were originally looking for. To solve this problem, researchers have come up with recommendation systems. A recommendation system provides relevant information to the users by taking into account their past preferences. Data is filtered and personally customized as per the user requirements. With more and more data available on the Internet, recommendation systems

S. Reddy (✉) · S. Nalluri · S. Kunisetti · S. Ashok · B. Venkatesh
VIT University, Vellore, Tamilnadu, India
e-mail: janani333333@gmail.com

have become really popular, due to their effectiveness in providing information in a short time-span. Recommender systems have been developed in various areas such as music, movies, news, and products in general. In today's age, a majority of organizations implement recommendation systems for fulfilling customer requirements. LinkedIn, Amazon, and Netflix are just a few to name. LinkedIn recommends relevant connections of the people the user might know among the millions that are subscribed on the portal. This way, the user does not have to run extensive searches for people manually. Amazon recommendation systems work such that they suggest correlated items that the customers can purchase. If a certain customer prefers buying books from the shopping portal, Amazon provides suggestions related to any new arrivals in previously preferred categories. In a very similar way, Netflix takes into account the types of shows that a customer watches, and provides recommendations similar to those. By the method in which recommendation systems work, they can be broadly classified into three categories—Content-based, Collaborative and Hybrid approach. A content-based recommendation system considers the user's past behavior and identifies patterns in them to recommend to recommend items that are similar to them. Collaborative filtering analyses the user's previous experiences and ratings and correlates it with other users. Based on the ones that have the most similarity, recommendations are made. Both content-based- and collaborative-based filtering have their own limitations. To overcome this, researchers suggested a hybrid approach which would combine the advantages of both the methods. This paper suggests a content-based recommendation system that utilizes genre correlation. The dataset used for this purpose is a Movie Lens dataset containing 9126 movies which are classified according to genres. There are a total of 11 genres. The ratings for these moves have been collected from 671 users. By taking into account the movies which received high ratings from the users, movies containing similar genres are recommended to them.

2 Background

Recommender systems are broadly classified into three types—collaborative filtering systems, content-based filtering systems, and hybrid systems [3]. Collaborative systems utilize inputs from various users and run various comparisons on these inputs [3]. They build models from the past behavior of the users [1]. Movie recommendation systems, for example, utilize the ratings of users for various movies [2], and attempt to find other like-minded users, and recommend movies they have rated well [3]. Collaborative filtering systems have two approaches—memory-based approaches and model-based approaches [3]. Memory-based approaches continuously analyze user data in order to make recommendations [3]. As they utilize the user ratings, they gradually improve in accuracy over time [3]. They are domain-independent and do not require content analysis [3]. Model-based approaches develop a model of a user's behavior and then use certain parameters to predict future behavior [3]. The use of partitioning-based algorithms also leads to better scalability and accuracy [3].

Content-based filtering systems analyze documents or preferences given by a particular user, and attempt to build a model around this data [3]. They make use of a user's particular interests and attempt to match a user's profile to the attributes possessed by the various content objects to be recommended [3]. They have the added disadvantage of requiring enough data to build a reliable classifier [1]. Content-based filtering systems are divided into three methods—wrapper methods, filter methods, and embedded methods [3]. Wrapper methods divide the features into subsets, run analysis on these subsets and then evaluate which of these subsets seems the most promising [3]. Filter methods use heuristic methods to rate features on their content [3]. Both these methods are independent of the algorithms used. In contrast, embedded methods are coupled with the algorithm used—feature selection is performed during the training phase [3]. Hybrid systems combine collaborative and content-based filtering systems, in order to optimize the recommender systems, and reduce the drawbacks present in each of the two methods [3]. Thus, it tries to stretch the benefits of one method to compensate for the disadvantages of the other [3]. There are three types of hybrid systems—weighted hybrid, mixed hybrid, and cross-source hybrid [3]. In weighted hybrid systems, a score is maintained for each object, finding the weighted sum with respect to the various context sources [3]. These are given different weights based on a user's preferences [3]. In mixed hybrid approaches, each source is used to rank the various items, and the top few items from each rank list are picked [3]. Cross-source hybrid methods recommend items that appear in multiple context sources [3]. These methods work on the principle that the more sources an item appears in, the more important the item [3]. Wakil et al. attempted to improve their recommendation system by filtering using emotions [4]. When a user watches a certain type of movie, certain emotions are triggered from within them [4]. In the same way, the emotions of a user can trigger the need to watch a certain type of movie [4]. They recognized that traditional user profiles do not take into account the user's emotional status, and designed an algorithm that utilizes emotion determination [4]. It analyses a color sequence chosen by the user in accordance with his emotions to determine current emotional state of the user [4]. Debnath et al. proposed a hybrid recommendation system that utilizes feature weighting [5]. They determined the importance of various features to each user, and accordingly assigned weights to these features [5]. They then found the weighted sum in order to predict which items would further interest the user [5].

3 Recommendation System Using Content-Based Filtering

The approach used for building the recommendation system is content-based filtering. As discussed earlier, content-based filtering analyses user's past behavior and recommends items similar to it based on the parameters considered. This aims at recommending movies to users based on similarity of genres. If a user has rated high for a certain movie, other movies containing similar genres are recommended by the system. The dataset used in for this purpose is subdivided into two sections.

One section contains the list of movies along with the genres that they have been categorized under. The other part of the dataset contains a list of ratings of movies that have been rated by the user on a scale of 1–5, with 5 being the highest. First, a combined dataset of movies, genres and their ratings has to be constructed for correlating genres with the ratings. For the sake of simplicity, the ratings have been converted to binary values. If the rating given by a particular user is greater than 3, it receives a value of 1, otherwise it receives a value of −1. The genres are also segregated in a binary format, maintaining a consistent approach. Out of the set of 11 genres that are present in total, if a movie has a certain genre, it receives the value of 1. If the genre is not present in the movie, it receives a value of 0. The user profile matrix provides a combined effect of the genres and ratings by computing the dot product of the genre and the ratings matrix. Again for the sake of consistency, a binary format is adopted. If the dot product is a negative value, 0 is assigned to it. For a positive value, 1 is assigned to it. After obtaining a dot product matrix of all the movies, a similarity measure is calculated by computing the least distance between the user under consideration and the others. The values which have the least deviation with respect to the current user's preferences are the ones that are recommended by the system. The algorithm adopted for building the recommendation system is given below:

Algorithm Step 1. Construct a data frame of the genre dataset with movie ID as the rows and genres as columns separated by pipeline character.

Step 2. Make a list of all the genres that are available in the dataset.

Step 3. Iterate through the previously made genre data frame. If a genre is present in a movie, value of 1 is assigned to the genre matrix.

Step 4. Read the ratings sheet and construct a ratings matrix which assigns 1 for movies which has rating more than 3 and −1 for movies which has ratings less than or equal to 3.

Step 5. Calculate the dot product of the two matrices—genre matrix and ratings matrix. This is the result matrix

Step 6. Convert the result matrix to a binary format. For a negative dot product value, assign 0, else assign a value of 1.

Step 7. Calculate the Euclidian distance between the current user and other users.

Step 8. Retain the rows which have the minimum distance. These are the recommended movies for the current user.

Offloading is a method of transferring resource-intensive application from portable device to remote server by considering different parameters. Offloading mechanisms involves three tasks before it get executed. They are partitioning, profiling, offloading decision.

4 Simulation Results

The genre matrix constructed with rows containing movies and genres separated by columns. There are a total of 11 genres in the dataset (Fig. 1).

The ratings matrix for each user corresponding to the movie ID is converted to a binary format. Every user has rated one or more than one movie (Fig. 2).

Using the genres matrix and ratings matrix, the result matrix is computed which is the dot product of the previous two matrices. The result is further converted in a binary format in Fig. 3. If the value of the dot product is more than 0, 1 is assigned to that cell otherwise 0 is assigned.

	1	2	3	4
1	Adventure	Animation	Children	Comedy
2	Adventure	Children	Fantasy	
3	Comedy	Romance		
4	Comedy	Drama	Romance	
5	Comedy			
6	Action	Crime	Thriller	
7	Comedy	Romance		
8	Adventure	Children		
9	Action			
10	Action	Adventure	Thriller	

Fig. 1 Genre matrix

	userId	movieId	rating
1	1	31	−1
2	1	1029	−1
3	1	1061	−1
4	1	1129	−1
5	1	1172	1
6	1	1263	−1
7	1	1287	−1
8	1	1293	−1
9	1	1339	1
10	1	1343	−1

Fig. 2 Ratings matrix

	col1	col2	col3	col4	col5	col6
1	0	0	1	1	1	1
2	0	0	1	1	1	0
3	0	0	0	1	1	0
4	0	1	0	1	1	0
5	0	0	1	1	1	0
6	0	0	1	1	1	1
7	1	0	1	1	1	1
8	0	1	1	1	1	1
9	1	0	0	1	1	1
10	1	0	1	1	1	0

Fig. 3 Result matrix

19	3
20	3
21	2.645751
22	3.316625

19	19	Ace Ventura: When Nature Calls (1995)	Comedy
20	20	Money Train (1995)	Action|Comedy|Crime|Drama|Thriller
21	21	Get Shorty (1995)	Comedy|Crime|Thriller
22	22	Copycat (1995)	Crime|Drama|Horror|Mystery|Thriller

Fig. 4 Euclidean distance

After computing the result matrix, the Euclidean distances with respect to the other users are obtained and the ones having the minimum value is recommended as represented in Fig. 4.

Figures 5 and 6 shows the output of the various movies that have been recommended to the users based on their previous behavioral patterns.

5 Conclusion and Future Work

The recommendation system implemented in this paper aims at providing movie recommendation based on the genres of the movies. If a user highly rates a movie of a particular genre, movies containing similar genres will be recommended to him. Recommendation systems are widely used in today's era of Web 2.0 for searching for reliable and relevant information. While simple recommendation systems recommend users based on a few parameters, complex ones take many parameters into consideration. By implementing machine learning in recommender systems, intel-

```
 1                                    Get Shorty (1995)
 2                                         Hush (1998)
 3                               Sleeping Beauty (1959)
 4                          Children of the Corn (1984)
 5                                     RoboCop 2 (1990)
 6 Iron Monkey (Siu nin wong Fei-hung ji: Tit Ma Lau) (1993)
 7                                     Spy Game (2001)
 8                     The Count of Monte Cristo (2002)
 9                          They Were Expendable (1945)
10                                       Havana (1990)
11                      Tunnel, The (Tunnel, Der) (2001)
12                             Brown Bunny, The (2003)
13                                  Jacket, The (2005)
14                                Vincent & Theo (1990)
15                           Mr. Bean's Holiday (2007)
16                                   Rescue Dawn (2006)
17                                 Single Man, A (2009)
```

Fig. 5 User 1 recommendations

Fig. 6 User 2
recommendations

```
1                     Sleeping Beauty (1959)
2                Children of the Corn (1984)
3                            Spy Game (2001)
4 The Count of Monte Cristo (2002)
5                              Havana (1990)
```

ligent recommendations can be made for customers. Given the potential of such systems, they have a huge commercial value. Several MNCs have been exploiting the potential of recommendation system to lure customers into using their products. This also impacts greatly on the field of data mining and web mining.

Mobile cloud computing (mcc) is able to save energy, improve application and experience of the users. All frameworks mentioned above have their own benefits and issues but still not up to level to address all issues related to security, energy and user experience. Security issues are key problem in mcc, they need to be focused more compare to other issues.

References

1. Ghuli, P., Ghosh, A., Shettar, R.: A collaborative filtering recommendation engine in a distributed environment. In: 2014 International Conference on Contemporary Computing and Informatics (IC3I). IEEE (2014)
2. Zhao, L., et al.: Matrix factorization + for movie recommendation. In: IJCAI (2016)
3. Bhatt, B.: A review paper on machine learning based recommendation system. Int. J. Eng. Dev. Res. (2014)
4. Wakil, K., et al.: Improving web movie recommender system based on emotions. (IJACSA) Int. J. Adv. Comput. Sci. Appl. **6**(2) (2015)
5. Debnath, S., Ganguly, N., Mitra, P.: Feature weighting in content based recommendation system using social network analysis. In: Proceedings of the 17th International Conference on World Wide Web. ACM (2008)