Movie Recommendation Systems through Genre Correlation-Based Content and Collaborative Filtering

Goshika Swapna[1], Velamala Venkataramana[2], Somireddy Spandana[3]

[1,2,3]Assistant Professor, Department of CSE, Malla Reddy Engineering College and Management Sciences, Hyderabad, Telangana.

Abstract

Recommendation systems play a pivotal role in suggesting resources such as books, movies, songs, and more to users based on data analysis. Movie recommendation systems, in particular, predict a user's preferences for movies by evaluating attributes found in their previously favored films. These systems are invaluable for organizations amassing data from numerous customers, aiming to deliver optimal suggestions. Various factors can influence the design of a movie recommendation system, including genre, actors, and directors. Recommendations can be made based on one attribute or a combination of multiple attributes. This paper presents a recommendation system that focuses on users' preferred movie genres. The approach employs content-based and collaborative-based filtering using genre correlation and utilizes the Movie Lens dataset.

Keywords: Recommender System, Clustering, Random Forest, Recommendation, RMSE, KNN, Softmax Regression, SVD, Genre Correlation.

I. INTRODUCTION

Recommender Systems has emerged as important factor for recent online shopping sites. As they make huge increase on the products sales, to find current trends, promotion opportunities for sellers. For users it helps to solve information overload, personalized recommendation, and finds new things in many online shopping sites. Recommender systems have been using Collaborative filtering as it0 s primal method. Profound work has been done to improve accuracy of recommender system. The basic necessity of recommender system is accuracy and speed to predict [1]. Recommendation systems or recommendation engines are a particular form of data filtering system that try to recommend products or any other information that are of interest by users. At starting, recommendation technology was relatively unrefined. It just recommended dissimilar products which other users had purchased, but the technology has become more sophisticated and is now an integral part of many online retailer0 s economic models. Complex algorithms have been used to analyze vast amount of data and determine what products the potential customers want to buy based on their shopping choices that a user makes. The user based approach enables the user to personalize the recommendations based on their taste [2], hence the system will generate more apt user centric recommendation. Personalization enables user to overcome data overload and can make searching more efficient and less time consuming [3]. Many websites have implemented their version of recommendation system but lack accuracy to some extent. Categorical jump in recommendation have also been detected where some recommended products have no relation with the searched products, thus making the recommendations erroneous. This system now only focuses movies as a product and for the future works all other categories of products are planned to be included thus making system much diverse in terms of products accommodated and products recommended [4]. Below figure (Fig:1) shows the design of the system.

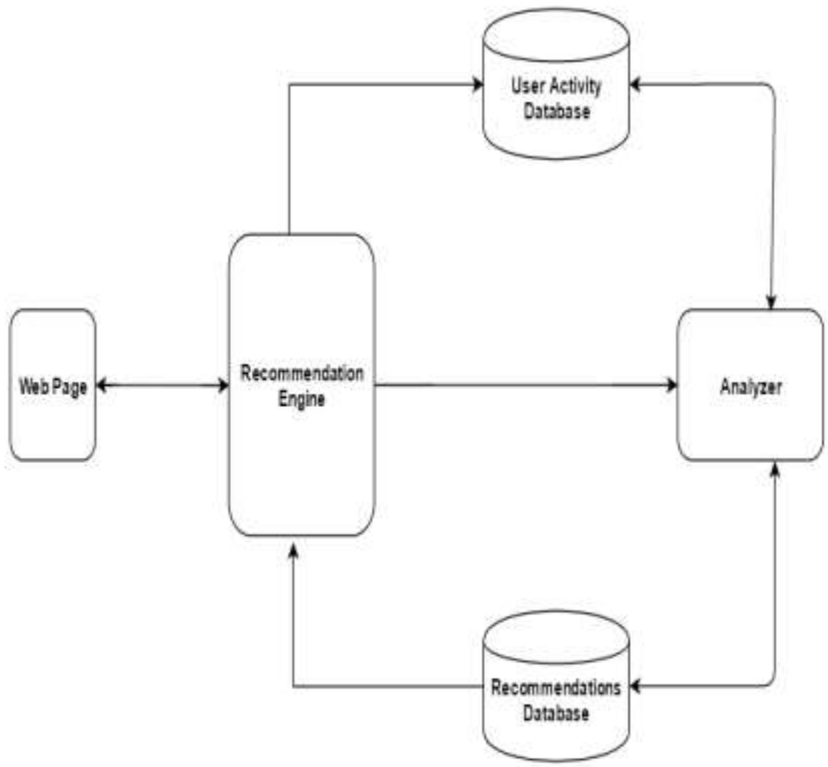

Fig. 1. System Design

Users interact with the system through a web page. Recommendation engine stores user activity details to a database. Depending upon the users activity, the analyzer groups the users based on their ratings given to a product. Groups created will have similar mindset users with similar tastes. Generated recommendation are provided back to the users. Analyzer does periodic and instant updates to generate the recommendations. Recommendation database hold the recommendations generated by the analyzer. Recommendations can be directly given to the user from the data available from the database [7] [8].

II. RELATED WORKS

A. Movie Recommendations from User Ratings [6]. This paper demonstrates a model for movie recommendations based on the user ratings. Clustering and Softmax Regression Classifier have been used to predict the recommendations. Clustering is done by implementing K-means algorithm. Users are clustered together to form groups. Classifier is used to predict the recommendation. Movie Lens dataset has been used for the implementation of the method. The method resulted in an RMSE score of 0.884.

B. Recommender System Based on Hierarchical Clustering Algorithm Chameleon [1]. Chameleon, the Hierarchical clustering algorithm is used with voting scheme to generate particular user rating. The system took set of users with their preference and using hierarchical cluster algorithm grouped the users into different clusters. The rating of a given item by particular user is generated by analyzing the rating of neighbors and then implementing voting scheme to all other users in that cluster for that specific item. By using the obtained outcome of voting scheme the system predicts rating of a user to particular product [1].

C. An Online Social Network-based Recommendation System [11]. Social network-based recommendation system used the information from user profiles and similar users by their connection. Here the implementation is the adapted version of the board game website. Authors have directed the use of recommender system for a network [11].

D. The Implementation of Knowledge-Based Recommender System for Electronic Commerce Using Java Expert System Library [9]. This paper deals with implementation of recommender system on e-shopping sites. Sometimes potential users would like to get recommendations on products to purchase in ecommerce application. Collaborative filtering is the dominant factor of automated recommendation engine. By using Java servlet and JESS implement the core engine and data base of item domain. The system collects the preference of a user on a specific product by explicitly asking to the user and analyze the knowledge base to discover the products based on user preference and meet the requirement of users [9].

E. Movie Rating Estimation and Recommendation [10] This paper deals with the movie rating prediction system and comparison of multiple methods implemented for it. Algorithms like Asymmetric SVD, Baseline predictor, SVD, KNN, Stochastic Gradient Descent, Integrated model, SVD++ and NMTF have been used[10]. RMSE is the estimation technique used for performance analysis.

3 Recommendation System Using Content and Collaborative -Based Filtering

The approach used for building the recommendation system is content and Collaborative -based filtering. As discussed earlier, content and Collaborative -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. S

tep 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.

Collaborative Filtering Algorithm

The users are clustered on the basis of ratings given by them for each movie. The dataset has been centered on zero by substracting the mean from individual rows. The resultant value of clustering will give clusters containing groups of users. The training set marked with the class labels is fed into the random forest classifier for learning. Test set is tested upon to get the predictions.

Random Forest Classifier Random forest is most accurate ensemble classifier and works efficiently on huge dataset. It can effectively predict the missing data accurately, even in situations where large portions of data are missing and without pre-processing[12]. It combines bagging and random feature selection. Random forest contains decision trees that are combined individual learners. Random subset of training data is used to generate trees. The test rows are passed through the forest after the forest have been trained. Each tree generate an output class we take the mode of that classes as the output of random forest [13][17]. Classification of labels using random forest classifier has been done as mentioned in Algorithm 3. In the method proposed Random forest classifier is used to predict the labels of users. The dataset has been divided into train and test set.The training set is given as an input to the classifier. The classifier will label the users into its respective classes which has been learned upon during the training phase.

Algorithm 2: Random Forest Classifier

Input: Training data(Td)

Output: class label

1) To form t classifiers:

for i = 1 to t do

Select Tdi from the training data Td randomly Generate a root node called Rni with Tdi Invoke GenerateTree(Rni)

end for

2) Generate Tree(Ni)

if Ni consists only one class instance then

return

else

Choose possible p% of the node(splitting features) in Ni randomly

For splitting elect the feature called F with the more gain of information

Build f child nodes of Ni , Ni1 ,..., Nif , here Fi possible values in can be in F that are Fi1 Fif

for j = 1 to Fi do

Replace Nij the content and Collaborative s of to TDj , here TDj is instances of Ni which is match to Fi Invoke GenerateTree(Ni)

end for

end if

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. 2). 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.3). 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. 4. 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. 2 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. 3 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. 4 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. 5 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- Content and Collaborative -Based Movie Recommendation System Using Genre Correlation igent 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] Utkarsh Gupta and Dr Nagamma Patil, "Recommender System Based on Hierarchical Clustering Algorithm Chameleon". Dept of Information Technology, National Institute of Technology Karnataka, 2015.

[2] Al Mamunur Rashid, George Karypis, and John Riedl,” Learning Preferences of New Users in Recommender Systems: An Information Theoretic Approach”, Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN-55455, 2008.

[3] Emmanouil Vozalis, Konstantinos G. Margaritis,”Analysis of Recommender Systems Algorithms”,University of Macedonia ,September, 2003.

[4] Daniar Asanov, ”Algorithms and Methods in Recommender Systems”, Berlin Institute of Technology Berlin, Germany, 2011.

[5] X. Su and T. M. Khoshgoftaar, ”A survey of collaborative filtering techniques”, Department of Computer Science and Engineering, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA, 3 August 2009.

[6] Hans BystrÃ ˝um ”Movie Recommendations from User Ratings”, Stanford University, 2013.

[7] S. S. Anand and B. Mobasher, ”Intelligent techniques for web personalization”, Proceedings of mm the 2nd Workshop on Intelligent Techniques in Web Personalization. 2003.

[8] Matteo Manca, Ludovico Boratto, Salvatore Carta,”Design and Architecture of a Friend Recommender System in the Social Bookmarking Domain”, Dipartimento di Matematica e Informatica Universit‘a di Cagliari via Ospedale 72, 09124 - Cagliari, Italy, August 27-29, 2014.

[9] In-Gook Chun, In-Sik Hong,”The Implementation Of Knowledge-Based Recommender System For Electronic Commerce Using Java Expert System Library ”, Division of Information Technology Engineering, SoonChunHyang University, 2001.

[10] Z. Bao, H. Xia, ”Movie Rating Estimation and Recommendation”, CS229 2012 project, Stanford University.

[11] Jorge Aranda, Inmar Givoni, Jeremy Handcock, Danny Tarlow,” An Online Social Network-based Recommendation System”, Department of Computer Science University of Toronto.

[12] Hastie,Trevor; Tibshirani, Robert; Friedman, Jerome ,” The Elements of Statistical Learning (2nd ed.).”, Springer. ISBN 0-387-95284-5, 2008.

[13] Ho, Tin Kam, ”The Random Subspace Method for Constructing Decision Forests” . IEEE Transactions on Pattern Analysis and Machine Intelligence , 1998.

[14] Ostrovsky, R., Rabani, Y., Schulman, L. J. and Swamy, C,”The Effectiveness of Lloyd-Type Methods for the k-Means Problem”. Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS’06). IEEE. pp. 165−174, 2006.

[15] Bahman Bahmani,Benjamin Moseley,Andrea Vattani,Ravi Kumar,Sergei Vassilvitskii ,”Scalable K-Means++”,Stanford University,2012.