International Journal of Computing Academic Research (IJCAR)
ISSN 2305-9184, Volume 7, Number 2 (April 2018), pp.29-37
© MEACSE Publications
http://www.meacse.org/ijcar

A Hybrid Movie Recommendation System Using Graph-Based Approach

Göksu Tüysüzoğlu[1] and Zerrin Işık[2]
[1]Graduate School of Natural and Applied Sciences, Dokuz Eylul University, Izmir, Turkey
[2]Department of Computer Engineering, Dokuz Eylul University, Izmir, Turkey

Abstract

Recommendation system is an assistive model for users with the intent of suggesting a set of new items to view (e.g., movie, news, research articles etc.) or buy (e.g., book, product etc.). Nowadays it has altered the way of seeking out the things of our interest by using information filtering approach. A movie recommendation system based on collaborative filtering handles the information provided by users, analyzes them, and suggests the best suited film to users according to their processed information. In the proposed system, a content-based movie recommendation is automatically made using a graph based approach according to past film preferences of users between 1995 and 2016; using demographic information of users, the recommendation list is updated. A combination of outputs of these two techniques reveals more precise recommendations concerning movies. The MovieLens dataset was used to explore the proposed hybrid system.

Keywords: content-based filtering approach, movie recommendation system, MovieLens, user-based collaborative filtering

Introduction

Recommendation system is an information filtering approach that collects ratings or preference information from users to make new suggestions to them for example in movie, video, book, music album recommendation or to recommend social elements such as people or groups to follow using a model built from the characteristics of an item or the users' demographic information and their social environment.

In literature two basic approaches, that are used for the construction of recommendation systems, are content-based and collaborative filtering techniques. A content-based recommendation system takes item's descriptions, which can be genre of a movie in case of movie recommendation or category of a book for book recommendation etc., into consideration so that similar items which are of particular interest to the user are likely to be suggested as a recommendation. The history of user interactions such as watched / rated movies or purchased items are collected to deduce user preferences. On the other side, a collaborative filtering handles information gathered from other people similar to the respective user who the system will recommend an item. The idea is that people who agreed in their evaluation of certain items in the past are likely to agree again in the future. Two versions of collaborative filtering approaches are commonly used as user-based and item-based filtering.

There are lots of studies which use Content-Based recommendation methods to result in concrete suggestions for users. Shu, Shen, Liu, Yi and Zhang [1] came up with a system which recommends learning resources to students based on their point of interests using a convolutional neural network which predicts the latent factors from the text information of the multimedia resources. The significant contribution of the applied algorithm is directly using text information in the content-based recommendation without tagging process. Soares and Viana [2] constructed content-based recommendation system to present new movie alternatives to the users of both Netflix and Movielens by using a number of metadata information of the movies comprising the title, genre, year of the release, and the list of directors and actors. In the study of Bergamaschi and Po [3], movie recommendations depend on the similarities between the plot of the videos that was watched by a user and the target movies. Two topic models, Latent Semantic Allocation (LSA) and Latent Dirichlet Allocation (LDA) are compared.

Collaborative Filtering methods also have a significant role in the recommendation step along with assistive filtering techniques such as content-based, knowledge-based or social ones. Recently, it has been applied to different recommendation systems, such as video recommendation system using a content-based technique (Deldjoo et al. [4]); book recommendation system based on combined features of content filtering, collaborative filtering and association rule mining (Tewari, Kumar and Barman [5]); suggesting recommendations to Last.fm and Movielens users by using social network and tag sources information by applying user-based technique (Zhou et al. [6]); point of interest recommendations for social media users using an author topic model (Jiang, Qian, Shen, Fu and Mei [7]).

Apart from the mentioned topics, there are a large variety of applications to construct movie recommendation systems. Sappadla, Sadhwani and Arora [8] compared two main recommendation filtering approaches (content-based and collaborative filtering) for different sizes of movie datasets and they found that user-based collaborative filtering resultswith the lowest mean squared error. In the study of Christakou and Stafylopatis [9], a combination of content-based and user-based collaborative filtering techniques using neural networks were proposed for movie recommendation.

Graph algorithms can also make recommendation phase easier by dealing with different types of contextual information. Huang, Chung, Ong, and Chen [10] developed a graph-based model that combines the content-based and the collaborative recommendation approaches to make recommendations for users of a digital library which is online Chinese bookstore. They implemented Hopfield net algorithm to exploit high-degree book-book, user-user and book-user associations and a two-layer graph structure was constructed using the similarity weights. Bogers [11] applied random walks over a contextual graph using the browsing process of a user on a movie database. Demovic et al. [12] made use of graph traversal algorithms to form a model which selects movies that are the best suited to the user's current interests.

This study proposes a hybrid system that is a combination of user-based and content-based recommendation techniques using a graph based approach to suggest new movie recommendations for users of the MovieLens database. The remaining parts of this paper are as follows. First, applied methodologies to construct the proposed model are given and then the used dataset and the application platform are explained. The findings from the applied methods are given in Results section. Finally, the main contributions of this study and planned future work are summarized.

Materials and Methods

In this section applied methodologies and datasets for experiments in addition to used platforms are presented.

Applied Methodologies. In the proposed system, two different recommendation schemes (user-based and content-based) are applied and their combination is used as the final recommendation for users. In the content-based recommendation model, past movie preferences of each user are taken into consideration so that which movie genre is the most suited for the respective user is inferred from the rating information. The content-based approach utilizes the concept of monotonic personal interest, in other words it is most likely that a person’s interest will not change in the near future (Salam and Najafi [13]). User-based recommendation model handles users’ demographic information (age, gender, occupation etc.) to recommend them movies which have not been seen yet for these users.

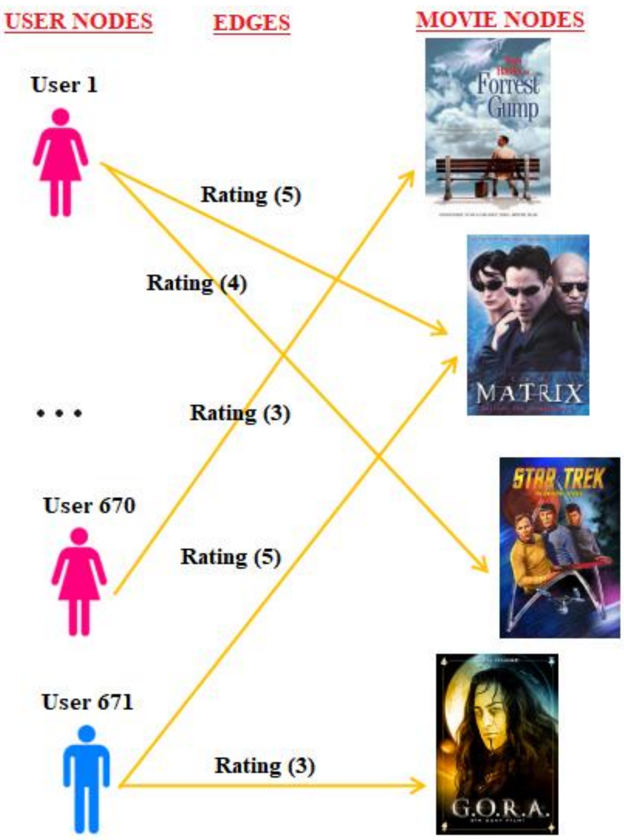

Fig. 1. Movie recommendation system as a graph.

The core part of the application is representing the system as a graph model. Graph G is composed of a pair of sets (V, E) where E is edges and V is vertices or nodes of the given graph. In order to construct movie recommendation system, movies and users should be placed into the vertices and edges should represent ratings (1 to 5) given by users to movies. Each node has its own features such that users have their respective IDs, age, occupation and gender information; movies have their respective IDs, title, year of the release, and genre information. Fig. 1 displays the visual representation of the movie recommendation system as a graph.

The first step is the extraction of some statistics from the constructed bipartite graph such as the most watched movie and its features; which user watched the most / the least movies; degree of each movie; and degree of each user.

In the latter step, all of the movies were distributed under their corresponding genres. A movie can be the member of more than one category. Then, the films per each category put in descending order according to the watch count.

The first neighbors of each movie node point out the users who rated / watched this movie and conversely, the movies that were rated by users are placed in the first neighbors of each user node. Using this information, movies which were watched and not watched by each user were extracted from the graph. This step is important because the most preferred film genres for each user are determined by looking through the frequency of the watched film genres. The film genres are ordered from the most preferred to the least one so that the proposed system suggests new movies to the user from the not watched film set of the corresponding user by selecting from the most preferred movie genre of him / her. This step can be stated as content-based recommendation technique since "genre" feature of the movie item is predicated on to make suggestion.

As a result of this technique five movie recommendations are made for each user from their most favored category and with the ones with highest watch counts.

The second recommendation technique takes the user's demographic data such as age and occupation into consideration. After extracting similar user groups, new film suggestions are made. Because the age attribute is continuous, it was discretized into five groups as the intervals of 7-17, 18-24, 25-35, 36-50, and 51-73. Later, the most watched film categories for each age group were identified.

As a result of two applied techniques, a combination of movie suggestion list was formed. The first three films are selected from content-based approach; fourth and fifth suggestions come from user-based approach.

Dataset Description and Used Platforms. The MovieLens database [14], that is available for public use, was chosen in this experimental study. The selected version of the dataset consists of 9125 movies, 671 users who rated movies between January 9, 1995 and October 16, 2016, 100004 ratings from the respective users. After removing the movies that are not rated by users, the remaining part covers 9066 movies. All selected users rated at least 20 movies. 18 genres are present (Action, Adventure, Animation, Children's, Comedy, Crime, Documentary, Drama, Fantasy, Film-Noir, Horror, Musical, Mystery, Romance, Sci-Fi, Thriller, War, Western) in the given context. Each dataset (Movies, Users and Ratings) has different attributes which are displayed in Table 1.

All of the implementations were applied in the RStudio using igraph package [15]. Three important methods from igraph package was utilized to construct the graph model: "graph_from_incidence_matrix" which creates a bipartite graph from an incidence matrix; "degree" / "ego_size" which calculates the size of the neighborhoods for the given vertices with the given order; "ego" which calculates the neighborhoods of the given vertices with the given order parameter.

Table 1. Used Datasets and Their Attributes

Dataset	**Attributes**
Movies	movieID, title, genres
Users	userID, movieID, rating, timestamp
Ratings	UserID, age, gender, occupation, zipcode

Results

Table 2 briefly reports the summary of the most watched / rated movie and the most / the least rated user according to the output from the degree information extracted from the nodes of the constructed bipartite graph.

Table2. Summary From the Calculation of the Degree of Graph Nodes

The *maximum* rated movie is Forrest Gump
The *maximum* rated movie's year of the release is 1994
The age of the user who rated *at most* is 50 with 2391 movies
The gender of the user who rated *at most* is male
The occupation of the user who rated *at most* is educator
The age of the user who rated *at least* is 24 with 20 movies
The gender of the user who rated *at least* is male
The occupation of the user who rated *at least* is technician

Table 3 demonstrates the ordered sequence of the first five movies under three genres as an illustration.

Table 3. Movies Under Categories as an Ordered Sequence According To the Watch Count

Movie Genre	**Movie Name**	**Movie Name**	**Movie Name**	**Movie Name**	**Movie Name**
Action	Star Wars: Episode IV - A New Hope	Jurassic Park	Matrix, The	Terminator 2: Judgment Day	Star Wars: Episode V - The Empire Strikes Back
Adventure	Star Wars: Episode IV - A New Hope	Jurassic Park	Toy Story	Star Wars: Episode V - The Empire Strikes Back	Back to the Future
Animation	Toy Story	Aladdin	Lion King, The	Beauty and the Beast	Shrek

Table 4. Movie Genre List from the Most Favored to the Least One for five users

User ID	**MovieGenre**	**MovieGenre**	**MovieGenre**	**MovieGenre**	**MovieGenre**
1	Adventure	Drama	Thriller	Action	Comedy
2	Drama	Comedy	Romance	Thriller	Action
3	Drama	Comedy	Action	Thriller	Adventure
4	Comedy	Action	Adventure	Drama	Thriller
5	Comedy	Drama	Romance	Adventure	Children

The distribution of genres for the first five users are illustrated in Table 4 as an ordered list by using the watching history of the corresponding users. The film category which has been watched at most / at least is found as Drama and Western, respectively.

Table 5 shows the suggested movie list from the process of content-based recommendation for the first five users.

According to the discretized age groups, each one selects movies from Drama category at most. In addition to age groups, occupation information is also significant to determine similar tastes for movies. Table 6 points out the resulting genre list of each user occupation.

The final recommendation list of the proposed hybrid system selects the first three films from content-based approach in addition to them fourth and fifth suggestions come from user-based approach where both the most watched movies from Drama category (obtained from age-groups) and the most watched ones from the respective movie categories in terms of user occupation given in Table 6 are chosen. Table 7 indicates the final recommendation list for five users using the hybrid and graph-based movie recommendation system.

Table 5. Final Recommendations for Five Users from the Content-Based Recommendation Technique

User ID	**MovieTitle**	**Movie Title**	**Movie Title**	**Movie Title**	**Movie Title**
1	Star Wars: Episode IV - A New Hope	Jurassic Park	Toy Story	Star Wars: Episode V - The Empire Strikes Back	Back to the Future
2	Shawshank Redemption, The	Fargo	American Beauty	Fight Club	Godfather, The
3	Fargo	Dances with Wolves	Apollo 13	Lion King, The	Godfather, The
4	Toy Story	Fargo	Ace Ventura: Pet Detective	Shrek	Dumb & Dumber (Dumb and Dumber)
5	Pulp Fiction	Toy Story	Back to the Future	Fargo	True Lies

Table 6. The most Favourite Movie Categories for User Occupations

Occupation	**The Most Favourite Movie Genre**	**Occupation**	**The Most Favourite Movie Genre**
administrator	Drama	marketing	Drama
artist	Drama	none	Comedy
doctor	Drama	other	Drama
educator	Drama	programmer	Drama
engineer	Drama	retired	Comedy

entertainment	Drama	salesman	Drama
executive	Drama	scientist	Drama
healthcare	Drama	student	Drama
homemaker	Drama	technician	Drama
lawyer	Action	writer	Drama
librarian	Drama		

Table 7. Final Recommendations as a Combination of Content-Based and User-Based Recommendation Techniques

User ID	**MovieTitle**	**MovieTitle**	**MovieTitle**	**MovieTitle**	**MovieTitle**
1	Star Wars: Episode IV - A New Hope	Jurassic Park	Toy Story	Star Wars: Episode V - The Empire Strikes Back	Waiting to Exhale
2	Shawshank Redemption, The	Fargo	American Beauty	Fight Club	Godfather, The
3	Fargo	Dances with Wolves	Apollo 13	Lion King, The	Godfather, The
4	Toy Story	Fargo	Ace Ventura: Pet Detective	Shrek	Waiting to Exhale
5	Pulp Fiction	Toy Story	Back to the Future	Fargo	Waiting to Exhale

Conclusions and Future Work

In this study, a combination of two techniques as user-based (demographic information) and content-based (past user rating information) to construct a movie recommendation system is proposed. The system is firstly represented as a bipartite graph and by determining the first neighborhoods of each node content-based information is drawn. On the other hand, user-based information from user node attributes is processed. As a result of the combination of two techniques, movie recommendations are automatically made for users. The proposed hybrid system integrates the power of both content-based and user-based filterings so that even if the user has no rating history beforehand, the system can make use of the demographic information of the respective user and a new suggestion can be made by using the preferences of users whose have similar demographic features. Furthermore, the execution time of the system speeds up by using a graph-based model.

As a future work, it is planned to develop the proposed model by calculating user and item similarities using Pearson Correlation or Jaccard Similarity metrics incorporating with the graph topology to improve the recommendation quality.

References

[1] J. Shu, X. Shen, H. Liu, B. Yi, and Z. Zhang, "A Content-Based Recommendation Algorithm for Learning Resources," in Multimedia Systems, March 2017, pp. 1-11, doi: 10.1007/s00530-017-0539-8.
[2] Soares and P. Viana, "Tuning Metadata for Better Movie Content-Based Recommendation Systems," Multimedia Tools and Applications, vol.74, Sept. 2015, pp. 7015-7036, doi: 10.1007/s11042-014-1950-1.
[3] S. Bergamaschi, and L. Po, "Comparing LDA and LSA Topic Models for Content-Based Movie Recommendation Systems," In International Conference on Web Information Systems and Technologies, Springer, Cham, Dec. 2015, pp. 247-263, doi: 10.1007/978-3-319-27030-2_16.
[4] Deldjoo Y. et al., "Content-Based Video Recommendation System Based on Stylistic Visual Features". Journal on Data Semantics, vol. 5, February 2016, pp. 99-113, doi: 10.1007/s13740-016-0060-9.
[5] A. S. Tewari, A. Kumar, A. G. Barman, "Book Recommendation System Based on Combine Features of Content Based Filtering, Collaborative Filtering and Association Rule Mining," In Advance Computing Conference (IACC), 2014 IEEE International, Febrauary 2014, doi: 10.1109/IAdCC.2014.6779375.
[6] J. Zhou et al., "Social Network and Tag Sources Based Augmenting Collaborative Recommender System," IEICE Transactions on Information and Systems, vol. 98, April 2015, pp. 902-910, doi: 10.1587/transinf.2014EDP7283.
[7] S. Jiang, X. Qian, J. Shen, Y. Fu, T. Mei, "Author Topic Model-Based Collaborative Filtering for Personalized POI Recommendations," IEEE Transactions on Multimedia, vol. 17, March 2015, pp. 907-918, doi: 10.1109/TMM.2015.2417506.
[8] P. Sappadla, Y. Sadhwani, P. Arora, "Movie Recommender System," Search Engine Architecture, Spring 2017, NYU Courant, 2017, https://pdfs.semanticscholar.org/767e/ed55d61e3aba4e1d0e175d61f65ec0dd6c08.pdf.
[9] C. Christakou, S. Vrettos, A. Stafylopatis, "A Hybrid Movie Recommender System Based on Neural Networks," International Journal on Artificial Intelligence Tools, vol. 16, Sept. 2007, pp. 771-792, doi: 10.1109/ISDA.2005.9.
[10] Z. Huang, W. Chung, T. H. Ong, and H. Chen, "A Graph-Based Recommender System for Digital Library," In Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries, July 2002, pp. 65-73, doi: 10.1145/544220.544231. Multimedia Tools and Applications, vol.74, Sept. 2015, pp. 7015-7036, doi: 10.1007/s11042-014-1950-1.
[11] T. Bogers, "Movie Recommendation Using Random Walks Over the Contextual Graph," In Proc. of the 2nd Intl. Workshop on Context-Aware Recommender Systems, 2010.
[12] L. Demovic, et al., "Movie Recommendation Based on Graph Traversal Algorithms," In Database and Expert Systems Applications (DEXA), 2013 24th International Workshop on, August 2013, doi: 10.1109/DEXA.2013.24.
[13] Z. Salam Patrous, and S. Najafi, "Evaluating Prediction Accuracy for Collaborative Filtering Algorithms in Recommender Systems," 2016.

[14] F. M. Harper, J. A. Konstan, "The MovieLens Datasets: History and Context," ACM Transactions on Interactive Intelligent Systems (TiiS), vol. 5, Jan. 2016, doi: 10.1145/2827872.
[15] G. Csardi, T. Nepusz, "The igraph software package for complex network research," InterJournal, Complex Systems, vol. 1695, pp. 1-9, 2006.