

TOWARDS A HYBRID PERSONALIZED MOVIE RECOMMENDER SYSTEM

GAYATRI GADIKAR
gayatrigadikar04@gmail.com

ABSTRACT

Recommender systems represent a powerful method for enabling users to filter through wide verity of information. Research in the recommender system is moving in the direction of a richer understanding of how recommender technology may be embedded in specific domains.

A recommendation system for movies is important in our social life to provide the enhanced entertainment .There are two major recommendation techniques –Collaborative and content based filtering but these filtering techniques are having some limitations thus an hybrid approach is often adopted. The proposed movie recommender system has an ability to recommend user specific movie by using their social networking site like Facebook data. This provides the generalized framework for personalized movie recommendation.

KEYWORDS: Movie Recommender System, Collaborative and content based filtering, hybrid.

I)INTRODUCTION

'Recommendation systems' are services which recommend users new items such as movies , brand-wise products, news articles, books and music they would like. Recommendation Systems based on AI technology have been explored especially around 1990s, when the WWW and internet services grew explosively all around the world.

Recommender systems have become an important research area in industry and academia. The interest in this area still remains high because it constitutes a problem-rich research area and because of the abundance of practical applications that help users to deal with information overloads and provide personalized recommendations, content, and services to them. Examples of such applications include recommending books, CDs, and other products at Amazon.com, movies by MovieLens, and news at VERSIFI Technologies (formerly AdaptiveInfo.com). Moreover, some of the vendors have incorporated recommendation capabilities into their commerce servers [1]

Two main paradigm used in recommender system are content based and collaborative filtering and our system proposes hybrid use of these two to recommend a movie to the user.

II) FILTERING ALGORITHM

A. COLLABORATIVE BASED RECOMMENDATION SYSTEMS:

Collaborative filtering Algorithm recommender system became one of the most researched techniques of recommender systems .if users shared the same interests in the past, they will also have similar tastes in the future. So for example, if user A and user B have a purchase history that overlaps strongly and user A has recently bought an item that B has not yet been, the basic rationale

is to propose this item also to B. The collaborative filtering technique recommends items based on user-based approach and item-based approach [2].

1) USER-BASED APPROACH:

In the User-based approach the user plays an important role. If certain majority of the customer have same taste then they join into the one group .Recommendations are given to user based on evaluation of items by other users form the same group, with whom he/she shares common preferences. If the item was positively rated by the community, it will be recommended to the user.

2) ITEM-BASED APPROACH:

Here in Item-Based Approach the items play an important role .Recommendations is based on evaluation of items the system generates recommendations with items in the neighborhood that a user would prefer.[3]

B. CONTENT BASED RECOMMENDATION SYSTEM

Here in Content-based recommender systems deal with profiles of users that are created at the beginning. A profile has information about a user and his taste which is based on how user rates the items. In the recommendation process, the engine compares items that were already rated by user with items he did not rate and looks for similarities. Those items that are mostly similar to the positively rated ones, and the one which are positively rated by the users are recommended to the users [2].

C. HYBRID BASED RECOMMENDATION SYSTEM

Here in Hybrid Recommendation it is a combination of both collaborative approach and content based approach .With the help of Hybrid Recommendation different types of problems can be easily overcome the problem such as Cold-Start problem can be handled using the hybrid recommendations[2].

III) PROPOSED SYSTEM/APPROACH

A Personalized movie Recommender System comprises of login Page, Home page, Registration form, My choice Movie tab and a search bar .The user can log in through Facebook details as well. When the user logs into the system, the collaborative filtering is activated to look for movies that are predicted to be highly rated by the user. Then, the user can go and rate those items. The user can post his/her movie choice as per the actor, director or genre through My choice. The Home page will display moving boxes with movies that have been added recently and highly rated.

IV) DATA RETRIEVAL & PRE-PROCESSING

a) The datasets freely available online consisted of large number of ratings, but they lacked the required features about the movie for the recommendation systems. Hence the dataset had to be preprocessed and modified before being implemented into the system. Movie data will

be extracted from MovieLens DataSet in JSON format or by using Jester or OMDb API in JSON format. These are popular datasets for collaborative filtering. The JSON data can be converted to CSV file to import it into the database. Movie details included the title, actor, cast, language, rating, director, plot, poster url etc. [4] [15]

b) LOGIN DETAILS: BY USING FACEBOOK:

By using Facebook Graph API user data will be collected. User's name, age, gender, language and liked movies will be extracted when user will be logged in using Facebook.

User's posts and comments containing movie tags will be collected. Natural Language processing is performed on the posts and comments, to check if the post is positive or negative. According to the data obtained from the processing, the database will be updated later. [4] (Twitter, and LinkedIn API also can be used to create user profile)

WITHOUT LOGIN TO SOCIAL NETWORKING SIGHT:

The user has to login the Movie recommender system and after that he/she has to fill a registration page with details like Name, Age, Gender, and Profession. The user also has an option to select a few genres that the user likes. This information about the user is used to solve the cold start problem that may arise. Thus, users having a cold start are firstly recommended movies popular in their own age group. Secondly, if the user has specified some genres that the user likes, content based recommendations are provided to the user. [4]

$$\bar{v}_u = \frac{\sum_{i \in S_u} v_{ui}}{|S_u|} \tag{5}$$

The weight that we previously mentioned can be calculated using Pearson correlation according to the following formula:

$$sim(i, j) = \frac{\sum_{\{u \in U | i \in S_u \& j \in S_u\}} (v_{ui} - \bar{v}_u)(v_{uj} - \bar{v}_u)}{\sqrt{\sum_{\{u \in U | i \in S_u \& j \in S_u\}} (v_{ui} - \bar{v}_u)^2 \sum_{\{u \in U | i \in S_u \& j \in S_u\}} (v_{uj} - \bar{v}_u)^2}} \tag{5}$$

Prediction formula is given below

$$p_{ai} = \frac{\sum_{\{j \in S_a | j \neq i\}} sim(i, j) \times v_{aj}}{\sum_{\{j \in S_a | j \neq i\}} |sim(i, j)|} \tag{5}$$

Item	I1	I2	I3	I4	I5
User					
U1	5	8		7	8
U2	10		1		
U3	2	2	10	9	9
U4		2	9	9	10
U5	1	5			1
User a	2		9	10	

c) PROCESSING

APPLY NATURAL LANGUAGE PROCESSING :

Facebook login will help to extract the information of user's like or dislike related to movies. The posts containing the tagged movie pages, implicit and explicit feedback are considered for NLP. The weight age to the every term (word, phrase and emoticons) is given which is used to perform NLP and calculate the approximate rating (disliking and liking) for a movie.

We can use any of the following for natural language processing and Text processing.

1. NLTK (one of the best toolkit to process data and make NLP applications)
2. orange (GUI interface toolkit developed using python works great!!)
3. scikit (machine learning tool)

a) Collaborative filtering: This method finds the sub set of users who have similar tests and preferences to the target user and use this sub set of offerings recommendations Main approaches 1) User based 2) Item Based

1) User Based collaborative filtering: By using Pearson correlation and Prediction

In this method, we predict the user behavior against a certain item using the weighted sum of deviations from mean ratings of users that previously rated this item and the user mean rate. First, we calculate the user mean rate using the following formula:

2) Item based collaborative filtering:

Item	I1	I2	I3	I4	I5
User					
U1	5	8		7	8
U2	10		1		
U3	2	2	10	9	9
U4		2	9	9	10
U5	1	5			1
User	2		9	10	

I decided to apply Item-based Collaborative Filtering to display similar items for the user once he or she selected a particular item using the adjusted cosine formula

$$sim(i, j) = \frac{\sum_{\{u \in U | i \in S_u \& j \in S_u\}} (v_{ui} - \bar{v}_u)(v_{uj} - \bar{v}_u)}{\sqrt{\sum_{\{u \in U | i \in S_u \& j \in S_u\}} (v_{ui} - \bar{v}_u)^2 \sum_{\{u \in U | i \in S_u \& j \in S_u\}} (v_{uj} - \bar{v}_u)^2}}$$

Additionally, we can predict how the user will rate this item using the previous similarity:

$$p_{ai} = \frac{\sum_{\{j \in S_a | j \neq i\}} sim(i, j) \times v_{aj}}{\sum_{\{j \in S_a | j \neq i\}} |sim(i, j)|}$$

(Candillier et al., 2007)

One of the issues involved with collaborative filtering is that it requires all the rated movies in order to consider them for ranking. If only a few rated items are available, the recommendation will be done with respect to the small set. It may face new user cold start problem or new item cold start problem

- b) Content based: Create a user profile to describe the type of movies that user prefers or likes. The information about user profile comprises of User name, age, movie and genre and actor/actress he/she like. Also, the properties of movie i.e. (Similarity and overlapping of the genres, overlapping of the Actors and directors, Similarity of the movie plots) are considered. The user is rated movies based on his user profile. Thus first movies are filtered by the genres liked by the user, followed by filtering these movies by the favorite actors and directors of the user, thereby giving the user recommendations based on what the user has liked before. The movies were filtered according the plot using similarity.[4] However, it suffers from the problem such as limited content analysis, over-specialization and new user problem.
- c) Hybrid Recommendation: Hybrid approaches combine collaborative and content-based methods to overcome certain limitations of these individual techniques. In a Pipelined Hybrid Recommendation System, the input is passed on to a recommender system and the output generated by this system is passed as input to the next recommender system. For example, first the movie recommendations are found out by collaborative filtering, and then these recommendations are passed to the content based recommender to generate movie recommendations which are similar in content. This approach allows the system to recommend movies that should be rated high but are not recommended by collaborative filtering approach since no or very few users have rated this movie [4]. Python inbuilt library 'CRABBE' is the best suited recommender in our case.[15][16]

d) CHALLENGES

- a) One of the challenges that commercial websites face is the new user or new item problem. Whenever a user logs into the system for the first time, the system does not have information about the preferences of that user to provide the recommendations and similarly. When a new movie is added to the system, it doesn't have the user ratings. Due to lack of ratings, the item may be placed lower in the recommended items list. To calculate the similarities for neighborhood, the historic ratings are necessary, and they are unavailable in the cold-start scenario. This makes it difficult to predict ratings for items.
- b) Getting complete user data through facebook API or other social networking sights is became difficult now.

e) CONCLUSION AND FUTURE WORK

This paper explains how quality movie recommendations can be constructed by taking minimum inputs from the user and by using Hybrid filtration technique. The huge amount of data and the computational power needed to bring accurate results should be considered in applying an online or even offline recommender system. However, there are other

recommender engines like Hunch which is used in eBay, one of the biggest leaders in online commerce. eBay was looking to compete with Amazon by acquiring Hunch engine due to its house built Item-based recommender system that does not compete very well. ChisDixon, the co-founders of Hunch and entrepreneur in the field of data mining and machine learning, refers to the value of using a history of data that can be linked to social commerce.

For example, Hunch uses Facebook accounts and twitter to refine what users like and build a graph that links people to objects. My6sense, a newsreader app, uses the same idea of Hunch where different API's like Twitter and other social media are integrated for recommendation purposes. These types of engines monitor the links you clicked, time spent reading, and whether the link was passed or shared. In addition, Forage is a Hunch API that recommends YouTube videos based on user's tweets.

In future I can think of developing a personalized mobile Recommender App which will recommend the movie as per the choice of an individual.

REFERENCES

- [1] Gediminas Adomavicius, Member, IEEE, and Alexander Tuzhilin, Member, IEEE "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 17, NO. 6, JUNE 2005
- [2] Mohammad Hamidi Esfahani, Farid Khosh Alhan - "New Hybrid Recommendation System Based On C-Means Clustering Method " - 5th conference of information Knowledge . IEEE-2013
- [3] Jaimeel Shah , Lokesh Sahu " A Survey of Various Hybrid based Recommendation Method" International Journal of Advanced Research in Computer Science and Software Engineering . Volume 4, Issue 11, November 2014 ISSN: 2277 128X
- [4] Khyati Aggarwal , Yashowardhan Soni "Movie Recommendations using Hybrid Recommendation Systems" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 4 Issue: 12 206 – 209
- [5] Ala Alluhaidan Grand Valley State University "Recommender System Using Collaborative Filtering Algorithm" Grand Valley State University ScholarWorks@GVSU Technical Library School of Computing and Information Systems 2013
- [6] Klesti Hoxha, Alda Kika, Eriglen Gani, Silvana Greca " Towards a Modular Recommender System for Research Papers written in Albanian" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No. 4, 2014
- [7] J. Beel, S. Langer, M. Genzmehr, and A. Nürnberger, "Introducing Docear's research paper recommender system", Proceedings of the 13th ACM/IEEE-CS joint conference, (JCDL '13), 2013, pp.459-460.
- [8] J. Lee, K. Lee, and J. G. Kim, "Personalized Academic Research Paper Recommendation System.", arXiv preprint arXiv:1304.5457(2013).

- [9] R. Torres, S. M. McNee, M. Abel, J.A. Konstan, and J. Riedl, "Enhancing Digital Libraries with Techlens", Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries (Tuscon, AZ, USA, 2004), pp. 228–236.
- [10] B. Gipp, J. Beel, and C. Hentschel, "Scienstein: A Research Paper Recommender System", In Proceedings of the International Conference on Emerging Trends in Computing (ICETiC'09), Virudhunagar (India), January 2009, pp. 309–315.
- [11] T. Strohman, W. B. Croft, and D. Jensen, "Recommending citations for academic papers.", In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '07, pages 705–706, New York, NY, USA, 2007. ACM.
- [12] T. Bogers, and A. van den Bosch, "Recommending scientific articles using citeulike", Proceedings of the 2008 ACM conference on Recommender systems (2008), 287–290.
- [13] J. Tang and J. Zhang, "A discriminative approach to topic-based citation recommendation", In T. Theeramunkong, B. Kijssirikul, N. Cercone, and T.-B. Ho, editors, Advances in Knowledge Discovery and Data Mining, volume 5476 of Lecture Notes in Computer Science, pages 572–579. Springer Berlin / Heidelberg, 2009.
- [14] C. D. Manning, R. Prabhaka and H. Sch tze, Introduction to information retrieval, New York: Cambridge University Press, 2008
- [15] <https://www.analyticsvidhya.com/blog/2016/06/quick-guide-build-recommendation-engine-python/>
- [16] <https://grouplens.org/datasets/movielens/100k/>