PULMONOLOGY

Development of Movie Mender: A Movie Recommender System Based on User Preferences

Sneha Patankar[1], Rohan Kulkarni, Neha Mehta[2], Aditya Deshmukh[3], Tanvi Jadhav[4]*
[1] Assistant Professor, Department of Computer Engineering, MIT College of Engineering, Pune
*[234] Students, MIT College of Engineering, Pune

ABSTRACT
In today's digital world where there is an endless variety of content to be consumed like books, videos, articles, movies, etc., finding the content of one's liking has become an irksome task. On the other hand digital content providers want to engage as many users on their service as possible for the maximum time. This is where recommender system comes into picture where the content providers recommend users the content according to the users' liking. In this paper we have proposed a movie recommender system MovieMender. The objective of MovieMender is to provide accurate movie recommendations to users. Usually the basic recommender systems consider one of the following factors for generating recommendations; the preference of user (i.e content based filtering) or the preference of similar users (i.e collaborative filtering). To build a stable and accurate recommender system a hybrid of content based filtering as well as collaborative filtering will be used.

KEYWORDS: Movies, Recommendation system, CBF- Content-based filtering, CF- Collaborative filtering, Hybrid systems

INTRODUCTION
Recommendation systems help users find and select items (e.g., books, movies, restaurants) from the huge number available on the web or in other electronic information sources. Given a large set of items and a description of the user's needs, they present to the user a small set of the items that are well suited to the description. Similarly, a movie recommendation system provides a level of comfort and personalization that helps the user interact better with the system and watch movies that cater to his needs. Providing this level of comfort to the user was our primary motivation in opting for movie recommendation system as our BE Project. The chief purpose of our system is to recommend movies to its users based on their viewing history and ratings that they provide. The system will also recommend various E-commerce companies to publicize their products to specific customers based on the genre of movies they like. Personalized recommendation engines help millions of people narrow the universe of potential films to fit their unique tastes. Collaborative filtering and content based filtering are the are prime approaches to provide recommendation to users. Both of them are best applicable in specific scenarios because of their respective ups and downs. In this paper we have proposed a mixed approach such that both the algorithms complement each other thereby improving performance and accuracy of the of our system.

LITERATURE SURVEY
MOVREC [10] is a movie recommendation system presented by D.K. Yadav et al. based on collaborative filtering approach. Collaborative filtering makes use of information provided by user. That information is analyzed and a movie is recommended to the users which are arranged with the movie with highest rating first. The system also has a provision for user to select attributes on which he wants the movie to be recommended.
Luis M Capos et al. [5] has analyzed two traditional recommender systems i.e. content based filtering and collaborative filtering. As both of them have their own drawbacks he proposed a new system which is a combination of Bayesian network and collaborative filtering. The proposed system is optimized for the given problem and provides probability distributions to make useful inferences.
A hybrid system has been presented by Harpreet Kaur et al. [9]. The system uses a mix of content as well as collaborative filtering algorithm. The context of the movies is also considered while recommending. The user - user relationship as well as user - item relationship plays a role in the recommendation.
The user specific information or item specific information is clubbed to form a cluster by Utkarsh Gupta et al. [12] using chameleon. This is an efficient technique based on Hierarchical clustering for recommender system. To predict the rating of an item voting system is used. The proposed system has lower error and has better clustering of similar items.
Urszula Kużelewska et al. [6] proposed clustering as a way to deal with recommender systems. Two methods of computing cluster representatives were presented and evaluated. Centroid-based solution and memory-based collaborative filtering methods were used as a basis for comparing effectiveness of the proposed two methods. The result was a significant increase in the accuracy of the generated recommendations when compared to just centroid-based method.
Costin-Gabriel Chiru et al. [3] proposed Movie Recommender, a system which uses the information known

about the user to provide movie recommendations. This system attempts to solve the problem of unique recommendations which results from ignoring the data specific to the user. The psychological profile of the user, their watching history and the data involving movie scores from other websites is collected. They are based on aggregate similarity calculation. The system is a hybrid model which uses both content based filtering and collaborative filtering.

To predict the difficulty level of each case for each trainee Hongli LIn et al. proposed a method called content-boosted collaborative filtering (CBCF).The algorithm is divided into two stages, First being the content-based filtering that improves the existing trainee case ratings data and the second being collaborative filtering that provides the final predictions. The CBCF algorithm involves the advantages of both CBF and CF, while at the same time, overcoming both their disadvantages.

There are various types of recommender systems with different approaches and some of them are classified as below:

Content-based Filtering Systems (CBF based systems):

In content-based filtering, items are recommended based on comparisons between item profile and user profile. A user profile is content that is found to be relevant to the user in form of keywords(or features). A user profile might be seen as a set of assigned keywords (terms, features) collected by algorithm from items found relevant (or interesting) by the user. A set of keywords (or features) of an item is the Item profile. For example, consider a scenario in which a person goes to buy his favorite cake 'X' to a pastry. Unfortunately, cake 'X' has been sold out and as a result of this the shopkeeper recommends the person to buy cake 'Y' which is made up of ingredients similar to cake 'X'. This is an instance of content-based filtering.

Advantages of content-based filtering are:

- They capable of recommending unrated items
- We can easily explain the working of recommender system by listing the Content features of an item.
- Content-based recommender systems use need only the rating of the concerned user,and not any other user of the system.

Disadvantages of content-based filtering are:

- It does not work for a new user who has not rated any item yet as enough ratings are required content-based recommender evaluates the user preferences and provides accurate recommendations.
- No recommendation of serendipitous items.
- Limited Content Analysis- The recommender does not work if the system fails to distinguish the items hat a user likes from the items that he does not like.

Collaborative filtering based systems (CF based systems):

Collaborative filtering system recommends items based on similarity measures between users and/or items. The system recommends items preferred by similar users. This is based on the scenario where a person asks his friends, who have similar tastes, to recommend him some movies.

Advantages of collaborative filtering based systems:

- It is dependent on the relation between users which implies that it is content-independent.
- CF recommender systems can suggest serendipitous items by observing similar-minded people's behavior.
- They can make real quality assessment of items by considering other peoples experience.

Disadvantages of collaborative filtering are:

- Early rater problem: Collaborative filtering systems cannot provide recommendations for new items since there are no user ratings on which to base a prediction.
- Gray sheep: In order for CF based system to work, group with similar characteristics are needed. Even if such groups exist, it will be very difficult to recommend users who do not consistently agree or disagree to these groups.
- Sparsity problem: In most cases, the amount of items exceed the number of users by a great margin which makes it difficult to find items that are rated by enough people.

SYSTEM DESCRIPTION

Owing to the various demerits of pure content-based and pure CF based systems, we have proposed a hybrid recommender system which is known as content-boosted collaborative filtering system. This hybrid system takes advantage from both the representation of the content as well as the similarities among users. The intuition behind this technique is to use a content-based predictor to fill the user-rating matrix that is sparsely distributed. A web crawler is used to download necessary movie content for our dataset. After the preprocessing the movie content database is stored. The dataset consists of a user-rating matrix. Content-based predictions are used to

train each user-rating vector in the user-rating matrix and convert it into a pseudo rating matrix which combines actual rating with the predicted ratings. Collaborative filtering is then applied to this full pseudo user-rating matrix to make recommendation for an active user.

Hybrid Algorithm

Step1: Use content-based predictor to calculate the pseudo user-rating vector 'v' for every user 'u' in the database.

$$v_{u,i} = r_{u,i} : \text{is user u rated item i}$$

$$v_{u,i} = r_{u,i} : \text{otherwise}$$

Step2: Weight all users with respect to similarity with the active user.

- Similarity between users is measured as the Pearson correlation between their ratings vectors.

Step3: Select n users that have the highest similarity with the active user.

- These users form the neighborhood.

Step4: Compute a prediction from a weighted combination of the selected neighbors' ratings.

In step 2, the similarity between two users is computed using the Pearson correlation coefficient, defined below:

$$P_{a,u} = \frac{\sum_{i=1}^{m}\left(r_{a,i} - \bar{r}_a\right) \times \left(r_{u,i} - \bar{r}_u\right)}{\sqrt{\sum_{i=1}^{m}\left(r_{a,i} - \bar{r}_a\right)^2 \times \sum_{i=1}^{m}\left(r_{u,i} - \bar{r}_u\right)^2}}$$

Where, $r_{a,i}$ is the rating given to item i by user a ;

$\bar{r}_a$ is the mean rating given by user a ;

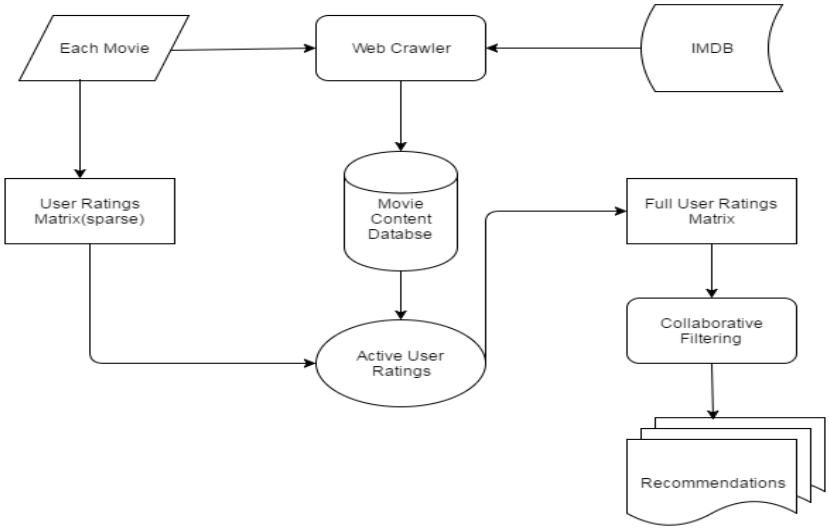
m is the total number of items .

In step 4, predictions are computed as the weighted averages of deviations from the neighbor's mean:

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^{n}\left(r_{u,i} - \bar{r}_u\right) \times P_{a,u}}{\sum_{u=1}^{n} P_{a,u}}$$

Where, $p_{a,i}$ is the prediction for the active user a for item i ;
$P_{a,u}$ is the similarity between users a and u ;
n is the number of users in the neighborhood .

Figure:



System Overview

CONCLUSION

A hybrid approach is taken between context based filtering and collaborative filtering to implement the system. This approach overcomes drawbacks of each individual algorithm and improves the performance of the system. Techniques like Clustering, Similarity and Classification are used to get better recommendations thus reducing MAE and increasing precision and accuracy. In future we can work on hybrid recommender using clustering and similarity for better performance. Our approach can be further extended to other domains to recommend songs, video, venue, news, books, tourism and e-commerce sites, etc.

REFERENCES

[1] James Bennett, Stan Lanning ; "The Netflix Prize", In KDD Cup and Workshop in conjunction with KDD, 2007.

[2] Mohammad Yahya H. Al-Shamri , Kamal K. Bharadwaj; "A Compact User Model for Hybrid Movie Recommender System " in International Conference on Computational Intelligence and Multimedia Applications 2007

[3] Costin-Gabriel Chiru, Vladimir-Nicolae Dinu , Ctlina Preda, Matei Macri ; "Movie Recommender System Using the User's Psychological Profile" in IEEE International Conference on ICCP, 2015.

[4] Christina Christakou, Leonidas Lefakis, Spyros Vrettos and Andreas Stafylopatis; "A Movie Recommender System Based on Semi-supervised Clustering ", IEEE Computer Society Washington, DC, USA 2015.

[5] Luis M. de Campos, Juan M. Fernández-Luna *, Juan F. Huete, Miguel A. Rueda-Morales; "Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks", International Journal of Approximate Reasoning, revised 2010.

[6] Urszula Kużelewska; "Clustering Algorithms in Hybrid Recommender System on MovieLens Data", Studies in Logic, Grammar and Rhetoric, 2014.

[7] Dietmar Jannach, Gerhard Friedrich; "Tutorial: Recommender Systems", International Joint Conference on Artificial Intelligence, Beijing, August 4, 2013.

[8] Gaurangi, Eyrun, Nan; "MovieGEN: A Movie Recommendation System", UCSB.

[9] Harpreet Kaur Virk, Er. Maninder Singh," Analysis and Design of Hybrid Online Movie Recommender System "International Journal of Innovations in Engineering and Technology (IJIET)Volume 5 Issue 2,April 2015.

[10] Manoj Kumar, D.KYadav, Ankur Singh, Vijay Kr. Gupta," A Movie Recommender System: MOVREC" International Journal of Computer Applications (0975 – 8887) Volume 124 – No.3, August 2015.

[11] Prerana Khurana , Shabnam Parveen; 'Approaches of Recommender System: A Survey'; International Journal of Computer Trends and Technology (IJCTT) – Volume 34 Number 3 - April 2016.

[12] Utkarsh Gupta1 and Dr Nagamma Patil2," Recommender System Based on Hierarchical Clustering Algorithm Chameleon" 2015 IEEE International Advance Computing Conference(IACC).