Movie Recommendation System

Based on content and collaborative filtering techniques

B Kalyani
Department of Computer Science Engineering
G Pullaiah College of Engineering and Technology
Kurnool, India
Kalyanib67@gmail.com

C Deeptanjali
Department of Computer Science Engineering
G Pullaiah College of Engineering and Technology
Kurnool, India
deepthi.chitrala@gmail.com

C Ayesha Shariff
Assistant Professor,
Department of Computer Science Engineering
G Pullaiah College of Engineering and Technology
Kurnool, India
cafshariff@gmail.com

Abstract

In today's modern world, where an endless variety of content can be consumed, such as books, videos, articles, movies, etc., finding one's liking content has become an irksome task. The recommender system enters into the picture where the content provider recommends the content to the user according to their preference. Personalized recommendation system can perform a major role, especially if the user doesn't have a specific target movie. In this paper, we analyze the content and collaborative filtering algorithms, we design and device a recommender model for movies. This also helps users to find their movies of choice efficiently and effectively based on other users ' movie experience without wasting a lot of time in useless browsing. Finally, we proved that our system provides better recommendation results.

Keywords: Data mining, Recommendation System, clustering, movies, Collaborative filtering, Content-based filtering

1. Introduction

With the rapid development of Internet technology the world has entered into the era of Web 2.0[1], this lead to information overload. Thus finding required data by user is difficult and hence is under research. As it a well-known fact that Movie is one of the popular sources of entertainment. So we propose a recommender system for providing list of movies that are similar to users preference [1,2]. Personalized recommendation trying to find the user's traits and tastes by gathering and evaluating past behavior in order to identify what type of individual the user is, what type of activities the user chooses, what type of stuffs the user wishes to share, and so on [4, 5, 6], and finally understand that user features and likings are based on system rules and recommend the movies that the user is interested in [7, 8]. A customized recommender is a kind of information filtering approach. It is an integrated system that combines a variety of data mining algorithms and data linked to users to meet users ' interests.

The recommendation systems can be classified in two broad categories [9]:

1. Collaborative filter
2. Content-based filter

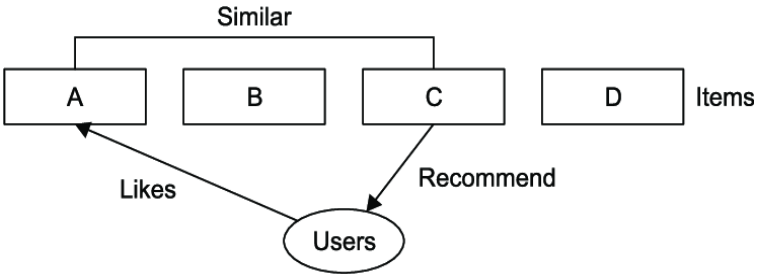
Every recommender of movie has a diverse custom variety and use condition, and diverse recommendation procedures are used for the similar data recommendation [1].

The model likely to be a hybrid recommendation system in the real application of the recommendation system. In short, the benefit of every recommendation procedures should be mingled with the recommended process to efficiently develop the effect of the recommendation. The vital research content in our paper is to aid users to naturally obtain movies interested in the enormous data of movie information using the content and the collaborative filter procedures and to implement a prototype movie recommender based on the content and collaborative filter procedures. Item's profile is a set of keywords (or features).

2. Literature Survey

2.1 Content based filtering

In this approach, items are recommended based on comparisons between the profile of item and the profile of user. A profile of user is content related to the user in the form of keywords (or features). A user profile can be seen as a set of assigned keywords (terms, features) collected by algorithm from items identified by the user as relevant (or interesting).



For example, consider a scenario in which a person goes to buy an item 'A'. Unfortunately, item 'A' has been sold out and as a result of this the shopkeeper recommends the person to buy item 'C' which is similar to item 'A'. This is an example of content-based filtering.

In case of Movie recommendation, the algorithm simply picks and recommend movie similar to what we like whether the recommendations will be based on genre, director of the movie, actors of the movie etc.

In the below example, the user watched some movie previously. For example Insidious, then the movie recommender going to recommend movies that are very similar to Insidious, it may recommend based on genre or director or the main actors. So the recommendation might contain a movie like conjuring.

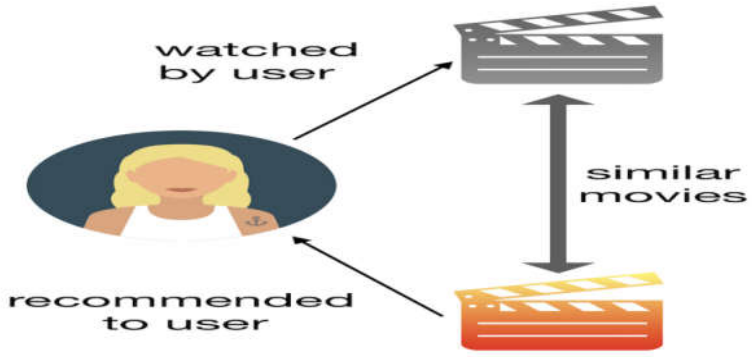


Fig: Content-based filter[21].

2.2 Collaborative based filtering

The collaborative filtering algorithm is classified as user-based and item-based collaborative filtering algorithm [5]. The two basic principles are quite similar, and this section mainly introduces the recommendation algorithm for collaborative filtering based on the user. The basic idea of the recommendation algorithm for collaborative filtering is to introduce users with similar interests.

For example, In the below figure First User loves movie A, B and third user likes movie A, B ,C so we can conclude that the preferences of first user and third user are very similar. So we can infer that the first user may also love item C, therefore item C would be recommended to the first user.

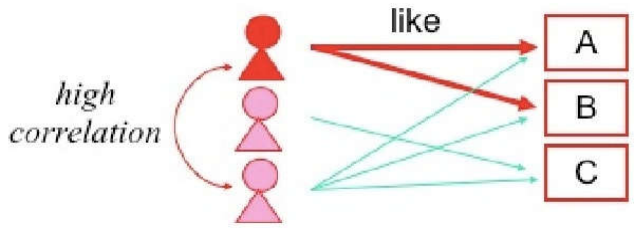


Fig: Collaborative-based filter.

In case of Movie Recommendation,this type of filter is based on the ratings of the users, and it will recommend movies that we have not yet watched for users like us. This filter considers both the movies watched and how they rated them to determine whether two users are similar or not. By looking at the common item, this type of algorithm basically predicts the ratings of a movie for a user who does not watched it.

In the below example,Lets say user A likes movie X and movie Y whereas user B like movie Y and movie Z .Now the Movie Recommender recommend movie Z to user A and movie X to user B.

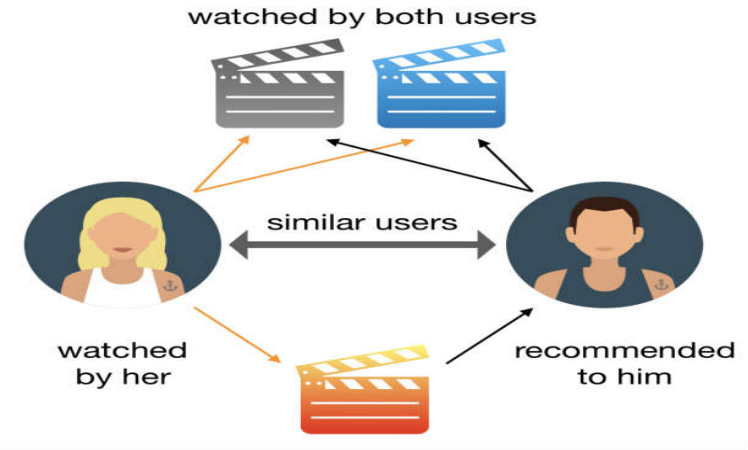


Fig: Collaborative-based filter [21].

3. RESEARCH METHODOLOGY

3.1KNN algorithm

K-NN algorithm is called algorithm of K nearest neighbor classification. The K-NN algorithm primary idea is that the sample is considered to be a part of some category if the majority of the most similar sample neighbors in the feature space are in a certain category.

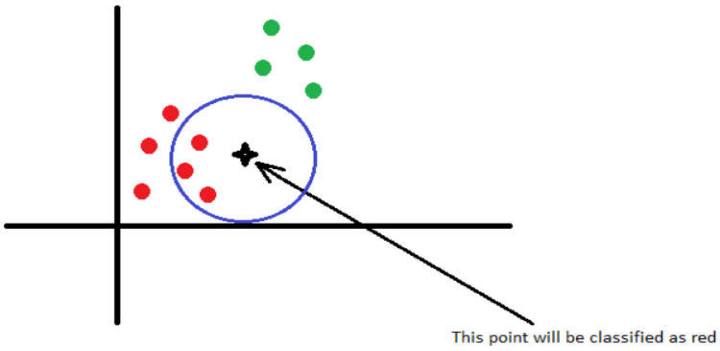


Fig: K-NN diagrammatic view

As presented in Figure, most of the neighbors closest to black point belong to the red point category so, we can conclude that black point belongs to the red point category.
To select neighbors, the K-NN collaborative filter approach, that is a collaborative filter approach mixed with the K-NN. The elementary steps of the algorithm are computing user similarity, selecting the nearest neighbor to KNN and computing the score[11,12].

3.2 User similarity computing

The similarity between users is calculated by evaluating the value of the items evaluated by two users. The regularly used method for computing user similarity is Cosine, Pearson Correlation and Euclidean distance similarity.

3.2.1 Cosine similarity

The technique calculates the similarity between two users by computing the cosine angle between the two user's vectors [19]:

$$sim(u,u') = cos(\theta) = \frac{\mathbf{r}_u \cdot \mathbf{r}_{u'}}{\|\mathbf{r}_u\| \|\mathbf{r}_{u'}\|} = \sum_i \frac{r_{ui} r_{u'i}}{\sqrt{\sum_i r_{ui}^2} \sqrt{\sum_i r_{u'i}^2}}$$

We can predict the rating of user- u for movie- i by taking a weighted sum of movie- i ratings from all other users (u's), where weight is the number of similarities between each user and user- u.

3.2.2 Pearson Correlation Similarity

The similarity of the Pearson correlation is ametric of the linear relationship between the two variables and **r** is the correlation coefficient. **[18].**
In according to find out the relationship between two variables X and Y, the following formula is used:

$$r = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2}\sqrt{(Y-\overline{Y})^2}}$$

Where, $\overline{X}$ = mean of X variable
$\overline{Y}$ = mean of Y variable

3.2.3 Euclidean Distance Similarity

When distance value between two Euclidean distances similarity is totally based on the distance from the items. This determines the distance among the items and in between the each point by forming coordinates to keep preference. We can conclude that two points

are not similar,if the distance between points is high. If the distance between the points is small, then we can say that two points are similar.

The Euclidean distance similarity formula is as below:

$$Euclidean\ Distance = d = \sqrt{\sum_{i=1}^{N}(Xi - Yi)^2}$$

3.3 KNN nearest neighbor selection

The algorithm picks a lot of users with the maximum similarity as the neighbor, after calculating the similarity among users. Set a fixed value K for neighboring selection,. Regardless of the value of neighboring user's similarity, select only the highest K similarity as neighbors.

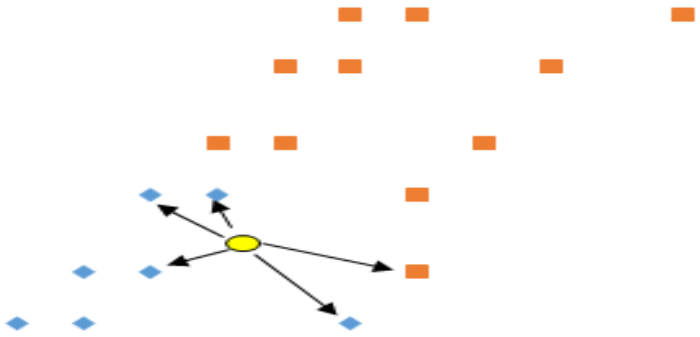


Fig: K-NN diagrammatic view [14]

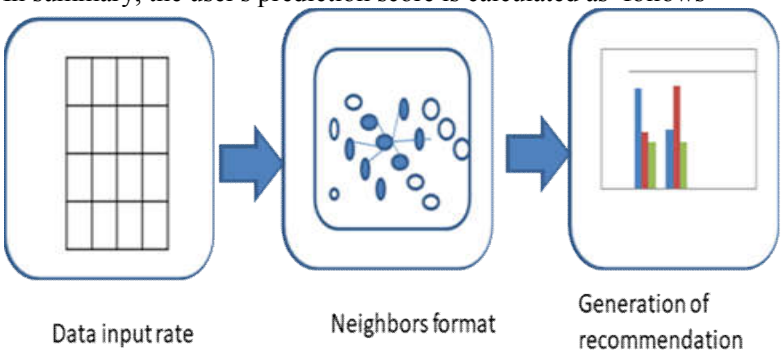
3.4 Predict score calculation

Once the user's neighbors have been determined, the score can be predicted according to the neighbor's score for the item [7].
The predictive rating of a particular user U of a particular movie i .That is computed by average weighted of all the neighbors score on the particular movie i, as follows:

$$P_{u,i} = A_u + \frac{\sum_{w=1}^{n}(R_{w,i} - A_w) \times sim(u,w)}{\sum_{w=1}^{n} sim(u,w)}$$

Where,
AW is the average rating of a user W ,who is neighbor to all other rated movies.
Sim (U, W) is the similarity between the neighbor user W and the target user U.
AU is the average rating of all the movies of a target user U.
R (W,i) is the rating of a target movie i, by the target user W.
Finally, n is the number of neighbors in the space.

In summary, the user's prediction score is calculated as follows



In this way, we can estimate the target user's prediction score for non- scored movies, and the user can recommend the N movies with the highest score. The KNN collaborative user- based filtering algorithm is used in this paper to implement the movie recommender [15].

4. Customized Movie Recommender design

The system is a simple Web Application where it is developed using different technologies. For User Interface we have used HTML, CSS and JavaScript and the main programming language is JAVA .The backend uses MYSQL for storage. It is very flexible and easy to maintain.

The user registration module in this system captures the complete and definite behavior and details of the user and the login module of user stores these behavior and details in the database of a user. The system will make the proper recommendation based on the user's behavior and past experiences or searched history of a user after logging into the system. As shown in the below fig:

Fig. The display of customized recommendation list [22]

K-Nearest Neighbor Merits and Demerits [16]

Merits:

1. Simple to understand
2. No data assumptions
3. Classification and regression can be applied
4. Works easily on problems of multiple classes

Demerits:

1. Intensive memory / Computationally costly
2. Sensitive to data size
3. The rare event(skewed) target variable does not work well
4. Struggle when there are a high number of independent variables.

5. Conclusion and Future Work

As long as massive information exists, the needs of the film amateur recommender system are going to be increase. This paper develops and implements a complete film recommendation system model based on the K-NN algorithm, collaborative filter approach and recommendation technology.[17]. we developed a customized recommendation engine for suggesting movies.

The recommendation system faces problems when the users change their taste and preferences. The system scalability can be improved by extending the analysis to more domains and sophisticated methods. Optimization of algorithms, query plan and transformations. We use combination of different algorithms to improve the accurateness in recommending.

References

[1] Cui, Bei-Bei. "Design and Implementation of Movie Recommendation System Based on Knn Collaborative Filtering Algorithm." *ITM Web of Conferences*. Vol. 12. EDP Sciences, 2017.*)*

[2] Peng, Xiao, Shao Liangshan, and Li Xiuran. "Improved Collaborative Filtering Algorithm in the Research and Application of Personalized Movie Recommendations", 2013 Fourth International Conference on Intelligent Systems Design and Engineering Applications, 2013.

[3] Munoz-Organero, Mario, Gustavo A. Ramíez-González, Pedro J. Munoz-Merino, and Carlos Delgado Kloos. "A Collaborative Recommender System Based on SpaceTime Similarities", IEEE Pervasive Computing, 2010.

[4] Al-Shamri, M.Y.H.. "Fuzzy-genetic approach to recommender systems based on a novel hybrid user model", Expert Systems With Applications, 200810.

[5] Hu Jinming. "Application and research of collaborative filtering in e-commerce recommendation system", 2010 3rd International Conference on Computer Science and Information Technology, 07/2010.

[6] Xu, Qingzhen Wu, Jiayong Chen, Qiang. "A novel mobile personalized recommended method based on money flow model for stock exchange.(Researc", Mathematical Problems in Engineering, Annual 2014 Issue.

[7] Yan, Bo, and Guanling Chen. "AppJoy : personalized mobile application discovery", Proceedings of the 9th international conference on Mobile systems applications and services - MobiSys 11 MobiSys 11, 2011.

[8] Davidsson C, Moritz S. Utilizing implicit feedback and context to recommend mobile applications from first use. 12, 04008 (2017) DOI: 10.1051/ 7120 ITA 2017 ITM Web of Conferences itmconf/201 4008 4 In: Proc. of the Ca RR 2011. New York: ACM Press, 2011. 19- 22.http://dl.acm.org/citation.cfm?id=1961639[doi:10.11 45/1961634.1961639]

[9] Kumar M, Yadav DK, Singh A, Gupta VK. A movie recommender system: Movrec. International Journal of Computer Applications. 2015 Jan 1;124(3).

[10] Bilge, A., Kaleli, C., Yakut, I., Gunes, I., Polat, H.: A survey of privacy-preserving collaborative filtering schemes. Int. J. Softw. Eng. Knowl. Eng. 23(08), 1085– 1108 (2013)CrossRefGoogle Scholar.

[11]] Hu Jinming. "Application and research of collaborative filtering in e-commerce recommendation system", 2010 3rd International Conference on Computer Science and Information Technology, 07/2010

[12] www.pacis2014.org

[13] Xie HT, Meng XW. A personalized information service model adapting to user requirement evolution. Acta Electronica Sinica, 2011,39(3):643-648 (in Chinese with English abstract).

[14] https://www.r-bloggers.com/k-nearest-neighbor-step-by-step-tutorial/

[15] Polat, H., Du, W.: Privacy-preserving top-n recommendation on distributed data. J. Am. Soc. Inf. Sci. Technol. 59(7), 1093–1108 (2008). http://dx.org/10.1002/asi.20831CrossRefGoogle Scholar.

[16] https://medium.com/@adi.bronshtein/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7Type equation here.

[17] Research.ijcaonline.org

[18] https://businessjargons.com/karl-pearsons-coefficient-of-correlation.html

[19] https://towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0

[20] Cui, Bei-Bei. "Design and Implementation of Movie Recommendation System Based on Knn Collaborative Filtering Algorithm." *ITM Web of Conferences*. Vol. 12. EDP Sciences, 2017.)

[21] https://towardsdatascience.com/how-to-build-from-scratch-a-content-based-movie-recommender-with-natural-language-processing-25ad400eb243

[22] https://towardsdatascience.com/prototyping-a-recommender-system-step-by-step-part-1-knn-item-based-collaborative-filtering-