International Journal of Research in Engineering, Science and Management
Volume-2, Issue-11, November-2019
www.ijresm.com | ISSN (Online): 2581-5792

Movie Recommendation System using AI and ML

Celeste Murnal[1], Deepika[2]
[1]*Student, Dept. of Computer Science and Engineering, New Horizon College of Engineering, Bangalore, India*
[2]*Sr. Assistant Professor, Dept. of Computer Science and Engg., New Horizon College of Engg., Bangalore, India*

***Abstract*: This paper presents an overview on Movie Recommendation System using AI and ML**

***Keywords*: movie recommendation system**

1. 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 is primary motivation in opting for movie recommendation system as my 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 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.

2. Filtering systems

A. Content based filtering system

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.

Fig. 1. displays the flow of information in a content based recommendation system. Relevant entities of an item and relations are kept together as input. Main features of items are extracted from item ontology. Features of items, user's ratings and user modeling data 30 are applied to content based recommender system. After applying, various recommended items are given as output.

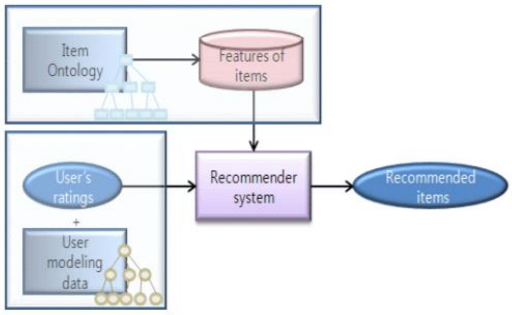

Fig. 1. Content-based filtering

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 that a user likes from the items that he does not like.

B. Collaborative filtering based system

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.

Fig. 2. Illustrates the flow chart of Collaborative Filtering Recommender System. It shows how collaborative filtering considers only numerical reviews given by different users and then gives recommended products as result. The user reviews are stored in a database to make further references and predictions. In the figure, User 1 and User 6 show similar behaviour and thus their profiles lie in the same neighborhood which indicates similar interests. Using this similarity, review about a product not rated by User 6 can be predicted using the reviews of User 1 that are available. Thus, a prediction regarding product C's review by User 6 is made using the available data. From these predictions, recommendations are extracted and suggested to the user.

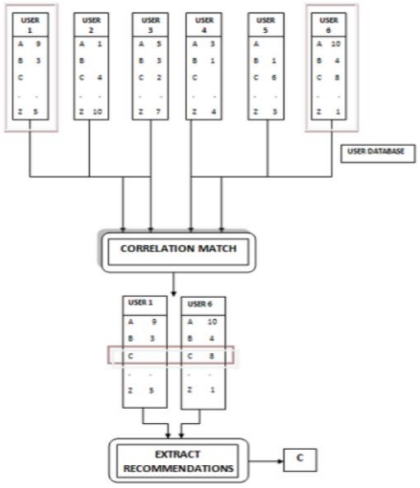

Fig. 2. Collaborative filtering

Active filtering is separated from passive filtering because using active filtering require the user to dedicate some time in order to rate the information items when, using passive filtering, users automatically provide data by only accessing the item. Another approach is the item based approach.

1) Active filtering (or explicit data collection)

Active filtering is a method for collaborative filtering because of its peer-to-peer matching approach. Various profiles from peers are matched to extract similar interests. This approach is based on the facts that peers exchange information such are ratings and appreciation of specific items. It mimics the natural approach of peers recommending shops to each other. This type of filtering is particularly effective in cases where people are not knowledgeable about the mass of information available to them. One of the main advantages of active filtering is that the information rating is provided by an actual person who has viewed the item with interest. Another advantage in heavily social-oriented systems is that it gives the opportunity for willing people to be heard and provide highly relevant information. The main disadvantage is that this system requires some action by the user and thus makes the data more expensive to obtain and rarer. Another incidence of having an action required is that the feedbacks provided might be biased, for example towards a negative or positive experience, depending on the target customer. Another issue of those content filtering systems comes from the averaging effect occurring in some specific situations. Over a range of similar items, the system will not know the differentiating characteristics between items. This ultimately often causes the most popular items to be recommended more often as they will have more ratings. The First-Rater problem occurs for new items with no previous rating and the Cold start problem occurs for new users with no previous preferences.

2) Passive filtering (or implicit data collection)

Passive filtering is user to collect information implicitly. Some examples are: purchasing an item using, saving printing, modifying, commenting repeatedly on an item Referring or linking to a site (in another context than only rating, for example social media) Number of times an item is queried Time measurements to determine if the user is scanning, reading or working with a document. The main advantage of passive filtering is that it broadens the population of user providing feedbacks. In effect, only some populations of users come back to the system to rate items whereas all connect to the system to access the item. Their behaviour during that phase is most likely able to provide information about their interest.

3) Item based filtering

In this filtering approach, items are rated and used as parameters for the matching instead of users. The items are grouped together and proposed to users. Users can then compare and rate them. User preferences are collected explicitly. Those preferences allows to group users by interest. The items are then selected using the ratings of a similar user.

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 behaviour.
- 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 exceeds the number of users by a great margin

which makes it difficult to find items that are rated by enough people.

3. 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 pre-processing 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.

A. Hybrid algorithm

Step-1: Use content-based predictor to calculate the pseudo user-rating vector 'v' for every user 'u' in the database.

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

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

Step-2: Weight all users with respect to similarity with the active user.

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

Step-3: Select n users that have the highest similarity with the active user.

- These users form the neighbourhood.

Step-4: Compute a prediction from a weighted combination of the selected neighbours' ratings.

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

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

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 neighbour's mean:

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

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 neighbourhood.

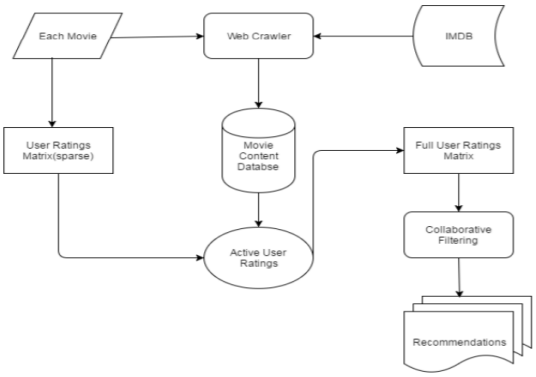

Fig. 3. System Overview

4. Conclusion and future work

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, 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, Volume 5, Issue 2, April 2015.