International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 6, Issue 2, March - April 2017 ISSN 2278-6856

Perspective Analysis Recommendation System in Machine Learning

Ayush Khare, Dr Naveenkumar Jayakumar

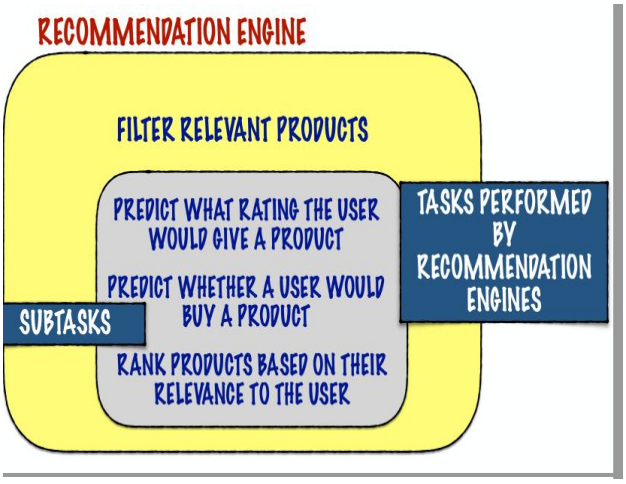
Bharti Vidyapeeth Deemed University College of Engineering
India, Pune

Abstract-*This paper explores the recommender system and the different ML Algorithm approaches that are used in order to provide an efficient system that provides internet users with requested information. Knowledge on these different techniques of obtaining information from users. This paper is a crisp yet primer introduction to Recommendation System and covers different aspects and techniques of recommendation systems like content based filtering, Collaborative filtering and comparison between them.*

Index Terms-Algorithm, Machine learning, filtering

I. INTRODUCTION

The continuous growth in the volume of digital information accessible and the number of visitors to the internet has created a possible problem of information overload which delays opportune access to items of importance on the internet. This problem has been partially solved by information retrieval systems such as AltaVista, Google and Devil Finder but there is a lack of systems to personalize and prioritize information data 1. This absence has created demand for recommender systems. According to Tejeda-Lorente; recommender system are information filtering systems that address the issue of information overload by filtering out important information portion of the great dynamically engendered data based on the interests, preference and perceived behavior regarding the item 2. Recommender systems are of great benefit to internet users and service provides. These systems also lower the cost of transaction of searching for and selecting items in an online shopping setting. Recommender setting are also beneficial as they improve the decision-making process. The picture below illustrates a RS -



II. RELATED WORK

Machine Learning (ML) is used to create algorithm used in recommender systems to provide with better recommendation for users. ML however lacks clear classification scheme for algorithm mainly because of different approaches and difference in suggested literature 3. As such, it is a daunting task to choose an ML algorithm that is suitable for a particular need when developing recommendation systems. In addition, searchers might find it thought-provoking to track the use of ML algorithm in recommendation system.

One approach to consider in ML is the collaborative approach which considers user information when processing data for recommendation. With such information, the recommender system is able to identify users that share the same music preference and then prefer information or product that I preferred by the same users.

Another technique that can be employed using the ML algorithm is content-based filtering approach. Content-based filtering approach bases its recommendation on the item information they can access. If the user is looking for a particular phone the recommender system browser phones with similar attributes from the data base and presents the finding to the user.

The third classification criterion of ML is the hybrid filtering approach which recommends items based on both the preference and item information. A good example is the social media where the recommending system is used to suggest user profile that have some relation to users i.e. use of collaborative filtering. After obtaining user profiles, the recommending system suggest profile as item information and thus the searching user is able to access information to search for 'n' such profiles i.e. this is based of content-based filtering. After the process is complete both sets of profiles are returned are recommendations[1].

While employing the hybrid or collaborative approach it is necessary for the research system to gather essential about the user in order to generate recommendations. This technique can be done by two techniques both implicitly or explicitly. Explicit user information gathering deals with the user having the full knowledge that they are issuing their information; while the other implicit information gathering filtering deals with collecting information indirectly i.e. without the knowledge of the users.

After deciding on the technique to use either implicit or explicit this it is important to look at learning phase that will incorporate the algorithm which will exploit user's information that are response from feedback gathered from the information gathering phase. After this phase, the recommending system is able to give appropriate prediction on requested information. This process can be direct based on the collected information which could be memory based or model based via system system's observed activities of the users.

It is important to use accurate and efficient recommendation techniques which is necessary for the system to provide feasible and credible recommendation to its individual users. This explains the significant of comprehending the features and potential of different methods of recommendation.

According to Lu et al. authors conduct systemic authors conduct systemic review of ML field to analyze the development of recommender system, which incorporates ML algorithm to see if there could be more systemic approach to engineering of recommender systems. Such reviews have two main objectives:

1. Identifying which ML algorithm are incorporated in recommender systems
2. Identifying and addressing issues in the development of the recommender systems that might be affected by software engineering.

It is also important to look at the filtering techniques that will be applied to prioritize data, one can use memory based and model based technique

Collaborative based technique

It makes use of rated items which helps recommending to other users with whom they share similar preferences. This model based technique relies on two mechanisms; user-based and item-based mechanism. User-based technique calculates the predicted ratings between users as the weighted average of ratings where weights are similarities of users on the elected target. Pearson correlation method which is employed to measure correlation between two variables is represented in the formula below:

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u) \cdot w_{a,u}}{\sum_{u \in U} |w_{a,u}|}$$

Where

$P_{a,i}$ = Predicted rating of active user a for product i

r_a = Average Rating of the active user a for any product

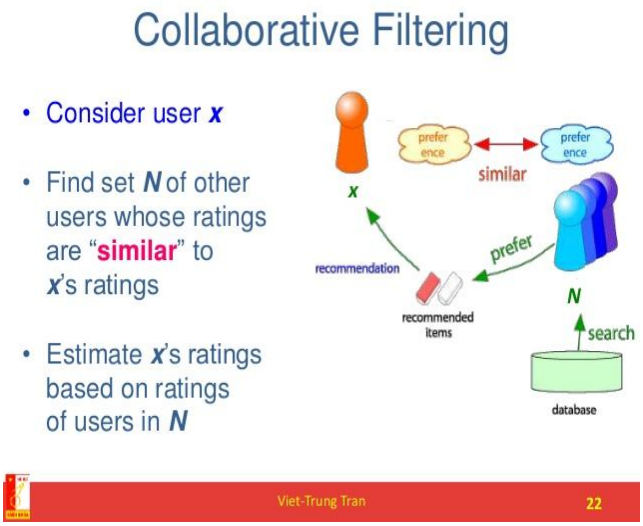
U = Set of nearest neighbors of active User a

$r_{u,i}$ = Rating of a user U For Product I

r_u = Average rating of U

$w_{a,u}$ = Similarity between user u and active user a

The pic below depicts the basic idea of collaborative filtering



Content based technique

Official word of Recommendation System (recommendationsystem.org) states "Content-based filtering, also referred to as cognitive filtering, recommends items based on a comparison between the content of the items and a user profile. The content of each item is represented as a set of descriptors or terms, typically the words that occur in a document. The user profile is represented with the same terms and built up by analyzing the content of items which have been seen by the user". Generally, products are represented in terms of descriptors or attributes manually and then a User Profile is generated and based on that recommendations are forwarded to the user.

III. RESEARCH TEST BASED COMPARISON OF BOTH SYSTEMS

Although the models and have similar (but not the same) output, they have different working mechanism and this is what can be used to distinguish each from the other. The Collaborative based technique makes use of previous user rating while Content Based technique makes use of defined user attributes. So, derive a test on a dataset of movies, which will recommend a user 5 -6 movies on the basis of his rating history.

Objective – The Objective is to use the given data in Movielens and recommend the User 5 movies which he hadn't watched or rated

Prerequisites

"MovieLens is run by GroupLens, a research lab at the University of Minnesota. By using MovieLens, you will help GroupLens develop new experimental tools and interfaces for data exploration and recommendation. MovieLens is non-commercial, and free of advertisements."

This file has 100K data points, each row is a rating given by 1 user for 1 movie at a particular date and time

Pandas is a python library for data analysis in a way that's similar to dataframe manipulation in R. We can read the

data from a csv, write to a csv, manipulate it into different shapes, subset the data based on conditions etc.

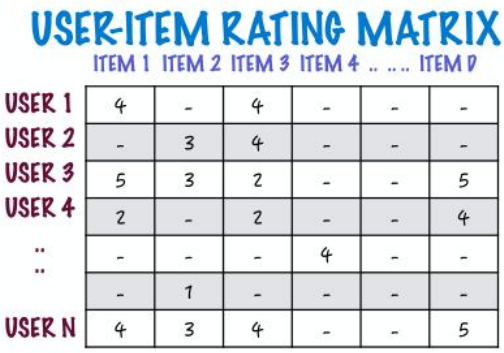
IV. COLLABORATIVE BASED FILTERING

Collaborative Filtering allows us to give a predicted rating on the basis of user behavior

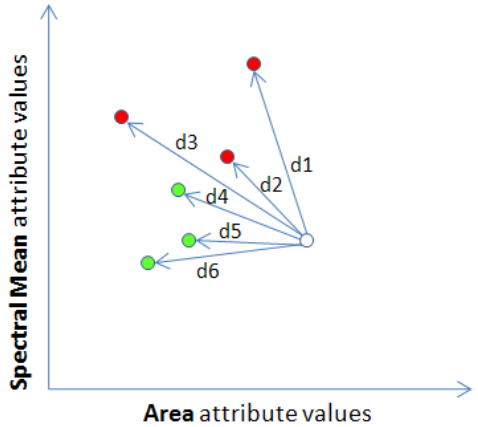
Algorithm

• User who give same ratings for a bunch of product are similar to each other and the K – nearest neighbors are the users who are most similar to each other

• So, we will find the weighted neighbors by forming a user – item matrix table



• We find the similarity by plotting all these users as vectors on an imaginary plane



• We can use the following methods to find the difference between two vectors (or tuples)

i) Euclidean Method

ii) Cosine Method

iii) Pearson correlation

• In the above problem, we used the Cosine relation to find out the distance

• After finding the K nearest neighbors we will find the weighted average rating of the product

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u) \cdot w_{a,u}}{\sum_{u \in U} |w_{a,u}|}$$

Where

$P_{a,i}$ = Predicted rating of active user a for product i

r_a = Average Rating of the active user a for any product

U = Set of nearest neighbors of active User a

$r_{u,i}$ = Rating of a user U For Product I

• The movies with the top 5 predicted ratings are printed

V. RESULTS

We got the following results when we implemented the Algorithm

```
Shawshank Redemption
Fargo
Usual Suspects
Forrest Gump
Orphan
The Perfect Murder
```

VI. IMPLICATIONS FROM RESULTS

Positives

• We can see a diversified range of movies among all of them not related to a particular genre, actor or target audience.

• Gives a clearer feedback about user behavior choices and more objective in nature.

Negatives

New items (movies) cannot be predicted those had been might have been of a user's interest but not received that much ratings in numbers to be qualified

VII. CONTENT BASED FILTERING

• A content based recommendation system works with the data that user provides with the data either explicitly or implicitly based that user is directed towards his/her recommendations

• In this example, each movie is stored as a vector of its attribute –

i. Target Audience

ii. Genre

• These attributes are solely described on the perspective of user analysis

• Each movie is stored in a 2d space and angles between vectors are calculated determine similarity between vectors

• We found the cosine values to find out the similarity between the items
• The reason for using cosine is that value of cosine increase with decrease in angle which signifies more similarity

Results

After the complete analysis of MovieLens data with above algorithm we got following results

Contact (1997)	509
Fargo (1996)	508
Return of the Jedi (1983)	507
Liar Liar (1997)	485
English Patient, The (1996)	481
Scream (1996)	478

VIII. IMPLICATIONS FROM RESULTS

Positives

• This algorithm is suitable for new products and also items which are not rated that much in numbers

Negatives

• We can clearly see these movies are all in commercial dramatic plane even if user had watched some artistic comedies as well the movies will be outnumbered by the algorithm
• A lot of manual work is involved in the algorithm
• The idea of commercial and artistic is not dependent on expert's perspective hence can't be objective

IX. CONCLUSION

The recommender system is an essential inclusion to technology as it opens up new prospects of recovering tailored information on the internet. The recommender system also helps alleviate the issue of information overload which is a very common occurrence in systems that deal with information retrieval and this advantage enables users to readily access products and services that are not available in other users on the system. This paper has explored the recommender system and the different ML algorithm approaches that are used in order to provide an efficient system that provides internet users with requested information. Knowledge on this different technique of obtaining information from users, ML algorithm will help software engineers with necessary knowledge and act as a blueprint to improve and revolutionize recommendation techniques.In the above research we found Collaborative filter better in the case of movies than the Content based Filter.

REFERENCES

[1] J. Lu, D. Wu, M. Mao, W. Wang and G. Zhang, "Recommender system application developments: A survey", Decision Support Systems, vol. 74, pp. 12-32, 2015.
[2] Á. Tejeda-Lorente, C. Porcel, E. Peis, R. Sanz and E. Herrera-Viedma, "A quality based recommender system to disseminate information in a university digital library", Information Sciences, vol. 261, pp. 52-69, 2014.
[3] T. MA, J. ZHOU, M. TANG, Y. TIAN, A. AL-DHELAAN, M. AL-RODHAAN and S. LEE, "Social Network and Tag Sources Based Augmenting Collaborative Recommender System", IEICE Transactions on Information and Systems, vol. 98, no. 4, pp. 902-910, 2015.