International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

A HYBRID RECOMMENDATION SYSTEM MODEL FOR OPTIMIZED RESULTS

Aditya Adkar[1], Gaurav Jaisinghani[2], Ashlesh Chincholkar[3],Prof. Riyaz Jamadar[4]

[1,2,3]*Students, AISSMS'S Institute of Information Technology, Pune, Maharashtra, India*
[4]*Professor, AISSMS'S Institute of Information Technology, Pune, Maharashtra, India*

ABSTRACT: *The advent of personalised technology came into being with a need for identifying customer requirements. One such important technology is Recommender Systems. Recommender Systems play a significant role in personalising decision making. The proposed system looks at different existing models with their advantages and drawbacks and amalgamates the advantages to build hybrid model that accomplishes the task of enhanced recommendation system. The proposed system also looks at optimizing recommendation by making the use of different algorithms and word clouds*.

KEYWORDS: ***Matrix-Factorization; Collaborative Filtering; TF-IDF Vectorizers; Word Cloud; Content Based Filtering***

1. INTRODUCTION

Recommendation Systems are an integral part of the technology ecosystem. They are used in a plethora of fields ranging from E-Commerce to Online streaming services.

The proposed system leverages advanced algorithms in ML to build a recommender that uses hybrid methodologies namely collaborative filtering and content based filtering. We also look at different solutions for the cold start problem which occurs in the system that requires collaborative filtering.

The first module in the proposed system uses User-User as well as User-Item based collaborative filtering. Collaborative Filtering uses a methodology that identifies the similarities between users particularly the User-User based Collaborative filtering. User-User based collaborative filtering is derived from the idea that users that rate or buy similar products generally have similar tastes. Consider a Subject 1 who buys object A, object B and object C. Subject 2 who has also bought object A and object B will now be recommended object C as the user profile of Subject matches that of Subject 1 based on purchase history.

The User Item Collaborative filtering on the other hand considers the ratings given by two different users to different products to recommend products. Collaborative filtering also takes into consideration memory based models. Content based filtering is a classified in the taxonomy of recommendation engines as the system that recommends on the basis of similarities in features of different items in a dataset. In our system we approach CB filtering using TF-IDF vectorizers.

The TF-IDF method assigns weight to various words in a given document and prioritizes the importance of words depending on its weight. TF or Term Frequency refers to the number of times a Term occurs in a given document. IDF or Inverse document frequency refers to the measure of significance of a particular word in the document. Finally in the content based method we use TF-IDF matrix to calculate important recommendations. The proposed system also aims to solve the infamous Cold Start problem by making use of popularity based recommendation. Cold Start problem refers to the non – ability of a system to be able to recommend due to the users non – existent history on the system. Popularity based modelling is the most common and one of the earliest ways of recommendation system modelling. It ranks products based on number of units sold and suggests the most sold product to a brand new user to get started with the process. Finally, we combine the results of all the three previous models and build a hybrid system that relinquishes the redundant recommendations and as a result the final recommendations from the system are much more enhanced than any of the individual module by itself. We also take a look at different Python libraries and visualisation techniques that help better our understanding of improvised recommendation.

2. COLLABORATIVE FILTERING USING MEMORY AND MODEL BASED APPROACHES

$$\text{Similarity}(p,q) = \cos\theta = \frac{p \cdot q}{\|p\|\|q\|} = \frac{\sum_{i=1}^{n} p_i q_i}{\sqrt{\sum_{i=1}^{n} p_i^2}\sqrt{\sum_{i=1}^{n} q_i^2}}$$

Fig- 1: Cosine Similarity

In the proposed system we aim to utilize collaborative filtering with memory based modelling and model based approaches. The Memory based approaches are immune to learning from gradient descent. A common method used in memory based approaches is the Cosine Similarity method. This method is used for User-User recommendation where the rows in a matrix denote the user's items and the columns denote item ratings. The cosine similarity is measured by the cosine of angle between two vectors. Figure 1 shows the formula for calculating cosine similarity. We can demonstrate the working of cosine similarity by considering the example of User 1 and User 2. The prediction of User 1's rating for a Product A can be calculated by taking weighted sum of Product A from all other users where weighting is similarity number between User 1 and all the other users. Memory based approaches like Cosine similarity are easy to use, but as there is no optimization they can quickly fall apart if the data is sparse and can suffer in terms of scalability.

For this purpose we can use Model based Collaborative Filtering. Model Based Collaborative Filtering consists of different techniques such as Matrix factorization where rows and columns represent Users and Ratings respectively.

One such MF algorithm used in the system is the Singular Value Decomposition or SVD algorithm. SVD is a method used to reduce a given dataset from K dimensions to N dimensions (N<K). It represents a matrix structure where the row represents a user and the column represents a rating. SVD allows the existing matrix to be broken into smaller matrices. Where one matrix represents a utility matrix, the other represents a orthogonal left matrix which represents the relation between users and latent factors. Two other matrices represent a diagonal and diagonal right matrix respectively. The latter two matrices are used to find the relation between items and latent factors. SVD therefore helps find recommendations on the basis of intersection of rows and columns in a matrix depending on the dataset.

To overcome the redundancies of method based techniques we use model based techniques in Collaborative Filtering.

Model based techniques are used instead to help solve the redundancies of memory based techniques. In this approach ML algorithms are used to predict user's rating of unrated items. There are various types of Clustering and Matrix Factorization algorithms that can be used to obtain model based systems.

2.1 Matrix Factorization in Collaborative Filtering

Matrix factorization or decomposition is a method which can be used to decompose the matrix to its constituent parts. Two different types of algorithms that we can use are Probabilistic Matrix Factorization and Non Negative Factorization.

Probability Matrix Factorization uses Bayesian Learning Techniques.

PMF considers two parameters A and B with a given dataset C.

With the required training we will get a revised matrix C*. That will also get ratings for user cells which were initially empty. We can therefore use these revised ratings to make predictions. On the other hand in Non - negative Factorization, two arrays A and B are considered. Dimensions of the array are defined by the rows in the dataset. The dimensions are also reduced in Non- negative factorization for yielding better results.

To sum up Matrix factorization is the decomposition of an a*b matrix into multiple matrices which are a*c and b*c. Matrix factorization is further divided into LU factorization, Cholseky decomposition and QR matrix decomposition. Matrix factorization can be of help in

recommendation by considering the intersection of rows which denote users and columns that denote the ratings.

MF algorithms are also used for reducing the dimensionality of a given dataset to perform on using methods such as PCA.

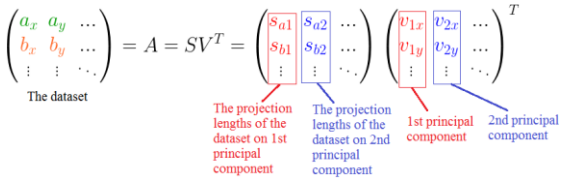

Fig- 2: Single Value Decomposition

3. CONTENT BASED FILTERING

Content Based Filtering refers to filtering based on the contents of products rather than the ratings based on different users as seen in Collaborative Filtering.

Content based filtering is optimal for people with large buying history as it is easy to encapsulate the data and its contents consumed by the user and by using labelled data match and recommend new products. However content based filtering requires more amount of labelled data to work with compared to Collaborative Filtering.

The algorithm used for Content Based Filtering in the given system is the TF-IDF algorithm.

This algorithm counts the frequency of a term occurring in a given document followed by Inverse Document Frequency which weighs how important a particular word is and based on the weightage and the occurrence of the word priorities are matched.

Content based filtering also uses word clouds to mine important words in a given document to get a general idea about the sentiment towards a particular product. This makes it easier to understand and visualize the complex system of content based recommending.

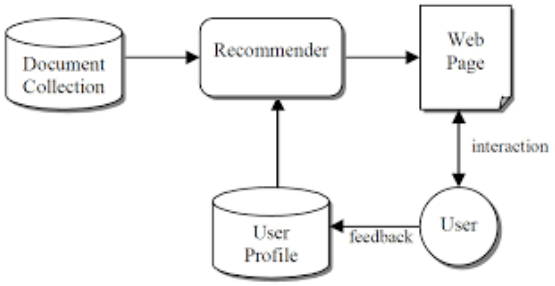

Fig-3: Content Based Recommender

Content based recommendation therefore works on large amounts of labelled data.

4. IMPLEMENTATION

For implementing the proposed system it is important that we have Jupyter notebook installed and important libraries like pandas and scikit learn which will help us with importing algorithms and help predict the accuracy score of the recommendations.

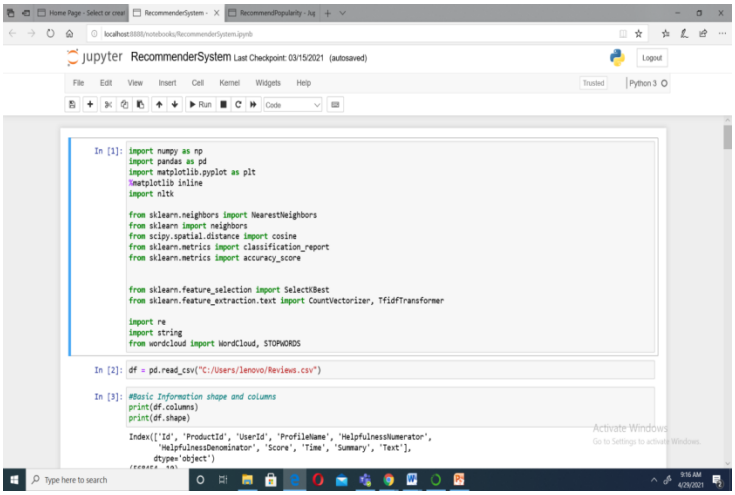

Fig- 4: Import Important Libraries

This is followed by computing the mean value and count and grouping them by products.

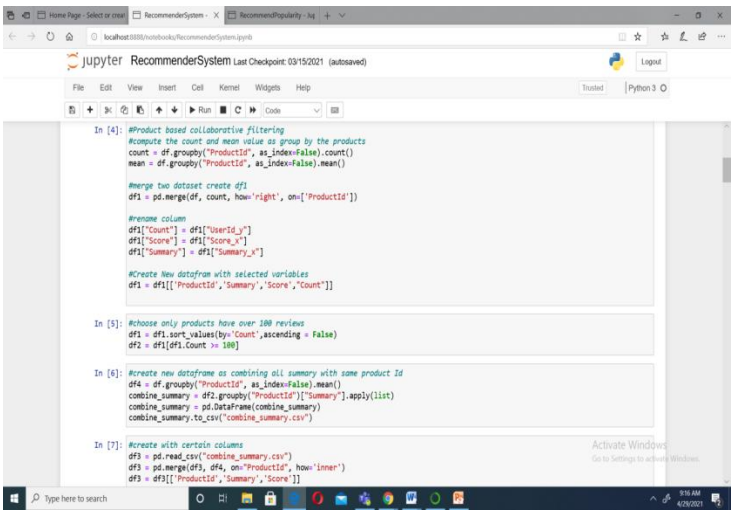

Fig-5: Grouping the data

Next step in the process is to clean the data and perform dimensionality reduction for better results.

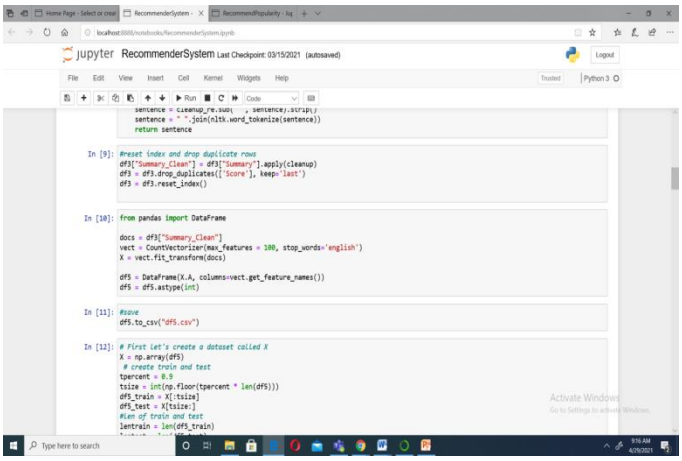

Fig- 5: Data cleaning and Dimensionality reduction

After applying the required algorithms we get the recommendations based on collaborative filtering and content based filtering.

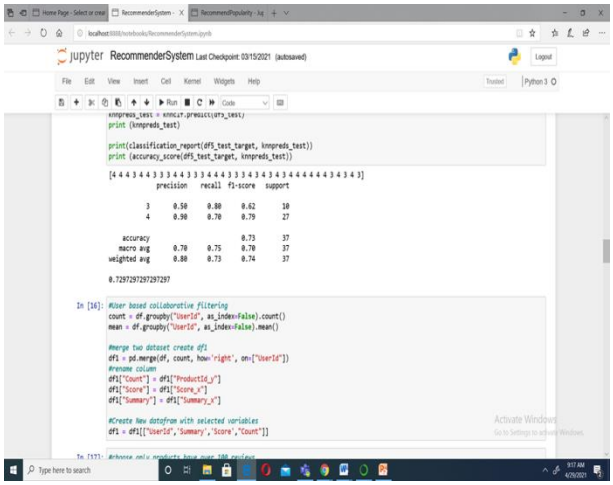

Fig- 6: Recommendation and Accuracy Score

5. CONCLUSIONS

The system built using two modules of Collaborative Filtering and Content Based Filtering was able to perform recommendations as required with an optimal accuracy score making it more effective than using any one of the two individually. We also tested different algorithms used by both the modules for the most relevant results. While Matrix factorization algorithms such as SVD and Cosine Similarity were used in the first module for Collaborative Filtering, TF-IDF algorithms were used in the second module for Content Based Filtering.

Finally we were able to amalgamate the two models to build a hybrid model which optimized the recommendation process.

6. REFERENCES

[1] Bogdan Walek, Petra Spackova. Content-based recommender system for online stores using expert system. 2018 IEEE First International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)

[2] Srikar Amara, R. Raja Subramanian. Collaborating personalized recommender system and content-based recommender system using TextCorpus. ICACCS 2020.

[3] Christos Troussas, Akrivi Krouska, Maria Virvou. Multi-algorithmic techniques and a hybrid model for increasing the efficiency of recommender systems. 2018 IEEE 30th International Conference on Tools with Artificial Intelligence

[4] Richa Sharma, Shalli Rani, Sarvesh Tanwar. Machine Learning Algorithms for building Recommender Systems. Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2019)IEEE Xplore Part Number: CFP19K34-ART; ISBN: 978-1-5386-8113-8.

[5] Liao Shaowen, Chen Young. An improved collaborative filtering model. 2017 International Conference on Smart Grid and Electrical Automation

[6] Chia-Yu Lin, Li-Chun Wang, Kun-Hung Tsai. Hybrid Real-Time Matrix Factorization for Implicit Feedback Recommendation Systems.

[7] Ha-Ram Won, Yunju Lee, Jae- Seung Shim, Hyunchul Ahn, A Hybrid Collaborative Filtering Model Using Customer Search Keyword Data for Product Recommendation. 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)