An Approach on Content-based, Collaboration-based & Hybrid Movie Recommender

AUTHORS:

Sankalp Kulkarni *,Student @ MIT ADT University*

Sasha Kumar *,Student @ MIT ADT University*

Mitrup Kabi *,Student @ MIT ADT University*

Rushab Kothari ***,****Student @ MIT ADT University* ***&***

Prof.Vishal Patil , *Asst. prof @ MIT ADT University*

Department of Computer Science and Engineering, MITSOE , MITADT University, Rajbaug, Loni kalbhor, Pune, India

---

ABSTRACT : A recommendation system is a filter mechanism that filters the data using different algorithms and recommends the most relevant items to the user. A recommendation system can be of various types such as movie recommender , music recommender , books recommender , etc. It first captures the past behavior of a customer and based on that, recommends products which the users might be likely to buy . Let's consider that there is a new user who has no past history with the site or UI , accesses the e-commerce site to get recommendations then what will the recommender do ? So here one of the possible solutions can be , recommend the best selling products , i.e. highly demanded products. Branded products etc. Another possibility is to recommend the products which can bring profit to the business or user. Here we will talk about the movie recommender ; which is based on two types of recommender engines , i.e. Content based and collaborative based engines , which combine gives a hybrid based recommender. These are the three main approaches of our recommender system. In content-based filtering, we try to profile the user's interests using information collected, and recommend items based on that profile. In collaborative filtering, we try to group similar users together and use information about the group to make recommendations to use. The main purpose of the project is to Research and learn about Content and collaborative based approach. Nowadays recommender systems have become an important part of our day to day life. Where one compares their work to others and gets recommended improved suggestions. Our project tries to collect data from various sources/datasets , and recommend it to the user.We have tried to study various aspects of the recommendation systems , By comparing a dataset of movies of nearly 10,000 movies. There exist several recommender systems in the market which recommend a wide variety of goods to users. These recommenders provide goods by taking the user's interest , history , location, and much more.Apart from everything individuality is common.

KEYWORDS: Simple recommenders, Content-based recommenders, Collaborative filtering engines, hybrid recommenders, dataset and prototypes .

INTRODUCTION

A movie Recommender system can be an information dividing system that works to analyze and estimate various aspects and data such as ratings, genres and scores to give preferences to the user and will provide the data required in a simple and sorted manner. The recommender system enhances the experience for News , websites , games , social media platforms , etc.

A content and collaborative-based recommendation system can also be a method of information sorting system that works to predict user preferences and provide suggestions that support them.The content on some platforms extends from movies, music, books and videos to friends. The two main approaches used for recommender systems are content based filtering , where we attempt to profile ; a client's interest utilizing gathered information and recommend the items to the client that supports the profile. Second , Collaborative filtering where we try to identify together and use information about different client IDs to create a recommender system. Every user has a different taste , likes and dislikes . For example let's say Sai likes hip-hop music , and Rajat likes traditional music. These two have a completely different taste of music , if these two IDs data is uploaded on recommender the recommended songs will be the combination of these two types of music. The type of music you want to focus on during exercise is severely different from that in which we listen to music while making dinner. They have to find new areas to see more about the customer, while still determining the majority of what is already known about the customer.

THE RECOMMENDER

The recommendation module is a different kind of data filtering. Preferences and ratings are used to select the item , that item will be given to a user. The recommendation module uses content-based, collaborative and hybrid filtering techniques .

A simple recommender provides generalized recommendations to each user related to any of the content in this case movies; supported by popularity based on genre and ratings.The idea behind this is that more acclaimed and probably famous movies are better likely to be watched and liked by the audience.

THE CONTENT BASED RECOMMENDER.

The content based recommender process checks for user interest and user based profiles and recommends similar items supported to selected items. This type of filtering uses Metadata for films , such as genre , ratings, director description , overview , Content etc.Content-based (cognitive filtering) recommendation systems recommend using items that support comparisons between the content and user profiles of those things.

Advantage of content-based filtering: -

1.This has the capability of mentioning unrated objects. It is easy to explain the work of the recommender system by cataloging the content features of an item.

2.Content-based recommender systems require only the rating of the concerned client, no one else's.

3.Operator of the system.

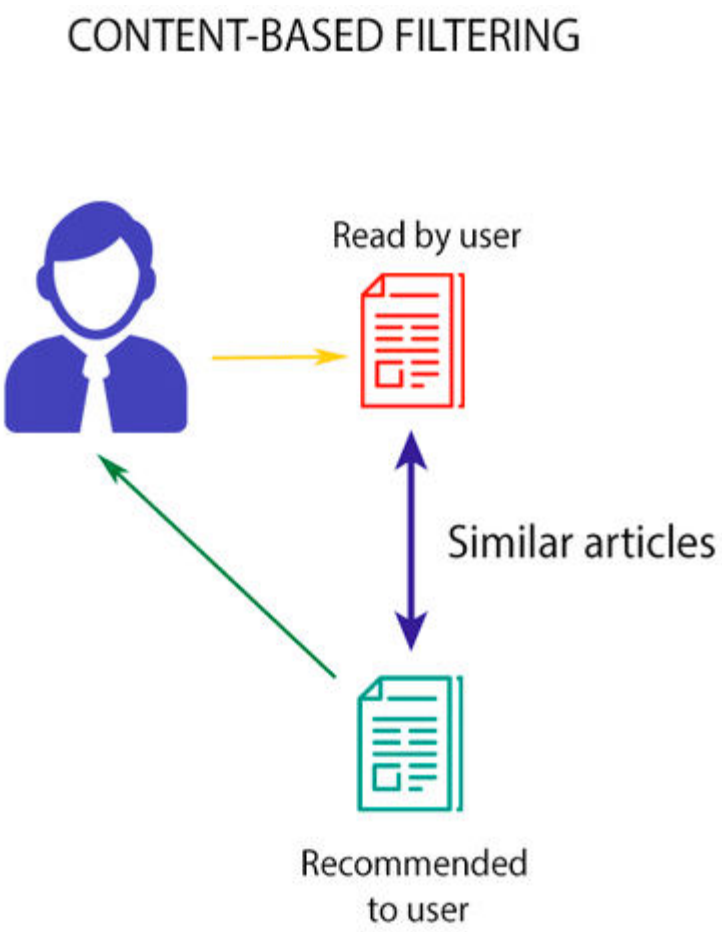
Disadvantage of content-based filtering: -

1.A new user who has not rated any item will not work for them, as enough ratings are required content-based recommender assesses the client inclinations and gives exact proposals.

2.No proposals of unexpected objects.

3. Restricted Content Study - If the system fails to distinguish the items the recommender does not work.

4.A client likes items that he/she does not like.



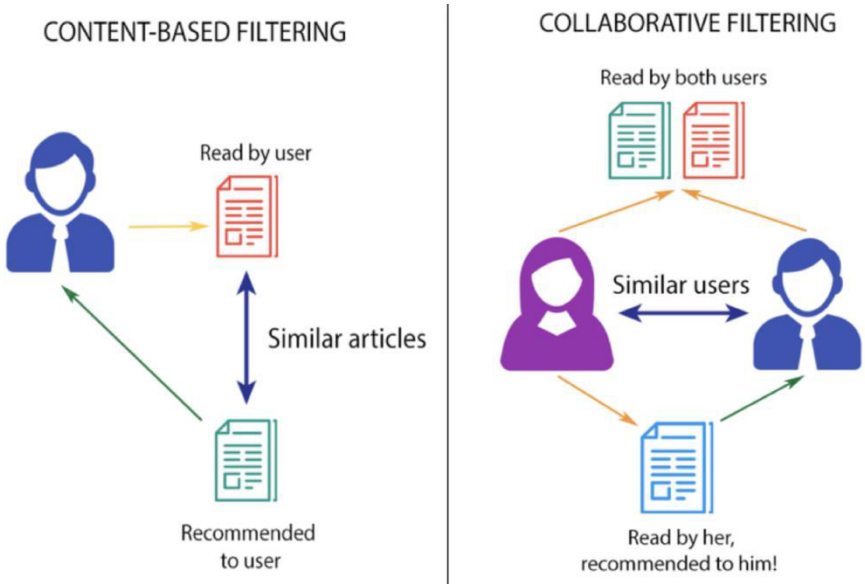
https://miro.medium.com/max/828/1*1b-yMSGZ1HfxvHiJCiPV7Q.png

COLLABORATIVE FILTERING

Collaborative filtering methods make recommendations and demonstrations for user-supported ratings and data from multiple users' preferences. These systems attempt to rating or preference and obtain that the client will provide an object based on the previous rating and preferences of different clients. These methods do not require metadata for objects like content-based equivalent.Collaborative filtering methods approach two filtering techniques, such as model-based or memory-based. These techniques define a model based on user-item interactions where the user or item representation has to learn from the interaction matrix. Memory-based filtering techniques do not interact with user-items and depend on match among clients or objects in terms of observed interactions.Model based filtering technique has several approaches, such as clustering technique, association technique, Bayesian network , neural network and so on. Memory based filtering

technique has two approaches, such as client based and object based.The collaborative uses recommended filtering in association with user or movie IDs.

Suppose there are two users ‘Krishna’ and ‘Veer’, Krishna likes movies Star trek, flash, Batman, Welcome and Veer likes movies flash, Batman, Welcome, Super Singh. Such as common movies flash, Batman and Welcome. User prefers; therefore, Star trek movie is being recommended for Veer and Super Singh movie is recommended for Krishna . We see collaborative filtering technology in action on various kinds of online video platforms like Amazon Prime Video, Netflix, Zee5, YouTube and Facebook.We explore two procedures for collaborative filtering, the k-nearest neighbor algorithm and also the latent factor procedure.



https://www.google.com/url?sa=i&url=https%3A%2F%2Fmedium.com%2Fanalytics-vidhya%2Fcollaborative-based-recommendation-system-using-svd-9adc5b6b3b8&psig=AOvVaw15H7rgHp0WbEwTyp8xgDOk&ust=1625594565488000&source=images&cd=vfe&ved=0CAoQjRxqFwoTCKCO9evBzPECFQAAAAAdAAAAABAN

Advantage of collaborative filtering-based filtering:-

1. It relies upon the connection between clients which infers that it is content-autonomous.

2.Collaborative Filtering recommender systems can propose unexpected items by detecting like-minded people's behavior.

3.They can make genuine quality evaluations of objects by considering other people's experience.

Tfidf Vectorizer :

Scikit-learn's Tdif-Vectorizer is a section of raw documents to a matrix of TF-IDF features.. This article shows you how to correctly use each module, the differences and some guidelines on what to use when.

IMPLEMENTATION

In this segment, we have explained the way to be implemented in our movie recommendation engine / system using python programming language using K-Means clustering library and K-Nearest Neighbor. The implementation of the system consists of many subsections which are standard processes to be followed while solving any machine learning problem.

These are as follows:

A. Dataset

B. Data Cleaning

C. Model Analysis

D. Model Building

A. Dataset

The movie dataset is hired in our research paper and collected from the Kaggle database. The Kaggle database provides datasets in the form of several varieties of movie content. The user rating data consists of many records and has a user ID,movie ID, rating, and timestamp. The Characterization of the movie's content information includes over 54058 records and includes movie ID, title, genre, language,release date, and more.

B. Data Cleaning

Description of movie content data consists approximately 25 columns. Most of these columns weren't necessary for our research paper and were therefore removed. The dataset contains many empty values and duplicate values that require to be resolved. Additionally, there are some entries for movies within the user rating files from movies, which didn't correspond to any movies within description of movie content data. These entries were erased for easy processing.

C. Data Analysis

We evaluated datasets to gain insight into movie datasets from Kaggle database that can help to develop and perform our engine / system using Matplotlib libraries in python programming. We acknowledged patterns for many rated movies, most rated genres, the amount of films in each genre, and therefore the number of rated movies in each rating category .

HYBRID SYSTEMS

The hybrid recommendation system is a blend of content-based or collaborative filtering methods. Content Based or collaborative filtering methods are executed individually and the outputs of both methods are recommended to jointly suggest better. The spread of distance between feature vectors can also be used to calculate the similarity of two objects: - Content-based similarity Weight from collaborative network.

Hybrid based approach uses a combination of generally two types of approaches such as content based and collaborative based OR content based or demographic based.

Nearest Neighbors Collaborative Filtering

The concept of nearest neighbors collaborative filtering means that users with rating behavior similar to this point share similar tastes and possibly exhibit similar rating behavior further. The algorithm first calculates relationships between clients employing a row vector within the rating matrix, treating the client as an indication for that client. Parity is calculated by either cosmic similarity or Pearson correlation. So, to guess the rating for a specific client for a specified film, we look for users who are almost like the very best for the present specific client, then take an equivalent average weighting of users' ratings as k, which is that the value of equality.

LITERATURE REVIEW:

Talking about the literature review , a recommender having content plus collaborative and hybrid kind of approach , has a bit different approach than the recommenders previously present. In 2007 a web based and knowledge-based intelligence movie recommendation system was offered using the hybrid filtering method . While in 2017 , a movie recommendation system supported style ratings coefficient or correlation purpose by the authors.. In 2013 a Bayesian network and trust model-based movie recommendation engine have been recommended to predict ratings for users and items, primarily from datasets to recommend users their choice and the other way around .In 2018, the authors built a recommendation engine by analyzing the ratings dataset collected from Kaggle to recommend movies for a user selected from Python.Initial research mainly concentrated on the content of the recommendation system that examined the features of the object to complete the recommendation task.Experiments verified that their approaches were more elastic and

precise.Bayesian networks are employed for model-based preferences based on their context. In 2007 Salakhuddinov and Minh proposed a collaborative filtering method , the probabilistic matrix factor, which can handle large-scale datasets.Also a simple movie recommender using kaggle datasets of 10000 movies namely 'movies_metadata.csv ' is very much efficient for recommending movies or just for a small scale purpose. Content based algorithm was more easy to understand and easy to implement and had less variety therefore researchers came up with a collaborative based algorithm which has more variety and a simple recommendation.. The collaborative filtering algorithm was distributed into portions for deeper study within the movie recommendation by Hurlkartal.

COSINE SIMILARITY:

Cosine similarity measures the similarity between two vectors of an scalar product space. It is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in roughly an equivalent direction. It is often used to measure document similarity in text analysis.Cosine similarity is a measure of similarity that can be used to compare documents or, say, give a ranking of documents with reference to a given vector of query words.

Tfidf Vectorizer :

Scikit-learn's Tdif-Vectorizer is a section of raw documents to a matrix of TF-IDF features.. This article shows you how to correctly use each module, the differences and some guidelines on what to use when.

IMPLEMENTATION

In this segment, we have explained the way to be implemented in our movie recommendation engine / system using python programming language using K-Means clustering library and K-Nearest Neighbor. The implementation of the system consists of many subsections which are standard processes to be followed while solving any machine learning problem.

These are as follows:

A. Dataset

B. Data Cleaning

C. Model Analysis

D. Model Building

A. Dataset

The movie dataset is hired in our research paper and picked up from the Kaggle database. The Kaggle database provides datasets in the form of several varieties of movie content. The user rating data consists of many records and has a user ID,movie ID, rating, and timestamp. The Characterization of the movie's content information includes over 50000 records and includes movie ID, title, genre, language,release date, and more.

B. Data Cleaning

Description of movie content data consists approximately 25 columns. Most of these columns weren't necessary for our research paper and were therefore removed. The dataset contains many empty values and duplicate values that require to be resolved. Additionally, there are some entries for movies within the user rating files from movies, which didn't correspond to any movies within description of movie content data. These entries were erased for easy processing.

C. Data Analysis

We evaluated datasets to realize insight into movie datasets from Kaggle database which will help to develop and perform our engine / system using Matplotlib libraries in python programming. We acknowledged patterns for many rated movies, most rated genres, the amount of flicks in each genre, and therefore the number of rated movies in each rating category.

The unnecessary columns in our code have been removed as mentioned above. The below images give an insight of it.

	adult	belongs_to_collection	budget	genres	homepage	id	imdb_id	original_language	original_title	overview	...	release_date
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...	30000000	[{'id': 16, 'name': 'Animation'}, {'id': 35, '...	http://toystory.disney.com/toy-story	862	tt0114709	en	Toy Story	Led by Woody, Andy's toys live happily in his ...	...	1995-10-30
1	False	NaN	65000000	[{'id': 12, 'name': 'Adventure'}, {'id': 14, '...	NaN	8844	tt0113497	en	Jumanji	When siblings Judy and Peter discover an encha...	...	1995-12-15
2	False	{'id': 119050, 'name': 'Grumpy Old Men Collect...	0	[{'id': 10749, 'name': 'Romance'}, {'id': 35, ...	NaN	15602	tt0113228	en	Grumpier Old Men	A family wedding reignites the ancient feud be...	...	1995-12-22
3	False	NaN	16000000	[{'id': 35, 'name': 'Comedy'}, {'id': 18, 'nam...	NaN	31357	tt0114885	en	Waiting to Exhale	Cheated on, mistreated and stepped on, the wom...	...	1995-12-22

Fig.(Original dataset showing all columns)

	adult	belongs_to_collection	budget	genres	homepage	tmdbId	original_language	original_title	overview
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...	30000000	[{'id': 16, 'name': 'Animation'}, {'id': 35, '...	http://toystory.disney.com/toy-story	862	en	Toy Story	Led by Woody, Andy's toys live happily in his ...
1	False	NaN	65000000	[{'id': 12, 'name': 'Adventure'}, {'id': 14, '...	NaN	8844	en	Jumanji	When siblings Judy and Peter discover an encha...
2	False	{'id': 119050, 'name': 'Grumpy Old Men Collect...	0	[{'id': 10749, 'name': 'Romance'}, {'id': 35, ...	NaN	15602	en	Grumpier Old Men	A family wedding reignites the ancient feud be...

Fig.(Newly obtained dataset showing deleted unnecessary columns)

CONCLUSION AND FUTURE WORK

In this research paper, we have implemented a movie recommendation engine / system using simple recommendations, content-based filtering, collaborative filtering, and hybrid systems. In addition, a movie recommendation engine has been developed using different method prediction methods. This model is implemented in the python programming language.

In future, we can try and test the system using more data and improve the accuracy of the system. In addition, we can try better to increase the accuracy of the recommendation systems.

REFERENCES:

[1] R. Ahuja, A. Solanki and A. Nayyar, "Movierecommender system using K-means clustering and K Nearest neighbor," in 9th International Conference onCloud Computing, Data Science & Engineering, 2019.

[2] N. Severt, "An Introduction to Recommender Systems,"Iterators, 6 Nov 2018. [Online]. Available:https://www.iteratorshq.com/blog/an-introductionrecommender-systems.

[3] Chavarriaga, Oscar, Florian-Gaviria, Beatriz, Solarteand Oswaldo, "A recommender system for students based on social knowledge and assessment data of competences," in School of Systems and ComputingEngineering (EISC),, Colombia , 2014.

[4] Katarya, R. Verma and O. Prakash, "An effective collaborative movie recommender system with cuckoo search," Egyptian Informatics Journal, vol. 18, no. 2, pp.105--112, 2017.

[5] Drachsler, H. Bogers, R. V. T Vuorikari, K. Duval, N. B. G .L. S. S. H. F. E Manouselis and Martin, "Issues and considerations regarding shareable data sets for recommender systems in technology enhanced learning," Procedia Computer Science, vol. 1, no. 2, pp. 2849--2858, 2010.

[6] T. Lorente, P. Carlos and S. Eduardo, "A quality based recommender system to disseminate information in a university digital library," Information Sciences, vol. 261, pp. 52--69, 2014.

[7] T. Tran, T. N, Atas, M. Felfernig and A. e. al, "Anoverview of recommender systems," in Journal ofIntelligent Information Systems, 2018.[8] Wu, C.-S. M. a. Garg, D. a. Bhandary and Unnathi, "Movie Recommendation system using collaborative filtering,"in 9th International Conference on SoftwareEngineering and Service Science, Beijing, China, 2018.

[9] C. Basu, H. Hirsh and W. Cohen, "Recommendation as classification : Using social and content-based information in recommendation," in Proceedings of the15th National Conference on Artificial Intelligence,New Jersey, 1998.

[10] H. W. Chen, Y. L. Wu, M. K. Hor and C. Y. Tang, "Fully Content-based movie recommender system with feature extraction using neural network," inInternational Conference on Machine Learning andCybernetics, Ningbo, China, 2017.

[11] B. R. Cami, H. Hassanpour and H. Mashayekhi, "Content-based movie recommender system based on temporal user preferences," in Iranian Conference onIntelligent Systems and Signal Processing , Shahrood,Iran, 2017.

[12] C. Davidsson and S. Moritz, "Utilizing implicit feedback and context to recommend mobile applications from first use," in Proceedings of the Workshop on ContextAwareness in Retrieval and Recommendation,California, Palo Alto, USA, 2011.

[13] W. Kebin and T. Ying, "A new collaborative filtering recommendation approach based on naive Bayesian Method," International Conference on SwarmIntelligence, vol. 6729, pp. 218--227, 2011.

[14] Z.-D. Zhao and M.-s. Shang, "User-based collaborative filtering recommendation algorithms on hadoop,"Proceedings of Third International Workshop on Knowledge Discovery and Data Mining, pp. 478--481,2010.