ACMEGRADE

MINI PROJECT ON

MOVIE RECCOMENDATION SYSTEM

Made by:

SHIVANI PRABU

ABSTRACT

Recommendation System is a system that seeks to predict or filter preferences according to the user's choices. Recommendation systems are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general, It is a simple algorithm whose aim is to provide the most relevant information to a user by discovering patterns in a dataset. The algorithm rates the items and shows the user the items that they would rate highly. An example of recommendation in action is when you visit Amazon and you notice that some items are being recommended to you or when Netflix recommends certain movies to you. They gradually learn your preferences over time and suggest new products which they think you'll

love.We can make this application using python language and collaborative based filtering algorithm.Collaborative filtering tackles the similarities between the users and items to perform recommendations.We include a data set with user id, ratings, item number and time spent. With these data we use mapping technique and correlation concept to match user id and ratings. The next movie recommendation should be based on the user's rating to watched movies.

CHAPTER 1
INTRODUCTION

1.1 Brief overview of movie reccomendarion system

Film proposals utilizing a number of procedures are widely targeted within the previous a few years. Models incorporate a proposal framework utilizing the ALS calculation, a suggestion smitten by the coefficient procedure, thing likeness based mostly synergistic separation. These procedures would like earlier information regarding the appraisals for the motion photos that square measure made by the shopper. These strategies significantly use film attentiveness datasets for assessment functions. Nonetheless, these frameworks aren't somewhat actual, and analysis is continuous to boost the continuing exhibition of those frameworks. self-addressed the suggestion framework Utilizing the rating and likeness among the 2 clients; the framework prescribes an issue to the shopper for the dynamic. At that time separate the film informational index into Associate in nursing unrated and evaluated take a look at set with the help of the KNN model. It will counsel the motion photos to the obscure shoppers through shopper tour of duty information, furthermore, it will create new and not thought film suggestions as indicated by the film's set of experiences and score. The info set during this approach is that the MYSQL data base. The tour of duty framework for a shopper can snap the client's outer and interior conduct qualities, and these attributes square measure place away within the shopper information base through a login module for the shopper. The to a lower place figure.1.Portrays their compelling technique of approach for a collective sifting approach utilizing KNN.Comparison with completely different calculations. In [4], Goutham Miryala projected an identical investigation of ALS on completely different calculations. still, it's seen that utilizing a additional broad making ready dataset of 80-20 (Training - Testing) yields a model that includes a lower RMSE once contrasted with the 60-40 (Preparing - Testing) dataset. The result shows that the upper regularization boundary expands RMSE and therefore the different method around. The ALS calculation is contrasted and SVD, KNN, and Normal Indicator, and therefore the outcomes show that ALS is that the best calculation for the suggestion framework.CNNs are inspired by the organization of the animal visual cortex and use a hierarchical pattern to recognize visual patterns. They consist of multiple layers, including convolutional layers, pooling layers, and fully connected layers.Convolutional layers apply a series of filters to the input image, which helps in extracting important features such as edges, shapes, and textures. Pooling layers reduce the spatial dimensions of the extracted features, making the network more robust to variations in the input. Fully connected layers at the end of the network combine all the extracted features to make a final prediction. CNNs are widely used in various applications such as image classification, object detection, facial recognition, and more, due to their ability to learn complex patterns from raw pixel data The most well-known sorts of suggestion frameworks square measure content-based and shared separation recommendation frameworks. In shared separation, the conduct of a gatheringof shoppers is employed to form proposals to completely different shoppers. The suggestion depends on the inclination ofvarious shoppers. An easy model would bring down a movie to a shopper smitten by the method that their companion treasured the film. There square measure 2 styles of communitarian models Memory-based ways and Model-based techniques. The top of memory-based strategies is that {they square straightforward to actualize and therefore succeeding suggestions are frequently straightforward to clarify. they're divided into two: User-based synergistic sifting: during this model, things square measure prescribed to a shopper smitten by the method that the things are most wellliked by shoppers just like the shopper. For

CHAPTER 2
LITREATURE SURVEY

Movie recommendation system is based on collaborative filtering approach. Collaborative filtering makes use of information provided by user. That information is analyzed and a movie is recommended to the userswhich are arrang ed with the movie with highest rating first. Luis M Capos et al has analyzed two traditional recommendation systems i.e. content based filtering and collaborative filtering. As both of them have their own drawbacks he proposed a new system which is a combination of Bayesian network and collaborative filtering. A hybrid system has been presented by Harpreet Kaur et al. The system uses a mix of content as well as collaborative filtering algorithm. The context of the movies is also considered while recommending. The user - user relationship as well as user - item relationship plays a role in the recommendation. The user specific information or item specific information is clubbed to form a cluster by Utkarsh Gupta et al. using chameleon. This is an efficient technique based on Hierarchical clustering for recommendation system. To predict the rating of an item voting system is used. The proposed system has lower error and has better clustering of similar items. Urszula Kużelewska et al. proposed clustering as a way to deal with recommendation systems. Two methods of computing cluster representatives were presented and evaluated. Centroid-based solution and memory-based collaborative filtering methods were used as a basis for comparing effectiveness of the proposed two methods. The result was a significant increase in the accuracy of the generated recommendations when compared to just centroid-based method. Costin-Gabriel Chiru et al. proposed Movie Recommendation, a system which uses the information known about the user to provide movie recommendations. This system attempts to solve the problem of unique recommendations which results from ignoring the data specific to the user. The psychological profile of the user, their watching history and the data involving movie scores from other websites is collected. They are based on aggregate similarity calculation. The system is a hybrid model which uses both content based filtering and collaborative filtering. To predict the difficulty level of each case for each trainee Hongli LIn et al. proposed a method called content boosted collaborative filtering (CBCF). The algorithm is divided into two stages, First being the content-based filtering that improves the existing trainee case ratings data and the second being collaborative filtering that provides the final predictions. The CBCF algorithm involves the advantages Of both CBF and CF, while at the same time, overcoming both their disadvantages

CHAPTER 3
METHODOLOGY

3.1 Aim of the project

To implement a recommendation for movies, based on the content of providing the most relevant
information to a user by discovering patterns in a dataset. The algorithm rates the items and shows the user
the items that they would rate highly.

3.1 Software requirements

Python is used for coding

3.2Python Syntax compared to other programming languages

Python was designed to for readability, and has some similarities to the English language withinfluence from mathematics. Python uses new lines to complete a command, as opposed to other programming languages which often use semicolons or parentheses. Python relies on indentation, using whitespace, to define scope; such as the scope of loops, functions and classes. Other programming languages often use curly-brackets for this purpose. Python is Interpreted Many languages are compiled, meaning the source code you create needs to be translated into machine code, the language of your computer's processor, before it can be run. Programs written in an interpreted language are passed straight to an interpreter that runs them directly. This makes for a quicker development cycle because you just type in your code and run it, without the intermediate compilation step.One potential downside to interpreted languages is execution speed. Programs that are compiled into the native language of the computer processor tend to run more quickly than interpreted programs. For some 10 applications that are particularly computationally intensive, like graphics processing or intense number crunching, this can be limiting. In practice, however, for most programs, the difference in execution speed is measured inmilliseconds, or seconds at most, and not appreciably noticeable to a human user. The expediency of coding in an interpreted language is typically worth it for most applications. For all its syntactical simplicity, Python supports most constructs that would be expected in a very high-level language, including complex dynamic data types, structured and functional programming, and object-oriented programming. Additionally, a very extensive library of classes and functions is available that provides capability well beyond what is built into the language, such as database manipulation or GUI programming. Python accomplishes what many programming languages don't: the language itself is simply designed, but it is very versatile in terms of what you can accomplish with it.

3.3 User based filtering

Imagine that we want to recommend a movie to our friend Stanley. We could assume that similar people will have similar taste. Suppose that me and Stanley have seen the same movies, and we rated them all almost identically. But Stanley hasn't seen 'The Godfather: Part II' and I did. If I love that movie, it sounds logical to think that he will too. With that, we have created an artificial rating based on our similarity. Well, UB-CF uses that logic and recommends items by finding similar users to the active user (to whom we are trying to recommend a movie). A specific application of this is the user-based nearest neighbor algorithm. This algorithm needs two tasks: In other words, we are creating a User-Item Matrix, predicting the ratings on items the active user has not see, based on the other similar users. This technique is memory based.

3.4 KNN algorithm

The K-nearest neighbors algorithm (K-NN) is a non-parametric classification method first developed by Evelyn Fix and Joseph Hodges in 1951, and later expanded by Thomas Cover.It is used for Classification and regression. In both cases, the input consists of the k closest training examples in data set. The output depends on whether k-NN is used for classification or regression:

- In k-NN classification, the output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.

- In k-NN regression, the output is the property value for the object. This value is the average of the values of k nearest neighbors. K-NN is a type of classification where the function is only approximated locally and all computation is deferred until function evaluation. Since this algorithm relies on distance for classification, if the features represent different physical units or come in vastly different scales then normalizing the training data can improve its accuracy dramatically.

Both for classification and regression, a useful technique can be to assign weights to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones.

CHAPTER 4

Module description

4.1 SYSTEM STUDY

A recommendation engine is a system that suggests products, services, information to users based on analysis of data. Notwithstanding, the recommendation can derive from a variety of factors such as the history of the user and the behaviour of similar users. Recommendation systems are quickly becoming the primary way for users to expose to the whole digital world through the lens of their experiences, behaviours, preferences and interests. And in a world of information density and product overload, a recommendation engine provides an efficient way for companies to provide consumers with personalised information and solutions.

4.1.1 BENEFITS

A recommendation engine can significantly boost revenues, Click-Through Rates (CTRs), conversions, and other essential metrics. It can have positive effects on the user experience, thus translating to higher customer satisfaction and retention. Let’s take Netflix as an example. Instead of having to browse through thousands of box sets and movie titles, Netflix presents you with a much narrower selection of items that you are likely to enjoy. This capability saves you time and delivers a better user experience. With this function, Netflix achieved lower cancellation rates, saving the company around a billion dollars a year. Although recommendation systems have been used for almost 20 years by companies like Amazon, it has been proliferated to other industries such as finance and travel during the last few years.

4.1.2 DIFFERENT TYPES

The most common types of recommendation systems are CONTENT-BASED and COLLABORATIVE FILTERING recommendation systems. In collaborative filtering, the behavior of a group of users is used to make recommendations to other users. The recommendation is based on the preference of other users. A simple example would be recommending a movie to a user based on the fact that their friend liked the movie.There are two types of collaborative models MEMORY-BASED methods and MODEL-BASED methods. The advantage of memory-based techniques is that they are simple to implement and the resulting

recommendations are often easy to explain. They are divided into two:

 User-based collaborative filtering: In this model, products are recommended to a user based on the

fact that the products have been liked by users similar to the user. For example, if Derrick and Dennis

like the same movies and a new movie come out that Derick like, then we can recommend that movie to

Dennis because Derrick and Dennis seem to like the samemovies.

 Item-based collaborative filtering: These systems identify similar items based on users' previous ratings. For example, if users A, B, and C gave a 5-star rating to books X and Y then when a user D buys book Y they also get a recommendation to purchase book X because the system identifies book X and Y as similar based on the ratings of users A, B, andC. Model-based methods are based on Matrix Factorization and are better at dealing with sparsity. They are developed using data mining, machine learning algorithms to predict users' rating of unrated items. In this approach techniques such as dimensionality reduction are used to improve accuracy. Examples of such model-based methods include Decision trees, Rule-based Model, Bayesian Model, and latent factor models.

 Content-based systems use metadata such as genre, producer, actor, musician to recommend

items say movies or music. Such a recommendation would be for instance recommending Infinity War

that featured Vin Diesel because someone watched and liked The Fate of the Furious. Similarly, you

can get music recommendations from certain artists because you liked their music. Content-based

systems are based on the idea that if you liked a certain item you are most likely to like something that

is similar to it.

4.2 DATA PRE-PROCESSING

For k-NN-based model, the underlying dataset ml-100k from the Surprise Python sci-unit was used. Shockmay be a tight call in any case, to search out out regarding recommendation frameworks. It's acceptable for

building and examining recommendation frameworks that manage unequivocal rating data.

4.3 MODEL BUILDING

Information is an element into a seventy fifth train take a look at and twenty fifth holdout take a look at. Grid Search CV completed over five - overlap, is employed to find the most effective arrangement of closeness live setup (sim_options) for the forecast calculation. It utilizes the truth measurements because the premise to get completely different mixes of sim options, over a cross-approval system.

4.4 DATA SET USED:

we are using the Movie Lens Data Set. This dataset was put together by the Group lens research group at the University of Minnesota. It contains 1, 10, and 20 million ratings. Movie lens also has a website where you can sign up, contribute reviews and get movie recommendations.

CHAPTER 5

Code for movie recommendation system

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

movie_ids_titles=pd.read_csv("movies.csv")

movie_ids_titles.head()

movie_ids_titles.shape

movie_ids_ratings=pd.read_csv("ratings.csv")

movie_ids_ratings.head()

movie_ids_ratings.shape

movie_ids_titles.drop(['Genre'],inplace=True,axis=1)

movie_ids_titles.head()

movie_ids_titles.drop(['timestamp'],inplace=True,axis=1)

movie_ids_titles.head()
```

```
merged_movie_df=pd.merge(movie_ids_ratings, movie_ids_titles, on='movie')

merged_movie_df.head()

merged_movie_df.groupby('title').describe

merged_movie_df.groupby('title')['rating'].mean().head()
```

5.1 CONCLUSION

In the last few decades, recommendation systems have been used, among the many available solutions, in order to mitigate information and cognitive overload problem by suggesting related and relevant items to the users. In this regards, numerous advances have been made to get a high-quality and fine-tuned recommendation system. Nevertheless, designers face several prominent issues and challenges. Although, researchers have been working to cope with these issues and have devised solutions that somehow and up to some extent try to resolve these issues, however we need much to do in order to get to the desired goal. In this research article, we focused on these prominent issues and challenges, discussed what has been done to mitigate these issues, and what needs to be done in the form of different research opportunities and guidelines that can be followed in coping with at least problems like latency, sparsity, context-awareness, grey sheep and cold-start problem

CHAPTER 6

REFERENCES

[1] Thakkar, Priyank & Varma (Thakkar), Krunal & Ukani, Vijay & Mankad, Sapan & Tanwar, Sudeep.(2019). Combining UserBased and Item-Based Collaborative Filtering Using Machine Learning: Proceedings of ICTIS 2018, Volume 2. 10.1007/978-981-13-1747- 7_17

[2] Wu, Ching-Seh & Garg, Deepti & Bhandary, Unnathi. (2018). Movie Recommendation System UsingCollaborative Filtering. 11-15. 10.1109/ICSESS.2018.8663822.

[3] Verma, J. P., Patel, B., & Patel, A. (2015). Big data analysis: Recommendation system with Hadoop framework. In 2015 IEEE International Conference on Computational Intelligence & Communication Technology (CICT). IEEE.

[4] Zeng, X., et al. (2016). Parallelization of the latent group model for group recommendation algorithm. In IEEE International Conference on Data Science in Cyberspace (DSC). IEEE.

[5] Katarya, R., & Verma, O. P. (2017). An effective collaborative movie recommender system with a cuckoo search. Egyptian Informatics Journal, 18(2), 105–112. DOI:10.1016/j.eij.2016.10.002

[6] Phorasim, P., & Yu, L. (2017). Movies recommendation system

[7] Sri, M. N., Abhilash, P., Avinash, K., Rakesh, S., & Prakash, C. S. (2018). Movie Recommender System

using Item-based Collaborative Filtering Technique.

[8] Research.ijcaonline.org