MOVIE RECOMMENDATION USING MACHINE LEARNING

Mrs Prasanna Pabba[1] , Sai Teja Asuri[2] ,N. Divya Sree[3],S. Sreechandana[4] , Akash Raj[5], S.Gaurav[6]

Assistant Professor[1],[2,3,4,5,6] Department of CSE, VNR Vignana Jyothi Institute of Engineering and Technology.

ABSTRACT

In today's world, everyone of any age enjoys going to the movies. In a way, this incredible medium connects us all. But one of the things that fascinates me the most is how different each of us is in terms of the movies we like and how we watch them. Some people prefer horror films, while others prefer action films, adventure films, suspense films, love stories, science fiction stories, and so on. Others, on the other hand, only watch movies starring or directed by legendary figures. Taking everything into consideration, assuming that everyone will enjoy the same film is extremely nuanced. Until now, the majority of Movie Recommendation methods have relied on either title-based recommendations or group evaluations from users, or on collaborative filtering. In this project, we use the KNN algorithm from machine learning to build a movie recommendation system based on the concept of content-based filtering. We can use the user's input to extract movies using k-nearest neighbours. This enables us to build Machine Learning models that assist in predicting appropriate films for individual users.

Some keywords to remember are movies, content-based filtering, the K-Nearest Neighbor (KNN) method, and cosine similarity.

1. INTRODUCTION

Since they were first introduced to society, movies have become a major way to pass the time. People's changing lifestyles and the rise of smartphones with cheaper internet connections have led to a rise in the number of people who use YouTube, Amazon Prime, Netflix, and other online movie streaming services like Ibomma. People can watch movies on all of these different kinds of screens. All of them are built on top of a system for making recommendations.

Due to an increase in the amount of content that these movie streaming sites offer, customers often get multiple recommendations, even if they're not likely to watch the video in question.

Because of this, a recommendation system that is based on what the consumer wants is very helpful.

The content-based system, the collaborative filtering system, and the hybrid recommendation system are the most common ways to group recommendation systems. The most common type of recommendation system is one that is based on the content. The way content-based systems work is based on how an object is classified or put into a category. If a user has seen a certain movie, the system will suggest other movies that are similar in terms of the director, the genre, and many other things. If users "A" and "B" have rated the same things in the past, it is assumed that they will continue to rate the same things in the future. This is the main idea behind collaborative filtering. Based on the abstract ideas of "closeness," "past," and "rating," the basic method can be used in a few different ways. The Hybrid system is made up of both the Content-Based Filtering System and the Collaborative Filtering System. But none of these methods are even close to being accurate, and more research is being done right now to improve how well they work in real time.

One of the goals of this work is to use machine learning to make predictions about movies that are similar to what people like based on their tastes.

2. Find out what's going on in the film industry right now so you can give users accurate and up-to-date advice.

4. For visualisation, many graphical methods and plotting libraries, such as Matplotlib, are used to make it easier to understand which movies other users liked.

5. The main reason we're making this system is to give customers accurate suggestions for movies based on the movies they've already chosen.

2. RELATED WORK

Recommendation systems are systems that are designed to recommend items to users based on a variety of factors. These systems forecast the most likely product that users are likely to buy and are interested in. Companies such as Netflix and Amazon use recommendation systems to assist their users in identifying the best product or movie for them.

The recommendation system handles a large volume of information by filtering the most important information based on data provided by a user and other factors that take into account the user's preferences and interests. It determines the match between the user and the item and infers the similarities between the two for recommendation.

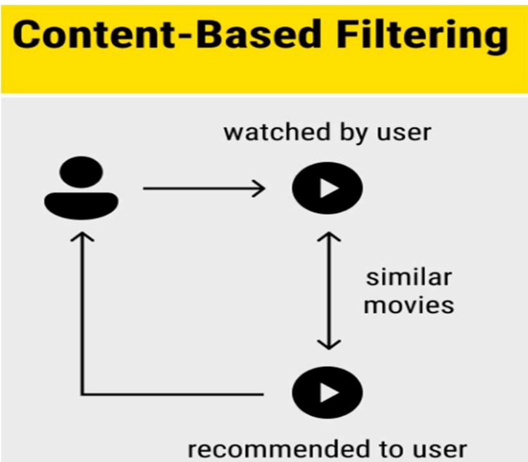
Collaborative Filtering is the process of predicting a user's interests by identifying preferences and information from many users. This is accomplished by filtering data for information or patterns using techniques that involve the collaboration of multiple agents, data sources, and so on. The underlying premise of collaborative filtering is that if users A and B have similar tastes in one product, they are likely to have similar tastes in other products as well.

Filtering based on content: Filtering based on content generates recommendations based on the user's preferences and profile. They attempt to match users with items that they have previously liked. The level of similarity between items is generally determined by the attributes of items that the user likes. Unlike most collaborative filtering models, which rely on ratings between the target user and other users, content-based models rely solely on ratings provided by the target user. In essence, the content-based approach uses various data sources to generate recommendations.

The following data sources are required by the simplest forms of content-based systems (these requirements can increase depending on the complexity of the system).

Item level data source — you'll need a reliable source of data related to the item's attributes. In our scenario, we have book price, number of pages, published year, and so on. The more information you have about the item, the better it will be for your system.

User level data source — you'll need some kind of user feedback based on the item for which you're making recommendations. This kind of feedback can be implicit or explicit. We're working with user ratings of books they've read in our sample data. The more user feedback you can collect, the better your system will be. complexity of the system you're attempting to construct



Hybrid Filtering:

Parallel and sequential architectures are the two that are most commonly used in hybrid recommendation systems. The input for the numerous recommendation systems is provided by the parallel design, and the combined output of those multiple recommendation systems is the final result. A single recommendation engine receives its input parameters from the sequential design, and that engine's output is then sent to the next recommender in the series.

3. METHODOLOGY

The purpose of this project is to design a recommendation system that would suggest movies to users not only on the basis of the titles of the films, but also on the basis of the films' genres and casts.

The content-based filtering serves as the foundation for the building of the recommendation system.

The information provided by the user is analysed and related items and parameters are suggested in the process of content-based filtering.

Customers will soon be able to submit their data set to the application in order to receive movie recommendations that are tailored to their individual preferences thanks to the new way that is being made available.

The following are some of the features that are included in the dataset: index, budget, genres, homepage, id, keywords, original language, original title, overview, popularity, production companies, production nations, release date, revenue, runtime, spoken languages, status, tagline, title, vote average, vote count, cast, crew, and director.

Flow of recommendation process:

Step-1:In order to execute this work, we utilised a dataset including the film's metadata. (open source)

Step-2:Carrying out exploratory data preprocessing (also known as EDA) on the information that was collected.

Step-3:By utilising Feature Extraction, the number of resources required for processing can be cut down significantly without the loss of any essential information.

Step-4:Input should be provided by the user

Step-5:To determine the degree of similarity between two movies, SKlearn's cosine similarity is the metric that is employed.

Step-6:Recommend similar movies in descending order

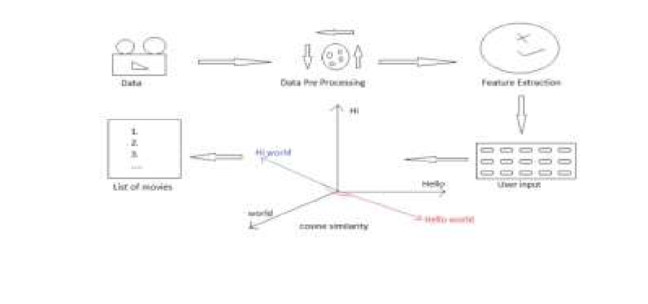


Fig 1: movie recommendation flow chart

4. EXPERIMENTAL SETUP AND RESULTS

4.1 KNN algorithm:

The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point.

Steps followed during KNN are:

Step 1 − In order to put any algorithm into motion, a dataset is required. Because of this, the very first thing that the KNN algorithm requires is for us to load both the training data and the test data.

Step 2 − Next, we will choose the value of K, which will be the set of data points that are clustered together the most closely. The value of K can be any integer.

Step 3 − For each point in the test data do the following :

3.1 − Calculate the distance between each row of the test data and each row of the training data using either the Euclidean distance, the Manhattan distance, or the Hamming distance, depending on the methodology that you choose to apply. The Euclidean method is the one that is utilised most frequently when determining distances because it is the most accurate.

3.2 − Now, based on the distance value, sort them in ascending order.

3.3 − After that, it will pick the first K rows from the array that has been sorted in the correct order.

3.4 − Now, it will decide which class the test point is a part of by looking over the row to see which class appears here the most often and basing its decision on that.

Step 4 − End

4.2.Cosine Similarity:

The degree of similarity between two data points on a plane can be evaluated using a statistic called the cosine similarity. One of the metrics that machine learning algorithms like the KNN employ to determine the distance between neighbours is called cosine similarity. When applied to textual data, it is used to determine the similarity of texts contained within a document. It is also used in recommendation systems, where it is used to recommend movies that have similarities to other movies.

$$\cos\theta = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|\|\vec{b}\|}$$

$$\|\vec{a}\| = \sqrt{a_1^2 + a_2^2 + a_3^2 + \cdots + a_n^2}$$

$$\|\vec{b}\| = \sqrt{b_1^2 + b_2^2 + b_3^2 + \cdots + b_n^2}$$

Cosine similarity

5. CONCLUSION

This program's major objective was to provide the user with unique movie recommendations that were tailored to their individual tastes and preferences. And the K-Nearest Neighbor (K-NN) clustering algorithm was put to use here. A method known as content-based filtering was implemented into this system so that it could provide users with recommendations. The findings of the system indicate that users would benefit from having recommendation capabilities integrated into digital movie libraries that have the contents describing the products and users' profiles of interest in movie Title,Cast,Genre readily available. Specifically, the findings indicate that users

6. REFERENCES

[1] Subramanyam Kunisetti and B Venkatesh.Content-Based Movie Recommendation System Using Genre Correlation,Conference Paper-November 2018.

[2] Rujhan Singla,Saamarth Gupta, Anirudh Gupta, Dinesh Kumar Vishwakarma.FLEX: A Content Based Movie Recommender,2020 International Conference for Emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020.

[3] Bagher Rahimpour Cami, Hamid Hassanpour and Hoda Mashayekhi. A content-based movie recommender system based on temporal user preferences, 2017 3rd Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS).

[4] Abinash Pujahari and Vineet Padmanabhan.An Approach to Content Based Recommender Systems Using Decision List Based Classification with k-DNF Rule Set,2014 International Conference on Information Technology.

[5]Niloy Ganguly and Pabitra Mitra.Feature weighting in .Conference Paper-January 2008.DOI: 10.1145/1367497.1367646 · Source: DBLP.

[6]Haili Wang, Nana Lou and Zhenlin Chao.A Personalized Movie Recommendation System based on LSTM-CNN,2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI).

[7] Meenu Gupta,,Aditya Thakkar,Aashish,Vishal Gupta and Dhruv Pratap Singh Rathore.Movie Recommender System Using Collaborative Filtering,Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020) IEEE Xplore Part Number: CFP20V66-ART; ISBN: 978-1-7281-4108-4.

[8] Suja Cherukullapurath and T.Sasipraba .A Machine Learning Based Implementation of Product and Service Recommendation Models,2021 7th International Conference on Electrical Energy Systems (ICEES) | 978-1-7281-7612-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICEES51510.2021.9383732.

[9]Raghavendra C K and Srikantaiah K.C. Similarity Based Collaborative Filtering Model for Movie Recommendation Systems,Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021) IEEE Xplore Part Number: CFP21K74-ART; ISBN: 978-0-7381-1327-2.

[10]Souptik Saha, S.Ramamoorthy and Eisha Raghav.User Centric and Collaborative Movie Recommendation System Under Customized Platforms,2021 3rd International Conference on Signal Processing and Communication (ICPSC) | 978-1-6654-2864-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICSPC51351.2021.9451672.

[11] Jamilu Maaruf Musa and XU Zhihong.Item Based Collaborative Filtering Approach in Movie Recommendation System Using Different Similarity Measures.

[12] Ching-Seh (Mike) Wu, Deepti Garg and Unnathi Bhandary.Movie Recommendation System Using Collaborative Filtering.

[13] Ramin Ebrahim Nakhli,Hadi Moradi and Mohammad Amin Sadeghi. Movie Recommender System Based on Percentage of View,5th Conference on Knowledge-Based Engineering and Innovation, Iran University of Science and Technology, Tehran, Iran.

[14] Nimish Kapoor, Saurav Vishal and Krishnaveni K S. Movie Recommendation System Using NLP

Tools, Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1.

[15]M. Chenna Keshava, P. Narendra Reddy,S. Srinivasulu and B. Dinesh Naik.Machine Learning Model for Movie Recommendation System,International Journal of Engineering Research & Technology (IJERT),ISSN: 2278-0181, Vol. 9 Issue 04, April-2020.

[16]H. Chen, Y. Wu, M. Hor and C. Tang, "Fully content-based movie
recommender system with feature extraction using neural network,"
2017 Int. Conf. on Machine Learning and Cybernetics (ICMLC),
Ningbo, 2017, pp. 504-509.

[17] Bagher Rahimpour Cami, Hamid Hassanpour, and Hoda Mashayekhi. User trends modeling for a content-based recommender system. Expert Systems with Applications,
87(30):209–219, 2017.

[18] Shouxian Wei, Xiaolin Zheng, Deren Chen, and Chaochao Chen. A hybrid approach for movie recommendation via tags and ratings. Electronic Commerce Research and Applications, 18:83–94, 2016.

[19] C. Basu, H. Hirsh, W. Cohen et al., “Recommendation as classification:
Using social and content-based information in recommendation,” in
AAAI/IAAI, 1998, pp. 714–720.

[20] P. Resnick and H. R. Varian, “Recommender systems,” Communications
of the ACM, vol. 40, no. 3, pp. 56–58, 1997.