CONTENT-BASED MOVIE RECOMMENDER SYSTEM WITH SENTIMENT ANALYSIS

PRAVALLIKA REDDY[1]

ASSISTANT PROFESSOR, DEPARTMENT OF CSE, BHOJ REDDY ENGINEERING COLLEGE FOR WOMEN, VINAY NAGAR, HYDERABAD-59

NEHA NAAZ[2]

UG SCHOLAR, DEPARTMENT OF CSE, BHOJ REDDY ENGINEERING COLLEGE FOR WOMEN, VINAY NAGAR, HYDERABAD-59

BANOTH POOJA SRI[3]

UG SCHOLAR, DEPARTMENT OF CSE, BHOJ REDDY ENGINEERING COLLEGE FOR WOMEN, VINAY NAGAR, HYDERABAD-59

SHIVARATHRI PRATHIMA[4]

UG SCHOLAR, DEPARTMENT OF CSE, BHOJ REDDY ENGINEERING COLLEGE FOR WOMEN, VINAY NAGAR, HYDERABAD-59

ABSTRACT

In this digital era, there are endless varieties of content to be consumed like books, videos, articles, movies, etc., it has become a hectic task to find the content of one’s liking. Whereas, the digital content providers want to get as many users on their service as possible for the maximum duration. Here the recommender system comes into play by which the content providers recommend users the stuff according to the users’ liking and preferences. In this paper, we have proposed a movie recommender system, “Movie Guide”. The goal of Movie Guide is to provide the precise movie recommendations to users. Generally, the basic recommender systems consider any of the below factors for generating recommendations; the choice of the user (i.e. content-based filtering) or the choice of similar users (i.e. collaborative filtering). To create a stable and on-point recommender system a hybrid of content-based filtering as well as collaborative filtering is to be used. Movies alike to the movie of user’s liking are recommended to user by the Content Based recommender system by analyzing the sentiments on the reviews given by the user for that movie.

KEYWORDS: Movies, Recommendation system, CBF- Content-based filtering, CF- Collaborative filtering, Hybrid systems

INTRODUCTION

In today’s world, the internet has become an important part of human life. The excessive available information often creates a complication for the user. To help users cope with this information explosion Recommendation systems (RS) are deployed. RS is mostly used in ecommerce applications and knowledge management systems such as tourism, entertainment, and online shopping portals. Our sole focus in this paper is on RS for movies which are an integral source of recreation and entertainment in our lives. Webbased portals are responsible for movie suggestions to users. Genres of movies like comedy, thriller, animation, and action are the easiest category to differentiate them. Another possible way to categorize movies can be achieved on the basis of metadata such as year, language, director, or cast. Most online video-streaming services provide a number of similar movies to the user to utilize the user’s previously viewed or rated history. Movie Recommendation Systems help us to search for our preferred movies and also reduce the trouble of spending a lot of time searching for favorable movies. The primary requirement of a movie recommendation system is that it should be very reliable and provide the user with the recommendation of movies that are similar to their preferences. In recent times, with an exponential increase in the amount of online data, RS is very beneficial for taking decisions in different activities of day-to-day life. RS is broadly classified into two categories: Collaborative filtering (CF) and Content-based filtering (CBF). Humans often tend to make decisions on the basis of facts, predetermined rules, and familiar information which is accessible over the internet and this predisposition of human behavior gave rise to the notion of CF. The idea of CF was introduced by Resnick et al. in net news, to help people find articles they liked in an enormous stream of available articles. CF emphasizes on helping people to make choices based on the standpoint of other people. Two users are considered like-minded when their ratings for items are similar whereas, in CBF, items are suggested through the similarity among the contextual information of the items. With the emergence of social media platforms like Quora, Facebook, Instagram, Twitter and many more similar social media services which authorize people to express their daily state of mind over the internet. Twitter is one of the most popular social network founded in 2006 where users can express their thoughts and emotions in limited characters. The Distinctive Selling Proposition of Twitter is that the users not only receive information based on their social links but also gain access to other userspecified information. “Tweets” are known as the source of information on Twitter which are of a limited character that keep users up to date on their best-loved topics, people, and movies. For the purpose of the present report, we propose a movie recommendation framework by amalgamating hybrid and sentiment scores from the database.

MATERIALS AND METHODS

One common approach when designing recommender systems is content-based filtering. On the illustration of the item and a profile of the user's preferences, Content-based filtering methods are based. These methods best fit to situations where there is known data on an item but not on the user. Content-based recommenders treat recommendations as a user-specific classification problem and learn a classifier for the user's likes and dislikes based on an item's features. To describe the items keywords are used and a user profile is generated to indicate the type of item the user likes. In other words, these algorithms try to recommend items that are similar to those that a user liked in the past, or is examining in the present. To generate this often-temporary profile it does not depend on a sign-in mechanism. Particularly, many candidate items are compared with the items rated by the user in the past, and the most-matching items are recommended. This method has its roots in information retrieval and information filtering research

METHODOLOGIES 1. CONTENT BASED FILTERING (CBF)

A Content-Based movie recommendation system utilizes the data provided by users like ratings, feedback, and reviews. To make recommendation to the user, a user profile is created using the user’s previous data. As the user takes more actions or provides more inputs on the recommendation system, the engine becomes more precise and robust. To implement a content-based filtering system following steps are to be followed: • Terms Allocation • Terms Representation • Learning Algorithm Selection • Provider Recommendation

CALCULATION OF SIMILARITY (CS)

Calculating the similarity between movies is the goal of content-based recommender systems. The content can be anything example; text, video, and image. In our project, each movie is represented by an attribute vector. Cosine Similarity is the most used measurement for document similarity. To calculate the similarity between two attributes, we have to calculate the cosine of the angle between the attribute vector using the given

cos(m,n) = m.n/||m||*||n||
Where, m.n= product (dot) of the vectors ‘m’ & ‘n’.
||m|| & ||n||= length of the two vectors ||m||*||n||= cross product of the two vectors.

SENTIMENT ANALYSIS (SA)

Sentiment analysis is one of the Natural Language Processing fields which is committed to the judgement of subjective opinions, views, or feelings collected from a collection of sources about a specific subject. In more accurate business terms, it can be encapsulated as “Sentiment Analysis is a set of tools to pinpoint and extract opinions and utilize them for the benefit of the business operation”. Such algorithms push deep into the text and find the substance those points out the feature towards the result in regular or its specific element.

The calculation is performed by: -
P (+ve | altogether liking of the movie) = P (altogether liking of the movie | +ve) * P (+ve) / P (altogether liking of the movie).

The data utilized in the study were taken from a content-based recommendation system using the TMDB 5000 movie dataset. These files have the metadata for more than 45,000 movies listed in the Full MovieLens Dataset. The dataset comprises of movies released on or before July 2017. Data points incorporate cast, crew, plot keywords, budget, revenue, posters, release dates, languages, production companies, countries, TMDB vote counts, and vote averages. This dataset also contains files having 26 million ratings from 270,000 users for all 45,000 movies. On a scale of 1-5, the ratings are given and have been gathered from the official Group Lens website.

CONCLUSION

From the researched papers, recommendation systems are a prominent technique and help to give an enhanced experience for the user as well as the company. These systems are of different types such as content-based, collaborative or hybrid, according to the system in which the developers are made. We review the several types of recommendation systems with their advantages and their drawbacks, in this paper. Content-based filtering methods become an advantage when dealing with a new user whereas these systems also have some drawbacks. The collaborative filtering methods are segregated into two parts in which, neighborhood methods are used to recommend plain content but they are unable to provide accuracy, and model-based methods improve the status of cold-start problems. Collaborative filtering systems are very popular because they have many booms. Hybrid systems overcome the drawbacks of both content-based and collaborative filtering systems, improve the outcome and make the system precise.

REFERENCES

[1] “Classification & regression by Random Forest,” A. Liaw et al- R News, vol. 2, no. 3, pp. 18–22, 2002..
[2] https://ieeexplore.ieee.org/document/8884146? denied
[3] https://scholar.google.co.in/scholar?q=survey+ paper+on+machine+learning+for+parkinson+disease+prediction&hl=en&as_sdt=0&as_vis=1 &oi=scholart#d=gs_qabs&u=%23p%3D5WsSs6TSE0MJ
[4] ‘‘Intraday prediction of Borsa Istanbul with convolutional neural networks and feature correlations,’’ Z. Cataltepe, H. Gunduz & Y. Yaslan- Knowl.-Based Syst., vol. 137, pp. 138–148, Dec. 2017.
[5] ‘‘Least squares support vector machine classifiers,’’ J. A. K. Suykens and J. Vandewalle Neural Process. Lett., vol. 9, no. 3, pp. 293–300, Jun. 1999.
[6] "Improved diagnosis of Parkinson’s disease with optimized crow search algorithm", D. Gupta, S. Sundaram, A. Khanna, A. E. Hassanien and V. H. C. de Albuquerque, Comput. Electr. Eng., vol. 68, pp. 412-424, May 2018.
[7] ‘‘Step length determines minimum toe clearance in older adults & people with Parkinson’s disease,’’ L. Alcock, B. Galna, R. Perkins, S. Lord, and L. Rochester- J. Biomech., vol. 7, pp. 30–36, Apr. 2018.
[8] ‘‘A pervasive & sensor-free deep learning system for Parkinsonian gait analysis,’ J. Ajay, C. Song, A. Wang, J. Langan, Z. Li, and W. Xu,’ in Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI), Las Vegas, NV, USA, Mar. 2018, pp. 108–111.
[9] “Support vector machines for multi-instance learning,” S. Andrews, I. Tsochantaridis, and T. Hofmann, in Proc. of Neural Information Processing Systems, vol. 15, 2003, pp. 577– 584.
[10] ‘‘Parkinson disease gait classification with ML approach,’’ N. M. Tahir and H. H. Manap- J. Appl. Sci., vol. 12, no. 2, pp. 180–185, 2012.
[11] ‘‘Using Kinect to classify Parkinson’s disease stages related to seriousness of gait impairment,’’ L. Dranca, L. de Abetxuko Ruiz de Mendarozketa, A. Goñi, A. Illarramendi, I. N. Gomez, M. D. Alvarado, and M. C. Rodríguez-Oroz- BMC Bioinf., vol. 19, no. 1, Dec. 2018, Art. no. 471.