Movie Recommendation Engine with Sentiment Analysis: A Survey

[1]Sonali Rathore, [2]Seema Narvare, [3]Satish Chadokar

Assistant Professor
Department of Computer Science and Engineering,
Shri Balaji Institute Of Technology & Management Betul(M.P.),

Abstract:

As Artificial Intelligence and Machine Learning have grown at a rapid pace in recent years, so has the amount of data on the internet. As a result, consumers find it difficult to select the precise information they desire, and learners find it difficult to suggest to users exactly what they require. Here, recommendation systems come into play to point consumers in the direction of the content based on their preferences. This study aims to explain the creation and implementation of Movie Recommendation Systems in the Context of Recommendation of Movies and TV Shows on Online Streaming Platforms.

Movie proposals in the Web climate are fundamentally significant for Internet clients. It completes thorough accumulation of client's inclinations, surveys, and feelings to help them find appropriate motion pictures advantageously. Recommendation systems (RSs) have garnered immense interest for applications in e-commerce and digital media. Recommendation System is a smart system that offers pertinent information regarding the decisions a user has made. Collaborative filtering and content-based filtering are two of its useful techniques. This paper is aimed to explain the making and implementation of Movie Recommendation Systems Using Machine Learning Algorithms, Sentiment Analysis and Cosine Similarity.

***Keywords*: Recommendation system, Content-Based filtering, Collaborating filtering, Sentiment Analysis, Cosine Similarity.**

Published in IJIRMPS (E-ISSN: 2349-7300), Volume 11, Issue 3, May-June 2023

I. Introduction

Recommendation systems are mostly used to assist consumers in receiving results that are tailored to their interests. By using a machine learning algorithm that has a target based on the viewer search, recommendation systems can also be utilized as a filtering strategy to isolate the best result from a set of anticipated results. Web-based models are required to provide consumers with movie recommendations. As Movies can be categorized based on genres like drama, comedy, action, drama, animation, and thrillers. Another technique to categorize movies is to use metadata such as actors, year of release, language, or director. By using the user's prior search terms and viewing history, the vast majority of online video-streaming services today provide a variety of related television shows and movies to the user.

These movie recommendation systems [1]-[2]-[3] assist users in finding the films or television shows of their choice, saving them time when making viewing decisions. The major objective while developing a movie recommendation system is to make it trustworthy and effective in order to give people choices that are specifically tailored to their needs. Basically, there are two categories of recommendation systems: content-based filtering (CBF) and collaborative filtering (CF).

As humans, we have a tendency to make judgements based on facts and data that we already have stored in our brain from searching the web, and this behavior gives rise to the notion of Collaborative Filtering. When two users' ratings are similar, they are regarded to be like-minded. While in the instance of content-based filtering, suggestions for results are made based on how contextually related two items are. In the modern world, the internet has become an essential component of daily life. Users also struggle with the issue of too much knowledge being available. Many applications, such as online shopping portals, netflix, amazon, have a recommendation system as their core feature. This paper focuses on movie recommendation system because movies are a significant source of leisure and entertainment in our lives. So, we propose a movie recommendation system by combining movie reviews datasets from various sources, such as Kaggle, where we take 5000+ different movies dataset till 2017 and Movie metadata dataset, as well as data from Wikipedia of movies from 2018 to 2020 and their reviews for sentiment analysis from TMDB website using TMDB API.

II. RELATED WORKS

Recommendation Systems are regarded as one of the most effective knowledge management engines, assisting us in filtering out irrelevant material and providing focused data based on feedback from previous data and similar data from user searches. Many recommendation systems have been presented to date, using various methodologies for computing such as Content-based filtering, Collaborative filtering, and hybrid models for recommendation. Sentiment Analysis is also used to increase the effectiveness of recommendations.

A. Efficient Bayesian Hierarchical User Modeling in Recommendation Systems

Providing users with individualised recommendations has been regarded as a serious challenge in the IR community since the 1970s. The techniques used to address this issue are classified into two types: content-based filtering and collaborative filtering. The scenario of content-based filtering is that a recommendation system screens a record stream and sends reports to the corresponding user that match a user profile. This study advances content-based recommendation research by improving the skill and feasibility of Bayesian hierarchical linear models, which have a strong theoretical background and great empirical performance on recommendation tasks [4][5]. This work makes no attempt to compare and contrast content-based and collaborative filtering or to handle the issue appropriately. According to this research, one compliments the other, and content-based filtering is especially useful for processing new reports/items with little or no user comments.

Due to the scarcity of information in IR applications, the generally employed EM method gradually merges data. This research proposes a novel quick learning method dubbed "Modified EM" for mastering a wide range of user profiles. This work uses a Bayesian hierarchical modelling technique to solve the cold start problem. Several experts have shown that this technique successfully balances shared and user-explicit data.

As a result, each user's initially inadequate performance is reduced. The EM algorithm is a popular boundary learning approach due to its ease of use and assembly assurance. However, a content-based recommendation system frequently works with reports in a very high-dimensional space, with each record represented by a relatively small vector. A comprehensive evaluation of the Expectation maximisation (EM) algorithm in this case reveals that, because to the limiting condition of the information elements, the EM method meets gradually. The "Modified EM algorithm," which is a better learning algorithm, is created by modifying the normal EM algorithm. Rather than calculating the mathematical answer for all of the user profile boundaries, For some element measurements, we infer the informative arrangement of the boundaries and employ the scientific arrangement rather than the mathematical arrangement determined at the E venture for those boundaries at the M advance. This drastically decreases processing at a single EM focus while also speeding up the learning algorithm. The Bayesian hierarchical modelling approach is becoming a prominent user profile learning approach due to its hypothetically argued capacity to help one user through information flow from diverse users via hyperpriors. This work investigated the drawbacks of the well-known EM-based learning methodology for Bayesian hierarchical linear models, as well as an improved learning method known as the Modified EM algorithm.

B. A hybrid recommendation strategy for a system of movie recommendations

A recommender system is a program that anticipates users' interests and prescribes relevant things or services to a given user based on information provided by the user and the items or services.

Hybrid systems use components from both approaches to improve performance and address weaknesses. This study suggests a hybrid strategy built on content-based filtering and collaborative filtering, which was initially used in MoRe, a framework for cinema recommendations. This is generally performed through the employment of two types. The first choice, replacement, calls for employing collaborative filtering as the primary expectation approach and switching to content-based when collaborative filtering forecasts aren't feasible. The second type of hybrid strategy put forth, called switching, bases its switching rule on the quantity of evaluations that are accessible to the dynamic user.

This research also carries out an empirical investigation of the hybrid approach to the main collaborative and content-based filtering strategies and makes insightful conclusions regarding their effectiveness. A Web-based recommendation tool called the MoRe framework collects user ratings for movies on a scale of one to five using a graphical user interface. In order for the system to start forecasting, a new user is required to submit a number of different assessments when they are added to it. Rashid et al. suggested a metric. The basis for deciding which movies are added to the consumer is [6] registered as log(popularity)*entropy. This research also includes a thorough examination of the hybrid recommendation strategy, which is a blend of both content-based filtering and collaborative filtering, as well as a diagram depicting the overall functioning structure of the MoRe system.

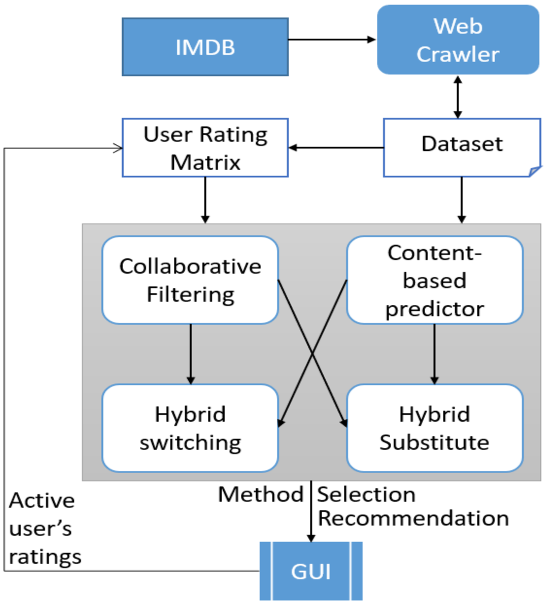

Fig 1. More system in general plan

The adaptation issue may be serious for excessively big datasets, and a similitudes pre-calculation stage may lower the runtime expectation cost even though collaborative filtering is still one of the most accurate recommendation systems. Conversely, content-based recommendations are supplied considerably more quickly but are less accurate than collaborative filtering. They employed the Substitute hybrid recommendation technique to get around this, which improves the precision and inclusiveness of collaborative filtering. According to the aforementioned research, the hybrid recommendation framework will enhance abilities while also expanding the versatility of movie suggestions. Consequently, a hybrid recommendation model is a suitable method for recommending movies.

III. Sentiment Analysis

Sentiment analysis is a subfield of natural language processing (NLP) that aims to automatically identify and extract subjective information from text data. It involves classifying texts or parts of texts as either positive, negative, or neutral with respect to a certain topic or subject. Sentiment analysis can be used to help businesses and organizations understand the sentiment of social media posts, customer reviews, and other forms of online or written communication. It can also be used to monitor brand reputation and customer satisfaction. There are various techniques and approaches for performing sentiment analysis, including rule-based methods, machine learning-based methods, and hybrid approaches that combine both. Some common challenges in sentiment analysis include dealing with irony, sarcasm, and negation, as well as handling subjectivity and cultural differences in language use.

Sentiment analysis is the process of interpreting, processing, summarising, and reasoning emotional text. This approach was used in [7] to calculate the polarity and trust of review sentences. Previously, Pang et al. classified the emotional polarity of movie commentary text into positive and negative using part of speech (POS), N-gram grammar (n-gram), and maximum entropy (ME) [8]. Turney investigated the polarity of text emotion using unsupervised learning in machine learning [9]. The word pair was extracted from the feedback using tags, and the emotional polarity of the text was then determined using the Pointwise Mutual Knowledge and Information Retrieval (PMI-IR) method by comparing the similarity between terms in the text and words in the corpus.

The authors of [10] suggested using the sentiment analysis model Valence Aware Dictionary and Sentiment Reasoner (VADER). Lexical features were coupled with five broad rules that combine grammatical and syntactic standards for expressing and emphasising emotion strength. Positive information is presumptively expected to have a positive impact, whereas negative information is presumptively expected to have a negative impact [11]. Based on this discovery, some research conducted user reviews using sentiment analysis and determined the polarity of the outcomes. After that, users were guided to the movies with the most encouraging content.

IV. PROPOSED SYSTEM

A content-based recommender system is proposed in this paper whose results are improved using sentiment analysis and cosine similarity score.

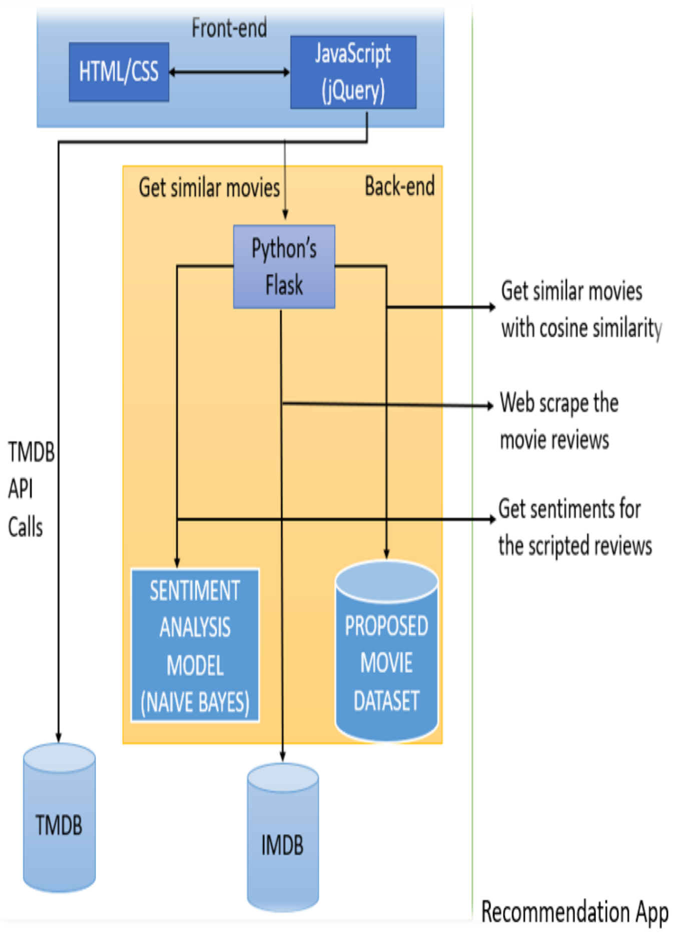

Fig 2. Proposed Movie recommendation system

- **DATASET USED**

The recommender system will be built using two databases: the first is a movie dataset of the top 5000 movies till 2017 and the movies metadata dataset, both of which are available on Kaggle, and the second is a Wikipedia database containing a list of movies from 2018 to 2020.

TMDB (The Movie Database) uses an API to retrieve the movie's finer details, such as title, class, runtime, rating, banner, and so on. https://www.themoviedb.org/documentation/pro Using the IMDB id of the film in the API, we intended to do web scratching to collect the surveys given by the customer on the IMDB website using beautifulsoup4 and perform conclusion examination on those audits. As previously stated, recommendation systems are classified into two types: collaborative filtering (CF) and content-based filtering (CBF). It is a human proclivity to make decisions based on facts, predetermined rules, and known data available on the internet, and this proclivity of human behavior gave rise to the concept of collaborative filtering (CF). We intended to use the following technologies to implement the Recommendation System. Sentiment analysis is a technique for determining a speaker's or author's point of view based on what he or she has said or recorded. It's regularly used to mine internet media (tweets, comments, polls, and so on) for opinions about a brand, item, or administration when doing so physically would be too difficult, costly, or time-consuming.

- **NAIVE BIAS:**

In general, the Multinomial Naive Bayes grouping computation will be a good starting point for opinion research projects. The Naive Bayes procedure's central idea is to identify the probability of classes assigned to messages by combining the probabilities of words and classes.

- **SENTIMENT ANALYSIS**

In this approach, we want to use NLP (Natural Language Processing) for sentiment analysis. We read the feedback from the.txt file into an ipynb file. To obtain user reviews from the IMDB website, we use web scraping. The text is processed using NLTK (Natural Language ToolKit) and the TFIDF (Term Frequency - Inverse Document Frequency) vectorizer. It aids in determining the significance of a word in a collection/corpus text. Because this tells us which terms appear the most in the text, if in the text exist, for example, Study and Studying, the program will treat them as the same. The text is analyzed while disregarding English stop words, which aids in avoiding terms that have no significance to the document (for example, 'the' or 'at'). The vectorizer turns all of the reviews to vectors before building a model and using the Multinomial Nave Bayes method to it. A fit transformation is used to generate a dictionary of known phrases for the supplied input text, and then the TFIDF for each found phrase is calculated. If the movie vector is "1," the review has positive polarity; if it is "0," the review has negative polarity.

- **SIMILARITY SCORE:**

How does it determine which item is most comparable to the item liked (or selected in our case) by the user**?** Here comes the similarity scores.

It is a numerical value ranges between zero to one which helps to determine how much two items are similar to each other on a scale of zero to one. This similarity score is obtained measuring the similarity between the text details of both of the items. So, similarity score is the measure of similarity between given text details of two items. This can be done by cosine-similarity.

- **COSINE SIMILARITY:**

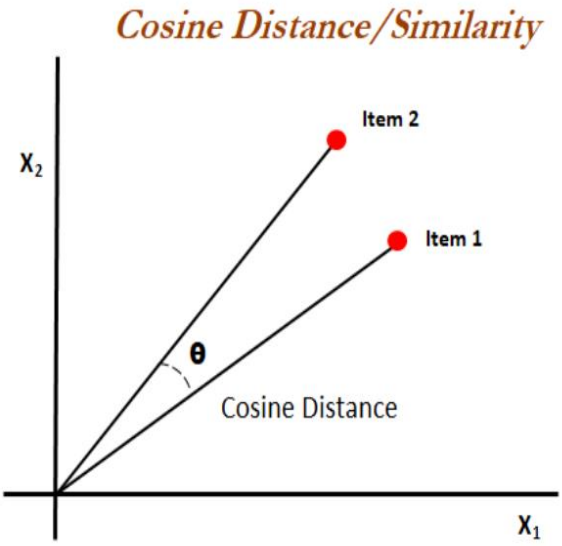

Fig 3. Cosine Distance/Similarity

Cosine similarity is a metric used to measure how similar the documents are irrespective of their size. Mathematically, it measures the cosine of the angle between two vectors projected in a multi-dimensional space. The cosine similarity is advantageous because even if the two similar documents are far apart by the Euclidean distance (due to the size of the document), chances are they may still be oriented closer together. The smaller the angle, higher the cosine similarity.

- **AJAX REQUEST:**

AJAX is stand for Asynchronous JavaScript and XML. AJAX requests are requests made by AJAX applications. Typically, it is an HTTP request done by (browser-resident) JavaScript that encodes the request and/or response data with XML.

V. CONCLUSION

The usage of recommendation systems can be a fantastic way to filter information and give users only the pertinent information. In this research, we present a recommender system that makes use of the TMDB data set and sentiment analysis. Metadata and a social network are essential factors utilized to recommend movies. The main use of Sentiment Analysis in the proposed model was to observe reviews of users for a particular movie. Sentiment analysis is helpful in gathering data on audience responses to a particular movie, and these responses are helpful in making more recommendations. In the future, we hope to improve the suggestion process by gathering more pertinent data regarding user opinions, preferences, etc. It will be possible to accomplish this and use all the facts to create a more potent recommendation system with the help of emerging technology. For our tryouts, we solely took into account films released up until 2020 using a static database. This structure can be studied in a fluid setting that frequently includes fresh films.

REFERENCES

1. Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6):734–749, 2005.

2. Cane WK Leung, Stephen CF Chan, and Fu-lai Chung. Integrating collaborative filtering and sentiment analysis: A rating inference approach. In Proceedings of the ECAI workshop on recommender systems, pages 62–66, 2006.
3. Hui Li, Jiangtao Cui, Bingqing Shen, and Jianfeng Ma. An intelligent movie recommendation system through group-level sentiment analysis in microblogs. Neurocomputing, 210:164–173, 2016.H. Poor, An Introduction to Signal Detection and Estimation. New York: Springer-Verlag, 1985, ch. 4.
4. Yu, K., Tresp, V., and Yu, S. Information filtering using a nonparametric hierarchical Bayesian framework. Pages 353–360 in SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and Development in Information Retrieval. The ACM Press published this book in 2004. (25)
5. P. Zigoris and Y. Zhang are the authors of this paper. User profiling using Bayesian adaptive methods with explicit and implicit input. Presented at the 2006 Conference on Information and Knowledge Management (27)
6. A. Rashid, I. Albert, D. Cosley, S. Lam, S. McNee, J. Konstan, and J. Riedl (2002) Learning new user preferences in recommender systems: Getting to know you. The proceedings of the international conference on intelligent user interfaces are published in this book (21)
7. B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? : sentiment classification using machine learning techniques,” in Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing—Volume 10 (EMNLP '02), pp. 79–86, Association for Computational Linguistics, Stroudsburg, Pennsylvania, USA, July 2002.
8. “Thumbs up or thumbs down?” by P. D. Turney, in Proceedings of the Association for Computational Linguistics' 40th Annual Meeting, pp. 417–424, Philadelphia, Pennsylvania, July 2002.
9. J. R. Priester and R. E. Petty, “The Gradual Threshold Model of Ambivalence: Relating the Positive and Negative Bases of Attitudes to Subjective Ambivalence,” Journal of Personality and Social Psychology, vol. 71, no. 3, pp. 431–449, 1996.
10. “An intelligent movie recommendation system through group-level sentiment analysis in microblogs,” Neurocomputing, vol. 210, pp. 164–173, 2016. H. Li, J.Cui,
11. B. Shen, and J. Ma, “An intelligent movie recommendation system through group-level sentiment analysis in microblogs,” Neurocomputing, vol. 210, pp. 164–173, 2016.
12. Aarushi Sharma, Kandyala Lekhana, Keerthana M, Keshav Anil Rathi, Dr. Sumithra Devi K A, “Movie recommendation system with sentiment analysis using AJAX,” Vol. 03, Issue. 04, 2021.
13. Raghav Mehta, Shikha Gupta, “Movie Recommendation Systems using Sentiment Analysis and Cosine Similarity,” vol. 7(01), pp. 16-22, 2021.