How to Cite:
Giridharan, N., Nathan, K. S., & Swetha, M. K. (2022). Movie recommendation system using machine learning. *International Journal of Health Sciences*, *6*(S8), 498–505. https://doi.org/10.53730/ijhs.v6nS8.9728

Movie recommendation system using machine learning

Mr. N. Giridharan
Assistant Professor, Computer Science and Engineering, K.S.Rangasamy College of Technology, Tiruchengode-637 215, Namakkal District, TamilNadu, India
Corresponding author email: giridharan@gmail.com

K. Senthil Nathan
Students, Computer Science and Engineering, K.S.Rangasamy College of Technology, Tiruchengode-637 215, Namakkal District, TamilNadu, India
Email: beingnathanz@gmail.com

M. K. Swetha
Students, Computer Science and Engineering, K.S.Rangasamy College of Technology, Tiruchengode-637 215, Namakkal District, TamilNadu, India
Email: smk937310@gmail.com

***Abstract*---**A recommendation system is a system that provides online users with recommendations for particular resources, such as books, movies, and music, based on a collection of data from the dataset. These recommendation systems are beneficial for companies that collect information from a large number of customers and aim to provide the best recommendations for the users.Withine-commerce, recommendation systems help in personalizing the user experience by providing items that the user is most likely to be interested in based on similar items that the user has been interested in or items that similar users have been interested in. Using datasets from kaggle to help us analyse the accuracy of our model.Many factors, including the genre of the movie, the actors, and even the director, should be considered while developing a movie recommendation system. The algorithms can recommend movies based on one or a combination of twoor more criteria by using collaborativeand content-based filtering techniques in machine learning.The cold-start problem, or the first lack of item reviews, is a new user issue that affects the most of these techniques. Here provided fresh user demographic data instead of rating history to produce suggestions in order to avoid the cold-start problem in this project.

***Keywords*---**MRS, CF, CBF, ML, movie recommendation.

Manuscript submitted: 27 March 2022, Manuscript revised: 9 May 2022, Accepted for publication: 18 June 2022

Introduction

Provided a machine learning framework that evaluates the use of a number of online suggestion attributes for recommendation generation, including user, item, and reviews. Experiments are carried out using the Movie Lens, Film Trust dataset, Kaggle, to assess the performance of the suggested architecture. Here utilized a K-Means architecture, which enables the substitution of particular algorithms, to reduce the dependency on the user recommendations in web mining.

People can find suitable movies with the use of recommendation systems, which are a collection of techniques and algorithms. To predict future actions based on historical data, they utilize a range of strategies and techniques, such as vectorization. In this project, the movie is recommended to different types of online users, and the movie recommendation system is built using the TMDB dataset and the Movie Lens open-source dataset. There are literally thousands of options available online that recommends particular resource to particular users online. Just a few examples include social networking, internet shopping, YouTube videos, and more. A user can customize a platform and find content that they like with the use of recommendation systems.

The two types of machine learning methods applied in recommender systems are content-based systems and collaborative filtering systems. Modern recommender systems include content-based algorithms and collaborative filtering. To predictthe users' reactions to movies they haven't yet seen or that are relevant to their recent browsing history in to provide them movie suggestions. On the basis of these ratings and past online activity, movies are then categorized and recommended to users. To do this, predicted future ratings using historical information on movies and user reviews.

Objectives

i. The major objectives of this work isto suggest movies to users based on their watching history and ratings by using datasets like MovieLens and TMDB.
ii. Using the machine learning algorithms like Collaborative Filtering, Content-Based Filtering, Singular Value Decomposition, our project proposes to develop movie recommendation system.

Literature Review

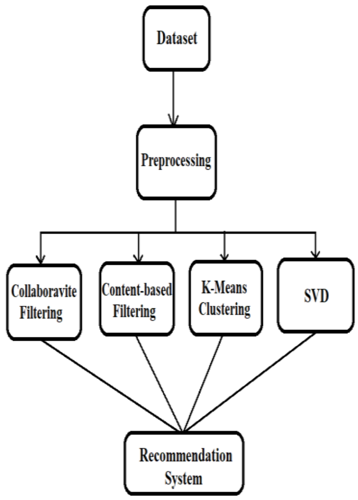
Jiang Zhang, Yufeng Wang, Zhiyuan Yuan, and Qun Jin, With the explosion of big data, have stated that realistic recommendation systems are now important in a variety of industries, such as e-commerce, social networks, and a number of web-based services[1]. Fangtao Li, Sentiment and subject lexicon extraction is essential for opinion mining.Earlier studies have demonstrated that supervised learning techniques are more effective for this task[2]. Lei Zhang, Interpreting people's ideas about a user's attributes has been suggested as a key functions of opinion mining. For instance, the statement "I appreciate the GPS function of Motorola Droid" indicates approval for the Motorola phone's "GPS function." The feature is "GPS function." [3].Kang Liu, has made offers based on a word-based translation model, this research suggests a novel method for extracting opinion objectives (WTM). First, use WTM to mine the relationships between opinion targets and opinion terms in a monolingual situation [4]. GuangQiu, has proposed opinion analysis, also known as sentiment analysis or opinion mining, has received a lot of interest recently due to its many useful applications and difficult research challenges [5]. Kanjar De, Partha Pratim Roy, and Sudhanshu Kumar have proposed for use in e-commerce and digital media, recommendation systems (RSs) have attracted a lot of attention. Collaborative filtering (CF) and content-based filtering (CBF) are examples of traditional approaches in RSs. These systems have some drawbacks, such as the requirement of prior user history and habits for executing the task of recommendation [6].Hsuan Chi and Chu-Hsing Lin, have proposed one of the most popular recommendation systems is for movies, and related technologies are always getting better [7]. Arjun Mukherjee, has proposed writing comments on news stories, blogs, or reviews has become a common activity on social media. In this study, examine customer feedback on reviews [8]. Zhiyuan Liu, In the Web 2.0 age, users frequently freely annotate online content with tags. There has been a lot of interest in automatic tag suggestion to simplify the annotation process [9].Qin Gao, has presented a framework for word alignment that can take into account manual alignments in part. The key component of the strategy is a brand-new semi-supervised technique that adds a constrained EM algorithm to the widely used IBM Models [10]. Shumeet Baluja, Users have a huge opportunity to discover content that interests them because to YouTube's ever expanding video library. The discovery of fresh content is unfortunately a challenging task because of the size of the video library and the complexity of searching videos [11].Partha Pratim Talukdar, has proposed a solution that offer a graph-based semi-supervised label propagation approach for collecting open domain labelled classes and their examples from a mix of unstructured and structured text sources [12]. GuangQiu has made an offer that the sentiment lexicon is crucial to the majority of sentiment analysis systems. However, because various words may be used in different areas, it is difficult, if not impossible, to compile and maintain a universal sentiment lexicon for all application domains. [13]. Robert C. Moore, themajority of methods for statistical machine translation are built on bilingual word alignment. The majority of current word alignment techniques are based on the generative models [14]. XiaowenDin as the user-generated content, such as blog postings, forum posts, and consumer reviews of lenses, is one of the key sources of information on the Internet [15].

Existing System

In the current recommendation system, recommender systems use knowledge discovery approaches to tackle the challenge of providing real customers with a specific product of the recommendations.For recommender systems, the recent growth in customers and items presents some significant challenges. For millions of customers and items, that these are generating by high-quality recommendations and performing several recommendations per second.Although Singular Value Decompositionbased recommendation systems must go through relatively expensive matrix factorization phases, they can rapidly deliver a high-quality movie recommendation system. Here proposed and experimentally test a method that may be used to gradually create a model to enable highly scalable recommender systems.The scalability issue in a practical movie recommendation system must be fixed in order to enable real-time recommendations by the use of content-based filtering with all of its key aspects.

Proposed Methodology

In the proposed work of movie recommendation system suggest by connecting users with those who share their interests, collaborative filtering with K-Means Clustering recommends products. It gathers user input in the form of reviews submitted by users for a certain item and looks for consistency in rating behaviors among users to identify groups of people with comparable tastes. A user profile in this case indicates the preferences a user has provided, either directly or implicitly. As an illustration, consider Amazon's implementation of the CF method, which proposes products based on customer reviews and purchase history. Each user, individually, has a list of things that have been either explicitly or implicitly rated.



Implementation

Dataset Analysis

Toget a better understanding of the movie dataset that could help in creating a movie recommender system using Python libraries, data from datasets in Kaggle, MovieLens, and TMDB were analyzed. Here, similarities such as the most popular films, the most popular genres, the number of films in each genre, and the number of films rated in each rating category were identified. These similarities were then analyzed using machine learning algorithms.

Data Pre-processing

Creating a database for movie recommendation system and processing data sets obtained from the MovieLens dataset and TMDB dataset. The use of the dataset determines the validity of the results, thus creating the database is an important step. It is possible to have a sufficient number of highly predicted items for recommendations to each user by using the datasets made available by several websites, which include users and items with considerable rating history.

Model Building

Python libraries are used to create the recommender system. In addition to the Pearson CorrelationSimilarity class, which employs the PearsonCorrelation Coefficient to assess the similarity between users' evaluations, the User Similarity class was used in this instance of collaborative filtering. CFg is the most popular recommendation technique, and computing similarity is an important factor of this technique. The co-occurrence of comparable events is simple, generally applicable method of determining similarity.It is easy to see that the position of user and item is interchangeable when their count size gap is not too much, thus proposinga vectorization model to show its performance. The similarity based on co-occurrence information suffers from cold start, so proposed a content-based similarity model based on which is another technology in NLP.

Model Evaluation

The movie recommender system constructed in this paper helps the understanding of how a recommender system works. To evaluate the accuracy and relevancy of the results produced by the system,here analyzed both the approaches differently. Then compare the Item based similarity coefficient results by mapping the MovieID. Using Python pandas libraries, to get the result. As obvious from the table, movies which are comparable are given a greater similarity metric. Next,assess the rating predictions on test data against the actual ratings as indicated in the training data. With recommender systems, the user and item cold-start problem are always expressed and so evaluated the system based on these issues and offer implementation design to resolve them.

Algorithm

Content Based Filtering

When making recommendations, content-based filtering compares features of related goods, services, or pieces of content as well as user information collected over time. When making recommendations, content-based filtering compares user profiles to the keywords and attributes that have been allocated to objects in a database such as the items in an online market. A user's behavior, such as purchases, ratings (likes and dislikes), downloads, products searched for on a website and/or added to a cart, and product link clicks, are used to form the user profile. Assigning attributes to database objects enables content-based filtering by providing the algorithm information about each object. These characteristics mainly depend on the goods, services, or information you're suggesting.Similarity metrics are calculated during recommendation using the item's feature vectors and the user's preferred feature vectors from prior records. The top few are then recommended. When making suggestions to a single user, content-based filtering does not require the data of other users.This method is entirely based on evaluating user preferences against product features. What's recommended are the products that have the greatest features in common with user interests. Because product features are so important in this system, it's necessary to talk about how users choose their favorite features.There are two methods that can be used in this situation (possibly in combination). To start, consumers can be provided a list of features from which they can select the ones that they best identify with. Second, the algorithm may remember which goods the user has previously selected and add those features to the user's data. Product features, on the other hand, can be identified by the product's developers.

Result

The dataset includes different ratings and tags that users have assigned to different movies as part of the MovieLens and TMDB datasets for movie recommendation systems. For inclusion, users were chosen at random. Each selected user had given at least 20 films a rating. MovieLens bases its recommendations on data provided by web users, such as movie reviews. The website makes use of several machine learning-based recommendation filtering techniques, including content-based filtering, collaborative filtering, and hybrid filtering. New users are prompted to rate how much they love watching various movie genres in the system (for example, movies with horror or horror comedies).Even before the user has given the website a big number of movie ratings, the algorithm can generate first recommendations based on the preferences gathered from the survey.When entering thewebsite, the user will be asked to sign in using mobile number or email id to create account in the website, and they will get a OTP for verification. User can select the genre of their choice and the movie will be recommended according to the user’s choice by using the content-based filtering algorithm in machine learning.

Conclusion

In this research, a collaborative filtering (CF) and content-based filtering (CBF) algorithm-based movie recommendation system based on rating and previously seen movie history is proposed. Discussed about what has been done to address these cold-start problems as well as what needs to be done in the form of various research possibilities and suggestions that may be used to address problems like context-awareness, grey sheep, and the cold-start problem.Our system's effectiveness has been evaluated using a large dataset that is used to recommend movies to users very effectively. This paper will offer recommendations and provide a framework for a movie recommendation system that takes into account for user opinion.

Future Work

Additionally, gaining a better result and high accuracy may enable the recommender system to produce a result that is even more accurate than the previous one. Hence having an imbalanced data issue in collaborative filtering. Similar user ratings are quite rare. So can group persons, items, or both based on their properties using clustering algorithms like K-Means. Then employ additional features in the hybrid technique to obtain more accurate predictions.

References

1. Jiang Zhang, Yufeng Wang, Zhiyuan Yuan, Qun Jin, "Personalized real-time movie recommendation system: Practical prototype and evaluation" published in Tsinghua Science and Technology, 2020, p. 180-191.
2. F. Li, S. J. Pan, O. Jin, Q. Yang, and X. Zhu, "Cross-domain coextraction of sentiment and topic lexicons," in Proceedings of the 50nd Annual Meeting on Association for Computational Linguistics, Jeju, Korea, 2020, pp. 410–419.
3. L. Zhang, B. Liu, S. H. Lim, and E. O'Brien-Strain, "Extracting and ranking product features in opinion documents." in Proceedings of the 23th International Conference on Computational Linguistics, Beijing, China, 2020, pp. 1462–1470.
4. K. Liu, L. Xu, and J. Zhao, "Opinion target extraction using word-based translation model," in Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju, Korea, July 2019, pp. 1346–1356.
5. G. Qiu, L. Bing, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," Computational Linguistics, vol. 37, no. 1, pp. 9–27, 2019.
6. Sudhanshu Kumar, Kanjar De, Partha Pratim Roy, "Movie Recommendation System Using Sentiment Analysis From Microblogging Data" published in IEEE Transactions on Computational Social Systems, 2020, p. 915-923.
7. Lin CH, Chi H, "A novel movie recommendation system based on collaborative filtering and neural networks. " In: International conference on advanced information networking and applications. Cham: Springer, 2019, p. 895–903.
8. T. Ma and X. Wan, "Opinion target extraction in chinese news comments." in Proceedings of the 23th International Conference on Computational Linguistics, Beijing, China, 2019, pp. 782–790.
9. A. Mukherjee and B. Liu, "Modeling review comments," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, Jeju, Korea, July 2019, pp. 320–329.
10. Z. Liu, X. Chen, and M. Sun, "A simple word trigger method for social tag suggestion," in Proceedings of the Conference on Empirical Methods in Natural Language Processing, Edinburgh, United Kingdom, 2019, pp. 1577–1588.
11. Q. Gao, N. Bach, and S. Vogel, "A semi-supervised word alignment algorithm with partial manual alignments," in Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR, Uppsala, Sweden, July 2019, pp. 1–10.
12. S. Baluja, R. Seth, D. Sivakumar, Y. Jing, J. Yagnik, S. Kumar, D. Ravichandran, and M. Aly, "Video suggestion and discovery for youtube: taking random walks through the view graph," in Proceedings of the 17th International Conference on World Wide Web, Beijing, China, 2019, pp. 895–904.

13. ShobhaRani, P., Sivalakshmi, P., Akarsh, A., Bhanuhithesh, C. H., & Aneesh, C. (2021). Characterization Of Plant Disease Prediction Using Convolutional Neural Network. Turkish Journal of Computer and Mathematics Education, 12(10), 1905-1912.
14. G. Qiu, B. Liu, J. Bu, and C. Che, "Expanding domain sentiment lexicon through double propagation," in Proceedings of the 21st International Jont Conference on Artifical Intelligence, Pasadena, California, USA, 2020, pp. 1199–1204.
15. Balasaranya, K. (2021). Real Time Emotion Recognizer And Classifier For Facial Expressions Based On Machine Learningapproach. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(10), 1958-1964.
16. Rinartha, K., & Suryasa, W. (2017). Comparative study for better result on query suggestion of article searching with MySQL pattern matching and Jaccard similarity. In *2017 5th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-4). IEEE.
17. Rinartha, K., Suryasa, W., & Kartika, L. G. S. (2018). Comparative Analysis of String Similarity on Dynamic Query Suggestions. In *2018 Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)* (pp. 399-404). IEEE.
18. Chanana, M. (2018). Empirical study: relationship between self efficacy and academic performance. *International Journal of Health & Medical Sciences*, *1*(1), 28-34. https://doi.org/10.31295/ijhms.v1n1.36