

# **Movie Recommendation System**

**REVIEW OF LITERATURE**

## **Machine Intelligence**

**BACHELOR OF TECHNOLOGY**

**Department of Computer Science & Engineering**

**V Semester Section G**

**SUBMITTED BY**

**Batch No: 10**

Student 1: Vijith B G

Student 2: Arnav Kumar

Student 3: Samay Jain

SRN: PES2UG20CS402

SRN: PES2UG20CS403

SRN: PES2UG20CS405

# **PES UNIVERSITY**

**(Established under Karnataka Act No. 16 of 2013)**

**100 Feet Ring Road, BSK III Stage, Bengaluru-560085**

## **REVIEW OF LITERATURE**

[1] H. Khatter, N. Goel, N. Gupta and M. Gulati, "Movie Recommendation System Using Cosine Similarity with Sentiment Analysis," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021, pp. 597-603, doi: 10.1109/ICIRCA51532.2021.9544794.

About the research paper:

Proposed recommendation system with sentiment analysis is very useful for the personalised recommendation and also helps in making business decision. Sentiment analysis really helps to highlight the weak and good point of the product so, that we can make better business decision. Here, Cosine similarity is used over many other available similarities for the recommendation system because it has better computing time and efficiency than others. Another advantage of cosine similarity is that it can still give smaller angle between two similar objects even if they are far apart by the Euclidean distance.

This approach can be used as a base for other recommender systems which recommends songs, books, news, videos etc. It can also be incorporated into various ecommerce websites. Also, the system can further be improved such that the user can also provide rating and comment on the movies.

Technique Used:

**Collaborative filtering:** technique that is used to filter those items which a user might like on the ground of reaction by similar users. It works by finding smaller group of users from the larger set of people. While content-based Filtering i.e., another approach uses an item feature in order to recommend other items with similar features or properties.

Limitations:

The existing models do not display all the information related to the movie. Most of the existing models do not consider Sentiment Analysis. The existing models don't contain any history of the users for providing the recommendation. The recommendations provided by the existing models are not personalized and does not cater to all the requirements of the user. Some of the existing models use very limited datasets.

[2] M. K. Kharita, A. Kumar and P. Singh, "Item-Based Collaborative Filtering in Movie Recommendation in Real time," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 2018, pp. 340-342, doi: 10.1109/ICSCCC.2018.8703362.

About the research paper:

Collaborative filtering is one of the most effective and adequate technique used in recommendation. The fundamental aim of the recommendation is to provide prediction of the different items in which a user would be interested in based on their preferences. Recommendation systems based on collaborative filtering techniques are able to provide approximately accurate prediction when there is enough data. User based collaborative filtering techniques have been very powerful and success in the past to recommend the items based on user's preferences. But there are also some certain challenges such as scalability and sparsity of data which increases as the number of users and items increases. In a large website, it is difficult to find the interested information in a certain time. But the recommendation system filters out information and items that are best suitable for us. Although there are different recommendation approaches, yet collaborative filtering technique is very popular because of the effectiveness. In this work, movie recommender system has been described, which basically uses item-based technique of collaborative filtering to provide the recommendations of items, which is dynamic and will learn from the positive feedback.

Technique Used:

Recommendation systems based on collaborative filtering techniques can provide approximately accurate prediction when there is enough data. User based collaborative filtering techniques have been very powerful and success in the past to recommend the items based on user's preferences. But there are also some certain challenges such as scalability and sparsity of data which increases as the number of users and items increases. In a large website, it is difficult to find the interested information in a certain time. But the recommendation system filters out information and items that are best suitable for us. Although there are different recommendation approaches, yet collaborative filtering technique is very popular because of the effectiveness.

Limitations:

Jun Wang et al; proposed the algorithm that unify the item based and content based collaborative filtering. They represented a fusion framework that is able to solve the prediction problem more accurate and also solve the problem related to the data sparsity and new user problem.

[3] M. M. Reddy, R. S. Kanmani and B. Surendiran, "Analysis of Movie Recommendation Systems; with and without considering the low rated movies," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1-4, doi: 10.1109/ic-ETITE47903.2020.453

About the Research paper:

This Paper shows the effect of ignoring the movies that have never got a rating above average using MovieLens-100k dataset. First, we find the predictions of the movies of the user under test considering all the movies, then we find the predictions of the movies of the user under test ignoring the movies that have never got an above -average rating and compares these predictions with the predictions we got previously when all the movies are considered. Consider the value '2.5' as average rating since the dataset we use has a rating range of 1 to 5. Comparison of the predictions we get when all the movies are considered, with the predictions that we get when movies that are always rated below average is done. Graphs of few random users gives a clear view of changes of predictions when movies that are always rated below average are ignored, these changes in the predictions are found to be negligible. As such movies do not have a significant contribution in rating prediction and therefore suggesting ignoring such movies.

Technique used:

Analysis of similarity measures like Pearson correlation coefficient, Jaccard similarity, Cosine1 similarity etc. has been done. Analysis of the Recommendation system's performance when different classifiers like Naive Bayes, Decision Tree, Logistic Regression, K-Nearest Neighbour., Artificial Neural Networks/Deep Learning, Support Vector Machine etc. are used has been good research area. Problems Recommendation system problems like Cold start problem, long tail problem, Sparsity, Shared account problem, grey sheep problem, Scalability etc. has been a good research area.

Limitations:

The existing models do not display all the information related to the particular movie. Most of the existing models do not consider Sentiment Analysis. The existing models don't contain any history of the users for providing the recommendation. The recommendations provided by the existing models are not personalized and does not cater to all the requirements of the user. Some of the existing models use very limited datasets.

[4] C. -S. M. Wu, D. Garg and U. Bhandary, "Movie Recommendation System Using Collaborative Filtering," 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), 2018, pp. 11-15, doi: 10.1109/ICSESS.2018.8663822.

About the Research paper:

A recommendation system or recommendation engine is a model used for information filtering where it tries to predict the preferences of a user and provide suggests based on these preferences. One goal of this paper is to design a movie recommendation system that considers the past movie ratings given by various users to provide suggestions to the user. We implemented this system using collaborative filtering algorithms and Apache Mahout framework. The second goal is to compare the performance and efficiency of user-based recommender system and item-based recommender system. This paper is organized as follows: First, a brief overview of a few relevant, recent research done in the space of recommender system will be discussed. Second, we will present the understanding on the technique of collaborative filtering. Third, the data preparation and data analysis approach using Mahout will be discussed. Finally, a qualitative evaluation on the techniques used will be presented.

Technique Used:

**Collaborative Filtering:** Collaborative filtering systems analyze the user's behavior and preferences and predict what they would like based on similarity with other users.

**Content Based Filtering:** Content-based systems consider the description and features of an item along with the user's preferences to provide suggestions.

**Hybrid System:** Hybrid recommendation systems are a combination of both collaborative and content-based filtering methods. In these type of systems, collaborative and content-based predictions are performed separately and then the results of both techniques are combined to provide recommendations.

Limitations:

- 1) New User Problem is concerned with the scenario when a new user is added to the recommender system. There are no ratings provided by him/her for any movie in the system yet. This is also called a User Cold Start.
- 2) There is a concern that no new item is recommended. It can be understood as an Item Cold Start problem.
- 3) Expensive to do at runtime as for each user, neighborhood is dynamic with new ratings from different users. Therefore, the former requires cache data storage, latter requires a dedicated processing server.

[5] R. E. Nakhli, H. Moradi and M. A. Sadeghi, "Movie Recommender System Based on Percentage of View," 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI), 2019, pp. 656-660, doi: 10.1109/KBEI.2019.8734976.

About the research paper:

In this paper, an Implicit Opinion Measure (IOM) is proposed to improve the performance of movie recommender systems on implicit feedback. To this end, the correlation between like probability and percentage of view of movies is shown and it was illustrated that there is a positive correlation between them. Furthermore, a recommender system has been developed using different prediction methods. The implemented recommender system works 5 times better than a random recommendation. For the evaluation of the recommender system, the average of precision of the top 20 recommendations for all of the users will be reported as performance of the recommender system. The precision is the ratio of number of common items of recommendation list and test list of the user to the size of recommendation list for the user.

Technique used:

Linear Regression and Random Forest Regression are used as the prediction algorithms for this method.

The residual method is a general concept and does not contain any specific algorithm. The main idea is to use a model to do a basic prediction and then use a second model to predict the error of the first level prediction. This can be applied to the next level predictions and deepen the model. The final prediction of the method is the summation of the first prediction with all of the predicted errors in the other levels.

Cosine similarity is used over many other available similarities for the recommendation system because it has better computing time and efficiency than others.

Limitations:

The existing models do not display all the information related to the particular movie.

Most of the existing models do not consider Sentiment Analysis. The existing models don't contain any past history of the users for providing the recommendation.

[6] S. Agrawal and P. Jain, "An improved approach for movie recommendation system," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2017, pp. 336-342, doi: 10.1109/I-SMAC.2017.8058367.

About the research paper:

In this paper, to improve the accuracy, quality and scalability of movie recommendation system, a Hybrid approach by unifying content based filtering and collaborative filtering; using Support Vector Machine as a classifier and genetic algorithm is presented in the proposed methodology. Existing pure approaches and proposed hybrid approach is implemented on three different Movie lens datasets and the results are compared among them. Comparative results depicts that the proposed approach shows an improvement in the accuracy, quality and scalability of the movie recommendation system than the pure approaches. Also, computing time of the proposed approach is lesser than the other two pure approaches. The result is that when the proposed methodology is individually implemented on three different Movie lens datasets it increases accuracy, scalability and quality of the movie recommendation system as compared to the implementation of pure content based approach or pure collaborative approach individually on three different Movie lens datasets.

Technique used:

Content-Based Filtering

Simple Support Vector Machine Algorithm

Adjusted K-Means Algorithm: In this algorithm, in input the numbers of clusters and items attribute features are given.

Genetic Algorithm

Cosine Similarity Measure: The cosine similarity for two vectors is a measure that calculates the cosine of the angle between them

[7] M. A. Hossain and M. N. Uddin, "A Neural Engine for Movie Recommendation System," 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT), 2018, pp. 443-448, doi: 10.1109/CEEICT.2018.8628128.

#### About the research paper:

There are numerous number of movies available over the world, all of those are not interesting and also impossible to watch for one user. That's why, a recommendation system is very important for user to find out the suitable product quickly. On the other hand, a recommendation system gives the flexibility of efficient searching rather than manually. In this way, recommendation system plays a prominent role to user. In this study, we have developed a scheme for a movie recommendation system named neural engine-based recommendation system (NERS) for users. In our recommended approach (NERS), we have incorporated data contents about user's interests via standard movie dataset, that helps us to make a neural engine called neural recommender (NR). We have used two sorts of data sets to make NR, one is general dataset associated with five different nature of data variables, and another one was based on user's choice pattern, where some of the volunteer user contributes their efforts to create it. After combining both data sets, NR engine was applying a neural network (NN), that's recognize user behavioral patterns and then forming a class database, where each class have constructed by using movie genres. In this way, we have initiated nine different grades of classes in the manner of various genres. Finally, two evaluation techniques were used to figure out the best solutions by selecting one or multiple class. For multiple classes, our system will combine information from selected classes and consider them as one for query purpose. At last, three estimators, mean square error(MSE), mean absolute error (MAE) and mean relative error (MRE), were exploiting to demonstrates prediction accuracy of our NERS approach. And, the simulation results show that, our system achieved better performance compare to other methods.

#### Technique Used:

A Hopfield network serve as content-addressable that means associative memory systems with binary threshold nodes. Hopfield networks also provide a model for understanding human memory. Including other functions, HNN have one more function called Transfer, which is used for transforming random logical numbers to the fixed form of digits (1,-1). After setup the network, i.e. design the input nodes and connections, calculating the weight metrics by the [equation \(1\)](#) and store them. Annotate that, the diagonal values will always be initial with '0'.

#### Limitations:

- 1) New User Problem is concerned with the scenario when a new user is added to the recommender system. There are no ratings provided by him/her for any movie in the system yet. This is also called a User Cold Start.
- 2) There is a concern that no new item is recommended. It can be understood as an Item Cold Start problem.



3) Expensive to do at runtime as for each user, the neighborhood is dynamic with new ratings from different users. Therefore, the former requires cache data storage, latter requires a dedicated processing server.

[8] B. R. Cami, H. Hassanpour and H. Mashayekhi, "A content-based movie recommender system based on temporal user preferences," 2017 3rd Iranian Conference on Intelligent Systems and Signal Processing (ICSPIS), 2017, pp. 121-125, doi: 10.1109/ICSPIS.2017.8311601.

About the research paper:

In this paper, we developed a content-based movie recommender system. The proposed method used a temporal preference model of the user to recommend favor movies to the users. For user modeling, the proposed method after extracting user interests from user profile, the priority of each interest is inferred as the user preference. We evaluated the proposed method with MovieLens dataset. Experimental results show that the proposed method improved the accuracy of movie recommendation. Also, it is able to recommend new movies based on their content. The Dataset used consists of 10 million ratings, 10000 movies, and 72000 users. We preprocessed the data and selected a subset of users with at least 100 ratings. To provide the content information for movies, we gathered the movie story-line and genre from IMDb for each movie item within dataset.

Techniques used:

To construct the content-based movie recommender system, we consider the user profile consisting of movie information (such as story – line and genre), which user rated in a specified time. The movie information is gathered from IMDb. We incorporate the profile of individuals into the interests extraction module and discover user interests. Each interest indicate a group of similar movies that are selected by the user in the past.jth group. After constructing user model, the prediction module is used to provide a recommendation list. This module calculates the likelihood of each new item, if it was assigned to each group to determine the selection probability of that item.

Limitations:

One limitation of content based filtering are that it can only make recommendations based on existing interests of the user, it does not consider the fact that what do other users think of an item, thus low quality item recommendations may occur sometimes.

Sometime, Content Overspecialization also occurs in which content similar to the one already present in the User's list is not recommended to the User.

[9] Anuja Gaikwad, Samradnyee Kudalkar, Ashwin Jethawa. Journal, IRJET. "IRJET-MOVIE RECOMMENDATION SYSTEM USING MACHINE LEARNING." IRJET, 2021. Volume: 08 Issue: 04 | Apr 2021, e-ISSN: 2395-0056, p-ISSN: 2395-0072

About research paper:

The proposed algorithm uses textual metadata of the movies like plot, cast, genre, release year and other production information to analyse them and recommend the most similar ones. The system only needs a movie which the user is interested in to come up with suitable recommendations. For evaluation, they ran the algorithm on a subset of all the movies present on the IMDb server. The paper analyzes application similarity measure for recommendations forecasting in recommendations systems. It is shown that used method for computing similarity measure in recommendations systems are cosine similarity measure. For the dataset, they have used three different data sets available in Movie Lens, which is generated by the group lens research team for the research work in the field of recommender system, to help developers to evaluate their recommendation systems. These are:

1. IMDB 5000 Movie Dataset
2. The Movies Dataset
3. List of movies in 2018
4. List of movies in 2019
5. List of movies in 2020

Technique used:

Content-based Filtering: A Content-Based movie recommendation system uses the data provided by users such as ratings, feedback, and reviews. A user profile is generated using this data which is then used to make recommendation to the user.

Term Frequency (TF) and Inverse Document Frequency (IDF) : TF-IDF is a statistical measure which determines how relevant a word is to a document in an accumulation of documents. It is most importantly used in automated text analysis and also in scoring words in machine learning algorithms for NLP.

Cosine Similarity: Similarity Score is a numeric value which ranges between Zeros to one. Which is used to determine the similarity of two items to each other on a scale of zero to one.

Limitations:

One main disadvantage of cosine similarity is that the magnitude of vectors is not taken into account, merely their direction. In practice, this means that the differences in values are not fully taken into account.

Other limitation found was that the model can only make recommendations based on existing interests of the user. In other words, the model has limited ability to expand on the users' existing interests.