

Future Movie Rating Prediction using Machine Learning: Survey

Akshata More¹ Sukanya Ladi² Gayatri Sinnarkar³ Prathamesh Shilarkhane⁴

Prof. Vandana. S. Rupnar⁵

^{1,2,3,4}Student ⁵Professor

^{1,2,3,4,5}Department of Computer Engineering

^{1,2,3,4,5}MMCOE, College of Engineering, Pune, Maharashtra, India

Abstract— Movies rating prediction using Machine learning is one of the most interesting and difficult tasks. Many websites are designed to estimate rating of released movies, Storing movies data and generation of simple statistics. Some websites and systems are using decision support systems, but they are largely limited. They can respond to simple queries like "What is the average date of release movies for a particular actor, directors and so on?", "How many movies had released in 10 days?", "how many awards had to achieved by an actor in the current year?" Still, they cannot answer complex queries like "Give a possibility of the forthcoming movie? Is it hit or not?" and so on. Hence, there is a need to develop a system that can answer such a question. This paper gives the survey of the different techniques used by the researcher for predicting movie ratings.

Key words: Movies Rating Prediction, Machine Learning

I. INTRODUCTION

The Indian movie industry produces the greatest number of movies per year at 1000/year, higher than any other country's movie industry. Conversely, a small number of movies taste success and are ranked high.

Given the low success rate, models and mechanisms to forecast consistently the ranking and/or box office collections of a movie that can help de-risk the business drastically and increase average returns. Various stakeholders such as actors, financiers, directors etc. can use these predictions to make more informed pronouncement. Some of the questions that can be answered using these forecasting models are:

- 1) Does the cast or director matter in the success or ranking of an Indian movie?
- 2) Is the genre of the Indian movie a key determinant of rank or success?
- 3) Does running time matter?

Furthermore, a DVD rental agency or a distribution house could use these predictions to determine which titles to stock or promote correspondingly.

II. LITERATURE REVIEW

Below are some techniques used by different researcher for movie rating prediction?

Ravinder Ahuja et al.[1] propose a method that uses dual training and dual prediction for sentiment classification at the time of addressing polarity shift. Here the original movie reviews and artificially generated opposite movie reviews are used, for training and prediction. Classification accuracy of the Baseline system and DT-DP system is compared using unigram as well as bigram features with Naïve Bayes and SVM as classification algorithms. The Open-PRNB toolkit is used for Naïve Bayes and LibSVM

toolkit with linear kernel and default penalty parameter is used for SVM.

Md. Asif Shahjalal et al.[2] presents A user rating based low-rank matrix factorization collaborative filtering based on movie recommender system. The main advantage of this method is that the features learned by the algorithm do not need to be human defined. It can select and learn the features all by itself.

Kemal Özkan et al.[3] used a deep learning approach to reveal the performance on box-office prediction of movies. The Inception-V3 model has exploited to estimate box-offices by considering the prediction problem as a classification task. On the basis of experimental studies, the relation between movie poster and box-office of movies have been explored and the system gives 33% prediction accuracy rate from conducting experiments on 1307 images.

Daniel Stanley Tan et al.[4] gives the extraction of film structure features from action films to build an arousal curve. The arousal curve is used as training data for building a Hidden Markov Model for predicting the rating of a movie evaluation of the model gives 70% accuracy, which shows that there is some form of correlation between the structure of a film and its perceived rating.

Anand Bhavne et al. [5] define Role of different factors in predicting movie success. The paper uses integration of both the classical and the social factors i.e. anticipation and user feedback and the study of interrelation among the classical factors will lead to more accuracy. The classical factors such as cast, producer, director etc. or the social factors in form of response of the society on various online platforms are used.

Guoshuai Zhao et al. [6] propose the factor of interpersonal rating behavior diffusion to deep understand users' rating behaviors. In the user-service rating prediction approach, author fuse four factors-user personal interest (related to user and the item's topics), interpersonal interest similarity (related to user interest), interpersonal rating behavior similarity (related to users' rating behavior habits), and interpersonal rating behavior diffusion (related to users' behavior diffusions)-into a unified matrix-factorized framework. The series of experiments are applied to the Yelp dataset and Douban Movie dataset.

G. Adomavicius et al.[7] gives an overview of the field of recommender systems and describes the current generation of recommendation methods that are usually classified into three main categories: content-based, collaborative, and hybrid recommendation approaches. This paper also describes various limitations of current recommendation methods and discusses possible extensions that can improve recommendation capabilities and make recommender systems applicable to an even broader range of applications. These extensions contain, among others, an improvement of accepting of users and items, incorporation

of the contextual information into the recommendation process, support for multi criteria ratings, and a provision of more flexible and less intrusive types of recommendations.

T. Jeon et al.[8] propose a prediction system which is based on the techniques of the recommendation system using a collaborative filtering and fuzzy system to solve the collaborative filtering problems. In order to verify the prediction system, Here the user's rating data about movies is used and predict the user's rating using this data. The correctness of this prediction system is determined by computing the predicted RMSE (Root mean square error) of the system against the actual rating about each movie. And predicted RMSE is compared with the existing system. Thus, this prediction system can be applied to the base technology of the recommendation system and also the recommendation of multimedia.

In [9] this paper proposed an approach by utilizing matrix value fact organization for predicting rating i by user j with the submatrix as k -most similar items specific to user i for all users who rated them all. In an attempt, previously predicted values are used for subsequent predictions. In order to investigate the accuracy of neighborhood methods, we applied our method to Netflix Prize. We have considered both items and users relationships on Netflix dataset for predicting movie ratings. We have conducted several experiments. The place of gas leakage.

O. Bora Fikir et al. [10] proposed the more precise method for rating prediction using each unrated user-item pair. Here, the similarity of items is defined as the number of corresponding reviews which a user has given and the differences between those viewpoints spreading in the network model. Using Movie-lens data set, experiments confirm that the presented algorithm has better performance than both collaborative filtering algorithm based on Pearson correlation coefficient and the original opinion spreading approach. Our approach produces a mean absolute error (MAE) 2.1% lower and root mean square error (RMSE) 2.5% lower than existing algorithms, which indicates a higher accuracy score. Get movies historical data form IDBM website and load into a database for movies rating prediction purpose.

III. PROPOSED SYSTEM

The system that can give rating prediction on the basis of Machine Learning is given below. The proposed system can give higher accuracy by using support vector machine compared to the previous method in rating prediction.

A. Load Movie Dataset

Classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches. Leaf node represents a classification or decision.

Get movies historical data form IDBM website and load into a database for movies rating prediction purpose.

B. Get a Common Factor Associated with Movie Rating Prediction

Identify or study on movies data and then decides which factors are common in all movies. Experiments were conducted with Weka 3.6.0 tool. Dataset of 1000 records with 8 attributes is used. The results of our experimental analysis in finding significant patterns for Movies rating prediction are presented in this section.

- Director Facebook Likes
- Movie Facebook Likes
- Actor 1st Facebook Likes
- Actor 2nd Facebook Likes
- Actor 3rd Facebook Likes
- Cast total Facebook Likes
- Language
- Num User for Reviews

C. Movie Rating Prediction

There are some conditions to predict Movie Rating Hit or Not.

E.g.

If(Director Facebook Likes>1.5k&& Movie Facebook Likes>1.5k&& Actor 1st Facebook Likes>1.5k&& Actor 2nd Facebook Likes<1.5k&& Language=="Hindi"&& Num User For reviews > 2.5k)

Then movies rating==5.0 or any more ratings then 5.1

Otherwise

Movies rating==4.9 or any less ratings then 4.9

And so on..

The system deals with existing movies data and performs analysis on that data. Decision tree builds

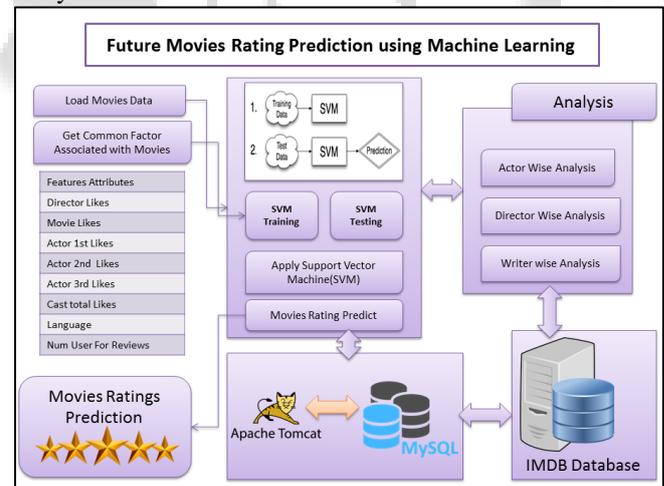


Fig. 1: System Architecture

D. Main Actor Wise Analysis

Display how many movies had released in the current year on the basis of the actor using historical data, and how many movies were to super hit or flop for a particular actor. Comparison between all actors will be done in actor wise analysis.

- Support vector machine
- Apply Support Vector Machine algorithm on following features to make a prediction of Movies Ratings.
- Director Facebook Likes
- Movie Facebook Likes

- Actor 1st Facebook Likes
- Actor 2nd Facebook Likes
- Actor 3rd Facebook Likes
- Cast total Facebook Likes
- Language
- Num User For Reviews

E. Director Wise Analysis

Same as Actor wise display how many movies had released in the current year on the basis of a director using historical data, and how many movies were to super hit or flop for the particular director.

Director can direct movies with other actors so we calculate director wise movies analysis.

Comparison between all directors will be done in director wise analysis.

IV. CONCLUSION

This paper gives a survey of different techniques used by researcher to predict the movie rating. By surveying different techniques we come to know that there is need to develop a system that can give the higher accuracy in prediction.

A technique of support vector machines for predicting numerical movie ratings, based on the attributes of a movie, such as actors, directors, budget or language is given. By using this system users come to know that if movies from certain directors or with certain actors have been successful in the past, then it is expected that also their future movies will be successful.

REFERENCES

- [1] Ravinder Ahuja, Willson Anand, "Sentiment Classification of Movie Reviews Using Dual Training and Dual Prediction", 2017 Fourth International Conference on Image Information Processing (ICIIP).
- [2] Md. Asif Shahjalal ; Zubaer Ahmad ; Mohammad Shamsul Arefin ; Mohammad Rubaiyat Tanvir Hossain , "A user rating based collaborative filtering approach to predict movie preferences", 2017 3rd International Conference on Electrical Information and Communication Technology (EICT).
- [3] Kemal Özkan; Osman Nuri Atak; Şahin Işık "Using movie posters for prediction of box-office revenue with deep learning approach", 2018 26th Signal Processing and Communications Applications Conference (SIU).
- [4] Daniel Stanley Tan; Solomon See; Thomas James Tiam-Lee "Automatic rating of movies using an arousal curve extracted from video features", 2014 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM).
- [5] Anand Bhave; Himanshu Kulkarni ; Vinay Biramane ; Pranali Kosamkar "Role of different factors in predicting movie success", 2015 International Conference on Pervasive Computing (ICPC).
- [6] Guoshuai Zhao; Xueming Qian ; Xing Xie "User-Service Rating Prediction by Exploring Social Users' Rating Behaviors", IEEE Transactions on Multimedia (Volume: 18 , Issue: 3 , March 2016)

- [7] G. Adomavicius and A. Tuzhilin, "Towards 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 (2005), 634-749.
- [8] T. Jeon, I. Cho, S. Lee, G. Baek, and S. Kim, "A movie rating prediction system of user propensity analysis based on collaborative filtering and fuzzy system," in Fuzzy Systems, 2009. FUZZ-IEEE 2009. IEEE International Conference on. IEEE, 2009, pp. 507-511.
- [9] O. Bora Fikir , İ lker O. Yaz , Tansel Özyer 2010 International Conference on Advances in Social Networks Analysis and Mining "A Movie Rating Prediction Algorithm with Collaborative Filtering" on. IEEE, 2009, pp. 507-511.
- [10] Jun Ai 1 , 2 , Linzhi Li 1 , Zhan Su 1 , 2 * , Chunxue Wu "Online-rating prediction based on an improved opinion spreading approach" 2017 29th Chinese Control And Decision Conference (CCDC).