IJIRIS: International Journal of Innovative Research in Information Security
Volume 09, Issue 03, June 2023
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives

Movie Popularity and Target Audience Prediction Using the Content-Based Recommender System

Raghavendra Rao B
Professor, Department Of CSE,
Raghavendrarao.cse@sairamce.edu.in
Aakash A G, Kamalakannan R,
UG Scholars, Department Of CSE,
Sri Sairam College of Engineering, Anekal, Bangalore
Deepika C, Dinakar D J
UG Scholars, Department Of CSE,
Sri Sairam College of Engineering, Anekal, Bangalore

Publication History
Manuscript Reference No: IJIRIS/RS/Vol.09/Issue03/JNIS10099
Research Article | Open Access | Double-blind Peer-reviewed |Article ID: IJIRIS/RS/Vol.09/Issue03/JNIS10099
Received: 02, June 2023 | Accepted: 20, June 2023 | Published: 23, June 2023
Volume 2023 | Article ID JNIS10099 http://www.ijiris.com/volumes/Vol09/iss-03/20.JNIS10099.pdf
ArticleCitation:Aakash,Kamalakannan,Deepika,Dinakar (2023).Movie Popularity and Target Audience Prediction using the Content –Based Recommender System. International Journal of Innovative Research in Information Security (IJIRIS), 10, 152-155 doi: https://doi.org/10.26562/ijiris.2023.v0903.20 BibTex key: **Aakash2023Movie**

Copyright: ©2023 This is an open access article distributed under the terms of the Creative Commons Attribution License; which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Abstract: Movies are an important part of our daily entertainment, with the global movie industry experiencing significant growth and capturing the attention of people of all ages. However, only a few movies achieve success, leading to pressure on movie production stakeholders. Therefore, researchers and moviemakers require expert systems that can accurately predict the probability of a movie's success prior to its production. Most research on predicting movie popularity has focused on post-production stages, but it's essential to predict a movie's success at an early stage to enable necessary changes to be made. To this end, a content-based movie recommendation system (RS) has been proposed that uses preliminary movie features such as genre, cast, director, keywords, and movie description. This RS output and movie rating and voting information are then used to create a new feature set, which is input into a CNN deep learning (DL) model to build a multiclass movie popularity prediction system. The study also proposes a system to predict the popularity of the upcoming movie among different age groups, dividing the audience into four categories: junior, teenage, mid-age, and senior. The study uses publicly available data from the Internet Movie Database (IMDb) and The Movie Database (TMDb). The multiclass classification model implemented in this study achieved 96.8% accuracy, outperforming all benchmark models. Overall, this study highlights the potential of predictive and prescriptive data analytics in information systems to support industry decision- making.

Key Words: Expert systems, content-based, recommendation system, deep learning, audience groups.

I. INTRODUCTION

The global movie industry is a fast-paced and highly profitable industry, with significant investment required to create both qualitative and quantitative content. As such, film producers and researchers are constantly looking for ways to accurately predict the likelihood of a movie's success prior to its production. This is a complex task, as the industry is massive and diverse, with a multitude of parameters involved in creating a successful movie.To address this challenge, the authors propose a three-module framework. The first module is a movie recommendation system that uses five essential features (genre, cast, director, keywords, and movie description) to recommend similar movies. The second module uses the rating and voting information of these similar movies to create a novel feature set, which is then used to predict the upcoming movie's popularity using a convolutional neural network (CNN). The popularity is divided into six classes: super-duper hit (SDH), super hit (SH), hit (H), above average (AA), average (A), and flop (F). The third and final module is designed to estimate the target audience for the upcoming movie, divided into four age groups: junior, teenage, mid-age, and senior. This is accomplished using a model built with fuzzy c-means and cosine similarity, which estimates the popularity of the upcoming movie among all age groups. The contributions of this study are twofold: firstly, it is among the first to use a recommendation system to predict a movie's popularity at its early production stage. Secondly, it proposes a model to estimate the popularity of an upcoming movie among different age groups using fuzzy c-means and cosine distance. Overall, the proposed framework has the potential to significantly reduce the risk involved in the movie industry.

Content Based Recommendation System

Content-based recommendation systems use metadata information or textual items to recommend items that are comparable in content to a reference item. This type of recommendation is useful in situations where items need to be similar, such as in movie recommendations. Content-based movie recommendation systems consider different movie attributes, such as genre, actors, and directors, to build a recommender system.

Correlations between these attributes are measured to find similarities between items, and a multi-attribute network is used to recommend items to users. Hybrid models have also been proposed, such as using genomic tags of movies with content-based filtering to recommend movies with similar tastes. Low-level features, such as colors and motion, have also been used in hybrid recommendation systems. A new movie recommendation system has been proposed to address the cold start problem for new items, using audio and visual descriptions extracted from movie videos to create a video genome. Sentiment analysis has also been incorporated into hybrid movie recommendation systems using movie tweets to understand public sentiment and user response. The sparsity of data is a significant challenge for recommendation system algorithms, but generative adversarial networks (GAN) have been used to deal with the sparsity of review data and ratings. GAN has also been used to rank movies based on user preferences, and LambdaGAN has been used to recommend top-N movies.

Movie Hit Prediction

Movie hit forecasting is a well-known problem in research, which can be broadly divided into two primary groups based on forecasting time. Significant work has been proposed for predicting movie success very late in the production stage before the movie's release or just after its official release. However, limited research has been carried out for movie hit forecasting at the initial or early stage of production, where only a few attributes or features are available, making the problem much more difficult. Defining the success of a movie is one of the most significant parts of this problem, and there are no benchmark models that define the success of a movie. Some earlier works have focused on whole box office revenue, while others have adopted the number of admissions as the parameter of success. Some works have measured success as profitability, which could be a numeric value of revenue or return on investment. Several works have divided movies into two classes (success or not) and used binary classifications, while others have considered the forecast a multiclass classification problem and tried to classify films into multiple discrete classes. Predictions are also made on continuous integral values of profit metrics, with values of these metrics containing logarithmic in some works. Machine learning models have been used for movie hit forecasting, including discriminant analysis, logistic regression, decision tree, and neural network models. Researchers have measured their performance to predict a movie's success. Some have proposed a multi-layer backpropagation architecture and a more quality increased neural network model. Researchers have also introduced the Cinema Ensemble Model to enhance forecast accuracy by comprising seven machine learning models and focusing on selecting attributes. In most of the works, similar movies are computed using a recommendation system from the existing movie. However, the recommendation system for the upcoming movie is sporadic. Content-based movie recommendation systems could be used to find out similar movies for an upcoming movie. Box office information of all these similar movies could be used to analyze the upcoming movie. Our research work has used a content-based movie recommendation system for an upcoming movie and successfully built a model to forecast movie popularity. We have also predicted the target audience from the information of the recommendation system. Most of the research works have focused on the movie hit prediction problem as a binary classification problem, sacrificing accuracy in the process. However, our research work has classified the movie popularity problem into six different classes and achieved high accuracy. Moreover, while most of the works only target the movie popularity prediction problem, research work regarding the forecast of the target audience of the upcoming movie is sporadic.

2. PROPOSED WORK

The system we propose consists of three main modules that are interconnected. The first module is a content-based movie recommendation system that generates a list of 10 movies that are most similar to all the movies in the dataset. The output of this module serves as the input for the second module, which predicts the likelihood of a movie becoming a hit. The third module groups the audience based on their age and predicts the most appropriate target audience for each movie. Finally, the entire system provides a prediction of the popularity of the upcoming movie, broken down by age group. A diagram of the overall process flow is shown in the figure.

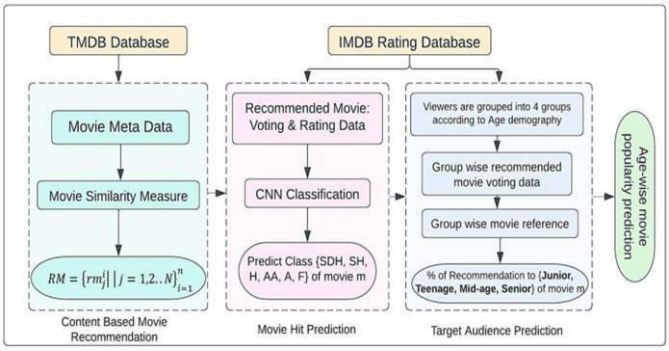

Fig -1: Overall process flow diagram

Content Based Movie Recommendation

The content-based movie recommendation module utilizes a content-based filtering approach to find movies that are similar to a given movie. This is achieved by computing the similarity between two movies based on their respective feature sets. Let F = (F1, F2, . . . , Fm) be the feature set. The similarity between two movies mi and mj with respect to feature Fk is computed as follows:

dist_Fki,j = similarity(Fki , Fkj)….. (1) disti,j = Um k=1 {distF ki,j }……(2)

Here, dist_Fki,j represents the distance vector between movies mi and mj. The goal of this module is to identify the N most similar movies to a given movie mi. To achieve this, we use the nearest neighbor algorithm and measure the overall similarity between movies mi and mj using measures such as cosine similarity:

$$c_distij = cosine_sim(distii, distij), \ldots..(3)\ c_disti = \{c_distij\}^{n} \ldots(4)$$

In equation (3), cosine similarity is used to reduce the m-dimensional distance vector to a one-dimensional distance. Equation (4) computes the similarity measure between a movie mi and all other n movies in the dataset. Figure 4 provides a block diagram of the content-based movie recommendation module.

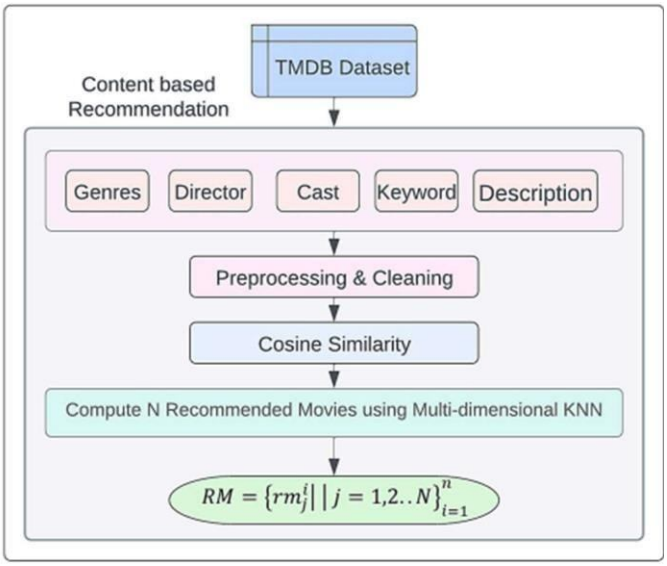

Fig-2: Movie Recommendation Module

Movie Hit Prediction

The second module of this research problem is focused on building a movie hit prediction system, which takes the output of the first module as its input. The recommended movie set RM for each movie mi is used as input, where RM is the union of all recommended movies rmi for j=1 to N. To achieve this, IMDb rating dataset has been used as an input, which serves as a training dataset for the deep learning model. The output of the module is a multiclass classification, which categorizes the movies into one of the six classes: super-duper hit (SDH), super hit (SH), hit (H), above average (AA), average (A), and flop (F). This classification is performed using the following equation:

$$X = f(RM, \text{IMDb rating data set})$$

where X is the output of the classification model, which represents the predicted class for each movie. The input data are processed and fitted into a deep learning model, which is trained to classify movies based on their attributes. This classification is performed by using a multiclass classification algorithm.

Target Audience Prediction

The third and final module of the proposed system aims to predict the target audience's preference based on their age.

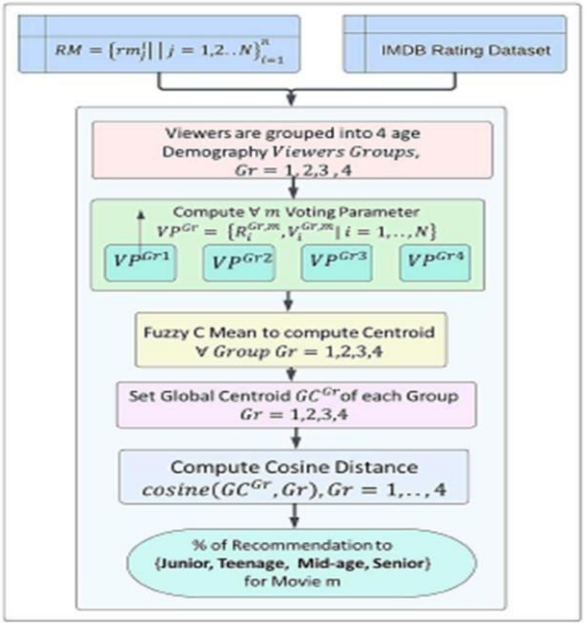

Fig-3: Target Audience Prediction Model

This module uses all the recommended movies obtained from the first module to predict the target audience for an upcoming movie. The system considers user ratings by each age group and the number of votes given by each group to the recommended movies. By analyzing all the voting and rating information from each group for all the recommended movies, the module predicts the popularity of the upcoming movie for each age group. The target audience prediction module takes input from the first module's output and the IMDb rating dataset.

3. Implementation and Methodology

The main goal is to study and to develop a model that predicts the popularity of a movie and its preference among different age groups using movie recommendations. Our objective is to classify movies into six different classes of popularity at an early stage of movie production: SDH, SH, H, AA, A, and F. Furthermore, we aim to regroup the target audience into four different age demographics: Junior, Teenage, Mid-Age, and Senior, and predict age-wise movie popularity. To achieve this objective, we employed a content-based movie recommendation system to find similar movies. We then utilized voting and rating information of each recommended movie to train a 1-D CNN deep learning model that classifies movies into six popularity classes. In the third module, we used recommended movie information and age-wise voting data, which we grouped into four different age groups, to predict the target audience's preference for each movie. Fuzzy C- Mean algorithm was used to calculate the movie preference for each age group. Our research framework has three main steps. Firstly, we acquire movie data and intrinsic features from the TMDb dataset and compute similar movies using a content based movie recommendation system. Secondly, we use the similar movie information and voting data from the IMDb beta set to predict the movie's popularity using a deep learning approach. Finally, we compute target audience prediction using fuzzy C- means algorithm. We have introduced a new dataset that contains voting and rating information of recommended movies, which is used to predict the movie's popularity class. Additionally, we have proposed new parameters called Global centroid for each age group and presented a novel approach for estimating the interest or popularity of upcoming movies among distinct age groups.

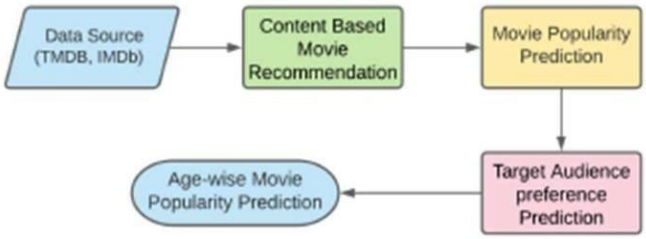

Fig-4: Framework Workflow

4. CONCLUSION

The study has proposed a novel approach for predicting the success probability of an upcoming movie. By using a content-based movie recommendation system and a CNN deep learning model, we have built a multiclass movie popularity prediction system that classifies movies into six classes. Additionally, we have proposed a system to predict the popularity of the upcoming movie among different age groups. The results of our study show that our model outperforms all the benchmark models with an accuracy of 96.8%. The potential of predictive and prescriptive data analytics in information systems to support industry decisions has been highlighted through this study. Overall, this study can be of great value to the film industry stakeholders who can leverage these predictions to make necessary changes to their upcoming movies and increase their chances of success.

REFERENCES

[1]. L. Sharma and A. Gera, "A survey of recommendation system: Research challenges," Int. J. Eng. Trends Technol., vol. 4, no. 5, pp. 1989–1992, 2013.
[2]. N. Das, S. Borra, N. Dey, and S. Borah, "Social networking in web based movie recommendation system," in Social Networks Science: Design, Implementation, Security, and Challenges. Cham, Switzerland: Springer, 2018, pp. 25–45.
[3]. P. Nagarnaik and A. Thomas, "Survey on recommendation system methods," in Proc. 2nd Int. Conf. Electron. Commun. Syst. (ICECS), Feb. 2015, pp. 1603–1608.
[4]. M. A. Hameed, O. Al Jadaan, and S. Ramachandram, "Collaborative filtering based recommendation system: A survey," Int. J. Comput. Sci. Eng., vol. 4, no. 5, p. 859, 2012.
[5]. B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithms," in Proc. 10th Int. Conf. World Wide Web (WWW), 2001, pp. 285– 295.
[6]. Y. Koren and R. Bell, "Advances in collaborative filtering," in Recommender Systems Handbook. Boston, MA, USA: Springer, 2015, pp. 77–118.