ISSN 2347 - 3983
Volume 8. No. 5, May 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter109852020.pdf
https://doi.org/10.30534/ijeter/2020/109852020

Movie Rating Prediction using Convolutional Neural Network based on Historical Values

Rudy Aditya Abarja[1], Antoni Wibowo[2]
[1] Computer Science Department, BINUS Graduate Program –Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, rudy.abarja@binus.ac.id
[2]Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, anwibowo@binus.edu

ABSTRACT

Movie rating is an important element to decide movie quality. It is like a summary to reflect the quality of all element inside a movie. People prefer to use rating as reference to decide before deciding to watch a movie or not. It is important to predict movie rating before it is released to maintain the objectivity of the movie rating. Many existing researches failed to address this problem because they used the post-release elements such as social media comments to predict movie rating. The other problem is the predicted rating is not intended for general people. Several researches used collaborative filtering, however the rating found was intended for specific people. To address limitations from previous researches, this study used historical values of the movie as features. Historical values could be generated from pre-released elements from the movie, it was created from relation between movie which based on movie similar attributes such as actor, director, genres, content rating, and production companies. By using historical values, objective prediction can be made even before the movie released. The proposed method was intended to make more accurate and general prediction for movie rating. In this study, usage of historical features and convolutional neural network (CNN) as model showed promising result.

Key words: Movie rating, rating prediction, historical values, convolutional neural network, CNN

1. INTRODUCTION

Movie is one of the most popular entertainment for people and it has become an integral part of our lives as a medium of relaxation and entertainment[1]. Not only as entertainment medium, movies have become a medium to learn various cultures all over the world. It has become the most important source of source of entertainment for people throughout the world, regardless of their diverse backgrounds, even movie can be considered a work of art that makes people crazy [2]. Every year many movies are released with various genres, story lines, and actors. In the past five years, US and Canada has released 765 movies per year in average. In 2019 alone, 835 movies were released in the United States and Canada, this number has increased by 70 movies compared to 2018[3]. With large number of movies released, people would need a guideline or measure to judge whether a movie is good or not so that they won't spend their money watching bad movies. People usually is not sure which movie to watch as entertainment in their free time [4].Moreover, watching bad movies can affect the mood of the audience [5].

As a guideline, most of the people will look for ratings or reviews that are obtained from people who have watched the movie. Numerical ratings have been shown to represent the contents of a review which iseasiertounderstandthan reviews[6]. Therefore, rating has an important role in determining whether a movie should be watched or not.

With the advancement of information technology and the easier access to the internet, people can easily access various information about movies. There are many websites that provide information about movies such as Internet Movie Database (IMDb), Rotten Tomatoes, Metacritic and The Movie Database (TMDb). These websites provide information about movies such as actors, directors, budgets, ratings and user comments. Among these websites, IMDb is the best consumer website that contains information about movies, such as: financial information, ratings, casts, reviews, crew, actors, directors, summaries, story lines etc. [4]. This site contains a large amount of data, which contains a lot of valuable information about general trends in movies [7].

An accurate movie rating prediction can help people to determine which movie to be watched. In addition, rating predictions are also beneficial for the economy. User ratings is form of Word of Mouth (WOM), ratingwhichis as statement made by consumers about a product that is available to other consumers on the internet[8]. User ratings are a good indicator for predicting the sales performance of a product in the future. Film industry experts agree that rating is a key factor in film success and helps film production companies and investors to gain financial success[9]. Companies can see which movies are likely to have good ratingsandmake strategies to utilize the movie to increase profits such as, by making merchandise such as, making merchandise for movies or create events and promotions related to the movie. Additionally, by utilizing the historical values obtained from previously released movies, rating predictions can be made before the film is produced. Film maker companies can make strategic plans and decisions related to the movie to be released to avoid loss.

Usually ratings and reviews come out after the movie is released and came from people who have watched the movie. However, the ratings generated when a new film is released tends to be biased because it comes from a group of people watching before the wider community watches[10]. One of the factors that could cause this bias is the possibility that the initial audience consists of viewers who are paid to give good ratings [11]. Given these problems, an objective rating is needed before the movie is released. The challenge of previous studies so far is to predict movie ratings before they are released in cinemas or even before they are produced. Ning, et al. mentioned in their study that many existing approaches fail to overcome this challenge because film rating predictions result from post-production factors such as commentary comments from social media[5]. Although there are several studies about movie rating prediction. The performance of rating prediction before the movie is released still needs to be increased. This study focuses on predicting rating based on data that can be obtained before the movie is released. Metadata attributes that were obtained from IMDb and TMDb were used as predictor. Attributes such as story line, artist, staff, director, genre, content rating, gross, and rating are used to create features.

One of the unique features that being used in this study is historical values. These features were created from relationship between a movie and previously released movies. It assumed that the predicted rating based on these historical values is certainly more objective than the rating from the audience that appeared when the movie was just released. Ning, et al. call this method a cohort rating prediction[5]. Cohort rating method looks for movies that have similarities based on existing historical attributes and values and makes a rating prediction based on those similarities. For example, moviewhichinvolvedJossWhedon(director) andRobert Downey Jr. (actor) would have similar rating with the movies which involved them before. Ning, et al. study showed that the addition of historical values in numerical features giving good result.

Many previous studies used basic machine learning model such as linear regression[12, 9, 13, 14], support vector regression (SVR), and k-NN[15, 16]. Ning, et al. used generative convolutional neural network (CNN) as model and compared it with several popular machine learning models. Ning, et al. study proved that deep learning models such as CNN and LSTM have better performance than basic machine learning models[5]. That study tried to maximize metadata attributes from movie and used several numerical features that contain historical values as features to predict movie rating before the movie is released.BesideNing, et al. study, CNN had goodtrackrecord in other case such as imageclassification[17, 18]. Nevertheless, optimizations were implemented in this study to increase CNN performanceandtooptimizeitformovieratingprediction. The usage of dropout method proved to increase CNN performance compared to baseline CNN.

2. PREVIOUS WORKS

Several methods to predict movie rating have been proposed based on several studies. Those methods were categorized based on their approach to collect data to predict movie rating either by (a) using collaborative filtering approach that utilize user account data, (b) only utilizing movie metadata, or (c) only utilizing social media data.

Previous studies [19, 16, 20]used a collaborative filtering approach. These studies utilized data from user data to make predictions, so the predicted rating is made for specific user. Lim & Teh used Netflix user data and Singular value decomposition (SVD) with variational Bayesian inference to predict movie rating [19]. Fikir, Yaz, & Özyeruse movie and user data from IMDb and use k-Nearest Neighborhood (kNN) as model [16]. Marović, et al. also used movie and user data from IMDb and compared content-based methods, collaborative methods, dan hybrid methods for collaborative filtering using different models [20]. The models that used in that study were regression tree, neural network, k-NN algorithm, personality diagnosis, SVD-kNN, dan latent variables (EM).

Beside using data from movie database, some studies used social media as data source. Oghina, et al.useIMDb data and social media data consisting of comments from Twitter and YouTube[12]. This study used linear regression to predict move rating. Twitter is one of favorite source to gather user comment, Schmit and Wubbenuse tweets data from Twitter as data source [13]. This study uses linear Regression (LR) dan support vector regression (SVR) as regression model and support vector classification (SVC) dan stochastic gradient descent classification (SGD) as classification model.

Another method proposed by Tan, et al.use video clip from the movie as a data source[21]. Features were extracted from frames which are taken from several parts of the movie. Hidden Markov model (HMM) was used as model to predict movie rating. Navarathna, et al. used videos taken from audience footage when watching a movie as data source [22]. Features are extracted from the audience movement while watching the movie. After that, SVR was used as model to predict movie rating. Khopkar and Nikolaev, 2017 conducted study to predict long-term movie rating [10]. The data source is taken from Movie Lens. This research focuses on the prediction of long-term rating by utilizing the rating given based on the initial rating given by the user. These studies have one thing in common, the source of the data is appeared after the movie was released. These studies did not focus on rating prediction before the movie was released.

Kabinsingha, et al. conducted an experiment to predict the Motion Picture Association of Amerika (MPPA) Rating of a movie based on film's attributes. MPPA rating is a rating given as a guide for parents about the content of a film. The data mining process was carried out to find terms that would affect MPPA rating[15].This study used kNN to predict MPPA Rating for a movie. Meanwhile, Hsu, et al. conducted a study to predict movie ratings based on movie attributes from IMDb such as genres, directors, actors, writers, country, film locations, and runtime[9]. Linear combination, multiple linear regression, dan neural networks are used as model in that study. Another study conducted by Zhu, et al. used the *douban.com* website as a data source[14]. Actors, directors, rating counts, wish counts, collect counts, comment counts, review counts, dates, and other data taken from a movie web page. This study used Gaussian mixture model (GMM) and linear regression as model. Another study by Basu was conducted using data derived from IMDb such as actors, actresses, directors, producers, screenwriters, and music directors in making movie rating prediction model based on artificial neural network (ANN) [23]. Meanwhile, Bristi, et al. conducted study for movie predictions by utilizing movies metadata from IMDb and Wikipedia[24]. Title, studio, director, screenplay, actor, actress, genre, country, year, and rating were used as features while bagging, random forest, decision tree, kNN and Naive Bayes as models.

Studies mentioned above[15, 9, 14, 23, 24] have the same weakness, weakness which are the features that being used were still in the form of raw metadata. This kind of approach did not produce strong relation between movies. Study by Ning, et al., 2018 took a different approach in processing data sources by categorizing movie metadata attributes into numerical, categorical, and topical features[5]. Numerical features were attributes in the form of numbers. Categorical features contained representation of categorical data such as genres and content ratings that are stored in the form of one hot encoding. Topical features contained a bag of words representation of the words contained in the movie storyline. Although the dataset was obtained from IMDb, feature extraction was conducted by considering historical aspects into account. In numerical features, there were historical values in the form of an average rating and gross of all related films. These values obtained from all movies released before the release date of each movie. The relationship between a movie and a previous movie was obtained from the similarity of the actor and director attributes. Generative convolutional neural network (CNN) was used as model.

Çizmeci and Ögüdücü perform the same approach as Ning, et al. However, IMDb, TMDb and Twitter were used as data source[25]. The features usedwere categorized into metadata features, extracted features from existing features, and social media features. Metadata features were data or attributes that contain information about movies, genres, budgets, revenue, release dates, languages, production companies, and production countries. Extracted features were features from existing attributes which obtained by utilizing the attributes in the metadata, such as the number of movies directed by the director, the number of movies played by the actor, the average rating of the movie directed by the directorand the actor, return of investment, day of release date, and month of release date. Social media features were data obtained from social media. These features consisted of the number of visits from the Wikipedia page of the movie, the number of hash tags on Twitter related to the movie, the number of mentions on Twitter related to the movie, the number of tweets on Twitter related to the movie, the subjectivity value of all the tweets related to the movie, the value of polarity (positive and negative sentiments), the number of followers from the director, and the number of followers from the actors. To preserve historical value, all Wikipedia data and tweets, except for the number of actor and director followers, were taken from the total social media related activities for 7 days before the movie's release date. This study by Çizmeci and Ögüdücü used Factorization Machines (FM) as model. The data collection methods used by Ning, et al. and Çizmeci & Ögüdücüare appropriate to obtain the historical value of the related movie[5][25]. However, Çizmeci and Ögüdücü did not mention about the method of processing the aggregate value of the metadata whether based on the movie’s release date or not, because the social media data collection is taken from 7 days before the movie release date. The dataset used was also less relevant because it is only based on movie data released in the United States during 2017.

Basuroy, et al. stated that the success of a movie is influenced by three factors, namely: star power, budgets, and critical review[26]. Star power is the popularity of the cast and director involved in the film. Budget is the amount of money spent to produce movies, the greater the budget, the more resources available to make movies. Critical review is a review given for the film therefore it can be described as a rating. Because of that, the rating has a relationship with star power and budget. Star power is obtained by finding the historical value of the cast or director obtained from films that involve the actor or director. Basuroy and Chatterjee state that the budget and MPPA rating with age ratings have a significant role in movie revenue[27]. Previous study also mentioned that the genre of the movie affects the number of viewers[28]. Because of that, genre could affect the movie rating indirectly, because more viewers would increase reviews quantity of the movie. Study by Ghiassi, et al. also mentioned the duration of the film does not have a direct influence on the success of the film, but the duration affects the decision of people going to watch the movie or not[28]. In addition, Ghiassi, et al. also mentioned that the film's release date has an influence on the number of viewers. Movies released in the holiday season and weekends would increase the viewers quantity. This is referred as seasonality factor.

The machine learning models used in the previous researches were mostly still using non-deep learning models, such as: SVD, k-NN, regression tree, linear regression, HMM, SVR, Factorization Machine, and simple neural network. There were only two studies that use deep learning models[5, 23]. Previous study by Ning, et al. proved that deep learning model performance is better than other models[5].

Based on the literature study, our research use data processing methods that used by Ning, et al. and Çizmeci & Ögüdücü[5, 23]. The results of Basu’s study showed that the addition of social media features did not provide a significant performance increase. Therefore, this research will focus on using metadata from movies as a feature. Based on performance comparison in Ning, et al. study[5]. CNN had better performance above another models, even it has better performance compared to LSTM as other deep neural network model.

3. PROPOSED METHOD

3.1 Data preparation

This study used datasets obtained from Kaggle. The datasets have comma separated value (CSV) format. There are two datasets, the first dataset source contains movie data from IMDb and the other contains movie data from TMDb. These datasets were combined into one dataset based on movie title and release year. After the dataset is collected, pre-processing process such as data cleaning, data transformation, and feature extraction are executed. The creation and selection of features aim to provide a good quality dataset that can improve the performance of the movie rating prediction.

3.1 Features

Features are created from movie attributes that have relationship with the success of a movie. The features are categorized into the following categories: historical features, numerical features, categorical features, topical features, and social media features.

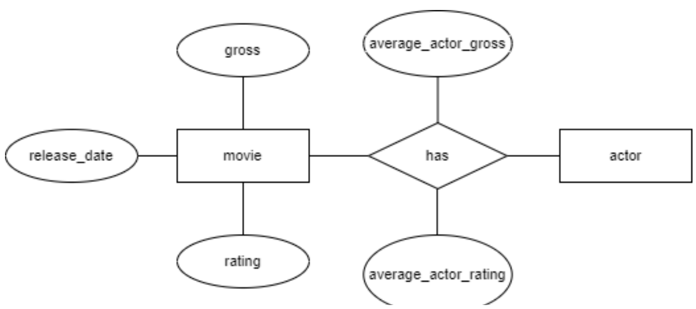

Figure 1: Relationship between Movie

Historical features. These features contain the historical value of a movie. Historical value is the performance value of related films that have been released previously. Related movies are all movies released before the release date of a movie and have the same relational attributes as the movie. This historical value includes the star power mentioned in the study [26]. Relationship attributes is needed to get the historical value of a movie, these attributes work as link to previously released movies. This relationship is described in Figure 1. The following attributes are used as relational attributes: actor, director, genres, production houses, and content rating.

Historical value describes the performance or success of all related movies that have been released previously. Performance can be taken from the rating and gross values. Rating of related movies certainly affects the rating of a movie. Based on previous research, movie rating is closely related to the financial success of a movie [9]. The gross attribute represents the gross income generated by the movie. With the existence of relational attributes, historical values in the form of rating and gross can be sought. Average rating and gross will be the historical value of a movie. All rating and gross average values that are based on all relational attributes will be used as new attributes. The average value is obtained from related films.

Table 1: Historical Features

Feature	Description
director_avg_rating	Average rating of all related movies directed by a director
director_avg_gross	Gross average of all relatedmovies directed by a director
director_sum _movies	The number of films a director has directed
actor_avg_rating	Average rating of all related movies directed by a director
actor_avg_gross	Gross average of all related movies directed by a director
actor_sum_movies	Number of films a director has directed
content_rating_avg_rating	Average rating of all related movies that have related content rating
content_rating_avg_gross	Average gross of all related movies that have related content rating
genres_avg_rating	Average rating of all related movies that have related genres
genres_avg_gross	Average gross of all related movies that have related genres
companies_avg_rating	Average rating of all related movies produced by all related production companies
companies_avg_gross	Average gross of all related movies produced by all related production companies
companies_sum_movies	Number of movies that have been produced by all related production companies

All average values are obtained from the average rating or gross of films released before the release date of a movie that has the same relational attributes. A movie can have more than one genre and production company. Average rating and average gross values are obtained from the average (mean) of the total average value of each genre or production company owned by movie, if one of the average values is null then the average value is not included in the total average calculation, as described in Equation (1). Null average data is not included in calculationbecause it can create imbalance in total average value. For example, the total average value of a movie become very small if thatmovie has two of three genres which contain null (0) average value. Thiscanbehappenbecause related genre does not haveany related movie released before the movie release date. Because of that we decided to count out the null average data and only calculate non null data.

$$Average = \frac{total\ non\ null\ average\ genre}{total\ non\ null\ genre} \quad (1)$$

Sum value is obtained from the number of movies released before the release date of a movie that has the same relational attributes. The value of *companies_sum_movies*is total of the number of movies produced by all related production companies.

Metadata features. Metadata features are the numerical attributes of a movie that give information about the movie. Some feature values are already in the dataset and some are aggregate values.

Table 2: Metadata Features

Feature	Description
budget	Budget spent on movie
duration	Run time duration of movie
total_companies	Number of production companies involved in making this movie
release_date_day	Day of movie release date
release_date_month	Month of release date
release_date_year	Year of release date
total_spoken_language	Number of spoken languages used in movie's audio
total_production_countries	Number of countries used for movie's production

Categorical features. In addition to numeric attributes, in the movie metadata there are also attributes in the form of text. Some text attributes contain explicit information to categorize movies and there are also attributes that are implicit and must first extract the information before it can be used to categorize movies. Categorical features are features that contain explicit categorical attributes.

Table 3: Categorical Features

Feature	Description
genres	Movie's genres
content_rating	Movie's content/ age rating

Because of*genres* and *content_rating*is text data; transformation is needed to get value from these attributes. There are 26 genres and 15 content ratings available in the dataset. Each genres and content ratings are stored as one attribute. One hot encoding process is performed to give a value of 1 if the movie has the value and 0 if the movie doesn't have the value.

Topical features. Topical features are features that extracted from text attributes as the results of the topic modeling process. The unsupervised machine learning process is carried out to get the topics from the existing set of terms. Terms are extracted from attributes in Table 4. These attributes consist of many terms that can be used to create term dictionary. These attributes will be combined and treated as one text attributes. After that, corpus is created by storing the number of occurrences of terms within a document to create Bag of Word (BOW) representation. Latent Dirichlet allocation (LDA) is used to create topical model and 20 topics is generated. Each topic becomes one attribute and is filled by a value generated from the level of similarity in a movie 's combined text attribute with a topic.

Table 4: Source of Topical Features

Attributes	Description
movie_title	Movie's title
plot_keywords	Keywords from movie's storyline
overview	Movie's storyline summary
tagline	Movie's tag line

Social features. Social features are feature that obtained from social media data. Previous study shows that the usage of social media data as predictor doesn't give significant performance boost in rating prediction accuracy[25]. IMDb dataset has several social media data such as actor, movie, director and total cast Facebook like count. This study uses actor and director Facebook count as features because these attributes can contribute to star value of a movie. Movie Facebook like count isn't used because it is assumed as post-release data.

Table 5: Social Features

Feature	Description
actor_likes	Number of Facebook likes for actor/ actress page
director_likes	Number of Facebook likes for director page
cast_total_likes	Total number of Facebook likes for entire movie cast

MinMaxScaler. All the mentioned features above are numeric data and have different scale. *Budget* and features that generated from *gross* have big amount compared to other features value. Because of that, data normalization is needed. MinMaxScaler is used to transforms all features value to new value that have same scale (0-1). MinMaxScaler formula is shown in Equation 2. Before dataset is processed by MinMaxScaler, all features with numeric data that has null value are filled by the median of the features.

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

count	3819.000000	count	3819.000000
mean	108.348259	mean	0.294838
std	22.830936	std	0.071347
min	14.000000	min	0.000000
25%	94.000000	25%	0.250000
50%	104.000000	50%	0.281250
75%	118.000000	75%	0.325000
max	334.000000	max	1.000000

Figure 2: Example of data transformed by MinMaxScaler

3.1 Model

Previous study shows that deep learning models have better performance in predicting movie rating than basic machine learning models [5]. That study uses generative convolutional neural network (CNN) as model and shows generative CNN outperform all baseline model. This study uses one-dimensional CNN (1DCNN) as model. This is because the dataset used is of tabular data type. Each data consists of only one row of data which means the data is one dimensional. The input layer isthefeatures in datasetandtheoutput layer is movie rating.

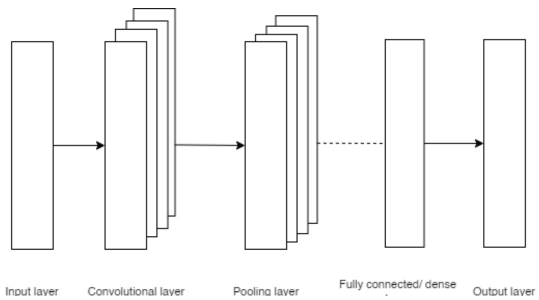

Figure 3: Architecture of 1DCNN

Convolutional layer and pooling layer still exist in 1DCNN and have different behavior compared to 2DCNN or 3DCNN.

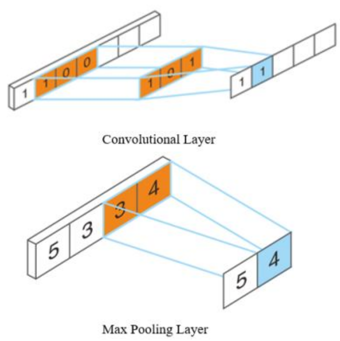

Figure 4: Convolutional and max pooling layer in 1DCNN

This study tries to optimize CNN by tuning the layer number, node number, activation method and uses dropout as regularization method. Dropout is proved as effective solution to solve common problem in deep learning model such as overfitting [29]. Figure 5 shows the deep neural network architecture that implemented dropout.

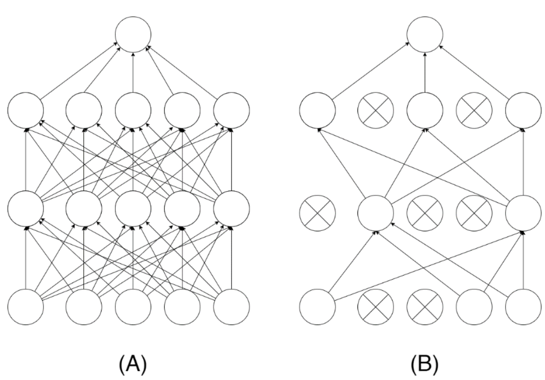

Figure 5: (A) show standard deep neural network model and (B) shows deep neural network model with dropout

4. EXPERIMENTS

There were four experiments that would be used in this study. The first experiment purpose would be to compare the performance of CNN model with other regression models in making predictions from tabular dataset. The second experiment would compare all previous models in the first experiment with modified CNN. Third experiment would be conducted to compare the extracted features performances in movie rating prediction. Fourth experiment would be conducted to compare several feature set performances and get the best feature set.

The dataset used in these experiments was the merged and pre-processed dataset from IMDb and TMDb dataset. There were 4.317 movies(released in 1916 – 2016) data in the dataset. The dataset was separated into training dataset (movies released from 2000 until 2013) and test dataset (movies released after 2013). Training dataset had 3.819 movies and test dataset has 498 movies. All the dataset already gone through pre-processing Different features set also used to explore how the data pre-processing process affects the movie rating prediction performance.

Mean squared error (MSE) and mean absolute error (MAE) were used as the performance metrics. MSE and MSE show the average of magnitude error for the predicted ratings. MAE formula is shown in Eq (3) and MSE is shown in Eq (4).

$$Mean\ absolute\ error\ =\ \frac{1}{N}\sum_{i=1}^{N}|f_i - \tilde{y}_i| \qquad (3)$$

$$Mean\ square\ error\ =\ \frac{1}{N}\sum_{i=1}^{N}(f_i - \tilde{y}_i)^2 \qquad (4)$$

This study uses 1DCNN as model. We created 1DCNN without optimization and used it as baseline model. This model was created for a baseline to compare how the optimization process affects the performance. Figure 6 shows the architecture of baseline CNN used in this study. RMSprop used as optimized method, batch size was set to 10 for training proses and epochs was set to 500.

Figure 6: Baseline CNN architecture

Gaussian process regressor (GPR), decision tree regressor (DTR), kNN regression (kNNR), and artificial neural network (ANN) were used as comparison model. All models were trained on 500 epochs, same as the epochs in the CNN. The dataset used in the first experiment was raw dataset with all movie attributes with numeric value, expect id, aspect ratio and average vote score. First experiment result could be seen in Table 6. The result in Table 6 shows that CNN had the best performance compared to other models with MSE 1.48 and MAE 0.94. The result shows that CNN had good prediction performance with tabular data as source.

Table 6: Model performance comparison

Model	MSE	MAE
GPR	39.95	6.21
ANN	31.07	5.44
DTR	1.99	1.06
kNNR	1.52	0.96
Baseline CNN	**1.48**	**0.94**

Figure 7 shows the architecture of modified CNN, after we optimized it. Relu was used as activation function, it changed from tanh activation function used in baseline CNN expect from the second convolutional layer. One max pooling layer was added and one dense layer was removed. Node number in remaining dense layer was modified and two drop out layers was added in this architecture as generalization method.

Figure 7: Optimized CNN architecture

Second experiment used all features extracted from the dataset (metadata, historical, categorical, topical and social) and all the features which had numeric value was processed by MinMaxScaler.

Table 7: Model performance comparison

Model	MSE	MAE
ANN	3.18	1.54
GPR	2.53	1.28
DTR	1.94	1.06
Baseline CNN	1.46	0.95
kNNR	1.30	0.87
Optimized CNN	**1.22**	**0.85**

Table 7 shows our optimized CNN got best result with MSE 1.22 and MAE 0.85. This result show that our modification process gave positive impact to CNN performance, our modification gave increase in performance by 0.24 in MSE and 0.1 in MAE compared to base CNN.

Third experiment conducted to observe the impact of features and pre-processing. Table 8 shows the performance of each feature used in this study. Optimized CNN used as model in this experiment.

Table 8: Features performance comparison

Features	MSE	MAE
Topical features	1.50	0.95
Metadata features	1.48	0.95
Social features	1.47	0.95
Categorical features	1.46	0.98
Historical features	**1.44**	**0.95**
Social features (minmax)	1.39	0.92
Metadata features (minmax)	1.36	0.92
Historical features (minmax)	**1.31**	**0.89**

Historical features had the best performance compared to other features by getting MSE 1.44 and MAE 0.95. The result also showed that using MinMaxScaler as data for numerical data give significant improvement. After processed by MinMaxScaler, historical features still got best performance with MSE 1.31 and MAE 0.89.

Table 9: Features set experiment

Features set	MSE	MAE
Numeric raw dataset	1.36	0.92
All features (without minmax)	1.34	0.90
Categorical + topical	1.40	0.92
Categorical + topical + social	1.33	0.90
Metadata + historical	1.29	0.88
Metadata + categorical + topical + social	1.28	0.91
Metadata + historical + social	1.26	0.88
Metadata + categorical + topical	1.25	0.85
Historical + categorical + topical + social	1.23	0.88
All features (minmax)	1.22	0.85
Metadata + historical + topical + categorical	**1.20**	**0.83**

Fourth experiment conducted to observe the impact of features selection in movie rating prediction performance, features were grouped in several features set. The model used in this war our optimized CNN. Fourth experiment results are shown in Table 9. The combination of historic, numeric, categorical and topical feature with MinMaxScaler gave the best performance with MSE 1.20 and MAE 0.83. Results showed that historical features give positive impact to rating prediction performance.

Features set which included historical features had better performance compared to the feature set that did not include it. We also found that social features did not always gave positive impact to rating prediction performance. The number of features set was also affecting the performance, in this experiment we found that the more features used the better prediction performance became. The features set which had 4 or 5 features had better performance than the features set which had less than 4 features. In addition, we found that our pre-processing gave positive impact in performance compared to use all numeric attributes from raw dataset as features.

5. CONCLUSION

Our study proved that one dimensional CNN (1DCNN)had promising result as model to predict movie rating based on the tabular dataset. Therefore, CNN could be used as model for small textual dataset. We also found that features have an important factor to building prediction model. Moreover, the usage of historical features as predictor gave positive impact to rating prediction performance. Historical values that created from previous movie show gave possibility to predict more accurate rating before the movie released.

Although our CNN in this study gave promising result, it still has many rooms to improve. Several optimization method and architecture tuning could be implemented in our model to increase its performance. Methods such as hybrid model and hyper parameter optimization could be implemented to achieve better performance. Beside the model, this study needs more exploration in feature extraction and selection, several features could be extracted from the metadata attributes. Better features also could give better performance in rating prediction. The usage of better and updated dataset also could provide better features.

REFERENCES

1. R. Parimi and D. Caragea, "**Pre-release Box-Office Success Prediction for Motion Pictures**," in *International Workshop on Machine Learning and Data Mining in Pattern Recognition. , , , .*, Berlin, Heidelberg, 2013. https://doi.org/10.1007/978-3-642-39712-7_44
2. S. Mundra, A. Dhingra, A. Kapur and D. Joshi, "**Prediction of a movie's success using data mining techniques**," in *Information and Communication Technology for Intelligent Systems*, Singapore, 2019. https://doi.org/10.1007/978-981-13-1742-2_22
3. "**2019 THEME Report**," Motion Picture Association of America (MPAA), Washington, D.C., 2019.
4. M. H. Latif and H. Afzal, "**Prediction of Movies popularity Using Machine Learning**," *IJCSNS International Journal of Computer Science and Network Security, VOL.16 No.8,* pp. 127-131, 2016.
5. X. Ning, L. Yac, X. Wang, B. Benatallah, M. Dong and S. Zhang, "**Rating prediction via generative convolutional neural networks based**," *Pattern Recognition Letters,* pp. 1-9, 2018.
6. S. M. Mudambi and D. Schuff, "**What makes a helpful review? A study of customer reviews on Amazon.com**," *MIS Quarterly, 34,* pp. 185-200, 2010. https://doi.org/10.2307/20721420
7. M. Saraee, S. White and J. Eccleston, "**A data mining approach to analysis and prediction of movie ratings**," *WIT Transactions On Information And Communication Technologies 33,* pp. 343-352, 2004.
8. R. Ullah, A. Zeb and W. Kim, "**The impact of emotions on the helpfulness of movie reviews**," *Journal of Applied Research and Technology 13.3,* pp. 359-363, 2015. https://doi.org/10.1016/j.jart.2015.02.001
9. P.-Y. Hsu, Y.-H. Shen and X.-A. Xie, "**Predicting movies user ratings with imdb attributes**," in *International Conference on Rough Sets and Knowledge Technology*, Cham, 2014.
10. S. S. Khopkar and A. G. Nikolaev, "**Predicting long-term product ratings based on few early ratings and user**," *Electronic Commerce Research and Applications 21,* pp. 38-49, 2017.
11. J. Lorenz, "**Universality in movie rating distributions**," *The European Physical Journal B 71.2,* pp. 251-258, 2009. https://doi.org/10.1140/epjb/e2009-00283-3
12. A. Oghina, M. Breuss, M. Tsagkias and M. de Rijke, "**Predicting IMDB Movie Ratings Using Social Media**," in *Advances in Information Retrieval*, Barcelona, 2012.
13. W. Schmit and S. Wubben, "**Predicting ratings for new movie releases from twitter content**," in *In Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Lisboa, 2015. https://doi.org/10.18653/v1/W15-2917
14. J. Zhu, Y. Guo, J. Hao, J. Li and D. Chen, "**Gaussian Mixture Model based prediction method of movie rating**," in *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*, Chengdu, 2016.
15. S. Kabinsingha, S. Chindasorn and C. Chantrapornchai, "**A movie rating approach and application based on data mining**," *International Journal of Engineering and Innovative Technology (IJEIT) Volume 2,* pp. 77-83, 2012.
16. O. B. Fikir, İ. O. Yaz and T. Özyer, "**A Movie Rating Prediction Algorithm with Collaborative Filtering**," in *2010 International Conference on Advances in Social Networks Analysis and Mining*, Odense, 2010. https://doi.org/10.1109/ASONAM.2010.64
17. S. Mothe, A. S. Teja, B. Kakumanu and R. K. Tata, "**A Model for Assessing the Nature of Car Crashes using Convolutional Neural Networks**," *International Journal of Emerging Trends in Engineering Research,* vol. 8, no. 3, pp. 859-863, 2020. https://doi.org/10.30534/ijeter/2020/41832020
18. R. I. M. Jr, A. S. Alon, M. A. F. Malbog and J. S. Gulmatico, "**Copra Meat Classification using Convolutional Neural Network**," *International Journal of Emerging Trends in Engineering Research,* vol. 8, no. 2, pp. 435-439, 2020. https://doi.org/10.30534/ijeter/2020/30822020

19. Y. J. Lim and Y. W. Teh, "**Variational Bayesian approach to movie rating prediction**," *Proceedings of KDD cup and workshop,* pp. 15-21, 2007.
20. M. Marović, M. Mihoković, M. Mikša, S. Pribil and A. Tus, "**Automatic movie ratings prediction using machine learning**," in *MIPRO 2011*, Opatija, 2011.
21. D. S. Tan, S. See and T. J. Tiam-Lee, "**Automatic rating of movies using an arousal curve extracted from video features**," in *International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, Puerto Princesa, Palawan, 2014.
22. R. Navarathna, P. Lucey, P. Carr, E. Carter, S. Sridharan and I. Matthews, "**Predicting movie ratings from audience behaviors**," in *IEEE Winter Conference on Applications of Computer Vision*, Steamboat Springs, CO, 2014.
23. S. Basu, "**Movie Rating Prediction System Based on Opinion Mining and Artificial Neural Networks**," in *International Conference on Advanced Computing Networking and Informatics*, Singapore, 2019.
24. W. R. Bristi, Z. Zaman and N. Sultana, "**Predicting IMDb Rating of Movies by Machine Learning Techniques**," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.*, Kanpur, 2019.
25. B. Çizmeci and Ş. G. Ögüdücü, "**Predicting IMDb ratings of pre-release movies with factorization machines using social media**," in *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, Sarajevo, 2018. https://doi.org/10.1109/UBMK.2018.8566661
26. S. Basuroy, S. Chatterjee and S. A. Ravid, "**How critical are critical reviews? The box office effects of film critics, star power, and budgets**," *Journal of marketing, 67(4),* pp. 103-117, 2003.
27. S. Basuroy and S. Chatterjee, "**Fast and frequent: Investigating box office revenues of motion picture sequels**," *Journal of Business Research, 61(7),* pp. 798-803, 2008. https://doi.org/10.1016/j.jbusres.2007.07.030
28. M. Ghiassi, D. Lio and B. Moon, "**Pre-production forecasting of movie revenues with a dynamic artificial neural network**," *Expert Systems with Applications, 42(6),* pp. 3176-3193, 2015. https://doi.org/10.1016/j.eswa.2014.11.022
29. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "**Dropout: A Simple Way to Prevent Neural Networks from Overfitting**," *Journal of Machine Learning Research 15 ,* pp. 1929-1958, 2014.
30. K. Lee, J. Park, I. Kim and Y. Choi, "**Predicting movie success with machine learning techniques: ways to improve accuracy**," *Information Systems Frontiers 20, no. 3,* pp. 577-588, 2018. https://doi.org/10.1007/s10796-016-9689-z
31. K. R. Apala, M. Jose, S. Motnam, C.-C. Chan, K. J. Liszka and F. d. Gregorio, "**Prediction of Movies Box Office Performance Using Social Media**," in *ASONAM '13: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Niagara Ontario Canada, 2013. https://doi.org/10.1145/2492517.2500232