

# CE7454 Project: Simultaneous Prediction of Box-office and Movie Rating using Multimodal Neural Network

Jie Li, Meng Shen, Ruihang Wang\*  
*School of Computer Science and Engineering*  
*Nanyang Technological University*  
Email: {jie006, meng005, ruihang001}@e.ntu.edu.sg

## I. MOTIVATION

The world's movie industry has witnessed an unprecedented boom in recent years. With heavy investments in today's movie industry, reliable predictions of a movie's box-office revenue before its theatrical release is essential for producers to reduce financial risk. From the perspective of viewers, however, movie rating is an important factor to measure the quality of a movie.

In this project, we are interested in predicting both box-office revenue and movie rating in an early stage, aiming to provide useful information for producers, distributors and viewers.

## II. PROJECT INTRODUCTION

The success of a movie has been considered as an unpredictable problem due to its social complexity, we are trying to develop our own model to disclose the relationship between movie-related variables.

Specifically, we will crawl raw data from IMDb. After preprocessing, the original dataset will be used to train a multimodal neural network for simultaneous movie rating and box-office revenue prediction. Finally, the potential relationship among movie-related variables will be analyzed and discussed for further investigation.

## III. EXISTING SOLUTIONS

It is not a whole new idea to predict the financial success of a movie by using computational models, several studies have attempted to develop approaches for box-office revenue prediction.

For example, movie trailers were used as input data for a linear classifier to predict the opening-week box-office revenues [1]. In [2], several nonlinear regression algorithms were employed for building box-office forecasting models. Somdutta Basu built a vanilla neural network for movie rating prediction [3].

## IV. PROPOSED SOLUTION

The majority of existing solutions, however, focus on either box-office revenue or movie rating prediction alone. The project will present an approach for simultaneous prediction

of box-office revenue and movie rating using special designed multimodal neural network. The raw data will be crawled from IMDb, including three types of inputs:

- 1) Movie Poster: image data containing potential information of a movie
- 2) Movie Genre: a binary vector of length 22
- 3) Movie Metadata: numerical data such as duration, budget, actors etc.

## V. MILESTONES

### A. Data Acquisition & Data Cleaning

To deliver this proposal, first we will crawl data from IMDb, together we will clean and normalize those data for further analytics, this work should be done before week 9.

### B. Apply Deep Learning Methods

Simultaneously, since we already have a plan on the data type in use, neural network designing could start at the same time, this one should be done before week 10, by the mean time we will have an effective neural network.

### C. Results Analytics

Finally, we will compare our model with different baseline architectures, according to the result, we will try to make some optimizations for our model to improve its accuracy. All deliverables and the final slides will be done in week 12.

## REFERENCES

- [1] Tadimari, A., Kumar, N., Guha, T., & Narayanan, S. S. (2016). "Opening big in box office? Trailer content can help." *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. doi: 10.1109/icassp.2016.7472183.
- [2] Kim, Taegu, et al. "Box Office Forecasting Using Machine Learning Algorithms Based on SNS Data." *International Journal of Forecasting*, vol. 31, no. 2, 2015, pp. 364–390., doi:10.1016/j.ijforecast.2014.05.006.
- [3] Basu, Somdutta. "Movie Rating Prediction System Based on Opinion Mining and Artificial Neural Networks." *International Conference on Advanced Computing Networking and Informatics Advances in Intelligent Systems and Computing*, 2018 pp. 41–47.

\*Authors are sorted by initials.