

Summer 6-26-2019

## **A Movie Weekly Box-office Revenues Prediction Model Based on Online Reviews**

Minglei Li

*School of Information Management and Statistics, Hubei University of Economics, China*

Guoyin Jiang

*School of Public Affairs and Administration, University of Electronic Science and Technology of China, China, jiangguoyin@uestc.edu.cn*

Junli Lei

*School of Information Management and Statistics, Hubei University of Economics, China*

Follow this and additional works at: <https://aisel.aisnet.org/whiceb2019>

---

### **Recommended Citation**

Li, Minglei; Jiang, Guoyin; and Lei, Junli, "A Movie Weekly Box-office Revenues Prediction Model Based on Online Reviews" (2019). *WHICEB 2019 Proceedings*. 82.

<https://aisel.aisnet.org/whiceb2019/82>

This material is brought to you by the Wuhan International Conference on e-Business at AIS Electronic Library (AISeL). It has been accepted for inclusion in WHICEB 2019 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# A Movie Weekly Box-office Revenues Prediction Model Based on Online Reviews

Minglei Li<sup>1</sup>, Guoyin Jiang<sup>2\*</sup>, Junli Lei<sup>1</sup>

<sup>1</sup> School of Information Management and Statistics, Hubei University of Economics, China

<sup>2</sup> School of Public Affairs and Administration, University of Electronic Science and Technology of  
China, China

**Abstract:** To predict the movie weekly box-office revenues, this paper proposes a new prediction model based on ensemble machine learning method. Firstly, we extract some important features from movie online reviews. Then, due to the limited ability of the single machine learning model, an ensemble machine XGboost is employed to predict the movie weekly box-office revenues in this paper. Finally, we collect the movie online reviews from Douban.com, and use about 600 movies to verify the performance of the model. The experimental results show that the effectiveness and practicability of this model.

Keywords: box-office revenues prediction, machine learning, online reviews, text analysis

## 1. RESEARCH QUESTION

The film industry in China has attracted more and more capital investment. However, the film industry is a high-investment and high-risk industry. Even so a lot of movies cost high production, but the box-office performance is poor. Therefore, the research of movie box-office revenues prediction is very important and meaningful <sup>[1]</sup>. The movie box-office revenues prediction is divided into early prediction and later prediction according to the timing of prediction. Compared with the early prediction, the later prediction, such as the weekly box-office revenues prediction, is more conducive to flexible selection of marketing strategies to improve movie revenue. And, the later prediction is more challenging. There are many factors that will affect the movie box-office revenues <sup>[2]</sup>. Many researches empirically revealed that the online movie reviews significantly influence the movie box-office revenues <sup>[3, 4]</sup>, but few literatures explore how to effectively predict the box-office revenues based on online movie reviews. This paper focuses on predicting the movie weekly box-office revenues based on the online movie reviews by using machine learning methods.

## 2. MAJOR RESEARCH FINDINGS

In this paper, a model based on **ensemble** machine learning model XGboost <sup>[5]</sup> is designed for movie **weekly box-office revenues prediction**. The model is named En-WBP. In En-WBP, we collect the movie online reviews from the Douban.com (a popular movie review website in China), and choose the related characteristics to predict the movie weekly box-office revenues: number of reviews in a week, proportion of one star, two stars, three stars, four stars and five stars in a week, sentiment average value of the reviews text in a week, proportion of extreme negative emotion, proportion of extreme positive emotion, similarity of the continuous text in a week, movie genre, movie release year, number of holiday days in a week and number of holiday days in the next week. We use these features to predict the movie weekly box-office releases in next week.

The core of prediction model En-WBP is XGboost. The XGboost is an extension of the gradient boosting decision tree algorithm. XGboost belongs to ensemble learning model, and its basic idea is to combine multiple Classification And Regression Trees (also known as CART).

The movie box-office data are collected from the Entgroup Chinese box-office website. The movie features are collected from the Douban.com. After removing the movies with missing data, we choose more than 600

---

\* Corresponding author. Email: jiangguoyin@uestc.edu.cn (Guoyin Jiang)

movies as the research object and get 3202 samples. The samples data is divided into two parts randomly. The one part, which is 80% of the total data, is the training set. The other one is the testing data. The model in this paper is developed based on XGboost library, and we implement the model by Python3 language.

We compare En-WBP with some familiar and excellent machine learning models including that linear regression model (LR), decision tree model (DT), ANN model and SVM regression model. The models of LR, DT, ANN and SVM are all implemented using Python's machine learning library scikit-learn. The comparison results are shown in Figure 1. It shows the prediction relative accuracy rate with 20% absolute error rate using different models. The prediction relative accuracy rate of En-WBP is the highest in the five prediction models for both training samples and testing samples. The results show that the effectiveness of our model.

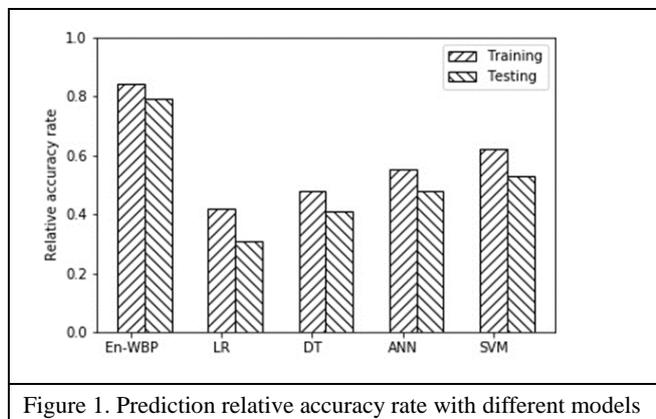


Figure 1. Prediction relative accuracy rate with different models

### 3. CONCLUSIONS

Movie box-office revenues prediction plays important role in the film industry operation. The main contributions of this paper are: (1) online movie reviews can be effectively used to predict the weekly movie box-office revenues, (2) we can conduct text mining to abstract abundant feature information form the online reviews, and (3) the provided ensemble machine learning model have the effectiveness and practicability in movie box-office prediction. From the result in this paper, it can be found that the movie online reviews has an important influence on the movie weekly box-office revenues, and the XGboost model can predict the movie weekly box-office revenues accurately. However, there are many other factors that affect the movie weekly box-office revenues. Our ongoing work is to consider more possible factors, and use some famous and more effective models to predict the movie box-office revenues.

### ACKNOWLEDGEMENT

This work was supported by a grant from the National Natural Science Foundation of China (No. 71671060); the Fundamental Research Funds for the Central Universities (Nos. ZYGX2017 KYQD185,2018-QR-05); and the Science and Technology Program of Sichuan (No. 2019JDR0011).

### REFERENCES

- [1] Ghiassi M, David L, Brian M. (2015). Pre-production forecasting of movie revenues with a dynamic artificial neural network. *Expert Systems with Applications*, 42(6):3176–3193.
- [2] Jong H L, Sun H J, Jaehong P. (2017). The role of entropy of review text sentiments on online wom and movie box office sales. *Electronic Commerce Research and Applications*, 22:42–52.
- [3] Minhoe H, Pilsung K, Sungzoon C. (2016). Box-office forecasting based on sentiments of movie reviews and independent subspace method. *Information Sciences*, 372:608–624.
- [4] Ho C B L, Li X. (2018). Impact of online word of mouth on channel disintermediation for information goods. *Journal of Management Information Systems*, 35(3):964–993.
- [5] Tianqi C, Carlos G. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, ACM, 785–794.