IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol.-7, Special Issue, 4, Feb 2019 E-ISSN: 2347-2693

Predicting Ticket Sales Using Web-Based External Factors and Box-Office Data

M. Keerthana[1*], J. Rabacca Cinthiya[2]

[1,2]M.Sc Computer Science, Idhaya College for Women, Kumbakonam, Tamilnadu, India

Corresponding Author: keerthana567@gmail.com

Available online at: www.ijcseonline.org

Abstract**—**Posting online reviews and rating their satisfaction purchased products has become an increasingly popular way to share the information for anonymous candidates who has interest in purchasing the product. In addition, people leave their interests and near-future purchasing plan on the web such as search history and search query volume. From this phenomenon, the prediction of sales performance is possible in many products by mining the data sets which are left on the web by consumers' online activities. In this paper, we focused on the movie ticket sales which word-of-mouth effect is prominent, and our goal is to forecast the sales performance of the near-weekend using box-office data and external factors such as online reviews, star ratings and search volume. For this work, we gather 1.7 million online reviews and movie ratings, and we also gather the daily search volume of movies' title for past three years. Using machine learning techniques and linear modeling, we develop a model for high-accuracy predicting of ticket sales on near-future. We also analyze a relationship between ticket sales performance on weekends and box-office data, online reviews, star ratings, and search volume. Through this work, we support to decide the ideal number of screens for a given weekend, thus it contributes to a substantial increase in the rate of profit on movie markets.

Keywords**—**Box-Office Data, Online Reviews, Star Ratings, Ticket Sales.

I. INTRODUCTION

With the development of Web 2.0 technologies and the diffusion of mobile Internet access, people can easily obtain and share the information through the Internet. The rise of social networking services (SNS) also accelerates the sharing of information and their opinion via the Internet. This phenomenon leads to significant changes in the pattern that people make a decision to purchase products. Internet users get information of products through various sources such as online communities, newsgroups and SNS before the new products is released. Then they use search engines in order to collect details in products. In this process, the customer makes decisions on products and purchases them through e-commerce services. After the purchase, the consumer spends time to write online reviews on multiple platforms for anonymous candidates who has interest in purchasing the product. This series of patterns mentioned above, would not be a new phenomenon anymore. From this pattern, the prediction of sales performance is possible in many products by mining the data sets which is left on the web by consumers' online activities. Based on these ideas, numerous studies have shown that search patterns and online reviews have a close relationship with the product's sales performance [1], [2]. Moreover, there have been several works to forecast new product sales by web data analysis in various categories [3], [4]. For the improvement of prediction accuracy, several studies introduced prediction models combined with machine learning techniques [5], [6]. Prior studies on the predicting sales performance with web data analysis have been treated in various categories (e.g. book, appliances, video games, etc.). In this paper, we present the prediction model for ticket sales on near weekend by analyzing box-office data, online reviews, star ratings and search term volume data of the past three years. The reasons that we choose the movie domain, first of all, is that it is easy to collect the precise measured data of the past sales performance. Second, the movie domain is familiar to purchasing patterns through word-of-mouth (WOM) as people recommend movies through online reviews and star ratings [1]. Third, it is relatively simple to crawl the web data, because reviews and star ratings are gathered in a few critic reviews websites and portal services. At last, in 2012, global theatrical market is a huge size market of $34.7 billion, and specifically the Korean film market size is $1.3 billion, which is about 3.8 percent of the global market [7]. Thus, a high revenue growth can be expected when the cinema decides the ideal number of screens through an accurate prediction of near-future ticket sales.

In previous works, they provided experimental results of the relationship between online reviews, search term volume data and box-office performance [1], [2]. Or they proposed prediction models for total revenue of newly released movies [3], [5], [8]. Compared with the previous work above, we propose a predicting model of ticket sales for every weekend by analyzing the time-series data of movies' internal factors and external factors. In summary, we make the following contributions:

- We present an empirical result, which maps a relationship between ticket sales performance on weekends and box-office data, online reviews, star ratings, and search query volume.
- We develop a model for high-accuracy predicting of ticket sales on the nearest weekend using several machine learning techniques and linear modeling.
- Using proposed prediction models we analyze how much the accuracy is improved by using external factors (e.g. volume of online reviews, star ratings and search volume), which reflect ticket sales performance.
- We help to decide the ideal number of screens on weekends, and then bring a direct effect with improving profit in movie markets through this advantage. The rest of this paper is organized as follows. Section II introduces studies related to our works. Section III explains the method of data acquisition, preprocessing and describes characteristics of data sets. Section IV gives the overall of proposed prediction system and an experiment result. The final section provides the conclusion and direction for future development.

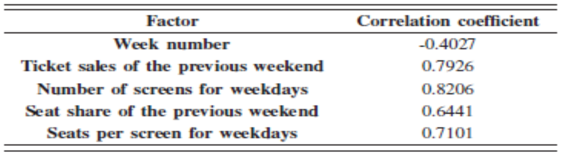
II. Methodology

1) Box-office Data: We gathered the following features from the box-office data and aligned the data on a weekly basis: Week number after the opening week, ticket sales of the previous weekend, seat share of the previous weekend, number of screens for weekdays, seats per screen for weekdays. Our goal is to determine the number of screens throughout the predicted ticket sales using our proposed prediction model. The South Korean film market determines the number of weekend screens on Wednesday to sell tickets in advance. With reference to this fact, we preprocess the features of the boxoffice data using the data available before Wednesday. TABLE II shows the correlation between our features and ticket sales of each weekend. Week number after the opening week shows negative correlation with ticket sales of the weekend. The correlation between ticket sales of the previous weekend and the number of screens weekdays are 0.7926, 0.8206. So these features have a strong correlation with ticket sales of weekend.

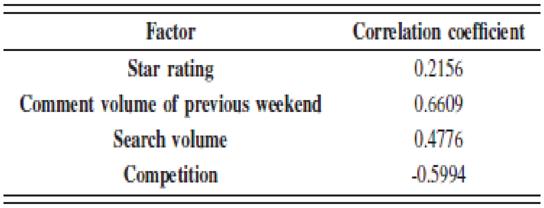
2) Online reviews, Star rates and Search Volume: In contrast to the box-office data of movies, we acquired the following external factors using online reviews, star ratings and search volume data: movie rating, comment volume of previous weekend, search volume and competition. We acquired search volume and competition using the formula explained previously.

III. Results and Discussion

TABLE I shows the correlation coefficient between the external factors and ticket sales of each weekend. As shown in the TABLE II, the volume of comments is more correlated with the ticket sales of the following boxoffice than the movie rating, and the search volume is not directly correlated with the weekend's ticket sales. High value of competition can make the ticket sales of the following weekend go down. The negative correlation value between competition and ticket sales supports this fact.



Factor	Correlation coefficient
Week number	-0.4027
Ticket sales of the previous weekend	0.7926
Number of screens for weekdays	0.8206
Seat share of the previous weekend	0.6441
Seats per screen for weekdays	0.7101



Factor	Correlation coefficient
Star rating	0.2156
Comment volume of previous weekend	0.6609
Search volume	0.4776
Competition	-0.5994

TABLE II: Correlation coefficient between the external factors and ticket sales of each weekend.

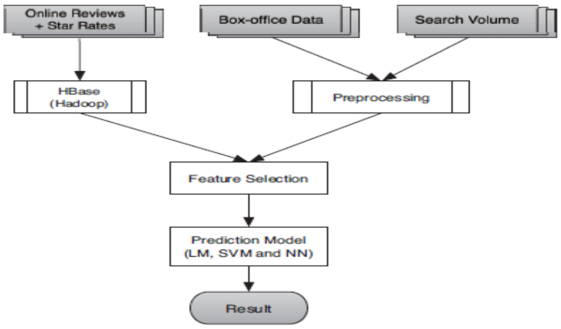


FIGURE 1: PROPOSED ARCHITECTURE

IV. Conclusion and Future Scope

In this paper, our goal is to forecast the sales performance of the near future using box-office data and external factor data such as online reviews, movie ratings and search volume. Especially, we focused on the movie ticket sales which WOM effect is prominent. Our prediction model forecasts the ticket sales of each weekend and can determine the ideal screen number based on our predicted ticket sales. We gather 1.7 million online reviews and movie ratings from South

Korea's biggest movie site, Naver movie service. We also gather search volume from Google Trends and Naver Trends. We use these data as external factors, which affects the ticket.

REFERENCES

[1] W. Duan, B. Gu, and A. B. Whinston, "Do online reviews matter ?An empirical investigation of panel data," vol. 45, pp. 1007–1016, 2008.

[2] S. Goel, J. M. Hofman, S. Lahaie, D. M. Pennock, and D. J. Watts, "What Can Search Predict ?"*WWW '10*, 2010.

[3] X. Yu, Y. Liu, X. Huang, and A. An, "Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 4, pp. 720–734, Apr. 2012. [Online]. Available: http: //ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5677530

[4] G. Kulkarni, P. K. Kannan, and W. Moe, "Using online search data to forecast new product sales," *Decision Support Systems*, vol. 52, no. 3, pp. 604–611, 2012.[Online]. Available: http: //dx.doi.org/10.1016/j.dss.2011.10.017

[5] L. Zhang, J. Luo, and S. Yang, "Forecasting box office revenue of movies with BP neural network," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6580–6587, Apr. 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S095741740800496X

[6] K. J. Lee and W. Chang, "Bayesian belief network for box-office performance: A case study on Korean movies," *Expert Systems withApplications*, vol. 36, no. 1, pp. 280–291, Jan. 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0957417407004228

[7] MPAA. 2012 theatrical statistics summary.[Online]. Available: http:// www.mpaa.org/resources/3037b7a4-58a2-4109-8012-58fca3abdf1b.pdf

[8] R. Sharda and D. Delen, "Predicting box-office success of motion pictures with neural networks," *Expert Systems with Applications*, vol. 30, no. 2, pp. 243–254, Feb. 2006. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0957417405001399

[9] E. V. Karniouchina, "Impact of star and movie buzz on motion picture distribution and box office revenue," *International Journal of Researchin Marketing*, vol. 28, no. 1, pp. 62–74, Mar. 2011. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0167811610000881

[10] H. Rui, Y. Liu, and A. Whinston, "Whose and what chatter matters? The effect of tweets on movie sales," *Decision SupportSystems*, vol. 55, no. 4, pp. 863–870, Nov. 2013.[Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0167923612003880

[11] J. Du, H. Xu, and X. Huang, "Box office prediction based on microblog," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1680–1689, Mar. 2014. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0957417413006866

[12] S. Moon, P. K. Bergey, and D. Iacobucci, "Dynamic effects among movie ratings, movie revenues, and viewer satisfaction," *Journal ofMarketing*, vol. 74, no. 1, pp. 108–121, 2010.

[13] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," *Advances in neural informationprocessing systems*, vol. 9, pp. 155–161, 1997.

[14] I. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *Journal of microbiological methods*, vol. 43, no. 1, pp. 3–31, 2000.