An Empirical Investigation on Movie Industry from 1980 to 2018

Choujun Zhan[1,2], and Jianjin Li[1], and Wei Jiang[1]
[1] Nanfang College of Sun Yat-Sen University, Guangdong 510970, China,
[2]School of Computer, South China Normal University, China
Email:zchoujun2@gmail.com and issyz@mail.sysu.edu.cn

***Abstract*—One of the critical problems in the emerging field of computational social science is how to achieve socially generated "big data" to build a "cleaned" dataset for accessing information about collective human behavior and. The movie is an essential cultural product and watching a movie is one of the most popular entertainment ways in our daily life. A favorite movie can attract the millions audience. Additionally, after years of development, the film industry has produced a large amount of data that can be supplied to the research about collective behavior propagation. Hence, movie industry is suitable candidate for investigating collective behavior propagation in social network. In this work, we construct a dataset including almost all the movies released in the United States from 1980 to 2017. By analyzing the US box office data, we found that "action", "documentaries" and "drama" are the most favorite movie genres in the past 38 years. We also find that the final total global gross of a movie is stronger related to its best weekly rank, namely, the global gross of the movie is almost proportional to its best weekly rank. Furthermore, we find that although the total box office is rising every year. However, with the birth of the Internet, fewer and fewer people are willing to go to the cinema to watch a movie. Our analysis can help understand how a movie become popular and help movie distributor to public audiences' taste and safety develop a plan for providing movie products.**

I. Introduction

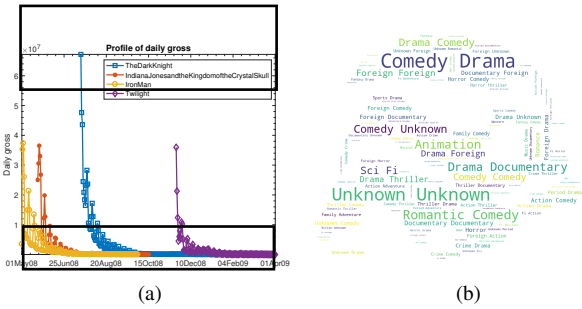
The movie is an important part of modern art and merchandise [1]. Watching a movie is still one of the favorite leisure activity for many individuals. In the United State, on average, each person watches five to eight movies per year. In 2017, A total of 1.2 billion movie tickets were sold, and 740 movies were released in the US. The annual box office of the United States reached 11 billion US dollars. It is estimated that the total US box office will reach 15 billion US dollars in 2020. The movie industry is a huge market in the U.S and the world. From 1980 to 2018, about 13,000 movies are released. Only a tiny fraction of movies can succeed and win huge box office. For example, less than 1,000 movies have a box office larger than 10 million, which means that the challenges of attracting the audience's heart in the modern world are highly fierce. Consequently, the mechanism that operates in society to determine the success of a movie is a topic of considerable importance in the movie industry and business [2]. This begs the question of how a movie becomes popular, which has been of much interest for researchers for years. Specifically, it is of significant interest to understand the factors that contribute to and determine the success of the box office growth of a product [3], [4]. Some of the factors were investigated in previous work [5], including word of mouth [6], online reviews [7], rates [8]. Some researchers focus on the prediction of box office [9], [10]. However, which movie will become popular and receive a huge box office is still unknown.

Here, we construct a new dataset including almost all the movies released by American distributors, such as Warner Bros, Universal, Buena Vista, Fox, etc. This dataset contains detailed information on 13,373 movies, which includes "movie title", "daily gross" or "weekly gross", "rank", "Budget", "Theater", "gross oversea" etc. In this work, based on this new database, we investigate the movie market with a big data approach to study the sales pattern and box office in American and the whole world. We quantified the growth of box office patterns, explore the daily gross and weekly gross, allowing us to uncover the demographic patterns of how a movie becomes successful. This empirical investigation on movie products focuses on the integration and use of knowledge about statistical methods, which may help the distributor to make a more safety plan for providing a movie product.

More and more movies are released every year, which makes the competition in the movie industry more and more fierce. The vast majority of movies have a total box office between 100,000 to a million(more than 80% of the total), and only a small number of movies can reach more than 10 million(less than 10% of the total). Most of the movies are only released less than ten weeks, and only a few hottest movies are leased more than 30 weeks. We find that the movie industry is monopolized by top 10 distributors, who released about 80% of all the movies and the fraction is growing year by year. We find that "action" movie is the most popular movie genre after 2010 and the second most popular movie genre includes Animation Comedy and Drama. Results show that the box office of action movie accounts for only 1% of the global total box office in the year 1980, while it is continuously growing over the last 30 years and reached about 38% of the world's total box in 2017. We also analyze the relationship between the total gross and released weeks of a movie. Results show that the total gross of a movie is highly correlated with the best weekly rank and the length of released time. The higher of the weekly best rank and the longer the released time, the bigger the box office of a movie. Finally, although the annual box office grows every year, we find that the number of people watching movie decrease after the year 2007, dropping to about 3/4 as 2007 in 2017. Now let's decrypt the reason

for the success of a movie.

II. Data description

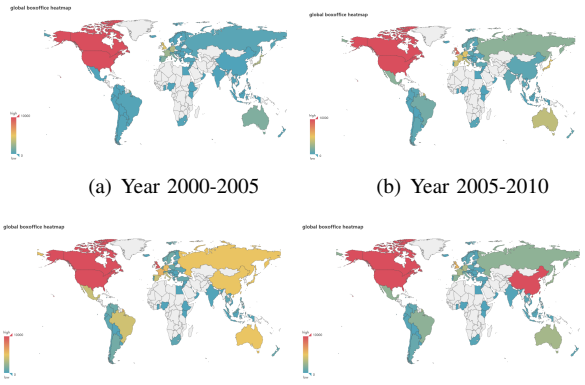


(a) (b)

Fig. 1. (a)Profiles of daily gross of four favorite movies: Batman, The Dark Knight, Indiana Jones and the Kingdom of the Crystal Skull; (b) Wordle of all the movie genres.(a)Profile of daily gross of four favorite movies: Batman, The Dark Knight, Indiana Jones and the Kingdom of the Crystal Skull; (b) Wordle of all the movie genres.

Box Office Mojo, founded by Brandon Gray in August 1998 and now belonged to Amazon, is one of the most famous and authoritative box office statistics sites in the United States for systemic computing of box office. We utilized web-crawler, circumvent the network anti-crawling mechanism, to gather data from Box Office Mojo. Spending months on data crawling and cleaning, we finally constructed a movie box office database containing information of 13,373 movies released from 1980 to 2018. This database contains a fruitful movie properties, including "title", "distributor", "released date", "MPAA rating", "runtime", "daily gross", "weekly gross", "rank", and etc. Figure 1 (a) shows daily gross profiles of four movies released in 2008, including "Batman 2: The Dark Knight", "Indiana Jones and the Kingdom of the Crystal Skull", "Iron Man", and "Twilight". There are various movie genres and Figure 1 (b) shows the wordle of all movie genres, such as action, Sci-fi, comedy, fantasy, etc. Furthermore, this database also includes the weekend and weekly gross of the USA movies released in 87 countries all over the world, such as France, UK, Japan, China, etc. In 2017, a total of 740 movies are released in the United States. The annual box office reaches a new record with a total $110.7 billion annual box office. Next, let's take an overlook of the global box office in different periods.

Thanks for a peaceful world, the movie industry has continuous growth in the last half-century. Based on the global box office database, we constructed a global box office heat map(Figure 2), which represents the distribution of box office in different countries in different periods. The red area represents a higher box office. Although the American movie industry still holds a dominant position, the movie industry in other countries become larger and larger. The average growth of the world movie industry is about 4% a year, and the movie industry in American has posted crazy growth period and become saturated, while China is the most fast growing area with an average growth rate of around 30 %. Hence, Hollywood place a great emphasis on the vast market



(a) Year 2000-2005 (b) Year 2005-2010

(c) Year 2010-2015 (d) Year 2015-2018

Fig. 2. Global box office heatmap in different periods.

potential of emerging movie markets in China. Figure 2 shows the cumulative gross of different countries in four different periods. Note that China market grows fast over the past 20 years (Figure 2 (d)). From Figure 2), we can see that North America remains the most significant film market, but China movie industry is growing dramatically [11], with box office reaching $7.9 billion in 2017, four times than box office in Japan. In the top five box office regions, Asian countries occupy four seats. The growth of the global box office in 2017 is almost driven by China alone. The Chinese film market grew by about $2.7 billion in 2017, surpassing the total box office growth over the world.

III. Statistical analysis of movie industry

Next, let study the movie industry based on our database.

A. The movie industry is monopolized by top 10 publisher

Figure 3 (a) shows the number of publishers and their distribution over the past 38 years. One can see, most of the distributors only distribute several movies (less than 5 movies). Only a few distributors distribute plenty of movies (more than 100 movies). Figure 3 (b) shows the top ten publishers in the past 38 years. We can see that the number of movies released by the top ten publishers, such as Warner Bros., all larger than 300 movies. Figure 3 shows the percentage of the number of movies released by the top 10 distributors over the whole released movies each year. We can see that the percentage continuously grew from 1980 to 2000 and now reaches about 80%. Hence, the movie industry is monopolized by the top 10 distributors. Now, let's try to find out the most popular movie genre from our database.

B. Action movie is the most popular movie genre

Recently, one media website (Fandango.com) launched voting about the most anticipated movie in 2018. About 30% of 8000 netizens select action movies, and 10% choose sci-fi movies. The voting result supports that action movie and sci-fi movie are the most popular movies. Next, we will find out whether the voting result also supported by the actual movie box office. Figure 4 (a) shows the total annual box office of

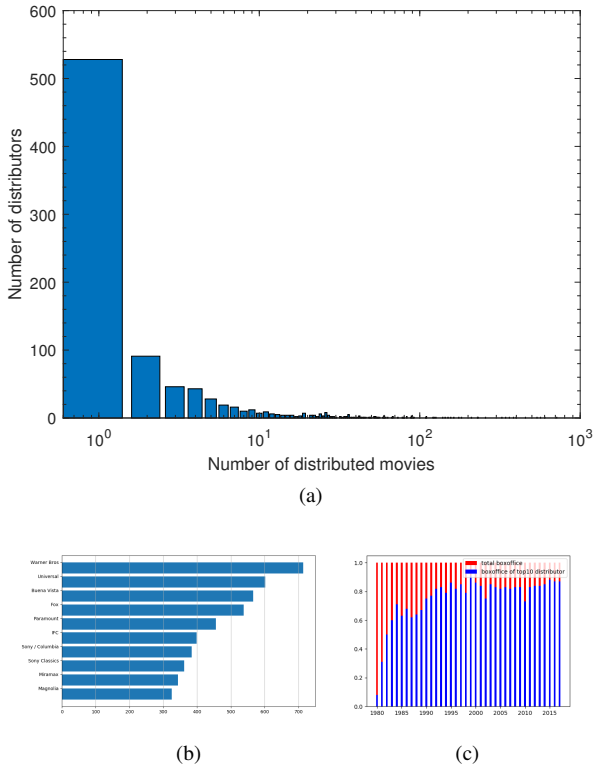


Fig. 3. (a) The probability of distribution of the released number of movies for each distributor; (b) top 10 largest distributor; (c) the partition of movies released by the top 10 distributors.

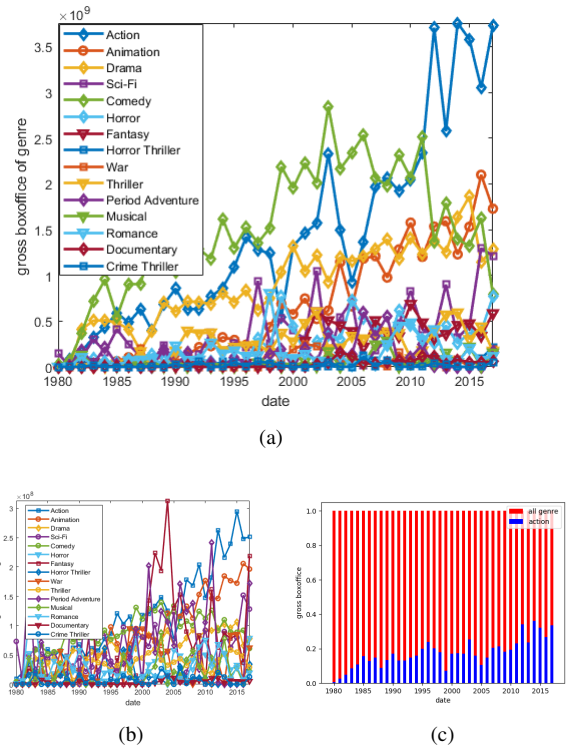


Fig. 4. (a) The annual cumulative box office of different movie genre from 1980 to 2017; (b) annual average box different movie genre from 1980 to 2017; (c) the fraction of box office of action movies.

different the movie genre. The result shows that the annual box office of action movie is far larger than others over the last five years, while the box office of action movie is almost two times of the second large movie genres (animation, drama, and comedy). Hence, the most favorite movie genre has changed from comedy to action movie. In the past In 2017, the box office of Action Movies reached an astonishing $4 billion (about 40% of the total US box office, shown in Figure 4 (c)). Before 2010, the gaps between the annual cumulative and average box office of the action movie and other movie genres are not very lager. However, after 2010, action movies become to dominate the movie industry. From Figure 4, we can easily find out that before 2010, comedy movie is the most favorite movie genre, while after 2010, action movie began to overtake the comedy movie and maintain the leading position until now. From this we can see the taste of the audience from funny comedy movies to stunt cool, fighting scenes thrilling action movies. We can claim a conclusion: action movies dominate the American movie industry.

C. Analysis of total gross and released weeks

Figure 5 (a) shows the box office distribution map of the United States over the past 38 years. It can be seen that the distribution of box office in the past 38 years is a double-normally distributed, and most of them are concentrated between 10,000 and 10 million. It can be seen from Figure 5 (b) that the majority of movies are very short-lived in the cinema,

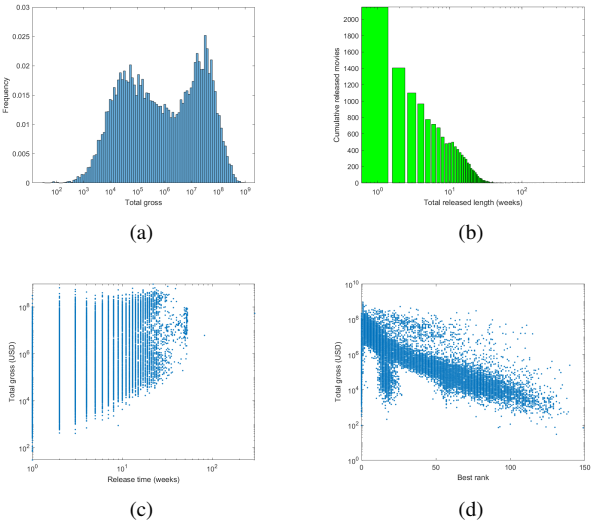
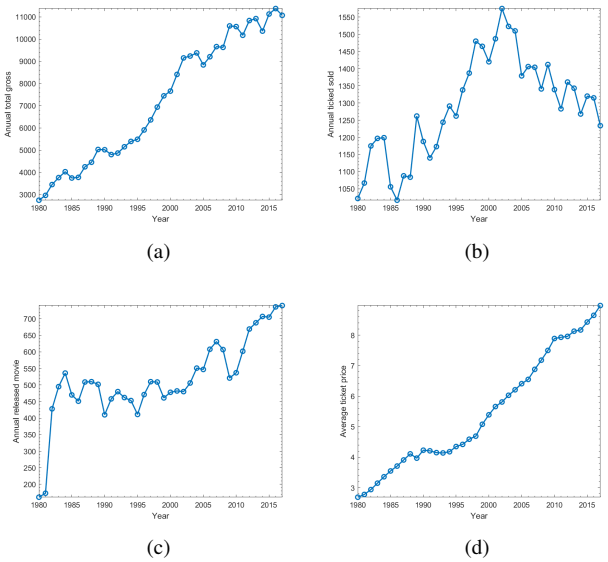


Fig. 5. Statistical results of total gross and released time of a movie: (a) the distribution of total movie gross; (b) the distribution of the released time of a movie; (c) the relationship between the total released time and the total gross; (d) the relationship between the best rank and the total gross.

most of them are within 10 weeks. And the distribution of released time follows a trend of the power law. There are also a small number of movies that can be released in cinemas for a long time. Figure 5 (c) illustrates that there is a positive correlation between the released time of the movie and its total

gross. The longer the released time, the bigger the total gross. We also analyze the relationship between the best weekly rank and the final global gross of a movie (shown in Figure 5 (d)). Results show that the higher the best weekly rank, the high probability that a movie can achieve success [12], [13].

D. The crisis of the movie box office



(a) (b) (c) (d)

Fig. 6. Movie genre in the USA from 1980-2018. (a) top ten genres (b) the dynamic growth profile of the top ten genres.

In 2017, a total of 740 movies are released in the U.S., and the box office of the U.S. reached 11 billion US dollar. A total of 1.2 billion movie tickets were sold. However, the annual sales of tickets are 22% lower than in 1995, which is the lowest in the 22 years since 1995. The per capita viewing rate in North America has fallen by 14 percent since 2007, dropping to less than four movie tickets per person per year (shown in Figure 6). The grow of total movie box office is based on the increasing ticket fare. However, even the number of movies is rising every year, this does not change the fact that the film has not attracted the audience 10 years ago. The market seems to be stalled and even new technologies can not drive the demand for viewing, and the downturn in the North American film market will continue.

IV. Conclusion

In this paper, one main contribution is that we construct a new database containing about 13,000 movies released in American and the other area in the world from 1980 to 2018. Then, based on big data methods, we investigate the whole movie industry by studying the movie release time, the total gross, the best rank, the movie genre, etc. We find action movie is the most popular movie genre after 2010. Moreover, the total gross of a movie is highly related to the released time and best weekly rank. However, in this work, we just consider one dataset including the time series information of the box office but without some text information, such as the Movie Synopsis, the scores in IMDB, etc. Additionally, this dataset only contains movie released in the US. In the future work, we plan to extend the dataset about the movie industry by considering more text information, pictures, and movies from other countries. Based on this new big and reliable dataset, we can future develop a model for the prediction of the box office of the movie and even the popular trend of movie genre based on state-of-art machine learning algorithm, such as deep learning, ensemble learning [14], [15]. Additionally, we can develop a prediction system for movie industry help distributor to make a plan for releasing movie product.

Acknowledgment

This work was supported by National Science Foundation of China (61703355) and Guangdong Youth University Innovative Talents Project(2016KQNCX223) and Guangdong Province Quality Project Grant ZL2013025.

References

[1] A. De Vany, "The movies," *Handbook of the Economics of Art and Culture*, vol. 1, pp. 615–665, 2006.
[2] R. K. Pan and S. Sinha, "The statistical laws of popularity: universal properties of the box-office dynamics of motion pictures," *New Journal of Physics*, vol. 12, no. 11, p. 115004, 2010.
[3] C. Zhan and C. K. Tse, "A universal model for growth of user population of products and services," *Network Science*, vol. 4, no. 4, pp. 491–507, 2016.
[4] C. Zhan and C. K. Tse, "A model for growth of markets of products or services having hierarchical dependence," *IEEE Transactions on Network Science and Engineering*, 2018.
[5] A. De Vany and C. Lee, "Quality signals in information cascades and the dynamics of the distribution of motion picture box office revenues," *Journal of Economic Dynamics and Control*, vol. 25, no. 3-4, pp. 593–614, 2001.
[6] Y. Liu, "Word of mouth for movies: Its dynamics and impact on box office revenue," *Journal of marketing*, vol. 70, no. 3, pp. 74–89, 2006.
[7] P. Boatwright, S. Basuroy, and W. Kamakura, "Reviewing the reviewers: The impact of individual film critics on box office performance," *Quantitative Marketing and Economics*, vol. 5, no. 4, pp. 401–425, 2007.
[8] M. Mestyán, T. Yasseri, and J. Kertész, "Early prediction of movie box office success based on wikipedia activity big data," *PloS one*, vol. 8, no. 8, p. e71226, 2013.
[9] K. J. Lee and W. Chang, "Bayesian belief network for box-office performance: A case study on korean movies," *Expert Systems with Applications*, vol. 36, no. 1, pp. 280–291, 2009.
[10] 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, 2006.
[11] Z. Luo, H. Liu, Y. Li, H. Wang, and L. Zhang, "Robust hybrid transceiver design for af relaying in millimeter wave systems under imperfect csi," *IEEE Access*, 2018.
[12] Q. Li, Z. Wu, and X. Xia, "Estimate and characterize pv power at demand-side hybrid system," *Applied Energy*, vol. 218, pp. 66–77, 2018.
[13] Z. Wu, H. Tazvinga, and X. Xia, "Demand side management of photovoltaic-battery hybrid system," *Applied Energy*, vol. 148, pp. 294–304, 2015.
[14] M. Zhao, T. W. Chow, Z. Zhang, and B. Li, "Automatic image annotation via compact graph based semi-supervised learning," *Knowledge-Based Systems*, vol. 76, pp. 148–165, 2015.
[15] M. Zhao, T. W. Chow, Z. Wu, Z. Zhang, and B. Li, "Learning from normalized local and global discriminative information for semi-supervised regression and dimensionality reduction," *Information Sciences*, vol. 324, pp. 286–309, 2015.