

The Big Data Applications in Film Industry Chain

Xinran Wang, Yan Wang, Jianping Chai, Xi Feng and Ziyu Liu

Communication University of China, Information and Engineering College
1149779627@qq.com, 31951253@qq.com, fengxi@cuc.edu.cn

Abstract

Nowadays, the audiences' consumption attitudes, consumption patterns and consumer groups are all in the great changes, thus it is necessary to improve the film's revenue by excellent script selecting, accurate market positioning, effective product marketing, and accurate forecasting of the box office. This paper introduced the application and benefit of big data in the film industry chain in terms of film making and investing, film publicity and distribution, film broadcasting and film audience, pointed out many challenges that big data encountered in China's film industry and finally provided useful suggestions for the practitioners in the film industry of all aspects.

Keywords: Big data, Movie, Film publicity and distribution, Film marketing, Film making and investing

1. Big Data in Film Industry

In China, film industry began the institutional reform since 1980s [1], with the rising number of film production institutions and cinemas year by year. However, it has been complicated to predict the return rate of the industry, even a film with a famous director was unable to ensure its revenue. *Heaven's Gate*, for instance, directed by Michael Cimino, a well-known director, was one of the most lose-money movies ever made. *Wind Talkers*, directed by a noted Chinese director, Wu Yusen, led to the bankrupt of Metro-Goldwyn-Mayer (MGM) [2]. Yet TV series *House of Cards* built on big data by Netflix subverted the situation, and led to the use of Big Data impressed by global film and television industry, which showed that using big data in the film and television industry was necessary in the Information Age.

Big data provides strong and powerful quantitative basis and decision reference for the film industry in the information era, which can be widely used in the film industry and bring limitless possibilities to its vigorous development. In June 2013, Google published the white paper *Quantifying Movie Magic with Google Search* [3]. The report chosen hundreds of films released in 2012 in the United States, and used three types of data index to construct a film box office prediction model, including the search volume of movie trailers, box office performance of the same series that released before and the seasonal characteristics of the film schedule. Besides, each type of index was subdivided into a number of sub-indices. This proposed model enabled to predict a movie's box office a month before the movie is screened to the public, with an accuracy of 94%, which indicated a strong correlation between the film-related search volume and the box office. Accordingly, several practical operation recommendations were given to optimize film marketing strategies by searching data.

In the Chinese market, "Idatage", a film analysis platform based on big data, made an attempt to study audience by a combination way of big data and questionnaire survey. There were a total sample of 150 thousand moviegoers from the first-tier, second-tier and third-tier cities across the country considered as a fixed sample to be tracked and monitored [4]. Audience viewing characteristics and preferences can be learned accurately and thoroughly by cross-analysis and synthetic treatment. Therefore, it became

possible to provide more personalized content, customized marketing and entertained propagation to different users. Entertainment Treasure, launched by Alibaba in March 2014, invested popular film and television by crowd-funding from fans. Till June 2015, Entertainment Treasure had raised a total financing of 560 million, investing nearly 20 films. *Fan economy*, turns out to be a key to the box office of film. By satisfying audience's tastes, film based on fans does obtain a considerable return rate with little investment. *Tiny Times 3*, for example, gained a box office of 521 million with 34.54 million investments in Entertainment Treasure.

A film industry chain includes the procedures of film investing and making, film publicity and distribution, film broadcasting and film audience, reflecting the economic ecosystem among companies in this industry [6]. As shown in Figure 1, big data plays an essential role in each link of the chain. Not only assisting producers in choosing scripts and investing the film shooting, big data also offers suggestions for film precise marketing, which favors cinema circuits picking and arranging their films. General preference of film audience becomes able to be detected in this way, so that the box office of film can be improved effectively.

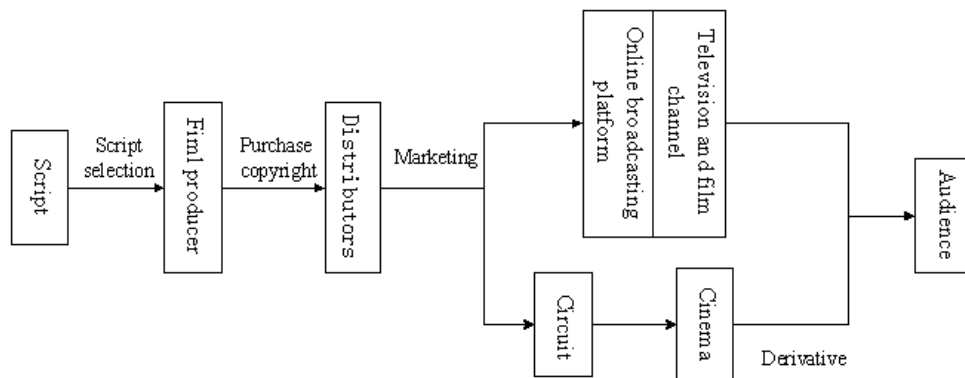


Figure 1. The Film Industry Chain

2. Big Data in Film Investing and Making

Film investing and making mainly accomplishes the process of script selection and adaption as well as investment cost and film shooting. Big data analysis plays a positive role in film investing and making, just like script selection, actors choosing, film shooting and the post producing, and providing an effective decision-making method. Based on target audiences' characteristics such as gender, age and regional distribution, the analyze finds out whether audience are interested in the film content and main performers and digs out their expected director and favored actors can be conducted, which provides accurate, scientific and quantitative recommendations for the decision-making of film producers and investment companies.

Since 2014, Intellectual Property, as a popular word, crashed the Chinese media hub. It is highly valued due to its primary market consisting of cumulative readers and fans. As to Intellectual Property scripts' choosing, big data does make a remarkable effect. The six most famous film companies in the United States are indeed half of data companies [8], with long preparation duration and conserved development process, and then make adjustments by surveying audience's preferences at times. All of these IP choosing and exploiting call for the support of big data. By mining and analyzing the *correlations* [9] between seven major user data (user basic characteristic data, user social generated data, user online interactive data, user consumption record data, user web public data, user offline behavior data and user location data) and six main film content elements (story, image, director, performances, sound and editing), modular and regular information can

be refined to feed back the secondary production of film and standardized production system, achieving prediction function in the field of business applications, which is the basic idea of movie content production in the big data era.

A lot number of films and TV series based on big data attained high box office in recent years, the most notable of which was TV series “House of Cards”, released in 2013 and produced by Netflix. Due to its success, the year was called *the first year of big data* [10]. Netflix, as the world’s largest online film renting service providers, mastered quantities of customers’ registration information and film order data. By analyzing and calculating of the database containing 30 million users' choice data, 4 million comments as well as 3 million times film theme search data, the films’ shooting contents, directors, actors and even play ways can all be decided by statistical preferences of tens of millions audiences [11]. Precise, meticulous, efficient and economic data guide producers from sighting, positioning and analyzing of the audiences to ultimately transforming into film works, so as to realize the content production decided by user demand.

Little Apple, as one of the Chinese films, gained good results as well by analyzing the viewing data of micro film *Old boys* played on Youku in 2010. The data was consist of the genders, ages and regional composition of fans, the viewing, commenting and reposting numbers of the micro film, and even when the audience paused, which specific episodes or plots viewers favored and which region’s users preferred reposting [12]. Liming Ke revealed that 75% of the information can be transformed into the film production [13]. Relying on the successful direction of big data, the trailers gained more than 100 million hits on the video website before the film released.

Till now, although some studies on modeling analysis of film elements data have been conducted, such as the “The Deep Film Access Project: Ontology and metadata design for digital film production assets” [14] studied by Atkinson S, the film production research based on big data haven’t formed a mature theoretical literature.

3. Big Data in Film Publicity and Distribution

Film distribution companies buy film copyrights from the production companies to bring them to screen in cinemas under its cooperated circuits. Distribution companies need more support from big data in the film marketing process. Big data from social media helps learn more clearly and completely about film audience’s composition, preferences, motivations and consumption characteristics, based on which precision marketing becomes possible.

In terms of film marketing, big data makes solid returns for fans economy and the development and operation of film derivatives. Letv, for instance, the company’s valuation reached 4.8 billion in September 2014, which had risen more than three times in a year. It’s the operations of films as *Tiny Times* series and *Old Boy* that performed the significant factor [15]. *The Avengers*, as another successful case, conducted a cross-analysis after releasing its trailer between user comments collected from social media and user characteristics including gender, occupation, frequency of going to cinemas, preference to comic books, etc. [16]. This kind of analysis did a favor to developing new consumer groups and understanding consumer's perceptions of the brand and product, which ensured a clear budget and the advertising creativity so as to achieve the goal of precision marketing before the movie released. With the help of a good publicity, the first-week sales of *The Avengers* reached 200 million dollars in the United States market after being released in May 2012. By September 2012, its total office box were more than 620 million dollars, ranked third in the world film box offices, second only to *Avatar* and *Titanic* [17].

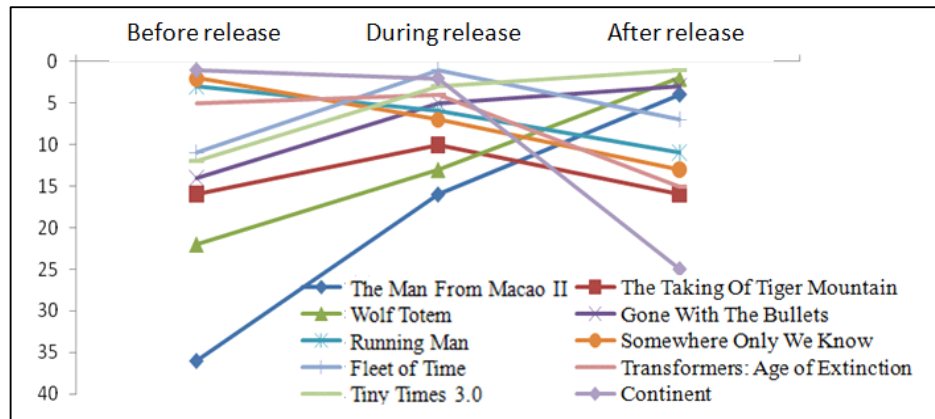


Figure 2. Comparison of Marketing Ability of Films in the Same Schedule

Diverse marketing strategies have a directly influence on films' box office performances. This paper used the entropy method and factor analysis method to compare and analyze 10 films released in 2014 with the same schedule on the marketing materials data, the quantity of user comments, main actors' web-influence, etc., from three aspects of before-releasing, on-showing and off-cinema, to reflect each film's marketing situation. Figure 2 showed the contrast marketing effect. It can be learned that 4 out of 10 films (*The Man From Macao II*, *Wolf Totem*, *Gone With The Bullets* and *Tiny Times 3.0*) had a rising marketing index, saying the film's marketing effect in the off-cinema period was better than that in the before-releasing period; another 3 films (*The Taking Of Tiger Mountain*, *Fleet of Time* and *Transformers: Age of Extinction*) performed the best in the on-showing period. Additionally, *The Man from Macao II* and *Wolf Totem* had similar marketing characteristics: low marketing index in the before-releasing period yet accomplished great progress in the latter two periods, speaking highly by media and customers, and eventually received a relatively satisfied box office performance, which made them be considered as stronger films. *Fleet of Time*, however, showed a distinguished marketing idea of an outstanding marketing performance in the on-showing period. During its shooting process, cumulative materials supported the implement of its marketing plan, together with a fine marketing situation in the before-releasing period, which resulted in its good grade in box office.

Under the directing of big data, film distribution can be more standardized, market positioning can be more precise, marketing can be more effective, derivatives can gain more market, and thus the efficiency of industry investment as well as the return rate can be improved, as a result, a more completed film industrial chain can be built.

4. Big Data in Film Broadcasting

Cinema circuit arranges the film schedule uniformity towards cinemas under its umbrella. Usually large circuit arranges more screening times for blockbusters while some low-cost films have to manage to obtain a higher row piece volume. With big data, the prediction of film box office becomes feasibly, which helps cinema circuits in choosing films for higher profits. Besides, big data can also be used in estimating the number of audience along with the viewing time and preferences of audience in order to arrange each film's row piece volume more reasonably.

Lots of researches on film box office prediction based on big data have been carried out. Jordi analyzed the feasibility of box office forecast from various data analysis [18]. Delen etc. constructed an estimation model of film box office based on the type of film, star, technical effect, release time and other characteristic parameters to establish a Web decision support system [19]; Ghiassi *et al.*, set up a dynamic artificial neural network prediction model according to the advertising budget, release time and seasonal

forecast variables [20]; Zheng Jian put forward a kind of the movie box office forecasting model based on feedback neural network [21], whose output can not only be more reliable to reflect the movie box office income during the on-showing period, but also point out the range of that; Jing Fei *et al.*, used machine learning methods to make statistical analysis by the information of Microblog, proposing a box office estimating method based on Microblog [22]. By these kinds of prediction and analysis methods based on big data, it is uncomplicated for cinema circuits to make more accurate choice of films.

This paper predicted the film box office assuming the mean of douban fans of the film's director and the former two actors had a similar trend to the sales trend. Using 67 films that were same type of "Movie Master" in year 2013-2014, the study adapts the release schedule, production countries, the mean of douban fans of the director and the former two actors and that of the director and the former five actors as affecting indicators, then uses machine learning to model the data in the database[23], and finally makes prediction and representation[24], with a high fitness of 97%. The estimated box office result can be seen in Table 1, showing that its sales estimation were 10 million yuan.

Table 1. Estimated Box Office of Film *Movie Master*

Schedule Setting			Predicted degree	Degree range		Predicted probability
Others	(avoiding the competition schedule)	the	G	10 million	-20	0.950
Summer			I	1 million	-5	0.938
Lunar New Year			I	1 million	-5	0.916

The films role piece plan recommended by big data helps in improving film box office. "Row Piece Treasure", a big data application software launched in June 2015, relying on the cinema's attendance history data, network data and survey data, can automatically calculated out the popularity of each film in the cinema even the movie preferences and behaviors of the audience to accomplish a precise and reasonable film row piece arrangement. In Shanghai, a well-known cinema's manager said, "after reasonable film row piece arrangement, the cinema's overall attendance rose more than 15%" [23]. Before this, TAOBAO website also launched the "Virtual Circuit Alliance". Fan Chi introduced that Virtual Circuit Alliance would arrange the film's row piece volume based on big data, for instance, the presale ranking list. "Besides the list, we will launch a row piece index to describe how many audience come to the theater every day, in what period of time they will come and which kind of movie they may choose, using big data to guide the theater row piece" [24]. *Switch*, *Badges of Fury* and *Tiny Times* released in June 2013, were ridiculed as "bad film trilogy" [25]. Yet these kind of "special" films has still achieved good box office, it was exactly the high row piece rate support those films' box office success.

Thus, big data does provide reasonable suggestions for cinema circuits to choose and arrange films, in order to make better profits from box office.

5. The Impact of Big Data on Movie Audiences

The audience is the final link of the film industry chain and also the key point, because the audience's viewing selection directly affects the film box office. The Internet produces a lot of users' film comments which play an important role on the audience's viewing choice. Fans, meanwhile, covered by certain film, are willing to pay for their favorite stars

without considering the quality of the film itself, which will also make a great contribution to the film's box office.

A film can attract viewers through publicity before its screening, but once be released, it depends on word of mouth (WOM) whether audience will choose the film while the propagation of information is rapid. Chakravarty and others studied the influence of internet users' reviews and the professional film critics' evaluations on film [26]. A large number of network users' comments effect the audiences' viewing options, and that's why WOM becomes a key factor into the box office. The WOM description of a film can reflect on the box office quickly, for example, *Monster Hunt*, *Pancake Man* and *Monkey King: Hero is Back*, their box office continued rising due to high reputations, which formed a long tail. Hence in the big data era, film box office ultimately depend on the quality, that is to say, content is king.

Fans, this particular target groups, can often push the movie box office with the development of fans film. Before released, *Tiny Times* series movies conducted an analysis based on survey of viewing data online, finding out that the main target consumption group of *Tiny Times* were made up of 40% high school students, 20% college students and 30% white-collars [27]. These groups have more obvious consumption preferences for brand and its products such as Adidas, Starbucks, Jumei, IKEA and durex. Therefore, the business implant or integrated marketing of *Tiny Times* made a close cooperation with these brands, which successfully explored the consumption demand and potential of these audiences as well as expanded the derivative value of the film. Some fans don't concern about the movie itself, they merely paid for their favorite stars. Although when *Tiny Times* just be released, its ideological content and artistic value were controversial, the precise command of meeting the audience need helped win the high box office.

As is seen, analysis of the film audience overall preferences using big data performs well in predicting the target audience and the performance of film box office, which provides the audience with the film related information services as well.

6. The Challenges film industry met in Big Data Era

Big data has arisen as a hot topic in the "Big Box Office Era" of China, which has a vital value in each link of the film industry chain, bringing limitless possibilities for the vigorous growth of the industry. Varieties of film and TV dramas works has successfully raised the return rate on the film using big data both in domestic and abroad market, creating a box office miracle one after another. Yet in China, it remains a long distance from applying the "big data" in real to support the film industry innovation comprehensively, while still facing with many difficulties.

Primarily, it can be difficult to get access to the data. The Chinese film industry does not have an industry tradition to establish a data sharing platform. Big data of domestic film, at present, are from dispersal sources, while data mining needs to attract cross-border cooperation, which can be a complicated work to integrate data from cinemas, search engines, video websites and social networking platforms and other multi-channels. Subsequently, the ability of processing big data is still weak. Nowadays, more than 90% of the Chinese film research firm tried to imitate the Google's box office prediction model by conducting a quantitative analysis based on the study of the "Baidu index" model, taking the theme of the film, creating team and other indicators into consideration. Yet the depth of the analysis and the overall data accuracy was not Satisfactory [28]. Furthermore, film is more than a technology, but an art. The analysis of big data may help film practitioners have deeper insights into the audience's thoughts to lead to a high film box office, but it can never replace the artist's creativity, intelligence and artistic in the film.

Over all, in the age of big data, both opportunities and challenges coexist in film industry, and there is still a long way to go for the application research on big data in film industry.

Acknowledgements

This paper is financially supported by Engineering Planning Project of Communication University of China (XNG1356), Engineering Planning Project of Communication University of China (XNG1412) and Outstanding Young Teacher Training Project of Communication University of China (YXJS201527).

References

- [1] Z. Jianming, "On the Marketing Process and Genre Analyses of China Commercial Film", Ji Lin University, (2013).
- [2] X. Kun, "How far is big data from films", New Business Week, vol. 13, (2013), pp. 84-85.
- [3] "Quantifying Movie Magic with Google Search. Google Whitepaper", <http://www.google.com/think/research-studies/quantifying-movie-magic.html>, (2013).
- [4] "'Montage' and 'Idatage' ---Films' big data era", Communications, vol. 17, (2014), pp. 4.
- [5] G. Jupeng, "Entertainment Treasure and TAOBAO Film Released Movie based on Big Data", <http://biz.zjol.com.cn/system/2015/06/15/020697173.shtml>, (2015).
- [6] "Film Industry Chain. Baidubaike", <http://baike.baidu.com/view/3051573.htm>.
- [7] "How to Do with the Film IP Under the Trend of Big Data", CDA Data Analyst. <http://cda.pinggu.org/view/14889.html>, (2015).
- [8] "American theatrical industry study team, Deng Lin He Jiajie, Ma Ping, Zhang Hui, Yang Li, Shen Shushu, Ge Van, Liu Jingyu, Tian Hui, Weng Chenxi, Lela Wang, Psoriasis, Qin Qian. 2013 American Movies in Mainland China Market Statistics Report", Commentary on Cultural Industry in China, vol. 2, (2014), pp. 265-305.
- [9] L. Tao, "The Thinking of Big Data and the Movie Creation of Content", Contemporary Cinema, vol. 6, (2014), pp. 9-14.
- [10] "The Year 2013 Was Named "The First Year of Big Data Era", Youth Journalist, vol. 3, (2013), pp. 15.
- [11] S. Hua, "Big Data Marketing Bring Thoughts Into our Domestic Network Homemade Drama in "House of Cards" Case", Youth Journalist, vol. 26, (2014), pp. 100-101.
- [12] "Internet Movie Era is Coming", Beijing Morning Post. <http://www.morningpost.com.cn/2014/0714/30303.shtml>, (2014).
- [13] "Internet Era: 75% Netizens Opinions Were Used in Filming "the Old Boy". Beijing Morning Post", http://tech.ifeng.com/internet/detail_2014_07/14/37318235_0.shtml, (2014).
- [14] S. Atkinson, J. Lehmann and R. Evans, "The Deep Film Access Project: Ontology and metadata design for digital film production assets, Big Data (Big Data)", 2014 IEEE International Conference on, doi: 10.1109/BigData.2014.7004445, (2014), pp. 1, 4, 27-30.
- [15] Z. Ling and Z. Zhao, "Why Letv Received a \$4.8 Billion Valuation", China Entrepreneur, vol. 20, (2014), pp. 32-33.
- [16] D. Xiaoxiao, "Cartoon Films Capture New Heights of Box Office – Analyses of Hollywood Movie Marketing Model in "The Avengers" Case", Chinese Film Market, vol. 5, (2013), pp. 21-24.
- [17] Z. Shunfeng, "How Big Data Power Movie Marketing? Marketing Success", <http://www.vmarketing.cn/index.php?mod=news&ac=content&id=5877>, (2013).
- [18] J. McKenzie, "Predicting Box Office with and without Markets: Do Internet users know anything? Information Economics and Policy", vol. 25, no. 2, (2013), pp. 70-80.
- [19] D Delen, R Sharda and P Kumar, "Movie forecast Guru: A Web-based DSS for Hollywood managers", Decision Support Systems, vol. 43, no. 4, (2007), pp. 1151-1170.
- [20] M Ghiassi, D Lio and B Moon, "Pre-production Forecasting of Movie Revenues with a Dynamic Artificial Neural Network. Expert Systems with Applications", vol. 42, no. 6, (2015), pp. 3176-3193.
- [21] Z. Jian and Z. Shangbo, "Modeling on Box-office Revenue Prediction of Movie Based on Neural Network", Journal of Computer, vol. 3, (2014), pp. 742-748.
- [22] J. Du, H. Xu and X. Huang, "Box Office Prediction Based on Microblog", Expert Syst. Appl, vol. 41, no. 4, (2014), pp. 1680-1689.
- [23] L. Y. Ying and H. Z. Xiao, "Turn Relational Data Model into an Object Data Model", Journal of Harbin University of Science and Technology, no. 5, (2001).
- [24] K. X. Feng, Z. H. Yu and L. Z. Jia, "Object-oriented Knowledge Representation Based on Database Technology", Journal of Harbin University of Science and Technology, no. 3, (2001).
- [25] Q. Yiqian, "How Theaters Arrange the Schedule? Big Data in Charge", Shaoxing daily. http://tech.gmw.cn/newspaper/2015-08/09/content_108450699.htm, (2015).

- [26] J. Taobao, “Push out a Booking List to Duide Cinema Circuits in Row Piece Arrangements”, “Virtual Circuit Alliance” tied to cinemas. Entgroup. <http://www.entgroup.cn/news/Markets/1223544.shtml>. (2015).
- [27] Z. Jian and Z. Shangbo, “The Movies of the Running Schedule of Summer Vacation in 2013”, Contemporary Cinema, vol. 10, (2013), pp. 4-10.
- [28] A. Chakravarty, Y. Liu and T. Mazumdar, “The Differential Effects of Online Word-of-Mouth and Critics' Reviews on Pre-release Movie Evaluation”, Journal of Interactive Marketing, vol. 24, no. 24, (2010), pp. 185-197.
- [29] S. Ruo, “The Big Data and Micro Marketing”, Contemporary Cinema, vol. 6, (2014), pp. 14-19.
- [30] X. Qi, “How “Big Data” Drive the Film Industry Innovation”, Journal of Tribune News Research, vol. 8, (2013), pp. 67-70.