

## Research Article

# Research on Graphic-Text Relationship in Film and Television Works Based on Big Data Model

Jun Liu <sup>1</sup> and YuQin Zhang <sup>2</sup>

<sup>1</sup>Jiangsu Normal University Kewen College, Xuzhou 221000, China

<sup>2</sup>Xuzhou Medical University, Xuzhou 221000, China

Correspondence should be addressed to YuQin Zhang; zhangyq@xzhu.edu.cn

Received 12 April 2022; Revised 5 May 2022; Accepted 17 May 2022; Published 24 June 2022

Academic Editor: Le Sun

Copyright © 2022 Jun Liu and YuQin Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rise of microblog, social network, mobile APP, location-based services, and other technologies, the global data has exploded. Big data provides a rich source for data mining and analysis, and the results mined from it will become more meaningful and even full of surprises. In this paper, through the problems in processing film and television data, the corresponding model is established to deal with the corresponding problems. According to the related technologies of big data processing and the filtering algorithm of the model, the processing strategies and methods are put forward. The middle also introduces the big data mining method. Finally, according to the different data characteristics and different data processing requirements of the intelligent analysis system of film and television data, comparing the data of film works and TV dramas, the experimental results show that, based on the big data model, the pictures and words in both film works and TV dramas are inseparable and indispensable. In the later research, we choose the appropriate data processing scheme to design and implement the data processing flow, and successfully apply the film and television big data processing strategy and method proposed in this paper to practice, which provides reliable data support for the content analysis of film and television works. The model-based filtering algorithm calculates the relationship between image and text, turns the point into the probability formula of node-to-meta path, and takes the approximate value to achieve efficient model learning.

## 1. Introduction

We put forward a “data usage quality model” in order to fill the gap that there are different data quality models to evaluate the quality of conventional data, but none of them is suitable for big data. This model consists of three data quality characteristics and can be integrated into any type of big data project because it is independent of any prerequisite or technology [1]. This paper introduces a big data model of the recommendation system based on social network data, which contains factors related to social network and can be applied to information recommendation of various social behaviors, thus improving the reliability of recommendation information [2]. This paper discusses the role of networked medical care and mobile cloud computing and big data analysis in its implementation, puts forward the

development momentum of networked medical applications and systems, and draws a conclusion that uses big data and mobile cloud computing technology to design networked medical systems [3]. This paper develops a big data analysis transformation model based on practice view, which reveals the causal relationship among big data analysis capability, IT-supported transformation practice, benefit dimension, and business value. The model also provides a strategic view of big data analysis and practical insights for managers [4]. Through the research of the big data analysis model, this paper establishes a big data management architecture model suitable for future pipeline system and illustrates the significance of big data in the development of pipeline industry [5]. This paper presents a new method for mining the big data using cloud model, which is a famous theory of uncertain artificial intelligence. Experimental results on actual

data sets show the effectiveness of this method [6]. Based on the accounting nature from the perspective of big data and cloud computing, this paper studies the construction of a multiple regression model related to management accounting. The final prediction model structure is the combination of straight line and curve prediction model structure [7]. In order to promote cultural soft power by establishing a good national image, this paper first studies what is China's national image; secondly, it emphasizes Wushu culture and advocates peace and harmony; finally, the measures to standardize Wushu film and television market are put forward [8]. Today, with the rapid development of economy and science and technology, China's film and television industry has made great progress, but the overseas market of Chinese film and television works is still backward. Therefore, this paper will put forward some strategies to solve the problems existing in the international distribution of Chinese film and television works [9]. This paper mainly studies the virtual expression in film and television animation art, analyzes the virtual technology in various aspects of film and television works shooting, and further studies the virtual aesthetics of animation art from the aspects of pictures and scenes, character image design, and design methods [10]. The basis of this paper is to combine Oracle database with GIS to study the distribution characteristics of film and television documents. In the aspect of database function, the system application test can meet the requirements of database retrieval and geographical location retrieval to quickly locate regional film and television literary forms [11]. This paper firstly investigates the attitudes of domestic audience towards foreign film and television works, and the results show that the audience has difficulties and lacks initiative in understanding foreign film and television works. Secondly, it describes the effects of foreign film and television works on teaching in culture and language, and finally puts forward effective practical strategies [12]. Using the framework of radical change, this study considers that both text conventions and readers' expectations of texts are increasing. This study considers the relationship between students' individual text features, especially the features they choose from school libraries. These proven relationships should help legitimize the inclusion of more graphic narrative text formats in school library collections [13]. This paper presents a conceptual framework based on cloud computing and discusses its future development based on on-demand service and decision support system. By combining the strategic perspective of information technology with the database, we can design a scientific perspective for a wider audience, thus providing new knowledge for service science [14]. This paper summarizes the content, scope, samples, methods, advantages, and challenges of big data, and discusses its privacy issues [15].

## 2. Big Data-Related Concepts

*2.1. Big Data.* The so-called big data refers to the data characterized by huge data volume, fast data speed, various data types, and low data value density. Due to the rapid

development of information technology, through mobile phones, computers, and other media, the growth rate of data sources has increased. Big data has three characteristics: diversity, large quantity, and fast speed. For data acquisition technology, because to ensure its accuracy and effectiveness, it is still facing great challenges.

*2.2. Characteristics of TV Data.* According to the characteristic analysis of TV data, TV ratings are affected by many factors. Some media think that there are two main factors, namely media and audience factors. Entering the era of science and technology, the rapid development of the Internet industry has brought great pressure to the traditional media industry, which leads to the mass media being defaulted as "traditional media." It breaks the monopoly position of traditional media on information dissemination and publishes a variety of electronic communication channels that are suitable for the public and can interact with each other. In recent years, new media has slowly broken the limitations of "traditional media." Let TV stations begin to consider the impact of Internet elements on ratings, such as the number of fans and searches of actors on platforms such as Weibo, the number of discussions on Weibo, the attention of the program, the number of searches by search engines, and the number of clicks on major video websites. Compared with the traditional industry and the prediction of film and television, the data mining of ratings has the following characteristics: many kinds of data, large quantity, short timelines, and fast technology update.

*2.3. Data Mining.* Data mining is a hot issue in the field of artificial intelligence and database. The so-called data mining refers to the nontrivial process of revealing hidden, previously unknown, and potentially valuable information from a large number of database data. Data mining is a decision support process, which can analyze the data of enterprises automatically, make inductive reasoning, dig out potential patterns from it, and help decision makers adjust market strategies, reduce risks, and make correct decisions.

### 2.4. Data Mining Methods

*2.4.1. Cluster Analysis.* Clustering is a multivariate statistical analysis method for classification. After classifying them according to their characteristics, individuals with the same characteristics are divided into the same set, and each set should have different characteristics without affecting each other. Clustering can find the seemingly irregular datasets that may have direct relationships.

*2.4.2. Classification.* Classification is one of the key technologies in the field of data mining. Data classification is to achieve the purpose of quickly distinguishing different types of data. Each record is tested by a certain calculation method so as to realize the prediction of quantity category.

**2.4.3. Sequential Mode.** Time series pattern can get the law of things' change. Time series mining has been applied in many industries. For example, the change of sunspot number increases or decreases with time. By collecting the historical change curve of sunspot number, we can predict the expected number of sunspots at the next time point to a certain extent. There is also an analysis of passenger flow and so on. As an important branch of data mining, time series patterns play an increasingly important role in social sciences and other fields.

**2.4.4. Bias Analysis.** Deviation refers to the obvious inconsistency between the detected data and the previous data or normal data. When confirming the deviation not caused by observation errors, a lot of useful information can be obtained by analyzing these deviations.

### 3. Model-Based Filtering Algorithm

**3.1. Model Description.** In the model, each node  $v_i$  has two identities, source and environment, which are represented by vectors  $s_i$  and  $c_i$ , respectively. Given a metapath  $P$ , a metapath instance from Node  $v_i$  to Node  $v_j$  is defined as an ordered node pair  $(v_i, v_j)$  obtained by sampling based on the proximity of the two nodes under the metapath  $P$ . Through the proximity degree, the probability of the occurrence of node  $v_i$  predicted by node  $v_j$  under the condition of  $P$  is calculated as shown in formulas (1), (2), and (3):

$$\text{prob}(v_j|v_i, p) = \frac{\exp(s_i \cdot c_j)}{\sum_{v_k \in V_{\text{emb}}} \exp(s_i \cdot c_j)}, \quad (1)$$

$$P_{i,d} = \bar{R}_i + \frac{\sum_{j \in NES_i} \text{sim}(i, j) * (R_{i,d} - \bar{R}_i)}{\sum_{j \in NES_i} (|\text{sim}(i, j)|)}, \quad (2)$$

$$W_{ij} = \frac{|N(i) \cap N(j)|}{\sqrt{|N(i)||N(j)|}}, \quad (3)$$

where  $V_{\text{emb}}$  is the node set that needs to be learned to represent. The process of expressing vector learning can be transformed into minimizing the KL divergence of empirical distribution and estimated distribution. The empirical distribution is obtained here by normalizing the proximity based on the metapath. KL divergence has two main properties: asymmetry and non-negativity. After simplification, formula (4) and formula (5) are obtained:

$$o_{ij} = \widetilde{\text{prob}}(v_j|v_i, p) * \log(\text{prob}(v_j|v_i, p)), \quad (4)$$

$$p(u, j) = \sum_{i \in N(u) \cap N(j,k)} W_{ij} r_{ui}. \quad (5)$$

The optimization of the above objective function is difficult because the denominator part of formula (1) needs

to traverse the whole  $V_{\text{emb}}$  set and calculate the inner product sum. Negative sampling technique is first proposed in word representation learning algorithm, which is a special case of noise contrast estimation in essence and is widely used in the optimization of representation learning. The objective function is converted into formula (6) and formula (7):

$$o_{ij} = \log \sigma(s_i \cdot c_j) + K * E_{v_n \sim P_{\text{neg}}(v_n)} [\log \sigma(-s_i \cdot c_j)], \quad (6)$$

$$\text{Distance}(d_i, d_j) = \sqrt{\sum_{k=1}^{k=n} (w_k * d_i - w_k * d_j)^2}, \quad (7)$$

where  $K$  is the sampling number of negative samples, the sampling rate of negative samples is set to  $\text{prob}_{\text{neg}(v)} \alpha d_v^{3/4}$ , and  $d_v$  is here the number of instances of metapath starting from a certain node  $V$ . The essence of negative sampling is to choose a certain method to adjust the sampling distribution.

**3.2. Model Learning.** It takes a lot of time to measure the proximity by counting metapath instances and traversing the whole set at the same time to get the exact distribution value, which is not feasible. Here, a random test is used to calculate the approximate empirical distribution value, and  $N$  nodes are randomly and independently sampled from the node set along a given metapath  $P$  with node  $v_i$  as the starting point so that the empirical distribution can be approximated as formula (8):

$$\widetilde{\text{prob}}(v_j|v_i, p) = \frac{\#\text{pathEndsAt}(v_j)}{N}. \quad (8)$$

With the above estimation method, we can design learning strategies to learn the model efficiently. Commonly used model learning optimization algorithms mainly include the gradient descent method, Newton method, quasi-Newton method, DFP algorithm, BFGS algorithm, and improved iterative scale method.

The calculation process is as follows:

$$\text{sim}(\vec{x}, \vec{y}) = \cos(\vec{x}, \vec{y}). \quad (9)$$

After simplification, it is

$$\frac{\vec{x} \cdot \vec{y}}{|\vec{x}| |\vec{y}|}. \quad (10)$$

Specifically, let the number of ordered node pairs  $(v_i, v_j)$  be sampled from node  $v_i$  along metapath  $P$ , and the rewritten global objective function is

$$o = \sum_j \sum_i \#s_{v_j, p}(v_i) (\log \sigma(s_i \cdot c_j)) + K * E_{\text{neg}} [\log \sigma(-s_i \cdot c_j)]. \quad (11)$$

There are two steps in sampling the learning model. Firstly, the node  $v$  is obtained from the node set sampling,

which follows the same distribution as that defined in negative sampling. Finally, node  $v_j$  is obtained by using model random walk sampling. The probabilities of this path sampling random walk guided by  $P$  in step  $n$  are as follows:

$$p(v_{k+1}|v_k) = \begin{cases} \frac{\text{Weight}(v_{k+1}|v_k)}{\sum_{v_j \in N(v_k, P_{k+1})} \text{Weight}(v_j, v_k), \text{VertexType}(v_{k+1})=P_{k+1}}, \\ 0, & \text{otherwise,} \end{cases}$$

$$\text{sim}(d_i, d_j) = \cos \theta = \frac{\sum_{k=1}^n (w_k * d_i - w_k * d_j)}{(\sum_{k=1}^n w_k^2(d_i))(\sum_{k=1}^n w_k^2(d_j))}. \quad (13)$$

In addition to checking the existence of a path instance, the above formula also uses a weight function to handle the sideband weighted case, which is defined as follows:

$$\text{Weight}(v_i, v_j) = \begin{cases} \{w_{ij}^\theta\}, & (v_i, v_j) \in E, \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

Because the weights of edges may have large errors, the function adds a content between 0 and 1 to the original value of the weights.

**3.3. Text Representation Learning for Tag Embedding.** Now, it is supposed that you already have a set of tags, and the user needs to select several tags from the set to index the important information of the document. If you think of indexes as links, a particular tag should not be directly linked to text, but more likely to be linked to a context, which refers to a fixed-length sequence of text. In two extreme cases, it is a single word or an entire document. Let the probability of occurrence be determined by  $t$  with  $w$ , and the condition is independent, that is,

$$\text{prob}(w_j|w_i, t_k) = \text{prob}(w_j|w_i) * \text{prob}(w_j|t_k). \quad (15)$$

At the beginning of this paragraph, aiming at the conjecture of tagging, we think that tags are related to a context, so we should simplify it appropriately and keep only the tags with the highest degree of relevance. Formula (8) is deformed into

$$\text{prob}(w_j|w_i, t_k) = \text{soft max}(x_i \cdot c_j) * \text{soft max}(h_k \cdot c_j). \quad (16)$$

The calculation of the function in the above formula needs to traverse the whole text set and label set, which is not desirable in efficiency. Therefore, negative sampling is used to reduce the complexity. We convert formula (16) into logarithmic form, such as formula (17):

$$O = \log \sigma(y_{ij}) + K * E_{\text{neg}}[\log \sigma(-y_{ij})] + \log \sigma(Z_{ij}) \\ + K * E_{\text{neg}}[\log \sigma(-z_{ij})]. \quad (17)$$

The problem of analyzing the similarity between two documents is to analyze the relationship between images and words. Let  $D_1$  and  $D_2$  be two documents, and  $X_1$  and  $X_2$  are the representation vectors corresponding to the title; then, the general definition of text similarity is as follows:

$$\text{sim}(d_1, d_2) = g(x_1 + x_2). \quad (18)$$

Here, we use a learning-based scheme to calculate images and words. Specifically, let  $s$  contain a fixed number of words quantity, as defined in formula (19):

$$x_s = \frac{1}{|s|} \sum_{w \in s} \frac{a}{a + \text{freq}(w)} x_w. \quad (19)$$

Among them,  $x$  is the representation vector of the text  $w$ , and the freq function can calculate the frequency of the text appearing in the text set.

### 3.4. From Text Representation to Document Representation.

Get the representation of the text first, and then get the representation of the document. The algorithm here consists of a smooth inverse frequency weighting process and a common component removal process. Compared with other weighting schemes, the algorithm has better performance and even approaches the results of sequential neural network processing in some experiments. Let  $s$  be a sentence,  $w$  be a word appearing in the sentence, and  $p(w)$  be the distribution probability of the word  $w$  in the whole document set, then the probability of  $w$  appearing in  $s$  can be defined as

$$\text{prob}(w|c_s) = \alpha p(w) + 1 - \alpha \frac{\exp(\langle c_s, v_w \rangle)}{z_{cs}^-}. \quad (20)$$

It can be seen that the two terms on the right side of formula (20) are aimed at two influencing factors, respectively, and then the probability of sentence occurrence is obtained from the probability of word occurrence as in formula (21):

$$\text{prob}(s|c_s) = \prod_{w \in s} \text{prob}(w|c_s). \quad (21)$$

Now, we need to maximize the above probability, which can be done by using maximum likelihood estimation. Formula (22) is obtained by taking logarithm of formula (21):

$$f_w = \log \left[ \alpha p(w) + 1 - \alpha \frac{\exp(\langle c_s, v_w \rangle)}{z_{cs}^-} \right]. \quad (22)$$

The estimation result can be approximated as formula (23):

$$\arg \max_{w \in s} f_w(\tilde{c}_s) \alpha \sum_{w \in s} \frac{\alpha}{\alpha + p(w)}, \quad (23)$$

$$\text{sim}(x, y) = p(x, y) = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{(n-1)s_x s_y}, \quad (24)$$

$$\text{Rating}_{ci} = \frac{\sum \text{rating}_{cij}}{m * n}. \quad (25)$$

The above approximate results show that the words with high frequency will be suppressed after the modification of sentence information. To get the representation of sentence information, it is necessary to remove the part of  $c_0$ , that is, to estimate the direction of  $c_s$ . This can be obtained by calculating the first principal component for  $c_s$

corresponding to a group of sentences. Because the number of sentences in short text is often very small, this paper directly replaces the corresponding sentence representation with the document representation and uses the above algorithm to calculate the document representation. On the basis of building a big data model, this paper studies the relationship between pictures and texts from the relationship between words and texts, which are related, and the filtering algorithm based on the model can mine the contents in a deeper level.

#### 4. Comparison of the Trends of Chinese Movies and TV Plays

4.1. *Box Office Analysis of Chinese Movies.* Big data shows that the film and television industry is developing rapidly, and the results are shown in Table 1.

Looking at Table 1, we can see that the box office revenue is increasing year by year, which indirectly shows that the quality of film and television works is gradually improving.

According to the growth curve of box office revenue in Figure 1, the box office revenue shows a linear growth trend in recent years, and the fitting function between box office revenue and time is obtained by linear regression fitting.

4.2. *Analysis of the Output and Number of Episodes of Chinese TV Dramas.* Big data shows that the number of TV dramas produced in China was 14,685 episodes in 2010 and 17,703 episodes in 2012, which reached its peak at this time. After that, the output of TV dramas in China has been decreasing, and as of 2018, it is only 13,310 episodes, as shown in Table 2.

It can be seen from the analysis that in recent years, the total output of TV dramas in China has decreased, but it is increasing in certain TV dramas. In fact, in recent years, there are many hot TV dramas, with more than 50 episodes of urban dramas and more than 60 episodes of costume dramas, and many dramas have “water injection.” The result is shown in Figure 2.

Looking at Figure 2, we can see that the output of TV dramas declined from 2010 to 2018, but the number of episodes of each TV drama gradually increased, which reflects the fierce competition between TV dramas and film works in recent years. Domestic film works began to emerge and become more and more popular with the audience, and both of them are constantly developing and growing.

4.3. *Film Production Growth and Distribution of High-Scoring Films.* While developing the domestic film market, we should pay close attention to the development trend of overseas film industry and learn from other people’s successful experiences. Based on the network data since 1960, this paper summarizes the number of movies in 10 years and makes statistics on the types and geographical distribution of high-scoring movies, as shown in Figures 3–5.

Figure 3 shows that from 1960 to 1990, China’s film production grew slowly, which was related to the national economic situation at that time; from 1990 to 2010, the film production began to develop rapidly. At this time, China’s

TABLE 1: Box office revenue from 2014 to 2018.

Time (year)	Total box office (100 million)
2014	296
2015	440
2016	457
2017	559
2018	600

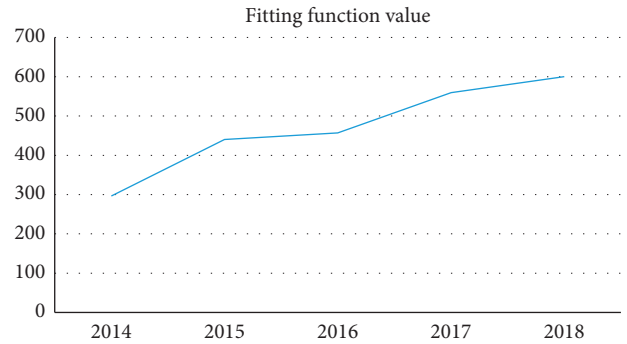


FIGURE 1: Comparison of fitting function values of movie box office in five years.

TABLE 2: TV drama production in recent years.

Time (year)	Output (column)	Number of sets
2010	436	14685
2012	506	17703
2014	429	15983
2016	330	14768
2018	310	13310

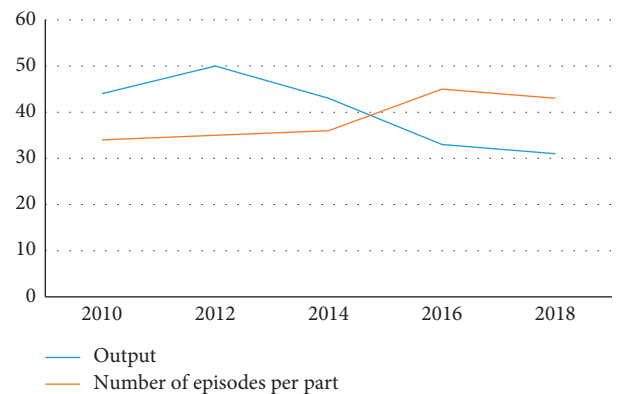


FIGURE 2: Comparison of TV drama output and number of episodes per episode.

economy began to develop rapidly, which also led to the development of the film industry, whether from the investment or the training of actors and other aspects. Secondly, in order to expand and internationalize China’s film market, it is necessary to retain the traditional high-quality culture of domestic films. From the output growth rate in Figure 3, we can see that movies are developing better than TV dramas, but the output of TV dramas is also increasing.

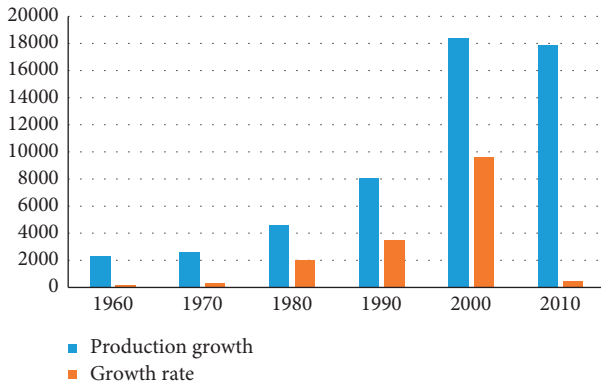


FIGURE 3: Growth in film production.

At present, China's media industry is in a very complicated environment, which has both international competitive pressure and domestic diversified disputes. By analyzing Figure 4, in recent years, documentaries have become more and more popular with audiences, probably because documentaries have a strong sense of reality and experience, while the quality of action films and science fiction films, which are popular in previous years, began to decline and gradually broke away from high-scoring films.

Figure 5 shows that the proportion and control of high-scoring films in mainland China need to be improved, and domestic films must have corresponding coping mechanisms for complicated environment so as to ensure that Chinese films have a place under the fierce international competition environment. Facing the whole international competition, we must integrate video resources, improve our competitive strength, and highlight our advantages. It can also break the boundaries of movies, combine other powerful domestic media across borders, and form a strong alliance group to cope with the global competitive pressure. In addition, we should go abroad to unite with film and television companies in other countries, set up transnational media with our giant film company as the core, and adopt the principle of fair risk to resist competitive pressure together.

**4.4. Movie Factor Analysis.** Taking movie investment, audience preferences, online novels and cartoons, online ratings, and types of people watching movies as keywords, relevant data are collected on the Internet, and relevant software is used to process the searched data. This companion probability of sphericity test is 0.00, which is less than the significance level of 0.05. The calculation results show that factor analysis can be applied, and the analysis results are shown in Table 3.

Table 3 shows that the biggest factor affecting the box office revenue in Chinese mainland is the production area, followed by whether the graphic content is adapted from famous novels and cartoons. At present, the quality of domestic films is inferior to that of foreign countries, which leads to a result that the box office continued to decline. Because of the mutual influence in the eyes of movies and

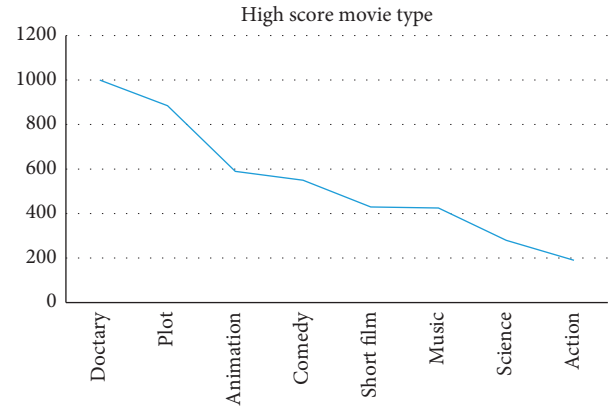


FIGURE 4: Types of high-scoring films in recent years.

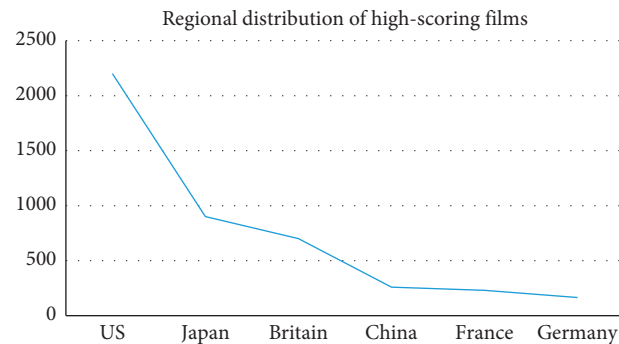


FIGURE 5: Regional distribution of high-scoring films in recent years.

TABLE 3: Factor analysis based on film.

Model	Error	Coefficient	$t$	Sig
Constant	12	205	17	87
Production	6.8	796	26	16
Investment	1.1	27	14	28
Graphic content	3.6	99	58	16
Score	1.6	188	15	38
Viewing type	2.8	94	34	32

TABLE 4: Factor analysis based on TV series data.

Model	Error	Coefficient	$t$	Sig
Constant	1.9	2.8	2.5	23
Production	265	464	1.7	102
Investment	861	324	1.4	174
Graphic content	623	370	1.5	142
Score	126	299	1.2	234
Viewing type	248	20	8.6	32

other works, films adapted from famous works very favorable in terms of effect.  $t$  represents the initial eigenvalue of the variance, and sig represents the final eigenvalue of the normalized coefficient.



FIGURE 6: Comparison between standard coefficient error and sig value of movies and TV dramas.

**4.5. Factor Analysis of TV Plays.** Similarly, considering the factors such as investment, ratings, episodes, whether adapted from well-known novels, network ratings, audience types, and so on, standardized factor analysis was carried out on the processed data by using software, and the accompanying probability given by the sphericity test was 0.007, which was less than the significant level (0.05). The calculation results show that it can be applied to factor analysis, and the analysis results are shown in Table 4.

The results in Table 4 show that the two positive factors that have the greatest influence on the ratings of domestic TV dramas are do they really come from online novels or scored on the Internet. In addition, it can be seen that the ratings are not necessarily positively related to production, in fact, the reputation of works that cost too much in the early stage is not very good, and the ratings of TV dramas with large investment are not necessarily good. The current situation is that TV dramas adapted from online novels have high praise and high ratings.

Figure 6 compares the ratings of standard coefficient errors and sig values of film works and TV dramas. The results show that the popularity of TV dramas is slightly higher than that of movies in recent years. One of the major factors may be that the adaptation of movie content conforms to the audience's interests. At present, the on-demand of TV dramas on emerging media platforms has begun to get used to it. Network base station use a variety of ways to obtain video resources, so there are majority online videos, mass multiuser watch that are film and television videos. The ordinate in Figure 6 represents the error value of the ratings standard coefficient of film works and TV dramas.

## 5. Conclusion

Nowadays, the competition in film and television industry is becoming more and more fierce, so we are urged to improve the shortcomings of the current film industry in order to achieve a good prospect. With the increasing requirements of reform and opening up to science and technology, it is extremely urgent to be in line with international standards. Based on big data analysis, China's film and television works are adapted from famous novels or cartoons. The relationship between pictures and texts in film content is

particularly important. The competition and cooperation between pictures and texts in the image era has a comprehensive impact on literary activities. As an important carrier art of symbols, literature has a long history, a relatively complete mechanism of creation, dissemination, acceptance and recreation, and a fixed creative group and acceptance group. What is more, language symbols have their own unique field, which cannot be replaced by other symbols. Charts and words cannot replace each other, they are stitched. This relationship will surely become a fine tradition in the film industry. In the future development, the two will become closer and closer, and jointly promote the development of the film and television industry, and the cultural treasures of the Chinese nation will also go to the world in all directions. Although this study has realized the application of model learning and made clear the relationship between images and words in a long time, it is not deep enough in the field of model mining. In order to make film and television works more perfect, it is necessary to conduct in-depth research on the methods based on the big data model in the later period.

## Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

## Acknowledgments

The Philosophy and Social Science Research Project of Jiangsu Universities: The Influence of the construction of national image in Film Works on Ideological and Political Education of College Students (No: 2020SJB1215).

## References

- [1] M. Jorge, Ae L. C. Ism, and R. Bibiano, "A data quality in use model for big data," *Future Generation Computer Systems*, vol. 63, pp. 123–130, 2016.
- [2] X. Han, "A big data model supporting information recommendation in social networks," *Second International Conference on Cloud & Green Computing IEEE Computer Society*, vol. 13, no. 5, pp. 533–549, 2012.
- [3] L. A. A. Tawalbeh, R. Mehmood, E. Benkhelifa, and H. Song, "Mobile cloud computing model and big data analysis for healthcare applications," *IEEE Access*, vol. 4, no. 99, pp. 6171–6180, 2016.
- [4] Y. Wang, L. A. Kung, and W. Wang, "An integrated big data analytics-enabled transformation model: application to health care," *Information & Management*, vol. 4, no. 48, pp. 64–79, 2017.
- [5] S. Dong and A. N. Yu, "Data analysis model for pipeline system and its application based on Big Data," *Oil & Gas Storage and Transportation*, vol. 36, no. 5, pp. 82–255, 2015.

- [6] J. Yang, W. P. Li, and J. P. Zhang, "Canonical correlation analysis of big data based on cloud model," *Journal on Communications*, vol. 24, no. 1, pp. 27–50, 2013.
- [7] Z. Hang, "The research based on big data management accounting model building," *Science Innovation*, vol. 4, no. 4, p. 228, 2016.
- [8] X. J. Wang and J. Y. Yang, "Orientation and development of Wushu film and television works in view of national image," *Journal of Wuhan Institute of Physical Education*, vol. 31, no. 4, pp. 153–170, 2011.
- [9] J. Zhao and Wang, "On international distribution strategy of Chinese film & television works," *Ecological Economy: English Version*, vol. 4, p. 7, 2019.
- [10] W. Jian, "The analysis of the virtuality of film and television animation art," *Advanced Research & Technology in Industry Applications IEEE*, vol. 37, no. 5, pp. 119–121, 2014.
- [11] Y. U. Hongwen, "The database design scheme of geographical distribution characteristics of film and television literature and," *Art Form*, vol. 5, no. 6, pp. 579–596, 2016.
- [12] X. Deng, "A survey on foreign film and television works and foreign language teaching and learning," *Journal of Nanyang Teachers College*, vol. 67, no. 15, pp. 244–367, 2007.
- [13] W. S. Stephens, "The influence of engagement with graphic narrative text formats on student attitudes towards the school library," *Unt Theses & Dissertations*, vol. 3, no. 2, pp. 32–58, 2014.
- [14] H. Demirkan and D. Delen, "Leveraging the capabilities of service-oriented decision support systems: putting analytics and big data in cloud," *Decision Support Systems*, vol. 55, no. 1, pp. 412–421, 2013.
- [15] S. Sagioglu and D. Sinanc, "Big data: a review," in *Proceedings of the International Conference on Collaboration Technologies and Systems (CTS)*, vol. 24, no. 22, pp. 556–675, IEEE, San Diego, CA, USA, May 2013.