# Analysis of the Polarization of Sentiment in Chinese Film Reviews and Related Factors

Haoyuan Qin
*Faculty of Engineering and Information Technology, The University of Melbourne, Victoria, 3010, Australia*

Keywords: Sentiment Polarization, Film Reviews, Natural Language Processing, Box Office Analysis.

Abstract: At a time when film production is increasingly targeting specific audiences, the phenomenon that some Chinese film reviews have more emotional outputs between different audiences rather than discussions about the content of the film has attracted widespread attention, but there is still a lack of specific qualitative and analytical indicators for emotional polarization in reviews. This paper analyzes the public data of Douban film reviews of 28 film samples, divides the emotional tendencies of their film reviews into three categories to analyze whether they have emotional polarization, and studies the correlation between polarization and other factors through quantitative analysis. This paper divides the types of film review polarization into three categories and analyzes that on the Chinese film review platform, the film actor factor is more likely to cause emotional polarization in film reviews than other factors and needs to be paid more attention to when taking advantage of consumers' emotional consumption behavior. Based on this, this paper makes the following suggestions: In the future, the film review emotional polarization index can be used as a new indicator for analyzing the controversy and content effectiveness of film content, and more film review data can be analyzed.



## 1 INTRODUCTION

In the context of the mature film industry today, the audience's evaluation of movies has formed a huge amount of film review data on the Internet. With the diversification of entertainment forms, the audience's aesthetic standards for film and television works have also changed. The audience's evaluation and the discussion in the comment area have become important indicators for some users to choose movies. The light-hearted and funny "popcorn movies" are still hot-selling, but as an important indicator for evaluating movies, the movie's comment area sometimes gives negative opinions. Although some comedy movies have achieved success at the box office, there are controversies about cultural criticism in the film review area, especially whether their jokes are vulgar or disrespectful, such as "making fun of the poor". Even in the comment area, some comments are not directed at the content of the movie but simply output emotions. On the other hand, although the quality of movie content varies, many works have been labeled as "vulgar" because of their overly deliberate, dramatic, and low-quality movie plots, and the film review emotions are polarized or one-sidedly positive/negative comments, but they also have a strong appeal because of their fast pace and easy to fascinate people. Therefore, if it can find out the factors related to the "polarization" and "one-sidedness" of film reviews, and the factors that make the reviews as a whole tend to output emotions rather than content, that is, the emotional polarization of film reviews, it may have a positive impact on the review of film production and the early publicity and promotion. In addition, since the emotional tendency is generally more extreme when the reviews are emotional outputs lacking content, studying the emotional polarization of film reviews may also provide more quantitative indicators for the analysis of the controversial degree of film content.

At present, with the rapid development of machine learning and natural language processing, many practical models have been put into use. As an important branch of natural language processing, sentiment analysis aims to understand and judge the emotional color contained in the text through computer algorithms and models, whether it is positive, negative, or neutral. Nowadays, many powerful and practical models have emerged, and some of them have even successfully entered the field of commercial applications. By using these natural language models, combined with methods such as the



Qin, H.
Analysis of the Polarization of Sentiment in Chinese Film Reviews and Related Factors.
DOI: 10.5220/0014151400004942

In *Proceedings of the 2nd International Conference on Applied Psychology and Marketing Management (APMM 2025)*, pages 703-709
ISBN: 978-989-758-791-7

sentiment dictionary of the film industry, the emotional tendency of film reviews can be obtained more accurately, and some emotional tendency patterns with statistical validity can be obtained by analyzing a large number of comments in the film review area. Specifically, when a film review is input, the model can quickly analyze the vocabulary, sentence structure, and context, thereby determining the emotional attitude expressed by the author when writing the review, whether it is praise, criticism, or neutral evaluation of the film (Ullah et al., 2023). In addition, due to the openness of review data, a large number of public user reviews can be easily obtained from domestic film review platforms. The openness and accessibility of this data provide conditions for an in-depth understanding of user feedback and emotional dynamics in the film market and also provide a strong reference for the development and optimization of the film industry.

In recent years, with the development of natural language processing technology, the research on combining film reviews with sentiment analysis mainly includes using machine learning methods to try to extract effective content from the words and sentences in emotionally extreme film reviews and improve them. Lee, Sang Hoon, and others use sentiment dictionaries to assist in analyzing the sentiment of film reviews and effectively improve the accuracy of models in the field of film review analysis (Lee et al., 2016). Because there is a correlation between the emotionality of film reviews and the effectiveness of the content, sentiment analysis can be used to classify film reviews to find more high-quality film reviews that are worth referring to. Sudhanshu Kumar and others combine sentiment analysis with traditional recommendation systems to filter and classify emotional reviews through sentiment analysis to recommend reviews and content that are closer to the user's emotional tendencies (Kumar et al., 2020; Soubraylu & Rajalakshmi, 2021). In terms of specific analysis methods, they include using hybrid models, pre-screening and classification based on text length, and building extreme word recognition dictionaries to perform more accurate sentiment analysis, etc., to improve the accuracy of analysis (Peng & Cheng, 2023). At present, there is relatively mature analysis software in the commercial market that can monitor the generation and decline of hot topics and even judge social media group polarization, which act as "accelerators" to jointly promote the generation of group polarization (Rabiee et al., 2024; Utmhikari, 2017). However, there is still a lack of research from a specific perspective on the correlation between audience emotional polarization and film-related factors.

The core issue of this study is what factors of the film are mainly related to the emergence of the emotional polarization phenomenon in the audience's evaluation in the comment area. Unlike existing film review sentiment analysis research, this study focuses on the audience emotional polarization phenomenon itself. This phenomenon is of great value because if it can be determined that the type of film or specific factors are correlated with the audience's emotional polarization or even causal relationship, it will have a positive effect on the filmmaker's marketing strategy formulation, helping them to accurately locate potential users and improve the success rate of marketing activities. Specifically, this study aims to define and classify the emergence of audience emotional polarization and analyze whether there is a correlation between the type of film, box office, ratings, and audience emotional polarization to provide possible new indicators for the analysis of film reviews.

# 2 METHOD

The data for this study comes from the public film review data uploaded to Kaggle in 2017, which is sourced from the Douban platform (Yuan et al., 2020). The data mainly includes the Chinese name, English name, review content, rating, number of likes, and other information about the movie. The earliest movie in the data is Iron Man 1 in 2008, and the rest of the movies are from 2012 and later. This data collection method avoids the copyright and technical difficulties that may be faced by self-collected data and can quickly obtain a large amount of real and public film review data.

## 2.1 Sentiment Tendency Analysis

This study aims to explore whether the type of movie affects the audience's emotional polarization. The analysis model selected is the Baidu voice sentiment tendency analysis model, which has been put into commercial use and is relatively stable. From the 28 movies in the database, 2000 reviews of each movie were randomly selected for analysis. Before data analysis, it first cleaned the comment confidence (confidence>0.5), then performed sentiment analysis on each comment, and finally combined statistical methods to analyze the sentiment distribution and its possible polarization phenomenon, and obtained the correlation analysis between the audience's sentiment

polarization phenomenon and the box office, ratings, and movie types. Specifically, after separating the extremes, if most of the extreme values are positive, it is considered to be a positive extreme; otherwise, it is a negative extreme. However, when the positive and negative extreme values are relatively balanced, they need to be discussed in categories. For example, comments with a positive sentiment bias greater than 0.99 are called positive extreme comments, and vice versa. After assuming that comments with a positive sentiment bias greater than 0.99 or less than 0.01 are extreme values, the number of extreme comments of each movie is counted and sorted, and it can be concluded that among the 28 movies, "Nine-story Demon Tower" and "Why the Flute is Silent" accounted for more than 66% of negative emotions, showing a typical negative emotional polarization phenomenon.

By identifying the opinion keywords and emotional words in the comments, it is judged whether there is a conflict of opinion between the comments. If there are a large number of comments that refute and argue with each other, and the sentiment tendencies of these comments are opposite (one is positive, the other is negative), it means that there may be emotional polarization in the comment area. For example, in the comment area about a certain movie, some users praised the actors for their excellent performance (positive), while other users strongly criticized the actors' acting skills (negative), and the two sides argued fiercely, which may be a manifestation of emotional polarization. At the same time, by counting the number and proportion of controversial comments, the phenomenon of comment emotional polarization can also be identified. For example, use Baidu's sentiment tendency analysis interface to calculate the proportion of extreme values (prob>0.99/non-middle 50%) and use distribution statistics indicators to count the proportion of different sentiment tendencies. If the proportion of controversial comments gradually increases, it means that the conflict of opinions in the comment area is intensifying, which may cause emotional polarization. Even if the emotional differences between groups are calculated in real-time, if the difference in emotional tendencies between different groups increases, it means that the emotional differentiation of commentators is intensifying. Such a possible phenomenon of emotional polarization can also be predicted in advance and may have an impact on real-time marketing strategies (Schwertman et al., 2004).

## 2.2 Correlation Analysis

Correlation is a statistical indicator used to measure the strength and direction of the linear relationship between two or more variables. When analyzing movie parameters, correlation can help analyze and understand the degree of correlation between different parameters (such as movie ratings, box office, number of moviegoers, etc.). Commonly used correlation indicators include the Pearson correlation coefficient, which measures the degree of linear correlation between two variables, and its value range is between -1 and 1. A value of 1 indicates a complete positive correlation, a value of -1 indicates a complete negative correlation and a value of 0 indicates that there is no linear correlation. This article uses the Pearson correlation coefficient to perform correlation analysis after one-hot encoding nonlinear parameters such as movie types and actors.

## 2.3 Word Frequency Analysis of Film Reviews

In the process of film review sentiment polarization analysis, the frequency statistics of text data is an important means to help identify the most representative keywords in the comments and reveal the core topics discussed by the audience under different movie types or emotional tendencies.

To ensure the accuracy of the analysis, this study first preprocessed the review text, including removing special characters and punctuation marks and using the stop word list of Harbin Institute of Technology to remove stop words from the review text to reduce the interference of invalid information. After the text is cleaned, this study uses word frequency statistics, TF-IDF (word frequency-inverse document frequency), and other methods to analyze the distribution of high-frequency words in each movie review and combines word cloud, table, and other visualization methods to display the distribution of keywords, to intuitively present the key content discussed in the comment area. For example, for movies with high box office but serious emotional polarization of film reviews, it is possible to observe whether high-frequency keywords are concentrated on specific topics (such as actors, plots, special effects, etc.) and further analyze whether these keywords are related to the emotional tendency of film reviews.

# 3 RESULT

## 3.1 Movie Genres and Emotional Tendencies

As shown in Fig. 1, in the study of movie genres and audience emotional tendencies, it was found that the average emotional tendencies of audiences for musicals, animations, and science fiction films showed a significant positive trend. In contrast, youth and romance films showed a different trend, with audience emotions tending to be more negative, and in the evaluation of such films, the phenomenon of emotional polarization was particularly prominent, that is, the audience's evaluation was more likely to be divided into two completely different poles, and the difference and opposition between positive and negative evaluations were more significant.

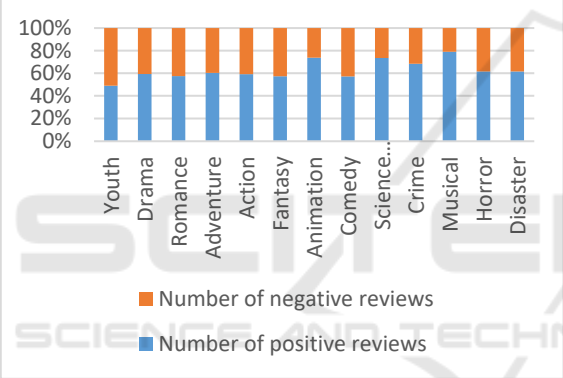


Figure. 1. Average emotional tendency of movies of different genres.

## 3.2 Analysis of Three Types of Sentiment Polarization

Since the quartiles can clearly and directly display the five important features of the data: the extremes, the upper and lower hinges (quartiles), and the median therefore using the quartiles to display the sentiment tendency score of each movie can directly determine the proportion of extreme values through the width of the upper and lower hinges (Guo, 2020). Based on the in-depth exploration and analysis of the sentiment analysis quartiles of 28 movies, it can be seen that the distribution area occupied by some movies in the quartiles is relatively wide, that is, the box area is relatively large. This feature means that in the audience's evaluation system of these movies, the distribution of positive and negative emotions tends to be balanced, and its sentiment tendency presents a diversified pattern, reflecting a large emotional discreteness. On the contrary, the distribution area of some other movies in the quartiles is highly concentrated, and the range is relatively narrow. The evaluation of these movies focuses more on the positive emotion dimension, and the proportion of negative emotions is relatively small. From the perspective of the theory of emotional polarization, this phenomenon may imply that such movies are less likely to cause negative reviews, and the audience group maintains a relatively consistent positive emotional attitude towards the movie as a whole, which in turn causes its distribution in the quadrant to show a relatively concentrated trend, which is in sharp contrast to those movies with dispersed emotional tendencies.

To analyze the three different types of emotional polarization tendencies more clearly, this study chose to observe the emotional analysis quadrants of three specific movies in detail (Fig. 2). It can be seen that the box of Chronicles of the Ghostly Tribe is generally lower and the median is almost zero, which means that the film review sentiment of this movie tends to be negative, and the most frequent words in its film reviews are discussions about the original work, actors, and directors. The box of The Mermaid is relatively large, and the median is relatively centered. An in-depth analysis of the main reasons behind it may be because the film focuses on the construction and portrayal of the relationship between actors at the core theme level of narrative and discussion. The word frequency analysis of the film reviews shows that the director's name, the leading actor's name, environmental protection, plot, and other representative keywords appear. The existence of these elements may, to a certain extent, reflect the wide distribution of audience evaluations in the emotional dimension, which increases the degree of dispersion of evaluations and makes the emotional tendency more diversified and decentralized. The box of Zootopia presents a smaller scale. This phenomenon intuitively reflects that its evaluation distribution has a highly concentrated characteristic and mainly tends to be in the positive evaluation category. Its degree of emotional polarization is more concentrated than that of the other two films. The audience's evaluation of the film is highly consistent, and the proportion of negative evaluations is extremely low. The words with a high evaluation rate in their comments include animals, plots, and stories, that is, discussions on the content of the movie.

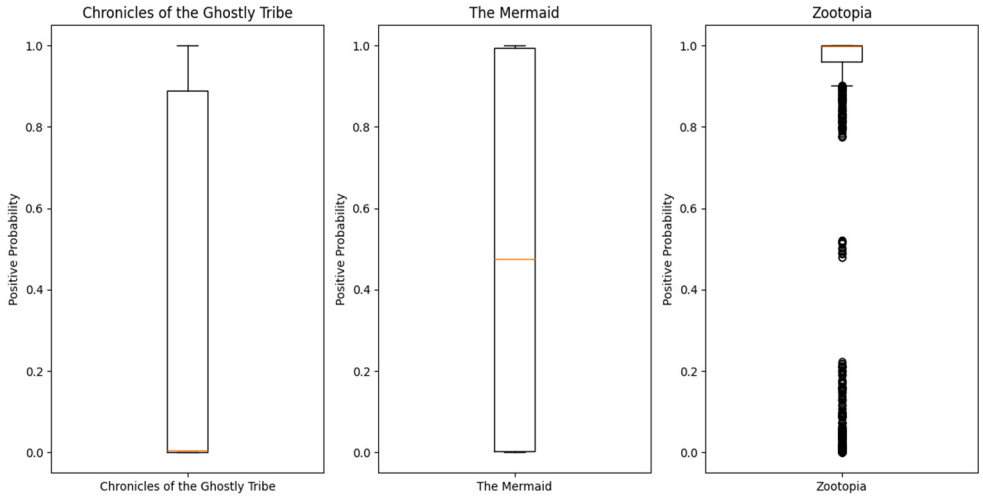


Figure. 2. Quartile chart of sentiment analysis of film reviews for the three films "Chronicles of the Ghostly Tribe", "The Mermaid", and "Zootopia".

This study uses the proportion of positive emotions in extreme emotional reviews greater than 66% and less than 33% as the classification of whether positive polarization and negative polarization occur. When the proportion of positive and negative extreme emotional reviews is similar, it is necessary to analyze specific movies. Generally, the proportion of noun entities in film reviews of films with emotional polarization will show obvious repetition and reduction. Among the 28 movies analyzed in this study, 9 movies have positive emotional polarization, 2 movies have negative emotional polarization, and the remaining 17 movies have similar proportions of positive and negative extreme emotional tendency reviews.

At the same time, combined with the movie ratings, 29.09, 49.65, and 85.98 correspond to the positive emotional tendency ratios of 28.9, 49.6, and 86.4 for the three movies "Nine-story Demon Tower", "The Mermaid", and "Zootopia", respectively. It can also be judged that the emotional tendency of film reviews and movie ratings have a high correlation.

## 3.3 Correlation Between Audience Needs and Emotional Polarity

The data analysis shows that the emotional polarization of the audience is weakly positively correlated with the box office. The viewing needs of Chinese audiences are mainly divided into seven categories: story, role, director/team, special effects, art, actor, and action (Li et al., 2022). Among all the movie data, the two movies with the most obvious positive and negative polarization are Zootopia and My Sunshine. After removing stop words using the stop word list of Harbin Institute of Technology, the word cloud diagram is generated using WorldCloud. In My Sunshine, which has the most extreme negative emotions, the most frequent categories are actor and story, including discussions about the actor's acting skills and comparisons with the plots of the TV series of the same name. In Zootopia, which has the most extreme positive emotions, the most frequent categories are character and story. Through a simple comparison of the two, the story of the movie is a more important factor affecting the audience's emotional tendency than other factors.

Whether it is a movie with positive or negative emotional polarization, the story factor has a large proportion in the discussion. Therefore, it can be considered that there is a strong correlation between the plot of a movie and the polarization of film reviews. When promoting a movie, special attention should be paid to avoid any points that make the audience feel abrupt or even disgusted. At the same time, the keyword review rate in negatively polarized film reviews is more concentrated than that in positively polarized film reviews, which may mean that the outbreak of negative emotional tendencies is also highly uncertain. Once there is an opinion leader at the core of the topic, negative emotions are likely to become polarized (Rabiee et al., 2024). In addition, compared with positively polarized film reviews, negatively polarized film reviews have more discussions about non-film content, such as actors and directors.

# 4 DISCUSSION

## 4.1 Movie Types and Emotional Tendencies

When comparing different types of movies, this study found that love and youth movies are more likely to have audience emotional polarization than other types of movies and tend to be negative, and negative comments are more focused on actors rather than roles than positive comments; the average emotion of the audience of musical movies tends to be positive. Combined with the results of word frequency analysis of film review data, the difference between positive and negative comments in the comments of love and youth movies with strong storylines is more significant, and the audience generally likes elements such as music and dance.

In terms of the content of the discussion, strong story movies are more likely to touch the audience's emotions and memories and are more likely to lead to polarization of audience evaluations. Therefore, when targeting such strong story movies, both in the early marketing and publicity, it is necessary to pay attention to the emotional discussions that such movies may bring and carry out targeted publicity, especially the publicity of actors. Appropriate publicity may bring higher attention to such movies more easily, but inappropriate publicity may also make such movies more likely to receive low ratings.

For musicals and dances, the data of this study shows that the average emotion of the audience of musicals and dances tends to be positive, but there are also obvious problems, such as a single discussion structure and a small number of comments. It may be necessary to pay more attention to other aspects during promotion, especially factors such as plot and actors that are more relevant to the audience's emotions, to attract the audience's attention.

## 4.2 Correlation Between Emotional Polarization and Film Indicators

The box office of movies with obvious audience emotional differentiation in the comment area is often relatively high, which may be because emotional polarization will cause more discussion and attention and then drive the box office to a certain extent. From the word frequency analysis, it can be seen that negative polarization is generally more targeted at specific topics than positive polarization. Therefore, the promotion of movies may be more targeted at some specific topics that are more likely to cause discussion, such as actor selection. Although it may be more likely to cause emotional polarization in the evaluation, the discussion will also increase accordingly and may further improve the box office performance of the movie.

# 5 CONCLUSION

This study explored the characteristics of audience emotional polarization and its related influencing factors by analyzing Chinese film reviews of 28 different types of films. The study showed that there was a certain correlation between the phenomenon of emotional polarization in film reviews and indicators such as the type, box office, and ratings of the film. Specifically, youth and romance films are more likely to cause emotional polarization, while science fiction and musical films tend to have positive emotional evaluations. This finding suggests that the type of film plays an important role in the phenomenon of emotional polarization.

In addition, the study also found that the phenomenon of emotional polarization in film reviews is closely related to the performance of film actors. Negatively emotionally polarized film reviews often focus on discussions about the actors' acting skills or their past reputations, while positive emotional polarization tends to focus on the characters or core plots of the film. Therefore, film production and publicity teams should focus on actor selection and plot design during the planning stage to reduce the risk of negative evaluations and enhance the audience's overall viewing experience.

Although this study conducted a detailed analysis of the phenomenon of emotional polarization in film reviews through natural language processing and statistical methods, the universality of the research conclusions still needs further study due to the limitations of sample size and film types. Future research can further explore the causes and potential impacts of the phenomenon of emotional polarization in film reviews by expanding the sample range, combining more types of film data, and introducing more accurate sentiment analysis models.

In short, the study of the phenomenon of emotional polarization in film reviews not only provides a new perspective on the emotional response of the audience for the film industry, but also provides a useful reference for research in related fields. In the future, with the continuous development of sentiment analysis technology, this research direction will have a wider application value.

# REFERENCES

A. Ullah, S. N. Khan, N. M. Nawi, Review on sentiment analysis for text classification techniques from 2010 to 2021. Multimedia Tools Appl. 82(6), 8137-8193 (2023)

G. Peng, X. Cheng, Research on the formation mechanism of social media group polarization in hot events. J. Inf. Resour. Manag. 13(2), 42-52 (2023)

H. Rabiee, B. T. Ladani, E. Sahafizadeh, A social network model for analysis of public opinion formation process. IEEE Trans. Comput. Soc. Syst. (2024)

J. Yuan, J. Shi, J. Che, C. Xu, J. Wang, Modeling and simulation analysis of public opinion polarization in a dynamic network environment. Concurrency Comput. Pract. Exper. 32(19), e5771 (2020)

L. Guo, Sentiment analysis of domestic film reviews based on deep learning, Ph.D. thesis, Guangxi Normal University (2020)

M. Li, Y. Zhang, S. Wang, Comparative study of Chinese audiences' demand and satisfaction for Chinese and foreign films based on Word2vec and TF-MONO algorithm. Intell. Eng. 8(2), 073-086 (2022)

N. C. Schwertman, M. A. Owens, R. Adnan, A simple more general boxplot method for identifying outliers. Comput. Stat. Data Anal. 47(1), 165-174 (2004)

S. H. Lee, J. Cui, J. W. Kim, Sentiment analysis on movie review through building modified sentiment dictionary by movie genre. J. Intell. Inf. Syst., Korea Intell. Inf. Syst. Soc. (2016)

S. Kumar, K. De, P. P. Roy, Movie recommendation system using sentiment analysis from microblogging data. IEEE Trans. Comput. Soc. Syst. 7(4), 915-923 (2020)

S. Soubraylu, R. Rajalakshmi, Hybrid convolutional bidirectional recurrent neural network based sentiment analysis on movie reviews. Comput. Intell. 37(2), 735-757 (2021)

Utmhikari, Douban movie short comments dataset. Kaggle (2017). Retrieved from https://www.kaggle.com/datasets/utmhikari/doubanmovieshortcomments/data.