Text Information Analysis and Sentiment Analysis -- A Case Study of Barrage Comments on Disney Animated Films

Xi Lan^{1,a,*}

¹School of Digital Economy, Guangdong University of Finance and Economics, Guangzhou, 510320, China a. 571737293@qq.com *corresponding author

Abstract: With the rise of digital media, video barrage—real-time user comments overlaid on videos—has become a prominent feature of online video platforms, particularly in China. This study examines Bilibili, a leading Chinese barrage video-sharing platform, which in 2023 acquired streaming rights to several Disney animated classics, including *Frozen*, *Zootopia*, and *Coco*. By analysing barrage data generated during the viewing of these films, the research aims to uncover valuable insights into users' feedback, emotional engagement, and interactive behaviours. The findings highlight the overall positive emotional resonance among viewers, indicating a favourable reception of the Disney animated films on the platform. The study explores the role of barrage data in enhancing the viewing experience. It provides fresh perspectives for the digital entertainment industry, guiding it towards becoming more user-friendly and innovative. Through in-depth analysis, this research contributes to understanding audience feedback, emotional experiences, and interactive behaviours, transforming individual viewing experiences into communal social events.

Keywords: Barrage Culture, Sentiment Analysis, Audience Interaction, Disney Animation, Digital Entertainment

1. Introduction

In the digital media landscape, video barrage has become a key interaction method, especially on Bilibili, a major Chinese platform celebrated for its barrage culture. With the 2023 acquisition of Disney's *Frozen, Zootopia*, and *Coco*, Bilibili has seen increased viewer engagement, highlighted by the barrage data. Originating in 2009 with a focus on ACG, Bilibili has since included films, music, and technology, becoming a diverse content hub. Its real-time barrage comment system enhances the viewing experience by enabling immediate feedback and shared viewer interactions, thus transforming personal viewing into a social event.

Barrage culture, a unique aspect of Bilibili, is deeply ingrained in China's internet culture. These real-time, on-screen comments, known as "barrage," appear as dynamic text, enhancing the viewing experience with immediate, shareable reactions and fostering a sense of community. This interactive element not only reflects instant responses to content but also embodies a collective cultural dialogue among viewers.

Disney animated films have long been globally beloved. Bilibili's 2023 acquisition of titles like *Frozen*, *Zootopia*, and *Coco* has enriched the platform, enabling users to engage with these films

[@] 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

within its social framework. The combination of Disney's emotional storytelling and Bilibili's interactive barrage system enhances the viewing experience, emphasizing communication and sharing. This setting allows for a communal viewing experience, and analysing the barrage data offers insights into audience feedback and emotional engagement.

While past research has explored the general dynamics of online video platforms, it has often overlooked the specific nuances of Bilibili's barrage culture and its impact on user interactions. This study uniquely focuses on the barrage comments for Disney animated films, which are known for their broad appeal and emotional depth, potentially revealing distinct user sentiments and preferences.

Bilibili's real-time barrage system offers a fresh perspective on audience reactions, an underexplored area that, when combined with Disney's emotionally charged narratives, provides a unique view of viewer engagement. As the digital entertainment sector expands, analysing barrage data becomes increasingly relevant, revealing viewer responses, emotional connections, and market demands. This analysis is crucial for driving industry innovation and enhancing user engagement.

2. Literature Review

2.1. Barrage reflects social culture and emotion

With the rapid rise of the internet and the proliferation of online culture, research surrounding the film industry, online barrage culture, and online public opinion has gained substantial attention in academia. For example, Lin Shilong et al. explored the factors influencing user barrage comment behaviour in online teaching videos using the Elaboration Likelihood Model (ELM), providing a new perspective on how audiences engage with content through real-time barrage comments.[1]

Barrage culture, an emerging social media interaction, is transforming audience engagement with films. Ye Xujie et al.'s research highlights its role in emotional expression within Reaction videos, while Liang Junjian and Long Pingru focus on its impact on collective emotions among youth watching mainstream dramas. [2] [3] These studies underscore the significant social and psychological role of barrage culture in film viewership.

In the field of online public opinion research, Huang Biying and Wu Bing demonstrated how online public sentiment, influenced by barrage emotions, can impact a film's box office performance and reputation.[4] Zhao Pengtao focused on text sentiment analysis of Chinese film barrage, providing new analytical tools for a deeper understanding of audiences' emotional attitudes and recommendations toward films.[5]

Although these studies offer a rich understanding of audience interaction with films and online content, there is still a lack of specific focus on the intersection between film content and barrage culture. This study aims to address that gap by analysing barrage data associated with Disney films *Frozen*, *Zootopia*, and *Coco* by analysing the barrage data of these movies.

2.2. Current application of Text Analysis Research

Text analysis, as a vital branch of Natural Language Processing (NLP), plays a critical role in extracting valuable information from large-scale text data. With the advent of the big data, techniques such as the Bag of Words model, TF-IDF, TextRank, Word2Vec, and BERT have become essential tools for extracting themes, emotions, and keywords.

In her research, Guo Qianqian utilized text analysis techniques including semantic networks, sentiment analysis, similarity analysis, and LDA topic models to deeply evaluate the intrinsic quality of corporate environmental information disclosure.[6] Zhao Junchang employed text analysis methods to calculate the textual similarity between audit reports, further examining the audit quality of the "Big Four" international audit firms.[7] Liu Qingquan improved the traditional TF-IDF algorithm to enhance the accuracy of keyword weight calculation for short texts.[8]

Text analysis has also been applied to analyse social media posts, such as stock-related Weibo data by Wang Peng, who quantified investor sentiment using deep reinforcement learning-based stock trading.[9] Zhong Shuyuan applied natural language processing technology to analyse learner feedback text on MOOC platforms, exploring behaviour patterns of the learner.[10] Zhao Yang proposed a hybrid model-based text analysis method to enhance the accuracy of movie recommendation systems.[11]

These studies demonstrate the growing importance and application of text analysis in various fields. With the continuous advancement of technologies such as deep learning, text analysis technology is expected to achieve significant improvements in accuracy and efficiency, further promoting research and application development in related fields.

2.3. Limitations of current research

The literature reviewed highlights extensive research on the film industry, online barrage culture, and the applications of text analysis, but there remains a gap in integrating these areas. While there has been significant exploration of audience engagement with films and the technologies used for text analysis, limited attention has been paid to how barrage culture specifically affects user interaction and sentiment during film viewing.

Moreover, although text analysis and data mining techniques are increasingly being applied to various text-based data sources, applying these tools to real-time barrage data—often characterized by its fast-paced, informal, and emotionally charged nature—presents unique challenges. Thus, this study aims to fill this gap by using advanced text analysis, web crawling, and data mining techniques to process and analyse the barrage data associated with Disney films *Frozen*, *Zootopia*, and *Coco*. This research will contribute valuable insights into how audience feedback, emotions, and interactive behaviours shape the reception and influence of these films.

3. Barrage Characteristic Analysis and Sentiment Analysis

3.1. Data Acquisition

3.1.1. Data Sources

Given the enduring appeal of Disney animations and Bilibili's acquisition of their streaming rights, this study focuses on the Chinese audience's reception of these films. We've chosen *Zootopia*, *Frozen*, and *Coco* as benchmarks for our analysis. These films have achieved global acclaim and particularly resonate with Chinese viewers. To comprehend audience reactions and the films' social media presence, we'll utilize Bilibili's barrage data as our primary resource. This approach allows us to explore how these films are perceived and discussed within China's digital sphere.

3.1.2. Data Collection

To collect the barrage data for *Zootopia*, *Frozen*, and *Coco* on Bilibili, we utilized Python's web crawling techniques. Using the requests library, we retrieved the XML content containing barrage data, while xml.etree.ElementTree was utilized to parse and extract the text. Altering the video cid values allowed us to gather comments from different films, collecting a total of 3,378 comments for *Frozen*, 8,377 for *Zootopia*, and 3,898 for *Coco*. The extracted data was saved into text files through Python's file I/O operations.

3.2. Data Preprocessing

3.2.1. Tokenization

We used the Jieba tokenizer for segmenting the text into words. Since the barrage dataset for this study is derived from movies that contain specific terminology, such as "Zootopia" in the film *Zootopia*, we customized Jieba's dictionary to recognize these terms during tokenization.

3.2.2. Data Cleaning and Stop word Filtering

After tokenization, the dataset still contained punctuation marks and common, non-meaningful words (e.g., ".", ",", "le"). We utilized the Harbin Institute of Technology (HIT) stop words list to filter these out, opting for HIT due to its thoroughness and broad coverage in the Chinese language, ensuring a more accurate data cleaning process.

3.3. Data Analysis

3.3.1. Word Frequency Analysis

Word frequency analysis, a staple of text analytics, reveals the dominant themes and emotional threads in audience discourse. It pinpoints the most common terms in barrage data, signalling the topics and reactions that capture viewers' attention. This method quantifies audience engagement, highlighting the video aspects that stir the most interest and emotion. By focusing on frequently mentioned elements, we measure the audience's response to specific content, providing a clear gauge of what truly resonates.

3.3.2. Part-of-Speech Analysis

Analysing part-of-speech (POS) helps identify the types of words viewers use, such as nouns, verbs, and adjectives, providing insights into the audience's attitudes and emotions. For instance, a high frequency of adjectives may indicate a strong emotional inclination towards the content, while a prevalence of verbs could suggest interest on specific actions.

3.4. Text Information Processing

3.4.1. Text Keywords and Weights

Analysis of keywords and their weights can reveal the focal points and emotions associated within different films. By identifying the most frequent keywords, we can determine the central topics of user discussions. Additionally, the weights of these keywords allow us to make preliminary judgments about users' emotional inclinations towards certain subjects or terms. In this study, we employed the TF-IDF method to calculate keyword weights, sorted them in descending order of weight, listing the top 10 keywords based on their prominence. Taking *Zootopia* as an example, a sample of these keywords is presented in Table 1.

Text Keywords	Frequency
hahaha	0.438516
cute	0.414527
ahh	0.410688
really	0.321450
fancy	0.164083

3.4.2. Text Word Cloud Generation

To visualizes frequency and importance of keywords, we generated word clouds. Word clouds assist users to comprehending the keywords intuitively by visualizing their respective weights within the text. Through word clouds, important information and key themes the text can be emphasized and captured. In this study, word clouds were created for the barrage data of three movies. Taking the barrage data of *Zootopia* as an example, the word cloud is illustrated in Figure 1.



Figure 1: Word Cloud of Zootopia

Legend: This word cloud visualizes the most frequent terms from the barrage comments for the film *Zootopia* on Bilibili. Key terms and their English translations are as follows:

- 哈哈哈哈 (hahaha): Laughter, indicating amusement or joy.
- 可爱 (ke'ai): Cute, expressing endearment towards characters.
- 朱迪 (Judy): A reference to the main character, Judy Hopps, a bunny.
- 尼克 (Nick): A reference to the main character, Nick Wilde, a fox.
- 动物 (dongwu): Animals, likely referring to the anthropomorphic characters in the film.
- 场面 (changmian): Scene, possibly indicating memorable or significant moments.
- 警察 (jingcha): Police, relating to the film's theme of law enforcement.

3.5. Text Sentiment Calculation

Sentiment analysis is crucial for deciphering user emotions, which is vital for improving user experience and tailoring services. It enables us to grasp attitudes and emotions towards specific content, facilitating more attuned responses. As immediate forms of interaction, barrage comments mirror authentic user sentiments. Analysing these offers a deeper dive into viewers' experiences and reactions, aiding in the refinement of movie content and services. This analysis is instrumental in uncovering user preferences, beneficial for curating movie recommendations and informing social media tactics, thereby boosting user engagement and satisfaction.

In this study, we employed the SnowNLP library for text sentiment calculation. SnowNLP is a Python library for processing Chinese text, offering extensive text analysis and sentiment calculation functionalities. With the SnowNLP library, we conducted sentiment analysis on barrage texts, computed the emotional tendency of the texts, and derived corresponding sentiment scores.

We input the collected barrage data into the SnowNLP library, where its sentiment analysis capabilities calculate the emotions expressed in the comments. SnowNLP adopts advanced natural language processing technologies to accurately identify the emotional tendencies expressed in the texts, including positive, negative, and neutral sentiments.

4. **Results and discussion**

4.1. Barrage Text Analysis

4.1.1. Word Frequency Analysis

The word frequency analysis from *Zootopia*'s barrage comments offers insights into audience reactions and interests. As shown in Table 2, "hahaha" leads with 965 instances, indicating the film's humour resonated well with viewers. "Cute," mentioned 457 times, reflects viewers' fondness for characters or scenes. "Ahh" (432 times) and "really" (428 times) suggest surprise and emphasis, marking emotionally engaging moments. "Fancy" appears 335 times, hinting at appreciation for the film's aesthetics. These findings underscore the film's success in engaging viewers on an emotional level and highlight the aspects of the movie that resonated most with the audience.

Table 2: Top 5 Word Frequencies in Zootopia Barrage Comments

Words	Frequency
hahaha	965
cute	457
ahh	432
really	428
fancy	335

4.1.2. Sentiment Analysis of Barrages

In the sentiment analysis of barrages, we conducted a detailed analysis of the barrage data for *Frozen*, *Coco*, and *Zootopia*. Initially, we observed that the proportion of positive emotions in all three films was relatively high, at 80.7%, 81.5%, and 81.0%, respectively. This indicates that the audience generally experienced positive emotions while watching these films, reflecting a fairly positive emotional resonance among viewers.

Notably, *Coco* stood out in terms of the proportion of positive emotions, reaching 81.5%. This may imply that the film elicited a more profound and positive emotional experience among viewers. In comparison, *Frozen* and *Zootopia* showed slight variations in this regard, but the differences were not significant. To better visualize the sentiment analysis of the barrage data, pie charts were generated for each film using Excel, further illustrating that the proportion of positive barrages in all three films is very high. The pie charts are shown in Figure 2 below.



Figure 2: Comment emotion for animated films. (A) Zootopia (B) Frozen (C) Coco

4.2. Analysis from the Perspective of Film Production

By analysing the sentiment of the barrage comments through Table 3, we conclude that Disney's *Zootopia*, *Frozen*, and *Coco* have all successfully garnered high positive sentiment, with scores ranging from 80.7% to 81.5%. This analysis underscores the films' broad appeal and their ability to elicit favourable emotional responses from viewers, highlighting their resonance within the audience.

Film Title Positive	Positive	Negative	Neutral	Pos.	Neg.	Neu.
	Inegative	Neutral	Comments	Comments	Comments	
Zootopia	0.806878442	0.315696552	0.5	2278	916	184
Frozen	0.815579724	0.312570734	0.5	2693	1036	169
Сосо	0.809985153	0.292842804	0.5	5488	2450	438

 Table 3: Barrage Sentiment Tendencies

This consistency in positive sentiment across the three films suggests a few key points for film production:

Firstly, it implies that the narrative craftsmanship, character development, and emotional storytelling in these films are of high quality and appeal to a broad audience. Secondly, the similar emotional impact of the films may also reflect a shared viewer demographic that appreciates the themes and values presented in Disney's animated features. Lastly, the close sentiment scores could be indicative of a consistent standard of production quality across these films, which is well-received by the audience.

In summary, sentiment analysis of barrage comments shows the three Disney films—*Zootopia*, *Frozen*, and *Coco*—have a similar emotional impact and are well-received, indicating their narratives and production quality meet audience expectations and preferences, key to their success.



Figure 3: Comparison Line Chart of the Proportion of Positive Comments in Three Movies.

5. Conclusion

This thesis provides an in-depth and systematic exploration of the role of barrage data in enhancing the film viewing experience. Through the acquisition, preprocessing, analysis, and sentiment calculation of barrage text data, we have uncovered valuable insights into the audience's focal points, emotional responses, and interactive behaviours. These findings offer practical implications for film producers, Bilibili, and the broader digital media industry.

For film producers, the analysis highlights the importance of crafting narratives and characters that resonate emotionally with viewers. By understanding the emotional language used in barrage comments, producers can fine-tune their storytelling to evoke the desired responses. For Bilibili, these

insights can inform content curation and platform features that enhance user engagement and satisfaction.

The sentiment analysis, facilitated by the SnowNLP library, reveals a rich tapestry of audience emotions, with a notable prevalence of positive sentiments. This suggests that the films have been successful in creating an emotional connection with their audience. The findings also point to the potential for further research to explore the nuances of sentiment across different film genres and the integration of more advanced text analysis techniques to glean deeper insights into audience preferences and behaviours.

This study, while insightful, has its limitations. Concentrating on a handful of Disney animations might not fully represent audience reactions to a broader film range. Future research should consider a more diverse film selection to reveal additional patterns and viewer preferences. Moreover, as language and communication trends on platforms like Bilibili continue to evolve, text analysis techniques will require regular updates to ensure they remain pertinent and precise.

In conclusion, this research has practical value for stakeholders in the film industry and digital media platforms. By analysing barrage data, filmmakers can better grasp audience emotions, fostering more compelling narratives. Platforms can use these findings to boost user experience and engagement. Future research can expand on these results, furthering our grasp of audience sentiment's role in the industry.

References

- [1] Lin Shilong, Lu Ting, Le Chengyi. Analysis of User Barrage Behaviour in Online Educational Videos on Bilibili Based on ELM [J]. Journal of Neijiang Normal University, 2024, 39(02): 44-51+57. DOI: 10.13603/j.cnki.51-1621/z.2024.02.008.
- [2] Ye Xujie, Zhao Yuxiang, Zhang Yan, et al. A Study on the Emotional Response Generation Mechanism of User Barrage Information Interaction Behaviour in Reaction Videos [J]. Journal of Information Resources Management, 2024, 14(02): 104-120. DOI: 10.13365/j.jirm.2024.02.104.
- [3] Liang Junjian, Long Pingru. Internal and External Aspects of Emotional Practice: A Study on Barrage in Mainstream Realist TV Dramas [J]. Modern Communication (Journal of Communication University of China), 2023, 45(05): 93-101. DOI: 10.19997/j.cnki.xdcb.2023.05.010.
- [4] Huang Biying, Wu Bing. The Impact of Barrage Emotional Characteristics on the Popularity of Science and Technology Videos [J]. Information Exploration, 2023, (10): 1-7.
- [5] Zhao Pengtao. A Study on Text Sentiment Analysis Methods Based on Chinese Movie Barrage [D]. Tianjin University of Science and Technology, 2022. DOI: 10.27359/d.cnki.gtqgu.2022.000889.
- [6] Guo Qianqian. A Study on the Evaluation of Environmental Information Disclosure Quality Based on Text Analysis [D]. Jiangxi University of Finance and Economics, 2023. DOI: 10.27175/d.cnki.gjxcu.2023.000447.
- [7] Zhao Junchang. A Re-examination of the Causal Relationship between the "Big Four" International Auditors and Audit Quality Based on Dual Machine Learning and Text Analysis [D]. Lanzhou University, 2022. DOI: 10.27204/d.cnki.glzhu.2022.002656.
- [8] Liu Qingquan. Application of Improved TFIDF Algorithm in Text Analysis [D]. Nanchang University, 2019. DOI: 10.27232/d.cnki.gnchu.2019.000283.
- [9] Wang Peng. Stock Trading Strategy and Implementation Based on Text Analysis and Reinforcement Learning Techniques [D]. Southwestern University of Finance and Economics, 2021. DOI: 10.27412/d.cnki.gxncu.2021.001728.
- [10] Zhong Shuyuan. A Study on Learner Feedback Text Analysis on MOOC Platforms Based on Natural Language Processing Technology [D]. Ningbo University, 2021. DOI: 10.27256/d.cnki.gnbou.2021.001165.
- [11] Zhao Yang. A Study on Movie Recommendation Based on Film Review Text Analysis [D]. Yanshan University, 2022. DOI: 10.27440/d.cnki.gysdu.2022.000459.