Proceedings of 3rd International Conference on Interdisciplinary Humanities and Communication Studies
DOI: 10.54254/2753-7064/52/2024.19111

Popularity Prediction of Bilibili Videos: Analysis of the Top Creators' Contents in 2023

Jian Zheng[1,a,*]

[1]*College of Software, Northeastern University (China), Shenyang, 110000, China*
a. 850314810@qq.com
**corresponding author*

Abstract: With the rise of video content platforms, Bilibili has become a major platform in China. By using data from 600 videos by top creators in 2023, this study examines the factors influencing video popularity on Bilibili. Key engagement metrics, such as bullet comments, likes and shares are analyzed alongside 3,576 comments for sentiment polarity. The analysis reveals a complex relationship between content characteristics and video popularity, providing insights for content optimization. The analysis demonstrates that certain content features, like the number of bullet comments and positive viewer sentiment, have a stronger impact on video popularity. However, other features, such as likes and shares, though reflective of user appreciation, show weaker correlations with actual view counts. This suggests that while multiple content characteristics influence engagement, not all have an equal or direct effect on driving video views.

Keywords: Bilibili, Video popularity, Engagement behavior, Sentiment analysis, Bullet comments.

1. Introduction

Bilibili has grown from an Anime, Comics, and Games (ACG) hub to a leading video-sharing platform in China. Its young user base and interactive features like bullet comments create a unique viewing experience. Content creators, also called uploader, are crucial, driving content production and influencing the broader social media space. The top 100 creators of 2023 highlight the platform's capacity for generating significant user engagement [1].

Video views are key for assessing a creator's influence on Bilibili. High viewership can lead to increased revenue from ads and platform recommendations. Engagement behaviors such as likes, comments, and bullet comments contribute to video popularity. Understanding these interactions helps reveal how content becomes viral on Bilibili [2, 3].

This study examines the content features that most influence video views, with a focus on metrics such as bullet comments, likes, and shares. It also investigates how engagement behaviors contribute to video popularity and explores the role of sentiment in comments. This research is significant as it offers insights into how content creators can optimize their videos to increase engagement and popularity on Bilibili, while also providing a deeper understanding of audience interaction patterns [4].

DOI: 10.54254/2753-7064/52/2024.19111

2. Methodology

2.1. Data Collection

Data were collected from 12 top creators, each contributing 50 videos, totaling 600 samples. The selection of top creators was based on their influence and engagement levels on the platform, ensuring a representative sample of high-performing content. Metrics such as views, bullet comments, likes, and shares were gathered to understand the factors driving video popularity. Additionally, 3,576 comments were collected to analyze sentiment, aiming to explore how viewer reactions contribute to overall video engagement and success [5].

2.2. Data Processing

Fields with characters like “ten thousand” were converted to numerical values. Outliers and missing values were removed to ensure data integrity.

3. Analytical Model

3.1. Model Selection

Several models, including linear regression, decision trees, and random forests, were tested for predicting video views. The random forest model was chosen because it aligns well with the study’s objective of understanding the complex interactions between multiple content features and engagement metrics. Its ability to handle non-linear relationships between variables makes it suitable for analyzing the diverse factors, such as bullet comments, likes, and sentiment that contribute to video popularity, offering a more accurate and robust prediction of view counts [6].

3.2. Features and Normalization

The key features include bullet comments, real-time comments on the video; likes, showing viewer approval; shares, indicating how often the content is distributed; and the positive comment ratio, reflecting the proportion of positive feedback. These features help explain what drives video popularity. The variables were normalized using min-max scaling, which ensures all feature values are transformed to a range of 0 to 1. This process improves model accuracy by preventing features with larger numerical ranges from dominating those with smaller ranges. Feature importance analysis highlighted bullet comments and positive comment ratios as strong predictors of video popularity [7].

3.3. Model Evaluation

Cross-validation yielded an R^2 value of 0.75, meaning the model can explain 75% of the variance in view counts, demonstrating a fairly strong predictive power. However, the model struggled with predicting outliers, likely due to factors such as sudden platform promotions or trending events, which were not captured in the dataset. For example, a video might go viral due to external factors like a sudden celebrity endorsement, which the model, focused on internal metrics, could not anticipate.

4. Analysis and Discussion

Sentiment data was derived using Natural Language Processing (NLP) and sentiment analysis. Of the 3,576 comments analyzed, 98.4% were neutral, 1.34% negative, and 0.25% positive. The classification, based on linguistic patterns and sentiment markers, was used to assess the relationship between viewer sentiment and video engagement. This analysis provides insights into how viewer sentiment might affect video engagement [8].

DOI: 10.54254/2753-7064/52/2024.19111

4.1. Analysis of Video Views

View counts displayed a long-tail distribution, with a few videos accumulating most of the views. The average view count was 163,965, but the median was only 121.65, highlighting this imbalance (Figure 1).

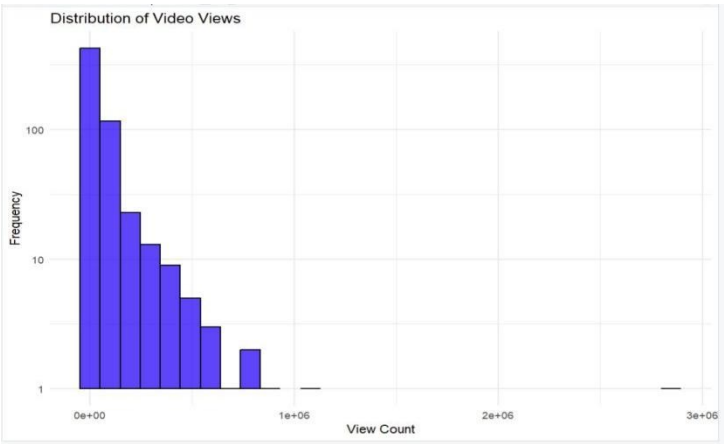

Figure 1: Distribution of Video Views

4.2. Correlation with Engagement Metrics

In Figure 2, bullet comments positively correlated with view counts $r = 0.102$, indicating that higher interaction through comments can boost viewership. Likes and shares showed weaker correlations, suggesting these behaviors reflect user appreciation more than they directly impact popularity.

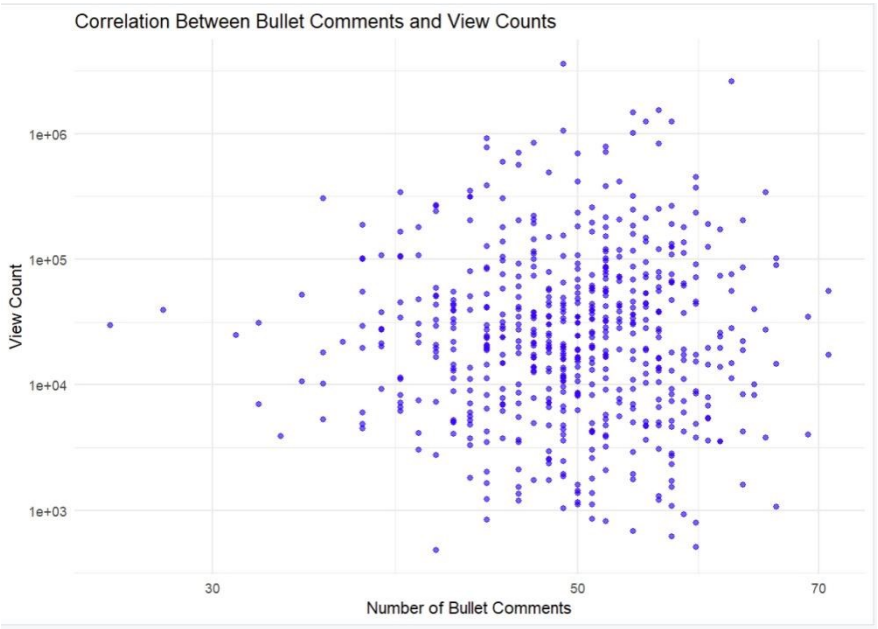

Figure 2: Bullet Comments vs. View Counts

DOI: 10.54254/2753-7064/52/2024.19111

4.3. Impact of Video Duration

First, the duration of each video was collected alongside engagement metrics like comments, likes, and view counts. Correlation analysis was performed to determine the strength of the relationships between video duration and these metrics.

A positive correlation was found between duration and comments ($r = 0.144$), and duration and likes ($r = 0.218$), suggesting that longer videos encourage more viewer interaction in the form of likes and comments.

However, the correlation between duration and view counts ($r = 0.012$) was negligible, meaning that the length of the video had no significant impact on the total number of views.

This analysis indicates that while longer videos may engage viewers more deeply through comments and likes, they do not necessarily drive higher viewership. The stronger interaction with longer videos could be due to the extended time viewers spend engaging with the content, allowing for more opportunities to like or comment. However, video length alone doesn't appear to be a determining factor in attracting more views.

4.4. Sentiment and Engagement

To elaborate on the sentiment analysis process, the comments were first pre-processed by removing irrelevant data such as stop words, punctuation, and special characters. Then, using a Natural Language Processing (NLP) algorithm, each comment was classified as positive, negative, or neutral based on linguistic features and sentiment markers. The correlation between positive comments and engagement metrics such as likes and shares was calculated. Positive comments showed a stronger correlation with engagement, while negative comments, though less frequent, were useful for identifying areas of content improvement [9].

5. Conclusion

This study identified several key factors contributing to video popularity on Bilibili. Among them, bullet comments emerged as a significant driver of engagement, providing real-time interaction that encourages viewers to participate more actively. Additionally, positive sentiment in comments was found to strongly correlate with likes and shares, indicating that fostering a positive viewer atmosphere can directly enhance engagement metrics. These findings underscore the importance of interactive and community-driven content in boosting video performance on the platform [10].

Based on these insights, creators can improve their video performance through specific strategies. First, designing content that encourages bullet comments can enhance real-time viewer interaction, which is essential for sustained engagement. Second, creators should engage with viewers by promoting a positive commenting environment, which is closely tied to increased likes and shares. Finally, producing content with lasting appeal, such as shareable moments or evergreen topics, can encourage viewers to share and save videos, extending the content's reach and relevance over time. This study has several limitations. It focused primarily on top creators, whose strategies and audience interactions may not represent the broader Bilibili ecosystem. Additionally, external factors such as platform promotions and trending topics were not fully considered, which could have affected the results. Future research should expand the sample to include a more diverse range of creators and incorporate external influences. Exploring how different types of content, such as educational or entertainment-focused videos, impact engagement metrics would also provide a deeper understanding of video popularity dynamics.

DOI: 10.54254/2753-7064/52/2024.19111

References

[1] *Jie, X., Yong, Z. (2022) Understanding user engagement on video-sharing platforms: A study on Bilibili. Journal of Social Media Studies, 35: 14–25.*

[2] *Wei, L., Qiang, S. (2021) The impact of interactive features on user engagement in social media videos. Media & Communication Research, 47: 78–89.*

[3] *Yong, C., Xia, L. (2020) Analyzing the role of bullet comments in video platforms: The case of Bilibili. New Media Research, 29: 102–110.*

[4] *Hua, Z., Rui, W. (2019) Video content analysis and prediction models in social media. Computational Social Science Journal, 22: 132–140.*

[5] *Sang, L., Min, K. (2020) Social influence and content engagement on video-sharing platforms. International Journal of Digital Media, 18: 55–68.*

[6] *Zhi, L., Lei, T. (2021) Predicting video virality using machine learning: A case study on Bilibili. Journal of Data Science and Analytics, 9: 44–58.*

[7] *Hao, Z., Jing, M. (2023) Sentiment analysis in user-generated content: Insights from Bilibili comments. Journal of Artificial Intelligence Research, 17: 89–102.*

[8] *Lin, C., Jun, W. (2020) Factors influencing user engagement on video-sharing platforms: A case study on Bilibili. Journal of Digital Media and Society, 15(3), 87–102.*

[9] *Yi, H., Shan, Z. (2021) The role of sentiment analysis in predicting video popularity on social media platforms. International Journal of Data Science and Analytics, 12(1), 45–58.*

[10] *Jian, W., Xin, L. (2022) Exploring the correlation between user sentiment and content performance on Chinese video-sharing platforms. New Media & Society, 24(5), 65–80.*