
AI translation · View original & related papers at
chinaxiv.org/items/chinaxiv-202310.00660

Analysis Methods and Issues of Online Comments in the Big Data Context: A Case Study of the 2019 “Stimulus Battlefield” Game Replacement Event (Postprint)

Authors: Sun Bo, Kong Weijian

Date: 2023-10-08T00:00:00+00:00

Abstract

In the context of big data, online comment data is characterized by massive volume, rich information sources, diverse presentation forms, and complex viewpoint composition, necessitating a more operational analytical framework to proceduralize and standardize online comment analysis. The 2019 replacement of the game ‘Stimulus Battlefield’ with ‘Peace Elite’ sparked intense discussions among netizens. Using this event as a case study and entry point for concrete exploration of online comment analysis methods can help identify and observe the characteristics and problems of current online comment analysis, providing valuable research references.

Full Text

Preamble

Analysis Methods and Issues for Online Comments in the Big Data Era: A Case Study of the 2019 *PUBG Mobile* Game Replacement Incident

Sun Bo, Kong Weijian

Institute of Tea-Horse Road Culture, Yunnan University, Kunming, Yunnan 650000

Abstract: In the big data era, online comment analysis faces challenges of massive data volume, diverse information sources, varied presentation formats, and complex viewpoint composition, necessitating a more operational set of analytical methods to standardize and proceduralize the work. The 2019 replacement of *PUBG Mobile* with *Peacekeeper Elite* triggered intense public discussion. Using this incident as a case study for exploring concrete analytical methods for online

comments helps identify and examine current characteristics and problems in online comment analysis, providing valuable research references.

Keywords: online comments; analysis methods; big data; PUBG Mobile; Peacekeeper Elite

Classification Code: C31

Document Code: A

Article ID: 1671-0134(2022)01-034-05

DOI: 10.19483/j.cnki.11-4653/n.2022.01.008

Citation Format: Sun Bo, Kong Weijian. Analysis methods and issues for online comments in the big data era: A case study of the 2019 PUBG Mobile game replacement incident [J]. *China Media Technology*, 2022(01): 34-37, 41.

Online comments, in essence, represent opinion-based information—evaluative opinions expressed by individuals or organizations on news events, social phenomena, or social issues via online media [1]. To a certain extent, they reflect public perspectives on specific events and phenomena, demonstrate citizens' willingness to participate in public affairs, and provide diverse reference points and observational perspectives for social development.

2. Problems Faced

Through observation of collected information during the analysis period, several primary problems emerge for online comment analysis under big data utilization.

2.1 Massive Data Volume

As exemplified by the “*PUBG Mobile* game replacement” incident, both the total dataset (160,921 entries) and single-platform volumes reach scales of thousands, tens of thousands, or even hundreds of thousands of comments. Previously, when online information sources were relatively limited, researchers primarily relied on portal websites, search engines, well-known forums, and message boards, supplemented by questionnaires and manual retrieval, resulting in relatively manageable workloads. Today, however, the application of web scraping technology and big data information collection systems inevitably confronts us with ever-increasing information volumes and platform carriers, making deduplication and information screening the first tasks required in big data analysis.

2.3 Diverse Presentation Formats

Beyond common text formats, information presentation forms such as images, audio, video, symbols, and even emojis have gained widespread application. In terms of interactive design, behaviors like supporting, opposing, liking, and sharing have become familiar and accepted by netizens. It would be unreasonable to exclude these formats and interactions from the definition of online comments simply because they may not directly express emotions and attitudes. In this

case study, numerous netizens used game screenshots and symbols to express their attitudes, posing challenges to big data collection systems that rely primarily on keyword-based retrieval schemes.

2.4 Expansion and Overlap of Comment Platforms

Some researchers categorize online comments as appearing in “network commentary columns (including op-eds on news sites, personal blogs, etc.), message board replies, and online forums (BBS)” [4], while others classify them from a platform perspective into five types: commentary channel reviews, news follow-up comments, forum comments, WeChat comments, and Weibo comments [5]. However, with the emergence of new platforms such as Weibo, WeChat, news clients, short video platforms, and other social applications, previous platform classifications reveal structural deficiencies. On one hand, if classified by platform type, network news comment platforms continue to expand—for instance, Sina’s big data collection system “Weire” categorizes information sources into 11 types, which appears cumbersome and lacks value for platforms with minimal information volume. On the other hand, from the perspective of platform and comment characteristics, it is difficult to argue that news follow-up comments on a portal website’s desktop version are entirely different from those on its mobile client, or that only comments beneath a Weibo post constitute online comments while the Weibo post itself does not. This demonstrates that online comment platforms are both expanding and overlapping, and cannot be classified solely by platform type in isolation; platform characteristics and other factors must be comprehensively considered.

4. Analysis Process

Addressing the aforementioned problems, we adopted the following analytical approach in studying the “*PUBG Mobile* game replacement” incident.

First, due to the timeliness of online information, influence of national laws and policies, and other uncontrollable factors, relying on a single platform as a data source cannot effectively present netizen attitudes during the analysis period, whether in terms of quantity or quality, nor can a single sampling method be applied. The total participation volume displayed for news articles (Weibo posts, forum threads, answers, etc.) cannot be directly used for analysis, as participation may include clicks, replies, likes, dislikes, forwards, and reports. The volume hierarchy follows this pattern: total participation > total comments > displayable comments > valid comments > actually relevant comments > sampled comments. Analysis sampling can only be conducted at the “displayable comments” level or below.

Second, beyond keyword-based collection, we employed image text recognition and conversion technology, video keyword extraction technology, and other methods to incorporate more diverse information presentation formats into the comment collection scope. Simultaneously, we utilized keyword restrictions, am-

ambiguous character restrictions, field length restrictions, and character format restrictions to exclude non-negative but unsubstantive comment expressions, such as pure emojis, kaomoji, single interjections, blocked words, irrelevant content, and meaningless text.

Finally, comprehensively considering platform characteristics, comment features, and other factors, combined with previous research conclusions and accounting for interactive design elements, we reclassified existing online comment platforms into the following four categories:

First Category: Comprising news comments and follow-up comments from traditional portal websites, media and self-media clients within the “two micros and one client” framework (Weibo, WeChat, and news clients), digital newspaper editions, and WeChat public account articles. Netizens primarily participate through “comment + reply + agree/disagree” mechanisms, with factual reports or articles as the foundation, where viewpoint expression is primarily manifested in the comments themselves—closest to traditional perceptions of online comments.

Second Category: Comprising Weibo and short video platforms within the “two micros and one client” framework. Netizens primarily participate through “Weibo (short video) + comment + forward + like” mechanisms, not necessarily relying on factual report subjects, as the Weibo (short video) itself may constitute a comment with equal status to its attached comments in terms of viewpoint expression. This category features the most diverse comment formats, and due to the forwarding function, frequently drives topic diffusion.

Third Category: Comprehensive and thematic forums such as Baidu Tieba. Netizens primarily participate through “post + reply” mechanisms, which may but need not be based on reports or information, with original posts and replies sharing equal weight in viewpoint expression. These platforms often exhibit conversational characteristics or discussion features, and due to this nature, comment content is relatively loosely structured.

Fourth Category: Professional Q&A communities and industry-specific Q&A platforms such as Zhihu. Netizens primarily participate through “question + answer + comment + agree/disagree” mechanisms, posing questions about specific topics with certain professionalism and detail. Viewpoint expression focuses more on answers, and since respondents are sometimes industry insiders or practitioners, these platforms often yield in-depth content rarely seen on other comment platforms.

Based on this reclassification, we applied different sampling methods to the four platform categories to approximate actual netizen viewpoints as closely as possible.

For the first category, due to significant variation in comment quantities across factual reports and articles, we employed quota sampling. Focusing on key factual reports and articles central to the incident’s public opinion dissemina-

tion, after deduplication assessment, we aggregated comments under articles on the same theme and selected 20 news items with the highest participation and interaction volumes. These were divided into four intervals based on total participation volume: 1-29 comments, 30-299 comments, 300-2,999 comments, and 3,000-9,999 comments, followed by statistical screening and sampling to obtain 300 valid comment samples.

Comment Quota and Sampling Distribution (First Category Platform)

Statistical analysis revealed eight categories of viewpoints and attitudes among netizen comments on this platform.

Online Comment Attitudes and Proportions (First Category Platform)

For the second category, using Sina Weibo as an example, we primarily employed judgment sampling. With total Weibo participation in the hundreds of thousands, we selected daily original popular Weibo posts under relevant discussions during the analysis period, using 800 popular Weibo posts as the sampling frame and selecting 200 valid samples based on interactive comment and forward volumes as reference standards.

Weibo Volume Trend Table (Second Category Platform)

Seven categories of viewpoints and attitudes emerged among netizen comments on this platform.

Online Comment Attitudes and Proportions (Second Category Platform)

For the third category, using Baidu Tieba as an example, we employed simple random sampling due to the equal weight of original posts and replies. The platform generated 3,649 relevant information entries during the period, with 300 valid samples selected.

[Figure 1: see original paper] Baidu Tieba Volume Trend Chart (Third Category Platform)

Seven categories of viewpoints and attitudes emerged among netizen comments on this platform.

Online Comment Attitudes and Proportions (Third Category Platform)

For the fourth category, using the Zhihu community as an example, we employed judgment sampling since viewpoints focus on answers. With 3,906 relevant entries on Zhihu, after deduplication and merging, we selected the 10 most relevant and highest-answered questions, extracting the 30 most upvoted answers for each question to meet large-sample standards, yielding 300 samples.

[Figure 2: see original paper] Zhihu Community Volume Trend Chart (Fourth Category Platform)

Eight categories of viewpoints and attitudes emerged among netizen comments on this platform.

Online Comment Attitudes and Proportions (Fourth Category Platform)

5. Analysis Conclusions

Overall, netizen comments across the four platform categories demonstrate certain commonalities while mutually validating the reliability of viewpoints. On nearly all platforms, a substantial user base considered *Peacekeeper Elite* to have poor gaming experience (e.g., graphics quality, weapon feel, action performance, voice guidance) and unsatisfactory game settings (e.g., loot box mechanics, waving goodbye gestures, blood color, victory mechanics). Compared to *PUBG Mobile*, many players expressed disappointment and dissatisfaction, engaged in sarcastic criticism, or chose to abandon the game and switch to international servers. Generally, no more than 40% of players held positive or neutral attitudes toward the game.

Simultaneously, each platform exhibited distinct structural characteristics in its comment composition. The first category, closest to traditional online comments, featured direct attitude expression that often escalated to attacks and sarcasm toward Tencent and Photon Studios. For instance, 30% of netizens expressed dissatisfaction with Tencent or Photon, believing *Peacekeeper Elite* to be a money-grabbing product (17%) with considerable sarcastic 吐槽 (13%). Comments on this platform were typically short and straightforward, easily generating and amplifying negative emotions, but suffered from poor user persistence and attention, making opinion leader formation difficult. Currently, the five traditional major portal websites (Sohu, NetEase, Sina, Tencent, Phoenix) display significantly fewer comments than actual participation volumes due to various factors, with netizen focus gradually shifting to self-media client comments represented by platforms like Toutiao.

Weibo, as the second category representative, presented diverse content with considerable 调侃 and 吐槽. Users were willing to share personal experiences and game-related peripherals (e.g., events, interactions, videos, moods) more than on other platforms, while showing the lowest abandonment rate, indicating relatively strong user stickiness. Various posts seeking teammates and friends for gaming were common, reflecting strong desire for exposure. Over half of users on this platform recalled and missed their gaming experiences, hoping for *PUBG Mobile*'s return (6%), while nearly one-tenth (9%) didn't care about game changes as long as they could play with companions.

Baidu Tieba, as the third category platform, combined characteristics of the first and second categories, featuring both straightforward attitudes and objective, rational thinking. Both original posts and replies could either output viewpoints on the game replacement or not, occasionally resulting in scattered and contradictory opinions. Due to loosely structured content, sustained discussion around a single topic with consistent viewpoint output was rare. However, players often sought help on Tieba when encountering game problems or having suggestions, resulting in numerous help and suggestion posts—though rarely

reaching professional Q&A levels, with 8% of players providing feedback and suggestions on game issues.

Zhihu community, as the fourth category platform, featured relatively high user sophistication, with users willing to elevate discussions to institutional and environmental levels, often seeking multi-angle, multi-causal perspectives. For instance, 5% of users on this platform attributed the poor gaming environment to combined effects of game companies, players, and government, requiring objective perspectives. Some users viewed *Peacekeeper Elite*'s emergence as an industry exploration and trial, identifying commercial capital drivers behind profit motives—representing profound viewpoints rarely seen on other platforms. Consequently, negative evaluations on this platform were often detailed and difficult to refute, though the relatively high user threshold resulted in a more niche audience with particular reference value for certain community members.

6. Limitations and Discussion

This analysis method for online comments exhibits several primary limitations:

First, comment sampling methods require further research. Ideally, exhaustive collection of all comment information followed by classification and quantitative presentation would yield the most accurate and reliable conclusions. However, this is impractical due to massive data volume, lacks operational feasibility, and is unnecessary. How to sample comment data becomes the primary consideration. Due to redundant comments, using simple random sampling or systematic sampling alone often fails to accurately represent opinion distribution. Our approach employed judgment sampling and quota sampling based on platform characteristics, focusing particularly on hot, key, and highly interactive comments. However, this involves obvious manual selection 痕迹, and issues such as whether judgment criteria, quota distribution, and sampling ratios are appropriate, as well as how to stratify in future incidents requiring stratified sampling or how to define clusters in cluster sampling, all require further research to develop more reasonable sampling methods.

Second, collection of non-textual comments urgently needs resolution. For the increasingly prevalent non-textual comment formats such as images, videos, and audio, although this study converted some into text through technical means, most such comments cannot yet be accurately collected and 统计 due to technical and resource limitations, resulting in comment information loss. Particularly with the rise of short video applications, short videos carrying opinions have become a non-negligible source of online comments, making the capture of video and other non-textual content an urgent problem for future online comment analysis in the big data era.

Third, the relevance of interactive participation requires quantification. Under big data algorithms, comments with high interactive participation are often identified as hot or key comments and recommended to homepages, top positions, or priority display. In this study, high likes, forwards, and replies served

as primary criteria for judgment sampling, under the assumption that they represent majority opinions. However, interactive volume only indicates emotional trends, not necessarily clear opinions or suggestions. The extent to which like and forward counts represent netizen viewpoints, or how to determine the representativeness of key comments through these metrics, requires more explicit quantitative algorithms or more reliable judgment criteria.

Fourth, heterogeneous factors such as user demographics should be incorporated into analysis. This study relied primarily on platform and comment characteristics, lacking analysis of netizens themselves. In reality, user research constitutes an important component of online comment analysis. Stratified profiling of netizens based on age, social class, education level, and internet usage habits can help explore the formation mechanisms of online comments from a macro perspective. User research and comment research are interconnected and mutually reinforcing, providing more fundamental understanding and revelation of online public opinion characteristics.

In summary, online comment analysis in today's big data era faces a new landscape: dramatically increased information volume, rich platform sources, diverse presentation formats, and complex viewpoint composition. Addressing this new situation requires continuous methodological refinement in online comment analysis.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv – Machine translation. Verify with original.