Thematic Analysis and Sentiment Evolution of Online Public Opinion on Hangzhou Asian Games: The Case of Sina Weibo

Wen Che
spe_chew@ujn.edu.cn
University of Jinan, China

Mingze Gao
University of Jinan, China

Shenghua Qi
University of Jinan, China

Article

Keywords: Hangzhou Asian Games, online public opinion, sentiment evolution, theme analysis, cross-dimensionality

Posted Date: July 15th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4598569/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: (Not answered)
Thematic Analysis and Sentiment Evolution of Online Public Opinion on Hangzhou Asian Games: The Case of Sina Weibo

Abstract: It is of great significance to study the emotional evolution characteristics of significant sports events, to understand and master the changing law of online public opinion topics, to increase the attention and influence of sports events, to promote the dissemination of sports culture, and to correctly guide the direction of online public opinion on sports events. By dividing the life cycle of "Hangzhou Asian Games" microblog topics, using the LDA topic model and SnowNLP sentiment analysis, we mined the topics of online public opinion, analyzed the characteristics of public sentiment, and portrayed the sentiment evolution of "Hangzhou Asian Games" microblog topics in three dimensions: time, space and gender. The results show that Sina Weibo users are mainly optimistic about the topic of "Hangzhou Asian Games." The economically developed provinces in the southern and eastern regions are more optimistic about the topic than the economically less developed provinces in the western and northeastern regions, with higher user participation. In each life cycle, the emotional focus and emotional tendency show positive emotion > neutral emotion > negative emotion, and there are differences in the public opinion attention and emotional state of different topics in different periods, regions, and genders. Compared with other public opinion events, the emotional change of sports events public opinion is distinctive and unique, and its emotional evolution process is closely related to the development of different themes at different phases of sports events.

Keywords: Hangzhou Asian Games; online public opinion; sentiment evolution; theme analysis; cross-dimensionality

Introduction

In the context of globalization and informatization, the Internet has become an essential place for people to obtain information, express their views, and communicate, especially in sports events, where Internet public opinion plays an indispensable role in understanding the public's spectator behaviour, shaping the image of the event, and optimizing the sports culture environment. Therefore, an in-depth analysis of the trends and characteristics of the application of online communication in sports events has far-reaching theoretical value and practical significance for multiple stakeholders such as event operators, spectators, and policymakers. The Asian Games is a comprehensive game organized by the Olympic Council of Asia and is one of Asia's largest and most influential sports events (Hao Hua 2018). Since the first Asian Games was held in New Delhi, India, in 1951, it has become an essential phase for Asian countries to show their sports strength and enhance their national image. In this 19th Asian Games in Hangzhou, the scale of the event, the participating teams, and the event's influence have reached an unprecedented height. Therefore, it is of great significance to take the 19th Hangzhou Asian Games as an example to study the evolution law of its online public opinion sentiment. Sina Weibo, the world's largest Chinese social media platform with massive users and active user participation, has become essential for people to obtain information, express their views, and exchange opinions (Zhu 2019). Public opinion on Weibo can directly or indirectly affect the athletes' performance and even the operation of the whole event.

Therefore, this study aims to investigate the sentiment evolution characteristics and topic change rules of the 19th Asian Games in Hangzhou online public opinion. Using Python crawler technology to obtain the data information on the topic "Hangzhou Asian Games" on the Sina Weibo platform, this study shows the public
sentiment change process in three dimensions of time, space, and gender based on the double integration of theme and emotion in each phase of the online public opinion of the 19th Asian Games in Hangzhou with the help of LDA theme model and SnowNLP sentiment analysis. Public sentiment change process. The significance of this paper is to reveal the evolutionary characteristics of emotional expression in the online ecosystem of sports events, enrich the theories and methods of sports event public opinion analysis, and provide a comprehensive and systematic understanding of the public's views on different phases and items of sports events, to provide a basis for decision-making.

Methodology

LDA Topic Modelling. Since many problems arise when extracting the required information from vast amounts of data in unstructured form (Madaan et al. 2018), topic models were introduced in text mining (Blei et al. 2003). Topic models have become one of the most developed and applied models in text classification and information retrieval (Blei 2012). As an advantage of topic modelling, large data corpora can interpret the underlying semantics more comprehensively and consistently (Arora et al. 2012). The most popular approach in topic modelling is the Latent Dirichlet Allocation LDA introduced by Blei and Jordan (Blei et al. 2001; Jelodar et al. 2019); LDA is interpreted as a generative probabilistic model based on a hierarchical Bayesian analysis of text to find the semantic structure of a corpus (Anandkumar et al. 2012). It has good stability in dealing with large-scale data because it gives parameters that are commonly used techniques in stochastic variational sentiment analysis and literature survey (Girolami and Kabán 2003; Lu et al. 2011), it has more comprehensive assumptions about the generation of text compared to other methods (Liu 2013), and it is widely used in the fields of Twitter Trend Data Analysis (Arun et al. 2010; Hidayatullah and Ma’arif 2017), Internet Commentary Analysis (Joung and Kim 2021), Agriculture (Li et al. 2020), Software Engineering (Junruo et al. 2020), Environment (Shah and Shah 2020), Deep Learning (Wang and Xu 2018), and Healthcare (Yun et al. 2020), so this study uses The effectiveness of LDA depends mainly on the "number of topics" (Hoffman et al. 2010), and the perplexity (Kobayashi 2014) and Coherence (Huang 2019) are common indicators to assess the number of topics, and related studies have also tested the reliability of perplexity and Coherence (Fang et al. 2016; Neishabouri and Desmarais 2020), so this study will use the perplexity and Coherence to assess the number of topics in LDA.

In the LDA model, data instances are considered to be generated by potential processes which depend on hidden variables. The dependencies of the latent generation process are shown in Fig. 1. The topic assignment \( Z_{d,n} \) depends on the topic proportion \( \theta_d \) of each document, and \( \theta_d \) depends on the prior probability hyperparameter \( \alpha \). The word \( W_{d,n} \) depends on the topic assignment \( \beta_k \) and \( \beta_k \) depends on the hyperparameter \( \eta \) (Newman et al. 2007). The joint probability distributions (hidden variables) obtained from Eq. (1) are modelled.
The notation used in Equation (1) is as follows: the subject is $\beta_{1:K}$, where each $\beta_k$ is the probability distribution of the word. The topic distribution of $d$th documents is $\theta_d$, where $\theta_1:D$ is the topic probability distribution of all $D$ documents. $z_{1:D,1:N}$ is the topic assignment of $N$ words in each $D$ document. The words extracted from each $D$ document are $W_{1:D,1:N}$. This paper uses $t$ as the candidate value of the best $K$. The posterior probability is shown in Equation (2), where $W_{1:D,1:N}$ denotes all observed words in all documents.

$$p(\beta_{1:K}, \theta_{1:D}, Z_{1:D,1:N}, W_{1:D,1:N}) = \prod_{k=1}^{K} p(\beta_k | \eta) \prod_{d=1}^{D} p(\theta_d | \alpha) = \left( \prod_{n=1}^{N} p(Z_{d,n} | \theta_d) p(W_{d,n} | \beta_{1:K}, Z_{d,n}) \right)$$  \hspace{1cm} (1)

The topic distribution of $d_{th}$ documents is $\theta_d$, where $\theta_{1:D}$ is the topic probability distribution of all $D$ documents. $z_{1:D,1:N}$ is the topic assignment of $N$ words in each $D$ document. The words extracted from each $D$ document are $W_{1:D,1:N}$. This paper uses $t$ as the candidate value of the best $K$. The posterior probability is shown in Equation (2), where $W_{1:D,1:N}$ denotes all observed words in all documents.

$$p(\beta_{1:K}, \theta_{1:D}, Z_{1:D,1:N} | W_{1:D,1:N}) = \frac{p(\beta_{1:K}, \theta_{1:D}, Z_{1:D,1:N}, W_{1:D,1:N})}{p(W_{1:D,1:N})}$$  \hspace{1cm} (2)

Perplexity: A metric used to evaluate the strengths and weaknesses of a language model. LDA is used to find the optimal number of topics, and a smaller perplexity means that the model is a better predictor of new text. Hence, the perplexity generally decreases as the number of potential topics increases(Ru 2022). Equation is (3): For a test set consisting of $M$ documents, the perplexity ($P$) is defined as follows(Wang et al. 2019). Where $p(w_d)$ is the probability of an observed word in document $d$. $N_d$ is the total number of words in document $d$.

$$p = \exp \left\{ \frac{-\sum_{d=1}^{M} \log p(w_d)}{\sum_{d=1}^{M} N_d} \right\}$$  \hspace{1cm} (3)

When the number of topics is high, the generated model tends to overfit, so you must rely on more than just perplexity to judge a model. We will combine this with changes in topic coherence to determine(Dewangan et al. 2020).

Coherence: Topic Coherence is an essential metric for assessing the quality of generated topics, which aims to ensure that the words in a topic are logically related. A higher coherence score usually means the topic is more meaningful(Röder et al. 2015). The calculation of topic coherence consists of the following steps: 1) Topic word selection: for each topic, the top $N$ words are selected. The value of $N$ is usually taken as 10, 20, or some other similar number. 2) Word pair construction: A collection of word pairs is constructed for each topic. For example, if the top $N$ words of a topic are, then the set of word pairs is.3) Calculate the coherence score of word
pairs: for each pair of words, calculate their coherence score. The calculation methods are Umass, Cv, and Uci.

Where. Umass is commonly used; its formula is (4)(5). Where $D(w_i)$ is the frequency of documents containing the word $w_i$, $D(w_i, w_j)$ is the frequency of documents containing both $w_i$ and $w_j$, and $D$ is the total number of documents in the corpus. 4) Average coherence scores: For each topic, the coherence scores of all its word pairs are averaged to obtain an overall coherence score. 5) Evaluating all topics:

Repeat the above steps to compute coherence scores for all topics in the document set coherence score. Finally, the average of the coherence scores for all topics is taken to assess the quality of the entire LDA model.

$$C = \sum_{ij} score_{Umass}(w_i, w_j)$$

(4)

$$score_{Umass}(w_i, w_j) = \log \frac{D(w_i, w_j) + 1}{D(w_j)}$$

(5)

SnowNLP Sentiment Analysis. SnowNLP is a class library written in Python that can quickly process Chinese text content; it is a natural language processing library for Chinese text(Pan et al. 2021). It is suitable for learning natural language processing or as an essential tool. It has the function of word segmentation, lexical annotation, sentiment analysis. Its advantage is that it can accurately capture the sentiment tendency in microblog texts, which provides a window for this study to gain an in-depth understanding of public sentiment attitudes and their changing trends. Therefore, in this study, SnowNLP is used to analyze the sentiment of microblog texts. Firstly, the text is preprocessed, and the sentiment analysis of each tweet is carried out using SnowNLP’s sentiments method, which will return a number between 0 and 1, with 0.5 as the cut-off point, more significant than being positive and less than being negative, and carry out the calculation of the percentage of comments on both sides. The fen denotes the sentiment tendency of the text (0 for negative and 1 for positive)(Chen et al. 2018). By analyzing the affective tendency of each microblog, we can understand the public's affective attitude towards the Asian Games and its changing trend, for example, whether the public's affectivity shifted after the occurrence of a particular event, the magnitude of the shift, and the change of positive and negative affectivity.

Research Framework Design. This paper takes the Hangzhou Asian Games as the research background and constructs the research framework (Fig.2). The research framework is divided into three parts:
Fig. 2 Research framework for public opinion evolution analysis

(1) Obtain blog posts and comment data on the topic of the Hangzhou Asian Games. According to the theory of the life cycle and the trend of the "Hangzhou Asian Games" topic given by Sina Weibo, we divided the public opinion cycle. (2) Apply the LDA topic model algorithm to thematically cluster the microblog text in different phases of the Asian Games public opinion evolution and extract the main discussion topics in each phase. (3) With the help of sentiment analysis methods, we analyze the sentiment characteristics and evolution laws of the Hangzhou Asian Games in each phase from three perspectives: time dimension, space dimension and gender dimension. In order to fine-grain multi-classification emotion label recognition and achieve multi-emotion classification, this paper refers to the emotion vocabulary ontology library of the Dalian University of Technology, which is used for text emotion classification (Zhang et al. 2018). Meanwhile, in order to better present the spatial dimension emotion evolution and gender dimension emotion tendency difference, SnowNLP is used to calculate the text emotion value.

**Data acquisition and preprocessing.** (1) Data collection. In this paper, we selected the Sina Weibo platform, set
the search keyword as "Hangzhou Asian Games", and set the time interval as "0:00 on 18 September 2023 - 23:00 on 15 October 2023", and we wrote a crawler to retrieve data through python software, and obtained a total of 282039 data. The content of the microblog post was crawled by a crawler written in Python software, and 282,039 pieces of data were obtained. In order to further obtain more data attributes, the information of the commenter of each comment is crawled twice, and the rest of the information of the commenter's microblog homepage is captured to obtain the "User ID" "IP territory" "Gender" "Province" "Nickname" "Posting source" "Posting time" "Text content" "Retweets " "Likes" "Comments" "Access time" and "Page address". (2) Data preprocessing. Delete the duplicated and missing serious data and invalid and irrelevant data from the data. Remove HTML tags, URLs, special symbols and other non-text information in the text to make it easy to analyze, and finally obtain 112989 valid texts. The deactivation thesaurus mainly includes the Chinese deactivation thesaurus of Renmin University of China, the deactivation thesaurus of Harvard University, the deactivation thesaurus of Baidu, the deactivation thesaurus of Machine Intelligence Laboratory of Sichuan University, and these four mainstream deactivation Chinese deactivation thesaurus are summarised, de-emphasized and sorted into 2,314 deactivated words and Chinese text lexical segmentation is carried out using the tool jieba(Yuan et al. 2020).

Results

Life cycle segmentation. The life cycle of online public opinion refers to the process of online public opinion from its emergence, development, and climax to decline, which can help us understand the dynamic characteristics and changing laws of online public opinion(Liu et al. 2021; Gao et al. 2021). Life Cycle Delineation The life cycle theory is mainly used to delineate the phases of online public opinion. This paper adopts the four-phase model of public opinion communication delineated by Moreover, it considers the recurrent nature of the development of public opinion on the Hangzhou Asian Games and divides the life cycle of public opinion into four phases.

Interaction volume is often an essential indicator of content popularity and user engagement on social media platforms. As one of China's largest social media platforms, Weibo's interaction volume is often used to analyze content and user behaviour. Interaction volume is the total number of interactions between microbloggers with #topic# tweets and microblogs, which includes the sum of the number of retweets, comments and likes of microblogs. The formula is as follows:

Engagement = number of blog posts + (retweets + comments + likes)

Each type of interaction reflects different levels of user engagement and concern. A high interaction volume usually implies that public opinion has a more incredible speed and breadth of dissemination on the platform. By observing the change in interaction volume, we can reveal the life cycle of the topic. Therefore, the trend of interaction volume of the topic "Hangzhou Asian Games" given by Sina Weibo (Fig.3) is chosen as the basis.

Fig. 3 Trend of "Hangzhou Asian Games" Topic Interactions

Fig.3 demonstrates that the development of public opinion on the Hangzhou Asian Games aligns with the life cycle theory. From 18 September to 21 September 2023, there is a small growth in the data. This phase is the preheating phase, which is mainly the phase of preparation and preview of the event, and public opinion mainly focuses on the preparations for the event, the participating teams, and the competition items. From 22
September - 2 October 2023 is the dawn phase; on 22 September, the network public opinion showed explosive growth; according to previous research, the network public opinion tends to be explosive growth after the opening ceremony of the event, but the opening ceremony of the Hangzhou Asian Games on 23 September 2023, this is because the Asian Games by the customary arrangements for some of the items in advance of the start of the competition. Some sports have a long competition time, such as football matches, often a long schedule; the system is mainly for the first group phase and then the game. So in the 22 September Hangzhou Asian Games women’s football group A match, the Chinese women’s football team 16-0 Mongolia women’s football team rushed to the hot search, so that the network public opinion in the opening ceremony before the phenomenon of discussion and attention to the rapid rise of the phenomenon.2023 3 October - 8 October is the climax phase, the event in the middle to late when the competition of various competitions is intense, the attention reaches its peak. The quantity and quality of online public opinion will peak in this phase. 9 October 2023 - 15 October 2023 is the recession phase, usually after the tournament. The attention of online public opinion will gradually decrease, but there will still be some people discussing the review and impact of the tournament. By dividing and analyzing the life cycle of online public opinion, the dynamic characteristics and changing laws of online public opinion can be better understood and grasped, which is of great significance for tournament organizers and relevant departments to do an excellent job in managing public opinion, responding to various situations promptly, and actively guiding the direction of public opinion. At the same time, different public opinion response strategies and measures can be adopted for different phases of online public opinion to effectively manage and guide online public opinion and create a good atmosphere for the event.

Number of Themes Selection. According to the coincidence score and perplexity to determine the number of topics of the LDA model, i.e., the perplexity degree value reflects the probability of belonging to a document under a single topic, the lower the value of the model describes the better the effect, according to the hand-time method, when there is a significant inflexion point, then it is judged to be the best number of topics. However, when the number of topics is vast, the generated model tends to overfit, so you can not rely solely on the degree of perplexity to judge whether a model is good or bad; it needs to be combined with the coherence of the trend, that is, the coherence of the model reflects the degree of fit to the overall set of documents, the higher the value of the model the better the results. In summary, it is necessary to combine the degree of perplexity, coherence trend, and actual needs (Fig.4) and, finally, to determine the optimal number of topics.
As seen in Fig.4, the number of phase 1 themes was determined to be 2, phase 2 themes to be 4, phase 3 themes to be 6 and phase 4 themes to be 4.

**LDA Theme Evolution.** In this study, the LDA topic model was applied to model the topics of the microblog texts in different communication phases of the Asian Games. Specifically, the LDA algorithm was used to extract the high-frequency feature words under each phase, which were taken as the representative words for each theme. Based on this, the change patterns of the main discussion themes at each phase during the evolution of public opinion were comparatively analyzed (Table 1)
<table>
<thead>
<tr>
<th>Subject No.</th>
<th>Thematic identifiers (partial)</th>
<th>Thematic overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic I - 1</td>
<td>Hangzhou (0.007), Asian Games (0.006), Asian Games (0.006), China (0.005), Countdown (0.003), Events (0.003), Hangzhou Asian Games Opening Ceremony (0.003), Competition (0.003), Opening Ceremony (0.002), Gold Medals (0.002)</td>
<td>Countdown to Hangzhou Asian Games Opening Ceremony</td>
</tr>
<tr>
<td>Topic I - 2</td>
<td>Go for it (0.018), Sun Yingsha (0.017), Luo Yunxi (0.014), Fan Zhendong (0.009), Gong Jun (0.005), Hangzhou Asian Games (0.005), Expectation (0.005), Wang Yibo (0.005), Li Wenhan (0.004), Torchbearer (0.004)</td>
<td>China’s sports flame passes on as Hangzhou Asian Games torch continues to carry on</td>
</tr>
<tr>
<td>Topic II - 1</td>
<td>Hangzhou Asian Games Opening Ceremony (0.006), Athletes (0.004), Motherland (0.003), Scene (0.003), China (0.002), Opening Ceremony (0.002), Sportsmen (0.002), Motherland (0.003), Happiness (0.003), Sportsmen (0.004)</td>
<td>Hangzhou Asian Games Opening Ceremony Shows China’s Charm and Strength</td>
</tr>
<tr>
<td>Topic II - 2</td>
<td>Wang Yibo (0.054), Breakdancing (0.040), Promo (0.022), Justified Burning (0.019), Asian Games (0.012), Sportsmen (0.008), Expectation (0.008), Events (0.008), Hangzhou Asian Games (0.007), Blessing (0.006)</td>
<td>Celebrity artists promote the Asian Games and cheer on the athletes.</td>
</tr>
<tr>
<td>Topic II - 3</td>
<td>Project (0.006), Hangzhou (0.006), Tower of Blades (0.005), China (0.005), Asian Games (0.005), eSports (0.004), Winning the Championship (0.004), Gold Medal (0.004), Asian Games (0.003), Yanny Wu (0.003)</td>
<td>eSports makes Asian Games debut to much fanfare.</td>
</tr>
<tr>
<td>Topic II - 4</td>
<td>Wang Chuqin (0.019), Sun Yingsha (0.019), Fan Zhendong (0.014), Final (0.013), Men’s Singles (0.010), Congratulations (0.010), Table Tennis (0.010), Champion (0.010), China (0.010), Go for it (0.008)</td>
<td>China’s table tennis team wins multiple gold medals in another success story.</td>
</tr>
<tr>
<td>Topic III - 1</td>
<td>Gold Medal (0.012), China (0.010), Final (0.008), China (0.008), Championship (0.008), China Women’s Basketball (0.008), Congratulations (0.007), Women’s Basketball (0.006), Asian Games (0.006), Women (0.006)</td>
<td>China’s women’s diving, basketball and dragon boat teams won titles at the Asian Games, showing the clanging rose Jiniang blooming.</td>
</tr>
<tr>
<td>Topic III - 2</td>
<td>China Men’s Basketball (0.006), Men’s Basketball (0.006), Men’s Basketball (0.005), Pan Pan (0.005), Hockey (0.004), Yao Ming (0.004), Council (0.004), Philippines (0.003), Food (0.003), Upper Field (0.003), Mum (0.002)</td>
<td>China’s men’s basketball team falls to Philippines, sparking discontent among spectators.</td>
</tr>
<tr>
<td>Topic III - 3</td>
<td>Badminton (0.012), Combined (0.011), Chen Yufei (0.010), Hangzhou Asian Games (0.008), Li Shifeng (0.008), Yasi (0.007), Shi Yuqi (0.007), Xiao Zhan (0.005), Siwei (0.005), Men’s Singles (0.005)</td>
<td>The Asian Games Chinese badminton team won gold medals in mixed doubles, men’s singles and women’s doubles,</td>
</tr>
<tr>
<td>Topic III - 4</td>
<td>Memorial (0.009), Successful (0.008), Conspicuous (0.006), Ha-ha-ha (0.006), Burning (0.005), Mascot (0.004), Lovely (0.004), Gold (0.004), Raman (0.004), Devotee (0.004)</td>
<td>Cute Asian Games mascot souvenir sparks internet users’ interest.</td>
</tr>
<tr>
<td>Topic III - 5</td>
<td>Goodbye (0.016), Future (0.012), Promotional Song (0.007), Moving (0.006), Aida Future (0.005), Dynamic (0.004), Kaburat (0.004), Moving (0.003), Theme (0.003), Official (0.003)</td>
<td>Zhou Shen’s emotional performance of the Hangzhou Asian Games promotional song “Aida Future” moved the crowd.</td>
</tr>
<tr>
<td>Topic III - 6</td>
<td>Closing Ceremony (0.020), Farewell (0.012), Memories of Jiangnan (0.012), Closing (0.011), Hangzhou (0.010), Asian Games (0.009), Chinese Women’s Volleyball Team (0.008), Asian Games (0.007), Numbers (0.005), Promotional Film (0.005)</td>
<td>Hangzhou Asian Games Closing Ceremony Memories of the South, the closing ceremony to “the most remembered is Hangzhou” as the overall theme.</td>
</tr>
<tr>
<td>Topic IV - 1</td>
<td>Xiao Zhan (0.084), Dream (0.027), Dream (0.022), Outline (0.022), Public Welfare (0.019), Brush (0.018), Amb</td>
<td>Xiao Zhaoshan as Asian Games Dream’s Charity Ambassador</td>
</tr>
</tbody>
</table>
As can be seen from Table 1, the preheating phase: 1) The Countdown to the opening ceremony of the Hangzhou Asian Games: In May 2022, affected by the epidemic, the Hangzhou Asian Games and the Asian Para Games decided to postpone the Games for one year successively, and Hangzhou took a series of optimization initiatives, such as more refined urban management, a significant leap in the city's infrastructure, and the "sports for the people" principle was put into practice, which stimulated public expectation and interest, let more people know that the Asian Games were about to start, and warmed up the online public opinion of the Asian Games. 2) The torch relay of the Hangzhou Asian Games and China's sports payload: The torch relay is an essential ceremony of the Asian Games, and the topic further strengthened the public's attention to the Asian Games and aroused the public sense of national pride and anticipation for the upcoming Asian Games.

dawn phase: 1) The opening ceremony of the Hangzhou Asian Games showed China's charm and strength: the opening ceremony was a major attraction of the Asian Games. The event was a sports event and a cultural feast with rich connotations. China once again demonstrated its profound cultural heritage and long historical legacy to the world. Through the opening ceremony, the public saw China's strength and charisma, which is also an essential reason for the rapid rise of online public opinion. 2) Star artists promoted the Asian Games and cheered for the athletes: the participation of stars further increased the popularity and heat of the Asian Games, and the public paid more attention to the Asian Games while paying attention to the stars. 3) ④ Sports officially entered the Asian Games. They received wide attention: e-Sports officially entering the Asian Games is a great victory, which not only means the further development of the e-sports industry but also the integration and breakthrough between traditional sports and e-sports. It attracts not only fans but also the audience from the public in online games, expanding the field and increasing the audience of the Asian Games. 4) China's Table Tennis Team Achieves Another Success: The success of the national ping pong team in the Asian Games is not only the glory of the gold medals but also the victory of the team spirit, unity, and cooperation. They showed a high level of skills and muscular strength in the game. With absolute strength and excellent performance, the Chinese table tennis team showed the Asian countries the power and toughness of Chinese table tennis and the athletes' deep friendship and mutual support.

Climax phase: 1) the Asian Games Chinese women's diving, women's basketball, and dragon boat teams won the championship, showing the resounding rose. Jinlan Bloom: The Chinese team in the Asian Games women's programme won the championship one after another, showing that Chinese women's perseverance and hard work have been in the difficulties and frustrations of the resilience of the forward. In the face of many challenges and external doubts, they never give up. It strengthens the public's concern and support for the Asian Games. 2) The Chinese men's basketball team's defeat to the Philippines triggered the audience's dissatisfaction: The Chinese men's basketball team suffered a heavy blow in the basketball event of the Asian Games, and the
result of the game is undoubtedly a heavy blow to the Chinese men's basketball team. This unexpected defeat triggered widespread public concern and worry, and people were confused about the prospects of the Chinese men's basketball team. The Chinese men's basketball team will face even more formidable challenges and a worse situation. This topic brought the online public opinion of the Asian Games to a climax. 3) The Asian Games Chinese badminton team won gold medals in mixed doubles, men's singles, and women's doubles: this time, the national badminton team demonstrated a good mentality, the strength not to give up or give up, and a positive and courageous spirit of hard work! 4) The cute Asian Games mascot souvenir triggered netizens to seek after it: it triggered the public's desire to consume it and attracted much attention. 5) Zhou Shen's emotional and moving singing of the Hangzhou Asian Games promotion song "Aida Future" touched the crowd: Hangzhou, as the host city of the Asian Games, attracted the world's attention. From promotion to release, the Asian Games song "Aida Future" was enthusiastically forwarded and praised by netizens, which undoubtedly strengthened the public's attention to and support for the Asian Games. The closing ceremony of the Hangzhou Asian Games, with the general theme of "The Most Memorable is Hangzhou" and the backdrop of the memory of the South River, announced the end of the Asian Games but also heralded the beginning of the next Asian Games in Japan, arousing the public's anticipation for the event, and bringing online public opinion to a final climax.

Recession phase: 1) Xiao Zhaobao served as the public welfare ambassador for the Asian Games to fulfil his dream: star artists sincerely participated in the public welfare cause and practised the social responsibility that stars should do with practical actions, and this sincerity touched the hearts of the people. It attracted more attention from the public and enabled the network public opinion of the Asian Games to maintain a certain degree of activity during the recession period. 2) The fire of the 4th Asian Para Games in Hangzhou was completed in Guangzhou: even if the Asian Games ended, the public's attention to sports did not end, which is the reason that enabled the network public opinion of the Asian Games to maintain a certain degree of activity during the recession period. 3) The Closing Ceremony of the Asian Games in Hangzhou, the Maiden of Hoegui Costume design plagiarism controversy: The emergence of this topic undoubtedly triggered a great deal of discussion, once again pushing the Asian Games into the limelight. 4) The Hangzhou Asian Games "Goodbye, Most Memorable Hangzhou" commemorative short film in tribute to all the participants and dedictees! : The emergence of this topic expresses heartfelt gratitude and high respect to all those who have contributed to the Asian Games, reminds the public of the beautiful moments of the Asian Games, and is the reason why the online public opinion of the Asian Games has been sustained in the period of recession.

According to the above topic clustering results, it can be seen that in the preheating phase of a sports event, public opinion topics are usually for the preparation of the upcoming event, warming up for the opening ceremony, and the prediction and discussion of the sports event. People may discuss topics about the participating teams, the status of the players, historical match-ups, and so on. This phase is characterized by anticipation and speculation, with people paying more attention but yet to be on a large scale. Topics in the outburst phase are usually controversial or unexpected events related to the opening ceremony of a sporting event and the programme of play. This may include controversial penalties during the game, mistakes or injuries to athletes, and inappropriate behaviour by spectators. Public opinion quickly erupts, sparking widespread controversy and discussion, with media coverage and discussions on social media heating up. Topics at the climax phase are usually related to the event's results, the creation of a champion and outstanding performances. People may discuss the champion's feats, the race's highlights, the athletes' outstanding performances. This phase is characterized mainly by praise and commendation and, to a lesser extent, disappointment over losing a race where expectations were too high. At this time, the public is very concerned about the results and performance of the race, and discussions remain active in media reports and on social media. The topic of conversation during the fading phase is usually the review and recap of the tournament. People may discuss the
highlights of the race, reflect on the results, and discuss the performance of the athletes. This phase is characterized by a gradual subsiding of public opinion as people's attention and fervour decline, and media coverage and discussions on social media gradually diminish.

**Visualization and analysis of public sentiment based on time dimension.** According to the different phases of the Hangzhou Asian Games process, combined with the results of LDA topic clustering, the ratio of the number of microblog data entries expressing different emotional polarities contained in each sub-topic to the total number of microblog entries of the sub-topic in that phase was calculated for each sub-topic within each phase (Fig.5).

![Fig. 5 Sentiment distribution of “Hangzhou Asian Games” microblog topic](image)

As can be seen in Fig.5, by analyzing the statistical results of microblog sentiment classification during the entire public opinion dissemination cycle of the Asian Games, it is found that from the perspective of the entire life cycle of public opinion, the proportion of positive sentiment in microblog texts is higher than that of negative sentiment or neutral sentiment in general. This indicates a phenomenon of positive emotions gathering in public opinion formation. Specifically, the proportion of positive emotions in the beginning and dawn phases is over 60%, and the proportion of neutral emotions is more significant than that of negative emotions. Secondly, entering the climax phase, the proportion of positive emotions declined slightly, and the proportion of negative emotions began to rise, especially in Topic III-2, where the proportion of negative emotions was as high as 30%.

In the men's basketball semifinals of the Hangzhou Asian Games, the Chinese men's basketball team was eliminated by the Philippines in a reversal of the 20-point lead. The result of the loss made netizens' jaws drop, and the users' comments showed a negative trend. Negative comments increased dramatically, with users expressing disappointment in the Chinese Men's Basketball team and worries about its prospects. Finally, entering the recession phase, the positive sentiment of Topic IV-1 and Topic IV-2 accounted for more than 70%, and the negative sentiment of Topic IV-3 was 30%, which can be seen due to the plagiarism fiasco of the design of the closing ceremony of the Asian Games in Hangzhou, Lotus Laurel Girls' Clothing Design, which resulted in some netizens doubting the level of the design of the closing ceremony of the Asian Games, Lotus Laurel Girls' Clothing Design and innovativeness, and at the same time there are also criticisms and accusations against the designers, which are considered to lack originality and creativity. It is believed that it lacks originality and creativity. This has hurt the overall image and reputation of the Hangzhou Asian Games. The incident also triggered discussions and reflections on protecting intellectual property rights and the spirit of originality in all walks of life.

**Spatial Dimension-Based Sentiment Visualization Analysis of Internet Users.** As a professional GIS software, ArcGIS provides a complete set of GIS tools and applications with powerful geospatial data processing and visualization capabilities. which can present the data two-dimensionally for various fields, including public
opinion trends, urban planning, defence and security (Fischer and Getis 2010). In order to better understand and reveal the geographic distribution characteristics and spatial evolution patterns of Internet users’ emotions, users are helped to display and convey geographic information and intuitively understand the relationship between data. In this paper, using ArcGIS software and referring to previous studies, we take the mean sentiment value of the location of microblog users as a regional sentiment evaluation index and use GIS, data mining and visualization techniques to statistically determine the spatial distribution of the overall mean sentiment value of the microblog comments at each phase of netizen sentiment within the public opinion cycle at the provincial scale (Fig. 6).

Fig. 6 Spatial distribution of affective state values at different phases and in each provincial-level administrative region at the provincial scale

We extracted samples from each province according to the location of the microblog posting and labelled each province with sentiment classification labels. Considering that the difference in the sample size of each province may affect the results, this study adopts the mean sentiment value as an indicator to reduce the effect of sample size. Since the sentiment values of all provinces are more significant than 0.7, the sentiment state is positive in order to be able to distinguish clearly, so the interval of the sentiment value is set to 0.70-1.00, and the larger the value of the sentiment state, the more positive the overall sentiment state of the province.

As can be seen in Fig. 6, during the preheating phase, the sentiment values of all provinces are less than 0.9, and the sentiment means of Macao, Shanxi, Qinghai, Taiwan, Chongqing, Guangdong, Hunan, Anhui, and Jiangxi are higher than 0.85. Taking the Qinling-Huaihe River line as the boundary, the spatial distribution of the sentiment means shows a more noticeable difference between the north and the south, with the ratio of the southern provinces’ sentiment mean values higher than 0.85 reaching 77.78 per cent and the number
significantly higher than that in the north. However, compared with the northern provinces, the emotional state is more optimistic, but there is less overall difference between different regions. The number is significantly higher than that in the north. However, the emotional state is more favourable than the northern provinces; the overall difference in the emotional mean value between different regions is more minor. In the outbreak phase, the overall emotional state value grew explosively, with the national average value of emotion increasing from 0.826 in the warm-up phase to 0.900, of which the average value of emotion of netizens in 24 provinces was higher than 0.9, presenting very high and positive emotions, while the average value of emotion in Beijing, Zhejiang, Gansu, and Tibet was lower than 0.85, with the difference between the northern and southern regions decreasing. The difference between the eastern and western regions is increasing. In the climax phase, the average value of emotions in Zhejiang and Beijing is lower than 0.85, and the emotional state of netizens in all country provinces is more stable. In the recession phase, the overall average value of emotions began to decline. The average value of emotions in Taiwan, Macau, Chongqing, Anhui, Hong Kong, Guizhou, Hunan, Hainan, and Hubei were higher than 0.9 in nine provinces, all of which were southern cities.

Only the average value of emotions of netizens in Zhejiang was lower than 0.8. Comprehensively, the four phases found that the average value of the emotions of netizens in Zhejiang, the venue of the Asian Games, in the four phases was lower than the national average. The reasons for this were that the hosting of the Asian Games may have brought some impacts to Zhejiang Province, such as the construction of competition venues, urban renovation, traffic congestion and other issues, which may have caused inconvenience and disturbance to the lives of residents, leading to their less positive emotional attitudes towards the Asian Games. Media reports and remarks on social media in Zhejiang Province may also affect the public's emotional attitude towards the Asian Games in Zhejiang Province. However, on the whole, netizens across the country have a positive emotional state towards the Hangzhou Asian Games. The Asian Games is an international sports event with high gold content, and it is a matter of great pride and celebration to be held in China, especially in a city like Hangzhou. At the same time, the athletes of the Chinese team in all disciplines are striving to win gold medals, which inspires netizens' patriotism and sense of unity; in addition, the Hangzhou Asian Games In addition, the design of the mascot and theme song of the Hangzhou Asian Games was also warmly welcomed and praised by netizens. The organizers of the Hangzhou Asian Games paid great attention to communication and publicity work with the public, actively interacting with netizens through various channels (e.g., social media) and providing various information, which triggered positive emotions among netizens.

**Visualization and analysis of public sentiment based on gender dimensions.** Gender is one of the critical factors influencing individuals' views, attitudes and behaviours. This section combines the results of the cycle division of the Hangzhou Asian Games, user gender and IP address information to count the number of male and female emotional states of each theme at each phase and the number of male and female participants in each province. The visual analysis of the gender dimension can provide an in-depth understanding of the different emotional responses and attitudes of different gender groups to the same event. It can also reveal the differences in attitudes and emotions between men and women when dealing with an event. Understanding the emotional responses and attitudes of different gender groups can help event organizers and relevant departments develop more effective publicity and public relations strategies to meet the needs and expectations of different gender groups (Figs 7 and 8).
Fig. 7 Distribution of gender sentiment on Weibo topics of the "Hangzhou Asian Games"

Fig. 8 Gender distribution of "Hangzhou Asian Games" microblog topics at different phases and by provincial administrative regions in China

Fig. 7 and Fig. 8 show that during the warm-up phase, 37,286 people participated in the microblogging "Hangzhou Asian Games" topic interaction, of which the total number of men and women who participated in the microblogging "Hangzhou Asian Games" topic interaction in Zhejiang, Guangdong, Sichuan and Beijing exceeded 2,000. The number of women participating in each province except Beijing was more than that of men. The total number of men and women in Zhejiang, Guangdong, Sichuan and Beijing who participated in the "Hangzhou Asian Games" topic interaction exceeded 2,000, and the number of female participants in all provinces except Beijing was higher than that of male participants. The positive sentiment of male and female participants in the Countdown to the Opening Ceremony of the Hangzhou Asian Games (Topic I-1) and the Passing of the Torch of the Hangzhou Asian Games (Topic I-2) was close to the ratio of positive emotions, and they were all looking forward to the opening of the Hangzhou Asian Games. In the dawn phase, the number of...
people participating in the topic interactions exploded to 108,982, with the number of girls accounting for 70% of the total number of people, of which the total number of men and women in Guangdong, Zhejiang, Jiangsu, Beijing, Shandong, Hebei and Shanghai exceeded 5,000. Guangdong reached 13,049, which shows that the opening of the Hangzhou Asian Games has made more people participate in the topic of celebrity artists promoting the Asian Games (Topic II -2), the number of male participants in the topic was 1,569 in contrast to the number of female participants, which was 4,626. This was because in order to increase the publicity for the Asian Games in Hangzhou, several male celebrity artists were invited by the Hangzhou Asian Games officials to take part in the topic interactions, which attracted countless female fans who were enthusiastic in their pursuit and the first time that eSports made its appearance in the Asian Games (Topic II -3) ushered in the outbreak of the phase. At the peak of topic interaction, the number of male and female emotions were 9078 and 10848, respectively, which can reflect the difference between eSports and other sports in that eSports has a high audience base in China, a low threshold of participation, and a high degree of mass participation. In the climax phase, the number of people participating in topic interactions in each province tends to be stable, and the gap is gradually narrowing, with Guangdong and Zhejiang remaining the top two in terms of male and female topic participation. The difference between this phase and the other phases is that, except for the sixth topic in the climax phase, the number of positive emotions in the remaining five topics is higher for men than for women. The negative emotions in the Chinese Men's Basketball team's defeat to the Philippines (Topic III-2) are higher, with more men Negative emotions are higher in the Chinese men's basketball team's defeat to the Philippines (Topic III-2), with more men expressing disappointment and regret, which is also related to the basketball programme itself, as men tend to pay more attention to fierce rivalry. The closing ceremony of the Hangzhou Asian Games, with the theme of "The Most Memorable Thing is Hangzhou" (Topic III-6), was a triumphant ending, with the topic reaching a small peak in popularity. In the recession phase, the number of male and female topic participants in each province declined again, with an average of 1,537 male and female participants per province. Xiao Zhan served as the Asian Games' dream public welfare ambassador (Topic IV-1), which made the interactive heat grow again, and the traffic flow brought about by the celebrity effect is not only embodied in the widespread popularity and influence but also the mighty publicity power possessed by the fan base. Lotus Gui girl clothing design plagiarism storm (Topic IV-3) so that the netizens of the negative emotions increased, but the positive emotions still occupied the majority, indicating that most of the netizens did not blindly follow the trend but calmly and objectively look at the problem, along with the Hangzhou Asian Games, "goodbye, the most memory of Hangzhou" commemorative short film to pay homage to all the participants and dedicatees! (Topic IV-4) The topic dropped to a low point, and public opinion gradually ended.

Conclusions

In this paper, we use Python to write a crawler to get the information of "Hangzhou Asian Games" topic on Weibo and construct a whole-process analysis framework from "data collection, data cleaning and preprocessing, division of public opinion cycle, thematic evolution and sentiment evolution", and make a map to reveal the discussion content and sentiment polarity development trend of each topic in the development cycle of Hangzhou Asian Games by province and gender by using the method of data visualization. Through the data visualization method, the data is mined for spatial and temporal patterns, and sentiment analysis maps are made by provinces and genders, which can reveal the discussion contents and sentiment polarity development trends of the topics at each phase in the development cycle of public opinion on the Hangzhou Asian Games. According to the evolution results, the emotional colour of online public opinion is closely related to the development phase of the sports event, the nature of the event and its scope of influence. In the preheating phase, online public opinion is dominated by anticipation and concern, and the emotional colour of the opinion field is generally favourable. At the down phase, netizens' attention and participation increased significantly. The
emotional colour of public opinion was diversified, ranging from affirmation and praise for the event and athletes’ performance to questioning and criticism of some controversial events. At the climax of the event, the emotional colour of online public opinion is more complex and varied, with intense emotional reactions involving the personal development of the athletes, teamwork, and the organization and operation of the event. In the recession phase, the emotional colour of online public opinion gradually smoothed out, and the topic of discussion gradually shifted from the event itself to evaluation and reflection on the event, as well as expectations and prospects for future events.

This study also has some limitations: firstly, only microblogs are selected as the data source, and a single sports event is the research object, which is difficult to reflect the comprehensiveness of the problem; in the future, the sample size should be enlarged to include a variety of social media platforms as well as different types of events; secondly, the process of emotion annotation does not take into account the influence of emoticons and network buzzwords in emotion expression, which may reduce the accuracy of emotion classification and prediction. Non-text features should be added to the sentiment computation model to improve the sentiment recognition ability; finally, there is a lack of validation of the study results, and a public sentiment attitude survey should be conducted for some microblogs to verify the accuracy of the previous analyses.

In the future, further strengthening the research and management of online public opinion is necessary to enhance sports events' organization and communication effect. On the one hand, technical means such as big data and artificial intelligence can be used to deeply mine and analyze online public opinion on sports events and monitor changes in public opinion in real time. On the other hand, the relationship between the degree of emotional inclination and age, city area, educational experience, occupation, number of followers, and fans, can be explored through the correlation analysis method. Drawing on successful domestic and international experience in operating sports events, strengthening the standardization and professionalization of event organization and operation, and improving the quality of the event will also help shape positive online public opinion and enhance the influence and public impact of the event. Overall, the impact of online public opinion on sports events is two-sided, with both challenges and opportunities. The study and management of online public opinion is an integral part of the operation of sports events and an important direction for future research on sports events.

Data availability

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

References


Blei DM, Ng AY, Jordan MIJomLr (2003) Latent dirichlet allocation. 3 (Jan):993-1022

Chen C, Chen J, Shi C Research on credit evaluation model of online store based on SnowNLP. In: E3S Web of Conferences, 2018. EDP Sciences, p 03039


Fischer MM, Getis A (2010) Handbook of applied spatial analysis: software tools, methods and applications. Springer,


Hidayatullah AF, Ma’arif MR Road traffic topic modeling on Twitter using latent dirichlet allocation. In: 2017 international conference on sustainable information engineering and technology (SIET), 2017. IEEE, pp 47-52


Joung J, Kim HMJJoMD (2021) Automated keyword filtering in latent Dirichlet allocation for identifying product attributes from online reviews. 143 (8):084501


Liu Z (2013) High performance latent dirichlet allocation for text mining. Brunel University School of Engineering and Design PhD Theses,


Wang Y, Xu WDSS (2018) Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. 105:87-95

Acknowledgements

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors thank Dr Wen Che for advising and proof reading, and to Dr Shenghua Qi for proof reading.

Competing interests

The authors declare no competing interests

Ethical approval

Ethical assessment is not required prior to conducting the research reported in this paper, as the present study does not have experiments on human subjects and animals, and does not contain any sensitive and private information.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.