

Research on Public Opinion on Twitter of 2022 Beijing Winter Olympics

-- Sentiment Analysis based on Support Vector Machine

Haotian Hou*

North China Electric Power University, Department of Foreign Studies, Baoding, Hebei, 071000

*Corresponding author's Email: houhaotian115@163.com

ABSTRACT

Beijing 2022 Olympic Winter Games are important landmark events in China, and their hosting is of great significance. This study uses Python data mining, Support Vector Machine algorithm, and #LancsBox to analyze the sentimental tendency of comments under related topics on Twitter and summarize netizens' opinions. The results show that Twitter users hold more negative sentiments about the Beijing Winter Olympics. By calculating the mutual information value, it is found that negative sentiment words are paired with high-frequency words more by chance.

Keywords: Beijing 2022 Winter Olympic Games, Sentiment Analysis, SVM

1. INTRODUCTION

President Xi Jinping proposed, "The Beijing Winter Olympics is a major landmark event at an important point in China's history."^[1] Beijing 2022 Winter Olympic Games and Winter Paralympic Games (from now on referred to as "Beijing Winter Olympic Games") are international comprehensive sports events. However, as the preparations and "Experience Beijing" series of test matches for the Beijing Winter Olympic Games are underway, there has been sporadic talk of a "boycott" and "diplomatic boycott" of Beijing Winter Olympic. 2012 London Organizing Committee Chairman Sebastian Newbold Coe recently stated that he opposes "The diplomatic boycott" would be "a meaningless and harmful gesture." A spokesman for British Prime Minister Johnson has also stated that he does not support a boycott of sporting events. Susanne Lyons, Chair of the U.S. Olympic Committee, said that a boycott would only unjustly "punish" athletes^[3].

Twitter is one of the most important social media globally, and its users come from every corner of the world. Using the sentiment analysis technology in Natural Language Processing to analyze the comments of tweets on Twitter is conducive to grasping the international public opinion on the topic on time and providing reference for China to organize a simple, safe and wonderful Winter Olympic Games. It is also conducive to promoting a community of human destiny.

In fact, it also promotes the interdisciplinary development of computer science and linguistics.

2. LITERATURE REVIEW

Sentiment analysis is currently a popular research area in computer science^[4]. It also integrates relevant knowledge in computer science and linguistics and can mine out the sentiment tendencies in texts, thus providing references for solving practical problems. Siqi Che and Xuepei Li combined sentiment dictionary and support vector machine algorithm to improve the model accuracy to a great extent and compared and analyzed the sentiment discourse in the letters to shareholders of Chinese and American companies under the perspective of evaluation system^[5]. Saif, He, Alani, Fernandez et al. improved the sentiment analysis method through a sentiment dictionary by static sentiment words and practiced it on the Twitter website. The results showed that using the SentiCircles method improved sentiment prediction accuracy^[6]. Dan Li et al. used deep learning algorithms to perform sentiment analysis on related topic opinion on microblog, which references the implementation of China's three-child policy and the introduction of related supporting facilities^[7]. The Beijing Winter Olympic Games is one of the hot topics this year. However, the visual analysis of the search results for the subject term "Beijing Winter Olympic Games" on CKNI(China National Knowledge Infrastructure), for example, shows that the number of published literature has been increasing year by year, even exponentially

before the opening ceremony. However, the total number of published articles is only 1278, and the research is still in the initial stage. In terms of the research area, 65% of the articles were in sports, and the value of the Beijing Winter Olympics is far from being limited to this area. Because of this problem, the present study tried to further explore the value of the Beijing Winter Olympics by applying natural language processing to Twitter comments.

Mainly research the following issues:

- 1) The emotional tendencies of Twitter netizens towards the Beijing Winter Olympics;
- 2) The textual features of Twitter netizens' comment discourse

Table 1 Trend of articles with the subject line "Beijing Winter Olympic Games"

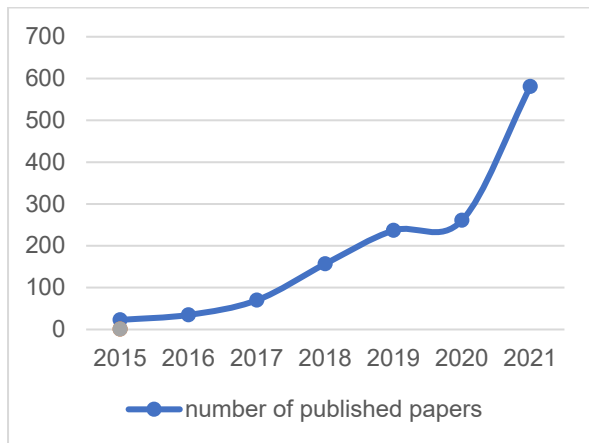
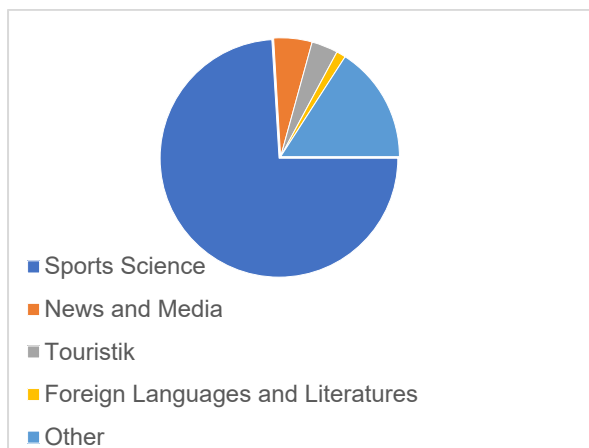


Table 2 Subject areas of articles with the subject line "Beijing Winter Olympics"



3. RESEARCH METHOD

3.1. SVM (Support Vector Machine)

Support vector machine is a new machine learning method based on statistical learning theory, which has excellent learning performance and is widely used in text

recognition, handwritten font recognition, face image recognition, etc. [8]. Support vector machines can classify some common emotional expressions [9]. This paper builds a training set for model construction to predict the sentiment polarity of word-level English texts.

3.2. Data

The author uses Python crawler technology to select users on Twitter for data crawling. Since the number of comments under individual users of Twitter in China is generally low, if individual users are selected for the study, it will greatly affect the accuracy of the research results, so the official media are decided to select in China and some of the official media abroad as the data source, and the crawling object is the direct comments under the topic. The comments of foreign netizens can reflect the attitude of foreign countries towards China's hosting of the Olympics. The time is from October 11, 2021, to November 27, 2021. The users include China's Beijing Winter Olympics official Twitter: @Beijing2022; some of China's official media, and the Ministry of Foreign Affairs spokesperson certified accounts: @SpokespersonCHN (Hua Chunying), @zlj517 (Lijian Zhao). Some foreign news media: @CNN, @nytimes (The New York Times), @torontostar (Toronto Star), the final crawl results show a total of 847 data, saved in Excel file format. We also crawled the tweets with the search terms "Olympic" and "China" and 783 used them to build the training set.

3.3. Data Pre-processing

To ensure the authenticity and reliability of the study results and the need for text processing, the raw data need to be screened and pre-processed in the following ways.

- 1) Converts all data to lowercase.
- 2) Remove comments that have little to do with this study, including discussions in the comments section about political topics that have nothing to do with the Beijing Winter Olympics, other hot topics, etc.
- 3) Remove data that cannot be text processed, such as emoticons, weblinks, punctuation, etc.
- 4) Replies to a large amount of repetitive text, such as the name of the user and the name of the participating topic.
- 5) Split the words according to the spaces, remove the deactivated words and extract the word stems.

After the above process is completed, the remaining 583 bar data is converted to .txt document format and saved.

The high frequency of the four words "China," "Beijing," "winter," "Olympic" in Figure 2 and Figure 3 indicates that the studied commentary discourse is closely related to the topic of the Beijing Winter Olympic Games and has the significance of continuing in-depth research.

From the positive utterance word cloud map, the opinions of Internet users can be categorized into the following three categories.

1) The Beijing Winter Olympics is an excellent opportunity for the athletes to show their sports style. For example, "I want to cheer for the athletes on the sideline with my family, welcome to Beijing.", "The Winter Olympics are The Winter Olympics are a gathering of the world's winter Olympic athletes, and the athletes are the stars." "This says a lot about the Chinese. Well done, China."

2) The belief that sports are borderless, pure, and shared by the people of the world. For example, "Sport knows no borders and the world is a better place.", "Olympics are supposed to be apolitical. "

3) Looking forward to coming to Beijing, to China. For example, "I think Beijing is beautiful in winter!", "I love going to Beijing2022 to see the glamorous city line sky view. "So cool, look forward to 2022 Beijing Winter Olympics".

Similarly, through the negative utterance word cloud map, the opinions of Internet users can be summarized and categorized into the following categories.

1) The worldwide outbreak of the covid-19 is still severe, and there are concerns about the safety of participating in the Beijing Winter Olympics. For example, "To be honest, it would not be safe.", "China has no rights to talking about Health. "There is a virus going on!"

2) It is believed that China does not respect human rights. "Will the human rights have a place on there?", "We demand that the Chinese government provide basic protection for the life, death, and human rights of the Chinese people."

The Beijing Winter Olympics has focused the world's attention and is an important opportunity for China to enhance its international discourse in a new era. Through the Beijing Winter Olympics, China has shown a more diverse, prosperous, and inclusive international image. National Ski Jumping Center "Snow Ruyi" in Zhangjiakou, the Beijing Winter Olympics mascot "Bing Dwen Dwen" and the Winter Paralympics mascot "Shuey Rhon Rhon" present a "Civilization" with a long history and cultural heritage. China should firmly grasp the image of a "civilized country" with a long history and cultural heritage, which should firmly grasp the opportunity of the Beijing Winter Olympic Games. It solidly promotes the preparation of the Beijing Winter Olympic Games and demonstrates the superiority of the

socialist system with Chinese characteristics in concentrating its efforts on major events; make "scientific and technological innovation" the core of supporting the Winter Olympic Games, and show the charm of sports through scientific and technological means; improve medical protection, strengthen epidemic prevention and control, and reduce the impact of epidemics on athletes. The epidemic will have the lowest impact on athletes^[10]. At the same time, we also need to use the Beijing Winter Olympics to promote the outreach of China's ice and snow culture with Chinese characteristics. Currently, some foreign media have started to make various politicized biased narratives about the Beijing Winter Olympics against the spirit of the Olympic Charter. Moreover, from the results of this study, most of the negative comments point to China's human rights issues, Hong Kong issues, etc. They are linked to the Beijing Winter Olympics, arguing that since China cannot solve the above problems, it cannot run the Winter Olympics either, and in fact, these comments are not in line with China's reality, thus also showing that some foreign netizens do not understand the actual situation of China. In this context, China should combine the relevant communication theories to let the world hear China's voice and see a real China. From the perspective of communication subject, the general public should be included in the category of communication subject, instead of relying only on the power of the government; from the perspective of communication content, we should pay attention to the differences between China's traditional ice and snow culture and modern international ice and snow culture, seek common ground while reserving differences, combine specific examples, get close to life, improve recognition, and meet the needs of the audience. From the perspective of communication channels, intelligence, sports events dissemination, educational dissemination, and other methods to improve the dissemination effect of ice and snow culture. From the perspective of the dissemination object, it is necessary to face international and focus on the masses^[11].

Holding the Winter Olympic Games under the new crown epidemic Beijing, facing great pressure and risk, Bach, the president of the Olympic Organizing Committee, has put forward a high evaluation for the Beijing Winter Olympic Games: "Overcoming the impact of the epidemic, the preparation for the Beijing Winter Olympic Games is almost a miracle." The Beijing Organizing Committee, the International Olympic Committee, and the International Paralympic Committee jointly released the second edition of the Beijing 2022 Winter Olympic and Paralympic Games Epidemic Prevention Manual, which includes relevant refinements based on issues that arose during the test events. The manual sets out six principles: vaccination, closed-loop management, which establishes a new liaison officer mechanism, detection, tracking and isolation, contact reduction and hygiene awareness, and a detailed plan for

health follow-up and nucleic acid testing of relevant personnel. All medical work for the Beijing Winter Olympics is ready^[12]. China has the strength and ability to host a simple, safe, and exciting Winter Olympic Games.

4.3. GraphColl

GraphColl is the most core feature of #LancsBox. It is important for natural language processing to identify co-occurrence forms of words in discourse. In the "GraphColl" module, we set the span as R5-L5, the statistical calculation method MI value ≥ 3 (MI value, i.e., mutual information value, can calculate the relationship tightness between two words to find the significant collocation words in the discourse, Hunston proposed that the mutual information value more excellent than "3" are significant collocation words^[13]), the frequency of collocation ≥ 5 , and the text display type "Type." The search terms were the most frequent actual words in the text: "Beijing" and "Olympic." The results are shown in the following figure.



Figure 4 GraphColl of "Beijing"



Figure 5 GraphColl of "Olympic"

In the Figure4, the top five words in descending order of MI values are: "meet" (6.7759, R, with the MI value and position of the word in order in parentheses, below), "welcome" (6.4540, L), "wish" (6.4540, L), "complete" (6.0964, R), "2022"(6.0543, L).

In the Figure5, the top five words in descending order of MI values are "motto" (6.7551, R), "spirit" (6.4921, R), "complete" (6.1701, R), "wish" (6.1701, L), "games" (6.0547, R), where the words meet and welcome express Beijing's warm welcome to the world as the host of the Winter Olympics. The words meet and welcome express Beijing's warm welcome to the world as the host of the

Winter Olympics, reflecting China's inclusive and open image. The words wish, complete, and2022 show China's desire to host a perfect Winter Olympic Games, showing the image of China as a responsible power.

From the GraphColl results, the words most closely related to the two high-frequency words are almost always positive or neutral, which infers that the co-occurrence of words carrying negative sentiment is more contingent with high-frequency words than words carrying positive sentiment in the prediction results of sentiment analysis.

5. CONCLUSION

The results of the sentiment analysis of Twitter comments related to the Beijing Winter Olympics show that

- 1) Foreign netizens have more negative sentiment toward the Beijing Winter Olympics and
- 2) Words that carry negative sentiment are more contingent in co-occurrence with high-frequency words in the corpus than words that carry the positive sentiment. This study does not analyze the coding process too much but only explores it in depth from the perspective of the results.

It can reflect that most of the negative emotions about the Beijing Winter Olympics do not come from the Winter Olympics itself but other issues, such as China's territorial issues and other sensitive topics.

There is still room for further exploration in analyzing public opinion about major events, and we hope to provide a reference for relevant national decisions.

REFERENCES

- [1] Ren Wanxia. Welcome the Winter Olympics together and look to the future together [J]. Beijing Observation,2021(11):10-11.
- [2] Yi Jiandong, Wang Daojie. On the value and significance of the Beijing 2022 Winter Olympics [J]. Sports and Science,2016,37(05):34-40+33.
- [3] Li Li, Ji Ye, Zhu Hong. "Diplomatic boycott" of the Winter Olympics? Purely forcibly "adding drama"! [N]. Xinhua Daily Telegraph,2021-11-26(007).
- [4] Feldman R. Techniques and applications for sentiment analysis[J]. Communications of the ACM,2013, 56(4): 82-89.
- [6] Saif H, He Y, Fernandez M, et al. Contextual semantics for sentiment analysis of Twitter[J]. Information Processing & Management,2016, 52(1): 5-19.
- [7] Li Dan, Li Liping, Li Dan. The Public Opinion Effect and Enlightenment of the Three-Child

- Policy—Analysis of Network Big Data Based on NLP [J]. *Research on Chinese Youth*,2021(10):46-53.
- [8] Ding Shifei, Qi Bingjuan, Tan Hongyan.Review of Support Vector Machine Theory and Algorithm Research[J]. *University of Electronic Science and Technology of China*,2011,40(01):2-10.
- [9] Li Mengnan, Wang Mingyan. A review of sentiment analysis methods and applications based on machine learning [J]. *Software Engineering*,2021,24(09):21-23+8.
- [10] Hu Jianqiu. The 2022 Beijing Winter Olympics and the multi-dimensional construction of the national image under the new crown pneumonia epidemic [J]. *Shandong Sports Science and Technology*,2020,42(06):8-13.
- [11] Zhang Lei, Xie Jun. Research on the communication strategy of ice and snow culture under the background of the 2022 Beijing Winter Olympics [J]. *Journal of Xi'an Institute of Physical Education*,2021,38(01):55-62.
- [12] Yang Na. Ensuring the safe holding of the Beijing Winter Olympics and Paralympics [N]. *China Women's Daily*,2021-12-24(002).
- [13] Hunston, S. *Corpora in Applied Linguistics*[M].Beijing: World Book Publishing Company Beijing Branch,2006:71.