Topic Mining and Sentiment Analysis of Twitter Texts Under the Belt and Road Initiative: Postprint

Authors: Changyu Zhao, Yaping Wu, Wang Jimin

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] The proposal of the "Belt and Road" Initiative has attracted widespread attention both domestically and internationally, with users from numerous countries expressing opinions, posting comments, and engaging in discussions on Twitter, the most representative social media platform. Extracting discussion topics and sentiment orientations regarding the "Belt and Road" from tweets can provide valuable references for government agencies to optimize publicity strategies and enhance the exposure and attention of the "Belt and Road" Initiative. [Method/Process] We collected over 60,000 tweets related to the "Belt and Road" in 2017, and conducted data preprocessing, data description, topic mining, and sentiment analysis separately for Chinese and English tweets, and performed cross-analysis of topics and sentiments to draw conclusions. [Results/Conclusion] The tweet topics in 2017 mainly revolved around the "Belt and Road" Summit Forum held in May. Specifically, Chinese tweets focused more on the planning and implementation of the Summit Forum, as well as security issues and leadership visits, with significant fluctuations in sentiment values, particularly large negative sentiment fluctuations regarding security issues. English tweets, in contrast, focused more on the fact that the Summit Forum was held and its economic effects, with smaller sentiment fluctuations; the proportion of positive sentiment values regarding economic aspects was significantly higher than that of negative and neutral sentiment values.

Full Text

Preamble

Twitter Text Topic Mining and Sentiment Analysis Under the Belt and Road Initiative

Zhao Changyu[1], Wu Yaping[2], Wang Jimin[1]

[1]Department of Information Management, Peking University, Beijing 100871
[2]Peking University Library, Beijing 100871

Abstract

[**Purpose/Significance**] The Belt and Road Initiative has attracted widespread attention both domestically and internationally, with users from numerous countries expressing opinions, posting comments, and engaging in discussions on Twitter, the most representative social media platform. Mining discussion topics and sentiment tendencies regarding the Belt and Road Initiative from these tweets can provide valuable references for government agencies to optimize publicity strategies and increase the initiative's exposure and attention. [**Method/Process**] This study collected over 60,000 tweets related to the Belt and Road Initiative in 2017, conducting data preprocessing, description, topic mining, and sentiment analysis separately for Chinese and English tweets, and performed cross-analysis of topics and sentiments to draw conclusions. [**Result/Conclusion**] The 2017 tweets primarily focused on the Belt and Road Forum for International Cooperation held in May. Chinese tweets showed greater attention to the planning and implementation of the forum, security issues, and leadership visits, with significant emotional fluctuations, particularly strong negative emotions regarding security concerns. English tweets focused more on the fact that the forum was held and its economic effects, with smaller emotional fluctuations and a significantly higher proportion of positive sentiment values in economic aspects compared to negative and neutral values.

Classification Number: TP391.1
Keywords: Belt and Road Initiative, Twitter, topic mining, sentiment analysis

1 Related Research

1.1 Research Status on the Belt and Road Initiative

As of April 2018, CNKI contained over 40,000 articles related to the Belt and Road Initiative, covering policy analysis, economic effect analysis, scientific research cooperation analysis, and other aspects. Among these, text analysis-related articles began to appear in 2015, with relatively small quantities, mostly focusing on news media reports such as *China Daily Africa Edition* and *The Washington Post* to extract semantic structural features and thematic viewpoints. For instance, Huang Yanqiu analyzed text data from *China Daily Africa Edition* to discuss issues of public diplomacy and communication paradigms, discovering during the analysis that the Belt and Road topic had moved beyond one-sided reporting to become an international issue, and finally proposed strategies and suggestions for publicity reporting [1]. Zhu Guisheng and other scholars used text data from the American *Washington Post* to conduct critical discourse analysis of Belt and Road-related reports, revealing from three levels—text, discourse practice, and social practice—how American media portrayed China's Belt and Road Initiative as a colonial expansion and profit-driven hegemonic image [2]. Research focusing on social media text is relatively scarce. Domestic data sources mainly rely on Sina Weibo, such as collecting approximately 360,000 Weibo posts from 2013-2016 and using spatial autocorrelation analysis to demonstrate response differences between core and peripheral cities, proposing suggestions to optimize the spatial structure of publicity [3]. Foreign data sources primarily use Twitter, such as collecting over 2,000 tweets on the Belt and Road theme in 2015 for regional division and keyword statistics to analyze information transmission processes [4]. However, current research is limited to structured traditional news texts, with shortcomings including small data volumes, relatively single data sources, and limited analytical dimensions.

1.2 Advances in Topic Mining Methods

Topic mining is the process of discovering research topics through the association relationships among text feature items in a document collection. Analyzing the evolution of topics along the temporal dimension can clarify their development trajectory and identify innovation points. Traditional topic mining methods mainly include word frequency analysis, co-word analysis, and citation analysis. Word frequency analysis, based on Zipf's Law, identifies topics through the frequency of keywords or subject terms. Although simple and easy to use, high-frequency and low-frequency words have strong subjectivity, leading to broad topic ranges and difficulty in normalization. Co-word analysis, based on statistical concepts, examines the co-occurrence relationship of two keywords in the same document, considering both word frequency and word relationships, but low-frequency words are not easily incorporated into topic discussions. Citation analysis, based on citation relationships, uses indicators such as citation rate, citation coupling, and co-citation for topic division, but faces issues like complex citation relationships and inconsistent citation formats. Overall, while traditional topic evolution analysis methods are easy to operate and widely applied, they are highly subjective and have limited research conclusions.

Subsequently, complex models combining machine learning and natural language processing emerged, including LSI, PLSI [5], and LDA. The Latent Dirichlet Allocation (LDA) model, proposed by D.M. Blei et al. [6] in 2003, is a technique for identifying common topics in a set of documents. It posits that each word in an article is generated by "selecting a topic with a certain probability, and then selecting a word from that topic with a certain probability" [7]. Around 2010, scholars began applying the LDA model to social media. M. Michelson et al. [8] used the LDA model to study the topics that Twitter users follow, Y.S. Hwang et al. [9] used the LDA topic model to study the patterns and methods of opinion leaders' discussion topics, and Y. Hu et al. [10] applied the LDA model to analyze social media comment data on current affairs to derive user viewpoints. Subsequently, improved models based on PLSA and LDA emerged, such as Q. Mei et al. [11] who modified the PLSA model by applying contextual information to create CPLSA (Contextual Probabilistic Latent Semantic Analysis), and S. Moghaddam et al. [12] who proposed ILDA (Incremental Latent Dirichlet Allocation), which adds text feature parameters to the LDA model to improve topic clustering accuracy. Additionally, Dynamic Topic Models (DTM) [13] and Online LDA (OLD) [14] have appeared. These developments demonstrate that numerous scholars have conducted extensive and in-depth research on the LDA model, which has been well-developed. Therefore, this paper selects the LDA method for tweet topic mining.

1.3 Advances in Sentiment Analysis Methods

Sentiment analysis is the process of identifying users' subjective emotions, opinions, and attitudes from text data [15], widely applied in public opinion monitoring and information prediction. Initial social media sentiment analysis was conducted based on Twitter data [16-18]. J. Bollen et al. [19] divided emotions into six dimensions based on Twitter data, analyzing the most representative emotions each day. P.S. Dodds et al. [20] attempted to explain patterns of human happiness from a sentiment analysis perspective. Sentiment analysis methods can be mainly divided into sentiment dictionary-based methods and machine learning-based methods. Dictionary-based sentiment analysis extracts feature words from the text to be tested, looks up the sentiment values of these feature words in a sentiment dictionary, and classifies sentiments based on accumulated sentiment values [21]. For sentiment dictionary selection, there are generally two approaches: one is to directly use existing sentiment dictionaries such as HowNet [22], SentiWordNet, and Inquirers [23]; the other is to build a dictionary through research data, such as R. Feldman et al. [24-25] who extracted sentiment words using partially manual annotation and bootstrapping based on existing sentiment dictionaries. Machine learning-based sentiment analysis methods first train a classifier based on a text collection and then use the classifier to classify new texts [26]. With the development of artificial intelligence and deep learning, many scholars have applied deep learning techniques to sentiment analysis. B. Pang et al. [27] first used machine learning methods for sentiment analysis of movie review texts. Zhang Zhihua [28] used a convolutional neural network model for sentiment analysis based on sentiment word vectors, conducting empirical research on English texts, with results showing that classification based on deep learning models has certain advantages.

However, for unstructured documents such as microblogs and social media, classification effects based on machine learning methods are not ideal. Due to the 140-character limit, tweets are short in length, typically expressing 1-2 sentences, and contain emotional symbols and internet slang that machine learning methods are relatively insensitive to. Machine learning-based sentiment analysis also requires extensive corpus training and manual intervention, which is time-consuming. Therefore, this paper adopts a sentiment dictionary-based approach for tweet processing.

2 Tweet Topic Mining and Sentiment Analysis Methods

2.1 Topic Mining Method

This article mines topics based on the LDA topic model algorithm. It is generally believed that the most critical aspect of applying LDA is determining the optimal number of topics, as the effectiveness of LDA topic extraction is directly related to the number of potential topics [29]. Scholars have proposed various methods to determine the optimal number of topics, including minimum perplexity algorithm, HDP algorithm, and Bayesian algorithm [30]. Considering these options, this paper adopts the Coherence method proposed by R. Michael et al. in 2015 as the evaluation standard for model quality, selecting the model with the maximum Coherence value to determine the optimal number of topics. The topic mining process is shown in Figure 1 [Figure 1: see original paper]. After data preprocessing and word segmentation, word frequency features are extracted to construct a document-word matrix. The Coherence value is used to determine the optimal number of topics, the LDA model is built, and topics are mined.

2.2 Sentiment Analysis Method

When applying sentiment dictionary-based sentiment analysis methods, this paper constructs and expands the sentiment dictionary based on existing research. For Chinese and English tweets, the following methods are respectively adopted:

(1) Chinese tweet sentiment analysis is based on the Dalian University of Technology Emotion Lexicon Ontology to expand sentiment words. This lexicon categorizes emotions into seven types: "joy, good, anger, sorrow, fear, evil, surprise," and defines emotion intensity. However, it does not address the relationship between sentiment words and degree adverbs, negation words, emoticons, etc., in sentences. In Chinese sentence patterns, both the positional relationship of negation and degree adverbs before or after sentiment words and the number of negation words affect the overall sentiment value. The authors comprehensively consider the effects of negation and degree adverbs on sentiment words in sentences, building separate negation word dictionaries and degree adverb dictionaries. Drawing on Yang Xi's [31] six sentiment word combination methods (see Table 1), the interaction between negation and degree adverbs is comprehensively considered to calculate tweet sentiment values. When calculating each tweet's sentiment value, each sentiment word is used as a benchmark to identify the positional relationship between negation and degree adverbs, and sentiment values in the text are accumulated. The seven-dimensional sentiment calculation formula is as follows:

$$E_{TW-i} = \sum_{j=0}^{K} e_i \cdot (-1)^N \cdot P$$

where i represents a certain category among the seven emotion categories, E_{TW-i} represents the sentiment value of a tweet in category i, K represents the number of sentiment words appearing in a tweet, e_i represents the emotion intensity of a sentiment word in category i, N represents the number of negation words related to that sentiment word, and P represents the weight value of degree adverbs.

(2) English tweet sentiment analysis uses SentiWordNet 3.0 built on WordNet to expand the sentiment dictionary. SentiWordNet 3.0 currently contains 117,659 words. Using a random walk model, each word under each Synset is assigned a PosScore (positive sentiment value) and NegScore (negative sentiment value). English sentiment analysis also constructs negation and degree adverb dictionaries and uses the emoticon dictionary built during Chinese sentiment analysis. The sentiment value calculation formula for each English tweet is:

$$E = \sum_{j=0}^{K} e_{pn} \cdot (-1)^N \cdot P + Pos - Neg$$

where E represents the positive or negative sentiment value of an English tweet, K represents the number of all sentiment words in a tweet, e_{pn} represents the sentiment intensity of a sentiment word in positive and negative dimensions, N represents the number of negation words related to that sentiment word, P represents the weight value of degree adverbs, Pos represents the number of positive emoticons, and Neg represents the number of negative emoticons.

(3) To more clearly present the sentiment tendency of tweets, the ternary sentiment polarity of each tweet is calculated. Ternary sentiment polarity includes positive, neutral, and negative, with the specific calculation formula as follows:

$$E_p = \sum_{j=0}^{K} e_p \cdot (-1)^N + Pos - Neg$$

where E_p represents the sentiment polarity of a tweet, with 1 indicating positive, 0 indicating neutral, and -1 indicating negative. e_p represents the polarity of a sentiment word, N represents the number of negation words related to that sentiment word, Pos represents the number of positive emoticons, and Neg represents the number of negative emoticons.

2.3 Topic-Sentiment Cross-Analysis

Topic-sentiment analysis combines the results of topic mining and sentiment analysis to obtain the trend of emotional tendencies under different topics over time. The specific implementation process is: through topic mining, the topic probability distribution of each tweet is obtained; through sentiment analysis, the sentiment value of each tweet under different topics is obtained; cumulative calculations are performed by topic to finally derive the changes in sentiment values under different topics over time.

3 Empirical Study

3.1 Tweet Collection and Data Overview

Using "OneBeltOneRoad," "OBOR," and "一带一路" as keywords, a total of 102,029 relevant tweets were collected between January 1 and December 31, 2017. Data samples are shown in Table 2 . To ensure experimental accuracy, low-relevance data from Twitter's fuzzy search and data in other languages such as French and Spanish were removed, ultimately retaining 63,907 tweets for in-depth analysis. This included 11,457 Chinese texts and 52,450 English texts, posted by 23,706 Twitter users. These users came from different regions and had different identities. Among known geographical information, users from North America (United States and Canada) accounted for 18%, Jakarta, Indonesia for 8%, Kuala Lumpur, Malaysia for 4%, and Beijing, China for 4%. Based on the collected user data, users can be categorized into media, journalists, politicians, experts, supporters, and general individual users, showing the following characteristics: media users initiate topics, while journalists, politicians, experts, and general individual users jointly promote topic discussions; supporters consistently express supportive or positive views; and media accounts post more than individual accounts. The monthly distribution of tweets on the Belt and Road theme in 2017 is shown in Figure 2 [Figure 2: see original paper].

It can be seen that the number of tweets on the Belt and Road theme remained stable from January to April 2017, increased significantly in May when China held the Belt and Road Forum for International Cooperation, reaching a peak; the number dropped sharply in June, then remained relatively stable but at a higher overall level than January-April, reflecting to some extent that the forum in May attracted high and concentrated attention from Twitter users and stimulated discussions in subsequent months.

To improve Chinese tweet segmentation accuracy, 360 official news reports related to the Belt and Road Initiative were collected. Using "new word discovery" and TF-IDF algorithms, the main keywords of each report were extracted to create a custom segmentation dictionary containing 8,242 words. The Jieba segmentation tool was used for word segmentation. For English tweets, lemmatization was adopted, considering a word's part-of-speech in the text before lemmatizing the word. After completing word segmentation, noun words in Chinese and English tweets were extracted, and tweet word clouds were generated based on word frequency, as shown in Figure 3 [Figure 3: see original paper].

It can be observed that Chinese word frequency is significantly higher than English word frequency, with most high-frequency words related to the Belt and Road Forum. The top ten Chinese and English high-frequency words are shown in Table 3 .

3.3 Topic Mining Results

Through experiments, the Coherence values of Chinese tweets under 3-15 topics were calculated to find the topic number with the maximum value as the optimal number of LDA topics. Experiments showed that when the number of topics is 6, the Coherence value is maximized, so the number of Chinese tweet topics was set to 6. Based on LDA model analysis, six hot topics in Chinese tweets were identified: Belt and Road and Korean Peninsula peace issues, Belt and Road and economic issues, Belt and Road and high-level visits at the forum, Belt and Road and cooperation and projects at the forum, Belt and Road and diplomatic strategy, and Belt and Road and domestic hotspots. The related keywords are shown in Table 4 .

The "Belt and Road and Korean Peninsula peace issues" topic highlights the tense surrounding situation caused by North Korea's missile test on April 29, 2017, with neighboring countries such as South Korea and Russia paying close attention, also indicating that Belt and Road development depends to some extent on the surrounding international environment. The "Belt and Road and economic issues" topic highlights the importance of economic development and financial cooperation, with the Belt and Road Initiative bringing opportunities for RMB internationalization, closely related to all aspects of people's lives. The third and fourth topics focus on high-level visits and cooperation and projects at the Belt and Road Forum, highlighting international attention and participation by heads of state from multiple countries, as well as the forum's promotion of globalization cooperation in politics, investment, education, culture, and other areas. The "Belt and Road and diplomatic strategy" topic highlights important diplomatic areas such as infrastructure, railways, trade, and investment, with facility connectivity being a priority area for Belt and Road construction, and cooperation projects facing both opportunities and challenges. The "Belt and Road and domestic hotspots" topic highlights the active response from domestic provinces and cities such as Beijing, creating opportunities for more domestic e-commerce companies to "go global."

The Coherence values of English tweets varying with topic number were calculated. Experiments showed that when the topic number is 4, the Coherence value reaches a maximum of 0.79, so the topic number was set to 4 for LDA topic mining. Based on this, four hot topics in English tweets were identified: Belt and Road and CPEC, Belt and Road and foreign cooperation, Belt and Road and economic effects, and Belt and Road and foreign policy. The related keywords are shown in Table 5 .

The "Belt and Road and CPEC" topic highlights China-Pakistan cooperation, where CPEC is the abbreviation for China-Pakistan Economic Corridor, known as the hub connecting the northern and southern Silk Roads, aimed at strengthening exchanges and cooperation in transportation, energy, ocean, and other fields between China and Pakistan to promote common development. PetroChina is an important force in building the China-Pakistan natural gas pipeline, providing a good demonstration for regional connectivity and longer-term strategic cooperation. The "Belt and Road and foreign cooperation" topic once again highlights the importance of infrastructure cooperation, with more discussions on cooperative countries such as Russia and India and international relations. The "Belt and Road and economic effects" topic highlights the economic blueprint of Belt and Road globalization development, including global power, global trade, economic cooperation, and joint connectivity. The "Belt and Road and foreign policy" topic involves overall diplomatic initiatives such as the "New Silk Road" and "economic corridor," as well as inter-regional cooperation policies, receiving widespread attention and heated discussion.

3.4 Sentiment Analysis Results

By judging sentiment polarity to identify positive, negative, and neutral ternary sentiment attitudes in texts and quantifying scores to obtain sentiment intensity, the monthly ternary sentiment analysis results for Chinese tweets on the Belt and Road theme in 2017 are shown in Figure 4 [Figure 4: see original paper].

Overall, among Chinese users' attitudes toward the Belt and Road Initiative, positive and neutral sentiments are higher than negative sentiment. Neutral sentiment was dominant from January to July, August was a turning point, and after September, users' neutral sentiment gradually became distinct, with positive sentiment rising steadily and linearly, reaching the highest point of 0.48 in December. Negative sentiment also increased slightly and tended to stabilize.

In comparison, emotions expressed in English tweets were relatively stable, with annual averages of neutral, positive, and negative sentiment values at 0.51, 0.37, and 0.12, respectively. Sentiment changes before and after the summit forum in May were not significant. The convening of the 19th National Congress in October attracted foreign media attention. The forum highlighted the influence of the Belt and Road Initiative, which reached a new level of development. It was in October that sentiment values showed a clear turning point, with sentiment tendencies becoming more defined, positive sentiment steadily increasing, surpassing neutral sentiment proportion in December, and negative sentiment also rising relatively slightly.

The sentiment analysis results for English tweets are shown in Figure 5 [Figure 5: see original paper].

3.5 Topic-Sentiment Cross and Evolution Analysis

As time progresses, users' discussion topics continuously change, and sentiment values under different topics also change. The topic sentiment evolution results for Chinese tweets in 2017 are shown in Figure 6 [Figure 6: see original paper].

It can be seen that the "Belt and Road and Korean Peninsula peace issues" topic shows significant emotional fluctuations, with negative sentiment values clearly rising from April-June and August-September. The reasons for this change are speculated to be related to the following: North Korea's sixth nuclear test in September 2017 was condemned by countries worldwide, and the Chinese government also expressed regret and opposition; the North Korean missile issue, apart from the September nuclear test, involved missile launches in February, March, April, September, and November, leading to tense situations around the Korean Peninsula. The "Belt and Road and economic issues" topic shows positive sentiment values higher than neutral and negative from March to the end of the year, with the proportion of positive sentiment values exceeding 50% by year-end. The "Belt and Road and high-level visits at the forum" and "Belt and Road and cooperation and projects at the forum" topics are both related to the forum. The high-level visits topic showed a negative peak in June, when Japanese Prime Minister Abe expressed Japan's participation in the Belt and Road plan, which many Chinese tweet users opposed. Regarding cooperation, users with positive and neutral attitudes were in the majority, while by year-end, the proportion of tweets with negative attitudes increased significantly. In November 2017, foreign media began to express concerns about Sri Lanka, Pakistan, and other countries, believing that excessive introduction of Chinese capital into these countries along the Belt and Road would lead to excessive economic dependence and loss of national decision-making power. In the "Belt and Road and diplomatic strategy" topic, neutral attitudes are mainstream, reaching over 70% at most times, possibly because most users reposted facts about China signing agreements with multiple countries and diplomatic strategies, which is relatively neutral. Under the "Belt and Road and domestic hotspots" topic, which includes many place names and regional information, tweet content mostly concerns achievements and publicity of certain domestic regions in Belt and Road projects. In this topic, positive sentiment values dominated after May, with low and stable negative sentiment proportions, indicating that Chinese users are relatively optimistic about Belt and Road and domestic development.

The topic sentiment evolution results for English tweets in 2017 are shown in Figure 7 [Figure 7: see original paper].

It can be seen that the "Belt and Road and CPEC" topic of China-Pakistan cooperation has attracted interest from foreign English tweet users, with most focusing on the economic effects of the China-Pakistan Economic Corridor. Domestic and foreign reports on CPEC are relatively optimistic, such as French media [32] estimating in July that CPEC created 300,000 local jobs. Sentiment fluctuations under this topic are small, with neutral sentiment in the majority, and positive sentiment values showing an overall improvement after October. The "Belt and Road and foreign cooperation" topic shows large sentiment fluctuations, possibly related to China frequently signing agreements with other countries after the May summit. By July, a total of 2,431 cooperation agreements had been signed with 61 countries, with new contract amounts of $71.42

billion. From July to the end of the year, the proportion of positive sentiment values was higher than the other two, and negative sentiment values regarding foreign cooperation also increased. The "Belt and Road and economic effects" topic shows positive sentiment value proportions higher than the other two in most time periods. According to the Ministry of Commerce website, in 2017, Chinese enterprises made $14.36 billion in non-financial direct investment in 59 countries along the Belt and Road [33], and English tweets gradually became optimistic about the economic effects brought by the Belt and Road Initiative. The "Belt and Road and foreign policy" topic shows gradually rising positive sentiment after the May summit forum. The forum strengthened policy communication and strategic alignment, signed multiple bilateral and multilateral cooperation documents and enterprise cooperation projects [34], and received positive responses.

4 Conclusion

The year 2017 was a year of breakthrough progress for the Belt and Road Initiative. The Belt and Road Forum for International Cooperation was held, the Mombasa-Nairobi Railway officially opened, the first production line of the Yamal liquefied natural gas project was put into operation, many projects were gradually implemented, new cooperation agreements were continuously signed, and the term "Belt and Road" became a hot word worldwide. News reports mostly presented facts from an official perspective, while domestic and foreign reactions were difficult to clarify. Numerous users worldwide discussed the Belt and Road Initiative on Twitter, which better reflects users' concerns and emotional tendencies. Based on LDA analysis of Chinese and English Twitter texts, Chinese tweets in 2017 had six hot topics: Belt and Road and Korean Peninsula peace issues, Belt and Road and economic issues, Belt and Road and high-level visits at the forum, Belt and Road and cooperation and projects at the forum, Belt and Road and diplomatic strategy, and Belt and Road and domestic hotspots. English tweets had four hot topics: Belt and Road and CPEC, Belt and Road and foreign cooperation, Belt and Road and economic effects, and Belt and Road and foreign policy. In comparison, Chinese tweet users focus on more micro-level issues, show high interest in the Belt and Road Forum, emphasize a "cooperation" attitude, and view the Belt and Road Initiative from a "cooperation" perspective, valuing the "cooperation" process. English tweet users focus on more macro-level topics, with more discussions on overall trends and development, evaluating the Belt and Road Initiative more from its economic effects and development situation. This indirectly reflects that Chinese tweets position the Belt and Road Initiative as a "regional cooperative development" project, while English tweets view it as an "economic cooperation" project.

From the perspective of Chinese tweet topic sentiment evolution patterns, except for the Korean Peninsula peace issue, positive and neutral emotions dominate other topics, with negative emotions accounting for the smallest proportion, and neutral emotions showing a downward trend after the May forum, making users' emotional tendencies more defined. English tweets also have positive and neutral sentiment values occupying the main position. Except for the Belt and Road and diplomatic policy topic, other topics show small emotional fluctuations, with fluctuations mainly appearing around October, indicating that important events in October such as the 19th National Congress had a significant impact on emotions. This paper attempts to present the content and emotional tendencies of international social media attention to the Belt and Road Initiative through topic mining and sentiment analysis methods. In the future, it will be necessary to expand to multilingual tweets such as Spanish and French to more comprehensively present international responses and focus points regarding the Belt and Road Initiative.

References

[1] Huang Yanqiu. A Study on China's Communication Strategy Towards Africa Under the Belt and Road Initiative from the Perspective of Constructivist International Relations: Based on Text Data from China Daily Africa Edition[D]. Wuhan: Central China Normal University, 2017.

[2] Zhu Guisheng, Huang Jianbin. The Belt and Road Strategy in the Vision of American Mainstream Media: A Critical Discourse Analysis of Related Reports in The Washington Post[J]. Press Circles, 2016(17): 58-64.

[3] Gong Yanhao, Zhen Feng, Xi Guangliang. Spatial Pattern of Attention and Response to the Belt and Road Initiative: An Analysis Based on Sina Weibo Data[J]. Areal Research and Development, 2018, 37(2): 29-35.

[4] Jia Shuang. The Belt and Road: Twitter Network Public Opinion Analysis and Countermeasures[D]. Nanjing: Nanjing University, 2016.

[5] Hofmann T. Probabilistic Latent Semantic Indexing[J]. Sigir Forum, 2017, 51(2): 211-218.

[6] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research, 2003, 3(1): 993-1022.

[7] Chen Xiaomei, Gao Cheng, Guan Xinhui. LDA Topic Model Method for Network Public Opinion Viewpoint Extraction[J]. Library and Information Service, 2015, 59(21): 21-26.

[8] Michelson M, Macskassy S A. Discovering Users' Topics of Interest on Twitter: A First Look[C]//Workshop on Analytics for Noisy Unstructured Text Data. New York: ACM, 2010.

[9] Yoo Sun H, Hong Jins. Opinion Leadership on Twitter and Twitter Use-Motivations and Patterns of Twitter Use and Case Study of Opinion Leaders on Twitter[J]. Korean Journal of Broadcasting and Telecommunication Studies, 2010, 24(6): 365-404.

[10] Hu Y, John A, Seligmann D D. Event Analytics via Social Media[C]//ACM

Workshop on Social and Behavioral Networked Media Access. New York: ACM, 2011: 39-44.

[11] Mei Q, Zhai C X. A Mixture Model for Contextual Text Mining[C]//Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2006: 649-655.

[12] Moghaddam S, Ester M. ILDA: Interdependent LDA Model for Learning Latent Aspects and Their Ratings from Online Product Reviews[C]//Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2011: 665-674.

[13] Blei D M, Lafferty J D. Dynamic Topic Models[C]//Proceedings of the 23rd International Conference on Machine Learning. New York: ACM, 2006: 113-120.

[14] Alsumait L, Barbara D, Domeniconi C. Online LDA: Adaptive Topic Models for Mining Text Streams with Applications to Topic Detection and Tracking[C]//Eighth IEEE International Conference on Data Mining. Washington, DC: IEEE, 2008: 3-12.

[15] Mikolov T, Sutskever I, Chen K, et al. Distributed Representations of Words and Phrases and Their Compositionality[C]//Advances in Neural Information Processing Systems. New York: Curran Associates, Inc., 2013: 3111-3119.

[16] Go A, Bhayani R, Huang L. Twitter Sentiment Classification Using Distant Supervision[R]. Cs224n Project Report. Palo Alto: Stanford University, 2009.

[17] Jansen B J, Zhang M, Sobel K, et al. Twitter Power: Tweets as Electronic Word of Mouth[J]. Journal of the American Society for Information Science & Technology, 2009, 60(11): 2169-2188.

[18] Tumasjan A, Sprenger T O, Sandner P G, et al. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment[C]//International Conference on Weblogs and Social Media, Icwsm 2010. Washington, DC: DBLP, 2010.

[19] Bollen J, Pepe A, Mao H. Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena[J]. Computer Science, 2009, 44(12): 2365-2370.

[20] Sheridan D P, Decker H K, Kloumann I M, et al. Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter[J]. PLOS ONE, 2011, 6(12): e26752.

[21] Hu M, Liu B. Mining and Summarizing Customer Reviews[C]//Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Seattle: ACM, 2004: 168-177.

[22] Zhang L, Ghosh R, Dekhil M, et al. Combining Lexicon-Based and Learning-Based Methods for Twitter Sentiment Analysis[EB/OL]. [2018-04-03]. https://www.hpl.hp.com/techreports/2011/HPL-2011-89.pdf.

[23] Sentiment Analysis Word Set Released by CNKI[EB/OL]. [2018-03-08]. http://www.keenage.com/html/c_{{{bulletin}}_{{2007}}}.htm.

[24] Feldman R. Techniques and Applications for Sentiment Analysis[M]. New York: ACM, 2013.

[25] Volkova S, Wilson T, Yarowsky D. Exploring Sentiment in Social Media: Bootstrapping Subjectivity Clues from Multilingual Twitter Streams[C]//Proceedings of ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics. Sofia: ACL, 2013: 505-510.

[26] Volkova S, Wilson T, Yarowsky D. Exploring Demographic Language Variation to Improve Multilingual Sentiment Analysis in Social Media[C]//Proceedings of Conference on Empirical Methods in Natural Language Processing. Seattle: ACL, 2013: 1815-1827.

[27] Pang B, Lee L, Vaithyanathan S. Thumbs Up?: Sentiment Classification Using Machine Learning Techniques[C]//Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10. Philadelphia: ACM, 2002: 79-86.

[28] Zhang Zhihua. Research on Sentiment Word Vectors and Text Sentiment Analysis Based on Deep Learning[D]. Wuhan: East China Normal University, 2016.

[29] Fan Yunman, Ma Jianxia. Research on Emerging Topic Detection Based on LDA and Emerging Topic Feature Analysis[J]. Journal of the China Society for Scientific and Technical Information, 2014, 33(7): 698-711.

[30] He Liang, Li Fang. Topic Discovery and Trend Analysis of Scientific and Technical Literature Based on Topic Models[J]. Journal of Chinese Information Processing, 2012, 26(2): 109-116.

[31] Yang Xi. Research on a Microblog Sentiment Analysis Model Based on Combination of Sentiment Dictionary and Rules[D]. Hefei: Anhui University, 2014.

[32] French Media Says China-Pakistan Economic Corridor Revitalizes Pakistan[EB/OL]. [2018-04-03]. http://news.163.com/17/0805/00/CR1LUS9A00018AOQ_{all}.html.

[33] Investment and Cooperation in Countries Along the Belt and Road in 2017[EB/OL]. [2018-04-03]. http://fec.mofcom.gov.cn/article/fwydyl/tjsj/201801/20180102699450.shtml.

[34] Belt and Road Forum for International Cooperation "Policy Communication" Parallel Thematic Meeting Signed 32 Cooperation Agreements[EB/OL]. [2018-04-03]. http://www.xinhuanet.com/2017-05/14/c_{1120970716}.htm.

Author Contributions

Zhao Changyu: Conceived the research topic, proposed the research framework, and wrote the paper.

Wu Yaping: Revised the paper and adjusted the framework.
Wang Jimin: Proposed the research ideas and revised the paper.

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