Revealing Public Opinion towards the COVID-19 Vaccine with Weibo Data in China: BertFDA-Based Model

Muni Zhuang 24520220157291(Cs), Yutong Fu 24520221154574(Cs), Rui Zhao 31520221154239(Ai), Jiajian Li 36920221153093(Ai), Wanqi Wu 24520221154584(Cs)

Abstract

The COVID-19 pandemic has created unprecedented burdens on people's health and subjective well-being. While countries around the world have established models to track and predict the affective states of COVID-19, identifying the topics of public discussion and sentiment evolution of the vaccine, particularly the differences in topics of concern between vaccine-support and vaccine-hesitant groups, remains scarce. Using social media data from the two years following the outbreak of COVID-19 (23 January 2020 to 23 January 2022), coupled with state-of-the-art natural language processing (NLP) techniques, we developed a public opinion analysis framework (BertFDA). First, using dynamic topic clustering on Weibo through the latent Dirichlet allocation (LDA) model, a total of 118 topics were generated in 24 months using 2,211,806 microblog posts. Second, by building an improved Bert pretraining model for sentiment classification, we provide evidence that public negative sentiment continued to decline in the early stages of COVID-19 vaccination. Third, by modeling and analyzing the microblog posts from the vaccine-support group and the vaccine-hesitant group, we discover that the vaccine-support group was more concerned about vaccine effectiveness and the reporting of news, reflecting greater group cohesion, whereas the vaccine-hesitant group was particularly concerned about the spread of coronavirus variants and vaccine side effects. Finally, we deployed different machine learning models to predict public opinion. Moreover, functional data analysis (FDA) is developed to build the functional sentiment curve, which can effectively capture the dynamic changes with the explicit function. This study can aid governments in developing effective interventions and education campaigns to boost vaccination rates.

Introduction

According to statistics from the National Health Commission of the People's Republic of China, as of 20 September 2022, China has reported a total of 3.44 billion doses of COVID-19 vaccines, with over 1.3 billion persons having completed the whole immunization (China Bureau of Disease Control and Prevention 2022). In 2021, China preliminarily established the national immunological barrier through mass vaccination against COVID-19 via the national system and entered the third stage of regular epidemic prevention and control (Wang et al. 2021). Although higher vaccination rates reduce the severity of breakthrough infections, there is evidence that the efficacy of one or two doses of the vaccine is decreased after six months, and COVID-19 variants strain may evolve frequently (Del Rio, Omer, and Malani 2022). This indicates that even if vaccination rates rise, the great majority of individuals will still be infected with COVID-19. The third dosage of vaccination, as well as the children's vaccine, is critical in averting a pandemic comeback. The understanding of the public's emotional reactions and willingness to receive the vaccine is critical for targeted decision making during the early stages of vaccination planning, so as to avoid vaccine hesitancy and improve the effectiveness of vaccination programs.

It is known that social media has become the major channel for people to express thoughts on COVID-19 vaccination with the emergence of the epidemic and the implementation of the lockdown policy (Tsao et al. 2021). Social media has had a tremendous impact on the public's attitude regarding vaccination. As a result, it is critical for governments, public health officials, and policymakers to understand the potential drivers that influence public sentiment regarding COVID-19 vaccination. In research related to the COVID-19 vaccine, using social media data for academic research has become an emerging trend. Social media provides a rich volume of real-time and cost-effective content including news, events, public comments, etc., (Hu et al. 2021), which has been widely used in health-related issues and public health crises (Lu and Brelsford 2014; Guo et al. 2021; Zhuang et al. 2021; Fang et al. 2021). However, research on the COVID-19 vaccine mainly employs the classic time series analysis based on discrete observation data. The dynamic change of the emotion function is frequently ignored. Moreover, traditional approaches typically employ the moving average to smooth the noise of high-frequency emotion, which makes it impossible to accurately generate the potential random process of actual observation. Finally, despite sentiment analysis models having been broadly applied in public opinion analysis, most of them employ traditional machine learning or simple sentiment analysis tools, ignoring the rich contextual semantic information hidden in the text, which makes the results of sentiment classification deviate greatly.

We developed the public opinion analysis framework based on FDA combined with the deep learning transfer model Bert. Specifically, the following questions are addressed:

1. How do we use the deep learning algorithm to capture the

profound semantic and emotional information behind the microblog posts more accurately?

2. How do we construct an intrinsic function to describe the dynamic characteristics of emotional evolution?
3. What quantitative measurements can be used to assess the continuity and popularity of topics?

We provides an actionable solution for depicting and predicting the dynamic characteristics of COVID-19 vaccination hesitancy. We used two years of social media data to detect subtle changes in emotions through the deep learning transfer model and explore the changes in topics in different periods through the calculation of topic continuity and popularity. The FDA obtains data with a higher signal-to-noise ratio and more accurately constructs the intrinsic public emotion to better investigate the dynamic evolution of emotion. Finally, we forecasted the public emotional evolution and the progress of vaccination. Our findings may also provide useful insights for the promotion of other vaccinations.

Related Work

Sentiment analysis is the field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language and is one of the most active research areas in natural language processing (Liu 2012). The dimensions and techniques of sentiment analysis in social media texts, which are task orientation, granularity, and methodology, were reviewed by Yue (Yue et al. 2019). Cambria et al. (Cambria et al. 2017) summarized sentiment analysis in social media texts using knowledge-based, statistical, and hybrid methods. Zhang et al. (Zhang, Wang, and Liu 2018) reviewed sentiment analysis tasks at the document, sentence, and aspect granularities. Nevertheless, not all thoughts are expressed clearly, especially implicit emotions that require natural language understanding (NLU), such as metaphors, sarcasm, and irony (Alswaidan and Menai 2020). Keyword-based and rule-based research methodologies are used in early sentiment categorization tasks (Hutto and Gilbert 2014; Loria et al. 2019), which require a great deal of manpower and time, so they are gradually replaced by traditional machine learning (Soleymani et al. 2017; Ravi and Ravi 2015; Hussein 2018; Medhat, Hassan, and Korashy 2014), although traditional machine learning's growth is quite limited (such as support vector machines, naive Bayes, k-nearest neighbors, hidden Markov models, conditional random fields, multi-layer perceptrons, etc.). At present, deep learning has become the mainstream sentiment analysis. Long short-term memory (LSTM) is the most used deep learning model, which is a special form of a recurrent neural network (RNN) with the capability of handling long-term dependencies (Tai, Socher, and Manning 2015), and the vanishing or exploding gradient in RNNs has been effectively alleviated during data transmission. However, when it comes to longer-term dependencies, LSTM is still powerless. Therefore, Vaswani et al. proposed the Transformer, a model architecture that eschews recurrence and instead relies entirely on an attention mechanism to draw global dependencies between the input and output (Vaswani et al. 2017), achieving a new state-of-the-art translation quality. Subsequently, Devlin et al. improved the fine-tuning-based approaches by proposing Bidirectional Encoder Representations from Transformers (Bert), which achieved new state-of-the-art results on 11 natural language processing (NLP) tasks and pioneered the use of emotion classification (Lee and Toutanova 2018).

Because social media text has characteristics such as short text, noise, multilingualism, metaphor, irony, and so on, many topic models have been developed to quickly mine hot themes from massive unstructured text data. Latent Dirichlet Allocation (LDA) is currently the most widely used probabilistic topic model. The application of sentiment analysis in the medical field has been faster and more extensive than ever since the outbreak and spread of the COVID-19 epidemic. Multiple factors, such as public knowledge, emotions, and personal health decisions, influence public acceptance of medical interventions involving infectious diseases and vaccines. Lyu et al. (Lyu, Le Han, and Luli 2021) used the LDA algorithm and an emotion lexicon to track topics and emotions in public discussions about the COVID-19 vaccine on Twitter. Hu et al. (Hu et al. 2021) used Twitter data to reveal the American public's opinion of the COVID-19 vaccine from a spatiotemporal perspective. Monselise et al. (Monselise et al. 2021) employed non-negative matrix factorization (NMF) to determine vaccine topics, then used VADER sentiment analysis libraries and sentence bidirectional encoder representations from transformer embeddings to identify emotional content and compared the embedding to different emotions using cosine similarity. Gbashi et al. (Gbashi et al. 2021) systematically scrutinized media communications (Google news headlines or snippets, and Twitter posts) using three standard computational linguistics models (i.e., TextBlob, VADER, and Word2Vec-BiLSTM) to understand the prevailing sentiments in Africa on the COVID-19 vaccine. Cruickshank et al. first examined the prevalence, dynamics, and content of websites shared in vaccination-related tweets. The research found that sharing websites is a common communication strategy, and its "bursty" pattern and inauthentic propagation strategy pose challenges to health promotion (Cruickshank et al. 2021). Ginosar et al. examined the content of YouTube videos shared in vaccine-related tweets before the COVID-19 vaccine rollout. The research discovered the role of cross-platform sharing of YouTube videos over Twitter as a strategy to propagate anti-vaccination messages (Ginossar et al. 2022).

Methodology

Sentiment Analysis Based on Bert

Bert is a new pre-training method for language representations. The semantic representation ability of the model is enhanced through the masked language model (MLM) and next sentence prediction task (NSP). Besides, it has achieved many NLP tasks depending on Transformer's powerful feature extraction and fine-tuning transfer learning abilities.

The Bert model lacks the training of emotion corpus in the pre-training stage, which leads to its poor performance

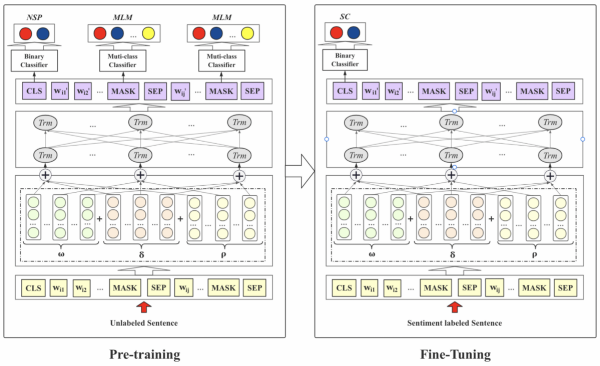

Figure 1: Structure of sentiment classification by Bert deep learning model.

in performing emotion classification tasks. To im-prove the accuracy and granularity of large-scale complex text in emotion classifica-tion tasks, we developed a new pre training task for Bert and introduced an improved pre-training corpus set $TB = \{TB_i \mid i \in \{1, 2, \ldots, M\}\}$. That is, in addition to the original Chinese Wikipedia corpus, we also added the public Sina Weibo and Baidu Tieba emotional corpus TW (https://github.com/SophonPlus/ChineseNlpCorpus), hoping that the model could learn more emotional information. At the same time, we also introduced the public Sina Weibo annotation set and a small number of emotional annotation sets of public health emergencies as the BERT in-depth pre-training corpus Finally, the final emotion classification was obtained by fine-tuning the model on the Weibo dataset $TC = \{TC_i \mid i \in \{1, 2, \ldots, M\}\}$The model is described as follows:

ω, δ, ρ denote token embeddings, segment embeddings, and position embeddings, respectively; Trm indicates the encoder unit of the transformer; $d'_i = \{w'_{ij} \mid j \in \{1, 2, \ldots, N_i\}\}$denotes the vector set, which combines the words of the document d_i and improved full-text semantic information.

In the pre-training phase (see Figure 1), after the segmented document $d_i = \{w_{ij} \mid j \in \{1, 2, \ldots, N_i\}\}$ was input into the model, each word was mapped into three vectors $W_{ij}(\omega + \delta + \rho)$, called word embedding. The residual network structure connects the multi-head mechanism and the feed-forward layer via the transformer encoder. The mul-ti-head method calculates the attention weight by performing numerous linear trans-formations on the input vector.

Thus, the transformer encoder captures and stores the semantic relationship and grammatical structure information of document D. It is connected with the output layer of Softmax to adapt to transfer learning under multitasking. In this paper, we first initialize the Bert model with pre-trained parameters, and then all of the parame-ters were fine-tuned using labeled sentimental classification data.

BertFDA Framework for Public Opinion Analysis

BertFDA is built with the goal of accurately simulating the evolution of large-scale network public opinion, grasping the evolution characteristics and laws of groups in real-time, and assisting government departments in rapidly forming an effective pub-lic opinion response mechanism. Figure 2 illustrates the process framework based on BertFDA. The description of the algorithm framework is as follows:

Step 1: Data gathering and preprocessing: The public opinion corpus of Weibo is crawled through web crawler technology to obtain the public opinion data related to the COVID-19 vaccine. It is preprocessed by format conversion, removal of stop words, and word segmentation to form an emotional corpus dictionary, and each word cor-responds to a unique index.

Step 2: Word embedding and LDA model: After inputting the corpus set TB into the Bert pre-training model, each word would be mapped to word embeddings $TB_{ij}(\omega + \delta + \rho)$.Then, it is input into the LDA model to improve the training of the topic vector μ. A better result μ' is obtained after iterative calculation, that is, the probability distributions of ι optimal topics and different “topic words” are derived.

Step 3: Building Bert’s sentiment classifier: The feature vector $TB_{ij}(\omega + \delta + \rho)$ output from Step 2 is introduced into the bidirectional transformer encoder, and then a single-layer neural network is constructed to connect the output vector corresponding to [CLS] in the transformer as the classifier to perform sentiment classification (SC). Simultaneously, the two pre-training tasks of MLM and NSP are retained and connected to the output vector corresponding to [MASK] and [CLS], respectively. Finally, the corpus TW is deeply pre-trained in the target field, and the COVID-19 vaccine corpus TC is finetuned to output the emotional classification and emotional value of the corpus

Step 4: FDA modeling: Taking the emotional time series as the input, cross-validation (CV) is used to estimate the number of basis functions, and the unde-termined coefficients of the model are obtained by the least-squares method. Ulteriorly, the intrinsic sentiment can be built. Finally, we obtain the sentiment evolution based on the function curve.

Step 5: Revealing public opinion and prediction: The public opinion of the COVID-19 vaccine is examined across four dimensions using the procedures outlined above: Sentiment classification and topic clustering using time series, topic emotion mixed analysis, and sentiment prediction with machine learning.

Experiments

Data Extraction and Preprocessing

Weibo has played a significant role in people’s lives as a source of information and communication. As a result, we selected Sina Weibo as the data source. The Wuhan Epidemic Prevention and Control Headquarters announced the “Wuhan lockdown” on 23 January 2020. Since then, the epidemic has spread, and the number of people discussing it has gradually increased. Therefore, we built a Python-based crawler architecture using the COVID-19 outbreak as a study background and search phrases such as “new crown vaccine” and “new crown vaccination.” It collects 2,597,823 microblog posts from 0 h on 23 January 2020 to 24 h on 23 January 2022 (732 days in total). The data include the

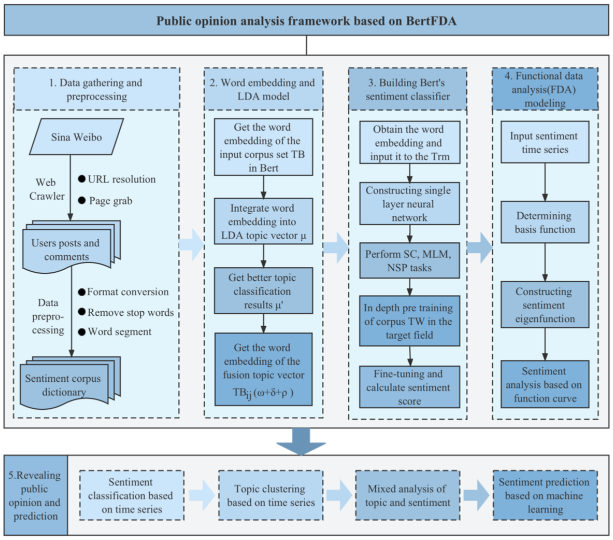

Figure 2: Public opinion analysis framework based on BertFDA.

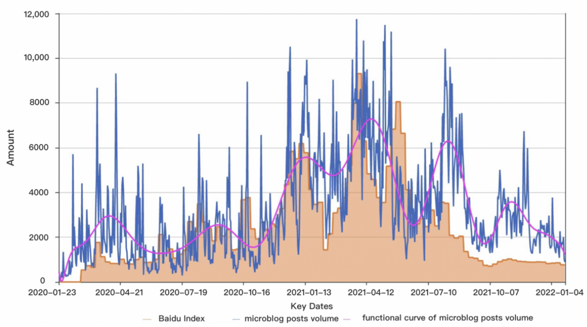

Figure 3: The number of Sina Weibo and Baidu index over the entire study timeline.

username, content, and the time of posts. As the first step in Section 3.4, the collected microblog posts are cleaned by detecting and processing duplicate and missing values, manually filtering irrelevant information such as advertisements and website links, and converting emoticons into text. For fur-ther analysis, the cleaned 2,353,435 valid microblog posts are integrated according to the time series and processed using function smoothing. In addition, we used the Baidu Index (http://index.baidu.com/accessed on 1 Otober 2021) to gauge public attention. As a result, we undertook an integrated analysis of public opinion in the Baidu Index throughout this study time, as well as the microblog post volume we crawled, using the search term "new crown vaccination." Figure 3 depicts the results.

The evolution pattern of the COVID-19 vaccine in the Baidu index is highly con-sistent with Sina Weibo. The public debate and interest in the COVID-19 vaccine peaked between November 2020 and August 2021. Combined with specific events, the peak in October 2020 is related to the massive discussion when Pfizer revealed that their COVID-19 vaccine provides 90% protection. The spike in November 2020 is related to the vaccine being administered to a British woman for the first time. The de-bate increased in waves after China approved the COVID-19 vaccine. In August 2021, it was approved for usage by school pupils aged 3 to 17, and the heat of discussion was reignited.

Tracking Topic over Time

Tracking topics over time allows regulatory authorities to more accurately predict and control emergency risks, resulting in more efficient information services and emergency management. Many researchers divide time into multiple phases to observe the dynamic evolution of topics. In order to examine the theme of sentiment evolution at the various stages. These methods could detect topic fermentation inflection points as well as people's overall sentiment tendencies. On the other hand, most time series studies of public health emergencies are divided according to the overall situation following the event, with rather coarse time unit particles. For the COVID-19 vaccination, we want to explore the theme characteristics and evolutionary laws of microblog posts. Based on the nature of public health emergencies, it is proposed that month be used as the time unit to better observe the public opinion trend. We vectorized each month's microblog posts to obtain feature vector representations of dimensions such as syntax, semantics, and theme, and then performed perplexity and coherence calculations on the optimized text vector to obtain the optimal number of topics, as outlined in step 2 of Section 3.4. In Table A1, we provide the results of the feature word extraction and topic distribution for a total of 24 months.

We can observe that the topic with the highest continuity is "Global development trend of the COVID-19", which occurred 14 times in 24 months and throughout the whole study period. The urgent need of the public for COVID-19 vaccines is closely related to the increasingly severe epidemic. Therefore, the public's most concerning issue is the global epidemic's development trend. The second topic is "The progress of China's COVID-19 vaccine research and development." It appeared 13 times, mainly in the early and middle stages of COVID-19. In terms of the popularity of the topic, its popularity in the early stages exceeded that of all other topics, especially in January and June 2020. In January 2020, the public paid special attention to the research and development plan for the COVID-19 vaccine. Good news about China's vaccine research and development came frequently, triggering heated public discussion. For example, the Chinese Center for Disease Control and Prevention took the lead in isolating the viral strain across the world, and the first batch of vaccines developed in the Zhejiang Province of China successfully induced antibody production and entered the stage of animal trials, etc. Five months later, significant progress had been made in the research and development of multiple vaccines in China.

The topics of "Epidemic prevention and control policies in China's provinces" and "COVID-19 vaccination doses in China" occurred nine and eight times, respectively. "Vaccination in China's provinces" related to this topic also appeared six times in the middle and late stages. In addition, with the continuous variation and spread of COVID-19, relevant discussions also appeared seven times in the middle and late stages. There are also some topics closely related to China's COVID-19 vaccine. In addition, Olympic champion Zhang Shan became the star spokesperson of the COVID-19 vaccine promotional film, calling on everyone to get the COVID-19 vaccine. As a consequence of these encouragement and publicity measures, China's vaccination rate for the COVID-19 vaccine rose rapidly. Therefore, the vaccination dose also became one of the topics of public concern and discussion. Another related topic is China's vaccination reaction. The public is very concerned about the safety, effectiveness, and side effects of the vaccine. The COVID-19 vaccines approved globally have more or less side effects and adverse reactions.

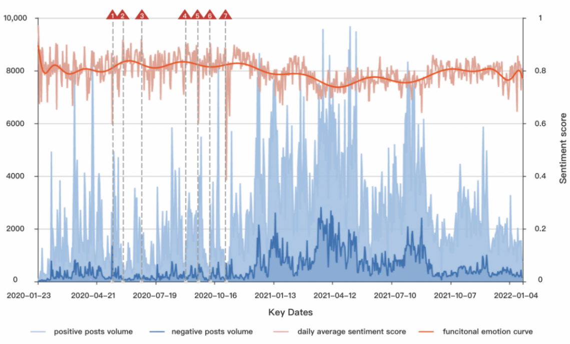

Figure 4: The smooth curve of average sentiment scores, and positive and negative microblog posts volume.

Tracking Sentiment over Time

To gain a comprehensive understanding of the evolution of public sentiment during the whole research period, we constructed the Bert sentiment classifier and performed fine-tuning based on the Bert pre-training model to summarize the sentiment in all 2,353,435 microblog posts, as described in steps 3 and 4 of Section 3.4. We observed 1,962,464 positive and 392,971 negative microblog posts, accounting for 83.3% and 16.7% of all, respectively.

Figure 4 depicts the smooth curve of daily average sentiment scores and positive and negative microblog post volume over a 732-day period beginning on 23 January 2020. This curve is obtained by the FDA method, where the time index is considered as the input and the original sentiment scores calculated by Bert as the output (see the orange curve). As for the number of basis functions, this is determined by minimizing the generalized cross-validation criterion (Liang et al. 2022). We confirmed that the optimal threshold for the positive and negative sentiment was 0.61 by F1-score (Fujino, Isozaki, and Suzuki 2021). The daily average sentiment score and its function smooth curve show that public sentiment was mostly positive. Nonetheless, public opinion began to gradually decline after the vaccination for COVID-19 in December 2020, and as the number of vaccinations increased, so did public dissatisfaction with the vaccine, which did not improve until most citizens had completed the entire vaccination course (October 2021).

The microblog posts volume in Figure 4 shows that as the COVID-19 vaccination work progresses, the public's positive and negative comments are increasing continuously. The cumulative number of vaccinations has gradually increased as a result of the state's vigorous vaccination promotion. Furthermore, the Dynamic COVID-Zero Strategy in China has effectively controlled the epidemic. It may have improved public confidence and mobilized the public's positive mood (Information Office of the State Council 2022a,b). However, people's willingness to be vaccinated is not as optimistic as expected in the early stages of vaccination in China. Many people are taking a wait-and-see approach, questioning the vaccine's safety and side effects. The conclusion is consistent with reference (Du et al. 2021). Moreover, some people exploit a tiny number of vaccine-related adverse events to disseminate unpleasant feelings and terror across the Internet. With the ongoing mutation of COVID-19, the efficacy of vaccination has been doubted, and a growing number of people have a negative attitude toward it.

Conclusions

Drawing on microblog posts from 23 January 2020 to 23 January 2022 (732 days in total), we examined public opinion on the COVID-19 vaccine in China. Firstly, the topic modeling of microblog posts was carried out through the improved LDA model, which effectively identified the keywords and topics discussed by the public at various stages of the COVID-19 vaccine's development and application. Secondly, the opti-mized Bert language pre-training model is utilized to analyze the sentiment of mi-croblog posts to describe the public's emotional changes to the vaccine at various phases. Through the distinction and modeling of positive and negative microblog posts, it reveals the differences in the topics concerning vaccine-support and vac-cine-hesitant groups. Finally, we use machine learning models to predict the evolu-tionary trends of public sentiment, providing practical significance and reference val-ue for the government's macro-control of COVID-19 vaccination.

References

Alswaidan, N.; and Menai, M. E. B. 2020. A survey of state-of-the-art approaches for emotion recognition in text. *Knowledge and Information Systems*, 62(8): 2937–2987.

Cambria, E.; Das, D.; Bandyopadhyay, S.; and Feraco, A. 2017. Affective computing and sentiment analysis. In *A practical guide to sentiment analysis*, 1–10. Springer.

China Bureau of Disease Control and Prevention. 2022. COVID-19 Vaccination. http://www.nhc.gov.cn/jkj/s7915/202209/879368f4fb544c28ae11a5387a519a5d.shtml. Accessed on 20 September 2022.

Cruickshank, I.; Ginossar, T.; Sulskis, J.; Zheleva, E.; Berger-Wolf, T.; et al. 2021. Content and dynamics of websites shared over vaccine-related tweets in COVID-19 conversations: computational analysis. *Journal of Medical Internet Research*, 23(12): e29127.

Del Rio, C.; Omer, S. B.; and Malani, P. N. 2022. Winter of Omicron—the evolving COVID-19 pandemic. *Jama*, 327(4): 319–320.

Du, S.-Y.; Dai, Y.-X.; Li, P.-W.; Zhao, N.; Li, S.; and Zheng, Y. 2021. Vaccinated or not? Survey on attitude toward 'approach-avoidance conflict' under uncertainty. *Hum.Vaccines Immunother*, 18: 1–6.

Fang, F.; Wang, T.; Tan, S.; Chen, S.; Zhou, T.; Zhang, W.; Guo, Q.; Liu, J.; Holme, P.; and Lu, X. 2021. Network Structure and Community Evolution Online: Behavioral and Emotional Changes in Response to COVID-19. *Frontiers in public health*, 9.

Fujino, A.; Isozaki, H.; and Suzuki, J. 2021. Multi-Label Text Categorization with Model Combination Based on f1-Score Maximization. *In Proceedings of the Third International Joint Conference on Natural Language Processing*, 7-12: I08–21162116.

Gbashi, S.; Adebo, O. A.; Doorsamy, W.; Njobeh, P. B.; et al. 2021. Systematic delineation of media polarity on COVID-19 vaccines in Africa: computational linguistic modeling study. *JMIR medical informatics*, 9(3): e22916.

Ginossar, T.; Cruickshank, I. J.; Zheleva, E.; Sulskis, J.; and Berger-Wolf, T. 2022. Cross-platform spread: vaccine-related content, sources, and conspiracy theories in YouTube videos shared in early Twitter COVID-19 conversations. *Human vaccines & immunotherapeutics*, 18(1): 1–13.

Guo, S.; Fang, F.; Zhou, T.; Zhang, W.; Guo, Q.; Zeng, R.; Chen, X.; Liu, J.; and Lu, X. 2021. Improving Google flu trends for COVID-19 estimates using Weibo posts. *Data Science and Management*, 3: 13–21.

Hu, T.; Wang, S.; Luo, W.; Zhang, M.; Huang, X.; Yan, Y.; Liu, R.; Ly, K.; Kacker, V.; She, B.; et al. 2021. Revealing public opinion towards COVID-19 vaccines with Twitter data in the United States: spatiotemporal perspective. *Journal of Medical Internet Research*, 23(9): e30854.

Hussein, D. M. E.-D. M. 2018. A survey on sentiment analysis challenges. *Journal of King Saud University-Engineering Sciences*, 30(4): 330–338.

Hutto, C.; and Gilbert, E. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, 216–225.

Information Office of the State Council. 2022a. People from Many Countries Spoke Positively of China's "Dynamic Zero" Epidemic Prevention Policy[EB/OL]. http://www.scio.gov.cn/37259/Document/1724016/1724016.htm.

Information Office of the State Council. 2022b. White Paper on China's Action against COVID-19. http://www.scio.gov.cn/zfbps/32832/Document/1681801/1681801.htm.

Lee, J. D. M. C. K.; and Toutanova, K. 2018. Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Liang, Z.; Weng, F.; Ma, Y.; Xu, Y.; Zhu, M.; and Yang, C. 2022. Measurement and Analysis of High Frequency Assert Volatility Based on Functional Data Analysis. *Mathematics*, 10(7): 1140.

Liu, B. 2012. Sentiment analysis: A fascinating problem. In *Sentiment Analysis and Opinion Mining*, 1–8. Springer.

Loria, S.; Keen, P.; Honnibal, M.; Yankovsky, R.; Karesh, D.; and Dempsey, E. 2019. TextBlob: simplified text processing. Secondary TextBlob: Simplified Text Processing 2014.

Lu, X.; and Brelsford, C. 2014. Network structure and community evolution on twitter: human behavior change in response to the 2011 Japanese earthquake and tsunami. *Scientific reports*, 4(1): 1–11.

Lyu, J. C.; Le Han, E.; and Luli, G. K. 2021. COVID-19 vaccine–related discussion on Twitter: topic modeling and sentiment analysis. *Journal of medical Internet research*, 23(6): e24435.

Medhat, W.; Hassan, A.; and Korashy, H. 2014. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4): 1093–1113.

Monselise, M.; Chang, C.-H.; Ferreira, G.; Yang, R.; Yang, C. C.; et al. 2021. Topics and sentiments of public concerns regarding COVID-19 vaccines: social media trend analysis. *Journal of Medical Internet Research*, 23(10): e30765.

Ravi, K.; and Ravi, V. 2015. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-based systems*, 89: 14–46.

Soleymani, M.; Garcia, D.; Jou, B.; Schuller, B.; Chang, S.-F.; and Pantic, M. 2017. A survey of multimodal sentiment analysis. *Image and Vision Computing*, 65: 3–14.

Tai, K. S.; Socher, R.; and Manning, C. D. 2015. Improved semantic representations from tree-structured long short-term memory networks. *arXiv preprint arXiv:1503.00075*.

Tsao, S.-F.; Chen, H.; Tisseverasinghe, T.; Yang, Y.; Li, L.; and Butt, Z. A. 2021. What social media told us in the time of COVID-19: a scoping review. *The Lancet Digital Health*, 3(3): e175–e194.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, C.; Han, B.; Zhao, T.; Liu, H.; Liu, B.; Chen, L.; Xie, M.; Liu, J.; Zheng, H.; Zhang, S.; et al. 2021. Vaccination willingness, vaccine hesitancy, and estimated coverage at the first round of COVID-19 vaccination in China: A national cross-sectional study. *Vaccine*, 39(21): 2833–2842.

Yue, L.; Chen, W.; Li, X.; Zuo, W.; and Yin, M. 2019. A survey of sentiment analysis in social media. *Knowledge and Information Systems*, 60(2): 617–663.

Zhang, L.; Wang, S.; and Liu, B. 2018. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4): e1253.

Zhuang, M.; Li, Y.; Tan, X.; Xing, L.; and Lu, X. 2021. Analysis of public opinion evolution of COVID-19 based on LDA-ARMA hybrid model. *Complex & Intelligent Systems*, 7(6): 3165–3178.