

Public Perception on “Hong Kong and Macao Vehicles Going North”: Text Mining on Sina Weibo

Fa Zhang*

Business School, Beijing Institute of Technology, Zhuhai
China
richter2000@163.com

Jian Chen

Business School, Beijing Institute of Technology, Zhuhai
China
61140710@qq.com

Abstract

To promote the integration and development of Guangdong, Hong Kong and Macao, the Chinese Government has introduced the policy “Hong Kong and Macao vehicles going north”. The public opinions are one of the bases for evaluating and optimizing policy. Sina Weibo is the main social media for Chinese to express their opinions. We collected Weibo posts related to the policy for sentiment analysis and topic modeling, and the results show that public sentiment is generally positive. There are differences in sentiment and topics between official and individuals. In terms of sentiment, official posts are more positive than individuals. In terms of topics, official posts mainly announce policy measures and report policy status. Individual posts focus on personal experience and social effects. To further understand the individual’s feelings, individual posts were analyzed based on the social exchange theory. The results show that in terms of benefits, individuals are mainly concerned with the convenience of life, and in terms of costs, they are mainly concerned with social losses. Convenience of life is mainly reflected in the using of cars, consumption, traveling, and customs clearance. Social losses are mainly reflected in the risks caused by Hong Kong and Macao vehicles, the unfairness caused by the inequality of the three regions, and the discomfort of crowding public resources. These findings will help in the evaluation and improvement of the policy.

CCS Concepts

• Applied Computing; • Law, society and behavior sciences; • Sociology;

Keywords

Hong Kong vehicles, sentiment, topic, benefit, cost

ACM Reference Format:

Fa Zhang and Jian Chen. 2024. Public Perception on “Hong Kong and Macao Vehicles Going North”: Text Mining on Sina Weibo. In *2024 the 7th International Conference on Information Science and Systems (ICISS) (ICISS 2024, August 14–16, 2024, Edinburgh, United Kingdom)*

*Place the footnote text for the author (if applicable) here.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICISS 2024, August 14–16, 2024, Edinburgh, United Kingdom

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-1756-7/24/08

<https://doi.org/10.1145/3700706.3700709>

2024), August 14–16, 2024, Edinburgh, United Kingdom. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3700706.3700709>

1 Introduction

The Guangdong-Hong Kong-Macao Greater Bay Area (GHMA), which covers nine cities in Guangdong Province as well as the Hong Kong and Macao Special Administrative Regions, is an important region for China’s reform and opening. The GHMA is characterized by “one country, two systems, three tariff zones and three jurisdictions”, and connectivity is the foundation for the integrated development of GHMA. On January 1, 2023, the Chinese government implemented the “Macao vehicles going north” and on July 1, 2023, the “Hong Kong vehicles going north”. “Hong Kong and Macao vehicles going north” means that Hong Kong (Macao) vehicle owners can drive into Guangdong after obtaining approval [1]. By the end of 2023, more than 57,000 Hong Kong and Macao vehicles had entered and left Guangdong [2].

The policy of “Hong Kong and Macao vehicles going north” has had a positive impact, and at the same time has brought about some challenges. There is a need for evaluation of the policy to understand its impact. Public opinion is an important basis for policy evaluation. By analyzing public opinions, the effectiveness of the policy can be measured, and its deficiencies identified, providing a basis for adjustment and optimization. In recent years, social media has become an important channel for the public to express their attitudes, views and opinions. Analyzing social media is an effective way to understand public opinion [3]. Mining of social media has many advantages and can get knowledge that cannot be obtained by questionnaires [4].

Insight into public attitudes through social media has a wide range of applications. During the COVID-19 pandemic, some researchers mined social media to understand the attitudes of the public [5–7]. Mining social media can also reveals new phenomena [8]. There is value in mining social media at all stages of public policy. At the policy implementation stage, it can help policy makers monitor people’s opinions [9] and evaluate policies [10–14].

Since the implementation of the “Hong Kong and Macao vehicles going north”, the public’s feelings have been reflected on social media. What are the overall feelings of the public, which aspects have had a positive impact, and which aspects have had a negative impact? Sina Weibo is the main social media for Chinese people to express their opinions. This paper collects Sina Weibo posts and mines the netizens’ attitudes, sentiments, and opinions to find the reasons behind the attitudes, to provide a basis for evaluating the policy, and to discover the public’s concerns.

2 Methods

2.1 Data acquisition and preprocessing

We used a crawler Octopus to collect Weibo posts. The keywords are “Hong Kong vehicles going north” OR “Macau vehicles going north” OR “Hong Kong and Macau vehicles going north” OR “Hong Kong and Macao single-license plate vehicles northbound” OR “Hong Kong people driving northbound” OR “Macao people driving northbound”. After manually deleting irrelevant posts, we got 1410 posts.

We wrote Python program to clean and transform the data. Including: calculating the date of posting, extracting user ID, and determining the type of user (official, individual). The date of the post is calculated according to the current time and post time. The user ID is extracted using a regular expression. To analyze the similarities and differences between official and individuals, two types of users are distinguished. Official users include government departments and news media. According to the user ID to query the user’s basic information, including username, profile, authentication, industry, etc. Use regular expression to determine whether the user is an official or an individual.

In addition, microblogs may include some content that is not related to the policy, such as super topics, hyperlinks, usernames, etc., which are filtered using regular expressions. After preprocessing, the fields of the dataset include publisher, user id, user type, content, cleaned content, posting date, posting time, posting source, number of comments, retweets, likes, etc.

2.2 Sentiment analysis and topic modeling

Sentiment analysis is a natural language processing technique used to identify and extract the emotional tone or attitude expressed in a text, such as positive, negative, or neutral. Lexicon methods look for emotion words in a text and determine the overall emotional tendency based on the emotional polarity and intensity of the words. Commonly used Chinese sentiment lexicons include DLUT, NTUSD, THUOCL, etc. Machine learning methods use models such as support vector machine and naive bayes to achieve emotion judgment. Deep learning uses deep neural networks such as CNN, RNN, BERT, GPT, etc. to achieve sentiment judgment [15].

We used Weiciyun to implement Chinese text sentiment analysis. Weiciyun (fenci.weiciyun.com) is a Chinese text online analysis tool. Weiciyun uses a combination of lexicon method and machine learning to achieve sentiment analysis, with a variety of built-in Chinese sentiment lexicons, such as DLUT, THUOCL, etc., or user can customize the lexicon, set synonyms, etc. It can provide sentiment analysis results and visualization.

We also used Weiciyun for LDA [16]. The number of topics is a hyperparameter and we determine the number of topics based on perplexity and coherence. First use elbow method to get the upper limit of the number of topics, and then determine the better number of topics based on the coherence. Combined with pyLDAvis visualization, the number of topics is determined.

2.3 Content analysis

To further understand the detailed impact of the policy, content analysis was conducted. Content analysis is a research method that

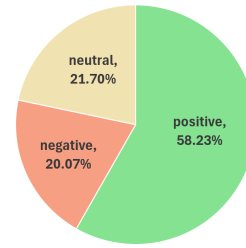


Figure 1: Sentiment distribution of all the posts

objectively, systematically, and quantitatively analyzes qualitative material [17]. All individual posts were used as a sample. Two coders were recruited. 10% of the sample was randomly selected for training and passed the coder reliability test.

Social Exchange Theory (SET) provides a framework for understanding individual and group behavior [18]. A multilevel category system was designed based on the theory, as shown in Table 1. The first-level categories include policy benefits and costs. Policy benefits are the benefits, satisfaction, or rewards perceived by individual. Policy costs are the costs, sacrifices, or losses perceived by individuals. Benefits and costs are divided into secondary categories, which are further divided into tertiary categories.

We used DivoMiner (<https://www.divominer.cn/>) to code each post. The coding results were analyzed and interpreted. Content analysis was used to answer two research questions:

RQ1: In what ways do the public feel positive affected by the implementation of the policy? Which aspects are the main ones?

RQ2: In what ways do the public feel negative affected by the implementation of the policy? Which aspects are the main ones?

3 Results

3.1 General feeling on social media

There was a total of 1,410 Weibo posts, including 787 official posts and 623 individual posts. The social media are mainly government information releases and media news reports, with relatively little individual content. All posts were analyzed for sentiment, and the distribution is shown in Figure 1. The proportion of positive sentiment is 58.23%, which indicates that the overall sentiment is positive, and the proportion of negative is 20.07%, which also has a certain percentage.

3.2 Analysis of official posts

There are 787 official posts, which are released by government departments and news media to play the role of public opinion guidance. The sentiment distribution of official posts is shown in Figure 2, in which positive posts accounts for 63.28% and negative accounts for 17.41%. The positive sentiment is higher.

Next, topic mining is performed on official posts. After compared the effect of different number of topics, we found the optimal number is 5. LDA analysis is carried out, and each topic, percentage, and high-frequency words are shown in Table 2.

Topic 1 is about the policy measures related to the application and filing of applications by Hong Kong vehicles going north. Topic

Table 1: The Category System

Category I	Category II	Category III
Benefits	Life	Car using
		Visit friends and relations
		Consume
	Economy	Travel
		Custom clearance
		Others
		Personal income
		Enterprise development
		Industry development
Society	Regional economy development	
	Interconnection of the Great Bay	
	Integration of the Great Bay	
Costs	Life	Government image promotion
		Traffic trouble
		Consume inconvenience
	Economy	Travel inconvenience
		Others
		Personal Income decreased
		Enterprise problems
		Industry problems
		Regional economy issues
	Society	Risks
		Unfair
		Resource competition
		Others

Table 2: The five topics for official posts

No.	Topic	Number (Percentage)of posts	Key words
1	Hong Kong vehicles applying to go north	171 (21.73%)	Mainland, Guangdong, Vehicle, Owner, Private, Entry, Record, Department, Processing, Approach, Transportation, Eligible, Procedures, Electronic, Logistics, Computer, Citizens, Validity, Period, Insurance, Public Security, Requirements
2	Hong Kong-Zhuhai-Macao Bridge clearance	201 (25.54%)	Bridge, Hong Kong and Macao, Policy, Vehicle, Clearance, Highway, Immigration, Traffic, Gongbei, Passenger, Flow, Border Checkpoint, Quantity, Single Day, Border Inspection, Daily Average, Data, Record, Historical, Peak, Departure, One-stop, News
3	Activities of Hong Kong people going north	128 (16.26%)	Hong Kong, Zhuhai, Shenzhen, Tourism, Consumption, China, Self-driving, Tour, Activity, Hour, Friends, Experience, Shenzhen-Hong Kong Life Circle, License, People, Economy, Returning to the Motherland, Trend, Everyone, Visitors
4	Activities of Macao people going north	136 (17.28%)	Macao, Hengqin, Customs, Cooperation, Drivers, Government, Quota, Inspection, Officers, Import, Issues, Information, Program, Equipment, Operation, Epidemic, Car, Body, RMB
5	Optimization of cross-border service	144 (18.30%)	Residents, the Greater Bay Area, Development, Traffic, Cross-border, Service, Personnel, Parking, Police, Situation, Optimization, License, Interconnection, Life, Zhuhai, Urban, Safety, Innovation, Management, Bureau

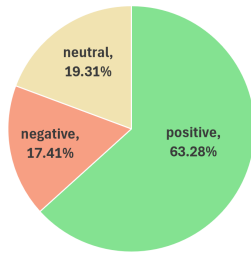


Figure 2: Sentiment distribution of official posts

2 is the report on the immigration situation of Hong Kong-Zhuhai-Macao Bridge. Topic 3 is about Hong Kong people’s consumption and tourism in the mainland. Topic 4 is about Macao people’s customs clearance and border control. Topic 5 is the interface and optimization of services related to Hong Kong and Macao vehicles going north. Among them, topic 2 and topic 1 account for the largest proportion, which shows that the official microblog mainly releases relevant policies and regulations and informs about the situation of immigration and customs clearance of the Hong Kong-Zhuhai-Macao Bridge.

Sentiment analysis of each topic is conducted separately, and the sentiment distribution of each topic is shown in Figure 3. It is worth noting that topic 2 has a higher positive sentiment, while topic 4 has a lower positive sentiment. That is, there is a greater difference in sentiment in terms of Hong Kong people going north and Macao people going north. The analysis of the posts on these two topics shows that topic 2 mainly reports on Hong Kong people going north for consumption, entertainment and exchanges, with a more positive tone. Topic 4 Macao people going north was more often reported on smuggling into the mainland, with less positive sentiment.

3.3 Analysis of individual posts

The 623 individual posts were analyzed for sentiment and the results are shown in Figure 4. The proportion of positive sentiment is

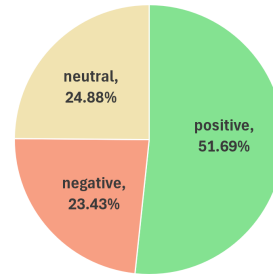


Figure 4: Sentiment distribution of individual posts

51.69% and negative 23.43%. It shows that most individuals present positive attitudes towards the policy, but there are also some negative attitudes that deserve attention.

Compared with the official microblogs, the proportion of positive sentiment of individual microblogs (51.69%), is significantly lower than the positive sentiment of official microblogs (63.28%). The proportion of negative sentiment of individual microblogs (23.43%) is significantly higher than the negative sentiment of official microblogs (17.41%). The chi-square test shows that there is a significant difference in the distribution between official and individual microblogs. Official microblogs show more positive and less negative emotions, which are somewhat different from personal feelings.

Set the number of topics as 5, we performed LDA analysis of individual posts. The theme, percentage, high-frequency words of each topic are shown in Table 3.

Compared with the five topics of the official microblogs, the three topics of bridge clearance, Hong Kong people going north and Macau people going north are common, and the Greater Bay Area interoperability is similar to the official topic 5. Topic 1 on automobile repair and maintenance is a blog post published by some individuals for commercial purposes, suggesting that automobile service-related industries are now mainly promoted.

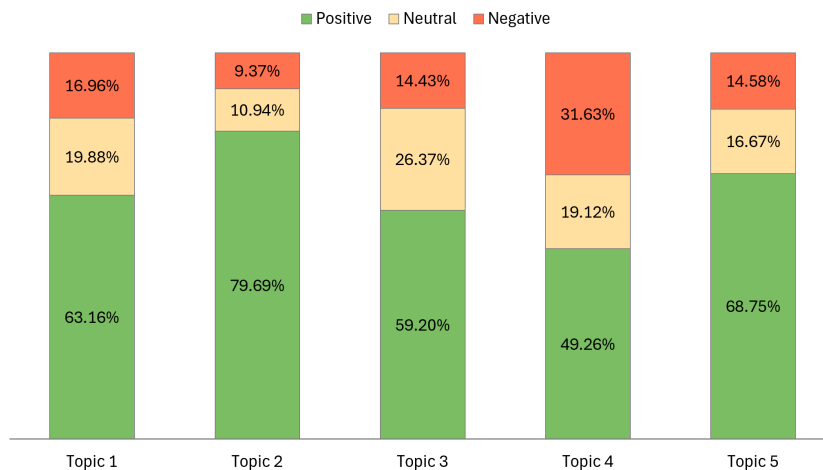
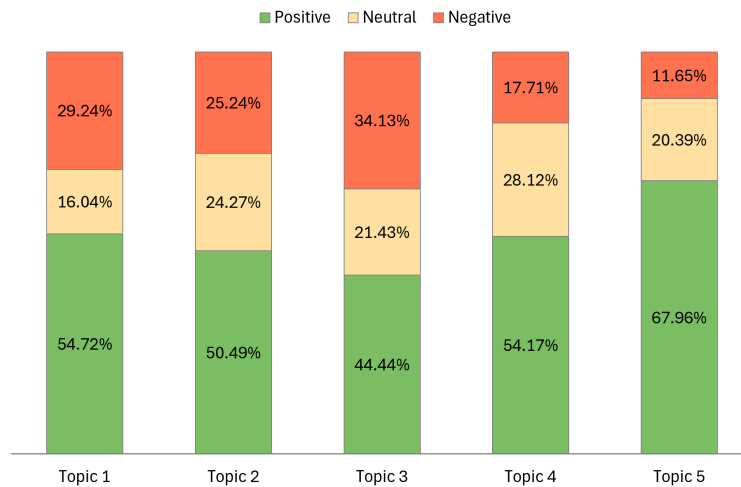


Figure 3: Sentiment distribution of official posts in each topic

Table 3: The five topics for individual posts

No.	Topic	Number (Percentage) of posts	Key words
1	Automotive Maintenance	106 (17.01%)	Automobile, Customs, Professional, Gongbei, Zhuhai, Market, Service Center, Driver, Car Show, Beauty, Traffic Police, Classic, Import, Xiangzhou District, Check, New Car, Record, Sofa, Limousine Wheel, Electric Vehicle, Wash, Calling, Card, Engine
2	Hong Kong-Zhuhai-Macao Bridge clearance	103 (16.53%)	Bridge, Vehicles, Immigration, Clearance, Highway, Residents, Single-Day Passenger Flow, Private Vehicles, Motorized Vehicles, Border Crossing Station, Passengers, Guangdong, Traffic Volume Record, Time, Data, Situation, Impact, Citizens, Import and Export Processing System, Passing, Compute
3	Activities of Hong Kong people going north	126 (20.22%)	Hong Kong, Mainland, Shenzhen, Hong Kong people, Consumer, Issues, Zhongshan, Experience, Friends, Self-drive, Sam, Cuisine, Travel, Craze, Company, Malls, Netizens, Lots, Easter Buying Shopping, Culture, Cantonese Habits, Lawyers, Prices
4	Activities of Macao people going north	96 (15.41%)	Hong Kong and Macao, Macao, Hengqin, Car Owners, Both Plates, Foshan, Entry, Tourism, Global, Life, Car Parking, World Project, Nanhai, Passing, Property License, Pedestrian Flow, Bay Area, Residential Design, Going Home, Free, Network, Neighborhoods
5	Interchange in the Greater Bay Area	103 (16.53%)	Policies, Greater Bay Area, Guangdong, Guangdong, Hong Kong and Macao, Services, Cross-Border Traffic, Government, Cooperation, Development, Landing, Integration, Plan, China, Two-Way Construction, National Story, Measures, Beijing, Insurance, Optimization, Driving, Suggestion, Regional Life

**Figure 5: Sentiment distribution of individual posts in each topic**

Sentiment analysis was conducted for the five topics, and the distribution of sentiment for each topic is shown in Figure 5.

The three common topics were compared between individual and official posts. Topic 2 is about the bridge clearance, where the personal positive sentiment is slightly lower than official sentiment. Topic 3 on Hong Kong people going north, official media positive sentiment is higher, while personal positive sentiment is significantly lower, suggesting that there is a difference between

official and personal in terms of Hong Kong people going north. Topic 4 on Macao people going north, official positive sentiment is lower while individual positive sentiment is slightly higher, suggesting that official reports on smuggling and other issues do not have a significant effect on individuals. Similarly for topic 5 on Guangdong-Hong Kong-Macao connectivity, official and individual sentiments are basically the same.

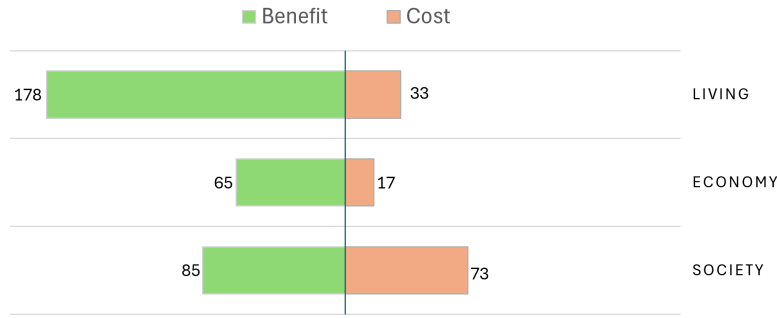


Figure 6: Public perceptions of benefits and costs in living, economy and society

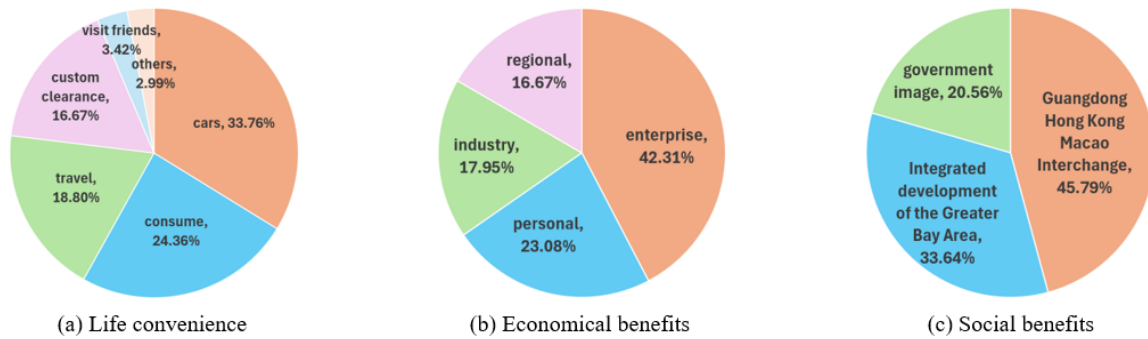


Figure 7: Perceptions of benefits in life, economy and society

3.4 Results of content analysis

All individual microblogs were coded according to the category system. The coding results is shown in Figure 6. There were 328 posts involving benefits and 123 involving costs, indicating that the policy effects were mainly positive. In the secondary category of the benefits, the most concern are life convenience, followed by social benefit, and then economic benefit, indicating that the benefits perceived by the public are dominated by the convenience of life. On the policy costs, society losses are mainly perceived, with fewer life and economic. The benefit and cost show an interesting asymmetry.

Based on the results of the third level of coding, public perceptions can be further refined. What specific aspects of life, economy, and society does the public feel a benefit from? In terms of convenience of life, the four main areas are cars driving and using, consumption, tourism, and customs clearance, with very little benefit from visiting friends, as shown in Figure 7(a). In terms of economical benefits, mainly in the enterprise related to vehicle sales and services, these companies felt an increase in business opportunities as well as a boost to the automotive industry. A small number of posts mentioned the promotion of regional economies such as Zhuhai and Shenzhen, as shown in Figure 7(b). In terms of social benefits, it mainly mentions the interchange of Guangdong, Hong Kong and Macao, as well as the integrated development of the Greater Bay Area. A small number of microblogs mention the enhancement of the government’s image, indicating that the

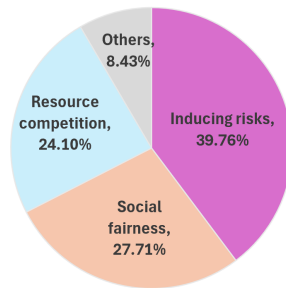


Figure 8: Perceptions of social losses

public’s political awareness is relatively weak, as shown in Figure 7(c).

The public’s perception of the policy costs is mainly in social losses, while the loss of life convenience and economic are not sensitive. For social losses, the public perception is mainly in three areas, as shown in Figure 8. First, it induced some risks. Hong Kong and Macao vehicles entering the mainland incur risks due to traffic violations, traffic accidents and cross-boundary smuggling. Second, vehicles from Guangdong, Hong Kong and Macao are not treated equally, triggering a sense of social unfairness. Third, Hong Kong and Macao vehicles entering the mainland take up public resources, generating a sense of discomfort.

4 Conclusions

This paper investigated the public’s attitudes, sentiments and opinions towards the policy of “Hong Kong and Macao Vehicles Going North”. We collected Sina Weibo posts for sentiment analysis and topic modeling. The results show that public sentiment is generally positive, and there are significant differences in sentiment and topics between official and individual posts. Based on social exchange theory, individual posts were analyzed, and the results show that individuals are mainly concerned about the convenience of life in terms of benefits and social losses in terms of costs.

This study has some limitations. First, Sina Weibo may not fully reflect the diverse perspectives of different populations and regions. Second, sentiment analysis and topic modeling have some limitations in interpreting human emotions and policy implications. In the future, there is a need to expand data sources and combine interviews and questionnaires to assess the actual impact of policies.

Acknowledgments

This research was funded by Guangdong Province Key Research Base of Humanities and Social Sciences (Grant No. 2022WZJD012) and the Study on the Development of the Guangdong-Hong Kong-Macao Greater Bay Area (XK-2023-007).

References

- [1] Government of Guangdong Province, “Circular of the Measures for the Administration of Entry and Exit of Hong Kong Motor Vehicles to and from the Mainland via the Hong Kong-Zhuhai-Macao Bridge-Zhuhai Highway Crossing,” Jun. 15, 2023. [Online]. Available: http://www.gd.gov.cn/zwgk/gongbao/2023/12/content/post_4173679.html
- [2] China News Service, “More than 57,000 Hong Kong and Macao private cars will enter and leave Guangdong by the end of 2023 under the ‘Northbound’ policy.” Accessed: Jun. 13, 2024. [Online]. Available: https://www.sohu.com/a/778482121_162758
- [3] K. Ravi and V. Ravi, “A survey on opinion mining and sentiment analysis: Tasks, approaches and applications,” *Knowl.-Based Syst.*, vol. 89, pp. 14–46, 2015, doi: 10.1016/j.knsys.2015.06.015.
- [4] X. Dong and Y. Lian, “A review of social media-based public opinion analyses: Challenges and recommendations,” *Technol. Soc.*, vol. 67, p. 101724, 2021, doi: 10.1016/j.techsoc.2021.101724.
- [5] K. Wang *et al.*, “A multi-country test of brief reappraisal interventions on emotions during the COVID-19 pandemic,” *Nat Hum Behav.*, vol. 5, no. 8, pp. 1089–1110, 2021, doi: 10.1038/s41562-021-01173-x.
- [6] J. C. Lyu, E. L. Han, and G. K. Luli, “COVID-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis,” *J. Méd. Internet Res.*, vol. 23, no. 6, p. e24435, 2021, doi: 10.2196/24435.
- [7] F. Zhang, Q. Tang, J. Chen, and N. Han, “China public emotion analysis under normalization of COVID-19 epidemic: Using Sina Weibo,” *Front Psychol.*, vol. 13, p. 1066628, 2023, doi: 10.3389/fpsyg.2022.1066628.
- [8] A. Field, C. Y. Park, A. Theophilo, J. Watson-Daniels, and Y. Tsvetkov, “An analysis of emotions and the prominence of positivity in #BlackLivesMatter tweets,” *Proc National Acad Sci.*, vol. 119, no. 35, p. e2205767119, 2022, doi: 10.1073/pnas.2205767119.
- [9] P. Singh, Y. Dwivedi, K. Kahlon, R. Sawhney, A. Alalwan, and N. Rana, “Smart Monitoring and Controlling of Government Policies Using Social Media and Cloud Computing,” *INFORMATION SYSTEMS FRONTIERS*, vol. 22, no. 2, pp. 315–337, 2020, doi: 10.1007/s10796-019-09916-y.
- [10] N. Adams-Cohen, “Policy Change and Public Opinion: Measuring Shifting Political Sentiment with Social Media Data,” *AMERICAN POLITICS RESEARCH*, vol. 48, no. 5, pp. 612–621, 2020, doi: 10.1177/1532673x20920263.
- [11] Y. Kang, Y. Wang, D. Zhang, and L. Zhou, “The public’s opinions on a new school meals policy for childhood obesity prevention in the US: A social media analytics approach,” *INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS*, vol. 103, pp. 83–88, 2017, doi: 10.1016/j.ijmedinf.2017.04.013.
- [12] W. Deng, J. Hsu, K. Löfgren, and W. Cho, “Who is leading China’s family planning policy discourse in Weibo? A social media text mining analysis,” *POLICY AND INTERNET*, vol. 13, no. 4, pp. 485–501, 2021, doi: 10.1002/poi3.264.
- [13] B. Shi, “Transmission mechanism of public concern in waste-sorting policy: Evidence from text mining,” *ENERGY & ENVIRONMENT*, vol. 35, no. 3, 2024, doi: 10.1177/0958305x231217634.
- [14] H. Wang and A. Cooper, “Public Opinion on Chinese Foreign Aid Policy: Calculated Opposition or General Discontent?,” *JOURNAL OF CONTEMPORARY CHINA*, vol. 32, no. 141, pp. 455–472, 2023, doi: 10.1080/10670564.2022.2090101.
- [15] J. Deng and F. Ren, “A survey of textual emotion recognition and its challenges,” *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 49–67, 2023, doi: 10.1109/taffc.2021.3053275.
- [16] D. Blei, A. Ng, M. Jordan, and J. Lafferty, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, Feb. 2003.
- [17] K. Krippendorff, *Content analysis: An Introduction to its methodology*, 4th Edition. Sage publications, 2019, doi: 10.4135/9781071878781.
- [18] R. Cropanzano and M. S. Mitchell, “Social Exchange Theory: An Interdisciplinary Review,” *J. Manag.*, vol. 31, no. 6, pp. 874–900, 2005, doi: 10.1177/0149206305279602.