Aftershock: Sentiment and Content Analysis of Weibo Posts about 2023 Health Insurance Reform in the Post Zero-Covid Context

Duoji Jiang[1,a,*]

[1]*Division of the Social Sciences, The University of Chicago, 5801 S. Ellis Ave, Chicago, USA*
a. duojijiang@uchicago.edu
*corresponding author

Abstract: This study investigates the public sentiment and thematic content of Weibo posts regarding the 2023 health insurance reform in China, a topic of heightened discussion in the aftermath of the country's zero-Covid policy. Utilizing a data-driven approach, the research analyzes netizens' reactions using phrase-level sentiment analysis and topic analysis techniques. The paper identifies the financial strains imposed by zero-Covid as a primary driver behind the reform, which led to reduced benefits and increased dissatisfaction among the population. Methodologically, the study employs Support Vector Regressor for sentiment analysis and Word2Vec for content analysis, with data sourced from an open-source Weibo crawler. The findings reveal a discrepancy in sentiment between official and unofficial accounts, pinpoint the sources of negative sentiment, and highlight the interconnection with the zero-Covid policy. The research provides insights into the public's perception of government policies during a transitional period in China's health insurance landscape.

Keywords: Zero-Covid, Social Media Analysis, Weibo, Public Opinion, Machine Learning

1. Introduction

Since February 2023, retired elderly people have taken to the streets to protest the government's new healthcare reform policy in several cities in China, mainly against the reduced benefit to their personal medical insurance account, while the reimbursement threshold rise. Despite the highly efficient censorship mechanism, netizens still widely expressed their dissatisfactions towards the healthcare reform on social media. This paper will apply word-associated sentiment analysis and topic analysis based on Weibo posts to analyze the sentiments of Weibo users. and further demonstrate how these negative sentiments and the medical insurance reform policy are inextricably linked with the Chinese authorities' zero-covid policy in 2020-2022.

1.1. Zero-Covid Costs and Medical Insurance Reform

Firstly, to understand the reasons for this large-scale dissatisfaction and the motivation behind the healthcare reform, the medical insurance reform needs to be briefly introduced. China's medical insurance system is mainly divided into two parts, urban employee basic medical insurance (UEBMI) and Urban and Rural Resident Basic Medical Insurance (URRBMI). About three-quarters of the people only enjoy the basic resident medical insurance, URRBMI. The coverage of this insurance is

significantly less than that of UEBMI. At the same time, this part of the insurance does not provide personal accounts that URRBMI provides [1][2]. Since the protest and online dissatisfaction are mostly towards the changes made in UEBMI, this article focuses on UEBMI, which covers about a quarter of the Chinese population, including civil servants, employees of the state-owned sector as well as employees of private enterprises with higher incomes, and retirees from these sectors. There are two unique features of the Chinese medical insurance system: Firstly, the medical insurance fund paid by employees and employers is managed by local governments, and law stipulates that when the medical insurance fund run into deficits, the governments are responsible for covering the deficit. Second, these funds are divided into two components, the shared fund and the personal accounts [3][4].

Secondly, it is also necessary to briefly introduce China's zero-covid policy. To put it simply, the zero-covid policy, or the more official term "dynamic zero-covid policy", refers to the realization of zero covid-19 infections through regular large-scale PCR tests and harsh lockdowns [5]. This policy is considered to have achieved great success in the early stages of the epidemic. However, when the main strain of covid-19 evolved to less harmful variant- omicron, it has led to huge controversy over whether the policy should be continued due to the economic cost of PCR tests and lockdowns. Eventually, the controversy grew into a series of protests that led to the discontinuation of zero-covid policy [6]. Evidence indicates that the main reason for the healthcare reform is that the governments ran into fiscal problems due to the financial cost of zero-covid, so that the payment to personal accounts must be reduced to support the shared funds. According to incomplete public statistics, in 2022, the Chinese government spent about 2%-6% of its fiscal expenditure on epidemic prevention and control. At the national level, while spending increased by 6.1%, fiscal revenue of the government in 2022 increased by merely 0.6% [7][8]. In addition, the medical insurance fund was responsible for a considerable part of the expenditure on vaccines and PCR tests. According to official reports, the medical insurance fund spent a total of 154.3 billion yuan on these two items from 2021 to 2022 [9]. The specific figures and ratios of reductions in payments to personal accounts differ by location. But taking Wuhan as an example, for a typical employee in Wuhan with a monthly income of 10,000 yuan, the total amount paid for medical insurance is about 1,000 yuan, and the proportion allocated to the personal account was reduced from 58% to 20%. In addition, the subsidy for individual accounts of retirees has been reduced from 260 yuan to 80 yuan [4]. Although the idea of allocating more payments to shared funds instead of personal accounts started essentially before the pandemic, it does not negate the relationship between the medical insurance reform and the zero-covid policy [10]. First, as mentioned above, a series of evidence points out the logical relationship between the zero covid expenditure and the medical insurance reform from the fiscal figures. Second, Implementing this controversial policy only three months after zero-covid led to street protests could be expected to cause dissatisfaction, thus revealing the urgency of the government fiscal problems. Thirdly, this paper mainly aims at analyzing negative sentiments expressed by Weibo users. Therefore, even if the causal evidence chain between the zero-covid and the healthcare reform might not be completely sufficient, it does not undermine the belief in such relationship among netizens.

1.2. Phrase-Level Sentiment Analysis

Phrase-level sentiment analysis is a subtype of sentiment analysis. Sentiment analysis refers to the use of machine learning or deep learning algorithms for prediction of sentiment of certain text. And phrase-level refers to refining the calculation of emotions to the level of phrases. Relevant literature points out that the leading implementations of phrase-level sentiment analysis are based on complex neural network algorithms, such as LSTM, BERT, etc. These deep-learning-based approaches could consider the context and linguistic structure of words and achieve higher accuracy [11]. In addition, the emotion analysis based on psychological frame of basic emotions has made achievements in

distinguishing emotional subcategories, specific categories differ by theory but generally includes anger, happiness, sadness, fear, surprise [12]. Limited by time and computing resources and a fairly limited amount of text for Deep Learning, this paper adopts a simplified phrase-level sentiment approach based on Support Vector Regressor. The idea is when measuring the sentiment of a word, the average sentiment score of the text containing such word is simply treated as the sentiment value of this word. Therefore, to be precise, the word-associated sentiment score in this research refers to the sentiment of this word in the context of the test data. This simplified method inevitably limits the accuracy of the model, but it can give the approximate emotional value of the words in the context under limitations. In addition, this article only considers the sentiment score in one dimension, the larger the value, the more positive, and vice versa.

1.3. Hypothesis

There are three hypotheses in this study: First, compared with non-official accounts, posts of official accounts would have more positive sentiments related to the healthcare reform. Second, the source of negative sentiments of unofficial users can be identified through analyzing word-associated sentiment values and word choices. Third, given the potential connection between the zero-covid and the healthcare reform, sources of negative sentiments are connected to the zero-covid as well.

2. Methodology

An open-source Weibo data crawling tool provided by GitHub User dataabc [13] that allowed avoiding limitations of the official Weibo API, is used to collect the research data, the Weibo posts and comments searched with keyword"医保改革"(Healthcare reform). The model Training data was open-source data shared by GitHub user Wansho [14]. Jieba, and open-source Chinese language segmentation tool is used to segment and tokenize the individual words. After which Word2Vec, and word-embedding approach based on Neural Network is applied to vectorize and represent the words and posts as combination of words. Support Vector Machine Regressor with Linear Kernal is used to predict the sentiment value of the words and posts in research data. Finally, Term Frequency and Inverse Document Frequency (TF-IDF) is used to calculate the significance score.

2.1. Jieba

Jieba is a Chinese text segmentation tool. Due to the nature of Chinese language that there is no space between words, identifying the word compositions is required. Jieba is one open-source tool that could parse Chinese text and segment the text into individual meaningful words. It realizes efficient word graph scanning based on the prefix dictionary, and generate a directed acyclic graph (DAG) composed of all possible word formations of Chinese characters in a sentence. Dynamic programming is used to find the maximum probability path and find the maximum segmentation combination based on word frequency. For words not included in the dictionary, the HMM model based on the word-forming ability of Chinese characters and the Viterbi algorithm is applied. Nevertheless, note that the algorithm might not be able to identify certain words, such as internet phrases, names [15].

2.2. Word2Vec

The word2vec algorithm is a word embedding model that applies neural network and back propagating to learn word associations from a large corpus of text. Which would vectorize the words by projecting each word onto high dimensions based on the embedded meaning relationships. Essentially words with similar embedded meanings would have a closer distance. In this study, the

number of dimensions is manually selected to 300, and the vector of each post is represented by the sum of the vectors of the words [16].

2.3. LinearSVR

Support vector machine (SVM) is a machine learning algorithm proposed by Vapnik on the basis of years of research on statistical learning theory. As a tool for solving several problems in data mining with the help of optimization methods, it can overcome the traditional difficulties such as "curse of dimensionality" and overfitting to a certain extent and has been widely used in the fields of text classification, forecasting and time series forecasting. Support Vector Regressor (SVR) is a subtype of SVM for regression problems. It makes predictions through establishing relationship between the vectors to be predicted and the support vectors in the training data with "kernel methods". Kernels with best performance are RBF kernel and sigmoid kernel [17]. However, limited by computing power as well as the size of the training text and the nature of high-dimensionality of text vectors, this study adopts linear kernel. It should be noted that since SVM treats text as the sum of words without order, it is less accurate in identifying sarcasm and multiple negations.

2.4. Word Significance score and word-associated sentiment score

Term Frequency and Inverse Document Frequency (TF-IDF) is a statistic approach that calculates the significance of a word in a corpus. Term Frequency (TF) is the number of times that one term appears in the document. Inverse Document Frequency (IDF) is the total number of documents in the corpus divided by the number of documents that contains the term, normalized with log function. TF-IDF is the product of TF and IDF. Significance score is calculated by adding up all TF-IDF score of this word in the corpus after filtering out words with TF-IDF score lower than 0.1. Word sentiment score is the numerical mean of the sentiment score of all posts including this word.

3. Data and Preprocessing

Training data is open-source pre-tagged Weibo posts, including 407058 of positive sentiments and 263995 of negative sentiment collected in 2019. For regression model training, I encoded all positive posts as having sentiment score of 1 and all negative posts as having sentiment score of 0. Research Data were Weibo posts scrapped by open-source tools, including 2471 Weibo posts related to healthcare reform from 2-03 to 2-21. Firstly, the parts of posts that does not involve own opinions of the accounts are deleted, which includes posts topics, marked with (##) and news titles marked with" 【】 " (Chinese version of punctuation " [ ] "). For example:

1. 【*Yunnan's medical insurance reform has benefited 3.76 million people*】 *On February 14, reporters learned from the Yunnan Provincial Medical Insurance Bureau…*
2. *#view the reform of medical insurance personal accounts rationally# That is why my medical insurance balance is not increasing…*

Then some posts without information value including those only containing reposting of videos or links are dropped. In addition, this study assumes that the text containing special characters '【|O】'are from official accounts, and the posts that do not contain these are from unofficial account. Research data are not manually labeled, so the accuracy and validity of these practices can only be intuitive but not guaranteed. After I used Jieba to segment the text, for the purpose of tuning the model, I used another set of labeled open-source Weibo texts as validation data, contains 1790 Weibo

posts about Mengniu Milk, of which 1010 are negative and 780 are positive, the model achieved accuracy of 0.88 on the validation data.[18]

4. Model Results

4.1. Word2vec

Table 1: Word2vec synonyms

keyword	Synonymous words, translated	Synonymous original in Chinese
Doctor (医生)	[Nurse, Patient, Hospital, examination]	[护士, 病人, 医院, 检查, 家属]
Sad (伤心)	[disappointed, heartbroken, grieved]	[难过, 失望, 心碎, 悲伤, 委屈]
Communist Party (共产党)	[rascal, traitor, Confucian, accomplice, outlaw]	[泼皮, 卖国, 儒家, 帮凶, 枉法]
Government (政府)	[law, developer, regulation, policy, people]	[法律, 开发商, 监管, 政策, 老百姓]
Bureaucrat (官员)	[corrupted, peasants, people, profiteers, developer]	[贪官, 农民, 民众, 奸商, 开发商]

Word2Vec vectorizes words by understanding their meanings in context, I trained the vectorizer with the training-data, as shown in table 1. I translated the words to English, but since no translation is perfect, I kept the original words in Chinese for more interested readers, this will also apply to later word tables. It returned outcomes turns to be normal for regular words such as doctor, and sad, but negative when I search for government, communist party, and bureaucrats. it is interesting to note how the "government" and "bureaucrats" are considered to have similar meanings to developers, which in Chinese context mostly relate to real estate. The results could be empirically examined through considering the local governments' financial dependency on selling land [19].
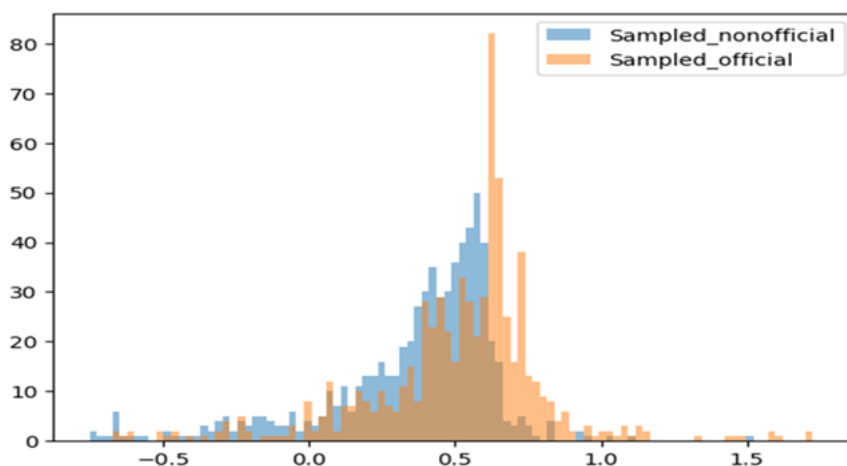
4.2. Distribution

Figure 2: Official vs Non-Official sentiment distribution

The distribution of sentiment scores of official accounts and unofficial accounts predicted by the LinearSVR model supported my hypothesis 1, that the official accounts do tend to have a more positive sentiments in the context of healthcare reform. In the graph, orange and blue referred to sentiment value distribution of official and unofficial accounts randomly sampled to same number of same amounts.

4.3. Phrase-level sentiment

Two lists of words constructed for analysis, the first being *Table2: non-official only words list*: where words only used by non-official accounts are selected and rank by the significance score. The second is *Table 3*: *Sentiment difference list:* words both used by official and non-official accounts and has a significance score of over 1.0 selected and ranked by the sentiment differences. Chinese words are translated and some particle and proposition words with no actual separate meaning emitted.

Table 2: Non-official only words list

Top Words	Employment	Rational	Flexible	Civil Servants	Delayed	Individual
Original Chinese	就业	理性	灵活	公务员	延迟	个人
Significance Score	9.75	9.63	9.44	7.64	7.31	6.86

Table 3: sentiment difference list

Top words	Transfer	Decrease	Medical Reform	Annual	Benefits	Shrink	loss	Personal account
Original Chinese	划入	减少	医改	年度	待遇	缩减	吃亏	个人账户
Sentiment difference	0.89	0.65	0.63	0.62	0.58	0.57	0.55	0.52

5. Analysis

First of all, it can be seen from list 2 that biggest difference of sentiments towards same words between official accounts and unofficial accounts lies in the dissatisfaction with the medical reform and the reduction of payment to personal accounts, which is largely due to the financial expenditure problem caused by the zero-covid. Part of the financial problem, including governments spending 2-6% of expenditure on epidemic control and total government spending increased by 6.1% when revenue increased by only 0.6%, as well as that the medical insurance funds were required to undertook considerable amount of vaccination and PCR test fees, have already been explained in the introduction section. What is more, increases in spending caused by the zero-covid has significantly worsened the debt problem that has plagued the Chinese government for a long time.[20] From 2019 to 2022, the ratio of national debt to GDP has increased from 57.2% to 76.9%, about 19.7%. It is particularly noteworthy that this ratio has increased by 10.8% in one year from 2019 to 2020. In comparison, during the 9 years from 2010 to 2019, this number increased by only about 2.59% per year.[21] In December 2022, China's Ministry of Finance further issued 750 billion yuan of special bonds.[22] Protests across the country in 2022 may have already dispelled myths about the effectiveness of the Chinese Communist Party's nationalist propaganda. Economic performance may not fully explain the legitimacy of the Chinese party-state, but people's tolerance for poor economic performance is Limited. Therefore, if the debt problem cannot be effectively controlled, the risk of public opinion brought about by the economic problem may appear in other issues besides the healthcare reform, undermining the legitimacy of the Chinese party-state.

Second, the significant difference in the sentiments of official and non-official accounts towards healthcare reform also reflects the limitations of the influence of the government propaganda. Other studies have also proved that nationalist soft propaganda such as movies and short videos can effectively enhance hatred towards other countries but cannot improve the public's evaluation of the party-state itself.[23] When the masses feel dissatisfied on a large scale, the propaganda system may only generate a counterproductive effect. See the second word "rational"(*理性*) in table 2. This word is one of the most significant sources of negative emotions of unofficial accounts. In the context of healthcare reform, the term mainly refers to an article in the Economic Daily, the official newspaper of the State Council of China, responding to mass dissatisfaction, " (Please) View on the Reform of Personal Medical Insurance Accounts Rationally (*理性看待医保个人账户改革*)[24]. It is certain that among users with negative sentiments, this propagandist article has deepened such sentiment.

Third, civil servants have become a center of negative emotions, as the term civil servants *(公务员)* appears in the 6th position in table 2. Examination the Weibo posts turned out that much dissatisfaction is expressed over the alleged privileges enjoyed by civil servants in healthcare reform.[25] In addition to UEBMI, civil servants enjoy certain extra benefits in medical insurance. According to New York Times, this part of the insurance enjoyed by civil servants has not been affected by this reform. Although state media later released rumor-refusing reports claiming that the insurance of civil servants was also affected as much as everyone else, given the experience of another state-media article mentioned above, this article perhaps did not produce a significant effect among users with negative sentiments as well. [4]

Fourth, the youth unemployment problem during the zero-covid may have exacerbated dissatisfaction and indirectly increased malice against civil servants. In the 12 years from 2009 to 2021, the youth unemployment rate increased by 1.57%.[26] But only one year after 2021, the unemployment rate increased by 8.55% and about one in five Chinese youth is unemployed [27]. By contrast, civil servants are known in China as "iron rice bowls,"(*铁饭碗*) meaning they can never be laid off [28]. From 2019 to 2023, the number of people who signed up for and passed the inspection of the central government's civil service examination increased from about 1.28 million to about 2.24 million, a growth rate of about 78.9%. Before 2022, the highest of such figure is 1.66million [29]. In fact, dissatisfaction with the employment issue is directly expressed in table 2, as the word with the highest significance score in is "employment"(*就业*).

Fifth, dissatisfaction with different government actions will be crossly expressed. In context, dissatisfaction with healthcare reform is mixed with dissatisfaction with delayed retirement. The 10th word in table 2, "delay"(*延迟*), in the context mainly refers to "delayed retirement"(*延迟退休*). It reveals how disparate grievances are intertwined in social media. Dissatisfaction with employment issues has influenced dissatisfaction with the privileges of civil servants, and dissatisfaction with delayed retirement is intertwined with dissatisfaction with the reduction of personal medical insurance accounts. In complex social media networks, the emotions generated by these different issues have complex cross-effects.

Sixth, what needs to be additionally addressed is that since Weibo users tend to be overall younger [30], and the main force of the current round of protests against the medical insurance reform is elderly retirees. Weibo texts cannot provide direct analytical insights on the healthcare reform protest events. However, by manual examination of posts including keyword "elderly people"(*老年人*) in the research data, it is found that quite a few netizens believe that this healthcare reform is to favor the elderly at the expense of the young. The differences of opinion among different age groups factor may also provide insights of other unstable factors, that is, it is more difficult to try to obtain positive evaluations among different groups.

6. Discussion and Conclusion

Firstly, limitations of this study should be admitted, which are mainly reflected in two aspects: model, and data. In terms of model selection, a relatively simple method is adopted for calculating the average word-level sentiment score as the numerical mean of all posts' sentiment score including such word. Then, the usage of SVR method that does not consider the order of words into count, as well as chosen the simplest (but fastest calculating) linear kernel during modelling. In terms of data preprocessing, firstly, a series of intuitive approaches were adopted, such as when judging posts with no information value, and when classifying official and unofficial accounts. Also. there is no manual correction of segmentations. For example, "flexible"(灵活) and "employment"(就业) that appear in the list 1 often appear in the form of "flexible employment"(灵活就业), refers to a category of employment in China including self-employed and contracted employees, but it is not in the dictionary.

The significance of this study is reflected in two aspects, the first is in terms of analyzing this incident. The results of this study confirmed the three hypotheses. By comparing the words usage and sentiment values of official accounts and unofficial accounts, this study successfully identified several main sources of negative emotions, and established the logical relationship between these sources of negative emotion to the zero-covid policy. The second is in terms of methodology. Although the natural language processing (NLP) field has achieved astonishing development in recent years, especially with some large-sample models based on neural networks, there are not many academic articles trying to adopt the NLP methods for social science analysis, especially with Chinese language. This paper demonstrates that ML-based sentiment analysis methods can provide some unique insights in interpreting social science problems.

References

[1] Ren et al; 2022; Did the Integrated Urban and Rural Resident Basic Medical Insurance Improve Benefit Equity in China?; Value in Health, Volume 25; page 1548-1558

[2] Su et al; 2018; Comparing the effects of China's three basic health insurance schemes on the equity of health-related quality of life: using the method of coarsened exact matching; Health and Quality of Life Outcomes volume 16, Article number: 41.

[3] Keith Bradsher; 2023; China's Cities Are Cutting Health Insurance, and People Are Angry; New York Times; https://www.nytimes.com/2023/02/23/business/china-health-insurance-protests.html?_ga=2.142182578.1425935494.1678497557-712904221.1672874390

[4] Keith Bradsher, 2023, How Health Insurance Works in China, and How It's Changing. https://www.nytimes.com/2023/02/23/business/china-health-insurance-explained.html

[5] Shawn Yuan; 2022; Zero COVID in China: what next?; World Report; Vol 399; Issue 10338, P1856-1857

[6] Dale, G; 2022; China's 'white paper' protest movement echoes freedom struggles across Asia and the world; The Conversation; https://theconversation.com/chinas-white-paper-protest-movement-echoes-freedom-struggles-across-asia-and-the-world-195487

[7] Xiaoqian Zhu; 2023; China's provinces spent almost £43bn on Covid measures in 2022; The Guardian https://www.theguardian.com/world/2023/feb/15/china-provinces-spent-almost-pounds-43bn-on-covid-measures-in-2022

[8] BBC; zhōngguó duōshěng "fángyì zhàngběn"tūxiǎn dìfāng cáizhèng shōuzhī máodùn guǎngdōng huāliǎo 711yìyuán [Many provinces in China highlight the contradiction between fiscal revenue and expenditure in their "epidemic prevention accounts," with Guangdong spending 71.1 billion yuan]; BBC; https://www.bbc.com/zhongwen/simp/chinese-news-64635752

[9] Reuters; 2023;China's medical insurance fund paid 150 bln yuan COVID vaccine expenses in 2021-22; Reuters; https://www.reuters.com/world/china/chinas-medical-insurance-fund-paid-150-bln-yuan-covid-vaccine-expenses-2021-22-2023-03-09/

[10] Zhu et al; 2017; Health financing and integration of urban and rural residents' basic medical insurance systems in China; International Journal for Equity in Health volume 16, Article number: 194

[11] E. Batbaatar, M. Li and K. H. Ryu, "Semantic-Emotion Neural Network for Emotion Recognition From Text," in IEEE Access, vol. 7, pp. 111866-111878, 2019, doi: 10.1109/ACCESS.2019.2934529.

[12] S. Zad, M. Heidari, J. H. J. Jones and O. Uzuner, "Emotion Detection of Textual Data: An Interdisciplinary Survey," 2021 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2021, pp. 0255-0261, doi: 10.1109/AIIoT52608.2021.9454192.

[13] https://github.com/dataabc/weibo-search

[14] https://github.com/wansho/senti-weibo

[15] Y. Ding, F. Teng, P. Zhang, X. Huo, Q. Sun and Y. Qi, "Research on Text Information Mining Technology of Substation Inspection Based on Improved Jieba," 2021 International Conference on Wireless Communications and Smart Grid (ICWCSG), Hangzhou, China, 2021, pp. 561-564, doi: 10.1109/ICWCSG53609.2021.00119

[16] T Mikolov, I Sutskever, K Chen, GS Corrado, J Dean; 2013; Distributed representations of words and phrases and their compositionality; Neural information processing systems

[17] ALEX J. SMOLA and BERNHARD SCHOLKOPF; 2004; A tutorial on support vector regression; Statistics and Computing vol 14: p199–222.

[18] https://github.com/wansho/senti-weibo

[19] Chen, Xingyan, and Shaohua Zhan. "Drawn-out Protests in China's Rustbelt: Land Revenue and the Limits to Bargained Authoritarianism." The China Quarterly 252 (2022): 1162-1182.

[20] Liu, Adam Y., Jean C. Oi, and Yi Zhang. "China's Local Government Debt: The Grand Bargain." The China Journal 87, no. 1 (2021).

[21] Trading Economics; China Government Debt to GDP; https://www.statista.com/statistics/270329/national-debt-of-china-in-relation-to-gross-domestic-product-gdp/

[22] Global Times;2022; China announces special bond worth 750 billion yuan to spur economic growth https://www.globaltimes.cn/page/202212/1281553.shtml

[23] Mattingly, Daniel C., and Elaine Yao. "How soft propaganda persuades." Comparative Political Studies (2022): 00104140211047403

[24] Lianhe Zaobao; 2023; zhōngguó guānméi xū mínzhòng"suàn dàzhàng"lǐxìng kàndài yībǎo gǎigé [China's official media urges the public to "look at the bigger picture" and take a rational view of the medical insurance reform]https://www.zaobao.com.sg/realtime/china/story20230221-1365266

[25] Zhou, Xinda; Liu, Denghui;2023;tuìxiū zhígōng zhìyí yībǎo gèzhàng zījīn biànshǎo zhànghù jiégòu diàozhěng yùzǔ [Retired workers question why there's less money in their individual medical insurance accounts as adjustments to the account structure face obstacles];Caixin.com;https://www.caixin.com/2023-02-09/101996174.html

[26] World Bank; China: Youth unemployment rate from 2002 to 2021; https://www.statista.com/statistics/811935/youth-unemployment-rate-in-china/

[27] National Bureau of Statistics of China; Monthly surveyed urban unemployment rate of people aged 16 to 24 in China from December 2020 to December 2022; https://www.statista.com/statistics/1244339/surveyed-monthly-youth-unemployment-rate-in-china/

[28] D. Xiang, D. Lei and Y. Wu, "Data Visualization and Its Implementation of Civil Service Jobs Based on Python," 2020 International Conference on Communications, Information System and Computer Engineering (CISCE), Kuala Lumpur, Malaysia, 2020, pp. 283-287, doi: 10.1109/CISCE50729.2020.00063.

[29] Huatu.com;lìnián guójiā gōngwùyuán kǎoshì bàomíng rénshù [Number of applicants for the National Civil Service Examination over the years];https://www.huatu.com/guojia/zhaokao/xxhz/

[30] Chen et al; 2015; Age Detection for Chinese Users in Weibo; WAIM 2015: Web-Age Information Management; pp 83–95