

# Sentiment analysis of LDA text surrounding public opinion on China's three-child policy

**Yuyong Guo**

Material shaping and controlling engineering, Xi'an University of Technology, Shaanxi, 710048, China

guoyuyong821129@163.com

**Abstract.** In a bid to understand Chinese social media tendency on the new “Three-Child Policy”, this paper analyzes netizens sentiment on Chinese microblogging platform Weibo based on the Latent Dirichlet Allocation (LDA) topic modeling. Specifically, comments from netizens crawling on Weibo were pre-processed so that 2624 valid comments from May 30, 2022 to June 15, 2022, were hereby obtained, sentiment analysis was thus completed by using the LDA. Further analysis showed a mostly negative and neutral sentiment towards the policy among netizens, largely because of the rising cost of child rearing and poor employment security for women.

**Keywords:** LDA, Three-child Policy, fertility.

## 1. Introduction

On May 31, 2021, the Political Bureau of the Communist Party of China (CPC) Central Committee reviewed a decision on improving birth policies to promote long-term balanced population growth (Three-child policy) and proposed to further improve childbirth policies, implementing the policy that allows married couples to have three children and supporting measures. The policy aims to enhance the willingness of the Chinese people to have children and raise the country's low fertility rate.

After the announcement, the policy immediately gave rise to heated repercussions on the Internet where netizens posted their comments on the three-child policy, which are solid reference for subsequent formulation of related policies. The research, with the help of tools such as LDA topic modeling and Sentiment Analysis, uses web crawlers to get web data like public opinions about the policy and discovers the focus of greatest concerns and netizens' sentiments regarding Three-child policy, so as to provide a solid basis for the next policy formulation.

## 2. Research Background

Three-Child Policy, whereby a couple can have three children, was a family planning policy in the People's Republic of China to actively respond to population aging and would be accompanied by supportive measures. After the official implementation, different regions in China have enacted a series of favorable policies, including employment, housing, children's schooling, among others depending on

local conditions. However, the policy met with mixed reviews on the social media networks. Analysis of the Internet public opinion can be of greater help in the formulation of subsequent supporting policies.

For the huge number of user comments including multifarious contents on Weibo, manual classification as usual is no longer an option. In this regard, researchers have proposed the idea of analysis with LDA topic modeling. For instance, Min Zhang utilized the model to achieve tendency judgment and the mining and analysis of hidden information [1], Guo et al. managed to mine online comments on hotel, providing guidance and suggestions for hotels to analyze user preferences in the same way [2], and Miao Qin et al. also succeed in analyzing the public opinion of car companies and summarizing the current situation of the auto industry based on LDA topic model [3]. Similarly, sentiment analysis can be carried out to determine sentiment tendencies by analyzing large amounts of text data [4]. However, the research is suffering from a problem that is LDA topic modeling is mostly used for commercial purposes, and there are few studies on text analysis of national policies. Theoretically, combining sentiment analysis to investigate the tendency of public opinion and people's willingness can provide effective theoretical support for the follow-up of policies.

### 3. Method

According to CSISC, studies targeting public opinion of the masses are usually realized by CATI computer-aided telephone surveys, intercept interviews and household interviews. However, the above methods are confronting with the problems of difficult data collection and high labor cost [5,6]. Researching online public opinion enjoys advantages such as rapid collection and large amount of data. For the three-child policy that is the focus of this paper, the text data left by users on the online platform is analyzed with the help of text analysis methods so as to determine users' sentiment and tendencies expressed on the internet. Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

#### 3.1. Data Collection and Pretreatment

On May 31, 2021, the topic "The three-child policy is coming" ranked the first trending on weibo, with 695,000 related discussions, which aroused public concern about it. In order to obtain timely and effective data, researchers used Python programming language to write code and crawl the number of relevant comments on the three-child policy on Weibo. For the crawling data, the throwing off of repeat comments and data cleaning were carried out first to obtain 2,624 effective comments from May 30, 2022 to June 15, 2022, then Jieba was used for word segmentation and part of speech, and finally Baidu stopwords list was used to remove stop words.

#### 3.2. Data Analysis

In this study, based on the data collection and data pretreatment mentioned above, text modeling is carried out, and a text word frequency matrix conforming to LDA topic is constructed. LDA topic modeling is carried out on the word frequency matrix of the document by selecting the optimal topic words and appropriate model parameters. Then, the contents of it are judged according to the characteristic words of each sub-topic, and similar topics are merged according to the contents. Finally, the sub-dimension of Weibo users' views on the three-child policy is obtained. Since the characteristic words extracted by the LDA topic model cannot fully interpret the meaning of the entire sentence like people, this study also supplements the interpretation with text analysis. Representative comment texts, combined with China's unique social and cultural background, are used to interpret the research objects. In this study, Python's SnowNLP tool package is used to conduct sentiment analysis on the collected texts. The closer the sentiment value is to 1, the more positive the attitude displayed by a text is, and the closer to 0, the more negative the attitude displayed by a text is (see Table 1). The sentiment values of the words shown in Table 1 are divided according to the sentiment tendency criteria into negative emotion (sentiment value < 0.3), neutral emotion (sentiment value between 0.3 and 0.7), and positive emotion (sentiment value > 0.7). The emotional tendency of users is explained according to the division result.

**Table 1.** Partial data Presentation.

Comments	Sentiment Values	Sentiment Tendencies
Until the employment problem is solved, it is useless to have a three-child policy any way you like. we can't support ourselves, much less our children.	0.131785079	negative
For teachers like me who have to work while taking care of children, we really cannot afford to respond to the three-child policy.	0.642471698	neutral
There are quite a lot of favorable housing policies for families with three children, what kind of families are having three children around you?	0.889798913	positive

### 3.3. Sentiment Tendency Analysis

In this paper, by conducting LDA topic modelling, the results are visually transformed, and the distance between subjects is reflected in a Cartesian coordinate system through multi-dimension scale analysis. After adjusting the hyper-parameters in the model, the topics of texts are established according to the principles of clear topic boundaries and explicable topic meanings [7]. As shown in Table 2, the focus online on the three-child policy is mainly on the cost of children's education, housing price, and protection of women's rights and interests.

**Table 2.** Keys words of topics.

Topics	Key Words
Children's education	Kindergarten, education, secondary school, schooling, parents, college entrance exam
Housing price and subsidy	House purchasing, subsidy, family, property market, purchase restriction, housing
Protection of women's rights and interests	Women, childbirth, work, maternity leave, workplace, marriage

Overall, positive comments on the three-child policy account for 13.16% of the overall comments, while neutral and negative comments are 48.48% and 38.20% respectively, which shows that the discussion on the three-child policy on the internet is mostly neutral and negative.

Positive comments are mainly about support and expectations for the three child policy, such as "the policy will go into effect strongly, waiting for detailed regulations" and "looking forward to seeing the early implementation of complementary policies". Users are concerned about the development of more specific follow-up policies and believe that the subsequent supporting policies can provide powerful safeguard for various aspects of people's livelihood.

Further research on the subject of 'cost of children's education', with 96.77% of the negative comments. Further research showed that this high proportion is due to the high cost of raising children in today's society, where the average direct cost of raising a child aged 0-17, without taking into account factors such as income penalties, is now 190,100 yuan [8], and the increase in the number of children raises the cost of parenting, making families more cautious about having children. Meanwhile, some commentaries assume that excising "excessive involution of study" and "involution of education" refer to the phenomenon that students and their parents compete with each other to obtain better educational resources, and the final return and pay are not equal. Such phenomenon includes but is not limited to highly stressed on-campus learning and highly expensive tutoring fees, etc. The above-mentioned

phenomena all reflected the excessively high cost of children's education, which is difficult to be borne by ordinary families.

On the topic of 'housing prices and subsidies', 90.91% of comments are neutral, while the remaining comments are mainly negative, with discussions focusing on terms such as 'subsidies' and 'restrictions on purchasing'. The increase in the number of children leads to an increase in the area needed for family housing and the need for children's "school district houses". After the three-child policy took effect, many cities in China responded on house purchasing subsidies. However, the internet is mostly on the fence about it. Housing is a significant contributor to the willingness of the childbearing age people to have three children [9], but not many users have a positive attitude towards this on the Internet, indicating that the current policy still needs to be improved and formulated to address issues such as high housing prices and inequality in housing areas.

Both of the above topics are actually included in the cost of childcare, and the high cost of childcare has an impact on the public's willingness to have children. The repeated comments such as 'the financial pressure of having children' and 'we can afford to have children but we can't afford to raise them' also prove that the current high cost of childcare is a major problem. Having three children is not simply a matter of increasing the number of children, but an issue closely related to education, housing and other livelihood policies.

In the topic of "women's rights and interests", the percentage of negative comments is 98.18%. "women's employment difficulties" and "workplace discrimination" are mentioned many times. Although the corresponding measures provide certain benefits, some users believe that this also increases the employment threshold for women, reduces the willingness of companies to recruit women, and makes it difficult for women to reconcile with their career development [10].

In summary, the reason for the large number of negative comments on the three child policy is that the current social environment is not friendly to the birth of three children, and most families cannot afford the pressure that comes with having three children.

### *3.4. Suggest*

After the sentiment analysis of the LDA topic modeling, this paper finds that the online sentiment towards the three-child policy is mainly negative and neutral, which is due to the low fertility intention of the Chinese people at present. According to the above analysis, the main reasons for the generally negative attitude towards the three-child policy are the high cost of raising children (education, housing prices, etc.), the lack of adequate protection policies for women, and the concern that the policy will be difficult to implement.

At present, in order to change the current situation, many low fertility countries in the world have taken certain measures to encourage fertility. These main measures are to improve the social environment friendly to multiple children, such as providing cash incentives for newborn families, offering education subsidies for school-age children, and giving tax breaks or relief for parents and the construction of childcare institutions and so on [11]. As for the development of subsequent supporting policies, the government should take the construction of a fertility-friendly social atmosphere as the purpose. First of all, to reduce people's reproductive costs, we should start with the supply of childcare resources and housing subsidies. Secondly, to properly allocate educational resources and avoid the issue of "involution" caused by excessive concentration of resources. Thirdly, in terms of the work dilemma for women, to encourage employers to recruit women by setting up a flexible working time system and reducing taxes to protect women's employment rights. Finally, to focus on publicizing good concepts of childbirth and family and promote the policy of equal rights for men and women in childbirth, so as to reduce the issue of father's absence caused by the traditional family values [10]. Through the joint linkage of above-mentioned measures, a fertility-friendly social atmosphere of "willing to have a child, afford to have a child and support a child".

#### 4. Conclusion and Prospect

The results of the online text analysis of the three-child policy show that the current attitude of users on Weibo towards the three-child policy is generally negative and neutral. Users have voiced concerns about the high cost of housing, the high cost of educating their children, the lack of job security for women, and the difficulty of implementing the policy, indicating a low propensity to have children. Supporting policies can be formulated from the above aspects, such as making reference to the birth incentive policies of low-fertility countries and strengthening the supply of resources for building child care. The three-child policy is not just a policy of 'to have' or 'not to have' or 'how many children we need to have'. It is a major initiative that affects people's livelihoods and the economy, and it is only through the provision of support measures in various areas that can strengthen people's willingness to have children.

This study only analyses the text data of comments on Weibo, and the conclusions may vary due to user comments on other online platforms. Subsequent analysis of the tendency of public opinion on different platforms can lead to more accurate conclusions.

#### References

- [1] Zhang, M. (2020). Sentiment Analysis of e-commerce reviews based on text mining. *Industrial & Science Tribune*, 19(02), 63-64.
- [2] Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism management*, 59, 467-483.
- [3] Qin, M., & Hu, E. (2022). Emotion Recognition and Prediction of Public Opinions about Automobile Enterprises. *Journal of Hubei University of Technology*, 37(02), 116-120.
- [4] Xin, M. (2013). Research on the Function and the Realization Mechanism Based on the Public Decision-Making of Government (master), Zhejiang University, Available from Cnki
- [5] Wang, D., & Tong, B. (2013). Review of Public Opinion Research in China. *Contemporary Communication*(02), 34-37.
- [6] Shen, H., & Luo, C. (2019). The theme, interest appeal and emotional expression of the middle class in microblog discourse. *Modern Media(Journal of Communication University of China)*, 41(08), 139-145.
- [7] Chun-hua, M. (2018). The Cost of Raising Children in China and Its Policy Implications. *Collection of Women's Studies*(05), 70-84.
- [8] Zhu, X. (2021). The Influence of Housing Factors on the Three-child Fertility Willingness —— Empirical Analysis Based on CGSS2017 Data. *Journal of Fujian Business University*(06), 59-66. doi:10.19473/j.cnki.1008-4940.2021. 06.003
- [9] Nie, J., & Dong, Z. (2021). Three-child Policy: Positive Influence, Obstacles and Coping Strategies. *Journal of Guangzhou University (Social Science Edition)* , 20(06), 77-84. Retrieved from <https://kns.cnki.net/kcms/detail/44.1545.C.20211125.1314.018.html>.
- [10] Yang, X. (2016). Low Fertility Rate Countries' Pronatalist Policies and the Enlightenment to China. *Northwest Population Journal*, 37(01), 7-12. doi:10.15884/j.cnki.issn.1007-0672.2016.01. 002
- [11] Yang, J. (2021). Inclusive Childbearing and Childrearing Policy: Concepts, Contents and Strategies. *Journal of Huazhong University of Science and Technology(Social Science Edition)*, 35(03), 75-85. Retrieved from <https://kns.cnki.net/kcms/detail/42.1673.C.20210602.1500.008.html>.doi:10.19648/j.cnki.jhustss1980.2021.03.08.