

Text Mining about Online and Offline Education During Coronavirus Pandemic

Tong Xiao^{1,a,*}

¹*School of Psychology, South China Normal University, Guangzhou, China*
a. shirleytongseven@gmail.com

**corresponding author*

Abstract: Due to the release of the pandemic policy in China on December 7, 2022, many educational systems switched from frequent online education to offline education. The subjects of previous studies focused on online education during the pandemic, while few studied the contrast between online and offline education, especially the comparison before and after the Chinese pandemic policy. To fill the research gap, the current study aims to explore public opinions about online and offline education under the circumstances. And the data was collected from Weibo in two phases: from June 7, 2022, to December 7, 2022 (6 months before the Chinese pandemic policy) and from December 7, 2022 to June 7, 2023 (6 months after the Chinese pandemic policy). Then the study performed word frequency analysis and topic modeling analysis. The public opinions of various periods and subjects were examined and contrasted. The final findings could contribute to a deeper comprehension of how Chinese citizens feel about online and offline education.

Keywords: online education, text mining, COVID-19

1. Introduction

During the past few years, the COVID-19 pandemic has influenced many educational institutions all around the world. Many countries have to close their educational institutions to check the widespread pandemic [1, 2]. The Chinese government once enacted a strict zero-COVID policy. To completely eliminate the number of infected individuals, several educational activities were required to routinely switch from offline to online [3]. Online education is a significant issue in China during the epidemic [4]. On December 7, 2022, the Chinese government issued the New 10 epidemic prevention policy which relaxed lockdowns and the requirement that most public locations must be entered with proof of a recent negative COVID-19 test. After the strict Chinese pandemic policy was lifted, individuals could resume their regular lives. Compared with the frequent online education in the last 3 years, people could study or teach offline normally, which is totally different from the past.

Online education has been the subject of several studies during the COVID-19 epidemic. For example, some researchers adopted time trend analysis and text mining to analyze topics about online education during the pandemic by using data from GDELT and Twitter [5]. In addition, some researchers also conducted a similar study in China. One previous study collected data from Weibo about online education in three phases: pre-pandemic, amid-pandemic, and post-pandemic. And public opinion during these three periods was analyzed and compared in the study [4]. Furthermore, studies are also concerned about the acceptance of online education, students' online learning

experience, online teaching, user satisfaction with online education platforms, online cheating, students' academic records of online education, and the impact of online education on carbon emissions [6-12].

During the pandemic, many educational systems adopted online education, which was very important in most educational activities. After the Chinese reopening policy, many educational activities could come back offline. Following a literature review, the current study discovered that many researchers focused on text mining of online education during the pandemic. However, few prior studies were concerned with changes in public opinions about online and offline education under different Chinese pandemic policies.

Therefore, based on the above findings, the study aims to extract data from Weibo to conduct text mining concerning online and offline education before and after the reopening policy in China. The following are four research questions.

Q1: What are the most referred words regarding online and offline education before and after the reopening policy in China?

Q2: What are the main topics regarding online and offline education before and after the reopening policy in China?

Q3: What are the differences in the most referred words between online and offline education?

Q4: What are the differences between the most referred words before and after the reopening policy in China?

2. Method

2.1. Data collection

The study employed a third-party crawler called Octopus to collect data from Weibo, one of the Chinese most popular social media platforms. The keywords were “epidemic, online education” or “epidemic, offline education”. To compare the public opinions before and after the Chinese reopening policy, the study chose December 7, 2022, as the time point and gathered data of 1 year before and after the time point. Finally, the study gathered a total of 23905 posts from June 7, 2022, to June 7, 2023. Table 1 illustrates the number of posts collected using the two keywords from different periods.

Table 1: The number of posts using the two keywords from different time periods.

Keywords	Time Period	Posts Count
epidemic, online education	June 7, 2022-December 7, 2022	9982
	December 7, 2022-June 7, 2023	4754
epidemic, offline education	June 7, 2022-December 7, 2022	5564
	December 7, 2022-June 7, 2023	3605

2.2. Data Analysis

The study analyzed data using three processes: data preprocessing, word frequency analysis, and topic modeling on RStudio 2022.07.2.

First, data preprocessing was conducted to tokenize and clean the text of each dataset. The first step was to tokenize the data including removing special characters and segmenting sentences into

individual words. JiebaR package was adopted to implement this process. Then, all stopwords like personal pronouns (e.g., I, we) were removed from each dataset. Second, the study applied word frequency analysis to each dataset to count and analyze the number of important words in the text. Finally, topic modeling was conducted. Topic modeling is a statistical approach for exploring hidden topics in the dataset. The topic model used in the study was Latent Dirichlet Allocation (LDA) including two Dirichlet distribution models with one representing the topic in the document and the other representing the words in each topic.

3. Results

3.1. Word Frequency Analysis

To address Q1, Q3, and Q4, the study applied word frequency analysis to the four datasets and obtained four word clouds as illustrated in Figures 1-4. Figures 1-2 display the two word clouds from the datasets related to online education, while Figures 3-4 display the two word clouds from the datasets about offline education.

The study compared the top five words about online education before and after the Chinese reopening policy (See Figures 1-2). Results indicated that “work” replaced “epidemic” as the top five words following the policy. The word “prevention and control” was out of the top five words and “development” ranked fourth. Besides, the study also explored the difference in offline education before and after the publication of the Chinese pandemic policy (See Figures 3-4). The findings revealed that the top five referred words changed significantly. Only two words remained in the top five words after the policy. Furthermore, the difference between online and offline education was explored (See Figures 1-4). According to the results, posts regarding offline education were more likely than those about online education to use the words "personnel", "offline", "student", and "health".



Figure 1: Word cloud about online education from June 7, 2022, to December 7, 2022

Note. The top five words are epidemic, prevention and control, work, education, and school.



Figure 2: Word cloud about online education from December 7, 2022, to June 7, 2023

Note. The top five words are work, epidemic, education, development, and school.



Figure 3: Word cloud about offline education from June 7, 2022, to December 7, 2022

Note. The top five words are epidemic, prevention and control, work, personnel, and offline.



Figure 4: Word cloud about offline education from December 7, 2022, to June 7, 2023

Note. The top five words are epidemic, school, work, student, and health.

3.2. Topic Modeling Analysis

With regards to Q2, the study conducted LDA-based topic modeling analysis on the four datasets. Each dataset could be divided into two topics. The results are shown in Figures 5-8.

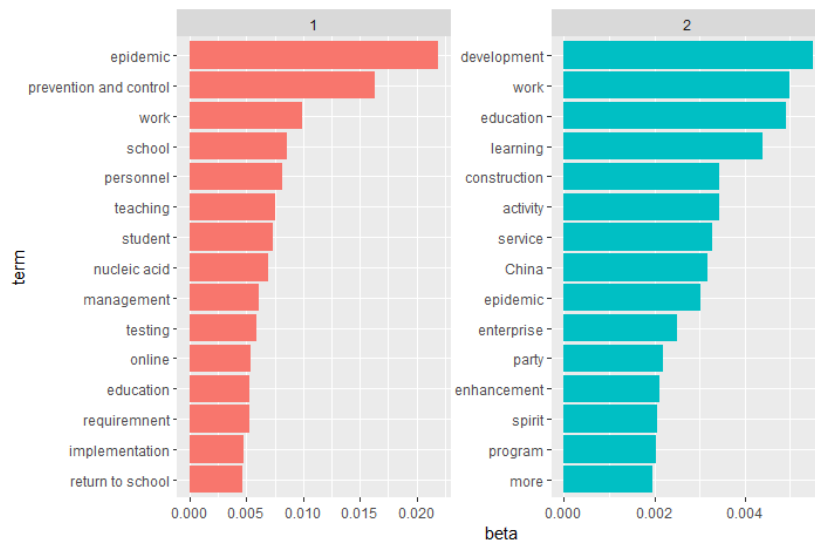


Figure 5: Topic modeling results about online education from June 7, 2022, to December 7, 2022

Note. The top five words of topic 1 are epidemic, prevention and control, work, school, and personnel. And the top five words of topic 2 are development, work, education, learning, and construction.

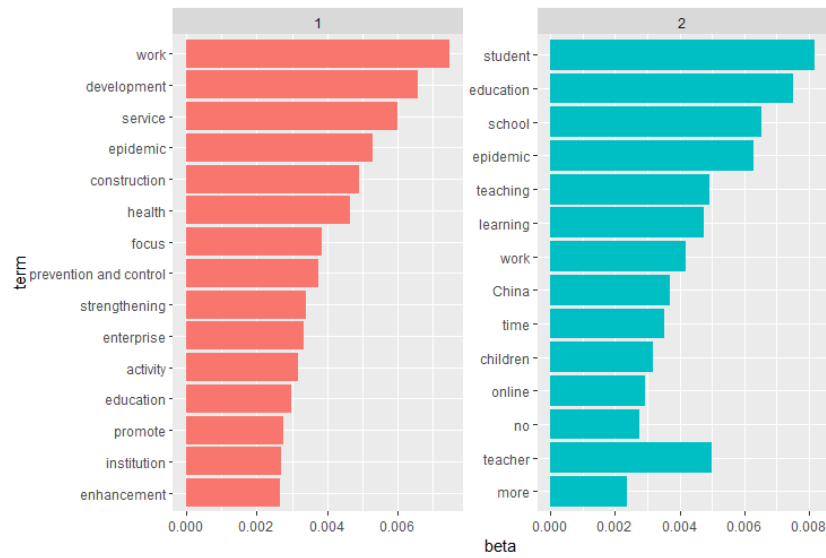


Figure 6: Topic modeling results about online education from December 7, 2022, to June 7, 2023
 Note. The top five words of topic 1 are work, development, service, epidemic, and construction. And the top five words of topic 2 are student, education, school, epidemic, and teaching.

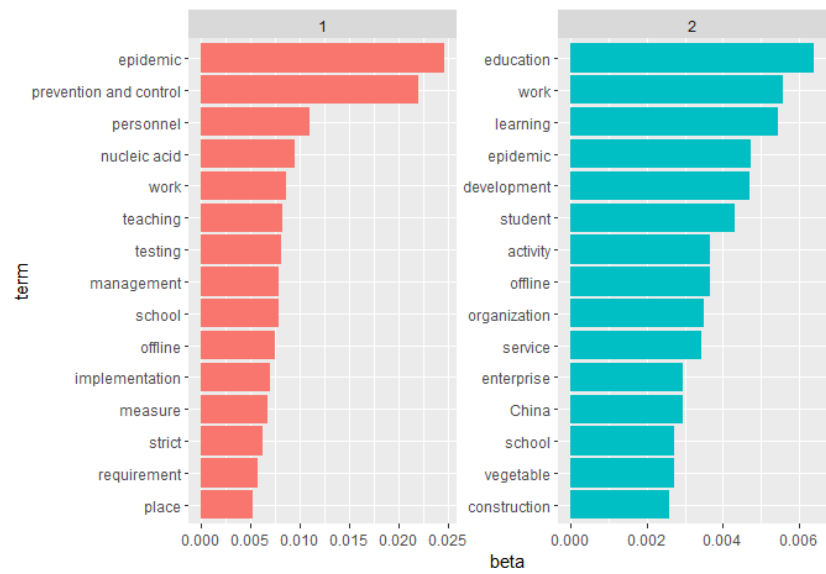


Figure 7: Topic modeling results about offline education from June 7, 2022, to December 7, 2022
 Note. The top five words of topic 1 are epidemic, prevention and control, personnel, nucleic acid, and work. And the top five words of topic 2 are education, work, learning, epidemic, and development.

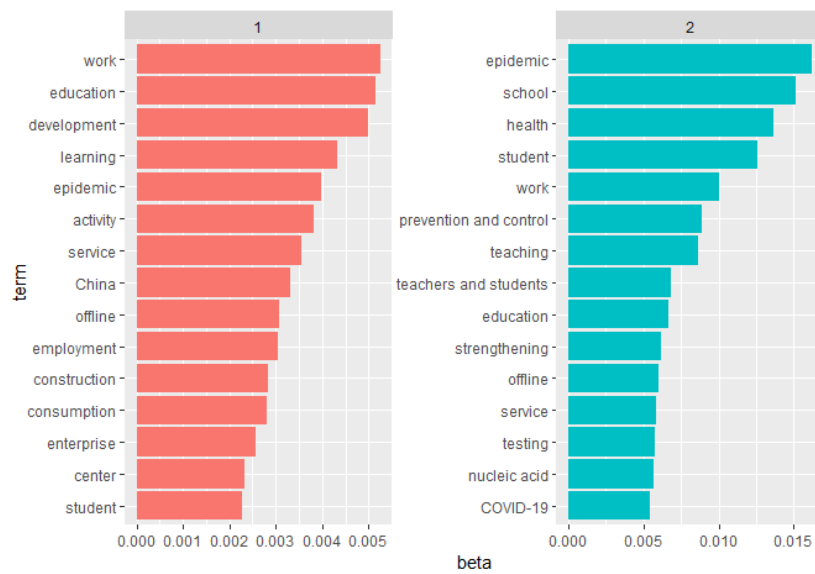


Figure 8: Topic modeling results about offline education from December 7, 2022, to June 7, 2023

Note. The top five words of topic 1 are work, education, development, learning, and epidemic. And the top five words of topic 2 are epidemic, school, health, student, and work.

4. Discussion

The study aims to examine public responses to online and offline education based on the Weibo platform. As a result, the total amount of posts was 23905 including four datasets about the two topics from different time intervals. Posts decreased compared to the time before the reopening policy was released. Then, the study conducted word frequency analysis on the four subsets. Through comparison between online and offline education, results showed that the words “personnel”, “offline”, “student” and “health” newly appeared among the top five referred words about offline education. This result may suggest that people are more concerned about their health during offline learning. Furthermore, the most frequently used terms also significantly differed before and after the reopening policy. The word “prevention and control” got out of the list and the words “development”, “student” and “health” got on the list. This result may infer that the public may pay more attention to the pandemic before the reopening policy. However, after the release of the Chinese reopening policy, the public may focus more on the nation’s development and students’ education. The findings reveal that changes in public opinion are consistent with different pandemic policies. Finally, the study conducted a topic modeling analysis and divided each dataset into two topics.

The present study has the following academic implications. First, the study performed research in the context of China, which was mostly disregarded in past academic studies about online education. Second, this is the first attempt to examine public perceptions of both online and offline education. By analyzing data before and after the reopening policy in China, researchers can know about the changes in people’s responses over the transition to normal offline education. Finally, the study adopted topic modeling for each dataset. Then researchers can observe the variation of public topics under different contexts across periods of time.

In spite of the above advantages, the study still has a few limitations. First, the study just conducted word frequency analysis and topic modeling analysis. Other analyses like sentiment analysis and social network analysis should be adopted to make a deeper analysis. Besides, the study lacked an analysis of the results of topic modeling. Researchers could make further explorations based on the findings. Finally, the current study did not distinguish the opinions of students and teachers. Future studies can be analyzed from two perspectives and compare the difference of opinions.

5. Conclusion

The current study aims to explore people's responses to online and offline education during the pandemic. First, the study extracted data from Weibo about two topics from different time intervals. Second, word frequency analysis was conducted to explore the difference in the most referred words between online and offline education, as well as the contrast before and after the Chinese reopening policy. Finally, the present study conducted a topic modeling analysis and divided each dataset into two topics. The results could contribute to a better understanding of Chinese people's responses to online and offline education under different pandemic policies.

References

- [1] S. R. Febriani, R. Widayanti, M. A. Amrulloh, and N. Mufidah, "Arabic Learning for Elementary School during COVID-19 Emergency in Indonesia," *OKARA: Jurnal Bahasa dan Sastra*, vol. 14, no. 1, Art. no. 1, May 2020, doi: 10.19105/ojbs.v14i1.3194.
- [2] P. K. Jena, "Online Learning During Lockdown Period For Covid-19 In India." *SocArXiv*, Jun. 22, 2020. doi: 10.31235/osf.io/qu38b.
- [3] W. Zhu, Q. Liu, and X. Hong, "Implementation and Challenges of Online Education during the COVID-19 Outbreak: A National Survey of Children and Parents in China," *Early Child Res Q*, vol. 61, pp. 209–219, 2022, doi: 10.1016/j.ecresq.2022.07.004.
- [4] M. Zhou and H. Mou, "Tracking public opinion about online education over COVID-19 in China," *Education Tech Research Dev*, vol. 70, no. 3, pp. 1083–1104, Jun. 2022, doi: 10.1007/s11423-022-10080-5.
- [5] Z. Fu et al., "Opinion Mining about Online Education Basing on GDELT and Twitter Data," in *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, Dec. 2020, pp. 741–745. doi: 10.1109/WIAT50758.2020.00114.
- [6] B. A. Swanson and A. Valdois, "Acceptance of online education in China: A reassessment in light of changed circumstances due to the COVID-19 pandemic," *Int J Educ Res Open*, vol. 3, p. 100214, 2022, doi: 10.1016/j.ijedro.2022.100214.
- [7] L. Yan, A. Whitelock-Wainwright, Q. Guan, G. Wen, D. Gašević, and G. Chen, "Students' experience of online learning during the COVID-19 pandemic: A province-wide survey study," *Br J Educ Technol*, vol. 52, no. 5, pp. 2038–2057, Sep. 2021, doi: 10.1111/bjet.13102.
- [8] S. M. Tsegay, M. A. Ashraf, S. Perveen, and M. Z. Zegeerish, "Online Teaching during COVID-19 Pandemic: Teachers' Experiences from a Chinese University," *Sustainability*, vol. 14, no. 1, p. 568, Jan. 2022, doi: 10.3390/su14010568.
- [9] T. Chen, L. Peng, X. Yin, J. Rong, J. Yang, and G. Cong, "Analysis of User Satisfaction with Online Education Platforms in China during the COVID-19 Pandemic," *Healthcare (Basel)*, vol. 8, no. 3, p. 200, Jul. 2020, doi: 10.3390/healthcare8030200.
- [10] M. S. Parsa and L. Golab, "Academic Integrity during the COVID-19 Pandemic: a Social Media Mining Study," *Proceedings of The 14th International Conference on Educational Data Mining*, 2021.
- [11] J. Geng, S. Xun, J. Yang, and N. Yang, "Online Education and Undergraduates' Academic Record during the COVID-19 Pandemic in China: Evidence from Large-Scale Data," *Sustainability*, vol. 14, no. 21, Art. no. 21, Jan. 2022, doi: 10.3390/su142114070.
- [12] Z. Yin, X. Jiang, S. Lin, and J. Liu, "The impact of online education on carbon emissions in the context of the COVID-19 pandemic – Taking Chinese universities as examples," *Appl Energy*, vol. 314, p. 118875, May 2022, doi: 10.1016/j.apenergy.2022.118875.