Analysis of sentiment changes in online messages of depression patients before and during the COVID-19 epidemic based on BERT+BiLSTM

Guo, Chaohui; Lin, Shaofu; Huang, Zhisheng; Yao, Yahong

published in
Health information science and systems
2022

DOI (link to publisher)
10.1007/s13755-022-00184-w

Guo et al.
Health Information Science and Systems (2022) 10:15
https://doi.org/10.1007/s13755-022-00184-w

Health Information Science and Systems

RESEARCH

Analysis of sentiment changes in online messages of depression patients before and during the COVID-19 epidemic based on BERT+BiLSTM

Chaohui Guo[1], Shaofu Lin[1*], Zhisheng Huang[2] and Yahong Yao[1]

Abstract
With the development of the Internet, more and more people prefer to confide their sentiments in the virtual world, especially those with depression. The social media where people with depression collectively leave messages is called the "Tree Hole". The purpose of this article is to support the "Tree Hole" rescue volunteers to help patients with depression, especially after the outbreak of COVID-19 and other major events, to guide the crisis intervention of patients with depression. Based on the message data of "Tree Hole" named "Zou Fan", this paper used a deep learning model and sentiment scoring algorithm to analyze the fluctuation characteristics sentiment of user's message in different time dimensions. Through detailed investigation of the research results, we found that the number of "Tree Hole" messages in multiple time dimensions is positively correlated to emotion. The longer the "Tree Hole" is formed, the more negative the emotion is, and the outbreak of COVID-19 and other major events have obvious effects on the emotion of the messages. In order to improve the efficiency of "Tree Hole" rescue, volunteers should focus on the long-formed "Tree Hole" and the user groups that are active in the early morning. This research is of great significance for the emotional guidance of online mental health patients, especially the crisis intervention for depression patients after the outbreak of COVID-19 and other major events.

Keywords: Adversarial training, BERT+BiLSTM, Time feature, Sentiment analysis, Depression

Introduction

Depression is one of the oldest and most common mental health diseases in mankind, and it often brings serious damage to a country [1–3]. In 2008, the World Health Organization listed major depressive disorder as the third cause of death and disease [4]. And it is predicted that it will ranked first in the global burden of disease in 2030. However, after suffering from depression, it is not only difficult for patients to realize it, but also difficult for doctors to recognize [5]. This phenomenon leads to depression patients usually accompanied by major pain, high morbidity and high mortality [6]. It can be seen that depression has become a serious social and medical problem [7]. Nearly half of depression patients have a sense of stigma [8, 9], which leads depression patients to be reluctant to seek help, delay in their illness, and even commit suicide [10]. With the continuous popularization of Internet terminals, more and more people like to confide their emotions in the virtual world [11], especially those with depression [12]. Usually, we call the social media where people with depression focus on their emotions as "Tree Hole."

In March 2012, a user of Microblog named "Zou Fan" committed suicide after leaving the last message on her Microblog account. After that, a large number of patients with depression left messages on her Microblog account, making it becoming the largest "Tree Hole" in Microblog, namely "Zou Fan Tree Hole". Based on the message data

*Correspondence: linshaofu@bjut.edu.cn
[1] Faculty of Information Technology, Beijing University of Technology, Beijing, China
Full list of author information is available at the end of the article

of "Zou Fan Tree Hole" and other Tree Holes, Professor Zhisheng Huang and a group of scholars have carried out research for monitor and early warn patients. Moreover volunteers such as the "Tree Hole Rescue Team" and social organizations have carried out online and offline suicide intervention and rescue [13], which has attracted great attention from the society to the "Tree Hole" population and even patients with depression. At present, there are more than 600 Tree Hole Rescue volunteers in the rescue community, including university professors and psychologists, covering all regions of the country [14]. From the end of July 2018 to the end of December 2020, the Tree Hole Rescue Team has provided assistance to 11,715 people at high risk of suicide and prevented 3629 suicides.

The outbreak of COVID-19 has led to an epidemic of fear, anxiety and depression. Patients with mental illness belong to a special population, who are more likely to experience aggravation of the disease, relapse of the disease, and increased risk of suicide when faced with acute stress events. This shows that people with mental problems are more emotionally affected by the new coronavirus epidemic. This phenomenon makes "Tree Hole" users become more active, and the number of comments increases. At the same time, new "Tree Holes" like Dr. Wenliang Li Microblog have also been formed.

Based on the message data of "Zou Fan Tree Hole", this paper used a deep learning model and sentiment scoring algorithm to analyze the fluctuation characteristics sentiment of user's messages in different time dimensions. The "Tree Hole" agent (i.e., an AI program) was used in the article to obtain the data of more than 2.3 million "Tree Hole" messages from 2012 to 2020. After preprocessing the data which includes cleaning, deduplication, and formatting, we used the Bert+BiLSTM deep learning model completes the sentiment classification task for each message data. Then this deep learning models and sentiment scoring algorithms was combined to analyze the fluctuations of the sentiment and number of messages in the "Tree Hole" at all time, day, week, month and year. It also analyzed the sentiment and quantity fluctuations of the "Tree Hole" messages before and after the outbreak of COVID-19.

The remainder of this paper is organized as follows. Section 2 introduces the literature work related to depression monitoring analysis based on Microblog "Tree Hole" and semantic analysis with machine learning and deep learning model. In Sect. 3, we describe the data set and present the proposed model and algorithm. Section 4 illustrates the designed experiment results and analysis, and Sect. 5 presents further discussions and suggestions. Finally, we draw a conclusion of this study in Sect. 6.

Related work

Research on Microblog data can help us find out the possible correlation factors between the onset of depression and suicide, so as to find suicidal people. At present, many papers have analyzed the message data of the "Tree Hole". For example, Gong et al. [15] analyzed the geographical distribution of the message data of the "Tree Hole" and found that the number of "Tree Hole" users in each region is related to the local GDP. Huang et al. [16] analyzed the change rule of the number of "Tree Hole" messages in the time dimension. Chen et al. [17] used the sentiment dictionary to score the sentiment of the "Tree Hole" message, and further analyzed the characteristics of the keywords in the message. However, most of the analysis of "Tree Hole" messages now only focuses on the number of messages, ignoring the regularity in the time dimension between the emotion of messages and the number of messages.

At present, the main sentiment analysis methods include three categories: sentiment classification methods based on sentiment dictionaries, sentiment classification methods based on machine learning, and sentiment classification methods based on deep learning. The sentiment classification method based on sentiment dictionary mainly uses Chinese sentiment dictionaries including NTUSD [18], How Net [19], and sentiment vocabulary ontology database [20]. Since most of the sentiment values of words defined in sentiment lexicons are based on general domains, Cai et al. [21] constructed a sentiment dictionary based on specific domains, and superimposed two classifiers of SVM and GBDT together, which is better than a single model. Lili et al. [22] optimized the context polarity algorithm of sentiment words by analyzing the dependency relationship of words, extracting entities and related attributes in the text to construct the domain ontology, and combining the domain ontology with the SBV (Subject-Verb) algorithm to obtain the sentiment tendency of the sentence.

Commonly used machine learning algorithms for text sentiment analysis include Maximum Entropy, Naive Bayes Model, and Support Vector Machine (SVM). Pang et al. [23] found that the use of SVM for text sentiment analysis is better than Naive Bayes and Maximum Entropy algorithms. Wikarsa et al. [24] studied an application that uses the Naive Bayes method to classify the sentiment of Twitter users.

Deep learning models are commonly used for sentiment classification include single-network models, multi-network combined models, and models with attention mechanisms. The representative deep learning models include Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM) and attention mechanism [25]. Deng [26] compared the performance of SVM, CNN and LSTM on sentiment classification tasks, and found that LSTM has the highest accuracy on the sentiment classification task of Microblog message text. Xu [27] used BiLSTM to complete the sentiment classification task of Microblog data and found that BiLSTM is superior to the LSTM model. Liu et al. [28] proposed a two-layer CNN-BiLSTM model (DASSCNN-BiLSTM) that combines attention mechanism and sentence ordering, which is superior to RNN, LSTM and BiLSTM in text sentiment classification tasks.

Szegedy et al. [29] proposed the concept of adversarial samples. Goodfellow et al. [30] put forward the concept of adversarial training for the first time, and gave a basic adversarial training method Fast Gradient Sign Method (FGSM), which is to add a disturbance to the original input sample and then use it for training. Miyato et al. [31] applied adversarial training to the text field for the first time, and improved the generalization performance of the text classification model by adding disturbances to continuous word vectors. Zhang et al. [32] used adversarial training methods and applied adversarial perturbation to the word embedding layer as a regularization technique to improve the quality of text representation.

Data set, models and algorithms

Data set and preprocessing

In this article, data such as message identity document (ID), time and text from 2012 to 2020 in "Zou Fan Tree Hole" were obtained through the "Tree Hole" agent. Due to some data fields were missing, we used distributed crawler technology to complete and correct the data. Finally, more than 2.7 million pieces of data from March 2012 to December 2020 were obtained, which is called "total data set". Considering the ethical issues of experimental data, during the process of data collection and use, we anonymized the message ID, user ID, user name and other information in the data. In order to protect the security of user information, we do not consider exposing experimental data to the public in the future. The data elements of the data set are shown in Table 1.

Science there were still some data to be standardized in the "total data set", such as " 2012-3-19 ", " 2015,1,1 ", " 0:12 ", etc. Therefore, the "total data set" was processed by the data normalization method, and the data items of "Date" and "Time" were split into finer-grained data items of "Year", "Month", "Day", "Hour" and "Minute". And according to the date of the message, the day of the week was calculated later. After the above data processing, a "processed data set" was formed, as shown in Table 2.

In order to meet the training needs of the sentiment classification model of the "Tree Hole" message text, we randomly separated 100,000 pieces of data from the "processed data set", and used public sentiment analysis tools to mark positive message texts as "1" and others as "0". The untrusted data was deleted by writing rules such as follows: (1) to delete data entries with full-symbol text; (2) to delete data entries in which the sentiment shown in the emoji does not match the marked sentiment; (3) randomly to select 200 pieces of data, analyze the characteristics of the text with errors, and delete the data entries with the same characteristics; (4) only to keep the message text and its emotional tags. After the above processing, we took the intersection of datasets processed by two researchers and balanced the amount of data for positive and negative labels. Finally, 60,000 pieces of data were obtained, forming a "labeled data set", as shown in Table 3.

Table 1 Total data set

Date	Time	Message-ID	User-ID	User-Name	Message	Other
2012-3-19	00:00	Null	165...711	Bob	A sigh.	Null
……	……	……	……	……	……	……
2020-12-28	16:42	458...335	622...929	Alan	I thought I could do it, but I couldn't.	It's hard

Table 2 Processed data set

Year	Month	Day	Hour	Minute	Week	User-ID	User-Name	Message
2012	03	19	00	00	1	165...711	Bob	A sigh.
……	……	……	……	……	……	……	……	……
2020	12	28	16	42	1	622...929	Alan	I thought I could do it, but I couldn't.

Sentiment classification model

In order to ensure better results in the short text sentiment classification task mentioned in this article, we divided the "labeled data set" into a training set, a validation set, and a test set at a ratio of 8:1:1. We used the training set to train 4 different deep learning models, and compared the training times and accuracy of the models under the sentiment classification task of "Tree Hole" short text messages on Microblog. The results are shown in Fig. 1. Through comparison, the Bert+BiLSTM model is most suitable for the short text classification task.

Model adversarial training

Adversarial training is to enhance the robustness of the model by attacking and defending. In the field of NLP, adversarial training is used as an effective method to improve the generalization ability of the model. Therefore, we used adversarial methods to train the Bert+BiLSTM classification model. We added a perturbation action r_{adv} to the original input sample x of the BiLSTM model as a confrontation example, and then used the confrontation example for training, as shown in formula (1).

$$\min_{\theta} E(x, y_{label}) \sim D\left[\max_{r_{adv}\in S} L(\theta, x + r_{adv}, y_{label})\right] \quad (1)$$

The formula is divided into two parts. The inner layer is the maximization formula, where x represents the sample input into the model, and r_{adv} represents the disturbance superimposed on the input. $L(\theta, x + r_{adv}, y_{label})$ is the loss function of the model, y_{label} is the label of the sample, and $x + r_{adv}$ means superimposing a disturbance r_{adv} on the sample x. The outer layer is the minimization formula, that is, when the disturbance is fixed, the neural network model with the smallest loss is trained to make the model robust.

Sentiment scoring algorithm

This article uses a trained sentiment classification model to calculate the positive probability $(\widehat{P})$ and negative probability $(\widehat{N})$ of all message texts. Then it is processed by the Softmax algorithm to make $\widehat{P} + \widehat{N} = 1$, so that the sentiment score of the message text can be obtained by calculating the expected value, as shown in formula (2).

$$S = P * \widehat{P} + N * \widehat{N} \quad (2)$$

In the formula, S represents the emotional score of the message text, P represents the weight of positive sentiment, and N represents the weight of negative sentiment. This article set the weight of positive sentiment to 1 and the weight of negative sentiments to 0. Therefore, formula (2) can be simplified to formula (3).

$$S = \widehat{P}, S \in [0, 1] \quad (3)$$

After the message text passes the sentiment classification model, the probability that the text is a positive sentiment

Table 3 Labeled data set

Message	Target
No day is happy	0
……	……
Happy new year, Fanfan	1

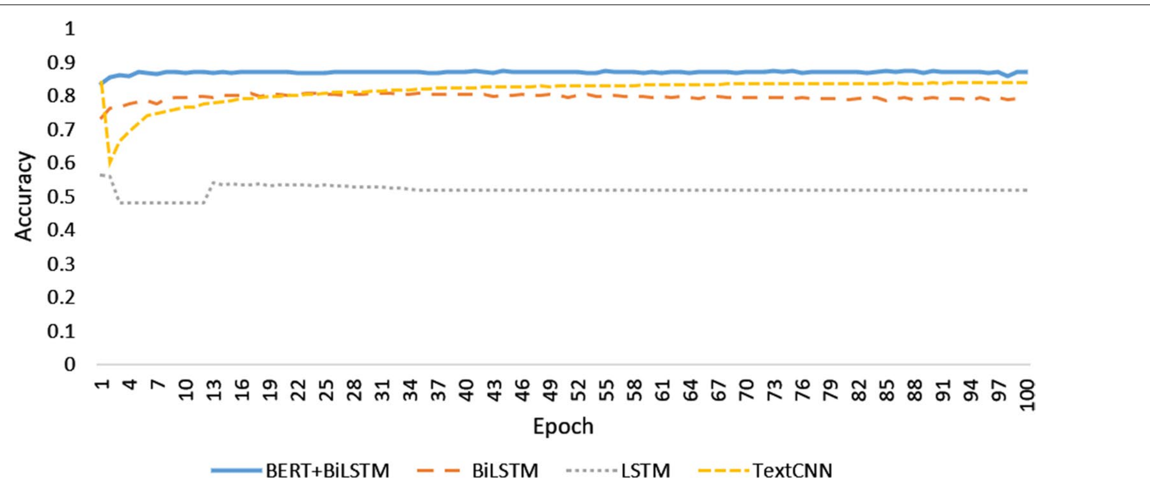

Fig. 1 The performance of the 4 models on the task of the short text classification

can be obtained ($\widehat{P}$). The probability of positive sentiment is multiplied by the weight, and the score of the sentiment of the message can be obtained. The higher the score of the sentiment of the message, the more positive the sentiment of the message. Through formula (3), the sentiment scoring algorithm of each time dimension can be obtained:

$$y = \frac{\sum_{i=1}^{n} S\left(Day\_hour = x\right)_i}{n}, \quad x \in [0, 23], \quad S \in [0, 1] \tag{4}$$

$$y = \frac{\sum_{i=1}^{n} S\left(Week\_day = x\right)_i}{n}, \quad x \in [1, 7], \quad S \in [0, 1] \tag{5}$$

$$y = \frac{\sum_{i=1}^{n} S\left(Month\_day = x\right)_i}{n}, \quad x \in [1, 31], \quad S \in [0, 1] \tag{6}$$

$$y = \frac{\sum_{i=1}^{n} S(Year\_month = x)_i}{n}, \quad x \in [1, 12], \quad S \in [0, 1] \tag{7}$$

$$y = \frac{\sum_{i=1}^{n} S(Year = x)_i}{n}, \quad x \in [2012, 2020], \quad S \in [0, 1] \tag{8}$$

Formulas (4), (5), (6), (7) and (8) are respectively used to calculate the emotional scores of "Tree Hole" messages in 1 day, 1 week, 1 month, 1 year and so far. Where x represents a certain time period, y represents the average emotional score of all messages in this time period, n represents the total number of message texts in this time period.

Experimental results and analysis

Analysis of "Tree Hole" message sentiment and number in a day

Formula (4) was used in this paper to calculate the sentiment score of "Tree Hole" messages in a day. The average change curve of sentiment and quantity of "Tree Hole" messages in a day were drawn according to the score to analyze the correlation between messages and sentiment, as well as the change of curve before and after the outbreak of COVID-19 epidemic, as shown in Figs. 2 and 3. Then the change law of the sentiment of the "Tree Hole" group in a day, and the impact of major events such as the outbreak of the epidemic on the sentiment of the depression group can be inferred.

From the sentimental and quantitative change curve of "Tree Hole" messages drawn in days, the following characteristics of "Tree Hole" people and the messages can be analyzed:

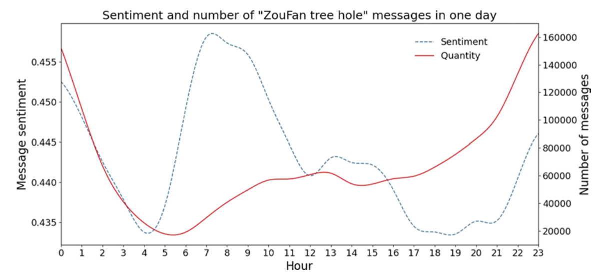

Fig. 2 Sentiment and number of "Zou Fan Tree Hole" messages in one day

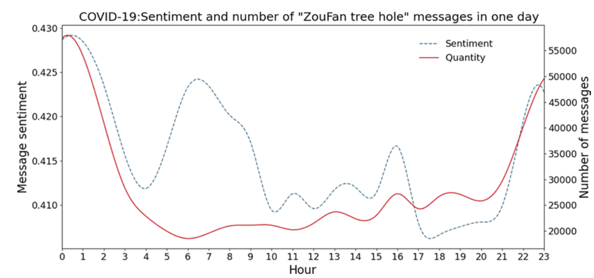

Fig. 3 COVID-19: sentiment and number of "Zou Fan Tree Hole" messages in one day

1. Before the outbreak of COVID-19 epidemic, the changes in number of "Tree Hole" messages in one day basically showed a trend of high at both ends and low in the middle. This rule is in line with the schedule of people who work and study during the day. That is, most "Tree Hole" users were used to pouring out their feelings to "Tree Hole" at night. Combined with the change trend of "Tree Hole" message sentiment in a day, it can be found that the closer the night was to the early morning, the more positive the message sentiment was, and the more the number of messages were. The highest point of sentiment in the "Tree Hole" in a day occurred between 7 and 8 AM. The sentiment between 0 and 4 AM showed a rapid downward trend.
2. After the outbreak of COVID-19 epidemic, the number curve of "Tree Hole" messages also showed a trend of high at both ends and low in the middle. The maximum number of "Tree Hole" messages appeared at 1 AM, the number of messages gradually decreased from 1 to 6 AM, and did not begin to rise until 8 PM The highest point of "Tree Hole" sentiment occurred between 0 and 1 AM every day.
3. Comparing the curve of the number and the sentiment of "Tree Hole" messages in one day before and after the outbreak of the epidemic, it could be found that the overall sentiment of "Tree Hole" after the outbreak of the epidemic was lower than that before the outbreak of the epidemic, and the sentiment fluc-

tuation was more intense. After the outbreak of the epidemic, there was no significant change in the fluctuation trend of the number of “Tree Hole” messages in one day, but the peak of the number of messages was delayed by 1 h.

Analysis of “Tree Hole” message sentiment and number in a week

Formula (5) was used in this paper to calculate the score of “Tree Hole” message sentiment in a week, and the change law curve of “Tree Hole” message sentiment and quantity in a week were drawn according to the score, as shown in Figs. 4 and 5. Then we analyzed the change law of the sentiment of the people in the “Tree Hole” in a week, and the impact of major events such as the outbreak of the epidemic on the sentiment of the depression.

From the sentimental and quantitative change curve of “Tree Hole” messages in weeks, the following characteristics of “Tree Hole” user group and the messages could be analyzed:

1. The sentiment of the message text in “Tree Hole” showed an obvious downward trend in a week, and only increased slightly on Friday. The number of messages also showed an obvious downward trend within a week. The fluctuation of the two curves in Fig. 3 is also roughly the same.
2. After the outbreak of COVID-19 epidemic, the sentiment of “Tree Hole” message text fluctuated greatly. Sentiment in Monday was low, but the sentimental change showed an upward trend until Thursday, when the sentiment reached the first peak in a week. Then the sentimental value decreased sharply, and reached their lowest values in a week on Friday and Saturday. But the sentimental value on Sunday rose rapidly, reaching the second peak in a week.
3. Comparing the curve of the number and the sentiment of “Tree Hole” messages in 1 week before and after the outbreak of the epidemic, it could be found that the overall sentiment of “Tree Hole” after the outbreak of the epidemic is lower than that before the outbreak of the epidemic. On the sentiment and quantity curve of message text in week, it could be clearly seen that the sentiment and quantity of message before and after the outbreak of the epidemic changed synchronously.

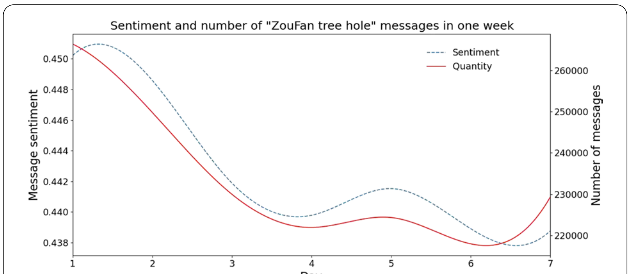

Fig. 4 Sentiment and number of “Zou Fan Tree Hole” messages in 1 week

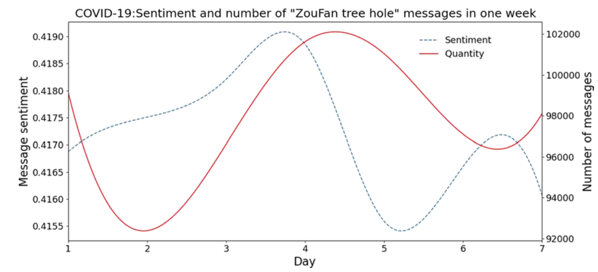

Fig. 5 COVID-19: Sentiment and number of “Zou Fan Tree Hole” messages in 1 week

Analysis of “Tree Hole” message sentiment and number in a month

Formula (6) was used in this paper to calculate the sentimental score of “Tree Hole” messages in 1 month, and change curve of sentiment and quantity of “Tree Hole” messages in 1 month was drawn according to the score, as shown in Figs. 6 and 7. Then the change law of the sentiment of the main user groups in the “Tree Hole” in 1 month, and the impact of major events such as the outbreak of the epidemic on the sentiment of the depression were analyzed.

From the sentimental and quantitative change curve of “Tree Hole” message text drawn in month, the following characteristics of “Tree Hole” people and the messages could be analyzed:

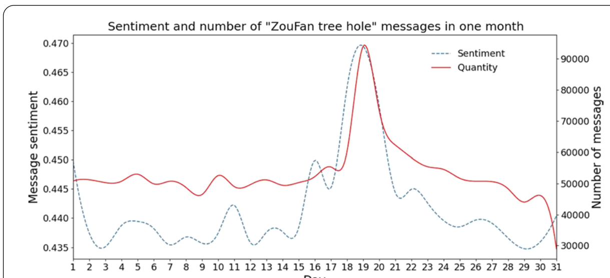

Fig. 6 Sentiment and number of “Zou Fan Tree Hole” messages in 1 month

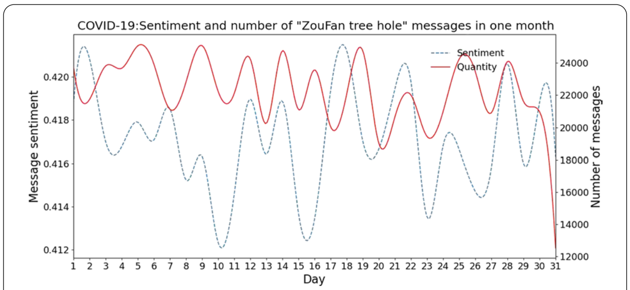

Fig. 7 COVID-19: sentiment and number of "Zou Fan Tree Hole" messages in 1 month

1. The sentiment of the message text in "Tree Hole" fluctuated basically in January, showing a rapid upward trend only in the days around the 19th, and then tended to be flat. The fluctuation of the number of message text within a month was similar to the fluctuation of sentiment, and there would be an explosive growth trend only on the 18th and 19th. The reason is that in the early morning of March 17, 2012, Zou Fan hanged herself, and on March 19, the death of Zou Fan was officially announced. Then the users of "Zou Fan Tree Hole" would post more positive messages around the 19th to commemorate her.
2. After the outbreak of COVID-19 epidemic, the fluctuation of sentiment and quantity of "Tree Hole" message text became irregular. The sentiment of the message text reached the lowest point around the 15th day of a month and reached the highest point of a month at the 30th day. The number of messages reached the highest point on the 8th day of a month and reached the lowest point in a month on the 31st.
3. Comparing the curve of the number and the sentiment of "Tree Hole" messages in 1 month before and after the outbreak of the epidemic, it can also be found that the overall sentiment of "Tree Hole" after the outbreak of the epidemic was lower than that before the outbreak of the epidemic, and the sentiment fluctuation was more intense. Comparing the number of messages in the two figures, it can be found that before and after the outbreak of the epidemic, the end of the month was the time with the least number of messages.

Analysis of "Tree Hole" message sentiment and number in a year

Formula (7) was used in this paper to calculate the sentimental score of "Tree Hole" messages in each month, and the change curve of sentiment and quantity of "Tree Hole" messages in each month was drawn according to the score, as shown in Figs. 8 and 9. Then the change law of the sentiment of people in the "Tree Hole" in each month, and the impact of major events such as the outbreak of the epidemic on the sentiment of the depression was analyzed.

From the sentimental and quantitative change curve of "Tree Hole" message drawn in year, the following characteristics of "Tree Hole" users groups and the messages can be analyzed:

1. The sentiment of messages in "Tree Hole" showed regular fluctuation in years. The sentiment of the year reached the most positive state in March and reaches the second peak from May to June. The change trend of the number of messages was basically consistent with the change trend of sentiment. The number of messages reached the peak in 1 year in March, and then decreased rapidly. The change of the number of messages tended to be gentle in May, and the number of messages showed a temporary increase in September.
2. After the outbreak of COVID-19 epidemic, the fluctuation trend of sentiment and quantity of "Tree Hole" message text was roughly the same. It showed an upward trend from January to March, a trough from April to May, and then rose slowly.

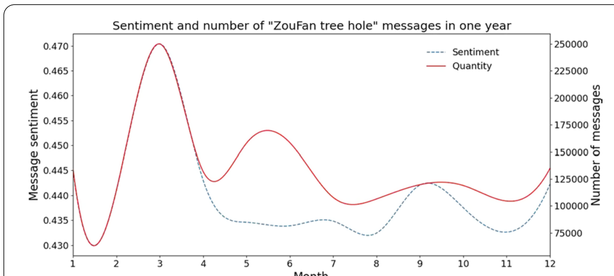

Fig. 8 Sentiment and number of "Zou Fan Tree Hole" messages in 1 year

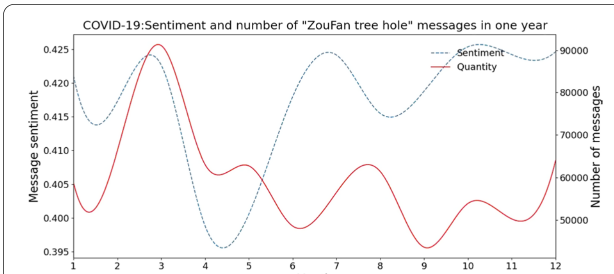

Fig. 9 COVID-19:Sentiment and number of "Zou Fan Tree Hole" messages in 1 year

3. Comparing the curve of the number and the sentiment of "Tree Hole" messages in year before and after the outbreak of the epidemic, it can also be found that the overall sentiment of "Tree Hole" after the outbreak of the epidemic is lower than that before the outbreak of the epidemic. The change trend of message number and sentiment was still similar. Before and after the epidemic, March was a very positive month for "Tree Hole" messages.

Analysis of messages sentiment and number from 2012 to 2019

Formula (8) was used to calculate the number of messages and average sentiment score of every year from the generation of "Zou Fan Tree Hole" to 2020. According to the score, the change curve of sentiment and number of "Tree Hole" messages in each year were drawn, as shown in Fig. 10. Then the change law of the sentiment of the main user groups in the "Tree Hole" in each year, and the impact of major events such as the outbreak of the epidemic on the sentiment of the depression was analyzed.

From the sentimental and quantitative change curve of "Tree Hole" message text drawn in year, we can analyze the following characteristics of people in "Zou Fan Tree Hole" and its message:

1. The sentiment of the messages in Zou Fan" Tree Hole" showed regular fluctuations in years. On the whole, the sentiment of "Zou Fan Tree Hole" messages from its formation to 2020 showed a downward trend year by year. Only the sentiment score in 2013 is more than 0.5, which means that the positive sentiment of the messages in that year is greater than the negative sentiment.
2. The number of messages in "Zou Fan Tree Hole" showed a fluctuating and increasing trend year by year, which is just opposite to the curve of sentiment score. Since 2016, the number of messages in "Zou Fan Tree Hole" has accelerated and reached the highest in 2020. The reason is that the suicide of Chinese online writer Pu Erding in 2016 pushed "Zou Fan Tree Hole" into the trending of Microblog.

Discussion and suggestion

Discussion

It can be seen from the results that the sentiment and quantity of "Tree Hole" messages show regular fluctuations. Therefore, we can infer the reasons for the regular fluctuation of sentiment and quantity of messages in different time dimensions as follows:

1. According to the analysis of the number of "tree hole" messages in 1 day and 1 week, the following inferences can be made. Most users in the "Tree Hole" may be people who are busy in the daytime such as students or staff. This group is used to pouring out their feelings in the "Tree Hole" after a day's work or study. Having good work and rest is conducive to maintaining positive sentiments. The more "Tree Hole" messages are, the more positive the sentiment is, and vice versa. This phenomenon shows that appropriate communication can eliminate some negative sentiments.
2. Comparing the sentiment of messages before and after the epidemic, it can be inferred from Figs. 4

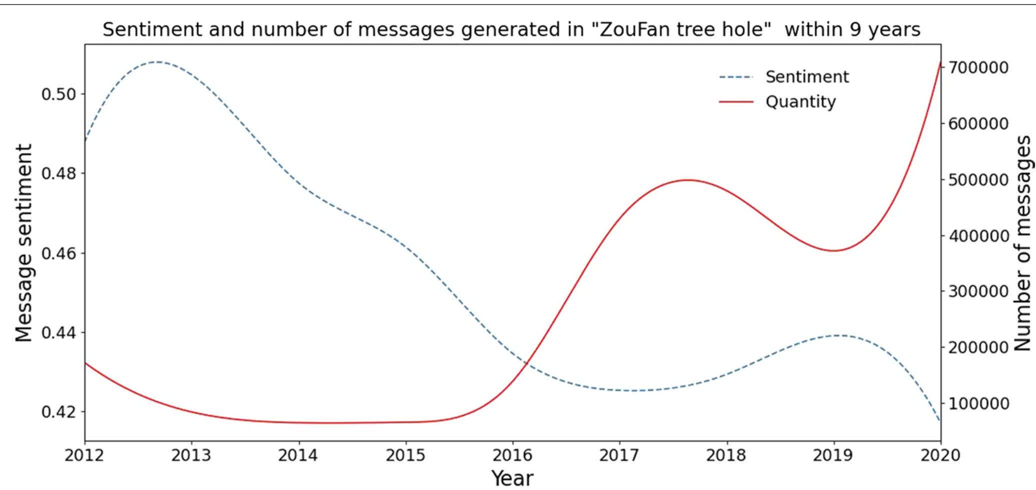

Fig. 10 Sentiment and number of messages generated in "Zou Fan Tree Hole" within 9 years

and 5. After the outbreak of COVID-19 epidemic, the overall sentiment of "Tree Hole" messages is far lower than before, which indicates that the outbreak of COVID-19 epidemic has a negative impact on the "Tree Hole" users. It makes people fear going out on weekends, and people at home on weekends are prone to negative sentiments.

3. According to the analysis of the number of messages in the "tree hole" and the trend of sentiment change in Sect. 4.5, we can make the following inferences. It can be observed that the quantity of messages in the "Tree Hole" shows an upward trend year by year, from the emergence of the "Tree Hole" to 2020, while the sentiment of messages in the "Tree Hole" shows a downward trend, which shows that the "Tree Hole" will gradually attract a large number of people with negative sentiments to leave messages. That is, the longer the "Tree Hole" occurs, the more dangerous it is. Moreover, the negative events such as the suicide of Pu Erding in 2016 and the outbreak of COVID-19 epidemic in 2019 will have a huge impact on the "Tree Hole" message.

Suggestion

Most users in the "Tree Hole" are students and staff, who prefer to enter the virtual world at night to talk about their sentiments. But staying up late and having other irregular habits will aggravate their negative sentiments. The volunteers such as the Tree Hole Rescue Team and so on should focus on people with irregular work and rest in the "Tree Hole" and improve the early warning sensitivity on those people. And in the implementation of crisis intervention and psychological counseling, they should guide patients to adjust their work and rest time to prevent the fermentation of their own negative sentiments.

Moreover, the Tree Hole Rescue Team should take the initiative to make positive comments in the "Tree Hole", especially at the time when the "Tree Hole" users are depressed, such as the early morning of every day, every Saturday, the middle and late days of every month. The longer the "Tree Hole" is formed, the lower the user's sentiment are. Therefore, early warning and rescue efforts should be strengthened for the "Tree Hole" that has been formed for a long time. In addition to paying attention to the special time period, the volunteers such as the Tree Hole Rescue Team and so on should also pay attention to the impact of major events on the mood of "Tree Hole" users, especially social celebrity suicide, epidemic outbreak and other events.

Conclusion and outlook

In this paper, we analyze the sentiment and quantity of the message text data in "Zou Fan Tree Hole" from multiple time dimensions, especially for the impact of the outbreak of the COVID-19 epidemic on the sentiment of "Tree Hole" users, and explains possible reason. We conclude that the sentiment and number of messages will fluctuate regularly as time goes by. "Tree hole" users with irregular work and the rest are more susceptible to negative emotions. However, in most cases, "Tree hole" emotions will be affected by the number of messages, and more communication can drive "Tree Hole" users to have positive emotions. Incidents such as the outbreak of the COVID-19 epidemic and the suicide of Chinese online writer called Pu Erding will have a negative impact on the emotions of "Tree Hole" users. The longer the "Tree Hole" is formed, the greater the proportion of people with negative emotions and the more concentrated depression patients.

This paper has completed a multi-dimensional time characteristic analysis of the sentiment and quantity of Microblog "Tree Hole" messages. In addition, we also analyzed the impact of the outbreak of the COVID-19 epidemic and Pu Erding suicide on "Tree Hole". However, we have not yet analyzed the impact of positive social events on "Tree Hole". Therefore, we will continue to collect more data to analyze events that have a positive impact on the sentiment of "Tree Hole" users.

Author details
[1]Faculty of Information Technology, Beijing University of Technology, Beijing, China. [2]Department of Computer Science, Vrije University Amsterdam, Amsterdam, The Netherlands.

Received: 9 January 2022 Accepted: 2 July 2022
Published online: 13 July 2022

References

1. Berto P, D'Ilario D, Ruffo P, et al. Depression: cost-of-illness studies in the international literature, a review. J Ment Health Policy Econ. 2000;3(1):3–10.
2. Gotlib IH, Hammen CL, editors. Handbook of depression. New York: Guilford Press; 2008.
3. Kessler RC, Aguilar-Gaxiola S, Alonso J, et al. The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. Epidemiol Psychiatric Sci. 2009;18(1):23–33.
4. World Health Organization. The global burden of disease: 2004 update. Geneva: World Health Organization; 2008.
5. Guan L, Hao B, Liu T, et al. A pilot study of differences in behavioral and linguistic characteristics between Sina suicide microblog users and Sina microblog users without suicide idea. Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi. 2015;36(5):421–5.
6. Cassano P, Fava M. Depression and public health: an overview. J Psychosom Res. 2002;53(4):849–57.
7. Wang R, Huang S. Research progress on the pathogenesis of depression. J Med Postgrad. 2014;27(12):1332–6. https://doi.org/10.16571/j.cnki.1008-8199.2014.12.023.

8. Levy B, Tsoy E, Brodt M, et al. Stigma, social anxiety, and illness severity in bipolar disorder: implications for treatment. Ann Clin Psychiatry. 2015;27(1):55–64.
9. Seeman N, Tang S, Brown AD, et al. World survey of mental illness stigma. J Affect Diserd. 2015;190(1):115–21.
10. Xue X, Li W. Research progress of stigma in patients with depression. J Clin Psychiatry. 2018;28(3):212–4.
11. Mislove A, Lehmann S, Ahn Y Y, et al. Understanding the demographics of Twitter users. Proceedings of the International AAAI Conference on Web and Social Media. 2011
12. Guo C, Lin S, Huang Z, et al. Mental Health Question and Answering System Based on Bert Model and Knowledge Graph Technology. Proceedings of the 2nd International Symposium on Artificial Intelligence for Medicine Sciences. 2021;472–476.
13. Huang Z, Hu Q, Gu G. Network intelligent robot and suicide monitoring and early warning. Chin Digital Med. 2019;14(3):2–6.
14. Guo C. AI tree cave rescue group: saving lives with AI. Manag Rese Sci Technol Achiev. 2021;16(5):66–7.
15. Gong J, Lin S, Huang Z. Research on data spatial characteristics of depression patients with microblog "tree hole". Chin Digit Med. 2020;15(4):70–4.
16. Huang Z, Min Y, Lin F, et al. Temporal characteristics of suicide information in social media. Chin Digit Med. 2019;14(3):7–10.
17. Chen P, Qian X, Huang Z, et al. Analysis of negative emotional characteristics of microblog "tree hole" message. Chin J Mental Health. 2020;5.
18. Ku LW, Chen HH. Mining opinions from the Web:beyond relevance retrieval. J Am Soc Inf Sci Technol. 2007;58(12):1838–50.
19. HowNet. Affective analysis word set[EB/OL].(2013-10-16)[2019-11-10]. http://www.keenage.com/html/c_index.html.
20. Xi L, Lin H, Pan H, et al. Construction of emotional vocabulary ontology. J Inf. 2008;27(2):180–5.
21. Cai Y, Yang K, Huang D, et al. A hybrid model for opinion mining based on domain sentiment dictionary. Int J Mach Learn Cybern. 2019;10(8):2131–42.
22. Lili D, Fanrong Z, Xiang Z. Analysing propensity of product reviews based on domain ontology and sentiment Lexicon. Comput Appl Softw. 2014;31(12):104–8.
23. Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. arXiv：cs/0205070, 2002.
24. Wikarsa L, Thahir SN. A text mining application of emotion classifications of Twitter's users using Naive Bayes method. 2015 1st International Conference on Wireless and Telematics (ICWT). IEEE, 2015;1–6.
25. Yadav A, Vishwakarma DK. Sentiment analysis using deep learning architectures: a review. Artif Intell Rev. 2020;53(6):4335–85.
26. Deng Y. Research on emotion analysis of Chinese microblog using deep learning technology. Southwest Jiaotong University, 2018.
27. Xu D. Emotion analysis of microblog based on deep learning. Changchun University of Technology, 2019.
28. Liu F, Xu M, Deng X. Affective analysis combining attention mechanism and sentence ordering. Comput Eng Appl. 2020;56(13):12–9.
29. Szegedy C, Zaremba W, Sutskever I, et al. Intriguing properties of neural networks. arXiv：1312.6199, 2013.
30. Goodfellow IJ, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. arXiv：1412.6572, 2014.
31. Miyato T, Dai AM, Goodfellow I. Adversarial training methods for semi-supervised text classification. arXiv：1605.07725, 2016.
32. Zhang X, Yu S, Wang Q, Xu B. Text representation and classification algorithm based on countermeasure training. Comput Sci. 2020;47(1):12–6.