

RESEARCH ARTICLE

Research on the Geospatial Characteristics of Emotional Expression in Micro-blog “Tree Hole”

Yahong Yao¹, Shaofu Lin^{1,*}, Zhisheng Huang², Ying Wu¹

¹Faculty of Information Technology, Beijing University of Technology, Beijing, China

²Department of Computer Science, Vrije University Amsterdam, Amsterdam, The Netherlands

ARTICLE DATA

Article History

Received 30 April 2022

Revised 25 November 2022

Accepted 06 March 2023

Keywords

“Tree Hole”

Knowledge graph

Sentiment analysis

Geospatial characteristics

ABSTRACT

This paper has finished a fine-grained sentiment analysis of micro-blog “Tree Hole” messages, and studied the distribution characteristics of emotional expression of the message makers in China's geospatial area, so as to provide support for the monitoring, early warning, intervention and rescue for patients with depression in the “Tree Hole” who have emotional deterioration and suicidal behavior. The Tree-Hole Intelligent Robot was used to capture the messages in the “Tree Hole”, and the standardized message data was obtained after data processed. Then an affective lexicon was set up together with the knowledge graph technology to count and analyze the emotional intensity scores of each province of China under the six emotions of joy, anger, sadness, disgust, fear and surprise, and the geospatial visualization results were presented. There are differences in the emotional expression of message makers in different provinces of China in “Tree Hole”. The emotional expression of the message makers in “Tree Hole” may be related to the economic condition, climatic condition, social background of the region, and may reflect the mental state of local netizens to a certain extent.

1. INTRODUCTION

Depression is one of the most common mental disorders faced by human society nowadays. The suicide rate of patients with depression has been extremely high in recent years. Around the world, the number of suicides per day in the world is about 3000 [1]. Suicide has gradually become a global public health problem. For adolescents, depression is one of the main causes of disability, and suicide is the second leading cause of death for adolescents aged 15–29 [2]. For adults, people with mental illness die prematurely, usually 20 years in advance.

“Tree Hole” refers to a public microblog which can be used for people to confide their secrets and express opinions anonymously and freely. At present, the largest “Tree Hole” in micro-blog comes from a user named “Zoufan”. After her death, many people, most of whom has mental illness, suicidal ideation and even preparing to commit suicide, chose to use her last micro-blog as the “Tree Hole” to express themselves.

There have been multiple studies on the “Tree Hole” crowd. However, these researches were basically limited to the analysis based on the time dimension. Jing XiaoMin et al. [3] studied the changes of the number of messages in tree holes at different time granularity. Guo Chaohui et al. [4] studied the variation

sentiment of “Tree Hole” before and after the COVID-19 epidemic. As for the analysis of spatial dimension, Gong Jingqiu et al. [5] studied the number distribution of tree holes in spatial dimension, but only studied the influence of economic factors. This study mainly studies the spatial distribution characteristics of emotional intensity in tree holes in 2019, gives a visual analysis, and analyzes climate, society and other factors behind the phenomenon.

In this paper, the Tree-Hole Intelligent Robot, which was used for monitoring “Tree Hole”, was used to capture the messages in the “Tree Hole”. Then the standardized messages were obtained by compiling and running a Python program to clean the data. And an emotional dictionary of negative emotion words was built by combining Chinese Emotion Word Ontology and negative word list. The negative emotion word dictionary, Chinese Emotion Word Ontology and messages were transformed into RDF data and loaded into GraphDB platform. In this experiment, by compiling and operating SPARQL query instructions, the emotional words, negative emotional words and their intensity of emotion in the messages of various provinces of China in 2019 were found out, then the emotional degree scores of 34 province in China under six emotions of joy, anger, sadness, fear, surprise and disgust were counted. And the visual results were shown. It is concluded that the emotional expression of the

*Corresponding author. Email: linshaofu@bjut.edu.cn

users in “Tree Hole” may be related to the economic conditions and social background of the geographical region.

2. DATA RESOURCES AND DATA PROCESSING

2.1. Data Resources

2.1.1. Messages of “Tree Hole”

The data of the experiment comes from the comment section of the last message of Microblog users named Zoufan. After the Tree-Hole Intelligent Robot which is provided by the founder of Tree Hole Rescue Team crawled the messages, 2356066 microblog messages from March 18, 2012 to August 31, 2020 were collected, forming the “Data set1” shown in Table 1.

2.1.2. Regional information of users in “Tree Hole”

Our purpose is to analyze the spatial distribution characteristics of users' intensity of sentiment, so their regional information is needed. Since the “Data set1” does not contain the regional information of users in the “Tree Hole”, we crawled the regional information of 2433333 microblog users by compiling the “Scrapy-Redis” distributed crawler, a sample of the results are shown in Table 2.

2.2. Data Processing

The degree of similarity compares the similarity of two things. After obtaining the message data (“Data set1”) and regional information of users in the “Tree Hole” (“Data set2”), we combined the data in “Data set1” and “Data set2” according to the keyword “User-ID” by writing and running a python program, and finally generated the “Data set3” with fields including

“Date-Time-User-ID-Message-Province”. At the same time, the message data without user geographic information was removed.

It was found that there were duplicate messages in “Data set3” through observation. Then the duplicate messages were removed by a running program written in Python. At the same time, messages of the user whose location is “overseas” were removed. A dataset which contains 696763 messages from 2012 to 2019 were obtained, as shown in the Table 3.

3. RESEARCH METHODS

3.1. Selection of Affective Lexicon

This paper uses the method based on affective lexicon to analyze the emotion of message text in the “Tree Hole”. Affective lexicon is an important resource of sentiment analysis [6]. At present, most of the research in the field of sentiment analysis is to classify sentiment into positive and negative categories, or add neutral categories [7]. There is little detailed classification of emotions, and the emotional intensity of texts is not considered. In this paper, the “Tree Hole” message texts are classified into fine-grained emotions, and they are specifically divided into six categories according to Ekman's theory, including joy, anger, sadness, fear, surprise and disgust [8].

In this experiment, we used the Chinese Emotion Word Ontology [9], which was labeled by Dalian University of Technology Information Retrieval to conduct the sentiment analysis. The Chinese Emotion Word Ontology is a Chinese ontology resource, which describes a Chinese vocabulary from the perspectives of speech type, emotional category, emotional intensity and polarity. There are 27466 Chinese words in the Chinese Emotion Word Ontology, and their emotions are divided into 7 categories and 21 sub categories. There are seven kinds of sentiments, including happiness, kindness, sadness, anger,

Table 1 | Data set1

Date	Time	Message-ID	User-ID	User-Name	Message	Other
2012/03/18	10:54	Null	1648943204	Jenny	Come back!	Null
2012/03/18	10:58	Null	1870301500	Rita	Good-bye	Null
...
2020/08/31	21:29	4544187703231624	5243886225	Mike	Are you inst?	Chenxiuji

Table 2 | Data set2

User-ID	User-Name	Province	City
1000349435	Cindy	Jiangsu Province	Nanjing
1001003271	Emma	Shanghai	Changning District
...
7498748940	Tom	Shandong Province	Null

Table 3 | Data set4

Date	Time	User-ID	User-Name	Message	Province
2019/05/02	21:35	2567122101	Rose	Fanfan, I don't want to take these medicines at all.	Sichuan Province
2019/05/02	21:36	2567122101	Rose	I want to crush all these medicines and throw them away.	Sichuan Province
2019/05/22	01:52	2567122101	Rose	Really, it's going to last.	Sichuan Province

Table 4 Examples of Chinese Emotion Word Ontology

CEW ID	Chinese Word	Word Class	Sense Number	Sense Order	Emotion Class	Strong Degree	Polarity
1000001	脏乱 (messy)	adj	1	1	NN	7	2
1000002	糟报 (bad news)	adj	1	1	NN	5	2
1027466	烛花 (snuff)	noun	1	1	PB	5	1

fear, disgust and surprise. Emotional intensity is divided into five levels, including 1, 3, 5, 7 and 9. Level 9 is the maximum intensity and 1 is the minimum intensity. In the experiment, we only used three fields of the ontology: ChineseWord, EmotionClass and StrongDegree (Table 4).

Ekman classifies emotions into six categories: happiness, sadness, anger, surprise, fear and disgust (Table 5). According to Ekman's way of emotional classification, we correspond the emotional categories of "happiness" and "kindness" to "joy" in Ekman emotional category, and the other emotional categories correspond to the other five emotional categories in Ekman one by one, as shown in the Table 6.

3.2. Construction of Negative Affective Lexicon

Since the Chinese Emotion Word Ontology does not contain any negative words, leading to the phenomenon that: only the word "happy" in the message can be recognized from "I am not happy", which will affect the accuracy of experimental results. This experiment chose to build a Chinese negative affective lexicon based on Chinese Emotion Word Ontology. The words in The Chinese Emotion Word Ontology were combined with 71 common negative words to form negative emotional words. Negative emotional words mean the phrase containing a negative prefix and a word in the affective lexicon. For example, "dirty" becomes "not dirty". At the same time, the emotional intensity value of negative emotional words is set as opposite number of the original emotional intensity value. Finally, a negative affective lexicon containing 1950086 negative emotional words was constructed.

3.3. Designing the SPARQL

In this experiment, message data and two affective lexicons were converted into RDF data and then loaded into a Knowledge Graph management platform, GraphDB. After that, The SPARQL language can be used to query the information we need. The emotional intensity of each province was obtained and counted by SPARQL query.

According to the structure of RDF data, the following SPARQL statements were designed. For example, query the emotional words contained in the "Tree Hole" messages in Shanxi Province in 2019 and the intensity of these emotional words:

```
PREFIX ztone: <http://www.ztonebv.nl/KG#>
select ?word ?degree
where {
?s ztone:EmotionClass ?emotion.
{FILTER regex(?emotion,"NE")} union {FILTER
regex(?emotion,"ND")} union {FILTER
regex(?emotion,"NN")} union {FILTER
regex(?emotion,"NK")} union {FILTER
regex(?emotion,"NL")}
```

Table 5 Emotion classification structure of Chinese Emotion Word Ontology

Number	Large Classification of Sentiment	Small Classification of Sentiment
1	Happiness	1. Happiness, 2. Relieved
2	Kindness	3. Respect, 4. Praise, 5. Believe, 6. Love, 7. Wish
3	Anger	8. Anger
4	Sadness	9. Sadness, 10. Disappointment, 11. Guilt, 12. Yearning
5	Fear	13. Panic, 14. Fear, 15. Shame
6	Disgust	16. Boredom, 17. Hatred, 18. Blame, 19. Jealousy, 20. Doubt
7	Surprise	21. Surprise

Table 6 Emotion classification system of Chinese Emotion Word Ontology

Number	Large Classification of Sentiment	Small Classification of Sentiment
1	Joy	1. Happiness, 2. Relieved, 3. Respect, 4. Praise, 5. Believe, 6. Love, 7. Wish
2	Anger	8. Anger
3	Sadness	9. Sadness, 10. Disappointment, 11. Guilt, 12. Yearning
4	Fear	13. Panic, 14. Fear, 15. Shame
5	Disgust	16. Boredom, 17. Hatred, 18. Blame, 19. Jealousy, 20. Doubt
6	Surprise	21. Surprise

```
?s ztone:ChineseWord ?word.
?s ztone:StrongDegree ?degree.
?s1 ztone:hasCity "山西".
?s1 ztone:hasDate ?date.
{FILTER regex(?date,"2019")}
?s1 ztone:hasMessage ?message.
FILTER (contains(?message, ?word))
}
```

By writing a python program, the emotional intensity of each province can be automatically queried and stored.

3.4. Emotion Scoring Algorithm

We have used the SPAQRL statement to find out all the messages, all the emotional words contained in these messages and the intensity values of these emotional words of the users in "Tree Hole" in 34 provinces of China in 2019. Finally, we

summed the intensity values of the emotional words, with the Equation (1) as the following.

$$A = \sum_{i=0}^n \text{strongdegree1} \quad (1)$$

Where n is the number of emotional words contained in all messages in the province in 2019, i is the variable used to count the number of emotional words and Strongdegree1 is the emotional intensity of emotional words.

We have used the SPAQRL statement to find out all the messages, all the negative emotional words contained in these messages and the intensity values of these negative emotional words of the users in "Tree Hole" in 34 provinces of China in 2019. Finally, we summed the intensity values of the negative emotional words, with the Equation (2) as the following.

$$B = \sum_{i=0}^n \text{strongdegree2} \quad (2)$$

Where n is the number of negative emotional words in all messages in this province, i is the variable used to count the number of negative emotional words and Strongdegree2 is the emotional intensity of negative emotional words.

$$E = (A + B) / C$$

Where E is the emotional intensity value of the province and C is the total number of messages in "Tree Holes" of the province in 2019.

4. EXPERIMENTAL RESULTS DISPLAY AND ANALYSIS

The emotional intensity of six emotional dimension, including joy, anger, sadness, fear, disgust and surprise in 34 provincial administrative regions in China, including 23 provinces, 5 autonomous regions and 4 municipalities directly under the Central Government and 2 special administrative regions are counted, and finally the geographical spatial distribution map of China of six emotional intensities are made. The darker the color of the area in the picture, the higher the value of emotional intensity of this kind of sentiment in the area. The experimental results show that there are differences in the emotional expression of message makers in different provinces of China in "Tree Hole".

- (1) It can be seen from the Figure 1 that the intensity of joy in Inner Mongolia Autonomous Region and Heilongjiang Province is higher than that in other provinces of China. Generally speaking, the emotional intensity of joy in

northeast of China is stronger than that in other provinces of China.

- (2) It can be seen from Figure 2 that the intensity of disgust in Jilin Province, Yunnan Province, Guangxi Province and Guizhou Province is higher than that in other provinces of China.
- (3) As can be seen from Figure 3, the intensity of sadness in Guangxi Province and Guizhou Province is higher than that in other provinces of China.
- (4) It can be seen from the Figure 4 that intensity of fear in Shaanxi Province and Gansu Province is higher than that in other provinces of China.
- (5) It can be seen from the Figure 5 that the intensity of surprise in Tibet, Qinghai Province, Jilin Province, Yunnan Province



Figure 2 | The distribution of emotional intensity of disgust in China's geographical region.



Figure 3 | The distribution of emotional intensity of sadness in China's geographical region.

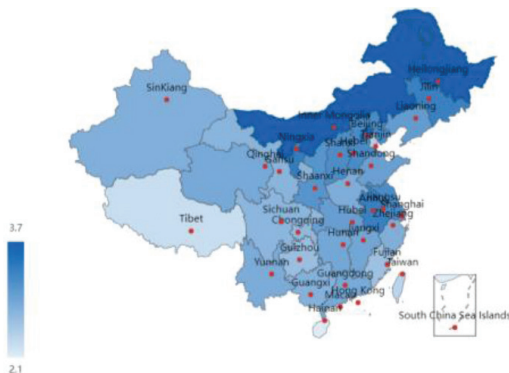


Figure 1 | The distribution of emotional intensity of joy in China's geographical region.



Figure 4 | The distribution of emotional intensity of fear in China's geographical region.

and Guizhou Province is higher than that in other provinces of China.

- (6) As shown in the Figure 6, the intensity of anger in Jilin Province, Ningxia Province, Tibet, Qinghai Province and Yunnan Province is higher than that in other provinces of China.

5. DISCUSSION AND SUGGESTION

5.1. Discussion

In this study, the geographical distribution of emotional intensity of users in “Tree Hole” in 2019 are visually analyzed. According to the experimental results, the following conclusions can be drawn:

- (1) The intensity of joy in northeast of China is stronger than that in other provinces of China. It can be inferred that has something to do with local economic condition. However, the reason why the areas such as Tibet where Level of economic development is also relatively low have low emotional intensity is unknown.
- (2) Negative emotions such as fear, surprise, sadness and anger are obviously higher in Ningxia Province, Tibet, Yunnan Province, Guizhou Province, Sichuan Province and Guangxi Province. It can be speculated that there may be the following reasons.
 - The average sunshine time in Sichuan Province and Yunnan Province is the shortest in China in 2019, and the humidity is also relatively high. Researches have shown that high temperature, high humidity, rainy weather and some abnormal weather events are not conducive to mental health of patients with depression [10].



Figure 5 | The distribution of emotional intensity of surprise in China's geographical region.



Figure 6 | The distribution of emotional intensity of anger in China's geographical region.

Therefore, it is reasonable to speculate that the concentration of negative emotions in these areas is partly related to their short illumination time, high humidity and heavy precipitation [11].

- A large number of negative emotions may be related to the vicious social events that occurred at that time. Studies have shown that the occurrence of major social events will affect public mood [12].

5.2. Suggestion

Based on the above analysis, the following suggestions are made: For economically underdeveloped areas, such as Southwest China, the mental health service system should be improved, increase mental health service resources, strengthen mental health education and popularize mental health-related knowledge [13].

The Tree Hole Rescue Team should strengthen the monitoring of users in “Tree Hole” whose location has high negative emotion values, including Sichuan Province, Yunnan Province, Guizhou Province, Jilin Province, Guangxi Zhuang Autonomous Region, Ningxia Hui Autonomous Region, etc. [14].

When a social vicious incident occurs in a certain region, the Tree Hole Rescue Team should strengthen the monitoring of local users and allocate more manpower for rescuing and monitoring.

6. CONCLUSION AND OUTLOOK

In this paper, the fine-grained sentiment analysis of “Tree Hole” message data is completed. But the sentiment analysis method relies on the sentiment dictionary, the accuracy of utilizing sentiment dictionary is not high enough. This paper only scores the emotional words contained in the “Tree Hole” message text. In fact, emoticons and modal particles in the messages can also express emotion. However, this part in the calculation of the emotional intensity of each province were not included in the calculation of the emotional intensity. And we will consider this in the follow-up work. The Chinese Emotional Vocabulary Ontology does not include all emotional words in life. In order to pursue better experimental results, we will continue to improve the emotional dictionary. This paper studies the distribution characteristics of the emotional expression of users in “Tree Hole” in China's geographic region, and the data is based on the location of users' personal information in micro-blog. The geographic information of the users in micro-blog that we crawled through a specific crawler is the personal information that users fill in the micro-blog, and there is the possibility that users fill in false geographic information, which leads to errors in the experimental results.

STATEMENTS AND DECLARATIONS

Conflict of Interest

All authors disclosed no relevant relationships.

Authors' Contribution

YY contributed in investigation and writing the original draft. SL contributed in guidance and reviewing the manuscript. ZH contributed in providing several figures and also reviewed the manuscript. YW visualized the project.

Acknowledgment

I thank Chaohui Guo for review and conceptualization.

REFERENCES

- [1] Hao Bai, Xiaoqing Ma, The incidence of depression in college students is increasing year by year, High incidence of freshmen and juniors, China Youth News. 2019.07. 29 (07).
- [2] Y.G. Huang, Y. Wang, H. Wang, Z. Liu, X. Yu, J. Yan, et al., Prevalence of mental disorders in China: a cross-sectional epidemiological study, *Lancet Psychiatry*. 6 (2019), 211–224.
- [3] X. Jing, S. Lin, Z. Huang, Extraction of characteristics of time in “Tree Hole” data, *J. Artif. Intell. Med. Sci.* 1 (2021), 43–48.
- [4] C. Guo, S. Lin, Z. Huang, C. Shi, Y. Yao, Research on the fluctuation characteristics of social media message sentiment with time before and during the COVID-19 epidemic, In: S. Siuly, H. Wang, L. Chen, Y. Guo, C. Xing (Eds.), *International Conference on Health Information Science*, Springer, Cham, 2021, pp. 3–14.
- [5] Gong Jingqiu, Lin Shaofu, Huang Zhisheng, Study on the spatial characteristics of depression patients' data on microblog “tree holes”, *China Digit. Med.* 15 (2020), 70–74.
- [6] Wang Shihong, Emotion analysis of Chinese microblogs using extended emotion lexicon, Nanjing Univ. Aeronaut. Astronaut. 2015.
- [7] Chen Xiaodong, Research on sentiment dictionary based emotional tendency analysis of Chinese microBlog, *Huazhong Univ. Sci. Technol.* 2012.
- [8] Guo Xianwei, The case of integrating emotional knowledge Weibo commented on the classification of emotions, *Chinese J. Comp.* 44 (2021), 564–578.
- [9] Chen Jianmei, The construction and application of Chinese emotion word ontology, Dalian Univ. Tech. China. 2009.
- [10] Wang Liguang, Liu Xiaodong, The influence of weather on mental diseases, *Jilin. Meteorological.* (2003), 9–12.
- [11] Kang Yanhai, Sheng Li, Influencing factors of suicide risk in elderly patients with first-episode depression, *Chinese J. Gerontol.* 42 (2022), 4.
- [12] Wang Yiju, Present situation and countermeasures of mental health education for left-behind children in rural areas, *Fundamental Educ. Forum.* (2022), 104–105.
- [13] Yan Penghao, Influencing factors and educational suggestions of left-behind children's personality development in rural areas, *Mental Health Education in Primary and Secondary School.* (2021), 12–14.
- [14] Zhao Duolu. Problems and solutions of mental health education for rural left-behind children, *Educator.* (2020), 18–19.