Advances in Educational Technology and Psychology (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/aetp.2022.060916
ISSN 2371-9400 Vol. 6 Num. 9

Identification and analysis of depression and suicidal tendency of Sina Weibo users based on machine learning

Lijing Sun[a], Yu Luo[b,*]

College of Psychology, Guizhou Normal University, Huaxi, Guiyang, Guizhou, 550025, China
[a]1074122560@qq.com, [b]yuluo@gznu.edu.cn
**Corresponding author*

Keywords: Sina Weibo users, Text, Depression, Suicidal ideation, Machine learning

Abstract: Recent years, with the development of big data, artificial intelligence, natural language processing and other technologies, the research on automatic mental health assessment driven by social network data has provided great convenience for the detection of depression and the suicidal tendency identification. In this study, machine learning and natural language processing technology had been adopted to identify depression of 203 Sina Weibo users. The recognition accuracy of 88.2% is achieved by using Gradient Boosting algorithm. Further, the suicidal tendency of 1204 Sina Weibo texts was identified and the Gradient Boosting algorithm was used to achieve an accuracy of 82.4% of Sina Weibo users with depression tendency. Then the suicidal ideation was analyzed by Beck Scale for Suicide Ideation (BSS) of Sina Weibo users with severe depression tendency. The results showed that most Sina Weibo users with severe depression tendency would hide their suicidal intention. It is also found that the word frequency related to suicidal ideation in the text of Sina Weibo users with severe depression tendency has no correlation with suicide intensity and suicide risk through statistical correlation analysis. The score of Beck Depression Inventory (BDI) of Sina Weibog users with severe depression tendency has a certain positive correlation with the score of suicidal ideation and suicide risk. The above identification and analysis of depression and suicidal tendency of Sina Weibog users will help quickly tap the depression and suicidal mood of users, so as to assist psychological workers and medical staff to carry out early warning intervention and avoid tragedy.

1. Introduction

Recent years, mental health problems have become one of the most serious and common public health problems in the world. According to the data provided by the World Health Organization, more than 350 million people around the world are affected by depression[1], 3000 patients with depression commit suicide every day. Thus, suicide caused by psychological problems is one of the three main factors of young people's death[2]. The above data shows that the current mental health problem has become an important problem affecting social development and personal health. However, the public's understanding of depression is not clear enough at present.

Depression is characterized by significant and lasting black mood. Its external performance is no different from that of ordinary people, so that many people who have suffered from depression do

not know they are ill. Thus the resulting deterioration of their condition and unhealthy psychological state for a long time by depression have a great impact on the life of patients. Therefore, earlier detection and intervention of depression is of great significance to improve personal quality of life and promote good social mentality. In recent years, with the development of big data, artificial intelligence, natural language processing and other technologies, more and more scholars begin to use data science and technology to explore the psychological and behavioral mechanisms of individuals or groups.Further, the automatic detection and analysis by the social network data generated by users could help identify their the mental health status and to achieve the purpose of automatic intervention and early warning.

In terms of automatic detection of depression, the language and behavior characteristics of Internet users was used to detect depression through classification and regression model[3], a new method was further proposed to detect depression through network behavior time-frequency analysis[4]. The research showed that users' depression could be predicted through social network data. In addition to the automatic detection of depression, many scholars have carried out early warning research on network suicide tendency aiming at the suicide risk caused by depression. Pete et al[5] classified social media texts, identified suicidal ideation of the social text, and analyzed the precursory characteristics of suicidal ideation to help explain the language used by social media users who perceive suicide. Deep neural network was also used to construct a personalized knowledge map for suicide in the study[6].The constructed personalized knowledge map was utilized to determine the key risk factors of personal suicidal ideation. The results showed that the accuracy of suicidal ideation detection based on social media could reach more than 93%.

The above research on depression and suicide risk has achieved great progress, but less research on suicidal tendency of users with depression has been carried out. At the same time, the earlier research has not integrated the depression of Sina Weibo users and the suicide risk caused by depression, and there is a lack of in-depth analysis on the internal logic and relationship between depression and its derived suicidal tendency. Therefore, this study aims at Sina Weibo users as the research object. Firstly, Sina Weibo users are detected and identified with depression tendency through machine learning method. Then each text of these users with depression tendency is taken as the object to identify the suicidal tendency of their text.Finally, the suicidal ideation of these users with severe depression tendency is analyzed so as to provide a research basis for further risk early warning and active intervention.

2. Automatic identification of Sina Weibo users with depression tendency

2.1. Data acquisition and preprocessing

This study selects Sina Weibo as the data collection platform. In terms of Sina Weibo text data tagging, the depression score of Sina Weibo users was obtained through the depression scale as the basis of tagging considering the problems that manual tagging embodied with the shortcomings of strong subjectivity, inaccuracy and large personal differences. In the process of data collection, Beck Depression Inventory(BDI)[7] and Self-rating depression scale(SDS)[8] were randomly distributed to a large number of Sina Weibo users to obtain the scale score of each user. According to the scale score, Sina Weibo users were divided into normal users and users with depression tendency. Sina Weibo users' texts are collected using Descendant collector with the knowledge and authorization of Sina Weibo users. In this study, texts with 102 users of depression tendency and 101 normal users are collected.

Sina Weibo text was preprocessed in python to remove invisible characters, redundant spaces, punctuation and other invalid contents, as well as contents containing regular expressions and emoticons. Jieba was used for Chinese word segmentation of Sina Weibo text, and the processing of

stop words was carried out on this basis.

2.2. Recognition of Sina Weibo users' depression tendency

In terms of Sina Weibo texts' feature extraction, term frequency-inverse document frequency (TF-IDF)[9] was taken as the tool of text feature transformation in the early stage. Numerous classifiers such as Neural network, Naive Bayes[10], Random forest[11], logistic regression, Adlboost[12] and Gradient Boosting[13] were studied,and their classification effects were compared after fully adjusting the parameters. During the construction of machine learning model, the input was the preprocessed users' text, and the output was whether the user was with depression tendency or not, which was a binary classification problem.

Table 1: Comparison of depression tendency recognition of Sina Weibo users with different machine learning models.

	Accuracy	Precision	Recall	F1 score
Neural network	78.40%	78.30%	78.40%	78.30%
Naive Bayes	73.50%	75.10%	73.50%	73.70%
Random forest	78.40%	79.30%	78.40%	78.60%
Adlboost	79.40%	79.30%	79.40%	79.40%
logistic regression	87.30%	88.80%	87.30%	87.40%
Gradient Boosting	88.20%	90.80%	88.20%	88.30%

In order to test the performance of different algorithms, in addition to the four indicators (accuracy, accuracy, recall and F1 score), simple cross validation(75% of the data is divided into training sets and 25% of the data is divided into test sets) was used to compare different algorithms, as shown in Table 1. As could be seen from table 1, the performance of Gradient Boosting is higher than that of other algorithms. The algorithm implementation process was completed by using Python machine learning package scikit learn.

Table 2: Comparison of accuracy of various algorithms under different cross validation methods.

	ANN	Naive Bayes	Random Forest	Adlboost	Logistic regression	Gradient Boosting
2-fold	75.4%	60.6%	71.9%	77.3%	49.8%	83.7%
3-fold	74.9%	58.6%	68.0%	76.8%	82.3%	84.2%
5-fold	79.8%	70.4%	72.9%	75.4%	84.7%	83.7%
10-fold	78.8%	70.0%	74.9%	73.4%	84.7%	84.2%
20-fold	77.3%	69.5%	69.0%	74.4%	84.7%	83.7%
75%-25%	76.5%	75.5%	80.4%	76.5%	87.3%	88.2%

In order to analyze whether different algorithms had statistically significant differences, simple cross validation , 2-fold cross validation, 3-fold cross validation, 5-fold cross validation, 10 fold cross validation and 20 fold cross validation were used to analyze the recognition accuracy of different algorithms. The results are shown in Table 2.

One way ANOVA was used to statistically test the accuracy of different cross validation methods for different algorithms in Table 3. It could be seen from table 3 that there is a significant difference in the accuracy of different algorithms using different cross validation methods, and different algorithms have a significant impact on the recognition accuracy.

Table 3: Results of one-way ANOVA.

	ANN	Naive Bayes	RF	Adlboost	LR	Gradient Boosting	*F*	*p*
Accuracy	77.12±1.91	67.43±6.47	72.85±4.48	75.63±1.51	78.92±14.35	84.62±8.26	4.350	0.004**

$*p < 0.05, **p < 0.01$

After the above analysis, it could be conclude that Gradient Boosting algorithm achieved the best classification accuracy of 88.2%. The confusion matrix of Gradient Boosting algorithm is shown in Figure 1. It could been seen from Fig 1 that among all 102 text data, 12 text data of subjects with depression tendency were incorrectly identified as text data of normal subjects.

		Predicted		
		Normal subjects	Subjects with depression tendency	Σ
Actual	Normal subjects	42	0	42
	Subjects with depression tendency	12	48	60
	Σ	54	48	102

Figure 1: Confusion matrix of Gradient Boosting algorithm.

3. Identification of suicidal tendency of Sina Weibo users with depression tendency

In addition to the recognition of depression, it is also of great significance to study the suicidal tendencies of Sina Weibo depression users for depression is often accompanied by significant and lasting depression, even pessimism and suicide attempts or behaviors. Thus, based on the previous analyse, this study further carried out the research on suicide tendency by using natural language processing technology and machine learning methods.

3.1. Data preprocessing

Table 4: Classification of suicide tendency of Sina Weibo text.

Text label type	Explanation	Suicidal tendency label
Strongly concerning	Convincingly showing serious suicidal thoughts or a desire to commit suicide completely, such as "I want to die" or "I want to commit suicide"; Disclose suicide plans and / or previous attempts, such as "I want to commit suicide", "if this happens, I will commit suicide", etc	2
Possibly concerning	May involve suicidal thoughts, but not very strong, no implementation plan, etc.	1
Safe to ignore	No reasonable evidence that there is a risk of suicide.	0

For the users with depression tendency collected previously, this study selected each text as the research object. A total of 1204 Sina Weibo texts were finally selected with Sina Weibo users' knowledge and authorization. As for the suicide tendency label of Sina Weibo text, according to the research of Bridianne et al (2015)[14], each Sina Weibo users with depression tendency was classified into suicide tendency grades, as shown in Table 4. Among all 1204 Sina Weibo texts, 933 of which were labeled as 0, 171 of which were labeled as 1 and 100 of which were labeled as 2.

Sina Weibo text was preprocessed in python to remove invisible characters, redundant spaces, punctuation and other invalid contents, as well as contents containing regular expressions and

emoticons. Jieba was used for Chinese word segmentation of Sina Weibo text, and the processing of stop words is carried out on this basis.

3.2. Suicidal tendency identification

Considering that there are many machine learning classifiers, and the effect of classifiers is different in different situations. Therefore, this study compared many classifiers such as neural network, k-nearest neighbor algorithm[15], random forest, logistic regression, adlboost and Gradient Boosting. In the process of building the machine learning model, the input was each Sina Weibo text of the pretreated Sina Weibo users with depression tendency, and the output is the Sina Weibo text suicide tendency grade, which is a three category problem.

Table 5: Comparison of identification of suicidal tendencies of Sina Weibos with depression based on different machine learning models.

	Accuracy	Precision	Recall	F1 score
ANN	77.90%	74.40%	77.90%	75.70%
Random Forest	80.10%	73.90%	80.10%	74.40%
KNN	80.60%	77.40%	80.60%	75.10%
Adlboost	75.10%	75.50%	75.10%	75.30%
Logistic regression	82.20%	79.30%	82.20%	78.20%
Gradient Boosting	82.40%	80.70%	82.40%	78.90%

In order to test the performance of different algorithms, in addition to four indicators (accuracy, accuracy, recall and F1 score), simple cross validation was used to compare different algorithms after fully adjusting the parameters, as shown in table 5. It could be seen from the table that Gradient Boosting was the highest than other algorithms in four indicators. The algorithm implementation process was completed by using Python machine learning package scikit learn.

Table 6: Comparison of cross validation accuracy of different algorithms.

	ANN	RF	KNN	Adlboost	LR	Gradient Boosting
2-fold	77.60%	77.70%	77.90%	71.70%	78.70%	78.70%
3-fold	77.20%	77.20%	78.20%	69.30%	78.70%	78.40%
5-fold	78.00%	77.70%	77.60%	70.10%	79.40%	79.70%
10-fold	78.30%	78.20%	78.20%	71.70%	79.80%	80.10%
20-fold	77.80%	78.20%	78.30%	70.30%	80.10%	80.20%
75%-25%	77.90%	80.10%	80.60%	75.40%	82.20%	82.40%

In order to analyze whether different algorithms have statistically significant differences, simple cross validation, 2-fold cross validation, 3-fold cross validation, 5-fold cross validation, 10 fold cross validation and 20 fold cross validation were used to compare the accuracy of different algorithms, as shown in table 6.

Table 7: Results of one-way ANOVA.

	ANN	RF	KNN	Adlboost	LR	Gradient Boosting	F	p
Accuracy	77.80 ±0.37	78.18 ±1.01	78.47 ±1.08	71.42 ±2.17	79.82 ±1.30	79.92 ±3.17	33.327	0.000 **

$*p < 0.05, ** p < 0.01$

The accuracy of different algorithms was statistically tested in combination with one-way ANOVA, as shown in table 7. It could be seen from table 7 that there is a significant difference

between different algorithms. Different algorithms have a significant impact on the recognition accuracy, and the average accuracy of Gradient Boosting algorithm is higher than that of other algorithms.

		Predicted			
		labelled as 2	labelled as 1	labelled as 0	∑
Actual	labelled as 2	16	2	27	45
	labelled as 1	5	15	59	79
	labelled as 0	10	3	465	478
	∑	31	20	551	602

Figure 2: Confusion matrix of Gradient Boosting algorithm.

The confusion matrix of Gradient Boosting algorithm is shown in Figure 2. According to the confusion matrix of Gradient Boosting algorithm, of all 478 Sina Weibo texts with 0 Tags, 10 were incorrectly identified as tag 2 and 3 texts were incorrectly identified as tag 1; Among all 45 Sina Weibo texts with tag 2, 27 texts were incorrectly identified as tag 0, and 2 texts were incorrectly identified as tag 1; Among all 79 Sina Weibo texts with tag 1, 59 texts were incorrectly identified as tag 0, and 5 texts were incorrectly identified as tag 2.

3.3. Analysis of suicidal ideation

Based on the above research on suicidal tendency, this study further explored the suicidal ideation of Sina Weibo users. After the construction of suicidal tendency index, the data of suicidal ideation was collected by issuing the Chinese version of Beck Scale for Suicide Ideation (BSS) to Sina Weibo users who suffered with severe depressive tendency.With the knowledge and authorization of Sina Weibo users, 43 valid questionnaires were collected, including 41 users with suicidal ideation, accounting for 95.35%. It can be seen that the vast majority of Sina Weibo users with severe depression had suicidal ideation, which was worthy of vigilance.

Due to the lack of data of few users, some relevant data of suicidal ideation of 38 Sina Weibo users with severe depression tendency was obtained. It was found that among the total 38 Sina Weibo users with severe depression tendency as well as suicidal ideation, 20 users had suicidal ideation related words in their Sina Weibo text, accounting for 52.62%. And about "do you let people know your suicidal thoughts?" ,Only 8% of users would frankly and actively say their thoughts of suicide in the recent week, while 22% of users would frankly and actively say their thoughts of suicide in the most depressed and melancholy time according to BSS(see Figure 3). This meant that although Sina Weibo users with severe depression tendency with suicidal ideation would not frankly and actively say their suicidal thoughts, some of them would express their suicidal thoughts hidden in real life at the platform of Sina Weibo. This was the research significance of using social networks to identify users' suicidal thoughts. On the other hand, although there were words related to suicidal ideation in the Sina Weibo text of some users with severe depression tendency, the outcome of BSS showed that they had no suicidal ideation. This might due to that the emergence of these suicidal ideation words was only the current suicidal ideation of users, not enough to form suicidal ideation.

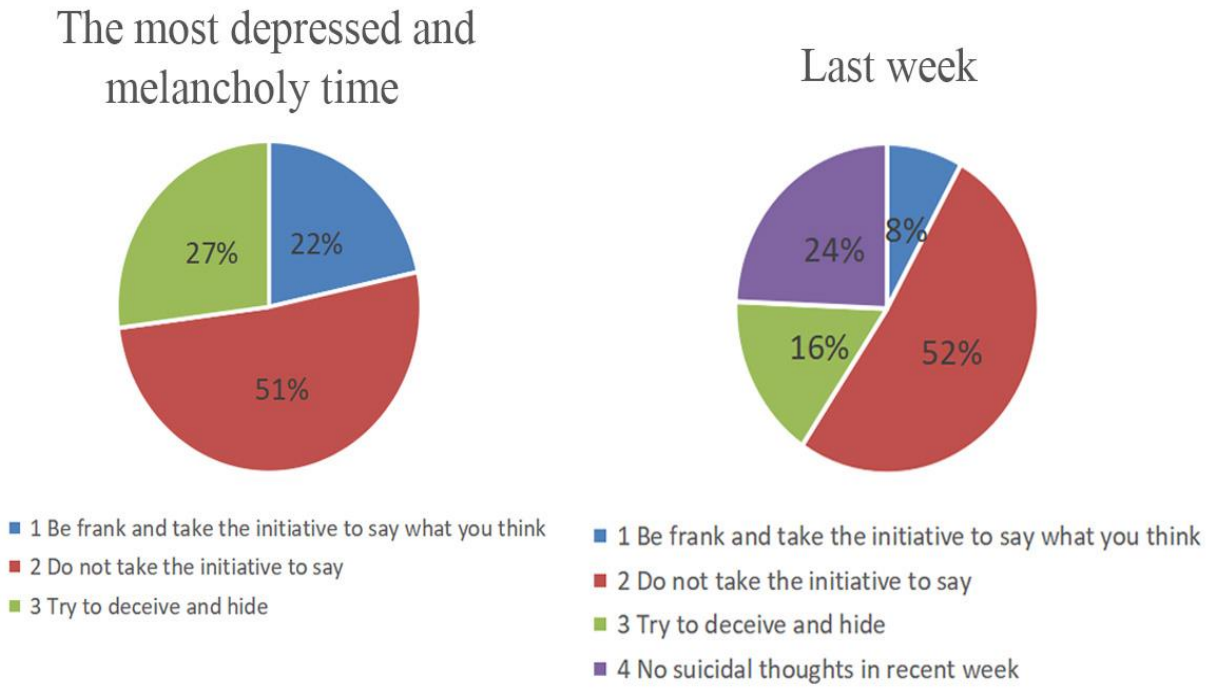

Figure 3: Option results of the question"do you let people know your suicidal thoughts?"

The correlation analysis was carried out on the word frequency and suicide intensity related to suicidal ideation in the texts of Sina Weibo users with severe depression tendency. Spearman coefficient was used to explore the correlation between the two factors. The analysis results are shown in Table 8 that the bilateral significance of the two factors was 0.389, greater than 0.05. So there was no correlation between the word frequency and suicide intensity.

Table 8: The correlation analysis between word frequency and suicide intensity related to suicidal ideation.

		Suicidal ideation intensity	Suicidal ideation word frequency
Suicidal ideation intensity	correlation coefficient	1.000	-0.146
	Sig. (bilateral)		0.389
	N	37	37
Suicidal ideation word frequency	correlation coefficient	-0.146	1.000
	Sig. (bilateral)	0.389	
	N	37	37

The correlation analysis was carried out on the word frequency and suicide risk related to suicidal ideation in the texts of Sina Weibo users with severe depression tendency. The analysis results are shown in table 9. Spearman coefficient was used to explore the correlation between the two factors. The analysis results show that the bilateral significance of the two factors was 0.375, greater than 0.05, which meant that there was no correlation between the word frequency and suicide risk.

Table 9: The correlation analysis between word frequency and suicide risk related to suicidal ideation.

		Suicidal ideation word frequency	Suicide risk
Suicidal ideation word frequency	correlation coefficient	1.000	-0.150
	Sig.(bilateral)		0.375
	N	37	37
Suicide risk	correlation coefficient	-0.150	1.000
	Sig.(bilateral)	0.375	
	N	37	37

The correlation between suicidal ideation intensity and the score of Beck depression questionnaire was analyzed using Pearson correlation coefficient. The result shows that the correlation coefficient of the two factors is 0.351, which was significantly correlated at the level of 0.05 (bilateral), as shown in Table 10. This meant that the depression tendency of Sina Weibo users with severe depression tendency had a certain positive correlation with their suicidal ideation.

Table 10: The correlation analysis between suicide ideation intensity and Beck Depression Questionnaire score.

		Suicidal ideation intensity	Beck Depression Questionnaire score
Suicidal ideation intensity	Pearson correlation coefficient	1	0.351*
	Sig.(bilateral)		0.033
	N	37	37
Beck Depression Questionnaire score	Pearson correlation coefficient	0.351*	1.000
	Sig.(bilateral)	0.033	
	N	37	37

$*p < 0.05, **p < 0.01$

The correlation between suicide risk and Beck Depression Questionnaire score is analyzed by Pearson correlation coefficient. The data showed that the correlation coefficient of the two factors was 0.429, which was significantly correlated at the level of 0.01 (bilateral), as shown in Table 11. This meant that the depression tendency of Sina Weibo users with severe depression tendency had a certain positive correlation with their suicidal ideation.

Table 11: The correlation analysis between suicide risk and Beck Depression Questionnaire scores.

		Suicide risk	Beck Depression Questionnaire score
Suicide risk	Pearson correlation coefficient	1	0.429**
	Sig.(bilateral)		0.008
	N	37	37
Beck Depression Questionnaire score	Pearson correlation coefficient	0.429**	1
	Sig.(bilateral)	0.008	
	N	37	37

$*p < 0.05, **p < 0.01$

4. Discussion

In this study, Sina Weibo users were selected as the research object, and 203 Sina Weibo users were identified with depression emotion through data mining, natural language processing technology and machine learning algorithm. Gradient Boosting algorithm was adopted to achieve the recognition accuracy of 88.2%. Furthermore, for Sina Weibo users with depression tendency, the suicidal tendency of 1204 Sina Weibo texts was identified, and the Gradient Boosting algorithm was used to achieve an accuracy of 82.4%. Several different machine learning algorithms were compared. On the basis of fully adjusting parameters, it was found that Gradient Boosting algorithm had the best recognition effect, because it could integrate the advantages of multiple algorithms as an integrated algorithm. At the same time,the reason why Gradient Boosting algorithm achieved the best classification accuracy among many machine learning algorithms was that it had good generalization ability and representation ability on densely distributed data, while the Sina Weibo text data in this study had a large number of similar synonyms.

In the research of suicidal tendencies and suicidal ideation of depressed Sina Weibo users, this study first explored the suicidal tendencies of depressed Sina Weibo users. Through the earlier data preparation, machine learning had achieved a relatively ideal recognition rate, that was, the text styles of depressed Sina Weibo users with different degrees of suicidal tendencies were different, and their suicidal tendencies could be inferred from the text. Based on the study of suicidal tendency, this study further conducted the suicidal ideation of Sina Weibo users with severe depressive tendency through the BSS and the word frequency related to suicidal ideation in the text. It was found that the word frequency of suicidal ideation of Sina Weibo users with severe depressive tendency was not significantly correlated with suicidal ideation and suicide risk, which meant it was not feasible to infer their suicidal ideation and suicide risk by word frequency only.

In order to further explore the reason why the word frequency of suicidal ideation of Sina Weibo users with severe depression tendency was significantly irrelevant to their suicidal intensity and risk, it was found that most users tend to hide their suicidal intention, whether in real life or in social networks during the in-depth analysis of BSS data. This outcome also indirectly confirmed the characteristics of patients with depression that almost of them were more inclined to close themselves, falling into their own world and couldn't extricate themselves. Therefore, more methods should be used to explore users' suicidal ideation by future research. In this regard, considering that most users would not express depression and suicide content, some researchers used social network nodes to mine depressed users. Researchers[16] regarded patients with depression as a node, built a graph network with this node as the center, and gave a model to calculate the depression status according to the attributes and connection weights of adjacent nodes in the network.

The research of mental health assessment based on social network also involves many problems, such as data privacy, information disclosure and so on. Data privacy is a continuing concern that once users are labeled with psychological problems, they may be discriminated against or ridiculed. Therefore, data protection and ownership framework agreements are needed to ensure that users will not be hurt.

5. Conclusion

Based on the identification and analysis of depression and suicidal tendency of Sina Weibo users based on machine learning, this study drew the following conclusions by combining the knowledge of psychology, data mining, natural language processing, machine learning and statistics. (1) Aiming at whether Sina Weibo users had depression tendency, the research labeled Sina Weibo users through depression questionnaire, and excavated the Sina Weibo text data of users. After preliminary data preprocessing and feature transformation, comparing different algorithms and fully

adjusting parameters, Gradient Boosting algorithm was used to identify the depression of Sina Weibo users, with an accuracy of 88.2%. (2) This study aimed at the suicidal tendency of Sina Weibo users with depression. Each Sina Weibo text was taking as the research object,thus three-level suicidal tendency was labeled. After preliminary data preprocessing and feature transformation, different algorithms were compared. Then Gradient Boosting algorithm was used to detect suicidal tendency, and the recognition accuracy reached 82.4% after fully adjusting its parameters. (3) There was no significant correlation between the word frequency related to suicidal ideation and suicide intensity as well as suicide risk in the text of Sina Weibo users with severe depression tendency. There was a significant positive correlation between the score of BDI and the score of suicidal ideation as well as suicide risk. Most Sina Weibo users with severe depression tendency will hide their suicidal intention.

References

[1] Peng Z H Q, Dang J. Multi-kernel SVM based depression recognition using social data [J]. International Journal of Machine Learning and Cybernetics, 2019, (1)(10):43-57.

[2] Liang X G S D J. Investigation of College Students' Mental Health Status via Semantic Analysis of Sina Sina Weibo [J]. Wuhan University Journal of Natureal Sciences, 2015, 2(20):159-164.

[3] Hu Q, Li A, Heng F, et al. Predicting Depression of Social Media User on Different Observation Windows [J]. 2015.

[4] Zhu C, Li B, Li A, et al. Predicting Depression from Internet Behaviors by Time-Frequency Features: IEEE/WIC/ACM International Conference on Web Intelligence, 2017 [C].

[5] Pete, Burnap, Gualtiero, et al. Multi-class machine classification of suicide-related communication on Twitter [J]. Online Social Networks & Media, 2017.

[6] Cao L, Zhang H, Feng L. Building and Using Personal Knowledge Graph to Improve Suicidal Ideation Detection on Social Media [J]. IEEE Transactions on Multimedia, 2020.

[7] T B A. Depression: Cause and treatment [M]. Philadephia: University of Pennsylyania Press, 1967.

[8] Shaver P R, Brennan K A. Measures of Depression and Loneliness [J]. Measures of Personality and Social Psychological Attitudes, 1991.

[9] Brauen T L. The SMART Retrieval System: Experiments in Automatic Document Processing [J]. 1971.

[10] Bermejo P, Gamez J A, Puerta J M. Speeding up incremental wrapper feature subset selection with Naive Bayes Classifier [J]. Knowledge-Based Systems, 2014, 55(jan.): 140-147.

[11] Breiman. Random forests [J]. MACH LEARN, 2001, 45(1) (-):5-32.

[12] Freund Y. Experiments with a new boosting algorithm [J]. icml, 1996.

[13] Friedman J H. Greedy Function Approximation: A Gradient Boosting Machine [J]. Annals of Statistics, 2001, 29(5):1189-1232.

[14] Bridianne, O'Dea, Stephen, et al. Detecting suicidality on Twitter [J]. Internet Interventions, 2015.

[15] Cover T, Hart P. Nearest neighbor pattern classification [J]. IEEE Transactions on Information Theory, 2003, 13 (1):21-27.

[16] Wang X, Zhang C, Ji Y, et al. A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network [J]. Lecture Notes in Computer Science, 2013.