Suicidal Ideation Detection via Social Media Analytics

Yan Huang[1,2], Xiaoqian Liu[1], and Tingshao Zhu[1(✉)]

[1] Institute of Psychology, Chinese Academy of Sciences, Beijing, China
hihihey@qq.com, tszhu@psych.ac.cn
[2] University of Chinese Academy of Sciences, Beijing, China

Abstract. Suicide is one of the increasingly serious public health problems in modern society. Traditional suicidal ideation detection using questionnaires or patients' self-report about their feelings and experiences is normally considered insufficient, passive, and untimely. With the advancement of Internet technology, social networking platforms are becoming increasingly popular. In this paper, we propose a suicidal ideation detection method based on multi-feature weighted fusion. We extracted linguistic features set that related to suicide by three different dictionaries, which are data-driven dictionary, Chinese suicide dictionary, and Language Inquiry and Word Count (LIWC). Two machine learning algorithms are utilized to build weak classification model with these three feature sets separately to generate six detection results. And after logistic regression, to get the final weighted results. In such a scheme, the results of model evaluation reveal that the proposed detection method achieves significantly better performance than that use existing feature selection methods.

Keywords: Suicidal ideation · Detection · Social media

1 Introduction

1.1 Suicidal Ideation and Social Media

Traditionally, mental health professionals usually rely on mental status examination through specialized questionnaires and the Suicidal Behaviors Questionnaire Revised (SBQ-R) (Osman et al. 2001) or patients' self-reported feelings and experiences to detect suicidal ideation, which have been considered insufficient, passive, and untimely (McCarthy 2010; Choudhury et al. 2013). In fact, patients with mental diseases do not often go to health care providers. Given the high prevalence and devastating consequences of suicide, it is critical to explore proactive and cost-effective methods for real-time detection of suicidal ideation to perform timely intervention.

In 2012, a 22-year-old girl nicknamed Zoufan left her last words on Weibo that she was suffering from depression and decided to die. She hanged herself in the dormitory, the police and medical staff could not save her life even though they did their best to

Y. Huang and X. Liu—These authors contributed equally to this work.

D. Milošević et al. (Eds.): HCC 2019, LNCS 11956, pp. 166–174, 2019.
https://doi.org/10.1007/978-3-030-37429-7_17

rescue. Netizens were generally saddened and regretted about this news, and more and more people are calling for greater attention to people with depression and preventing them from being suicidal.

Previous studies have found that suicidal ideation of a person can be discovered from his/her communication through analysis of his/her self-expression (Stirman and Pennebaker 2001; Li et al. 2014). With the advancement of Internet technology, more and more people share their lives, experiences, views, and emotions in those online venues (Kaplan and Haenlein 2010; Robinson et al. 2015). There has been an increase of research on the effect of social media use on mental health (Pantic 2014). Research has shown that people with mental health problems, especially those with depression, have high social media use (Davila et al. 2012; Rosen et al. 2013). Analyzing people's behavior in social media may provide a genuine, proactive, and cost-effective way to discover the intention or risk of an individual to commit suicide.

Wang et al. (2013) developed a depression detection model with characteristics of users' language and behavior as input features. 180 users' social media data were used to verify the models. These results demonstrate that it is feasible to detect individuals' suicidal ideation through social media.

1.2 Related Work

In the past decade, there has been increasing efforts toward automatic identification of suicidal ideation from social media data (O'Dea et al. 2015; Paul and Dredze 2014; Wang et al. 2015). Suicidal ideation identification is a binary classification problem – to determine if an individual has a suicidal ideation or not. As a result, the majority of studies in this field use a classification machine learning approach to build suicidal ideation detection models, in which feature selection is critical. Existing models use different predictive features, including LIWC-based (Pestian et al. 2012; Li et al. 2014), dictionary-based (Lv et al. 2015), and social media behavior based (e.g., the total number of followers, the number of blogs published after midnight) (Guan et al. 2015). Each has its advantages and disadvantages. The knowledge-experience-driven approach is gradually accumulated in practice, and the data-driven approach is derived from factual analysis under a large number of data records. The former is more targeted and flexible than the latter, but experience is easily limited by subjective judgment and data insight is more objective. However, there is a lack of studies that investigate which feature selection method might be more effective to select predictive features.

This study is aimed to propose a new suicidal ideation detection method based on multi-feature weighted fusion (Called MFWF method hereafter), to verify if this new method provides better model performance than existing suicidal prediction model. If this assumption is true, we will get higher accuracy of suicidal ideation. Based on the method of big data analysis, we can automatically identify the blogger's suicidal ideation by analyzing their blog posts, which can also be used to assist manual screening, and ultimately improve the efficiency of mental health services.

2 Methods

In this research, we aim to advance knowledge about correlations between linguistic features of textual social media content and suicidal ideation, and provide an optimized weighting method that combines multiple features and logistic regression to detect suicidal ideation. The proposed MFWF method works as follows.

First and foremost, collect the data from blog posts with suicidal ideation. Then linguistic features are extracted through three ways, including data-driven dictionary that automatically generated by n-gram (Liu et al. 2019), Chinese suicide dictionary, and Language Inquiry and Word Count (LIWC) (Pennebaker et al. 2001). Two machine learning algorithms are utilized to build weak classification model with these three features sets separately to generate six detection results. Input them into logistic regression model to get different parameters that corresponded to weighting results of different linguistic feature. Finally, get the judgment on whether there is suicidal ideation (Fig. 1).

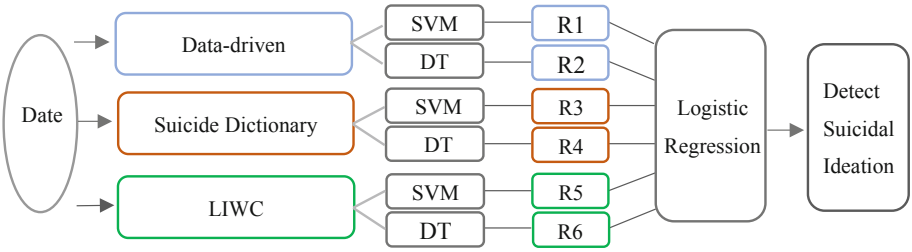

Fig. 1. The multi-feature weighted fusion method

2.1 Data Collection

The data source for this study is from Sina Weibo, China's largest open social media platform, with current monthly active users reaching 465 million.

After Zoufan's death, people still paid attention and left messages below her last blog, and some of the messages revealed suicidal thoughts. Our research team crawled 65,352 messages from there through Sina Weibo API. They were posted by 23,000 bloggers. Considering that extremely short blogs may contain little useful information, we removed the blogs containing fewer than 5 words (excluding punctuation and Arabic numerals). The average length of the remaining blogs was 29.63 words.

We established a team of three experts who specialized in psychology and suicidal behavior, to assess whether there is suicidal ideation and manually labelled each of the collected blogs. In order to ensure the accuracy of the results, each expert was trained before the assessment, and the consistency of the assessment results was verified on a small sample. Then multiple rounds of assessment began, and the label would be confirmed only if the three experts make a consistent judgment, otherwise re-determine.

The labeling results of individual coders were examined for consistency. Finally, we had 8,548 blogs labeled as "suicide" (i.e., with suicidal ideation). To avoid the

common data imbalance problem in machine learning, we randomly selected 10,000 blogs from the ones labelled as non-suicide as negative training samples.

2.2 Feature Extraction

We used three kinds of dictionaries for linguistic feature extraction, including (1) automated machine learning dictionary driven by N-gram. N-gram is an algorithm based on statistical language model. Its basic idea is to operate the contents of the text in a sliding window of size N in bytes to form a sequence of byte segments of length N. Each byte segment is called gram, and the frequency of occurrence of all grams is counted and filtered according to a preset threshold to form a key gram list, which is the vector feature space of the text. N-gram is based on the assumption that the occurrence of the nth word is related to the first n-1 words, and is not related to any other words. The probability of each word can be calculated by statistical calculations in the corpus. The probability that the entire sentence appears is equal to the probability product of the occurrence of each word. A 3-gram algorithm was utilized to process blog content and identify one-character, two-character, and three-character terms in the blog corpus. We removed stop words and considered the remaining terms as candidate terms. (2) Chinese suicide dictionary, a suicide-related lexicon that developed by some Chinese experts, it was based on knowledge and experience. (3) SCMBWC, the Simplified Chinese Micro-Blog Word Count tool (Gao et al. 2013), it is a Chinese version of the commonly used Language Inquiry and Word Count (LIWC) (Pennebaker et al. 2001), a classic tool for sentiment analysis. The set of selected terms using the above three kinds of dictionaries were called feature set A, B and C.

To constitute feature set A, a Chi-square test was then performed on the candidate terms to identify those that are mostly related to the blogs labelled as suicide. Chi-square test is a widely used hypothesis testing method that examines whether two categorical variables are independent of each other. In order to find terms closely related to suicide, we first propose a null hypothesis that each candidate term is independent of the "suicide" class. The Chi-square value of a candidate term is calculated by Eq. (1), P value <0.05.

$$X2_t = \frac{N(A_tD_t - B_tC_t)^2}{(A_t + C_t)(A_t + B_t)(B_t + D_t)(B_t + C_t)} \tag{1}$$

Where N denotes the total number of blogs in the corpus. For each term t, A_i is the number of blogs that contain t and are labeled as "suicide"; B_t is the number of blogs that contain t and are labeled as "non-suicide"; C_t is the number of blogs that do not contain t and are labeled as "suicide"; and D_t is the number of blogs that do not contain t and are labeled as "non-suicide". A larger Chi square value $X2_t$ implies that the term t is more related to suicidal ideation.

After calculating the Chi-square value of each candidate term, we selected the top 2,000 terms with the highest Chi-square values to constitute feature set A, such as "negative energy", "miserable", and "rubbish", as the predictive features for our proposed suicidal ideation detection model. Quite some of those feature terms do not exist

in any existing emotion- or suicide-related dictionaries, which shows that the proposed feature selection method can identify new suicide-related terms from social media content, which may help with the expansion of existing dictionaries.

We use Term Frequency–Inverse Document Frequency (TF-IDF) (Salton and McGill 1986) for feature extraction, which is a statistical method used to assess the importance of a word for a file set or one of the files in a corpus. The importance of a word increases proportionally with the number of times it appears in the file, but it also decreases inversely with the frequency it appears in the corpus. The main idea of using TF-IDF is that if a word or phrase appears in a class of blog posts with a high frequency TF and rarely appears in other categories, the word or phrase is considered to be suitable for classification.

Each blog f_j, after word segmentation, would be represented as a feature vector with 2,000 dimensions corresponding to 2,000 feature terms. The TF-IDF value of each feature term s_i was calculated as the feature value by Eqs. (2)–(4).

$$\text{tfidf}_{i,j} = tf_{i,j} \times idf_i \tag{2}$$

$$\text{tf}_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{3}$$

$$\text{idf}_i = \log \frac{|D|}{|j : s_i \in d_i|} \tag{4}$$

Where $n_{i,j}$ in Eq. (3) denotes the number of times a feature term s_i appeared in the blog f_j, and the denominator is the sum of occurrences of all feature terms appeared in f_j. In Eq. (4), D is the total number of blogs in the corpus, and the denominator is the number of blogs containing s_i.

Feature set B contains 295 terms, which were manually selected by a panel of domain experts based on their correlations with suicide. Feature set C contains 102 linguistic features, including emotional word categories such as positive words, sad words, and angry words. Among the 102 linguistic features extracted, 88 corresponded to the 88 word categories of SCMBWC; 3 features were related to the categories indicating past, now and future, respectively, such as “yesterday”, “today” and “tomorrow”; and the remaining 11 features were related to different kinds of punctuations, including period, comma, colon, semicolon, question mark, exclamation point, dash, quotation mark, apostrophe, parent, and other punctuations. The three original different predictive feature sets are presented in Table 1.

Table 1. The original predictive feature sets

	Number of features	Selection methods
Feature set A	2,000	Learned from real blog data
Feature set B	295	Chinese suicide dictionary
Feature set C	102	SCMBWC

2.3 Model Construction with Machine Learning Algorithms

We selected two algorithms, SVM (Support Vector Machine) and DT (Decision Tree), which not only the most commonly used classification machine learning algorithms, but also frequently used in the psychological character, emotion, and suicidal ideation detection (Gamon et al. 2013; Huang et al. 2014; Guan et al. 2015; Iliou et al. 2016).

SVM is an important example of "kernel methods" in statistical learning, one of the key areas in machine learning. It is a discriminative classifier for binary classification formally defined by a separating hyperplane. The operation of the SVM algorithm is based on finding the hyperplane that separates m-dimensional data into two classes and gives the largest distance to the nearest training samples of the two classes.

DT uses a treelike graph to model the reasoning process of mapping an input feature set to one of the predefined class labels. We chose the CART (classification and regression trees) algorithm in this study. CART constructs a binary decision tree by splitting a node into two child nodes repeatedly, beginning with the root node that contains the whole learning sample.

2.4 Measures of Model Performance

The performances of the constructed suicidal ideation detection models were evaluated using four metrics that have been commonly used in evaluating classification models, including precision (P), recall (R), F-measure (F), and accuracy (A), as defined by Eqs. (5)–(8). Precision is the percentage of all blogs detected as suicidal ideation that are indeed suicide ones; recall is the percentage of blogs in the entire corpus with suicidal ideation that are correctly identified; F-measure is a harmonic mean of precision and recall; and accuracy is the percentage of correctly detected blogs with and without suicidal ideation among the total number of blogs examined.

$$\mathrm{P} = \frac{TP}{TP + FP} \tag{5}$$

$$\mathrm{R} = \frac{TP}{TP + FN} \tag{6}$$

$$\mathrm{F} = \frac{2 \times P \times R}{P + R} \tag{7}$$

$$\mathrm{A} = \frac{TP + TN}{TP + FP + TN + FN} \tag{8}$$

Where TP denotes the number of blogs that are correctly classified as suicidal ideation; TN denotes the number of blogs that are correctly classified as non-suicidal ideation; FN denotes the number of blogs with suicidal ideation but classified as non-suicidal ideation incorrectly; and FP denotes the number of blogs without suicidal ideation but classified as suicidal ideation incorrectly.

3 Suicidal Ideation Detection Results

After calculation, we found the model constructed by SVM achieved better performances than DT, and feature set A made the best predictive performance. Therefore, the four metrics value of SVM from feature set A were chosen as the result of MFWF method.

We then compared the performances with the existing data-driven results (Liu et al. 2019) to draw final conclusion of this study. Table 2 shows the performances of the two different detection method.

Table 2. Precision, Recall, F-measure, and Accuracy of two detection methods

Detection method	Precision	Recall	F-measures	Accuracy
MFWF method	**0.89**	**0.88**	**0.88**	**0.89**
Data-driven method	0.88	0.78	0.83	0.85

The results show that the suicidal ideation detection based on MFWF method ($p < 0.01$), achieved better performances than data-driven method. It represented that the multi-feature fusion weighting method proposed in this study is an improvement for the data-driven method with single feature extraction. It's worth noting that recall of MFWF method increased from 0.78 to 0.88 by 12.8%, which also indicated that the rate of missed detection decreased. It's of great significance for suicidal ideation detection.

4 Discussion

There are two major findings and contributions of this research. First, on suicidal ideation detection, the multi-feature fusion method is better than the single feature extraction method, as it effectively integrates the existing methods. not only considered data and expert opinion (Chinese suicide dictionary), but also combines the features of emotional expression (LIWC). Second, general linguistic features extracted by or traditional suicide dictionary or natural language processing tool such as SCMBWC, may not be very effective. People may use some suicide-related terms that do not exist in any existing suicide lexicons, as we observed in this study. It suggests that the predictive features for a suicidal ideation detection model for social media should take into consideration suicide-related terms used in social media.

The results of experiments show that a suicidal ideation detection system can successfully discover suicidal ideations from social media content with a reasonably high accuracy. The promising results demonstrate the practical feasibility and value of real-time detection of suicidal ideation and prevention on social media, which may help reduce suicidal intentions and attempts.

Based on the conclusions of this study, we can better use social media content analysis to prevent suicide in practice. For example, evaluating blog posts on social media on a daily basis by machine. If machine screening detected suicidal ideation, we

would conduct a manual assessment. Once we confirmed the same conclusion, early warning and intervention would be implemented in a timely manner. Which is also the reason why we want to achieve more accurate and efficient machine screening result.

There are some limitations of our study. First, we used blogs collected from a single social media site. There are other types of textual social media content, such as discussion messages in online communities that may possess different characteristics. Therefore, the generalizability of the findings of this research needs to be taken with caution. Second, given the scope of this study, we did not include other potential feature selection methods, such as feature selection through deep learning, which is worth to be explored. Third, the units of the analysis in this study are individual blogs. We did not consider any significant changes in emotions, self-expressions, and other relevant features across multiple contiguous blogs, which may provide additional helpful cues for assessment. These limitations provide us opportunities for future research.

Acknowledgement. This study was partially supported by the Key Research Program of the Chinese Academy of Sciences (No. ZDRW-XH-2019-4).

References

Choudhury, M.D., Gamon, M., Counts, S., Horvitz, E.: Predicting depression via social media. In: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, Cambridge, Massachusetts, USA, 8–11 July, pp. 128–137 (2013)

Davila, J., Hershenberg, R., Feinstein, B.A., Gorman, K., Bhatia, V., Starr, L.R.: Frequency and quality of social networking among young adults: associations with depressive symptoms, rumination, and co-rumination. Psychol. Popul. Media Cult. **1**(2), 72–86 (2012)

Gao, R., Hao, B., Li, H., Gao, Y., Zhu, T.: Developing simplified Chinese psychological linguistic analysis dictionary for microblog. In: Imamura, K., Usui, S., Shirao, T., Kasamatsu, T., Schwabe, L., Zhong, N. (eds.) BHI 2013. LNCS (LNAI), vol. 8211, pp. 359–368. Springer, Cham (2013). https://doi.org/10.1007/978-3-319-02753-1_36

Gamon, M., Choudhury, M.D., Counts, S., Horvitz, E.: Predicting depression via social media. AAAI (2013)

Guan, L., Hao, B., Cheng, Q., Yip Paul, S.F., Zhu, T.: Identifying Chinese microblog users with high suicide probability using internet-based profile and linguistic features: classification model. Jmir Mental Health **2**(2), e17 (2015)

Huang, X., Zhang, L., Liu, T., Chiu, D.: Detecting suicidal ideation in Chinese microblogs with psychological lexicons. In: Proceedings of Intl Conf on Ubiquitous Intelligence and Computing, and Intl Conf on Autonomic and Trusted Computing, and Intl Conf on Scalable Computing and Communications and ITS Associated Workshops, pp. 844–849 (2014)

Iliou, T., et al.: Machine Learning Preprocessing Method for Suicide Prediction. In: Iliadis, L., Maglogiannis, I. (eds.) AIAI 2016. IAICT, vol. 475, pp. 53–60. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-44944-9_5

Kaplan, A.M., Haenlein, M.: Users of the world, unite! The challenges and opportunities of social media. Bus. Horiz. **53**(1), 59–68 (2010)

Li, T.M.H., Chau, M., Yip, P.S.F., Wong, P.W.C.: Temporal and computerized psycholinguistic analysis of the blog of a Chinese adolescent suicide. Crisis **35**(3), 168 (2014)

Lv, M., Li, A., Liu, T., Zhu, T.: Creating a Chinese suicide dictionary for identifying suicidal ideation on social media. Peerj **3**(10.7177), e1455 (2015)
McCarthy, M.J.: Internet monitoring of suicidal ideation in the population. J. Affect. Disord. **122** (3), 277–279 (2010)
O'Dea, B., Wan, S., Batterham, P.J., Calear, A.L., Paris, C., Christensen, H.: Detecting suicidality on Twitter. Internet Interv. **2**, 183–188 (2015)
Osman, A., Bagge, C.L., Guitierrez, P.M., Konick, L.C., Kooper, B.A., Barrios, F.X.: The suicidal behaviors questionnaire-revised (SBQ-R): validation with clinical and nonclinical samples. Assessment **5**, 443–454 (2001)
Pantic, I.: Online social networking and mental health. Cyberpsychology Behav. Soc. Netw. **17** (10), 652–657 (2014)
Paul, M.J., Dredze, M.: Discovering health topics in social media using topic models. Plos One **9** (8), e103408 (2014)
Pennebaker, J.W., Francis, L.E., Booth, R.J.: Linguistic inquiry and word count: LIWC2001. Lawrence Erlbaum Associates, Mahwah (2001)
Pestian, J.P., et al.: Sentiment analysis of suicide notes: a shared task. Biomed. Inform. Insights **5** (Suppl 1), 3–16 (2012)
Rosen, L.D., Whaling, K., Rab, S., Carrier, L.M., Cheever, N.A.: Is Facebook creating "iDisorders"? The link between clinical symptoms of psychiatric disorders and technology use, attitudes and anxiety. Comput. Hum. Behav. **29**, 1243–1254 (2013)
Salton, G., McGill, M.J.: Introduction to Modern Information Retrieval. McGraw-Hill, New York (1986)
Stirman, S.W., Pennebaker, J.W.: Word use in the poetry of suicidal and nonsuicidal poets. Psychosom. Med. **63**(4), 517–522 (2001)
Wang, X., Zhang, C., Ji, Y., Sun, L., Wu, L., Bao, Z.: A depression detection model based on sentiment analysis in micro-blog social network. In: Li, J., et al. (eds.) PAKDD 2013. LNCS (LNAI), vol. 7867, pp. 201–213. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40319-4_18
Wang, X., Li, A., Zhu, T.: Digital detection of suicidal ideation on social media. Int. J. Emerg. Ment. Health Hum. Resil. **17**(3), 661–663 (2015)
Liu, X., et al.: Proactive Suicide Prevention Online (PSPO): machine identification and crisis management for Chinese social media users with suicidal thoughts and behaviors. J. Med. Internet Res. (2019). https://doi.org/10.2196/11705