Comput Syst Sci & Eng (2019) 2: 65–78

International Journal of
Computer Systems
Science & Engineering

Incorporating stress status in suicide detection through microblog

Yuanyuan Xue[1,2], **Qi Li**[1], **TongWu**[1], **LingFeng**[1], **Liang Zhao**[3], **FengYu**[3]

[1] *Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China*
[2] *Department of Computer Technology and Application, Qinghai University, Xi'ning 810016, China*
[3] *Institute of Social Psychology, Xi'an Jiaotong University, Xi'an 710049, China*

Suicide has been a perplexing social problem around the world for a long time. Timely sensing hidden suicide risk and offering effective intervention are highly desirable and valuable for individuals and their families. Psychological studies prove that stress status, suicide-related expressions, and social engagement are reliable predictors of suicide risk. However, existing clinical diagnosis can only provide effective treatments to a restricted number of people because of its limited capacity. With the popular usage of social media like microblogs, a new channel to touch the inner world of many potential suicides arises. In this paper, we explore to automatically detect individual's suicide risk via a microblog platform. Referring to psychology theories, we take one's stress, self-concerns, suicide-related expressions, last words, social interaction, and emotional traits throughout the posting period on microblogs into account, and construct a 6-dimensional microblog feature space. We examine the differences of these features between the suicide group and the non-suicide group, through a set of real on line blogs posted by those who committed suicide and those who have no suicide intention. The observations reveal the same tendency as psychological theories suggested. To seek the causal relationship between these features and suicide risk, we describe a fuzzy cognitive map (FCM) classification model for suicide risk detection. We test the performance of the FCM classification model on a set of suicide and non-suicide users' real blogs from the Sina Weibo. The results show that the proposed model is effective and efficient on detecting users' suicide risk through Micro-blog, and yields better performance than other machine learning algorithms on small data set, with precision, recall and F1-measure increased by 9.7%, 15.8% and 13% respectively over second algorithm. The results also reveal stress feature vector plays more important role than other feature vectors and can effectively improve the performance of suicide risk detection.

Keywords: Suicide risk detection, microblog, stress status.

1. INTRODUCTION

1.1 Motivation

Suicide is a serious public health problem that can have lasting harmful effects on individuals, families, and communities. Each year in the world about 1 million individuals complete suicide, 10 to 20 million individuals survive from suicide attempts, and 50 to 120 million people are deeply influenced by the suicide-related behaviors of what happened to their relatives and friends [1]. In recent years, suicide rate continues to be on the rise with the intensification of contradictions between rapid development of society and people's adaptive capacity. The U.S. suicide rate rose 24% over the past 15 years [2]. In China and Japan, suicide has become the primary cause of death for young people [3], [4]. Given this, it is very much desirable to effectively detect and prevent suicide behaviors at its early stage.

Enlightened by the phenomenon that more and more people record their experiences and express themselves on social media like Twitter, researchers explore the use of this new communication media in suicide analysis [5]–[7]. Barak *et al.* [5] confirmed that the textual expressions of the people with suicide risks on the Internet are nearly consistent with their real emotions offline. A few individual suicide plans published on Twitter had drawn much attention from society and public, and had successfully been intervened [7]. Detecting suicide risks via social media offers the advantages of reaching massive population, low-cost, and real-time, compared to the traditional psychological, physiological, and biochemical scale and signal measurements.

The social media studies range from online suicide notes [6],

[8]–[10], community forums [11]–[15] to microblogs [6], [16], [16]–[23], where users' linguistic expressions, emotional traits, posting and interaction behaviors on the social media were captured. Researchers employ the information that reflects user's true emotions and feelings for suicide risk analysis. Techniques used in suicide risk analysis include emotion and sentiment analysis, opinion mining, natural language processing, and machine learning.

This study aims at microblogs for suicide risk detection due to its equality, freedom, fragmentation, and extensive use in the world. We could touch users' real inner world by analyzing their posting contents and behaviors on this kind of platform. It has confirmed that we can detect users' mental stresses, sentiments, anomalous situations and unknown threat behaviors [24]–[28]. In this study, we draw inspirations from psychological research results that it is the underlying long-standing mental stress that causes suicide behavior [29]–[32].

1.2 Stress is the Key Predictor of Suicide Ideation

Stress, by definition, is the psychological state of confusion and threat caused by various stimulate events and inherent requirements in life, which manifests as psychosomatic nervous or discomfort [33]. Excessive stress may compress individual's capability of emotional experiences and information-processing. That is, with a high stress level, individual's information-processing capacity and resolving power decrease, which simplifies decision making and causes an emotional experience in either-or and absolute ways [34]. When an individual feels stressed under the action of negative life events, his/her basic character of susceptibility will be launched, which tends to generate significant negative emotional response like depress, anxiety, angry, sorrow, and despair. These persistent traumas will make an individual more vulnerable and response capability damaged, resulting in a lower stress reduction capability and intensive suicide ideation [35]. Rich and Bonner [29] put forwards a stress-vulnerability model of suicide ideation and behavior, verifying that facing with life stress or negative life events, individuals usually feel depressed, loneliness, and hopeless. For university students, life stress and suicide ideation are positively correlated each other [31]. Stress in study and interpersonal relationship constitutes the largest impact factor for students [32]. In reality, people with suicide ideation constantly suffer from severe stress associated with negative or great traumatic events in their daily life [30].

1.3 Our Work

Stress causes the fluctuation, recurrence, and duration of manifested negative emotions and emotion fluctuations. Hence, going to the underlying stress origin and assessing individual's stress level offer us another channel to judge whether one is in the risk of extreme behavior or not. This is the focal different point from the existing work.

Beyond well-recognized evidential symptoms such as emotional trait, social interaction, and suicide-related statement such as last words on microblogs [19]–[23], [36], this paper involves stress into suicide risk analysis on microblogs. We leverage six sets of features on microblogs, including *stress* feature, *emotion* feature, *social interaction* feature, *self-concern* feature, and *suicide-related expression* feature throughout the whole posting period, and *last word* feature within the latest week. Based on this 6-dimensional feature space, we design a fuzzy cognitive map model (FCM) to learn which features have causations with suicide risk and how well they could be. Our experiments on 65 suicides and 65 non-suicides show that: 1) Stress feature vector plays the most determinative role in suicide risk detection than other features. 2) FCM classification model achieves the best performance, i.e., 80.8% in precision, 86.4% in recall and 83.5% in F1-measure, than other machine learning methods including Decision Tree, Naive Bayesian, Random Forest, and SVM. 3) For all features, stress root-mean-square deviation and mean stress level get the top 2 information gain, better than other features, which coincide with a view that the volatility and degree of stress are good indicators for suicide behavior detection.

To our knowledge, this is the first attempt in the literature that conducts microblog-based suicide risk detection from the stress perspective, as we believe it is the underlying long standing stress that causes manifested negative emotions and emotion fluctuations, leading to the final suicide consequence.

The reminder of the paper is organized as follows. We review some closely related work in Section 2. We construct a 6-dimensional microblog feature space in Section 3, followed by a FCM model for suicide ideation detection in Section 4. We report our performance study in Section 5, and discuss implications of the study in Section 6. We conclude the paper in Section 7.

2. RELATED WORK

2.1 Traditional Suicide Detection in Psychology

Suicide has been intensively studied in the field of psychology. Traditional measurements rely on questionnaires and face-to-face diagnosis to assess whether one is in the risk of suicide. Psychological scales and physical detection instruments are usually adopted to measure individual's physiological and biochemical signals for suicidal ideation prediction. To those who are identified with suicide risks, professional psychologist counseling and targeted treatments are offered.

As wearable technologies develop, some diagnosis and treatment centers utilize wearable devices in clinical trials to analyze and diagnose one's mental illness, such as bipolar disorder [37] and mental stress [38]. Through the wearable devices, patients' physiological signals could be automatically collected and uploaded to the treatment centers, where psychologists could rate patients' mental illness and take proper actions to help the patients at a high suicide risk.

While both approaches are professional and accurate, they are applicable to a small group of patients in the treatment centers. Real-time monitoring and helping massive population with suicide ideation outside the treatment centers are expensive and hard. Particularly, for the people who are suffering but tend to hide in most thoughts and refuse to seek helps from others, the approaches cannot function.

2.2 Suicide Detection via Social Media

In recent years, enlightened by the phenomenon that more and more people record their experiences and express themselves on social media like Twitter, researchers investigate the use of this new communication channel for suicide prediction, for its advantages of reaching massive population, low-cost, and real-time. The social media studies range from online suicide notes, community forums, to microblogs. Techniques used in suicide risk analysis include emotion and sentiment analysis, opinion mining, natural language processing, and machine learning.

2.2.1 Suicide Notes

Marcinczuk *et al.* [8] constructed an annotated suicide-related corpus through analyzing suicide notes. A suicide note classifier was built using machine learning techniques in [9], whose experimental result showed its superiority to psychologist in distinguishing fake suicide notes from real ones.

To get feasible indicators of suicidal behaviors, Desmet and Hoste [10] employed natural language processing and machine learning techniques and sensed 15 different emotions from suicide notes. Its performance study showed that fine-grained emotion detection benefits from classifier optimization and a combined lexicon-semantic feature set. Tim *et al.* [6] studied a 13-year-old youth's 193 blogs using the Chinese Linguistic Inquiry and Word Count (CLIWC), and found several key features from his suicidal notes, including posting frequency, progressive self-reference and positive-to-negative emotion words ratio.

2.2.2 Community Forums

The high rate of online social network use offers a novel venue to reach hard-to-reach young lesbian, gay, and bisexual (young LGB) individuals who have higher rates of suicide ideation. Silenzio *et al.* [11] used mySpaceCrawler to map social connections between LGB self-identified individuals between 16 and 24 years old participating in the online social network (www.MySpace.com, a popular site for adolescents and young adults, particularly sexual minority adolescents with over 189,000,000 registered users worldwide). A descriptive analysis of the structural characteristics known to affect diffusion within such networks was conducted. Finally, Silenzio *et al.* [11] conducted Monte Carlo simulations of peer-driven diffusion of a hypothetical preventive intervention within the observed social network. Huang *et al.* [12] conducted a lexicon-based keyword matching method over MySpace.com to check whether users have an intent to commit suicide.

For users who talked about mental health issues in the online forum called Reddit, Choudhury *et al.* [13] used logistic regression to analyze their shift tendency from mental health sub-communities to a suicide support sub-community. Three linguistic and interactional measures, namely *linguistic structure, interpersonal awareness*, and *interaction*, were taken into consideration. Linguistic structure aspect includes the fraction of nouns, verbs, and adverbs in posts and comments; automated readability index, a measure to gauge the understand ability of text; and linguistic accommodation, a process by which individuals in a conversation adjust their language styles according to that of others. Interpersonal awareness aspect includes the proportions of first person singular and plural, second and third person pronouns. Interaction aspect includes volumes of posts and comments authored, post length, length of comments authored, volume of comments received on shared posts, length of comments received, mean vote difference, and response velocity, given by the time elapsed between the first comment and the time the corresponding post was shared.

Tim *et al.* [14] proposed a collective intelligence system, which combined text affect analysis and summarization techniques, to identify suicide expressions in a Chinese web forum. Its affect analysis method classified a thread based on the main post only, whereas collective intelligence approach examined the comments from the replying users to derive the mainstream opinion. Masuda *et al.* [15] examined online forums in Japan and discovered that the number of communities to which a user belongs to, the intransitivity, and the fraction of suicidal neighbors in the social networks contributed the most to suicide ideation.

2.2.3 Microblogs

Sueki [16] examined the correlation between suicide-related tweets and suicidal behaviors based on a cross-sectional survey, where participants answered a self-administered online questionnaire, containing questions about Twitter use, suicidal behavior, depression and anxiety, and demographic characteristics. The survey result showed that Twitter logs could help identify suicidal young Internet users. Jashinsky *et al.* [17] used a list of search keywords and phrases relevant to suicide risk factors to filter potential suicide-related tweets, and group these at-risk tweets by state. It then compared the numbers of suicide tweeters in different states against the national suicide data, and observed a strong correlation between state Twitter-derived data and actual state age-adjusted suicide data, demonstrating that Twitter could be viewed as a viable tool for real-time monitoring of suicide risk factors on a large scale. Guan *et al.* [18] investigated linguistic and behavior features of the posts on a Chinese microblog, coming from 33 users who have committed suicide. The behavioral features included self-description, self-reference, group reference, interaction, openness, originality, transitivity, nocturnal activeness, and adoption of negative emoticons. The linguistic features contained 88 items coming from the Simplified Chinese Microblog Word Count Dictionary. The result showed that the suicide group had more self-mentions in their blogs than the controlled group.

O'Dea *et al.* [19] applied Support Vector Machine (SVM) and Logistic Regression (LR) methods to classify tweets into three levels (*strongly concerning, possibly concerning,* or*safe concerning*) on Twitter based on the weighted term frequency in human coded tweets. Its findings confirmed that Twitter tends to be used by individuals to express suicidality and that such posts evoked a level of concern that warranted further investigation.

Huang *et al.* [20] extracted linguistic features from 53 known suicidal blogs on Weibo, a Chinese microblog, based on a psychological lexicon dictionary. An SVM classifier was applied to detect suicidal ideation, who's F-measure reached 68.3%. A topic model combined with machine learning algorithms was further derived for suicide ideation detection. It was found that Latent Dirichlet Allocation (LDA) based approach outperforms the Chinese Linguistic Inquiry and Word Count (CLIWC) lexicon based approach in suicide prediction on Weibo [21], [36].

Aiming at classifying text relating to communications around suicide on Twitter, Burnap *et al.* [22] built a few baseline classifiers (SVM, J48 Decision Tree, and Naive Bayes) to distinguish between the more worrying content such as suicidal ideation and other suicide-related topics such as reporting of a suicide, memorial, campaigning and support. Three sets of features were extracted from the text. 1) Features representing lexical characteristics of the sentences used, such as the Parts of Speech (POS), and other language structural features such as the most frequently used words and phrases. References to self and others are also captured with POS. 2) Features representing sentiment, affective and emotional features such as fear, anger, and general aggressiveness, and levels of the terms used within the text. 3) Features representing idiosyncratic language expressed in short, informal text such as social media posts within a limited number of characters. Burnap *et al.* [22] further deployed an ensemble classifier using the Rotation Forest algorithm and a Maximum Probability voting classification decision method for further improvement. The performance study showed that the later could achieve an F-measure of 0.728 overall and 0.69 for the suicidal ideation class.

Based on eight basic emotion categories (joy, love, expectation, anxiety, sorrow, anger, hate, and surprise), Ren *et al.* [23] proposed three accumulated emotional traits (i.e., emotion accumulation, emotion covariance, and emotion transition) as the special statistics of emotions expressions in blog streams for suicide risk detection. The emotion accumulation trait is the summarization of emotion distribution within L continuous blog articles, the emotion covariance trait is the connection between different emotion categories, and emotion transition is the patterns in emotion changes. Throughout the study, 4 suicide and 4 non-suicide individuals' emotional traits were illustrated for comparison. A linear regression algorithm based on the three accumulated emotional traits was employed to examine the relationship between emotional traits and suicide risk. The experimental result showed that by combining all of three emotion traits together, the proposed model could generate more discriminative suicidal prediction performance.

3. A 6-DIMENSIONAL MICROBLOG FEATURE SPACE FOR SUICIDE RISK DETECTION

The quality of microblog feature space has important effect on suicide risk recognition. Guided by the psychological investigation about the predictors of suicidal behavior and existing good work, we explore a six-dimensional microblog feature space $(\overline{FS}, \overline{FC}, \overline{FU}, \overline{FW}, \overline{FO}, \overline{FE})$ to address users *stress, self-concerns, suicide- related expressions, last words, social interaction,* and*emotional traits* throughout the posting period.

3.1 Stress Feature Vector $\overline{FS}$

Long-standing stress is a good predictor of suicide risk according to the psychological study [29], [30]. Here, we measure long standing stress in terms of *stressful intervals*.

3.2 Stressful Moments

Previous work demonstrated the feasibility of stressful blog detection on a microblog platform (e.g. Twitter, Tencent Weibo, and Sina Weibo) [24], [25], [39]–[42]. Let (t, b) represent a blog b posted at a time moment t. We apply the stress detection function $stress(t, b) = (t, sc, sl)$ [41] to each blog, whose posting content (linguistic text, emoticons, repetitive exclamation or question punctuations, shared music/picture genre), posting time and frequency, as well as social interaction with friends (being liked/reposted/cared and comment-response acts beneath the blog) are examined. The function returns the stress category $sc \subseteq C_{ategory}$ and stress level sl from blog b, where $C_{ategory}$={"study", "family", "inter-personnel", "self-cognition", "affection", "health", "finance", "profession", "unknown"} and $L_{evel} = \{0, 1, 2, 3, 4, 5\}$, corresponding to {"none", "very weak", "weak", "moderate", "strong", "very strong"} stress level. b is called a**stressful blog**, and t is called a **stressful moment**, if and only if its detected stress level is over zero (i.e., $sl > 0$).

3.3 Stressful Intervals

Based on and beyond stressful moments, we look for stressful intervals, since long-standing repetitive stress has a more serious impact on one's physical and psychological health than acute momentary stress due to sustained high levels of chemicals released in the response, playing a causal role in suicide action.

However, deriving stressful intervals from stressful moments is not a trivial task, facing two challenges. On one hand, a blog is limited to 140 characters, short enough for one to completely precisely express the endured stress. Correctly sensing one's stressful moment from a short blog is challenging. On the other hand, users posting behaviors are quite random and sparse. Sometimes they may post a lot of blogs in a day, and sometimes no blogs have been posted for a few days. For instance, in a suicide case, suffering from study stress, a college student posted 18 stressful blogs in succession on the day of suicide. Among his total 200 blogs in a year, around 100 were posted in the month before his suicide. Such an imbalanced posting behavior with data missing and aggregation requirements challenges stressful intervals detection.

To remedy wrong and missed interpretations of stressful moments, we propose to examine precedent and follow-up moments consecutively within a certain period under the assumption that the occurring frequency of stressful moments in a stressful interval is more than the occurring frequency of stressful moments in a non-stressful interval. In other words, the rate of posting stressful blogs in a stressful interval is more than the rate of posting stressful blogs in a non-stressful interval.

3.3.1 Aggregating Momentary Stress in Unified Time Units

To fairly compute users' posting rates, we first need to uniform time unit (which can be day, week, or month, and it is day in this study). For blogs posted in the same time unit (day), we aggregate the detected stress results by the average of the stress levels and union of the stress categories.

Formally, suppose a user posts n blogs consecutively $(t_1, b_1), (t_2, b_2), \ldots, (t_n, b_n)$ at time moments $t_1, t_2, \ldots, t_n$, respectively, where $(t_1 <_t t_2 <_t \cdots <_t t_n)$. Through the stress detection function $stress(t_i, b_i) = (t_i, sc_i, sl_i)$ (where $1 \leq i \leq n$, we obtain $(t_1, sc_1, sl_1), (t_2, sc_2, sl_2), \ldots, (t_n, sc_n, sl_n)$. For the blogs posted at the same time unit (day) T, $(t_j, b_j), \ldots, (t_{j+k}, b_{j+k})$ (where $t_1 \leq_t t_j \leq_t t_{j+k} \leq_t t_n$), we aggregate their stress detection results into one (T, Sc, Sl), where $(Sc = sc_j \cup \cdots \cup sc_{j+k})$ and $Sl = \lceil (sl_j + \cdots + sl_{j+k})/(k+1) \rceil$.

Let $(T_1, Sc_1, Sl_1), (T_2, Sc_2, Sl_2), \ldots, (T_m, Sc_m, Sl_m)$ be the stress sequence detected at the unified time units $T_1, T_2, \ldots, T_m$, where $(T_1 <_t T_2 <_t \cdots <_t T_m)$ and $(m \leq n)$. For the temporal continuity, we set $(Sc_i = \emptyset)$ and $(Sl_i = 0)$ when no blogs were posted at time unit $T_i (1 \leq i \leq m)$. In a similar way as stressful moment, we define T_i as a **stressful time unit** if and only if $Sl_i > 0$.

3.3.2 Definition of Stressful Interval

Definition 1. Let $\mathrm{I} = T_1, T_m$ be a time interval, starting at time unit T_1 and ending at time unit $(T_m >_t T_1)$. The temporal length of I is the number of time units in I, denoted as $|I| = m$. Let $(T_1, Sc_1, Sl_1), (T_2, Sc_2, Sl_2), \ldots, (T_m, Sc_m, Sl_m)$ be a stress sequence detected consecutively at the unified time units in $T_1, T_2, \ldots, T_m$ in I. I is called a **stressful interval**, if and only if it satisfies the following three conditions.

(1) The temporal span is not less than threshold δ, i.e., $|I| = m \geq \delta$.

(2) The proportion of stressful time units in I is not less than threshold τ, i.e., $|\{T_i | (1 \leq i \leq m) \wedge (Sl_i > 0)\}|/|I| \geq \tau$.

(3) There does not exist a longer time interval $I' = T'_1, T'_m$, such that $I' = T'_1, T'_m$ temporally encloses I; T_1, T_m (i.e., $T'_1 \leq_t T_1 <_t T_m \leq_t T'_m$) and meanwhile $|\{T'_i | (1 \leq i \leq m) \wedge (Sl'_i > 0)\}|/|I'| \geq \tau$.

Based on the results of comparative experiment for thresholdand δ and θ, in the study, $\delta = 7$ days, and $\tau = 80\%$. □

According to the definition, stress appears to be more concentrative in stressful intervals than in non-stressful intervals.

3.3.3 Characteristics of Stressful Intervals

Many times, users with suicidal thoughts have to go through different stress concurrently or successively, and react strongly or weakly. During some periods, their stress feelings fluctuate sharply, but slope gently during some other periods. In some suicidal cases, there exists a certain period before suicidal behavior, when the stress levels of the suicides go to the extreme, reaching the peak throughout the whole posting period. Different stressful intervals sensed throughout the posting period signify users' different echoes to the stress.

Centered on *stress interval*, we consider six features in the stress feature vector $\overline{FS} = (FS_1, FS_2, FS_3, FS_4, FS_5, FS_6)$ for suicide risk analysis. They characterize the intensity of user's long-standing stress.

Let $\mathcal{I}$ be the set of discontinuous stressful intervals detected throughout the posting period. Assume $I \in \mathcal{I}$ is a stressful interval, where $I = T_1, T_m$ and $(T_1, Sc_1, Sl_1), (T_2, Sc_2, Sl_2), \ldots, (T_m, Sc_m, Sl_m)$ be a stress sequence detected consecutively at its unified time units $T_1, T_2, \ldots, T_m$ respectively.

(1) *Total number of stressful intervals* $FS_1 = |\mathcal{I}|$.

Figure 1 compares the six features of detected stressful intervals from 65 suicides and 65 non-suicides on the Sina Weibo. The 65 suicides have passed away, as verified by news or media. We randomly selected 65 non-suicide users, whose self-description tags contain no suicide-related terms on the Sina Weibo, and whose blogs are checked by three psychological students to ensure they do not have potential suicidal tendencies. Among the users, the longest posting period is up to 5 years and 6 months, and the shortest period is 1 year. A suicide user posted 5102 blogs maximally, 130 minimally, and 779 averagely. A non-suicide user posted 3581 blogs maximally, 105 minimally, and 678 averagely.

In Figure 1 (a), about half of suicide users experienced more stressful intervals within the more recent1year, while most non-suicide users have a few or even none of stressful intervals. For the non-suicide users, the average number of stressful intervals is 3.2. While for the suicide users, the average number of stressful interval is 7.1, which is over twice as much as non-suicide users. We also observe that there are 62% suicide users with more than 5 stressful intervals, and only 25% for non-suicide users; as well as 31% suicide users with more than 10 stressful intervals, and only 9% for non-suicide users. These phenomena indicate the suicides had more frequent stress feelings than the non-suicides.

(2) *Mean stress level on average*

$$FS_2 = \lceil \sum_{I \in \mathcal{I}} \mathrm{meanLevel}(I)/|\mathcal{I}| \rceil.$$

Function *meanLevel(I)* returns the mean stress level in I. It reflects the average stress level experienced by a user in I. It is a modest way to discern user's stress level during this period. $meanLevel(I) = \lceil (Sl_1 + Sl_2 + \cdots + Sl_m)/m \rceil$.

(3) *Root-mean-square deviation of stress levels on average*

$$FS_3 = \lceil \sum_{I \in \mathcal{I}} deviate(I)/|\mathcal{I}| \rceil.$$

Function *deviate(I)* computes the root-mean-square deviation of stress levels from the mean stress level in I. It depicts the fluctuation degree of user's stress levels in I.

$$deviate(I) = \lceil \sqrt{\frac{1}{m} \sum_{1 \leq i \leq m} (Sl_i - meanLevel(I))^2} \rceil.$$

(4) *Peak stress level on average*

$$FS_4 = \lceil \sum_{I \in \mathcal{I}} peakLevel(I)/|\mathcal{I}| \rceil.$$

Function *peakLevel(I)* indicates the user's maximal stress level in I. $peakLevel(L) = \max(Sl_1, \ldots, Sl_m)$.

(5) *Span on average* $FS_5 = \lceil \sum_{I \in \mathcal{I}} span(I)/|\mathcal{I}| \rceil$.

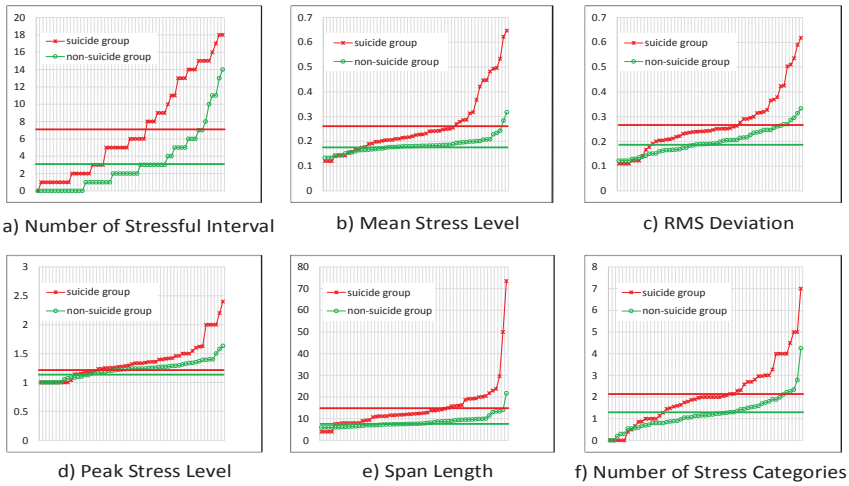


Figure 1 Comparing stressful intervals related features ($\overline{FS}$) of 65 suicides' and 65 non-suicides' blogs on Sina Weibo, where x-axes denote user ID, and y-axes denote features values, horizontal red lines correspond to the average value of suicides, and horizontal red lines correspond to that of non-suicides.

Function *span(I)* returns the temporal duration of the stressful interval I. The longer the duration, the higher stress intensity one has to bear. $span(I) = |I| = m$.

(6) *Number of different stress categories on average*

$$FS_6 = \lceil \sum_{I \in \mathcal{I}} categoryNum(I)/|\mathcal{I}| \rceil.$$

Function $categoryNum(I)$ returns the number of different stress that one undertakes in I. Suffering multiple stress at the same time is quite harmful to one's mental and physical health, leading to nervous breakdown easily. $categoryNum(I) = |Sc_1 \cup Sc_2 \cup \cdots \cup Sc_m|$.

From Figure 1 (b)-(f), we observe obvious differences between suicide users and non-suicide users. For the suicide users, the values of mean stress level, root-mean-squared deviation of stress levels, peak stress level, span length, and number of different stress categories are significantly higher than those of non-suicide users.

As shown in Figure 1 (b), the average value of mean stress level of the suicide group (0.26) is 44% higher than that of the non-suicide group (0.18). We also find that most non-suicide users' mean stress levels are more evenly distributed around the average value (0.18) except a few individual points. While for the suicide group, those values change obviously, with 30% higher than its average value (0.26) and 0.65 maximally. That implies suicide users are more likely to suffer greater stress than non-suicide users.

In Figure 1 (c), the average root-mean-squared deviation (RMS) of the suicide group (0.27) is 42% higher than that of the non-suicide group (0.19), which probably means suicide users stand more intense stressful feeling fluctuations. Specially, there are5individualseven with a value of over 5, indicating that their stressful feelings fluctuate sharply during posting periods.

As shown in Figure 1 (d), the average peak stress level of the suicide group (1.36) is only 11% higher than that of the non-suicide group (1.22), with no obvious difference. However, we can also find that most non-suicide users' peak levels are around the average level (1.22). While about half of suicide users' peak levels are far from the average level (1.36).

Figure 1 (e) shows that the average value of the suicide users' stressful interval span (15) is almost twice as much as that of the non-suicide users (8). Moreover, we note that most non-suicide users' stressful interval span last about 7 days, accounting for 60% or so; and most suicide users' stressful interval span last from 10 to 20 days, accounting for 67%. This observation possibly indicates that most normal people could recover from stress feelings after about 7 days. While for the suicide users, who are more susceptible to mental stress, they suffer higher and long-term stress and take longer time to recover from stressful situation than non-suicide users.

As to the number of different stress categories, from Figure1 (f), we can learn that the average value in the suicide group (2.1) is nearly double of that in the non-suicide group (1.2). We can infer that for most normal users, they only undergo single stress type at same time in most cases. While for the suicide users, about 66% of them suffer more than 2 stress categories during the same period, even to 7 maximally. This may reveals that most suicide users are vulnerable to be trapped into suicidal thoughts when suffering many kinds of stresses for a long time.

All these observations indicate that suicide users endure more intensive, frequent, and fluctuated stress than non-suicide users. So, it is reasonable to use these features as predictors for suicide detection through microblog.

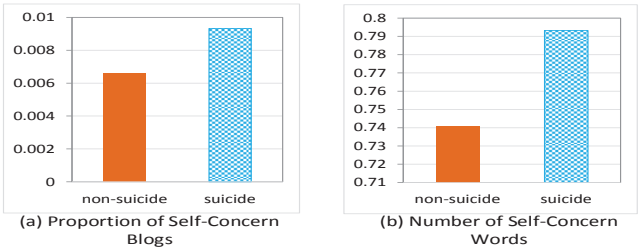


Figure 2 Comparing self-concern features $\overline{FC}$ of 65 suicides' and 65 non-suicides' blogs on Sina Weibo.

3.4 Self-Concern Feature Vector

The frequency of self-references implies one's self-concern level in social networks, and a high frequency indicates a high self-concern level. White and Mazlack [43] compared suicide notes with non-suicides' blogs, and discovered that the suicides possessed more self-concerns, and used more first-person pronouns (such as "I", "me", "my", and "mine") in their notes than non-suicides. This phenomenon is also mirrored in microblogs.

In this study, we examine the words related to self-concerns in one's blogs, including "I", "me", "my", and "mine", "am", "we", "us", "our", and "ourselves", etc. We call a blog **self-concern-blog**, if it contains one or more self-reference words. Let CBdenote the set of self-concern blogs throughout the posting period. Function $selfWOrdNum(cb)$ returns the number of self-concern words in the self-concern-blog $cb \in CB$. We compute the following two elements to characterize the self-concern feature vector $\overline{FC} = (FC_1, FC_2)$ on microblogs.

(1) *Proportion of self-concern blogs over the whole blogs throughout the posting period* $FC_1 = |CB|/|B|$

(2) *Number of self-concern words per self-concern-blog on average*

$$FU_2 = \lceil \sum_{sb \in SB} swordNum(sb)/|SB| \rceil$$

(3) Proportion of suicide-related words over all the words per suicidal-blog on average $FU_3 = \sum_{sb \in SB} \frac{swordNum(sb)}{wordNum(sb)}/|SB|$

In Figure 2, we can observe some differences between suicide users and non-suicide users. For the suicide users, the values of proportion of self-concern blogs and number of self-concern words are much higher than those of the non-suicide users. It reveals that the suicides are prone to self-preoccupation and collective attention in social media by using first person singular and first person plural. Literature suggests that pronoun usage can reveal an individual's mental well-being in social media [44].

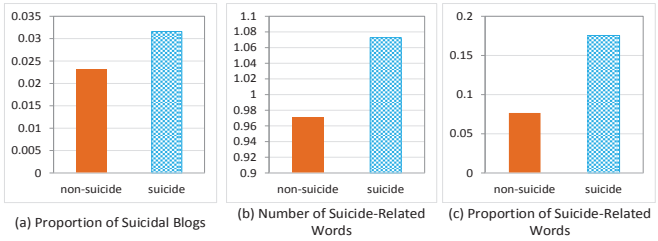
3.5 Suicide-Related Expression Feature Vector

Suicide-related expression is one of the important observable indexes of suicide ideation [45]. Suicides tend to express rather than repress their despair suicidal feelings on their blogs. Table 1 gives some real example blogs taken from Sina Weibo, containing suicide-related words.

Table 1 Example Posts Containing Suicide-Related Words on Sina Weibo.

I really don't want to touch anything, because I'm so **sad** and **cannot control myself.**
Where is the **dead-end path**?
My heart is so cold! Where I could find a warm world?
Every day seems to **suffer**.
Why I am so **uncomfortable** and **sad**, and could only talk to strangers.

To identify individual's suicide-related expression from microblogs, we take the Chinese social media based suicide dictionary [46] for reference. The dictionary lists 2167 suicide related words and phrases, belongingto5categories (i.e., suicidal thoughts, self-injury, physical and mental status, life event, and situational mood). These categories measure suicidal ideation from different perspectives. Moreover, words and phrases in each category are assigned a weight value (from1 to3) to mark their correlation degrees with suicide risks. The higher the weight is, the closer the correlation is. Table2 shows parts of the dictionary.



(a) Proportion of Suicidal Blogs (b) Number of Suicide-Related Words (c) Proportion of Suicide-Related Words

Figure 3 Comparing suicide-related expression features $\overline{FU}$ of 65 suicides' and 65 non-suicides' blogs on Sina Weibo.

We call a blog **suicidal blog** if it contains one or more words/phrases (e.g., the words of boldface type in Table 1) in the dictionary.

Let B and SB denote the set of blogs and suicidal-blogs throughout the posting period, respectively. Function $wordNum(sb)$ returns the total number of words in the suicidal blog sb, and $swordNum(sb)$ returns the number of suicide-related words in sb, where $(sb \in SB)$. We assess one's suicidal ideation via the following three elements in the suicide-related expression feature vector $\overline{FU} = (FU_1, FU_2, FU_3)$.

1. *Proportion of suicidal-blogs over the whole blogs*

$$FU_1 = |SB|/|B|$$

2. *Number of suicide-related words per suicidal-blog on average*

$$FU_2 = \left\lceil \sum_{sb \in SB} swordNum(sb)/|SB| \right\rceil$$

3. *Proportion of suicide-related words over all the words per suicidal-blog on average*

$$FU_3 = \sum_{sb \in SB} \frac{swordNum(sb)}{wordNum(sb)}/|SB|$$

In Figure 3, we can observe that the values of proportion of suicidal-blogs, number of suicide-related words, and proportion of suicide-related words in the suicide group are much higher than those in the non-suicide group, which means the suicide are more likely to use suicide-related words to hint their hopeless feelings in microblog. These observations imply that suicide-related expression could be valuable signs for suicide risk detection.

3.5 Last Word Feature Vector $\overline{FW}$

Besides examining suicide-related expressions throughout the whole posting period, we pay attention to the presence of last

Table 2 A Chinese suicide dictionary oriented to social media [45].

Category	Word Number	Example Words (weight=1)	Example Words (weight=2)	Example Words (weight=3)
Suicidal Thought	586	fade, termination, regret ante mortem, inanition isolation, destiny, unrest gloomy, dim, flee, downfall, etc.	senseless, burden, confuse disillusionment, despair, hard nothing more to say, deserve discard, death, out of misery, etc.	pass away, last wish, afterlife nightmare, purgatory, moksha leave this world, missed life as good as death, extremity, etc.
Self-Injury	88	drug, self-mutilation self-destructing, dosage self-burning, self-injury drowning, scar, lacerate, etc.	carbon, jumping off a building take poison, knife, hit and kill autosadism, Prozac, bleed, fast cutthroat, sword, sleeping pills, etc.	drink pesticide, hang oneself hara-kiri, jump into the sea throw oneself into a river, euthanasia, cut one's wrists, etc.
Physical Mental Status	929	unbearably, crazy, sorrow difficult, help, cruel, grief abandon, awfully, tragic decay, fear, darkness, sad, etc.	guilt, fault, indifference, crying helpless, loneliness, lost mind self-abasement, out of control incurable, languish, chilling, etc.	come to a crisis, grievance insomnia, breakdown, tearful collapse, fragility, numbness torment, depression, stressed, etc.
Life Event	312	Heartbreaker, frustration reality, strike, misfortune disaster, love to the end eternal wound, separate, etc.	heavy burden, single person forgiveness, cheat, selfishness betrayal, despise, depravation derailment, unfair, lose myself, etc.	pressure, stress, lost, injure disappointment, hurt, failure breaking up, stressing, loser underdog, die for love, etc.
Situational Mood	252	anger, damn it, fury, hate abuse, sordid, speechless rage, impudicity, sarcasm fuck off, curse, kick myself, etc.	disgusting, brutality, merciless atonement, too late to regret atone for one's crime, victim encumber, grieving parents, etc.	regret, hatred, bad blood repent, compunctious, anger get angry, pissed off, hate take offence, be burned up, etc.

words, particularly those post within the latest one week. According to the interpersonal theory, last words are recognized as an important observable risk factor of suicide ideation. Many times, before committing suicide, people are prone to leave some words conveying complex innermost feelings like regret, guilty, wishes towards their families and friends, or funeral [45]. In recent years, people who are considering suicide tend to post last words on microblogs to catch others' attentions. As an example, on February 12, 2012, in the Fujian province of China, a lady was lovelorn and left a message of attempting suicide on the weibo microblog. Right after the release, the message immediately received netizens wide attentions. They spread the message on the microblog platform and wrote encouraging words under the message, trying to deter her suicide action. Some enthusiastic netizens finally found and saved the lady [47]. Table 3 gives some example blogs containing last words from a few suicides. These words (of boldface type) reveal strong suicidal signals, and have weight 2 or 3 in the suicide dictionary [46].

As last words are usually released by individuals with suicidal ideation a few days before completing suicide, we check the existence of last words within the latest one week's blogs. We call a blog a **last-word-blog**, if it contains one or more last words in the dictionary. Let B' and LB denote the set of blogs and suicidal-blogs within the latest one week, respectively. Function $wordNum(lb)$ returns the total number of words in the last-word-blog $lb \in LB$, and $lwordNum(lb)$ returns the number of last words in lb.

Four elements in association with the last word feature vector $\overline{FW} = (FW_1, FW_2, FW_3, FW_4)$ are computed to measure suicidal ideation risks.

1. *Proportion of last-word-blogs over the whole blogs within the most recent one week* $FW_1 = |LB|/|B'|$

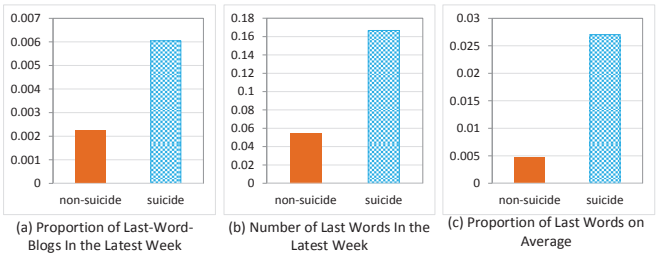
2. *Number of last words per last-word-blog within the most recent one week on average*

$$FW_2 = \lceil \sum_{lb \in LB} lwordNum(lb)/|LB| \rceil$$

3. *Proportion of last words over all the words per last-word-blog on average*

$$FW_3 = \sum_{lb \in LB} \frac{lwordNum(lb)}{wordNum(lb)}/|LB|$$

In Figure 4, we can observe significant differences between suicide users and non-suicide users. For the suicide users, the values of proportion of last-word-blogs, number of last words per last-word-blog, and proportion of last words are significantly higher than those of the non-suicide users. This probably reveals the phenomenon of that most suicide users choose to leave last words to express sorrow and regret feelings on microblog before committing suicide. These observations also support the view of using last words as traits for suicide risk detection.



(a) Proportion of Last-Word-Blogs In the Latest Week (b) Number of Last Words In the Latest Week (c) Proportion of Last Words on Average

Figure 4 Comparing last word-related features $\overline{FW}$ of 65 suicides' and 65 non-suicides' blogs on SinaWeibo.

Table 3 Example Blogs Containing Last Words on Sina Weibo.

Goodbye, and **sorry**. If someone care about me, please don't be **sorry** for me. Believe that this is a **relief** for me. Don't **blame** me. I know this is **selfish behavior**. It's the **last time**. Please believe that this is **not an impulse**, but a thought for a long time. In short, I would like to thank everyone who has brought happiness to me. **Goodbye**, my friends, I can no longer say "**as long as I am alive**..."
Too much **lies** and **deceit**, too much **doubt** and **suspicion**, too much **mockery** and**subversion**, this world is really **tiring**. Well, **this ends now**!
So be it, I go, no **goodbye**. And then **give up** only to **let go**. It's time to make a **complete farewell**, which should be **thoroughly terminative**.
I **don't want to live anymore**!**Goodbye**, this world!**Goodbye**, all those who care about me. **I'm gone**, don't miss me. Maybe it will not happen, because I really **useless**, right? When you see this tweet, **I have gone**! I've thought that I **really sorry** about hurting my **parents** that night. Actually I don't want to do like that. This time I'm really **ready to leave**. Still **angry** with me? Please don't be **angry** with a **death** and it's not worth it. **Goodbye**, brother! If there is an **afterlife**, I don't want to be friends with you.

3.6 Social Interaction Feature Vector $\overline{FO}$

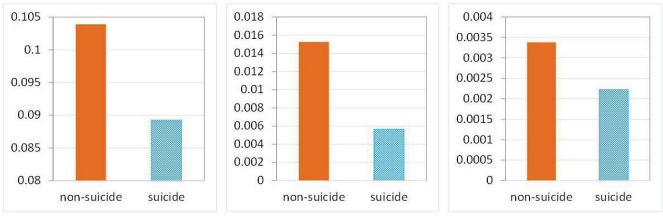
Social isolation is a significant and reliable predictor of suicide ideation. Desperate person tends to have a weak social network, and thus gets weak social supports. Psychologists including Mandelli believe that weak social networks are usually related to mental disorder, higher depressed severity, and suicide tendency, which indirectly reflect low social support levels. Low social support levels inevitably increase the risk of depression and disorder when people endure stressful life events [48].

With the extensive use of the microblog platform, people are keen on sharing thoughts and interacting with like-minded companions. Especially when individual trapped in trouble, if he/she could get the support, understanding, and acceptance from family, friends and peers, his/her stress and follow-up negative emotions could be offset, which could probably avoid extreme suicidal behavior. In a word, users' online social interaction reveals his/her social isolation, and thus social support level to a certain extent.

Three *direct* ways for one to interact with others on microblogs are via @-mention, @-reply, and blog-forward mechanisms. We call a user's blog **interaction-blog**, if it involves at least one of the above interaction activities. Let MB, AB and FB denote the set of interaction-blogs via @-mention, @-reply, and blog-forward mechanisms throughout the posting period, respectively.

We use the following three elements to measure one's social engagement in the social interaction feature vector $\overline{FO} = (FO_1, FO_2, FO_3)$.

1. *Proportion of @-mention based interaction-blogs over the whole blogs throughout the posting period* $FO_1 = |MB|/|B|$



(a) Proportion of @-mention Blogs (b) Proportion of @-reply Blogs (c) Proportion of Forwarding Blogs

Figure 5 Comparing social interaction-related features $\overline{FO}$ of 65 suicides' and 65 non-suicides' blogs on Sina Weibo.

2. *Proportion of @-reply based interaction-blogs over the whole blogs throughout the posting period* $FO_2 = |AB|/|B|$

3. *Proportion of blog forward based interaction-blogs over the whole blogs throughout the posting period* $FO_3 = |FB|/|B$

In contrast to other observations, as shown in Figure 5, we can observe that there exist obvious differences between suicide users and non-suicide users. But this time, for the suicide users, the values of proportion of @-mention, proportion of @-reply, and proportion of blog forward are significantly lower than those of the non-suicide users. These observation may imply that suicide users have lower degrees of social activity and social support than non-suicide users in social media. This result is consistent with literature [48], which suggests social activity and social support are good predictors for suicide risk.

3.7 Emotion Feature Vector $\overline{FE}$

Regarding the emotion features, we follow the most recent good work [23], which considered eight emotion categories (joy, love, expectation, anxiety, sorrow, anger, hate, and surprise), and computed three accumulated emotional traits (i.e., emotion accumulation, emotion covariance, and emotion transition) as the special statistics of emotions expressions in the most recent L (around 1000) blogs for suicide risk detection.

1. *Emotion accumulation* FE_1 is the summarization of emotion intensities in the most recent L blogs. Emotion intensity ranges between 1 and 5 [23].

2. *Emotion covariance* FE_2 is the connection between different emotion categories. Ren *et al.* [23] employed Spearman's rank correlation coefficient for every two emotion categories in continuous L blogs, and plotted the mean of emotion correlation coefficients as emotion covariance value.

3. *Emotion transition* FE_3 depicts the patterns in emotion changes. Ren *et al.* [23] deployed an emotion transition matrix, where the row entries correspond to emotions in the previous time point, and the column entries corresponds to emotions in the current time point. It counted the mean of emotion transition matrices within ten continuous blogs as the emotion transition value.

In this study, L is set to be 1000.

4. SUICIDE RISK DETECTION USING FUZZY COGNITIVE MAP (FCM)

As described in previous section, we construct a 6-dimensional microblog feature space ($\overline{FS}$, $\overline{FC}$, $\overline{FU}$, $\overline{FW}$, $\overline{FO}$, $\overline{FE}$) to capture users stress, self-concerns, suicide- related expressions, last words, social interaction, and emotional traits from microblogs. However, to what extent can these features be used to measure the suicide risk? To evaluate the relationships between them, we build a fuzzy cognitive map (FCM) classification model for suicide risk detection.

FCM is a fuzzy feedback dynamic system with a strong capability of fuzzy reasoning. Fuzzy technology provides a framework for modeling the interface between human conceptual categories and data. Choosing FCM as classification model for suicide risk detection is because it is suitable for text classification with small amount of data and uncertain topics, due to its numerical reasoning and emphasizing feedback [49]. In addition, it also describes the causal relationships between categories and between features, well explaining why some instances have high or low eigenvalues.

FCM is mainly composed of concept nodes and directed edges among them. Each directed edge has a weight denoting the degree of the causal relationship between two connected concept nodes. Given a digraph D, let $C = \{c_1, c_2, \ldots, c_n\}$ be a set of concept nodes denoting vertexes in D, and $E \quad (c_i, c_j) \rightarrow w_{ij}$ be a set of mappings, $w_{ij} \in E$, $c_i, c_j \in C$, where w_{ij} denotes the causal effect degree of c_i on c_j. $E(C \times C) = (w_{ij})_{n\times n}$ is a connection matrix of D. Let $X \quad c_i \rightarrow x_i$ be a mapping function, where x_i corresponds to the status of concept node c_i, $x_i(t)$ gives the status value of c_i at time point t, and $x_i(t+1)$ gives the status value of c_i at time point $t+1$. In FCM model, the reasoning rule of concept nodes at anytime t can be formulated by:

$$x_i(t+1) = f\left(\sum_{i=1}^{n}, i \neq j w_{ij} \cdot x_j(t)\right) \tag{1}$$

where f is an activation function to guarantee output values are mapped into [0, 1].

The dynamic feedback mechanism of FCM provides a theoretical basis to our FCM classification model. As shown in figure 6, the FCM classification model for suicide risk detection has two kinds of concept nodes: attribute nodes and label nodes. The former corresponds to the six-dimensional microblog features, and the latter represents the final detection results: suicide and non-suicide. During the classification procedure, the values of attribute nodes remain stable, and the values of label nodes keep changing and iterating along with the reasoning process evolving, until reaching to a convergence status, and then getting a certain category.

We employ the reasoning rule proposed by [50]. Let $L = \{l_1, l_2, \ldots, l_m\}$ be a set of label nodes. The modified reasoning rule is defined as:

$$L_i^{t+1} = f\left(L_i^t + \sum_{i=1, i\neq j}^{k} w_{ij} \cdot L_j^t\right) \tag{2}$$

where l_i^t is the status value of label node l_i at time point t.

Activation function. The six-dimensional microblog features belong to different scopes, not suitable for comparing to each other in the same domain. To constrain these feature values into a same range and to achieve comparability among them, we adopt an improved Sigmoid function as the activation function [51]. The modified Sigmoid function can compensate the great disparity in weight and enhance the comparability as comparing a feature with higher value to a feature with lower value. It can be formulated by:

$$f(g) = \frac{1}{1 + e^{-c(g_i - m_i)t_i}} \tag{3}$$

where g_i denotes the value of the ith attribute node, m_i is the median value in the range of the th attribute. Moreover, $t_i = \frac{R_{\max}}{R_i}$, where R_i is the difference between the maximum and minimum value in the range of the ith attribute, $R_{\max}$ is the max value among $\{R_1, \ldots, R_n\}$. Parameter c is used to determine the curve slope. Then, attributes with different scopes are evenly distributed into [0, 1].

Objective function. To evaluate whether a classifying procedure reaches a steady state, we draw an error mechanism into the FCM classification model as an objective function. As shown in Figure 6, by testing the error feedback of objective function, the FCM classification model can estimate when to terminate the classifying procedure and to get a stable result based on a threshold. We define the objective function as:

$$Error(w) = \frac{1}{2}\sum_{t=1}^{T}\sum_{m=1}^{M}(L'_m(t) - L_m(t))^2 \tag{4}$$

where $L'_m(t)$ is the output value of the tth iteration of the mth label node, $L_m(t)$ is the true status value of this node, T and M correspond to the number of iterations and the number of label nodes, respectively. When objective function value $Error(w)$ is less than a small enough number ε, the model converges to a stable state, then the algorithm terminates. In this paper, ε is set to be 0.01.

Weight of the connection matrix. The FCM model constantly adjusts the weight of the connection matrix to make the value of label node infinitely close to or equal to its true values. To learn the weight of the connection matrix, we employ an improved genetic algorithm. Let the connection matrix $E(C \times C) = (w_{ij})_{n\times n}$ convert to a chromosome matrix $\hat{E} = w'_{11}, w'_{12}, \ldots, w'_{1n}, w'_{21}, w'_{nn}$. For a given population of chromosomes, each individual's fitness is calculated, and then those with higher fitness are selected for further crossover and mutation. In this paper, an adaptive crossover mutation operator is used to make the crossover and mutation probabilities automatically change with fitness. Let

$$F_{dif} = |Fitness_{high} - Fitness_{low}| \tag{5}$$

where $Fitness_{high}$ and $Fitness_{low}$ denote the average value of individuals greater than and less than the average fitness of the population. Then adaptive crossover mutation operator can be formulated by:

$$P_{ci} = k_1 \cdot \left(\frac{1}{1 + e^{(-F_{dif})}} + \frac{1}{2}\right) \tag{6}$$

$$P_{mi} = 1 - \frac{1}{1 + e^{(k_2 \cdot F_{dif})}} \tag{7}$$

where P_{ci} and P_{mi} represent the crossover and mutation probabilities of the ith population, respectively. k_1, k_2 are parameters, with values of 0.8 and -80.

5. PERFORMANCE EVALUATION

5.1 Dataset

We extract 130 users' blogs from Sina Weibo, including 65 suicide users and 65 non-suicide users, where suicide users are verified by news, and non-suicide users with no suicide intentions are confirmed by 3 psychological college students. The suicide users are aged from 13 to 34 years old, and the proportion of 15 to 34 year old users is 98.15%. The distributions of female and male suicides are relatively even, with 53.70% female suicides and 46.30% male suicides. A suicide user posted 5102 blogs maximally, 130 minimally, and 779 averagely. A non-suicide user posted 3581 blogs maximally, 105 minimally, and 678 averagely. The longest posting period is up to 5 years and 6 months, and the shortest posting period is 1 year. We randomly select 60% from suicides and non-suicides respectively for model training, and the rest 40% for testing. The 6-dimensional microblog feature space for each user is computed and initialized to prepare for suicide-or-not classification.

5.2 Basic Experiment

Taking the well-known Decision Tree, Naive Bayesian, Random Forest, and SVM as the baseline methods, we compare and evaluate the performance of the FCM classifier in suicide risk detection.

All six groups of features are taken into account in this section. The basic parameters for stressful interval extraction are set to $\delta = 7$ days and $\tau = 80\%$. To tune the weights of the connection matrix for FCM classifier, we adopt genetic algorithm and process multiple iterations (from 200 to 2000 iterations). To avoid over-fitting or under-fitting, we use 10-fold cross validation for all classifiers.

Table 4 lists the general performance for all classification methods. The proposed FCM model achieves the best performance with a precision of 80.8%, recall of 86.4% and F1-measure of 83.5%. It achieves improvement in precision, recall and F1measure with 9.7%, 15.8% and 13% respectively over second-placed Random Forest, indicating the proposed model reaches better performance on small data sets. In baseline classification methods, Random Forest achieves a relatively high performance with 71.1% precision, 70.6% recall and 70.5% F1-measure. Decision Tree shows similar performance as Naive Bayesian, while SVM works not well compared with other models.

5.3 Influence of Feature Vectors on Suicide Risk Detection

To evaluate the contribution of each feature vector group, we conduct tests on FCM classification model by removing one of the six feature vectors respectively: stress feature vector ($\overline{FS}$), self-concern feature vector ($\overline{FC}$), suicide-related expression feature vector ($\overline{FU}$), last word feature vector ($\overline{FW}$), social interaction feature vector ($\overline{FO}$), and emotion feature vector ($\overline{FE}$). Figure 7 illustrates that the classification performance without stress feature vector decreases most, with precision, recall and F1-measure decreased by 18.5%, 17.7% and 18.2%, respectively, followed by last word and suicide-related expression feature vectors with F1-measure decreased by 13.8% and 12.1% respectively during suicide risk prediction. Self-concern and social interaction feature vectors have relatively smaller effects on suicide risk detection, with F1-measure decreased by 2.3% and 5.7% respectively. This result demonstrates that stress feature vector plays more important role than other feature vectors to suicide risk detection.

To further evaluate the influence of every single feature, we calculate information gain for each feature. As shown in Figure 8, top five features with the highest information gain include root-mean-square deviation (IR), mean stress level (IL), proportion of suicidal-blogs (NSU), proportion of @-mention (NA), and peak stress level (IS), with the value of 0.65, 0.64, 0.61, 0.58, and 0.54 respectively. Three features of stress feature vector are on the top five list. Besides, the average information gain of each feature vector group are 0.52 for $\overline{FS}$, 0.24 for $\overline{FC}$, 0.46 for $\overline{FU}$, 0.44 for $\overline{FW}$, 0.25 for $\overline{FO}$, and 0.43 for $\overline{FE}$, where stress feature vector ($\overline{FS}$) has the maximum value. These results could supplement the explanation of why stress feature vector plays the most significant role in detecting suicide risk on microblog. The results also prove that among all features used for suicide detection, root-mean-square deviation and mean stress level are more important than other features, probably matching a view that the volatility and level of stress are good predictors for suicide prediction.

5.4 Influence of Parameter Settings

To assess the effect of stressful interval selection on suicide risk detection performance, we conducted a comparative experiment for two parameters of stress interval: the proportion of stressful time units τ and the temporal span δ.

First, the effect of different values of τ on detection performance is tested. Keep δ constant at 7 days, and then increase τ from 0.1 to 1, with an increment value of 0.1. The result in Figure 9 (a) shows that when τ is 0.8, the performance of detection result achieves the most effective, with F1-measure of 0.835. When τ is 0.6, the performance of detection result reaches the minimum, with F1-measure of 0.46. We also find that the performance of detection result improves along with the increase of τ from 0.6 to the summit (τ at value of 0.8), and then gradually declines. This is probably because with the increase of τ, more stressful posts in microblog will be captured in stressful intervals, thereby improving the accuracy of detection result. However, when τ continues to increase, many low-density stressful intervals will be left out, missing many stressful posts and then reducing the performance of detection.

Next, we test the effect of different values of δ on the performance of suicide detection. Similarly, we keep τ constant at 0.8, and then increase δ from 7 to 42 days, with an increment value

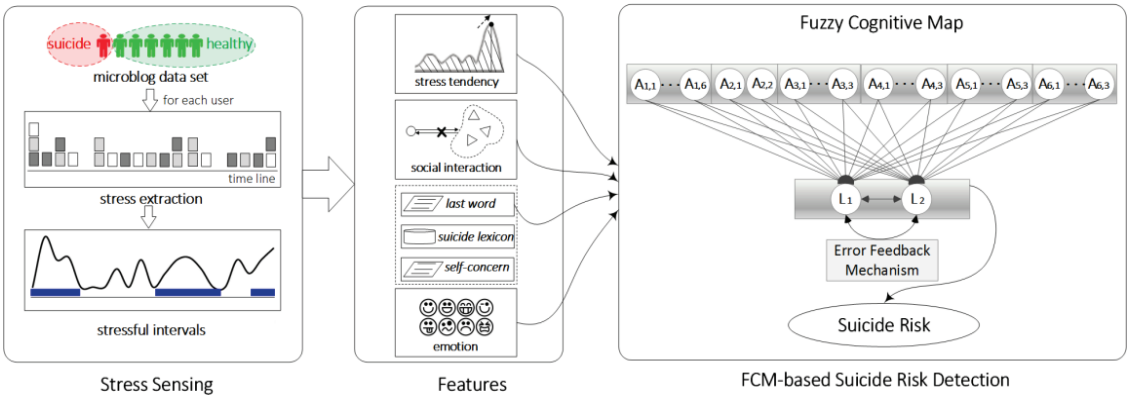


Figure 6 The FCM classification model, where nodes A_{11}, A_{12}... represent microblog features, and L_1 and L_2represent suicide risk or not.

Table 4 Results of Basic Experiments.

	FCM	Decision Tree	Naïve Bayesian	Random Forest	SVM
Precision	80.80%	67.90%	73.20%	71.10%	63.40%
Recall	86.40%	67.90%	68.80%	70.60%	63.30%
F1-Measure	83.50%	67.90%	67.10%	70.50%	63.20%

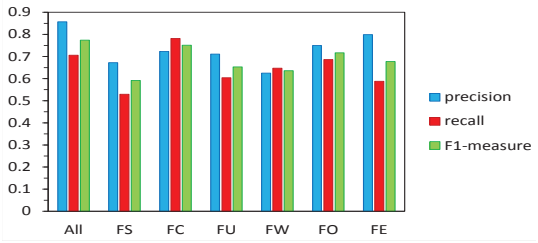


Figure 7 Effect of feature vectors on suicide risk detection.

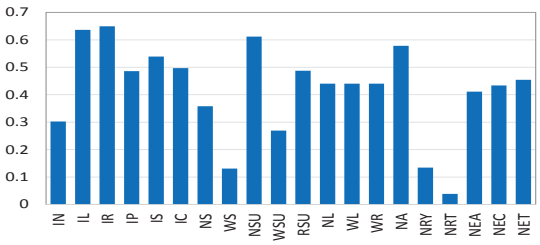
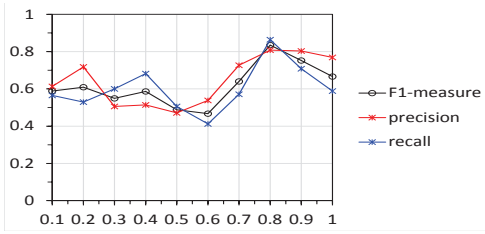


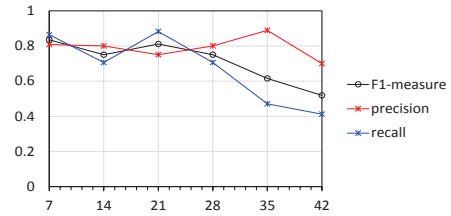
Figure 8 Impact of each feature on suicide risk detection, where IN denotes stressful interval number, IL denotes mean stress level, IR denotes RMS deviation, IP denotes peak stress level, IS denotes span length, IC denotes stress category number, NS denotes self-concern blogs ratio, WS denotes self-concern words number, NSU denotes suicidal blogs ration, WSU denotes suicidal words number, RSU denotes suicidal word ratio, NL denotes last words blog ratio, WL denotes last words number, WR denotes last words ratio, NA denotes @-mention ratio, NRY denotes @-reply ratio, NRT denotes forwarding ratio, NEA denotes emotion accumulation, NEC denotes emotion covariance, and NET denotes emotion transition.

of 7. The result in Figure 9 (b) shows that when δ is 7, the performance of detection result achieves the best, with F1-measure of 0.835. When δ is 7, 14 and 21, F1-measure changes within a reasonable range. However, as δ is beyond 21, a clear downward trend occurs in F1-measure value. By observing users' stressful interval span feature (shown in Figure 1 (e)), we find that about 84% suicide users' average stressful interval spans are between 7-21 days, only 9% more than 21 days. Therefore, when δ is set to be more than 21 days, most users' stressful intervals will be left out for their short-term, missing the effective information and resulting in a reduced detection accuracy.

In this study, τ and δ are set to be 0.8 and 7 days, respectively.



(a) Performance change with proportion of stressful time units (τ), where x-axes denotes τ



(b) Performance change with temporal span (δ), where x-axes denotes δ

Figure 9 Comparison of different proportion of stressful time units and temporal span in stressful interval.

6. DISCUSSION

Suicide is one of the most complex and difficult human behaviors to understand. Suicidal people find their problems to be over-

whelming to the point that suicide seems to be the best solution even though they don't necessarily want to die. Due to the great loss caused by suicide, this study makes an effort to analyze one's risk of suicidal ideation or suicide through his/her behaviors on the social media microblog. While the experimental result appears promising, there still remain a number of challenges to be tackled in the further work.

6.1 Considering both Suicide Risk Factors and Protective Factors

The study focuses on examining negative risk factors like stressful intervals, negative emotions, last words, etc. that tend to increase one's suicide risk. Some protective factors, which make it less likely someone will engage in suicidal ideation or behavior, are not incorporated into the analysis approach. For example, coming holidays, friends' visits, and personality revealed through the microblog etc. could help alleviate one's suicide ideation.

6.2 Examining Traces for Result Measurement

The experimental data encloses ones who have completed suicide, making the performance evaluation easy. Confronted with people who have not committed suicide, measuring their suicide risks generated by the approach is hard. One way for doing this could be checking and matching the traces and tendencies of ones' behaviors which may finally lead to final suicide action on the microblog. This will constitute another research question worthy investigation.

7. CONCLUSION

In this paper, we study the feasibility and effectiveness of suicide risk detection using fuzzy cognitive map and microblogs. A 6-dimensional microblog feature space is constructed to capture one's stress, self-concerns, suicide-related expressions, last words, social interaction, and emotional traits throughout the posting period on microblogs. We examine the differences of these features between the suicide group and the non-suicide group, through a set of real online microblogs posted by those who committed suicide and those who have no suicide intention. We then describe a fuzzy cognitive map (FCM) classification model upon the microblog feature space to detect users' suicide risk. The experimental results show that the FCM classifier model outperforms other machine learning methods (Decision Tree, Naive Bayesian, Random Forest, and SVM). The results also prove that all of the extracted features contribute to the suicide detection, and among them stress-related features play the most significant role than the rest of features.

ACKNOWLEDGMENTS

The work is supported by National Natural Science Foundation of China (61872214, 61532015, 61521002), Chinese Major State Basic Research Development 973 Program (2015CB352301), and and Open Research Fund Program of State key Laboratory of Hydroscience and Engineering (sklhse-2017-A-05).

REFERENCES

1. L. Wang, M. Phillips, Z. Huang, Y. Zhang, Y. Zhao, and G. Yang, "Evaluation on the accuracy of reported suicides in the Chinese population," *J. of Epidemiology*, pp. 889–892, 2003.
2. "Suicide rates rise sharply in U.S." TARAPARKER-POPE, 2013.
3. L. Panand Z. Wang, "Domestic research situation about suicide in china," *J. of Behavioral Medical Science*, pp. 669–670, 2005.
4. The Guardian, "Chambers a japan: ending the culture of the 'honourable' suicide," 2010.
5. A. Barak and O. Miron, "Writing characteristics of suicidal people on the internet: A psychological investigation of emerging social environments," pp. 507–524, 2005.
6. M. Tim, C. Michael, S. Paul, and W. Paul, "Temporal and computerized psycholinguistic analysis of the blog of a Chinese adolescent suicide," pp. 168–175, 2014.
7. K. Fu, Q. Cheng, P. Wong, and P. Yip, "Responses to a self-presented suicide attempt in social media: Asocial network analysis," p. 406, 2013.
8. M. Marcinczuk, M. Zasko-Zielinska, and M. Piasecki, "Structure annotation in the polish corpus of suicide notes," in *Proc. of TSD*, 2011, pp. 419–426.
9. J. Pestian, H. Nasrallah, P. Matykiewicz, A. Bennett, and A. Leenaars, "Suicide note classification using natural language processing: A content analysis," pp. 19–28, 2010.
10. B. Desmet and V. Hoste, "Emotion detection in suicide notes," p. 6351C6358, 2013.
11. V. Silenzio, P. Duberstein, W. Tang, N. Lu, X. Tu, and C. Homan, "Connecting the invisible dots: Reaching lesbian, gay, and bisexual adolescents and young adults at risk for suicide through online social networks," p. 469C474, 2009.
12. Y. Huang, T. Goh, and C. Liew, "Hunting suicide notes in web 2.0 preliminary findings," in *Proc. of ISMW*, 2007, pp. 517–521.
13. M. Choudhury, E. Kiciman, and M. Dredze, "Discovering shifts to suicidal ideation from mental health content in social media," in *Proc. of CHI*, 2016, pp. 2098–2110.
14. M. Tim, C. Ben, C. Michael, W. Paul, and S. Paul, "Collective intelligence for suicide surveillance in web forums," pp. 29–37, 2013.
15. N. Masuda, I. Kurahashi, and H. Onari, "Suicide ideation of individuals in online social networks," 2014.
16. H. Sueki, "The association of suicide-related twitter use with suicidal behavior : Across-sectional study of young internet users in japan," vol. 170, p. 155C160, 2015.
17. J. Jashinsky, S. Burton, C. Hanson, J. West, C. Giraud-Carrier, M. Barnes, and T. Argyle, "Tracking suicide risk factors through twitter in the US," pp. 51–59, 2014.
18. L. Guan, B. Hao, and T. Zhu, "How did the suicide act and speak differently online? behavioral and linguistic features of china's suicide microblog users," pp. 994–1008, 2014.
19. B. O'Dea, S. Wan, P. Batterham, A. Calear, C. Paris, and H. Christensen, "Detecting suicidality on twitter," pp. 183–188, 2015.
20. X. Huang, L. Zhang, T. Liu, D. Chiu, T. Zhu, and X. Li, "Detecting suicidal ideation in Chinese microblogs with psychological lexicons," in *Proc. of UIC-ATC-SCALCOM*, 2014, pp. 844–849.
21. X. Huang, X. Li, L. Zhang, T. Liu, D. Chiu, and T. Zhu, "Topic model for identifying suicidal ideation in Chinese microblog," in *Proc. of the 29th Pacific Asia Conf. on Language, Information and*

Computation, 2015, pp. 553–562.
22. P. Burnap, G. Colombo, and J. Scourfield, "Machine classification and analysis of suicide-related communication on twitter," in *Proc. of the ACM Conf. on Hypertext & Social Media*, 2015, pp. 75–84.
23. F. Ren, X. Kang, and C. Quan, "Examining accumulated emotional traits in suicide blogs with an emotion topic model," *J. of Biomedical and Health Informatics*, pp. 1384 – 1396, 2016.
24. Y. Xue, Q. Li, L. Feng, G. Clifford, and D. Clifton, "Towards a microblog platform for sensing and easing adolescent psychological pressures," in *Proc. of Ubicomp*, 2013.
25. Y. Xue, Q. Li, L. Jin, L. Feng, D. Clifton, and G. Clifford, "Detecting adolescent psychological pressures from micro-blog," in *Proc. of HIS*, 2014.
26. W. Li, Y. Li, and Y. Wang, "Chinese microblog sentiment analysis based on sentiment features," in *Proc. of the Conf. on Web Technologies and Applications*, 2016, pp. 385–388.
27. A. Coronato and G. Paragliola, An anomolous situation detection system for cognitive impaired people, International Journal of Computer Systems Science and Engineering, vol. 30, no. 1, 2015.
28. C. Choi, J. Choi, and P. Kim, Abnormal behaviour pattern mining for unknown threat detection, International Journal of Computer Systems Science and Engineering, vol. 32, no. 2, 2017.
29. A. Rich and R. Bonner, "Concurrent validity of a stress-vulnerability model of suicide ideation and behavior: a follow-up study," pp. 265–270, 1987.
30. V. Kraiij, E. Arensman, and P. Spinhoven, "Negative life events and depression in elderly person: Ameta-analysis," *J. of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 57, pp. 87–94, 2005.
31. S. Kwok and D. Shek, "Social problem solving, family functioning, and suicidal ideation among Chinese adolescentsin-hongkong,"vol.44, no. 174, pp. 391–406, 2009.
32. J. Luo, L. Yang, W. Zhang, Y. Wang, and K. Jiang, "A study on suicide attitude, suicide ideation, life event and coping style of college students," *J. of Chinese Medical Ethics*, pp. 57–59, 2004.
33. T. Yang, *Healty Behavior Theory and Research*. People's Medical Publishing House, 2007.
34. A. Zautra, J. Reich, and M. Davis, "The role of stressful events in the relationship between positive and negative affects: Evidence from field and experimental studies," pp. 927–951, 2000.
35. D. Wasserman, "Suicide: An unnecessary death," pp. 189–194, 2003.
36. L. Zhang, X. Huang, T. Liu, A. Li, Z. Chen, and T. Zhu, "Using linguistic features to estimate suicide probability of Chinese microblog users," in *Proc. of HCC*, 2015, pp. 549–559.
37. A. Greco, G. Valenza, A. Lanata, G. Rota, and E. Scilingo, "Electrodermal activity in bipolar patients during affective elicitation," *J. of Biomedical and Health Informatics*, pp. 1865 – 1873, 2014.
38. Q. Xu, T. New, and C. Guan, "Cluster-based analysis for personalized stress evaluation using physiological signals," *J. of Biomedical and Health Informatics*, pp. 275–281, 2015.
39. Q. Li, Y. Xue, J. Jia, and L. Feng, "Helping teenagers relieve psychological pressures: a micro-blog based system," in *Proc. of EDBT*, 2014, pp. 660–663.
40. H. Lin, J. Jia, Q. Guo, Y. Xue, Q. Li, J. Huang, L. Cai, and L. Feng, "User-level psychological stress detection from social media using deep neural network," in *Proceedings of the ACM International Conference on Multimedia, MM'14, Orlando, FL, USA, November 03 -07, 2014*, 2014, pp. 507–516.
41. H. Lin, J. Jia, Q. Guo, Y. Xue, and etal., "Psychological stress detection from cross-media microblog data using deep sparse neural network," in *Proc. of ICME*, 2014, pp. 1–6.
42. L. Zhao, J. Jia, and L. Feng, "A category-dependent time-aware feature space for teenagers' stress detection from micro-blog," in *Proc. of the 4th IFIP Intl. Conf. on Artificial Intelligence in Theory and Practice*, 2015.
43. E. White and L. Mazlack, "Discovering causality in suicide notes using fuzzy cognitive maps," in *Proc. of the Conf. on Midwest Artificial Intelligence and Cognitive Science Conference*, 2011, p. 142.
44. M. Choudhury, S. Counts, and E. Horvitz, "Predicting postpartum changes in emotion and behavior via social media," in *Proc. of the Conf. on Human factors in computing systems*.
45. J. Sabbath, "The suicidal adolescent: The expendable child," pp. 272– 285, 1969.
46. M. Lv, A. Li, T. Liu, and T. Zhu, "Creating a Chinese suicide dictionary for identifying suicide risk on social media," p. https://peerj.com/articles/1455/, 2015.
47. F. News, http://news.66163.com/2012-02-17/599138.shtml, 2012.
48. L. Mandelli, F. Nearchou, C. Vaiopoulos, C. Stefanis, S. Vitoratou, A. Serretti, and N. Stefanis, "Neuroticism, social network, stressful life events: Association with mood disorders, depressive symptoms and suicidal ideation in a community sample of women," p. 38C44, 2014.
49. G. Zhang, Y. Liu, and Y. Wang, "Reference algorithm of text categorization based on fuzzy cognitive maps," p. 155, 2007.
50. D. Zhu, B. Mendis, and T. Gedeon, "A hybrid fuzzy approach for human eye gaze pattern recognition," in *Proc. of the Conf. on Neural Information Processing of the Asia-PPacific Neural Network Assembly*, 2008, p. 655.
51. Y. Zhang, "The modeling and control of dynamic system based on fuzzy cognitive maps," Ph.D. dissertation, Dalian University of Technology, China, 2012.