# Building and Using Personal Knowledge Graph to Improve Suicidal Ideation Detection on Social Media

Lei Cao ⓘ, Huijun Zhang, and Ling Feng

*Abstract*—A large number of individuals are suffering from suicidal ideation in the world. There are a number of causes behind why an individual might suffer from suicidal ideation. As the most popular platform for self-expression, emotion release, and personal interaction, individuals may exhibit a number of symptoms of suicidal ideation on social media. Nevertheless, challenges from both data and knowledge aspects remain as obstacles, constraining the social media-based detection performance. Data implicitness and sparsity make it difficult to discover the inner true intentions of individuals based on their posts. Inspired by psychological studies, we build and unify a high-level suicide-oriented knowledge graph with deep neural networks for suicidal ideation detection on social media. We further design a two-layered attention mechanism to explicitly reason and establish key risk factors to individual's suicidal ideation. The performance study on microblog and Reddit shows that: 1) with the constructed personal knowledge graph, the social media-based suicidal ideation detection can achieve over 93% accuracy; and 2) among the six categories of personal factors, *post, personality,* and *experience* are the top-3 key indicators. Under these categories, *posted text*, *stress level*, *stress duration*, *posted image*, and *ruminant thinking* contribute to one's suicidal ideation detection.

*Index Terms*—Personal knowledge graph, social interaction, social media, suicidal ideation detection.

## I. INTRODUCTION

SUICIDAL ideation is a problem which is on the rise across all the nations in the world. According to different reports and statistical figures, the lifetime prevalence of suicidal ideation of the entire world population is around 9 per cent [1], and this figure is significantly higher for individuals in the age group of 18 to 25 years old. There are a number of causes behind why an individual might suffer from suicidal ideation, and involvement of personality traits in susceptibility to suicide has been investigated for a long time. However, due to the limitations of mass group reaching and the diversity of conceptual and methodological approaches, the extent of their independent contributions remain is still have been difficult to establish. The risk factors for the immediate precursor to suicidal ideation are still not well-known especially in developing countries [1]. In any circumstance, for those with suicidal ideation, the earlier they can be detected, the better the chance of suicide prevention [2].

In the literature, psychologists have developed a number of suicide risk measurements (such as Suicide Probability Scale [3], Adult Suicide Ideation Questionnaire [4], Suicidal Affect-Behavior-Cognition Scale [5], etc.) to assess individual's suicidal ideation. As this kind of methods requires people to either fill in a subjective questionnaire or participate in a professional interview, it is only applicable to a small group of people. For those who are suffering but tend to hide inmost thoughts and refuse to seek helps from others, the approach cannot work [6]–[9].

With social media (like Twitter, online forums, and microblogs) becoming an integral part of daily lives nowadays, more and more people go to social media for information acquisition, self-expression, emotion release, and personal interaction. The large-scale, low-cost, and open advantages of social media offer us the unprecedented opportunity to capture individual's symptoms and traits related to suicidal ideation [10]–[16]. Nevertheless, analysis and detection of individual's suicidal ideation through social media are not trivial, facing a number of challenges from the following two aspects.

### A. Data-Aspect Challenges

Data implicitness and data sparsity are two critical data-aspect problems, challenging social media-based solutions for a long time.

**Data Implicitness.** Due to the unique equality, freedom, fragmentation, and individuality characteristics of social media, users linguistic and visual expressions on social media are implicit, reserved, and even anti-real. To illustrate, let's compare users' normal posts with their commenting posts in a hidden *microblog tree hole*. Here, a microblog tree hole refers to a microblog space, whose author has committed suicide, and other microblog users usually with suicidal ideation tend to share and post their inner real feelings and thoughts as comments under the last post of the passed one. One such a tree hole on Sina Weibo (the largest microblog platform in China) has collected over 1700000 commenting posts from more than 160000 suicide attempts under the last post of microblog user called *Zoufan*, who committed suicide due to depression on March 17, 2012.

| User | Normal Posts | Hidden Tree Hole Posts |
|------|--------------|------------------------|
| 1 | 2018-06-15 12:31:53<br>"What to buy after getting the salary?"<br><br>2018-06-18 21:45:22<br>"I sang a song today. It's ugly. Everyone laughed at me. Who cares?"<br><br>2018-06-19 08:12:13<br>"Ohh! Today is a new day. Good things happen one by one." | 2018-06-17 17:11:05<br>"In fact, I am dead, I am a worthless person. My work is a failure."<br><br>2018-06-18 19:31:27<br>"No place to go, timid, no matter where I go, I am still like this! Incurable! The place I want to go to heaven."<br><br>2018-06-19 05:45:00<br>"No meaning to live. I am like a corpse, who can understand me?" |
| 2 | 2019-05-25 22:15:10<br>"My birthday! Thanks for the happy moment!"<br><br>2019-05-26 07:00:00<br>"Breakfast is delicious!" | 2019-05-25 23:55:15<br>"Today is my birthday, but I seem to have passed away."<br><br>2019-05-26 01:29:11<br>"I want to kill myself, living with my parents without a job. So hard to find a satisfactory one." |

Fig. 1.    Two users' normal posts versus their hidden tree hole posts on Sina Weibo.

TABLE I
STATISTICS OF USERS' NORMAL POSTS VS. HIDDEN TREE HOLE POSTS ON SINA WEIBO BASED ON 3652 USERS FROM MAY 1, 2018 TO APRIL 30, 2019

| Perspective | Normal Posts | Hidden Tree Hole Posts |
|-------------|--------------|------------------------|
| Prop. of posts containing *self-concern* words | 43% | 50% |
| Prop. of posts containing *others-concern* words | 8% | 6% |
| Prop. of posts containing *suicide-related* words | 31% | 95% |
| Prop. of *self-concern* words per post | 4% | 9% |
| Prop. of *others-concern* words per post | 8% | 1% |
| Prop. of *suicide-related* words per post | 0.02% | 0.3% |
| Total post number | 252,901 | 190,087 |

Fig. 1 lists two users' example posts on microblog. From their normal posts, we can sense a joyful feeling. However, if reading their commenting posts in the hidden tree hole, we are touched by their severe hopelessness and suicidal thoughts.

To further examine the differences between users' normal posts and their commenting posts in the hidden tree hole, we crawled and analyzed 3652 individuals' posts from May 1, 2018 to April 30, 2019 on Sina Weibo. As Table I shows, their posting contents were quite different. The hidden tree hole posts focused more on oneself through more self-concern words (like "I," "me," "my," "mine," "am," etc.) rather than others-concern words, and they used more suicide-related words than the normal posts, demonstrating users' willingness and directives to express their inner suicidal thoughts in the hidden tree hole. On the contrary, the users were reluctant to show their feelings in the open normal posts, and they used more others-concern (such as

"they," "their," "she," "he," etc.) words than the tree hole posts. Such phenomena challenge the suicide risk detection through the normal posts.

**Data Sparsity.** Apart from data implicitness, subjective to habits, characters, emotions, etc., some people may not want to express themselves actively on social media. Particularly, when immersed in hopelessness, loneliness, and suicidal thoughts, people tend to be self-enclosed and have very low motivations to write or leave something on social media.

To illustrate, we identified 3652 users which made both normal posts and hidden tree hole posts on Sina Weibo, and their tree hole posts contained suicide-related words. We hypothesized that this group of users have suicidal ideation. For comparison purpose, we also randomly crawled another 3677 active users, who only posted normal blogs on microblog, and their blogs contained no suicide-related words at all. We called this group of users *ordinary users*. From Fig. 2, we can find that users with suicidal thoughts made much less normal posts than ordinary users. Overall, the majority of this group of users had less than 1 normal post per week, verifying the serious data sparsity problem.

Moreover, due to violent emotional fluctuations, individuals with suicidal ideation tend to delete their previous posts related to suicidal ideation, striving to hide the true inner intentions. We counted the number of normal posts from the 7329 users (3652 with suicidal ideation and 3677 ordinary users) in one year from 2018.05.01 to 2019.04.30 twice (one on 2019.04.30, and the other on 2019.08.31), and discovered that (1) around 45.81%

TABLE II
STATISTICS OF DELETED NORMAL POST NUMBERS BASED ON 7329 USERS (3652 WITH SUICIDAL IDEATION AND 3677 NORMAL USERS) FROM MAY 1, 2018 TO APRIL 30, 2019 ON SINA WEIBO

|  | # Users with Suicidal Ideation | # Ordinary Users |
|---|---|---|
| Number of users who deleted their previous normal posts after 3 months | 1,673 | 326 |
| Proportion of users who deleted their previous normal posts after 3 months | 45.81% | 8.87% |
| Proportion of deleted normal posts after 3 months | 35.70% | 3.95% |
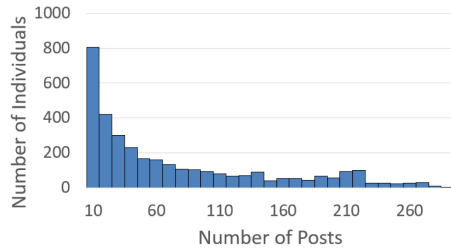


(a) Users with suicidal ideation
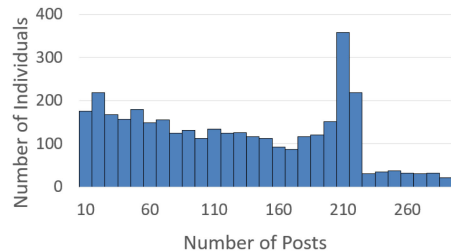
(b) Ordinary users

Fig. 2. Distributions of normal post numbers from 7329 users (3652 with suicidal ideation and 3677 ordinary users) from May 1, 2018 to April 30, 2019 on Sina Weibo.



Fig. 3. Examples of normal posts by user 1 and his two followed users on Sina Weibo.

of the users with suicidal ideation deleted their previous normal posts after 3 months, while only 8.87% of ordinary users did it. (2) around 35.70% of normal posts were deleted by the users with suicidal ideation. For ordinary users, the figure was just about 3.95%. In other words, individuals with suicidal feelings tend to delete much more normal posts than ordinary ones (Table II).

The above data-aspect challenges hinder the detection performance greatly.

**Our Work.** To address the data-aspect challenges, we drew inspirations from previous psychological studies that individuals tend to find and follow one's own kind in the networks, looking for similar people, comfort each other, find a way to get rid of the difficulties, or even commit suicide together [17]. In such an emotional sharing and resonance process, negative suicidal emotion and influence could be developed and fast propagated among potentially suicidal individuals [18]. As in Fig. 3, we can notice similar negative feelings from user 1 and his two followed friends. In our crawled 7329 users, each user follows 5 neighbour users (as friends) on average.

Motivated by the findings, we leveraged individual's friends information via the follower-following relationship on microblogs to enrich users data and cope with data implicitness and sparsity problems in suicidal ideation detection.
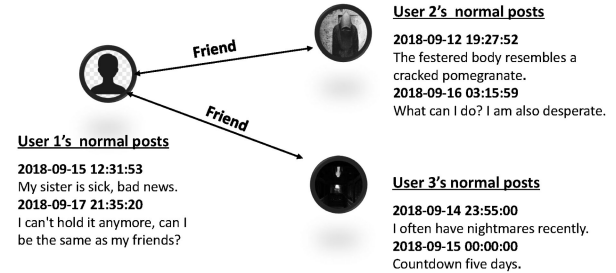
### B. Knowledge-Aspect Challenges

Although there has been a large body of research addressing the task of suicidal ideation based on social media, the detection performance remains to be constrained due to the disparity between the informal language used by social media users and the concepts defined by domain experts in medical knowledge bases [19]–[21]. Moreover, methods based on domain knowledge have been successful in many fields [22]–[24]. To fill the gap, recently, [19] incorporated domain specific knowledge into a learning framework to predict the severity of suicide risk for a Redditor user. Specifically, it employed two medical lexicons (TwADR-L and AskaPatient) to map social media contents to medical concepts [25], and identified contents with negative emotions based on anonymized and annotated suicide notes available from Informatics for Integrating Biology and the Bedside (i2b2) challenge. It further developed a suicide risk severity lexicon using medical knowledge bases and suicide ontology to detect cues relevant to suicidal thoughts and actions, and used language modeling, medical entity recognition and normalization and negation detection to create a gold standard dataset of 500 Redditors developed by four practicing psychiatrists [19].

The work [19] proved that utilizing domain specific knowledge is a promising approach for suicidal ideation analysis. In reality, there could be a large number of causes behind why an individual might suffer from suicidal ideation. It is of extreme importance that the suicidal ideation detection method should be able to recognize the warning signs or symptoms around the individual who is suffering, and involve them to improve the detection performance. However, few contribution measurements of different indicators in relation to individual's *personal information* (display image on social media, gender, age, etc.), *personality* (pursuing perfection, ruminant thinking, or interpersonal sensitivity), *experience* (stress or previous suicide attempt), negative emotion and emotion transition pattern, and social interaction impact have been established for suicidal ideation detection. In

the literature, so far there has no social media-based *personal* suicide-oriented knowledge framework being developed and incorporated into suicidal ideation analysis.

**Our Work.** Beyond looking at individual's series of posting contents and social interactions on social media [26], [27], we built a structured personal suicide-oriented knowledge graph, and instantiated the personal knowledge graph by applying deep learning techniques [28] to 7329 microblog users (3652 with suicidal ideation and 3677 without suicidal ideation) and 500 Redditors (108 without suicidal ideation, 99 with suicide indicator, 171 with suicidal ideation, 77 with suicide behavior and 45 with actual attempt).

We further built a two-layered attention mechanism (property attention and neighbour attention) to propagate model messages through the designed personal knowledge graph, and adaptively tune the weights (contributions) of different risk factors (properties and neighbour users) to individuals' suicidal ideation.

Finally, we incorporated and unified the achieved personal suicide-oriented knowledge graph with graph neural networks to facilitate suicidal ideation detection. Our experimental results on microblog and Reddit showed that with the personal suicide-oriented knowledge graph and the two attention mechanisms, suicidal ideation detection can achieve over 93% of accuracy and F1-measure. Among the six categories of personal factors, *posts, personality,* and *experience* are the top-3 key indicators. Under these categories, *posted texts*, *stress level*, *stress duration*, *posted image*, and *ruminant thinking* contribute the most to one's suicidal ideation detection.

In summary, the contributions of the study are three-fold.

- We build and unify a high-level suicide-oriented knowledge graph with graph neural networks for suicidal ideation detection on social media.
- We design a two-layered attention mechanism that explicitly reasons and establishes key risk factors to individual's suicidal ideation.
- We deliver a social media-based suicide-oriented knowledge graph. Together with the labeled dataset (https://github.com/bryant03/Sina-Weibo-Dataset) with 3652 (3677) users with (without) suicidal ideation from Sina Weibo, they could facilitate further suicide-related studies in the computer science and psychology fields.

## II. LITERATURE REVIEW ON DETECTING SUICIDAL IDEATION

### A. Traditional Questionnaire-Based Suicide Risk Assessment

Traditional methods rely on questionnaires and face-to-face diagnosis to assess whether one is in the risk of suicide [29]. Some widely used psychological measurements include Suicide Probability Scale (SPS) [3], Depression Anxiety Stress Scales-21 (DASS-21) [30], [31], Adult Suicide Ideation Questionnaire [4], Suicidal Affect-Behavior-Cognition Scale [5], functional Magnetic Resonance Imaging (fMRI) signatures [32], and so on. While this kind of approaches are professional and effective, they require respondents to either fill in a questionnaire or participate in a professional interview, constraining its touching to suicidal people who have low motivations to seek help from professionals [6]–[9]. A recent study found out that taking

a suicide assessment may bring negative effect to individuals with depressive symptoms [33].

### B. Suicide Risk Detection From Social Media

Recently, detection of suicide risk from social media is making great progress due to the advantages of reaching massive population, low-cost, and real-time [34]. [35] reported that suicidal users tend to spend a lot of time online, have great likelihood of developing online personal relationships, and great use of online forums.

**Suicide Risk Detection from Suicide Notes.** [36] built a suicide note classifier used machine learning techniques, which performs better than human psychologists in distinguishing fake online suicide notes from real ones. [37] hunted suicide notes based on lexicon-based keyword matching on MySpace.com (a popular site for adolescents and young adults, particularly sexual minority adolescents with over 1 billion registered users worldwide) to check whether users have an intent to commit suicide.

**Suicide Risk Detection from Community Forums.** [38] applied textual sentiment analysis and summarization techniques to users' posts and posts' comments in a Chinese web forum in order to identify suicide expressions. [39] examined online forums in Japan, and discovered that the number of communities which a user belongs to, the intransitivity, and the fraction of suicidal neighbors in the social network contributed the most to suicidal ideation. [40] built a logistic regression framework to analyze Reddit users' shift tendency from mental health sub-communities to a suicide support sub-community. heightened self-attentional focus, poor linguistic coherence and coordination with the community, reduced social engagement and manifestation of hopelessness, anxiety, impulsiveness and loneliness in shared contents are distinct markers characterizing these shifts. Based on the suicide lexicons detailing suicide indicator, suicidal ideation, suicide behavior, and suicide attempt, [11] built four corresponding semantic clusters to group semantically similar posts on Reddit and questions in a questionnaire together, and used the clusters to assess the aggregate suicide risk severity of a Reddit post. [19] used Reddit as the unobtrusive data source to conduct knowledge-aware assessment of severity of suicide risk for early intervention. Its performance study showed that convolutional neural network (CNN) provided the best performance due to the discriminative features and use of domain-specific knowledge resources, in comparison to SVM-L that has been used in the state-of-the-art tools over similar datasets. [41] introduced a new evaluation measure to move beyond suicide risk classification to a paradigm in which prioritization is the focus. The proposed joint ranking approach outperformed the logistic regression under the new evaluation measure. [42] used document embeddings generated by transfer learning to classify users at suicide risk or not. [43] employed an LSTM-CNN combined model to evaluate and compare to other classification models.

**Suicide Risk Detection from Blogs.** [44] used search keywords and phrases relevant to suicide risk factors to filter potential suicide-related tweets, and observed a strong correlation between Twitter-derived suicide data and real suicide data, showing

that Twitter can be viewed as a viable tool for real-time monitoring of suicide risk factors on a large scale. The correlation study between suicide-related tweets and suicidal behaviors was also conducted based on a cross-sectional survey [45], where participants answered a self-administered online questionnaire, containing questions about Twitter use, suicidal behaviour, depression and anxiety, and demographic characteristics. The survey result showed that Twitter logs could help identify suicidal young Internet users.

Based on eight basic emotion categories (joy, love, expectation, anxiety, sorrow, anger, hate, and surprise), [46] examined 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. 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.

Natural language processing and machine learning techniques, such as Latent Dirichlet Allocation (LDA), Logistic Regression, Random Forest, Support Vector Machine, Naive Bayes, Decision Tree, etc., were applied to identify users' suicidal ideation based on their linguistic contents and online behaviors on Sina Weibo [12], [47]–[50] and Twitter [51]–[57]. Deep learning based architectures like Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory Neural Network (LSTM), etc., were also exploited to detect users' suicide risk on social media [13]–[15], [58]. [16] detected users' change points in emotional well-being on Twitter through a martingale framework, which is widely used for change detection in data streams.

Very recently, [59] designed suicide-oriented work embeddings to capture the implicit suicidal expression and proposed a suicide risk detection model based on LSTM and ResNet. [27] applied the machine learning algorithms to detect suicidal profiles in Twitter based on features extracted from users accounts and tweeting behaviors. [26], [60] proposed a multipronged approach and implemented different neural network models such as sequential models and graph convolutional networks, which were trained on textual content shared in Twitter, tweeting histories of the users and social network formed between different users posting about suicidality.

Different from the previous work, in this study, we looked beyond individual's posting contents and social interactions on social media, and incorporated the constructed personal knowledge graph into suicidal ideation detection. The performance study showed that the proposed method can achieve over 93% of accuracy and F1-measure.

## III. FRAMEWORK

### A. Building the Personal Suicide-Oriented Knowledge Graph

We define the terminology for representing the nodes and edges of the personal suicide-oriented knowledge graph, and then convert the extracted data from individuals' social media
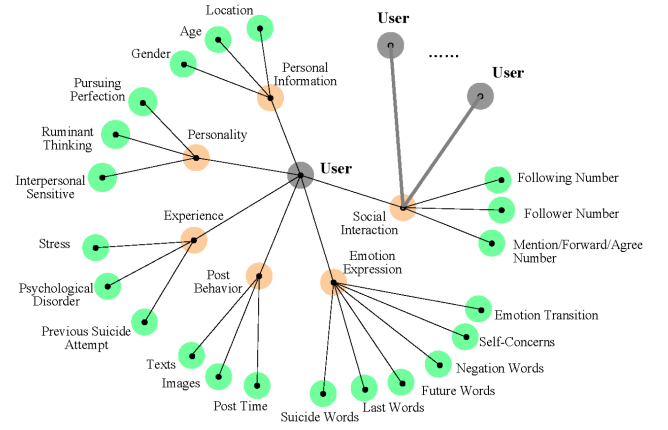


Fig. 4. Ontology of the social media-based suicide-oriented knowledge graph.

accounts to instantiate such a personal knowledge graph. Inspired by the psychological investigation into the predictors of suicide risk [61]–[67], we extracted and analyzed individuals' relevant social media behaviors from the following six perspectives (i.e., *personal information, personality, experience, post behavior, emotion expression,* and *social interaction*), as shown in Fig. 4.

*1) Personal Information (Gender, Age, Location):* Based on the previous study result that suffering women are three times more likely than suffering men to attempt suicide [61], we captured user $u$'s personal details including gender, age, and location from his microblog account, and involved them into the personal knowledge graph.

We used a 3-dimensional vector to describe user's gender information. $Gender(u) = (1,0,0), (0,1,0),$ or $(0,0,1)$, representing *female*, *male*, or *unknown*, respectively.

We used a scalar from zero to one to represent user's age. The scalar is obtained by dividing the user's age by the maximal user age in the collected dataset. $Age(u) = \frac{u.age}{max\_user\_age}$, where $max\_user\_age = 65$.

Like user's gender expression, we used a vector of length 8 to represent user $u$'s location. $Loc(u) = (1,0,0,\dots,0),$ $(0,1,0,\dots,0), (0,0,1,\dots,0), \cdots,$ or $(0,0,0,\dots,1)$. Each element of the vector represents a certain geographic direction. It takes value 1 if the user is in the corresponding location, and 0 otherwise. Totally 8 location directions (i.e., *east, south, north, south-west, north-west, middle, north-east*, plus an *unknown* location) were considered.

*2) Experience (Stress, Psychological Disorder, Previous Suicide Attempt):* Suicidal ideation often occurs when an individual is no longer able to cope with some kind of difficult or overwhelming situation [62]. Some of the commonly agreed causes of suicidal ideation include situations where an individual experiences continuous stressful periods of high stress levels and different stress categories [63]. To detect user's stress (stress periods, stress levels, and stress categories) within the year, we applied the algorithm [68] to user's posting behaviors, and captured a series of stressful periods $\mathcal{S}(u)$, each of which is of the form $s = (s_p, s_l, s_c)$, where $s \in \mathcal{S}(u)$, $s_p$ is a temporal stress period which should be over five days, $s_l$ is the stress level 1

| Type | Words |
|------|-------|
| Perfection-Related | perfect, perfectionist, perfection, stupid, wrong, upset, mood, struggle, goal, realistic, not enough, lose, loser, failure, imperfect, compete, force, negative, disappoint, despair, letdown, useless, always, never, still, upset, problem |
| Ruminant Thinking-Related | regret, repent, rue, penitent, confess, hate, self-blame, grievance, complaint, injustice, unfairness, growl, discontent, lose, loser, owe, sinner, die |

(*weak stress*) or 2 (*strong stress*), and $s_c$ is the stress category, which can be *study, work, family, interpersonal relation, romantic relation,* or *self-cognition*. We aggregated user $u$'s stressful periods into the total number of stress periods, the average stress level, and the number of different stress categories that user $u$ experienced throughout the year.

$$Stress(u) = (sNum(u), sLevel(u), sCatNum(u)),$$

where $sNum(u) = |\mathcal{S}(u)|$, $sLevel(u) = \frac{\sum_{s \in \mathcal{S}(u)} s_l}{|\mathcal{S}(u)|}$, and $sCatNum(u) = |\{s_c \mid s \in \mathcal{S}(u)\}|$.

Furthermore, people who are suffering from psychological disorders like depression or bipolar disorder are also more prone to suffering from suicidal ideation. People who previously tried to attempt suicide also tend to have a higher suicidal risk than those who did not. Like [69], we set $Disorder(u) = 1$, if and only if user $u$ posted something on microblog like *"I have depression/bipolar disorder ⋯"*, and 0 otherwise. $Attempt(u) = 1$, if and only if user $u$ posted something on microblog like *"I've committed suicide #number# times ⋯"*, and 0 otherwise.

*3) Personality (Pursuing Perfection, Ruminant Thinking, Interpersonal Sensitive):* Involvement of personality factors in susceptibility to suicidality has been the subject of research since the 1950 s. Psychological findings showed that personality dimensions are significantly associated with suicide-related behaviors, and specific personality factors impact suicide uniquely for each gender [64]. Such personality types like *pursuing perfection, ruminant thinking,* and *interpersonal sensitivity* could be useful markers of suicide risk [65].

To assess the extent of one's *pursuing perfection* characteristic, we identified 36 potential perfectionist users from the "Perfectionism" subreddit (https://www.reddit.com/r/perfectionism/), whose posts contained sentences like "*like many perfectionists, I feel... ,*" "*I definitely have unhealthy perfectionism,*" and so on. We calculated top 100 high-frequency words from their posts. We remove stop words and words representing things like examination. and acquire such words like "*no enough,*" "*perfect,*" "*loser,*" "*failure,*" "*imperfect,*" "*compete,*" and so on. We took them as seed words, and enrich the perfection-related lexicon with synonymous or similar words. We also made a ruminant thinking-related lexicon consisting of very destructive negative words such as "*regret,*" "*repent,*" "*rue,*" "*penitent,*" "*self-blame,*" etc. Table III lists typical perfection-related and ruminant thinking-related words that we used in the study.

We measured user's pursuing perfection and ruminant thinking based on the mean proportion of perfection-related and ruminant-related words per post, respectively. Let $P(u)$ denote the set of posts made by user $u$. For each post $p \in P(u)$, assume function $pwordNum(p, u)$ and $rwordNum(p, u)$ return the number of perfection-related and ruminant-related words in $p$, respectively.

$$Perfect(u) = \frac{1}{|P(u)|} \sum_{p \in P(u)} \frac{pwordNum(u,p)}{|p|}$$
$$Ruminat(u) = \frac{1}{|P(u)|} \sum_{p \in P(u)} \frac{rwordNum(u,p)}{|p|}$$

We measured user $u$'s interpersonal sensitivity by counting how many times the user experienced stressful periods in the stress category of *inter-personnel relation*.

$$Sensitive(u) = |\{s | s \in \mathcal{S}(u) \wedge s_c = \text{"} interpersonal\ relation\text{"}\}|$$

where $\mathcal{S}(u)$ is a set of stressful periods detected from user $u$'s posts, and $s = (s_p, s_l, s_c)$ is one of it in $\mathcal{S}(u)$.

*4) Post Behavior (Text, Image, Post Time):* An individual may post texts and images anytime at will. These linguistic and/or visual contents, as well as the posting time, may reveal personal thoughts or feelings. Since Bert [70] and ResNet [71] are common used pre-trained models which perform well on textual feature extraction [72], [73] and visual feature extraction [74], [75]. For each post, we encoded its linguistic and visual contents into a 768-dimensional and a 300-dimensional vector, respectively, through a pre-trained Bert model and a 34-layer ResNet. When the image was missing, we filled in with a default null image, which is usually displayed at the position for image wanted in applications. When the post contains multiple images, we took the average feature vector of the images as the visual content representation of the post. We mapped the user's absolute post time into the hour from 0 to 23. Assume user $u$ made totally $n$ posts $post_1(text_1, image_1, ptime_1), \ldots,$ $post_n(text_n, image_n, ptime_n)$, where $text_i \in \mathbb{R}^{768 \times 1}$, $image_i \in \mathbb{R}^{300 \times 1}$, and $ptime_i \in \{0, 1, 2, \ldots, 23\}$ for $1 \leq i \leq n$.

To further learn user $u$'s posts representation, we concatenated $(text_i, image_i, ptime_i)$ as a 1069-dimensional vector $postvec_i$, where $postvec_i = text_i || image_i || ptime_i \in \mathbb{R}^{1 \times 1069}$, and then applied LSTM to extract textual and visual information from $postvec_1, \ldots, postvec_n$. Let the hidden dimension size of LSTM be 300.

$$out_i, h_i = LSTM(postvec_i, h_{i-1}).$$

The last output of LSTM was $out_n \in \mathbb{R}^{1 \times 300}$. To reduce the computational complexity of subsequent calculations, a fully connected layer was used to generate a 30-dimensional vector as the value of user's *post behavior* property:

$$PostBehavior(u) = ReLU(out_n W_0 + b_0) \in \mathbb{R}^{1 \times 30},$$

where $out_n \in \mathbb{R}^{1 \times 300}$, $W_0 \in \mathbb{R}^{300 \times 30}$, and $b_0 \in \mathbb{R}^{1 \times 30}$.

Details of learning user $u$'s post behavior representation are illustrated in Fig. 5.
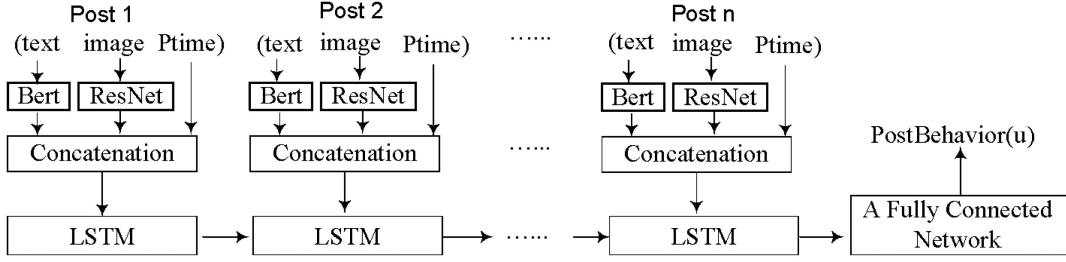
Fig. 5. Framework for learning user $u$'s post behavior.

*5) Emotion Expression (Suicide Words, Last Words, Future Words, Negation Words, Self-Concerns, Emotion Transition):* Suicides tend to express rather than repress their despair suicidal feelings, less future words, more negation and self-referenced first-person pronouns such as *I, me, my, myself,* and *mine*, etc. in their posts [66]. The existence of last words (which express regret, guilty, wishes towards their families and friends, or funeral), particularly in those recent posts, is also recognized as an important observable risk factor of suicidal ideation [76]. From the Chinese Suicide Dictionary [77] and Chinese Linguistic Inquiry and Word Count [78], we extracted 586 about suicide, 125 about last words, 86 about future words, 665 about negation words, and 36 words/phrases about self-concern. We calculated the mean word frequencies (i.e., mean word proportions over the total words per post) as the property values of $SuicideProp(u)$, $LastWordProp(u)$, $FutureProp(u)$, $NegProp(u)$, and $SelfProp(u)$ for user $u$. Here, only two most recent weeks' posts were considered in the computation of $LastWordProp(u)$.

Apart from the specific words, we drew the inspiration from the study [46] that there is a significant difference in emotion transition between suicide users and non-suicide users. That is, love emotion is more regularly followed by anxiety or sorrow emotion among the suicide people, while love is more regularly followed by joy emotion among the non-suicide people. Hereby, as in [46], we considered eight emotions (*love, joy, expectation, anxiety, sorrow, anger, hate,* and *surprise*), and counted two types of emotion transitions from love to joy ($love$-$joy(u)$) or from love to anxiety/sorrow ($love$-$anxiety/sorrow(u)$) within the ten most recent continuous posts as the property value of emotion transition: $EmotionTrans(u) = (love$-$joy(u), love$-$anxiety/sorrow(u))$.

For user $u$, if there existed a pair of posts, where the first one contained one or more love lexicons, and the second later post contained one or more joy lexicons based on the Chinese emotion lexicons DUTIR (http://ir.dlut.edu.cn/), the value of $love$-$joy(u)$ was increased by 1. The value of $love$-$anxiety/sorrow(u)$ was computed in the same way.

*6) Interaction (Following Number, Follower Number, Mention/Forward/Agree Number, Neighbour Users):* Social isolation is a significant and reliable predictor of suicidal ideation. When trapped in trouble, if one 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 reality, desperate people tend to have a weak social network, and thus get weak social supports [67]. Here, one's online social engagement is reflected from his following/follower number, mention/forward/agree number, and neighbour users. Specially, in order not to be affected by "Zombie fan" (fake fans in social media), we filtered out fans who posted less than three original posts in the last six months when counting the follower number.

$Interact(u) = (FollowingNum(u), FollowerNum(u), IntActNumb(u), Users(u))$, where the first three elements take non-negative integer values, and $Users(u)$ is a set of neighbour users which user $u$ follows.

Formally, we describe the personal suicide-oriented knowledge graph as $G = \{V, E\}$, where $V$ and $E$ are node set and edge set respectively. A user node can have multiple property nodes, and may link to another user node via the following relationship on the social media. The weight of a social interaction-user edge implies the suicidal emotion influence between the two users on the social media, and the weight of a user-property edge represents the contribution of the property to the suicidal ideation. These weights will be computed during the following learning process.

Finally, the statistic of knowledge graph is shown in Table IV.

### B. Suicidal Ideation Detection Based on the Personal Knowledge Graph

We designed a property attention mechanism and neighbour attention mechanism to reason the key factors related to suicidal ideation, and discover the suicidal emotion influence among the neighbour users on the social media.

*1) Property Attention:* Excepting user $u$'s neighbour users, we concatenated all the instantiated properties from $u$'s personal knowledge graph, and obtained a 61-dimensional property vector $P_u = (p_1, p_2, \ldots, p_n) \in \mathbb{R}^{61 \times 1}$, where $n = 61$.

To compute the significance of different properties, we enforced $P_u$ with a property attention vector $\alpha \in \mathbb{R}^{1 \times 61}$ and acquired a new property vector $P'_u \in \mathbb{R}^{61 \times 1}$:

$$P'_u = P_u \times \alpha^T,$$

$$\alpha = softmax((P_u)^T W_1 + b_1),$$

where $W_1 \in \mathbb{R}^{61 \times 61}$ and $b_1 \in \mathbb{R}^{1 \times 61}$ are two trainable parameters.

*2) Neighbour Attention:* We imposed neighbour influences upon user $u$'s property representation $P'_u$. Considering different neighbour influences upon user $u$'s suicidal ideation, for instance, some people may easily be influenced by elders, and

TABLE IV
STATISTICS OF USERS' PERSONAL SUICIDE-ORIENTED KNOWLEDGE GRAPHS ON WEIBO

| Category | Properties | Users with Suicidal Ideation | Ordinary Users without Suicidal Ideation |
|---|---|---|---|
| Personal Information | Gender (*male,female,unknown*) | (21.3%, 78.5%, 0.2%) | (50.6%, 42.3%, 7.1%) |
| | Age on average | 25.8 | 28.3 |
| | Location (*east, south, north, south-west, north-west, middle, north-east, unknown*) | (18.8%, 10.5%, 7.4%, 6.6%, 3.0%, 7.0%, 3.3%, 43.5%) | (25.0%, 15.0%, 15.7%, 5.8%, 2.9%, 7.9%, 5.3%, 22.3%) |
| Personality | Pursuing Perfection | 0.25% | 0.17% |
| | Ruminant Thinking | 0.086% | 0.052% |
| | Interpersonal Sensitive | 1.3 | 1.0 |
| Experience | Number of Stressful Periods | 2.1 | 1.8 |
| | Psychological Disorder | 0.1% | 0.03% |
| | Previous Attempt | 3.5% | 0.1% |
| Post Behavior | Number of Texts | 252,901 | 491,130 |
| | Number of Images | 93,461 | 260,667 |
| Emotion Expression | Proportion of Suicide Words Per Post | 0.034% | 0.016% |
| | Proportion of Last Words Per Post | 0.0013% | 0.000023% |
| | Proportion of Future Words Per Post | 0.031% | 0.045% |
| | Proportion of Negation Words Per Post | 0.012% | 0.009% |
| | Proportion of Self-Concern Words Per Post | 0.079% | 0.029% |
| | Emotion Transition (*love-joy, love-anxiety/sorrow*) | (0.1, 0.5) | (0.3, 0.1) |
| Social Interaction | Following Number | 207.0 | 378.1 |
| | Follower Number | 566.9 | 1515.3 |
| | Mention/Forward/Agree Number | 3.4 | 10.9 |
| | Number of Neighbour Users | 4.3 | 5.1 |

else may trust their peers more, we adopted a neighbour attention mechanism similar to the graph attention mechanism [28].

To reduce computational complexity, we firstly figured out user $u$'s initial hidden state $h_u \in \mathbb{R}^{1 \times 60}$ from his property representation $P'_u$ through a fully connected layer:

$$h_u = tanh((P'_u)^T W_2 + b_2),$$

where $W_2 \in \mathbb{R}^{61 \times 60}$ and $b_2 \in \mathbb{R}^{1 \times 60}$ are trainable parameters.

Let $\mathbb{N}_u$ be the set of direct neighbour users linked with user $u$ on the social media. For each neighbour pair $(u, \widetilde{u})$ (where $\widetilde{u} \in \mathbb{N}_u$), we obtained the attention coefficient $c_{u,\widetilde{u}} \in \mathbb{R}^{1 \times 1}$ through another fully connected layer with concatenate operation $||$:

$$c_{u,\widetilde{u}} = tanh((h_u || h_{\widetilde{u}}) W_3 + b_3),$$

where $W_3 \in \mathbb{R}^{120 \times 1}$ and $b_3 \in \mathbb{R}^{1 \times 1}$ are two trainable parameters.

In this way, we obtained a vector of user $u$'s neighbour attention coefficients:

$$C_u = [c_{u,1}, c_{u,2}, \ldots, c_{u,|\mathbb{N}_u|}] \in \mathbb{R}^{1 \times |\mathbb{N}_u|}$$

where $|\mathbb{N}_u|$ is the number of $u$'s direct neighbours.

A softmax function was then applied to compute a series of scores $\mathcal{B}_u = [\beta_{u,1}, \beta_{u,2}, \ldots, \beta_{u,|\mathbb{N}_u|}] \in \mathbb{R}^{1 \times |\mathbb{N}_u|}$ to represent the influences of neighbour users $\mathbb{N}$ upon user $u$. The higher the score is, the bigger the suicidal influence is.

$$\mathcal{B}_u = softmax(C_u).$$

By means of $\mathcal{B}_u$, we aggregated information from neighbour users and updated $u$'s hidden state $h_u$ accordingly into $h'_u \in \mathbb{R}^{1 \times 60}$:

$$h'_u = \sigma \left( \sum_{\widetilde{u} \in \mathbb{N}_u} \beta_{u,\widetilde{u}} h_{\widetilde{u}} + h_u \right).$$

Once obtained, we utilized a fully connected layer to get the user's final representation $r_u \in \mathbb{R}^{1 \times 60}$:

$$r_u = tanh(h'_u W_4 + b_4),$$

where $W_4 \in \mathbb{R}^{60 \times 60}$ and $b_4 \in \mathbb{R}^{1 \times 60}$.

*3) Suicidal Ideation Detection:* Finally, a fully connected layer and softmax function were applied to detect suicidal ideation of user $u$:

$$[y_1, y_0] = softmax(r_u W_5 + b_5),$$

where $y_1, y_0$ represent the possibility of user $u$ with or without suicidal ideation, $W_5 \in \mathbb{R}^{60 \times 2}$ and $b_5 \in \mathbb{R}^{1 \times 2}$ are trainable parameters.

## IV. EXPERIMENTS

### A. Datasets

In this study, we evaluate our personal suicide-oriented knowledge graph based suicidal ideation detection methods on two types of datasets: microblog posts from Sina Weibo and forum posts from Reddit.

*1) Sina Weibo Dataset:* On March 17, 2012, a user which screen name was "Zoufan" left his last word on Sina Weibo and then committed suicide. Since then over 160000 users gathered here and posted more than 1700000 comments, and the numbers keep growing today, forming a large microblog tree hole. The majority of these commenting posts spoke out the posters' minds, disclosing their tragic experiences, hopeless thoughts, and even plans of suicide.

We crawled all the users' commenting posts from May 1, 2018 to April 30, 2019, and computed each user's expression degrees related to suicide according to the Chinese social media-based suicide dictionary [66]. The dictionary lists 2168 suicide-related words and phrases, falling into 13 categories (suicidal thought, self-harm, psychache, mental illness, hopeless feeling, somatic complain, self-regulation, negative personality, stress,

TABLE V
STATISTIC OF DATA COLLECTED FROM SINA WEIBO

| | # Users | # Normal Posts | # Normal Posts with Images | # Neighbour Users to be followed |
|---|---|---|---|---|
| Users with suicidal ideation | 3,652 | 252,901 | 93,461 | 4.3 |
| Ordinary users without non-suicide | 3,677 | 491,130 | 260,667 | 5.1 |
| Total users | 7,329 | 744,031 | 354,128 | 4.7 |

trauma/hurt, talking about others, shame/guilt, anger/hostility) Each word/phrase is assigned a weight from 1 to 3, indicating its binding degree with suicide risk.

Assume user $u$ posted a set of comments $P(u)$ in the tree hole. We define $u$'s expression degree related to suicide by counting his total suicide-related weighted words/phases in $P(u)$.

$$UserDegree(u) = \sum_{post \in P(u)} postDegree(post)$$
$$postDegree(post) = \sum_{w \in post} weight(w)$$

where $weight(w)$ is the weight of word/phrase $w$ in $p$ according to the suicide dictionary [66]. A word/phrase not in the suicide dictionary has weight 0.

We ranked and selected top 4000 users who had high expression degrees related to suicide. After that, we recruited four PhD researchers to manually scan the 4000 users and keep 3652 users, who clearly mentioned suicide thought, suicide plan, or self-injury at least five times in different days. For example, if a user expressed clear suicidal thoughts like *"At this moment, I especially want to die. I feel very tired. I really want to be free."*, *"Heh, even Weibo cant give me a reason to keep on going."* or *"I don't want to do anything for the last five days of my life."* more than 5 times in different days, then we label him/her at suicide risk. Finally, we labeled these 3652 users with suicidal ideation at suicide risk.

As comparison, we randomly selected 3677 ordinary users on Sina Weibo, subject to the following three conditions. Firstly, they were active users, each making over 100 normal posts on the open microblog during the same temporal period. Secondly, neither of their posts contained any suicide-related word or phrase. Thirdly, they did not make any commenting post in any hidden tree hole. In the study, we regarded the ordinary users without suicidal ideation. If two users had a following-follower relationship on Sina Weibo, we built an edge linking the two neighbour users. Table V illustrates the data collected from Sina Weibo.

The training set, validation set, and test set contained 6129, 600, and 600 users, who were exclusively and randomly chosen from the total 7329 microblog users, respectively. Users with suicidal ideation and ordinary users without suicidal ideation occupied nearly half in each set.

*2) Reddit Dataset:* The Reddit dataset was crawled from the popular online forum Reddit, and contained 500 users in total. There were five categories of users, i.e., Supportive, Suicide Indicator, Suicidal Ideation, Suicidal Behavior, and Actual Attempt with user number of 108, 99, 171, 77, and 45, respectively. From Supportive to Actual Attempt, the risk of suicide gradually increases. We split the Reddit dataset into training set, validation set, and test set with the number of 303, 99, and 98, respectively.

TABLE VI
CONFUSION MATRIX USED TO EVALUATE THE PERFORMANCE

| Detected / Actual | Positive | Negative |
|---|---|---|
| Positive | True Positive (TP) | False Negative (FN) |
| Negative | False Positive (FP) | True Negative (TN) |

### B. Experimental Setup

We compared the performance of our personal suicide-oriented knowledge graph based method with the following four methods.

- **CNN** [19]: A Convolutional Neural Network takes embeddings of user posts as input.
- **LSTM+Attention** [15]: An self-attention mechanism based Long Short-Term Memory model which captures the contextual information between suicide-related words and others.
- **SDM** [59]: An hierarchical attention network based on Long Short-Term Memory module and ResNet module.
- **Text+History+Graph** [26]: A multipronged approach includes bidirectional Long Short-Term Memory module and Graph Neural Network. The textual content shared in Twitter, the historical tweeting activity of the users and social network formed between different users posting about suicidality are concerned to identify and explore users' suicidal ideation.

As Reddit dataset only contained posts from each user, during the experiment, we deleted the ResNet module and the user profile module of the SDM method, the Social Graph module of the Text+History+Graph method, as well as the Neighbour Attention module, Personal Information property and Social Interaction property of our KG-based method. We took Bert contextual word embeddings [70] as our word embeddings.

Five metrics (accuracy, precision, recall, F1-measure, macro-F1-measure) were adopted in the performance evaluation.

$$Precision = \frac{TP}{TP+FP}, Recall = \frac{TP}{TP+FN}$$
$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$
$$F1\text{-}measure = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

where true positive (TP), false positive (FP), true negative (TN), and false negative (FN) are defined in the following confusion matrix (Table VI).

As toward Reddit dataset there were more than two classes (actually 5), we employed a commonly used multi-class macro-F1-measure, which calculates F1-measure for each class, and then finds their un-weighted mean.

TABLE VII
PERFORMANCE COMPARISON ON THE SINA WEIBO DATASET

| Method | Acc. | F1. | Prec. | Rec. |
|---|---|---|---|---|
| CNN | 86.77% | 84.47% | 84.19% | 84.77% |
| LSTM+Attention | 88.89% | 88.10% | 88.56% | 87.65% |
| SDM | 91.35% | 90.97% | 90.11% | 91.85% |
| Text+History+Graph | 91.56% | 91.81% | 91.85% | 91.77% |
| KG-based method | **93.74%** | **93.69%** | **93.75%** | **93.64%** |

TABLE VIII
PERFORMANCE COMPARISON ON THE REDDIT DATASET

| Method | Acc. | macro.F1. | Prec. | Rec. |
|---|---|---|---|---|
| CNN | 52.31% | 52.61% | 53.05% | 52.19% |
| LSTM+Attention | 55.12% | 54.49% | 53.16% | 55.89% |
| SDM | 60.58% | 60.44% | 60.31% | 60.58% |
| Text+History+Graph | 59.61% | 59.02% | 58.42% | 59.64% |
| KG-based method | **65.92%** | **65.93%** | **65.35%** | **66.21%** |

TABLE IX
PERFORMANCE OF PROPERTY AND NEIGHBOUR ATTENTION MECHANISMS

| Method | Acc. | F1. | Prec. | Rec. |
|---|---|---|---|---|
| avg-weight-property | 92.55% | 93.23% | 93.55% | 92.91% |
| avg-weight-neighbour | 92.17% | 93.02% | 93.17% | 92.88% |
| random-weight-property | 90.12% | 89.56% | 89.49% | 89.63% |
| random-weight-neighbour | 91.86% | 92.16% | 92.55% | 91.77% |
| property+neighbour attention | **93.74%** | **93.69%** | **93.75%** | **93.64%** |

TABLE X
PERFORMANCE OF USING DIFFERENT GRAPH NEURAL NETWORK (GNN) MODELS

| Method | Acc. | F1. | Prec. | Rec. |
|---|---|---|---|---|
| GCN | 92.26% | 92.69% | 92.86% | 92.54% |
| GraphSAE | 92.49% | 91.16% | 90.89% | 91.43% |
| GAT-inspired (property+neighbour attention) | **93.74%** | **93.69%** | **93.75%** | **93.64%** |

## C. Effectiveness of Involving Personal Suicide-Oriented Knowledge Graphs in Suicidal Ideation Detection

Table VII shows that, with the considerations of visual information from posts and users' profile information, the SDM method achieved better performance compared with that of CNN method and LSTM-Attention method. After using the social graph of users, Text+History+Graph method outperformed the above methods slightly in accuracy and F1-measure. Furthermore, involving personal suicide-oriented knowledge graphs into suicidal ideation detection could improve the detection accuracy, F1-measure, precision, and recall by 2.18%, 1.88%, 1.9%, 1.87% over the Text+History+Graph method, demonstrating the effectiveness of the personal knowledge graph based method.

Since five-class classification is a harder task than two-class classification, and the modalities of the Reddit dataset are less than that of the Sina Weibo dataset, the overall performance of all the methods dropped, compared with that on the Sina Weibo dataset, as shown in Table VIII. Without the contribution of social graph, the performance of the Text+History+Graph method dropped to 59.61% and 59.02% in accuracy and F1-measure, respectively, which were worse than the SDM method. With the help of personal suicide-oriented knowledge graph, the KG-based method still outperformed the rest methods.

## D. Effectiveness of Property and Neighbour Attention Mechanisms

Property attention and neighbour attention are two crucial modules of our personal knowledge graph based method. They aim to reason about the key properties and find the most influential neighbour users to user's suicidal ideation, reflected from their associated property weights and neighbour weights, respectively. To investigate the effectiveness of the two attention mechanisms, we replaced each learnt attention weight with an average or random value in the range of [0,1], respectively. Table IX shows that the two attention mechanisms were effective, achieving the best performance compared to the alternative average and random weight solutions, and the average solution was better than the random solution.

It is worth mentioning here that the property+neighbour attention strategy drew inspirations from the Graph Attention networks (GAT) [28], which incorporates the neighbor attention mechanism. This also contributes to its best performance than those of the other two well-known GNN models (GCN [79] and GraphSAGE [80]), which are commonly used for neighbour information aggregation. As illustrated in Table X, after replacing the neighbour attention module with the GCN method or GraphSAGE method, there are obvious declines about more than 1.25%, in accuracy and 1% in F1-measure on the Weibo dataset, since the neighbour attention mechanism can derive the different influence from one's different neighbour users.

## E. Contributions of Social Neighbours and Personal Suicide-Oriented Knowledge Graph to Suicidal Ideation Detection

The effectiveness of the personal suicide-oriented knowledge graph prompted us to further investigate the impact of different personal factors upon suicidal ideation detection. We used *information gain* to assess the change of information amount involving a certain personal factor (property) based on *entropy* and *conditional entropy*. A bigger information gain implies a bigger significance of the factor.

Formally, let $F$ be a category, and $f$ be a vector value in the domain of category $F$, i.e., $f \in Dom(F)$. Let $y \in \{y_1, y_0\}$ denote the detection result (with/without suicidal ideation). $Prob(.)$ represents probability and $H(.)$ is the entropy.

$InfoGain(y|F) = H(y) - H(y|F)$
where $H(y) = -\sum_{y \in \{y_1, y_0\}} Prob(y) log Prob(y)$,
$H(y|F) = \sum_{f \in Dom(F)} Prob(F = f) H(y|F = f)$,
$H(y|f) = \sum_{y \in \{y_1, y_0\}} Prob(y|F = f) log Prob(y|F = f)$.

We firstly considered the six categories of personal properties (*Personal Information, Personality, Experience, Post Behaviors, Emotion Expression*, and *Social Interaction*) in one's personal suicide-oriented knowledge graph, and then delved into the concrete properties of the key categories.

To ease the computation, for a property which takes continuous values (like *Age*, *Pursuing Perfection*, *Ruminant Thinking*, *Interpersonal Sensitive*, *Stress Duration*, *Stress Level*, *Stress Category*, *Suicide Words*, *Last Words*, *Future Words*, *Negation Words*, *Self-Concerns Words*, *Following Number*, *Follower Number*, and *Mention/Forward/Agree Number*), we used the mean to divide all its values into two classes, and transformed the domain of the property into $\{0, 1\}$. The domains of discrete property values (like *Location*, *Gender* and *Emotion Transition*) remained unchanged.

We mapped the *Texts* property values under the *Post Behavior* category to three classes based on the emotional polarities given by SnowNLP (https://github.com/isnowfy/snownlp). Assume user $u$ wrote a sequence of texts $Texts(u)$ in his posts.

$$Class(Texts(u)) = \begin{cases} 0 \text{ if } EP(Texts(u)) \leq -0.3; \\ 1 \text{ if } -0.3 < EP(Texts(u)) < 0.3; \\ 2 \text{ if } EP(Texts(u)) \geq 0.3 \end{cases}$$

where $EP(Texts(u)) = \frac{\sum_{t \in Texts(u)} SnowNLP(t)}{|Texts(u)|}$.

In a similar way, we mapped the *Images* property values under the *Post Behavior* category to four classes based on brightness and proportion of warm color, calculated by RGB. Assume user $u$ posted a sequence of images $Images(u)$.

$$Class(Images(u)) = \begin{cases} 0 \text{ if } B(Images(u)) < 0.5 \wedge \\ \quad W(Images(u)) < 0.5; \\ 1 \text{ if } B(Images(u)) < 0.5 \wedge \\ \quad W(Images(u)) \geq 0.5; \\ 2 \text{ if } B(Images(u)) \geq 0.5 \wedge \\ \quad W(Images(u) < 0.5; \\ 3 \text{ if } B(Images(u)) \geq 0.5 \wedge \\ \quad W(Images(u)) \geq 0.5 \end{cases}$$

where $B(Images(u)) = \frac{\sum_{i \in Images(u)} RGB-Bright(i)}{|Images(u)|}$, and $W(Images(u)) = \frac{\sum_{i \in Images(u)} RGB-Warm(i)}{|Images(u)|}$.

Fig. 7 shows the average information gain of each category in the detection of suicidal ideation. From the result presented, we can see that all the listed categories brought positive impact on suicidal detection, and the top-3 categories were *Post Behavior*, *Personality*, and *Experience*, whose information gains were 0.33, 0.28, and 0.21, respectively.

We further looked into all the categories, and examined the contributions from their respective properties. As shown in Fig. 8, the top-5 key properties were *posted text, stress level, stress duration, image,* and *ruminant thinking*.

To be more illustrative in the detection of suicidal ideation, we conducted a further feature selection experiment. From the result presented in Fig. 9, we note that after removing the top-$x$ key properties in information gain (where $x = 1, 3, 5, 10$), both accuracy and F1-measure kept declining, verifying the effectiveness of the key properties in suicidal risk analysis.

### F. Impact of Data Noise on Suicidal Ideation Detection Performance

Due to the less restrictive and free-style nature of the social media, the presence of data noise is quite common, and may
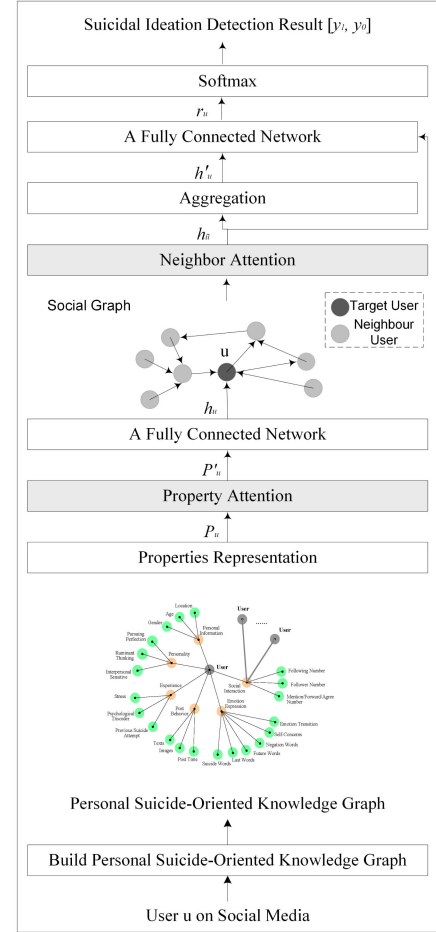


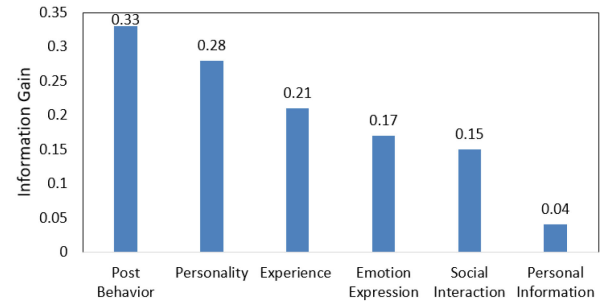Fig. 6. Architecture of the personal knowledge graph based suicidal ideation detection.



Fig. 7. Information gains of the top-3 categories.

have a misleading effect on the experimental results. To examine the impact of data noise on the suicidal ideation detection performance, we consider two types of data noise: (1) **anti-real posts** (that is, users didn't fill in real information of themselves on the social media.) (2) **a few posts** (that is, users made only a few posts on the social media), and conducted experiments on a few datasets, each possessing different levels of data noise.

*1) Users With Anti-Real Posts:* Thanks to the existence of the tree hole, we could glimpse some anti-real sentiment expression from a user's normal open post as follows. For a user who made posts in both the tree hole and on the open microblog platform, if
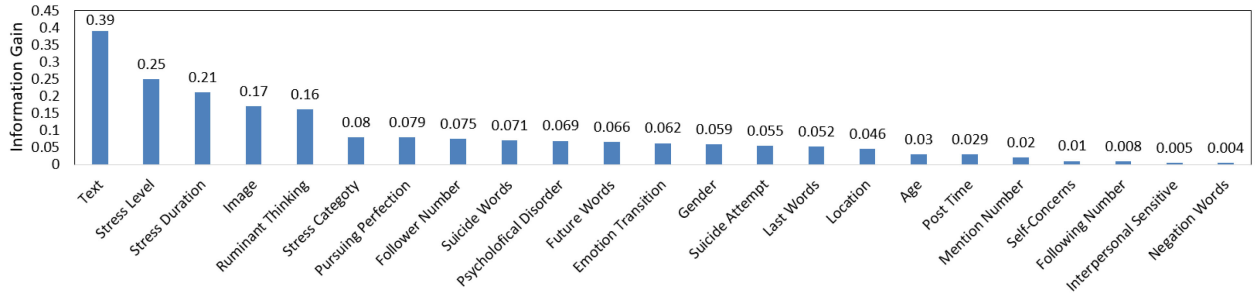
Fig. 8.    Information gains of all the properties from personal suicide-oriented knowledge graph.

TABLE XI
DERIVED FIVE TEST DATASETS CONTAINING DIFFERENT PROPORTIONS OF USERS WITH ANTI-REAL POSTS

|  | $TD_1$ | $TD_2$ | $TD_3$ | $TD_4$ | $TD_5$ |
|---|---|---|---|---|---|
| # Users with suicidal ideation | 43 | 100 | 200 | 300 | 257 |
| # with anti-real posts | 43 (100%) | 43 (43%) | 43 (21.5%) | 43 (14.3%) | 0 (0%) |
| # without anti-real posts | 0 | 57 | 157 | 257 | 257 |
| # Ordinary users (random selection) | 43 | 100 | 200 | 300 | 257 |
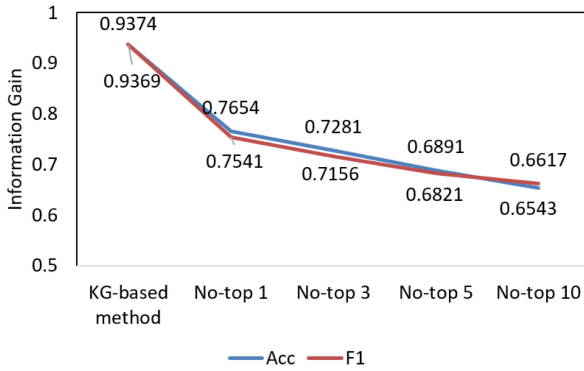


Fig. 9.    Results of the feature selection experiment through removing the top-$x$ key properties in information gain (where $x = 1, 3, 5, 10$).

TABLE XII
DETECTION PERFORMANCE ON THE FIVE DERIVED TEST DATASETS
CONTAINING DIFFERENT PROPORTIONS OF USERS WITH ANTI-REAL POSTS

| Method | $TD_1$ | $TD_2$ | $TD_3$ | $TD_4$ | $TD_5$ |
|---|---|---|---|---|---|
| CNN | 65.75% | 73.73% | 82.97% | 86.77% | 90.12% |
| LSTM+Attention | 67.27% | 76.54% | 84.23% | 88.89% | 92.12% |
| SDM | 74.98% | 82.31% | 87.59% | 91.35% | 94.03% |
| Text+History+Graph | 75.61% | 83.59% | 88.64% | 91.56% | 94.06% |
| KG-based method | **80.08%** | **88.11%** | **90.76%** | **93.74%** | **95.45%** |
| –On suicidal users | 66.51% | 82.33% | 88.06% | 93.97% | 97.58% |
| –On ordinary users | 93.65% | 93.89% | 93.46% | 93.51% | 93.51% |

TABLE XIII
DERIVED TEST DATASET CONTAINING USERS WITH LESS THAN 5 POSTS
THROUGHOUT THE YEAR

|  | $TD_6$ |
|---|---|
| # Users with suicidal ideation | 88 |
| # Ordinary users | 88 |

(1) he expressed despair, depression, anxiety, or suicidal ideation in the tree hole, but showed uplift feelings on the open posts within the 24 hours; and (2) the above situation happened more than twice within the latest one month, we annotated the user with anti-real posts.

Our test Sina Weibo dataset contains 300 users with suicidal ideation and 300 ordinary users. From the 300 users with suicidal ideation, we identified totally 43 users with anti-real posts. By varying the proportions of users with suicidal ideation in the test dataset, we derived 5 test datasets with different levels of data noise. In each derived test dataset (Table XI), as ordinary users made no posts in the tree hole, we randomly picked up the same number of ordinary users from the 300 ordinary test users as the control group.

Table XII lists the performance of all the methods on the five different test datasets. Our knowledge graph based method outperforms all the other methods in the four metrics. The less proportions of the users with anti-real posts, the higher the detection performance. When the proportion of the users with suicidal ideation and anti-real posts dropped to 21.5%, 14.3%, and 0%,

the detection accuracy of the knowledge graph based method on the suicidal group reached 88.06%, 93.97%, and 97.58%, respectively. The detection performance for the ordinary group remained unchanged, around 93.65% in accuracy and 93.89% in F1-measure.

*2) Users With Only a Few Posts:* From the total 600 test users (300 ordinary users and 300 users with suicidal ideation), we filtered out 88 users with suicidal ideation and another 88 ordinary users without suicidal ideation, and all of them had less than 5 posts throughout the year (Table XIII).

It is not surprising that the performance of all the trained models (in Table XIV) dropped, given such a small number of test posts. Among the five detection methods, our knowledge graph based method achieved the highest overall accuracy of 89.73% on all the users, 92.04% on the suicidal group, and 87.50% on the ordinary group.

TABLE XIV
DETECTION PERFORMANCE ON THE DERIVED TEST DATASET CONTAINING
USERS WITH ONLY A FEW POSTS

| Method | Acc. | F1. | Prec. | Rec. |
|---|---|---|---|---|
| CNN | 72.35% | 72.38% | 72.12% | 72.65% |
| LSTM+Attention | 74.56% | 74.91% | 74.98% | 74.84% |
| SDM | 85.49% | 85.36% | 85.46% | 85.26% |
| Text+History+Graph | 86.88% | 86.52% | 86.49% | 86.55% |
| KG-based method | **89.73%** | **89.49%** | **89.92%** | **89.49%** |
| –On suicidal users | 92.04% | 90.00% | 88.04% | 92.04% |
| –On ordinary users | 87.50% | 89.53% | 91.67% | 87.50% |

TABLE XV
PERFORMANCE ON THE NEW DATASET

| Method | Acc. | F1. | Prec. | Rec. |
|---|---|---|---|---|
| CNN | 87.56% | 87.65% | 87.61% | 87.69% |
| LSTM+Attention | 88.86% | 87.99% | 88.23% | 87.77% |
| SDM | 91.78% | 90.87% | 90.28% | 91.47% |
| Text+History+Graph | 92.01% | 91.62% | 91.76% | 91.48% |
| KG-based method | **94.53%** | **94.31%** | **94.45%** | **94.18%** |

*3) Experiments on a Small-Sized New Dataset:* Apart from examining the impact of data noise through the six datasets derived from our Weibo test dataset, we also built a small-sized new dataset, containing 85 suicide users (who had passed away as verified by news from 2012 to 2014) and another 85 ordinary users randomly sampled from our test dataset. It is expected that the data noise of this new dataset was less than that of our Sina Weibo dataset. Comparatively, suicide users in this new dataset had more supportive information.

1) They made more posts (average 83.2 in their last year) than the suicidal ideation users in our Weibo dataset (average 69.2 in the latest one year).

2) They experienced more stressful periods (average 6.0), showed more negative expressions (23% posts with negative emotions given by SnowNLP) in the last year than the suicidal ideation users (average 69.2 posts, 2.1 stressful periods, 12% posts with negative emotion in the latest one year).

With the help of the supportive information, our knowledge graph based method could deliver a very high performance, 94.53% in accuracy and 94.33% in F1-measure, as shown in Table XV.

To sum up, the proposed knowledge graph based suicidal ideation detection model outperforms the other baseline methods, when facing challenging few-posts and anti-real-posts on microblog.

### G. Ablation Study

As the introduction of the user knowledge graph is essential in this study, we conducted an ablation study to see the performance with only the user's post behavior (Without-KG method). As shown in Table XVI, XVII there are huge declines about 4.15%, 7.58% in accuracy and 4.48%, 6.91% in F1-measure on both datasets. The mediocre performance of the Without-KG method introduced the effectiveness of personal suicide-oriented knowledge graph.

TABLE XVI
RESULTS OF THE ABLATION STUDY ON THE SINA WEIBO DATASET

| Method | Acc. | F1. | Prec. | Rec. |
|---|---|---|---|---|
| KG-based method | **93.74%** | **93.69%** | **93.75%** | **93.64%** |
| Without-KG method | 89.59% | 89.21% | 89.68% | 88.74% |

TABLE XVII
RESULTS OF THE ABLATION STUDY ON THE REDDIT DATASET

| Method | Acc. | macro.F1. | Prec. | Rec. |
|---|---|---|---|---|
| KG-based method | **65.92%** | **65.93%** | **65.35%** | **66.21%** |
| Without-KG method | 58.34% | 59.02% | 59.41% | 58.63% |

TABLE XVIII
TRAINING AND INFERENCE TIME COSTS OF THE METHODS ON SINA WEIBO
DATASET

| Method | Training Time | Inference Time(batch) |
|---|---|---|
| CNN (text) | 56.7 mins | 37.9 seconds |
| LSTM+Attention (text) | 282.4 mins | 49.6 seconds |
| SDM (text+image) | 362.3 mins | 55.4 seconds |
| Text+History+Graph (text+graph) | 286.8 mins | 50.8 seconds |
| KG-based method (text+image+graph+KG) | 386.5 mins | 56.4 seconds |

### H. Training and Inference Time Costs

Table XVIII shows the training time costs of the five methods until convergence (getting the best result on the validation set) on a computer server with 4 NVIDIA GTX 1080Ti GPU. As our KG-based method utilized the most modalities (including text, image, social graph, and knowledge graph), its training time is the most. CNN requires the least training time, as it does not use the time-consuming sequential module. The inference time cost per batch (batch size is 16) of all the five methods are close except for CNN.

## V. DISCUSSIONS

### A. Ethical Considerations

Keeping ethical considerations in mind is essential for the task of suicidal ideation detection. All the data (including name, age, posts, etc.) was crawled from the public social media, and was only used for research. We anonymized the data before labeling. There was no interaction or intervention with the subjects.

### B. Limitations

*1) Insufficiency of Microblog Data:* While the detection result is promising, the construction of users' personal suicide-oriented knowledge graphs, particularly personality-related factors, is still very preliminary in the current study, leaving a few important factors out of the study due to resource limitations. For instance, according to the psychological study [81], individual's *parenting rearing style* is a persistent factor that affects an individual's mental health. It is widely believed that individuals who are poorly educated are more likely to develop suicidal ideation than individuals who are actively educated. Parental

rearing style is of great significance to the formation and development of individual's personality. Positive parental rearing style like emotional warmth and understanding, rather than negative excessive interference and excessive protection, contributes to one's healthy growth [82]. Completely relying on social media to analyze such personal knowledge is limited, and some other channels need to be explored to identify whether one was positively or negatively brought up.

*2) Noise of Microblog Data:* Due to the less restrictive and free-style nature of social media like microblog, it is quite possible that someone may have a few more accounts, or may not want to fill in real information on the social media, such as Weibo. In this study, we empirically examined the impact of data noise on suicidal ideation detection performance. Two types of data noise were considered: (1) users didn't fill in real information of themselves on the social media; and (2) users made only a few posts on the social media. We conducted experiments on a few derived datasets and another new dataset, and the results were expected. That is, the lower data noise exhibits, the higher detection performance can be achieved. This raised a very interesting question: "*can we firstly take a look at the data first before feeding them to train a detection model?*" In addition, user identification and inference of users real thoughts and feelings will also be quite desirable, deserving further investigation.

## VI. CONCLUSION

In this paper, we built and used a personal suicide-oriented knowledge graph for suicidal ideation detection on social media. A two-layered attention mechanism was deployed to explicitly reason and establish key risk factors to individuals' suicidal ideation. The performance study on 7329 microblog users (3652 with suicidal ideation and 3677 without suicidal ideation) show that: 1) with the constructed personal knowledge graph, the social media-based suicidal ideation detection can achieve over 93% accuracy, outperforming the state-or-art approach; and 2) among the six categories of personal factors, *post, personality,* and *experience* were the top-3 key indicators. Under these categories, *posted text*, *stress level*, *stress duration*, *posted image*, and *ruminant thinking* contribute the most to one's suicidal ideation detection.

While the paper shows some promising results of the proposed method, leveraging social media to accurately identify personal properties is still limited. More reliable data sources and domain-specific expert knowledge need to be used and integrated into suicidal ideation detection. Dynamic maintenance of user-centric knowledge graph also deserves deep investigation for building really handy solutions.

## ACKNOWLEDGMENT

## REFERENCES

[1] M. Nock *et al.* "Cross-national prevalence and risk factors for suicidal ideation, plans and attempts," *Brit. J. Psychiatry*, vol. 192, no. 2, pp. 98–105, 2008.

[2] N. Jacob, J. Scourfield, and R. Evans, "Suicide prevention via the internet," *J. Crisis*, 2014, pp. 1–7.

[3] C. Bagge and A. Osman, "The suicide probability scale: Norms and factor structure," *Psychol. Reports*, vol. 83, no. 2, pp. 637–638, 1998.

[4] K.-w. Fu, K. Y. Liu, and P. S. Yip, "Predictive validity of the chinese version of the adult suicidal ideation questionnaire: Psychometric properties and its short version." *Psychol. Assessment*, vol. 19, no. 4, pp. 422–429, 2007.

[5] K. M. Harris *et al.*, "The ABC's of suicide risk assessment: Applying a tripartite approach to individual evaluations," *PLoS One*, vol. 10, no. 6, 2015, pp. 1–21.

[6] H. Christensen, P. Batterham, and B. O'Dea, "E-health interventions for suicide prevention," *Int. J. Environ. Res. Public Health*, vol. 11, no. 8, pp. 8193–8212, 2014.

[7] C. A. Essau, "Frequency and patterns of mental health services utilization among adolescents with anxiety and depressive disorders," *Depression Anxiety*, vol. 22, no. 3, pp. 130–137, 2005.

[8] D. J. Rickwood, F. P. Deane, and C. J. Wilson, "When and how do young people seek professional help for mental health problems?" *Med. J. Aust.*, vol. 187, no. S7, pp. S35–S39, 2007.

[9] H. D. Zachrisson, K. Rödje, and A. Mykletun, "Utilization of health services in relation to mental health problems in adolescents: A population based survey," *BMC Public Health*, vol. 6, no. 1, pp. 1–7, 2006.

[10] S. Ji *et al.*, "Suicidal ideation detection: A review of machine learning methods and applications," *IEEE Trans. Computat. Social Syst.*, 2020, pp. 1–13.

[11] A. Alambo *et al.*, "Question answering for suicide risk assessment using reddit," in *Proc. IEEE 13th Int. Conf. Semantic Comput.*, 2019, pp. 468–473.

[12] Q. Cheng, T. M. Li, C. L. Kwok, T. Zhu, and P. S. Yip, "Assessing suicide risk and emotional distress in chinese social media: A text mining and machine learning study," *J. Med. Internet Res.*, vol. 19, no. 7, 2017, Art. no. e 243.

[13] J. Du *et al.*, "Extracting psychiatric stressors for suicide from social media using deep learning," *BMC Med. Infor. Decis. Mak.*, vol. 18, no. 2, pp. 78–87, 2018.

[14] R. Sawhney *et al.*, "Exploring and learning suicidal ideation connotations on social media with deep learning," in *Proc. 9th Workshop on Comput. Approaches Subjectivity, Sentiment Social Media Anal.*, 2018, pp. 167–175.

[15] G. Coppersmith, R. Leary, P. Crutchley, and A. Fine, "Natural language processing of social media as screening for suicide risk," *Biomed. Inform. Insights*, vol. 10, 2018, doi: 1178222618792860.

[16] M. J. Vioulés, B. Moulahi, A. J., and S. Bringay, "Detection of suicide-related posts in twitter data streams," *IBM J. Res. & Develop.*, vol. 62, no. 1, pp. 7:1–7:12, 2018.

[17] J. Robinson, M. Rodrigues, S. Fisher, E. Bailey, and H. Herrman, "Social media and suicide prevention: Findings from a stakeholder survey," *Shanghai Arch. Psychiatry*, vol. 27, no. 9, pp. 27–35, 2015.

[18] M. Gould, P. Jamieson, and D. Romer, "Media contagion and suicide among the young," *Amer. Behav. Scientist*, vol. 46, no. 9, pp. 1269–1284, 2003.

[19] M. Gaur *et al.*, "Knowledge-aware assessment of severity of suicide risk for early intervention," in *Proc. World Wide Web Conf.*, 2019, pp. 514–525.

[20] L. Brisset *et al.*, "Language barriers in mental health care: A survey of primary care practitioners," *J. Immigrant Minority Health*, vol. 16, no. 6, pp. 1238–1246, 2014.

[21] L. Nie *et al.*, "Bridging the vocabulary GAP between health seekers and healthcare knowledge," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 2, pp. 396–409, Feb. 2015.

[22] G. Ning, Z. Zhang, and Z. He, "Knowledge-guided deep fractal neural networks for human pose estimation," *IEEE Trans. Multimedia*, vol. 20, no. 5, pp. 1246–1259, May 2017.

[23] F. Xue *et al.*, "Knowledge based topic model for multi-modal social event analysis," *IEEE Trans. Multimedia*, vol. 22, no. 8, pp. 2098–2110, Aug. 2020.

[24] C. Chaudhary, P. Goyal, D. N. Prasad, and Y.-P. P. Chen, "Enhancing the quality of image tagging using a visio-textual knowledge base," *IEEE Trans. Multimedia*, vol. 22, no. 4, pp. 897–911, Apr. 2020.

[25] L. Nut and C. Nigel, "Normalising medical concepts in social media texts by learning semantic representation," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, 2016, pp. 1014–1023.

[26] P. P. Sinha *et al.*, "Suicidal—A multipronged approach to identify and explore suicidal ideation in twitter," in *Proc. 28th ACM Int. Conf. Infor. Knowl. Manag.*, 2019, pp. 941-950.

[27] A. Mbarek, S. Jamoussi, A. Charfi, and A. B. Hamadou, "Suicidal profiles detection in twitter," in *Proc. of the 15th Intl. Conf. Web Infor. Syst. Technol.*, 2019, pp. 289–296.

[28] P. Velickovic *et al.*, "Graph attention networks," in *Proc. 6th Int. Conf. Learn. Represent.*, 2018, pp. 1–12.

[29] J. P. Pes tian *et al.*, "A machine learning approach to identifying the thought markers of suicidal subjects: A prospective multicenter trial," *Suicide Life-Threatening Behav.*, vol. 47, no. 1, pp. 112–121, 2017.

[30] J. R. Crawford and J. D. Henry, "The depression anxiety stress scales (DASS): Normative data and latent structure in a large non-clinical sample," *Brit. J. Clin. Psychol.*, vol. 42, no. 2, pp. 111–131, 2003.

[31] J. D. Henry and J. R. Crawford, "The short-form version of the depression anxiety stress scales (DASS-21): Construct validity and normative data in a large non-clinical sample," *Brit. J. Clin. Psychol.*, vol. 44, no. 2, pp. 227–239, 2005.

[32] M. A. Just *et al.*, "Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth," *Nat. Hum. Behav.*, vol. 1, no. 12, 2017, Art. no. 911.

[33] K. M. Harris and M. T.-T. Goh, "Is suicide assessment harmful to participants? findings from a randomized controlled trial," *Int. J. Ment. Health Nurs.*, vol. 26, no. 2, pp. 181–190, 2017.

[34] S. R. Braithwaite *et al.*, "Validating machine learning algorithms for twitter data against established measures of suicidality," *JMIR Mental Health*, vol. 3, no. 2, 2016, Paper e 21.

[35] K. M. Harris, J. P. McLean, and J. Sheffield, "Suicidal and online: How do online behaviors inform us of this high-risk population?" *Death Stud.*, vol. 38, no. 6, pp. 387–394, 2014.

[36] J. Pestian *et al.*, "Suicide note classification using natural language processing: A content analysis," *Biomed. Informatics Insights*, vol. 3, pp. BII–S 4706, 2010.

[37] Y.-P. Huang, T. Goh, and C. L. Liew, "Hunting suicide notes in web 2.0-preliminary findings," in *Proc. 9th IEEE Int. Symp. Multimedia Workshops.*, 2007, pp. 517–521.

[38] T. M. Li *et al.*, "Collective intelligence for suicide surveillance in web forums," in *Proc. Pacific-Asia Workshop Intell. Secur. Infor*. Berlin, Germany: Springer, 2013, pp. 29–37.

[39] N. Masuda, I. Kurahashi, and H. Onari, "Suicide ideation of individuals in online social networks," *PLoS One*, vol. 8, no. 4, pp. 1–8, 2013.

[40] M. De Choudhury *et al.*, "Discovering shifts to suicidal ideation from mental health content in social media," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 2016, pp. 2098–2110.

[41] H.-C. Shing, P. Resnik, and D. W. Oard, "A prioritization model for suicidality risk assessment," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 8124–8137.

[42] N. Jones *et al.*, "Analysis of online suicide risk with document embeddings and latent dirichlet allocation," in *Proc. 8th Int. Conf. Affect. Comput. Intell. Interact. Workshops Demos.*, 2019, pp. 1–5.

[43] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of suicide ideation in social media forums using deep learning," *Algorithms*, vol. 13, no. 1, pp. 1–19, 2020.

[44] J. Jashinsky *et al.*, "Tracking suicide risk factors through twitter in the US," *Crisis*, vol. 35, no. 1, pp. 51–59, 2014.

[45] H. Sueki, "The association of suicide-related twitter use with suicidal behaviour: A cross-sectional study of young internet users in japan," *J. Affect. Disord.*, vol. 170, pp. 155–160, 2015.

[46] F. Ren, X. Kang, and C. Quan, "Examining accumulated emotional traits in suicide blogs with an emotion topic model," *J. Biomed. Health Inform.*, vol. 20, no. 5, pp. 1384–1396, 2015.

[47] 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," 2014, *arXiv:1407.0466*.

[48] L. Zhang *et al.*, "Using linguistic features to estimate suicide probability of chinese microblog users," in *Proc. Int. Conf. Human Centered Comput.*, 2015, pp. 549–559.

[49] X. Huang *et al.*, "Topic model for identifying suicidal ideation in chinese microblog," in *Proc. 29th Pacific Asia Conf. Lang., Infor. Comput.*, 2015, pp. 553–562.

[50] L. Guan, B. Hao, Q. Cheng, P. S. Yip, and T. Zhu, "Identifying chinese microblog users with high suicide probability using internet-based profile and linguistic features: Classification model," *JMIR Mental Health*, vol. 2, no. 2, 2015, Paper e 17.

[51] B. O'Dea, S. Wan, P. J. Batterham, A. L. Calear, C. Paris, and H. Christensen, "Detecting suicidality on twitter," *Internet Interv.*, vol. 2, no. 2, pp. 183–188, 2015.

[52] G. Coppersmith, R. Leary, E. Whyne, and T. Wood, "Quantifying suicidal ideation via language usage on social media," in *Proc. Joint Stat. Meetings Proc., Stat. Comput. Sect.*, 2015, pp. 1–15.

[53] J. Parraga-Alava, R. A. Caicedo, J. M. Gómez, and M. Inostroza-Ponta, "An unsupervised learning approach for automatically to categorize potential suicide messages in social media," in *Proc. IEEE 38th Int. Conf. Chilean Comput. Sci. Soc.*, 2019, pp. 1–8.

[54] S. Fodeh *et al.*, "Using machine learning algorithms to detect suicide risk factors on twitter," in *Proc. IEEE Int. Conf Data Mining Workshops*, 2019, pp. 941–948.

[55] S. Ji, X. Li, Z. Huang, and E. Cambria, "Suicidal ideation and mental disorder detection with attentive relation networks," 2020, *arXiv:2004.07601*.

[56] A. Malhotra and R. Jindal, "Multimodal deep learning based framework for detecting depression and suicidal behaviour by affective analysis of social media posts," *EAI Endorsed Trans. Pervasive Health Technol.*, vol. 6, no. 21, pp. 1–9, 2020. [Online]. Available: https://doi.org/10.4108/eai.13-7-2018.164259

[57] H. A. Bouarara, "Detection and prevention of twitter users with suicidal self-harm behavior," *Int. J. Knowl.-Based Org.*, vol. 10, no. 1, pp. 49–61, 2020.

[58] R. Sawhney, P. Manchanda, R. Singh, and S. Aggarwal, "A computational approach to feature extraction for identification of suicidal ideation in tweets," in *Proc. ACL Student Res. Workshop*, 2018, pp. 91–98.

[59] L. Cao *et al.*, "Latent suicide risk detection on microblog via suicide-oriented word embeddings and layered attention," in *Proc. Conf. Empir. Methods Nat. Lang. Process.*, 2019, pp. 1718–1728.

[60] R. Mishra *et al.*, "Snap-batnet: Cascading author profiling and social network graphs for suicide ideation detection on social media," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Student Research Workshop*, 2019, pp. 147–156.

[61] N. A. Cooperman and J. M. Simoni, "Suicidal ideation and attempted suicide among women living with hiv/aids," *J. Behav. Med.*, vol. 28, no. 2, pp. 149–156, 2005.

[62] C. Steele, "The psychology of self-affirmation: Sustaining the integrity of the self," *Adv. Experimental Social psychol.*, vol. 21, no. 2, pp. 261–302, 1988.

[63] A. R. Rich and R. L. Bonner, "Concurrent validity of a stress-vulnerability model of suicidal ideation and behavior: A. follow-up study," *Suicide Life - Threatening Behav.*, vol. 17, no. 4, pp. 265–270, 1987.

[64] V. Blüml, N. D. Kapusta, S. Doering, E. Brähler, B. Wagner, and A. Kersting, "Personality factors and suicide risk in a representative sample of the german general population," *PlOS One*, vol. 8, no. 10, 2013, paper no. e 76646.

[65] T. S. Greenspon, "Is there an antidote to perfectionism?" *Psychol. Sch.*, vol. 51, no. 9, pp. 986–998, 2014.

[66] M. Lv, A. Li, T. Liu, and T. Zhu, "Creating a chinese suicide dictionary for identifying suicide risk on social media," Peerj, 2015. [Online]. Available: https://peerj.com/articles/1455/

[67] L. Mandelli *et al.*, "Neuroticism, social network, stressful life events: Association with mood disorders, depressive symptoms and suicidal ideation in a community sample of women," *J. Psychiatry Res.*, pp. 38–44, 2014.

[68] Q. Li *et al.*, "Analyzing and identifying teens stressful periods and stressor events from a microblog," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 5, pp. 1–15, 2017.

[69] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in twitter," in *Proc. Workshop Comput. Linguistics Clin. Psychol.: From Linguistic Signal Clin. Reality*, 2014, pp. 51–60.

[70] H. Xiao, "Bert-As-Service," 2018. [Online]. Available: https://github.com/hanxiao/bert-as-service

[71] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[72] C. Du, H. Sun, J. Wang, Q. Qi, and J. Liao, "Adversarial and domain-aware bert for cross-domain sentiment analysis," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 4019–4028.

[73] Y.-C. Chen, Z. Gan, Y. Cheng, J. Liu, and J. Liu, "Distilling knowledge learned in bert for text generation," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 7893–7905.

[74] Z.-J. Zha, J. Liu, D. Chen, and F. Wu, "Adversarial attribute-text embedding for person search with natural language query," *IEEE Trans. Multimedia*, vol. 22, no. 7, pp. 1836–1846, Jul. 2020.

[75] P. Dai, H. Zhang, and X. Cao, "Deep multi-scale context aware feature aggregation for curved scene text detection," *IEEE Trans. Multimedia*, vol. 22, no. 8, pp. 1969–1984, Aug. 2019.

[76] J. Sabbath, "The suicidal adolescent: The expendable child," *J. Amer. Acad. Child Psychiatry*, 1969, pp. 272–285.

[77] M. Lv, A. Li, T. Liu, and T. Zhu, "Creating a chinese suicide dictionary for identifying suicide risk on social media," *PeerJ*, vol. 3, 2015, Paper e 1455.

[78] C.-L. Huang *et al.*, "The development of the chinese linguistic inquiry and word count dictionary." *Chin. J. Psychol.*, 2012, pp. 185–201.

[79] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. 5th Int. Conf. Learn. Represent.*, 2017, pp. 1–14.

[80] W. L. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, Dec. 4–9, 2017, pp. 1024–1034. [Online]. Available: http://papers.nips.cc/paper/6703-inductive-representation-learning-on-l% arge-graphs

[81] N. Darling and L. Steinberg, "Parenting style as context: An integrative model," *Psychol. Bull.*, vol. 113, no. 3, pp. 487–496, 1993.

[82] K. Lai and C. Mcbride-Chang, "Suicidal ideation, parenting style, and family climate among hong kong adolescents," *Int J. Psychol.*, vol. 36, no. 2, pp. 81–87, 2001.

**Huijun Zhang** is the Ph.D. candidate with the Department of Computer Science and Technology, Tsinghua University, Beijing, China. Her research interests include computational psychology and sentiment analysis.

**Ling Feng** is a Professor of Computer Science and Technology with Tsinghua University, China. Her research interests include computational mental healthcare, context-aware data management and services toward ambient intelligence, data mining and warehousing, and distributed object-oriented database systems.

**Lei Cao** is the Ph.D. candidate with the Department of Computer Science and Technology, Tsinghua University, Beijing, China. His research interests include computational psychology and sentiment analysis.