PACIS 2023 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

7-8-2023

Suicide Risk Prediction for Users with Depression in Question Answering Communities: A Design Based on Deep Learning

Haoyue Fan
Beijing Institute of Technology, xkfanhaoyue@126.com

Qiuju Yin
Beijing Institute of Technology, yinqiuju@bit.edu.cn

Zhijun Yan
Beijing Institute of Technology, yanzhijun@bit.edu.cn

Junwei Kuang
Beijing Institute of Technology, kuang_junwei@163.com

Suicide Risk Prediction for Users with Depression in Question Answering Communities: A Design Based on Deep Learning

Completed Research Paper

Haoyue Fan
Beijing Institute of Technology
Zhongguancun South Street, Haidian District, Beijing, 100081, P.R. China
xkfanhaoyue@126.com

Qiuju Yin
Beijing Institute of Technology
Zhongguancun South Street, Haidian District, Beijing, 100081, P.R. China
yinqiuju@bit.edu.cn

Zhijun Yan
Beijing Institute of Technology
Zhongguancun South Street, Haidian District, Beijing, 100081, P.R. China
yanzhijun@bit.edu.cn

Junwei Kuang
Beijing Institute of Technology
Zhongguancun South Street, Haidian District, Beijing, 100081, P.R. China
kuang_junwei@163.com

Abstract

*In the field of public health, suicide risk prediction is a central and urgent problem. Existing researches mainly focus on user's current post but overlook historical post. In light of the psychological characteristics, we argue that it is valuable to consider users' historical post in addition to current post for predicting suicide risk. Based on this rationale, we propose a deep learning-based suicide risk prediction framework - **D**ynamic **H**istorical **I**nformation based **S**uicide **R**isk **P**rediction (DHISRP) - by considering the user's current post content and historical post content. To capture the dynamic and complicated information of historical post, we design a unit based on long short-term memory (LSTM), named RNLSTM. We also conduct experiments to compare with the benchmark model to prove the effectiveness of our model, and perform ablation experiments to verify the significance of each component in the prediction framework in this study.*

Keywords: Suicide risk prediction; Deep learning; Text mining; LSTM

Introduction

Depression has received much attention in recent years as a result of common and serious mental disease. Due to the stigmatization of depression, many patients with depression worry about social discrimination and are reluctant to admit their identity of depression (Descalzi et al. 2017). Patients with depression are very likely to seek help anonymously on social media rather than directly from someone close to them (Cheng et al. 2017). Anonymous communication on social media helps to improve the willingness of personal privacy information disclosure (Posey et al. 2010). However, about 200,000 people in China still commit suicide to escape depression in every year (Huang et al. 2019). The WHO claimed that early identification of suicide risk has been developed as a national suicide prevention strategy (WHO 2018).

Thus, depressed patients at risk of suicide could be identified earlier based on their data on social media. Early identification of depressed patients with suicidal risk and providing appropriate timely intervention by psychologists are critical to preventing patients' suicide.

On social media, patients with depression can anonymously post their concerns without worry about stigmatization of depression; and psychologists as well as other users can provide information support and emotional support to them, which is helpful to improve their mental condition. However, many countries are facing a shortage of psychologists, especially children psychologists (Wu and Pan 2019). During the medical treatment of patients with depression, psychologists also may not effectively track the changes of patients' psychological state, and thus cannot treat them in time. Especially, it is impossible for psychologists to always pay attention to the problems posted by patients on social media. The application of artificial intelligence techniques to prediction problems offers chances to timely identify depressed patients with suicidal risk. The content posted by users on social media can reflect their emotions. For people with depression, their emotional distresses are fluctuating. We can consider users' latest post and historical posts together to more accurately understand users' psychological state, which is helpful for predicting users' suicide risk.

In order to prevent suicide of patients with depression, researchers have begun to identify severe depression through negative self-expression on social media and provide timely professional medical support. However, the existing research on suicidal risk prediction has a few limitations. First, existing literatures mainly identify suicide factors by the examination of psychological, and demographic variables of individuals through assessment methods, such as Adult Suicide Ideation Questionnaire (Fu et al. 2007), Suicide Probability Scale (SPS) (Bagge and Osman 1998), Depression Anxiety Stress Scales-21 (DASS-21) (Henry and Crawford 2011) and so on. In these studies, people's mental situation data are collected by questionnaires or interviews, which is a complex and difficult task, especially for those who have suicide intention. More importantly, people with suicide risk are reluctant to seek the help of psychologists (Bernert et al. 2020), and their mental situation data can hardly be collected by questionnaires or interviews. Second, previous suicide risk prediction methods only predict suicide risk based on people's current online post, and people's historical post information is ignored. However, the mental state of people will not change suddenly and greatly (Choudhury et al. 2016). People usually have been in emotional distress for a certain time before they have suicidal thoughts. Thus, it is necessary to consider people historical psychological status and incorporate people's online historical post information to predict suicide risk. Third, machine learning (ML) technology has been adopted in previous studies to predict suicide risk and obtain good results (Jie et al. 2021; Rink and Harabagiu 2010; Zhang and Wang 2015), but traditional ML models can only capture the nature from the post-level to predict, and the dynamic and complex characteristics of users' posts sequence is ignored. Therefore, this study aims to propose a new innovative suicide risk prediction method based on deep learning methods.

To address the above limitations of existing research, we propose a deep learning-based prediction method by considering users' current post content and historical post content. Taking into account users' current posts and historical posts at the same time can more comprehensively reflect users' mental state, which helps to better predict people's suicide risk. At the same time, current posts can better reflect users' latest psychological situation than historical posts, and they definitely play different roles in predicting suicide risk. In order to distinguish the difference between current posts and historical posts, we propose a two-level deep learning method based on long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997) and adopt the attention mechanism (Bahdanau et al. 2014).

The major contributions of this study are as follows. First, from a theoretical perspective, we contribute to IS methodology by designing a suicide risk prediction method that integrates users' current posts and historical posts information. Second, we propose a deep learning model, Dynamic Historical Information Based Suicide Risk Prediction, which includes an improved LSTM, named RNLSTM, to predict suicide risk. DHISRP model performs better than other benchmark methods significantly. Third, from a practical perspective, the proposed prediction method can automatically identify users with suicide risk based on based on their posts, which can help to reduce the labor cost of manual tagging of risky users in social media. Overall, this study emphasizes the importance of considering both current and historical post information to predict suicide risk for people with depression and provides a promising deep learning-based method to achieve this.

Literature Review

Suicide Risk Prediction Using Social Media

In recent years, a large number of scholars have devoted themselves to preventing the risk of suicide by analyzing the content posted on social media by people with depression. As summarized in Table 1, most researches focused on texts posted on social media (Chau et al. 2020), although other researches obtained data through questionnaires (Lin et al. 2020), short message service (SMS) (Nobles et al. 2018) and electronic health records (Simon et al. 2018). Research data for suicide risk prediction can be classified into two categories: static and dynamic data. Static data refers to cross-sectional online text data or accessible data. Existing studies were mainly based on static text data. For example, Shah et al. (Shah et al. 2020) proposed a hybrid method to create feature sets based on computational features and linguistic features in order to detect suicidal ideation in static data from Reddit website. Cheng et al. (Cheng et al. 2017) concluded Chinese related to suicide risk or emotional distress, and used SVM method to identify suicide risk based on Weibo posts. Dynamic data refers to multi-period online text data or medical record. For example, Simon et al. (Simon et al. 2018) used a logistic regression model to study electronic health records of many years to predict the risk of suicide. Choudhury et al. (Choudhury et al. 2016) applied the propensity matching score method to identify suicidal ideation based on uses' posts on Reddit website, and found factors associated with suicidal ideation through causal inference methods.

From the perspective of method category, most of the previous studies used traditional machine learning method, that is, classifying data through feature engineering. Support vector machine, Bayesian and logistic regression are common methods adopted in suicide risk prediction (Birjali et al. 2017; Desmet and Hoste 2018; Lin et al. 2020). In recent years, with the advent of the deep learning boom, some scholars have begun to use deep learning methods to predict suicide risk. Different from traditional machine learning methods, deep learning methods do not need to extract text features, but encode the text into a computer recognizable language and then directly predicts. For example, Chau et al. (Chau et al. 2020) applied genetic algorithm (GA) to feature selection in feature engineering, and used SVM and rule-based classification to identify emotionally distressed users in Chinese social media. Nobles et al. (Nobles et al. 2018) proposed a deep neural net (DNN) to study the language differences of users with or without suicide risk based on short message service. Tadesse et al. (Tadesse et al. 2019) combined LSTM and CNN into a new model and applied it to detect suicidal ideation in Reddit website. The experimental results show that the effects of a single LSTM or CNN, or the composite model LSTM-CNN are superior to the traditional machine learning methods based on feature engineering.

Author	Method Category	Methods	Representation	Data
Birjali et al. (2017)	Machine Learning	SVM	Static	Online social text
Choudhury et al. (2016)	Statistical	Propensity Score Matching	Dynamic	Online social text
Desmet et al. (2018)	Machine Learning	SVM	Static	Online social text
Simon et al. (2018)	Machine Learning	Logit, Lasso	Dynamic	Electronic health records
Cheng et al. (2017)	Machine Learning	SVM	Static	Online social text
Lin et al. (2020)	Machine Learning	Logit, SVM, Decision Tree	Static	Questionnaires
Shah et al. (2020)	Machine Learning	SVM, Bayes	Static	Online social text
Nobles et al. (2018)	Deep Learning	DNN	Static	SMS
Tadesse et al. (2019)	Deep Learning	LSTM, CNN	Static	Online social text
Chau et al. (2020)	Machine Learning	SVM, GA	Static	Online social text

Our study	Deep Learning	LSTM	Dynamic	Online social text

Table 1. ***Summary of the literature about suicide risk prediction on social media***

Approach with Deep Learning

As a frontier technology of artificial intelligence, deep learning has made tremendous progress in the past decade. Different from other methods of artificial intelligence, deep learning is characterized by deep autonomous learning based on neurons to conduct a variety of prediction tasks, such as image sequence representation (Goodfellow et al. 2013), machine translation (Sutskever et al. 2014), and text analysis (Zhang et al. 2021). Meanwhile, deep learning has been successfully applied in various fields and achieved good results, such as, finance (Liu et al. 2020), medical care (He et al. 2021), and transportation, etc. In the field of IS, existing studies usually used some frameworks constructed by deep learning basic models, such as CNN, RNN, LSTM and so on, to achieve learning tasks in different scenarios. For example, Tadesse et al. (Tadesse et al. 2019) applied a framework, which was combined by LSTM and CNN, to identify suicidal ideation.

With the further development of deep learning, some scholars began to combine it with optimization algorithms or attention mechanism to improve the accuracy and interpretability of deep learning models. He et al. (He et al. 2021) improved the deep learning model LSTM according to multi-attention at three level of feature, visit and user. Ahmad et al. (Ahmad et al. 2020) combined CNN, BILSTM and structural equation model (SEM) for psychometric measures including novel representation and demographic embedding, and a multitask learning mechanism. Others have improved the basic model according to the characteristics of the data.

In this study, we mainly focus on improvements to the basic model LSTM. LSTM is a special kind of neural networks, which is designed to learn long-term dependency for sequential data. It is introduced by Hochreiter and Schmidhub (Hochreiter and Schmidhuber 1997), which has been improved and popularized in many fields. LSTM includes five components, which are input gate, output gate, forget gate, memory cell and hidden state. In order to better apply LSTM to dynamic historical information, Pham et al. (Pham et al. 2017) and Baytas et al. (Baytas et al. 2017) proposed two different time-aware LSTMs, which added a decreasing function to the forget gate and memory cell of LSTM respectively. Differently, Xie et al. (Xie et al. 2021) improved LSTM by added increasing and decreasing function to three gates.

Previous studies of suicide risk prediction have some limitations in the following two aspects. First, most research on suicide risk prediction based on social media platforms have mainly focused on static online social text. Based on the confirmation of existing research in the field of psychology, the single-issue online social text is one-sided that cannot fully reflect the suicide risk of users. It is necessary to consider people's online historical post information to predict suicide risk. Second, neither basic machine learning methods nor deep learning methods can accurately capture the complex features of dynamic historical posts. Therefore, in this study, we focus on dynamic online social text to more accurately predict whether users have a suicidal risk, and adopt an improved LSTM method based psychological theory which is good at capturing sequence characteristics.

A New Suicide Risk Prediction Method

In this section, we formulate the suicide risk prediction problem, and detail our proposed DHISRP method. Figure 1 illustrates the overall framework of the proposed method. DHISRP includes two main components: content encoder and user encoder. The function of content encoder module is to convert the text content of users' posts into digital vectors through deep learning methods. User encoder module is used to integrate the results of a user's current post and historical posts obtained by content encoder.

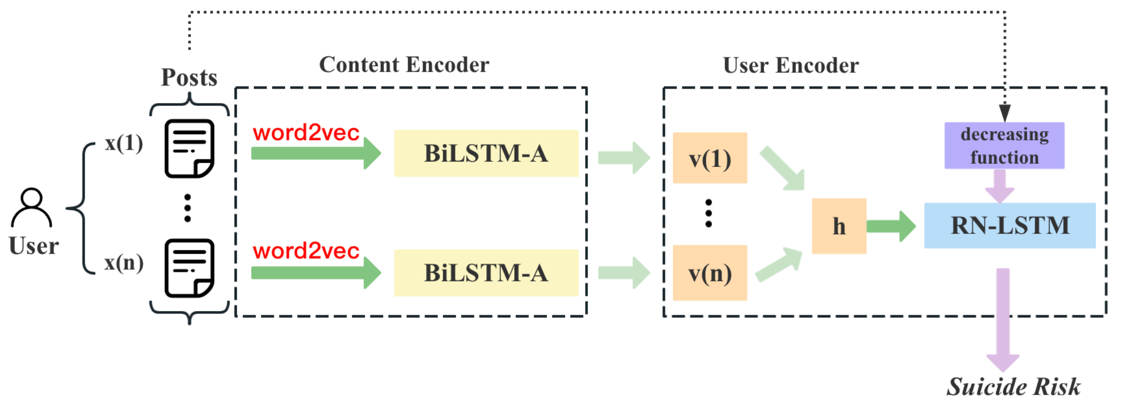

Figure 1. The Framework of DHISRP Suicide Risk Prediction Method

Problem Definition

Let U denote a set of users, where each user is denoted by u_i ($u_i \in U$, i = 1, 2, ..., U). In the entire time windows, a user u_i may have posted M_i ($M_i \geq 1$) questions. We denote the posted question sequence of user u_i as $x_i = < x^1{}_i, x^2{}_i, ..., x^{M}{}_i>$, where x^{M_i} represents the newest posted question by user u_i. Each posted question consists of N words, and the words sequence of the posted question x^{M_i} is denoted as $x^{M_i} =< w^1{}_{iM}, w^2{}_{iM}, ..., w^N{}_{iM} >$, where represents the whole words in the question x^{M_i}.

Given a posted question x^{M_i} question posted by user u_i, we can observe a sequence of posted questions by user u_i. Therefore, the objective of this study is to predict whether user u_i who having posted the last question $x^{M}{}_i$ isthe posted question x^{M_i} at risk of suicideis risky based on the posted question sequence.

Content Encoder

The suicide behavior caused by emotional distress will not suddenly emerge. Users may have emotional change before the suicide crisis, and seek help by posting on social media. In order to more accurately predict the suicide risk indicated by the latest posts, we comprehensively consider users' emotional changes by combining their historical posts with latest posts, and then define the first part of DHISRP method, content encoder module. As shown in Figure 2, the representation of users' posts mainly includes three layers: word embedding, BLSTM, attention mechanism. All posts go through three layers of content encoder module respectively, and finally merged into a potential representation.

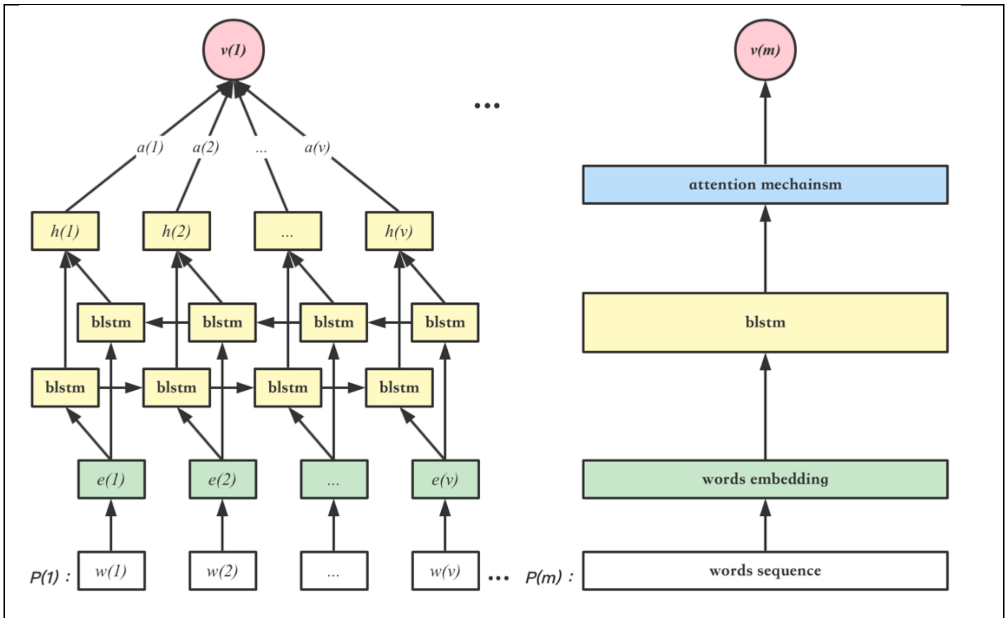

Figure 2. The Representation of Users' Posts

In word embedding layer, users' posts are transformed into low dimensional vectors. First, we use word segmentation technology to preprocess users' posts, and get the words sequence $x = < w_1, w_2, \ldots, w_N >$ of each post. Then Word2Vec is adopted to vectorize the words in word sequences, and obtain word vectors $e = < e_1, e_2, \ldots, e_N >$ of each post.

In BLSTM layer, the deep learning model BLSTM is used to extract semantic features of posts. To more sufficiently capture posts' contextual information, we choose BLSTM, which gets the context of words from both forward and backward directions, instead of LSTM. The input of BLSTM model is the result word vectors $\boldsymbol{e}$ of word embedding layer. Through the learning of forward and backward semantic features in BLSTM model, the forward hidden features $\overrightarrow{\boldsymbol{h}} = < \overrightarrow{h^1}, \overrightarrow{h^2}, \ldots, \overrightarrow{h^N} >$ and the backward hidden features $\overleftarrow{\boldsymbol{h}} = < \overleftarrow{h^1}, \overleftarrow{h^2}, \ldots, \overleftarrow{h^N} >$ are obtained. Then, we get hidden features $\boldsymbol{h} = < h^1, h^2, \ldots, h^N >$ of each post by element-wise sum to connect forward and backward hidden features.

In the attention layer, the final representation of each post is obtained by using the attention mechanism to weight their hidden features $\boldsymbol{h}$. The amount of textual data posted by users is large, but most of the data are redundant for text classification, and only a small part of the data is crucial (Bahdanau et al. 2014). Thus, it is necessary to apply attention mechanisms to capture more important features. Firstly, we compute feature attention weight $\boldsymbol{a} = < a^1, a^2, \ldots, a^N >$ as Equation (1). Then, the weighted hidden features $\boldsymbol{v}$ can be calculated as Equation (2):

$$\boldsymbol{a} = softmax\,(\,\boldsymbol{W}_a \times \tanh\,(\,\boldsymbol{h}\,)) \tag{1}$$

$$\boldsymbol{v} = \boldsymbol{h} \times \boldsymbol{a}^T \tag{2}$$

Then, the weighted hidden features of all words are concatenated as the final representation of each post and outputted from the content encoder.

User Encoder

Generally, the content of the latest post can better reflect users' current emotional state. However, due to the limitations of language expression, the content of one latest post cannot comprehensively represent a user's emotional state. We have to consider users' current post and historical posts together to more accurately understand users' psychological state, which is helpful for predicting users' suicide risk. At the same time, two characteristics of historical posts are considered. First, the latest post plays more important role in suicide risk prediction than historical posts, because the latest post has more salient recency effect. Second, for patients with depression, the more replies to their posted questions, the better for them. The effectiveness of different replies to historical posts in predicting suicide risk is diverse (Zhou et al. 2022). Therefore, we design an improved LSTM to capture various potencies of posted questions from users' question sequences, named Response-Number based LSTM (RNLSTM).

In the RNLSTM, we design a monotonic decreasing function $e^{-(n-\lambda)}$ to weight historical posts according to the number of replies received. When a user's posts are replied, his or her emotional distress will be alleviated, and more replies are more helpful for them. We assume that the number of replies to a post expected by the user is λ, and the number of responses actually received is n. When n is less than λ, the function $e^{-(n-\lambda)}$ will strengthen the impact of the post, otherwise attenuate its impact. We modify gate mechanism of the basic LSTM by adding the function $e^{-(n-\lambda)}$ to each historical post. Different from the basic LSTM, the weight parameters of the hidden state for each historical post in the RNLSTM are not shared. The specific equations are as follows:

$$\boldsymbol{i}^{(t)} = \sigma(\boldsymbol{W}_{xi}\,\boldsymbol{v}^{(t)} + \boldsymbol{W}_{hi}(\boldsymbol{h}^{(t-1)} \times e^{-(n-\lambda)}) + \boldsymbol{b}_i) \tag{3}$$

$$\boldsymbol{f}^{(t)} = \sigma(\boldsymbol{W}_{xf}\,\boldsymbol{v}^{(t)} + \boldsymbol{W}_{hf}(\boldsymbol{h}^{(t-1)} \times e^{-(n-\lambda)}) + \boldsymbol{b}_f) \tag{4}$$

$$\boldsymbol{o}^{(t)} = \sigma(\boldsymbol{W}_{xo}\,\boldsymbol{v}^{(t)} + \boldsymbol{W}_{ho}(\boldsymbol{h}^{(t-1)} \times e^{-(n-\lambda)}) + \boldsymbol{b}_o) \tag{5}$$

$$\boldsymbol{c}^{(t)} = \boldsymbol{i}^{(t)} \circ tanh\,(\boldsymbol{W}_{xc}\,\boldsymbol{v}^{(t)} + \boldsymbol{W}_{hc}\,\boldsymbol{h}^{(t-1)} + \boldsymbol{b}_u) + \boldsymbol{f}^{(t)} \circ \boldsymbol{c}^{(t-1)} \tag{6}$$

$$\boldsymbol{h}^{(t)} = \boldsymbol{o}^{(t)} \circ tanh\,(\boldsymbol{c}^{(t)}) \tag{7}$$

where W_x, W_h and b are weight parameters corresponding to the multiplicative terms with values between 0 and 1. σ (·) and $tanh$ (·) are two different activation functions. The operator ∘ is element wise multiplication. The three gates, input gate, forget gate and output gate, are introduced the attenuation function $e^{-(n-\lambda)}$ to amplify or discount the hidden state $h^{(t-1)}$ of the past information from time step t-1. The memory cell $c^{(t)}$ and the hidden state $h^{(t)}$ from time step t will improve with the modification of the coefficient for the hidden state $h^{(t-1)}$ in three gates.

Evaluation

In this section, we have conducted an empirical evaluation of the proposed DHISRP method. We compared the experimental results of DHISRP and that of benchmark methods utilizing the real social media data.

Data Set

In this study, we collected user Q&A information from YiXinLi platform (www.xinli001.com), one of mental question answering communities (MQACs) in China. YiXinLi platform provides online solutions for people who need psychological help, and facilitates the promotion of psychological institutions. By 2022, it has more than 20 million users, and has established close cooperative relationships with about 790 psychological institutions across the country. There have many online sections divided by different topics on YiXinLi platform, and one of which is about suicide relief topic. Users can post their questions or emotional distress in this section and ask users for help, as well as directly ask questions to some specific psychologists. According to the content of posts, employees of the platform mark the dangerous posts as suicide risk.

We chose online suicide relief section as the research context, and used Python to crawl the data of user Q&A information posted in this section from January 2020 to March 2021. This dataset comprises 3687 posted questions, including posted content, time, number of replies. Then, we obtained 1019 posted questions by removing the questions posted by YiXinLi platform employees from the sampled datasets. Of the 1019 data samples, each user posted an average of 2-3 posts, and 356 posts were labeled as suicide risk. Employees of the YiXinLi platform marked the dangerous posts as suicide risk based on the content of posts.

Baseline Models

To systematically evaluate the effectiveness of our proposed model, we select a set of standard baseline methods to compare, including traditional machine learning methods and alternative deep learning methods. All selected baseline methods are most prevalent in existing literatures. For traditional machine learning methods, we first vectorize Chinese word segmentations of posts using the N-gram mode. Then, we choose Gaussian Naïve Bayes (NB), logistic regression (LR), and support vector machine (SVM) to classification, respectively. For alternative deep learning methods, we first vectorize word segmentations of posts by word embedding of Word2Vec model. Then, we apply the recurrent neural network (RNN), the long short-term memory (LSTM), the Bi-directional LSTM (BLSTM) and the gated recurrent unit (GRU), respectively.

Evaluation metrics

To evaluate the performance of these methods in predicting suicide risk, we choose four metrics that have been commonly used for binary classification models from prior research (Belsher et al. 2019; Chau et al. 2020; Tadesse et al. 2019), including precision (P), recall (R), F1-score (F), and accuracy (A). These metrics evaluate models from different dimensions. Precision measures the proportion of posts the model classified as having suicide risk that actually have suicide risk; recall is the proportion of actually posts having suicide risk in the datasets that the model can identify; F1-score is a comprehensive evaluation metric of precision and recall; accuracy assesses the proportion of correctly predicted posts to the total datasets. Definitions of four metrics are as follows,

$$\text{precision} = TP / (TP+FP)$$

$$\text{recall} = TP / (TP+FN)$$

$$\text{F1-score} = 2^* \text{precision}^* \text{recall} / (\text{precision} + \text{recall})$$

$$\text{accuracy} = (TP+TN) / (TP+FP+TN+FN)$$

where TP denotes the number of posts correctly identified as suicide risk; FP denotes the number of posts without suicide risk that are incorrectly classified as suicide risk; TN is the number of correctly classified no suicide risk posts, and FN is the number of posts with suicide risk that are identified as no risk.

Results

Experimental results

We randomly split each dataset into 75% for training, 25% for validation. For fair comparison, we perform random search for each method to optimize hyperparameter selection. All results are averaged over ten independent runs of the experiments, with a different data splitting for each run.

We first compare our proposed DHISRP with traditional machine learning methods and summarize the results in Table 2. We repeated ten experiments for each method to perform paired-sample t-tests to verify the statistical significance of our model DHISRP outperforming other models. Table 3 shows the results of t-tests between DHISRP and those traditional machine learning methods. From the results, we can conclude that our DHISRP outperforms all traditional machine learning methods across almost all evaluation metrics, which demonstrates the superiority of our model. The recall is 24.43% ($p < 0.001$) higher than that of NB model with the best result among three baseline models, the F1 score is 6.32% ($p < 0.001$) higher than that of SVM, and the accuracy is 3.72% ($p < 0.01$) higher than that of LR and SVM. Serious depression may lead to suicide due to untimely counseling. The 24.43% improvement in recall could potentially save the lives of many depression patients, if appropriate prevention strategies are implemented. Even though DHISRP has relatively lower precision than SVM model, misclassifying a few posts to be at suicide risk could not pose many adverse effects.

Methods	Precision	Recall	F1-score	Accuracy
NB	57.14	46.38	51.20	71.09
LR	57.35	56.52	56.94	72.04
SVM	**66.67**	28.99	40.40	72.04
DHISRP (Ours)	57.63	**82.14**	**66.35**	**73.49**

Table 2. DHISRP Versus Traditional Machine Learning Models (%)

Methods	Precision	Recall	F1-score	Accuracy
DHISRP vs NB	0.111	< 0.001***	< 0.001***	< 0.001***
DHISRP vs LR	0.469	< 0.001***	< 0.001***	< 0.01**
DHISRP vs SVM	< 0.01**	< 0.001***	< 0.001***	< 0.01**

Table 3. *P*-value of T-test for DHISRP Against Traditional Machine Learning Models

In addition, we compare our model with other deep learning methods and show the results in Table 4. The results of paired-sample t-tests for DHISRP against baseline deep learning models are given in Table 5. We can conclude that our DHISRP outperforms all these deep learning methods across four evaluation metrics ($p < 0.001$ for all methods and metrics), which shows the effectiveness of our model due to its framework integrating historical and current information. The precision is 1.81% higher than that of GRU model with the best result among four deep learning baseline models, the recall is 14.28% higher than that of BLSTM, the F1 score is 6.66% higher than that of BLSTM, and the accuracy is 2.15% ($p < 0.01$) higher than that of LSTM. Compared with traditional machine learning methods, all deep learning baseline methods performs significantly better on the recall, and DHISRP employs the structure of deep learning to improve the predicting ability.

Methods	Precision	Recall	F1-score	Accuracy
RNN	45.45	61.91	52.41	68.18
LSTM	48.93	64.29	59.67	71.34
BLSTM	53.27	67.86	59.69	70.83
GRU	53.84	66.67	59.57	71.21
DHISRP (Ours)	**57.63**	**82.14**	**66.35**	**73.49**

Table 4. DHISRP Versus Alternative Deep Learning Models (%)

Methods	Precision	Recall	F1-score	Accuracy
DHISRP vs RNN	< 0.001***	< 0.001***	< 0.001***	< 0.001***
DHISRP vs LSTM	< 0.001***	< 0.001***	< 0.001***	< 0.001***
DHISRP vs BLSTM	< 0.001***	< 0.001***	< 0.001***	< 0.001***
DHISRP vs GRU	< 0.001***	< 0.001***	< 0.001***	< 0.001***

Table 5. *P*-value of T-test for DHISRP Against Alternative Deep Learning Models

Ablation Study

In this paper, DHISRP is an integrated method composed of content encoder and user encoder, which is designed by some deep learning components. We conduct an ablation study to investigate the performance influenced by each component in DHISRP:

BLSTM scanned word sequence of the current post, and used the basic BLSTM for classification.

A-BLSTM extended BLSTM by adding attention, which can capture more important word segmentations for classification.

BLSTM-LSTM scanned each users' post separately using the basic BLSTM, and then integrated the current post and historical posts by the basic LSTM for classification.

A-BLSTM-LSTM extended BLSTM-LSTM by adding attention in the module of BLSTM.

BLSTM-RNLSTM improved BLSTM-LSTM by substituting RNLSTM presented in Figure 1 for the basic LSTM.

The outcome of the ablation study is listed in Table 6. The results of paired-sample t-tests for DHISRP against each component in DHISRP are given in Table 7. BLSTM and was significantly inferior to most of other models in all evaluation metrics, which may be due to only the current posting content of the user is contained, making BLSTM inefficient in identifying complicated depression emotion. Compared with BLSTM, A-BLSTM, which added a layer of attention, achieved improvement especially in Recall metric (increase 1.19%), indicating of the advantage of the attention mechanism on embedding sentences by word-level. BLSTM and A-BLSTM used only the current posting content to capturing the features of sentences.

Methods	Precision	Recall	F1-score	Accuracy
BLSTM	51.27	67.86	59.69	70.83
A-BLSTM	52.25	69.05	59.49	70.08
BLSTM-LSTM	54.31	75.00	63.00	71.97
A-BLSTM-LSTM	55.36	75.00	63.64	72.72
BLSTM-RNLSTM	**59.09**	77.38	67.01	72.34
DHISRP (Ours)	57.63	**82.14**	**66.35**	**73.49**

Table 6. Ablation Study Results (%)

Methods	Precision	Recall	F1-score	Accuracy
Ours vs BLSTM	$< 0.001^{***}$	$< 0.001^{***}$	$< 0.001^{***}$	$< 0.001^{***}$
Ours vs A-BLSTM	$< 0.01^{**}$	$< 0.001^{***}$	$< 0.001^{***}$	$< 0.01^{**}$
Ours vs BLSTM-LSTM	$< 0.01^{**}$	$< 0.001^{***}$	$< 0.001^{***}$	$< 0.01^{**}$
Ours vs A-BLSTM-LSTM	0.064	$< 0.001^{***}$	$< 0.01^{**}$	$< 0.05^{*}$
Ours vs BLSTM-RNLSTM	0.129	$< 0.001^{***}$	$< 0.05^{*}$	$< 0.05^{*}$

Table 7. *P*-value of T-test for DHISRP Against Ablation Study Results

Different from BLSTM and A-BLSTM, BLSTM-LSTM and A-BLSTM-LSTM were two-level deep learning model, which contained historical posting content, and significantly improved in four evaluation metrics, demonstrating the positive impact of historical posting content on reducing uncertainty about depression. Similarly, when encoding text data into sentence vectors, A-BLSTM-LSTM with attention mechanism performs better than BLSTM-LSTM without attention mechanism in three evaluation metrics, which proves the effectiveness of adding attention mechanism.

BLSTM-RNLSTM was significantly superior to other models except our model DHISRP in all evaluation metrics. Component RNLSTM was an improved LSTM that codes a user by considering the current posted question and historical posted questions, especially considering the response of historical posted questions. Compared with BLSTM-LSTM and A-BLSTM-LSTM, component RNLSTM played an indispensable role in predicting suicide risk problem based on complicated text data (Recall: 2.38% more than BLSTM-LSTM; F1-score: 4.01% more than BLSTM-LSTM), and its importance exceeded the attention mechanism added in sentence vector coding (Recall: 2.38% more than A-BLSTM-LSTM; F1-score: 3.37% more than A-BLSTM-LSTM). Our model DHISRP incorporated a two-level deep learning model, attention mechanism, and the improved LSTM model, and its performance was superior to most of other models. Although the performance of BLSTM-RNLSTM was better than that of DHISRP in precision metric, the improvement of it was not statistically significant ($p = 0.129$). Overall, the above results clearly prove that DHISRP can improve the performance of suicide risk prediction models by incorporating the current posted question, historical posted questions and the response of historical posted questions.

Discussion and Conclusion

Predicting suicide risk is an important topic in the field of public health. In this paper, we propose a deep learning-based suicide risk prediction method. Accurately predicting suicide risk can not only prevent suicide, but also improve the happiness index of society.

This research makes several contributions as follows. First, we integrated users' current posts and historical posts information on social media to grasp their mental, then accurately predict suicide risk. Past researches on identifying suicide risk based on social media text have been more predictive based on a single post. A recent study empirically demonstrated that historical post information can have an impact on the negative emotion of users. Thus, on the basis of considering the current post, this study added historical post information of users, including the content of historical posts and the responses to historical posts. The analysis of the experimental results shows that the suicide risk prediction model considering the historical post information of users performs better.

Second, we designed and developed a deep learning-based suicide crisis prediction method DHISRP to improve the prediction accuracy. Particularly, DHISRP was a two-layer model based on LSTM, and each layer corresponds to the content encoder based on posts and the user encoder based on users' history posts. In the past, machine learning based suicide risk prediction method didn't integrate the user's all posts to capture suicide risk. By comparing with the state-of-the-art machine learning based suicide risk prediction method, we evaluated the effectiveness of our DHISRP method.

Third, we proposed an improved model RNLSTM based on LSTM, which improves the gate mechanism of LSTM by using the number of historical replies as the weight of historical information. RNLSTM was applied to the user encoder of DHISRP method to integrate the user's historical posting information. Traditional machine learning models were unable to dynamically consider the impact of historical posts on users' mental changes. The analysis of the ablation experiment results proves the feasibility of RNLSTM for historical posting content considering different reply situations.

This research also offers several practical implications. First of all, the suicide risk prediction model DHDIL proposed in this paper can identify people with suicide risk at an early stage. Early identification of such individuals allows for proactive intervention to avoid the irreversible situation of people suicide. Secondly, there are generally fewer psychologists, which leads to an imbalance between supply and demand, and psychologists cannot track the changes of patients in the treatment process in time. The system can monitor the emotional changes of patients in real time and implement effective treatment measures for patients in a time. Third, the suicide risk prediction model proposed in this paper can be directly used in social media to automatically predict the suicide risk of all questions posted by users, which created practical insights. It can reduce the labor cost of manual tagging in social media, and more importantly, it can avoid errors caused by manual tagging.

However, our study has some limitations, and it can be developed from three directions. First, only the number of replies to historical posts is considered in the suicide risk prediction model considering historical post information. Responses to historical posts can be further subdivided into emotional support and social support in terms of content. In future research, it would be interesting to explore how to integrate the reply content of historical posts into the framework of suicide risk prediction. Second, only the data of YiXinLi platform have been used to train the model proposed in this paper. We do not know the performance of users on other social media, such as Weibo and WeChat. Future research can correlate the data of users on YiXinLi platform with the data on other social media for suicide risk prediction. Third, the prediction framework of this study is applicable to analyzing text for suicide risk prediction. We did not verify the effectiveness of the prediction model for other research content. In practical applications, there are many demands for personalized predictions through text analysis. We can verify and improve our model, and apply it to more scenarios in the future.

References

Ahmad, F., Abbasi, A., Li, J., Dobolyi, D. G., Netemeyer, R. G., Clifford, G. D., and Chen, H. 2020. "A Deep Learning Architecture for Psychometric Natural Language Processing," *ACM Transactions on Information Systems* (38:1), pp. 1-29.

Bagge, C., and Osman, A. 1998. "The Suicide Probability Scale: Norms and Factor Structure," *Psychological Reports* (83:2), pp. 637-638.

Bahdanau, D., Cho, K., and Bengio, Y. 2014. "Neural Machine Translation by Jointly Learning to Align and Translate," *arXiv preprint arXiv:1409.0473*.

Baytas, I. M., Xiao, C., Zhang, X., Wang, F., Jain, A. K., and Zhou, J. 2017. "Patient Subtyping Via Time-Aware Lstm Networks," *23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 65-74.

Belsher, B., Smolenski, D. J., Pruitt, L. D., Bush, N. E., Beech, E. H., Workman, D. E., Morgan, R. L., Evatt, D. P., Tucker, J. A., and Skopp, N. 2019. "Prediction Models for Suicide Attempts and Deaths: A Systematic Review and Simulation," *JAMA psychiatry* (76:6), pp. 642-651.

Bernert, R. A., Hilberg, A. M., Melia, R., Kim, J. P., and Abnousi, F. 2020. "Artificial Intelligence and Suicide Prevention: A Systematic Review of Machine Learning Investigations.," *International Journal of Environmental Research and Public Health* (17:16), pp. 1-25.

Birjali, M., Beni-Hssane, A., and Erritali, M. 2017. "Machine Learning and Semantic Sentiment Analysis Based Algorithms for Suicide Sentiment Prediction in Social Networks," *Procedia Computer Science* (113), pp. 65-72.

Chau, M., Li, T. M. H., Wong, P. W. C., Xu, J. J., Yip, P. S. F., and Chen, H. 2020. "Finding People with Emotional Distress in Online Social Media: A Design Combining Machine Learning and Rule-Based Classification," *Management Information Systems Quarterly* (44:2), pp. 933-955.

Cheng, Q., Li, T. M. H., Kwok, C. L., Zhu, T., Yip, P. S. F., and Cheng, Q. 2017. "Assessing Suicide Risk and Emotional Distress in Chinese Social Media: A Text Mining and Machine Learning Study," *Journal of Medical Internet Research* (19:7), p. e243.

Choudhury, M. D., Kiciman, E., Dredze, M., Coppersmith, G., and Kumar, M. 2016. "Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media," in: *34th Annual CHI Conference on Human Factors in Computing Systems*. pp. 2098-2110.

Descalzi, G., Mitsi, V., Purushothaman, I., Gaspari, S., Avrampou, K., Loh, Y., Shen, L., and Zachariou, V. 2017. "Neuropathic Pain Promotes Adaptive Changes in Gene Expression in Brain Networks Involved in Stress and Depression," *Science Signaling* (10:471), p. 1549.

Desmet, B., and Hoste, V. 2018. "Online Suicide Prevention through Optimised Text Classification," *Information Sciences* (439), pp. 61-78.

Fu, K. W., Liu, K. Y., and Yip, P. S. F. 2007. "Predictive Validity of the Chinese Version of the Adult Suicidal Ideation Questionnaire: Psychometric Properties and Its Short Version," *Psychological Assessment* (19:4), pp. 422-429.

Goodfellow, I. J., Bulatov, Y., Ibarz, J., Arnoud, S., and Shet, V. 2013. "Multi-Digit Number Recognition from Street View Imagery Using Deep Convolutional Neural Networks," *arXiv preprint arXiv:1312.6082*.

He, L., Liu, H., Yang, Y. C., and Wang, B. 2021. "A Multi-Attention Collaborative Deep Learning Approach for Blood Pressure Prediction," *ACM Transactions on Management Information Systems* (13:2), pp. 1-20.

Henry, J. D., and Crawford, J. R. 2011. "The Short-Form Version of the Depression Anxiety Stress Scales (Dass-21): Construct Validity and Normative Data in a Large Non-Clinical Sample," *British Journal of Clinical Psychology* (44:2), pp. 227-239.

Hochreiter, S., and Schmidhuber, J. 1997. "Long Short-Term Memory," *Neural Computation* (9:8), pp. 1735-1780.

Huang, Y., Wang, Y., and Wang, H. 2019. "Prevalence of Mental Disorders in China: A Cross-Sectional Epidemiological Study," *The Lancet Psychiatry* (6:3), pp. 211-224.

Jie, X. A., Li, Y. A., Dz, B., and Yl, A. 2021. "Dncp: An Attention-Based Deep Learning Approach Enhanced with Attractiveness and Timeliness of News for Online News Click Prediction," *Information & Management* (58:2), pp. 1-11.

Lin, G. M., Nagamine, M., Yang, S. N., Tai, Y. M., Lin, C., and Sato, H. 2020. "Machine Learning Based Suicide Ideation Prediction for Military Personnel," *IEEE Journal of Biomedical and Health Informatics* (24:7), pp. 1907-1916.

Liu, R., Mai, F., Shan, Z., and Wu, Y. 2020. "Predicting Shareholder Litigation on Insider Trading from Financial Text: An Interpretable Deep Learning Approach," *Information & Management* (57:8), pp. 1-17.

Nobles, A. L., Glenn, J. J., Kowsari, K., Teachman, B. A., and Barnes, L. E. 2018. "Identification of Imminent Suicide Risk among Young Adults Using Text Messages," *The ACM CHI Conference on Human Factors in Computing Systems*, pp. 1-11.

Pham, T., Tran, T., Phung, D., and Venkatesh, S. 2017. "Predicting Healthcare Trajectories from Medical Records: A Deep Learning Approach," *Journal of Biomedical Informatics* (69), pp. 218-229.

Posey, C., Lowry, P. B., Roberts, T., and Ellis, S. 2010. "Proposing the Online Community Self-Disclosure Model: The Case of Working Professionals in France and the Uk Who Use Online Communities," *European Journal of Information Systems* (19:2), pp. 181-195.

Rink, B., and Harabagiu, S. 2010. "Utd: Classifying Ssemantic Relations by Combining Lexical and Semantic Resources," in: *Proceedings of the 5th international workshop on semantic evaluation*. pp. 256-259.

Shah, F. M., Jahan, S. A., Haque, F., Mahmud, Z., and Nur. 2020. "A Hybridized Feature Extraction Approach to Suicidal Ideation Detection from Social Media Post," *IEEE-Region-10 Symposium (TENSYMP)*, pp. 985-988.

Simon, G. E., Johnson, E., Lawrence, J. M., Rossom, R. C., Ahmedani, B., Lynch, F. L., Beck, A., Waitzfelder, B., Ziebell, R., and Penfold, R. B. 2018. "Predicting Suicide Attempts and Suicide Deaths Following Outpatient Visits Using Electronic Health Records," *American Journal of Psychiatry* (175:10), pp. 951-960.

Sutskever, I., Vinyals, O., and Le, Q. V. 2014. "Sequence to Sequence Learning with Neural Networks," *Advances in neural information processing systems* (27).

Tadesse, M. M., Lin, H., Xu, B., and Yang, L. 2019. "Detection of Suicide Ideation in Social Media Forums Using Deep Learning," *Algorithms* (13:7), pp. 1-19.

WHO. 2018. "National Suicide Prevention Strategies: Progress, Examples and Indicators." World Health Organization: Geneva, Switzerland.

Wu, J. L., and Pan, J. 2019. "The Scarcity of Child Psychiatrists in China," *Lancet Psychiatry* (6:4), pp. 286-287.

Xie, J., Zhang, B., Ma, J., Zeng, D. D., and Ciganic, J. L. 2021. "Readmission Prediction for Patients with Heterogeneous Medical History: A Trajectory-Based Deep Learning Approach," *ACM Transaction on Managemnt Information Systems* (13:2), pp. 1-27.

Zhang, D., and Wang, D. 2015. "Relation Classification Via Recurrent Neural Network," *arXiv e-prints arXiv:1508.01006*.

Zhang, Z., Wei, X., Zheng, X., and Zeng, D. D. 2021. "Predicting Product Adoption Intentions: An Integrated Behavioral Model-Inspired Multiview Learning Approach," *Information & Management* (58:7), pp. 1-12.

Zhou, J., Zhang, Q., Zhou, S., Li, X., and Zhang, X. 2022. "Unintended Emotional Effects of Online Health Communities: A Text Mining-Supported Empirical Study," *Management Information Systems Quarterly* (47:1), pp. 195-226.