Suicidal Ideation Cause Extraction From Social Texts

DEXI LIU$^{1,2}$, QI FU$^{1,3}$, CHANGXUAN WAN$^{1,2}$, XIPING LIU$^{1,2}$, TENGJIAO JIANG$^{1,2}$, GUOQIONG LIAO$^{1,2}$, XIANGQING QIU$^{1,2}$, AND RENQIANG LIU$^{1,2}$

$^1$School of Information Management, Jiangxi University of Finance and Economics, Nanchang 330013, China
$^2$Jiangxi Province Key Laboratory of Data and Knowledge Engineering, Jiangxi University of Finance and Economics, Nanchang 330013, China
$^3$School of Communication and Electronics, Jiangxi Science and Technology Normal University, Nanchang 330013, China

Corresponding author: Qi Fu (jxsyfq@yeah.net)

This work was supported in part by the National Natural Science Foundation of China under Grant 61762042, Grant 61972184, and Grant 61662027; in part by the Key Project in the Science and Technology Program of the Jiangxi Provincial Department of Education under Grant GJJ180252 and Grant GJJ180198; and in part by the Knowledge Innovation Fund for the Graduate Students of Jiangxi Province under Grant YC2017-B069.

ABSTRACT Suicide has become a major public health and social concern in the world. Suicidal ideation cause extraction (SICE) in social texts can provide support for suicide prevention. This article summarizes the research on suicidal ideation causes (SICs) through the use of psychological and sociological analysis. Then, a social text-based SIC dataset is constructed and analyzed statistically for various features. A CRF model is provided along with Char-BiLSTM-CRF, which uses concatenation of word embeddings and character embeddings as word representation inputs. Then, the effects on the task are explored by the word (W), part of speech (POS), dependence relationship (DP), suicidal psychology (PCS), emotion (ET), and language (LG) features in the CRF model. The experiment shows that the word features worked best. POS and DP can be somehow covered by word features. PCS, ET and LG features can improve the effect of SICE. It also shows that Char-BiLSTM-CRF is better than CRF in general, but CRF still has advantages in terms of precision. Adding character embeddings and CRF layers can significantly improve the extraction using Char-BiLSTM-CRF. The experiment also compared three-word embeddings with Word2vec, ELMo and BERT. Compared with Word2vec, the F-value of ELMo is increased by approximately 5% on average, and compared with ELMo, the C_F and E_F of BERT are increased by 3.5% and 2.3%, respectively. Finally, the challenge of SICE is discussed based on the experimental results of Char-BiLSTM-CRF.

INDEX TERMS Suicidal ideation, suicidal ideation cause extraction (SICE), social text, CRF, char-BiLSTM-CRF.

I. INTRODUCTION Suicide is considered one of the most serious mental health problems globally. According to a World Health Organization report, nearly 800,000 people commit suicide each year, not to mention those who attempt suicide or have suicidal ideation. In 2016, statistics showed that suicide was the second leading cause of death in the 15-29 age group worldwide [1]. Suicide causes a heavy burden on individuals, families, communities, and even countries. It has become a major public health and social concern.

The US National Institute of Mental Health believes that suicide can be divided into three levels: suicidal ideation, suicide attempts and completed suicide. Bonner and Rich [2] pointed out that suicidal ideation means to end one’s life mentally or cognitively. Beck et al. [3] defined suicidal ideation as a desire or plan to commit suicide but no real attempt, which is an important indicator for suicide risk assessment. Suicide prevention is a worldwide research hotspot, especially for government work, mental health work and social work. The key to effective suicide prevention is to assess suicidality, such as suicidal ideation or suicide attempts over time, which can provide scientific grounds and more guidance for suicide prevention work by further analyzing the suicidal ideation cause (SIC). However, as the initial stage of suicidal behavior, suicidal ideation is typically concealed, universal, contingent, and differs among individuals, which greatly challenges monitoring and cause analysis.

The associate editor coordinating the review of this manuscript and approving it for publication was Arianna Dulizia.
In detecting suicidal ideation, traditional methods mainly rely on professional knowledge among psychologists and self-reported questionnaire surveys. A large number of suicidal ideation measures and evaluation tools and scales have been developed. These methods are effective and easy to operate, but are prone to false negatives caused by participants’ deliberate concealment, coupled with their intrusiveness, high cost, poor timeliness, and difficulties with continuous tracking. They are difficult to carry out on a large scale and over a long-time. With regard to the analysis of the SIC and the suicide cause, existing work has mainly focused on theoretical and empirical research. However, problems such as geographical bias, limitations of representation, short time span, deliberate concealment of subjects, and the reliability and validity of the scale limit empirical research.

The rise of social media has provided new opportunities for psychological research and new channels for monitoring suicidal ideation and cause analysis. Individuals in psychological crisis often show distrust and resistance to traditional mental health services and institutions. Research shows that individuals in psychological crisis tend to seek help from informal resources, such as social media, rather than from psychological experts [4], [5]. A large amount of research has confirmed that it is feasible and effective to use social media [6], [7], such as Twitter, Facebook, and Weibo, to predict suicidal ideation, and a small number of researchers have also started to use social media data to analyze the causes of mental health problems [8] and have even carried out suicide interventions using social media platforms [9].

Using the behavior and content posted by users on social media platforms to detect users’ suicidal ideation helps make timely interventions and reduce the risk of suicide. The intervention measures must be user-centered, which requires psychological experts to read the posts on social media (referred to as “social text” in this paper), and obtain the SIC, as shown in Fig. 1.

However, due to the massive amount of social media content, fast updates, and the large number of users, it is difficult for psychologists to obtain the SIC through manual reading.

Suicidal ideation cause extraction (SICE) from social text can help psychological experts to get better and faster understanding of the intervention objects and reduce their workload. At the same time, it also provides support for statistical analysis of SIC using large-scale data and allows comparing SIC in different countries, regions, populations and time periods. However, according to our systematic review process, there is no relevant research in either the field of psychology or sociology, and the methods are mostly case analysis or empirical analysis. There is a correlation between emotion and suicidal ideation, and the extraction of emotional cause is a concern among researchers.

A. THE CAUSE OF SUICIDAL IDEATION

Suicidal ideation is associated with many factors, such as population economics and psychosocial behavior. It involves
the individual’s internal and external environment, including the direct or indirect effects of biological, genetic, psychological, social, environmental, cultural, external pressure, and others. The specific factors include academic studies, interpersonal relationships, environment, emergencies, race and religion, mental state, marital status, financial status, health issues and others. The selection principle varies according to the researcher’s perspective (biological, social, psychological, environmental) and the characteristics of the individuals being studied.

1) AGE
Kok et al. [10] explored the factors of suicide from the perspective of adolescents (ages 15-24) and summarized them as heterosexual relations, race and religion, academic pressure, competitive pressure, global economy, family education, original family (single parent family), and others. Chatterjee and Basu [11] explored the internal and external factors of suicidal ideation among female Indian college students and summarized four major factors, such as academic factors, interpersonal factors, environmental factors and emergency factors. Luo et al. [12] analyzed suicidal ideation among China’s urban and rural elderly population based on data from the “Tracking Survey on the Status of China’s Urban and Rural Elderly Population” from the 2010 China Aging Science Research Center (CASRC) and concluded that sociodemographic variables, physical health, mental health, and family were important factors leading to suicidal ideation in the elderly.

2) REGION
Maskill and Hodges [13] summarized the causes of suicide in New Zealand as marital status, family composition, education level, religion, socioeconomic status, occupational status, mental disorders, substance use disorders, physical diseases, and others. In the analysis of suicide in Australia, Hassan [14] explained suicide as an act involving a complex relationship between psychiatric and sociological factors and summarized eight categories of SIC, including psychosocial and social environmental factors, marital status (single, divorced), economic status, occupation (low income, low autonomy, low training opportunities), immigration and race, relationship issues (unhappy love, family and marriage issues, guilt), self-worth issues (economic and unemployment issues, feel life is a failure) and health problems.

3) COMPREHENSIVE PERSPECTIVE
Chesney et al. [15] systematically reviewed 407 suicide articles at a macro level and attributed suicide cause to depression, mental health causes such as schizophrenia, eating disorders, learning disabilities, loneliness, and childhood behavioral disorders (behavioral disorders and oppositional defiant disorder, attention deficit, hyperactivity), dementia, personality disorders, substance use disorders, and other factors. In addition, some causes of suicide are summarized from the conclusions of various media, psychological counseling and treatment platforms or institutions. Mental Health Daily [16] summarizes 15 causes of suicide: mental illness, painful experience, bullying, personality disorders, substance abuse, eating disorders, unemployment, social isolation/loneliness, interpersonal problems, genetics/family history or family with a history of suicide, existence of crisis/value deficiency, terminal illness, chronic pain, financial problems, and side effects of drugs including various prescription drugs such as antidepressants. Suicide.org (Suicide Survivors Forum) [17] lists 29 causes of suicide in detail, including: the death of a loved one, divorce, separation or break up of a relationship, a serious loss, such as loss of a job, house or money, disease or serious illness, a terminal illness, intense emotional pain, and others.

In short, as an extreme and complex behavior, suicide is caused by many factors [18], [19]. The traditional research on SIC can be summarized into two areas: theoretical research and empirical research.

Theoretical research mainly explores the intricate relationships among various biological, psychological, and sociological factors, and the main models used include the Diathesis-stress Model [20], the Aggregation Optimization Model [21], the Stress-Vulnerability Model of Suicidal Ideation and Behavior [22], and the Biopsychosocial Model of Suicidal Ideation [23]. Many theories have been used to study SIC, such as psychoanalytic theory (Sigmund Freud, 1895), biological theory [24], sociological theory [25], escape theory [26], evolutionary theory [27], attachment theory [28], theory of personality development stages [29], interpersonal theory of suicide [30], strain theory of suicide [31], and others. Most of the theoretical studies are done from the perspective of their respective disciplines, with narrow focus and lack of empirical analysis.

Based on the scale of suicidal ideation, empirical research is carried out mainly by questionnaires and statistical methods to analyze logistic regression relations among single and multiple factors, and the relationship between suicidal attitude and factors that influence it is discussed from the sample data [32], [33]. In recent years, some scholars have also used structural equation modeling to overcome the complex interaction between latent variables and multiple variables that cannot be analyzed by other statistical methods [34], [35]. In the empirical study, there are still some problems, such as regional bias in the sample, limited representativeness, short time span, deliberate concealment by subjects, and the reliability and validity of the scale.

Traditional studies on SIC, both in theoretical and empirical research, obtain the cause from questionnaires, interviews, or researchers’ analysis of patients. Although some researchers also used social media data to study SIC [8], [36], the extraction of SIC is still based on manual work. At present, according to our systematic review process, there is no research that uses computer technology to extract SIC automatically from notes (diaries, testaments, microblog posts, etc.) posted by patients, and it is difficult to expand
the selection of patient samples and maintain continuous tracking.

B. DETECTION OF SUICIDAL IDEATION BASED ON SOCIAL MEDIA

In recent years, a great amount of progress has been made in the detection of suicidal ideation from social media data by utilizing computer technology, which are mainly based on related knowledge bases (Emotional Dictionary, Linguistic Inquiry and Word Count (LIWC), WordNet, HowNet, etc.) and various features (user attributes, user behavior, published content, user emotion or psychology, etc.) by training a machine learning model with labeled samples.

The features used in related works can be classified into four categories.

1) ATTRIBUTE FEATURES

The attribute features, which are usually filled in when registering for social media, describe the basic personal information about social media users, such as gender, age, region, nickname, occupation, picture, etc. [37]–[39]. They also include some basic user settings, such as whether users enable private messages, whether users allow others to leave messages, whether users enable the geographical indicator on their account, whether users include “me/myself” in their self-description, etc. Guan et al. [40] used attribute features, such as the user’s gender, username length, number of favorites, number of fans, and self-description text length, to detect the user’s suicidal ideation, and found that there was a correlation between these attribute features and suicide risk.

2) BEHAVIORAL FEATURES

Behavioral features describe the behavior of users, including posting, forwarding, commenting, liking, following and other behaviors. Previous studies have shown that users with different degrees of suicidal tendencies, users who have committed suicide, and those without suicidal ideation behave differently when using Weibo [41], [42]. The common behavioral features used for suicidal ideation detection include the average number of days users stay active, the average number of comments on a single post, the average number of times a user’s single post is forwarded, the average number of likes on a single post, the total number of interaction posts, the frequency of activities at night, the social activity (the number of friends/fans), and other factors. [40]. For example, Huang et al. [6] conducted statistical analysis on the behavioral features of users with and without suicidal ideation and found that the post type (original, forwarding), post time and social relationships mentioned (@ others) are effective features for suicide risk detection.

3) CONTENT FEATURES

Content features mainly refer to user-generated content. A large number of studies have found that users with different mental health and suicide behaviors have different language expressions and diction styles. For example, Sueki [43] studied suicide-related posts and suicide behaviors of Twitter users and concluded that some specific phrases were closely related to users’ suicidal ideation, such as “want to commit suicide” and others.

In the detection of suicidal ideation, content features mainly include the frequency and weight of the content in the texts, such as keywords, phrases [43], N-Gram [7], [44], words [45], and TF-IDF [44], part of speech (POS) [6], [44], and syntactic dependency [6], [44]. Some artificially conducted knowledge bases are also used to analyze suicidal ideation, such as LIWC [46]–[49] and its simplified Chinese version (SCLIWC) or a specific version of the Simplified Chinese Micro-Blog Word Count Dictionary (SCMBWC) [40] for a specific data type. The topics [6], [49], [50] and word embeddings [6] are also used as content features for improving semantic accuracy.

4) EMOTIONAL AND PSYCHOLOGICAL FEATURES

Emotional and psychological features describe the emotions, feelings or psychology in the posts, not only involving positive and negative emotions, such as joy, anger, sadness and happiness but also including friendliness, aggression, guilt, compunction, insecurity, fear, self-pity, self-mockery and other fine-grained emotions and personality traits. Studies have found that emotional and psychological activities and personality traits have a strong correlation with suicidal ideation or suicide behavior. For example, a mild manic state is related to multiple suicide attempts [51], and people with high suicide risk have a higher sense of guilt or shame [52], [53] and have a higher proportion of posts with anger, sadness, fear, disgust and loneliness [7].

Some researchers also use emotional features and other psychological features with personality features in the detection of suicidal ideation. Ren and Kang [54] used blog data to analyze the relationship between emotional features (such as anxiety and sadness) and suicide crises and used an emotional topic model to detect potential emotions in blogs to improve the detection of suicide risk. Huang et al. [6] and Lv et al. [46] used and expanded an emotional dictionary to detect suicide risk among Weibo users. There are several similar studies to detect suicidal ideation or suicide risk in text using emotional dictionaries or by analyzing text emotion. Burnap et al. [44] used SentiWordNet to study twitter posts, and Aladag et al. [48] established an emotional matrix to study Reddit posts.

5) SUICIDAL IDEATION DATASET

The generalized suicidal ideation dataset includes questionnaires, suicide notes, suicide blogs, electronic health records, online social texts [55], and other information, but the suicidal ideation dataset in this article refers to a collection of social texts expressing suicidal ideation or suicidal tendencies. It mainly comes from Reddit, Twitter, ReachOut, Sina Weibo and other social network platforms or forums.
C. EXTRACTION OF EMOTIONAL CAUSE

There is a strong correlation between emotion and suicidal ideation and their cause. In recent years, some researchers shifted their attention from emotional text analysis to the extraction of emotional cause. The widely used methods are reviewed below.

1) KNOWLEDGE-BASED AND RULE-BASED METHODS

Following Talmy’s [60] viewpoint, emotion is a kind of internal response triggered by understanding external events. Lee et al. [61] and [62], Chen et al. [63], Neviarouskaya and Aono [64], Li and Xu [65] extracted the events that trigger an emotion through linguistic clues. At the same time, some scholars have proposed the commonsense-based method (CB) [65], [66], which constructs rules by combining a linguistic model, psychological knowledge and other domain knowledge.

2) FEATURE-BASED MACHINE LEARNING METHODS

These methods consider emotion cause extraction to be a candidate sentence or clause classification problem [63], [65], [67], [68] or a sequence labeling problem [69], [70]. The classification problem is to classify whether the candidate sentence or clause is the cause through its own features and contextual features. Compared with the classification model, the sequence labeling model can not only conveniently add insentence features and contextual features but also consider the sequence features between the cause clauses. Various features are used for emotional cause extraction including word features, word category features (such as the categories of words in the LIWC or Emotional Dictionary), POS features, grammatical features, word embeddings features, distance from emotional words, context features and linguistic rule features, among other.

3) DEEP LEARNING METHODS

Knowledge-based and rule-based methods have several disadvantages such as the heavy workload of constructing the knowledge and rules, the limited coverage, and the poor adaptability of different corpuses, whereas feature-based machine learning methods require rich experience for feature selection. Gui et al. [71] and Mu [72] et al. used deep learning technology to obtain the semantic representation through a word embedding model called Word2vec and then measured the effect of words for detecting emotional cause through an attention mechanism. It captures semantic continuity by integrating contextual semantic information. Xia et al. introduced a relative position augmented embedding learning algorithm, and transformed the task from an independent prediction problem to a reordered prediction problem where the dynamic global label information is incorporated [73]. To consider the relationships among sentences, they also proposed an emotion cause extraction framework (RTHN) (RNN-Transformer converter hierarchical network) [74]. To avoid manually annotating emotional
clauses, they also proposed a joint extraction model of both emotional cause clauses and emotional clauses based on hierarchical BiLSTM and interactive multitask learning [75] and a joint extraction model based on a 2D square matrix.

Other research on works of emotional cause extraction based on deep learning also include Tang et al. [76], Fan et al., Song et al. [77], Bi and Liu [78], and Yu et al. [79].

Although SICE is similar to emotional cause extraction, there are still many differences. First, in emotional cause extraction, the exaction object is usually events with sentence-level or clause-level granularity, and the location of the emotional cause is usually closely related to the position of emotional words. In SICE, the extracted objects include events, emotions, psychophysiological disorders, and other, and the extraction granularity is smaller, such as word-level and phrase-level. Second, although there is a Chinese Suicide Dictionary [46] built by the Institute of Psychology of the Chinese Academy of Sciences, the suicide words are more complex, and their relationship with SICE is not clear. Third, emotional words mainly describe the emotional tendency and are independent of the emotional cause. However, it is more complicated with suicide words because they cover a wider range of words, such as suicidal tendency or suicidal ideation words (such as death, to die, not want to live, give up, want to commit suicide), psychological words (such as annoyance, resentment, regret, sadness, disappointment, loneliness), physiological words (such as illness, neurasthenia, schizophrenia, cancer, malignant lymph nodes cancer), interpersonal relationship words (such as mom, love you, friends, family, parents), and others. What’s even harder is that some of these words can be cause words or clue words themselves.

III. CONSTRUCTION OF THE SIC DATASET
To our knowledge, there are no publicly available corpus resources for SIC in social texts, and we constructed an SIC dataset based on our proposed annotation scheme under the guidance of psychological professionals.

We constructed an SIC dataset based on the Weibo suicidal ideation dataset [9] introduced in Section II-B-5, which contains only Chinese posts. Since this dataset is a collection of posts with suicidal ideation that were manually annotated, we no longer know whether the posting users have suicidal ideation or not.

A. DATA PREPROCESSING
The preprocessing of the dataset mainly includes data cleaning and data transformation. Data cleaning includes removing post ID, @username, illegal characters, hyperlinks, emoticons, and non-Chinese characters. The data transformation mainly refers to the transformation of Chinese-Simplified and Chinese-Traditional text. The data preprocessing also includes Chinese word segmentation, part-of-speech tagging, and dependency parsing, and LTP was used, which is a language cloud platform from Harbin Institute of Technology. Although there are similar tools such as jieba word segmentation, THULAC, and others, LTP has more complete functions than other tools, including Chinese word segmentation, part-of-speech tagging, named entity recognition, dependency syntax analysis, semantic role tagging, and others. It is a popular and widely used platform for Chinese text processing.

B. ANALYSIS OF SUICIDAL IDEATION CAUSE
Manual annotation of the SIC dataset is based on the main classifications of “Chinese Classification and Diagnostic Standards for Mental Disorders” (CCMD-3) [80] and other previous research. Under the guidance of Professor Renheng Wu, executive director of the Jiangxi Psychological Society, and three other psychological professionals, the SIC was classified into two groups: individual internal cause and external environmental cause. The internal causes included physical factors (including physical diseases, eating disorders, nonorganic sleep disorders, nonorganic sexual dysfunction, genetic factors, etc.) and psychological factors (three types according to degree: general psychological problems that require psychological consultation, severe psychological problems that require psychiatrist treatment, and psychiatric patients who need psychiatrist intervention). External cause included chronic stress (including economic stress, work stress, academic progression pressure, and interpersonal stress), stress events and emergencies (including severe or acute stress disorder, acute stress psychosis, posttraumatic stress disorder, adaptation disorder, sharp and severe mental strikes, and traumatic psychological events), psychoactive substances (including alcohol, opioid, cannabis, hypnotic, anxiolytic, anesthetic, stimulant, hallucinogen, tobacco, etc.), and culture (including religion, race, immigration, etc.). Because the boundaries between some classes are not obvious and correlated, the annotation is subjective. For example, mental health problems such as depression have both psychological factors and physiological factors of the brain’s organic changes.

C. MANUAL ANNOTATION SCHEME
Because of the informality and colloquialism of language in social text, the SIC is usually expressed implicitly, that is, there are no linguistic clue words such as “because”, “as”, “so”, which increases the difficulty of annotation work.

The annotation scheme we proposed adopts the principle of “Terse and concise, No reasoning, Continuity of cause words, and Unique category”. “Terse and concise” refers to annotating only the smallest string that can reflect the SIC. “No reasoning” requires that the annotation is based on literal meaning only, without guessing or reasoning, and without considering indirect causes. “Continuity of cause words” requires that the words covered by the same cause must be continuous and uninterrupted. “Unique category” means that each cause belongs to only one cause category. The tagging symbol refers to the classical tagging method of 4-tag (B: begin, the first part of the cause; E: end, the last part of the cause; M: middle, the middle part of the cause; S: the cause only contains a single word).
The annotation work was completed by four graduate students who had received psychological training and passed the test. The annotation results were checked, and inconsistent annotations were discussed and voted on by the group. Annotations that remained controversial were sent to a psychological expert for arbitration. Consistency in detection of the cause annotation adopted a standard Kappa, and the average Kappa value between two annotators was 0.617. This finding showed that there was a difference in the artificial judgment of the SIC, and informal expressions in Chinese posts increased the complexity and ambiguity of semantic understanding.

In the end, 5,994 posts with causes were annotated in 13,013 posts. Examples from the SIC dataset are below:

Ex.1: He/ECSI_B doesn’t/ECSI_M need/ECSI_M me/ECSI_M anymore/ECSI_E; I can probably die. (cause type: External cause E, chronic stress ECS, interpersonal stress ECSi)

Ex.2: I am inferior/IPG_S and useless/IPG_S. I don’t want to live anymore. (cause type: Internal cause I, psychological factors IPS, general psychological problems IPSg)

D. STATISTICAL ANALYSIS OF THE SIC DATASET

1) DISTRIBUTION OF CAUSES

Although a post is limited to 140 words, there may be multiple suicidal ideation causes. The number of SICs contained in each post is shown in Table 1. As seen from Table 1, 53.9% of the posts with suicidal ideation do not explicitly express the cause of suicidal ideation. In the posts with SICs, there are 44.8% that contained a single cause, whereas 55.2% contained multiple causes. More SICs in a post meant more difficulty for SICE.

2) POS DISTRIBUTION IN SIC

Statistics for part of speech (POS) distribution in the SIC can help us understand the distribution when expressing SICs. There were 13,792 SIC phrases with 26,756 words, whereas there were 223,272 words in the non-SIC phrases. The POS statistical results used to express SIC are shown in Fig. 2 (statistics in terms of words). In the expression of an SIC, the verb (v) accounts for the largest proportion, accounting for 33.5%, whereas the other POS ranked in the order of adjective (a), noun (n), and adverb (d).

Comparing the POS in an SIC with that in a non-SIC, it was found that the adjectives (a) showed the greatest difference (except for abbreviations (j) and conjunctions (c)). The former was very rare, whereas the latter was a function word, accounting for 20.8% (PinS) and 4.8% (PinN) respectively, and the value of PinS/PinN was 4.3. Generally, the distribution of POS in the SIC and non-SIC was quite different.

3) THE DISTRIBUTION OF DEPENDENCY IN SIC

The analysis of dependency in SIC can help us understand the expression of the SIC and the structural information in the phrases or clauses used. The statistical results are shown in Fig. 3 (statistics in terms of words). Fig. 3 shows that in SIC, the verb-object phrase (VOB) and coordinate phrase (COO) account for 24% and 22.2%, respectively. By comparing dependency in the SIC with that in non-SIC, we found that the biggest difference is the core relationship (HED) (except the indirect object (IOB) and right adjunct (RAD)). The former was very rare, whereas the latter was a function word) that accounted for 4.8% (DPinS) and 2.1% (DPinN) respectively, and the value of DPinS/DPinN was 2.3, followed by verb object phrases (VOB), accounting for 24% (DPinS) and 11.2% (DPinN), and the value of DPinS/DPinN was nearly 2.2.

4) THE DISTRIBUTION OF EMOTIONAL WORDS IN SIC

There is a strong correlation between emotion and suicidal ideation. Negative emotion is one of the important causes for suicidal ideation. The statistics of the distribution of...
emotional words in SIC is helpful to understanding the correlation between them.

In our study, the National Taiwan University Sentient Dictionary (NTUSD) and Dalian University Technology Sentiment Dictionary (DUTSD) are employed as sentiment dictionaries. NTUSD divides 11,086 words into 2,810 positive emotion words and 8,276 negative emotion words. The DUTSD included 11,107 positive emotion words and 10,713 negative emotion words after removing the emotional words with polarity ambiguity. The distribution of emotional words in SIC is shown in Table 2.

Table 2 shows that the distributions of emotion words in the DUTSD and NTUSD are basically the same in the SIC dataset. In general, the proportion of emotion words in the SIC was higher than that in the non-SIC. In terms of phrases in the SIC, the number of negative emotion words was significantly larger than positive emotion words. For example, by using DUTSD, the proportion of positive emotion words was 2.4%, whereas negative emotion words was 3.3 times higher than positive emotion words(nearly 7.9%). Using the NTUSD, the proportion of positive emotion words was 4.5%, whereas negative emotion words was 5.3 times of positive emotion words (as high as 23.9%).

There are some differences between the distribution of emotional words in SIC and non-SIC, especially negative emotion words. Specifically, negative emotion words in SICs were 7.2 times (based on DUTSD) and 5.6 times (based on NTUSD), respectively, more than those in non-SIC.

5) THE DISTRIBUTION OF LINGUISTIC FEATURES IN SIC
The analysis of the distribution of language features in SICs is helpful for understanding language style and the diction used in the expression of an SIC. Based on the Simplified Chinese LIWC (SC-LIWC), which is a translation and revision of LIWC by Taiwanese scholar Jinlan Huang et al. [81]

under the authorization of Pennebaker, the founder of LIWC, we statistically analyzed the distribution of language features. SC-LIWC includes 71 word categories, such as emoticons, first-person singular and plural.

The distribution of SC-LIWC categories in SICs is shown in Fig. 4 (statistics in terms of words). In addition to the function words, “funct”, the higher proportion in SICs were emotion processing words “effect” (9.6%), negative emotion words, “negemo” (7.5%), cognitive processing words, “cogmech” (6.1%), physiological process words, “bio” (4.8%), verbs (4.8%), “sad” words (4.7%) and “health” words (3.8%).

Comparing the SC-LIWC words in SIC and non-SIC, we found that “sad” words were the most different, accounting for 4.7% (LIWCinS) and 0.23% (LIWCinN). The value of LIWCinS/LIWCinN reached a peak of 20.5, followed by “anger” words, “health” words, “negemo” words, and “anx” words, as shown in Table 3.

In general, the distributions of SC-LIWC in SICs and non-SICs were quite different. According to the definition of SICs, the SC-LIWC words with higher difference rankings, such as “sad”, “anger”, “negemo” and “anx”, belong to psychological factor IPS, which is an internal cause of suicidal ideation.

6) THE DISTRIBUTION OF SUICIDE WORDS IN SIC
For emotional cause extraction, the causes are usually located by emotional words, and emotional cause clauses usually appear near the emotional words. A simple idea is that there could also be a similar relationship between suicide words and SICs. Suicide words indicate the tendency for suicide, and the SIC is the cause of this tendency. Studying the distribution of suicide words around the SIC can help to further understand this relationship.

Based on the Chinese Suicide Dictionary developed by the Institute of Psychology, Chinese Academy of Sciences, the distribution of suicide words around SICs was calculated, as shown in Table 4. The distribution showed that suicide words appeared in 67.9% of phrases in SICs, and 64.3% of them had one suicide word, whereas approximately 87.9% of the posts contained suicide words outside the SICs, and even as many as 68.8% of the posts contained at least two suicide words outside the SICs. This finding is quite different from

### Table 2. The distribution of emotional words in SICs.

|                | DUTSD (positive) | DUTSD (negative) | NTUSD (positive) | NTUSD (negative) |
|----------------|------------------|------------------|------------------|------------------|
| The number of  | 644              | 2,110            | 1,217            | 6,385            |
| emotional words|
| in SIC (proportion) | (2.4%)          | (7.9%)          | (4.5%)          | (23.9%)          |
| The number of  | 3,438            | 2,361            | 6,470            | 9,527            |
| emotional words|
| in non-SIC (proportion) | (1.5%)         | (1.1%)         | (2.9%)         | (4.3%)         |
The above statistics showed that the distinction between suicide words and SICs was not clear. Further analysis showed that suicide words in the Suicide Dictionary included not only the words that described suicidal ideation but also the words (cause) that led to suicidal ideation, such as “depression”, “illness”, and “cancer” and even some interpersonal words. This finding showed that it was difficult to extract SIC by suicide words, and it also reflected that the SICE was more challenging than emotional cause extraction.
7) THE DISTRIBUTION OF CAUSE CLUE WORDS

Clue words such as “because”, “as”, and “make” are good indicators for detecting the SIC. It has been shown in emotional cause extraction that linguistic clue words play an important role in the connection between emotional keywords and emotional causes. Researchers use linguistic clue words to construct linguistic rules to extract emotional causes in text [63], [69]. The cause clue words used in this paper are shown in Table 5. The cause clue words were divided into four categories according to grammar and function: conjunctions and prepositions, causative-verbs, perception-verbs, and other words. In the SIC dataset, the closest distance between cause clue words and cause phrases was calculated. The distance was based on the number of clue words. It was the closest distance between the cause clue word and the word that began or ended the cause phrase, that is, the distance from the word marked “B”, “E” or “S”, as shown in Table 6. In the SIC dataset, 1,187 posts (19.8%) did not have any cause clue words, and these posts contained 1,958 cause phrases, accounting for 14.2% of the total number of cause phrases.

Although 80.2% of the posts in the SIC dataset had cause clue words, it was found that many of them had different uses and did not always express the cause. For example, in the example of the cause clue word “make” in the SIC dataset given below, only (make or let) in the first example is a clue word that truly expresses the cause.

**Ex.3:** I feel guilty/IPSg_S of wanting to kill makes me want to kill myself. (cause type: Internal cause IPS, psychological factors IPSg, general psychological problems IPSg, “makes” is the cause clue word)

**Ex.4:** I don’t know what to do, let him go, and I’ll be lonely/IPSg_S to bear. He may not understand how terrible the night is coming, and many suicidal thoughts will come out of my mind. (cause type: Internal cause IPS, psychological factors IPS, general psychological problems IPSg, “let” is the not cause clue word)
To summarize up, our research is similar to emotion cause extraction in that both of them are extract causes from the text and use statistical analysis, and some of the results are similar in the literature [68]–[70], [72], [82]. For example, the features that have significant differences between cause phrases and noncause phrases are verbs (v), adjectives (a) and nouns (n) in POS, and core HED (verb) and verb object phrases (VOB) in dependency [82]. However, there are still some differences or challenges in SICE, which include the following three issues.

a: THE GRANULARITY OF EXTRACTED OBJECTS
The extracted objects of emotional causes are usually events, and the granularity is usually sentence-level or clause-level. The location of the emotional cause is usually close to the emotional words. However, the extracted objects from SICs include more fine-grained cause phrases or words, such as events, emotions, psychophysiological disorders, etc.

b: THE ROLE OF EMOTIONAL WORDS AND SUICIDE WORDS
First, emotion dictionaries have much bigger vocabularies than suicide dictionaries. Although different emotion dictionaries contain different emotional words, the definition of emotional words is relatively clear. Most emotional words are annotated with word-level granularity. However, suicide dictionaries are very rare at present, and there are far more words used for suicidal ideation than a word-level expression. The suicide dictionary built by the Institute of Psychology of the Chinese Academy of Sciences aims to detect suicidal ideation, which includes a large number of words, phrases and even sentences.

Second, emotional cause extraction usually involves extracting emotional causes around emotional words. Emotional words and emotional causes do not overlap, and the boundaries are clear. However, suicide words include words not only describing suicidal ideation but also leading to suicidal ideation, and even some interpersonal relationship words, which have unclear boundaries with suicidal ideation and more complex relationships.

c: THE EFFECT OF CAUSE CLUE WORDS
The dataset in emotional cause extraction generally uses more formal news text data, such as literature [71], and there are clear syntax rules between the clue words, emotional words and emotional causes. However, in addition to the different roles of suicide words and emotional words discussed above, the relationship between cause clue words and SICs also become ambiguous because of the informal expression of posts. In addition, nearly 20% posts that contain SICs have no cause clue words. There is a large amount of ambiguity in the use of cause clue words, which makes it even more difficult to analyze SICs through clue words or syntactic analysis.

IV. CRF MODEL AND FEATURE ENGINEERING
As the types of SICs are complex and interrelated, to simplify the problem, only the SIC is extracted in our work, without distinguishing its specific type.

The task of SICE can be regarded as a sequential labeling problem. At present, the classic sequential labeling models are mainly the Hidden Markov Model (HMM), the Maximum Entropy Markov Model (MEMM) and the Conditional Random Field (CRF), among others. CRF does not have the strict independence assumptions like HMM, and it can accommodate arbitrary context information and has flexible feature design. Compared with MEMM, CRF calculates the conditional probability of the global optimal output node so it also overcomes the shortcomings of label bias in the MEMM model. Therefore, CRF is widely used in sequential labeling problems and performs well. This paper attempts to use CRF to extract SICs.

Given the text, \( x = \{x_1, x_2, \ldots, x_t\}, x_i (i = 1, \ldots, t) \) is the word (including punctuation) that make up the text, \( t \) is the number of words in \( x \), and the label sequence \( y = \{y_1, y_2, \ldots, y_t\} \) is the label of the SIC on \( x \), and its calculation is shown in formula (1).

\[
P(y|x) = \frac{1}{Z(x)} \exp \left( \sum_{i,k} \lambda_k I_k (y_{i-1}, y_i, x, i) + \sum_{i,1} \mu_i s_i (y_i, x, i) \right)
\]

where,

\[
Z(x) = \sum_y \exp \left( \sum_{i,k} \lambda_k I_k (y_{i-1}, y_i, x, i) + \sum_{i,1} \mu_i s_i (y_i, x, i) \right)
\]

Feature function, \( s_i (y_i, x, i) \), only depends on the current position, which is called a state feature, that is, the cause labeling sequence is only related to the label of the current word in \( x \), and the features of \( x \) (such as POS, syntax, etc.) are considered. Feature function, \( I_k (y_{i-1}, y_i, x, i) \), depends on the labels of the current and previous position, which is called a transfer feature. The labeling sequence, \( y \), which makes \( P(y|x) \) the largest, is the final labeling result.

The CRF model takes the definition of the feature, the selection of the feature function and the design of the feature template into consideration. In this paper, the open source tool CRF++ (https://www.softpedia.com/get/Science-CAD/Taku-CRF.shtml) was used. There are four main parameters that need to be set in the experiments. The control variable method is used and after many experiments, the parameters with better experimental results are selected based on the verification dataset. The final main parameter settings and meanings are as follows.

1. -a(CRF-L2 or CRF-L1): This parameter is for the normalized training algorithm, and the CRF-L2 training algorithm is finally chosen.
2. -f(NUM): This parameter means that the feature with a frequency less than NUM is ignored in the training process, and the final choice is f=4.
3. -c(float): This parameter is the hyperparameter of the CRF. It adjusts the balance between overfitting and underfitting during training. The final choice is c=1.
4. \(-p\text{(NUM)}\): This parameter is for devices with multiple CPU resources to train with NUM threads. The final choice is \(p=1\).

A. CANDIDATE FEATURES

According to the results of the statistical analysis of the SIC dataset in section III-D, the CRF combines six kinds of three classes of atomic features: content features (Word (WD), part of speech (POS), Dependency (DP)), emotional and psychological features (Psychological Characteristics of Suicide (PCS), emotional features (ETs)), and linguistic features (LGs).

Word (WD). The WD of \(x_i\) refers to the word itself in the \(x_i\) position, which is recorded as \(W_i\).

Part of speech (POS). The POS of \(x_i\) refers to the POS of the word in the \(x_i\) position, which is recorded as \(P_i\).

Dependency (DP). The DP of \(x_i\) refers to the word that \(x_i\) depends on and the dependency, which is recorded as \(D_i\).

Psychological Characteristics of Suicide (PCS). The PCS of \(x_i\) refers to whether \(x_i\) belongs to "suicide words", which is recorded as \(S_i\). The Chinese Suicide Dictionary published by the Institute of Psychology of the Chinese Academy of Sciences is adopted here.

Emotional Features (ETs). The ET of \(x_i\) refers to the emotion polarity of \(x_i\), which is recorded as \(E_i = (E_{ai}, E_{pi})\), where \(E_{ai}\) and \(E_{pi}\) respectively indicate whether \(x_i\) is a negative or positive emotion word. Emotional polarity was determined by the NTUSD Emotion Dictionary.

Linguistic Features (LGs). The LG of \(x_i\) refers to the word category of \(x_i\) in the SC-LIWC, which is recorded as \(L_i\). There are 71 word categories in the SC-LIWC, which are recorded as \(L_i\). The only five LIWCinS/LIWCinN values were adopted, such as "sad", "anger", "health", "negemo" and "anx". The LG of \(x_i\) only considers whether \(x_i\) belongs to these five categories. Since \(x_i\) may belong to multiple categories at the same time, \(L_i\) is expressed as \((L_{a1}, L_{a2}, L_{b1}, L_{b2})\), which, respectively, indicates that \(x_i\) belongs to "sad", "angle", "health", "negemo" and "anx".

In addition to the abovementioned six atomic features, we also combine these atomic features in two ways: the same types of features are combined in a given window to form the context of the feature and the different types of atomic features are combined to form mixed features.

For example, given a window length of 5, the context of \(x_i\) is "\(x_i-2, x_i-1, x_i, x_i+1, x_i+2\)". The WD of \(x_i\) includes unitary WD \((W_{i-2}, W_{i-1}, W_i, W_{i+1}, W_{i+2})\), binary WD \((W_{i-2}W_{i-1}, W_{i-1}W_i, W_iW_{i+1}, W_{i+1}W_{i+2})\), ternary WD \((W_{i-2}W_{i-1}W_i, W_{i-1}W_iW_{i+1}, W_iW_{i+1}W_{i+2})\), and more relations. Similarly, the POS includes unitary POS \((P_{i-2}, P_{i-1}, P_i, P_{i+1}, P_{i+2})\), binary POS, ternary POS, and more, as do the other types of features.

Consider that POS, dependency, and word category (whether positive or negative emotion words, whether suicide words, what kind of type it belong in the LIWC) are dependent on the word itself, and compared with the POS and dependency, the ambiguity of a word category is relatively small. Therefore, the mixed features only consider the combination of word, POS, and dependency.

Based on the experiments, the candidate features we finally designed are shown in Table 7.

| Type of feature | Candidate features (The context window \(l\) is set according to the validation set) |
|-----------------|--------------------------------------------------------------------------------|
| WD \(W_{i-2} (2 \leq l \leq 2), W_{i-1}W_{i+1} (0 \leq l \leq 1), W_iW_{i+2}, W_{i+1}W_{i+2}, W_{i+2}W_{i+3}\) |
| POS \(P_{i-2} (1 \leq l \leq 1), P_{i-1}P_i, P_iP_{i+1}\) |
| DP \(D_{i-2} (2 \leq l \leq 2), D_{i-1}D_{i+1} (0 \leq l \leq 1), D_{i+1}D_{i+2}\) |
| PCS \(S_{i-1} (0 \leq l \leq 1), S_{i+1}\) |
| ET \(E_{i-2} (2 \leq l \leq 2), E_{i-1}E_{i+1} (0 \leq l \leq 1), E_{i+1}E_{i+2}\) |
| LG \(W_{i-2}W_{i-1}W_iW_{i+1}W_{i+2}\) |

Mix features \(W_{i-2}P_{i-1}P_iP_{i+1}P_{i+2}\), where \(W_{i-2}\) refers to the word dependent on \(W_i\) by dependency \(D_{i+1}\).

B. FEATURE SELECTION STRATEGY

The feature selection strategy based on the greedy method was adopted for the CRF. First, we selected the feature type that resulted in the best single type features. Then, we added the remaining five types of features and selected the feature type with the best results the combination of all six types of features was examined.

V. CHAR-BILSTM-CRF MODEL

The Char-BiLSTM-CRF model did not require feature design and achieved good results in sequential labeling tasks, such as part-of-speech tagging, chunking, and named entity recognition. Therefore, we attempted to use this model to extract the SICs. Char-BiLSTM-CRF [83]–[85] uses a three-layer architecture neural network model, as shown in Fig. 5.
A. CHAR LAYER
The Char layer extends the input of the word embeddings, and the char embeddings and the word embeddings are concatenated as the input layer of this model, which can represent the internal structure of the word, making up for the deviation in word embeddings caused by Chinese word segmentation or insufficient training data.

B. BiLSTM LAYER
The BiLSTM layer can capture the long-distance dependence of words in a sentence, and can learn more about the context of the words.

C. CRF LAYER
The CRF layer tries to optimize the output label sequence. The label of the current word is not only affected by the features of the word itself and the context but also by the label of the context. For example, the label “M” must be preceded by “B” or “M”, and the label of “E” must be preceded by “M” or “B”. These labels are mutually constrained. BiLSTM can’t fully consider the dependence between labels, which may result in irregular label sequences. The CRF layer can learn the label transition probability of the entire sentence and fully consider the dependencies between the context labels.

The hyperparameter settings used in the Char-BiLSTM-CRF model in this paper are shown in Table 8.

| Parameters        | Description                        | Value |
|-------------------|------------------------------------|-------|
| dim_word          | the dimension of word embeddings   | 300   |
| dim_char          | the dimension of char embeddings   | 100   |
| nepochs           | the count of epoch                 | 15    |
| dropout           | dropout                            | 0.5   |
| batch_size        | the size of the batch              | 20    |
| lr                | learning rate                      | 0.001 |
| lr_decay          | the decay of the learning rate     | 0.9   |
| hidden_size_char  | hidden size of lstm on char embeddings | 100  |
| hidden_size_lstm  | hidden size of lstm on word embeddings | 300  |

We also compared three-word embedding models: Word2vec, ELMo and BERT. Word2vec is a kind of distributed, low-dimensional, and dense word embedding representation that can fully consider the context of words and map words with similar semantics to a close position in vector space. In view of Word2vec’s insufficient understanding of the contextual meaning of words, ELMo achieves representation of different representations of the same word in different contexts through pretraining and fine tuning. BERT further enhances the generalization ability of the word embedding model, fully expresses the character level, word level, sentence level and even the relationship between sentences, and better solves the problem of polysemy that Word2vec cannot solve.

In this article, Word2vec is pretrained on Chinese Wikipedia, ELMo is pretrained on 13,013 posts with suicidal ideation and uses Word2vec for initialization. The dimensions of the vectors trained by both Word2vec and ELMo are set to 300. The main training parameters of the ELMo model are shown in Table 9.

| Parameters        | Description                        | Value |
|-------------------|------------------------------------|-------|
| Pro_dim           | the dimension of word embeddings   | 300   |
| nepoch            | the count of epoch                 | 10    |
| dropout           | dropout                            | 0.1   |
| batch_size        | the size of the batch              | 128   |
| learning_rate     | learning rate                      | 0.2   |
| unroll_steps      | the window size of the language model | 10    |

BERT is pretrained by Google on 3.3 billion words and 2.5 billion Wikipedia and 800 million text corpora. The word vector dimension is set to 768. At the same time, when using BiLSTM to obtain the char embeddings to concatenate, parameter dim_char is set to 256.

VI. EXPERIMENT ANALYSIS
A. DATASET AND EVALUATION METRICS

1) DATASET
After manual annotation of the suicidal ideation dataset in Section III-C, a total of 5,994 posts containing SIC were obtained (subsequent work will not take posts without SIC into account). They were randomly divided into the training set, the validation set, at an 8:1:1 proportion. The distribution of posts and labels in the SIC dataset are listed in Table 10.

It can be seen from Table 10 that the dataset has a label category imbalance problem, and the number of labels “O” is much larger than the number of other labels. This experiment has not yet dealt with this imbalance problem, which will also be one of the follow-up research projects.

2) EVALUATION METRICS
We used precision, recall, and F1-score as the evaluation metrics. Because an SIC may contain more than one word, to investigate the performance of the SICE model in detail, this paper uses four methods in two categories to determine whether the results of model extraction are correct.

a: EVALUATION GRANULARITY IS THE CAUSE PHRASE
   a. Complete Match: For an SIC that contains multiple words, each part (B, M, E) of the cause is required to match exactly. The corresponding precision, recall and F1 values are recorded as C_P, C_R, and C_F, respectively.
   b. Partial Match: As long as there is an overlap between the labeled cause and the ideal cause, it is considered correct. This metric mainly investigates whether the location of the cause is correct, and the corresponding precision, recall and F1 values are denoted P_P, P_R, and P_F, respectively.

b: EVALUATION GRANULARITY IS WORDS
   a. Exact Match: Taking the words in the SIC as the evaluation granularity does not consider whether the whole cause phrase
is correctly labeled. The corresponding precision, recall and F1 values are recorded as E_P, E_R, and E_F, respectively.

b. Fuzzy Match: Unlike the exact match method, the fuzzy match method does not distinguish the label category, such as B, M, E, or S. The fuzzy match is equivalent to the evaluation of binary classification problems. The corresponding precision, recall and F1 values are recorded as F_P, F_R, and F_F, respectively.

### B. EVALUATION RESULTS AND ANALYSIS

The evaluation results from the CRF and Char-BiLSTM-CRF using the SIC dataset are shown in Table 11.

#### 1) FEATURE ANALYSIS

Word features should be used as the most basic feature. In the CRF model, the effects of various types of atomic features on the task are validated. The word feature, W, performs best, and the C_F value reaches 0.661, which is followed by the POS features and the dependency features, DP, and their C_F value are 0.659 and 0.455, respectively. In general, the word feature, W, is the best and most basic feature, because the expression of SIC depends more on the vocabularies themselves.

The expert knowledge-based features can effectively improve the performance of the CRF model. Both “W + ET” and “W + LG” are 0.8% higher than W's C_F. When examining combinational features, the best combination of atomic features does not produce the best results. For example, the combined features of “W + POS” and “W + DP” perform obviously worse than the combined features of “W + ET” and “W + LG”. This finding shows that the POS features and DP features can be somehow covered by the word feature itself. This finding is easy to explain in that the POS of a word and its dependency in a sentence depend more on the word’s meaning itself. On the other hand, the linguistic features (LGs) and emotional features (ETs) using the LIWC dictionary and the emotion dictionary perform better than POS and DP because they have expert knowledge, which also shows that the SIC is closely related to the expression of emotions or feelings. Moreover, the performance of “W + ET” and “W + LG” are almost the same, but the performance of “W + ET + LG” and “W + LG + ET” declined based on “W + ET” and “W + LG” respectively, which shows that in SICE, the LGs and ETs are basically equivalent.
The suicide dictionary has an auxiliary effect on the CRF model. Compared with “W + ET”, after adding the PCS features, the C_F value increased by approximately 1%, which indicates that although the suicide words are not well-distinguished in terms of SIC and non-SIC phrases, they are still good indicators for SICE. In addition, based on combining all the features, “W + ET + PCS + DP + LG + POS”, removing PCS significantly reduced extraction performance, especially recall, which dropped by 1.9% on average.

Dependency features can improve the performance of CRF models. Compared with “W + ET + PCS”, the combined feature “W + ET + PCS + DP”, produced no significant improvement in C_F, but produced a 0.3% - 0.7% improvement in P_F, E_F and F_F. In addition, based on all features “W + ET + PCS + DP + LG + POS”, precision, C_P, was basically unchanged after removing DP. However, recall, C_R, decreased by 2.3%. This finding shows that adding DP can improve the performance of SICE by capturing the SIC and the dependencies between SIC and other words.

Finally, the CRF model achieved the best results after combining all six types of features, that is, “W + ET + PCS + DP + LG + POS”, where C_P, C_R, and C_F are 0.755, 0.620, and 0.681, respectively.

We also tried to add features used in the CRF model to the Char-BiLSTM-CRF model in two ways, adding features in the CRF layer and adding features in the Char layer, but the performance was not significantly improved, indicating that the Char-BiLSTM-CRF model had a better understanding of semantics and how to make full use of feature engineering in deep learning models remains to be further studied.

2) MODEL ANALYSIS

Compared with the CRF model, the neural network-based models are shown in Table 11 (the default word embeddings is Word2vec). Although the neural network based models perform well in a variety of tasks, they do not have a stability advantage if the model structure is not optimized for a specific task. For example, the basic BiLSTM in Table 11 is not better than the CRF. Specifically, compared to BiLSTM, the C_P, C_R, and C_F of the CRF were 9.2%, 2.4% and 5.4% higher, respectively.

The improved Char-BiLSTM-CRF model has more obvious advantages:

1) The Char-BiLSTM-CRF model is generally better than the CRF model, but CRF still has advantages in precision. The Char-BiLSTM-CRF model performed better than the CRF model on four harmonic average values of C_F, P_F, E_F, and F_F, and it mainly improved recall. For example, the C_R, P_R, E_R and F_R of the Char-BiLSTM-CRF model were 4.8%, 4.7%, 5.1%, and 5.2% higher than that of the CRF model, respectively. However, the experimental results also showed that the C_F, P_F and P_P of the Char-BiLSTM-CRF model were 0.2%, 0.5%, and 1.1% lower than that of the CRF model, respectively. This finding showed that although the neural network-based model performed well on classification tasks, the CRF still had a competitive advantage on sequence labeling tasks, especially in situations where training samples were very limited, such as the SICE task.

2) Adding character embeddings can significantly improve the performance of SICE.

From the comparison of the evaluation results of Char-BiLSTM with BiLSTM and Char-BiLSTM-CRF with BiLSTM-CRF, adding the character embeddings made the F1 value of each evaluation metric increased by approximately 2% on average, and recall increased by approximately 4% on average. It showed that compared with using only word embeddings, adding character embeddings enhanced the semantics of words. Therefore, it significantly improved the performance of SICE because with informal expressions, such as Weibo posts, the accuracy of word segmentation is reduced, and the text representations in terms of words becomes less accurate due to word segmentation errors and the increase in unknown words. The use of character embeddings can effectively alleviate this problem.

3) Adding the CRF layer can optimize sequence labeling results.

Comparing BiLSTM with BiLSTM-CRF and Char-BiLSTM with Char-BiLSTM-CRF, after adding the CRF layer, the precision of C_P and P_P based on taking a complete SIC as the evaluation granularity increased significantly. In particular, the complete match precision, C_P, increased by 8.9%, and C_F increased by 6.1%. This finding showed that CRF better captured the dependency relationship between adjacent labels through learning transition probabilities, and made up for the poor performance of sequence detection by BiLSTM and Char-BiLSTM in the SICE.

4) After replacing the embedding of Word2vec with the well-trained ELMo and BERT, the Char-BiLSTM-CRF model achieved better results. Compared with the Char-BiLSTM-CRF using Word2vec, the C_F, P_F, E_F and F_F of Char-BiLSTM-CRF (ELMo) increased by 4.7%, 5.6%, 5.4% and 5.7%, respectively. The explanatory ELMo changed dynamically according to the context, which can better understand the contextual meaning of words to effectively identify the cause of suicidal ideation. In the same way, compared with the Char-BiLSTM-CRF model using Word2vec, the C_F, P_F, E_F and F_F of Char-BiLSTM-CRF (BERT) increased by 8.2%, 5%, 7.7% and 5.9%, respectively. Based on ELMo, the C_F and E_F of Char-BiLSTM-CRF (BERT) increased by 3.5% and 2.3%, respectively. This finding indicated that BERT was more advantageous with the exact match and BERT understood the semantics of words better than ELMo, especially when dealing with polysemy.

3) DEEPER-ANALYSIS OF THE CHAR-BILSTM-CRF MODEL

By analyzing the results of Char-BiLSTM-CRF, it was found that the model performed better in cases such as the short cause, concise expressions and abstract SICs. Due to the greater commonalities between these expressions, there were more samples with similar semantics and structure in the dataset, and the model could be fully trained, such as the
following example in the corpus (Note: the prediction results are outside the brackets, and the manual or ideal answers are in the brackets).

Ex.5: 走饭1，| 我| 坚持/B(B) 不/M(M) 去/E(E) 了/| 我| 想/I/I 去/I/I 找/I/I 我/妈/I，| 我| 去/I/I 找/I/I 你/I，| 我/看/I/I 见/I/I 是/I/I 不/I/I 起/I/I 就/I/I 好/I/I 了/。

Zou Fan, I can’t/B(B) hold/M(M) on/M(M) any/M(M) longer/E(E). I want to find my mother and see that whether it is better if I die.

Ex.6: 每天/| 都/I/I 不/I/I (B) 开心/E(E) 所以/I/I (S(S) ) (M) 却/I/I 竟然/I/I (M) 干/了/E(E) 嘛/I/I (M) 的/I/I (M) [[微笑/I/I]]。

I’m not/B(B) happy/E(E) every day, so what’s the purpose of living?

For the examples they are difficult to predict, but the features can be summarized as follows:

(1) The causes are more specific and unusual, and they cannot be fully trained by the model. For example:

Ex.7: 8月/|5号/| 我/I/I 打/B(B) 疫苗/M(M) 了/I/I | 我/I/I (M) 的/I/I (M) 男孩/I/I (E(E) 小孩/I/I (M) 从来/I/I 没/I/I 想/I/I 过/I/I 自己/I/I 会/I/I 这/I/I 么/I/I 痛苦/I/I (S(S) )。

I had/B(B) an/I/I abortion/E(E) on/August/|5/| and/I/I I/I/I never/I/I thought/I/I I/I/I would/I/I be/I/I so/I/I miserable/S(S)。

Ex.8: 四/I/I 个月/| 前/I/I 丢失/I/I 不/I/I (B) 能/I/I (M) 生育/I/I (E) 背负担/I/I 各种/I/I 生活/I/I 压力/I/I (S(S) )/I/I 我/I/I 得/I/I 了/I/I 后悔/I/I (B) 抑郁/I/I (S(S) )/I/I 因此/I/I 放弃/I/I (M) 了/I/I (M) 最/I/I (M) 心/I/I (M) (M) 男孩/I/I (E) 现/I/I 在/I/I 感觉/I/I 自己/I/I (M) 会/I/I 就/I/I 是/I/I 赞成/I/I (S(S) )。

Four months ago, I was found to be/I/I infertile/E(E), and I was under various pressures/S(S). I have an/I/I (M) depression/S(S), and I/I/I gave/I/I up/I/I my/I/I (M) beloved/I/I (M) boyfriend/I/I (M). Now/I/I it/I/I ’s/I/I a/I/I burden/S(S) to/I/I alive.

(2) The long posts with irregular expressions contained many causes. This problem is similar to the problem encountered with manual annotation. There are great challenges with both manual annotation and model learning when dealing with long posts. The sentence structure is not clear, and multiple causes are complicated. In the following example, the whole sentence had no notation in Chinese:

Ex.9: 我/I/I 结婚了/I/I 了/I/I /M 父母/I/I 的/I/I (M) 意愿/I/I (M) 结婚/I/I (E) 了/I/I /M 可是/I/I /M 我/I/I 融入/I/I (B) 不/I/I (M) 到/I/I (M) 新/I/I (M) 的/I/I (M) 家庭/I/I (E) 也/I/I 不/I/I 想/I/I 即使/I/I 现/I/I 在/I/I 我/I/I 怀/I/I 育/I/I 四/I/I 个月/I/I 了/I/I 我/I/I 也/I/I 不/I/I (B) 想/I/I (M) 改变/I/I (M) 从前/I/I (M) 的/I/I (M) 生活/I/I (E) 也/I/I /M 想/I/I 到/I/I 要/I/I 改变/I/I 很多/I/I 我/I/I 就/I/I 觉/I/I (S(S) )/I/I 感觉/I/I 这/I/I 个/I/I 的/I/I 家庭/I/I (S(S) )/I/I 是/I/I (M) 个/I/I (M) 感/I/I (S(S) )/I/I 自/I/I 失/I/I (M) 了/I/I (M) 自由/I/I (E) 膨/I/I 胀/I/I 了/I/I (M) 不/I/I (B) 能/I/I (M) 再/I/I (M) 回到/I/I (M) 从前/I/I (M) 不/I/I (B) 想/I/I (M) 离开/I/I (M) 父母/I/I (E) 可/I/I 不/I/I 敢/I/I 说/I/I 有时/I/I 想/I/I (M) 还/I/I (M) 就/I/I 好/I/I (M) 了/I/I 在/I/I 孩子/I/I 出生/I/I 的/I/I 那/I/I 刻/I/I 就/I/I 再/I/I 也不/I/I 用/I/I 面对/I/I 这/I/I 一切/I/I 就/I/I 好/I/I (M) 了/I/I。

I got married according/M to/I/I my/I/I (M) parents/I/I (M) wishes/I/I (E). But/I/I I/I/I can’t/I/I (B) fit/I/I in/I/I (M) the/I/I (M) new/I/I (M) family/I/I (E) (E), and I/I/I don’t/I/I want/I/I to/I/I. Even/I/I if/I/I I/I/I pregnant/I/I four/I/I months/I/I now/I/I, I/I/I don’t/I/I (B) want/I/I to/I/I change/I/I (M) my/I/I (M) life/I/I (E). I/I/I get/I/I annoyed/S(S) at/I/I the/I/I thought/I/I of/I/I changing/I/I a/I/I lot/I/I. I/I/I think/I/I this/I/I family/I/I is/I/I (S(S) )/I/I a/I/I (M) burden/S(S). I/I/I lost/I/I (B) freedom/I/I (E) and/I/I became/I/I forward-looking/I/I. I/I/I can’t/I/I (B) go/I/I (M) back/I/I (M) to/I/I (M) the/I/I (M) past/I/I (E) and/I/I I/I/I don’t/I/I (B) want/I/I to/I/I (M) leave/I/I (M) my/I/I (M) parents/I/I (E), but/I/I I/I/I dare/I/I not/I/I say/I/I. Sometimes/I/I I/I/I just/I/I feel/I/I dead.

At the moment of the baby’s birth, it would be nice not to have to face it all again.

(3) There are some rare causes in the corpus, such as “smoking”, “father’s illness”, “Zoloft”, “buried in marriage”, “gastrorrhagia”, “not tall” and others. Similarly, these causes lead to the model not being fully trained.

In addition, while analyzing failed examples labeled by Char-BiLSTM-CRF, we found that there were still some problems with the manual annotation. On the one hand, there were subjective factors in the process of manual annotation. On the other hand, based on the principle of “Terse and concise”, manual annotation usually only marked the smallest substring reflecting the cause of suicidal ideation, whereas Char-BiLSTM-CRF easily predicted a much longer result that contained more complete semantic words, including the annotated answer:

Ex.10: 阿饭/， | 最近/ | 我/I/I 身体/I/I 不/I/I (B) 好/I/I (E) | 太/I/I (M) 疲劳/I/I (M) | 好/I/I (E) | 我/I/I 生/I/I (B) 也/I/I (M) 涕/I/I (M) | 精/I/I (S(S) )/I/I (E) 的/I/I (M) | 兴/I/I (M) | 雨/I/I (M) | 过/I/I (M) | 了/I/I | 现/I/I 在/I/I 我/I/I 这/I/I 个/I/I 世界/I/I 下/I/I。

I don’t know why people such as me who are abandoned/B(S) by/I/I the/I/I world/E/E still/I/I appear/I/I in/I/I this/I/I world.

VII. CONCLUSION AND FUTURE WORK

SICE can help psychological consultants or researchers quickly understand the causes of psychological crisis through social media and provide a basis for intervention or decision-making. In this paper, based on texts with suicidal ideation in Chinese posts, the definition of the SICE task is given based on related works on SIC and the suggestions of psychological experts. To generate the SICE dataset, we manually annotated the causes in the posts with suicidal ideation collected and annotated by the Psychology Institute of the Chinese Academy of Sciences. We proposed two models, the CRF model and the Char-BiLSTM-CRF model, to extract SICs in posts. We also analyzed the role of various features in the CRF model in detail and compared the traditional CRF model based on feature engineering and the Char-BiLSTM-CRF model based on a neural network. Finally, examples are given to illustrate the challenge of SICE.

The experiment shows that although SICE needs to be further improved, the precision and recall of both proposed models can still reach 0.933 and 0.798. This finding shows that it is highly feasible to monitor large-scale psychological crises and analyze the cause of the crisis through social media.

At the same time, we also see that SICE from social texts still faces great challenges, and we need to conduct more in-depth research on data annotation, the extraction model and making better use of expert knowledge or common sense. The future works we will focus are listed as follows:

(1) Automatic or semiautomatic data annotation remains to be explored. This article uses manual annotation, which is time-consuming, labor-intensive, and subjective. Exploring automatic or semiautomatic annotation methods and few shot
or zero shot learning for SICE will be one of our next research directions.

(2) The generation of suicidal ideation has strong timeliness and is comprehensive. It needs more consideration of timeliness and other factors.

In this research, we only extracted SICs from a single post. However, the generation of suicidal ideation is often gradual and for many reasons. Therefore, how to utilize more information posted by users, how to investigate the time dimension, and how to comprehensively consider the potential causes are the second direction for future research.

(3) Introduce expert knowledge and common sense into the model.

As seen from the examples in Section VI-B-3, “smoking”, “father’s illness”, “Zoloft” (an antidepressant drug), “buried in marriage”, “gastrorrhagia” and other causes need to be connected with depression, disease, family relationships, etc. With the help of expert knowledge or common sense, extraction would be improved. At the same time, the specific cause should be further generalized and abstracted to solve the problem of insufficient model training through expert knowledge or common sense.

(4) Optimize the model to deal with the samples with long sentences, irregular sentences, long causes and mixed causes.

The examples in Section VI-B-3 also show that when the sentence is long and irregular or the cause is long and has multiple reasons, the model’s ability to capture the SIC decreases. The reason is that the semantics of the sentence or SIC cannot be clearly detected. It may be because the deep learning model with character-level embeddings or word-level embeddings cannot fully learn the semantics at the sentence level or phrase level. Therefore, making targeted improvements to the model is a direction to improve the extraction.

ACKNOWLEDGMENT

Thanks to Prof. T. Zhu of the Institute of Psychology, Chinese Academy of Sciences for his suicidal ideation dataset and Suicide Dictionary.

REFERENCES

[1] World Health Organization (WHO). Suicide. Accessed: Sep. 2, 2019. [Online]. Available: https://www.who.int/zh/news-room/fact-sheets/detail/suicide
[2] R. L. Bonner and A. R. Rich, “Toward a predictive model of suicidal ideation and behavior: Some preliminary data in college students,” Suicide Life-Threatening Behav., vol. 17, no. 1, pp. 50–63, Mar. 1987.
[3] A. T. Beck, M. Kovacs, and A. Weissman, “Hopelessness and suicidal behavior: An overview,” J. Amer. Med. Assoc., vol. 234, no. 11, pp. 1146–1149, 1975, doi: 10.1001/jama.1975.03360240050026.
[4] R. P. Auerbach, J. Alonso, W. G. Axinn, and P. Cuypers, “Mental disorders among college students in the World Health Organization World Mental Health Surveys.” Psychol. Med., vol. 46, no. 14, pp. 2955–2970, 2016.
[5] T. Ao, S. Shetty, T. Sivili, C. Blanton, H. Ellis, P. L. Geltman, J. Cochran, E. Taylor, E. W. Lankau, and B. Lopes Cardozo, “Suicidal ideation and mental health of Bhutanese refugees in the united states,” J. Immigrant Minority Health, vol. 18, no. 4, pp. 828–835, Aug. 2016.
[6] X. Huang, X. Li, T. Liu, D. Chiu, T. Zhu, and L. Zhang, “Topic model for identifying suicidal ideation in Chinese microblog,” in Proc. Pacific–Asia Conf. Lang. Inf. Comput., 2015, pp. 553–562.
[7] G. Coppersmith, K. Ngo, R. Leary, and A. Wood, “Exploratory analysis of social media prior to a suicide attempt,” in Proc. 3rd Workshop Comput. Linguistics Clin. Psychol., 2016, pp. 106–117.
[8] N. Berry, F. Lobban, M. Belousov, R. Emsley, G. Nenadie, and S. Bucci, “WhyWeTweetMH: Understanding why people use twitter to discuss mental health problems,” J. Med. Internet Res., vol. 19, no. 4, p. 107, 2017.
[9] Z. Tan, X. Liu, X. Liu, Q. Cheng, and T. Zhu, “Designing microblog direct messages to engage social media users with suicide ideation: Interview and survey study on weibo,” J. Med. Internet Res., vol. 19, no. 12, p. e381, Dec. 2017.
[10] J. K. Kok, C. C. Gan, and L. Y. Goh, “Youths talk about death: Causes for suicide,” Int. J. Social Sci. Humanity, pp. 145–149, 2011.
[11] I. Chatterjee and J. Basu, “Perceived causes of suicide reasons for living and suicidal ideation among students,” J. Indian Acad. Appl. Psychol., vol. 36, no. 2, pp. 311–316, 2010.
[12] M. Luo, J. Li, and Y. He, “Research on suicide ideation of elderly people,” (in Chinese), Sci. Res. Aging, vol. 3, no. 7, pp. 41–57, 2015.
[13] C. Maskill and I. Hodges, “Explaining patterns of suicide,” Minstry Health, Wellington, New Zealand, Dec. 2005. [Online]. Available: https://www.moh.govt.nz
[14] R. Hassan, “Social factors in suicide in Australia,” Trends Issues Crime Criminal Justice, no. 52, 1996.
[15] E. Chesney, G. M. Goodwin, and S. Fazel, “Risks of all-cause and suicide mortality in mental disorders: A meta-review,” World Psychiatry, vol. 13, no. 2, pp. 153–160, 2014.
[16] M. H. Daily. 15 Common Causes of Suicide, Why People Kill Themselves. [Online]. Available: https://mentalhealthdaily.com/2014/07/23/15-common-causes-of-suicide-why-people-kill-themselves
[17] Suicide.org (Suicide Survivors Forum). Suicide Causes. [Online]. Available: http://www.suicide.org/suicide-causes.html
[18] R. C. Oconnor and M. K. Nock, “The psychology of suicidal behaviour,” Lancet Psychiatry, vol. 1, no. 1, pp. 73–85, 2014.
[19] G. Turecki and D. A. Brent, “Suicide and suicidal behaviour,” Lancet, vol. 387, no. 10024, pp. 1227–1239, 2016.
[20] D. E. Schotte and G. A. Clum, “Problem-solving skills in suicidal psychiatric patients,” J. Consulting Clin. Psychol., vol. 55, no. 1, pp. 49–54, 1987.
[21] V. Cernat. (Nov. 5, 2000). A Coherence Optimization Model of Suicide. [Online]. Available: http://cogprints.org/1081/1/coherencesuicide.htm
[22] 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.
[23] J. Cheng, “A review of current research on mechanism of suicide,” (in Chinese), J. Hebei Polytech. Univ. (Social Sci. Ed.), vol. 4, no. 8, pp. 29–32, 2008.
[24] D. Wasserman and M. Li, Suicide: An Unnecessary Death. China, Beijing: China Light Industrial Press, 2003, pp. 189–194.
[25] M. Wray, C. G. Colen, and B. A. Pescosolido, “The sociology of suicide,” Rev. Sociol., vol. 37, no. 1, pp. 505–528, 2011.
[26] P. J. Dean, L. M. Range, and W. C. Goggin, “The escape theory of suicide in college students: Testing a model that includes perfectionism,” Suicide Life Threatening Behav., vol. 26, no. 2, pp. 181–186, 1996.
[27] D. de Catananzo, “Reproductive status, family interactions, and suicidal ideation: Surveys of the general public and high-risk groups,” Ethol. Sociobiol., vol. 16, no. 5, pp. 385–394, Sep. 1995.
[28] J. C. Lessard and M. M. Moretti, “Suicidal ideation in an adolescent clinical sample: Attachment patterns and clinical implications,” J. Adolescence, vol. 21, no. 4, pp. 383–395, Aug. 1998.
[29] P. R. Fortes, D. S. Sandhu, and R. Longwedge, “Understanding adolescent suicide: A psychosocial interpretation of developmental and contextual factors,” Adolescence, vol. 37, no. 148, pp. 805–814, 2002.
[30] J. D. Ribeiro and T. E. Joiner, “The interpersonal-psychological theory of suicidal behavior: Current status and future directions,” J. Clin. Psychol., vol. 65, no. 12, pp. 1291–1299, Dec. 2009.
[31] J. Zhang, “A Preliminary Study of ‘incompatible stress theory’ for suicide (review),” (in Chinese), Chin. Mental Health J., no. 11, pp. 63–67, 2005.
[32] X. Wu, H. Zheng, and Q. Li, “Suicidal ideation and associated factors among university students,” (in Chinese), Chin. J. Public Health, vol. 32, no. 3, pp. 359–362, 2016.
[33] L. Zhou, “Analysis of influencing factors of suicidal idea in college students,” (in Chinese), Sci. Social Psychol., no. 10, pp. 27–30, 2015.
[34] Z. Zhou et al., “Influencing factors of suicide ideation among migrant service workers in Shenzhen city,” (in Chinese), Chin. J. Public Health, vol. 32, no. 7, pp. 948–952, 2016.
[83] Z. Huang, W. Xu, and K. Yu, “Bidirectional LSTM-CRF models for sequence tagging,” 2015, arXiv:1508.01991. [Online]. Available: http://arxiv.org/abs/1508.01991
[84] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, “Neural architectures for named entity recognition,” in Proc. North Amer. chapter Assoc. Comput. Linguistics, 2016, pp. 260–270.
[85] X. Ma and E. Hovy, “End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF,” 2016, arXiv:1603.01354. [Online]. Available: http://arxiv.org/abs/1603.01354

DEXI LIU received the Ph.D. degree in computer software and theory from Wuhan University, Wuhan, China, in 2007. He is currently a Full Professor with the School of Information Management, Jiangxi University of Finance and Economics. His research interests include web information management, information retrieval, social computing, and natural language processing. He is a Senior Member of the China Computer Federation.

QI FU received the M.S. degree in computer science and technology from East China Jiaotong University, Nanchang, China, in 2007. She is currently pursuing the Ph.D. degree with the School of Information Management, Jiangxi University of Finance and Economics. Her research interests include machine learning and social media processing.

CHANGXUAN WAN received the Ph.D. degree in computer science and technology from the Huazhong University of Science and Technology, Wuhan, China, in 2003. Since 1989, he has been with the Jiangxi University of Finance and Economics, where he is currently a Professor. His research interests include data mining, sentimental computing, web data management, and information retrieval.

XIPING LIU received the Ph.D. degree in management science and engineering from Jiangxi Normal University, Nanchang, China, in 2010. He is currently a Professor with the School of Information Management, Jiangxi University of Finance and Economics. His research interests include big data analysis, data mining, and database technology.

TENGIJAO JIANG received the Ph.D. degree in management science and engineering from Jiangxi Normal University, Nanchang, China, in 2015. She is currently an Associate Professor with the School of Information Management, Jiangxi University of Finance and Economics. Her research interests include big data analysis, data mining, and database technology.

GUOQIONG LIAO received the Ph.D. degree in computer software and theory from the Huazhong University of Science and Technology, Wuhan, China, in 2003. He was a Postdoctoral Research Fellow with the Huazhong University of Science and Technology from 2003 to 2005. He is currently a Professor with the School of Information Management, Jiangxi University of Finance and Economics. His research interests include data mining and social network analysis.

XIANGQING QIU received the M.E. degree in software engineering from Fuzhou University, Fuzhou, China, in 2010. He is currently pursuing the Ph.D. degree with the School of Information Management, Jiangxi University of Finance and Economics, Nanchang. His research interests include natural language processing and affective computing.

RENQIANG LIU received the B.S. degree in educational technology from Gannan Normal University, Ganzhou, China, in 2017. He is currently pursuing the M.S. degree in computer technology with the Jiangxi University of Finance and Economics, Nanchang. His current research interests include data mining, machine learning, and social media processing.