MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190181

Unveiling Online Suicide Behavior: What Can We Learn About Mental Health from Suicide Survivors of Reddit?

Ashwin Karthik Ambalavan[a], Bilel Moulahi[b], Jérome Azé[c], Sandra Bringay[c,d]

[a] *Arizona State University, Tempe, AZ, USA*
[b] *Acelys Informatique, Montpellier, France*
[c] *LIRMM, University of Montpellier, CNRS, Montpellier, France*
[d] *University Paul Valéry Montpellier, Montpellier, France*

Abstract

Suicide is a growing public health concern in online communities. In this paper, we analyze online communications on the topic of suicide in the social networking platform, Reddit. We combine lexical text characteristics with semantic information to identify comments with features of suicide attempts and methods. Then, we develop a set of machine learning methods to automatically extract suicide methods and classify the user comments. Our classification methods performance varied between suicide experiences, with F1-scores up to 0.92 for "drugs" and greater than 0.82 for "hanging" and "other methods". Our exploratory analysis reveals that the most frequent reported suicide methods are drug overdose, hanging, and wrist-cutting.

Keywords:

Suicide, attempted; Natural Language Processing; Social Media

Introduction

Social media platforms, such as Twitter, Facebook, and Reddit, are bringing important challenges for a variety of social- and health-related phenomena. The widespread use of these modern means of communication creates a potential to identify and characterize users behaviors by analyzing content shared online. Recent studies have shown that people are more likely to seek support from informal online resources, rather than seeking formal treatment from professionals [1] and reported a positive correlation between suicide rates and the volume of social media posts related to suicidal ideation [2]. This research suggests that online sources may contain valuable information about specific cohorts, such as vulnerable populations, and their suicidal behavior.

In this paper, we focus on the content analysis of forums dedicated to a specific population -- suicide survivors, or which is defined here as those who have had suicide attempts. Our work aims to contribute to the literature by understanding and analyzing communications on the topic of suicide in the social networking platform, Reddit.[1] As of February 2018, Reddit is a highly trafficked website[2] with more than 542 million monthly visitors, 725.85 million comments, and 6.89 billion up-votes from its users.[3] Reddit provides users anonymity while sharing and discussing information about almost everything, without any bias on the expressed opinions and feelings. In contrast to other social media, it provides domain-specific discussion forums that are moderated by a number of users and that carry health-related knowledge expressed by various cohorts.

In our work, we analyze a serious Reddit thread[4] detailing extremely personal and traumatic experiences of suicide survivors. This thread is moderated by 41 users and includes only serious comments about the topic of suicide attempts. We argue that the study of this dataset is an important step towards the development of automated methods to identify suicide risk factors that could help mental health professionals and psychiatrists to understand and prevent suicide.

Our main objective is to identify methods used to commit suicide such as drug overdose, hanging and to quantify their evocation in social media.

Our research contributions include: (1) extraction and analysis of linguistic characteristic and key aspects of the language used by suicidal individuals to describe suicide methods; and (2) development of a set of machine learning algorithms to extract methods used by suicide attempters and classify the Reddit comments.

State of the Art

Suicide is one of the most common health issues impacting the world's population. According to the Centers for Disease Control and Prevention [3], more than 40,000 suicides were reported in the United States in 2012, which positioned Suicide as the 10th leading cause of death in the country.

A rich collection of work has been done on social media as an effort to identify and understand communications about mental health problems, including depression, mental disorder and suicide. While most of the research from literature is done on suicide notes [4], more recent work demonstrated that evaluating suicidal risk factors in social networks can be used to prevent suicide and detect suicidal ideation in its early stages.

For example, the Durkheim[5] project studied the online activities and shared content of group of US war veterans on Twitter, Facebook, and LinkedIn in order to identify markers of harmful behaviour. The group developed linguistics-driven prediction models to estimate the risk of suicide using text from the clinical notes [5]. Results showed that people who committed suicide frequently recorded behaviours indicative of fear, agitation and delusion, with around 65% accuracy.

[1] www.reddit.com
[2] https://redditblog.com/2015/12/31/reddit-in-2015/
[3] https://en.wikipedia.org/wiki/Reddit
[4] https://redd.it/4e8oip/
[5] http://www.durkheimproject.org/research/

In the same line of research, Sueki [6] examine a panel of young (early 20s) Twitter users to evaluate the association between suicide-related tweets and suicidal behaviour. The authors investigated the linguistic features of suicidal ideation and identified the most important markers of future suicide. For example, phrases such as *"want to commit suicide*" were found to be strongly associated with lifetime suicide attempts, while phrases that suggest suicidal intent, such as *"want to die*," were found to be less strongly associated. For a complete summary of literature on suicidal thoughts and behaviours, the reader may refer to the recent pioneering work of [7], who conducted a meta-analysis of 365 studies from the past 50 years.

Most closely related to the current paper, Ghitsis [8] investigate the linguistic characteristics of Reddit posts that need urgent attention. The authors focussed on subreddits in which users post comments about *Addiction, Anxiety, Dementia, Depression, self-harm* and *suicide ideation.* However, the work does not focus specifically on serious suicide posts, and, most importantly, does not investigate the classification of methods of suicides, which in turn could allow the understanding of several potential warning signs.

With the present work, we consider a broad range of linguistic, lexical, and semantic features to facilitate the task of methods classification. We specifically address the following research questions:

1. How do suicidal individuals communicate about the method used during suicide attempt experiences?
2. What are the most popular methods used by suicidal individuals to end their life?
3. How to recognize a comment dealing about suicide method?

Methods

Step 1: Data acquisition

Our dataset consists of all the comments of the Reddit thread, "*[Serious] Suicide survivors of Reddit, what was your first conscious thought after you realized that you hadn't succeeded?*"[6] We have scraped all the comments using the Python Reddit API Wrapper. The topic has been tagged with the keyword "[Serious]", which means that it is monitored by a group of moderators who remove any off-topic and irrelevant comments in the thread, keeping only serious comments about suicidal ideation. In Reddit, all the comments are organized as a big tree of comments with the parent comment as the top level node and the comments about the parent comment as their child nodes. Table 1 reports statistics about the dataset.

The lack of informed consent given by social media users for data usage leads to ethical questions. In particular, confidentiality with respect to the publication of research results is an issue. We adhere to the guidelines of [9]. Results are presented with a degree of detail that does not permit drawing conclusions on individual users.

Step 2: Data Preprocessing

We initially preprocessed the data to remove HTML tags, white spaces, pipe symbols and carriage returns. We expand English word contractions such as won't, can't, to will not, cannot, respectively using regular expressions. These steps are essential to tokenize words properly and for negation identification. We lemmatize the words using both the Python's NLTK Wordnet lemmatizer.

Table 1– Dataset statistics

Total number of comments	6,229
Number of top-level comments	1,833
Number of sentences	12,782
Average number of sentences per comment	7
Number of unique users	3,58

Step 3: Feature Extraction

In the following, we define the features used in the classification process: Trigrams, NLTK POS Tags and Customised POS Tags.

Trigrams: Lemmas on their own have very little meaning and cannot be used as an independent criteria for annotating a sentence according to a method. Hence, we used trigrams to explore sentences and find recurring patterns. Some examples include: have to kill, going to bleed, tried to slit.

NLTK POS Tags: We tokenized and tagged all the words using NLTK POS tags. We obtained a total of 5,229 verbs, 10,901 adjectives and 24,815 nouns. We used NTLK POS-tagger.

Customised POS Tags: We have also manually defined Customized POS Tags described in Table 2. Due to space restrictions, we only reported the most frequent words associated with these tags. All the Customized POS Tags are self-explanatory, except for 'Ability' categories which represent the words that usually tend to be used by people who are expressing support need such as 'I need support'.

Table 2– Customized POS Tags

Customized POS TAG	Meaning	Common Tokens
FPRP	First Personal Pronouns	I, my, myself
SPRP	Second Personal Pronouns	You, him, he, it
WPRP	Group First Personal Pronouns	We, us, our
NEG	Negations	Not, note, never
SWR	Swear words	Wtf, ass, poop
INTSFR	Intensifiers	Really, so, very
NEGINTSFR	Negative intensifiers	Awfully, horrifyingly
POS	Positive words	Adore, affirmation
NEG	Negative words	Depression, sacrifice
ABLT	Ability words	Handle, support
METHOD	Suicide methods	Hang, drug, shotgun
POSEMO	Positive Emojis	Lol, LMAO, ROFL

extracted from an internal API using 8 online synonym dictionaries[7] and we keep a synonym only if it is find at least 3 in 3 dictionaries. After manually removing irrelevant tokens, we obtain a list of 218 tokens mapped to 15 suicide methods as mentioned in Table 3.

[6] https://redd.it/4e8oip/

[7] Reverso www.reverso.net, Bab.la fr.bab.la/dictionnaire, Atlas dico.isc.cnrs.fr, Thesaurus www.thesaurus.org, Orto-lang www.cnrtl.fr/synonymie/, SensAgent dictionnaire.sensagent.com/synonyme/en-fr/, The FreeDictionary www.thefreedictionary.com and the Synonym www.synonym.com, all retrieved on July 13, 2018

Table 3– Most frequent tokens associated with each suicide method label and its distribution in the dataset

Methods	Most frequent tokens	Distributions
Alcohol	Drunk, Liquor, boost	9,9
Bleed	Bleeding, blood	6
Carbon monoxide	Monoxide, gas-poisoning	0,2
Cut	Slit, cleave, tear	10,5
Starve	Starvation, famine	0,2
Disease	Illness, Sickness	4,7
Drown	Drowning, underwater	1,6
Drug	Overdose, pills, dope	36,1
Electrocute	Electrocution, electrocute	0,1
Gun	Shotgun, shoot, revolver	6
Hang	Hanging, noose, strangle	17,8
Hypothermia	Hypothermia	0,1
Jump	Leap, jump-over, jumping	4,2
Vehicle	Collision, speed, car	2,5
Suffocate	suffocation	0,1

FPRP, SPRP, WPRP and ABLT are self defined. POS and NEG list have been compiled from 4 lexicons [1,10-12].

The tag 'NEGINTSFR' is used when a word occurs in both the Negative words dictionary and the Intensifier dictionary such as 'awfully', 'horrifyingly'.

To build the METHOD list, first, we defined a list of 34 suicide methods obtained from dedicated mental health websites[8]. Then we enriched this list by using synonyms

The POSEMO list contains regular expression rules to detect positive social networking emotions including emojis and common slang such as *lol, lmao*... We also use regular expressions in order to include repeated letters as *looooool, lol.*

Step 4: Classification

We applied supervised machine learning algorithms in order to automatically predict and extract suicide methods at the comment level. Because a comment could be annotated with one or more of the labels listed in Table 3, the task is a multi-label classification task. Every comment containing at least one annotation was extracted. We obtained a training dataset of 874 annotated comments.

We used a One VS Rest Classifier with four different models: Support Vector Machine, Logistic Regression, SGD and Perceptron.

We perform automatic cross validation with 611 comments used as training set and 263 comments used as test set (30% of the dataset).

We tuned the hyper-parameters of the model for the four different models. Global system performance is measured with micro-averaged F1-score because of the use of a skewed dataset.

[8] https://en.wikipedia.org/w/index.php?title=Suicide_methods&oldid=788573333 http://regretfulmorning.com/2011/08/the-7-most-common-drugs-people-overdose-on/ both retrieved on July 13, 2018 Reverso www.reverso.net, Bab.la fr.bab.la/dictionnaire, Atlas dico.isc.cnrs.fr, Thesaurus www.thesaurus.org, Orto-lang www.cnrtl.fr/synonymie/, SensAgent dictionnaire.sensagent.com/synonyme/en-fr/, The FreeDictionary www.thefreedictionary.com and the Synonym www.synonym.com, all retrieved on July 13, 2018

Results

POS Tagged Sentence Sequence:

We tagged all the sentences with the Customised POS Tags mentioned above and look for regularities. In the following, we list two prominent examples:

- 'FPRP', 'VBD', 'TO', 'METHOD' (First Personal Pronoun followed by a verb, a preposition and the method) : "I decided to hang", "I tried to - slit, strangle, jump, or drown".
- 'FPRP', 'VBD', 'INTSFR', 'NEGATIVE' (First Personal pronoun followed by a verb, an Intensifier and a Negative Word) gives us the emotional state the person: "I felt more alone", "I became more depressed", "I was completely helpless"...

Customized POS Tagged Sequences are very useful to identify and extract patterns to characterize the way individuals communicate about the methods they used to commit a suicide attempt (RQ1).

Method Labelling

The distribution of the labels in the training set is also presented in Table 3. It is apparent that some methods to commit suicide are much more frequent compared to others. Some of the most frequent labels are, in order, *drug, hang, cut* and *alcohol*. It is interesting that most Reddit users who have commented on this submission preferred to commit suicide through *drug overdose* (RQ2).

Comment classification

Table 4 gives the comparison of the 4 used estimators along with their classification features (RQ3).

Table 4 – Classifier performances

Estimator	Accuracy	Micro-precision	Micro recall	F1-Score
Logistic Regression	0.452	0.726	0.534	0.615
Perceptron	0.631	0.863	0.709	0.778
SGD Classifier	0.642	0.856	0.722	0.783
Linear SVC	0.684	0.912	0.722	0.806

As seen in many different NLP tasks, Linear SVC has the best performance in comparison to Logistic Regression, Perceptron and SGD Classifier. Table 5 gives the classification report of Linear SVC for each individual suicide method label.

Label frequency determines classifier performance considerably. For all comments with support above 20, we can see that we have an F1-score of at least 50%. Unsurprisingly, classifiers for rare examples such as *Carbon Monoxide, Electrocute, Drown* perform worst because of the low number of samples. Classifier performance would likely improve for the low-frequency labels if more training data were obtained, without the need for new features.

Some labels (*gun, jump* and *disease*) perform better than would be expected from frequency alone, indicating that they are easier to learn than others, possibly because they are lexicalized more often, or more consistently.

Table 5 – Classification performance per methods

Methods	Precision	Recall	F1-score	Support
Alcohol	0.97	0.57	0.71	53
Bleed	1	0.52	0,69	42
Carbon monoxide	0.00	0	0	0
Cut	0.94	0.59	0,72	51
Disease	1	0.39	0,56	23
Drown	0	0	0	7
Drug	0.89	0.95	0,92	154
Electrocute	0	0	0	0
Gun	1	0.6	0,75	15
Hang	0.86	0,79	0,82	71
Hypothermia	0.94	0,81	0,87	21
Jump	1	0,5	0,67	2
Vehicle	1	1	1	1
Suffocate	1	0,29	0,44	7

Discussion

Based on the feature extraction process, we find that there are many prominent methods used by leading individuals to attempt suicide. Our analysis reveals that most of these causes appear in sentences either as a negative word or after an intensifier word. In table 6 we report some examples of comments with the number of votes as scored by Reddit. The user names of the authors are removed in order to preserve privacy.

Now, it would be interesting to combine this information about suicide methods with other dimensions such as sentiment analysis or causes such as "bullying" (User 1), "ptsd," "depression," and "divorce" (User 3). These others dimensions could be extracted from the text with a similar approach to the one used for the methods. Such signs could be used as an important signal to prevent suicide.

Table 6– Examples of comments

Author	Comment	Score
User 1	Well I **had attempted to hang myself** when **I was 15**. I have a **birth defect** (deformed legs) and **just could not take the bullying anymore**. So...	7,455
User 2	**I'am a diabetic.** I can remember deciding **to use insulin to go**. Figured that passing out and dying of a seizure due to hypoglycemia would be a **quick and easy way to go.**	1,460
User 3	**I deal with ptsd and depression** at the time I was dealing with my issues that I did after coming home from over seas **my divorce** and **I was getting close to losing my apartment I got drunk as helle got in a fight...**	709

On the other hand, the distribution of suicide method labels in the dataset (Table 3) shows that drug overdose is the most common method used to commit suicide followed by hanging, cutting, and alcohol poisoning, respectively. Since drug overdose is a growing cause of suicide, suitable precautionary methods should be taken to avoid it as much as possible. Some suggestions would be to identify the most common drugs used to commit suicide and to prescribe them to individuals only after throughly assessing their mental health.

The Reddit comment classification reveal that classifier performance depends on frequency of labels in the dataset. In order to limit the annotation effort, methods such as active learning [13] could be used to select the example to annotate in order to improve the model more rapidly.

Conclusions

In this paper we developed a set of methods based on natural language processing and machine learning in order to study the suicidal behavior of individuals who attempted suicide. We built a set of linguistic, lexical, and semantic features that help in capturing relevant language clues. These new features have been successfully used to improve the classification of suicidal thoughts, experiences, and, most importantly, suicide methods. We found that POS tagging features are important to identify and understand users' communications in forums especially when used as features within a machine learning method.

We also created and annotated a dataset that can be used by researchers to investigate others questions about suicide attempts. For future research, we plan to undertake large-scale experimental evaluation to assess the language of suicide across different types of data, such as microblogs and forums. It may be also instructive to determine whether users' classification could reveals similar profiles, suggesting similar suicide methods and comparable risk factors. The classification of the language used could also be informative.

References

[1] D. Rickwood, F.P. Deane, C.J. Wilson, and J. Ciarrochi. Young people's help-seeking for mental health problems. *Australian e-Journal for the Advancement of Mental Health (AeJAMH)*, **4** (2005) 218-251.

[2] P. Burnap, W. Colombo, and J. Scoureld. Machine classication and analysis of suicide-related communication on twitter. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media,* HT '15, pages 75-84, New York, NY, USA, 2015. ACM.

[3] S.L. Murphy, K.D. Kochanek, J. Xu, and E. Arias. Mortality in the United States. *NCHS Data Brief.*, **229** (2014) 1-8.

[4] J.P. Pestian, P. Matykiewicz, and J. Grupp-Phelan. Using natural language processing to classify suicide notes. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, BioNLP '08, pages 96-97, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics.

[5] C. Poulin, B. Shiner, P. Thompson, L. Vepstas, Y. Young-Xu, B. Goertzel, B. Watts, L. Flashman, and T. McAllister. Predicting the risk of suicide by analyzing the text of clinical notes. *PLoS ONE*, **9** (2014).

[6] H. Sueki. The association of suicide-related twitter use with suicidal behaviour: A cross-sectional study of young internet users in japan. *Journal of Affective Disorders*, **170** (2015) 155-160.

[7] J.C. Franklin, J.D. Ribeiro, K.R. Fox, K.H. Bentley, E.M. Kleiman, X. Huang, K.M. Musacchio, A.C. Jaroszewski, B. P. Chang, and M.K. Nock. Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin,* **143** (2017) 187-232.

[8] G. Gkotsis, A. Oellrich, T.J.P. Hubbard, R.J.B. Dobson, M. Liakata, S. Velupillai, and R. Dutta. The language of mental health problems in social media. In *Proceedings of*

the 3rd Workshop on Computational Content Analysis of Texts by Suicide Survivors of Reddit 15 Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, CLPsychNAACL-HLT, pages 63-73, 2016.

[9] R.T Kraut, J. Olson, M. Banaji, A. Bruckman, J. Cohen, and M. Couper. Psychological research online. Report of board of scientic affairs' advisory group on the conduct of research on the Internet. *American Psychologist*, **59** (2004) 108-117.

[10] B. Liu, M. Hu, and J. Cheng. Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of the 14th International Conference on World Wide Web,* WWW'05, pages 342-351, New York, NY, USA, 2005. ACM.

[11] S.M. Mohammad and P.D. Turney. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence,* **29** (2013) 436-465.

[12] M. Ueda, K. Mori, T. Matsubayashi, and Y. Sawada. Tweeting celebrity suicides: Users' reaction to prominent suicide deaths on twitter and subsequent increases in actual suicides. *Social Science & Medicine*, **189** (2017) 158-166.

Address for correspondence

Sandra Bringay, E-mail: bringay@lirmm.fr

Phone: +336 83 24 7933