Detecting Suicidal Victims through Corpus Analysis of Sympathizers Using Twitter Data

Anika Sharma
Department of Computer Science
Golisano College of Computing and Information Sciences
Rochester Institute of Technology
Rochester, NY 14586
axs7639@rit.edu

***Abstract*—Society that has become digitally dependent in many ways, especially in the way of communication. From people expressing their daily routines, ideas, and opinions, the phenomenon of social media offers ample amount of public data to discern and gain awareness from. The type of discernment intended through this project is to examine and detect sympathizers of victims who imply suicidal behavior through one of the largest social media platforms used today, Twitter. We are to perform analysis upon tweets collected in order to gain accurate insights on what type of patterns in tone and text are exemplified by catalysts of informal forms of support.**

I. Introduction

As much as social media helps and eases the process of communication in today's world, the harm that such public platforms can impose upon individuals may be substantial. Suicide is known to be the 18th leading cause of death worldwide. While many people who have suicidal ideation may be able to seek help either by initiation from themselves or through the intervention of concerned friends, family, or other worried individuals surrounding them, there still exists suffering that goes unnoticed and uninitiated by the sufferer themselves. This is where social media as well as an intimate support system play an imperative role in identifying which individual may be giving a subliminal cry for help through the words and emotions they choose to express on the World Wide Web.

In this project, we intend to draw attention towards these more hidden cries for help with the aid of Twitter. This social media platform holds an enormous amount of information to study from, all of which is publicly available. We use data which is in the form of tweets to analyze specific words, tone, or behavior that is helpful in deciphering whether suicidal thoughts are being conveyed or not, through the implications set by sympathizers.

II. Related Work

Roberts et al. [1] we are able to observe a proper end to end implementation of emotion detection through Twitter data. In this paper, the authors set out to to build a system that analyzes emotion based off of a more challenging medium than most social media practices, micro-blogging. Micro-blogging techniques, such as those used in Twitter, offer more limitations when it comes to analysis. This is due to the facts that (1) there is a limit on the length of characters a user may utilize and (2) the platform enables users to express themselves in real time and convey their immediate reactions, which can produce a more emotionally charged statement than normal. With this information, the authors build a system where a manually annotated corpus of tweets are created in order to categorize the expression as one of the seven emotions of: *ANGER, DISGUST, FEAR, JOY, LOVE, SADNESS,* and *SURPRISE*. Now with using this annotated corpus, we develop a baseline method to ultimately try and identify emotions automatically within a given corpus. This method we wish to construct is done with binary SVM classifiers as our machine learning model such that each classifier performs independently on a single emotion. The final observations this experiment leads to is the acknowledgement that there exits a difference between a topic-specific linguistic style and an emotion-specific linguistic style. While keeping this concluding thought in mind, we are aware that there can be improved machine learning systems built, both through unsupervised and supervised learning, such that automatic detection of emotions can output a more clear result.

Analysis done at Worcester Polytechnic Institute indicate another method of automating emotion detection. We know that Twitter data offers a wide-range of varying emotions and is also publicly and freely available to observe as an analyst. Hasan et al. [2] observes a study where the authors are determined to find a supervised method of modeling the emotions conveyed in Twitter messages, all while performing at a high rate of accuracy. The modeling work for this is done through using the well-established Circumplex model. The representation of emotions this model provides enables the characterization of individuals experiences based on two dimensions of valence and arousal. The methodology used here is that Twitter messages are used as input data, hashtags are used as labels, and the supervised classifiers are trained to

detect multiple levels of emotion. Using this Circumplex logic allows for the automation process to begin as early as the annotating data stage. This meaning, virtually no manual effort is happening even from the initial building of this system, letting a fully automated end to end detection system come to fruition. It is found that when performing an automated annotation technique upon different classification methods such as Decision Trees, SVM, KNN, and Naive Bayes, we are led to over 90% in accuracy as a result. One finding from this experiment is that utilizing hashtags and other conventional markers on Twitter messages does indeed prove to be useful for sentiment analysis. Further work is intended where we may integrate other components into the background information such as, sleep data, physical activity levels, and food information in hopes of discovering more unique and informative insights.

One important pattern of behavior we must pay mind to when studying tendencies of suicidal thoughts is the stress level of a person. Lin et al. [3] ideate a comprehensive scheme to measure a users stress level from their social media habits. We learn from the authors here that while stress levels are feasible to detect on social media, so far there has been little attention given in the social network analysis field towards examining exact stressors and stress levels of a user on social media. When we dig a little deeper, we may find that this lack of stress-related analysis stems from challenges related to stressor subject identification, stressor event detection, and data collection and representation. With this in mind, we build a system in which a benchmark data set is first created such that we are able to withdraw and concentrate on strong stress-related features. Once these features are decided upon, we will build a hybrid model combining multi-task learning with convolutional neural network (CNN) where a stressor event along with it's subject can be discovered on social media posts. It is imperative that this model is capable of representing the relatedness of stressor events and their respective stressor subjects. Lastly, we resort to an expert defined look-up table where detected subjects and their events are stored such that an estimation process is allowed for the stressor subject and stress level and a proper detection of stress may be determined. For future work, the plan is to consider personal stress coping abilities which may vary per individual, however will aid in fine-tuning measurements of stress levels and also enable for more personalized care to be extended upon users.

III. Data

The data used as the basis for analysis and understanding of the problem statement was initially obtained through Datasift, for the use within an existing system built by Homan et al. [4]. The processing allows for all Twitter data spanning across a 16-county region centered around a mid-sized Northeastern United States city to be acquired. The time period of the tweets gathered date from June 30, 2013 to July 1, 2014.

$$\textit{Number of Users} = 86{,}137$$
$$\textit{Number of Tweets} = 6{,}793{,}552$$

Table 1. Statistics of user and tweet numbers for data [4]

This data was initially utilized in the creation of a Natural Helping model. In this case, the main interest in utilizing the data was to identify where networks of help propagate, such that we may be able to detect central helpers. The importance of such an analysis corresponds to the understanding of how strongly informal helping is present within communities (i.e. friends and family support), as well as examining sustainability within a said community. Another nature of this natural helping analysis allows for a clearer perception on whether a given society is strong enough to combat anti-social influences, such as economic hindrances or violent behaviors.

The model implemented in order to perform the helping analysis utilized that of language-use graphs. This implies the construction of a graph where each user is represented as a node and each form of communication, in this case the tweet content, is an edge. It is of interest to build the graph such that each edge is representative of helping and positive nature. Detection of whether communication between users was uplifting and of helping fashion was determined based off of a manual set of words. The vocabulary fed into the model mainly consists of [thank, thanks, thank you, ty]. The intuition behind the set of word identifiers stems from the fact that in social interactions amongst humans, when one person helps another the usual response from the person who is being helped comprises words of gratitude.

From this, a variety of sub-graphs were generated. Firstly based from the construction of one comprehensive graph analyzing the entirety of the raw data, then scaling down to sub-graphs which represent that of direct mentions towards users who exude help (labeled through the '@' symbol as used in Twitter), reciprocal helping graphs; where the behavior of helping is mutual amongst users, and central helper graphs; where one user who is connected strongly to multiple users is deemed as a central source of help (i.e. two or more edges).

Given this data and existing analysis performed under the objective of detecting distributions of help within a network, we are able to translate this model into the study of suicidal implications within the network as well. The Natural Helping model may serve as a strong reference point in the building of a death-related system to better discern victims of death as well as study the

communication and behaviors of the informal support system that surround them.

IV. Methods

In analysis upon the given problem statement of detecting victims of death, a variety of experiments were performed upon the 7 million tweets dataset which serves as the basis of the model. These experiments included that of processing, cleaning, parsing, and manipulating the data itself. The core design of the model involves analyzing friends of people who have died, such that they are labeled the sympathizers and are part of an informal support system. If analysis is done upon people who are offering sympathy or condolences to a specific person, it is most likely that, that person has died and further examination into the language and tone pattern through their 280 character limit can be performed more discreetly on whether suicidal ideation was implied before death.

A. *Corpus Creation*

In preparation for analysis, a baseline of keywords were constructed in order to adhere to during the parsing and processing through the data. Much similar to the vocabulary manually constructed in the Natural Helping analysis, our Death Word Categorical vocabulary consisted of the following keywords:

* **HEAVEN**
* **LOSS**
* **MISSED**
* **PRAYERS**
* **RIP**

Table 2. Keywords of Death Word Categorical Vocabulary

As the entirety of the raw data is parsed through in correspondence to these listed keywords, we construct a table which maps each keyword to the respective Twitter username and the text contained in each tweet where the keyword was initially detected.

Username	Keyword	Text
@jayhorrr	Heaven	...one time I took a heaven...
@19CRG	Loss	...was a loss for the team...
@JasonGianotti	Missed	Missed my chance earlier...
@johnnyScarpelli	Prayers	Praying for you and your entire family...
@cris-Munna	RIP	Rip baby boy i lovee you...

Table 3. Preliminary Corpus structure

B. *Mutual Friends Detection*

After gaining a sense of all the users who utilized keywords related to the Death Word Vocabulary, we are interested in retrieving more information on these now labeled sympathizers. With this, we now examine and study more of who are these sympathizers mentioning in their tweets, and do their mentions match the mentions of any other user. If so, the other user and themselves are linked to the mutual mention, hence also becoming mutual friends of a possible victim (the original mutual mention.) As this is also imperative information to construct in order to examine first-hand, we create another tabular form with this data contained.

The aim of mutual friends detection is to see the language and tone used by sympathizers when a person has died. In a sense, we are confirming on whether the initial manually created Death Word Categorical vocabulary containing the five keywords holds true to it's purpose, in detecting if someone has died.

Username	Keyword	Mutual Friends
@DropDeadDiva11	Heaven	[@tataxoxo]
@JakeeClark2	Loss	[@Sweezy-F-Baby,@nicknugget69,@doublemm5]
@Z-kess1	Missed	[@JohnCarr-23,@TPantoja24,@Rs668000]
@DanLohrs	Prayers	[@HeyyLana,@DeeJayJona,@1in-emilyon]
@Mackomore	RIP	[@crattelade,@PaigeSynesael]

Table 4. Mutual Friends Detection

With this mutual friends corpus created, we are now able to better examine and identify users who constitute as central death proponents between a linkage of sympathizers.

C. *Corpus Annotation and Analysis*

In this step, we are to manually go through the table generated from automating the extraction of all data containing the keywords (Table 3) as well as confirming through manual annotations on whether the tweet was indeed implying a death or not. Simultaneously, we are interested in any linkage between multiple users and a possible central person of death, as we can determine through reference of Table 4.

D. *SVM Analysis*

In the last step of implementation for the model, we ran the annotated data through a classification algorithm. The keyword set which was manually created (features) as well as annotated results (labels) were fed into a support vector machine in order to analyze the accuracy of our system. The accuracy achieved on our data using this machine learning model was 91.6%

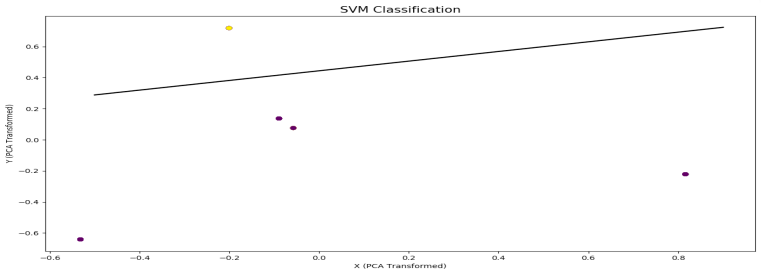

Fig. 1: Visualization of SVM Hyperplane classification

V. Results

1. ***Keyword Frequency:***

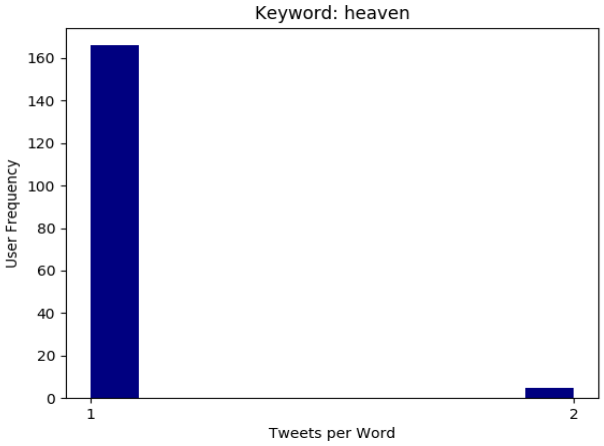

Fig. 2: Histogram plotted to observe frequency of keyword ***'heaven'***

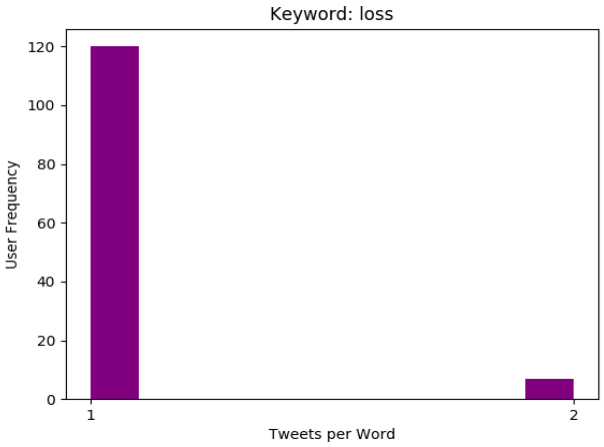

Fig. 3: Histogram plotted to observe frequency of keyword ***'loss'***

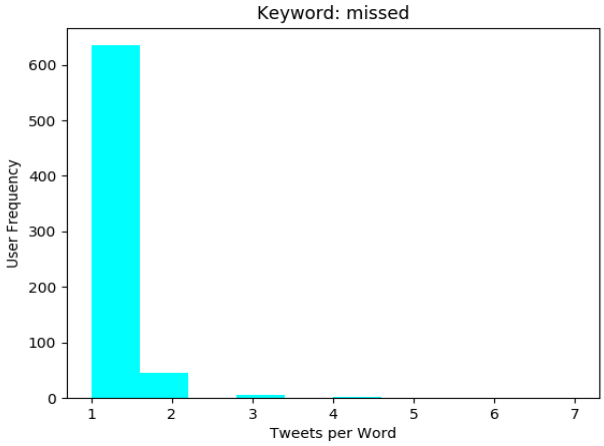

Fig. 4: Histogram plotted to observe frequency of keyword ***'missed'***

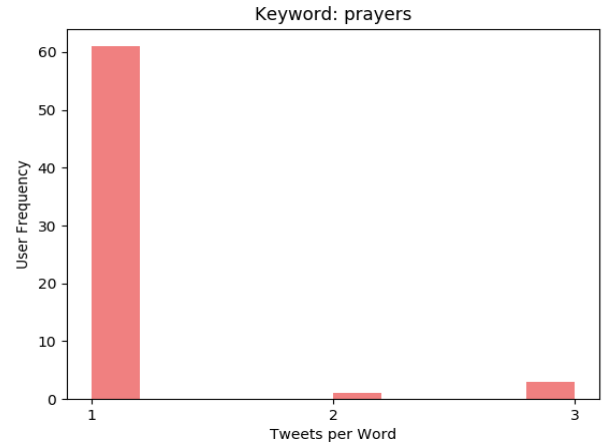

Fig. 5: Histogram plotted to observe frequency of keyword ***'prayers'***

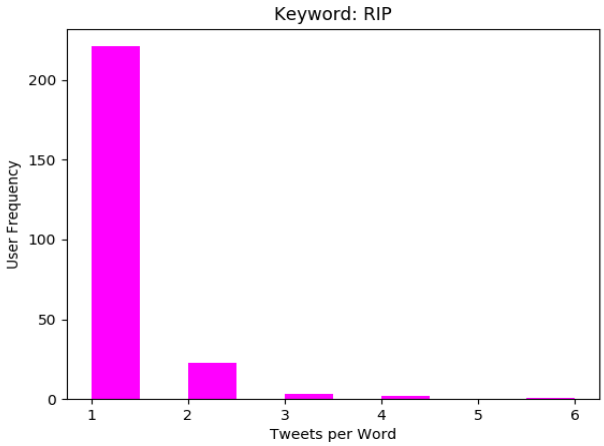

Fig. 6: Histogram plotted to observe frequency of keyword ***'RIP'***

2. ***Corpus Annotations:***

After careful manual analysis of the created corpus, we may note a few observations that were recorded during the process. Firstly, the corpus created contained 1,425 tweets related to death, which is all recorded in a manner similar to Table 3. Then, we examine over the tweet text and label whether the given tweet is indeed death related or not. The purpose of doing so is such that, many people who use death-related terms may be using them loosely and in everyday conversational manners. After this careful inspection of tweets, key takeaways consist of how many tweets talk of an actual death, what other main subjects do the keywords correlate with, and an example of such a non-death popular related topic are noted, per each keyword:

Keyword	% Death Related in Corpus	Non-death related topic	Example
Heaven	2%	Food	"This coffee is heaven"
Loss	10%	Games/Sports	"Miami and Florida state will both end up with a loss leaving UO and bama left"
Missed	8%	Missing someone who is still alive	"@Brt_ergusonmanuelmissedgoatoo!"
largePrayers	21%	Well wishes to people	" My heart and prayers go out to those still fighting nationalcancerday"
RIP	11%	Joking/Sarcasm	"if you jinx it I will kill you...then RIP to you"

Table 5. Annotation Observations

3. ***Mutual Friends Inspections:***

After manually processing through and labeling each of the 1,425 existing rows within the corpus, we now focus on the messages which are labeled "yes", as in death related, to further investigate upon mutual friends of the user whose tweet was labeled in agreement of death. By this, the results achieved for the 55 out of 1,425 "yes" labels are deaths categorized in terms of:

- Unsure on whether death can be inferred and how death may have occurred (if suicide may be implied or not)
- Death of family member
- Personal death
- Celebrity death
- Death of inanimate object
- Death of animal

We have now completed investigation upon possible reasons of a death related tweet.

VI. Conclusion

In all, by the several iterative experiments performed upon the research of detecting victims of death through sympathizers, we were able to successfully construct a corpus based model which is helpful for further investigation. Therefore, a baseline approach is built such that we are allowed information on what type of language is most frequently used by catalysts of an informal support system as well as the language and words chosen to display sympathy to someone who has most likely passed.

VII. Future Work

As a foundation of analysis has been built through the completion of this project, further projection is imperative and will help to achieve more insightful results. Some additional directions to take on from here are:

- Acquire data from different social media platforms (i.e. Reddit, Instagram, Facebook, etc.)
- Further analysis on central death person and delve into tone or behavior patterns they use themselves through their tweets
- Further experimentation using various Machine Learning models (i.e. Decision Trees, Regression Analysis, etc.) for additional classification techniques

Acknowledgment

I would like to thank Dr. Christopher Homan for being a very helpful source of guidance during the completion of this project.

References

[1] K. Roberts, M. A. Roach, J. Johnson, J. Guthrie, and S. M. Harabagiu, "Empatweet: Annotating and detecting emotions on twitter." in *Lrec*, vol. 12. Citeseer, 2012, pp. 3806–3813.

[2] M. Hasan, E. Rundensteiner, and E. Agu, "Emotex: Detecting emotions in twitter messages," 2014.

[3] H. Lin, J. Jia, L. Nie, G. Shen, and T.-S. Chua, "What does social media say about your stress?." in *IJCAI*, 2016, pp. 3775–3781.

[4] C. M. Homan, R. Shah, S. Khasbag, A. Ptah, S. Kaori-Mei, Y. He, M. Funchess, and A. M. White, "Discovering natural helping in social media through language, network structure, and community knowledge," 2020.

[5] "Suicidedata, howpublished = https://www.who.int/mental$_h$*ealth/preventio* *Accessed* : 2020 − 01 − 30.″

[6] S. Rice, J. Robinson, S. Bendall, S. Hetrick, G. Cox, E. Bailey, J. Gleeson, and M. Alvarez-Jimenez, "Online and social media suicide prevention interventions for young people: a focus on implementation and moderation," *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, vol. 25, no. 2, p. 80, 2016.

[7] J. Robinson, M. Rodrigues, S. Fisher, E. Bailey, and H. Herrman, "Social media and suicide prevention: findings from a stakeholder survey," *Shanghai archives of psychiatry*, vol. 27, no. 1, p. 27, 2015.

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

[9] Y. Wang and A. Pal, "Detecting emotions in social media: A constrained optimization approach," in *Twenty-fourth international joint conference on artificial intelligence*, 2015.

[10] S. C. Guntuku, A. Buffone, K. Jaidka, J. C. Eichstaedt, and L. H. Ungar, "Understanding and measuring psychological stress using social media," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 13, no. 01, 2019, pp. 214–225.

[11] H. Almeida, A. Briand, and M.-J. Meurs, "Detecting early risk of depression from social media user-generated content." in *CLEF (Working Notes)*, 2017.

[12] I. Hemalatha, G. S. Varma, and A. Govardhan, "Sentiment analysis tool using machine learning algorithms," *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, vol. 2, no. 2, pp. 105–109, 2013.

[13] J. Robinson, G. Cox, E. Bailey, S. Hetrick, M. Rodrigues, S. Fisher, and H. Herrman, "Social media and suicide prevention: a systematic review," *Early intervention in psychiatry*, vol. 10, no. 2, pp. 103–121, 2016.