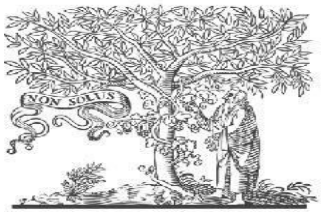
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

IJIEMR Transactions, online available on 2nd Jan 2021. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-09&issue=ISSUE-12

DOI: 10.48047/IJIEMR/V09/I12/161

Title: DETECTION OF SUICIDE-RELATED POSTS IN TWITTER DATA STREAMS
Volume 09, Issue 12, Pages: 941-948
Paper Authors

K. SAHITHI, SHETTY RENUKA, NEELAM UMA RANI, E.LAXMAN

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

DETECTION OF SUICIDE-RELATED POSTS IN TWITTER DATA STREAMS

K. SAHITHI[1], SHETTY RENUKA[2], NEELAM UMA RANI[3], E.LAXMAN[4]

[1,2,3,] B TECH Students, Department of CSE, Princeton Institute of Engineering & Technology For Women, Hyderabad, Telangana, India.

[4] Assistant Professor, Department of CSE, Princeton Institute of Engineering & Technology For Women, Hyderabad, Telangana, India.

Abstract: Suicidal idea identification in online interpersonal organizations is an arising research zone with significant difficulties. Ongoing examination has demonstrated that the freely accessible data, spread across web-based media stages, holds important pointers for successfully identifying people with self-destructive aims. The critical test of self destruction anticipation is understanding and identifying the intricate danger factors and cautioning signs that may hasten the occasion. In this paper, we present another methodology that utilizes the web-based media stage Twitter to measure self destruction cautioning finishes paperwork for people and to recognize posts containing self destruction related substance. The primary innovation of this methodology is the programmed ID of unexpected changes in a client's online conduct. To recognize such changes, we consolidate common language preparing procedures to total social and literary highlights and pass these highlights through a martingale system, which is broadly utilized for change recognition in information streams. Investigations show that our content scoring approach viably catches cautioning signs in content contrasted with conventional AI classifiers. Moreover, the utilization of the martingale structure features changes in online conduct and shows guarantee for distinguishing social changes in danger people.

I. Introduction

According to the World Health Organization (WHO), it is estimated that 800,000 people worldwide die by suicide each year with at least as many suicide attempts [1]. The grief felt in the aftermath of such an event is compounded by the fact that a suicide may be prevented. This reality of suicide has motivated WHO member states to commit themselves to reducing the rate of suicide by a significant percent by 2020 [2]. In an effort to educate the public, the American Foundation for Suicide Prevention (AFSP) [3] has identified characteristics or conditions that may increase an individual's risk. The three major risk factors are: 1) health factors (e.g., mental health, chronic pain), 2) environmental factors (e.g., harassment, and stressful life events), and 3) historical factors (e.g., previous suicide attempts and family history). Additionally, the time period preceding a suicide can hold clues to an individual's struggle. The AFSP categorizes these warning signs as follows: 1) talk (e.g., mentioning being a burden or having no reason to live), 2) behavior (e.g., withdrawing from activities or sleeping too much or too little), and 3) mood (e.g., depression or rage). Identifying these risk factors is the first step in suicide prevention. However, the social stigma surrounding mental illnesses means that at-risk individuals may avoid professional assistance [4]. In fact, they

may be more willing to turn to less formal resources for support [5]. Recently, online social media networks have become one such informal resource. Research has shown that at-risk individuals are turning to contemporary technologies (forums or micro-blogs) to express their deepest struggles without having to face someone directly [6, 7]. As a result, suicide risk factors and warning signs have been seen in a new arena. There are even instances of suicide victims writing their final thoughts on Twitter, Facebook, and other online communities [8, 9]. We believe that this large amount of data on people's feelings and behaviors can be used successfully for early detection of behavioral changes in at-risk individuals and may even help prevent deaths. Social computing research has focused on this topic in recent years. However, few initiatives have been concerned with the real-time detection of suicidal ideation on Twitter. Previously proposed detection methods rely heavily on manually annotated speech, which can limit their effectiveness due in part to the varying forms of suicide warning signs in at-risk individuals. Many of these methods also focus on the messages published by individuals at a specific time, independent of the whole context, which may be represented by the sequence of publications over time. In this paper, we address the challenge of real-time analysis of Twitter posts and the detection of suicide-related behavior. To process the stream of an individual's online content, we implement a martingale framework, which is widely used for the detection of changes in data stream settings. The input into this framework is a series of behavioral features computed from each individual Twitter post (tweet). These features are compared to previously seen behavior, in order to detect a sudden change in emotion that may indicate an elevated risk of suicide. The main contributions of this paper are twofold. First, using research from the field of psychology, we design and develop behavioral features to quantify the level of risk for an individual according to his online behavior on Twitter (speech, diurnal activities, size of social network, etc.). In particular, we create a feature for text analysis called the Suicide Prevention Assistant (SPA) text score. Second, we monitor the stream of an individual Twitter user and his behavioral features using an innovative application of a martingale framework to detect sudden behavioral changes.

II. Related work

The definition and identification of risk factors and warning signs lie at the core of suicide prevention efforts. In this paper, we have chosen to reference the risk factors defined by the American Psychiatric Association (APA) [13] and the warning signs identified by the American Association of Suicidology (AAS) [14]. These resources represent a level of consensus between mental health professionals and also provide a rich discussion of the differences between suicide risk factors and warning signs. For further reading, we direct the reader to the work of [14]. As highlighted by [14], warning signs signify increased imminent risk for suicide (i.e., within minutes, hours, or days). According to the APA, suicide warning signs may include talking about dying, significant recent loss (death, divorce, separation, or broken relationship), change in personality, fear of losing control, suicide plan, suicidal thoughts, or no hope for the future. As discussed in the following paragraphs, recent research has shown the emergence of such signs on social networking sites. Most of the research

at the intersection of behavioral health disorders and social media has focused on depression detection in online communities, specifically Major Depressive Episodes (MDE). However, the risk factors for suicide defined by the APA [13] go far beyond depression alone. It is important to remember that depression does not necessarily imply suicidal ideation. Rather, suicide should be thought of as a potential end symptom of depression. While mental health issues such as depression, suicidal ideation, and self-mutilation are defined medically as separate illnesses with overlapping symptoms, the approaches proposed to detect them online can be quite similar. The approaches vary in the data they are treating, i.e., Facebook posts, Twitter tweets, Reddit forum threads, etc., and the specific event they are attempting to predict. Moreno et al. [7] first demonstrated that social networking sites could be a potential avenue for identifying students suffering from depression. The prevalence rates found for depression disclosed on Facebook corresponded to previous works in which such information was self-reported. On a larger scale, Jashinsky et al. [15] showed correlation between Twitter-derived and actual United States per-state suicide data. Together, these works established the presence of depression disclosure in online communities and opened up a new avenue for mental health research. De Choudhury et al. [6] explored the potential to use social media to detect and predict major depressive episodes in Twitter users. Using crowd-sourcing techniques, the authors built a cohort of Twitter users scoring high for depression on the CES-D (Center for Epidemiologic Studies Depression Scale) scale and for other users scoring low. Studying these two classes, they found that what is known from traditional literature on depressive behavior also translates to social media. For example, users with a high CES-D score posted more frequently late at night, interacted less with their online friends, and had a higher use of first-person pronouns. Additionally, online linguistic patterns match previous findings regarding language use of depressed individuals [16]. More recently, De Choudhury et al. [10] have shown that linguistic features are important predictors in identifying individuals transitioning from mental discourse on social media to suicidal ideation. The authors showed a number of markers characterizing these shifts, including social engagement, manifestation of hopelessness, anxiety, and impulsiveness based on a small subset of Reddit posts. Coppersmith et al. [17] examined the data published by Twitter users prior to a suicide attempt and provided an empirical analysis of the language and emotions expressed around their attempt. One of the interesting results found in this study is the increase in the percentage of tweets expressing sadness in the weeks prior to a suicide attempt, which is then followed by a noticeable increase in anger and sadness emotions the week following a suicide attempt. In the same line of research, O'Dea et al. [18] confirmed that Twitter is used by individuals to express suicidality and demonstrated that it is possible to distinguish the level of concern among suicide-related tweets, using both human coders and an automatic machine classifier. These insights have also been investigated by Braithwaite et al. [19], who demonstrated that machine learning algorithms are efficient in differentiating people who are at a suicidal risk from those who are not. For a more detailed review of the use of social media platforms as a tool for suicide prevention, the reader may refer to the recent systematic survey by Robinson et

al. [20]. These works have shown that individuals disclose their depression and other struggles to online communities, which indicates that social media networks can be used as a new arena for studying mental health. Despite the solid foundation, the current literature is missing potential key factors in the effort to detect depression and predict suicide. Currently, few works analyze the evolution of an individual's online behavior. Rather, the analysis is static and may take into consideration one post or tweet at a time while ignoring the whole context. Additionally, an individual's online "speech" is often compared to other individuals and not to their own linguistic style. This is a disadvantage because two individuals suffering the same severity of depression may express themselves very differently online.

III. Methodology

A general framework for detecting suicide-related posts in social networks: Here we present the proposed framework for the analysis and real-time detection of suicide-related posts on Twitter. First, we introduce the real-time detection problem. Then, we define our online proxy measurements (behavior features) for suicide warning signs. Finally, we describe the approach we implement for detecting behavioral change points. Problem statement Sudden behavioral change is one of the most important suicide warning signs. As reported by the AFSP, a person's suicide risk is greater if a behavior is new or has increased, especially if it is related to a painful event, loss, or change. Considering this in conjunction with social media, where users constantly publish messages and deliberately express their feelings, we address suicide warning sign detection as a real-time data stream mining problem. Given a series of observations over time (tweets, messages, or blog posts), the task is to detect an abrupt change in a user behavior that may be considered as a suicide warning sign. In the field of data stream mining, this can be specifically seen as change point detection problem. However, unlike retrospective detection settings which focus on batch processing, here we are interested in the setting where the data arrives as a stream in real time. To address this challenge, we chose an approach employing a martingale framework for change point detection. This algorithm has been successfully applied to detecting changes in unlabeled data streams, video-shot change detection and, more recently, in the detection of news events in social networks. To the best of our knowledge, this is the first attempt to apply the martingale framework on a multidimensional data stream generated by Twitter users. In the following section, we start by introducing and describing the proxy measurements for suicide warning signs that we use to assess the subject's level of suicide risk. As previously mentioned, these warning signs will be the input into the martingale framework. Suicide warning signs in online behavior To identify online behaviors that may reflect the mental state of a Twitter user, we established two groups of behavioral features: user-centric and post-centric features. User-centric features characterize the behavior of the user in the Twitter community, while post-centric features are characteristics that are extracted from the properties of a tweet. These features have been shown to successfully aid in determining the mental health of a user [6]. Table 1 shows a detailed description of the features we selected. The AAS identifies withdrawing from friends, family, or society as one of the warning signs of suicide. With the

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

user-centric behavioral features, we aim to capture changes in a Twitter user's engagement with other users. The friends and followers features can quantify an individual's interaction with his or her online community, such as a sudden decrease in communication. On the other hand, they can also reflect an expansion of an individual's online community. This is relevant, as at-risk individuals have also been shown to increase their time online developing personal relationships. It is important to note that we have chosen the terms friends and followers to represent the unidirectional relationships that are inherent on Twitter. We acknowledge that this term may not apply for certain user accounts such as celebrities and news outlets. Additional features include volume, replies, retweets, and links, which were all identified by De Choudhury et al. [6] as markers for mental health. These measures can help to quantify the number of interactions a user has with their friends and followers for it could be the case that an individual's social network remains stable while their interactions increase or decrease. The final user-centric feature, questions, may also indicate a user's attempt to engage with others online. Post-centric behavioral features are characteristics originating from the post itself. One important piece of information is the hour at which the tweet is published (time feature). Late-night activity can be an indication of unusual rhythms in sleep (insomnia and hypersomnia) [6] and can predict future episodes of depression. In addition to the time feature, we address the text of the post (text score), which holds the most vital information pertaining to an individual's current mood and mental health.

Feature	*Definition*
User-centric features	
Friends	The total amount of friends at the time of post. A friend is defined as another Twitter user that the author is following online (out-link).
Followers	The total amount of followers at the time of the post. A follower is another Twitter user that is following the author online (in-link).
Volume	The number of tweets per hour; retweets are included in this feature.
Replies	The number of tweets per day directed at another Twitter user. An author may post a tweet destined for another user by including "@" plus the user's screen name.
Retweets	The number of tweets per day that are retweeted. A retweet is defined as a tweet previously composed by another Twitter user and that is re-published or shared.
Links	The number of tweets per day that include a URL
Questions	The number of tweets per day posing a question.
Post-centric features	
Time	The hour at which the tweet was published.
Text score	Text score (based on NLP)/distress classifier

To classify the text of the post, we propose two different approaches. The first approach is a natural language processing (NLP) method that combines features generated from the text, based on an ensemble of lexicons. These lexicons are composed of linguistic themes commonly exhibited by at-risk individuals. The second approach, called the distress classifier, is based on machine learning. Although machine learning is commonly used to classify text, the supervised algorithms require annotated datasets, which may be costly in terms of time and potential annotator error. Additionally, traditional machine learning methods are difficult to apply in this context because of the nature of depression and distress in general. Two individuals suffering from depression may not express their symptoms in the same way, which translates to texts exhibiting the same level of depression or distress having vastly different content. This means it is difficult for the algorithm to find the concept mapping between the textual features and the level of depression/distress.

Feature extraction for text scoring: Classification with machine learning In general, the challenge of categorizing tweets is traditionally addressed using text classifiers that rely on machine learning. Therefore, we considered it important to also test a classifying algorithm as a benchmark against our NLP approach detailed above. We took inspiration

from and chose to categorize tweets according to different levels of distress. Again, although distress is not equivalent to suicide ideation and major depressive episodes, it is an important risk factor in suicide and one that is highly observable from micro-blog text. In addition to the four features used in the previous section (fsymptoms; fswear; fintensifiers, and ffirst pronouns), we split the text into n-grams, which are commonly used in text classification tasks and are popular as a base feature for sentiment analysis of tweets. The character limitation (140 maximum) of tweets lends itself to a choice of shorter n-grams, particularly unigrams and bi-grams. Design of a martingale-based approach for emotion change detection in the previous sections, we described the user-centric and post-centric behavioral features we extract from a user's online content. In this section, we present the approach we use to process these behavioral features and detect sudden emotional changes. If we consider the features as a series of multi-dimensional observations over time, the challenge of detecting an abrupt change in behavior resembles the classic problem of change point detection often seen in the field of data stream mining. We implemented a martingale framework, which is an online real-time, non-parametric change point detection model. Let X ¼ fx1; x2; x3 ... xng be a sequence of unlabeled m-dimensional data points with new data points, xi, arriving in a sequence. The m dimensions correspond to the values of the user-centric and post-centric behavioral features identified in the previous section. When a new tweet is published, it is first characterized by this set of features. With this information, the tweet is then run through a hypothesis test to determine if its features represent a prominent change in the data stream. More formally, the test is as follows: H0, if there is no change in the data stream (i.e., no marked emotional change), and H1 otherwise. The full martingale framework can be broken down into three steps. The first step is to calculate the strangeness measure, which quantifies for each specific user how much a tweet is different from previous ones. Next, a statistic is defined to rank the strangeness measures of the tweets. Finally, using this statistic, a family of martingales is defined in order to detect movements in the tweet stream and run the hypothesis test.

IV. Conclusion

In this paper, we designed and evaluated a novel approach to monitor the mental health of a user on Twitter. Building off existing research, we worked to translate and quantify suicide warning signs in an online context (user-centric and post-centric behavioral features). In particular, we focused on detecting distress-related and suicide-related content and developed two approaches to score a tweet: an NLP-based approach and a more traditional machine learning text classifier. To detect changes in emotional well-being, we considered a Twitter user's activity as a stream of observations and applied a martingale framework to detect change points within that stream. Our experiments show that our NLP text-scoring approach successfully separates out tweets exhibiting distress-related content and acts as a powerful input into the martingale framework. While the martingale values "react" to changes in online speech, the change point detection method needs improvement. We were able to detect the true change point for one validation case, but the approach needs to be more robust with respect to parameter setting and positive changes in speech. For future

research, we plan to further explore the impact of martingale parameters on the change detection effectiveness. We also hope to expand the approach to include image processing and other social media outlets in order to assess the effectiveness in other settings. Another interesting perspective is to consider more fine-grained emotion classes such as anger, sadness, fear, etc., instead of considering four levels of distress. However, overall, we believe our initial work presents an innovative approach to detecting suicide-related content in a text stream setting.

V. References

1. "Preventing suicide: A global imperative," World Health Organization, Geneva, Switzerland, 2014. [Online]. Available: http://www.who.int/mental_health/suicide-prevention/ world_report_2014/en

2. "Mental health action plan 2013–2020," World Health Organization, Geneva, Switzerland, 2013. [Online]. Available: http://www.who.int/mental_health/publications/action_plan/en

3. American Foundation for Suicide Prevention (AFSP). [Online]. Available: https://afsp.org

4. P. Corrigan, "How stigma interferes with mental health care," Amer. Psychologist, vol. 59, no. 7, pp. 614–525, 2004.

5. D. Rickwood, F. P. Deane, C. J. Wilson, et al., "Young people's help seeking for mental health problems," Aust. e-J. Adv. Mental Health, vol. 4, no. 3, pp. 218–251, 2005.

6. M. De Choudhury, M. Gamon, S. Counts, et al., "Predicting depression via social media," in Proc. 7th Int. AAAI Conf. Weblogs Social Media, Boston, MA, USA, 2013, pp. 128–137.

7. M. Moreno, L. Jelenchick, K. Egan, et al., "Feeling bad on Facebook: depression disclosures by college students on a social networking site," Depression Anxiety, vol. 28, no. 6, pp. 447–455, 2011.

8. J. F. Gunn and D. Lester, "Twitter postings and suicide: An analysis of the postings of a fatal suicide in the 24 hours prior to death," Suicidologi, vol. 17, no. 3, pp. 28–30, 2012.

9. V. Kailasam and E. Samuels, "Can social media help mental health practitioners prevent suicides?" Current Psychiatry, vol. 14, no. 2, pp. 37–51, 2015.

10. M. De Choudhury, E. Kiciman, M. Dredze, et al., "Discovering shifts to suicidal ideation from mental health content in social media," in Proc. 2016 CHI Conf. Human Factors Comput. Syst., San Jose, CA, USA, 2016, pp. 2098–2110.

11. M. De Choudhury, S. Counts, and E. Horvitz, "Predicting postpartum changes in emotion and behavior via social media," in Proc. SIGCHI Conf. Human Factors Comput. Syst., New York, NY, USA, 2013, pp. 3267–3276.

12. M. T. Lehrman, C. O. Alm, and R. A. Proano, "Detecting distressed and non-distressed affect states in short forum texts," in Proc. 2012 Workshop Lang. Social Media, Montreal, QC, Canada, 2012, pp. 9–18.

13. American Psychiatric Association, "Practice guideline for the assessment and treatment of patients with suicidal behaviors," Amer. J. Psychiatry, vol. 160, no. 11, pp. 1–60, 2003.

14. M. D. Rudd, A. L. Berman, T. E. Joiner, et al., "Warning signs for suicide: Theory, research, and clinical applications," Suicide LifeThreatening Behav., vol. 36, no. 3, pp. 255–262, 2006.

15. J. Jashinsky, S. H. Burton, C. L. Hanson, et al., "Tracking suicide risk factors through Twitter in the US," J. Crisis Intervention Suicide Prevention, vol. 35, no. 1, pp. 51–59, 2014.

16. S. S. Rude, E. M. Gortner, and J. W. Pennebaker, "Language use of depressed and depression-vulnerable college students," Cogn. Emotion, vol. 18, no. 8, pp. 1121–1133, 2004.

17. G. Coppersmith, et al.,"Exploratory analysis of social media prior to a suicide attempt," in Proc. 3rd Workshop Comput. Linguistics Clinical Psychol., 2016, pp. 106–117.

18. B. O'Dea, S. Wan, P. J. Batterham, et al., "Detecting suicidality on twitter," Internet Interventions, vol. 2, no. 2, pp. 183–188, 2015.

19. S. R. Braithwaite, C. Giraud-Carrier, J. West, et al., "Validating machine learning algorithms for twitter data against established measures of suicidality," JMIR Mental Health, vol. 3, no. 2, 2015, Art. no. e21.

20. J. Robinson, G. Cox, E. Bailey, et al., "Social media and suicide prevention: A systematic review," Early Intervention Psychiatry, vol. 10, no. 2, pp. 103–121, 2016.