Social Media and Machine Learning in Suicide Prevention

Kelly Soberay, MA & Nora Mund, BA
For the Military Suicide Research Consortium
March 15 2017

Statement of the Problem

Suicide is a leading cause of death in military personnel (WHO, 2016). Current practices in predicting suicide attempts are limited in timing and accuracy (Bentley et al., 2016; Chang et al., 2016; Franklin et al., 2016; Ribeiro et al., 2016). Identifying individuals at risk for suicide with the use of machine learning in social media posts and medical databases are new approaches to suicide prevention.

Summary of the relevant literature

Social media has become a platform for individuals to express suicidal thoughts, behaviors, and intent (Ahuja et al., 2014; O'Dea et al., 2017*)*.Identifying individual users of social media who may be at risk for suicide using human coding (O'Dea et al., 2015; Mowery et al., 2017) and machine learning (Abboute et al., 2014; Varathan & Talib, 2014) is possible. However, the application is not yet practical (Christensen, Batterham, & O'Dea, 2014), ethical (Lavot, Ben-Zeev, & Neville, 2012), acceptable (Orme, 2014), or necessarily warranted (McGee et al., 2013) for outreach purposes related to social media posts.

The use of machine learning (ML) is relatively new in the field of clinical psychology, specifically in its use to predict suicide risk. ML implements algorithms to classify complex problems. Recent meta-analyses demonstrated that the ability to predict suicide attempts requires a complex combination of hundreds of risk factors which ML may be better suited to analyze over traditional techniques (Franklin et al., 2016). ML studies support that this approach is promising in providing discriminative accuracies for suicide attempters (Delgado-Gomez et al., 2012; Mann, Ellis, Waternaux, & Liu, 2008). Research in progress attempts to explore ML longitudinally and with larger number of suicide attempters to explore the utility and accuracy of suicide prediction.

Another approach to predicting increased rates of suicidal posts on social media is the use of an autoregressive integrated moving average (ARIMA) data analysis, to identify and predict trends in communication about depression and suicide (McClellan et al., 2016). Timely dissemination of public health messages could help agencies influence health behaviors when these trends are identified. In turn this could facilitate recognizing and responding to anticipated and unanticipated events that may influence an increase of social media postings on suicidal thoughts, behaviors, and intent.

Twitter has received extensive attention on analyzing social media content. This is likely due to the large amount of postings which are all available publicly; other social media can be protected by privacy settings. Studies involving Twitter use automatic and coded linguistic analyses of suicide-related Twitter posts. When comparing matched non-suicide related Twitter posts to suicide-related posts, strong linguistic suicide posts were differentiated by higher word counts, increased use of self-referencing, greater anger, and increased focus on the present related to time of death (Baddeley, Daniel, & Pennebaker, 2011; Fernandex-Cabana et al., 2012; O'Dea et al., 2017). However, studies have also found that positive emotions near the time of death were linked to death by suicide (Lester, 2009; Lester & McSwain, 2011). Increased use of content related to self-referencing may be related to a sense of hopelessness and social withdrawal (O'Dea et al., 2017).

A study of 2.73×10^9 tweets in a 12 week period investigating emotion mapping and mental health outcomes found that rage and suicide rates were significantly correlated (Larsen et al., 2015). The availability of emotion mapping to detect changes can provide information regarding burden and fluctuation in suicidal thoughts and behaviors. This may be useful in the identification of suicide related tweets (Larsen et al., 2015). Research supports that there is a virtual community within social media outlets that can have a level of social contagion; individuals with higher risk for suicide-related behaviors have a high degree of connectivity with other authors of suicide content (Colombo, Burnap, Hodorog, & Scourfield, 2016). The risk of contagion through social networking sites is especially high for adolescents and younger populations.

There is case evidence that shows Facebook postings were linked to self-harm, substance misuse, and suicidal intent (Barrett et al., 2016). Recently, Facebook announced that algorithms are being tested to identify content related to suicide or self-injury for prevention efforts. Groups and prevention organizations are already using Facebook for outreach purposes, however this is only on a scale of individuals that are a part of the group forum or that reach out personally. No research has been conducted on this prevention method.

Gaps in the literature

There are several different venues for social media, none of which are standardized. Research is emerging in other social networking outlets such as Tumbler (Cavazos-Rehg et al., 2016) and responses to YouTube videos (Lewis et al., 2012) where individuals may express suicidal behaviors and intent to die by suicide. Currently this literature is limited in its coding for themes and has yet to include a model of prediction of suicide or near-death attempts. Identification is not accurate enough to create or model an intervention from social media platforms.

New social media platforms, such as Snapchat and Instagram, where an individual can post pictures of suicide-related thoughts and behaviors, are becoming increasingly common (Chhabra & Bryant, 2016). Some of the media platforms will evolve and combine multiple platforms like microblogging in China (Weibo). The growth of discussion forums and the cross-cultural influence of globalized social media could account for the difficulties in generalization. Individuals are also known to use multiple platforms and handles. Other countries are also creating ML algorithms and classification models within their social media sites (Guan et al., 2015; Won et al., 2013).

Further development of AI/ML models is needed to determine not only the feasibility of utilizing such models for risk detection, but whether the methods are valid and reliable (O'Dea et al., 2017). There are some ethical concerns regarding whether or not it is acceptable to contact users due to posting content. Additional research may be needed using machine learning with virtual communities, followers or friends, and sharing of information (such as retweeting links) with suicide-related content (Colombo, Burnap, Hodorog, & Scourfield, 2016).

Recommendations

Although AI/ML methods appear promising for identification of suicide risk expressed via social media all current research has been conducted with civilian samples and questions remain regarding reliability and validity of models. Work is expanding rapidly in these areas, but there are no clear evidence-based approaches to recommend to the Department of Defense which might assist in identifying high risk for suicide service members based on the content of their social media posts.

References

Abboute, A., Boudjeriou, Y., Entringer, G., Azé, J., Bringay, S., & Poncelet, P. (2014). Mining Twitter for suicide prevention. In E.Métais, M. Roche, & M. Teisseire (Eds.), *Natural language processing and information systems* (Vol. 8455, pp. 250–253). Cham, Switzerland: Springer International Publishing.

Ahuja, A. K., Biesaga, K., Sudak, D. M., Draper, J., & Womble, A. (2014). Suicide on Facebook. *Journal of Psychiatric Practice, 20*(2), 141–146.

Baddeley, J. L., Daniel, G. R., & Pennebaker, J. W. (2011). How Henry Hellyer's use of language foretold his suicide. *Crisis*, *32*(5),288–292.

Barrett, J. R., Shetty, H., Broadbent, M., Cross, S., Hotopf, M., Stewart, R., & Lee, W. (2016). "He left me a message on Facebook": Comparing the risk profiles of self-harming patients who leave paper suicide notes with those who leave messages on new media. *British Journal of Psychiatry Open, 2*(3), 217–220.

Bentley, K. H., Franklin, J. C., Ribeiro, J. D., Kleiman, E. M., Fox, K. R., & Nock, M. K. (2016). Anxiety and its disorders as risk factors for suicidal thoughts and behaviors: A meta analytic review. *Clinical Psychology Review*, *43*, 30–46. http://doi.org/10.1016/j.cpr.2015.11.008

Cavazos-Rehg, P. A., Krauss, M. J., Sowles, S. J., Connolly, S., Rosas, C., Bharadwaj, M., & Bierut, L. J. (2016). An analysis of depression, self-harm, and suicidal ideation content on Tumblr. *Crisis*.

Chang, B. P., Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., & Nock, M. K. (2016). Biological risk factors for suicidal behaviors: A meta-analysis. *Translational Psychiatry*, *6*, 887.

Chhabra, N., & Bryant, S. M. (2016). Snapchat toxicology: social media and suicide. *Annals of emergency medicine*, *68*(4), 527.

Christensen, H., Batterham, P., & O'Dea, B. (2014). E-health interventions for suicide prevention. *International Journal of Environmental Research and Public Health, 11*(8), 8193–8212.

Colombo, G. B., Burnap, P., Hodorog, A., & Scourfield, J. (2016). Analysing the connectivity and communication of suicidal users on twitter. *Computer communications*, *73*, 291-300.

Delgado-Gomez, D., Blasco-Fontecilla, H., Sukno, F., Socorro Ramos-Plasencia, M., & Baca-Garcia, E. (2012). Suicide attempters classification: Toward predictive models of suicidal behavior. *Neurocomputing*, *92*, 3–8. http://doi.org/10.1016/j.neucom.2011.08.033

Fernández-Cabana, M., García-Caballero, A., Alves-Pérez, M. T., García-García, M. J., & Mateos, R. (2012). Suicidal traits in Marilyn Monroe's fragments: An LIWC analysis. *Crisis, 34*(2), 124–130.

Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., Jaroszewski, A. C., & Nock, M. K. (2016). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin*. Advance online publication.

Guan, L., Hao, B., Cheng, Q., Yip, P. S., & Zhu, T. (2015). Identifying Chinese microblog users with high suicide probability using internet-based profile and linguistic features: Classification model. *JMIR mental health*, *2*(2), e17.

Larsen, M. E., Boonstra, T. W., Batterham, P. J., O'Dea, B., Paris, C., & Christensen, H. (2015). We feel: mapping emotion on Twitter. *IEEE journal of biomedical and health informatics*, *19*(4), 1246-1252.

Lehavot, K., Ben-Zeev, D., & Neville, R. E. (2012). Ethical considerations and social media: A case of suicidal postings on Facebook. *Journal of Dual Diagnosis, 8*(4), 341–346.

Lester, D. (2009). Learning about suicide from the diary of Cesare Pavese. *Crisis, 30*(4), 222–224.
Lester, D., & McSwain, S. (2011). A text analysis of the poems of Sylvia Plath. *Psychological Reports, 109*(1), 73.

Lewis SP, Heath NL, Sornberger MJ, Arbuthnott AE. Helpful or harmful? An examination of viewers' responses to nonsuicidal self-injury videos on YouTube. *Journal of Adolescent Health*, *51*(4):380–385. DOI: 10.1016/j.jadohealth.2012.01.013

Mann, J., Ellis, S., Waternaux, C., & Liu, X. (2008). Classification trees distinguish suicide attempters in major psychiatric disorders: A model of clinical decision making. *Journal of Clinical Psychiatry*, *69*, 23–31.

McGee, J., Caverlee, J., & Cheng, Z. (2013). Location prediction in social media based on tie strength. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management* (pp. 459-468). ACM.

McClellan, C., Ali, M. M., Mutter, R., Kroutil, L., & Landwehr, J. (2016). Using social media to monitor mental health discussions− evidence from Twitter. *Journal of the American Medical Informatics Association*, ocw133.

Mowery, D., Smith, H., Cheney, T., Stoddard, G., Coppersmith, G., Bryan, C., & Conway, M. (2017). Understanding depressive symptoms and psychosocial stressors on Twitter: A corpus-based study. *Journal of Medical Internet Research*, *19*(2), e48.

O'Dea, B., Larsen, M. E., Batterham, P. J., Calear, A. L., & Christensen, H. (2017). A linguistic analysis of suicide-related Twitter posts. *Crisis*.

O'Dea, B., Wan, S., Batterham, P. J., Calear, A. L., Paris, C., & Christensen, H. (2015). Detecting suicidality on Twitter. *Internet Interventions*, *2*(2), 183–188.

Orme, J. (2014). Samaritans pulls 'suicide watch' Radar app over privacy concerns. *The Guardian*. Retrieved from http://www.theguardian.com/society/2014/nov/07/samaritans-radar-app-suicide-watch-privacy-twitter-users

Ribeiro, J. D., Franklin, J. C., Fox, K. R., Bentley, K. H., Kleiman, E. M., Chang, B. P., & Nock, M. K. (2016). Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: A meta-analysis of longitudinal studies. *Psychological Medicine*, 46, 225–236. DOI: http://doi.org/10.1017/S0033291715001804

Varathan, K. D., & Talib, N. (2014, August). Suicide detection system based on Twitter. *Science and Information Conference (SAI), 2014*,785-788.

World Health Organization. (2016). Suicide data. Geneva, Switzerland: Author.

Won, H. H., Myung, W., Song, G. Y., Lee, W. H., Kim, J. W., Carroll, B. J., & Kim, D. K. (2013). Predicting national suicide numbers with social media data. *PloS one*, *8*(4), e61809.