Enhancing Depression Symptoms Detection using Social Media Data

Silas P. Lima Filho
silasfilho@ufrj.br
Federal University of Rio de Janeiro
Rio de Janeiro

Jonice Oliveira Sampaio
jonice@ppgi.ufrj.br
Federal University of Rio de Janeiro
Rio de Janeiro

ABSTRACT

A clear and well-documented LaTeX document is presented as an article formatted for publication by ACM in a conference proceedings or journal publication. Based on the "acmart" document class, this article presents and explains many of the common variations, as well as many of the formatting elements an author may use in the preparation of the documentation of their work.

KEYWORDS

depression, social networks, social media

ACM Reference Format:
Silas P. Lima Filho and Jonice Oliveira Sampaio. 2020. Enhancing Depression Symptoms Detection using Social Media Data. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Depression is one of the most related mental diseases in the world. Some people call it the century illness due to its dangerousness. It can lead, in extreme situations, to suicide [1]. World Health Organization (WHO) presents a number of around 300 million people from different ages who suffer some kind of depression[1]. Some of these symptoms are, for example, depressed mood in most of the day, lost of interest in regular activities, weight loss, and insomnia.

Sometimes depression is wrongly understood as a emotional state which reflects sadness. Independently of above examples, the fact is that the mental disease depression is wrongly defined and acknowledged. Depression is one of the most related mental disease in the world and some people call it as the century illness. World Health Organization (WHO) presents a number of around 300 million people from different ages who suffer some kind of depression [2]. In Brazil, the Health Ministry presents a number of 11,5 million people who are affected by depression [3]. Depression is classified on 11th international Disease Classification (ICD 11) as a disease when it is diagnosed in someone behavior. Some of these symptoms are for example, depressed mood in most part of the day, lost of interest in regular activities, weight loss and insomnia. Depression can happen triggered by an event where the person has lost something e.g. job, some close person etc. This disease is also dangerous in order of its extreme consequences. Depression, according to ICD 11, can lead to suicide ideation and consequently to suicide [1].

With the wide spread and with more people with depression, it becomes a challenge identify a person who can be in a real state of depression. Depression also afflicts people who are in specific location where the access to a professional is **costly impending**. Thus, identify and attend someone who could be a potential depressive patient, in a fast and unobtrusive way, seems very helpful.

With more affected people, it becomes a challenge to identify a person who can be in a real state of major depression disease. It also afflicts people who are in a specific location where access to a professional is hindered due to cost. Thus, it is helpful to recognize and attend someone who could be a potential depressive patient, in a fast and unobtrusive way.

In the context of online diagnosis, *infodemiology* and *digital disease detection* are correlated terms to describe the use of digital platforms and tools to improve society health. They can be translated as efforts to tackle epidemics detection, identify individuals at risk, and communicate candidate urgent illness. The use of technology could directly support institutions, professionals, and even to aid people to make themselves aware of some disease [6].

Psychology already has methods to deal with depression identification. One of these methods is interview with psychologist. Through this conversation it could be possible to the professional, find some clues to identify someone with depression symptoms.

Social media has been employed on academia to monitor people's behaviour and their personal choices. With this in mind, sounds interesting to investigate if it is possible to identify signs, symptoms of depressive behaviour on social media platforms.

As said above, the task of identifying some disease, even if it is not depression, can be challenging. Due to the plant offers of data types, select what is the most effective, precise and reliable technique, method analysis could require a great quantity of time of research.

Thus, we can abbreviate our effort as the following question: *Is it possible to identify psychological diseases symptoms from people who are social media users?* Our research questions can be described as follow *It would be possible to identify psychological diseases symptoms using social media?* Is there a disturb diagnosis method only using data from social media?

[1] www.who.int/en/news-room/fact-sheets/detail/mental-disorders
[2] www.who.int/en/news-room/fact-sheets/detail/mental-disorders
[3] www.blog.saude.gov.br/index.php/materias-especiais/52516-mais-de-onze-milhoes-de-brasileiros-tem-depressao

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 BACKGROUND LITERATURE

For the literature selection we have applied a systematic literature review in order to have a deeper insight from the most actual research which tackles the depression diagnosis in social media.

From a computational perspective, the problem of dealing with depression and social media can be understand as search for people who suffer the symptoms of depression. Using different methods to identify if someone have depression or not, or if some person have one of the symptoms of depression. A good amount of articles relies on natural language processing (NLP) to make a systemic analysis over the text in social media publications.

In [14], the authors have applied text classification technique using Convolutional Neural Networks to classify depression using text analysis. The authors in [8] also have used neural networks in order to find patterns on periods when the risk of suicide attempt is increasing in SMS texts.

The work in [7] is a qualitative study that try to understand how is the behavior and reception of chinese population about depression. It is a qualitative study and differs from the prior ones.

De Choudhury is an author that we would like to highlight. She is one author who have developed many articles and researches about measurement of depression on social media context. Her works vary on some types of measurement and implications of depression in some contexts of affected persons. We will cite some of these works due to the importance and relevance of her effort to contribute to this area. In [4], the authors start to analyze might afflicted people by depression. They have used crowdsourcing to obtain data from twitter by people who were clinically diagnosed with depression. With this data, they have constructed a corpus and created a probabilistic model. The trained model would be relevant to indicate if a not seen post indicates depression. Similar to previous work, Tsugawa et al [11] have applied the same analysis. However the users were selected from Japan. They try to replicate the results from [4].

In [9], they present how activities on Facebook are associated with the depressive states of users and how depressive moods. Their goal was to raise awareness to depression at the university where the study was conducted.

On [2], the authors explore self-disclosures posts in Instagram. In this article, the authors use posts from people who tagged their post with #depression in order to understand what kinds of sensitive disclosures do people make on Instagram.

In [5], the authors bring an approach to understand what is the behaviour of users in TrevorSpace. TrevorSpace is a social media platform (social network site - SNS) where their users are people from LGBTQ. This platform aims to prevent and avoid suicide among these users community. Even though suicide and depression can occur inside different groups and ages, the authors confirm that some groups can face this problem more frequently. The LGBTQ is assured to be one of these groups who face it. The authors choice on that SNS is justified by completeness of the network. Usually the task of selecting a certain group from a general network or social media 'esbarra' on selecting properly the users related to analysis. Thus, 'split' not related users to the desired ones is a reflection needed task. In this paper the authors don't have this problem because the whole network and its users are from the studied analysis group.

The author in [12] conduct an observational study to understand the interactions between clinically depressed users and their ego-network when contrasted with a group of users without depression. They identify relevant linguistic and emotional signals from social media exchanges to detect symptomatic cues of depression.

In [13], authors incorporates temporal analysis of user-generated content on social media for capturing symptoms. They developed a statistical model which emulates traditional observational cohort studies conducted through online questionnaires by extracting and categorizing different symptoms of depression by modeling user-generated content in social media.

In [3], they detected eight basic emotions and calculated the overall intensity (strength score) of the emotions extracted from all past tweets of each user. After that, created a time series for each emotion of every user to generate a selection of descriptive statistics for these time series.

3 PROPOSED METHODOLOGY

The proposed methodology aims to combine concepts from computer science area, and from psychology topic area called psychometrics. Our methodology can be divided in three stages and it is represented in Figure 1. It has as premise a given dataset of publications in a social media platform (e.g. twitter, reddit, instagram, etc.), we should be capable to apply all the stages for any kind of social media.

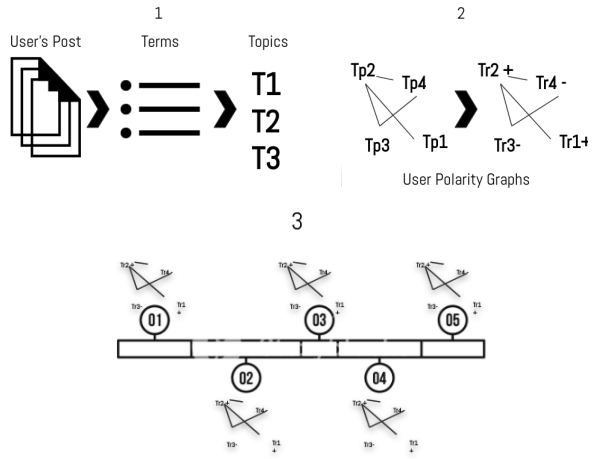

Figure 1: Workflow stages for the proposed methodology

As most of the presented articles in Section 2, we also intend to apply text analysis over the datasets, specifically the technique that we want to employ is topic identification. Based on the work of [10], we aim to first identify what are the main topics of each user in a social media platform. After the topics identification, for each user, we could map the terms into a polarity graph. The polarity graph could help to identify if the most used words and terms of a user tend to be positive or negative. Related work has shown that depressive people used to manipulate more negative words. This

reflects the low self-vision that this group express of themselves. With the polarity graphs of each user, in third stage we intend to analyze how these graph have evolved over the time. In this manner, we could be capable to check if someone discourse has been turned into more positive or negative. This could help to better explain the phenomena of depression inside social media. The time series analysis seems to be important, given that someone is considered depressed if he has manifested a set of symptoms for a certain period of time. Also, with the graph of topics, we can apply social network metrics to understand how the users are connected, and how the people around them are affected by their discourse or behavior.

In the psychological approach, we intend to apply well defined questionnaires from psychometrics area on users selected from the computational approach. Psychometrics represent the theory and technique of measuring mental processes and it is applied in the fields of psychology and education. These questionnaires could be incorporated to social media users to corroborate the classification of potential users with depression.

4 EXPECTED CONTRIBUTIONS

This work still in development of methodology, with some ideas of analysis that could be made. The expected contributions are related to computer science and psychology. For the computer science, we expect that employing computational metrics like social network analysis and topic classification could support people to acquire a more proper understanding of how the phenomena of depression happen in social media, and also how the social media reflects the real life. For the health research point of view (psychology, medicine), our approach could improve how the diagnosis of depression is performed. This could aid people with few resources to enjoy a better health by the use of technology.

REFERENCES

[1] American Psychiatry Association Apa. 2013. *DSM-V-TR - Manual Diagnóstico e Estatístico de Transtornos Mentais.* Number 1. 59–66 pages. https://doi.org/1011769780890425596 arXiv:arXiv:1011.1669v3

[2] Nazanin Andalibi, Pinar Ozturk, and Andrea Forte. 2017. Sensitive Self-disclosures, Responses, and Social Support on Instagram: The Case of #Depression. In *Proceedings of the 2017 {ACM} {Conference} on {Computer} {Supported} {Cooperative} {Work} and {Social} {Computing} - {CSCW} '17.* ACM Press, Portland, Oregon, USA, 1485–1500. https://doi.org/10.1145/2998181.2998243

[3] Xuetong Chen, Martin D Sykora, Thomas W Jackson, and Suzanne Elayan. 2018. What about Mood Swings? Identifying Depression on Twitter with Temporal Measures of Emotions. (2018), 8. https://doi.org/10.1145/3184558.3191624

[4] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Social Media As a Measurement Tool of Depression in Populations. In *Proceedings of the 5th Annual ACM Web Science Conference (WebSci '13).* ACM, New York, NY, USA, 47–56. https://doi.org/10.1145/2464464.2464480

[5] Christopher M Homan, Naiji Lu, Xin Tu, Megan C Lytle, and Vincent M B Silenzio. 2014. Social Structure and Depression in TrevorSpace. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14).* ACM, New York, NY, USA, 615–625. https://doi.org/10.1145/2531602.2531704

[6] Eric Horvitz and Deirdre Mulligan. [n.d.]. POLICY FORUM Data, privacy, and the greater good. ([n. d.]). https://doi.org/10.1126/science.aac4520

[7] Guo Li, Xiaomu Zhou, Tun Lu, Jiang Yang, and Ning Gu. 2016. SunForum: Understanding Depression in a Chinese Online Community. (2016). https://doi.org/10.1145/2818048.2819994

[8] Alicia L Nobles, Jeffrey J Glenn, Kamran Kowsari, Bethany A Teachman, and Laura E Barnes. 2018. Identification of Imminent Suicide Risk Among Young Adults Using Text Messages. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18).* ACM, New York, NY, USA, 413:1—-413:11. https://doi.org/10.1145/3173574.3173987

[9] Sungkyu Park, Inyeop Kim, Sang Won Lee, Jaehyun Yoo, Bumseok Jeong, and Meeyoung Cha. 2015. Manifestation of Depression and Loneliness on Social Networks: A Case Study of Young Adults on Facebook. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15).* ACM, New York, NY, USA, 557–570. https://doi.org/10.1145/2675133.2675139

[10] Diogo Nolasco Ferreira Sousa. 2016. *Automatic Research Areas Identification in C&T.* Ph.D. Dissertation. Universidade Federal do Rio de Janeiro.

[11] Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. 2015. Recognizing Depression from Twitter Activity. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15* (2015), 3187–3196. https://doi.org/10.1145/2702123.2702280

[12] Nikhita Vedula and Srinivasan Parthasarathy. 2017. Emotional and Linguistic Cues of Depression from Social Media. *Proceedings of the 2017 International Conference on Digital Health - DH '17* (2017), 127–136. https://doi.org/10.1145/3079452.3079465

[13] Amir Hossein Yazdavar, Hussein S Al-Olimat, Monireh Ebrahimi, Goonmeet Bajaj, Tanvi Banerjee, Krishnaprasad Thirunarayan, Jyotishman Pathak, and Amit Sheth. 2017. Semi-Supervised Approach to Monitoring Clinical Depressive Symptoms in Social Media. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 (ASONAM '17).* ACM, New York, NY, USA, 1191–1198. https://doi.org/10.1145/3110025.3123028

[14] Xiaoli Zhao, Shaofu Lin, and Zhisheng Huang. 2018. Text Classification of Micro-blog's "Tree Hole" Based on Convolutional Neural Network. In *Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence (ACAI 2018).* ACM, New York, NY, USA, 61:1—-61:5. https://doi.org/10.1145/3302425.3302501