Online User Depression Detection Using Text Sequence With Neural Network

Palivela Jyothi

Research scholar (CSE) in International School of Technology and Sciences for Womens (ISTS women's college) in NH-16, East Gonagudem, Rajanagaram, Andhra Pradesh, 533294
.E-mail: palivelajyothi22@gmail.com

Athili Venkataraju

Associate professor and hod (CSE) in International School of Technology and Sciences for Womens (ISTS women's college) in NH-16, East Gonagudem, Rajanagaram, Andhra Pradesh, 533294
.E-mail: venkataraju.athili@gmail.com

Abstract: Depression is a psychological disorder that affects over three hundred million humans worldwide. A person who is depressed suffers from anxiety in day-today life, which affects that person in the relationship with their family and friends, leading to different diseases and in the worst-case death by suicide. With the growth of the social network, most of the people share their emotion, their feelings, their thoughts in social media. If their depression can be detected early by analyzing their post, then by taking necessary steps, a person can be saved from depression-related diseases or in the best case he can be saved from committing suicide. In this research work, a hybrid model has been proposed that can detect depression by analyzing user's textual posts. Deep learning algorithms were trained using the training data and then performance has been evaluated on the test data of the dataset of reddit which was published for the pilot piece of work, Early Detection of Depression in CLEF eRisk 2019. In particular, Bidirectional Long Short Term Memory (BiLSTM) with different word embedding techniques and metadata features were proposed which gave good results.

**Keywords: depression, social media, early detection, deep
Learning.**

I. INTRODUCTION

According to World Health Organization (WHO), more than 300 million people worldwide are suffering from depression, which equals about 4.4 percent of the global population. While forms of depression are more common among females (5.1 percent) than males (3.6 percent) and prevalence differs between regions of the world, it occurs in any age group and is not limited to any specific life situation. Depression is therefore often described to be accompanied by paradoxes, caused by a contrast between the self-image of a depressed person and the actual facts. Latest results from the 2016 National Survey on Drug Use and Health in the United States report that, during the year 2016, 12.8 percent of adolescents between 12 and 17 years old and 6.7 percent of adults had suffered a major depressive episode (MDE). Precisely defining depression is not an easy task, not only because several sub-types have been described and changed in the past, but also because the term "being depressed" has become frequently used in everyday language. In general, depression can be described to lead to an altered mood and may also be accompanied, for example, by a negative self-image, wishes to escape or hide, vegetative changes, and a lowered overall activity level. The symptoms experienced by depressed individuals can severely impact their ability to cope with any situation in daily life and therefore differ drastically from normal mood variations that anyone experiences. At the worst, depression can lead to suicide. WHO estimates that, in the year 2015, 788,000 people have died by suicide and that it was the second most common cause of death for people between 15 and 29 years old worldwide. In Europe, self-harm was even reported as the most common cause of death in the age group between 15 and 29 and the second most common between 30 and 49, again in results obtained by WHO in 2015. Although the severity of depression is well-known, only about half of the individuals affected by any mental disorder in Europe get treated. The proportion of individuals seeking treatment for mood disorders during the first year ranges between 29–52 percent in Europe, 35 percent in the USA, and only 6 percent in Nigeria or China. In addition to possible personal reasons for avoiding treatment, this is often due to a limited availability of mental health care, for example in conflict regions. Via a telephone survey in Germany, researchers found out that shame and self-stigmatization seem to be much stronger reasons to

not seek psychiatric help than actual perceived stigma and negative reactions of others. They further speculate that the fear of discrimination might be relatively unimportant in their study because people hope to keep their psychiatric treatment secret. Another study amongst people with severe mental illness in Washington D. C. showed that stigma and discrimination indeed exist, while they are not "commonly experienced problems" but rather "perceived as omnipresent potential problems". While depression and other mental illnesses may lead to social withdrawal and isolation, it was found that social media platforms are indeed increasingly used by affected individuals to connect with others, share experiences, and support each other. Based on these findings, peer-to-peer communities on social media can be able to challenge stigma, increase the likelihood to seek professional help, and directly offer help online to people with mental illness. A similar study in the USA came to the conclusion that internet users with stigmatized illnesses like depression or urinary incontinence are more likely to use online resources for health-related information and for communication about their illness than people with another chronic illness. All this emphasizes the importance of research toward ways to assist depressed individuals on social media platforms and on the internet in general. This paper is therefore focused on ways to classify indications of depression in written texts as early as possible based on machine learning methods.

1.1 Objective of the project

Depression is ranked as the largest contributor to global disability and is also a major reason for suicide. Still, many individuals suffering from forms of depression are not treated for various reasons. Previous studies have shown that depression also has an effect on language usage and that many depressed individuals use social media platforms or the internet in general to get information or discuss their problems. This paper addresses the early detection of depression using machine learning models based on messages on a social platform. In particular, a convolutional neural network based on different word embeddings is evaluated and compared to a classification based on user-level linguistic metadata. An ensemble of both approaches is shown to achieve state-of-the-art results in a current early detection task. Furthermore, the currently popular ERDE score as metric for early detection systems is examined in detail and its drawbacks in the context of shared tasks are illustrated. A slightly modified metric is proposed and compared to the original score. Finally, a new word embedding was trained on a large corpus of the

same domain as the described task and is evaluated as well.

II. LITERATURE SURVEY

"Delay and failure in treatment seeking after first onset of mental disorders in the world health organization's world mental health survey initiative,"

Data are presented on patterns of failure and delay in making initial treatment contact after first onset of a mental disorder in 15 countries in the World Health Organization (WHO)'s World Mental Health (WMH) Surveys. Representative face-to-face household surveys were conducted among 76,012 respondents aged 18 and older in Belgium, Colombia, France, Germany, Israel, Italy, Japan, Lebanon, Mexico, the Netherlands, New Zealand, Nigeria, People's Republic of China (Beijing and Shanghai), Spain, and the United States. The WHO Composite International Diagnostic Interview (CIDI) was used to assess lifetime DSM-IV anxiety, mood, and substance use disorders. Ages of onset for individual disorders and ages of first treatment contact for each disorder were used to calculate the extent of failure and delay in initial help seeking. The proportion of lifetime cases making treatment contact in the year of disorder onset ranged from 0.8 to 36.4% for anxiety disorders, from 6.0 to 52.1% for mood disorders, and from 0.9 to 18.6% for substance use disorders.

By 50 years, the proportion of lifetime cases making treatment contact ranged from 15.2 to 95.0% for anxiety disorders, from 7.9 to 98.6% for mood disorders, and from 19.8 to 86.1% for substance use disorders. Median delays among cases eventually making contact ranged from 3.0 to 30.0 years for anxiety disorders, from 1.0 to 14.0 years for mood disorders, and from 6.0 to 18.0 years for substance use disorders. Failure and delays in treatment seeking were generally greater in developing countries, older cohorts, men, and cases with earlier ages of onset. These results show that failure and delays in initial help seeking are pervasive problems worldwide. Interventions to ensure prompt initial treatment contacts are needed to reduce the global burdens and hazards of untreated mental disorders.

"The stigma of psychiatric treatment and help-seeking intentions for depression,"

The stigma of mental illness has often been considered a potential cause for reluctant willingness to seek help for mental problems, but there is little evidence on this issue. We examine two aspects of stigma related to seeing a psychiatrist and their association with help-seeking intentions for depression: anticipated discrimination by others when seeking help and desire for social distance

from those seeking help. Representative population survey in Germany 2007 (n = 2,303), containing a depression vignette with a question on readiness to seek psychiatric care for this problem, a focus group developed scale anticipated discrimination when seeing a psychiatrist (ADSP), and a scale on desire for social distance from someone seeing a psychiatrist (SDSP).

We further elicited previous contact to psychiatric treatment, depressive symptoms, and socio-demographic data. Both scales had good internal consistency (Cronbach's alpha ADSP 0.87, SDSP 0.81). Exploratory factor analysis of all items revealed a distinct factor representing the social distance scale and three factors "anticipated discrimination", "anticipated job problems" and "anticipated shame" derived from the ADSP scale. In both the general population and in those with current depressive syndrome, personal desire for social distance significantly decreased willingness to seek psychiatric help, but anticipated discrimination by others did not. Other factors related to likely help-seeking were female gender and previous contact to psychiatric treatment or to psychotherapy. Contrary to expectations, anticipated discrimination from others was unrelated to help-seeking intentions, while personal discriminatory attitudes seem to hinder help-seeking. Our findings point to self-stigmatization as an important mechanism decreasing the willingness to seek psychiatric help.

"Naturally occurring peer support through social media: the experiences of individuals with severe mental illness using YouTube,"

Increasingly, people with diverse health conditions turn to social media to share their illness experiences or seek advice from others with similar health concerns. This unstructured medium may represent a platform on which individuals with severe mental illness naturally provide and receive peer support. Peer support includes a system of mutual giving and receiving where individuals with severe mental illness can offer hope, companionship, and encouragement to others facing similar challenges. In this study we explore the phenomenon of individuals with severe mental illness uploading videos to YouTube, and posting and responding to comments as a form of naturally occurring peer support.

We also consider the potential risks and benefits of self-disclosure and interacting with others on YouTube. To address these questions, we used qualitative inquiry informed by emerging techniques in online ethnography. We analyzed n=3,044 comments posted to 19 videos uploaded by individuals who self-identified as having schizophrenia, schizoaffective disorder, or bipolar disorder. We found peer support across four themes: minimizing a sense of isolation and providing hope; finding support through peer exchange and reciprocity; sharing strategies for coping with day-to-day challenges of severe mental illness; and learning from shared experiences of medication use and seeking mental health care.

These broad themes are consistent with accepted notions of peer support in severe mental illness as a voluntary process aimed at inclusion and mutual advancement through shared experience and developing a sense of community. Our data suggest that the lack of anonymity and associated risks of being identified as an individual with severe mental illness on YouTube seem to be overlooked by those who posted comments or uploaded videos. Whether or not this platform can provide benefits for a wider community of individuals with severe mental illness remains uncertain.

III.   METHODOLOGY

In this study, we first focused on four types of factors such as emotional process, temporal process, linguistic style and all (emotional, temporal, linguistic style) features together for the detection and processing of depressive data received as Facebook posts. We then apply supervised machine learning approaches to study each factor types independently. The classification techniques such as 'decision tree', 'k-Nearest Neighbor', 'Support Vector Machine', and 'ensemble' are deemed suitable for each type (refer to Fig. 1).
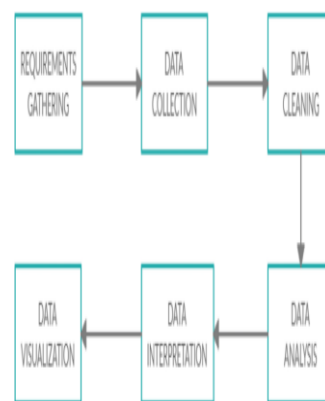

Fig. 1 Methodology

IV.   DATA SET EXPLORATION

We worked on Facebook users' comments for depressive behavioral exploration and detection. We collected data from the social network [26]. Preparing of social network data, in particular Facebook user's comments is one of the primary challenges which bear information on whether or not they could contain depression bearing content. To tackle this issue we use NCapture for collecting data from Facebook [27, 28]. For qualitative data analysis, NCapture is a powerful tool in the world today. It is intended to enable to arrange, break down and discover knowledge in unstructured data like open-ended survey responses, social media, interviews, articles and web content. Furthermore it gives a place to arrange and deal with material to discover knowledge in a more proficient way [29].

V. DATA SET PREPARATION

After collecting the raw data from Facebook, it was analyzed by using LIWC Software [7, 8]. LIWC is the heart of the text analysis strategy and can process text on a line by line. Our primary dataset contains total 21 columns where 13 columns represent the linguistic style (articles, prepositions, auxiliary verbs, conjunctions, personal pronoun, impersonal pronouns, verbs, negation etc.) information, 5 columns represent the emotional (positive, negative, sad, anger and anxiety) information, 3 columns represent the temporal process (past, present and future) information and each column gives the individual information's about depressive behavior (refer to Table 1).

Precision is the proportion of true positives to the cases that are anticipated as positive. It is the level of chosen cases that are right.

$$Precision\,(P) = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall is the proportion of true positives to the cases that truly positive. It is the level of chosen cases that are selected.

$$Recall\,(R) = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

F-measure is the mean of Precision and Recall. It takes both false positives and false negatives into a record. F-measure is calculated as:

$$F - Measure = 2\frac{PR}{P+R}$$

VI. CONCLUSION

This work has been used to examine the currently popular ERDEo metric for early detection tasks in detail and has shown that especially ERDE5 is not a meaningful metric for the described shared task. Only the correct prediction of few positive samples has an effect on this score and the best results can therefore often be obtained by only minimizing false positives. A modification of this metric, namely ERDE% o , has been proposed that is better interpretable in the case of shared tasks that require information to be read in chunks. Exemplary scores using this score have been shown in comparison to ERDEo scores for the experiments in this work. Previous experiments for the eRisk 2017 task for early detection of depression have been continued by examining additional user-level metadata features and evaluating a convolutional neural network as text-based depression classifier. State-of-the-art results have been reported for the eRisk 2017 dataset using these two approaches. A new fastText word embedding has been trained on a large corpus of reddit comments. The analysis of the resulting word vectors has shown that the model has learnt some features specific to this domain and is viable for general syntactic questions in the English language as shown based on the standard word analogy task. As the results presented in this paper are optimized to obtain the best performance on the eRisk 2017 task for comparison to previously published results and among these models, future work will have to show how these models perform on yet unseen data. This has first been done during the eRisk 2018 task, which used the old dataset as training data and contained 820 new test subjects. In addition, eRisk 2018 contained an additional task aimed at the early detection of anorexia that this team has also participated in. The five submitted predictions achieved the best F1 and ERDE50 scores in both tasks and the CNN without metadata in particular achieved the best results in the new anorexia task. The same working notes paper for this second participation has also been used to evaluate the modified ERDE% o metric for all participants and again shows how especially the original ERDE5 metric favors systems that correctly predict test users with only few documents in total regardless of their overall performance

VII. REFERENCES

[1] Depression and Other Common Mental Disorders: Global Health Estimates. World Health Organization,2017,https://www.who.int/mental_health/management/depression/prevalence_global_ health_estimates/en/.

[2] A. T. Beck and B. A. Alford, Depression: Causes and Treatment. 2nd ed. Philadelphia, PA, USA: Univ. Pennsylvania Press, 2009.

[3] Key Substance Use and Mental Health Indicators in the United States: Results from the 2016 National Survey on Drug Use and Health, Rockville, MD: Center for Behavioral Health Statistics and Quality: Substance Abuse and Mental Health Services Administration, 2017. [Online]. Available: https://www.samhsa.gov/data/

[4] E. S. Paykel, "Basic concepts of depression," Dialogues Clinical Neuroscience, vol. 10, no. 3, pp. 279–289, 2008.

[5] Global Health Estimates 2015: Deaths by Cause, Age, Sex, by Country and by Region, 2000 2015.WorldHealthOrganization,2016,https://www.who.int/entity/healthinfo/global_burden_disease/GHE2015_Deaths_Global_2000_2015.xls

[6] J. Alonso, M. Codony, V. Kovess, M. C. Angermeyer, S. J. Katz, J. M. Haro, G. De Girolamo, R. De Graaf, K. Demyttenaere, G. Vilagut, et al., "Population level of unmet need for mental healthcare in europe," British J. Psychiatry, vol. 190, no. 4, pp. 299–306, 2007.

[7] P. S. Wang, M. Angermeyer, G. Borges, R. Bruffaerts, W. T. Chiu, G. De Girolamo, J. Fayyad, O. Gureje, J. M. Haro, Y. Huang, et al., "Delay and failure in treatment seeking after first onset of mental disorders in the world health organization's world mental health survey initiative," World Psychiatry, vol. 6, no. 3, 2007, Art. no. 177.

AUTHORS PROFILE

Palivela Jyothi completed her B.Tech in 2013 in Computer Science and Engineering and has interest in Software Testing, Operting System, DBMS, Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression Indications in Text Sequences.

Athili Venkataraju He is working as an associate professor in ISTS women's engineering college.