First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Detection of Depression or Sentiment Analysis

Chaitanya Suryawanshi[1], Taufik Tamboli[1], Saurav Tayade[1], Prashant Yeole[1], Prof. Niyamat Ujloomwale[1]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

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. 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. Social networks have been developed as a great point for its users to communicate with their interested friends and share their opinions, photos, and videos reflecting their moods, feelings and sentiments. This creates an opportunity to analyze social network data for user's feelings and sentiments to investigate their moods and attitudes when they are communicating via these online tools.

Keywords: Social network, Emotions, Depression, Sentiment analysis.

I. INTRODUCTION

According to World Health Organization (WHO) , more than 300 million people worldwide are suffering from depression, which equals about 4.4% of the global population. While forms of depression are more common among females (5.1%) than males (3.6%) 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% of adolescents between 12 and 17 years old and 6.7% 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 The proliferations of internet and communication technologies, especially the online social networks have rejuvenated how people interact and communicate withbeach other electronically.

The applications such as Facebook, Twitter, Instagram and alike not only host the written and multimedia contents but also offer their users to express their feelings, emotions and sentiments about a topic, subject or an issue online. On one hand, this is great for users of social networking site to openly and freely contribute and respond to any topic online; on the other hand, it creates opportunities for people working in the health sector to get insight of what might be happening at mental state of someone who reacted to a topic in a specific manner. In order to provide such insight, machine learning techniques could potentially offer some unique features that can assist in examining the unique patterns hidden in online communication and process them to reveal the mental state (such as 'happiness', 'sadness' ,'anger', 'anxiety', depression) among social networks' users. Moreover, there is growing body of literature addressing the role of social networks on the structure of social relationships such as breakup relationship, mental illness ('depression', 'anxiety', 'bipolar' etc.), smoking and drinking relapse, sexual harassment and for suicide ideation 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. In this study, we examine various linguistic cues which help to detect emotion cause events: the position of cause event and experiencer relative to the emotion keyword: emotional process like positive emotion (e.g.'happy', 'love', 'nice'), negative emotion (e.g. 'worthless ','loser', 'hurt', 'ugly', 'nasty'), sadness (e.g. 'worry', 'crying' ,'grief', 'sad'), anger (e.g. 'stop', 'shit', 'hate', 'kill', 'annoyed')and anxiety (e.g. 'worried', 'fearful'). A temporal process like present focus (e.g. 'today', 'is', 'now'), past focus (e.g. 'ago', 'did', 'talked') and future focus (e.g. 'shall', 'may', 'will' ,'soon'). Linguistic words like articles (e.g. 'a', 'an', 'the'),prepositions (e.g. 'for', 'in', 'of', 'to', 'with', 'above'), auxiliary verbs (e.g. 'do', 'have', 'am', 'will'), conjunctions (e.g. 'and' ,'but', 'whereas'), personal pronoun (e.g. 'I', 'them', 'her' ,'him'), impersonal pronouns (e.g. 'it', 'it's', 'those'), verbs (e.g. 'go', 'good') and negation (e.g. 'deny', 'dishonest', 'no ','not', 'never').

II. LITERATURE REVIEW

Previous studies have already shown that depression also has an effect on the language used by affected individuals. For example, a more frequent use of first person singular pronouns in spoken language was first observed in 1981. An examination of essays written by depressed, formerly-depressed, and non-depressed college students at University of Texas confirmed an elevated use of the word "I" in particular and also found more negative emotion words in the depressed group. Similarly, a Russian speech study found a more frequent use of all pronouns and verbs in past tense among depression patients. A recent study based on English forum posts observed an elevated use of absolutist words (e.g. absolutely, completely, every, nothing1) within forums related to depression, anxiety, and suicidal ideation than within completely unrelated forums as well as ones about asthma, diabetes, or cancer .The knowledge that language can be an indicator of an individual's psychological state has, for example, lead to the development of the Linguistic Inquiry and Word Count (LIWC) software.

By utilizing a comprehensive dictionary, it allows researchers to evaluate written texts in several categories based on word counts. A more detailed description of LIWC. With a similar purpose, Differential Language Analysis Toolkit (DLATK) an open-source Python library, was created for text analysis with a psychological, health, or social focus. Driven by the growing availability of data, for example through social media, and the technological advances that allow researchers to work with this data, ethical considerations are becoming more and more important in the field of Natural Language Processing (NLP). Based on these developments, NLP has changed from being mostly focussed on improving linguistic analysis towards actually having an impact on individuals based on their writings. Still, a proper discussion about ethics in NLP has only been started in 2016 by Hovy and Spruit . Although Institutional Review Boards (IRBs) have been well-established to enforce ethical guidelines on experiments that directly involve human subjects, the authors note that NLP and data sciences in general have not constructed such guidelines. They further argue that language "is a proxy for human behavior, and a strong signal of individual characteristics" and that, in addition, "the texts we use in NLP carry latent information about the author and situation" . On top of this direct connection to the individual, they also describe the social impact of NLP research . A demographic bias in the selection of training texts can lead to the exclusion of specific groups, overgeneralization based on false positives can have serious consequences depending on the task, and research results can potentially cause or confirm biases and ultimately discrimination by topic overexposure. Even if all these factors are considered, they conclude that dual-use problems can exist for any research if results are used in a different way than originally intended.

The work described in this paper belongs to the area of Natural Language Processing (NLP) and text classification in particular. The origins of text classification tasks can be found in early research to automatically categorize documents based on statistical analysis of specific clue words in 1961 . Later, similar research goals lead to rule based text classification systems like CONSTRUE in 1990 and finally the field began to shift more and more to machine learning algorithms around the year 2000 In addition to text categorization, machine learning was also a driving force in other text-based tasks like sentiment analysis, which is focussed on extracting opinions and sentiment from text documents . It was first used in combination with machine learning to find positive or negative opinions in movie reviews and was then extended to other review domains , as well as completely different areas like social media monitoring and general analysis of consumer attitudes. More recently, deep learning has been utilized for text classification in addition to its more common, usages in image classification. State-of-the-art results in several text-based tasks could, for example, be achieved by transfer learning methods like Universal Language Model Fine-tuning (ULMFit) and the Google research project Bidirectional Encoder Representations from Transformers(BERT) for the training of language representations, which includes ULM Fit and several other methods. Based on these Especially the availability of social media messages enabled researchers to extract population-based health information that made it possible to track disease symptom.

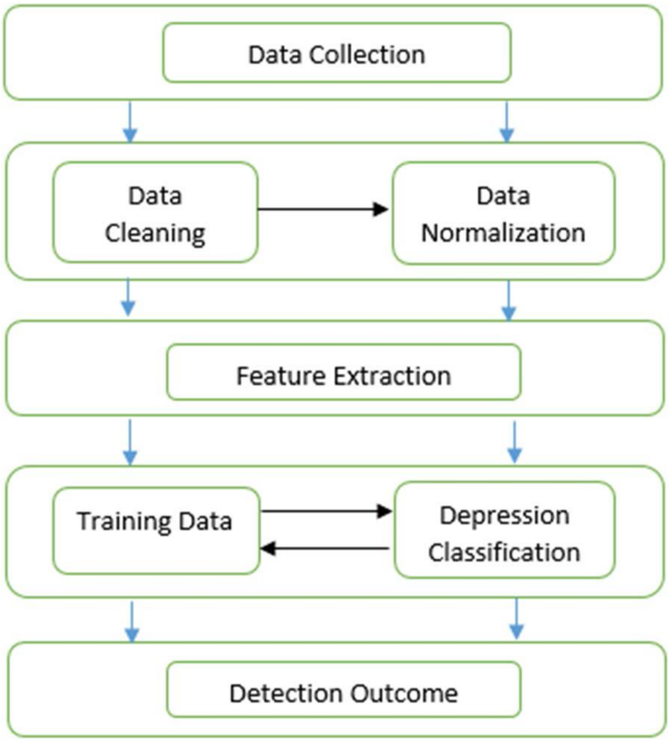

Fig. 1 A methodological overview of Facebook data analysis for depression analysis

There is growing body of literature that analyses the properties of depression . Choudhury et al.
argue that depression constitutes a genuine test in individual and general wellbeing. Considerable number of individuals experiences the ill-effects of despondency and just a division gets sufficient treatment every year. They also investigated the possibility to utilize online networking to identify and analyze any sign of significant depression issue in people. Through their web-based social networking postings, they quantified behavioral credits identifying with social engagement, feeling, dialect and semantic styles, sense of the self-system, and notices of antidepressant medications.
Choudhury et al. considered online networking
as a promising instrument for public health, concentrating on the utilization of Twitter presents on fabricating predictive models about the forthcoming impact of childbirth on the conduct and disposition of new mothers. Utilizing Twitter posts, they measured postpartum changes in 376 mothers along measurements of social engagement, feeling, informal community, and phonetic style. O'Dea et al. examined that Twitter is progressively researched as methods for recognizing psychological well-being status, including depression and suicidality in the population. Their investigation revealed that it is conceivable to recognize the level of worry among suicide- related tweets, utilizing both human coders and a programmed machine classifier.
Zhang et al. have shown that if individuals with a
high danger of suicide can be recognized through online Islam *et al. Health Inf Sci Syst (2018) 6:8* Page 3 of 12 networking like microblog, it is conceivable to actualize a dynamic intervention system to save their lives. Many researchers have demonstrated that utilizing user-created content (UGC) accurately may help decide individuals' psychological wellness levels. For instance, Aldar wish and Ahmad examined that the utilization of Social Network Sites (SNS) is expanding these days, particularly by the more youthful eras. Because the accessibility of SNS enables clients to express their interests, sentiments and offer day by day schedule .
Nguyen et al. utilized machine learning and statistical strategies to separate online messages amongst depression and control groups utilizing temperament,
psycholinguistic procedures and substance subjects
removed from the posts created by individuals from
these groups. Park et al. investigated states of mind and practices toward online web-based social networking in view of whether one is discouraged or not. They directed semi-organized up close and personal meetings with 14 dynamic Twitter users, half of whom were discouraged and the other half non-discouraged. Other than they examined a few plan implications for future social networks that could better suit users with depression and give bits of knowledge towards helping discouraged users address their issues through online web-based social

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020 - 162-169

networking. Bachrach et al. studied how user's activity on Facebook identifies with their identity, as measured by the standard Five Factor Model. They analyse relationships between user's identity and the properties of their Facebook profiles. For instance, the size and thickness of their friendship network, number of transferred photographs, and number of occasions went to, number of gathering enrolment's, and number of times the user has been tagged in photographs.

Ortigosa et al. have exhibited a new strategy for sentiment examine in Facebook that suggests that starting from messages composed by users,

as to extract data about the users' assessment extremity (positive, unbiased or negative), as transmitted in the messages they write; and to show the users' standard conclusion extremity and to distinguish huge passionate changes. In the context of Facebook mining, Holleran found initial evidence that depression is a major contributor to the overall global burden of diseases. In other related work, Wang et al. and Shen et al. examined various depression-related features, and built amultimodal depressive model to detect the depressed users. Although, some of the above reported work has discussed emotional process, temporal process, linguistic style to detect depression, the following shortcomings are observed in the existing literature:

There are few individual studies that have applied SVM, KNN, Decision Tree and Ensemble separately. There are no well-known studies that have combined all these techniques together at same dataset to investigate the variations in technique-based findings. There is no significant study that has applied the abovementioned machine learning techniques on Facebook data for depression detection. To address the above-listed shortcomings, we make an attempt to detect depression from Facebook comments with the present work; expand the scope of social media-based depression measures, describing the different features of Facebook user comments. applied machine learning approaches that can use those measures for the detection of individuals who are suffering with depression.

Year	authors	Data
1981	University of Texas	more frequent use of first person singular pronouns in spoken language
2017	Almeida, H., Briand, A., Meurs, M.J	Detecting early risk of depression from social media user-generated content. In: Proceedings Conference and Labs of the Evaluation Forum CLEF
2018	Cacheda, F., Fernandez, D., Novoa, F., Carneiro, V.:	Artificial intelligence and social networks for early detection of depression.
2017	Trotzek, M., Koitka, S., Friedrich, C.M	Linguistic metadata augmented classiers at the clef 2017 task for early detection of depression. In: Proceedings Conference and Labs of the Evaluation Forum CLEF
2014	Prieto, V.M., Matos, S., Alvarez, M., Cacheda, F., Oliveira, J.L.:	Twitter: a good place to detect health conditions.
2017	Aldarwish MM, Ahmad HF	Predicting depression levels using social media posts. In: 2017 IEEE 13th international Symposium on Autonomous decentralized system

Table 1. Summary of literature review

III. SYSTEM ARCHITECTURE

There are a number of ways to analyze the information, but the reality is that mental health, specifically depression, is a subjective and complex topic. While it may be possible to quantify the degree to which one might be depressed based on a Tweet, the only real question that matters for this project is, is an individual exhibiting linguistic markers indicative of depression? Knowing the question and the subjective nature of mental illness, a binary classification model made the most sense for this project. While a logistic regression model made sense as a benchmark model, a Long Short Term Memory network (LSTM) model wound up being the most robust for the project at hand. A recurrent neural network allows information to be passed from one step of a network to another, and are ideal for sequences, lists, and other language processing problems. A LSTM is capable of learning long-term dependancies and work incredibly well on a large variety of problems. The LSTM + CNN model takes in an input and then outputs a single number representing the probability that the tweet indicates depression. The model takes in each input sentence, replaces it with its embeddings, and then runs the new embedding vector through a convolutional layer. The convolutional layer passes the structure that it learns from the sequential data into a LSTM layer. The output of the LSTM layer is then fed into a Dense model for prediction.

Once the model was designed and built, the issue then became refining the model to achieve the best results.

Fig. 2 A methodological overview of tweets analysis for depression analysis

The system will send the tweets to be analyzed and stored the results in the database. The tweets will be analyzed in all three models. The system will return the predicted sentiments which are Positive, Negative or Neutral. When the system returns two Positive results and one Negative or Neutral result, the system will take the Positive predicted sentiment as for the Overall Predicted Sentiment, same as for two Negative results and two Neutral results. The tweets are then analyzed using three different techniques which are Naïve Bayes Classifier technique, NLP techniques and Deep Learning technique

After the sentiment of each user tweets is calculated, the depression percentage is then calculated based from the total positive and total negative tweets. If the users have a high percentage of positive tweets, it will classify the users as an optimistic person that implies the user is no depression related. Meanwhile, users that have a high percentage of negative tweets, it will classify the users as an optimistic person that can implies the users might be depression related

IV. CONCLUSION

In this paper exhibited the capability of using
or measuring and detecting major depression among its users. To give a clear understanding of our work, numbers of research challenges were stated at the start of this paper. The analytics performed on the selected dataset, provide some insight on the research challenges : What depression is and what are the common factors contributing toward depression. While we feel moody, sad or low from time to time, few people encounter these emotions seriously, for drawn out stretches of time (weeks, months or even years) and in some cases with no apparent reason. Despondency is something other than a low state of mind—it's a genuine condition that influences someone's physical and emotional feelings. Depression can influence any of us anytime. However, some phases or events make us more vulnerable to depression. Physical and emotional changes associated with growing-up, losing a loved one, beginning a family, retirement may trigger some emotional influx that could lead toward depression

for few people. What are the factors to look for depression detection in social networking comments? It is important to remember that depressive emotions have several signs and symptoms spread across various categories as reported in Based on signs and symbols divided dataset into 5 emotional variables (positive, negative, sad, anger, anxiety), 3 temporal categories (present focus, past focus and future focus), 9 standard linguistic dimensions (e.g., articles, prepositions, auxiliary verb, adverbs, conjunctions, pronoun, verbs and negations) How to extract these factors from social sites comments ? To extract the above-mentioned factors, we applied Linguistic Inquiry and Word Count (LIWC) on our dataset. The LIWC2015 Dictionary is the heart of the text analysis strategy. It processes our comments on a 'line by line' basis within and across columns of spreadsheet and accesses a single text within a spreadsheet and analyse each line sequentially and reads one target word at a time. What is the relationship between these factors and attitudes toward depression? The relationship between the above-mentioned issues with the attitudes towards depression are varies from person to person. When are the most influential time to communicate within depressive Indicative Facebook user? In this study, got 54.77% depressive indicative Facebook users communicate with their friends from midnight to midday and 45.22% from midday to midnight.

V. REFERENCES

[1]. Depression and Other Common Mental Disorders: Global Health Estimates. World Health Organization, 2017.

[2]. A. T. Beck and B. A. Alford, Depression: Causes and Treatment. Second Edition. University of 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. OnlineAvailable: https://www.samhsa.gov/data/

[4]. E. S. Paykel, "Basic concepts of depression," Dialogues in Clinicaln 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. World Health Organization, 2016.

[6]. J. Alonso, M. Codony, V. Kovess, M. C. Ange rmeyer, S. J. Katz, J. M. Haro, G. De Girolamo, R. De Graaf, K. Demyt tenaere, G. Vilagut et al., "Population level of unmet need for mental healthcare ineurope," The British Journal of Psychiatry, vol. 190, no. 4, pp. 299

[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, p. 177, 2007.

[8]. A. Rahman, S. U. Hamdani, N. R. Awan, R. A. Bryant, K. S. Dawson, M. F. Khan, M. M.-U.-H. Azeemi, P. Akhtar, H. Nazir, A. Chiumento et al., "Effect of a multicomponent behavioral intervention in adults impaired by psychological distress in a conflictaffected: A randomized clinical trial," JAMA, vol.316, no. 24, pp. 2609–2617, 2016.

[9]. G. Schomerus, H. Matschinger, and M. C. Angermeyer, "Them stigma of psychiatric treatment and help-seeking intentions for depression," European Archives of Psychiatry and Clinical Neuroscience ,vol. 259, no. 5, pp. 298–306, 2009.

[10]. R. Whitley and R. D. Campbell, "Stigma, agency and recovery amongst people with severe mental illness," Social Science &Medicine, vol. 107, pp. 1 – 8, 2014.

[11]. K. Gowen, M. Deschaine, D. Gruttadara, and D. Markey, "Young adults with mental health conditions and social networking websites:Seeking tools to build community." Psychiatric Rehabilitation Journal, vol. 35, no. 3, pp. 245–250, 2012.

[12]. Haberler G. Prosperity and depression: a theoretical analysis of cyclical movements. London: Routledge; 2017.

[13]. Guntuku SC, et al. Detecting depression and mental illness on social media: an integrative review. Curr Opin Behav Sci. 2017;18:43–9.

[14]. De Choudhury M, et al. Predicting depression via social Media. In: ICWSM, vol. 13. 2013. p. 1–10.

[15]. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. In: Proceedings of the SIGCHI conference on human factors in computing systems. New York: ACM; 2013.

[16]. O'Dea B, et al. Detecting suicidality on Twitter. Internet Interv. 2015;2(2):183–8.