International Journal of Computer Engineering in Research Trends
Multidisciplinary, Open Access, Peer-Reviewed and fully refereed
Research Paper | Volume-5, Issue-2 ,2018 Regular Edition | E-ISSN: 2349-7084

Machine Learning for Depression Diagnosis using Twitter data

Krishna Shrestha[#1]
[#]CSE Department, Amity University Mumbai, India
[1]shresthakrishna1997@gmail.com

Available online at: http://www.ijcert.org

Received: 07/February/2017, Revised: 07/February/2017, Accepted: 14/February/2017, Published: 06/March /2017

Abstract: - World Health Organisation reports that Depression is the most prevalent mental illness and major causes of disability in the world. Though effective treatment for Depression is known, it does not reach the majority of the sufferers in both rich as well as poor countries. In an attempt to address this issue, numerous scientists and researchers are working upon the development of Machine Learning models that shall identify the stage of depression of Twitter user from the users' public tweets and other activities on Twitter. This paper: (1) provides background on depression, use of Twitter for predictions and machine learning; (2) reviews previous studies that employed machine learning for identifying depression; and (3) attempts to guide to future work on the topic.

Keywords:-Machine Learning, Artificial Intelligence, Twitter data, Depression detection, Public Health

1. Introduction

Mental illness is among the most prevalent yet overlooked issues. According to World Health Organisation, around 20% of child and adolescent population and 23% of the total human population have one or more mental illness, making neuropsychiatric disorders the major cause of disability all over the world [1]. Among all, Depression is among the most prevalent mental health disorders. More than 300 million individuals or 4.4% of world's population is estimated to suffer from depression [2]. However, its prevalence varies by the WHO region and gender, from a minimum of 2.6% among the Western Pacific males to a maximum of 5.9% among the African females.

Despite being massively prevalent, it is observed that the many countries lack essential facilities to tackle depression. It is found that only 87% countries offer some primary care for mental health [3]. Effective treatments for depression reaches less than 10% of the sufferers globally [4]. While 76-85% sufferers from low- and middle-income countries receive no treatment. The scenario isn't much brighter in high-income countries where 35-50% patients with mental illness are deprived of treatment [5]. It is also found that 30% countries do not have any specialised program to deal with mental health. 28% nations do not allocate funds for mental health-related initiatives in their yearly budget [3]. All these facts highlight the need for modern techniques for identifying those suffering or at risk of depression.

Researchers have been continually working hard to find novel methods for the depression diagnosis. In a study in 1976, it was observed that 50% of patients were diagnosed with depression post 6 months of the first onset of the depressive episode [6]. In the year 1982, Oxman et al. showed that it is possible to classify patients as depressed and not-depressed by analysing the language of their speech [7]. Studies conducted by Brown et al. demonstrated that lack of social support and lower self-esteem are major factors associated with greater incidences of depression [8]. The year 2003 saw two different independent studies correlating depression with negative cognition. One predicted depression from negative cognitive biases in resolving ambiguous verbal information [9]. The other showed a method to predict the onset, number and duration of depressive episodes from negative cognition styles and contemplation of subjects in response to stress [10]. Work of Cloninger et al. highlighted the role of personality traits in the vulnerability of onset of depression in future [11]. All these works provided grounds for exploring additional means for diagnosis of depression by employing linguistics and social media mannerism.

Over time, the Internet became increasingly useful and witnessed the rise of social networking. With greater accessibility and availability of high-speed Internet, use of social networking became a trend among all age groups. Smartphone revolution made Internet access much more comfortable and also sharing content on social networking platforms. This increased the use of social networking. All these ultimately lead to the present scenario where numerous social networking platforms are essential and nearly indispensable part of our daily life. The social networking platforms are rich in content in the form of text, images, videos. Users express their views and opinions and share them with their connections on social media. This wealth of data attracted the attention of researchers finding means to reach out to huge fractions of the population. A study found similarities between users' website and their self-judged personalities [12]. This proved social networking data useful for providing insights.

In the meantime, the field of psycholinguistics also experienced advancement with the development of computational methods and applications. Pennebaker, Mehl and Niederhoffer used LIWC to detect the state of depression from text [13]. Later, studies showed the estimation of depression and neuroticism by application of Latent Dirichlet Allocation (LDA) on topics obtained from essays written by college students [14]. Such works made it possible to generate insights from text data available on social networking. Studies utilised Facebook status updates to detect depression [15]. Another study created a regression classifier by fitting Facebook status updates and personality survey results as features in a regression model to find the degree of depression [16].

It must be noted that much of the Facebook user data is private and accessible only to the user's connections. This constraint makes the useful wealth of data inaccessible. So, to develop a method to find depressed individuals in the population, the researchers turned to Twitter. According to Pew Research, nearly 24% of Internet users use Twitter [17]. It is also found that Twitter is more popular in 18-29 years age group [17]. User posts on Twitter are public unless opted otherwise. However, a study suggests privacy and stigmatization concerns may intimidate some users from sharing disease-related information on social media [18]. As an alternative, users' Internet usage patterns obtained from online logs can be analysed to detect depression [19].

Machine learning is a trending concept in which the program learns to classify or make predictions based on patterns found in the training dataset. During its training, the program attempts to find patterns in given data. If the training dataset is already annotated as per our choice, it's called Supervised Learning. In Supervised Learning, the program may learn to classify things or make predictions according to our annotations. Machine Learning is called Unsupervised Learning when the training dataset doesn't contain labels. In Unsupervised Learning, the program finds patterns in the given training data on its own and does future predictions or classifications according to its findings in the given training data. With enormous ability to model almost any data and make predictions at a superhuman pace, machine learning found itself employed in every single place possible. It is used for a variety of purposes like stock trading, predicting residential prices, reading MRI reports, in self-driving cars, etc. In the current paper, we shall see the use of Supervised Learning for diagnosis of depression.

2. Aim

This paper intends to:

(1) Provide background on depression, use of Twitter for predictions and machine learning;

(2) Review previous studies that employed machine learning for identifying depression; and

(3) Attempt to guide future work on the topic

3. Discussion

Numerous studies explored means to make predictions out of Twitter data. But, only a few successfully implemented machine learning algorithms to detect depression in Twitter users. In this section, we attempt to discuss and assess the methods adopted in the past studies.

A. Data Collection and Sampling

Different studies adopted different strategies for acquiring the necessary Twitter data. Most of them used Twitter API to fetch public tweets. However, attributes used differed as per strategy adopted. Some studies directly utilised the dataset generated for CLPsych 2015 Shared Tasks. The rest either: (1)conducted surveys to find suitably depressed as well as non-depressed individuals and fetched their public Twitter activity for

 https://doi.org/10.22362/ijcert/2018/v5/i2/v5i208

specified time duration with their permission; or (2)directly mined all public tweets in English language containing either word "depression" or some suitable

Conducting surveys possess another challenge – finding subjects. Finding prospecting subjects directly or online forums may be suitable only for smaller studies. De Choudhary et al. used Amazon's Machine Turk (Turk) interface to conduct large-scale studies on crowd workers [20]. They asked the crowd workers to take standardised clinical depression survey along with sharing their depression history, demographics and username of their public Twitter profile. The similar approach was later adopted by Reece et al. [21]. They discarded data samples where the crowd worker spent insufficient time in completing the survey. Also, auxiliary screening tests were included in the review to filter out those entries whose depression scores didn't correlate much across the scales [20].

TABLE 1. COMPARISON OF DATASET USED IN MAJOR STUDIES

Study	**Content**	**Size**
Park et al. 2012	1-week worth tweets prior and up to date of survey	69 users (Depressed = 23) 5,706 tweets
De Choudhary et al. 2013	All tweets posted from exactly 1 year before the survey date (normal) or first onset of depressive episode (for user who reported depression)	476 users (Depressed = 171) 2,157,992 tweets
Tsugawa et al	Normalised frequency of words used by 50 users in tweets	50 users 14,757 words
CLPsych 2015 Shared Tasks	3000 most recent public tweets per user of age- and gender-matched control users	1,148 users (Depressed = 327) 3438 tweets
Reece et al. 2017	All participants' Twitter posts to the most recent	204 users (Depressed = 105) 279,951 tweets

B. Choice of Depression Scale

Most of the studies used Centre for Epidemiological Studies Depression (CES-D) Scale for estimating the degree of depression in the subjects. CES-D is a 20-questions long questionnaire designed to measure the extent of depression in the general population [22]. The scale ranges from zero to sixty. Depending upon the CES-D score, the likelihood of depression may be: low (0-15), mild to moderate (16-22) or high (23-60) [23]. Park et al. chose 22 as the threshold CES-D score to improve specificity and false-positive diagnosis [24]. The same threshold was followed by the consecutive studies. Studies even used CES-D along with additional auxiliary screening like Beck Depression Inventory (BDI) [20, 21] and Kellner Symptom Questionnaire [21]. The individuals whose depression scores didn't correlate across the main CES-D scale and its auxiliary scales were filtered out. Tsugawa et al. employed Zung's Self-rating Depression scale in their study [25].

C. Feature Extraction

Most of the studies came up with their unique approach to feature extraction. Here, we discuss the feature extraction methods used by the major studies.

1) Park et al., 2012: They performed sentiment analysis on tweet text using Linguistic Inquiry Word Count (LIWC) [24]. LIWC is a text analysis program that categorises words into multiple psychologically meaningful categories and sub-categories. It returns scores based on the number of the word belonging to its categories and subcategories. Out of the obtained scores, they eliminated categories with very high multicollinearity by examining the bivariate correlation between independent variables and condition numbers. The remaining categories were accepted as features for the next step.

2) Tsugawa et al., 2013: This study attempted to implement Bag of Words approach. The frequencies of words used by users in tweets are used as the feature [25].

3) *De Choudhary et al., 2013:* In their study, De Choudhary et al. developed a set of dynamic parameters to characterise behavioural difference between depressed and non-depressed based on Twitter activity data [20]. They included the following features [20]:

(i) Volume, Reply, Retweet, Links, Question-centric and Insomnia index as measure of Engagement

 https://doi.org/10.22362/ijcert/2018/v5/i2/v5i208

(ii) Followers, Followees, Reciprocity, Prestige ratio, Graph density, Clustering coefficient, Two-hop neighbourhood, Embeddedness and Ego components as measures of Egocentric Social Graph

(iii) Positive affect (PA), Negative affect (NA), Activation and Dominance as measures of Emotion

(iv) 22 Linguistic Styles as in LIWC

(v) Depression lexicon and Antidepressant use as measure of Depression Language

These features can prove various facts available in literature related to behavioural cues of depression, like increased activity during night-time, greater usage of first-person pronouns, etc. [20]. The values of these features are collected daily from each user for one year. These features are used for constructing a time series per measure per user. Out of such time series, feature vectors are constructed for depression prediction framework. These feature vectors contain features like Mean frequency, Variance, Mean momentum, Entropy along with self-reported information like age, gender, education level and income of users. This gives 4 numbers per measure, i.e. 188 features in all for an individual user. Each feature vector is standardised to zero mean and unit variance.

4) *Nadeem et al., 2016:* Like Tsugawa et at 2013, this study attempts to implement Bag of Words technique on tweet text to identify depression [26]. As a feature to classifier model, they used the frequency of words in each tweet.

5) Reece et al., 2017: They utilised tweet count per user per day, word count per tweet, retweet, reply and results of labMT, LIWC and ANEW sentiment analysis as features for their machine learning models [21].

D. Machine Learning Models for Depression Identification

Among all studies, use of classifier to categorise the user as either "depressed" or "not depressed" seems to be a common practice. However, some early studies attempted to predict depression scores using multiple regression [22, 23].

For Supervised learning-based classifiers, the specified number of tweet selected randomly formed the training and testing datasets. Training dataset was mostly annotated manually. To avoid over fitting, De Choudhary et al. implemented Principal Component Analysis (PCA) [20]. In their study, Nadeem et al. evaluated the performance of four classifiers by implementing a Decision Tree, a Linear Support Vector Classifier, a Logistic Regression based classifier and a Naïve Bayes classifier [26]. Reeve et al. observed that Random Forest classifier worked the best with their data [21].

TABLE 2. COMPARISON OF MACHINE LEARNING OUTCOME OF MAJOR STUDIES

Experiment	**Algorithm**	**Result**
Park et al. 2012	Multiple Stepwise Regression	Depressed people are more likely to use negative words while referring to social words
Tsugawa et al. 2013	Multiple Regression	Medium correlation (about 0.45) between selected scale and predicted score
De Choudhary et al. 2013	Support Vector Machine	Accuracy ~ 70% Precision = 0.74
Nadeem et al. 2016	Naïve Bayes Linear SVM Logistic Regression Decision Tree	Accuracy = 86% (using unigram Naïve Bayes classifier)
Reece et al. 2017	Random Forest Hidden Markov Model	Only 1 false positive prediction for every 10 depression diagnoses

E. Direction for future studies

The major scope for future development lies in finding the novel technique for automated large-scale implementation of machine learning in the diagnosis of Depression. It shall be great to automate this and generate reports regularly. That may help to check the spread of Depression.As techniques of Natural Language Processing is new and still in development stage, we may expect the algorithms to get better in the future. We may also want the accuracy of these algorithms to improve further to be reliably used with large-scale data.

As against the two categories of classification in the current algorithms, it shall be good to have an additional category for those vulnerable to the onset of depression shortly, or for example those with a CES-D score between 16-22.The sentiment analytics programs used currently and in previous studies face issues due to the informal language used in social media posts (example: "tireddddd"). These issues significantly lower the performance of NLP tools and techniques developed for use on formal language [27]. We may hope to see new NLP algorithms with the ability to make sense out of such informal text being used for depression diagnosis on Twitter data.

References

[1] World Health Organization (WHO), "10 Facts on Mental Health," *World Health Organisation*, 2015. [Online]. Available: http://www.who.int/features/factfiles/mental_health/mental_health_facts/en/. [Accessed: 31-Dec-2017].

[2] World Health Organization, "Depression and other common mental disorders: global health estimates," 2017.

[3] R. Detels and C. C. Tan, "The Scope and Concerns of Public Health," *Oxford Textb. Glob. Public Heal.*, vol. 1, pp. 3–18, 2015.

[4] World Health Organization (WHO), "Depression (Fact Sheet)," *World Health Organisation*, 2017. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs369/en/. [Accessed: 31-Dec-2017].

[5] S. W. H. Assembly, "The global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level," *Sixty-fifth World Health Assembly*, no. May, pp. 2–5, 2012.

[6] G. Winokur, "Duration of Illness before Hospitalization (Onset) in the Affective Disorders," *Neuropsychobiology*, vol. 2, no. 2–3, pp. 87–93, 2008.

[7] T. E. Oxman, S. D. Rosenberg, and G. J. Tucker, "The language of paranoia," *Am. J. Psychol.*, 1982.

[8] G. W. Brown, B. Andrews, T. Harris, Z. Adler, and L. Bridge, "Social support, self-esteem and depression.," *Psychol. Med.*, vol. 16, no. 4, pp. 813–31, 1986.

[9] S. S. Rude, C. R. Valdez, S. Odom, and A. Ebrahimi, "Negative Cognitive Biases Predict Subsequent Depression," *Cognit. Ther. Res.*, vol. 27, no. 4, pp. 415–429, 2003.

[10] M. S. Robinson and L. B. Alloy, "Negative cognitive styles and stress-reactive rumination interact to predict depression: A prospective study," *Cognit. Ther. Res.*, vol. 27, no. 3, pp. 275–291, 2003.

[11] C. R. Cloninger, D. M. Svrakic, and T. R. Przybeck, "Can personality assessment predict future depression? A twelve-month follow-up of 631 subjects," *J. Affect. Disord.*, vol. 92, no. 1, pp. 35–44, 2006.

[12] S. Vazire and S. D. Gosling, "e-Perceptions: Personality impressions based on personal websites," *J. Pers. Soc. Psychol.*, vol. 87, no. 1, pp. 123–132, 2004.

[13] J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer, "Psychological Aspects of Natural Language Use: Our Words, Our Selves," *Annu. Rev. Psychol.*, vol. 54, no. 1, pp. 547–577, 2003.

[14] P. Resnik, A. Garron, and R. Resnik, "Using Topic Modeling to Improve Prediction of Neuroticism and Depression in College Students," *Proc. 2013 Conf. Empir. Methods Nat. Lang. Process.*, no. October, pp. 1348–1353, 2013.

[15] M. A. Moreno *et al.*, "Feeling bad on Facebook: Depression disclosures by college students on a social networking site," *Depress. Anxiety*, vol. 28, no. 6, pp. 447–455, 2011.

[16] H. A. Schwartz *et al.*, "Towards Assessing Changes in Degree of Depression through Facebook," *Proc. Work. Comput. Linguist. Clin. Psychol. From Linguist. Signal to Clin. Real.*, pp. 118–125, 2014.

[17] Shannon Greenwood, Andrew Perrin, and Maeve Duggan, "Demographics of Social Media Users in 2016 | Pew Research Center," *Pew Research*, 2016. [Online]. Available: http://www.pewinternet.org/2016/11/11/social-media-update-2016/. [Accessed: 01-Jan-2018].

[18] J. Mikal, S. Hurst, and M. Conway, "Ethical issues in using Twitter for population-level depression monitoring: A qualitative study," *BMC Med. Ethics*, vol. 17, no. 1, 2016.

[19] R. Kotikalapudi, S. Chellappan, F. Montgomery, D. Wunsch, and K. Lutzen, "Associating internet usage with depressive behaviour among college students," *IEEE Technol. Soc. Mag.*, vol. 31, no. 4, pp. 73–80, 2012.

[20] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting Depression via Social Media," *Proc. 7th Int. AAAI Conf. Weblogs Soc. Media*, vol. 2, pp. 128–138, 2013.

[21] A. G. Reece, A. J. Reagan, K. L. M. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, "Forecasting the onset and course of mental illness with Twitter data," *Sci. Rep.*, vol. 7, no. 1, pp. 1–11, 2017.

[22] T. K. Houston, L. A. Cooper, H. T. Vu, J. Kahn, J. Toser, and D. E. Ford, "Screening the Public for Depression Through the Internet,"

Psychiatr. Serv., vol. 52, no. 3, pp. 362–367, 2001.

[23] B. A. Husaini, J. A. Neff, J. B. Harrington, M. D. Hughes, and R. H. Stone, "Depression in rural communities: Validating the CES-D scale," *J. Community Psychol.*, vol. 8, no. 1, pp. 20–27, 1980.

[24] Park Minsu, Chiyoung Cha, and Meeyoung Cha, "Depressive Moods of Users Portrayed in Twitter," *ACM SIGKDD Work. Healthc. Informatics*, pp. 1–8, 2012.

[25] S. Tsugawa *et al.*, "On estimating depressive tendencies of Twitter users utilizing their tweet data," *Proc. - IEEE Virtual Real.*, 2013.

[26] M. Nadeem, M. Horn, and G. Coppersmith, "Identifying Depression on Twitter," *CoRR*, pp. 1–9, 2016.

[27] A. Ritter, S. Clark, Mausam, and O. Etzioni, "Named Entity Recognition in Tweets: An Experimental Study," *Proc. 2011 EMNLP (EMNLP 2011)*, pp. 1524–1534, 2011.