



A Multiclass Depression Detection in Social Media Based on Sentiment Analysis

89

Raza Ul Mustafa, Noman Ashraf, Fahad Shabbir Ahmed, Javed Ferzund, Basit Shahzad, and Alexander Gelbukh

Abstract

Depression is a common mental health disorder. Despite its high prevalence, the only way of diagnosing depression is through self-reporting. However, 70% of the patients would not consult doctors at an early stage of depression. Meanwhile people increasingly relying on social media for sharing emotions, and daily life activities thus helpful for detecting their mental health. Inspired by these a total of 179 depressive individuals selected from Twitter, who have reported depression and they are on medical treatment. A sample of their recent tweets collected ranges from (200 to 3200) tweets per person. From their tweets, we selected 100 most frequently used words using Term Frequency-Inverse Document Frequency (TF-IDF). Later, we used the 14 psychological attributes in Linguistic Inquiry and Word Count (LIWC) to classify these words into emotions. Moreover, weights were assigned to each word from happy to unhappy after classification by LIWC and trained machine learning classifiers to classify the users into three classes of depression High, Medium, and Low. According to our study, better features selections and their combination will help to improve performance and accuracy of classifiers.

Keywords

Depression · Twitter · Social media · Deep learning · Machine learning · Neural network

89.1 Introduction

The World Health Organization (WHO) predicts that by the year 2030 there will be 322 million people estimated to be suffering from depression [1]. Depression leads to mood disruption, uncertainty, loss of interest, tiredness, and physical issues [2]. Despite this, there is no laboratory test for diagnosing this type of illness. The subjects in this study identified their mental illness either by self-diagnosing or by being diagnosed by friends or family members. Symptoms expressed by a depressed person are anxiety, restlessness, hopelessness, and misery, which can frequently lead to thoughts of self-harm and suicide. People suffering from depression need continuous support from their family, friends, relatives and neighbors [3].

With the development of Internet usage, many people have started sharing their personal feelings and mental illness on social platforms. Their activities on Social Media (SM) have encouraged many researchers to prevent this mental illness and detect its early stage before severe consequences. Many studies have identified these individuals from their proposed methods using Natural Language Processing (NLP) techniques [4]. Even with recent significant progress in the field, the challenges are still there. This research aims to use a different methodology for the early detection of depressive individuals. We considered diagnosed depressive users from Twitter for analysis and classified them into three classes High (H), Medium (M), Low (L) depress stage. We selected Twitter for its simplicity for the data collection on a certain topic. The most significant conversations are centered around

R. U. Mustafa (✉) · J. Ferzund
Department of Computer Science, COMSATS University Islamabad,
Islamabad, Pakistan
e-mail: jferzund@ciitsahawal.edu.pk

N. Ashraf · A. Gelbukh
IPN - Computing Research Center, Mexico City, Mexico
e-mail: nomanashraf@sagitario.cic.ipn.mx; gelbukh@gelbukh.com

F. S. Ahmed
Yale School of Medicine, New Haven, CT, USA

B. Shahzad
Department of Software Engineering, National University of Modern
Languages NUML, Islamabad, Pakistan

a hashtag, which helps to detect people with similar interests. First, we considered a set of a dataset from twitter discussing depression in their tweets. Then manually selected 179 depressive users who have tweeted about their mental illness and they are on treatment. Later, we collected their recent tweets and extracted word frequency. Regarding the correlation, we focused on the LIWC dictionary and classified collected word frequency into 14 psychological attributes. Finally, we assigned weights to each word classified by LIWC based on a scale of happiness ranging from unhappy to happy (1–9) [5] a proposed method for the classification of depressive users into three classes. For classification, we used Neural Network (NN), Support Vector Machine (SVM), Random Forests (RF) and 1D Convolutional Neural Networks (1DCNN). A suggested classification approach can be used to detect similar patterns on Twitter for timely handling of severe consequences. Our study has three main contributions. (1) A proposed method for the classification of documents such as tweets of 179 diagnosed users as 179 documents, and classified them into three classes of depression *H*, *M*, *L*. (2) We investigate and report the performance of several Machine Learning (ML) classifiers commonly used in NLP tasks, in particular, to detect mental disorder. iii) Finally, we have naturally annotated data that we have separated from normal users.

The rest of the paper is organized as follows. In Sect. 89.2, we discussed the related work. In Sect. 89.3, we introduced the methodology. Evaluation of the proposed approach and results are discussed in Sect. 89.4. Finally, a conclusion is drawn in Sect. 89.5.

89.2 Related Work

Depression is a severe public health challenge [6–8]. SM has been used for extracting psychological attributes from the text posted by its users. Billing and Moos [9] studied the role of stress in depression. The research provides strong evidence that SM environments contain a crucial source of information for dealing with depressive individuals. Choudhury et al. [10] used tweets to engage with the problem. They developed a statistical model that may be used by healthcare agencies for the detection of depressive users on SM before the illness progresses towards a serious level. The attributes used in that study were user social activity, negative effects in tweets, highly clustered ego network, and evidence of suicidal thoughts in the text. Similarly, Moreno et al. [11] demonstrated that Facebook status updates could contain symptoms of major depressive episodes [12, 13]. Studies to date have improved the efficiency of the statistical model and conducted surveys on homogeneous samples of individuals [14, 15]. However, the gap of finding new methods for the detection of depression from SM and to

increase the efficiency of already proposed methods are still there. Our study analyzed diagnosed depressive individuals from Twitter. Later, we used the potential of LIWC to detect emotions from text and classified the documents into *H*, *M*, and *L* classes of depression.

89.3 Methodology

We used Twitter Developer, Application Programming Interface (API) [16], for public data. We developed an application that fetches data using hashtags, query strings, and specific user data. We started collecting tweets in 2016 and continued until July 2019. We have 1,56,511 tweets that contain 19,89,890 words. We converted the raw tweets into useful text. The first step in this approach is pre-processing. *Pre-processing* is a way of cleaning data. It involves data transformation, instance selection, normalization, and feature extraction. We removed unwanted text from the data, i.e., stops words, links, punctuation marks, and special characters. Thus, the representation of data in a high-quality format is the first and foremost step before running any analysis. Then we converted sentences into tokens a process called tokenization. *Tokenization* is the process of breaking a large string of data into smaller units that may include phrases and words often called tokens. These tokens are used to conduct quality analysis of the data. Of the two approaches to tokenization (phrase and word tokenization), word-level tokenization is considered more effective due to the resulting statistical significance [17]. In this process, for instance, the sentence ‘previous depressions triggered by coming out bad relationship or even worse relationship’ was separated into the tokens ‘previous’, ‘depressions’, ‘triggered’, ‘by’, ‘coming’, ‘out’, ‘bad’, ‘relationship’, ‘or’, ‘even’, ‘worse’, ‘relationship’, etc. The algorithms used to tokenize a sentence separates the tokens based on the spaces between words and the built-in dictionary.

After tokenization, we assigned weights to the tokens based on their relative effectiveness. This process is known as feature weighting. A standard function to compute the weights is TF-IDF [18]. The TF-IDF scheme is based on two parts: term frequency (TF) and inverse document frequency (IDF). TF is used to count the tokens represented in a document. It gives a complete count of term occurrences. One hundred most frequently used words using the TF-IDF collected from 179 users. The total number of words collected were 17,900. Later, we used LIWC which classified the words into 14 psychological attributes such as social, family, friends, religion, death, feel, health, sexual, risk, positive emotions, negative emotions, anxiety, anger and sad.

Finally, we assigned weights 1 to each word classified by LIWC based on a scale of happiness ranging from unhappy

to happy (1–9) for further categorical classification such as H, M, L users documents. Repetition of words was removed from the set of 17,900 words that makes 96 unique words for 179 users. After sorting the words in ascending order the categories based on weights are (1–3.9) = H, (4–6.9) = M, and (7–9) = L. A *H* depressive user is more concerned in his/her interests, feeling worthless or guilty, difficulty with decision-making, and thoughts of suicide. These users have used words such as ‘sh*t’, ‘panic’, ‘guilty’, ‘suicide’, ‘killing’, ‘dead’, and ‘anxiety’. Users with Premenstrual Dysphoric Disorder (PMDD) have symptoms of anxiety, fatigue, irritation, and mood swings. We classified words of this class as *M*. The words most frequently used by this class of depressed users are ‘valentine’, ‘s*x’, ‘friends’, ‘soul’, ‘religion’, and ‘f**king’. Some signs of fatigue, believing that someone is harming you, seasonal affective disorder (SAD), situational depression, and a typical depression were categorized as *L*. The words used by such users include ‘bless’, ‘lover’, ‘heaven’, and ‘passion’ etc.

A string has made in such a way if word found in the document of respective user tweets then it is replaced by 1 otherwise 0 making a string of (0,1) of length 96 for each user. The algorithm 1 has used for such purpose.

Algorithm 1 Multi-class depression detection

Input: sw = string words, iw = input words, sd = string document, ww = word weight, and A = matrix

Output: Depression class of the tweet in the form of H, M and L

1. For $I \leftarrow 0$ to n
2. do $A[0,i] \leftarrow sw_i // 96$ words
3. do $A[1,i] \leftarrow 0 //$ initialize all with zeros
4. For $i \leftarrow 0$ to n
5. input iw_i
6. If ($iw_i == x_i$)
7. Then $A[1,i] \leftarrow 1$
8. $H \leftarrow 0, M \leftarrow 0, L \leftarrow 0$
9. For $j \leftarrow 0$ to n
10. If $ww[j] \geq 1$ and $ww[j] \leq 3.9$
11. Then $H \leftarrow H+1$
12. Else if $ww[j] = 4$ to 6.9
13. Then $M \leftarrow M+1$
14. Else $L \leftarrow L+1$
15. If $H > M$ and $H > L$
16. Then $MaxVal \leftarrow H$
17. Else if $M > H$ and $M > L$
18. Then $MaxVal \leftarrow M$
19. Else $MaxVal \leftarrow L$

Where iw refers to input words, sd is used for the document string, which is usually a combination of 200 to 3200 tweets per user, ww is the weight assigned to each word, and A denotes the matrix. The function takes iw , sd , ww , and

matrix A . The matrix contains two rows, the first is dedicated to unique string words and the second is reserved for the occurrence flag. In the first row, we have initialized 96 string words. The corresponding occurrence flag is initially set to 0. We classified words in such a way that each input word is searched for in each user’s tweet repository. The corresponding occurrence flag is set to 1 if the input word is located in each user’s tweet text. Finally, we made a document that has combinations of 0,1 for 179 distinct users. On line 8 of the above code, H, M, and L counters are initialized with value 0. The third loop, at line 9, contains a series of if statements to maintain the count of words that belong to each of the intensity levels, i.e., H, M or L. Thereafter, lines 15 to 19 are used to determine which intensity level has the highest count among the three. Here, the maximum value is the total number of words used by a depressed person from each of the H, M, and L classes.

We used Keras, a Python library for experiments that wraps the efficient numerical libraries Theano and TensorFlow. Theano is open-source numerical computational library, very valuable for fast numerical computations. We adopted the one-vs-all technique to differentiate the different level of depressed users. First High instances classified from Medium and Low, in the second step, Medium instances classified from High and Low, and finally Low separated from High and Medium.

89.4 Results and Discussion

We used 1-DCNN, NN, SVM, RF to evaluate the appropriateness of our data representation and to train models. The performance of selected classifiers are listed in Table 89.1. Where H, M, L presents comparison. Three evaluation measures (precision, recall and f-measure) are used to evaluate the performance of classifiers. The mathematical definition of these measures with respect to a positive class is defined in Eqs. (89.1), (89.2), and (89.3) respectively.

$$\text{Recall } (R) = \frac{\text{no of CPP}}{\text{no of PE}} \dots \quad (89.1)$$

$$\text{Precision} = \frac{\text{no of CPP}}{\text{no of PP}} \dots \quad (89.2)$$

$$F - \text{score} = \frac{2 \times P \times R}{P + R} \dots \quad (89.3)$$

In Eqs. (89.1) and (89.2), CPP, PE and PP stand for correct positive predictions, positive examples and positive predictions respectively.

Table 89.1 Overall area under curve (AUC), precision, recall and f-score

Model	Class	AUC	Precision	Recall	F-Score
1DCNN	H vs. M L	0.91	0.92	0.86	0.89
1DCNN	M vs. H L	0.83	0.85	0.54	0.66
1DCNN	L vs. H M	0.86	0.93	0.78	0.85
NN	H vs. M L	0.89	0.78	0.78	0.78
NN	M vs. H L	0.89	0.83	0.83	0.83
NN	L vs. H M	0.88	0.83	0.83	0.83
SVM	H vs. M L	0.86	0.77	0.93	0.84
SVM	M vs. H L	0.91	0.90	0.81	0.85
SVM	L vs. H M	0.86	0.93	0.78	0.85
RF	H vs. M L	0.80	0.90	0.60	0.72
RF	M vs. H L	0.83	0.85	0.54	0.66
RF	L vs. H M	0.83	0.84	0.84	0.84

89.5 Conclusion

In this study, we have extracted useful information from the tweets posted by diagnosed depressed individuals on Twitter. The identification and classification of word selections in the classes of H, M, and L depression constitute major findings. We utilized the top 100 words used by depressive users to build a classifier that has classified users with an accuracy of 91%. In the future, we are interested in extracting further, more detailed information from depressive Twitter user Tweets, such as emojis, pictures, gifs that are embedded in their writings.

References

- Murray, C.J., Lopez, A.D.: Alternative projections of mortality and disability by cause 1990–2020: global burden of disease study. *Lancet*. **349**(9064), 1498–1504 (1997)
- Hur, N.W., Kim, H.C., Waite, L., Youm, Y.: Is the relationship between depression and c reactive protein level moderated by social support in elderly?-Korean Social Life, Health, and Aging Project (KSHAP). *Psychiatry Investig.* **15**(1), 24 (2018)
- Liu, A., Liu, B., Lee, D., Weissman, M., Posner, J., Cha, J., Yoo, S.: Machine learning aided prediction of family history of depression. In: 2017 New York Scientific Data Summit (NYSDDS), pp. 1–4. IEEE (2017)
- Tadesse, M.M., Lin, H., Xu, B., Yang, L.: Detection of depression-related posts in reddit social media forum. *IEEE Access*. **7**, 44883–44893 (2019)
- Salton, G., Buckley, C.: Term-weighting approaches in automatic text re-trieval. *Inf. Process. Manag.* **24**(5), 513–523 (1988)
- Calvo, R.A., Milne, D.N., Hussain, M.S., Christensen, H.: Natural language processing in mental health applications using non-clinical texts. *Nat. Lang. Eng.* **23**(5), 649–685 (2017)
- Khalil, R.M., Al-Jumaily, A.: Machine learning based prediction of depression among type 2 diabetic patients. In: 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), pp. 1–5. IEEE (2017)
- Shen, G., et al.: Depression detection via harvesting social media: a multimodal dictionary learning solution. In: IJCAI (2017)
- Billings, A.G., Moos, R.H.: Coping, stress, and social resources among adults with unipolar depression. *J. Pers. Soc. Psychol.* **46**(4), 877 (1984)
- De Choudhury, M., Gamon, M., Counts, S., Horvitz, E.: Predicting depression via social media. In: Seventh International AAAI Conference on Weblogs and Social Media (2013)
- Moreno, M.A., Jelenchick, L.A., Egan, K.G., Cox, E., Young, H., Gannon, K.E., Becker, T.: Feeling bad on Facebook: depression disclosures by college students on a social networking site. *Depress. Anxiety*. **28**(6), 447–455 (2011)
- Park, M., Cha, C., Cha, M.: Depressive moods of users portrayed in twitter. In: Proceedings of the ACM SIGKDD Workshop on Healthcare Informatics (HI-KDD), vol. 2012, pp. 1–8 (2012)
- Deshpande, M., Rao, V.: Depression detection using emotion artificial intelligence. In: 2017 International Conference on Intelligent Sustainable Systems (ICISS), pp. 858–862. IEEE (2017)
- Seah, J.H., Shim, K.J.: Data mining approach to the detection of suicide in social media: a case study of Singapore. In: 2018 IEEE International Conference on Big Data (Big Data), pp. 5442–5444. IEEE (2018)
- Dieris-Hirche, J., Bottel, L., Bielefeld, M., Steinbuechel, T., Khyayan, A., Dieris, B., te Wildt, B.: Media use and internet addiction in adult depression: a case-control study. *Comput. Hum. Behav.* **68**, 96–103 (2017)
- Twitter Developer Application Programming API. <https://developer.twitter.com/>. Online. Accessed 1 July 2016
- Sebastiani, F.: Machine learning in automated text categorization. *ACM Comput. Surv.* **34**(1), 1–47 (2002)
- Ramos, J.: Using TF-IDF to determine word relevance in document queries. In: Proceedings of the First Instructional Conference on Machine Learning, vol. 242 (2003)