

Survey on Detection of Mental Illnesses by Analysing Twitter Data

Aksharaa Sundarrajan*, M. Aneesha

Student, SRM Institute of Science and Technology

*Corresponding Author Email: aksharaasundarrajan185@gmail.com

Abstract

Mental illnesses are serious problems that place a burden on individuals, their families and on society in general. Although their symptoms have been known for several years, accurate and quick diagnoses remain a challenge. Inaccurate or delayed diagnoses result in increased frequency and severity of mood episodes, and reduce the benefits of treatment. In this survey paper, we review papers that leverage data from social media and design predictive models. These models utilize patterns of speech and life features of various subjects to determine the onset period of bipolar disorder. This is done by studying the patients, their behaviour, moods and sleeping patterns, and then effectively mapping these features to detect whether they are currently in a prodromal phase before a mood episode or not.

Keywords: Bipolar Disorder Detection, Prodromal phase, Sentiment Analysis, Emotion Analysis, Social Media, Mental Disorder.

1. Introduction

Twitter and other social media platforms, are being increasingly used by several people to share their thoughts and opinions with the world. The posts on these sites are made in a naturalistic environment in the course of daily activities. Additionally, people suffering from mental health problems often withdraw from social situations and activities.

Depleted self-esteem, minimal energy, pain with no legible cause and loss of interest in usually enjoyable activities are all major symptoms for depressive disorders. Six out of ten people who commit suicide have had depression or similar mood disorders.

The next mental disorder we tackle in this paper is Post-traumatic stress disorder (PTSD). PTSD develops post a person experiencing a traumatic incident. These incidents may include warfare, traffic accidents, sexual assault, or other threats to life amongst others. Symptoms for PTSD include disturbing thoughts, destructive feelings or attempts to avoid trauma-related cues, dreams related to the incident, mental or physical distress to trauma-related cues, alterations in how a person thinks and feels, and mostly an increase in the fight-or-flight response.

Bipolar disorder causes periods of depression and periods of elevated mood. A period of depression is called manic depression and a period of elevated mood is called hypomania.

The three disorders listed above are common conditions ranked by the World Health Organisation as amongst the top ten causes of global disability amongst adults. Individuals suffering from these disorders are known to make decisions poorly with little regard to the consequences. Suicide possibility is very high in those suffering from these diseases at higher than 15 percent. Self-harm is known to transpire at greater than 30–40 percent. Treatment commonly includes counselling and medication.

To understand the burden of the disease properly, one needs to understand the delay between the initial occurrences of the symptoms to the treatment being given. This could take as long as 9 years[3] (for Bipolar disorder) and is much higher for Depression. There are plenty of physical symptoms that often come with depression. This is perhaps why non-psychiatrist physicians often can't recognize or treat depression in contrast to psychiatric physicians. Research states that about two-thirds of cases presented to non-psychiatrist physicians are not diagnosed rightly, but these numbers have reduced in more recent studies.

Early detection of the first symptoms of a relapse is extremely important in preventing a full blown episode. According to Sierra et al. [8], the prodrome is nothing but the time gap between the most initial indication of the disease and when the indications are most heightened.

Early curative intervention- can stop the initial stages from turning into full blown mood episodes. According to Swann et al. [14], the patient responds to diagnosis much better during the starting stages of the disorder. Prodrome identification has three parts: recognition that one has a mental disorder; agreeing to proper care; and understanding the pathological implications. Good insight about the same helps the diagnosed understand their situation and functioning much better [8]. So it can be extremely beneficial to the patient if the disorder is detected in its earlier stages.

2. Psychiatric Evaluation of the Disorder Through Time

1) Depression: According to Fava et al. [15], Hays et al. examined and found these four indicator patterns: (i) sudden-onset depressions (ii) gradual-onset depressions; (iii) neurotic-onset depressions, (iv) 'fluctuating onset' depressions. Hopkinson took interviews of 43 patients suffering from 'depressive psychosis' around 30% showed a prodrome characterized by 'stress and slight anxiety', and other symptoms like impaired concentration and indecision. Family members were asked to observe the patient and to provide additional information. Fifteen subjects reported

experiencing at least one prodromal symptom prior to the onset of the depressed mood. Generalized anxiety was present in 13 cases and irritability in nine. Other common symptoms were fatigue, initial delayed insomnia and impaired work initiative.

2) Post Traumatic Stress Disorder: Meta-analysis indicates that roughly 25% of PTSD cases present with delayed onset, were noted after at least 6 months of the instigating damaging event. In a military population, Grieger et al. [16] found that the most of the subjects who tested PTSD-positive seven months after serious combat injury did not meet diagnostic threshold for testing positive at one month post-injury. In cases of PTSD followed by mild traumatic brain injury (TBI), the percentage of cases having delayed onset can be higher. Bryant et al. found that of those subjects meeting PTSD criteria at 2 years after a TBI, 44.1% gave no account of PTSD at 3 months. The factors contributing to delayed-onset PTSD in the absence of mild TBI are not completely understood.

3) Bipolar Disorder: In the early days, programs present for the recognition and regulation of hypomania episodes were based on individual prodrome evaluation and sporadic use of lithium. In 1999, Baldesserani et al. and Hirschfield (2003) et al. found the gap between onset of illness and proper lithium treatment to average at about 9-10 years [11]. In 2005, Kennedy et al. found that the occurrence of the first manic episode in bipolar I disorder was maximum in the 21 to 25 year age band, it steadily diminished from 25 to 40 years, and made a second smaller peak from 40 to 55 years [14]. This shows that psychiatrists wait a long time before they diagnose their patients. This is probably because the traditional frontline treatment for BD is medication and clinicians may be wary to treat patients before a definitive diagnosis is made.

3. Sentiment Analysis in Twitter

A. Why use twitter

Twitter is one of the most popularly used social media platform that lets us have access to public accounts more easily. Tweets are casual messages that are limited to 140 characters of text which are packed with emotions and sentiments users feel throughout the day. They sometimes express the user's views or thoughts about different subjects. To study the emotions of the users, we can weigh the emotional sentiment of the tweet. Sentiment classification is of large scope and a lot of advanced work is being carried on in the field.

Often people who suffer from various health conditions consider themselves unfit to go out and express emotions physically. So they take to social blogging sites like twitter to express their thoughts and observations. Patients prefer using Twitter as it contains several community portals and support psychotherapists which means help is available at anytime, anywhere. Patients discuss about their life freely on twitter which enables a huge insight for us into their life patterns and thought process.

The omnipresence of social media gives us a great chance to improve the quality of data available to doctors and researchers, ensuring a more knowledgeable and a more capable mental health field. Analysing mental health on publicly available Twitter data, can show us how meticulous implementation of simple natural language processing techniques can give us intuition into specific disorders as well as mental health writ large.

B. Collecting Patient data

The range of papers used in this section are very varied and are used for various purposes. Some for detecting various mental health problems and others for classifying emotions.

[2], Coppersmith et al. use the Twitter REST API to gather all public data available on Twitter. They gathered data from users who had self-identified as being diagnosed with some mental health issue X.

[6] Chang et al. use the method of subconscious crowd sourcing to collect their data. They first collected community support pages

for the particular mental illness. After collecting adequate number pages, they utilized the Twitter REST API to obtain the followers of the pages. The followers' accounts then became their main pool from which they chose both patient and expert candidates. This was done by looking at their profile description and selecting self-reported patients and experts. They were then manually labelled them into three categories: 1) Patient, 2) Expert, any professional that can treat mental illness, 3) Non-related, any user who doesn't fall in any of the categories mentioned above. Lastly, after having obtained the final list of patients and experts, they retrieved 3200 of their most recent tweets using the Twitter REST API.

[3] Choudhury et al. used a 2 part approach to collect data from mothers of new-borns to study post-partum depression. First, mothers were selected by searching for keywords associated with the announcements of their baby's birth on Twitter. Next the job of manual verification was given to Amazon's Mechanical Turk Interface to design human intelligence tasks (HITS) which would help improve the reliability of the data.

[7] Park et al. searched the Twitter API for the keyword 'depression' and collected tweets over a 2 month period in 2009 that were in English and published in the United States.

C. Pre-processing the data

Pre-processing of data is an important part of the data mining process. Analysing data that has not been cleaned and screened can result in misleading results. Hence it is very important to ensure that the data is of top quality before analysing it.

[12] Huang et al. pre-process their data by: Removing noisy data by adding filters like 1) time of diagnosis identification (maybe in days, months, years). 2) Language (English only). 3) Active user filter (should have tweeted a minimum of 100 times). 4) Not selecting users whose accounts contain more than 50% of hyperlinks. 5) Time and Location information (if available). Time and location information are extracted here to take into account when the tweet was sent out and adjust the time according to time zones respectively.

[2], [5] Coppersmith et al. remove noisy data by: ensuring that the selected usernames have 75% English tweets and that each account has a minimum of 25 tweets.

[4] Go et al. pre-process by: 1) Removing emoticons from the tweets 2) Using equivalence class tokens (username and URL) in place of all words starting with @ symbol and http respectively. 3) Removing letters occurring more than 2 times in a row and replacing them with two occurrences. 4) Removing repeated tweets. 5) Removing retweets. 6) Removing tweets containing both positive and negative emoticons.

[7] Park et al. pre-process by: Selecting a random set of 1000 tweets (from the 2 month period selected), parsing through them manually and choosing viable candidates for the study.

Apart from being an important part of the data mining process, pre-processing has an important and substantial role in each of these cases. Huang et al. [12] need a large amount of textual data to analyse the tweets and sentiments hence they only extracted data from Active users and those users whose accounts contain less than 50% hyperlinks. They also require time and location data to study the sleeping patterns of the subjects since insomnia is a symptom of Bipolar Disorder. Coppersmith et al. [2], [5] require their data to be free of noise like mentions, hyperlinks, etc. to analyse the emotions and moods of their subjects. Go et al. remove tweets containing both positive and negative emotions because their study mainly focusses on classifying tweets into positive negative and neutral, so introducing tweets having contrasting emotions would confuse the classifier. Park et al. wanted to study the accounts of people who were depressed and found it best to take a random selection of tweets to avoid biases.

D. Classifying emotions and sentiments

To diagnose a person with a mental illness, we need to be able to classify their moods and emotions first so that we can check for

symptoms. Hence it is essential that we classify the tweets and detect the various emotions felt by the person.

[4] Go et al.'s work helped classify tweets into positive, negative and neutral. They use various machine learning algorithms like Maximum Entropy, Naïve Bayes and Support Vector Machines, and check the performance of each algorithm in classifying the tweets into positive negative and neutral emotions.

[1] Argueta et al. prepared the segmented and normalized data (pre-processed data) to extract patterns of emotion. This was done using a graph based approach. The precedent obtained is used to find the sentiments expressed in a microblog post. Algorithms that require very less supervision and are applicable to Plutchik's wheel of emotions are presented. The paper also has commendable capacity to explore new languages.

[12] Huang et al. and [6] Chang et al. used a combination of [4] Go et al. and [1] Argueta et al.'s work.

[3] Choudhury et al. consider four emotional states for their users: positive effect (PA), negative effect (NA), activation and dominance. To detect PA and NA they use LIWC lexicon. They used the ANEW lexicon for calculating activation and dominance. The papers cited above used various algorithms and methods to map the emotions of twitter users. Go et al.'s [4] approach was to first train the dataset and then test it on a random collection of tweets which showed high accuracy. Argueta et al. [1] attempted to classify the data by segmenting it, using the segmented data to create a graph, reducing the graph and finally extracting emotions from the graph. Choudhury et al.[3] mainly wanted to look at the moods of the subjects, the degree of physical intensity in an emotion (activation) and the degree of control in an emotion (dominance). They needed to know this so that they could then detect energy levels of the patients and also some symptoms commonly faced by clinically depressed people.

E. Detecting symptoms of the prodrome

As discussed before, people suffering from mental health disorders typically display some symptoms before being diagnosed in their prodromal phase. Some of the symptoms for detecting bipolar disorder in its prodromal phase are: insomnia, over talkativeness, euphoria, severe mood swings, etc. Several papers have found unique approaches to detect these symptoms for various mental disorders. Some of them are:

[2] Coppersmith et al. study the Twitter data of patients of PTSD, Seasonal Affective disorder (SAD), Bipolar disorder (BD) and depression. After collecting the pre-processed data, they quantified various attributes using various automated methods. They then extracted these features for subsequent training of the model. They used the LIWC lexicon for gathering data about the position from the diagnosed person's writing. They extracted and studied the tweet rate difference, social features, sleep patterns and exercise patterns of the subjects. They also categorised the tweets into positive, negative and neutral classes.

[6] Chang et al. understood that there are several behavioural and linguistic features that are important for BD or Borderline Personality disorder (BPD) detection. TF-IDF was used to capture the words most frequently used by the patients, LIWC was used to consider the psychological terms most frequently used by the patients. Pattern of life features revealed the emotional patterns and behavioural tendency of users by finding out the polarity, emotions and interactions they make online. The authors calculated emotional scores of each tweet by using the classifier

suggested by [1] Argueta et al., and transformed their 8 emotions into 8 emotional scores. They also predicted the age and gender of the person. Using all of this data, they were able to study the user and detect prodromal symptoms.

[12] Huang et al.'s work mainly concentrates on people suffering from BD. Their approach can be divided into 2 parts. 1) Psychological feature extraction and 2) Phonological feature extraction. They use an approach very similar to that of [6] Chang et al. (mentioned above) for the part of their model that dealt with psychological feature extraction. For the phonological feature extraction part, they first converted all the tweets into IPA format, then they created a dictionary for each phoneme and assigned an energy value for each proceeding to calculate the energy value of each tweet. 3) The energy level results were then calculated after normalization.

d) [3] Choudhury et al. checked the users online engagement and created an ego-centric graph based on the data available. They then took into account the findings from the emotion classification data and characterized the linguistic style in posts by users. They also looked into the time of the posts made by users to understand their sleeping patterns and made an aggregation of their behaviour. The best way to diagnose a mentally ill patient is by studying his moods, and pattern of life. Coppersmith et al. [2] analysed the talking patterns of patients, their activities, sleep patterns, twitter activeness and social features. Chang et al. [6] added on to the work of Coppersmith by studying the mood polarities, emotion scores and delving deeper into the study of the social circles of the patients. They also took into account the age and gender of the subject as according to several studies, they affect the disorder a person is diagnosed with. Huang et al. [12] added on to the work of Chang et al. [6] by extracting the phonological features of the subject to get their energy level at the time of typing and posting each tweet.

F. Model Training and Cross-Validation

This section deals with applying algorithms and classifiers to the datasets so as to train the model to able to detect patterns in the data. Cross validation methods are used to correct the model and make it free of biases.

[12] Huang et al. used a collection of supervised machine learning approaches and found the most efficient. They use Random Forest Classification to train their model. RFC is capable of performing regression tasks as well as classifying tasks. They also use 10-fold cross validation to check their model's accuracy and precision.

[5]Coppersmith et al.: They tried several classifiers and checked their accuracy by measuring how efficient the system is in differentiating between diagnosed users and normal users.. If the accuracy in detecting the difference is high, the classifier is selected. They also perform leave-one-out cross validation for evaluation of the performance of the classifiers.

[3] Choudhury et al., leveraged the various unique qualities of the subjects to build an SVM classifier that could forecast before the reported beginning of depression of the subject, their likelihood of depression.

G. Summary Table

Sr No.	Title	Purpose of Paper	Accuracy/Findings	Limitation
1	Detection of the Prodromal Phase of Bipolar Disorder from Psychological and Phonological Aspects in Social Media [12]	Detecting the prodromal phase of bipolar in an efficient and reliable manner	Applying BDPLF and Phonological feature extraction on a time period of 3 months – 98.3%	The symptoms are often so similar for various diseases like depression, BD etc that the psychologist needs to interview the patient before a consensus diagnosis can be given
2	Quantifying Mental Health Signals in Twitter [2]	Presenting a normal method for data gathering of mental illnesses quickly and cheaply, with focus on PTSD, BD, Depression and SAD	Precision at false alarm (10&20% respectively) BD – 64%, 82% Depression -48%, 68% PTSD – 67%, 81% SAD – 42%, 65%	The people who were willing to reveal to the public their mental disorder, may not be a precise representation of the entire population of people diagnosed with that mental disorder.
3	Subconscious Crowdsourcing: A Feasible Data Collection Mechanism for Mental Disorder Detection on Social Media [6]	Applying subconscious crowdsourcing to gather rich data from Twitter and use it to distinguish patients from non-patients	For BD PLF - 72% ,LIWC- 22%, TF-IDF- 22% For BPD TF-IDF- 0%, LIWC- 48%, PLF – 82%	Relied a lot on manual data collection.
4	Predicting Depression via Social Media [3]	Leveraging the behavioural cues to build a statistical classifier that gives an evaluation of the possibility of depression before the reported diagnosis	Accuracy – around 70% Precision – 74%	Relied a lot on manual data collection.
5	Depressive Moods of Users Portrayed in Twitter [7]	Providing evidences that online social networks provide meaningful data for capturing depressive moods on twitter	People with regular jobs were more likely to not be depressed ($p<0.05$) Participants with college degrees were less depressed than people with lower education ($p<0.05$)	Did not find gender or age to play a significant role in determining depression
6	Measuring Post Traumatic Stress Disorder in Twitter [5]	Basically to detect post traumatic stress disorder in the military individuals of the US by building a PTSD classifier	1%- extra percentage of PTSD tweets in military areas 0.7%- more PTSD tweets in deployment regions	Other classifiers that can improve accuracy like geographical ones should be used or for a more specific group of users

4. Conclusion

In this survey paper, we have discussed the various approaches for mental disorder detection using data from Twitter. We discussed detecting mental disorders in the prodromal phase and otherwise using sentiment analysis and feature extraction. These symptoms are taken into account and tracked by studying the subject's Twitter data and general pattern of life. These studies have been carried out for various mental disorders. Several methods and algorithms have been used to detect and classify the users suffering from these disorders. There are some caveats to all the papers discussed which are listed in the limitations column of the results table but, all these papers have made great strides in detecting mental disorders at an early stage which can be very helpful for the patients.

5. Acknowledgement

We would first like to thank our research advisor Mr. Karthick Nanmaran of the Computer Science and Engineering Department at SRM University for his support and guidance.

References

- [1] C. Argueta, E. Saravia, and Y.-S. Chen. Unsupervised graph-based patterns extraction for emotion classification. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, pages 336–341. ACM, 2015.
- [2] G. Coppersmith, M. Dredze, and C. Harman. Quantifying mental health signals in twitter. In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pages 51–60, 2014.
- [3] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz. Predicting depression via social media. In ICWSM, page 2, 2013.
- [4] A. Go, R. Bhayani, and L. Huang. Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1:12, 2009.
- [5] G. A. C. C. T. Harman and M. H. Dredze. Measuring post traumatic stress disorder in twitter. In Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM).
- [6] Chun-Hao Chang, Elvis Saravia, and Yi-Shin Chen. 2016. Subconscious Crowdsourcing: A feasible data collection mechanism for mental disorder detection on social media. In Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on. IEEE, 374–379.
- [7] M. Park, C. Cha, and M. Cha. Depressive moods of users portrayed in twitter. In Proceedings of the ACM SIGKDD Workshop on healthcare *informatics (HI-KDD), pages 1–8, 2012.
- [8] Pilar Sierra, Lorenzo Livianos, Sergio Arques, Javier Castelló, and Luis Rojo. 2007. Prodromal symptoms to relapse in bipolar disorder. Australian & New Zealand Journal of Psychiatry 41, 5 (2007), 385–391.
- [9] Oliver D Howes, Samuel Lim, George Theologos, Alison R Yung, Guy M Goodwin, and Philip McGuire. 2011. A comprehensive review and model of putative prodromal features of bipolar affective disorder. Psychological medicine 41, 8 (2011), 1567–1577.
- [10] Yen-Hao Huang, Lin-Hung Wei, Yi-Shin Chen. 2017. Detection of the Prodromal Phase of Bipolar Disorder from Psychological and Phonological Aspects in Social Media. Archived at: <https://arxiv.org/pdf/1712.09183.pdf>.

- [11] Giovanni A. Fava, M.D., and Robert Keler, M.D., Ph.D. 2006. Prodromal Symptoms in affective disorders. *American Journal of Psychiatry* 148, 823–830.
- [12] M Berk, S Dodd, P Callaly, L Berk, P Fitzgerald, AR De Castella, S Folia, K Folia, S Tahtalian, F Biffin, et al. 2007. History of illness prior to a diagnosis of bipolar disorder or schizoaffective disorder. *Journal of affective disorders* 103, 1 (2007), 181–186.
- [13] Giovanni A. Fava, Eliana Tossani. 2007. Prodromal stage of major depression. *Early Intervention in Psychiatry* (2007) 1: 9–18
- [14] JS Berg, TA Grieger, JL Spira. 2005. Psychiatric symptoms and cognitive appraisal following the near sinking of a research submarine. *Mil Med* 170:44–47.
- [15] S.V.Manikanthan and D.Sugandhi “ Interference Alignment Techniques For Mimo Multicell Based On Relay Interference Broadcast Channel ” *International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)* ISSN: 0976-1353 Volume- 7 ,Issue 1 –MARCH 2014.
- [16] T. Padmapriya and V. Saminadan, “Priority based fair resource allocation and Admission Control Technique for Multi-user Multi-class downlink Traffic in LTE-Advanced Networks”, *International Journal of Advanced Research*, vol.5, no.1, pp.1633-1641, January 2017.