Analysis of Mental Health signal on social media platforms

Comparing behavior on Twitter and Reddit

Quang Pham

School of Science and Technology

Temple University

Abstract

Mental health problems have become one of the general concerns in our society, especially when there is one person in every five American adults experienced mental health issues and one in ten young people experienced a period of major depression. The burden of mental illness on our society is very high since we need a lot of time, money and effort to detect and treat the disease (Ronald C.Kessler,2011). With the development of social media, a platform allowing users to exchange information which includes billion of behaviors, details, ideas, and thoughts, we have a great chance to detect some symbols of mental health problems through analyzing data from social media and lead to a method to detect mental illness of users through social media (N Couldry,2012). I had studied mental health discourse on two popular social media platforms: Reddit and Twitter. The two platforms are very different in the way they operate, and it is the main reason the behavior of users on two platforms are quite different and it leads to the different task of analyzing on each kind.

1 Introduction

Social media sites have begun to emerge as popular platforms wherein information seeking and sharing practices are achievable. There are diverse papers on a wide range of topics based on data from social media: political science (TB Ksiazek,2010) or Social Science (D Zang, 2010). Since the major tasks of mental health research are to study the behavior of patients, their communication with others, their social status and their received support, using social media as a tool can help us overcome the lack of data available and give us the chance to developing a good diagnoses method and effective treatment for mental health disorders(J.A. Naslund, 2016) . Specifically, I perform analysis on two popular social media platforms: Reddit and Twitter.

First, Reddit is an online social media system that quite similar to a forum, where users can share texts, images or other types of media as a post in the forum. Each post is the place where users can interact with others through votes or comments. One difference of Reddit from other social media platforms is that user can be anonymous with a one-time used account. Because Reddit does not have strict rules on create a new account and provide information, as well as the media shared on the forum, especially texts, doesn't have limitation, Reddit, like other social networks, has become a good place for user to express self-disclosure, seeking for social- support and asking for other useful information to treat mental health problems(ADJ Kramer,2014). From the Reddit data, I can analyze the language and content characterize of self- disclosure relate to mental health issues and what kinds of social support is available on Reddit (M De Choundhury,2014).

The second part of my work is based on data from Twitter. Twitter is a popular social media which is very similar to our real life social. On Twitter, each user can have a number of friends, followers or followees and information is shared through tweets. By using Twitter data, I can analyze social pattern, self- expression, behaviors and symptom of mental illness patients(G Coppersmith,2014) The main difference of using Twitter data is that user on Twitter is not anonymous, so that we can understand more deeply about each user through their past data, and also give us a chance to predict their mental health problems through their tweets.

2 Related Work

There is some related work had been done in the past to show the potential in using social media data to analyze health problems. The result from De Choudhury(2014) has proved that social media data can be used to extract relevant mental health literature. From (Paul&Dredze 2011), arising in the usage of social media to discuss health problem has been proved. Also, the difference in language use has been observed in the data from students who have depression compare to students who do not (Rude et al, 2004) and the frequency of negative emotion words and anger has been recorded (Park et al,2012).

3 Reddit platform analysis

3.1 Data

On Reddit, each website link is created when a user creates contents of links or text posts. Each user is a “redditor” and they can vote for each post with the value of “up” or “down”, which also helped in ranking the post on the network by counting “upvotes” and “downvotes”. Posts on Reddit are organized by areas of interest, called “subreddits”, can be many subjects: science, film, music, etc. I used Reddit’s official API to download posts, comments on Reddit based on three subreddits: Depression, and Mental Health since this is the seed attributes relate to Mental Health and go further with SucideWatch .For each post, I got the textual content, id, timestamp, author id and the number of upvotes and downvotes for each post. Also, work on the distribution by the time all posts show that nearly 90% of the comment to any post is created during three days after the
post uploaded.

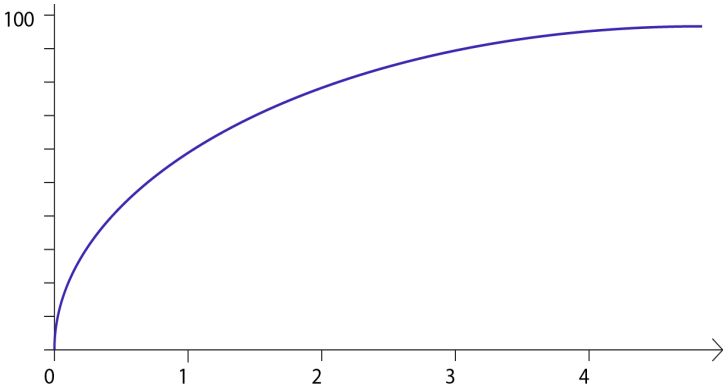

Figure 1. CDF of numbers of comment in days

Number of posts	7672
Average posts by user	0.62
Average upvotes	821.8
Average comments	112.17
Average title length	77.4
Average post length	668.78

Table 1. Statistics of the data

3.2 Methodology and Results

3.2.1 Self-disclosure and information sharing

I used NLTK package to perform sentimental analysis on the content of posts belongs to the three subreddits. For each content, I eliminated all the stopword and streamed them into the original form and count the frequency of each word. These words can be used for several intentions in our normal communication: emotional words, society related words, time,

occupations, cognitive verbs and inhibition words (DA Balota,2012). Some words belong to these subgroups has a high correlation with symptoms of mental health illness, such as anxiety(emotional), hate (emotional), fell (cognitive verbs), hate (cognitive verbs), escape (inhibition words) and avoid (inhibition words) (M.Choundry 2014). Since user on Reddit shares their story and emotion through a post on a forum and each post should be posted into a subreddit which is a cluster for all post with similar contents, all posts in mental health subreddits should be related to mental health problems. A high percentage of negative words recorded would show that these users on the three subreddits are suffering on some negative emotions like instability, loneliness or helpless (Rude et al 2004). Also, there is a significant amount of first-person pronoun (I, we) show that these posts are actually a story/thought from the writer, which similar to the fact that high-self attentional can be a symbol of mental health problems (W Bucci,1981).

Word	**Count**	**Word**	**Count**	**Word**	**Count**	**Word**	**Count**
Like	2726	Life	1756	Know	1502	Day	1023
Feel	2621	Get	1722	Even	1487	Work	842
Want	2426	Year	1679	Think	1428	Good	722
People	2100	Anxiety	1525	One	1363	Bad	699
Depression	1820	Time	1512	Stop	1329	Avoid	672

Table 2: Frequency of most popular 16 words

One more thing to notice that information on these subreddits can belong to these purposes: sharing symptoms or treatment for mental illness, a story to explain since when their have trouble and get ill or asking for help and support if they have been mentally triggered recently. Here is some example of posts on subreddit Mental Health:

‘‘I’m proud of you for getting out of bed today, and even if you haven’t yet, I’m still proud of you that you’re still here living to see another day. I just kinda figured somebody out there needed to hear this.”

“I'd look into EMDR therapy if you haven't tried that yet. Here's a study talking about EMDR for TRD‘‘

These results show that users on reddit often share their personal problems on subreddits, even if the story is very tragic and the range of information is very wide, from symptoms or treatment to challenging problems they have to face, or just some emotion they want to share.

3.2.2 Users interactive and the ways of support

There are two ways that user can interact with others on Reddit: using comments or giving post upvote and downvote. we have two features can be used to evaluate the networks: using the number of comments and using the difference between upvotes and downvotes, which is "Karma". There are a very high set of correlation coefficients between the two above variable and some features of the post such as length of the post, number of negative emotions, number of first-person pronouns, and number of words related to symptoms or treatment of mental illness(ADI Kramer,2014). These coefficients show that a long post can have a huge chance to be famous on the forum, especially when the post story belongs to the creator and the information of the post is highly relating to mention health. It looks like the community will care and support the story that they have understand more deeply and more tragic than the story which is short and not so tragic. One more thing to notice from the data is that there are two main ways the supporting comments can be: giving emotional comment or giving information comments. The percentage of giving emotional comment is about 70%, which show that most of the users are willing to support others in emotional ways, while the 30% have some knowledge about mental illness and they could be the answers for users who confused with their illness and need more information.

Giving emotional comment:

"I feel you bro! Hope everything will be better"

Giving information comment:

"Depression not mean feeling sad. Depression is a medical term for functioning below necessary levels"

4 Analysis on Twitter platform

4.1 Data

Twitter is a popular social network where each user can have a profile contains his or her personal information such as name, age, description and each user can have a wall where they can share their thought or emotion through tweets. Tweets are recorded by time and tweets could be many things from text content, picture, video or emoji. Users on Twitter can connect through friends or follower status and they can interact with each other through comments, retweets, likes, mention and direct message. Most of the data, except direct message, are public and can be accessed through Twitter API. Using the API and method to find Diagnosed group and Control Group (SB Blessing,2012). I achieved data from Twitter and separated it into two mentioned groups.

Diagnosed Group To download data for the diagnosed group, I searched for users who had public that they have been diagnosed with some types of mental illness such as PTSD or bipolar depression. One reason users public this statement is because this is a step in the process to treat the illness, so if user upload this information, it is a high chance that they have started fighting with the illness, also it can be a chance that user want to ask for support from the community and they want others to understand that they have trouble in their life because of mental illness (MW O'Hara,1986). The Tweet will be downloaded if it contains the phrase "I was diagnosed with" plus PTSD or depression. After having the tweets, I performed a filtering step to ensure the user actually have a mental illness, since there are cases when the tweet is just a joke or the tweet is not talking about the creator, but some other people. After having a list of users which had truly public their status online, I downloaded their most recent tweets, up to 3000 tweets per user, and only choose users whose main language is English.

The same method is used for **Control Group,** but the difference is for control group, I searched for the phrase that expresses a happy emotion such as "I feel happy today" or "love the weather", etc. The same filtering step is used, but I also check if among all tweets in the past, had the user ever mentioned the word depression or mental health. I will only take users who never used the word depression because of it is the only way to check if the user doesn't have depression based on social media data.

Diagnosed Group		Control Group	
Number of users	700	Number of users	1200
Numbers of tweet	~242.000	Number of tweets	~600,000

Table 3. Amount of data from Diagnosed Group and Control group

This method of downloading data is a good choice if we want to download a huge amount of data from tweets that contain users who public their diagnosed status online. However, it still has some drawbacks. The first thing is that for Control Group, we cannot ensure 100% if the user does not have a mental illness because the only value we have is text data from their tweets, not a diagnosis from the doctor. Even for the diagnosed group, we cannot ensure the diagnosis is true, even for most of the time no one wants to fake their mental health problems on social media.

4.2 Methodology and Results

4.2.1 Linguistic style and information sharing

Using the same technique as for the Reddit platform, I create a frequency table for most popular words diagnosed user used on Twitters. There is some similarity on the way words used, as "feel", "life" or "depression". However, since the length of a tweet is limited, the content of each tweet is very straight forward and for the diagnosed group, the group of words related to depression has high frequency and emotional words also be used frequently. Also in data, the percentage of first-person pronoun is very high (70%), show that there is a high attention to self-ness, which is also a symptoms of mental illness (AD Kramer,2014). Compare the statistic to the control group, user without depression will less likely used words related to depression, as well as the percentage of first pronoun appeared in each tweet is lowers. Also, there is a time when user reduces the frequency of the third person pronoun and start using the first pronoun among the diagnosed group, which can lead to the assumption that the time when first-person pronoun increased is the time the illness could start.

Depression related words	Depression, anxiety, addictive, sad, attack, nervousness
Emotional words	Fun, happy, enjoy, love, like, suffer, hate, angry, great

Table 4. Example of depression related words and Emotional words.

Using Non-negative matrix factorization, I realized that diagnosed users could have many favorite topics to talk about such as politics, sports, etc. However, there are four categories which are appeared in almost every diagnosed user and have a very high correlation coefficient with the sentimental score given by NLTK package: Symptoms, Disclosure, Treatment, Relationship and life. These four categories are similar to the four that I found on Reddit's data, despite the fact that Reddit has a different way of operation from Twitter. This result can prove that user mental health illness will share the symptoms, feelings and emotional state to seek for supports from their friends and help them overcome the loneliness and emotional attacks (N Lin,2013). Compare to the Control Group, the percentages of Symptoms and Treatment categories are much higher.

Symptoms	Anxiety, disable, depression, mental, PTSD, trauma, stress, attack
Disclosure	Fun, play, held, want, leave, suffer, sorry, answer, enjoy, hate, love
Treatment	Medication, drugs, doctor, diagnosed, effective, therapy, inhibitor
Relationships and life	Home, woman, family, she, him, girl, men, friend, someone, kid

Table 5. Some examples of words in each four main categories.

4.2.2 Network of user analysis

Based on the friend and follower list of each user, I have built a network of social media relationship for each user in order to demonstrate their connection on social network. For each user, I have the name of all users from their follow and followees list, and each of them will be node in the network. The network is a directed graph, where a link is added if user A follow user B.

Mean value	Diagnosed group	Control group
Number of followers (In degree)	28.4	48.7
Number of followees (Out degree)	18.2	42.2
Prestige ratio	0.87	0.613
2-hop neighborhood	112.4	197.4
Clustering coefficient	0.023	0.011

Table 6. Statistics of network of user relationship on Twitter

As in table 6, users in the diagnosed group have lower number of in degree and out degree compare to the control group. This fact can show that users in the diagnosed group do not want to be socialized as much as a normal user in control group and they will not receive and transmit much information because the in and out degrees of the network are low. Also diagnosed user will not have a wide range of relationship because the average degree of their network is small. Diagnosed group networks have smaller a 2-hop neighborhood, which shows that the connection of two users in the diagnosed group is weaker than the control group because information will need more steps to spread in the diagnosed group than needed in control group(Y Li,2009). The statistics of prestige ration and clustering coefficient show that in the network of diagnosed user, every node often have a quite similar number of neighbors, and these nodes is strongly connected than the network of control group. This detail will explain the fact that user with mental illness will tend to connect with user who they can trust can talk to share emotions and thoughts and users with mental illness often cluster together because they are the ones who understands mental health problems more than normal users(L Cooper-Patrick,1997). That is why users with mental health problems will have more closed and smaller networks, but all people in the networks can be very similar, so that they can give each other enough support and help.

4.2.3 Users activities analysis

For each tweet on Twitter, there are several attributes recorded such as time, location, number of mentions and like, etc. Analyzing these features give some results which are related to symptoms of mental illness. User from the diagnosed group often has a decreasing trend of activities by time of the day, especially at late night. Some symptoms of mental illness are tiredness, sleeping long hours (which means sleep earlier and longer), so the lower activity rate may lead to the fact that user is suffering from exhaustion, lack of time due to sleep disturbance, etc.(MAgarun,1997). As we can see in Figure 2 and 3, a user in the diagnosed group often have their activity decreased from 8 p.m to 10 a.m the next day, while user in control group has their activity increased from 9 p.m to 4 a.m. Moreover, users are less active when we observe the number of posts by days, as well as the number of likes, retweets or other interactions among the network. This can suggest the idea that user having loss connection on their relationships when

they have mental illness problems (WE Broadhead, 1990). Users themselves also may not want to be active on social media due to the illness.

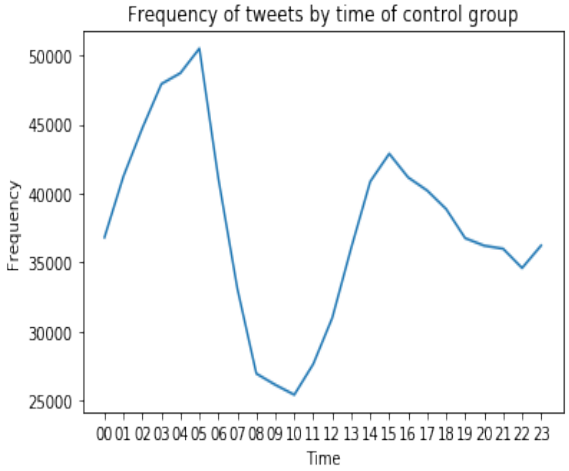

Figure 2

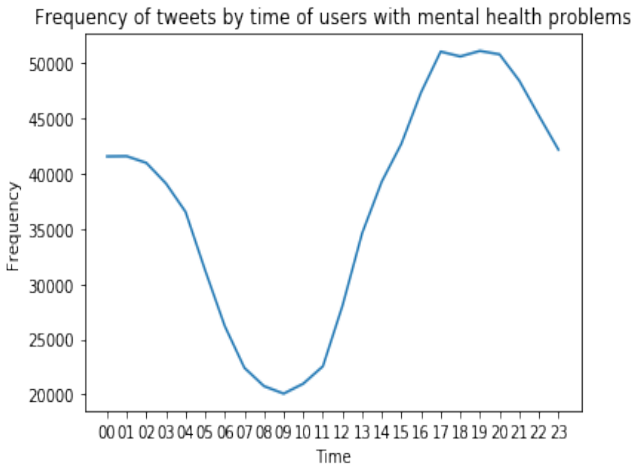

Figure 3

4.2.4 Predicting user with mental illness

4.2.4a Data preprocessing

With the data from Diagnosed group and Control group, I have about 2000 users from both group and nearly 850.000 tweets and I transformed the dataset of 850.000 tweets into a new dataset of textual statistics features of 2000 users. First of all, I only used tweets that have a negative label to perform the aggregate function. The label is evaluated by NLTK framework, with the accuracy of 88% on 500 examples. As I have mentioned on the previous parts, there are there important attributes of every tweet that have high correlation coefficients with symptoms of mental illness: time uploaded, the use of first person pronoun and the number of words belong to the four categories in Table 5. For each user, I will count the number of words belong to each category used in different time frames in a day such as from 0 to 2a.m, 2 to 6a.m or 8p.m to 10 a.m. These time frames are selected from the daily schedule of sleep and work of our people. I will also count number of first-person pronoun used among all tweets. Besides these above features, each user will have features relate to his or her social media profile and status such as number of followers, number of tweets, likes, age, etc. All attributes are normalized before used as input for the model.

4.2.4.b Models and scores

The first model I used is Logistic Classifier to classify the two labels: user with depression or not have depression. The accuracy of this model is ~80% on 1400 train cases and 600 test cases. When I checked the coefficient matrix of the model, it turns out that some features have weight 0, such as age, the mean number of words belong to relationship group in the time from 8a.m to 11a.m. On the other hand, some feature has a very high weight, such as average number of disclosures between 2 to 6a.m. Words from Symptom and Treatments has quite similar weights, despite is was posted in different time a day.

The second model I used is Adaboost classifier and the accuracy is ~90%. By used this model, I had combined the results of weak learners and reduce the variance, which helped a lot in increasing the accuracy(RE Schapire, 2013) . When compare to the same model but only used dataset contain attributes related to profiles information and recorded numbers of likes, tweets or retweets, the accuracy of using statistics of text information perform much better, which once again proves that there is a high correlation between mental illness symptoms and them linguistic styles and time of activities of user on Twitter.

5 Discussion

Despite the high accuracy of the models, there are several drawbacks with these works. First of all, the dataset is quite small, and it could lead to overfitting when using the model in real life. Secondly, the method to download data is based on the phrase “I was diagnosed with…”, so this means we can already know the label for these users and also the user has already been diagnosed, so we are just trying to do the work again. With the intention to detect user without any diagnosis have a mental illness or not, or model cannot give the accuracy of 90% if we only use this method of collecting and processing data. Even if the label given by the model is negative, we still need professional advice for the diagnosis. However, if we add some more survey question for the user who want to use the model, such as have you ever suffer from harassment/traumatic event/drug or alcohol addiction, we can have a higher confidence in giving the prediction (J Elith, 2006). One another approach is that we try to find the proof of traumatic /addiction or some events that can start the illness in the past inside the set of tweets.

6 References

Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Chatterji, S., Lee, S., Ormel, J., ... & Wang, P. S. (2009). The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. *Epidemiology and Psychiatric Sciences*, *18*(1), 23-33.

Couldry, N. (2012). *Media, society, world: Social theory and digital media practice*. Polity.

Ksiazek, T. B., Malthouse, E. C., & Webster, J. G. (2010). News-seekers and avoiders: Exploring patterns of total news consumption across media and the relationship to civic participation. *Journal of Broadcasting & Electronic Media*, *54*(4), 551-568.

Zeng, D., Chen, H., Lusch, R., & Li, S. H. (2010). Social media analytics and intelligence. *IEEE Intelligent Systems*, *25*(6), 13-16.

Naslund, J. A., Aschbrenner, K. A., Marsch, L. A., & Bartels, S. J. (2016). The future of mental health care: peer-to-peer support and social media. *Epidemiology and psychiatric sciences*, *25*(2), 113-122.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788-8790.

De Choudhury, M., & De, S. (2014, May). Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Eighth International AAAI Conference on Weblogs and Social Media*.

Coppersmith, G., Dredze, M., & Harman, C. (2014). Quantifying mental health signals in Twitter. In *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality* (pp. 51-60).

Stephanie S. Rude, Eva-Maria Gortner, and James W. Pennebaker. 2004. Language use of depressed and depression-vulnerable college students. *Cognition & Emotion,* 18(8):1121–1133, December.

Stephanie S. Rude, Eva-Maria Gortner, and James W. Pennebaker. 2004. Language use of depressed and depression-vulnerable college students. *Cognition & Emotion,* 18(8):1121–1133, December.

Aramaki, E., Maskawa, S., & Morita, M. (2011, July). Twitter catches the flu: detecting influenza epidemics using Twitter. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 1568-1576). Association for Computational Linguistics.

Balota, D. A. (2012). The role of meaning in word recognition. In *Comprehension processes in reading* (pp. 31-54). Routledge.

Bucci, W., & Freedman, N. (1981). The language of depression. *Bulletin of the Menninger Clinic*, *45*(4), 334.

Blessing, S. B., Blessing, J. S., & Fleck, B. K. (2012). Using Twitter to reinforce classroom concepts. *Teaching of Psychology*, *39*(4), 268-271.

O'Hara, M. W. (1986). Social support, life events, and depression during pregnancy and the puerperium. *Archives of General Psychiatry*, *43*(6), 569-573.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788-8790.

Lin, N., Dean, A., & Ensel, W. M. (Eds.). (2013). *Social support, life events, and depression*. Academic Press.

Li, Y., Chen, C. S., Song, Y. Q., Wang, Z., & Sun, Y. (2009). Enhancing real-time delivery in wireless sensor networks with two-hop information. *IEEE Transactions on industrial informatics*, *5*(2), 113-122.

Cooper-Patrick, L., Powe, N. R., Jenckes, M. W., Gonzales, J. J., Levine, D. M., & Ford, D. E. (1997). Identification of patient attitudes and preferences regarding treatment of depression. *Journal of general internal medicine*, *12*(7), 431-438.

Ağargün, M. Y., Kara, H., & Solmaz, M. (1997). Sleep disturbances and suicidal behavior in patients with major depression. *The Journal of clinical psychiatry*.

Schapire, R. E. (2013). Explaining adaboost. In *Empirical inference* (pp. 37-52). Springer, Berlin, Heidelberg.

Elith*, J., H. Graham*, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., ... & Li, J. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, *29*(2), 129-151.

Broadhead, W. E., Blazer, D. G., George, L. K., & Tse, C. K. (1990). Depression, disability days, and days lost from work in a prospective epidemiologic survey. *Jama*, *264*(19), 2524-2528.