Twitter Analysis for Depression on Social Networks based on Sentiment and Stress

Xiaohui Tao, Ravi Dharmalingam, Ji Zhang, Xujuan Zhou[1], Lin Li[2], and Raj Gururajan[1]
School of Sciences, University of Southern Queensland, Australia
[1]School of Management and Enterprise, University of Southern Queensland, Australia
[2]School of Computer Science and Technology, Wuhan University of Technology, China
{xiaohui.tao, u1081062, ji.zhang, xujuan.zhou, raj.gururajan}@usq.edu.au, [2]cathylilin@whut.edu.cn

***Abstract*—Detecting words that express negativity in a social media message is one step towards detecting depressive moods. To understand if a Twitter user could exhibit depression over a period of time, we applied techniques in stages to discover words that are negative in expression. Existing methods either use a single step or a data subset, whereas we applied a multi-step approach which allowed us to identify potential users and then discover the words that expressed negativity by these users. We address some Twitter specific characteristics in our research. One of which is that Twitter data can be very large, hence our desire to be able to process the data efficiently. The other is that due to its enforced character limitation, the style of writing makes interpreting and obtaining the semantic meaning of the words more challenging. Results show that the sentiment of these words can be obtained and scored efficiently as the computation on these dataset were narrowed to only these selected users. We also obtained the stress scores which correlated well with negative sentiment expressed in the content. This work shows that by first identifying users and then using methods to discover words can be a very effective technique.**

***Index Terms*—Twitter, depression, sentiment, stress, topic model**

I. Introduction

Mental health has become a very important issue for the well-being of people in our society. It is observed that some people are not hesitant to express their emotions in social media. As messaging via social media has become a popular medium for communication, the ability to detect depression in its content becomes very beneficial.

Our purpose is to find words used in tweets that can help predict if a user is expressing negative sentiment or experiencing depression. By analysing these tweets, we can form a predictive model that can be applied to detect similar sentiment in other tweets. Using the results of our experiments, this will give us a deeper understanding of the relationship between the semantic used and the mental status of the user.

Twitter data can be quite large and to analyse the entire dataset can introduce a measure of complexity that requires more sophisticated hardware and processing methods which some other experiments have relied on. Instead of analysing a large dataset or focussing on a smaller dataset, our method here is to first identify users which had expressed highly negative sentiments based on a set criteria and then applying further processing to obtain the words that of use to us. This trimming of the data to be processed allows us to not only focus on the more negative users but also greatly simplifies its processing.

We looked at both sentiment and stress analysis whereas some research have focused on a single aspect of a person's tweets. Both sentiment and stress are related which gives us a complementary perspective of a user's mental state and in combination allows us to see this as a longer term pattern of behaviour. As people are not always negative or stressful, our analysis also looked at the positive sentiment and relaxation scores as this gives us a good insight of the overall behaviour of the user over a period of time.

While just identifying words has its uses, we propose that quantifying the strength of the sentiment gives us the ability to measure the degree of sentiment in a tweet. We experimented by adjusting parameters when grouping the words into topics and applying an adaptive algorithm to produce the results. To analyse the tweets, a few techniques were employed, which when used together, forms a cohesive method which can be applied to other datasets. This includes analysing negative and positive sentiment, the tensile strength and relaxation and topic modelling. Our contributions to the work as shown in this paper are that words expressing sentiment can be scored and ranked. These words can be obtained in an efficient manner.

We begin this paper by doing a literature survey, followed by our methods to process the data and measuring the sentiment which then allowed us to identify the users of interest. We next present the framework of our methodology with the experimental results. The paper ends with a conclusion and our plans for further research.

II. Related Works

Research has been done to detect depression in heterogeneous information networks like Twitter, Facebook or Instagram, and how activities on social media may help detect depression. We will be examining work that has been done to identify depression from Twitter tweets using methods and techniques such as Natural Language Processing, Text Mining, Sentiment Analysis and Machine Learning. As social media is used by many people to express their views, this has sparked the interest of many researchers in how such information can be extracted from social media.

The degree of depression can be measured in Twitter [11]. Machine learning was used with an accuracy of 69%. Extracting features such as frequency of words (i.e. bag of words), ratio of positive to negative words, number of words, tweets with URL or first person pronouns can be used to predict depression with a reasonable level of accuracy.

Features that expresses the sentiment [3] were identified of the person doing the tweeting. Features of the post included the emotional state, the time of the post, the linguistic style (e.g. first person) and the n-grams. This correlates well with what psychologists [1] have stated regarding depressed persons. For example, the time of posts can indicate sleep deprivation or active at certain times of day or n-grams (mostly unigrams) can indicate the unwillingness to express words fully.

Twitter is used by some people as a medium to express their emotions. The words that they express can be detected to determine if they are in a negative emotional state or leading to that. Apart from the content, the context like the number of tweets or time of the posting are almost as important. To get a good of view of the many efforts that are being done on Twitter sentiment analysis, many studies [5] have been surveyed. The authors give a good overall view of Twitter, the current methods of processing tweets and a summary of approaches and methods that have been used from lexicon analysis to machine learning.

Lexical analysis is an important aspect of sentiment analysis. Lexical decision lists [7] was used to detect depression and PTSD. Stemming or spell checking was not done, instead n-grams and decision lists were used to determine depression. Using bag-of-words is a common technique where some words are more prevalent is message types that we want to classify like depressive messages or abusive messages. This was the technique done by [11] to analyse the semantic features. Adapting a modified bag of words with the 140 character limit of Twitter can be used to determine the polarity of the data and the preference of 1-gram over n-gram words and looking at the different grammar used in Twitter can help us to better extract sentiment from tweets.

Instead of looking at only vocabulary based approach, behaviour [6] in Twitter can be used to detect depression or symptoms of depression such as repetitive thoughts which are negative with expressions of worry and rumination and tweets regarding sleep, pain and suicide can be indicators of depression.

III. Framework

The Twitter raw data was collected from three months around the Australian election in 2016 and contains about 7.2 million tweets. The average number of words are 7 and the longest tweet is about 39 words.

A. Data pre-processing

The messages were filtered for only tweets originating from Australia and in the English language with emoticons and URLs removed. This reduced the number of useful tweets to about 6.1 million. Stop words are common words used in the English language like "the", "a" or "that". These were removed from the tweets as they do not contribute to the understanding of tweets. Stemming was then applied to reduce words to its word stem or its root. For example, "toys" to "toy" or "analysis" to "analys". The purpose of this is to let us to analyse the meaning or intent of the words used by the author of the tweets even if they have used different variations or spelling of that word as Twitter users are prone to.

B. Measuring sentiment, stress and relaxation

Useful information can now be derived by processing the tweets by measuring the sentiment of the tweets and the magnitude of it's stress and relaxation.

Two Java programs were used to mine opinions that measures the sentiment and the stress and relaxation of tweets. These programs have been written to have "human-level accuracy" for the short tweets texts. All the tweets strengths were measured on a scale of -1 to -5 on the negative scale and from +1 to +5 on the positive scale.

The first program, SentStrength [1] written by M. Thelwall, K. Buckley, G Paltoglou et al. measures the positive and negative sentiment. That is, -1 as slight negative to -5 as extremely negative; and +1 as slight positive to +5 as extremely positive. This program works well for messages with short texts like Twitter.

The second program, called TensiStrength [2], measures the stress and relaxation of tweets. The measurements are -1 is slight stress to -5 is very high stress; +1 is slight relaxation and +5 is highly relaxed.

C. Selecting users

We next selected users by filtering users who had tweeted more than fifty messages and who had an average negative sentiment of less than -2.5 from which thirteen users were selected. The scores of these users were calculated and grouped into two week intervals.

A Java program developed by *Dat Quoc Nguyen. jLDADMM: A Java package for the LDA and DMM topic models* [3] was applied to provide topic modelling and document clustering using the Latent Dirichlet Allocation (LDA) algorithm. The number of topics was set to ten. LDA calculates the probabilities of the similarities of the topics, which LDA will identify, of the tweets.

The objective is to obtain a set of words which can be used to predict that a user is showing symptoms of depression. The 140 character limit (at that time) of Twitter forces us to rethink and to modify methods of data mining, natural language processing, sentiment analysis and machine learning. When compared to content in other medium, Twitter content contains more usage of word shortening, slang, spelling and emojis.

D. Sentiment, Stress and Relaxation Analysis

Rather than having either negative or positive polarity; these tools produced both positive and negative scalings.

[1] http://sentistrength.wlv.ac.uk

[2] http://sentistrength.wlv.ac.uk/TensiStrength.html

[3] http://jldadmm.sourceforge.net/

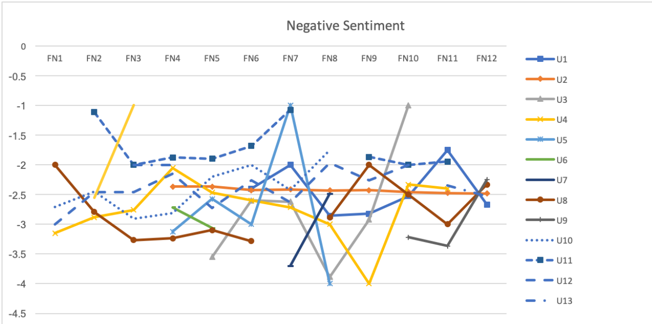

Fig. 1. Negative Sentiment

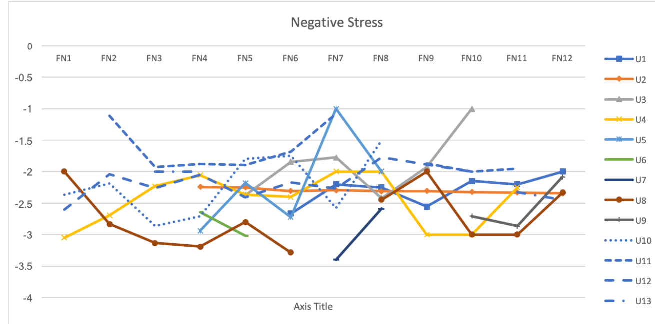

Fig. 2. Stress measurement

The reason being is that, as human beings, we can have mixed emotions [2]. Perhaps one is stronger than the other at certain times. A dictionary of terms was used and scores were assigned to them to determine sentiment and stress [9], [10].

Sentiment analysis was used to select the users and to discover words that were written that expressed negative sentiment. This allows us to determine the emotional state of the user who has a pattern of sending tweets over a period of time. It may be signs of depression but it can also be a response to other emotional attitudes.

Stress is an emotional mental reaction which can be caused by external physical occurrences or a result of anxiety, worry, pressure or fear. Prolonged stress can lead to depression [12] which in our case, is something we want to detect.

E. Topic Modelling

From the selected users, Latent Dirichlet Allocation (LDA) [4] was used to group these words into topics. We limited the number of topics to five with ten words for each topic. This limitation was set as the number of words in each tweet is small.

IV. Experimental Results and Analysis

1) Sentiment Analysis: The thirteen users were selected based on their negative sentiment analysis values. [4] Looking at the plot of the negative sentiment in Fig. 1, where "$U\#$" refers to a user with id $\#$ and "FN" refers to "fortnight", almost all of the users do express negative sentiment but they vary. There is no pattern that can be seen to determine a typical behaviour of someone exhibiting negative sentiment. However, the up and down shifts of sentiment does fit into a pattern of changing moods. As examples, *User 2* has an almost constant negative value; *User 4* did express very negative views throughout and *User 5* had very erratic changes to the sentiment which could indicate mood swings. This is confirmed in the standard deviation in Table I. While most users vary around a mean of -2.4 and a standard deviation of 0.4, there are some who are relatively large.

[4]While we are focussed on analysing the negative sentiment, it is worthwhile to also look at the positive sentiments expressed. However, the work is not reported in this paper due to space limit.

Users do express positive sentiment but the negative sentiment is generally maintained for the duration of the sample period. For example, User 10, while negative in the earlier weeks, became positive in later weeks.

User	σ
User 1	0.41847557
User 2	0.041402144
User 3	1.005755363
User 4	0.51834159
User 5	1.102116957
User 6	0.246709556
User 7	0.856377023
User 8	0.485509914
User 9	0.606263724
User 10	0.404307739
User 11	0.365630107
User 12	0.372841374
User 13	0.262734162

TABLE I
Standard Deviation for Negative Sentiment

2) Stress and Relaxation: In the experiments, our selection was based on sentiment. However, measuring stress and relaxation can give us a different but related view of the user. By comparing Fig. 1 and Fig. 2, it can be observed that the patterns and values are somewhat similar. From the results, we can conclude that stress can give rise to a person having negative emotions and wanting to express these emotions in words.

The similarity with the sentiment is quite apparent as users with a high level of stress have low relaxation values. The standard deviation in Table II shows a similar pattern.

Human behaviour is not constant, that is, we are not always in either a positive or negative state - we have mixed emotions and moods. However, for those who may be suffering from depression, there is a pattern that more negative behaviour is displayed; and the users' measure of positivity and relaxation is lower than negativity or stress. A pattern of behaviour is not readily apparent, if the users are constantly tweeting or are always very negative. For the majority of users, their sentiment do not vary much as can be seen in the standard deviation. This leads us believe that (at least during the sample period) users do not change their emotional state. From the adaptive algorithm, a list of negative sentiment words were produced

User	σ
User 1	0.32621526
User 2	0.023627791
User 3	0.511596414
User 4	0.453457839
User 5	0.891671646
User 6	0.284964033
User 7	1.276167031
User 8	0.410342621
User 9	0.397513677
User 10	0.792639122
User 11	0.310705977
User 12	0.281479749
User 13	0.144171851

TABLE II
STANDARD DEVIATION FOR STRESS

with a score attached to each of them. This is the final outcome of our experiments.

V. CONCLUSION

We set out to find a set of words that can be used as a predictor for negative sentiment which can be a sign of depression. This must be looked holistically in relation to patterns of behaviour over a period of time and whether positive sentiments are expressed as well, as it coexists with negative sentiment. By processing tweets and by experimenting with parameters in the model, adjusting thresholds, scores, number of tweets and topics and selecting the users, we can obtain a list of words with associated scores that statistically when used in tweets are more probable to be used by people who are depressed.

In future work, we would like to experiment and examine the results using stress and the aspects of positive sentiment and relaxation. An interesting adaptation is to analyse sentiment by processing emoticons.

ACKNOWLEDGMENT

The authors thank M. Thelwall, K. Buckley, G. Paltoglou et al for allowing the use of their SentStrength and TensiStrength programs for sentiment scoring, and to Dat Quoc Nguyen for use of his jLDADMM program for LDA analysis.

REFERENCES

[1] Aaron T. Beck, M.D. and Brad A. Alford, Ph.D. *Depression: Causes and Treatment*. University of Pennsylvania Press, February 2009.

[2] Raul Berrios, Peter Totterdell, and Stephen Kellett. Eliciting mixed emotions: a meta-analysis comparing models, types, and measures. *Frontiers in Psychology*, 6:428, 2015.

[3] Munmun De Choudhury, Scott Counts, and Eric Horvitz. Social media as a measurement tool of depression in populations. In *Proceedings of the 5th Annual ACM Web Science Conference*, WebSci '13, pages 47–56, New York, NY, USA, 2013. ACM.

[4] Munmun De Choudhury, Sanket S. Sharma, Tomaz Logar, Wouter Eekhout, and René Clausen Nielsen. Gender and cross-cultural differences in social media disclosures of mental illness. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 353–369, New York, NY, USA, 2017. ACM.

[5] Anastasia Giachanou and Fabio Crestani. Like it or not: A survey of twitter sentiment analysis methods. *ACM Comput. Surv.*, 49(2):28:1–28:41, June 2016.

[6] P. Nambisan, Z. Luo, A. Kapoor, T. B. Patrick, and R. A. Cisler. Social media, big data, and public health informatics: Ruminating behavior of depression revealed through Twitter. In *System Sciences (HICSS), 2015 48th Hawaii International Conference on*, pages 2906–2913, Jan 2015.

[7] Ted Pedersen. Screening twitter users for depression and ptsd with lexical decision lists. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 46–53, Denver, Colorado, June 5 2015. Association for Computational Linguistics.

[8] C. H. Tai, Z. H. Tan, and Y. S. Chang. Systematical approach for detecting the intention and intensity of feelings on social network. *IEEE Journal of Biomedical and Health Informatics*, 20(4):987–995, July 2016.

[9] Mike Thelwall. Tensistrength: Stress and relaxation magnitude detection for social media texts. *CoRR*, abs/1607.00139, 2016.

[10] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. Sentiment strength detection in short informal text. *Journal of the Association for Information Science and Technology*, 61(12):2544–2558, 2010.

[11] Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. Recognizing depression from twitter activity. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 3187–3196, New York, NY, USA, 2015. ACM.

[12] T. Plieger, M. Melchers, C. Montag, R. Meermann and M. Reuter (2015) Life stress as potential risk factor for depression and burnout.