Analysis of Mental Health Expression on Twitter

Michael Billot, University of Colorado at Colorado Springs

Abstract—**Since 2006 Twitter has existed as a platform which allows users to broadcast brief textual messages of no more than 140 characters. These short pieces of text are known as tweets. The most common purposes of tweets are daily conversations, information sharing, news critiques, and updates about a Twitter user's life. By facilitating such content Twitter promotes a wide array of emotional expression. In this research Twitter is queried for tweets containing the keyword "depressed." To begin analysis, a collection of personal, expressive tweets will be gathered. These collected tweets will contain content where the Twitter user appears to be sincerely writing about their depression. Analysis will be done by using human judges to score these expressive tweets along the Profile of Mood States (POMS) six dimensions of mood. A corpus of words will be produced based on the magnitude of scores for the six mood dimensions: tension, anger, depression, vigor, confusion, and fatigue.**

I. Introduction

THE amount of information presented by Twitter is staggeringly vast. The number of daily posts has continuously grown since it's introduction in 2006. In November of 2010, gigatweet was unable to continue its counting of tweets due to technical changes made by Twitter. However, over the course of the year prior, gigatweet documented a sustained increase in tweets from a rate of around 300 tweets per second to a rate of over 1000 tweets per second[1]. Within this mountain of information there are a large number of tweets where users are discussing and writing about mental health issues. An interesting subset of these tweets are those in which a Twitter user explicitly shares his or her feelings about an experience or affliction with a mental health issue. A desire to further understand the nature of these tweets is what motivates this research.

II. Motivation

The primary motivation of this research is to better understand the degree at which mental health issues are expressed by Twitter users. The expressive tweets scored by POMS will illustrate which dimensions of mood are being experienced by people sharing content about personal mental health. A measure of mood could be complemented by lexical analysis to draw conclusions about how a Twitter user's language is reflective of his emotional condition. Machine learning techniques using the collection of tweets could help automate the detection of similar ones in the future. The most profound product of this research might be a computational tool that automatically detects users who are habitually expressing negative sentiments or mental health problems. Automatic detection would be invaluable for a longitudinal analysis of these twitter users. A long term analysis could show the reasons why people choose to user Twitter as platform to share mental health issues. A long term analysis could also give insight into the benefits and positive effects that people experience from their expressive writing on Twitter.

III. Related Work

Bollen et al. performed sentiment analysis research of all public Twitter posts over a period of four months [1]. They used a syntactic term based approach to measure the sentiment of tweets via a psychometric instrument called POMS. They found that spikes in certain dimensions of sentiment could be correlated with critical events such as the 2008 presidential election and stock market fluctuations. This approach showed that supervised learning is not the only viable way to perform sentiment analysis of Twitter.

//Bollen and Pepe analyzed the mood expressed in 10,741 emails to the future [2]. To score the moods of the emails, they used an extended Profile of Mood States metric. They extended POMS original 65 adjectives with WordNet 3.0 synonyms. The extended list of POMS words were then stemmed using the Porter Stemmer. They scored the words

Pak and Paroubek used Twitter as a corpus for sentiment analysis and opinion mining [3]. Their method studied the POS tag distribution differences between positive, negative, and neutral tweets. A multinomial Naive Bayes classifier based on POS tags and n-grams was used. They concluded that a Twitter user's emotion is reflected in the syntactic structures of their tweets.

Lu et al. created a framework which automatically constructs a context dependent sentiment lexicon [4]. An unambiguous, gold standard sentiment lexicon is used as the basis. The polarity of these sentiments are propagated into other aspect-word pairs through language clues, a synonym antonym dictionary, and overall review ratings. In conclusion they found the framework could successfully learn new aspect dependent sentiments. Also, the coverage and accuracy of the general lexicon was greatly improved by the sentiment lexicon.

IV. Approach

Searching Twitter for posts containing the keyword "depressed" returns many tweets. Between February 11, 2011 and March 9, 2011 over 247,000 tweets were returned by continuously querying Twitter for "depressed." A large portion of this data is not relevant to the analysis of mental health expression. Only a subset of these tweets contain examples of users sincerely expressing their depression. The initial task is to isolate at least 1000 of these relevant tweets in which users seem to be expressing the emotional impact of depression.

The next step will be to score the set of at least 1000 expressive tweets. Scoring will be done by a group of human

[1] http://gigatweeter.com/analytics

judges. The tweets will be scored along the six dimensions of mood that are used in the Profile of Mood States scoring. It is important to ensure the validity of the scoring by confirming that the human judges are all scoring with like minds. A sample of tweets will be presented to the judges to see how well their POMS scores agree with one another. If their POMS scoring is in agreement, then POMS should be a good measure to score the emotions and moods expressed in tweets.

The collection of POMS scored tweets will then be lexically analyzed. Through analysis a lexicon of common words and syntactic groups of words could be shown to be common features of these tweets. The weights of words in the lexicon could correspond to the POMS scores of their containing tweets. This lexicon of features would be used to help accurately identify similar tweets. The similarity of these tweets could be determined by scoring them on how relevant their words are to the content in the lexicon.

Once emotionally expressive tweets can be selected for with some accuracy, then Twitter users expressing such emotions will be monitored over time. The idea is to observe users who are continuously and habitually writing tweets that are related to their depression or negative feelings. Studying this type tweeting behavior could give an understanding about what effect users experience by sharing such emotions publicly on Twitter. By studying them over time it would be possible to see fluctuations in their sentiments.

One other potential angle of approach would be to observe the networks that Twitter users are embedded in. A twitter user's network would include those that she follows or her followers. Perhaps a user who is continuously tweeting about depression and negative sentiments is influenced by their peers on Twitter. The idiom "misery loves company" could be a phenomenon among groups of followers on Twitter. Alternatively, users might express feelings of depression as a way to call for help and support from their friends. Either way, there is a lot that can be learned about the group interaction and the role it plays in mental health expression on Twitter.

A. Extending POMS

The POMS test consists of sixty-five mood related adjectives. Each of these sixty-five words are related to one of this six dimensions of mood. The purpose of the POMS list of words is to find the words in tweets. For example, if the word "angry" is contained in a tweet, then an anger-hostility mood is probably expressed in the tweet. However, the sixty-five words have limited coverage of the wide range of words that can be expressive of mood. The POMS list of words will be extended with WordNet synonyms. The extended list should contain more words that are representative of the six dimensions of mood.

B. Scoring Tweets

The goal of scoring tweets is to score them along the six dimensions of mood expressed in POMS. The two options for scoring methods are automatic scoring and human judge scoring. The automated scoring will be done by using the extended POMS word list. For each tweet, a six dimension POMS mood vector will be calculated. Each vector will be of the form tension-anxiety, anger-hostility, depression-dejection, confusion-bewilderment, fatigue-inertia, vigor-activity. If a tweet contains a word from the extended POMS list, then the respective dimension in its mood vector will be incremented.

The human judge scoring will involve scoring individual tweets on a scale of one to five along the six mood dimensions. Compared to the automatic scoring, human judges will be able to spot subtle expression of mood in the text.

C. Lexicon Construction

The Twitter mood lexicon will be a set of features that are most common in tweets. The features will be weighted based on their POMS scores. The feature weights could be based on a combination of the automatic scores and human scores. For example, all the words in a tweet scored as "anger-hostility" would contribute to their "anger-hostility" weight in the lexicon.

V. Future Work

There are two major possible applications for the Twitter mood lexicon. Understanding the mood trends in a population. Secondly, understanding an individual's mood. Investing an individuals mood could be most valuable when studied over a period of time.

VI. Acknowledgment

NSF grants CNS-0958576 and CNS-0851783 have funded the research reported in this paper.

References

[1] J. Bollen, A. Pepe, and H. Mao, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena," *CoRR*, vol. abs/0911.1583, 2009.

[2] A. Pepe and J. Bollen, "Between conjecture and memento: shaping a collective emotional perception of the future," *CoRR*, vol. abs/0801.3864, 2008.

[3] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, N. C. C. Chair), K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias, Eds. Valletta, Malta: European Language Resources Association (ELRA), may 2010.

[4] Y. Lu, M. Castellanos, U. Dayal, and C. Zhai, "Automatic construction of a context-aware sentiment lexicon: an optimization approach," in *Proceedings of the 20th international conference on World wide web*, ser. WWW '11. New York, NY, USA: ACM, 2011, pp. 347–356. [Online]. Available: http://doi.acm.org/10.1145/1963405.1963456