Predicting Depression in Facebook Users using Language Analysis

Anonymous Author(s)
Affiliation
Address
email

Abstract

In this work, we explore variations of topic-modeling in an attempt to understand depression symptoms and use that to predict depression. We propose a model drawing features from user behavior and topic modeling and demonstrate that use of language features helps in predicting risk of depression.

1 Introduction

Identifying depression symptoms is a challenging problem faced by health practitioners. About 25 million adults suffer from symptoms of depression in the United States [] and only ...% of these is detected. A lot of these go undetected because often people don't seek medical help when confronted with these symptoms. Symptoms that suggest depression include insomnia, urge to cry, loss of appetite, weight loss/gain and since these can occur due to changes in routine, work demands, depression is often mistaken and causes it to go unreported. Also, so far, the clinical assessments for depression (e.g. The Minnesota Multiphasic Inventory, MMPI) have been based on patient's self accounts of depression associated symptoms and this relies heavily on patients' ability to recognize and report symptoms. Suppression of these symptoms is also an issue among individuals in positions that demand lack of depression (e.g. pilots, army officials).

Alternative methods for evaluating depression are on the rise, which are more far-reaching than the traditional methods. With the outbreak of the social network, there is data from social networks which contain a trace of online activities. With careful examination of this data, people under higher risk of depression can be identified and counseled before they are stricken by a massive depression episode. Some of these methods include the strategic placement of online apps in sites frequented by large number of people and collecting data about changes in their behavior without too much intrusion. One such app is the Facebook app. This app presents a few questions regarding their daily routine, sleeping and eating habits and then uses that information to predict if the person has a risk of depression.

Though the app is not that intrusive, it only can predict depression for users that use the app truthfully. Hence, predicting risk of depression using user statuses, change in the choice of words and tone is a more non-intrusive way of predicting depression and is even more far-reaching than the app. To this end, in this work, we examine Facebook statuses of users and employ computational linguistics methods for predicting risk of depression. We use supervised machine learning methods to find language features that correlate heavily with depression risk as predicted by the app to learn language indicators that are suggestive of depression. We explore variations of topic modeling—*seeded topic models* [] and *SH-LDA* [] for understanding the text. We demonstrate that these linguistic tools help capture depression symptoms and help in predicting risk of depression.

The rest of the paper is organized as follows: 1)

2 Problem Statement

3 Pre-processing

The data that was provided to us were Facebook status messages of users. Status messages posted in Facebook are inherently noisy and thus deserves careful processing. We list the various preprocessing techniques that we applied on the status data to get rid of the noise from it.

(i) **Emoticons**: We replaced the emoticons using an emoticon dictionary that we constructed from [**?**]. For example, whenever we encountered any of these symbols `:-)` `:)` `:o)` `:]` `:3` `:c)` `:>` `=]` `8)` `=)` `(:`, we replaced it with the word **smile**.

(ii) **Stopwords, Punctuations & Numbers**: Stopwords, punctuations and alphanumeric characters from the status messages were removed as they do not carry much information on their own.

(iii) **Acronyms**: Short messages esp. the staus messages posted in Facebook tend to contain lots of acronyms like LOL meaning laugh out loud, ROFL meaning rolling on the floor laughing, etc. We construct an acronym dictionary and use it for replacing acronyms with their expanded versions.

(iv) **Expressions**: Expressions such as “hahahahhahahahaha”, “hehehehhehhehehhehehe”, “yayyyyyyyyyyyy” are very common across status messages in Facebook. The best way to deal with these words or expressions is to user regular expressions for detection of such patterns and replace them with words like **laugh**, **happy**, etc. Further words containing repeated characters (with 3 occurences or more) like “happpppyyyyyyyy” were reduced to **happyy** as it is not common or perhaps impossible to have three consecutive occurences of the same letter in a word in English language. Therefore any such occurence were reduced to a sequence of two consecutive letters and then our spell-checker (described next) was applied on it to get the correct word.

(v) **Spelling correction**: For words not found in the vocabulary after applying the previous preprocessing steps, we correct them using a spell-checker that was built using ideas from [**?**]. Essentially, in the spell-checker we are trying to find the correction c, out of all possible corrections, that maximizes the probability of c given the original incorrect word w:

$$\text{argmax}_c P(c|w)$$

which is equivalent to

$$\text{argmax}_c P(w|c)P(c)$$

Here $P(c)$ is the language model, which is the likelihood of occurrence of c in an English text. We use an unigram model to calculate the probability of c from the Brown corpus. To provide better estimates for $P(c)$, we could have used a bigram model here, however in that case we would probably require a “status message” corpus. Next, we need to figure out $P(w|c)$, the probability that w would be typed in a text when the user meant c, which is the error model. Finally, we need to enumerate all feasible values of c and then choose the one that gives the best combined probability score.

4 Model

4.1 Baseline

The baseline system had the following features:

1. Unigram and Bigram: Bag of words approach has been known to provide a significant feature set for many of the language modeling task. But we wanted to include only those words that were highly indicative of the label that we are trying to predict. For this, we first ranked the unigrams according to their frequency and took the top 6000 unigrams. Out of these, we selected those which correlated more than 0.1 as per Pearson’s rank correlation with the label we are tring to predict. This gave us a total of 112 unigrams. We did a similar filtering for the bigrams and got a set of 38 bigrams. These 112 unigrams and 38

bigrams corresponded to a set of 150 features where each feature value was the count of that unigram (or bigram) present in the concatenated statuses of that user.

2. Gender: Kenneth et.al [ref] studied that one of the potential risk factors for major depression is female sex. We included the gender as a binary feature.
3. NRC word-emotion association lexicon: The lexicon is a list of English words and their associations with eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust). We included eight features corresponding to the eight emotions and the feature value for a particular emotion was the sum of the association measures of all the words in concatenated status with that emotion.
4. Time window: We included 5 features related to the time of the status update (1) frequency of status updates per day, (2) number of statuses posted between 6-11 am, (3) number of statuses posted between 11-16, (4) number of sta- tuses posted between 16-21, (5) number of statuses posted between 21-00, and (6) number of statuses posted be- tween 00-6 am.
5. Topic modeling: We used vanilla LDA as implemented by the mallet toolkit. We concatenated all the statuses of a single user into one document. We need this for both the users in training dataset and the test dataset. We then ran the LDA on these documents to get a posterior topic distribution over 50 topics. This topic distribution was then used as 50 features in our model.

4.2 Seeded LDA

When we ran LDA on the Facebook statuses, we observed that words that can represent a similar disposition are spread across multiple topics.

Observing the top topic terms output by LDA, we observe that there are some words that occur across topics causing the topic distribution values to get distributed across multiple topics. This way, no one topic gets a high score and this is not helpful in determining what topics are predictive of *depression*.

Table **??**

topic 1: smile, sad, laugh, wink, playful, love, day, crying, surprise, god, feel, movie, book, mood, woohoo, perfect, place, un
topic 2: happy, glad, excited, fun, energy, pleasant, love, awesome
topic 3: love, life, heart, surprise, joyful, smile
topic 4: unpleasant, unhappy, sad, irritated, hate, jealous, gloomy, disappointed
topic 5: hate, awful, fuck, indecision, bored

Table 1: Seed words for *SeededLDA*

We observed that LDA on the Facebook statuses

topic 1: work, bad, tired, stress, time, cry, pain
topic 2: happy, glad, excited, fun, energy, pleasant, love, awesome
topic 3: love, life, heart, surprise, joyful, smile
topic 4: unpleasant, unhappy, sad, irritated, hate, jealous, gloomy, disappointed
topic 5: hate, awful, fuck, indecision, bored

Table 2: Seed words for *SeededLDA*

These two topics for instance demonstrate similar feeling and if the topic distribution for statuses get distributed across these two topics, we would not get a clear indication of what the underlying feeling of the user was when this was posted. Also, note that smile and sad occur in the same topic, so we cannot distinguish between these using the topic distribution.

So, to mitigate this we try to use SeededLDA, to guide topic models to learn topics that are of specific interest to us.

On examining some statuses and the top topic terms produced by LDA, we find that, words such as love, like, happy, playful, etc denote happiness and hence are not likely to correlate with depression.

On the other hand, words such as hate, cry, disappointed, sad, unhappy, etc are likely to correlate with depression.

Also words that represent work and related tension, such as stress, tired, time, busy, can correlate with depression symptoms.

So, we seeded the topics with words that we think correlate with depression.

— work, stress, tired, ... — happy, glad, excited, love, — sad, disappointed, hate, unhappy,

SeededLDA finds the words that are related to the seed words and places those words in the same topic. This helps because we don't need to identify all the words corresponding to depression and lack of depression. SeededLDA will gather these words as placing the related words in the same topic.

SeededLDA parameters:

Total 10 topic: 5 seeded topics, 5 regular topics Iterations: 500

Top topic terms after running SeededLDA:

— happy, glad, pleased, love, life, energy, awesome, surprise, — sad, disappointed, hate, unhappy, dishearten, awful, lonely, hate,

4.2.1 Seeded LDA with DSM Features

4.3 SHLDA

4.4 Word2Vec

5 Seeded LDA

6 SH-LDA

7 Empirical Evaluation

8 Discussion