Analysis of People's Emotions and Mental Health During a Pandemic Using Twitter

Amrit Alok
University of Massachusetts Amherst
aalok@umass.edu

Shruti Chanumolu
University of Massachusetts Amherst
schanumolu@umass.edu

Code submitted by Amrit Alok

Abstract

COVID-19 has been declared as a pandemic and its impact on the world is unprecedented. In addition to causing damage to various sectors such as economy, healthcare, and border immigration, the virus is significantly impacting the mental health of people due to strict lockdown, stay-at-home guidelines, people being in quarantine, and social distancing. During this time, to express emotions and thoughts, most of the people are using social media platforms. This presents a unique opportunity where the world is encompassed by a common phenomenon and talking about it. Twitter is one of the social media platforms which is used prominently everywhere with 500 million tweets every day. Due to this, We chose to analyze tweets about COVID-19 to understand its impact on the change of the user's emotions and mental health over time. Also, our goal is to use Twitter data to develop a model to predict users at the risk of depression. The developed model will be used to statistically analyze the risk of depression in twitter users as the pandemic progresses. We found that trust and joy had been two prominent emotion. Trust has stayed the same over time but joy decreased. Another interesting emotion has been fear which decreased over time.

1 Introduction

The COVID-19 pandemic has been a global catastrophe that has expanded up to 215 countries/territories (wor). With global cases of more than 3.5 million and 239,740 deaths (at the time of writing) (wor), its impact is insurmountable. The virus is not only detrimental to our physical health but has significantly impacted our mental health. The number of people impacted due to the loss of their family members, medical treatment, and quarantine will be much higher than the number of infected individuals. On top of that, strict lockdown, stay-at-home guidelines, people being in quarantine, and social distancing is causing anxiety, depression, distress, and other mental health issues.

Social media such as Twitter has become an integral part of our lives with more than 500 million tweets are shared every day. Since the spread of pandemic has been to 215 countries or territories which means it has engulfed the entire globe which means more and more people are talking about it. There has not been a single event in the last couple of decades where a common phenomenon has been discussed globally for such a long period. It provides a unique opportunity to analyze how people are reacting to the pandemic.

The study of emotional and mental impact due to COVID-19 is still in its early stages. We want to analyze twitter data and derive how people across the world are expressing their opinion and emotions. To understand the impact of the pandemic on mental health our study comprises the following three stages.

- Study eight basic emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) expressed by people. Since emotions are a major indicator of mental health.

- Develop a model to predict signs of depression from tweets and run this model on COVID-19 tweet data to study depression among users during the pandemic.

2 Related Work

In the past, the analysis of the mental health of a person was done based on clinical interviews. It was identified that the relative frequency of first-person singular pronouns spoken in interviews is a strong predictor of future depressive symptoms(Zimmermann et al., 2017). The Linguistic Inquiry and Word Count (LIWC) (Tausczik and Pennebaker, 2010) has been widely used in research to derive the emotion expressed or the mental health of a person. LIWC consists of words arranged in a hierarchy of emotions and the word representing it. One of the first works to use tweets to analyze the mental health of a person has been by (Choudhury et al., 2013). They created their dataset using crowd-sourcing and labeling the tweets with the psychometric test result, CES-D. They constructed various feature sets from the tweet such as the time the user has posted the tweets, the network of a user. They achieved an accuracy of 72%. It showcased that social media texts can be used to say whether a person is depressed or not. There has been a shared task by the 2015 ACL Workshop CLPsych (Coppersmith et al., 2015) which created a dataset for tweets for depression and PTSD. On this task, the best result was achieved by (Resnik et al., 2015). They used supervised topic modeling and bag-of-words features.

The work by (Weerasinghe et al., 2019) expands on the idea of using tweets to predict depression by removing the tweets which actively mentions depression. With this approach, they found that the machine learning model has learned features to predict whether a person is depressed or not even if they do not mention it. We think their work is in the right direction and should be expanded.

We did not come across any work which has analyzed the mental health of people during the COVID-19 crisis.

3 Dataset

We have collected tweets in two categories. First dataset is for depression prediction model and another is for analyzing COVID-19 tweets. Depression dataset is used to train the model to predict whether a tweets shows depression or not. The COVID-19 dataset is used to perform statistical results and analyze the emotion and mental health.

3.1 Depression Dataset

Train/Test dataset: It consists of 20K tweets retrieved using the Twitter scraping tool TWINT with keywords `'depress*`, `'feeling down'` and `'suicidal'` annotated as 'Depression'. These tweets are combined with 20K randomly sampled positive tweets(Sentiment value=4) from the Kaggle twitter sentiment dataset (https://www.kaggle.com/ywang311/twitter-sentiment) and labeled as 'Not Depressed'. The Train/Test dataset will be used to develop our depression classifier detailed in the next section. Collected tweets are split into training, testing, and validation sets with a ratio of 60%:20%:20% detailed in Table 1.

Dataset / Annotation	Depression	No Depression	Total
Train	6,739	12,000	28,739
Validation	2,246	4,000	6,246
Test	2,247	4,000	6,247
Total	11,232	20,000	31,232

Table 1: Dataset for Depression

3.1.1 Problems and Drawbacks

Our analysis of the depression-related tweets showed that some of them were informational or promotional for instance:

- "Are You Depressed? Signs and Symptoms of Depression in Women: http://t.co/h2YR0SYZdW depression"
- "What does depression feel like? Mental Health Advocate"

We excluded all promotional, awareness, and informational tweets as they don't indicate if the user is suffering from depression. To do so, we manually identified tweets that have first-person pronouns. After filtering tweets with first-person pronouns we only retrieved 11,232 tweets. Our annotation method labels a tweet from a person who is depressed if he claims to be depressed in his/her tweet, therefore, it targets self-diagnosis of depression. This method does not include depression diagnosed users who haven't explicitly mention depression-related keywords in their tweets. Further for non-depressed tweets we only collected positive sentiment tweets, but there can also be negative sentiment tweets that don't indicate depression for instance, "I don't like this dress". From the above observation, we realized that our data is biased. Therefore we randomly sampled 10K negative tweets from the sentiment140 dataset(https://www.kaggle.com/ywang311/twitter-sentiment) (sentiment value=4) and manually selected 5000 tweets that

Annotator	1	2	Our method
1	100%	87.3%	74.3%
2	87.3%	100%	70.6%
Our method	74.3%	70.6%	100%

Table 2: Inter Annotation Agreement

Dataset	Count
Global	123,382
United States	67,624

Table 3: Count of tweets related with COVID-19

correspond to negative sentiment but not depression. To validate our annotation method, we randomly selected 2000 tweets from the depression dataset and sent the subset to two annotators to compare results as shown in 2. (The rules for annotation are provided in the Appendix). Through this annotation exercise, we observed that our method labeled positive experiences such as recovery from depression for example "it was painful to cope with depression, glad i am recovered #seek help" also as depressed resulting in the disagreement with the annotators. Since there was at least a 70 % inter annotation agreement, we decided to proceed with our depression dataset.

3.2 COVID-19 Dataset

It consists of around 123K COVID-19 related tweets collected by (Müller and Salathé, 2019) using keywords *'wuhan', 'ncov', 'coronavirus', 'covid', 'sars-cov-2'* from 21 Jan - 19 April 2020. We randomly sample these tweets from the dataset because the entire dataset is too large to be downloaded and processed. We also collected geographic location(3) and timestamp corresponding to each tweet in the data. Each tweet is limited to 280 characters. This dataset will be used to analyze emotions and topic distribution and depression statistics for the COVID-19 pandemic.

4 Method and Experiments

Our method comprises of the following tasks:

4.1 Preprocessing

The tweets are converted to lowercase and the links to websites and user mentions `''` were removed from the tweets since they don't provide valuable information for analysis. We expanded the contraction in the tweet such as `"I'm", "That's"` because the first person pronouns has been indicated to identify with depression. The `'#'` was removed from the hashtags, this will help us to analyze the emotion or topic being expressed using hashtags e.g. "blessed" will become "blessed". The tweets were tokenized using NLTK (Loper and Bird, 2002) TweetTokenizer library and all words specific to Twitter were removed such as `('rt', 'at_user', and 'url')` along with the stopwords from the NLTK library for English excluding "not, no" and first person pronouns. We chose to experiment with Lemmatization because it gives the base form of the word and we have a higher chance of finding that word in the lexicon dictionary (AIL). The tokenized tweets were Lemmatized only for nouns and verbs using the NLTK POS tagging feature and then passing the tag to WordNet Lemmatizer for the corresponding word. For Task-1, we will be removing the words from the tweets which are related to depression e.g. `'depress*', 'feeling down', 'suicidal', and 'therapy'` to reduce the bias from the classifier.

4.2 Depression detection model

We experimented with different features and machine learning model architectures to identify linguistic markers in tweets that show signs of depression in user

4.2.1 Features

We used four feature sets on our analysis.

- **Bag-of-Words(BOW):** BOW words representation of each tweet was created using a vocabulary of 20,510 words using only. We only used word that occurred in more than 1% of the tweets.

- **Word Clusters(WC):** The language of social media is different than formal language. People tend to express their emotions using similar text which represents the same meaning such as *Hi, Hello, Yo, Hey*. These words represent the same thing - greeting. Having a cluster of similar words and treating them as a single token will help us to identify the common theme in the text. We use the set of 1000 hierarchical clusters created by (Owoputi et al., 2012) that are based on English tweets. These clusters are computed using Brown Clustering

(Brown et al., 1992) which assigns words to classes based on the frequency of word co-occurrence resulting in a hierarchical set of classes that are grouped semantically and syntactically. We replaced words in tweets by their cluster identifier if they have one.

- **POS Tags:** The grammar of a sentence has been an important indicator of depression. We used the POS tagger from the NLTK library which works great for grammatically correct sentences. However, another POS tagger, Tweet NLP Project (Gimpel et al., 2011) which is more suited for twitter POS tagging. The NLTK POS tagger does not recognize the emojis so we labeled them with `<EMJ>` We use four feature sets in our analyses. Most of the systems submitted to the CLPsych workshop [11] and other previous studies [17] show that bag-of-words features perform well. Resnik et al.'s system [34], which used a supervised topic modeling approach, performed best in the Shared Task. We, therefore, include these two feature sets in our analyses. Previous studies have shown that people with depression tend to use more personal pronouns [28, 47] and past tense verbs [40] in their writing. While De Choudhury et al. [13] included the frequencies of pronouns as a feature, differences in other parts of speech constructs were not analyzed. Therefore we included part-of-speech (POS) tags in our analyses. One drawback of using sparse feature sets like bag-of-words is that the models could overfit and the analyses of these features become difficult. To overcome this we used clusters of related words. This creates a dense feature matrix and allows the model to generalize to previously unseen words.

- **AffectVec:** Previous word embedding representation such as word2vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014) have been extensively used to learn dense vector representation of words. In which GloVe has been about distributional properties of words. However, these distributions do not represent any meaning behind them. AffectVec (Raji and de Melo, 2020) mitigated this issue by modifying the vector space to better account for the sentiment. First, they obtained a regular word vector and then adapted the original vector space to better reflect emotional ties. This was achieved by minimizing below objective function to obtain new vectors $v'_w \in V'$:

$$
\begin{aligned}
&l(V, V') \\
&\quad = \sum_{(u,w)\in O} \max(0, \epsilon - d(v'_u, v'w)) \\
&\qquad + \sum_{(u,w)\in s} d(v'_u, v'_w) + \sum_{(u,w)\in A} 1 - d(v'_u, v'_w) \\
&\qquad + \sum_{w\in V} \sum_{u\in N(w)} \max(0, d(v'_u), v'_w), d(v_u, v_w))
\end{aligned} \tag{1}
$$

Here, $d(v, v') = 1 - \cos(v - v')$ is the cosine distance and 0 denotes the null vector. Additionally, O is a set of word pairs with opposite sentiment polarity, A is a set of antonym word pairs with the opposite meaning. $N(w)$ is is a function that yields a set of related words for w by retrieving nearest neighbors in the original vector space. Finally, ϵ is a hyperparameter. All of the representation of the vector was converted to an embedding and passed to the model.

4.2.2 Supervised Learning Models

Our baseline supervised model takes BOW features as input and predicts depression by features. as and word cluster features as an input to the RandomForest classifier to detect depression symptoms in tweets. Besides, we also used a CNN-LSTM model to train our network for both GloVe and AffectVec embeddings. The model for them is shown in figure 1 and 2

```
Model: "sequential_6"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_6 (Embedding)      (None, 3015, 239)         4901890
_________________________________________________________________
conv1d_6 (Conv1D)            (None, 3015, 32)          22976
_________________________________________________________________
max_pooling1d_6 (MaxPooling1 (None, 1507, 32)          0
_________________________________________________________________
dropout_9 (Dropout)          (None, 1507, 32)          0
_________________________________________________________________
lstm_6 (LSTM)                (None, 150)               109800
_________________________________________________________________
dropout_10 (Dropout)         (None, 150)               0
_________________________________________________________________
dense_6 (Dense)              (None, 1)                 151
=================================================================
Total params: 5,034,817
Trainable params: 132,927
Non-trainable params: 4,901,890
```

Figure 1: The model used for CNN-LSTM training with GloVe Embedding

Layer (type)	Output Shape	Param #
embedding_5 (Embedding)	(None, 3015, 25)	512750
conv1d_5 (Conv1D)	(None, 3015, 32)	2432
max_pooling1d_5 (MaxPooling1	(None, 1507, 32)	0
dropout_5 (Dropout)	(None, 1507, 32)	0
lstm_5 (LSTM)	(None, 150)	109800
dropout_6 (Dropout)	(None, 150)	0
dense_5 (Dense)	(None, 1)	151

Total params: 625,133
Trainable params: 112,383
Non-trainable params: 512,750

Figure 2: The model used for CNN-LSTM with AffectVec Embedding

4.3 Emotion detection model

To map the tweet to the emotion being expressed, we used the NER sentiment score and which provides the score for close to 6000 words on how strongly a word represents an emotion between 0 and 1, where a score of 1 represents the word representing a particular emotion strongly.

We will be comparing the NER score with AffectVec (Raji and de Melo, 2020) which provides a score for 76427 words across 239 emotions which range from -1 to 1. The word embeddings value such as Glove does not represent any meaning behind the values. Raji and de Melo (2020) idea was to create a word embedding in which value can be interpreted. The paper provides more details about the score calculation and their interpretation.

4.3.1 Word Affect Intensities (AIL) Paper

There have been numerous works to create valence or sentiment lexicons which can be used to represent the emotion and sentiment. The most prominent work has been General Inquirer (Stone and Hunt, 1963), LIWC (Pennebaker et al., 2001), ANEW (Nielsen, 2011) etc. However, none of them provide a unified score to represent the emotions. We chose AIL because it provides a score for words which can be mapped to eight different emotions. The decision to use eight emotions follows the work by (PLUTCHIK, 1980) who suggested that some emotions are more basic than others. The AIL lexicons have been created considering analyzing the social media texts such as Twitter along with common English terms. It includes the terms which are associated with the emotions to various degrees and even includes some terms which may not predominantly convey that emotion.

The score for the words was calculated using Best-Worst Scaling (BWS) (Louviere and Woodworth, 1991). In this approach, annotators are given a tuple of size 4 and asked which item is the best (highest in terms of the property of interest) and which is the worst. The annotators were given the task as:

- the word that is associated with the MOST anger
- the word that is associated with the LEAST anger

To measure the accuracy of the annotations, Mohammad (2018) split the annotated data into two halves and calculated the reliabilities (measured by Pearson correlation and Spearman rank correlation) over 100 trials. The correlation between two sets was calculated and its average was above 0.9 indicating good reproducibility.

Since the AIL lexicons were tuned to analyze the social media posts and provided the score for eight emotions with their reliability, we found it suitable to analyze the emotions of people during a pandemic. This will give a granular view on how people are reacting during the pandemic than detecting positive and negative emotions.

4.3.2 Approach

To identify the emotion expressed in the tweet, we used the NER sentiment lexicon. The NER sentiment lexicons are only for words and do not include emojis. Shoeb et al. (2019) worked on to map the score of the emojis across four emotions *anger, fear, joy, sadness*. To include emojis in our calculation, we focused only on 4 emotions.

To analyze the emotions, we created eight dictionaries for eight different emotions. The preprocessed tokenized words from the tweet were looked up in the dictionary across different emotions dictionary and their score was added. With this, we had the score of each tweet across the different categories of emotions. We selected the score with the maximum value and mapped the corresponding emotion to the tweet. If the maximum score was 0, then it was labeled as ”Neutral”.

To analyze the emotions, we captured the tweet date, location, identified emotion. This data was plotted on a time-series to analyze the pattern. We also identified the prominent emotion expressed and plotted them in a time-series manner.

We plotted 4 concerned emotions *anger, fear, joy, sadness* vs other 5 using NER lexixcons and EmoTag.

For completeness, we are providing details about the break-up of emotions expressed 5 for in the United States without using EmoTag.

5 Results

5.1 Depression detection

5 shows the results for different supervised learning models. We see that both the CNN-LSTM models perform better than the baseline BOW model. Also by incorporating POS tags and Word clustering features into BOW features, the model performance has improved. Despite the improvement, we find the accuracy for all models to be quite low. 4 shows all the significant BOW features divided into non-depressed and depressed word clouds. The low accuracy of the BOW model can be attributed to the overlap of words in the two classes. However, I see more emphasis given to self pronouns and and word 'life' in depression class. These tokens are more general features that distinguish between the two classes.4 shows that individuals with depression tweeted fewer abbreviations such as "lol" and "haha", and tweeted less about happy or gratitude expressing words like 'thank you'. Further, some swear words can be seen in the depressed class features. These patterns become more apparent in feature analysis of word clusters and POS tags, therefore, the improved performance. The AffectVec embedding captures the underlying emotions of a word and our understanding of improved accuracy can be explained on its ability to distinguish between polar opposite emotions, therefore better suited for emotions related task.

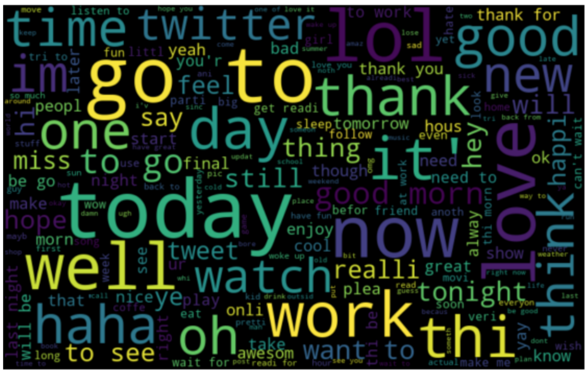

Figure 3: Bag-of-Words features corresponding to the non-depressive class. Size represents information gain

Figure 4: Bag-of-Words features corresponding to the depressive class. Size represents information gain

5.2 Emotion Identification

Our expectation was the most prominent emotion expressed will be anger and fear. It was surprising to find that the top-3 emotions expressed were trust, fear, and joy. Upon analysis of the tweets, we found that trust was identified with texts which reported facts, news, and people commenting towards administration and showing solidarity. The tweets marked with joy were mostly related to expressing positive emotion by spending time with family, motivating others, expressing gratitude.

We compared the identification of emotions using AffectVec (Raji and de Melo, 2020) and plotted them on the time-series as well. They provide emotions expressed by a word across 239 emotions. However, to compare against NER+EmoTag, we only considered 4 emotions. We found a similar

Emotion	Percentage
Trust	22.8%
Neutral	19.0%
Fear	15.9%
Joy	13.7%
Anticipation	10.0%
Sadness	9.8%
Anger	4.7%
Disgust	4.4%
Surprise	3.8%

Table 4: Different Emotions Expressed by People during the COVID-19 Pandemic

Model	Accuracy	Precision	Recall	F1 Score
BOW	0.45	0.53	0.45	0.48
BOW+NLTK+WC	0.49	0.54	0.47	0.50
AffectVec+CNN+LSTM	0.63	0.61	0.67	0.64
GloVe+CNN+LSTM	0.59	0.62	0.61	0.61

Table 5: 5-fold cross-validated results for the classification task on the depression dataset

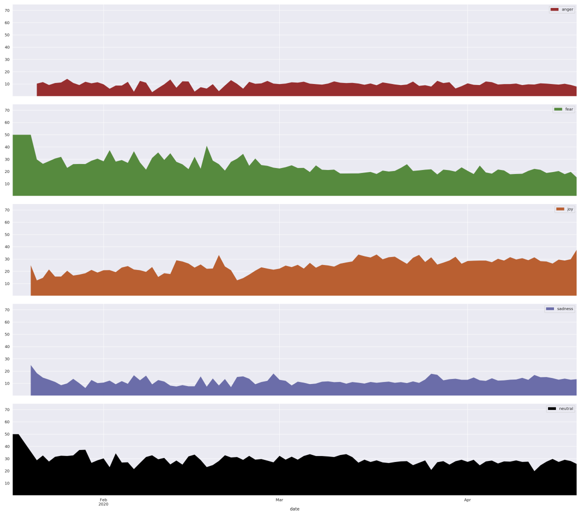

Figure 5: The trend of emotions (anger, fear, joy, sadness, and others: top to bottom) expressed in the United States during COVID-19 in using NER Sentiment and EmoTag Lexicons

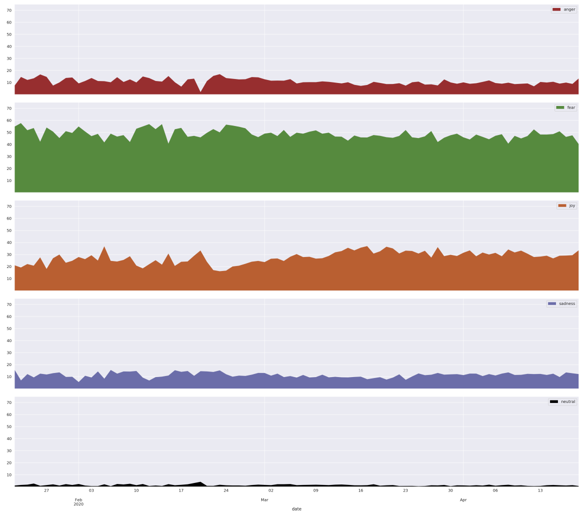

Figure 6: The trend of emotions (anger, fear, joy, sadness, and others: top to bottom) expressed in the United States during COVID-19 in using AffectVec

trend 6 for the tweets collected in the United States. The neutral/other category has a very low percentage due to large numbers of words represented compared to the NER lexicon.

We also identified the emotions expressed in the tweets across 239 categories and found them somewhat accurate and promising. Due to the recent publication of the AffectVec, more time needs to be spent on experimenting with them.

5.2.1 Issues and Challenges:

To confirm the emotions identified with tweets are accurate or not, we randomly selected 1000 tweets and did the exploratory analysis. Identifying emotion for shorter tweets was not correct, for example, International Medical Transport - labeled as ”joy” but it is difficult to interpret the emotion for shorter texts. Sarcastic tweets were also identified, however, it is difficult for humans as well to identify a sarcastic text without background information.

6 Future Work

The most important task will be to create a balanced dataset which contains more tweets from a particular user to identify depression on a user level. Our dataset had very few tweets from the same user which made the task a bit challenging. This inhibited us to create topics related to depression on a user level and use them as a feature. Our next approach will be to combine the embeddings, emotion, and topics as the prior probability to use as a feature in a neural network to identify depression in users.

To identify the correct emotion, identifying sarcasm and humor based on the context of the tweet or background information about the prominent entity in the tweet will help us to more accurately identify the emotion.

Since the accuracy of our depression model was low, we will work on improving it and study the impact of COVID-19 on mental health.

7 Conclusion

We found that the embeddings based CNN-LSTM model performs better than simple BOW models. This means that the deep learning-based models learn the underlying semantics of the text which can help us to identify the depression or mental health of a user.

To identify the emotions expressed in a tweet, we found that the lexicons based approach works better by including the emojis. A lexicon dictionary containing a large number of words will help us to identify the emotion expressed more accurately.

From the results, we saw that the major emotions expressed during the pandemic varied over time. Fear was decreasing among people which could mean that people were getting used to the environment but how this will pan in the long run will need to be by including recent tweets.

References

Coronavirus disease (covid-19) pandemic.

Peter F Brown, Peter V Desouza, Robert L Mercer, Vincent J Della Pietra, and Jenifer C Lai. 1992. Class-based n-gram models of natural language. *Computational linguistics*, 18(4):467–479.

Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *ICWSM*.

Glen Coppersmith, Mark Dredze, Craig Harman, Kristy Hollingshead, and Margaret Mitchell. 2015. CLPsych 2015 shared task: Depression and PTSD on twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 31–39, Denver, Colorado. Association for Computational Linguistics.

Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. 2011. Part-of-speech tagging for twitter: Annotation, features, and experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*, HLT '11, page 42–47, USA. Association for Computational Linguistics.

Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1*, ETMTNLP '02, page 63–70, USA. Association for Computational Linguistics.

Jordan J Louviere and George G Woodworth. 1991. Best-worst scaling: A model for the largest difference judgments. *University of Alberta: Working Paper*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc.

Saif Mohammad. 2018. Word affect intensities. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Martin M Müller and Marcel Salathé. 2019. Crowdbreaks: Tracking health trends using public social media data and crowdsourcing. *Frontiers in public health*, 7.

Finn Årup Nielsen. 2011. A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.

Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, and Nathan Schneider. 2012. Part-of-speech tagging for twitter: Word clusters and other advances. *School of Computer Science, Carnegie Mellon University, Tech. Rep*.

James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

ROBERT PLUTCHIK. 1980. Chapter 1 - a general psychoevolutionary theory of emotion. In Robert Plutchik and Henry Kellerman, editors, *Theories of Emotion*, pages 3 – 33. Academic Press.

Shahab Raji and Gerard de Melo. 2020. What sparks joy: The AffectVec emotion database. In *Proceedings of The Web Conference 2020*, WWW â™20, pages 2991–2997, New York, NY, USA. ACM.

Philip Resnik, William Armstrong, Leonardo Claudino, and Thang Nguyen. 2015. The university of Maryland CLPsych 2015 shared task system. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 54–60, Denver, Colorado. Association for Computational Linguistics.

Abu Awal Md Shoeb, Shahab Raji, and Gerard de Melo. 2019. EmoTag – towards an emotion-based analysis of emojis. In *Proceedings of the International Conference on Recent Advances in Natural Language*

Processing (RANLP 2019), pages 1094–1103, Varna, Bulgaria. INCOMA Ltd.

Philip J. Stone and Earl B. Hunt. 1963. A computer approach to content analysis: Studies using the general inquirer system. In *Proceedings of the May 21-23, 1963, Spring Joint Computer Conference*, AFIPS '63 (Spring), page 241–256, New York, NY, USA. Association for Computing Machinery.

Yla R. Tausczik and James W. Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54.

Janith Weerasinghe, Kediel Morales, and Rachel Greenstadt. 2019. "because... i was told... so much": Linguistic indicators of mental health status on twitter. *Proceedings on Privacy Enhancing Technologies*, 2019(4).

Johannes Zimmermann, Timo Brockmeyer, Matthias Hunn, Henning Schauenburg, and Markus Wolf. 2017. First-person pronoun use in spoken language as a predictor of future depressive symptoms: Preliminary evidence from a clinical sample of depressed patients. *Clinical psychology psychotherapy*, 24 2:384–391.