

The Role of Personality, Age and Gender in Tweeting about Mental Illnesses

Daniel Preoțiuc-Pietro^{1,2}, Johannes Eichstaedt¹, Gregory Park¹, Maarten Sap¹
Laura Smith¹, Victoria Tobolsky¹, H. Andrew Schwartz^{2,3} and Lyle Ungar^{1,2}

¹Department of Psychology, University of Pennsylvania

²Computer & Information Science, University of Pennsylvania

³Computer Science, Stony Brook University

danielpr@sas.upenn.edu

Abstract

Mental illnesses, such as depression and post traumatic stress disorder (PTSD), are highly underdiagnosed globally. Populations sharing similar demographics and personality traits are known to be more at risk than others. In this study, we characterise the language use of users disclosing their mental illness on Twitter. Language-derived personality and demographic estimates show surprisingly strong performance in distinguishing users that tweet a diagnosis of depression or PTSD from random controls, reaching an area under the receiver-operating characteristic curve – AUC – of around .8 in all our binary classification tasks. In fact, when distinguishing users disclosing depression from those disclosing PTSD, the single feature of estimated age shows nearly as strong performance (AUC = .806) as using thousands of topics (AUC = .819) or tens of thousands of n-grams (AUC = .812). We also find that differential language analyses, controlled for demographics, recover many symptoms associated with the mental illnesses in the clinical literature.

1 Introduction

Mental illnesses, such as depression and post traumatic stress disorder (PTSD) represent a large share of the global burden of disease (Üstün et al., 2004; Mathers and Loncar, 2006), but are underdiagnosed and undertreated around the world (Prince et al., 2007). Previous research has demonstrated the important role of demographic factors in depression risk. For example, while clinically-assessed depression is estimated at 6.6% in a 12-month interval for U.S. adults (Kessler et al., 2003), the prevalence in

males is 3-5%, while the prevalence is 8-10% in females (Andrade et al., 2003). Similarly, prevalence of PTSD among U.S. adults in any 12-month period is estimated at 3.5% (Kessler et al., 2005b) – 1.8% in males and 5.2% in females – yet this risk is not distributed evenly across age groups; prevalence of PTSD increases throughout the majority of the lifespan to reach a peak of 9.2% between the ages of 49-59, before dropping sharply to 2.5% past the age of 60. (Kessler et al., 2005a).

Large scale user-generated content provides the opportunity to extract information not only about events, but also about the person posting them. Using automatic methods, a wide set of user characteristics, such as age, gender, personality, location and income have been shown to be predictable from shared social media text. The same holds for mental illnesses, from users expressing symptoms of their illness (e.g. low mood, focus on the self, high anxiety) to talking about effects of their illness (e.g. mentioning medications and therapy) and to even self-disclosing the illness.

This study represents an analysis of language use in users who share their mental illness through social media, in this case depression and PTSD. We advocate adjusting for important underlying demographic factors, such as age and gender, to avoid confounding by language specific to these underlying characteristics. The age and gender trends from the U.S. population are present in our dataset, although imperfectly, given the biases of self-reports and social media sampling. Our differential language analyses show symptoms associated with these illnesses congruent with existing clinical theory and consequences of diagnoses.

In addition to age and gender, we focus on the important role of inferred personality in predicting

mental illness. We show that a model which uses only the text-predicted user level ‘Big Five’ personality dimensions plus age and gender perform with high accuracy, comparable to methods that use standard dictionaries of psychology as features. Users who self-report a diagnosis appear more neurotic and more introverted when compared to average users.

2 Data

We use a dataset of Twitter users reported to suffer from a mental illness, specifically depression and post traumatic stress disorder (PTSD). This dataset was first introduced in (Coppersmith et al., 2014a). The self-reports are collected by searching a large Twitter archive for disclosures using a regular expression (e.g. ‘I have been diagnosed with depression’). Candidate users were filtered manually and then all their most recent tweets have been continuously crawled using the Twitter Search API. The self-disclosure messages were excluded from the dataset and from the estimation of user inferred demographics and personality scores. The control users were selected at random from Twitter.

In total there are 370 users diagnosed only with PTSD, 483 only with depression and 1104 control users. On average, each user has 3400.8 messages. As Coppersmith et al. (2014b) acknowledge, this method of collection is susceptible to multiple biases, but represents a simple way to build a large dataset of users and their textual information.

3 Features

We use the Twitter posts of a user to infer several user traits which we expect to be relevant to mental illnesses based on standard clinical criteria (American Psychiatric Association, 2013). Recently, automatic user profiling methods have used on user-generated text and complementary features in order to predict different user traits such as: age (Nguyen et al., 2011), gender (Sap et al., 2014), location (Cheng et al., 2010), impact (Lampos et al., 2014), political preference (Volkova et al., 2014), temporal orientation (Schwartz et al., 2015) or personality (Schwartz et al., 2013).

3.1 Age, Gender and Personality

We use the methods developed in (Schwartz et al., 2013) to assign each user scores for age, gender and personality from the popular five factor model of personality – ‘Big Five’ – (McCrae and John, 1992), which consists of five dimensions: extraversion, agreeableness, conscientiousness, neuroticism and openness to experience.

The model was trained on a large sample of around 70,000 Facebook users who have taken Big Five personality tests and shared their posts using a model using 1-3 grams and topics as features (Park et al., 2014; Schwartz et al., 2013). This model achieves $R > .3$ predictive performance for all five traits. This dataset is also used to obtain age and gender adjusted personality and topic distributions.

3.2 Affect and Intensity

Emotions play an important role in the diagnosis of mental illness (American Psychiatric Association, 2013). We aim to capture the expression of users’ emotions through their generated posts. We characterize expressions along the dimensions of *affect* (from positive to negative) and *intensity* (from low to high), which correspond to the two primary axes of the circumplex model, a well-established system for describing emotional states (Posner et al., 2005).

Machine learning approaches perform significantly better at quantifying emotion/sentiment from text compared to lexicon-based methods (Pang and Lee, 2008). Emotions are expressed at message-level. Consequently, we trained a text classification model on 3,000 Facebook posts labeled by affect and intensity using unigrams as features. We applied this model on each user’s posts and aggregated over them to obtain a user score for both dimensions.

3.3 Textual Features

For our qualitative text analysis we extract textual features from all of a user’s Twitter posts. Traditional psychological studies use a closed-vocabulary approach to modelling text. The most popular method is based on Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2001). In LIWC, psychological theory was used to build 64 different categories. These include different parts-of-speech, topical categories and emotions. Each user is thereby

represented as a distribution over these categories. We also use all frequent 1-3 grams (used by more than 10% of users in our dataset), where we use point-wise mutual information (PMI) to filter infrequent 2-3 grams.

For a better qualitative assessment and to reduce risk of overfitting, we use a set of topics as a form of dimensionality reduction. We use the 2,000 clusters introduced in (Schwartz et al., 2013) obtained by applying Latent Dirichlet Allocation (Blei et al., 2003), the most popular topic model, to a large set of Facebook posts.

4 Prediction

In this section we present an analysis of the predictive power of inferred user-level features. We use the methods introduced in Section 3 to predict nine user level scores: age, gender, affect, intensity and the Big Five personality traits.

The three populations in our dataset are used to formulate three binary classification problems in order to analyse specific pairwise group peculiarities. Users having both PTSD and depression are held-out when classifying between these two classes. To assess the power of our text-derived features, we use as features broader textual features such as the LIWC categories, the LDA inferred topics and frequent 1-3 grams.

We train binary logistic regression classifiers (Pedregosa et al., 2011) with Elastic Net regularisation (Zou and Hastie, 2005). In Table 1 we report the performance using 10-fold cross-validation. Performance is measured using ROC area under the curve (ROC AUC), an adequate measure when the classes are imbalanced. A more thorough study of predictive performance for identifying PTSD and depressed users is presented in (Preoțiu-Pietro et al., 2015).

Our results show the following:

- Age alone improves over chance and is highly predictive when classifying PTSD users. To visualise the effect of age, Figure 1 shows the probability density function in our three populations. This highlights that PTSD users are consistently predicted older than both controls and depressed users. This is in line with findings from the National Comorbidity Survey and replications (Kessler et al., 2005a; Kessler et al.,

Feature	No.feats	C-D	C-P	D-P
Random	-	.5	.5	.5
Age	1	.557	.742	.801
Gender	1	.559	.513	.522
Age + Gender	2	.590	.747	.8
Big5	5	.784	.813	.777
Age + Gender + Big5	7	.783	.844	.806
Affect + Intensity	2	.583	.519	.559
LIWC	64	.824	.854	.781
Topics	2000	.851	.901	.819
Unigrams	5432	.858	.911	.809
1-3 grams	12677	.859	.917	.812

Table 1: Predictive performance using Logistic Regression in ROC area under the curve (AUC) between controls (C), depressed (D) and PTSD (P).

2005b). As a consequence, factoring in age in downstream analysis is necessary, as language changes with age on social media.

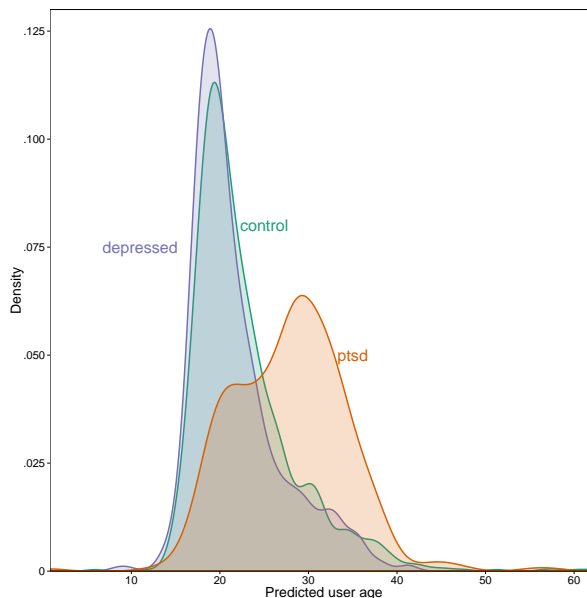
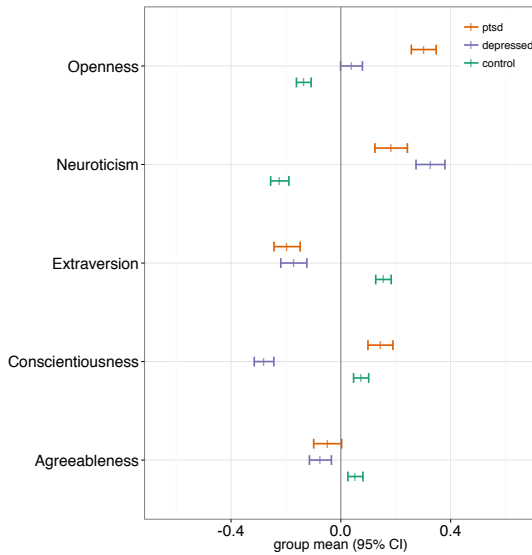


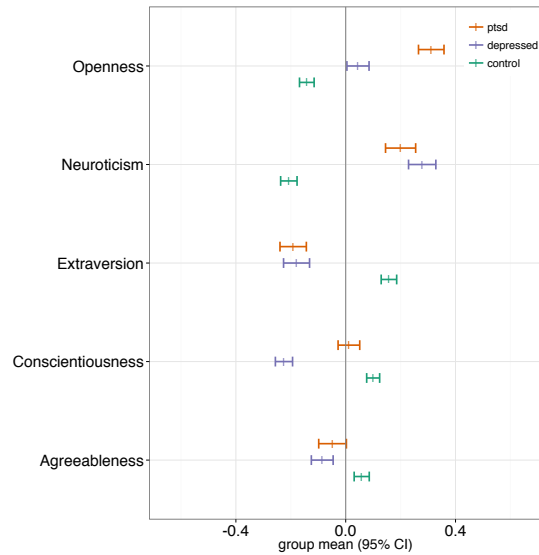
Figure 1: Age density functions for each group.

- Gender is only weakly predictive of any mental illness, although significantly above chance in depressed vs. controls ($p < .01$, DeLong test¹). Interestingly, in this task age and gender combined improve significantly above each individual prediction, illustrating they contain complementary information. Consequently, at least when analysing depression, gender should

¹A non-parametric test for identifying significant differences in ROC curves (DeLong et al., 1988)



(a) Not controlled for age and gender.



(b) Controlled for age and gender.

Figure 2: Big Five personality means (grand mean centered) and confidence intervals for each group.

be accounted for in addition to age.

- Personality alone obtains very good predictive accuracies, reaching over .8 ROC AUC for classifying depressed vs. PTSD. In general, personality features alone perform with strong predictive accuracy, within .1 of >5000 unigram features or 2000 topics. Adding age and gender information further improves predictive power (C-P $p < .01$, D-P $p < .01$, DeLong test) when PTSD is one of the compared groups.

In Figure 2 we show the mean personality scores across the three groups. In this dataset, PTSD users score highest on average in openness with depressed users scoring lowest. However, neuroticism is the largest separator between mentally ill users and the controls, with depressed having slightly higher levels of neuroticism than PTSD. Neuroticism alone has an ROC AUC of .732 in prediction depression vs. control and .674 in predicting PTSD vs. control. Controls score higher on extraversion, a trait related to the frequency and intensity of positive emotions (Smillie et al., 2012). Controlling for age (Figure 2b) significantly reduces the initial association between PTSD and higher conscientiousness, because PTSD users are likely to be older, and conscientiousness tends to increase with

age (Soto et al., 2011). After controlling, depressed users score lowest on conscientiousness, while PTSD and controls are close to each other.

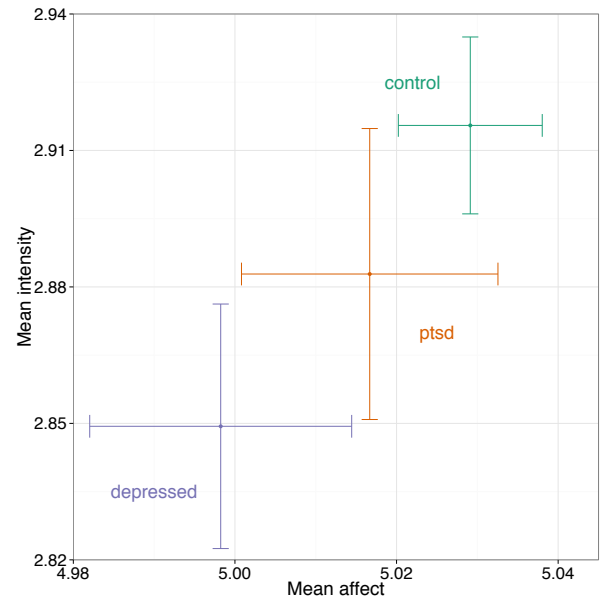


Figure 3: Users mapped on the emotion circumplex, consisting of affect (valence) and intensity (arousal).

- Average affect and intensity achieve modest predictive performance, although significant (C-D $p < .001$, D-P $p < .001$, DeLong test) when one of the compared groups are depressed. We

use the two features to map users to the emotion circumplex in Figure 3. On average, control users expressed both higher intensity and higher (i.e. more positive) affect, while depressed users were lowest on both. This is consistent with the lowered (i.e. more negative) affect typically seen in both PTSD and depressed patients, and the increased intensity/arousal among PTSD users may correspond to more frequent expressions of anxiety, which is characterized by high arousal and lower/negative affect (American Psychiatric Association, 2013).

- Textual features obtain high predictive performance. Out of these, LIWC performs the worst, while the topics, unigrams and 1-3 grams have similarly high performance.

In addition to ROC AUC scores, we present ROC curves for all three binary prediction tasks in Figures 4a, 4b and 4c. ROC curves are specifically useful for medical practitioners because the classification threshold can be adjusted to choose an application-appropriate level of false positives. For comparison, we display methods using only age and gender; age, gender and personality combined, as well as LIWC and the LDA topics.

For classifying depressed users from controls, a true positive rate of ~ 0.6 can be achieved at a false positive rate of ~ 0.2 using personality, age and gender alone, with an increase to up to ~ 0.7 when PTSD users are one of the groups. When classifying PTSD users, age is the most important factor. Separating between depressed and PTSD is almost exclusively a factor of age. This suggests that a application in a real life scenario will likely overpredict older users to have PTSD.

5 Language Analysis

The very high predictive power of the user-level features and textual features motivates us to analyse the linguistic features associated with each group, taking into account age and gender.

We study differences in language between groups using differential language analysis – DLA (Schwartz et al., 2013). This method aims to find all the most discriminative features between two groups by correlating each individual feature (1-3 gram or topic)

to the class label. In our case, age and gender are included as covariates in order to control for the effect they may have on the outcome. Since a large number of features are explored, we consider coefficients significant if they meet a Bonferroni-corrected two-tailed p -value of less than 0.001.

5.1 Language of Depression

The word cloud in Figure 5a displays the 1-3 grams that most distinguish the depressed users from the set of control users.

Many features show face validity (e.g. ‘depressed’), but also appear to represent a number of the cognitive and emotional processes implicated in depression in the literature (American Psychiatric Association, 2013). 1-3 grams seem to disclose information relating to illness and illness management (e.g. ‘depressed’, ‘illness’, ‘meds’, ‘pills’, ‘therapy’). In some of the most strongly correlated features we also observe an increased focus on the self (e.g. ‘I’, ‘I am’, ‘I have’, ‘I haven’t’, ‘I was’, ‘myself’) which has been found to accompany depression in many studies and often accompanies states of psychological distress (Rude et al., 2004; Stirman and Pennebaker, 2001; Bucci and Freedman, 1981).

Depression classically relies on the presence of two sets of core symptoms: sustained periods of low mood (dysphoria) and low interest (anhedonia) (American Psychiatric Association, 2013). Phrases such as ‘cry’ and ‘crying’ suggest low mood, while ‘anymore’ and ‘I used to’ may suggest a discontinuation of activities. Suicidal ideations or more general thoughts of death and dying are symptoms used in the diagnosis of depression, and even though they are relatively rarely mentioned (grey color), are identified in the differential language analysis (e.g. ‘suicide’, ‘to die’).

Beyond what is generally thought of as the key symptoms of depression discussed above, the differential language analysis also suggests that anger and interpersonal hostility (‘fucking’) feature significantly in the language use of depressed users.

The 10 topics most associated with depression (correlation values ranging from $R = .282$ to $R = .229$) suggest similar themes, including dysphoria (e.g. ‘lonely’, ‘sad’, ‘crying’ – Figures 6b, 6c, 6f) and thoughts of death (e.g. ‘suicide’ – Figure 6h).

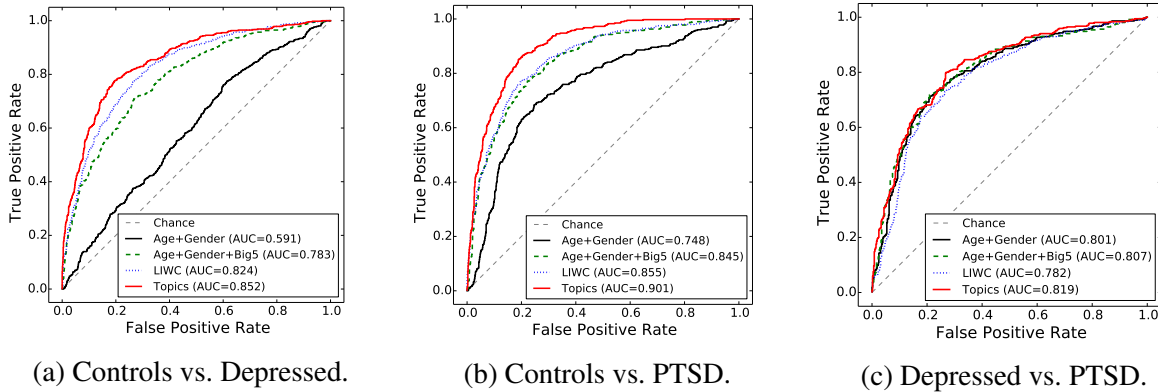


Figure 4: ROC curves for prediction using different types of features.

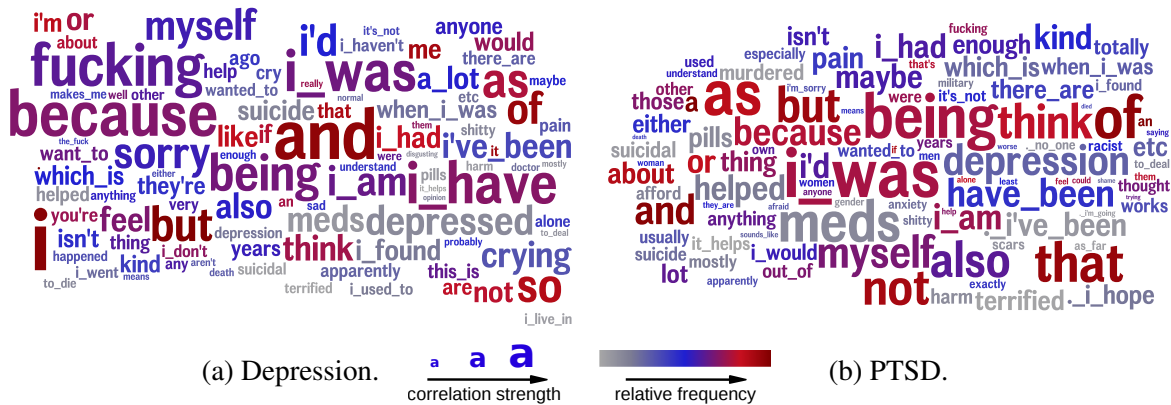


Figure 5: The word clouds show the 1-3 grams most correlated with each group having a mental illness, with the set of control users serving as the contrastive set in both cases. The size of the 1-3 gram is scaled by the correlation to binary depression label (point-biserial correlation). The color indexes relative frequency, from grey (rarely used) through blue (moderately used) to red (frequently used). Correlations are controlled for age and gender.

5.2 Language of PTSD

The word cloud in Figure 5b and topic clouds in Figure 7 display the 1-3 grams and topics most correlated with PTSD, with topic correlation values ranging from $R = .280$ to $R = .237$. On the whole, the language most predictive of PTSD does not map as cleanly onto the symptoms and criteria for diagnosis of PTSD as was the case with depression. Across topics and 1-3 grams, the language most correlated with PTSD suggests ‘depression’, disease management (e.g. ‘pain’, ‘pills’, ‘meds’ – Figure 7c) and a focus on the self (e.g. ‘I had’, ‘I was’, ‘I am’, ‘I would’). Similarly, language is suggestive of death (e.g. ‘suicide’, ‘suicidal’). Compared to the language of depressed users, themes within the language of

users with PTSD appear to reference traumatic experiences that are required for a diagnosis of PTSD (e.g. ‘murdered’, ‘died’), as well as the resultant states of fear-like psychological distress (e.g. ‘terrified’, ‘anxiety’).

5.3 PTSD and Depression

From our predictive experiments and Figure 4c, we see that language-predicted age almost completely differentiates between PTSD and depressed users. Consequently, we find only a few features that distinguish between the two groups when controlling for age. To visualise differences between the diseases we visualize topic usage in both groups in Figure 8. This shows standardised usage in both groups for each topic. As an additional factor (color), we include

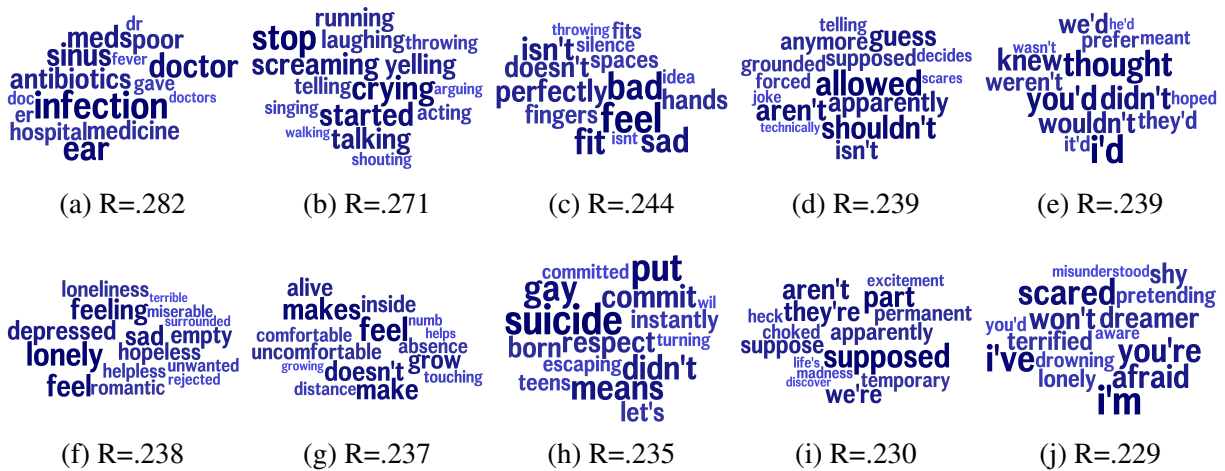


Figure 6: The LDA topics most correlated with depression controlling for age and gender, with the set of control users serving as the contrastive set. Word size is proportional to the probability of the word within the topics. Color is for display only.

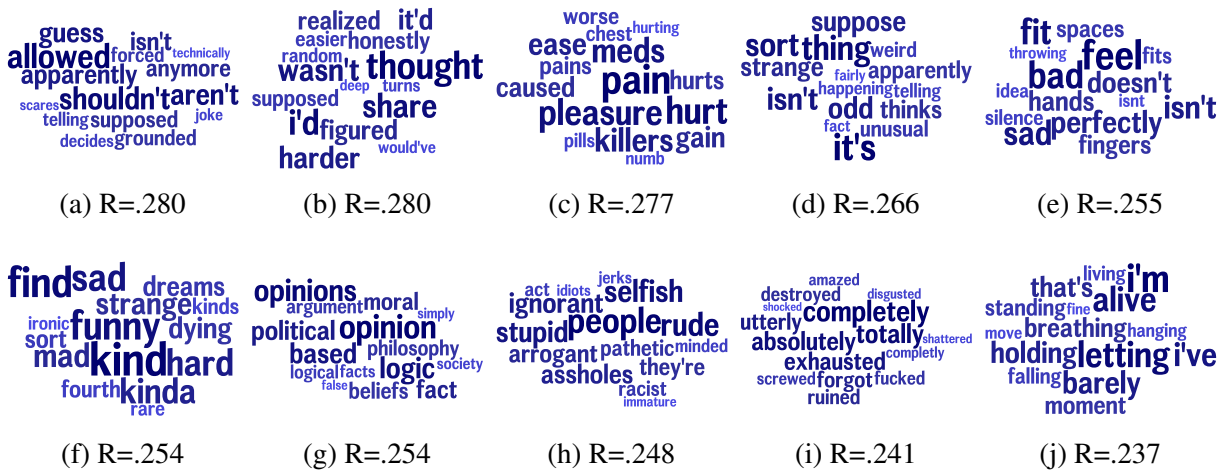


Figure 7: The LDA topics most correlated with PTSD controlling for age and gender, with the set of control users serving as the contrastive set. Word size is proportional to the probability of the word within the topics. Color is for display only.

the personality trait of neuroticism. This plays the most important role in separating between mentally ill users and controls.

The topics marked by arrows in Figure 8 are some of the topics most used by users with depression and PTSD shown above in Figures 6-7. Of the three topics, the topic shown in Figure 6h has 'suicide' as the most prevalent word. This topic's use is elevated for both depression and PTSD. Figure 6f shows a topic used mostly by depressed users, while Figure 7c highlights a topic used mainly by users with PTSD.

6 Related Work

Prior studies have similarly examined the efficacy of utilising social media data, like Facebook and Twitter, to ascertain the presence of both depression and PTSD. For instance, Coppersmith et al. (2014b) analyse differences in patterns of language use. They report that individuals with PTSD were significantly more likely to use third person pronouns and significantly less likely to use second person pronouns, without mentioning differences in the use of first person pronouns. This is in contrast to the strong differences in first person pronoun use among depressed individuals documented in the literature

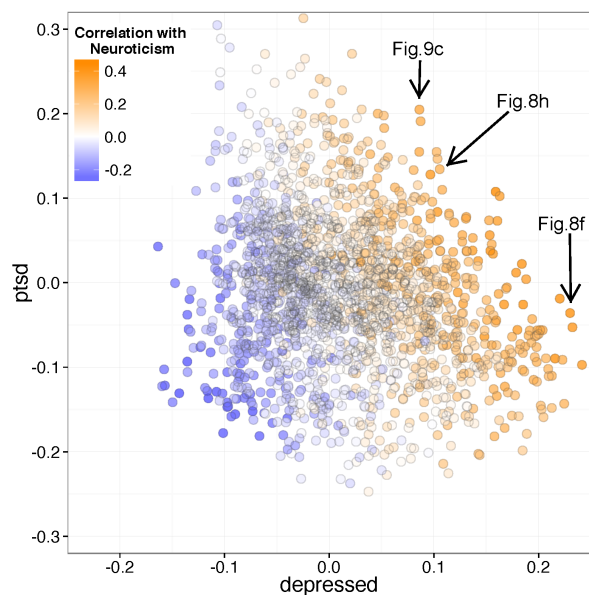


Figure 8: Topic usage (z-scored) for depressed and PTSD users. Color shows correlation of each topic to neuroticism. Labeled topics can be found in Figures 6- 7.

(Rude et al., 2004; Stirman and Pennebaker, 2001), confirmed in prior Twitter studies (Coppersmith et al., 2014a; De Choudhury et al., 2013) and replicated here. De Choudhury et al. (2013) explore the relationships between social media postings and depressive status, finding that geographic variables can alter one’s risk. They show that cities for which the highest numbers of depressive Twitter users are predicted correlate with the cities with the known highest depression rates nationwide; depressive tweets follow an expected diurnal and annual rhythm (peaking at night and during winter); and women exhibit an increased risk of depression relative to men, consistent with known psychological trends. These studies thus demonstrate the utility of using social media outlets to capture nuanced data about an individual’s daily psychological affect to predict pathology, and suggest that geographic and demographic factors may alter the prevalence of psychological ill-being. The present study is unique in its efforts to control for some of these demographic factors, such as personality and age, that demonstrably influence an individual’s pattern of language use. Further, these demographic characteristics are known to significantly alter patterns e.g. pronoun use (Pennebaker, 2011). This highlights the utility of controlling for these

factors when analysing pathological states like depression or PTSD.

7 Conclusions

This study presented a qualitative analysis of mental illness language use in users who disclosed their diagnoses. For users diagnosed with depression or PTSD, we have identified both symptoms and effects of their mental condition from user-generated content. The majority of our results map to clinical theory, confirming the validity of our methodology and the relevance of the dataset.

In our experiments, we accounted for text-derived user features, such as demographics (e.g. age, gender) and personality. Text-derived personality alone showed high predictive performance, in one case reaching similar performance to using orders of magnitude more textual features.

Our study further demonstrated the potential for using social media as a means for predicting and analysing the linguistic markers of mental illnesses. However, it also raises a few questions. First, although apparently easily predictable, the difference between depressed and PTSD users is largely only due to predicted age. Sample demographics also appear to be different than the general population, making predictive models fitted on this data to be susceptible to over-predicting certain demographics.

Secondly, the language associated with a self-reported diagnosis of depression and PTSD has a large overlap with the language predictive of personality. This suggests that personality may be explanatory of a particular kind of behavior: posting about mental illness diagnoses online. The mental illness labels thus acquired likely have personality confounds ‘baked into them’, stressing the need for using stronger ground truth such as given by clinicians.

Further, based on the scope of the applications – whether screening or analysis of psychological risk factors – user-generated data should at minimum be temporally partitioned to encompass content shared before and after the diagnosis. This allows one to separate mentions of symptoms from discussions of and consequences of their diagnosis, such as the use of medications.

References

- American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders: DSM-5*. American Psychiatric Association, 5 edition.
- Laura Andrade, Jorge J Caraveo-anduaga, Patricia Berglund, Rob V Bijl, Ron De Graaf, Wilma Vollebergh, Eva Dragomirecka, Robert Kohn, Martin Keller, Ronald C Kessler, et al. 2003. The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys. *International Journal of Methods in Psychiatric Research*, 12(1):3–21.
- David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022.
- Wilma Bucci and Norbert Freedman. 1981. The language of depression. *Bulletin of the Menninger Clinic*.
- Zhiyuan Cheng, James Caverlee, and Kyumin Lee. 2010. You are where you tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th ACM Conference on Information and Knowledge Management*, CIKM, pages 759–768.
- Glen Coppersmith, Mark Dredze, and Craig Harman. 2014a. Quantifying mental health signals in twitter. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, ACL.
- Glen Coppersmith, Craig Harman, and Mark Dredze. 2014b. Measuring post traumatic stress disorder in twitter.
- Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Social media as a measurement tool of depression in populations. In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 47–56. ACM.
- Elizabeth R. DeLong, David M. DeLong, and Daniel L. Clarke-Pearson. 1988. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*, 44(3):837–845.
- Ronald C Kessler, Patricia Berglund, Olga Demler, Robert Jin, Doreen Koretz, Kathleen R Merikangas, A John Rush, Ellen E Walters, and Philip S Wang. 2003. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *Journal of the American Medical Association*, 289(23):3095–3105.
- Ronald C Kessler, Patricia Berglund, Olga Demler, Robert Jin, Kathleen R Merikangas, and Ellen E Walters. 2005a. Lifetime prevalence and age-of-onset distributions of dsm-iv disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, 62(6):593–602.
- Ronald C Kessler, Wai Tat Chiu, Olga Demler, and Ellen E Walters. 2005b. Prevalence, severity, and comorbidity of 12-month dsm-iv disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, 62(6):617–627.
- Vasileios Lampos, Nikolaos Aletras, Daniel Preoțiuc-Pietro, and Trevor Cohn. 2014. Predicting and characterising user impact on Twitter. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, EACL, pages 405–413.
- Colin D Mathers and Dejan Loncar. 2006. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS medicine*, 3(11):e442.
- Robert R McCrae and Oliver P John. 1992. An introduction to the five-factor model and its applications. *Journal of Personality*, 60:175–215.
- Dong Nguyen, Noah A. Smith, and Carolyn P. Rosé. 2011. Author Age Prediction from Text Using Linear Regression. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, ACL.
- Bo Pang and Lillian Lee. 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends Information Retrieval*, 2(1-2):1–135, January.
- Gregory Park, H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Michal Kosinski, David J Stillwell, Lyle H Ungar, and Martin EP Seligman. 2014. Automatic personality assessment Through Social Media language. *Journal of Personality and Social Psychology*.
- Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- James W. Pennebaker, Martha E. Francis, and Roger J. Booth. 2001. *Linguistic Inquiry and Word Count*. Lawrence Erlbaum Associates.
- James W Pennebaker. 2011. The secret life of pronouns. *New Scientist*, 211(2828):42–45.
- Jonathan Posner, James A Russell, and Bradley S Peterson. 2005. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(03):715–734.
- Daniel Preoțiuc-Pietro, Maarten Sap, H Andrew Schwartz, and Lyle Ungar. 2015. Mental Illness Detection at the World Well-Being Project for the CLPsych 2015

- Shared Task. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, NAACL.
- Martin Prince, Vikram Patel, Shekhar Saxena, Mario Maj, Joanna Maselko, Michael R Phillips, and Atif Rahman. 2007. No health without mental health. *The Lancet*, 370(9590):859–877.
- Stephanie Rude, Eva-Maria Gortner, and James Pennebaker. 2004. Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8):1121–1133.
- Maarten Sap, Gregory Park, Johannes Eichstaedt, Margaret Kern, Lyle Ungar, and H Andrew Schwartz. 2014. Developing Age and Gender Predictive Lexica over Social Media. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, EMNLP, pages 1146–1151.
- H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, and Lyle H Ungar. 2013. Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PLOS ONE*.
- H Andrew Schwartz, Greg Park, Maarten Sap, Evan Weingarten, Johannes Eichstaedt, Margaret Kern, Jonah Berger, Martin Seligman, and Lyle Ungar. 2015. Extracting Human Temporal Orientation in Facebook Language. In *Proceedings of the The 2015 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies*, NAACL.
- Luke D Smillie, Andrew J Cooper, Joshua Wilt, and William Revelle. 2012. Do extraverts get more bang for the buck? Refining the affective-reactivity hypothesis of extraversion. *Journal of Personality and Social Psychology*, 103(2):306.
- Christopher J Soto, Oliver P John, Samuel D Gosling, and Jeff Potter. 2011. Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology*, 100(2):330.
- Shannon Wiltsey Stirman and James W Pennebaker. 2001. Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, 63(4):517–522.
- TB Üstün, Joseph L Ayuso-Mateos, Somnath Chatterji, Colin Mathers, and Christopher JL Murray. 2004. Global burden of depressive disorders in the year 2000. *The British Journal of Psychiatry*, 184(5):386–392.
- Svitlana Volkova, Glen Coppersmith, and Benjamin Van Durme. 2014. Inferring User Political Preferences from Streaming Communications. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, ACL.
- Hui Zou and Trevor Hastie. 2005. Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B*, 67:301–320.