Cell Phone Communication Patterns and their Relationships to Anxiety and Depression

Trevor Wexner
University of Pennsylvania
wexnert@seas.upenn.edu

Abstract

Given today's widespread adoption of cell phone technology, the relationship between cell phone use and mental health has become a crucial topic in psychological research. In this paper, we explored the difference in cell phone communication patters between people who suffer from anxiety or depression and people who do not. Specifically, we used the call and text logs of 211 individuals to construct classifiers predicting anxiety and depression. Our classifiers achieved out of sample AUC scores of .74 for anxiety and .57 for depression. Moreover, we found that there was a significant difference between the control group and non-control groups in the amount of consistent phone calls someone has with the same counterparties ($p = .0035$). The control group tended to have more consistent communication with the same individuals.

1 Introduction

Recent years have seen a substantial increase in the number of people who use smart phone technology. According to the Pew Research Center, 77% of American adults own a smart phone, up from 35% in 2011. Additionally, the center estimates that 95% of adults in the United States own a cell phone of some kind [1]. Thus, any negative impact of cell phone use on mental health could affect the vast majority of Americans. Moreover, any adverse effects of smart phone usage would have implications for a rapidly growing fraction of the population.

Thus, it has become the focus of much recent psychological literature to attempt to understand the relationships between mental health and various patterns of smart phone use [2]. These works have examined how a variety of smart phone features - from time on the smart phone, to social media use, to overall smart phone addiction - relate to psychological well-being. Numerous mental health indicators have been used, including anxiety, depression, and satisfaction with life [3],[4],[5]. The consensus across much of this work is that there is some negative association between smart phone use and mental health.

Our work aims to expand upon these analyses by examining the relationship between cell phone communication patterns and anxiety and depression. Specifically, we use call and text data to build classifiers predicting anxiety and depression. We also conduct significance testing to understand the importance of various calling and texting behaviors as they relate to these diagnoses.

Whilst previous studies have been conducted that attempt to predict depression from calling and texting behavior, these efforts have either been aimed at predicting within-person changes in depression [6] or have been smaller sample studies that do not account for consistency of communication [7]. Our study expands on these works by predicting both anxiety and depression on a larger sample as well as accounting for consistency and regularity of communication. Specifically, we consider how consistently people communicate with the same parties in addition to how regularly spaced their communications are. Accounting for consistency of communications enhances our understanding of how interpersonal and temporal phone contact varies across people with anxiety and depression.

2 Methodology

2.1 Variables and Data Processing

Data Collection

Our dataset consisted of the complete call and text logs of 211 participants. This set of individuals was enrolled in the study by Northwestern University Institutional Review Board approved protocol. Specifically, participants downloaded the "Purple Robot" app [8], which collected their phone sensor data for 6 weeks. The app captured the logs of all incoming, outgoing, and missed calls and texts for each participant during this time period. Additionally, at the outset of the study, participants filled out a survey with their basic demographic and mental health information. Depression was measured through the PHQ-9 scale [9], and anxiety was measured through the GAD-7 scale [10]. Clinically determined diagnoses were then assigned to each patient based on these scores.

Participation was limited to residents of the United States who were at least 18 years of age and literate in English. Individuals also needed to own of an Android phone and have access to Wi-Fi for at least three hours per day [11].

Variables

The raw data available for our study were the the communication logs of each participant for the duration of his/her enrollment in the study as well as corresponding demographic and mental health information. In particular, communication logs included the communication type (`Phone` or `SMS`), communication direction (`Incoming`, `Outgoing`, `Missed`, or `Other`), timestamp, and an encoded contact name related to each exchange. Mental health information included PHQ-9 and GAD-7 scores as well as a corresponding label (`Anxious`, `Depressed`, `Depressed-Anxious`, or `Control`). From this information we engineered a set of features for the analysis with the goal of capturing the different types of cell phone communication behaviors. These metrics are described in detail below.

Measures of Overall Quantity of Communication

`avg_calls_per_day`, `median_calls_per_day`, `avg_texts_per_day`, `median_texts_per_day`

Measures of Incoming Relative to Outgoing Communication

`median_outgoing_calls_per_day`
`missed_to_total_call_ratio`
`incoming_to_outgoing_text_ratio`
`incoming_to_outgoing_call_ratio`

Quantity of Texting Relative to Calling

`text_to_call_ratio`

Intraday Communication Variability

We split each day into four hour bins (i.e. `00-04, 04-08, ..., 20-24`), then examine a person's communication behavior within each bin. This temporal decomposition aims to provide insight into how someone's texting and calling behavior changes based on time of day. We choose four hour bins because they are smallest interval of time by which we could group without producing overwhelmingly sparse results. For each time bin, we consider the following variables:

`Pct_Calls`, `Pct_Texts`, `text_to_call_ratio`, `missed_to_total_call_ratio`, `incoming_to_outgoing_call_ratio`

Here the percentages denote the proportion of daily activity that occurs during a given bin. The values of all of these variables are computed each day. Then, the median is taken across all days to extract typical behavior and control for outliers.

Interday Communication Variability

`mean_call_time`, `sd_call_time`, `mean_text_time`, `sd_text_time`

Consistency of Communication with the Same People

`Num_Freq_Contacts`

This measure is defined as the number of people with whom someone communicates via a connected phone call at least 5 days per week (on average). Five days per week was chosen as the threshold for consistent communication because it was the highest frequency that yielded a distribution of values not too sparse to use.

Spacing between Communications

`mean_time_btwn_calls`, `median_time_btwn_calls`, `max_time_btwn_calls`, `mean_time_btwn_texts`, `median_time_btwn_texts`, `max_time_btwn_texts`, `avg_phone_entropy`, `median_phone_entropy`, `avg_text_entropy`, `median_text_entropy`

Missing Data

Missing values were imputed with medians, and corresponding one-hot encodings were created for each column with missing data.

2.2 Significance Testing

We then conducted significance testing on this enhanced set of features. Specifically, we examined how any above defined calling or texting behavior was significantly indicative of anxiety or depression.

Due to the large number of variables, we reduced the dimensionality of the dataset prior to testing. First, we eliminated from consideration any of the indicator variables of missing data that did not have at least 20 observations present and 20 observations missing. This elimination ensured that any remaining results were robust and not based on extremely small sample size. We then used sparse PCA to cut the number of remaining variables [12]. We did not use any supervised dimensionality reduction technique because we did not want to bias the remaining variables for testing by choosing only factors which we already observed to be correlated with outcomes. Moreover, we chose sparse PCA, rather than other unsupervised dimensionality reduction procedures, because its sparse output allowed for a more intuitive interpretation of results. We then took the non-zero variables chosen by sparse PCA and combined them with `Num-Freq-Contacts` as well as measures of spacing in communication. These added variables ensured that our initial hypothesis would be tested: that there are relationships between consistency and spacing of cell phone communication and mental health diagnoses.

We subsequently ran the below tests with the Benjamini-Hochberg correction on the reduced set of variables [13].

Test	Captures
Mann-Whitney-U Test for Anxiety and Depression	Is the distribution of a given variable the same for control group as for groups with anxiety/depression?
Mann-Whitney-U Test for Control vs Non-Control Groups	Is the distribution of a given variable the same for control group as for non-control groups?
t-Test for Significance of Correlation	Is the correlation of PHQ-9 or GAD-7 with a given variable significant? Assumes $r \sim N[0, 1]$
Kruskal-Wallis Test	Do all groups come from same distribution of a given variable?

Python's `scipy.stats` package was for all testing [14], [15]. False discovery rate was set to .05 in the Benjamini-Hochberg correction. The variables tested were:

`Num_Freq_Contacts, mean_time_btwn_texts, max_time_btwn_calls, median_time_btwn_texts, 20-24_Pct_Texts, avg_phone_entropy, median_phone_entropy, 16-20_Pct_Calls, 08-12_Pct_Texts, max_time_btwn_texts, 08-12_Pct_Calls`

2.3 Predictive Modelling

Having examined the significance of various relationships, we then moved on to our second task: using communication patterns to predict anxiety and depression. To accomplish this task, we built classifiers of the clinical labels assigned to each patient.

Because there were only 211 observations, our dataset was too small to benefit from the flexibility of deep learning. Instead, we used other common machine learning classifiers known to train well with fewer observations. In particular, we constructed random forest and logistic regression models for both anxiety and depression. We then chose the more accurate model for final use. While other techniques such as SVM, boosting, and Python's auto-ML ensemble method [16] were briefly explored, we did not pursue these in depth. We instead opted for logistic regression and random forest because of the interpretability of their results.

To construct our classifiers, we first split our dataset into 20% for testing and 80% for training. To ensure balanced training and testing sets, random sampling is stratified by diagnosis, with each stratum being divided 80/20. This same stratified random sampling method is used in all further random sampling.

Dimensionality Reduction
Because of the large number of explanatory variables relative to observations as well as the many probable redundancies among these predictors, we do dimensionality reduction on the training set. For consistency with our methodology used during significance testing, we again performed sparse PCA. In order to determine the sparsity parameter as well as the number of components to keep, we used 10-fold cross validation on the training set. A logistic regression predicting anxiety was constructed at each split to evaluate the parameters. We did not fit a separate model for depression because of the high correlation between the two diagnoses with respect to the inputs.

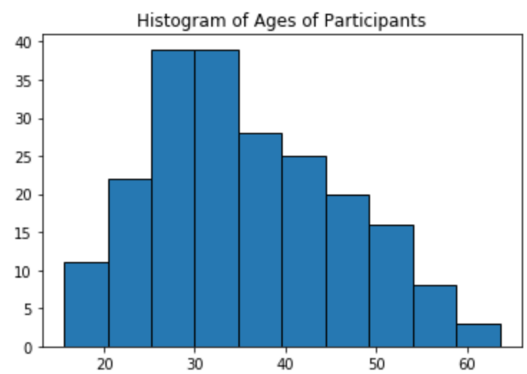
Model Construction
We then used our optimal parameters in a sparse PCA on the entire training set and reduced the size of the input space. We subsequently constructed logistic regression and random forest models for both anxiety and depression and used our test set to evaluate results.

3 Results

3.1 Descriptive Statistics

Below we highlight some descriptive trends in the dataset which provide additional insight into the underlying population and its behaviors.

Ages of the participants were roughly normally distributed with a slight left skew. The mean age was 38, minimum age 18, and maximum age was 66. A full histogram of the ages of participants can be seen below.



Roughly 81% of the participants were female, and mental health diagnoses were split roughly evenly among the four groups.

Additionally, the following were initial empirical observations based on means, medians, and correlations. Note that these were not necessarily directly tested for significance.

- The median value of text-to-call-ratio was lower in the control group than in non-control groups
- People in the control group were on average, older than people in non-control groups (41.5 years in control vs. ∼37 for others)
- There was a negative correlation between age and text-to-call ratio, which could explain the discrepancy in mental health across ages ($r = -.23$)
- People in the control group waited longer, on average, between text communications
- People with anxiety tended to have higher entropy of communication (for both texting and calling)
- People in the control group had, on average, more frequent contacts than those with anxiety or depression

- People with more frequent contacts appear to have lower GAD scores. Moreover, people with very high GAD scores do not have frequent contact (see Figure 2)

The below figures delineate some of these observations in more detail.

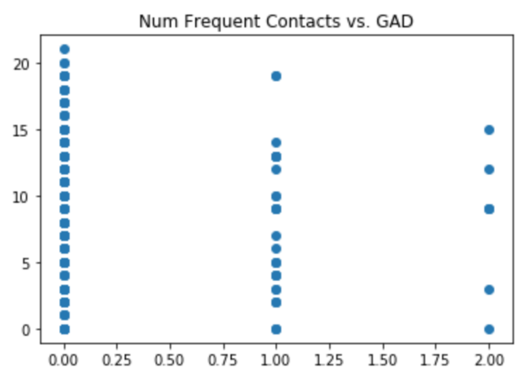
*Median Values Across Groups**

	text_to_call_ratio	AGE	mean_time_btwn_texts
CONTROL	3.317028	40.5	0.602082
DEPRESSED	4.487031	37.5	0.515000
ANXIOUS	4.697834	36.5	0.556662
DEPRESSED_ANXIOUS	4.511421	35.5	0.470133

	median_text_entropy	avg_phone_entropy	Num_Freq_Contacts
CONTROL	1.840052	0.922779	0.362069
DEPRESSED	1.852701	0.936327	0.200000
ANXIOUS	2.131270	1.136852	0.090909
DEPRESSED_ANXIOUS	2.059382	0.994910	0.211538

*Except for `Num-Freq-Contacts`, which is summarized by mean

Figure 2



3.2 Significance Testing

Sparse PCA Parameter Selection

We used only the first 3 components output by sparse PCA to ensure that a small number of variables would be considered for testing. Additionally, the first 3 principal components explained a large majority of variance in the corresponding standard PCA. We selected the sparsity parameter by choosing the value such that the size of the remaining set, when combined with consistency variables, was roughly 10. Having 10 input variables into the Benjamini-Hochberg procedure corresponds to having a smallest significant p-value of .01. This p-value is the smallest cut off for significance in many practices, so we aim to use it here as our minimum threshold in the Benjamini-Hochberg correction.

Significance Results

Of the four tests run the only significant results (at $\alpha = .05$) were from the Mann-Whitney-U test for control vs. non-control groups and the Kruskal-Walace test. Both of these tests were only significant for one variable: `Num-Freq-Contacts`. That is, the Mann-Whitney-U Test suggests that there was a significant difference between control and non-control groups in the number of frequent contacts made ($p = .0035$). The Kruskal-Wallis test backs these results, indicating that the at least one mental health cohort differs from the rest in its pattern of regular interpersonal contact ($p = .007$). These results are consistent with our preliminary empirical observation that people who have more frequent call contact with the same parties tend to be less anxious or depressed.

No significant correlations were found between any of the 11 tested variables and GAD or PHQ scores. Moreover, no significant differences were found in the distributions of any of these variables for the presence vs lack of a particular mental health diagnosis.

3.3 Classification

For the models of both depression and anxiety, logistic regression performed better out of sample than random forest ($\sim 6\%$ margin). Thus, our results below expound the performance of the logistic regression. For all models, we used a sparse PCA with sparsity parameter .025 and included the first seven components to transform the inputs. We then added `Num-Freq-Contacts` as an additional variable because of its significant relationship to outcome.

Prediction of Anxiety

Our logistic regression classifier was able to predict anxiety to 74% accuracy out of sample, which was a 21% improvement over guessing the majority class. This accuracy corresponded to an AUC score of .74. The coefficients in the model are:

Variable	Coeff
`sd-call-time`: -12.7	-0.3253
`12-16-Connected-Calls-Missing`: 4.89; `08-12-Texts-Missing`: -7.93	-0.1407
`12-16-Connected-Calls-Missing`: 12.7	-0.3253
`Outgoing-Texts-Missing`: -11.79; `12-16-Connected-Calls-Missing`: 1.03	-0.2282
`incoming-to-outgoing-call-ratio`: 4.85; `20-24-Outgoing-Texts-Missing`: -8.03	-0.2025
`20-24-Outgoing-Texts-Missing`: -12.7	-0.4019
`00-04-Incoming-Calls-Missing`: 12.7	-0.1151
`Num-Freq-Contacts`	0.4436

Here the positive prediction outcome is `NOT ANXIOUS`, so higher values of coefficient correspond to a higher probability of not having anxiety. The first seven rows correspond to the components output from sparse PCA and are displayed with original variable as well as PCA-derived eigenvalue. According to the model, the measures most predictive of anxiety were `Num-Freq-Contacts`, `20-24-Outgoing-Texts-Missing`, `12-16-Connected-Calls-Missing`, and `sd-call-time`.

All else held constant, someone having more frequent phone communication with the same parties had the most influence on someone being labelled as not anxious. Similarly, people who did not send text messages between 8:00 PM and midnight were deemed less likely to be anxious, while those who did not have any connected calls between noon and 4:00 PM were considered more likely to be anxious.

Prediction of Depression

Our out of sample predictions for depression were less accurate; our model achieved only a 7% improvement over guessing the majority class and had an AUC score of .57. Due to this only marginal improvement above guessing, we omit any analysis of the classifier's coefficients. The reliability of such an interpretation would be limited because of accuracy.

4 Discussion

Limitations

There are several limitations of this study. Some of these limitations are inherent in the dataset. For example, the scores used to measure anxiety (GAD-7 and PHQ-9) are based on self-reported survey questions. Thus, these scores may be deflated. Additionally, the vast majority of participants were female, which introduces potential bias if there are discrepancies in communication patterns across genders. Lastly, the majority of participants were over the age of 30. However, many active smart phone users are under 30 and likely have different cell phone communication habits than older cohorts. Therefore, results valid for those over 30 with anxiety might not generalize to younger generations.

There are also some limitations to our methodologies, namely in significance testing. We tested only a very small subset of all variables in our augmented dataset. This reduction ensured that the set of tested predictors was small enough to allow for meaningful results. Moreover, in selecting which variables to test, we used an unsupervised dimensionality reduction method. We did not test variables most correlated with outcomes because we did not want to bias the results. However, our lack of testing certain predictors does not imply that these factors were insignificant. For example, we empirically observed that people with lower text to call ratios appeared less likely to be depressed or anxious. This result agrees with intuition and could potentially offer further insight into the behavior of people with anxiety or depression. Nevertheless, we were unable to test this feature for significance, so its importance remains inconclusive.

Suggestions for Future Research

Per the above limitations, it would be beneficial for some follow-up study to validate these results on a younger population where the gender of participants is more evenly distributed. Additionally, future studies may want to explicitly test the effects of text-to-call ratio.

Future work may also consider more sophisticated modelling techniques. For simplicity and interpretability we used random forest and logistic regression models. However, using a clustering technique or aggregation of individual models along with some sort of interpretability heuristic (e.g. LIME [17]) might yield more accurate results while still giving an understanding of feature importance.

Conclusions

In this paper we set out to better understand the relationship between cell phone communication behavior and anxiety and depression. While we were able to gain some insight into the call and text patterns of persons suffering from anxiety, similar revelations for depression proved elusive.

An additional goal of this study was to explore whether temporal regularity and spacing of someone's communications has any relation to anxiety or depression. We did not find any significant results in this area.

Still, our ability to predict anxiety to 74% accuracy out of sample exclusively by using call and text data suggests that someone's cell phone communication behavior is highly indicative of anxiety. One of the main predictors in our model that caused someone to be labelled as not anxious was having not sent any text messages between 8:00PM and midnight. This result agrees with intuition: people who are awake later and sending text messages are potentially sleeping less, less present in their surroundings, and more fixated on their phone screen. All of these factors have been linked to anxiety [18], [19], [20].

Most importantly, the main result of our analysis is that people who do not suffer from anxiety or depression tend to have frequent and consistent phone calls with the same parties. In particular, people who spoke on the phone with more individuals an average of at least five times per week were significantly more likely to be in the control group. Similarly, the variable "number of frequent contacts" had the most influence on our model predicting someone to not have anxiety. These results are consistent with intuition; people who have regular verbal interaction with the same parties likely feel less isolated, and, therefore less depressed or anxious. Consistency of interpersonal contact does, indeed, matter for understanding anxiety.

5 Acknowledgements

This paper would not be possible without the help of Professor Lyle Ungar, Tony Liu, and Salvatore Giorgi.

6 References

1. Demographics of Mobile Device Ownership and Adoption in the United States. Pew Research Center: Internet, Science Tech. February 05, 2018. Accessed May 13, 2019. https://www.pewinternet.org/fact-sheet/mobile/.
2. Pratap, A, Atkins, DC, Renn, BN, et al. The accuracy of passive phone sensors in predicting daily mood. Depress Anxiety. 2019; 36: 72– 81. https://doi.org/10.1002/da.22822
3. Lepp, Andrew, Jacob E. Barkley, and Aryn C. Karpinski. The Relationship between Cell Phone Use, Academic Performance, Anxiety, and Satisfaction with Life in College Students. Computers in Human Behavior 31 (2014): 343-50. doi:10.1016/j.chb.2013.10.049.
4. Rosen, L. D., Whaling, K., Rab, S., Carrier, L. M., Cheever, N. A. (2013). Is Facebook creating "iDisorders"? The link between clinical symptoms of psychi-

atric disorders and technology use, attitudes and anxiety. Computers in Human Behavior, 29(3), 1243-1254.

5. Alhassan, Aljohara A., Ethar M. Alqadhib, Nada W. Taha, Raneem A. Alahmari, Mahmoud Salam, and Adel F. Almutairi. The Relationship between Addiction to Smartphone Usage and Depression among Adults: A Cross Sectional Study. BMC Psychiatry 18, no. 1 (2018). doi:10.1186/s12888-018-1745-4.

6. Pratap, A, Atkins, DC, Renn, BN, et al. The accuracy of passive phone sensors in predicting daily mood. Depress Anxiety. 2019; 36: 72– 81. https://doi.org/10.1002/da.22822

7. Ghandeharioun, Asma, Szymon Fedor, Lisa Sangermano, Dawn Ionescu, Jonathan Alpert, Chelsea Dale, David Sontag, and Rosalind Picard. Objective Assessment of Depressive Symptoms with Machine Learning and Wearable Sensors Data. 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), 2017. doi:10.1109/acii.2017.8273620.

8. Center for Behavioral Intervention Technologies. PurpleRobot. 2015.

9. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613

10. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097

11. Saeb, Sohrab, Emily G. Lattie, Konrad P. Kording, and David C. Mohr. ”Mobile Phone Detection of Semantic Location and Its Relationship to Depression and Anxiety.” JMIR MHealth and UHealth 5, no. 8 (2017). doi:10.2196/mhealth.7297.

12. Hui Zou, Trevor Hastie Robert Tibshirani (2006) Sparse Principal Component Analysis, Journal of Computational and Graphical Statistics, 15:2, 265-286, DOI: 10.1198/106186006X113430

13. Benjamini, Yoav, and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 57, no. 1 (1995): 289-300. doi:10.1111/j.2517-6161.1995.tb02031.x.

14. ”Statistical Functions (scipy.stats).” Statistical Functions (scipy.stats) - SciPy V1.2.1 Reference Guide. Accessed May 13, 2019. https://docs.scipy.org/doc/scipy/reference/stats.html.

15. ”Scipy.stats.pearsonr.” Scipy.stats.pearsonr - SciPy V1.2.1 Reference Guide. Accessed May 13, 2019. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.pearsonr.html#scipy.stats.pearsonr.

16. Feurer et al. Efficient and Robust Automated Machine Learning. Advances in Neural Information Processing Systems 28 (NIPS 2015).

17. Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. Why should I trust you?: Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM (2016).

18. Alhassan, Aljohara A., Ethar M. Alqadhib, Nada W.Taha, Raneem A. Alahmari, Mahmoud Salam, and AdelF. Almutairi. The Relationship between Addiction toSmartphone Usage and Depression among Adults: A Cross Sectional Study. BMC Psychiatry 18, no. 1(2018). doi:10.1186/s12888-018-1745-4.

19. ”Sleep Disorders.” Anxiety and Depression Association of America, ADAA. Accessed May 13, 2019. https://adaa.org/understanding-anxiety/related-illnesses/sleep-disorders.

20. Hoge, Elizabeth A., Eric Bui, Luana Marques, Christina A. Metcalf, Laura K. Morris, Donald J. Robinaugh, John J. Worthington, Mark H. Pollack, and Naomi M. Simon. ”Randomized Controlled Trial of Mindfulness Meditation for Generalized Anxiety Disorder.” The Journal of Clinical Psychiatry 74, no. 08 (2013): 786-92. doi:10.4088/jcp.12m08083.