Current and Future Psychological Health Prediction using Language and Socio-Demographics of Children for the CLPysch 2018 Shared Task

Sharath Chandra Guntuku[1,3]**, Salvatore Giorgi**[2] **and Lyle Ungar**[1,2,3]
[1]School of Medicine, University of Pennsylvania
[2]Department of Psychology, University of Pennsylvania
[3]Computer & Information Science, University of Pennsylvania
{sharathg@sas,sgiorgi@seas,ungar@cis}.upenn.edu

Abstract

This article is a system description and report on the submission of a team from the University of Pennsylvania in the 'CLPsych 2018' shared task. The goal of the shared task was to use childhood language as a marker for both current and future psychological health over individual lifetimes. Our system employs multiple textual features derived from the essays written and individuals' socio-demographic variables at the age of 11. We considered several word clustering approaches, and explore the use of linear regression based on different feature sets. Our approach showed best results for predicting distress at the age of 42 and for predicting current anxiety on Disattenuated Pearson Correlation, and ranked fourth in the future health prediction task. In addition to the subtasks presented, we attempted to provide insight into mental health aspects at different ages. Our findings indicate that misspellings, words with illegible letters and increased use of personal pronouns are correlated with poor mental health at age 11, while descriptions about future physical activity, family and friends are correlated with good mental health.

1 Introduction

Studying early markers of well-being is a significant emerging frontier in child development research, examining the strengths, assets and abilities to establish positive developmental trajectory for children (Masten and Coatsworth, 1998). Humans are affected by experiences early in their childhood in ways that shape their life course. Language can be very useful in predicting well-being in the short term (Schwartz et al., 2013b). Predictions about the long-term future using language is rather unexplored by the NLP community, and can aid a variety of applications aimed at the understanding of early life markers and development of preventative care.

The CLPsych 2018 shared task explores the predictive ability of language to elucidate a person's long-term well-being. The competition uses a corpus of individuals, who were surveyed at various points in their life since their birth to monitor their health and socioeconomic status. At age 11, the participants wrote short essays on where they saw themselves at age 25, fourteen years in the future; these essays are used to predict aspects of their mental health, measured by depression syndrome, anxiety syndrome, and the total Bristol Social Adjustment Guide (BSAG) score (Stott and Sykes, 1963). The two sub tasks are to predict these aspects of a) current mental health at age 11 (Task A), and b) future mental health at ages 23, 33, and 42 (Task B). Additional non-linguistic variables, including gender and childhood parental social class were also provided.

For our participation in this shared task, we treat the task as a regression problem using standard regularised linear regression algorithm (i.e. Ridge Regression). We use a wide range of automatically derived textual features (based on word clustering and other pre-trained models) to obtain different representations of the language used by individuals. Our regression model returns a continuous score for each aspect of mental health for each individual. The results are measured on Disattenuated Pearson Correlation (shown as r_{disatt} in the results of our paper) between the predictions and the actual survey outcomes. This metric is similar to a Pearson correlation, but it accounts for measurement error and thus yields values with larger variance. The measurement error (accounted for by its inverse, reliability) is taken from the literature on the reliability of the psychological distress questionnaires (0.77; (Ploubidis et al., 2017)) and of similar language-based predictions (0.70; (Park et al., 2014)). The metric is thus:

$$r_{disatt} = \frac{r_{Pearson}}{\sqrt{.77 * .70}} \qquad (1)$$

Parameter	Train	Test
Number of individuals	9218	1000
Female	49.12%	47.1%
Professional occupations	4.95%	5.9%
Managerial and technical occupations	15.71%	18.1%
Skilled non-manual occupations	8.23%	8.8%
Skilled manual occupations	51.73%	48.4%
Partly-skilled occupations	14.39%	14.6%
Unskilled occupations	4.94%	4.2%

Table 1: Descriptive statistics of socio-demographics at age 11 for the individuals in training and test datasets.

In addition to the shared task we also looked at characterizing language for each mental health indicator using both open and closed vocabulary approaches.

2 System Overview

In our approach, we aggregate the word counts in all of an individual's posts, irrespective of the word order within (a bag-of-words approach). Each individual in the dataset is thus represented by a distribution over words. We then use automatically derived groups of co-occurring words (or 'topics') to obtain a lower dimensional distribution for each individual. These topics, built using automatic clustering methods from separate large datasets, capture a set of semantic and syntactic relationships (e.g. words reflecting depression, pronouns etc). In addition, we use the socio-demographics of each individual.

2.1 Data

This study has undergone IRB ethics review at the University of Pennsylvania and has been deemed exempt. The shared task uses data from the National Child Development Study (Davie et al., 1972), which is a British birth cohort study following an initial 17,416 babies born in Britain in one week in March 1958. The study was augmented in subsequent childhood sweeps by immigrants to Great Britain born in the studys target week, bringing to the total NCDS sample to 18,558. Surviving members of this birth cohort have been surveyed on eight further occasions in order to monitor their changing health, education, social and economic circumstances, of which the data for ages 11, 23, 33 and 43 are shared in this task.

When the children of the NCDS were eleven years old in 1969 they were asked to write an es-

Statistic / Outcome	Mean	Std. Dev	[Min-Max]
Age 11: BSAG Score	8.07	8.70	[0 - 61]
Age 11: Anxiety	0.53	1.18	[0 - 12]
Age 11: Depression	1.00	1.51	[0 - 10]
Age 23: Distress	0.93	1.46	[0 - 9]
Age 33: Distress	0.70	1.37	[0 -9]
Age 42: Distress	1.03	1.62	[0 - 9]

Table 2: Descriptive statistics of mental health aspects at multiple ages for the individuals in the training dataset.

	Age 11: Depression	Age 11: BSAG	Age 23: Distress	Age 33: Distress	Age 42: Distress
Age 11: Anxiety	.12	.37	.05	.04	.04
Age 11: Depression		.71	.05	.05	.04
Age 11: BSAG			.06	.05	.03
Age 23: Distress				.39	.32
Age 33: Distress					.44

Table 3: Pearson inter-correlations between mental health aspects at multiple ages for the individuals in the training dataset. All correlations are significant at $p < .05$, Benjamini-Hochberg corrected, two-tailed t-test.

say about what they thought their life would be like at age 25. 10,511 essays were then restored and transcribed from historic records (see (Davie et al., 1972) for details of the transcription process). The statistics of both the training and test datasets shared, which excludes any essays that contained fewer than 50 words, are presented in Table 1. The descriptive statistics of the mental health outcomes for the training dataset are presented in Table 2. The inter-correlations between mental health aspects at multiple ages are shown in Table 3.

3 Features and Methods

We briefly summarize the features used in our prediction task. The entire pipeline of feature extraction, out of sample prediction (for the shared task) and language insights used the Differential Language Analysis ToolKit (DLATK) Python package (Schwartz et al., 2017).

3.1 Features

Unigram Features (unigrams) We use unigrams as features in order to capture a broad range of textual information. First, we tokenized the essays into unigrams using a modified version of Chris Potts' *HappyFunTokenizer* (Manning et al., 2014) which captures social media content such

as emoticons and hashtags[1]. We use the unigrams mentioned by at least 1% of individuals in the training set, resulting in 1,147 features (out of 55,486 features).

UnigramMeta   After extracting unigrams, we calculate two meta features for each individual: a) average length of unigrams, and b) number of unigrams per essay. These features were shown to predict depression in social media individuals (Guntuku et al., 2017c).

Word2Vec Word Clusters (W2V)   Neural methods have recently been gaining popularity in order to obtain low-rank word embeddings and obtained state-of-the-art results for a number of semantic tasks (Mikolov et al., 2013b). These methods, like many recent word embeddings, also allow to capture local context order rather than just 'bag-of-words' relatedness, which leads to also capture syntactic information. We use the skip-gram model with negative sampling (Mikolov et al., 2013a) to learn word embeddings from a corpus of 400 million tweets also used in (Lampos et al., 2014). We use a hidden layer size of 50 with the Gensim implementation.[2] We then apply spectral clustering on these embeddings to obtain hard clusters of words. We create 200 hard clusters i.e. one word can belong to only one topic. The importance score associated with every word represents how central the word is in its cluster. Clusters are computed using spectral clustering over a word-word similarity matrix generated by Word2Vec. These features were shown to predict income and personality of users on social media (Lampos et al., 2014; Guntuku et al., 2017a). These clusters are available online[3].

LDA Word Clusters (LDA)   A different type of clustering is obtained by using topic models, most popular of which is Latent Dirichlet Allocation (Blei et al., 2003). LDA models each post as being a mixture of different topics, each topic representing a distribution over words, thus obtaining soft clusters of words. We use the 2000 clusters introduced in (Schwartz et al., 2013a), which were computed over a large dataset of posts from 70,000 Facebook users. These features were

shown to predict multiple user traits like depression (Schwartz et al., 2014), personality (Schwartz et al., 2013a), other demographic and psychological traits (Jaika et al., 2018) on social media. These clusters are available online[4]

Linguistic Inquiry and Word Count (LIWC)   LIWC (Pennebaker et al., 2007) is a dictionary comprising 64 different categories (e.g., topical categories, emotions, parts-of-speech) which are manually constructed based on psychological theory. We use LIWC to represent the language of each individual as normalized frequency distributions of these categories, by counting the words associated with each category for each user and normalizing them based on the total number of words that the user posted . These features were shown to predict user traits across multiple modalities such as essays, social media and blogs (Boyd and Pennebaker, 2017). LIWC has also been used to understand the relationship between a persons social media activities and real life behaviors, such as substance use (Ding et al., 2017).

NRC Emotion Lexicon (NRCEmot)   The NRC Emotion Lexicon (Mohammad and Turney, 2013) is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). The annotations were manually done by crowdsourcing. We use NRC Lexicon to represent the language of each individual as normalized frequency distributions of these emotions.

Personality   We used automatic text-regression methods (Schwartz et al., 2013a) to assign to each individual scores on the Big Five personality traits. This personality model was trained on a sample of over 70,000 Facebook users, using tokens and topics extracted from status updates as features, achieving a validation predictive performance of $r = 0.35$ on average for all five traits. Personality is shown to influence multiple user attributes such as likes (Guntuku et al., 2016a), emotions (Guntuku et al., 2015a,b) and mental health (Guntuku et al., 2017b).

Socio-Demographics   We used the gender and social class of children collected at the age of 11 as additional features.

Current Psychological Health				
Feature	Age 11/ Metric	Anxiety (BSAG)	Depression (BSAG)	Total BSAG Score
LIWC	r_{disatt}	0.154	0.305	0.407
	MAE	0.757	1.089	6.369
LDA	r_{disatt}	0.130	0.329	0.430
	MAE	0.756	1.080	6.313
NRCEmot	r_{disatt}	0.041	0.154	0.203
	MAE	0.763	1.113	6.658
Personality	r_{disatt}	0.030	0.103	0.130
	MAE	0.766	1.118	6.749
SocioDemographics	r_{disatt}	0.073	0.243	0.307
	MAE	0.764	1.106	6.554
W2V	r_{disatt}	0.168	0.317	0.387
	MAE	0.754	1.091	6.428
unigramsMeta	r_{disatt}	0.107	0.265	0.323
	MAE	0.761	1.103	6.544
unigrams	r_{disatt}	0.152	0.370	0.477
	MAE	0.750	1.072	6.241

Table 4: Performance (measured by Disattenuated Pearson Correlation, r_{disatt} and Mean Absolute Error, MAE) of different features at predicting current mental health aspects (Task A).

3.2 Methods

Task A and B We stratified individuals into five-folds. In this five-fold cross validation setting, we tried linear regression with ridge regularization. We used the implementation from Scikit-Learn (Pedregosa et al., 2011) which uses Stochastic Gradient Descent for inference. Parameter tuning plays a vital role in good performance of regression algorithms. We measure Pearson correlation on our training set using 5 cross-fold validation and optimize parameters using grid search for each feature set individually. The performance was measured by calculating Disattenuated Pearson's Correlation r_{disatt} and Mean Absolute Error (MAE) over the aggregated predictions from the five-folds.

Language Insights In addition to Task A and B we also tried to identify language that characterizes each of the mental health outcomes using both an open and closed vocabulary approach. For the open vocabulary approach we used Differential Language Analysis (DLA) (Schwartz et al., 2013a). Here we individually correlate the unigram features against each of our outcomes (age 11 anxiety, depression and BSAG score, age 23 distress, age 33 distress and age 44 distress) via ordinary least squares regression. We only considered unigrams used by at least .1% of users (5,457 total features).

For the closed vocabulary approach we used

LIWC categories and applied the same analysis (univariate correlations via ordinary least squares regression). In both approaches we added gender as a covariate in the regression model but this produced few (or zero) significant ($p < 0.05$) results for distress outcome at various ages. We also applied a Benjamini-Hochberg correction (Benjamini and Hochberg, 1995) to the significance threshold in order to compensate for multiple comparisons.

4   Results and Discussion

Task A The results of our methods at predicting current mental health on a cross-validation setting are presented in Table 4.

For total BSAG score, unigrams show the best performance followed by LDA clusters, LIWC and Word2Vec clusters. It is interesting that both LDA and Word2Vec clusters perform well, even though trained on datasets from a different modality than essays (i.e. social media). unigram-Meta and SocioDemographic features rank next in performance, which is interesting considering they are a very low dimensional representation. For Depression, the performance of different features is relatively similar with the exception that Word2Vec clusters have marginally better performance than LIWC. Predicting Anxiety yields the lowest performance of all three aspects of mental health, with minor changes in rank order of different features.

NRCEmot and language predicted Personality features do not perform well, specifically for predicting Anxiety, possibly because the difference in both the modality on and the time at which these features are built when compared to the essays being analyzed. NRCEmot was primarily developed for identifying emotion-related words on Twitter. The huge difference in the language of Twitter and essays written by the children in this sample would have led to poor generalisation of NRCEmot. The Personality model was also built on another social media platform – Facebook; considering the time period in which the model was built and that in which the essays were written, drift in language (Biber and Finegan, 1989; Jaidka et al., 2018; Wijaya and Yeniterzi, 2011) apart from modality differences would have led to poor generalization of the feature space.

At the time of submission, we did not evaluate the performance of unigram features, and sub-

Future Psychological Health				
Feature	Distress at/ Metric	Age 23	Age 33	Age 42
LIWC	r_{disatt}	0.152	0.066	0.088
	MAE	1.074	0.948	1.202
LDA	r_{disatt}	0.226	0.141	0.134
	MAE	1.067	0.934	1.206
NRCEmot	r_{disatt}	0.075	–	–
	MAE	1.074	0.948	1.203
Personality	r_{disatt}	0.102	–	0.019
	MAE	1.075	0.951	1.201
SocioDemographics	r_{disatt}	0.325	0.215	0.207
	MAE	1.053	0.918	1.201
W2V	r_{disatt}	0.213	0.128	0.130
	MAE	1.066	0.940	1.203
unigramsMeta	r_{disatt}	0.056	–	0.027
	MAE	1.075	0.951	1.2
unigrams	r_{disatt}	0.234	0.134	0.140
	MAE	1.067	0.937	1.206

Table 5: Performance (measured by Disattenuated Pearson Correlation, r_{disatt} and Mean Absolute Error, MAE) of different features at predicting future distress (Task B).

mitted the predictions from LDA topics for total BSAG score and Depression, and prediction from Word2Vec clusters for Anxiety on the test set.

Task B The results of our methods at predicting future mental health on a cross-validation setting are presented in Table 5. Predicting future distress is a much tougher task when compared to predicting current mental health aspects, as also seen by the performance metrics.

Surprisingly SocioDemographics outperform all other language features in the prediction of future distress. Socio economic status is known to affect health over individual's life course as suggested by prior research (Smith, 2007), and in this cohort it is seen to outperform the language of essays that children wrote about their impression of their future self.

Among language features, performance of predicting distress worsens with increase in the time from when the child wrote the essays and the time at which the prediction is being made (i.e. r_{disatt} at Age 23 > r_{disatt} at Age 33 $\simeq$ r_{disatt} at Age 42). For predicting distress at Age 23 and 42, unigrams rank best followed by LDA and Word2Vec clusters. For Age 33, LDA clusters outperform unigrams and W2V. Also it should be noted that the mental health aspects at age 11 and not strongly correlated with the mental health aspects at age 23 and 33 (Table 3) which potentially indicate that

the linguistic characteristics of the essays that the children wrote at age 11 might not be able to accurately reflect their future mental health.

Considering the complexity of the task involved, it can be hypothesized that the relationship between the language features and the outcomes is non-linear, potentially consisting of multiple latent variables. Using stacked auto-encoders to capture the non-linearity in the task could potentially improve the modeling performance (Guntuku et al., 2016b). Further, simpler text selection/categorization techniques like representing all misspelled words/words not in a dictionary/punctuation by a single category might be worth exploring, thereby reducing the feature space to consist of dimensions which contribute to the modeling task (Preoţiuc-Pietro et al., 2017).

Language Insights Table 6 shows the inter-correlations between meta-language features, mental health and socio-demographics. Here we see that higher social classes are correlated (significantly, though with a low effect size) with increased word usage and increase word length (Ling, 2005). All age 11 mental health measures are negatively correlated with word length and word totals. Males have higher depression and BSAG at age 11 while females have higher distress at age 23, 33 and 42.

Figure 1 shows the results of our open vocabulary approach (DLA). Here color represents the words frequency in the corpus (darker for more frequent) and size represents correlation strength. Misspelled words like 'will', 'wen', 'marid', 'mared', 'old' are associated bad psychological health at age 11, while words like 'house', 'saturday', 'friends', 'playing' are associated with the language of those with good psychological health (Ginsburg et al., 2007). Language of individuals with bad psychological health at age 11 is also associated with words containing letters which were illegible to transcribe (as indicated by ∗), and several spelling errors ('marid', 'mared', 'houes', 'gow') which are not found in language of mentally healthier children (Crum et al., 1993). It is interesting that the words 'and' and 'will' seem like low-hanging fruit for validating this approach.

Distress at ages 23 and 33 is positively correlated with daily activities of life 'shopping', 'hairdresser', 'sewing', 'school' whereas words associated with sports 'football', 'training', 'cricket', 'boat' etc are negatively correlated with distress

Words correlated with bad psychological health

Words correlated with good psychological health

Age 11 (a) Anxiety   (b) Depression   (c) Total BSAG score   Distress at (d) Age 23   (e) Age 33

Figure 1: Unigrams correlated with anxiety, depression, BSAG score and distress at each age. All correlations are significant at $p < .05$, Benjamini-Hochberg corrected, two-tailed t-test. The top row shows words which are positively correlated with high scores and the bottom row shows words which are negatively correlated with high scores on anxiety, depression, BSAG and distress.
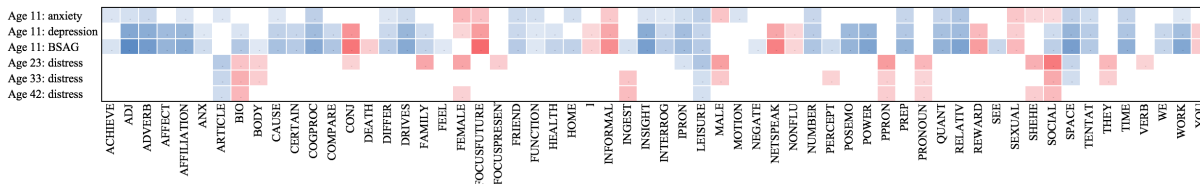

Figure 2: LIWC categories correlated with anxiety, depression, BSAG score and distress at each age. All correlations are significant at $p < .05$, Benjamini-Hochberg corrected, two-tailed t-test. Blue cells show a negative correlation, red show positive correlations while white cells are not significant.

	Avg unigram length	Total unigrams	SC I: Professional	SC II: Mangerial	SC IIIN: Non-manual	SC IIIM: Manual	SC IV: Partly-skilled	SC V: Unskilled	Female
Age 11: Anxiety	-.04	-.05		-.04		.02	.02		
Age 11: Depression	-.10	-.13	-.06	-.09	-.04	.04	.05	.08	-.08
Age 11: BSAG	-.12	-.15	-.07	-.10	-.05	.05	.06	.08	-.15
Age 23: Distress		.03	-.04	-.04			.04	.02	.23
Age 33: Distress		.02	-.03					.03	.16
Age 42: Distress		.03							.15
Avg unigram length		.04	.05	.06	.03	-.02	-.05	-.06	.03
Total unigrams			.04	.04	.03	-.03	-.03	-.03	.22

Table 6: Pearson inter-correlations between mental health aspects, socio-demographics and meta-language features. All correlations are significant at $p < .05$, Benjamini-Hochberg corrected, two-tailed t-test.

(Ortega et al., 2008). It is interesting that the words 'husband', 'school' etc. were associated with good psychological health at age 11 whereas they are associated with bad psychological health at age 23. There is very little research in the NLP community on the language markers of future mental health and this shared task opens up this promising line of research. It should be noted that these are words from the language of essays that the children at age 11 wrote to the prompt:

'Imagine you are now 25 years old. Write about the life you are leading, your interests, your home life and your work at the age of 25. (You have 30 minutes to do this).' (Power and Elliott, 2005). It is interesting that several insights about their future mental health can be gleaned using responses to such prompts.

The results of the LIWC analysis are in Figure 2. Here red cells are positively correlated with the outcome (more distress, anxiety, etc.),

blue cells are negatively correlated (less distress, anxiety, etc.) and white cells are not significant after correction for multiple comparisons. Here we see 'posemo', 'family' and 'affiliation' are all protective at age 11 (Kellam et al., 1977). Bad mental health is associated with both the 'i' and 'informal' categories at age 11 with pronoun usage and with pronoun usage at older ages. While 'leisure' is protective at all ages, no categories are associated with mental illness at every age. This is consistent with the linguistic manifestation of several mental health conditions (e.g. depression (Schwartz et al., 2014; **?**)).

5 Conclusions

This paper reported on the participation of a team from the University of Pennsylvania in the CLPsych 2018 shared task on identifying current and future mental health of children based on language from essays they wrote.

Our methods were based on linear regression using different types of word clusters. The methods we presented were designed to be as task agnostic as possible, and thus, our approach showed best results for predicting distress at the age of 42 and for predicting current anxiety on Disattenuated Pearson Correlation, and ranked fourth in the future health prediction task. Our method did not perform well compared to other teams in predicting current mental health. Fitting more complex non-linear models might have yielded better performance for that subtask. It is interesting that SocioDemographic features outperformed all language features in predicting future distress. Next, normalized word counts (unigrams) performed best at most subtasks. In addition to the subtasks presented, we attempted to provide insight into mental health aspects at different ages. Our findings show that a) mental health aspects at age 11 correlate poorly with mental health at ages 23 and 33 for the children in this cohort; b) males have higher depression scores when compared to females at age 11, while females have higher distress at ages 23, 33 and 42; c) mental health measures are negatively correlated with word length and total number of words used in the essay; d) misspellings, words with illegible letters and increased use of personal pronouns ('I') are correlated with poor mental health at age 11, while descriptions about future physical activity, family and friends are correlated with good mental health.

For future work, since the Socio Demographic performed best, we could apply methods such as User-Factor Adaptation which focus on the author of the content in addition to the content (Lynn et al., 2017; Zhu et al., 2018). It would also be interesting to investigate if word clusters trained on historical sources (for e.g. Google books) might yield reliable feature representations when studying mental health aspects at different ages to emulate the linguistic associations of elderly, for whom data from other platforms such as social media is be scarce.

References

Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300.

Douglas Biber and Edward Finegan. 1989. Drift and the evolution of english style: A history of three genres. *Language*, pages 487–517.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022.

Ryan L Boyd and James W Pennebaker. 2017. Language-based personality: a new approach to personality in a digital world. *Current Opinion in Behavioral Sciences*, 18:63–68.

Rosa M Crum, James C Anthony, Susan S Bassett, and Marshal F Folstein. 1993. Population-based norms for the mini-mental state examination by age and educational level. *Jama*, 269(18):2386–2391.

Ronald Davie, Neville Butler, and Harvey Goldstein. 1972. From birth to seven: the second report of the national child development study.(1958 cohort). *London Longmans 1972. 198 p. 1 ref*.

Tao Ding, Warren K Bickel, and Shimei Pan. 2017. Multi-view unsupervised user feature embedding for social media-based substance use prediction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2275–2284.

Kenneth R Ginsburg et al. 2007. The importance of play in promoting healthy child development and maintaining strong parent-child bonds. *Pediatrics*, 119(1):182–191.

Sharath Chandra Guntuku, Weisi Lin, Jordan Carpenter, Wee Keong Ng, Lyle H Ungar, and Daniel Preoţiuc-Pietro. 2017a. Studying personality through the content of posted and liked images on twitter. In *Proceedings of the 2017 ACM on web science conference*, pages 223–227. ACM.

Sharath Chandra Guntuku, Weisi Lin, Michael James Scott, and Gheorghita Ghinea. 2015a. Modelling the Influence of Personality and Culture on Affect and Enjoyment in Multimedia. ACII.

Sharath Chandra Guntuku, J Russell Ramsay, Raina M Merchant, and Lyle H Ungar. 2017b. Language of adhd in adults on social media. *Journal of attention disorders*, page 1087054717738083.

Sharath Chandra Guntuku, Michael James Scott, Huan Yang, Gheorghita Ghinea, and Weisi Lin. 2015b. The cp-qae-i: A video dataset for exploring the effect of personality and culture on perceived quality and affect in multimedia. In *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*, pages 1–7. IEEE.

Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. 2017c. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, 18:43–49.

Sharath Chandra Guntuku, Joey T Zhou, Sujoy Roy, Lin Weisi, and Ivor W Tsang. 2016a. Who likes what, and why? insights into personality modeling based on imagelikes'. *IEEE Transactions on Affective Computing*.

Sharath Chandra Guntuku, Joey Tianyi Zhou, Sujoy Roy, Weisi Lin, and Ivor W Tsang. 2016b. Understanding deep representations learned in modeling users likes. *IEEE Transactions on Image Processing*, 25(8):3762–3774.

Kokil Jaidka, Niyati Chhaya, and Lyle Ungar. 2018. Diachronic degradation of language models: insights from social media. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*.

Kokil Jaika, Sharath Chandra Guntuku, and Lyle H Ungar. 2018. Facebook vs. twitter: Cross-platform differences in self-disclosure and trait prediction. In *ICWSM*.

Sheppard G Kellam, Margaret E Ensminger, and R Jay Turner. 1977. Family structure and the mental health of children. *Archives of General Psychiatry*, 34(9):1012–1022.

Vasileios Lampos, Nikolaos Aletras, Daniel Preoţiuc-Pietro, and Trevor Cohn. 2014. Predicting and Characterising User Impact on Twitter. EACL.

Rich Ling. 2005. The sociolinguistics of sms: An analysis of sms use by a random sample of norwegians. In *Mobile communications*, pages 335–349. Springer.

Veronica Lynn, Youngseo Son, Vivek Kulkarni, Niranjan Balasubramanian, and H Andrew Schwartz. 2017. Human centered nlp with user-factor adaptation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1146–1155.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.

Ann S Masten and J Douglas Coatsworth. 1998. The development of competence in favorable and unfavorable environments: Lessons from research on successful children. *American psychologist*, 53(2):205.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. In *Proceedings of Workshop at the International Conference on Learning Representations*, ICLR.

Tomas Mikolov, Wen tau Yih, and Geoffrey Zweig. 2013b. Linguistic Regularities in Continuous Space Word Representations. In *Proceedings of the 2010 annual Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL, pages 746–751.

Saif M Mohammad and Peter D Turney. 2013. Nrc emotion lexicon. *NRC Technical Report*.

FB Ortega, JR Ruiz, MJ Castillo, and M Sjöström. 2008. Physical fitness in childhood and adolescence: a powerful marker of health. *International journal of obesity*, 32(1):1.

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*, 108(6):934–952.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine Learning in Python. *JMLR*, 12.

James W Pennebaker, Roger J Booth, and Martha E Francis. 2007. Linguistic inquiry and word count: Liwc [computer software]. *Austin, TX: liwc. net*.

GB Ploubidis, A Sullivan, M Brown, and A Goodman. 2017. Psychological distress in mid-life: evidence from the 1958 and 1970 british birth cohorts. *Psychological medicine*, 47(2):291–303.

Chris Power and Jane Elliott. 2005. Cohort profile: 1958 british birth cohort (national child development study). *International journal of epidemiology*, 35(1):34–41.

Daniel Preoţiuc-Pietro, Sharath Chandra Guntuku, and Lyle Ungar. 2017. Controlling human perception of basic user traits. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 2335–2341.

H Andrew Schwartz, Johannes Eichstaedt, et al. 2014. Towards assessing changes in degree of depression through facebook. In *CLPsych*.

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, et al. 2013a. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):e73791.

H Andrew Schwartz, Salvatore Giorgi, Maarten Sap, Patrick Crutchley, Lyle Ungar, and Johannes Eichstaedt. 2017. Dlatk: Differential language analysis toolkit. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 55–60.

Hansen Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Richard E Lucas, Megha Agrawal, Gregory J Park, Shrinidhi K Lakshmikanth, Sneha Jha, Martin EP Seligman, et al. 2013b. Characterizing geographic variation in well-being using tweets. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM.

James P Smith. 2007. The impact of socioeconomic status on health over the life-course. *Journal of Human Resources*, 42(4):739–764.

Denis Herbert Stott and Emily G Sykes. 1963. *Bristol Social-adjustment Guides*. University of London Press.

Derry Tanti Wijaya and Reyyan Yeniterzi. 2011. Understanding semantic change of words over centuries. In *Proceedings of the 2011 international workshop on DETecting and Exploiting Cultural diversiTy on the social web*, pages 35–40. ACM.

Yi Zhu, Sharath Chandra Guntuku, Lin Weisi, Gheorghita Ghinea, and Judith A Redi. 2018. Measuring individual video qoe: A survey, and proposal for future directions using social media. *ACM Transactions on Multimedia Computing, Communications and Applications*.