The University of Chicago

Assessing Personality Traits Through Social Media Language

By

Chun Hu

May 2021

A paper submitted in partial fulfillment of the requirements for

the Master of Arts degree in the Master of Arts in

Computational Social Science

Faculty Advisor: Jon Clindaniel

Preceptor: Elizabeth L Huppert

Abstract

With the emergence of online social media, people increasingly leverage these platforms to build connections and express themselves. In particular, our language reflects the increasing emphasis on personal agency and expression. In this paper, we explored whether language can be a strong predictor of Big Five personality traits on Yelp, a site in which there is no expectation of building a socially desirable profile as do other social media. The linear models of lexical correlates suggest that users that produce high-quality reviews show higher Extraversion, Agreeableness, Openness, Conscientiousness, and lower Neuroticism. Feature importance suggests that low-quality reviews show a higher degree of egocentricity while high-quality reviews are more socially oriented.

Keywords: Big Five, social media, LIWC, topic modeling, LDA

Assessing Personality Traits Through Social Media Language

Recent years have seen a dramatic change in the way people communicate and obtain information. Consumption of traditional media is declining, while online social media has become the main source of information. Increasingly, people rely on digital platforms to build connections and stay informed of what's trending throughout the world. Facebook, for instance, maintains over 1.85 billion daily active users in 2020[1].

Inevitably, the changes in the way of social interaction lead to new form of self-expression. While these platforms provide people with opportunities to construct social networks tailored to their own needs and interests, they also put the challenges of self-presentation on the same screens. These include an increasing emphasis on the expression of personal agency and identity. On the bright side, social media provide people with control over their presentations of self in social interactions and the opportunity to form social networks that solicit positive feedbacks. However, on the other hand, the pressure to present a socially desirable profile has never been greater (Holtgraves, 2004). The unique formulation of social contexts online provides social scientists with interesting data to examine the experiences of social practices.

Social Grooming

Originally a word to describe the social activity in which animals bond and reinforce their social structure and build relationships, "social grooming" has now been used to describe how people seek social information and interaction using social media and networking sites (Tufekci, 2008). In typical social grooming, people maintain friendships, read public posts in their networks, keep track of comments and interactions, and show interests of social events. Social grooming is closely associated with our perception of self. For instance, researchers have shown that through social grooming, students gain positive emotional support and increase their sense

[1] https://www.statista.com/statistics/346167/facebook-global-dau/

of well-being (Suphan & Mierzejewska, 2016). Grooming behaviors are showcased in developing close emotional relationship and trust in social bonding.

In particular, language plays a key role in the expression of self in social grooming behaviors. In order to broadcast a polished self-image, people learn to speak what is appropriate and valued by the social media audience. In social interactions, Fu & Lee (2007) has shown that preschoolers have already learned how to articulate feelings and thoughts about others, and to display flattery languages and behaviors with an emerging sensitivity to the social contexts. This suggests that humans have an instinctive tendency to manipulate their communication based on social context. In the digital world, a recent study that investigates uncivil language on social media has found that social network size is a negative predictor of incivility (Kim, 2020). Specifically, researchers found that Twitter users with a larger network size tend to use fewer uncivil remarks when they engage in more positive interactions. Hence, it is not surprising that our language on social media has become a main source of data for social scientists to study our self-expression and identity formation. Moreover, language provides the possibility of building novel measurement methods. If researchers can quantify language, they can provide us with rich information on human activities at an unexpected scale and scope.

Literature Review

Research on social media languages has surged with the emergence of computational text analysis. In psychology, researchers are interested in finding language correlates with personality. Most studies have focused on modeling the relation at a broad level, such as using broad personality domains such as the Big Five (Yarkoni, 2010). The Big Five theory is based on the semantic associations between words and psychological perspectives, and classifies personality into five dimensions: Openness to experience (inventive/curious vs.

consistent/cautious), Conscientiousness (efficient/organized vs. extravagant/careless), Extroversion (outgoing/energetic vs. solitary/reserved), Agreeableness (friendly/compassionate vs. challenging/callus), and Neuroticism (sensitive/nervous vs. resilient/confident). People who score high in Openness enjoy new experiences and have a wide range of interests. They see themselves as original, artistic, and have an active imagination. People who score high in Conscientiousness tends to show good impulse control and act dutifully. They are achievement-oriented, well organized, and mindful of details. People with high Extraversion are more assertive, sociable, and energetic. They are described as outgoing and gain energy from social interactions. People with higher Agreeableness tend to be friendly, affectionate, cooperative, and altruistic. They also tend to be more compassionate. Last but not least, Neuroticism is associated with one's emotional stability and degree of negative emotion. People who score high on Neuroticism often experience emotional instability and unpleasant emotions like sadness, anger, and anxiety. These people are prone to mood swings and experience high levels of stress.

There are a variety of measures and tests around the effectiveness of Big Five personality traits. Cobb-Clark & Schurer (2012) demonstrated that the Big Five model is stable for working-age adults over a four-year time period. During the time period, mean personality changes are small and constant across different age groups. The Big Five model has been widely used to measure the personality traits in various domains including online communities, health (Orji et al., 2017), e-commerce (Huang & Yang, 2010), and job performance (Barrick & Mount, 1991). All studies showed a significant true score correlation and validity, and implications for future research and practice.

Big Five and Social Media

Research has shown a strong and consistent association between language and the Big Five personality traits. Specially, the use of language is reflective of personality dimensions. Some of the earliest findings indicate that individuals who are high on Neuroticism use first-person singulars (e.g. I, me, mine) more frequently, and those who score high on Extraversion use more positive emotion words (Pennebaker & King, 1999). Recent studies have confirmed the associations on social media. Gregory et al. (2015) compared their language-based personality results with self-reported assessments for more than 66,000 Facebook users. They found stable patterns of correlations over 6-month intervals. Schwartz et al. (2013) found similar associations with 75,000 Facebook users, and reported that people scored higher on Neuroticism tend to use negative phrases disproportionately.

Other research has suggested that Big Five personality traits are reflected in the use of social media. Seidman (2013), for instance, showed in a 184-undergraduate Facebook study that self-presentation behaviors and motivations were best predicted by low conscientiousness and high neuroticism, controlling demographic variables. Conscientious people are cautious in online self-presentation, while neuroticism is strongly associated with presenting an ideal self-image. Ryan & Xenos (2011) pointed out that Facebook itself is a social phenomenon and represents a particular type of personality traits. The study investigated the differences in personality influences among users and non-users of Facebook for 1324 adults and found out that Facebook users are more extraverted and narcissistic that non-users. Facebook non-users are more conscientious. This illustrated that social grooming behavior itself is indicative of the personality traits. An investigation on consumer services on SNS further suggested that extraversion is positively associated with one’s publishing proportion and neuroticism is positively related to the expression of angers in posts (Bai, Zhu & Cheng, 2012). In conclusion, personality factors are strongly related to individuals' use of social networking sites, and the language in social media use is reflective of the Big Five personality traits.

Our Study

Research on social media languages and behaviors shed light on the novel measurement methods for personality assessments. The benefits of building language-based models are two-fold. First, interactions on social media are more reflective of natural social activities than in experiment settings, which provide invaluable data into human behaviors. Second, a comprehensive self-reported personality assessment often takes 15-20 minutes to complete; this is often impractical in social media domains. However, a language-based model can approximate personality traits within seconds, which enables efficient and cost-effective large-scale analysis.

While the technique of language-based assessment has been validated on various social media and social networking sites, it has not been applied to other social platforms. An example would be Yelp, one of the most influential customer review sites in the U.S. Yelp users are constantly encouraged to participate and interact in their location-based community, and they provide feedback to others' reviews through upvotes and comments. However, what differentiates Yelp from other social media is that there is no expectation of building a socially desirable profile and often users can go completely anonymous. That means, grooming behaviors are alleviated on the platform.

As discussed earlier, grooming behaviors lead to certain style of languages. With no expectation of social grooming, languages on Yelp are expected to be different from that on other social media. According to Kim (2020), fewer social grooming factors and smaller social network size lead to a higher tendency of incivility when commenting or posting. Social networking site users also become less expressive without grooming factors. We expect reviews on Yelp to be more honest reflection of Big Five personality traits through their languages. In addition, if the language-based assessment proves to be effective on these social platforms, it would have wider applicability on other online communities as well. Hence, the primary goal of this research is to investigate whether language is a strong predictor of Big Five personality traits on Yelp.

Methods

An ideal approach of the investigation would be to compare the personality assessment by language-based models and the self-reported Big Five measures. However, it is unrealistic to obtain the external assessments because user identifiers are removed in the Yelp dataset. Hence, this paper intends to be an exploratory analysis of the association between language and personality on this platform. In addition, this paper will examine the important language features that distinguish between high-quality and low-quality Yelp reviews. The intuition is that users that produce high-quality and those that product low-quality reviews will have different personality traits.

Dataset

Yelp releases annual open datasets about businesses, reviews, and users. The 2020 datasets (pre-pandemic) include information about 209,000 businesses across 10 metropolitan areas, along with over 8,000,000 user reviews. We chose the pre-pandemic (before March) dataset to exclude the impact of the outbreak on the investigation. There are two relevant datasets. The review dataset contains full review text data including the user id who wrote the review and the business id for the business that the review is written about. The review dataset also includes the number of useful, funny, and cool votes each review received. The user dataset includes the user's friend mapping and other metadata about the user. Specifically, each user is described with the number of reviews written, the number of fans, the years in *Elite* status, the average rating of all reviews, and the number of compliments received. The more stars and compliments a user received for the reviews, the higher chance he/she will be rewarded with the *Elite* status.

Research has found significant correlations between user-related features and their Big Five personality traits (Kosinski et al., 2014). Take an online profile as an example, the number of likes and the number of friends and joined groups are positively correlated to scores on Openness and negatively correlated to Conscientiousness. Hence, the user-related features in the Yelp dataset could be indicative of users' personality.

Prediction Labels

We select two groups of users based on their review performance and profile status. The first group of users is considered to produce "low-quality" reviews and is defined by the following criteria:

1. The user has made at least two reviews in the database. This criterion is used to prevent bot-generated accounts.
2. The average rating of his/her reviews is less than the lower quantile of possible scores.
3. The user has received no compliments (e.g. upvotes, stars, useful rating…) for his/her reviews.

The second group of users is considered to produce "high-quality" reviews and is defined by the following characteristics:

1. The average rating of his/her reviews is higher than the upper quantile of possible scores.
2. The user has received at least one-year *Elite* status.

The information of the select groups of users will be merged with their reviews in the review dataset based on user ids. Our goal here is to construct two datasets that differentiate maximally in their features and reviews; this allows us to have efficient computation and comparison between their languages and Big Five personality traits. As mentioned earlier, the intuition behind the criteria is that users that produce "high-quality" reviews will have different Big Five personality traits than those that produce "low-quality" reviews. Note that the low-quality and high-quality here is based on the quantile of possible scores and relevant user-related features. It is possible that certain reviews in the "low-quality" groups actually received fair ratings. However, since their average rating is less than the lower quantile of possible scores, the reviews are not valid enough on this platform. We expect that the quality of language is indicative of users' personality, and thus the reviews written by the two groups of users will show varying Big Five personality traits.

Language Correlates with Personality

The first part of the analysis is to compute linear models of Big Five personality traits through correlations with language features. Most psychological studies have used a closed-vocabulary, word counting approach to analyze the lexical components of language. The most popular implementation is the Linguistic Inquiry and Word Count (LIWC) tool (Tausczik & Pennebaker, 2010), which has shown to be effective on social media language analysis (Golbeck, Robles, & Turner, 2011). This tool counts word frequencies for over 50 psychologically relevant categories; a list of categories and variables used in our model are shown in Table 1. LIWC2007 dictionary contains 4,500 words and word stems divided into 55 categories. Since this approach uses predefined categories of words, it is considered to be a "closed-vocabulary" method.

LIWC has been widely used to analyze languages in social communities. Bazelli et al. (2013) explored the Big Five personality traits of users on Stack Overflow, the largest developer community that shares programming knowledge, and found that the top contributors in the community scored high on Extraversion. Researchers studying Facebook, for instance, also confirmed the correlation with LIWC categories. They showed that people scored high on Neuroticism use more acronyms and first-person singulars, and those scored high on Openness use more quotations (Sumner et al., 2011).

Large-scale data from social media also enabled the investigation of language in single-word use. Yarkoni (2010) investigated Big Five personality traits as a function of single word use and identified pervasive correlations of personality and LIWC categories with 406 Google bloggers. The categories that have a significant correlation with the Big Five traits are used in our model (Table 2). I chose to use correlations in this study because the language of bloggers is most similar to that of our investigation. Since bloggers were free to write topics of their choice without the pressure of social demands and they were usually not aware of being analyzed in relation to personality, their data also provides a naturalistic view of the influence of personality on language. While the specific correlation values are slightly different across studies, the identified categories converged strongly with other findings. Neuroticism is positively correlated with negative emotion words, including anxiety/fear, sadness, and anger. Conversely, Extraversion is positively correlated with positive emotions and social interactions. Similarly, Agreeableness is positively correlated with community-related language and positive emotion, e.g. family and friends, and negatively correlated with negative emotion words and swear words.

Classification of Reviews

We train classifiers separately using lexical and topical features to see whether we can differentiate between low-quality and high-quality reviews. Lexical features are described previously using the word count of LIWC categories. Topical features are constructed using topic modeling. Topical features are represented as clusters of words that tend to appear in similar contexts. The *Latent Dirichlet Allocation* (LDA) model identifies clusters of semantically related words and learns a set of topics in an unsupervised fashion. For instance, one example of a topic identified by LDA in our language sample includes words such as burger, wing, serve, seat, and wine. These words often co-occur with each other and so are defined in the same group.

The number of topics is approximated using the Silhouette analysis, which was originally used to select the optimal number of clusters of unsupervised clustering algorithms. The Silhouette method measures how close each point in one cluster is to points in the neighboring clusters and produces a number in the range of [-1, 1]. A higher value indicates that the sample is well matched to its own cluster, and vice versa. Topical features are constructed through the probability of the user to use the topic. Park et al. (2015) have defined the probability as:

$$\rho(topic, user) = \sum \rho(topic|word) \times \rho(word|user)$$

where $\rho(topic|word)$ is the probability of topic given the word and $\rho(word|user)$ is the user's use of that word.

The LDA model has been used widely to capture psychological processes in social media language. Past research has built a 50-topic LDA model on roughly 3 million tweets data to find the linguistic signal for detecting depression (Resnik et al., 2015). The results are promising: with a recall rate of 75%, 3 of 4 individuals who are clinically diagnosed with depression are successfully detected. The LDA model uncovers meaningful latent structures in their languages.

Other research has found the relations between Big Five personality traits and topic preferences through massive social network information data (Liu, Wang & Jiang, 2016). Their model suggests significant correlations in personality-specific topics that cannot be identified in the closed-vocabulary form. They reported a negative correlation between Conscientiousness and entertainment topics, a positive correlation between Extraversion and travel, and between Neuroticism and horror movies. In our research, while we are unable to verify the correlation between topics and Big Five personality traits, we examine whether low-quality and high-quality reviews have different distributions of topics, which may be indicative of different personality traits.

Models

The primary goal of this paper is twofold: (1) compute linear models for each Big Five personality trait using the attested correlations with LIWC categories and compare the personality traits for the two classes of reviews, (2) build classifiers based on lexical and topical features and determine important features that differentiate between high-quality and low-quality reviews. For the latter task, the dataset is divided into training and test set with the test set ratio equals 0.25. Three classifiers – Logistic Regression, Support Vector Classifier, and Random Forest – are trained with careful hyperparameter tuning and 10-fold cross-validation.

Past work looking at languages on social media often used Logistic Regression and Support Vector Classifier as the learning algorithm. Terentiev and Tempest (2014) have trained the two classifiers to differentiate high scoring comments under posts on Reddit. Their results suggest that Support Vector Machine with a linear kernel and L1 regularization achieved the highest training set accuracy (99.38%) and Logistic Regression achieved the highest test set accuracy (84.08%). Random Forest Classifier was also shown to be a good predictor for pair-wise comparison of text, with a test set accuracy at 95.13% (Simpson).

The performance of models is evaluated based on accuracy rate, F1, and AUC score. The accuracy rate measures the percentage of reviews that have the same predicted classes as their designated classes by true score. F1 score is a sum of precision and recall of the performance on the test set, and the AUC score represents the degree of separability and tells how much our model is capable of distinguishing between two classes of reviews at different probability thresholds.

Results

Preliminary Analysis

Based on the criteria to create "high-quality" and "low-quality" groups, we identified 199,331 users that produce low-quality reviews and 566,099 users that produce high-quality reviews. 190,000 users were randomly selected for each group after shuffling the dataset to ensure balanced classes. The user information is merged with their reviews based on user id. The final data include 59,892 low-quality reviews and 59,098 high-quality reviews. Note that each user produced at least 2 reviews, so the number of reviews is higher than the number of users selected.

Language Correlates with Personality

We computed linear models of the Big Five traits using LIWC correlates. Our models suggest that there are significant differences in personality traits between users who wrote the two classes of reviews. On average, users that produce high-quality reviews receive higher scores on Extraversion than those who write low-quality reviews (High-quality: M = 0.036, SD = 0.016; Low-quality: M = 0.030, SD = 0.015, $t(44)$ = -46.48, p = .000), higher scores on

Agreeableness (High-quality: M = 0.073, SD = 0.019; Low-quality: M = 0.066, SD = 0.020, t(44) = -44.60, p = .000), higher scores on Conscientiousness (High-quality: M = -0.017, SD = 0.012; Low-quality: M = -0.025, SD = 0.014, t(44) = -61.96, p = .000), higher scores on Openness (High-quality: M = -0.010, SD = 0.027; Low-quality: M = -0.104, SD = 0.026, t(44) = -17.56, p = .000), and lower scores on Neuroticism (High-quality: M = 0.021, SD = 0.013; Low-quality: M = 0.028, SD = 0.015, t(44) = 53.31, p = .000). All differences are significant at $p < .001$. In particular, Openness reveals the biggest difference between two groups.

Lexical Model

Three classifiers were trained using the lexical features and the performance is summarized in Table 3. SVC with a non-linear kernel obtained the highest accuracy, F1 score, and AUC score. All classifiers achieved an accuracy rate of over 75% for both training and test datasets. This suggests that our models are able to distinguish between high-quality and low-quality reviews based on the lexical components. In addition, all classifiers obtained an F1 score of over 78%. The F measure looks at the weighted average of precision and recall, and our high score indicates a well-trained classification model. An average AUC score of over 77% also indicates that our models are good at distinguishing between the two classes of reviews at different probability thresholds.

In order to understand the important features that help distinguish between the two classes, we calculated the permutation feature importance. The function looks at the drop in the model accuracy score when a single feature value is randomly shuffled and is calculated with different permutations of the feature. The top 3 features for the classifiers are: i, pronoun, and you (Figure 1). According to Table 2, i and pronoun are negatively correlated with Openness (r =

-0.16); you is positively correlated to Extraversion (*r* = 0.16) and negatively correlated to Neuroticism (*r* = -0.15). These are salient traits that distinguish between the two groups of users.

Topical Model

The *Latent Dirichlet Allocation* (LDA) model from Gensim package was used for topic modeling. The Silhouette method suggests that the optimal number of topics is 6, while the clustering score of 0.004 indicates that the sample is very close to the decision boundary between two neighboring classes. This suggests that the language in the two classes of reviews are not distinctively different in their topics. This is confirmed in the performance of model (Table 3). All classifiers trained with topical features achieved an average accuracy rate of just above 50%. SVC with a non-linear kernel and Random Forest obtained test accuracy at around 57%. The low F1 score and AUC score also confirmed that the models can only distinguish between the topics in two classes of reviews just a bit higher than chance.

Feature importance reveals that for both Logistic Regression and SVC, topic 1 is the main feature that distinguishes the two classes (Figure 1). The top 10 words in this topic are amaze, vega, locate, excite, mussel, invite, freeze, grit, stumble, and refresh. One explanation is that topic 1 makes up the largest proportion of topics in the reviews – it contributes to 21.75% of the topics in low-quality reviews and 26.05% of the topics in high-quality reviews. Hence, more linguistic features are extracted in this topic. Another reason could be that the two classes of reviews have diverse discussions on words on this topic. Most of the top words express positive emotion and are centered around food and services.

Discussion

Our lexical model is able to distinguish between high-quality and low-quality reviews with a relatively high accuracy rate. All models achieved an average accuracy rate of over 75%

on the test set. Among the three models, SVC with a non-linear kernel has the best performance based on the evaluation metrics. The linear models of the Big Five personality traits suggest different linguistic and psychological processes in the languages of these two classes of reviews. Users that produce high-quality reviews show higher Extraversion, Agreeableness, Conscientiousness, Openness, and lower Neuroticism than those that produce low-quality reviews.

The important lexical features that help classify the two classes of reviews are first-person singular (e.g. I, me, mine), pronoun, and second person (you, your). Interestingly, words that represent linguistic processes obtain higher weights than those that have psychological processes. A potential reason could be that personal pronouns are frequently used in speech and function to distinguish between the speaker and others (McGregor, 2010). Hence, the difference between the two types of users is more salient. First person singular (I, me, mine) is positively correlated with Neuroticism and negatively correlated with Openness. This confirms previous findings that individuals scored higher on Neuroticism tend to use more first-person singular more frequently (Pennebaker & King, 1999). This also suggests that users that produce low-quality reviews are more self-centered in their reviews.

On the other hand, second person (you, your) is positively correlated with Extraversion (r = 0.15), and negatively correlated with Neuroticism (r = -0.15) and Openness (r = -0.16). Since individuals higher on Extraversion tend to use more positive emotion words and those lower on Neuroticism use fewer negative phrases, we conclude that individuals that use second person more frequently have more positive phrases in their reviews. This suggests that users that produce high-quality reviews talk about the other person more frequently and use more positive phrases. Another important feature that contributes to our classification is first person plural (e.g.

we, us, our). These phrases are positively correlated to Extraversion and Agreeableness. Individuals higher on Agreeableness use more communal and positive emotion words, and this further confirms the difference in personality traits between the two groups of users.

Hence, in summary, the feature important method in our lexical classifier suggests that users that produce high-quality reviews show higher Extraversion, Agreeableness, and lower Neuroticism. This confirms the result of our linear models. In terms of their uses of pronouns, low-quality reviews suggest a higher degree of egocentricity (i.e., use of first-person singular) while high-quality reviews suggest more social orientation (i.e., use of first plural pronouns and second person). The sense of egocentricity let individuals single themselves out and take credit for the event instead of addressing the contribution of the group, which could be one of the reasons that the reviews received extremely low ratings (McGregor, 2010).

On the other hand, our topical model is only able to distinguish between the two classes of reviews just above chance. SVC with a non-linear kernel and Random Forest obtained the highest test accuracy at around 57%. This suggests that low-quality and high-quality reviews have similar topics despite their ratings. The most distinctive topic has a collection of words about food, location, and services. However, it is unclear what exactly these words refer to. The other topics are also mixed in word combinations. This is contrary to previous research that has successfully employed the LDA model to examine the psychological processes and social network information on social media.

One reason that we did not manage to detect meaningful topics could be that we included reviews about all kinds of businesses on Yelp, including arts & entertainment, education, restaurant, hotels, public services, and etc. Hence, the reviews are so diverse in topics that there is no distinctive pattern. Previous research on Yelp reviews was able to extract meaningful subtopics when they restricted the business category to “Restaurant” (Huang, Rogers & Joo, 2014). With 158,000 restaurant reviews, which is 2% of our total reviews, researchers managed to identify 8 relevant topics in the text: lunch, healthiness, décor, service, location, value, and different food categories.

The diverse reviews in our dataset hinder us from detecting meaningful correlations with users’ Big Five personality traits. Past research has suggested that food-related personality traits of food involvement and food neophobia are indicative of customer’s satisfaction and loyalty. In addition, these traits interact with novelty seeking intentions and influence food consumption and subsequent satisfaction (Eertmans et al., 2005). Hence, in future research, if we are able to narrow down the range of businesses to obtain a more focused list of topics, we may enable automated detection of customer sentiments and related business activities.

Conclusion and Future Work

In this research, we explored the possibility to assess Big Five personality traits through languages in Yelp reviews. We constructed two groups of reviews that vary in ratings and quality and demonstrated that we are able to distinguish between the two classes based on lexical features. These lexical features describe the linguistic and psychological processes within the languages and are closely correlated with the personality traits. Combining the results of linear models and classification models, we conclude that users that produce high-quality reviews show higher Extraversion, Agreeableness, Openness, Consciousness, and lower Neuroticism traits than users that produce low-quality reviews.

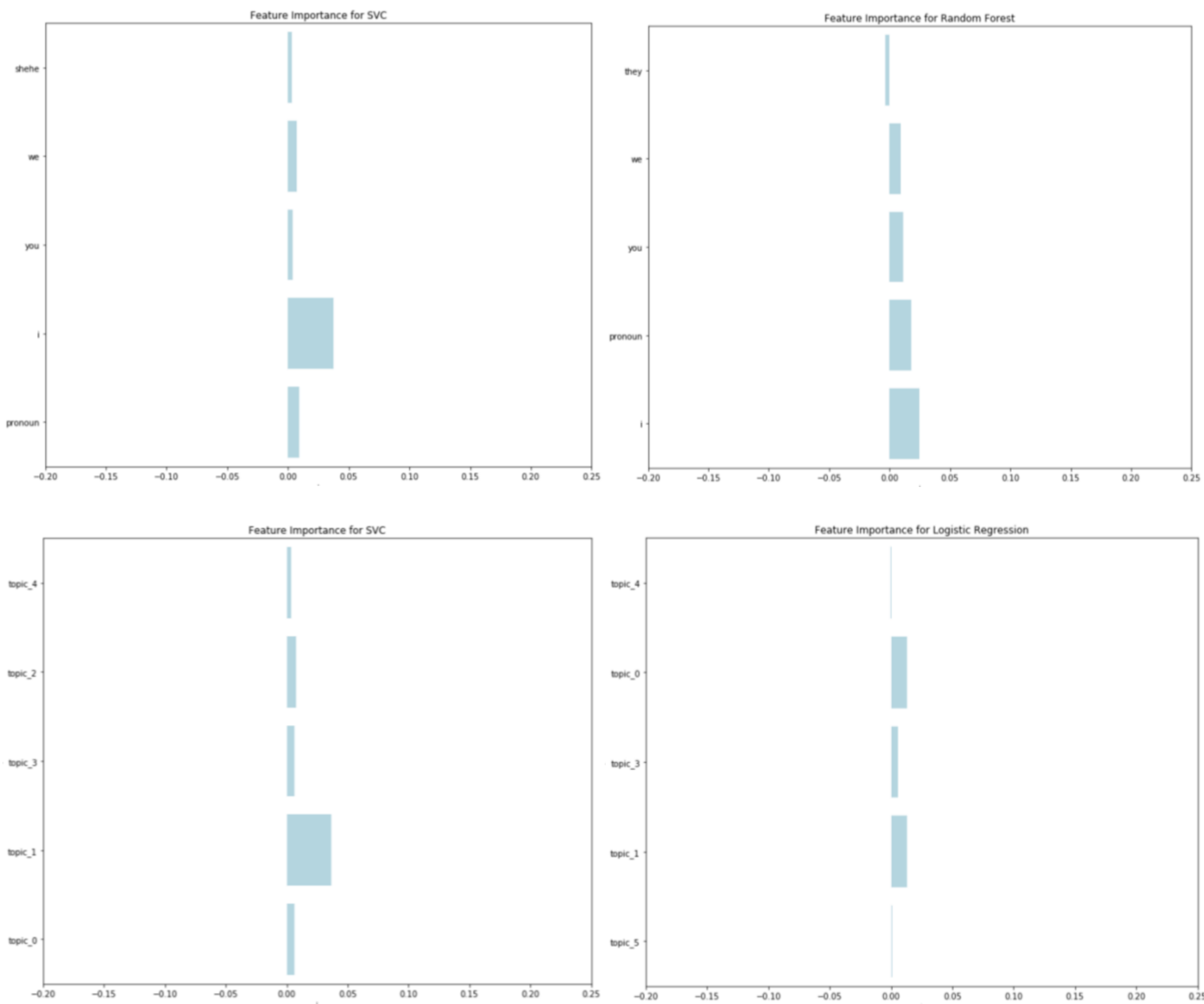
We also show that pronouns (i.e., first-person singular, second person) play a key role in classifying the two classes of reviews. Specifically, users that produce low-quality reviews show more egocentric characteristics while users that produce high-quality reviews are more socially oriented. While it is surprising that linguistic features are more prominent than other psychological processes, e.g. emotion words, swear words, in their contribution to predicting Big Five personality traits, the reason might be that the pronouns occur more frequently in our dataset. We encourage future research to examine the mechanisms of linguistics processes to learn more about the underlying processes. The linguistic features are also indicative of narrower personality dimensions. For instance, a high degree of egocentricity could indicate narcissism and mental disability and thus the detection of self-centered language could be helpful in psychology researches. In addition, future work should look further into the correlations between psychological processes and Big Five personality traits to provide more insights into language in reviews.

We also suggest potential reasons why our topical model did not manage to distinguish between the two classes of reviews. One possibility is that there may be no distinctive pattern in topics with all the business categories. Our most distinguishing topic touched upon several categories, including food, location, services, and feelings. In addition, we did not have the chance to validate the correlation between our topics and Big Five personality traits due to the lack of self-reported assessments. Future researchers, if they have the capability to obtain external personality measures, can dive into the topic by finding out the associations and validating our results. Last but not least, food-related personality is also an interesting topic to investigate for Yelp businesses. Automated detection of customer sentiments and satisfaction by personality traits can provide invaluable data for businesses to gain real-time feedback and determine the best strategies.

Our primary goal in this research was to explore whether language can be a strong predictor of Big Five personality traits on Yelp, a site in which there is no expectation of building a socially desirable profile as do other social media. While we trained our models exclusively on Yelp, they can be applied to other social platforms where the social media effect is downplayed (e.g. Reddit, Foursquare, etc.). With the ever-increasing digital trace humans now leave on online social media sites, we believe that digital text data will become one of the most important tools to help us digest social networks and human behaviors, and thus will provide us with more valuable insights in future research.

Figure 1.

Feature Importance for Lexical and Topical Classifiers



Note: *Top Left*: SVC lexical model; *Top Right*: Random Forest lexical model; *Bottom Left*: SVC topical model; *Bottom Right*: Logistic Regression topical model. The top features for lexical models are i, pronoun, and you. The top feature for topical models is topic_1.

Table 1.

LIWC 2007 Category Variable Information

Category	Abbreviation	Examples
Linguistic Processes		
Total pronouns	pronoun	I, them, itself
1st pers singular	i	I, me, mine
1st pers plural	we	We, us, our
2nd person	you	You, your, thou
Articles	article	A, an, the
Past tense	past	Went, ran, had
Present tense	present	Is, does, hear
Negations	negate	No, not, never
Numbers	number	Second, thousand
Swear words	swear	Damn, piss, fuck
Psychological Processes		
Social processes	social	Mate, talk, they, child
Family	family	Daughter, husband, aunt
Friends	friend	Buddy, friend, neighbor
Affective processes	affect	Happy, cried, abandon
Positive emotion	posemo	Love, nice, sweet
Negative emotion	negemo	Hurt, ugly, nasty
Anxiety	anx	Worried, fearful, nervous
Anger	anger	Hate, kill, annoyed
Sadness	sad	Crying, grief, sad
Cognitive processes	cogmech	Cause, know, ought
Causation	cause	Because, effect, hence
Discrepancy	discrep	Should, would, could
Tentative	tentat	Maybe, perhaps, guess
Certainty	certain	Always, never
Inhibition	inhib	Block, constrain, stop
Inclusive	incl	And, with, include
Exclusive	excl	But, without, exclude
Perceptual processes	percept	Observing, feeling
See	see	View, saw, seen

Hear	hear	Listen, hearing
Feel	feel	Feels, touch
Biological processes	bio	Eat, blood, pain
Body	body	Cheek, hands, spit
Sexual	sexual	Horny, love, incest
Space	space	Down, in, thin
Time	time	End, until, season
Personal Concerns		
Work	work	Job, majors, xerox
Achievement	achieve	Earn, hero, win
Leisure	leisure	Cook, chat, movie
Home	home	Apartment, kitchen, family
Money	money	Audit, cash, owe
Religion	relig	Altar, church, mosque
Death	death	Bury, coffin, kill
Spoken Categories		
Assent	assent	Agree, OK, yes

Table 2.

Correlations between Big Five personality traits and LIWC categories

Category	Extraversion	Agreeableness	Conscientious	Neuroticism	Openness
Total pronouns		0.11			-0.21
1st pers singular				0.12	-0.16
1st pers plural	0.11	0.18			-0.1
2nd person	0.16			-0.15	-0.12
Articles			0.09	-0.11	0.2
Past tense		0.1			-0.16
Present tense					-0.16
Negations			-0.17	0.11	-0.13
Numbers	-0.12	0.11			-0.08
Swear words		-0.21	-0.14	0.11	
Social processes	0.15	0.13			-0.14
Family	0.09	0.19			-0.17
Friends	0.15	0.11		-0.08	
Affective processes	0.09				-0.12
Positive emotion	0.1	0.18			-0.15
Negative emotion		-0.15	-0.18	0.16	
Anxiety				0.17	
Anger		-0.23	-0.19	0.13	
Sadness			-0.11	0.1	
Cognitive processes			-0.11	0.13	-0.09
Causation	-0.09	-0.11	-0.12	0.11	
Discrepancy			-0.13	0.13	-0.12
Tentative	-0.11		-0.1	0.12	
Certainty			-0.1	0.13	
Inhibition	-0.13			0.09	
Inclusive	0.09	0.18			0.11
Exclusive			-0.16	0.1	
Perceptual Processes	0.09		-0.1		-0.11
See		0.09			
Hear	0.12		-0.12		-0.08
Feel		0.1		0.1	

Biological Processes	0.14	0.09			-0.09
Body	0.1	0.09			-0.04
Sexual	0.17	0.08			
Space		0.16		-0.09	-0.11
Time		0.12	0.09		-0.22
Work	-0.08				
Achievement	-0.09		0.14		
Leisure	0.08	0.15			-0.17
Home		0.19			-0.2
Money		-0.11			
Religion	0.11				
Death		-0.13	-0.12		0.15
Assent			-0.09		-0.11

Note: Spearman correlations values from Yarkoni's study of personality. Blank cells indicate values not significant at $p < .05$. Highlighted cells are values significant at $p < .001$. Green cells represent positive correlation values and orange cells represent negative correlation values.

Table 3.

Model Performance on Evaluation Metrics

Features	Metrics	Logistic Regression	SVC (non-linear)	Random Forest
Lexical (LIWC)	Accuracy	0.7716	0.7930	0.7838
	F1 Score	0.7813	0.8095	0.7956
	AUC Score	0.7724	0.7988	0.7856
Topical (LDA)	Accuracy	0.5595	0.5690	0.5747
	F1 Score	0.6084	0.5715	0.5601
	AUC Score	0.5607	0.5690	0.5760

Note: Accuracy here refers to test accuracy.

References

Bai, S., Zhu, T., & Cheng, L. (2012). Big-five personality prediction based on user behaviors at social network sites. *arXiv preprint arXiv:1204.4809.*

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta‐analysis. *Personnel psychology*, 44(1), 1-26.

Bazelli, B. et al. (2013). On the personality traits of stackoverflow users. *Software Maintenance (ICSM)*.

Cobb-Clark, D. A., & Schurer, S. (2012). The stability of big-five personality traits. *Economics Letters*, 115(1), 11-15.

Eertmans, A., Victoir, A., Vansant, G., & Van den Bergh, O. (2005). Food-related personality traits, food choice motives and food intake: Mediator and moderator relationships. Food Quality and Preference, 16(8), 714-726.

Fried, D., Surdeanu, M., Kobourov, S., Hingle, M., & Bell, D. (2014, October). Analyzing the language of food on social media. In *2014 IEEE International Conference on Big Data (Big Data)* (pp. 778-783). IEEE.

Fu, G., & Lee, K. (2007). Social grooming in the kindergarten: The emergence of flattery behavior. *Developmental Science*, 10(2), 255-265.

Golbeck, J., Robles, C., & Turner, K. (2011, May). Predicting personality with social media. In Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI ’11, Vancouver, BC, 253–262.

Hirsh, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of research in personality*, 43(3), 524-527.

Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin*, 30(2), 161-172.

Huang, J., Rogers, S., & Joo, E. (2014). Improving restaurants by extracting subtopics from yelp reviews. iConference 2014 (Social Media Expo).

Huang, J. and Yang, Y. 2010. The relationship between personality traits and online shopping motivations. *Social Behavior and Personality*.

Kern, M. L., Park, G., Eichstaedt, J. C., Schwartz, H. A., Sap, M., Smith, L. K., & Ungar, L. H. (2016). Gaining insights from social media language: Methodologies and challenges. *Psychological methods*, 21(4), 507.

Kim, B. (2020). Effects of social grooming on incivility in COVID-19. *Cyberpsychology, Behavior, and Social Networking.*

Kim, J. W., & Chock, T. M. (2015). Body image 2.0: Associations between social grooming on Facebook and body image concerns. *Computers in Human Behavior*, 48, 331-339.

Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., & Graepel, T. (2014). Manifestations of user personality in website choice and behavior on online social networks. *Machine learning*, 95(3), 357-380.

Liu, Y., Wang, J., & Jiang, Y. (2016). PT-LDA: A latent variable model to predict personality traits of social network users. *Neurocomputing*, 210, 155-163.

McGregor, S. (2010). The analysis of personality through language: Narcissism predicts use of shame-related words in narratives (Doctoral dissertation).

Orji, R. et al. (2017). Towards personality-driven persuasive health games and gamified systems. *Proceedings of SIGCHI Conference.*

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... &

Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of personality and social psychology*, 108(6), 934.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. Journal of Personality and Social Psychology, 77, 1296–1312. http://dx.doi.org/10.1037/0022-3514.77.6.1296

Resnik, P., Armstrong, W., Claudino, L., Nguyen, T., Nguyen, V. A., & Boyd-Graber, J. (2015). Beyond LDA: exploring supervised topic modeling for depression-related language in Twitter. In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: *From Linguistic Signal to Clinical Reality* (pp. 99-107).

Ryan, T., & Xenos, S. (2011). Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage. *Computers in human behavior*, 27(5), 1658-1664.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., ... & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9), e73791.

Schwartz, H. A., Sap, M., Kern, M. L., Eichstaedt, J. C., Kapelner, A., Agrawal, M., ... & Kosinski, M. (2016). Predicting individual well-being through the language of social media. In *Biocomputing 2016: Proceedings of the Pacific Symposium* (pp. 516-527).

Seidman, G. (2013). Self-presentation and belonging on Facebook: How personality influences social media use and motivations. *Personality and individual differences*, 54(3), 402-407.

Simpson, S. Reconstructing Thread Structure For Reddit Comments.

Sumner C, Byers A, Shearing M (2011) Determining personality traits & privacy concerns from Facebook activity. *In: Black Hat Briefings*. pp. 1–29.

Suphan, A., & Mierzejewska, B. I. (2016). Boundaries between online and offline realms: how social grooming affects students in the USA and Germany. *Information*, *Communication & Society*, 19(9), 1287-1305.

Terentiev, A., & Tempest, A. (2014). Predicting Reddit Post Popularity Via Initial Commentary. *nd): n. pag.*

Tufekci, Z. (2008). Grooming, gossip, Facebook and Myspace. Information, Communication & Society, 11, 544–564. doi:10.1080/13691180801999050

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality*, 44(3), 363-373.