2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing

Predicting Personality from Twitter

Jennifer Golbeck*, Cristina Robles*, Michon Edmondson*, and Karen Turner*
*University of Maryland;{jgolbeck,crobles,michonk8,kturner}@umd.edu

***Abstract*—Social media is a place where users present themselves to the world, revealing personal details and insights into their lives. We are beginning to understand how some of this information can be utilized to improve the users' experiences with interfaces and with one another. In this paper, we are interested in the personality of users. Personality has been shown to be relevant to many types of interactions; it has been shown to be useful in predicting job satisfaction, professional and romantic relationship success, and even preference for different interfaces. Until now, to accurately gauge users' personalities, they needed to take a personality test. This made it impractical to use personality analysis in many social media domains. In this paper, we present a method by which a user's personality can be accurately predicted through the publicly available information on their Twitter profile. We will describe the type of data collected, our methods of analysis, and the machine learning techniques that allow us to successfully predict personality. We then discuss the implications this has for social media design, interface design, and broader domains.**

***Index Terms*—personality, social media**

I. Introduction

Social networking on the web has grown dramatically over the last decade. In January 2005, a survey of social networking websites estimated that among all sites on the web there were roughly 115 million members [14]. Just over five years later, Twitter alone has exceeded 200 million members. In the process of creating social networking profiles, users reveal a lot about themselves both in what they share and how they say it. Through self-description, status updates, photos, and interests, much of a user's personality comes out through their profile.

For decades, psychology researchers have worked to understand personality in a systematic way. After extensive work to develop and validate a widely accepted personality model, researchers have shown connections between general personality traits and many types of behavior. Relationships have been discovered between personality and psychological disorders [42], job performance [4] and satisfaction [24], and even romantic success [46].

This paper attempts to bridge the gap between social media and personality research by using the information people reveal in their online profiles. Our core research question asks whether social media profiles can predict personality traits. If so, then there is an opportunity to integrate the many results on the implications of personality factors and behavior into the users' online experiences and to use social media profiles as a source of information to better understand individuals. For example, the friend suggestion system could be tailored to a user based on whether they are more introverted or extraverted.

Previous work has shown that the information in users' Facebook profiles is reflective of their actual personalities, not an "idealized" version of themselves [3]. We expect Twitter to have similar characteristics, and that plus a broad user base of 200 million people makes it an ideal platform for study.

We administered the Big Five Personality Inventory to 279 subjects through a Twitter application. In the process, we gathered their 2000 most recent public Twitter posts (tweets). This was aggregated, quantified, and passed through a text analysis tool to obtain a feature set. Using these statistics, we were able to develop a model that can predict personality on each of the five personality factors to within between 11% and 18% of the actual values.

The ability to predict personality has implications in many areas. Existing research has shown connections between personality traits and success in both professional and personal relationships. Social media tools that seek to support these relationships could benefit from personality insights. Additionally, previous work on personality and interfaces showed that users are more receptive to and have greater trust in interfaces and information that is presented from the perspective of their own personality features (i.e. introverts prefer messages presented from an introvert's perspective). If a user's personality can be predicted from their social media profile, online marketing and applications can use this to personalize their message and its presentation.

We begin by presenting background on the Big Five Personality index and related work on personality and social media. We then present our experimental setup and methods for analyzing and quantifying Twitter profile information. To understand the relationship between personality and social media profiles, we present results on correlations between each profile feature and personality factor. Based on this, we describe the machine learning techniques used for classification and show how we achieve large and significant improvements over baseline classification on each personality factor. We conclude with a discussion of the implications that this work has for social media websites and for organizations that may utilize social media to better understand the people with whom they interact.

II. Background and Related Work

A. The Big Five Personality Inventory

The "Big Five" model of personality dimensions has emerged as one of the most well-researched and well-regarded measures of personality structure in recent years. The models five domains of personality, Openness, Conscientiousness, extroversion, Ageeableness, and Neuroticism, were conceived

978-0-7695-4578-3/11 $26.00 © 2011 IEEE
DOI

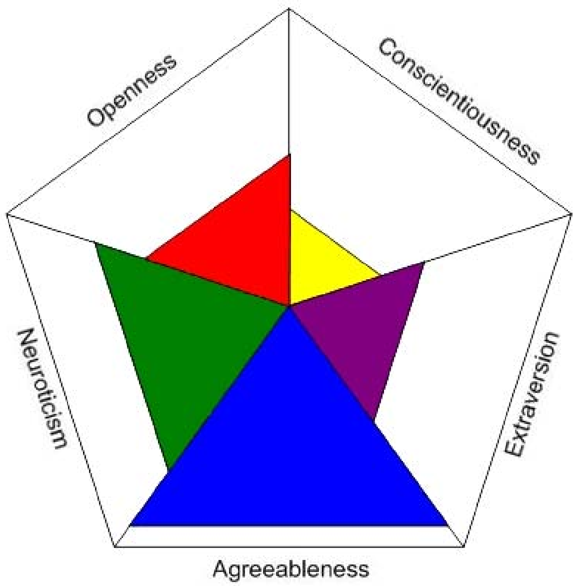

Fig. 1: A person has scores for each of the five personality factors. Together, the five factors represent an individual's personality.

by Tupes and Christal [47] as the fundamental traits that emerged from analyses of previous personality tests [29]. McCrae & Costa [28] and John [21] continued five-factor model research and consistently found generality across age, gender, and cultural lines [29]. Additional research has proved that different tests, languages, and methods of analysis do not alter the models validity [29], [10], [21], [27]. Such extensive research has led to many psychologists to accept the Big Five as the current definitive model of personality [43], [34]. It should be noted that the models dependence on trait terms indicates that the Big Five traits are based on a lexical approach to personality measurement [43], [9], [10], [16]. The Big Five traits are characterized by the following:

- Openness to Experience: curious, intelligent, imaginative. High scorers tend to be artistic and sophisticated in taste and appreciate diverse views, ideas, and experiences.
- Conscientiousness: responsible, organized, persevering. Conscientious individuals are extremely reliable and tend to be high achievers, hard workers, and planners.
- extroversion: outgoing, amicable, assertive. Friendly and energetic, extroverts draw inspiration from social situations.
- Agreeableness: cooperative, helpful, nurturing. People who score high in agreeableness are peace-keepers who are generally optimistic and trusting of others.
- Neuroticism: anxious, insecure, sensitive. Neurotics are moody, tense, and easily tipped into experiencing negative emotions.

B. Applications of the Big Five

Much work has been done with personality as it relates to our lives and the choices we make. In terms of relationships with others, many relationships have been identified. Personality type is linked to whom users choose to friend on Facebook. [45] found that extraversion, agreeableness, and openness all correlated with friendship selection. Personality features have also been tied to many aspects of romantic relationships, including partner choice, level of attachment and success [8], [46]. In terms of interpersonal conflict, studies have associated Big Five traits with coping responses, vengefulness, and rumination [32],[5]. Social relationships aside, personality also relates to preferences. Rentfrow and Gosling [39] is one of many studies that found that personality is a factor that relates to the music an individual prefers to listen to. Jost et al. [23] also found that the personality type of an individual was able to predict whether they would be more likely to vote for McCain or Obama in 2008. Research has also found personality differences between self-professed "dog people" and "cat people" [37], [17]. Within the context of marketing and advertising, Big Five personality traits have been shown to accurately predict a consumers preference for national brands or independent brands [48]. Studies like this show a promising future for the integration of personality analysis and consumer profiling.

Many studies have demonstrated the usefulness of personality profiles within the professional context. Hodgkinson and Ford [20] found that personality traits affect job performance and satisfaction, and Barrick and Mount [4] correlated specific traits with occupational choices and proficiency. Big Five dimensions have proved valid predictors for team performance [31], counterproductive behaviors [41], and entrepreneurial status [49], among many other factors. [6] also revealed relationships between personality and behavior among managers, and Barrick and Mount found recurring personality profiles among both high-autonomy and low-autonomy positions in the workforce [5].

In the space of Human-Computer Interaction, one of the pioneering studies on the connection between personality and interface preference was presented in [30]. Users listened to audio readings of five book reviews which were written from the perspective of introverts vs. extroverts. Subjects were able to identify the personality differences between the reviews and showed an attraction to those which were closest to their own personality type. When the personality type matched, subjects were even more likely to buy the book being reviewed.

This work was extended into ideas of Graphical User Interface design in [25]. Different GUIs were developed to represent introverted vs. extroverted personality types. As in [30], subjects could identify the personality differences and preferred the interface that matched their own personality type.

C. Personality Research and Social Media

To the best of our knowledge, our work is among the first to look at the relationship between profile information provided in social networks and personality traits. However, there have

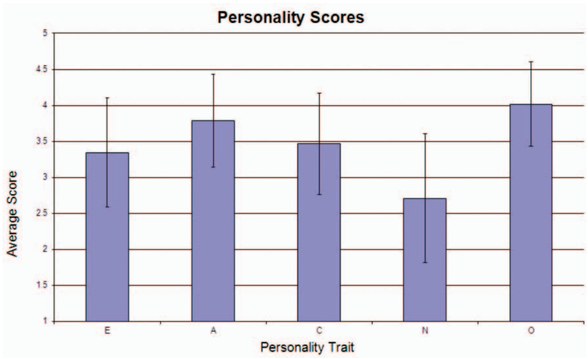

Fig. 2: Average scores on each personality trait shown with standard deviation bars.

been a few previous studies on how personality relates to social networking more generally.

It has been shown in [40] that extroversion and conscientiousness positively correlate with the perceived ease of use of social media websites. extroversion was also shown to have a positive correlation with perceived usefulness of such sites. Not surprisingly, extroversion was also shown to correlate with the size of a user's social network in several studies [2], [44], [45]. There have also been mixed results for other personality traits. Work in [45] showed that individuals with high agreeableness scores were selected more often as friends and that people tended to choose friends with similar agreeableness, extroversion, and openness scores. This was not repeated in [44], but a correlation between openness and number of friends.

III. Data Collection

We created a Twitter application with two functions. First, it administered a 45-question version of the Big Five Personality Inventory [22] to users. Subjects would take the test and for each, we collected the most recent 2,000 tweets from the user (or all tweets if they had less than 2,000).

We had fifty subjects who were recruited through posts on Twitter, Facebook, and relevant mailing lists. Twitter does not collect or release demographic information about its users and, since we would have no general baseline for comparison, we did not collect it for our subjects.

Average scores on the personality test are shown in figure 2 and in table I.

For each user, we began by collecting a simple set of statistics about their accounts and their tweets. These included the following:

- Number of followers (people following the user)
- Number of following (people the user follows)
- Density of the social network
- Number of "@mentions" - An @mention is when a user mentions the name of another user by adding an @ to the front of the username, as is convention on Twitter
- Number of replies - Using the Twitter API, we could see how many of the user's tweets were direct replies to other user's tweets.
- Number of hashtags - Hashtags (e.g. #cscw2012) are a way of tagging a tweet to be part of a given topic or event. They are also used in "games" where users come up with tweets to go with a tag (e.g. #firstdraftmovielines is used with altered first movie lines created by users).
- Number of links
- Words per tweet

For the number of @mentions, replies, hashtags, and links, we used the raw numbers and the average per tweet.

Our primary analysis was a basic processing of the text of the tweets. This was done by merging the collected tweets for a given user into a single "document" and analyzing that.

Previous research has shown that linguistic features can be used to predict personality traits [26], [36]. . Data collected in [36] was used in both studies. They had three separate sources of text, ranging from an average of 1,770 words to over 5,000 words per person.

There is potential to apply these linguistic analysis methods to help predict personality by analyzing a person's tweets. However, the text samples used in earlier studies are much larger than are available to us through any twitter posting. Aggregating many tweets from a user gives more information, but as a series of disconnected statements rather than a coherent document as was used in other studies. Thus, it is unclear if Twitter text will be as connected to personality as was the case in other work. Tweets are much different sources of text. Each one is limited to 140 characters, and a compilation of tweets from a given user is more a stream of disjointed thoughts than a coherent narrative as is found in the text used in previous personality studies. Thus, it was not entirely clear whether tweets would be a useful source of data for this type of analysis.

There were an average of 1914 words per user, and the distribution is shown in figure 3. The number of words ranged from 50 to 5724. These came from an average of 142.2 tweets, with one using having a maximum of 350 tweets and another with a minimum of 4.

Following the methods used in [26], [36] as well as other studies of social media behavior, such as [13], we utilized two main tools to analyze the content of users' tweets. The first is that Linguistic Inquiry and Word Count (LIWC) tool [35]. LIWC produces statistics on 81 different features of text in five categories. These include Standard Counts (word count, words longer than six letters, number of prepositions, etc.), Psychological Processes (emotional, cognitive, sensory, and social processes), Relativity (words about time, the past, the future), Personal Concerns (such as occupation, financial issues, health), and Other dimensions (counts of various types of punctuation, swear words). We excluded the Standard Counts and Other Dimension features to eliminate what is likely to be noise on the type of text we have. The exceptions are that we included word count, words per sentence, and swear word counts since these reflect verbosity and tone of

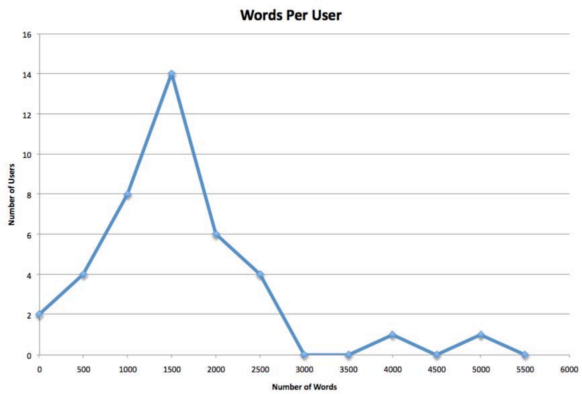

Fig. 3: Number of words per user.

the user. For the other three categories, the values are given as the percentage of words in the input that match words in a given category. For example, it counts the number of "social" words such as "talk", "us", and "friend", or "anxiety" words like "nervous", "afraid", and "tense". Correlations between these features and personality traits (e.g. anxiety words and neuroticism scores) would not be surprising. This produced 79 text features.

In addition, we ran the text again the MRC Psycholinguistic Database, a list of over 150,000 words with linguistic and psycholinguistic features of each word. These include: Kucera-Francis written frequency, number of categories, and number of samples; Brown verbal frequency; Familiarity rating; Meaningfulness via Colorado norms and via Paivio Norms; Concreteness; age of acquisition; Thorndike-Lorge written frequency; and the number of letters, phonemes, and syllables. We computed the average non-zero score for each feature over all the words from each user.

In addition, we performed a word by word sentiment analysis of each user's tweets. Using the General Inquirer dataset [1], which provides a hand annotated dictionary that assigns words sentiment values on a -1 to +1 scale, we computed a score for each user that was the average sentiment score for all words used in their list of tweets.

IV. Personality and Twitter Behavior Correlations

We began by running a Pearson correlation analysis between subjects' personality scores and each of the features obtained from analyzing their tweets and public account data. These are shown in table II.There are a number of significant correlations here, however none of them are strong enough to directly predict any personality trait. Correlations that were statistically significant for $p < 0.05$ are bolded.

Many of the correlations make intuitive sense. For example, conscientiousness is negatively correlated with words about death (e.g. "bury", "coffin", "kill") and with negative emotions and sadness, suggesting conscientious people tend to talk less about unhappy subjects. At the same time, the trait is positively

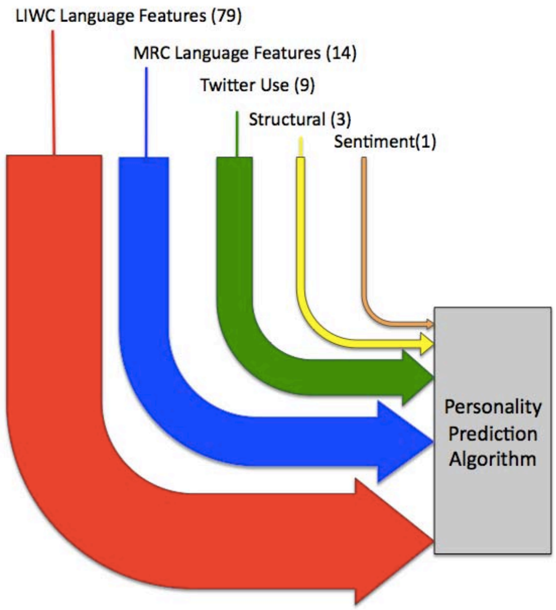

Fig. 4: Features used for predicting personality.

TABLE I: Average scores on each personality factor on a normalized 0-1 scale

	Agree.	Consc.	Extra.	Neuro.	Open.
Average	0.697	0.617	0.586	0.428	0.755
Stdev	0.162	0.176	0.190	0.224	0.147

correlated with the use of "you", indicating the same people tend to talk about or to others. Agreeable people also tend to use "you" a lot, but are less likely to talk about achievements and money.

However, there are not such intuitive explanations for other correlations. For example, the number of parentheses used is negatively correlated with both extraversion and openness. It is unclear why this is the case, or if these are perhaps falsely significant data points. However, since our focus in this paper is on predicting personality rather than on focusing on any particular correlation, we do not assign much weight to any of these connections. A space of future work would be to probe more deeply into these correlations over a larger data set.

V. Predicting Personality

To predict the score of a given personality feature, we performed a regression analysis in Weka [18]. We used two regression algorithms: Gaussian Process and ZeroR, each with a 10-fold cross-validation with 10 iterations. Two algorithms had similar performance over the personality features. Results are shown in table III.

We found that Openness was the easiest to compute and neuroticism was the most difficult, consistent with the results

TABLE II: Pearson correlation values between feature scores and personality scores. Significant correlations are shown in bold for $p < 0.05$. Only features that correlate significantly with at least one personality trait are shown.

Language Feature	Examples	**Extro.**	**Agree.**	**Consc.**	**Neuro.**	**Open.**
"You"	(you, your, thou)	0.068	**0.364**	**0.252**	-0.212	-0.020
Articles	(a, an, the)	-0.039	-0.139	-0.071	-0.154	**0.396**
Auxiliary Verbs	(am, will, have)	0.033	0.042	**-0.284**	0.017	0.045
Future Tense	(will, gonna)	0.227	-0.100	**-0.286**	0.118	0.142
Negations	(no, not, never)	-0.020	0.048	**-0.374**	0.081	0.040
Quantifiers	(few, many, much)	-0.002	-0.057	-0.089	-0.051	**0.238**
Social Processes	(mate, talk, they, child)	**0.262**	0.156	0.168	-0.141	0.084
Family	(daughter, husband, aunt)	**0.338**	0.020	-0.126	0.096	0.215
Humans	(adult, baby, boy)	0.204	-0.011	0.055	-0.113	**0.251**
Negative Emotions	(hurt, ugly, nasty)	0.054	-0.111	**-0.268**	0.120	0.010
Sadness	(crying, grief, sad)	0.154	-0.203	**-0.253**	0.230	-0.111
Cognitive Mechanisms	(cause, know, ought)	-0.008	-0.089	**-0.244**	0.025	0.140
Causation	(because, effect, hence)	0.224	**-0.258**	-0.155	-0.004	**0.264**
Discrepancy	(should, would, could)	0.227	-0.055	**-0.292**	0.187	0.103
Certainty	(always, never)	0.112	-0.117	-0.069	-0.074	**0.347**
Perceptual Processes						
Hearing	(listen, hearing)	0.042	-0.041	0.014	**0.335**	-0.084
Feeling	(feels, touch)	0.097	-0.127	**-0.236**	**0.244**	0.005
Biological Processes	(eat, blood, pain)	-0.066	0.206	0.005	0.057	**-0.239**
Body	(cheek, hands, spit)	0.031	0.083	-0.079	0.122	**-0.299**
Health	(clinic, flu, pill)	**-0.277**	0.164	0.059	-0.012	-0.004
Ingestion	(dish, eat, pizza)	-0.105	**0.247**	0.013	-0.058	-0.202
Work	(job, majors, xerox)	0.231	-0.096	**0.330**	-0.125	**0.426**
Achievement	(earn, hero, win)	-0.005	**-0.240**	-0.198	-0.070	0.008
Money	(audit, cash, owe)	-0.063	**-0.259**	0.099	-0.074	0.222
Religion	(altar, church, mosque)	-0.152	-0.151	-0.025	**0.383**	-0.073
Death	(bury, coffin, kill)	-0.001	0.064	**-0.332**	-0.054	0.120
Fillers	(blah, imean, youknow)	0.099	-0.186	**-0.272**	0.080	0.120
Punctuation						
Commas		0.148	0.080	**-0.24**	0.155	0.170
Colons		-0.216	-0.153	**0.322**	-0.015	-0.142
Question Marks		**0.263**	-0.050	0.024	0.153	-0.114
Exclamation Marks		-0.021	-0.025	**0.260**	**0.317**	**-0.295**
Parentheses		**-0.254**	-0.048	-0.084	0.133	**-0.302**
Non-LIWC Features						
GI Sentiment		0.177	-0.130	-0.084	-0.197	**0.268**
Number of Hashtags		0.066	-0.044	-0.030	-0.217	**-0.268**
Words per tweet		**0.285**	-0.065	-0.144	0.031	0.200
Links per tweet		-0.061	-0.081	**0.256**	-0.054	0.064

found in [26], which used similar methods to ours for analyzing larger text corpa.

We found mixed results in this analysis. In our previous work studying personality on Facebook [15], we had fewer and weaker correlations, but were able to predict all personality traits to within roughly 11%. This analysis of Twitter data yielded similar results for openness and agreeableness, but less impressive results for conscientious, extraversion, and neuroticism.

We believe a larger sample size would produce much better results. With only fifty subjects, building effectively training algorithms is difficult. However, the results we obtained even with this small sample show promise that these analysis techniques can be useful for computing personality on Twitter and other micro-blogging sites.

VI. Discussion

Our results show that we can predict personality to within just over 10%, a resolution that is likely fine-grained enough for many applications. The difference between being 65% vs 75% extraverted, for example, is likely small enough that the error would not have many practical implications. Furthermore, in many cases, simply identifying a person as being on one side of the scale vs. the other (e.g. introverted vs. extraverted) is likely enough to offer features that are beneficial.

Since this research relies heavily on text analysis, the nature of text on Twitter raises an interesting challenge. We analyzed it with standard tools, but the standards of editing down and misspelling words on Twitter make it likely that there are language features that could be missed. Hhow to apply these tools on Twitter is an interesting question for future research.

The important question that comes from this research is how the results can be used. Drawing on research results like those discussed as related work, there is potential to integrate previous personality results into social media as a way to enhance the accuracy of certain features or the user's experience.

The research on interfaces and personality presented above showed that users preferred interfaces designed to represent personalities that most closely matched their own [30], [25]. This has significant implications for this work. With the ability to infer a user's personality, social media websites, e-commerce retailers, and even ad servers can be tailored to reflect the user's personality traits and present information such that users will be most receptive to it. For example, the presentation of ads could be adjusted based on the personality of the user. Similarly, product reviews from authors with personality traits similar to the user could be highlighted to increase trust and perceived usefulness by the user. Customized website "skins" could be created for different user personality types, as suggested in [7]. Our methods provide a straightforward way to obtain personality profiles of users without the burden of tests, and this will make it much easier to create personality-oriented interfaces.

This same idea can be extended even further to advertising. While results of integrating personality to marketing have been mixed, some work has demonstrated connections between marketing techniques and consumer personality [33]. For e-commerce marketers, both those who advertise on social media sites and elsewhere, utilizing social media profiles as a way to determine consumer personality can make it easy to implement existing techniques that benefit from this knowledge of consumer background.

Consider Twitters friend recommendation feature, where people are suggested to the user as potential friends. Previous results on personality and relationships may indicate that people with certain personality types are more likely to add one another as friends. By integrating these findings into Twitters friend recommendation feature we would be able to more accurately predict who else on Twitter a user might be more likely to add.

Recommender systems may also benefit from integrating predicted personality values. Results showing correlations between personality and music taste are well established in the literature [39], [11], [19], [38]. Inferring personality traits from Twitter profiles may allow recommender systems to improve their accuracy by recommending music, and possibly other items, that are tailored to the user's personality profile. Collaborative filtering algorithms find people with similar tastes to the user, and then recommend items that those people like. This similarity is typically computed over shared ratings, but with personality information available, this could be used to give more weight to users who share similar personality traits. Such techniques have been successful when used with trust relationships [12]. An alternate use of personality in recommender systems would be to identify types of items that are liked by individuals with certain personality traits and to give those items greater consideration based on the user's personality profile. Developing these algorithms and evaluating them is a space open for future work.

VII. Conclusions

In this paper, we have shown that a users' Big Five personality traits can be predicted from the public information they share on Twitter. Our subjects completed a personality test and through the Twitter API, we collected publicly accessible information from their profiles. After processing this data, we found many small correlations in the data. Using the profile data as a feature set, we were able to train two machine learning algorithms - ZeroR and Gaussian Processes - to predict scores on each of the five personality traits to within 11% - 18% of their actual value.

With the ability to guess a user's personality traits, many opportunities are opened for personalizing interfaces and information. We discussed some of these opportunities for marketing and interface design above. However, there is much work to be pursued in this area.

One area that deserves attention is the connection between personality and the actual social network. We considered two structural features - number of friends and network density -

TABLE III: Mean Absolute Error on a normalized scale for each algorithm and personality trait.

	Agree.	Consc.	Extra.	Neuro.	Open.
ZeroR	0.129980265	0.146204953	0.160241663	0.182122225	0.11923333
GaussianProcess	0.130675423	0.14599073	0.160315335	0.18205923	0.11922558

but we did not look at personality scores between friends. Understanding the connections between personality, tie strength [13], trust [14], and other related factors is an open space for research. By improving our knowledge of these relationships, we can begin to answer more sophisticated questions about how to present trusted, socially-relevant, and well-presented information to users.

VIII. Acknowledgments

Research was sponsored, in part, by the National Science Foundation (DBI-0850373) and by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] General inquirer, 1997. http://www.wjh.harvard.edu/ inquirer/.

[2] A. Acar and M. Polonsky. Online Social Networks and Insights into Marketing Communications. *Journal of Internet Commerce*, 6(4):55–72, 2008.

[3] M. Back, J. Stopfer, S. Vazire, S. Gaddis, S. Schmukle, B. Egloff, and S. Gosling. Facebook Profiles Reflect Actual Personality, Not Self-Idealization. *Psychological Science*, 21(3):372, 2010.

[4] M. Barrick and M. Mount. The Big Five personality dimensions and job performance: A meta-analysis. *Personnel psychology*, 44(1):1–26, 1991.

[5] M. Barrick and M. Mount. Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. *Journal of Applied Psychology*, 78(1):111–118, 1993.

[6] S. Berr, A. Church, and J. Waclawski. The right relationship is everything: Linking personality preferences to managerial behaviors. *Human Resource Development Quarterly*, 11(2):133–157, 2000.

[7] W.-P. Brinkman and N. Fine. Towards customized emotional design: an explorative study of user personality and user interface skin preferences. In *EACE '05: Proceedings of the 2005 annual conference on European association of cognitive ergonomics*, pages 107–114. University of Athens, 2005.

[8] T. Chamorro-Premuzic. *Personality and Romantic Relationships*, volume Personality and Individual Differences. Blackwell Publishing, 2007.

[9] B. De Raad. *The Big Five personality factors: The psycholexical approach to personality*. Hogrefe & Huber G
"ottingen, 2000.

[10] J. Digman. Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1):417–440, 1990.

[11] S. Dollinger. Research Note: Personality and Music Preference: Extraversion and Excitement Seeking or Openness to Experience? *Psychology of Music*, 21(1):73, 1993.

[12] T. DuBois, J. Golbeck, J. Kleint, and A. Srinivasan. Improving Recommendation Accuracy by Clustering Social Networks with Trust. In *Recommender Systems & the Social Web*, 2009.

[13] E. Gilbert and K. Karahalios. Predicting tie strength with social media. In *Proceedings of the 27th international conference on Human factors in computing systems*, pages 211–220. ACM New York, NY, USA, 2009.

[14] J. Golbeck. *Computing and Applying Trust in Web-based Social Networks*. PhD thesis, University of Maryland, College Park, MD, USA, April 2005.

[15] J. Golbeck, C. Robles, and K. Turner. Predicting personality with social media. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, CHI EA '11, pages 253–262, New York, NY, USA, 2011. ACM.

[16] L. Goldberg. From Ace to Zombie: Some explorations in the language of personality. *Advances in personality assessment*, 1:203–234, 1982.

[17] S. D. Gosling, C. J. Sandy, and J. Potter. Personalities of self-identified dog people and cat people. *Anthrozoos: A Multidisciplinary Journal of The Interactions of People & Animals*, 23:213–222(10), September 2010.

[18] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten. The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18, 2009.

[19] C. Hansen and R. Hansen. Constructing personality and social reality through music: Individual differences among fans of punk and heavy metal music. *Journal of broadcasting & electronic media*, 35(3):335–350, 1991.

[20] G. Hodgkinson and J. Ford. *International Review of Industrial and Organizational Psychology, 2008*. Wiley-Interscience, 2008.

[21] O. John. The Big Five factor taxonomy: Dimensions of personality in the natural language and in questionnaires. *Handbook of personality: Theory and research*, 14:66–100, 1990.

[22] O. D. John. Big five inventory, 2000.

[23] J. Jost, T. West, and S. Gosling. Personality and ideology as determinants of candidate preferences and Obama conversion in the 2008 US presidential election. *Du Bois Review: Social Science Research on Race*, 6(01):103–124, 2009.

[24] T. Judge, C. Higgins, C. Thoresen, and M. Barrick. The big five personality traits, general mental ability, and career success across the life span. *Personnel psychology*, 52(3):621–652, 1999.

[25] A. Karsvall. Personality preferences in graphical interface design. In *NordiCHI '02: Proceedings of the second Nordic conference on Human-computer interaction*, pages 217–218, New York, NY, USA, 2002. ACM.

[26] F. Mairesse, M. Walker, M. Mehl, and R. Moore. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30(1):457–500, 2007.

[27] R. McCrae. Why I advocate the five-factor model: Joint factor analyses of the NEO-PI with other instruments. *Personality psychology: Recent trends and emerging directions*, pages 237–245, 1989.

[28] R. McCrae and P. Costa. *Personality in adulthood: A five-factor theory perspective*. The Guilford Press, 1990.

[29] R. McCrae and O. John. An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215, 1992.

[30] C. Nass and K. M. Lee. Does computer-generated speech manifest personality? an experimental test of similarity-attraction. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 329–336, New York, NY, USA, 2000. ACM.

[31] G. Neuman, S. Wagner, and N. Christiansen. The relationship between work-team personality composition and the job performance of teams. *Group & Organization Management*, 24(1):28, 1999.

[32] T. O'Brien and A. DeLongis. The interactional context of problem-, emotion-, and relationship-focused coping: The role of the big five personality factors. *Journal of personality*, 64(4):775–813, 1996.

[33] G. Odekerken-Schr
"oder, K. De Wulf, and P. Schumacher. Strengthening outcomes of retailer-consumer relationships:: The dual impact of relationship marketing tactics and consumer personality. *Journal of Business Research*, 56(3):177–190, 2003.

[34] D. Peabody and B. De Raad. The Substantive Nature of Psycholexical Personality Factors: A Comparison Across Languages. *Journal of Personality and Social Psychology*, 83(4):983–997, 2002.

[35] J. Pennebaker, M. Francis, and R. Booth. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 2001.
[36] J. Pennebaker and L. King. Linguistic styles: Language use as an individual difference. *Journal of personality and social psychology*, 77(6):1296–1312, 1999.
[37] R. Perrine and H. Osbourne. Personality characteristics of dog and cat persons. *Anthrozoos: A Multidisciplinary Journal of The Interactions of People & Animals*, 11(1):33–40, 1998.
[38] D. Rawlings and V. Ciancarelli. Music preference and the five-factor model of the NEO Personality Inventory. *Psychology of Music*, 25(2):120, 1997.
[39] P. Rentfrow and S. Gosling. The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, 84(6):1236–1256, 2003.
[40] P. Rosen and D. Kluemper. The Impact of the Big Five Personality Traits on the Acceptance of Social Networking Website. *AMCIS 2008 Proceedings*, page 274, 2008.
[41] J. Salgado. The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment*, 10(1&2):117–125, 2002.
[42] L. Saulsman and A. Page. The five-factor model and personality disorder empirical literature: A meta-analytic review* 1. *Clinical Psychology Review*, 23(8):1055–1085, 2004.
[43] D. Schmitt, J. Allik, R. McCrae, and V. Benet-Martinez. The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology*, 38(2):173, 2007.
[44] J. Schrammel, C. Köffel, and M. Tscheligi. Personality traits, usage patterns and information disclosure in online communities. In *BCS HCI '09: Proceedings of the 2009 British Computer Society Conference on Human-Computer Interaction*, pages 169–174, Swinton, UK, UK, 2009. British Computer Society.
[45] M. Selfhout, W. Burk, S. Branje, J. Denissen, M. van Aken, and W. Meeus. Emerging Late Adolescent Friendship Networks and Big Five Personality Traits: A Social Network Approach. *Journal of personality*, 78(2):509–538, 2010.
[46] P. Shaver and K. Brennan. Attachment styles and the "Big Five" personality traits: Their connections with each other and with romantic relationship outcomes. *Personality and Social Psychology Bulletin*, 18(5):536, 1992.
[47] E. Tupes and R. Christal. Recurrent personality factors based on trait ratings. *Journal of Personality*, 60(2):225–251, 1992.
[48] S. Whelan and G. Davies. Profiling consumers of own brands and national brands using human personality. *Journal of Retailing and Consumer Services*, 13(6):393–402, 2006.
[49] H. Zhao and S. Seibert. The big five personality dimensions and entrepreneurial status: A meta-analytical review. *Journal of Applied Psychology*, 91(2):259–271, 2006.