Making Use of Derived Personality: The Case of Social Media Ad Targeting

Jilin Chen[1]*, Eben Haber[1], Ruogu Kang[2], Gary Hsieh[3], Jalal Mahmud[1]**

[1]IBM Research –Almaden, chenjilin@gmail.com, eben@acm.org, jumahmud@us.ibm.com
[2]Human-Computer Interaction Institute, Carnegie Mellon University, ruoguk@cs.cmu.edu
[3]Human Centered Design & Engineering, University of Washington, garyhs@uw.edu

Abstract

People's social media text has been shown to have limited prediction power of their personality. In this work, we conducted a survey study and a field study to explore the feasibility of using predicted personality traits derived from social media text for the purpose of ad targeting. In the survey study, we measured people's personalities and their responses to an advertisement tweet. We found that people with high openness and low neuroticism responded more favorably to a targeted advertisement, thus demonstrating the effects of the personality traits themselves. In the field study, we sent the advertisement tweets to real-world Twitter users, and found the same effects on users' responses using personality traits derived from users' tweet text. These results suggest that the derived personality traits had the same effects as the personality traits measured by traditional personality questionnaires, and can indeed improve ad targeting in real-world settings.

Introduction

The rise of social media platforms such as Facebook and Twitter has also brought the rise of social media ads. Marketers send social media users targeted ads about product and services, hoping to get clicks, followers, and eventually purchases. Ad targeting in this context is often done through second-guessing users' interest, based on keyword matching and social network structures (Miller et al. 2010).

Recent research has indicated that personality can also influence whether people would accept a suggested product or service (Golbeck et al. 2013, Hirsh et al. 2012, Hu et al. 2011), and thus be a promising venue to further improve ad targeting. However, as the traditional way of measuring personality requires people to complete personality questionnaires (e.g. Goldberg et al. 2006), it is challenging to obtain such personality measurements at a large scale for commercial use.

Motivated by this challenge, prior research has attempted predicting personality through lexicon features extracted from people's social media posts (Golbeck et al. 2011a, Golbeck et al. 2011b, Gou et al. 2014, Sumner et al. 2012). The results have been decidedly mixed: while these lexicon features indeed contain predictive information of personality (Golbeck et al. 2011a, Golbeck et al. 2011b), the predicted personality may only mildly correlate with measurements from questionnaires, and can sometimes be only slightly better than random chance (Gou et al. 2014, Sumner et al. 2012).

In this work we investigate the feasibility of using these derived personality traits to improve ad targeting on social media. To this end, we created a Twitter-based travel information service, and hypothesized that Twitter users of high openness and low neuroticism would respond more favorably to unsolicited advertisements of our service. Openness and neuroticism are two of the Big 5 personality traits (Goldberg 1993), and we focused on these two traits as they are particularly relevant to our use case of ads.

We conducted two studies based on our use case and the hypothesized effects of personality. In a survey study, we asked people to report their likely responses to a tweet advertising travel information service, and measured their personality using traditional questionnaires. In the field study, we sent the advertisement tweets to travelling Twitter users, and derived the users' personality traits from their past tweets via a lexicon-based approach (Pennebaker et al. 2007, Yarkoni et al. 2010). The two studies' results were consistent: people with high openness and low neuroticism indeed responded more favorably, despite the different approaches we obtained the personality traits in the two studies.

This work makes an important contribution to practice. For the first time, we have demonstrated that derived personality traits, despite their potential inaccuracies, can still substantially improve ad targeting in a real-world setting.

* Current affiliation: Google Inc. ** Current affiliation: Couchbase, Inc.

Figure 1. Account Profile of TravelersLikeMe.

In other words, social media companies like Twitter may be able to derive personalities from their users' public posts and improve the effectiveness of their advertisements. This work also furthers our understanding of personality itself, by shedding light on the mechanism behind the observed effects, and by providing a meaningful contrast between personality measured from questionnaires and personality derived from social media text.

We organize the rest of this article as follows. We first introduce the research background and our three research hypotheses. We then describe our use case, i.e. the travel information service, and the two user studies with their respective results. We end the paper with discussions of our findings and ethical considerations.

Research Background and Hypothesis

The Big 5 personality traits have long been shown to affect various human behaviors (Goldberg 1993). Recent research has indicated that Big 5 personality can also affect people's acceptance of advertisements and product suggestions. For instance, in marketing contexts, Hirsh et al. (2012) surveyed people about various marketing messages and found people respond more positively to messages tailored to their personality. In the music domain, Hu et al. (2011) have shown that music recommendations are more successful when they leverage the correlations between people's personality and their music preferences. Similarly, Golbeck et al. (2013) discovered correlations between personality and movie preferences among Netflix users.

The Effects of Openness and Neuroticism

For scope, in this work we focused on the effects of openness and neuroticism, which are two of the Big 5 personality traits (Goldberg et al. 1993).

The trait of openness reflects appreciation for art, emotion, adventure, unusual ideas, curiosity, and variety of experience. In the context of online marketing, people with

Figure 2. Advertisement Reply from TravelersLikeMe.

high openness has been found to be more intellectually curious, open to new ideas, and therefore more likely to try out innovative forms of shopping such as e-commerce (Chen et al. 2011).

The trait of neuroticism reflects the tendency to experience various unpleasant emotions, including anger, anxiety, depression, and vulnerability. In particular, it has been found that people with high neuroticism are more likely to feel vulnerable and insecure, and thus less likely to be trusting of others (Evans et al. 2008).

In social media marketing, advertisements are delivered as social media posts from the merchant or service provider to individual end users. In our use case, our Twitter account represents our newly created unknown brand, and sends out unsolicited advertisement tweets to people. As a result, we hypothesize that people with high openness will be more willing to give our service a try even though they do not know us at all, while people with high neuroticism will be less trusting of us and thus be less likely to respond positively. This reasoning gives us two hypotheses:

H1: The higher openness (measured) a Twitter user has, the more successful the targeted ad would be;

H2: The higher neuroticism (measured) a Twitter user has, the less successful the targeted ad would be.

We test these two hypotheses in our survey study, where we present the ads of our service in a hypothetical use case, and ask people to report their likely responses. The *success* of an advertisement is measured by indications of positive responses, such as clicking the link embedded within the ads and following our Twitter account.

Deriving Personality from Social Media Writing

The measurement of personality has traditionally relied on the use of personality questionnaires (e.g. Goldberg et al. 2006). However, on social media, many people are simply not willing to spend the extra effort in completing such questionnaires, making measurement difficult. Deriving personality from people's writing therefore becomes an attractive option, as it requires no extra effort from the end users.

A rich body of research exists on relating people's social media writing to their personality. Yarkoni analyzed blogs and showed that people's word use reliably correlated with their personality (Yarkoni et al. 2010). Golbeck et al. (2011a), (2011b), Gou et al. (2014) and Sumner et al. (2012) all studied using people's text snippets on Facebook and/or Twitter to predict their personality. All these prior

	Mean	Std. Dev.	Correlations					
			1	2	3	4	5	6
1. Likelihood to click the link	2.58	1.38						
2. Likelihood to follow the account	2.33	1.26	**.71**					
3. Likelihood to reply and ask for details	2.05	1.21	**.61**	**.54**				
4. Likelihood to report spam	2.52	1.37	**-.39**	**-.37**	-.15			
5. Openness personality	3.36	0.95	**.27**	**.22**	.12	**-.19**		
6. Neuroticism personality	3.09	1.09	**-.37**	**-.27**	**-.24**	.14	**-.22**	
7. Interest on the topic of travelling	3.20	1.22	.16	**.22**	-.06	-.06	-.02	.06

Table 1. Descriptive Statistics of Variables from the Survey Study

Significant correlations are shown in bold. The three positive responses (i.e. click, follow, reply) are reliably correlated to each other, and are all negatively correlated with the negative response (i.e. report spam).

works primarily used lexicon-based features extracted from text, such as the Linguistic Inquiry and Word Count (LIWC) dictionary (Pennebaker et al. 2007).

Reports on the accuracy of such lexicon-based prediction have been mixed. In relation to personality measurements from questionnaires, the predicted personality values have been reported to moderately correlate (0.4~0.6) (Golbeck et al. 2011b), weakly correlate (0.05~0.2) (Guo et al. 2014), or be close to random chance (Sumner et al. 2012). These mixed results leave open the question whether personality derived in this way can be reliably used for ad targeting on social media.

In this work, we derived openness and neuroticism traits via the same lexicon-based approach as in prior work, and hypothesized that the derived personality traits can indeed make a difference:

H3: Targeting Twitter users of high openness (derived) and low neuroticism (derived) can substantially improve the success of the ads.

We test this hypothesis in our field study, where we send the advertisements of our service to real-world traveling users on Twitter and measure their reactions. Similar to the survey study, the *success* is measured by people's positive actions, such as clicking the embedded link, and following our account.

It should also be noted that a few prior studies have used derived personality traits for targeting Twitter users for answering questions and for spreading information (e.g. Lee et al. 2014). However, because these studies' primary goal was to maximize the overall targeting accuracy, they have included the derived personality traits as generic features along with many other features in complex machine learning models. It is therefore difficult to tell from these studies how much the personality traits contributed, or why it was so.

Use Case: TravelersLikeMe

We created TravelersLikeMe, a Twitter-based travel information service that recommends attractions to travelers. Our two studies centered on the use case of advertising this service to Twitter users.

We chose travel as the topic domain for two reasons. Firstly, many people travel (e.g., according to ustravel.org, there were 1.6 billion person-trips for leisure in the U.S. alone in 2012) and they often need advice on food, accommodations, places to go, etc. Secondly, people often express their intent to travel on social media, so that we can accurately target our advertisement to those people and avoid spamming other Twitter users. In particular, we focus on Twitter users visiting New York City (NYC), because empirically we have found NYC among the most popular destinations mentioned on Twitter.

The identity of TravelersLikeMe centers on its Twitter account, @travelerslikeme (Figure 1). In our use case, this account identifies Twitter users who in their recent tweets have publicly stated their intent to visit NYC in the near future, and sends a reply tweet to each of these users. The reply tweet suggests a few interesting activities in NYC, invites the users to sign up for the service by following the @travelerslikeme account, and contains a web link that further describes the details of service (Figure 2).

Such proactive replies are a common practice for businesses to engage Twitter users (Chen et al. 2013), and prior research showed that Twitter users are often tolerant toward such practice when the replies are relevant to their own tweets (Lee at al. 2014). We adopt this approach to advertise TravelersLikeMe instead of buying display ads directly from Twitter, because we need to fully control the targeting process and track the response (or the lack of such) for each user we target, while Twitter ads only has a limited selection of targeting criteria. For example, it does not reveal who have seen the advertisement without taking any further action.

Because our tweets are unsolicited advertisements, people might consider our practice as spamming. To alleviate this concern, in our field study we tried to ensure maximal accuracy of our targeting so as to avoid disrupting Twitter users, and we indeed provided a meaningful service to users who signed up. We discuss these details in the field study section, and discuss ethical considerations of our approach near the end of this paper.

	Predicting the likelihood to click the link		Predicting the likelihood to follow the account		Predicting the likelihood to reply and ask for details		Predicting the likelihood to report spam	
	Std. Coef.	Sig.	Std. Coef.	Sig.	Std. Coef.	Sig.	Std. Coef.	Sig.
Openness personality	.236	**	.262	*	.092		-.181	
Neuroticism personality	-.358	***	-.299	**	-.250	*	.119	
Interest on the topic of travelling	.184	*	.253	**	-.046		-.071	
R^2	22.1%		16.8%		6.7%		5.6%	

Table 2. Predicting Survey Responses to Hypothetical Advertisement Tweets

Significance levels: *** p<.001, ** p<.01, * p<.05. Consistent with our hypotheses H1 and H2, for predicting the likelihood to click and follow, openness is a significant positive predictor, and neuroticism is a significant negative predictor, even when we control for the positive effect of participants' interest on the travel topic. The predictors appear incapable of predicting likelihood to reply or report spam.

Survey Study

The purpose of the survey study was to understand in a controlled setting if openness and neuroticism measured through personality questionnaires (Goldberg et al. 2006) would have the hypothesized effects on people's responses (i.e. H1 and H2).

Methodology

We recruited participants from Amazon Mechanical Turk (MTurk). The survey took about 6 minutes to complete, for which we paid USD$0.50 to each participant. To obtain informed responses, we required that each participant be living in United States, have a few months' experience using Twitter and have a basic understanding of Twitter concepts such as followers and replies. To ensure those qualifications and overall response quality, the survey included a few questions testing participants' attentiveness to the survey and their knowledge of Twitter. In particular, we ensured that each participant understood how replies work on Twitter, as it is a core of the use case.

The survey presented our use case to the participants as a hypothetical scenario:

Imagine you are going to New York City for fun next week, and have posted a tweet: "Going to NYC next week! Excited!" Soon after you post the tweet, you receive an @reply from TravelersLikeMe, a Twitter account you haven't heard of. And the account seems to be a travel tip recommendation service according to its Twitter profile.

Along with the above text description, we showed participants the profile of the TravelersLikeMe account (Figure 1) and the reply tweet (Figure 2) that they would have seen in the use case.

Based on the use case, the survey asked participants to rate on a 5-point Likert scale the likelihood that they would 1) click the embedded link within the tweet, 2) follow the TravelersLikeMe account on Twitter, 3) reply to the tweet asking for more details, and/or 4) report the tweet as spam, where 1 means extremely unlikely and 5 means extremely likely.

We also asked each participant to provide a short written explanation for their overall ratings, so as to gain some insights into their decision process.

The survey then measured each participant's openness and neuroticism traits through 20 questions adopted from the IPIP scales for Big 5 personalities, where each trait is measured through 10 questions (Goldberg et al. 2006). The resulting measures of both traits appeared reliable, with Cronbach's alphas above 0.80.

Near the end of the survey, we asked participants to rate on a 5-point Likert scale their general interest on in the topic of traveling, so as to control the effect of their topic interest. The survey also included qualification questions as well as questions about demographics and general Twitter usage.

Participants

We recruited 133 participants from MTurk. In the end we included 113 participants in the analysis, after removing incomplete responses and participants who failed the qualification questions. Among the participants 52% were male. A majority of participants were between 22 to 34 years old (57%), although we also had some participants under 21 years old (11%) or above 55 years old (7%). In terms of Twitter usage, a majority of participants self-reported to have used Twitter for at least a year (80%), followed at least 50 people (65%), have posted at least 100 tweets (53%), and were visiting Twitter at least once a week (68%).

Measurements

From the survey we obtained the reported likelihood of four responses from the participants, i.e. their likelihood to click the link, follow the account, reply to the tweet, and report spam. We obtained from each participant a short written explanation for their reported likelihoods as well.

For each participant, we also had measures of the openness and neuroticism traits and a self-reported interest level on the topic of traveling.

		Mean	Std. Dev.	Correlations								
				1	2	3	4	5	6	7	8	9
1	**Click response (Phase I)**	0.063	0.243									
2	**Follow response (Phase I)**	0.104	0.306	**.17**								
3	**Click response (Phase II)**	0.068	0.252	-	-							
4	**Follow response (Phase II)**	0.047	0.213	-	-	**.13**						
5	**User account age**	904.5	516.2	.02	-.03	.02	.00					
6	**User tweet count**	12349.9	19041.0	-.03	**-.07**	**-.07**	**-.06**	**.25**				
7	**User followee count**	571.9	1177.3	.01	**.16**	-.03	.01	**.08**	**.18**			
8	**User follower count**	1294.4	6940.8	-.01	.00	**.04**	-.01	**.12**	**.10**	**.30**		
9	**User topic interest**	0.317	0.253	.03	**.06**	.02	**.06**	**.09**	-.03	**.12**	**.08**	
10	**Openness-neuroticism index**	0.000	1.000	**.10**	**.09**	**.09**	**.10**	**.31**	**-.15**	.04	**.07**	**.27**

Table 3. Descriptive Statistics of Variables from the Field Study

Significant correlations are shown in bold. The openness-neuroticism index was already standardized during its computation and therefore had a mean of zero and a standard deviation of one. Much larger portion of targeted users followed our account in Phase I than in Phase II, presumably because only in Phase I did our account proactively follow them and some of these users followed back for etiquette.

In our analysis here, we have omitted other measures from the survey such as gender, age, and the self-reported Twitter usage statistics, because none of them had any significant correlation or prediction power toward the four types of responses.

Results

We show the descriptive statistics in Table 1. The top four variables capture responses to our advertisement tweets. Overall the responses were mixed. The means of participants' estimated responses to our advertisement (click, follow, reply, or report spam) were all slightly lower than neutral, and with substantial variations covering the range between extremely likely and extremely unlikely. The three positive responses (i.e. click, follow, reply) were reliably correlated to each other, and were all negatively correlated with the negative response (i.e. report spam). By examining the written explanations associated with the responses, we found that the observed correlations may be largely attributed to a single deciding factor: participants who considered our tweet as spam in their explanations consistently rated clicking, following and replying as unlikely or worse.

To understand the effect of the personality traits, we applied linear regressions to predict the participants' responses (variable 1-4 in Table 1), using their openness, neuroticism as predictors, while controlling for their interests on the topic of traveling.

We show the regression results in Table 2. The predictors had meaningful strength on predicting the likelihood to click and follow (22.1% and 16.8% R^2, respectively), and appeared incapable of predicting likelihood to reply or report spam (6.7% and 5.6% R^2, respectively). The likelihood to reply or report spam was more difficult to predict, perhaps because these decisions are inherently more confounded in the study: a few participants have explained that they were not sure if they should reply because our tweet did not ask for replies explicitly; several participants considered the tweet as spam but did not rate reporting spam as likely, attributing their inaction to "laziness" or unfamiliarity with Twitter's spam policy.

For predicting the likelihood to click and follow, openness was a significant positive predictor ($p < .01$ for clicking on the link and $p < .05$ for following the account), neuroticism was a significant negative predictor ($p < .001$ for clicking on the link and $p < .01$ for following the account), even when we controlled for the positive effect of participants' interest on the topic of traveling.

These results support our hypotheses H1 and H2.

The Role of Personality in Participants' Written Explanations

To better understand the role of personality traits in people's decision process, we grouped participants into high and low groups for each of the two traits, and found a number of written explanations that may have suggested the effect of personality.

Participants with high openness have expressed curiosity and willingness to try out the service given the uncertainty:

I might find it weird that it messaged me, but I might still follow and/or click on the link. Plus I like knowing about cool things to do when I'm out and about. (P1)

I would be curious as to what they had to offer in terms of travel advice. (P2)

I would be interested in the information they could provide as I have never been there. (P3)

In contrast, participants with high neuroticism expressed lack of trust, and were afraid of malicious intents:

I pretty much never click on links that I don't know of/trust. I limit how many non-friends I actually follow. (P4)

I don't like to click on random links because it could be spam or a virus. I don't want to take the risk of getting my account hacked. (P5)

I am wary of bit.ly links from everyday people, I'd be even more wary of a bit.ly link from some random company. (P6)

These explanations are consistent with the mechanism behind our hypotheses H1 and H2.

We also identified participants with both high openness and high neuroticism. Some of them were optimistic, like the case of high openness only, while some were cautious, like the case of high neuroticism only. A few participants of both high openness and high neuroticism showed both curiosity and cautiousness at the same time:

I don't like to click on links from twitter users because a lot of them are spam or virus. I would ask for more info first, might follow after seeing their answer. May or may not report it as spam. (P7)

If someone were to reply at me based off a tweet, I would be likely to see what they put, but I would not follow the account unless I liked what they put up for me to view.(P8).

Field Study

The purpose of the field study was to understand whether openness and neuroticism derived from people's tweets in a lexicon-based approach would substantially improve the success of the ads in a real-world setting (i.e. H3).

Methodology

According to our use case, we targeted Twitter users who have stated their intent to visit NYC in the near future in their public tweets. In this deployment the targeting was done through a semi-automated process, described below.

We started by monitoring Twitter traffic through Twitter streaming API, collecting public tweets that mention at least one term related to NYC (e.g. "New York", "NYC", "Manhattan", "Big Apple") and one term related to traveling (e.g. "travel", "visit", "going", "flight", "hotel"). We then passed the collected tweets to a custom-built machine learning filter that removes tweets that merely indicate a wish to travel (e.g. "I wish I could visit NYC someday") and tweets that describe a trip in the past (e.g. "I visited NYC last year"). This process gave us about 500 Twitter users per day, each authoring a tweet stating some travel intent related to NYC.

We then, on a daily basis, manually filtered those users based on their recent tweets and Twitter profile, further excluding users who may not be appropriate to target. We first excluded all users that clearly represent companies and organizations. Among the rest, we excluded people who may be visiting upstate NY but not NYC, people who live in NYC and the vicinity already, people who will visit more than 6 months in the future, and people who are obviously under 18 years of age. This manual filtering gave us about 200 users per day, and the human judgment ensured that virtually all the targeted users were indeed traveling to NYC and had sufficient reasons to be interested in our travel information service.

Figure 3. Recommendation from TravelersLikeMe.

Within 24 hours of a targeted user stating the intent to travel in a tweet, the TravelersLikeMe account sent the user a tweet advertising its service, in reply to the user's original tweet. Our tweet message was randomly selected from three different versions, each suggesting different activities. The first version is the same as what was used in the survey study (Figure 2), suggesting "*fun bars, broadway shows, and even free kayaking*", the second version suggests "*social hotspots, cozy neighborhoods & themes tours*", while the third version suggests "*luxury hotels, fine dining haunts & designer shops*". The purpose of this design was to understand if the suggested activities affect people's response. However, as our post-hoc analysis showed no significant difference across the three versions, we omit this comparison from the rest of the paper.

As outcomes we measured if each targeted user clicked the embedded link in our tweet, followed our account as suggested, and/or replied to our tweet. We were not able to track how many users reported our account for spam, although the number must have been small because Twitter did not warn us or ban our account.

For the whole field study we ran the above process for 30 days. We further break the study into two phases. In Phase I (the first 10 days), the TravelersLikesMe account not only sent the reply tweet and but also followed each targeted user immediately, while in Phase II (the remaining 20 days), the account only replied to users without following them. The purpose of this two-phase design was to understand if users' responses would differ depending on whether they are followed, given prior evidences that some Twitter users would follow back their followers simply for etiquette (Cha et al. 2010).

Users who indeed signed up for the service by following our account periodically received recommendations of NYC attractions from other travelers (Figure 3). As the recommendation service itself is not the focus of this work, we generated recommendations in a naive way, by crawling positively opinionated tweets about NYC attractions and drawing quotes from those tweets as recommendations. Despite the simplicity, 22% of our participant-followers complimented our service in the field study.

Participants

Over the course of the field study we targeted 5,917 Twitter users, 2,043 in Phase I and the rest 3,874 were in Phase

II. As shown in Table 1, an average targeted user created his or her Twitter account 2~3 years ago, has 12,250 tweets, 572 followees and 1,294 followers. The medians of tweets, followees and followers are substantially lower, at 5,075 tweets, 300 followees and 297 followers.

Measurements

To understand the effect of the derived personality traits, we derived the openness and neuroticism traits for each user in a lexicon-based approach, using the user's past tweets as the input. The approach is based on the LIWC dictionary (Pennebaker et al. 2007).

A trait is computed from a number of word counts. Each word count corresponds to the use of words in a LIWC word category that is known to correlate with the trait. For instance, the neuroticism trait is known to be correlated to the use of anxiety words (e.g. "afraid", "nervous") (Golbeck et al. 2011a, Yarkoni et al. 2010). Given a vector containing the correlation coefficients, and a vector containing word counts of the corresponding word categories, the trait is computed as the dot product of the two vectors, i.e. a linear combination of the word counts weighted by the correlation coefficients.

While several prior works have reported such correlations (Golbeck et al. 2011a, Golbeck et al. 2011b, Sumner et al. 2012, Yarkoni et al. 2010), in this work we have adopted significant correlations from Yarkoni et al. (2010)[1], because the correlations were based on a substantially larger corpus in comparison to other similar works, and because their effectiveness for deriving personality traits has been independently validated by other researchers (Guo et al. 2014).

For our users in the field study the derived openness and neuroticism traits turned out to have a strong negative correlation ($r = -.66$, $p < .001$). The strong negative correlation caused a substantial multicollinearity problem when we included both traits as independent variables in regressions. When we included either trait, openness had a significant positive effect on both dependent variables, while neuroticism had a significant negative effect on both dependent variables. This situation made it impossible to clearly separate the effects of the two traits. Given the situation, we analyzed the two traits using principal component analysis (PCA), a known way to understand the internal structure of personality (Van der Linden et al. 2010). We found that the top principal component from the PCA can capture 83% of the total variance, and have therefore used the component as the single independent variable to represent both openness and neuroticism. We will further discuss the implication of this result near the end of this paper.

[1] We adopted the correlations in Table 1 from Yarkoni. We used correlation coefficients significant at a FDR of 0.05.

Independent variable

Openness-Neuroticism Index: A combination of openness and neuroticism traits derived from a user's past tweets. It is computed as the top principal component in the PCA of the two derived traits (see the preceding paragraphs for rationales and computation details).

The value of this variable is quantitatively close to derived openness minus derived neuroticism. When we replace this variable with either derived openness or negative derived neuroticism, we have obtained similar overall results and slightly smaller effect sizes.

In this way, conceptually a high value on the openness-neuroticism index means that the user is high on openness and/or low on neuroticism in the derived personality.

Dependent Variables

The dependent variables represent each targeted user's responses to our advertisement tweet, and measure the success of our advertisement tweets.

Click: A binary variable, where 1 means the user clicked the embedded link within a week we sent the advertisement tweet, and 0 means that the user did not click.

Follow: A binary variable, where 1 means the user has followed the TravelersLikeMe account on Twitter within a week we sent the advertisement tweet, and 0 means that the user did not follow.

We have omitted replies from the analysis, because overall merely 1.3% users replied to our tweet, making the positive samples too sparse for analysis.

Control Variables

We computed a number of control variables that might have affected targeted users' responses.

User account age: The time since the user's Twitter account was created, measured in days.

User tweet count: The number of tweets the user had posted on Twitter previously.

User followee count: The number of other Twitter users that the targeted user followed.

User follower count: The number of other Twitter users that followed the targeted user.

User topic interest: An approximate measure of the user's topic interest on our advertisement tweet, based on text similarity between a user's past tweets and a number of keywords related to our advertisement. More specifically, for a targeted user we collected all the user's past tweets and represented the user as a bag-of-words vector, where each dimension of the vector represents the number of times the user has used a particular word in past tweets. We created another bag-of-words vector from terms related to NYC (e.g. "New York", "NYC"), terms related to traveling (e.g. "travel", "visit"), and terms related to activities suggested in the advertisement tweet (e.g. "bars", "din-

	Phase I: TravelersLikeMe proactively follows targeted users								Phase II: TravelersLikeMe does not follow targeted users							
	Predicting the likelihood to click the link				Predicting the likelihood to follow the account				Predicting the likelihood to click the link				Predicting the likelihood to follow the account			
	Baseline Model		Proposed Model		Baseline Model		Proposed Model		Baseline Model		Proposed Model		Baseline Model		Proposed Model	
	Std. Coef.	Sig.	Std. Coef.	Sig.	Std. Coef.	Sig.	Std. Coef.	Sig.	Std. Coef.	Sig.	Std. Coef.	Sig.	Std. Coef.	Sig.	Std. Coef.	Sig.
User account age	.099		-.042		-.024		-.131		.191		.093		.063		-.100	
User tweet count	-.199		-.072		-.410	**	-.292	**	-.500	***	-.357	**	-.633	***	-.398	**
User followee count	.062		.060		.852	***	.854	***	-.885	**	-1.01	**	.193	*	.224	*
User follower count	-.079		-.131		-2.01	**	-2.05	***	.064		.063		-.948		-1.41	
User topic interest	.085		-.005		.123	*	.068	*	.068		-.017		.148	**	.059	*
Openness-neuroticism index			.411	***			.396	***			.393	***			.429	***
Deviance	956.8		939.8		1290.3		1277.0		1878.1		1862.9		1443.4		1418.8	
Deviance difference			17.0***				13.3***				15.2***				24.6***	

Table 4. Predicting Responses to Advertisement Tweets in the Field Study

Significance levels: *** p<.001, ** p<.01, * p<.05. The openness-neuroticism index turns out to be a significant positive predictor for both response types (click and follow) across the two study phases, even when we control for the effects of a number of other factors. These results suggest that focusing on people with high derived openness and low derived neuroticism in ad targeting can indeed generate more positive responses from Twitter users, confirming our hypothesis H3.

ing"). We then weighed both bag-of-words vectors using TF*IDF [20], and computed their cosine similarity as the user's topical interest measure. Intuitively, a user with a higher measurement has posted more content relevant to the advertisement in the past, and should therefore be more interested in the content of the advertisement.

Results

We show the descriptive statistics in Table 3. In the table, all the independent variables were aggregated across the two study phases, because they were from user populations selected using the same procedure, and had no statistically significant differences. As for the dependent variables, the click rate of our links was 6~7% for both phases, and only the follow rate was different: 10.4% of the targeted users followed our account in Phase I, while only 4.7% did so in Phase II. This big difference is likely because only in Phase I did our account proactively follow the users, and some of those users in turn followed our account back simply for etiquette (Cha et al. 2010). To understand the effect of the personality traits, we applied logistic regressions to predict the two dependent variables (i.e. click and follow). We used logistic regressions because the dependent variables are binary, in which case logistic regressions are more appropriate than more commonly used ordinary linear regressions.

For each phase and each dependent variable, we applied logistic regressions to build two prediction models, the baseline model using only the control variables as predictors, and the proposed model using both the control variables and the independent variable (i.e. openness-neuroticism index). Comparing the two models allows us to evaluate if the derived personality traits can provide significant additional prediction power to users' responses.

Table 4 presents the logistic regression models, grouped by the two study phases and the two dependent variables (i.e. the two response types). In the table, a model's deviance statistic (i.e. -2 log-likelihood) reflects its goodness-of-fit, and the deviance difference between a baseline model and a proposed model reflects how much more variance the proposed model explains in comparison to the baseline model, similar to the change of R^2 in ordinary linear regressions. For both study phases and both dependent variables, the deviance difference was highly significant ($p < .001$), suggesting that the proposed models had better fits than the baseline models in all cases. We therefore rely on the proposed models for interpreting our results.

A few control variables had significant effects across the two study phases consistently. Users with more tweets were less likely to follow our account, while those with higher topic interest and those who have more followees were more likely to follow.

There were also a few differences between the two phases. In Phase I when our account proactively followed users, popular users (i.e. those with more followers) were much less likely to follow back, while in Phase II when our account stopped following, this effect largely disappeared. This difference suggested that our tactic of following users disproportionally attracted less popular users to follow us back, while more popular users are much more difficult to be influenced. Meanwhile, in Phase II users with more tweets and more followees were less likely to click our links, perhaps because their Twitter streams were so overloaded that without us following them, which often produces a notification from Twitter, they would often miss our tweet entirely and thus fail to respond.

The effects of our derived personality traits, the openness-neuroticism index, have been strong and stable. Across the models in Table 4 its standardized coefficient is around 0.4. In other words, holding all the control variables constant, increasing the openness-neuroticism index by one standard deviation would have increased the odds-ratio of positive response by $e^{0.4}$=1.49, i.e. the advertisement tweet would be 1.49 times more likely than before resulting in a click (or follow) than resulting in no action.

These results strongly support our H3, that targeting Twitter users of high derived openness and low derived neuroticism can greatly improve the success of the ads.

Discussion

Practical Impact for Ad Targeting and Beyond

To the best of our knowledge, this paper is the first to show that personality traits derived in a lexicon-based approach can indeed improve social media ad targeting in a real world setting. In comparison to prior work that analyzed correlations on self-reported survey data (e.g. Hirsh et al. 2012), this work is a big step forward, because by making interventions in real world and controlling for other contributing factors in regressions, we can make a stronger inference on the causal relationship between personality traits and user responses, and can also demonstrate a much more tangible benefit to practitioners.

To illustrate this benefit in a practical marketing context, Figure 4 shows the click rate and follow rate if we had only targeted the most promising candidates according to the openness-neuroticism index in Phase II. Overall, we see

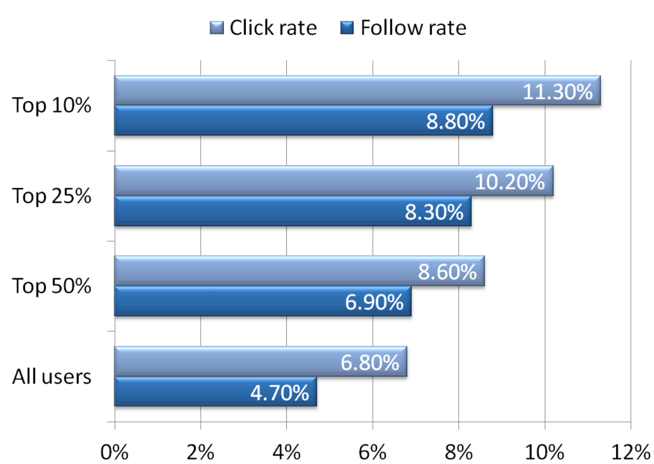

Figure 4. Effects of Openness-Neuroticism Index in Phase II.

If in Phase II we only target the 10% users with the highest openness-neuroticism index, we can increase the click rate from 6.8% to 11.3%, and the follow rate from 4.7% to 8.8%

that the more selective we are, the higher the boost in effectiveness is. For example, while overall the follow rate is 4.7%, if we focus on the top 10% users with the highest openness-neuroticism index, the follow rate increases to 8.8%, which gives a high lift ratio of 187%. The lift ratio here is a common measure of success in marketing, as it tells marketers how much benefit they can gain by focusing their effort on a segment of the population.

Because our observed effect is likely due to the general curiosity and lack of skepticism of certain Twitter users, our results can potentially be generalized broadly to social media advertisement (e.g. display ads on Facebook and promoted tweets on Twitter), regardless of topics or subject domains. It is reasonable to also expect that the observed effect would be weaker for major established brands, because the general public is already familiar with these brands, so that curiosity and skepticism toward the unknown is much less relevant.

More broadly, our results also demonstrate the effectiveness of derived personality traits despite the mixed results on their accuracy in prior work (Golbeck et al. 2011b, Guo et al. 2014, Sumner et al. 2012). There are numerous ways that we may leverage the derived personality in future work to better engage with people. For example, by deriving people's neuroticism from their online posts, we may be able to identify people more prone to anxiety, and encourage their friends to provide them more preempt emotional support.

Structure of Big5 Personality Traits

In the field study the derived openness and neuroticism traits were so strongly correlated ($r = -.66$, $p < .001$) that we were forced to combine the two traits into a single openness-neuroticism index. How do we interpret that?

Combining multiple Big 5 traits into a higher-order variable is not a new practice. Notably, a number of recent works have proposed a general factor of personality (GFP) that combines all Big 5 traits into a single factor, based on their common correlations (e.g. Musek et al. 2007, Van der Linden et al. 2010). For instance, van der Linden et al. (2010) conducted a meta-analysis of more than 200 personality studies, validated the GFP as a coherent linear combination of Big 5 traits, and found that the GFP can strongly predict job performance. Given that openness has a stable positive loading in the GFP while neuroticism has a stable negative loading, it is possible to view our openness-neuroticism index as a component of the GFP. On the other hand, Ashton et al. (2009) have argued that while Big 5 traits indeed sometimes correlate strongly, making their combinations a valid quantitative solution, in general the five traits are orthogonal so that a combined variable cannot be viewed as a valid theoretical construct. They further attributed the occasional correlation of Big 5 traits to the sporadic use of certain words that correspond to multiple traits (e.g. the word "friendly" corresponds to high levels of two Big 5 traits extraversion and agreeableness). It is thus possible that the strong negative correlation between the derived traits in our field study is simply due to the word use pattern in people's tweets.

Because openness and neuroticism only had a mild negative correlation when measured through questionnaires in the survey study, the strong negative correlation from the field study is perhaps due to the lexicon-based personality derivation. We therefore prefer the latter interpretation based on Ashton et al., and consider further explorations of this issue as future work.

Ethical Considerations

Our field study involved sending unsolicited advertisements as replies to Twitter users. While we have seen this approach being used in the wild on Twitter, it is morally contentious that we would not advocate its general adoption for commercial gains. In addition, as this approach is in direct competition with Twitter's paid ads, it is unlikely that Twitter or any other ad-supported social media platform would allow such approach at a truly large scale.

Our foremost motivation behind this approach was to be able to target Twitter users and track the results in a realistic field study setting, so as to allow our research. To minimize disruptions we cause, we manually ensured that our replies were relevant to every targeted user, and provided a meaningful service to users. These measures have paid off: while prior works have reported Twitter accounts being banned for spamming after sending fewer than 200 unsolicited replies (Nichols et al. 2012), our account was able to complete the whole study, sending thousands of replies. Meanwhile, 22% of our participant-followers complimented our service on Twitter, showing that our study indeed provided benefits to our participants regardless of our research contribution. It is our belief that future field studies on Twitter and social media in general should similarly be considerate of users' time and opinions.

Conclusion

We created a Twitter-based travel information service, and found that Twitter users of high openness and low neuroticism responded more favorably to unsolicited advertisements of our service. Two studies showed that this result holds true no matter if the two traits were measured through questionnaires or derived from people's tweets. This work contributes to research by demonstrating the feasibility of using derived personality for ad targeting, and by furthering our understanding of personality itself.

References

Ashton, M.C., Lee, K., et. al. 2009. Higher order factors of personality: Do they exist? Personality & Social Psychology Review.

Cha, M., Haddadi, H., et al. 2010. Measuring user influence in Twitter: The million follower fallacy. In ICWSM 2010.

Chen, J., Cypher, A., et al. 2013. CrowdE: Filtering tweets for direct customer engagements. In ICWSM 2013.

Chen, T. 2011. Personality traits hierarchy of online shoppers. International Journal of Marketing Studies, 3(4).

Evans, A. M., & Revelle, W. 2008. Survey and behavioral measurements of interpersonal trust. Journal of Research in Personality, 42(6).

Golbeck, J. Robles, C., et al. 2011. Predicting personality from Twitter. In IEEE SocialCom.

Golbeck, J. Robles, C., et al. 2011. Predicting personality with social media. In.CHI 2011 Extended Abstracts.

Golbeck, J., & Norris, E. 2013. Personality, movie preferences, and recommendations. In ASONAM 2013.

Goldberg, L. R. 1993. The structure of phenotypic personality traits. American Psychologist 48 (1).

Goldberg, L. R., Johnson, J. A., et al. 2006. The International Personality Item Pool and the future of public-domain personality measures. Journal of Research in Personality, 40.

Gou, L., Zhou, M. X., et al. 2014. KnowMe and ShareMe: Understanding automatically discovered personality traits from social media and user sharing preferences. In CHI 2014.

Hirsh, J. B., Kang, S. K., et al. 2012. Personalized persuasion tailoring persuasive appeals to recipients' personality traits. Psychological science, 23(6).

Hu, R., & Pu, P. 2011. Enhancing collaborative filtering systems with personality information. In RecSys 2011.

Lee, K., Mahmud, J., et al. 2014. Who will retweet this?: Automatically identifying and engaging strangers on Twitter to spread information. In IUI 2014.

Miller, C. C., & Vega, T. 2010. After building an audience, twitter turns to ads. The New York Times.

Musek, J. 2007. A general factor of personality: Evidence for the Big One in the five-factor model. Journal of Research in Personality, 41(6).

Nichols, J., & Kang, J. H. 2012. Asking questions of targeted strangers on social networks. In CSCW 2012.

Pennebaker, J. W., Chung, C. K., et al. 2007. The development and psychometric properties of LIWC2007. LIWC.net.

Sumner, C., Byers, A., et al. 2012. Predicting Dark Triad Personality Traits from Twitter usage and a linguistic analysis of Tweets. In ICMLA 2012.

Tausczik, Y. R. & Pennebaker, J. W. 2010. The Psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology 29 (1), 24-54.

Van der Linden, D., te Nijenhuis, J., et al. 2010. The general factor of personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. Journal of Research in Personality, 44(3).

Yarkoni, T. 2010. Personality in 100,000 words: A large-scale analysis of personality and word usage among bloggers. Journal of Research in Personality.