Predicting Perceived Brand Personality with Social Media

Anbang Xu[1], Haibin Liu[1], Liang Gou[1], Rama Akkiraju[1],
Jalal Mahmud[1], Vibha Sinha[1], Yuheng Hu[2], Mu Qiao[1]

1. IBM Research - Almaden, 2.University of Illinois at Chicago
{anbangxu, haibinliu, lgou, akkiraju, jumahmud, vibha.sinha, mqiao}@us.ibm.com, yuhenghu@uic.edu

Abstract

Brand personality has been shown to affect a variety of user behaviors such as individual preferences and social interactions. Despite intensive research efforts in human personality assessment, little is known about brand personality and its relationship with social media. Leveraging the theory in marketing, we analyze how brand personality associates with its contributing factors embodied in social media. Based on the analysis of over 10K survey responses and a large corpus of social media data from 219 brands, we quantify the relative importance of factors driving brand personality. The brand personality model developed with social media data achieves predicted R^2 values as high as 0.67. We conclude by illustrating how modeling brand personality can help users find brands suiting their personal characteristics and help companies manage brand perceptions.

Introduction

Brand has personality because people tend to associate human attributes with brands. Brands are often consumed in a social setting where a brand's personality creates brand differences that can satisfy people's self-expression and social needs (Dittmar 1992; Markus and Nurius 1986). People use brands to define how young or old they are, how masculine or feminine they are, how upscale or downscale they are, and how different or similar to members of their social groups. These brand personality traits can effectively assess how brands are perceived in people's minds (Aaker 2012), yet they are not captured in *human* personality measures.

The social computing research community has recently become interested in predicting human personality from social media (Chen, et al. 2014; Golbeck, et al. 2011) and developing corresponding personalized systems (Gou, et al. 2014). However, in many cases, assessing human personality is only one side of the coin. Research shows that the relationship between human and brand personality im-

pacts users' preferences, satisfaction and their social interactions with others (Cialdini and Trost 1998; Elliott and Wattanasuwan 1998). Assessing brand personality is essential in order to gain a full picture of user behavior.

Until now, to accurately gauge the perceived brand personalities, survey tests have to be taken by users. This makes it impractical to apply brand personality analysis on a large scale in many recommender systems and social media domains. Despite significant research efforts in conceptualizing brand personality and its contributing factors, little is known of the relationship between brand personality and social media, where the latter has the power to shape the perceived personality of a brand. We take an initial step towards predicting brand personality with social media.

This paper aims to investigate the relationship between brand personality and its contributing factors on social media. We collect 10,950 survey responses to obtain 219 brands' perceived personality as ground truth. Our model is built on a large corpus of social media data gathered from three factors for the brands. In total, 1,996,214 brand follower descriptions (*User Imagery*), 312,400 employee reviews (*Employee Imagery*), and 680,056 brand official tweets (*Official Announcement*) are collected from Twitter and Glassdoor (glassdoor.com). Word use features are extracted from the factors, and each factor is used to model brand personality separately. Surprisingly, results show that *User Imagery* and *Employee Imagery* factors are equally important in predicting brand personality, while *Official Announcement* has significantly less predictive power. With the factors combined, our brand personality model achieves predicted R^2 as high as 0.67.

In this paper, we first review the marketing theories that we followed in measuring brand personality and developing its contributing factors embodied in social media. Then, we describe the comparison results between the factors. Next, we present the brand personality model including the factors together. The paper concludes by discussing our findings toward implications for practice and research.

Brand Personality

The term brand personality, first introduced by Martineau in 1958 (Martineau 1958), refers to a set of human characteristics associated with a brand. For example, Apple is perceived to be young, while IBM is perceived to be older. Within thirty years, brand personality became widely accepted as an effective way to capture users' perceptions of brands, which affect users' preferences, and their self and social identities (Cialdini and Trost 1998; Dittmar 1992; Elliott and Wattanasuwan 1998; Markus and Nurius 1986).

A great number of studies have been carried out to measure brand personality. Researchers initially relied on qualitative methods, such as photo-sorts, free associations, and psychodramatic exercises (Gardner and Levy 1955), but these open-ended techniques are often abandoned in the later stages of research as marketers look for more quantitative ways to detect and enumerate differences among their brands. Also, researchers attempted to use human personality scales developed in psychology to directly measure brand personality (Goldberg 1990). However, these scales are not adequate and powerful enough to describe the personality of a brand.

Dimensions Measuring Brand Personality

In 1997, Aaker developed *brand personality scales* (Aaker 1997). She analyzed the individual ratings of 37 brands on 114 personality traits by 613 respondents recruited in the United States. As a result, *brand personality scales* are made up of 42 traits. These traits are grouped into five dimensions: *Sincerity*, *Excitement*, *Competence*, *Sophistication* and *Ruggedness* (see table 1). *Sincerity* encapsulates traits related to *family-oriented*, *small-town*, *wholesome*, *sincere*, and *friendly*. *Excitement* denotes traits described as *daring*, *young*, *trendy*, *imaginative*, *unique* and *independent*. *Competence* is represented by traits referred to *reliable*, *secure*, and *successful*. *Sophistication* is characterized by traits such as *upper-class* and *good-looking*. *Ruggedness* is typified by traits such as *masculine* and *outdoorsy*.

Brand personality scales have been demonstrated to be a reliable, valid and generalizable scale for assessing brand personality. Since 1997, most of the marketing literature has adopted Likert scale surveys based on Aaker's scale to assess brand personality.

However, the inherent limitation of survey-based approaches is the flexibility. Conducting surveys is often time-consuming and labor-intensive. It is expensive to assess brand personality frequently. Also, survey-based approaches suffer from non-response and sampling related deficits, and carry the risk of experimenter bias.

Factors Driving Brand Personality

Perceptions of brand personality traits can be possibly formed and influenced by at least three factors: *User Imagery*, *Employee Imagery*, and *Official Announcement*.

User Imagery and *Employee Imagery* are the set of human characteristics associated with typical users and employees of a brand. Based on stereotyping theory (Harré and Lamb 1986), customers may develop generalized beliefs about groups of users and employees in which all individuals from one group are regarded as having the same set of leading attributes. Customers' beliefs about users and employees may affect their perceptions of the corresponding brand (Wentzel 2009). A wide range of user and employee-generated content has been shared on social media, offering a new opportunity to capture the *User Imagery* and *Employee Imagery* of a brand.

Official Announcement refers to marketing messages, which are designed specifically to make consumers aware of a brand and develop a positive attitude towards it (Schultz 1992). Marketing messages were often distributed to consumers through a variety of media channels such as social media, TV, and radio.

In theory (Aaker 2012; Parker 2009), *User Imagery* is the primary factor driving brand personality. Yet, little research has examined the relative importance of the factors, especially when they are embodied in social media.

Computational Methods Assessing Human Personality

Numerous research efforts have been focused on modeling *human* personality (Gosling, Gaddis and Vazire 2007), emotion (Bollen, Pepe and Mao 2009; Xu, et al. 2012), values (Chen, et al. 2014), satisfaction (Sibona and Choi 2012), engagement (Shami, et al. 2015) mental health (De Choudhury, et al. 2013), political orientation (Cohen and Ruths 2013), and dietary choices (Kulshrestha, et al. 2015). A person's personality can be estimated based on his/her textual data such as essays (Mairesse, et al. 2007) and online posts (Golbeck, et al. 2011; Gou, et al. 2014; Yarkoni 2010). For instance, Yarkoni modeled people's personality using their blogs (Yarkoni 2010). The model was built on word use features extracted from the blog content. Our research method is similar to prior work. We extracted word use features for a brand related to three aspects, *User Imagery*, *Employee Imagery* and *Official Announcement*.

However, in many cases, assessing human personality is only one side of the coin. For instance, in order to gain a full picture of user behavior, both user and brand personality are essential. We take an initial step towards predicting brand personality with social media.

Research Questions

RQ1: The existing marketing literature suggests brand personality can be derived from three factors: *User Imagery*, *Employee Imagery*, and *Official Announcement*. As these factors are represented in social media, which one is more important in predicting brand personality?

RQ2: How and in what combination can these factors predict brand personality?

Method

To address our research questions, we obtained the ground truth of perceived brand personality by conducting a survey on 219 brands. The social media data was collected to model these brands' personality.

Brand Selection

Two criteria guided the choice of brands: First, well-known brands were selected so that a national sample could be used to gather survey data. Specifically, these brands were listed among the top 1,000 Fortune companies ranked by gross revenue (http://fortune.com/global500; http://www.geolounge.com/fortune-1000-companies-2014-list). They should have both corporate offices and considerable markets in the United States. Second, a large variety of brands were covered to enhance the generalizability of the prediction model across product categories (Katz 1960). We systematically chose brands spanning a wide variety of categories such as restaurants, clothing, automobiles, electronics, and financial services. In total, 219 brands were selected (see examples in Table 1).

Ground Truth Collection

We conducted a survey to acquire ground truth for brand personality modeling. 50 survey responses were collected for each brand. We obtained 10,950 valid responses on 219 brands from 3,060 participants. The survey procedure is described below.

Participants Recruitment

Amazon MTurk was used to recruit participants for two reasons: a) MTurk reaches a more diverse population than traditional student samples and community samples (Buhrmester, et al. 2011), allowing researchers to gain generalizability to broader populations. For instance, MTurk workers have a similar income distribution compared with the general U.S. population (Ipeirotis 2010). b) MTurk allows for a rapid collection of survey responses. Most responses can be obtained within a few days.

Three criteria were considered in selecting the participants: First, participants should be familiar with a brand in order to describe their perceptions of the brand. In the survey, participants were asked to assess how familiar they were with a brand. Participants who were not familiar with it were excluded from the study. Second, participants should not share common interests with a brand. We removed the responses from participants who reported they or their family members have ever been employed by the company of that brand. Third, all participants were required to reside in the United States, to be consistent with the criterion used in the brand selection. Thus, we only recruited MTurk workers identified as living in the United States.

Participants' Background

All 3,060 participants were 18 or older. 16.5% of participants at age 20–24, 18.5% at age 25–29, 16.7% at age 30–34, 12.7% at age 35–39, 10.8% at age 40–44, 8.0% at age 45–49, 7.1% at age 50–54, 4.4% at age 55–59, and 1.9% at age 65 or older. 66.0% of participants were female. Although not uniformly distributed, each group was well represented in the sample.

In the study, most participants (97.0%) were familiar with brands shown in the survey. Familiarity rating was obtained by having participants rate a brand on a 7-point scale ranging from *extremely familiar* (7) to *extremely unfamiliar* (1). Following prior work (Malär, et al. 2012), we did not consider the responses with a familiarity rating below 4. As a result, the mean was 6.0 with standard deviation 0.84.

Questionnaire

Participants responded an online standardized questionnaire with regard to their perceptions of one brand (Aaker 1997). They rated how descriptive the 42 traits were of the brand in general, using a 7-point scale anchored at *not at all descriptive* (1) to *extremely descriptive* (7). The traits were arranged in random order to control order effects. Duplicated questions (reverse-scored items) were included in the questionnaire to help filter low quality responses. For example, we asked participants to rate how descriptive both "young" and "old" are of a brand. "old" is the reverse-scored item of "young". There were two reverse-scored items randomly added in a survey. If participants' ratings were contradictory, their responses were regarded as invalid and they were not allowed to work on other survey tasks. 27% of participants failed the check and their responses were removed. We kept posting survey tasks on MTurk until a brand had 50 valid responses.

Similar to prior work (Aaker 1997), participants were allowed to describe their perceptions on multiple brands. After participants completed the questionnaire for one brand, they were allowed to choose to work on the questionnaire for another brand. Participants were paid $0.1 for completing a questionnaire. The questionnaires were completed within a few days. On average, participants answered 3.6 questionnaires.

Dimension	Trait	Top 5 rated brands	Mean	STD	Distribution
Sincerity	**down-to-earth**	*Cracker Barrel, Old Navy, IHOP, LEGO, Dick's*	4.45	0.64	
	family-oriented	*Walt Disney, LEGO, Toys R Us, Six Flags, IHOP*	4.88	0.95	
	small-town	*Cracker Barrel, Dairy Queen, Dollar General, Denny's, Kmart*	3.3	0.74	
	honest	*Cracker Barrel, Amazon, PetSmart, Walgreens, PayPal*	4.87	0.45	
	sincere	*PetSmart, Cracker Barrel, Walt Disney, Farmers Insurance, Walgreens*	4.58	0.46	
	real	*Campbell Soup, Home Depot, Amazon, Ford Motor, Whole Foods*	5.25	0.34	
	wholesome	*Whole Foods, Walt Disney, LEGO, Cracker Barrel, Campbell Soup*	4.44	0.7	
	original	*Walt Disney, LEGO, Apple, Amazon, Google*	4.84	0.54	
	cheerful	*Walt Disney, Toys R Us, LEGO, Six Flags, IHOP*	4.64	0.68	
	sentimental	*Walt Disney, Campbell Soup, Tiffany & Co., Cracker Barrel, LEGO*	3.62	0.64	
	friendly	*Walt Disney, PetSmart, LEGO, Cracker Barrel, IHOP*	5.11	0.58	
Excitement	**daring**	*Victoria's Secret, Red Bull, Six Flags, Urban Outfitters, Samsung*	4.04	0.64	
	trendy	*Apple, Forever 21, Sephora, Starbucks, Samsung*	4.71	0.86	
	exciting	*Walt Disney, Six Flags, Apple, Electronic Arts, BMW*	4.49	0.7	
	spirited	*Walt Disney, Six Flags, ESPN, Progressive Insurance, Forever 21*	4.45	0.59	
	cool	*Apple, LEGO, Urban Outfitters, Samsung, Burberry*	4.57	0.77	
	young	*Toys R Us, Forever 21, The Children's Place, MTV, LEGO*	4.18	0.89	
	imaginative	*Walt Disney, LEGO, Google, Apple, Mattel*	4.59	0.7	
	unique	*Walt Disney, LEGO, Apple, Tiffany & Co, Burberry*	4.53	0.64	
	up-to-date	*Samsung, Amazon, Google, Apple, Intel*	5.37	0.53	
	independent	*Apple, Volvo, BMW, Netflix, Uber*	4.7	0.37	
	contemporary	*Apple, Starbucks, Google, Samsung, Sephora*	4.83	0.56	
Competence	**reliable**	*Amazon, Volvo, Google, UPS, Samsung*	5.43	0.46	
	hard-working	*UPS, AutoZone, FedEx, Home Depot, General Electric*	4.95	0.5	
	secure	*PayPal, Amazon, Intel, Volvo, Marriott*	4.97	0.49	
	intelligent	*Apple, Google, IBM, Intel, Microsoft*	4.91	0.61	
	technical	*Intel, Samsung, Apple, IBM, Sony*	4.24	1.24	
	corporate	*IBM, Merrill Lynch, Walmart, Microsoft, Capital One*	5.64	0.52	
	successful	*Walt Disney, Google, Tiffany & Co, Amazon, Apple*	5.85	0.42	
	leader	*Google, Apple, Amazon, Walt Disney, Intel*	5.14	0.53	
	confident	*Mercedes-Benz, Louis Vuitton, BMW, Victoria's Secret, Ralph Lauren*	5.27	0.42	
Sophistication	**upper-class**	*Mercedes-Benz, Louis Vuitton, Tiffany & Co., BMW, Bloomingdale's*	4.07	1.01	
	glamorous	*Tiffany & Co., Mercedes-Benz, Victoria's Secret, Sephora, Louis Vuitton*	3.56	1.01	
	good-looking	*Victoria's Secret, BMW, Sephora, Burberry, Tiffany & Co.*	4.33	0.88	
	charming	*Walt Disney, Tiffany & Co., Pottery Barn, Burberry, Children's Place*	4.05	0.71	
	feminine	*Victoria's Secret, Ann Taylor, L'Oreal, Revlon, Sephora*	3.56	1.02	
	smooth	*Audi, BMW, Apple, Mercedes-Benz, Starbucks*	4.19	0.54	
Ruggedness	**outdoorsy**	*Columbia Sportswear, The North Face, Dick's, Sports Authority*	3.16	0.96	
	masculine	*ESPN, Dick's, Advance Auto Parts, Home Depot, Lowe's*	3.81	0.83	
	Western	*Cracker Barrel, Buffalo Wings, Wells Fargo, Arby's, AM Eagle Outfitters*	3.7	0.43	
	tough	*Goodyear, Ford Motor, Home Depot, Dick's, General Motors*	3.65	0.78	
	rugged	*The North Face, Home Depot, Ford Motor, Dick's, Columbia Sportswear*	3.28	0.8	

Table 1. Five dimensions measure brand personality, and each consists of several personality traits (Left columns). Middle column shows top-rated brands in our survey. Right columns show the descriptive statistics and rating distribution of the trait.

Brand Personality Scales Analysis

The average participants' ratings of a brand was used to measure the brand's personality (Aaker 1997; Malär, et al. 2012). Consistent with prior work, all traits within each of the five dimensions have relatively high correlation values ($\mu = 0.60$, $\sigma = 1.0$), and the average correlation for all pairs of 42 traits was low ($\mu = 0.20$, $\sigma = 0.32$). In addition, on a 7-point scale, the standard error of an estimation of a brand's personality trait was from 0.15 to 0.25 with a mean of 0.20.

Our survey results of brand personality aligned with prior findings. For instance, we found that ESPN was the most *masculine* brand, and Walt Disney was the most *family-oriented* brand (see Table 1). Also, the most significant difference between Apple and IBM was the trait *young* ($t = 10.5$, $p < .01$). In other words, Apple was perceived to be much younger than IBM. These results indicate the ground truth of brand personality was appropriately obtained.

Social Media Data Collection

The social data was collected from three contributing factors of brand personality. 1,996,214 brand follower descriptions (*User Imagery*), 312,400 employee reviews (*Employee Imagery*), and 680,056 brand tweets (*Official Announcement*) were collected from Twitter and Glassdoor.

User Imagery

A brand's Twitter account often has followers. These followers are very likely to be using and/or liking the particular brand. We considered a set of brand followers as *User Imagery* represented on social media. For each brand, we first identified its Twitter account and sampled 20,000 followers from the account. We then collected followers' self-description (http://support.twitter.com/articles/166337-the-twitter-glossary), which is a short description in a follower's public profile. Since a significant proportion of Twitter accounts are fake or inactive (Thomas, et al. 2011), only considering followers with description is an effective way to reduce noise in the data. In total, we collected 1,996,214 followers' description by querying the Twitter API, and the average description length was 12.1 words.

Employee Imagery

Glassdoor.com is a social media platform, where current and former employees can post reviews about their employers. In the reviews, employees often provide statements about working conditions, company culture, management style, and so on. These reviews were used to capture *Employee Imagery*. We obtained 312,400 employee reviews. Brands, on average, had about 1,400 employee reviews. The average review length was 86.85 words.
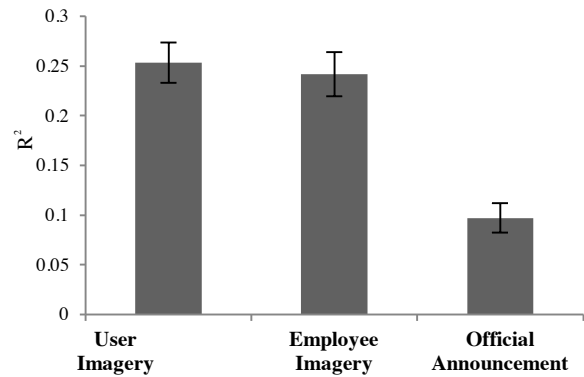

Figure 1. The comparison of the predicted R^2 values among the contributing factors: User Imagery, Employee Imagery, and Official Announcement. Each factor was used to predict brand personality individually.

Official Announcement

Twitter allows companies to create their own accounts and push intended information to the public. We used the tweets from a brand Twitter account as its official announcements. Due to the limitations of the Twitter API, if an account had more than 3,200 tweets, we were only able to collect the last 3,200 tweets. Thus, 680,056 tweets were obtained and the average tweet length was 14.0 words.

Feature Extraction from Social Media Data

To make a fair comparison among *User Imagery*, *Employee Imagery*, and *Official Announcement*, we used the same features from *Linguistic Inquiry and Word* (LIWC) to characterize each factor. LIWC, a dictionary developed in psycholinguistic field, has been widely used in psychological (Pennebaker, et al. 2001) and social computing research (Chen, et al. 2014; Xu and Bailey 2012) to quantify the linguistic and psychological features of a text document.

In our study, a text document was one single follower self-description, employee review, or brand official tweet. Over 60 LIWC features were extracted by counting the number of words in each document that match a word in a LIWC category (60 categories in total). For each brand, we used seven descriptors of the distribution of documents over each LIWC feature: mean, 5th to 95th percentile, variance, skew, kurtosis, minimum, and maximum.

Brand Personality Modeling

Dependent and Predictive Variables

Brand personality scales have 42 traits (see Table 1); each trait was regarded as a dependent variable. To address the first research question (**RQ1**), models were developed separately based on *User Imagery*, *Employee Imagery* and *Official Announcement* factors. A model with each factor had 420 predictive variables (60 LIWC features x 7 descriptors = 420 predictors). The factors were used together

to predict *brand personality scales* for the second research question (**RQ2**). The model had 1,260 predictive variables (420 predictors x 3 factors = 1,260 predictors). In our study, each brand was treated as one observation, so that 219 observations were included.

Modeling Techniques and Performance Measure

Lasso regularized regression was employed to modeling brand personality. Since the number of predictors (e.g. 1,260 predictors) exceeds the number of observations (219 brands) with a high collinearity between predictors, Lasso is able to seek for a sparse solution by shrinking the coefficients of weak and/or correlated predictors to zero (Tibshirani 1996). In other words, Lasso regression can select a set of best explanatory predictors.

We followed a standard process to perform Lasso implemented by glmnet[1]:

- Used 10-fold cross-validation (initial cross-validation) to repeatedly split the data into training and testing sets.

- For each split, glmnet performed another 10-fold cross-validation on the training set to determine the optimal values for lambda. Then, glmnet refitted the model with the training set and the optimal lambda, and made predictions from the testing set. The lambda values were computed for a split. Their values were from 0.0058 to 0.0953 with a mean of 0.0193. Once the model was refitted with a training set, the features were selected by Lasso for the split. The number of selected features in a model was from 29 to 148 with a mean of 81.3. All these selected features were used to make predictions for the corresponding testing set.

- The model performance was measured by the predicted R^2, calculated by the initial cross-validation. Predicted R^2 was computed by systematically removing each subset from the data set, estimating the regression equation, and determining how well the model predicts the removed subset. Predicted R^2 can avoid overfitting the model and can be more useful than adjusted R^2 for comparing models because it is calculated using observations not included in model estimation (Montgomery and Peck 1982). Larger values of predicted R^2 suggest a model has greater predictive ability. Since there were 42 dependent variables (42 personality traits), a predicted R^2 value was calculated for each dependent variable (personality trait).

Comparison with Individual Factors (RQ1)

One-way ANOVA revealed significant differences in predicted R^2 values among the three factors (F (2, 123) =

Figure 2. Prediction results of brand personality at dimension level using User Imagery, Employee Imagery, and Official Announcement separately.

22.88, $p < 0.001$), as shown in Figure 1. Post-hoc tests between paired factors showed that R^2 values predicted from *Official Announcement* factor were significantly lower than from *User Imagery* and *Employee Imagery* factors ($p < 0.01$). This indicates that *Official Announcement* has significantly less predictive power than *User Imagery* and *Employee Imagery* factors[2].

Interestingly, there was no statistically significant difference in predicted R^2 values between *User Imagery* and *Employee Imagery* ($p = 0.56$). Previous research in marketing often emphasizes *User Imagery* as the key driver of brand personality (Aaker 2012). However, our results reveal that *Employee Imagery* is as important as *User Imagery* in predicting brand personality.

Analysis of Brand Personality Dimensions

We examined how the three factors affect the predicting performance in the five dimensions of brand personality.

Predicted R^2 values by *Employee Imagery* were significantly higher than predicted R^2 values by *User Imagery* in *Competence* and *Sophistication* dimensions ($p < 0.05$). In contrast, in *Sincerity* and *Ruggedness* dimensions, predicted R^2 values by *User Imagery* were significantly greater than predicted R^2 values by *Employee Imagery* ($p < 0.05$). As the results show, although *User Imagery* and *Employee Imagery* factors showed no difference overall in predicting brand personality, these two factors impacted prediction performance very differently in different personality dimensions (see Figure 2).

---

[1] It was implemented by glmnet (http://cran.r-project.org/web/packages/glmnet). Since the multivariate Gaussian model selects the same predictors for the dependent variables and our focus is not to examine the relationship among these variables, we adopted univariate Gaussian model to select the most effective predictors for each dependent variable.
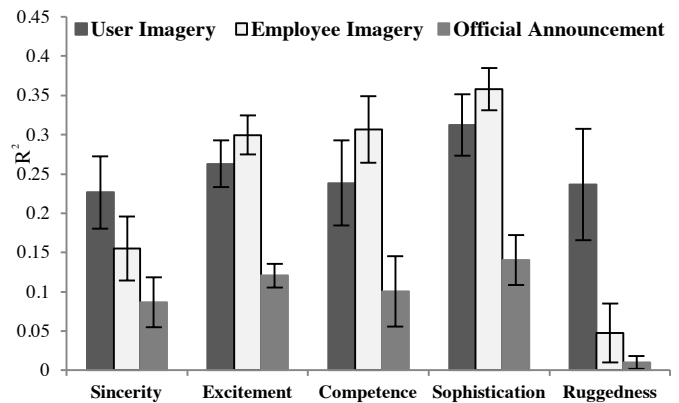
[2] We also undersampled the *User* data until the number of user descriptions was equal to the number of tweets (*Announcement*) for each brand. There were no statistically significant differences in predicted R^2 values between the undersampled and full-size data ($p = 0.32$, t-test). R^2 values from *Announcement* was still lower than from the undersampled *User* data ($p < 0.01$). So we only reported the results from the full-size data.
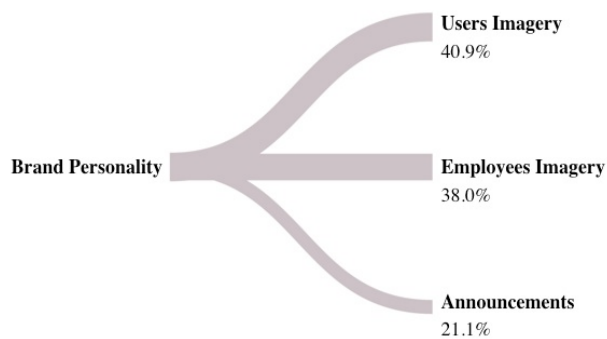
Figure 3. The relative importance of three contributing factors in predicting brand personality in the final model. The factor weight was computed by summing the absolute values of the coefficients belonging to it.

Additionally, the *Official Announcement* factor had significantly lower R^2 values than other factors across all the brand personality dimensions ($p < 0.05$).

Prediction with Combined Factors (RQ2)

The model combing three factors *User Imagery*, *Employee Imagery*, and *Official Announcement* together predicted the brand personality traits with R^2 values as high as 0.67, and achieved a MAE as low as 0.0807 on a continuous 0–1 scale. Each personality trait had a predicted R^2 value and the range of R^2 values was from 0.18 (MAE = 0.1215) to 0.67 (MAE = 0.0807) with an average value of 0.42 (MAE = 0.1065). This resolution is acceptable for most applications (Chen, et al. 2014; Gou et al, 2014).

The prediction performance was consistent and reasonably accurate across the five brand personality dimensions. There were no significant differences in predicted R^2 values among the five dimensions (F (4, 37) = 2.12, p = 0.098). More specifically, looking at individual traits, a majority (60%) of the traits' R^2 values were higher than 0.4, and 26% of the traits' R^2 values were above 0.5. The *technical* trait had the highest R^2 value (0.67), followed closely by *feminine* (0.58), *charming* (0.57), *small-town* (0.54), *cheerful* (0.55), *daring* (0.55), *trendy* (0.53), *cool* (0.52), *young* (0.52), *intelligent* (0.50), *smooth* (0.50), etc.

The R^2 values of *Western* and *honest* traits were below 0.20 ($R^2_{Western}$ = 0.19; R^2_{honest} = 0.18). *The relatively low R^2 value suggests* that predictive variables in the current model only accounts for a small proportion of variability for the prediction of these two traits.

In addition, we analyzed the relative importance of three factors *User Imagery*, *Employee Imagery,* and *Official Announcement* on the final regression model. After the Lasso regression fitted the data, the *β* coefficients associated each factor were obtained. These coefficients enabled us to investigate the relative power of the factors, when they were used together in predicting brand personality. The weight

of a factor was calculated by summing the absolute coefficient values of the variables belonging to it[3] (Gilbert and Karahalios 2009). Specifically, we first computed a factor's weight for each brand personality trait individually, and then averaged over the traits. Figure 3 shows that *User Imagery* and *Employee Imagery* had a dominant influence in predicting brand personality, while *Official Announcement* had the least important influence. This is consistent with results in regression analysis using each factor separately.

Analysis of Word Use

We examined the influence of various LIWC categories on the predictive model (see Figure 4). The weight of a category was computed by summing the standardized coefficients of the predictive variables belonging to it (Gilbert and Karahalios 2009). Specifically, for each individual brand personality trait, a category's weight was calculated by summing the corresponding coefficients across the three factors. Then, the weights were averaged over the traits.

We found *Personal Concerns* had the most predictive power in predicting brand personality. Most brand personality traits had correlations with *Personal Concerns* words such as *leisure activity* (e.g. *sport*, *TV & movie*), *financial issues* (e.g. *money*), and *metaphysical issues* (e.g. *death*). For instance, we found that *sport* words offered the most significant influence within the model for *Ruggedness*. A greater proportional use of *sport* words was likely to increase the chance for a brand to be perceived as *masculine* (β = 0.38), *rugged* (β = 0.20), *tough* (β = 0.18), and *outdoorsy* (β = 0.09). In contrast, more *TV & Movie* words were likely to decease the chance to be perceived as *masculine* (β = - 0.14), and *outdoorsy* (β = - 0.13). Also, *money* words were negatively correlated to the perception of *family-oriented* (β = -0.14); *death* words were positively correlated to the perception of *hard-working* (β = 0.1).

Linguistic Processes had a significant influence on regression models for *Competence* and *Sophistication* dimensions. For example, a high frequency of *prepositions* was likely to enhance the perception of *Competence* including *corporate* (β = 0.16), *secure* (β = 0.12), *successful* (β = 0.08), *reliable* (β = 0.06), *technical* (β = 0.06), and *leader* (β = 0.06). The frequent use of *first & second person pronouns* had a negative correlation with the perception of *Sophistication* such as *upper-class* (β = -0.20). Similarly, *swearing* words were negatively correlated with *upper-class* (β = -0.04) and *smooth* (β = -0.09), but they

---

[3] When predictors are highly correlated, Tikhonov regression shrinks coefficients of correlated variables together, leading to a more stable behavior than Lasso regression (Grave, Obozinski and Bach 2011). We applied Tikhonov regression to compute the coefficients for the predictors, and the aggregated results were consistent with Lasso regression. So we only reported the results from Lasso.
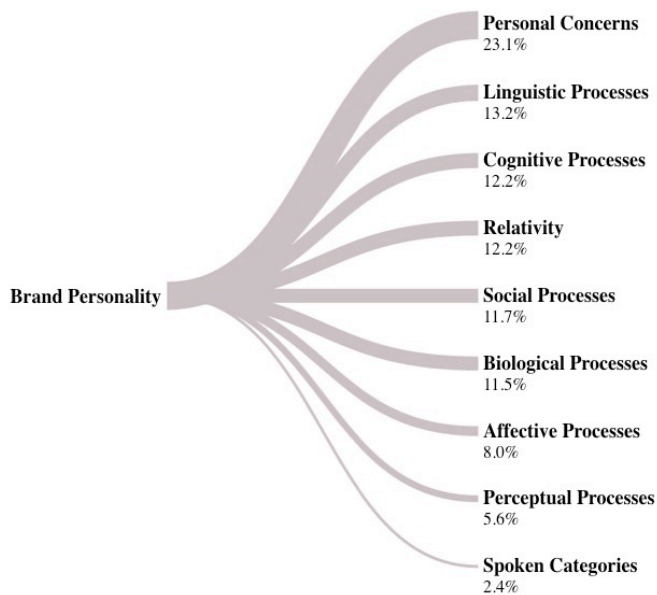
Figure 4. The predictive power of LIWC categories. The weight of a LIWC category was calculated by summing the absolute values of the coefficients belonging to it.

were positively correlated with *young* ($\beta = 0.03$). Many *prepositions* and fewer *first & second person pronouns* are often found in official documents and academic writings (Herring 1996). This indicates that a formal language style could enhance the perception of *Competence* and *Sophistication*.

We found *Affective Processes* and *Cognitive Processes* exerted a significant influence on the model for *Excitement*. Affective words correlated positively with *exciting* ($\beta = 0.21$), *cool* ($\beta = 0.12$), *young* ($\beta = 0.11$), *daring* ($\beta = 0.11$), and *imaginative* ($\beta = 0.08$), while *anger* words correlated negatively with *exciting* ($\beta = -0.08$), *imaginative* ($\beta = 0.08$), *trendy* ($\beta = -0.08$), *spirited* ($\beta = -0.06$), *cool* ($\beta = -0.05$), and *young* ($\beta = -0.05$). In the *Cognitive Processes* category, we observed that words related to *certainty* (e.g. always, never) had a positive correlation with the perception of *unique* ($\beta = 0.10$). This reflects that the use of certainty words is an indicator of improved critical thinking (Carroll 2007).

Social Processes had relatively strong predictive power in predicting *Sincerity*. A great proportional use of *social* words increased a brand's chance to be considered as *cheerful* ($\beta = 0.18$), and *friendly* ($\beta = 0.11$). Similarly, *friends* and *family* words positively correlated the perception of *wholesome* ($\beta = 0.16$), *family-oriented* ($\beta = 0.12$), *sincere* ($\beta = 0.12$), *cheerful* ($\beta = 0.09$), *original* ($\beta = 0.05$), and *friendly* ($\beta = 0.05$).

Biological Processes was observed to be most predictive of *Sophistication*. The frequency of *sexual* words was positively correlated with *charming* ($\beta = 0.06$), while the frequency of *eating* words was negatively correlated with

good-looking ($\beta = -0.18$), *feminine* ($\beta = -0.14$), *glamorous* ($\beta = -0.12$), and *upper-class* ($\beta = -0.10$). One possible explanation is that *eating* words can be associated with body size, and as such they may be related to perception of physical appearance (Toma and Hancock 2012).

Discussion

Implications for Recommender Systems

We foresee many opportunities to apply brand personality modeling in personalized systems. Research in social psychology has shown that material possessions have a profound symbolic significance for their owners, as well as for other people. For instance, understanding symbolic meanings of material objects and making appropriate choices are critical for the creation and maintenance of users' personal and social identities (Dittmar 1992). Brand personality modeling can be used to quantify symbolic meanings of products at scale, and allow recommender systems to consider products as symbolic communication to satisfy users' individual and social needs. Consider red wine, for example, few customers can actually tell the taste differences between brands. Yet, wine brands have different personalities (symbolic meanings), are served in a social setting, can make a powerful statement about those who drink them. In this case, it is vital for a recommender system to understand brand personality, so the system can help users shape their personal images through brand choices. Moreover, future recommender systems could use human and brand personality models together to quantify the associations between user and brand personality, and use these associations to optimize product recommendation services.

Recent job recommender systems have started suggesting jobs and career advices based on users' personality (e.g. http://good.co). Considering the complexity of brand personality could enhance job recommendation services. Research shows that personality congruence between employees and their companies affects employees' attitudes, behaviors, and productivity (Kristof-Brown, Zimmerman and Johnson 2005). People would be most comfortable and successful in companies that share their personalities. Future systems could use brand and human personality together to suggest better fits for both employees and employers. While people seeking jobs can use such a system to find companies that match well with their personalities and interests, hiring managers can also utilize it to screen candidates from the recruiting perspective.

Brand personality model can also be applied to social analytic tools for brand management. In practice, brand managers often have an intended brand personality and devote extensive resources on marketing activities; however, they often fail to ensure consumers perceive the brand

as intended (De Chernatony 1999). An analytic tool could be developed to help brand managers *assess brand personality*, *detect perception gaps*, and *improve perceptions*. 1) *Assess brand personality*. The system could help managers constantly assess the perceived brand personality and monitor the trend of the changes, and thus greatly tighten the feedback loop. 2) *Detect perception gaps*. The system may help to detect the gaps between perceived and intended personality of a brand by summarizing and highlighting the differences in notable personality dimensions. 3) *Improve perceptions*. Our model assigns weights to contributing factors of brand personality. According to these weights, the system could suggest actions from these factors to bridge the perceptions gaps in personality dimensions.

Implications for Brand Personality Modeling

We define the importance of the factors of brand personality as manifested in social media. Our results show that *User Imagery* and *Employee Imagery* factors are equally important in predicting brand personality, while *Official Announcement* has significantly less predictive power. Interestingly, these weights do not always match with previous work (Aaker 2012; Parker 2009), indicating a new paradigm of how people's perceptions of brands are derived from social media. Rather than focusing on the design of marketing campaign messages, future research should investigate how to identify and leverage individual employees' and users' capability in the context of social media.

We observe *User Imagery* are more likely to predict traits in *Sincerity* and *Ruggedness,* while *Employee Imagery* are more likely to estimate traits in *Competence* and *Sophistication.* Prior work shows *Sincerity* and *Ruggedness* are related to *brand affect* (Sung and Kim 2010), an immediate and spontaneous process. In contrast, *Competence* and *Sophistication* are associated with *brand trust*, a well-thought and consideration process (Sung and Kim 2010). One interpretation is that, compared with *User Imagery*, *Employee Imagery* may take a longer process to have influence on the perceptions of a brand. This suggests a new research direction to quantify the latency effects of contributing factions on the perceived brand personality.

Our models offer reasonably accurate predictions of brand personality using three factors: *User Imagery*, *Employee Imagery*, and *Official Announcement*, but additional factors can be included. According to the instrumental-symbolic framework (Lievens and Highhouse 2003), people's perceptions can be associated with two types of attributes: *symbolic* and *instrumental*. The factors used in our model are *symbolic* attributes, and future studies can examine whether the perceived brand personality can be affected by *instrumental* attributes (e.g. product-related attributes), which are related to people's basic needs to maximize benefits and minimize costs. In addition, we only extract word use features from two social media sites Twitter and Glassdoor. Additional data sources (e.g. Facebook, LinkedIn) and features (e.g. social, semantic, and temporal features) can be integrated to improve the performance.

Among the LIWC categories, *Personal Concerns* makes the greatest contribution to brand personality. Personal concerns by nature are individual, personal, and contextual (Pennebaker, et al. 2001). They are related to problems that people are facing, or goals they are trying to achieve during certain stages in their lives. One explanation is that users or employees may express their concerns directly or indirectly related to a brand on social media. These concerns may evoke various types of responses and perceptions, and impact the perceived brand personality.

Limitations

A limitation of the present research is the possible interaction effects between LIWC categories and contributing factors. For example, the effects of certain word use on brand personality may depend on which factor the text is taken from. A future study can be designed to investigate the effects of word use in different factors.

We expect that our findings would not be restricted to one single culture or social platform. The specific coefficients of predictive variable may not allow for a direct prediction for other cultures or platforms, but the factors and their relative importance may be generalizable. It would be interesting to examine how the prediction model can be applied in different cultures and social platforms.

Conclusion

We investigate brand personality from digital traces in social media. Three contributions are made by this research:

- We define the relative importance of *User Imagery*, *Employee Imagery*, and *Official Announcement* factors on the perceived brand personality in the context of social media. Our results show that *User Imagery* and *Employee Imagery* factors are equally important in predicting brand personality, while *Official Announcement* has significantly less predictive power.
- A computational approach to assess brand personality. Using three factors *User Imagery*, *Employee Imagery*, *Official Announcement* together predicts the perceived brand personality with R^2 values as high as 0.67. Among word use categories, words related to *Personal Concerns* makes the greatest contribution in the prediction.
- Implications of the brand personality prediction technique, as well as the relationship between brand personality and its contributing factors on social media. Modeling brand personality opens vast new opportunities for the development of recommender systems in order to help users find brands suiting their personal characteristics and to help companies manage brand perceptions.

References

Aaker, D.A. 2012. Building strong brands. Simon and Schuster.

Aaker, J.L. 1997. Dimensions of brand personality. Journal of Marketing Research 34(3):347-356.

Bollen, J.; Pepe, A; and Mao, H. 2009. Modeling public mood and emotion: twitter sentiment and socio-economic phenomena. In ICWSM.

Buhrmester, M.; Kwang, T.; and Gosling, S.D. 2011. Amazon's mechanical turk: a new source of cheap, yet high-quality, data? Perspectives on Psychological Science 6:3-5.

Carroll, D.W. 2007. Patterns of student writing in a critical thinking course: a quantitative analysis. Assessing Writing 12(3):213-227.

Chen, J.; Hsieh, G.; Mahmud, J.; and Nichols, J. 2014. Understanding individuals' personal values from social media word use. In CSCW.

Cialdini, R.B., and Trost M.R. 1998. Social influence: Social norms, conformity and compliance. The Handbook of Social Psychology 2:151–192.

Cohen, R., and Ruths, D. 2013. Classifying political orientation on twitter: It's not easy!. In ICWSM.

De Chernatony, L. 1999. Brand management through narrowing the gap between brand identity and brand reputation. Journal of Marketing Management 15(1-3):157-179.

De Choudhury, M.; Gamon, M.; Counts, S.; and Horvitz, E. 2013. Predicting depression via social media. In ICWSM.

Dittmar, H. 1992. The social psychology of material possessions: To have is to be. Hemel Hempstead, Harvester Wheatsheaf.

Elliott, R., and Kritsadarat, W. 1998. Brands as symbolic resources for the construction of identity. International Journal of Advertising 17(2):131-144.

Gardner, B.B., and Levy, S.J. 1955. The product and the brand. Harvard Business Review 33(2):33-39.

Gilbert, E., and Karahalios, K. 2009. Predicting tie strength with social media. In CHI.

Golbeck, J.; Robles, C.; and Turner, K. 2011. Predicting personality with social media In CHI Extended Abstracts.

Goldberg, L.R. 1990. An alternative "description of personality": the big-five factor structure. Journal of Personality and Social Psychology 59(6):1216-1229.

Gosling, S.D.; Gaddis, S.; and Vazire, S. 2007. Personality impressions based on facebook profiles. In ICWSM.

Gou, L; Zhou, M.; and Yang, H. 2014. KnowMe and ShareMe: understanding automatically discovered personality traits from social media and user sharing preferences. In CHI.

Grave, E.; Obozinski, G.R.; and Bach, F.R. 2011. Trace lasso: a trace norm regularization for correlated designs. In NIPS.

Harré, R., and Lamb, R. 1986. The dictionary of personality and social psychology. Oxford: Blackwell.

Herring, S.C. 1996. Computer-mediated communication: Linguistic, social, and cross-cultural perspectives. John Benjamins Publishing.

Ipeirotis, P.G. 2010. Demographics of mechanical turk. New York University.

Katz, D. 1960. The functional approach to the study of attitudes. Public Opinion Quarterly 24:163-204.

Kristof-Brown, A.L.; Zimmerman, R.D.; and Johnson, E.C. 2005. Consequences of individuals' fit at work: A meta-analysis of person-Job, person-organization, person-group, and person-supervisor fit. Personnel Psychology 58(2):281-342

Kulshrestha, J.; Zafar, M.B.; Noboa, L.E.; Gummadi, K.P.; and Ghosh, S. 2015. Characterizing information diets of social media users. In ICWSM.

Lievens, F., and Highhouse, S. 2003. The relation of instrumental and symbolic attributes to a company's attractiveness as an employer. Personnel Psychology 56(1):75-102.

Mairesse, F.; Walker, M.A.; Mehl, M.R.; and Moore, R.K. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. Journal of Artificial Intelligence Research 30(1):457-500.

Malär, L.; Nyffenegger, B.; Krohmer, H.; and Hoyer, W.D. 2012. Implementing an intended brand personality: a dyadic perspective. Journal of the Academy of Marketing Science 40(5):728-744.

Markus, H., and Nurius, P. 1986. Possible selves. American Psychologist 41(9):954–969.

Martineau, P. 1958. The personality of the retail store. Harvard Business Review 36:47-55.

Montgomery, D.C., and Peck, E.A. 1982. Introduction to linear regression analysis. New York: John Wiley & Sons.

Parker, B.T. 2009. A comparison of brand personality and brand user-imagery congruence. Journal of Consumer Marketing 26(3):175-184.

Pennebaker, J.W.; Martha E.F.; and Roger J.B. 2001. Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Associates.

Schultz, D.E. 1992. Integrated marketing communications. Journal of Promotion Management 1(1):99-104.

Shami, N.S.; Muller, M.,; Pal, A.; Masli, M.; and Geyer, W. 2015. Inferring employee engagement from social media. In CHI.

Sibona, C. and Choi, J.H. 2012. Factors affecting end-user satisfaction on facebook. In ICWSM.

Sung, Y., and Kim, J. 2010. Effects of brand personality on brand trust and brand affect. Psychology & Marketing 27(7):639-661.

Thomas, K.; Grier, C.; Song, D.; and Paxson, V. 2011. Suspended accounts in retrospect: an analysis of twitter spam. In SIGCOMM.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B (Methodological): 267-288.

Toma, C.L., and Hancock, J.T. 2012. What lies beneath: The linguistic traces of deception in online dating profiles. Journal of Communication 62(1):78-97.

Wentzel, D. 2009. The effect of employee behavior on brand personality impressions and brand attitudes. Journal of the Academy of Marketing Science 37(3):359-374.

Xu, A. and Bailey, B. 2012. What do you think?: a case study of benefit, expectation, and interaction in a large online critique community. In CSCW.

Xu, A., et al. 2012. Learning how to feel again: towards affective workplace presence and communication technologies. In CHI.

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.