Journal of Management Information Systems

ISSN: 0742-1222 (Print) 1557-928X (Online) Journal homepage: https://www.tandfonline.com/loi/mmis20

Generating Business Intelligence Through Social Media Analytics: Measuring Brand Personality with Consumer-, Employee-, and Firm-Generated Content

Yuheng HuYUHENG HU, Anbang XuANBANG XU, Yili HongYILI HONG, David GalDAVID GAL, Vibha SinhaVIBHA SINHA & Rama AkkirajuRAMA AKKIRAJU

To cite this article: Yuheng HuYUHENG HU, Anbang XuANBANG XU, Yili HongYILI HONG, David GalDAVID GAL, Vibha SinhaVIBHA SINHA & Rama AkkirajuRAMA AKKIRAJU (2019) Generating Business Intelligence Through Social Media Analytics: Measuring Brand Personality with Consumer-, Employee-, and Firm-Generated Content, Journal of Management Information Systems, 36:3, 893-930, DOI: 10.1080/07421222.2019.1628908

To link to this article: https://doi.org/10.1080/07421222.2019.1628908

Generating Business Intelligence Through Social Media Analytics: Measuring Brand Personality with Consumer-, Employee-, and Firm-Generated Content

YUHENG HU, ANBANG XU, YILI HONG, DAVID GAL, VIBHA SINHA, AND RAMA AKKIRAJU

YUHENG HU (yuhenghu@uic.edu) is an assistant professor of information and decision sciences at the College of Business Administration, University of Illinois at Chicago. Prior to joining academia, he was a researcher at IBM Research. Dr. Hu's research interests focus on social computing and machine learning. His current and past projects include characterizing the effect of social media sharing for crowdfunding campaigns, and understanding and predicting people's behaviors, sentiments, and engagement with public events and life events.

ANBANG XU (anbangxu@us.ibm.com) is a Manager and Research Staff Member in the USER group at IBM Research - Almaden. He received his Ph.D. in Computer Science from the University of Illinois at Urbana-Champaign. His work combines human-computer interaction and artificial intelligence. His recent focus is on the design of chatbots and personality-analytics systems for user engagement on social media. Dr. Xu has been an associate chair and served on program committees of several top-tier conferences. He is an Associate Editor of *ACM Transactions on Interactive Intelligent Systems*.

YILI HONG (hong@asu.edu; corresponding author) is an associate professor (with tenure) of information systems at the W. P. Carey School of Business, Arizona State University. He received his Ph.D. in information systems from the Fox School of Business at Temple University. His research focuses on online markets and consumer uncertainty. Dr. Hong's papers have appeared or are forthcoming in top academic journals. He is the winner of the 2014 ACM SIGMIS Best Dissertation Award, runner up of the INFORMS ISS Nunamaker-Chen Dissertation Award and 2012 ICIS Best Paper Award, among other distinctions.

DAVID GAL (davidgal@uic.edu) is professor of marketing at the University of Illinois at Chicago. He previously served on the faculty of Northwestern University. His research areas include the psychology of decision-making, identity and behavior, and innovation and creativity. He is currently working on a book, *The Power of the Status Quo*.

VIBHA SINHA (vibhash@us.ibm.com) is a Senior Technical Staff Member at IBM Watson, working on technology that can infer various user characteristics such as personality, needs, emotions, and skills from his/her digital contributions. Before this,

Journal of Management Information Systems / 2019, Vol. 36, No. 3, pp. 893–930.

ISSN 0742–1222 (print) / ISSN 1557–928X (online)
DOI: https://doi.org/10.1080/07421222.2019.1628908

she was with IBM Research India, leading the Productivity Tools group in the Programming Tools and Technologies department. She worked on developing usable research tools that help in automation/simplification of various software development and maintenance tasks and further promote reuse of information across projects. Her research interests include human-computer interaction and software analytics. She holds an M.S degree in Electrical Engineering from Stanford University.

RAMA AKKIRAJU (akkiraj@us.ibm.com) is a Director, Distinguished Engineer, Master Inventor, and IBM Academy Member, at IBM's Watson Division where she leads the AI operations teams and also heads the AI mission of enabling natural, personalized and compassionate conversations between computers and humans. She has worked on agent-based decision support systems, electronic market places, and semantic Web services, for which she led a World-Wide-Web (W3C) standard. She has co-authored 4 book chapters, over 50 technical papers, and has over a dozen issued patents with over 20 pending. She is the recipient of three best paper awards in AI and Operations Research and has received multiple awards and honors at IBM. Ms. Akkiraju has been named by *Forbes* as one of the 'Top 20 Women in AI Research' in 2017. She and her team's work have been featured by various media outlets. She holds a Master's degree in Computer Science and serves as the President for ISSIP, a Service Science professional society.

ABSTRACT: Social media platforms provide an enormous public repository of textual data from which valuable information can be extracted. We show that firms can extract business intelligence from social media data bearing on an important business application, measuring brand personality. Specifically, we develop a text analytics framework that integrates different distinct sources of social media data generated by consumers, employees, and firms, to measure brand personality. Based on Elastic-Net regression analyses of a large corpus of social media data, including self-descriptions of 1,996,214 consumers who followed the sample of brands on social media, 312,400 employee reviews of the brands' firms, and 680,056 brand official tweets, we develop a brand personality model that achieves prediction accuracy as high as 0.78. Among key insights, we find that the profile of individuals who choose to associate with brands on social media is an important predictor of brand personality; this provides the first real-world evidence for a consumer identity-brand personality link. We also identify a link between an organization's internal corporate environment as perceived by employees and brand personality as judged by consumers. We further illuminate the practical implication of our predictive model by building a cloud-based information system that allows managers and analysts to explore and track personality of their own brands and their competitors' brands.

KEY WORDS AND PHRASES: social media analytics, business intelligence, consumer-generated content, employee-generated content, firm-generated content, brand personality.

Introduction

Humans tend to attribute human personality traits to brands [3]. Through their personality, brands are thought to satisfy people's self-expression and social needs [1, 4, 23, 72, 73]. Individuals can use the brands they purchase or are associated with 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 they are to members of their social groups. Moreover, research assumes a close relationship between human and brand personality [66] and suggests that the relationship impacts individuals' preferences, satisfaction, and their social interactions with others [18, 27, 107].

Recognizing the potential importance of brand personality to business strategy, researchers have devoted significant attention to modeling brand traits for a number of applications, including the development of recommendation systems and perceptual maps (e.g., Hauser and Koppelman [43], Lehmann et al. [62], John et al. [51], and Nunes [80]). Researchers typically assess brand personality through consumer surveys that can include both compositional and decompositional methods. In the former, consumers are asked to rate their perception of brands on specific traits whereas in the latter consumers sort brands into categories from which traits are inferred [48, 94]. However, the prevailing existing methods employed by market research firms are costly, slow, static, and the results might be quickly outdated, making the application of brand personality analysis in perceptual mapping, in tracking of brand trends over time, and in many other domains impractical (e.g., Steenkamp and Van Trijp [95] and Culotta and Cutler [20]).

Recent years have witnessed a widespread increase in the quantity of online content offering the promise of more automated methods that extract informational value for businesses. And researchers from multiple disciplines have expended research effort in using text analytics methods to understand unstructured data, such as in Information Systems (e.g., Fan and Gordon [29], Chen et al. [16], and Wixom et al. [108]), Computer Science (e.g., Cohen and Ruths [19] and De Choudhury et al. [22]), and Marketing (e.g., Fader and Winer [28], Lee and Bradlow [61], and Markus and Kunda [20]). Meanwhile, much research attention in this space has focused on predicting human personality from online user-generated content [6, 17, 37] and developing corresponding personalized recommender systems [40]. However, until now, a minimal amount of attention has been paid to the use of various sources of unstructured secondary observational data to model brand personality.

Paralleling efforts to extract human personality from unstructured textual data, this paper aims to extract brand personality from user-generated (consumers and employees) and firm-generated social media content and to investigate what types of content have predictive power [36]. Our model is built on a large corpus of social media data gathered from three independent sources, each representing a different factor of information about brands. In particular, we collect data on the brand's own communications through collecting 680,056 brand official tweets from Twitter (termed *Official Announcements*), data on internal organizational environment through collecting 312,400 employee reviews of the organization (termed *Employee Imagery*) from Glassdoor (glassdoor.com), and data on brand-users through collecting the self-descriptions of 1,996,214 users who choose to associate with brands through following them on Twitter (termed *User Imagery*). Further, we obtain 10,950 survey responses to measure 219 brands' perceived personality as the "ground truth" measure of brand personality.

We built a predictive model of brand personality based on textual analysis through extracting word features from the three above-mentioned content sources. The analysis of text through word feature extraction has become a popular method of assessing individual psychology [6]. Research has shown that the use of specific words by individuals is statistically associated with psychologically important variables, such as individuals' emotions and personality [26, 100]. We extend this approach of measuring human personality [6] to the current study of brand personality.

Using this approach, our results show that features extracted from consumer-generated content (User Imagery) and employee-generated content (Employee Imagery) are the most important factors in predicting brand personality, whereas features extracted from firm-generated content (Official Announcements) have significantly less predictive power. With the factors combined, our brand personality model achieves prediction accuracy as high as 78.5 percent.

This paper contributes to the IS literatures on social media and text analytics. Beyond the abundant IS literature on social media [47, 86, 91, 98], prior research has primarily focused on developing predictive models for human personality from user-generated content [6], this study takes an initiative in using text analytics to model firm-level constructs, that is, brand personality. Specifically, we take advantage of a multitude of archival sources of social media data, to build a model and a system for text analytics at the firm level. This paper also contributes substantively to the literature on measuring brand personality. The importance of *User Imagery* in predicting brand personality is particularly significant as it provides the first real-world evidence linking consumer identity to brand personality. This is significant because the perceived importance of brand personality stems from the fact that it is believed to influence consumer preferences and behaviors through allowing consumers to express either their actual identity or an identity they wish to project [69]. In other words, the establishment of a link between consumer identity and brand in a real-world context is important because it is the very basis for studying brand personality in the first place. The importance of *Employee Imagery* in predicting brand personality is also significant because it suggests that brand personality does not arise *ex nihilo* but is related to the corporate environment in which the brand originates.

To further illuminate the practical implication of our predictive model, we build an information system that allows managers and analysts to explore and track brand personality of their own brands and their competitors' brands. The back end of our system handles data collection, storage, and large-scale computation using a big data computation platform (Spark), NoSQL database technology (MongoDB), and various programming languages (Python, Scala). The front end of the system is hosted on IBM's Cloud Platform and provides users an easy-to-use web interface. Similar to recent IS research that has modeled business proximity with archival data for competitive industry intelligence [90], our model and system provide actionable insights for businesses to gain insights into their own and their competitors' brands, which helps prescribe business strategies.

In this paper, we first provide a theoretical overview on brand personality and social media analytics. We then describe the data that we collected and our text analytics framework of brand personality measurement. We then describe the integrated cloud-based system we built based on the prediction model. We conclude by discussing implications of our findings.

Background

Brand Personality

The term brand personality, first introduced by Martineau in 1958, refers to a set of human characteristics associated with a brand [84]. For example, the Apple brand is perceived to be young, while the IBM brand is perceived to be old. Since its conceptualization, brand personality has become widely accepted as an effective way to capture users' perceptions of brands that are assumed to reflect users' self and social identities, which in turn affect their preferences [18, 23, 27, 72]. Brand is an important predictor for perceived website quality [64]. and brand personality is also presumed to be an important determinant of brand equity [54]. When competitors can easily copy product characteristics, a strong brand personality is viewed as increasingly valuable to building brand equity and, thus, is viewed as having crucial business value [103].

Existing Brand Personality Scales and Their Limitations

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 [32], but these open-ended techniques were gradually abandoned as researchers looked for more quantitative ways to detect and enumerate differences among brands. Later, researchers attempted to use human personality scales developed in psychology to directly measure brand personality [38].

In 1997, Aaker developed *brand personality scales* [3]. She analyzed the individual ratings of 37 brands on 114 personality traits by 613 respondents recruited in the United States. With this approach, *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 as *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 scale for assessing brand personality [3, 4]. Since Aaker's article, most of the extant literature has adopted Likert scale surveys based on Aaker's scale to assess brand personality and new scale development broadly follows methods based on those used by Aaker [3, 30, 7].

Table 1. Five Dimensions of Brand Personality, Each Comprised of Several Personality Traits

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 and 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	

(continues)

Table 1. Continued

Dimension	Trait	Top 5 rated brands	Mean	STD	Distribution
	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 and 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 and 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	

(continues)

Table 1. Continued

Dimension	Trait	Top 5 rated brands	Mean	STD	Distribution
Sophistication	**upper-class**	*Mercedes-Benz, Louis Vuitton, Tiffany and Co., BMW, Bloomingdale's*	4.07	1.01	
	glamorous	*Tiffany and Co., Mercedes-Benz, Victoria's Secret, Sephora, Louis Vuitton*	3.56	1.01	
	good-looking	*Victoria's Secret, BMW, Sephora, Burberry, Tiffany and Co.*	4.33	0.88	
	charming	*Walt Disney, Tiffany and 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	

Note: Middle column shows top-rated brands in our study. Right columns show the descriptive statistics and rating distribution of the trait.

However, the inherent limitation of survey-based approaches is their lack of flexibility and scalability [45]. Conducting surveys is often time-consuming, labor-intensive, and expensive. This means that brand personality surveys cannot scale due it the costs and their results can become quickly outdated. Thus, a main motivation of this work is the development of an automated text-analytic prediction framework to effectively and efficiently measure brand personality with massive, widely-available social media data.

Social Media for Brand Analytics

The surge of social media usage by both consumers and firms offers a promising data source to understand consumer behaviors [87, 96, 99, 115], which can then be used to inform firm strategy, and to create business value [65, 109]. Social media analytics has seen applications in many business functions, and research on social media analytics represents a growing stream of literature in the information systems field [12, 24, 42, 67, 81]. For example, recent research in this domain includes examination of the impact of user-generated content of a brand's Twitter followers on brand image [78, 58], and product sales [36]. Other research examines the social structure of a brand's Twitter follower base [91], regardless of whether the base creates or consumes content, for estimating brand perceptions [20, 35). Despite the advantages of using social media for analytics, the noise, volume, and ambiguity of such data sources pose substantial challenges for the development of robust and effective algorithmic solutions [114]. It is therefore no surprise that, to date, a minimal amount of research has examined brand personality using social media data. In this paper, we provide a first attempt to develop automated solutions using online social media content for predicting and measuring brand personality.

Theoretical Overview

Consumer Identity and Brand Personality

The underlying rationale for developing and investigating the brand personality construct is the belief that consumers express themselves symbolically through brands [52, 93, 105]. Through making sense of brand personality, researchers and firms can better understand consumers' symbolic use of brands. Notwithstanding this view, academic researchers have had difficulty demonstrating associations between the traits of individual consumers and the brand personality of their preferred brands. For example, research performed in the 1960s and 1970s, theorizing that brands tend to be used for self-expression, hypothesized that consumers would prefer self-congruent brands; that is, brands with personalities similar to their own. However, a minimal amount of empirical evidence supports this hypothesis (for reviews, see Kassarjian [52] and Sirgy [93]).

Scholars put forth several explanations to account for this perceived anomaly. One was that these early studies were based on the false logic that the self-concept is stable

[52, 93]. Subsequent research provided evidence for a malleable self-concept and showed that different facets of the self were more or less salient depending on the context (e.g., [72]). For example, one's identity as an athlete and, therefore, one's perception of oneself as physically strong is likely to be more in focus when playing sports; whereas, one's identity as a scholar and, therefore, one's perception of oneself as intellectual, is likely to be more salient when writing an academic article.

Another explanation put forth to explain early failures to identify links between individual personality and brand personality was that individuals ascribe different degrees of importance to different facets of their personality [71, 104]. This suggests that simply possessing a personality trait does not imply that the individual will be motivated to express it. For the individual to be motivated to express a personality trait, the trait would have to be important to the individual's identity.

A third explanation that researchers considered for the lack of association between individual personality traits and brand preference was that the brand personality scales used in early research were adapted from human personality scales and that they otherwise lacked sufficient validation in the brand context [4]. A final explanation considered for the lack of association between individual personality traits and the brand personality of preferred brands was that brand preference is not merely a form of self-expression but that it can fulfill other needs such as the need to belong to a particular group or to present an aspirational version of oneself (e.g., Kassarjian [52] and Gal [31]).

By addressing the limitations of earlier research that attempted to demonstrate an association between human personality and brand personality, [4] demonstrated that individuals preferred brands with personalities that were congruent with active aspects of their self-concept. Aaker's investigation differed from prior work in using the validated brand personality typology she had developed [3] instead of using ad hoc or human personality scales to categorize brands. In addition, when examining whether consumer traits predicted brand preference, Aaker focused on consumer personality traits that were either chronically accessible or that were made temporarily accessible through activation by situational cues [8, 88].

Nonetheless, it must be stated that the existing evidence suggests that, overall in the literature, the link between consumer identity and brand personality is relatively fragile. Moreover, such a link has never been demonstrated in the real-world observational setting, which tends to be much noisier than a controlled lab context. This is important as the brand personality construct is considered important primarily on account of its theorized ability to illuminate consumers' symbolic use of brands [3] Without real-world evidence of a link between consumer identity and brand personality, the raison d'être of the brand personality construct can be called into question.

Factors Driving Brand Personality

Individuals' perceptions of brand personality can be conceptualized as a host of associations they have about the brand that are assumed to form through either direct or indirect contact with the brand [85]. One factor central in forming individuals' perceptions of brand personality is the profile of the people that are associated with the brand, most notably its users. For example, an individual's perception of the Apple brand is associated with users of Apple products. Therefore, in this context, we formally define *User Imagery* as "the set of human traits associated with the typical user of a brand" [p. 1, 3]. The traits of these users can be transferred to the brand [1, 2, 75], and the link between user traits and brand personality is bidirectional such that users are also attracted to brands that reflect their actual or desired identity [3].

A second factor that is important in generating brand personality is the internal organizational environment, such as the company culture, of the firm that owns the brand. Although, logically, it is not essential that the personality of a brand (as perceived by consumers) be related to the internal environment of the organization, such as the organization's culture, it is frequently found that the brand is embodied by the organization [10]. In this vein, it is assumed that the organization's culture is reflected in the brand's conduct and through the marketing mix [55, 97].

A third key factor that we propose to influence perceptions of brand personality is how a brand projects its identity through its explicit communication strategy [2]. This factor is perhaps the most intuitive driver of brand personality and the one of which the organization has the tightest and most direct control.

In the present research, we attempt to capture these influences through three main factors that we operationalize through three different data sources. The first factor is *User Imagery*, which reflects the traits of individuals who choose to associate with a brand, and which we derive from the profiles of individuals who choose to follow brands on Twitter. The second is *Employee Imagery*, which reflects how employees perceive the internal corporate environment of the brand, and which we derive from employee reviews of the brands on Glassdoor. The third is *Official Announcements*, which reflects the messaging put out by the brand, and which we derive from the brand's official announcements on Twitter.

We extracted word use features for a brand from unstructured text data from three sources of information about brands available through social media for measuring these factors, namely (a) user-generated content, in the form of profiles and tweets of consumers who follow brands on Twitter (*User Imagery*); (b) employee-generated content, in the form of employee reviews on Glassdoor (*Employee Imagery*), and (c) firm-generated content, in the form of official brand tweets (*Official Announcements*). Prior research has shown that the use of particular words by individuals is statistically associated with psychologically important features, such as where individuals are directing their attention, their emotions, their social relationships, their personal traits, their values, and their thinking styles [100]. Although, to date, such textual analysis has been focused on examining the psychology of individuals and predicting their behaviors [46, 47, 112], in the present research we extend this method to brands.

Research Method: Data-Analytic Framework

In this section, we document the research methods and detailed procedures. We first obtained the "ground truth" of perceived brand personality (i.e., our baseline measure of brand personality) by conducting a survey on 219 brands. Next, we collected secondary data from social media and constructed a prediction model based on this data to model these brands' personality.

Ground Truth Collection

As a standard approach, we conducted a large-scale survey to acquire ground truth for brand personality modeling. The survey procedure is described in the following section.

Brand Selection

Two criteria guided our choice of brands. First, well-known brands were selected so that a national sample could be used to gather survey data. Specifically, we first collect brands that were listed among the top 1,000 Fortune companies ranked by gross revenue (http://www.geolounge.com/fortune-1000-companies-2014-list). These firms 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 [53]. Our initial set of brands cover a wide variety of categories such as restaurants, clothing, automobiles, electronics, and financial services. We randomly chose up to 10 brands from each category. Finally, 219 brands were selected (see examples in Table 1).

Participants Recruitment

Amazon Mechanical Turk (MTurk) has seen an increasing usage in IS research [13, 15]. In this research, we used MTurk to recruit participants primarily because MTurk reaches a more diverse population than traditional student samples and community samples [11], allowing researchers to gain generalizability to broader populations. For instance, MTurk workers have a similar income distribution compared with the general U.S. population [49].

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. Therefore, we only recruited MTurk participants who are identified as living in the United States.

We take multiple strategies to ensure the quality of the survey. First, we set up a pre-task screening exam to avoid bots. We ask simple questions like "What is 12-8?" as well as adding reCAPTCHA.[1] Next, we select workers whose approval rating (the rate that previous task requesters have approved tasks that the worker has completed) is at least 99 percent. According to the prior research, higher approval rating can necessarily ensure high-quality output from the MTurk participants [82].

All 3,060 participants were 18 or older. 16.5 percent of participants at age 20–24, 18.5 percent at age 25–29, 16.7 percent at age 30–34, 12.7 percent at age 35–39, 10.8 percent at age 40–44, 8.0 percent at age 45–49, 7.1 percent at age 50–54, 4.4 percent at age 55–59, and 1.9 percent at age 65 or older. 66.0 percent of participants were female. Although not uniformly distributed, each group was well represented in the sample. Each participant was paid $US 0.05 per survey.

In the study, most participants (97.0 percent) were familiar with brands shown in the survey. Familiarity ratings were obtained by having participants rate a brand on a 7-point Likert-type scale ranging from *extremely unfamiliar* (1) to *extremely familiar* (7). Following prior work [69, 70], 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.

Survey Questionnaire

Each participant responded to an online standardized questionnaire with regard to their perceptions of one brand [3]. The participant rated how descriptive the 42 traits were of the brand in general, using a 7-point Likert-type scale anchored at *not at all descriptive* (1) to *extremely descriptive* (7). The traits were arranged in random order to control for 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. The "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. A total of 27 percent of participants failed the sanity checks and their responses were removed. We kept posting survey tasks on MTurk until a brand had 50 valid responses.

Similar to prior work [3], 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. The questionnaires were completed within a few days. Fifty survey responses were collected for each brand. We obtained 10,950 valid responses on 219 brands from 3,060 participants. On average, participants answered 3.6 questionnaires.

Brand Personality Scales Analysis

The average of participants' ratings of a brand was used to measure the brand's personality [3, 70]. 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 seven-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 [3]. 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, the statistical difference indicates that Apple was perceived to be much younger than IBM. Overall, these results support the validity of the ground truth of brand personality we obtained.

Social Media Data Collection

We collected social media data from three publicly-available archival sources. Specifically, we collect 1,996,214 brand followers' descriptions and tweets as *user imagery*, 312,400 employee reviews from Glassdoor as *employee imagery*, and 680,056 brand tweets as *official announcements*). We present more details in the following sections.

User Imagery

As mentioned earlier, user imagery refers to the set of human personality traits associated with the typical user of a brand. Numerous research efforts have been focused on using content generated by individuals to model human personality [6, 39], emotion [9, 46], values, satisfaction [92], engagement [89, 110], chronic conditions [63], political orientation [19], and dietary choices [59]. Specifically, an individual's personality can be estimated based on his or her textual data such as essays [68] and online posts [37, 40, 111]. For instance, Yarkoni modeled individuals' personality using their blogs [111], and similarly, [6] modeled individuals' personality using their microblogs. The model was built on word use features extracted from the microblog content.

Inspired by these works, we considered the profiles of 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,[2] which is a short self-description in a follower's public profile, as well as the most recent two months of tweets. In total, we collected 1,996,214 followers' profile by querying the Twitter API, and the average description length was 12.1 words.

Employee Imagery

On the other hand, *employee imagery* refers to how employees perceive the internal corporate environment of the brand. To build this factor, we consider information on Glassdoor.com which 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. The brands, on average, had about 1,400 employee reviews. The average review length was 86.85 words.

Official Announcement

Finally, *Official Announcement* refers to the messaging put out by the brand. 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.

Modeling Techniques and Performance Measures

Brand personality scales have 42 traits (see Table 1); each trait was regarded as a binary outcome variable. Initial models were developed separately for each data source, namely for User Imagery, Employee Imagery, and Official Announcements. The purpose of this was to identify and compare the performance of each of these sources both in predicting brand personality overall and in predicting individual brand personality traits. We present the model performance with the combined data sources next.

Features

We treat prediction of the binary outcome as a binary classification task. One popular way to address this is to use supervised learning approaches. To construct such an supervised model, we need to first extract features that materialize User Imagery, Employee Imagery, and Official Announcements factors. We consider several features in the following list.

a. *Linguistic feature*: The first set of features is about linguistics of User Imagery, Employee Imagery, and Official Announcements. We used regular expressions to match the same features from *Linguistic Inquiry and Word Count* (LIWC) to characterize each source. LIWC, a dictionary developed in the psycholinguistic field, has been widely used in psychology (e.g., Pennebaker et al. [83]), information systems (e.g., Hong et al. [46], Huang et al. [47], and Yin et al. [112]), and social computing research (e.g., Chen et al. [17]) to quantify the linguistic and psychological features of a text document by counting words in psychologically meaningful categories such as psychological processes like social processes, cognitive processes, pronouns, etc. In social media, sometimes the words are with in-regular format. Thus, we apply a standard dictionary-based text normalization technique.[3] 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 7 descriptors of the distribution of documents over each LIWC feature: mean, 5th to 95th percentile, variance, skew, kurtosis, minimum, and maximum. Thus, the total number of LIWC features is 420.

b. *Bag of words feature*: Next, we construct the bag of words (BOW) vectors for representing User Imagery, Employee Imagery, and Official Announcements. BOW features are among the most fundamental features using in text mining and analytics [56]. After applying tokenization, removal of stop words, stemming, we measure 2-grams and measure their weights by term frequency-inverse document frequency (TF-IDF).
c. *Topical feature*: Our third type of features consists of topical features. We apply a popular topic model LDA to distill topic distribution from the text resources of User Imagery, Employee Imagery, and Official Announcements. In practice, we set the number of topics K = 20 for each factor.

Model

Our proposed brand personality model employs the Elastic-Net regularization-based approach [116]. We choose the Elastic-Net because the number of predictors (e.g., 420 LIWC predictors plus 120 topic predictors plus features from 2-grams) greatly exceeds the number of observations (219 brands). This is known as large p small n problem in statistics and can cause a high collinearity between predictors in our data and insufficient degrees of freedom to estimate the full model [106]. In this case, the Elastic-Net can outperform many other predictive approaches while seeking for a sparse solution by shrinking the coefficients of weak and/or correlated predictors [25]. As a result, the Elastic-Net regression can select a set of best explanatory predictors whereas other methods cannot [116, 50]. Besides, since the elastic-net imposes on a combination of Lasso (least absolute shrinkage and selection operator) and ridge penalties, it provides more reproducible prediction than using multiple regressions [101, 60].

Specifically, we use the following process to train, test, and validate our model:

- We used 10-fold cross-validation (initial cross-validation) to repeatedly split the data into training and testing sets by the standard 80 percent training-20 percent test.
- For each split, our model first uses Elastic-Net to perform another 10-fold cross-validation on the training set to determine the optimal values for lambda (the

shrinkage parameter in Elastic-Net for model selection). Then, the Elastic-Net 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 Elastic-Net 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.

Prediction Performance

Metrics

The model's prediction performance was measured by the predicted R^2, 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 the out-of-sample observations not included in model estimation [76]. 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 outcome variable. As these 42 dependent variables correspond to five dimensions of brand personality, we average the predicted R^2 of the corresponding traits for each of these dimensions.

To examine the performance of our proposed model, we compare it with several baselines: a) Logistic Regression, trained with the same set of features used in our model, b) SVM, trained with the same set of features used in our model, and c) K-nearest neighborhood, using the same set of features. We also consider examining the individual and joint impact of User Imagery, Employee Imagery, and Official Announcements.

Brand Personality Prediction

We first examine the prediction performance on each individual brand personality traits. The results are shown in Figure 1, which are based on combining features from all three *User Imagery, Employee Imagery*, and *Official Announcements* together. A majority (71 percent) of the traits' R^2 values were higher than 0.5, and 31 percent of the traits' predicted R^2 values were above 0.6. The *technical* trait had the highest R^2 value (0.78), followed closely by *feminine* (0.75), *young* (0.70), *cheerful* (0.66), *cool* (0.66), *good-looking* (0.65), *trendy* (0.65), *charming* (0.64), *intelligent* (0.62), *small-town* (0.62), *smooth* (0.60), *glamorous* (0.60), and *daring* (0.60), among others. The R^2 values of some traits (e.g., *Western* and *honest)* were relatively low ($R^2_{Western}$ = 0.28;

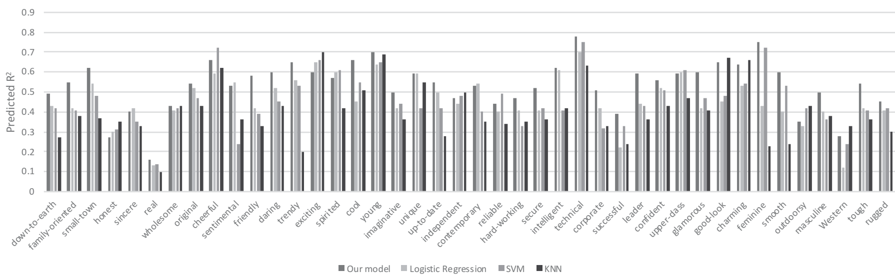

Figure 1. The comparison of the predicted R2 values with the contributing factors: user imagery, employee imagery, and official announcement combined for predicting 42 personality traits.

R^2_{honest} = 0.27; R^2_{real} = 0.16). The relatively low predicted R^2 value suggests that predictive variables in the current model only account for a small proportion of variability for the prediction of these traits. Compare to other baselines, our model outperforms them in most traits. The range of performance improvement margin is from 12.5 percent to over 50.5 percent.

Next, we examine how our proposed model compares with other baselines in predicting five brand personality dimensions. The results are shown in Figure 2. Note that we average over the corresponding traits for each dimension. There were no significant differences in predicted R^2 values among the five dimensions ($F(4, 37) = 2.12$, $p = .098$). Similar to the prediction performance of individual traits, the prediction performance of our model was consistent and reasonably accurate across the five brand personality dimensions. In particular, our model achieves the predicted R^2 as high as 0.643 and a mean absolute error (MAE) as low as 0.0767 on a continuous 0–1 scale. The range of predicted R2 values was from 0.426 (MAE = 0.1315) to 0. 643 (MAE = 0. 0767) with an average value of 0.54 (MAE = 0.0844). It is also clear that our model consistently outperforms other baselines in predicting all brand personality dimensions with an improvement margin ranging from 23 percent to 45 percent. This resolution compares

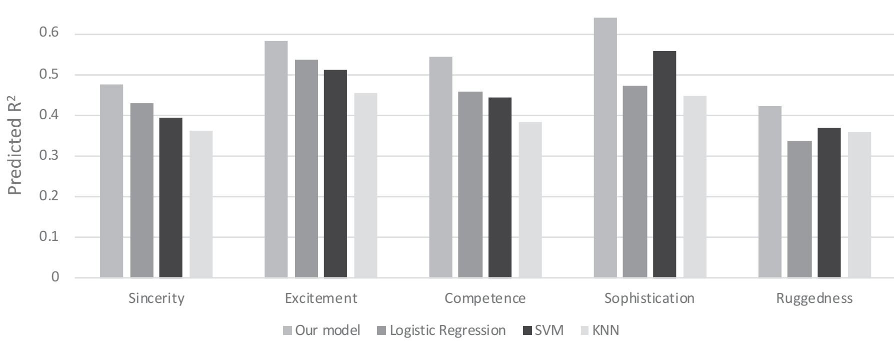

Figure 2. The comparison of the predicted R2 values with the contributing factors: user imagery, employee imagery, and official announcement combined for predicting 5 brand personality dimensions.

favorably to predictions of human personality from individual writing samples and is acceptable for most applications [17, 40].

Effect of Individual Factors

Next, we look into the effect of three data sources, *User Imagery, Employee Imagery*, and *Official Announcements*. The results are shown in Figure 3. Predicted R^2 values by *Employee Imagery* were significantly higher than predicted R^2 values by *User Imagery* in *Competence* and *Sophistication* dimensions ($p < .05$). 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* ($p < .05$). As the results show, variables from *User Imagery* and *Employee Imagery* impacted prediction performance very differently in different personality dimensions. Additionally, the *Official Announcements* variables had significantly lower R^2 values than other factors across all the brand personality dimensions ($p < .05$).

We next analyze the relative importance of the three factors. After the Elastic-Net regression fitted the data, the β coefficients associated with 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.[4] Specifically, we first computed a factor's weight for each brand personality trait individually, and then averaged over the traits. Figure 4 shows that *User Imagery* and *Employee Imagery* had a dominant influence in predicting brand personality, while *Official* Announcements had the least important influence. Furthermore, one-way analysis of variance (ANOVA) revealed significant differences in predicted R^2 values among the three data sources ($F(2, 123) = 22.88$, $p < .001$). Post-hoc tests between paired factors showed that R^2 values

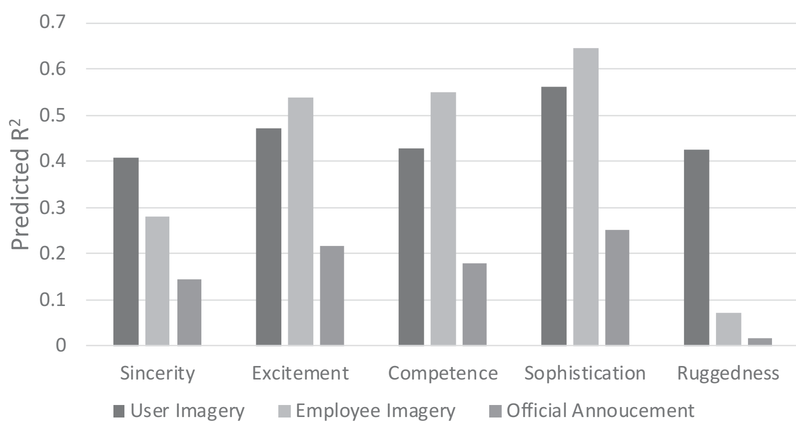

Figure 3. The comparison of the predicted r2 values among the contributing factors: user imagery, employee imagery, and official announcement. Each factor was used to predict brand personality individually.

predicted from the *Official Announcements* variables were significantly lower than those predicted from the User Imagery and Employee Imagery variables ($p < .01$). This indicates that *Official Announcements* has significantly less predictive power than User Imagery and Employee Imagery variables.[5]

Varying Parameter

An interesting and important parameter in Elastic-Net regularization is α, which determines the balance between L2-norm (α = 0) and L1-norm (α = 1) regularization [116]. The α is close to 0 if many predictor variables have small or medium predictive power, while α is close to 1 if only several predictors have large power. To explore the effect of α on the model prediction performance, we varied 100 different values of α, ranging from 0 to 1 with a step of 0.01. For each value, we ran Elastic-Net regression with 10-fold cross-validation. Figure 5 shows that the model achieved the highest accuracy when α was equal to 0.01. This indicates that brand personality is likely to be derived from a set of mixed types of non-dominant features in different factors.

Effect of Different Features

Since our model uses 3 categories of features, namely LIWC features, POW features, and topic features. It is interesting to examine their effect in predicting 5 brand personality dimensions. The results are shown in Figure 6. There are several observations. First, we find that the prediction power of each individual feature varies. Overall, LIWC features perform the best except in predicting the competence dimension. Conversely, the combination of the features often achieves the better prediction results. In particular, combining all 3 set of features achieves the

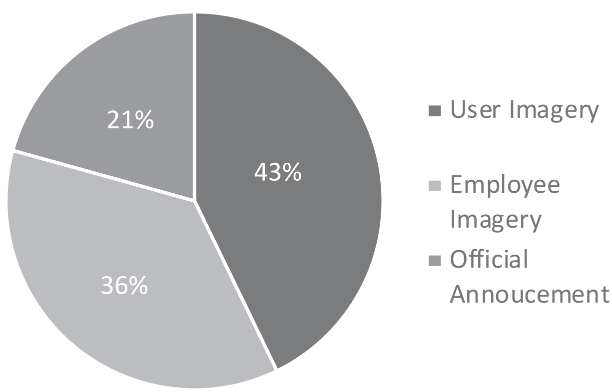

Figure 4. The predictive power of three contributing factors: user imagery, employee imagery, and official announcement. *Note*: The factor weight was computed by summing the absolute values of the coefficients belonging to it.

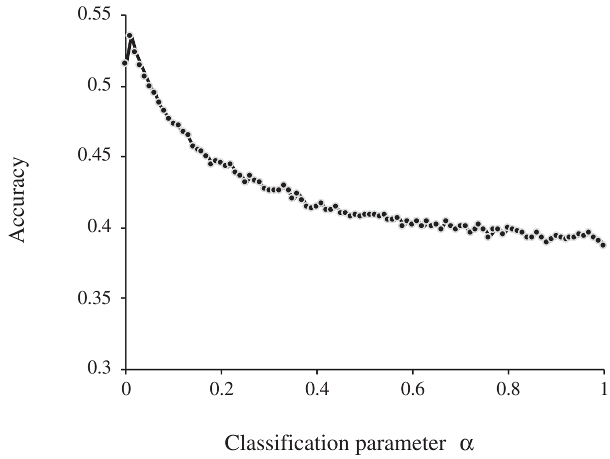

Figure 5. Prediction performance changes according to the value of alpha in elastic-net regularization.

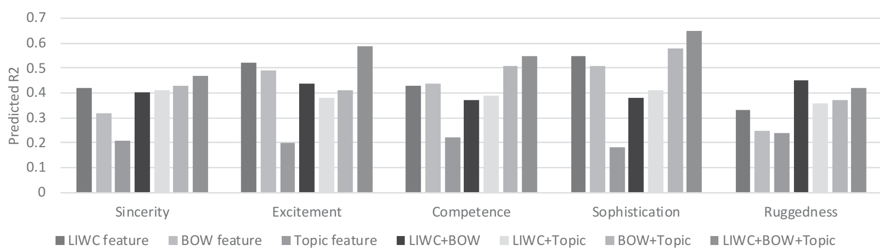

Figure 6. Prediction performance with respect to different set of features.

best performance with an average predicted R^2 around 0.536 and an average improvement over other features and their combinations of 36.4 percent.

Analysis of Word Use

We now examine the influence of various LIWC categories on the predictive model (see Figure 7 and Table 2). The weight of a category was computed by summing the standardized coefficients of the predictive variables belonging to it [34]. Specifically, for each individual brand personality trait, a category's weight was calculated by summing the corresponding coefficients across the three sources. 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 and 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

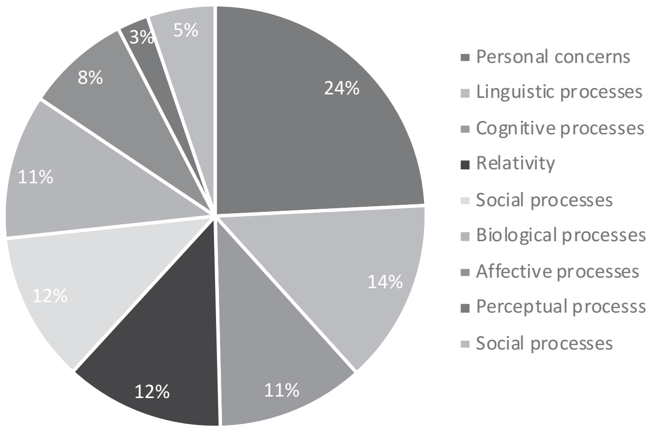

Figure 7. Prediction performance with respect to LIWC categories.

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 and 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, the results indicate that 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 and 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 were positively correlated with *young* (β = 0.03). Many *prepositions* and fewer *first and second person pronouns* are often found in official documents and academic writings [44]. This result indicates that a formal language style could enhance the perception of *Competence* and *Sophistication*.

We found that *Affective Processes* and *Cognitive Processes* exerted a significant influence on the model for *Excitement*. Affective words correlated positively with *exciting* (β = 0.21), *cool* (β = 0.12), *young* (β = 0.11), *daring* (β = 0.11), and *imaginative* (β = 0.08), while *anger* words correlated negatively with *exciting* (β = −0.08), *imaginative* (β = 0.08), *trendy* (β = −0.08), *spirited* (β = −0.06), *cool* (β = −0.05), and *young* (β = −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* (β = 0.10). This reflects that the use of certainty words is an indicator of improved critical thinking [14].

Table 2. Prediction Accuracy Calculated for Each Personality Trait of LIWC feature

	Trait	Accuracy	User Imagery	Employee Imagery	Announcement
Sincerity	**down-to-earth**	0.49	money_v, number_m, article_s, metaphysical_q, future_k, grooming_v, human_s, preposition_v, hear_s	tentative_q, music_m, money_q, religion_v, ositive_affect_k	assent_k, tv_v, past_m, discrepancy_q, tv_q, preposition_m
	family-oriented	0.55	unmatched_s, grooming_v, money_q, article_m, 1stperson_s, past_s, unmatched_v, sad_k, money_v, cognition_m, causation_m	money_m, family_q, preposition_m, assent_m, family_v, certainty_q, past_v, sexual_m, inhibition_m, article_q, symptom_s, death_m	NA
	small-town	0.62	inclusion_m, metaphysical_m, time_s, motion_q, number_m, achievement_q, symptom_v, sad_k, job_k, 1stperson_plural_s, inhibition_k,	affect_m, tentative_q, 1stperson_v, unmatched_s, touching_m, space_k, inclusion_k, certainty_q, sexual_v, family_v, article_m, death_v, down_q,	preposition_q, touching_q, tv_q
	honest	0.27	commun_q, hear_s, money_v, see_q	NA	tv_v, assent_k, anger_q, past_m
	sincere	0.40	unmatched_v, 1stperson_plural_v, future_k, human_s, commun_s, discrepancy_k, anxiety_k	music_m, certainty_q, commun_s, family_q, money_v, exclusion_k, negative_affect_s, filler_q, preposition_m, inhibition_m, ref_other_k	tv_v, 1stperson_plural_q, money_m, assent_k, sexual_s, achievement_k, anger_q, hear_k, 3rdperson_q,

(continues)

Table 2. Continued

Trait	Accuracy	User Imagery	Employee Imagery	Announcement
real	0.16	commun_q, optimism_k, inclusion_m, sad_s, 1stperson_plural_v, human_s, future_k, past_s, down_m, unmatched_q, negation_s, inhibition_m, ositive_affect_k	NA	NA
wholesome	0.43	money_v, friend_m	family_v, certainty_q	space_m, past_m, achievement_k, family_s
original	0.54	money_v, tv_v, religion_k, see_q, future_k, commun_q, sleep_k	causation_m, money_q, present_m, human_m, 2ndperson_s, article_k, sexual_m, certainty_q, negative_affect_k	pronoun_k, tv_k, optimism_q, certainty_q, sleep_m, family_q, inclusion_v, time_k
cheerful	0.66	social_m, money_v, ref_other_m, friend_m, article_m, motion_m	money_q, sexual_q, friend_m, music_q	tv_q, occupation_k, job_q
sentimental	0.53	money_v, down_m, see_m, article_m, hear_k, see_k	present_m, family_v, certainty_q, article_q	NA
friendly	0.58	money_v, social_m, see_q, sad_k, article_m, number_m, home_q, past_s, nonfluency_k	money_q, home_q, human_k, music_q, tentative_q, ref_other_k, money_k, family_q, religion_v, past_v, death_m	tv_v, inhibition_m, achievement_k, preposition_m, discrepancy_q, past_m

Excitement	**daring**	0.60	commun_q, social_k, motion_s, filler_q swear_k, money_q, causation_q, exclusion_k	sport_q, eat_q, affect_m, death_v, symptom_k, family_m, touching_m, human_s, assent_m, see_k	eat_s, filler_k, affect_k, certainty_k, occupation_k, optimism_q
	trendy	0.65	2ndperson_m, money_v, friend_v, anger_q, sleep_v, causation_q, past_k	death_m, sexual_v, touching_m, down_m, filler_v, unmatched_s	affect_k, number_k, certainty_q
	exciting	0.60	money_q, inclusion_v, causation_q, money_s, commun_q, motion_k, friend_v, anger_q, nonfluency_k	affect_m, eat_q, sexual_q, friend_v, negative_affect_k, sensation_q, future_q, exclusion_k, present_q, positive_feelings_k, 2ndperson_s	certainty_q, 2ndperson_v, occupation_k, religion_m, optimism_q, down_q, sleep_m, sport_m, certainty_k
	spirited	0.57	money_v, sensation_k, anger_q, inclusion_v	death_m, 1stperson_plural_v, sexual_v, eat_q, money_m, ositive_affect_q	family_m, eat_s
	cool	0.66	money_v, religion_v, friend_v, causation_q, anger_q	affect_m, eat_q, metaphysical_m, family_m, death_m, friend_q, present_k	1stperson_plural_k, eat_s, occupation_k, insight_m, tv_s
	young	0.70	inclusion_v, social_m, friend_k, sleep_v, tv_k, motion_k, anger_q, friend_v, music_k	money_v, eat_q, human_m, affect_m, death_m, sexual_q, time_m, up_m, preposition_s, inhibition_q, commun_m, human_s, article_m, job_s, swear_q	negation_q, occupation_k, human_q, see_k, down_q, ositive_affect_k, insight_q, tv_s

(continues)

Table 2. Continued

	Trait	Accuracy	User Imagery	Employee Imagery	Announcement
	imaginative	0.50	anger_q, see_k, sensation_k, causation_q	affect_m, causation_m, affect_q, eat_q	pronoun_k, eat_s, sleep_m, leisure_k, certainty_k, affect_k
	unique	0.59	motion_s, money_v, swear_k	causation_m, money_q, future_m, down_v, negative_affect_k, sport_q, certainty_m, friend_q, touching_m, sexual_m, death_m, death_v	inclusion_v, achievement_q, sleep_m, see_k, 1stperson_plural_k, eat_s, school_v, filler_m
	up-to-date	0.55	present_s, tv_k, money_k, commun_s, time_k	death_v, sexual_v, number_k, 1stperson_q, exclusion_s, negation_k, up_s, sport_m, preposition_q, unmatched_s, future_m, present_k	2ndperson_q (-0.09), certainty_k, eat_s, negation_q, metaphysical_k, 1stperson_plural_k, affect_k, leisure_k, friend_k, sleep_m
	independent	0.47	ref_other_v, commun_s, preposition_k, causation_k, filler_k, space_k, insight_s, achievement_v, exclusion_k	number_k, assent_q, anxiety_q, future_q, sport_k, present_k, metaphysical_v, negative_affect_k, tv_q, money_k	1stperson_plural_k, hear_m, music_q, inclusion_v, sleep_m
	contemporary	0.53	friend_v, touching_m	death_m, sexual_v, motion_v, death_v	occupation_q, affect_k, 1stperson_k, past_s
Competence	**reliable**	0.44	commun_q, music_s, 1stperson_plural_v, hear_s, future_k, sexual_k, past_k, negation_s, optimism_k	tv_q, music_m, certainty_q, negative_affect_k, money_m, death_m, exclusion_k, family_v, certainty_m, present_m, commun_q, negation_s	NA

hard-working	0.47	inclusion_m, commun_q, future_k, eat_m, hear_s, 1stperson_plural_v, discrepancy_s, music_s, space_s, sensation_s, optimism_s, negation_s	music_m, family_v, sexual_m, article_m, up_m, death_m, exclusion_k, tv_q, tentative_q, number_k, symptom_m, article_k, achievement_q, commun_s	money_v, anger_q, human_q, assent_k, optimism_m, preposition_m, causation_q
secure	0.52	sexual_s, touching_m, preposition_k, filler_k, inhibition_m	preposition_q, commun_k, swear_q, music_m, physicial_k, up_m, down_v	1stperson_q, family_s, discrepancy_m, sleep_m, assent_k, inhibition_q
intelligent	0.62	discrepancy_k, anger_q, past_k	music_m, future_q, tv_q, down_q, swear_q, negation_k, pronoun_k	job_q, 1stperson_q, article_m, assent_k
technical	0.78	inhibition_k, past_k, anxiety_v, see_s, preposition_q, cognition_k, sport_q	number_k, negation_k, present_k, grooming_m, touching_q, hear_k	job_q, eat_s, sensation_k, assent_q, space_m
corporate	0.51	touching_m, occupation_v, human_k	preposition_q, certainty_k, grooming_q, death_m, physicial_k, negation_k	sexual_m
successful	0.39	money_s, commun_q, cognition_s, optimism_q, inhibition_m, symptom_v	human_m, swear_q, cognition_q, tv_q, leisure_m, preposition_q, touching_m, present_k, positive_feelings_m, money_k, future_q, money_m	optimism_m
leader	0.59	inhibition_m, discrepancy_k, see_v	exclusion_s, music_m, present_k, 1stperson_plural_q, number_k, preposition_q, physicial_k	tv_m, sleep_m, optimism_m

(continues)

Table 2. Continued

	Trait	Accuracy	User Imagery	Employee Imagery	Announcement
	confident	0.56	commun_q, inhibition_v	1stperson_plural_q, touching_m, number_m, sexual_v, down_m, exclusion_k, school_q	sleep_m
Sophistication	**upper-class**	0.59	1stperson_m, anger_q, sleep_v, 3rdperson_s, hear_q, tv_s, article_q, grooming_v, time_s, 1stperson_plural_k, touching_v, negation_s	friend_q, tentative_q, present_m, job_k, eat_q, causation_m, up_q, human_m, family_q, swear_q	2ndperson_q, motion_q, motion_k, past_m, space_m, sad_k, number_k, exclusion_m
	glamorous	0.60	sport_q, see_k	eat_q, death_m, family_q, tentative_q, certainty_m, number_m, anger_v	job_v, see_q
	good-look	0.65	sleep_v, anger_q	eat_q, present_m, sexual_v, death_m	eat_s
	charming	0.64	ref_other_m, friend_m, see_k, money_v, down_m, inclusion_k, grooming_s, preposition_k, negation_s, anger_q, friend_k	present_m, friend_v, up_m, inclusion_s, symptom_s, metaphysical_q, sexual_m, ref_other_k, time_k, certainty_m	family_m, past_q, tv_v, space_m, certainty_q, friend_k, anxiety_k, ositive_affect_k, inclusion_v, religion_m, negation_k, sleep_m, see_q
	feminine	0.75	money_v, number_m, article_s, metaphysical_q, future_k, grooming_v, human_s, preposition_v, hear_s	tentative_q, music_m, money_q, religion_v, ositive_affect_k	assent_k, tv_v, past_m, discrepancy_q, tv_q, preposition_m

	smooth	0.60	unmatched_s, grooming_v, money_q, article_m, 1stperson_s, past_s, unmatched_v, sad_k, money_v, cognition_m, causation_m	money_m, family_q, preposition_m, assent_m, family_v, certainty_q, past_v, sexual_m, inhibition_m, article_q, symptom_s, death_m	NA
Ruggedness	**outdoorsy**	0.35	inclusion_m, metaphysical_m, time_s, motion_q, number_m, achievement_q, symptom_v, sad_k, job_k, 1stperson_plural_s, inhibition_k	affect_m, tentative_q, 1stperson_v, unmatched_s, touching_m, space_k, inclusion_k, certainty_q, sexual_v, family_v, article_m, death_v, down_q	preposition_q, touching_q, tv_q
	masculine	0.50	commun_q, hear_s, money_v, see_q, optimism_k	NA	tv_v, assent_k, anger_q, past_m
	Western	0.28	unmatched_v, 1stperson_plural_v, future_k, human_s, commun_s, discrepancy_k, anxiety_k	music_m, certainty_q, commun_s, family_q, money_v, exclusion_k, negative_affect_s, filler_q, preposition_m, inhibition_m, ref_other_k	tv_v, 1stperson_plural_q, money_m, assent_k, sexual_s, achievement_k, anger_q, hear_k, 3rdperson_q
	tough	0.54	commun_q, optimism_k, inclusion_m, sad_s, 1stperson_plural_v, human_s, future_k, past_s, down_m, unmatched_q, negation_s, inhibition_m,	NA	NA
	rugged	0.45	money_v, friend_m	family_v, certainty_q	space_m, past_m, achievement_k, family_s

Note: The regression results for predicting brand personality using the predictors from: *User Imagery, Employee Imagery*, and *Official Announcement*. The predicted R^2 was reported with each personality trait (Left columns) and the Right columns show the predictors that were consistently selected in the cross-validation of the prediction. The predictors are listed in descending order of their standardized beta coefficients' absolute value, and the predictors with negative coefficients are highlighted in red.

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* (β = 0.18), and *friendly* (β = 0.11). Similarly, *friends* and *family* words positively correlated the perception of *wholesome* (β = 0.16), *family-oriented* (β = 0.12), *sincere* (β = 0.12), *cheerful* (β = 0.09), *original* (β = 0.05), and *friendly* (β = 0.05).

Biological Processes was observed to be most predictive of *Sophistication*. The frequency of *sexual* words was positively correlated with *charming* (β = 0.06), while the frequency of *eating* words was negatively correlated with *good-looking* (β = −0.18), *feminine* (β = −0.14), *glamorous* (β = −0.12), and *upper-class* (β = −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 [102].

Online System for Managing and Tracking Brand Personality

We further illuminate the practical implication of our predictive model by building a cloud-based information system that allows managers and analysts to explore and track personality of their own brands and their competitors' brands. For the details of the system prototype, please read our online Supplemental Appendix A.

General Discussion

Summary

Our investigation was aimed at examining the business value of social media analytics by predicting and measuring brand personality. Using a large-scale consumer survey to establish the "ground truth" of brand personality and millions of unstructured text documents from three different secondary sources, we have shown that brand personality can be predicted with a reasonable level of resolution; in particular, its prediction compares favorably to human personality prediction from writing samples that have been deemed acceptable for most applications [17, 40].

From a methodological standpoint, we make several advances. First, we are the first to develop a predictive model of brand personality using consumer- and brand-generated content. In doing so, we utilize Elastic-net regression, which is a relatively novel regression method in statistics, and which has, to date, seen only very limited applications in the information systems literature. We justified the motivation for using this approach to address our research question and proved its prediction performance. Second, we employ automated textual analysis to the problem of predicting brand personality. Previously, textual analysis has focused on human personality [6]; however, our findings demonstrate its relevance to the assessment of brand personality.

From a theoretical perspective, we provide important contributions to the literatures on user-generated content, social media, and brand personality. By combining consumer-, employee-, and firm-generated content from social media platforms, we are able to gain important insights into the prediction and measurement of brand personality, a concept of strategic business value. As information systems and information technology are now deeply embedded in many business functions, our research showcases the importance of IS research in generating impactful and actionable insights for important business strategies [16, 29, 33, 79, 113]. In terms of the brand personality literature, we have provided the first real-world evidence demonstrating a brand personality — consumer identity link through showing that user-generated content (textual data from user profiles of those who choose to associate with a brand) can be used to predict and measure brand personality (which was obtained independently through consumer ratings). This finding has important conceptual implications as it provides support for the basic rationale of the brand personality construct. We also identified a link between brand personality and an organization's internal environment. We did this through showing that employee-generated content (employee evaluations of a company) are predictive of the firm's brand personality (as judged by consumers). In principle, a company with a cutthroat, backstabbing culture can adopt a sincere brand identity. However, practitioners [74] and scholars [55, 97] alike have argued that a firm's brand inherently reflects its culture and behavior. For example, James Martin [74], founder of the highly regarded Martin advertising agency admonished firms to "Be the Brand." Our findings lend support to this idea and argue for future research to examine what underpins this link.

Implications for System Design

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 [23]. Brand personality modeling can be used to quantify symbolic meanings of products at scale and allow recommender systems to consider products' symbolic meaning to satisfy users' individual and social needs. Consider red wine, for example, where few customers can actually tell the taste differences between brands. Yet, wine brands have different personalities (symbolic meanings), are served in a social setting, and 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 providing career advice 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 [57]. People tend to 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 these systems to find companies that match their personalities and interests well, hiring managers can also utilize these systems to screen candidates from the recruiting perspective.

Similar to recent IS research that has modeled business proximity with archival data for competitive business intelligence [90], brand personality modeling 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 [21]. An analytic tool could be developed to help brand managers assess brand personality, detect perception gaps, and improve perceptions. More specifically, such a tool could help managers assess the perceived brand personality of their brand and monitor it over time. Managers might also be able to view a perceptual map identifying the personality of their brand in relation to the personality of other brands in perceptual space. Likewise, such a tool could be used to detect the gaps between the perceived and targeted personality of a brand by summarizing and highlighting the differences in notable personality dimensions. Moreover, because our model assigns weights to different variables that predict brand personality, the tool could suggest actions to help bridge the perceptions gaps in personality dimensions.

Limitations and Future Directions

Our approach does not explicitly model the dynamic aspects of brand personality, which likely changes in response to marketing actions and changes in the social and cultural environment, among other factors. This opens up an array of potential future research questions to investigate, including examining how brand personality changes in response to changes in product development, re-positioning initiatives, advertising campaigns, and crisis management. Our model also offers the possibility to dig further into the drivers of brand personality change through examining how marketing actions influence specific user imagery and employee imagery variables (which, in turn, affect brand personality).

Our model variables are currently extracted from two social media sites, Twitter and Glassdoor, and rely on feature extraction using LIWC. Additional data sources (e.g., Facebook, LinkedIn) and features (e.g., social, semantic, and temporal features) can be integrated to improve the model performance. In addition, product reviews and product attributes, from sites such as Amazon can be integrated into the model.

We expect that our findings will 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 type of data that is predictive of brand personality and the establishment of a brand personality — consumer identity link is likely to be generalizable. It would be interesting to examine how the prediction model can be applied in different cultures and social platforms.

CONCLUSION

In this work, we developed the first prediction model to measure brand personality from multiple archival sources of social media content, including content generated by the consumers, employees, and firms. This model and resulting classifications provide a proof of concept for the idea that brand personality can be extracted from social media analytics with high accuracy, paralleling similar efforts to extract human personality from writing samples rather than from individual survey ratings data. Our high-performance model also provides the first real-world evidence for a brand personality-consumer identity link, thereby supporting the fundamental notion that consumers use brands symbolically to communicate their identity. In addition, we show that brand personality tends to reflect the internal environment of the firm originating the brand rather than to arise *ex nihilo*.

In today's business environments, to gain competitive intelligence, firms need to gather data not just internally but externally as well. Understanding and accurately measuring fundamental business concepts and metrics could substantially enhance a firms' strategic decision making, such as product positioning, new product development, and customer relationship management. Given the amount of data being generated via social media in terms of both quantity and format variety, we hope the present research encourages researchers and practitioners to utilize state-of-the-art text mining and machine learning methods to conduct social media analytics, thereby gaining actionable insights into important business concepts (such as brand personality) and into using such data to gain competitive business intelligence on human-brand interactions.

NOTES

1 https://www.google.com/recaptcha/intro/v3.html

2 http://support.twitter.com/articles/166337-the-twitter-glossary

3 http://people.eng.unimelb.edu.au/tbaldwin/etc/emnlp2012-lexnorm.tgz

4 When predictors are highly correlated, Elastic-Net regression shrinks coefficients of correlated variables together (e.g., $\alpha < 1$), leading to a more stable behavior than Lasso regression [41].

5 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 ($p = .32$, *t*-test). R^2 values from *Announcements* was still lower than from the undersampled *User* data ($p < .01$). In this paper, we only reported the results from the full-size data for brevity.

Supplemental Material

Supplemental material for this article can be accessed on the publisher's website.

REFERENCES

1. Aaker, D.A. *Building Strong Brands*. Simon and Schuster, New York City, New York, 2012.

2. Aaker, D.A.; and Joachimsthaler, E. *Brand Leadership*. Simon and Schuster, New York City, New York, 2012.

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

4. Aaker, J.L. The malleable self: The role of self-expression in persuasion. *Journal of Marketing Research*, *36*, 1, (1999), 45-57.

5. Abbasi, A.; Zhou, Y.; Deng, S.; and Zhang, P. Text analytics to support sense-making in social media: A language-action perspective. *MIS Quarterly*, *42*, 2, (2014), 427–464.

6. Adamopoulos, P.; Ghose, A.; and Todri, V. The impact of user personality traits on word of mouth: text-mining social media platforms. *Information Systems Research*, *29*, 3, (2018), 612-640.

7. Ambroise, L.; Ferrandi, J.M.; Merunka, D.; Valette-Florence, P.; Zine-Danguir, S.; and Jolivot, A.G. Modeling and measuring brand personality: A cross-cultural application. In *The Ninth Cross-Cultural Research Conference*, Half Monn Bay, Jamaica, 2003, 29–34.

8. Bhattacharjee, A.; Berger, J.; and Menon, G. When identity marketing backfires: Consumer agency in identity expression. *Journal of Consumer Research*, *41*, 2, (2014), 294–309.

9. Bollen, J.; Mao, H.; and Pepe, A. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain, AAAI, 2011.

10. Brexendorf, T.O.; and Kernstock, J. Corporate behaviour vs. brand behaviour: Towards an integrated view?. *Journal of Brand Management*, *15*, 1, (2007), 32–40.

11. Buhrmester, M.; Kwang, T.; and Gosling, S.D. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*, 1, (2011), 3–5.

12. Bulgurcu, B.; Van Osch, W.; and Kane, G.C. The rise of the promoters: User classes and contribution patterns in enterprise social media. *Journal of Management Information Systems*, *35*, 2, (2018), 610–646.

13. Burtch, G.; Hong, Y.; Bapna, R.; and Griskevicius, V. Stimulating online reviews by combining financial incentives and social norms. *Management Science*, *64*, 5, (2017), 2065–2082.

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

15. Chen, D.L.; and Horton, J.J. Research note-Are online labor markets spot markets for tasks? A field experiment on the behavioral response to wage cuts. *Information Systems Research*, *27*, 2, (2016), 403–423.

16. Chen, H.; Chiang, R.H.; and Storey, V.C. Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*, 4, (2012).

17. Chen, J.; Hsieh, G.; Mahmud, J.U.; and Nichols, J. Understanding individuals' personal values from social media word use. In *Proceedings of the 17th ACM Conference On Computer Supported Cooperative Work and Social Computing*, Baltimore, Maryland,USA, ACM, 2014, 405–414.

18. Cialdini, R.B.; and Trost, M.R. Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, and G. Lindzey (eds.), *The Handbook of Social Psychology*. New York, NY: McGraw-Hill, 1998, pp. 151–192

19. Cohen, R.; and Ruths, D. Classifying political orientation on Twitter: It's not easy!. In *Seventh International AAAI Conference on Weblogs and Social Media*, Dublin, Ireland, AAAI, (2013).

20. Culotta, A.; and Cutler, J. Mining brand perceptions from twitter social networks. *Marketing Science*, *35*, 3, (2016) 343–362.

21. De Chernatony, L. (1999). Brand management through narrowing the gap between brand identity and brand reputation. *Journal of Marketing Management*, *15*, 1–3, (2016), 157–179.

22. De Choudhury, M.; Gamon, M.; Counts, S.; and Horvitz, E. Predicting depression via social media. *In Seventh International AAAI Conference on Weblogs and Social Media*, Dublin, Ireland, AAAI, 2013.

23. Dittmar, H. *The Social Psychology of Material Possessions: To Have Is to Be*. Harvester Wheatsheaf and St. Martin's Press, London, 1992.

24. Dong, W.; Liao, S.; and Zhang, Z. Leveraging financial social media data for corporate fraud detection. *Journal of Management Information Systems*, *35*, 2, (2018), 461–487.

25. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.; Gruber, B.; Lafourcade, B.; Leitão, P.J.; and Münkemüller, T. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*, 1 (2013), 27–46.

26. Easley, R.F.; Devaraj, S.; and Crant, J.M. Relating collaborative technology use to teamwork quality and performance: An empirical analysis. *Journal of Management Information Systems*, *19*, 4, (2003), 247–265.

27. Elliott, R.; and Wattanasuwan, K. Brands as symbolic resources for the construction of identity. *International journal of Advertising*, *17*, 2, (1998), 131–144.

28. Fader, P.S.; and Winer, R.S. Introduction to the special issue on the emergence and impact of user-generated content. *Marketing Science*, *31*, 3, (2012), 369–371.

29. Fan, W.; and Gordon, M.D. The power of social media analytics. *Commun. ACM*, *57*, 6, 74–81. (2014).

30. Freling, T.H.; Crosno, J.L.; and Henard, D.H. Brand personality appeal: Conceptualization and empirical validation. *Journal of the Academy of Marketing Science*, *39*, 3, (2011), 392–406.

31. Gal, D. Identity-signaling behavior. *The Cambridge Handbook of Consumer Psychology*, (2015), 257-281.

32. Gardner, B.B.; and Levy, S.J. The product and the brand. *Harvard Business Review*, *33*, 2, (1955), 33–39.

33. Ge, R.; Feng, J.; Gu, B.; and Zhang, P. Predicting and deterring default with social media information in peer-to-peer lending. *Journal of Management Information Systems*, *34*, 2, (2017), 401–424.

34. Gilbert, E.; and Karahalios, K. Predicting tie strength with social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Boston, MA, USA, ACM, 2009, 211–220.

35. Goel, S.; and Goldstein, D.G. Predicting individual behavior with social networks. *Marketing Science*, *33*, 1, (2013), 82–93.

36. Goh, K.Y.; Heng, C.S.; and Lin, Z. Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. *Information Systems Research*, *24*, 1, (2013), 88–107.

37. Golbeck, J.; Robles, C.; and Turner, K. Predicting personality with social media. *In CHI'11 Extended Abstracts on Human Factors in Computing Systems*, Vancouver, Canada, ACM, 2011.

38. Goldberg, L.R. An alternative" description of personality": The big-five factor structure. *Journal of Personality And Social Psychology*, *59*, 6, (1990), 1216.

39. Gosling, S.D.; Gaddis, S.; and Vazire, S. Personality impressions based on facebook profiles. *ICWSM*, *7*, (2007), 1–4.

40. Gou, L.; Zhou, M.X.; and Yang, H. KnowMe and ShareMe: Understanding automatically discovered personality traits from social media and user sharing preferences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto, Canada, ACM, 2014.

41. Grave, E.; Obozinski, G.R.; and Bach, F.R. Trace lasso: A trace norm regularization for correlated designs. In *Advances in Neural Information Processing Systems*, (2011), 2187–2195.

42. Gunarathne, P.; Rui, H.; and Seidmann, A. Whose and what social media complaints have happier resolutions? Evidence from Twitter. *Journal of Management Information Systems*, *34*, 2, (2017), 314–340.

43. Hauser, J.R.; and Koppelman, F.S. Alternative perceptual mapping techniques: Relative accuracy and usefulness. *Journal of Marketing Research*, *16*, 4, (1979), 495-506.

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

45. Hong, Y.; Chen, P.Y.; and Hitt, L.M. Measuring product type with dynamics of online product review variances: A theoretical model and the empirical applications, (2014). Available at SSRN2506328.

46. Hong, Y.; Huang, N.; Burtch, G.; and Li, C. Culture, conformity and emotional suppression in online reviews. *Journal of the Association for Information Systems*, *17*, 11,(2016), 308–329.

47. Huang, N.; Hong, Y.; and Burtch, G. Social network integration and user content generation: Evidence from natural experiments. *MIS Quarterly*, *41*, 4(2016), 17-001.
48. Huber, J.; and Holbrook, M.B. Using attribute ratings for product positioning: Some distinctions among compositional approaches. *Journal of Marketing Research*, *16*, 4, (1979), 507–516.
49. Ipeirotis, P.G. Demographics of Mechanical Turk. *In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, New York, NY, USA, ACM, 2010, 135-143.
50. Jia, J.; and Yu, B. On model selection consistency of the elastic net when p□ n. *Statistica Sinica*, (2010), 595-611.
51. John, D.R.; Loken, B.; Kim, K.; and Monga, A.B. Brand concept maps: A methodology for identifying brand association networks. *Journal of Marketing Research*, *43*, 4, (2006), 549–563.
52. Kassarjian, H.H. Personality and consumer behavior: A review. *Journal of Marketing Research*, *8*, 4, (1971), 409–418.
53. Katz, D. The functional approach to the study of attitudes. *Public Opinion Quarterly*, *24*, 2, (1960), 163–204.
54. Keller, K.L. Conceptualizing, measuring, and managing customer-based brand equity. *Journal of Marketing*, *57*, 1, (1993), 1–22.
55. Keller, K.L.; and Richey, K. The importance of corporate brand personality traits to a successful 21st century business. *Journal of Brand Management*, *14*, 1–2, (2006). 74–81.
56. Kosala, R.; and Blockeel, H. Web mining research: A survey. *ACM Sigkdd Explorations Newsletter*, *2*, 1, (2000). 1–15.
57. Kristof-Brown, A.L.; Zimmerman, R.D.; and Johnson, E.C. Consequences of individuals fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology*, *58*, 2, (2005). 281–342.
58. Kuksov, D.; Shachar, R.; and Wang, K. Advertising and consumers' communications. *Marketing Science*, *32*, 2, (2013). 294–309.
59. Kulshrestha, J.; Zafar, M.B.; Noboa, L.E.; Gummadi, K.P.; and Ghosh, S. Characterizing Information Diets of Social Media Users. *Ninth International AAAI Conference on Web and Social Media*, University of Oxford, Oxford, UK,2015.
60. Le Cessie, S.; and Van Houwelingen, J.C. Ridge estimators in logistic regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *41*, 1, (1992), 191–201.
61. Lee, T.Y.; and Bradlow, E.T. Automated marketing research using online customer reviews. *Journal of Marketing Research*, *48*, 5, (2011), 881–894.
62. Lehmann, D.R.; Keller, K.L.; and Farley, J.U. The structure of survey-based brand metrics. *Journal of International Marketing 16*, 4, (2008), 29–56.
63. Liu, X.; Zhang, B.; Susarla, A.; and Padman, R.Go to YouTube and See Me Tomorrow: The Role of Social Media in Managing Chronic Conditions. *Available at SSRN 3061149*, (2018). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3061149, July 2019
64. Lowry, P.B.; Vance, A.; Moody, G.; Beckman, B.; and Read, A. Explaining and predicting the impact of branding alliances and web site quality on initial consumer trust of e-commerce web sites. *Journal of Management Information Systems*, *24*, 4, (2008), 199–224.
65. Luo, X.; and Zhang, J. How do consumer buzz and traffic in social media marketing predict the value of the firm?. *Journal of Management Information Systems*, *30*, 2, (2013), 213–238.
66. Maehle, N.; and Shneor, R. On congruence between brand and human personalities. *Journal of Product and Brand Management*, *19*, 1, (2010), 44–53.
67. Mai, F.; Shan, Z.; Bai, Q.; Wang, X.; and Chiang, R.H. How does social media impact Bitcoin value? A test of the silent majority hypothesis. *Journal of Management Information Systems*, *35*, 1, (2018), 19–52.
68. Mairesse, F.; Walker, M.A.; Mehl, M.R.; and Moore, R.K. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, *30*,(2007), 457-500.
69. Malär, L.; Krohmer, H.; Hoyer, W.D.; and Nyffenegger, B. Emotional brand attachment and brand personality: The relative importance of the actual and the ideal self. *Journal of Marketing*, *75*, 4, (2011), 35–52.

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

71. Markus, H. Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, *35*, 2, (1977), 63.

72. Markus, H.; and Kunda, Z. Stability and malleability of the self-concept. *Journal of Personality and Social Psychology*, *51*, 4, (1986), 858.

73. Markus, H.; and Nurius, P. Possible selves. *American Psychologist*, *41*, 9, (1986), 954.

74. Martin, D.N. *Be the Brand: How to Find a Powerful Identity and Use It to Drive Sales*. Richmond, VA :Oaklea Press, 2000.

75. McCracken, G. Who is the celebrity endorser? Cultural foundations of the endorsement process. *Journal of Consumer Research*, *16*, 3, (1989), 310–321.

76. Montgomery, D.C.; Peck, E.A.; and Vining, G.G. *Introduction to Linear Regression Analysis*. Chicago, IL: John Wiley and Sons, 2012.

77. Nallapati, R.; Zhai, F.; and Zhou, B. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

78. Naylor, R.W.; Lamberton, C.P.; and West, P.M. Beyond the "like" button: The impact of mere virtual presence on brand evaluations and purchase intentions in social media settings. *Journal of Marketing*, *76*, 6, (2012), 105–120.

79. Netzer, O.; Feldman, R.; Goldenberg, J.; and Fresko, M. Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, *31*, 3, (2012), 521–543.

80. Nunes, M.A.S.N. Recommender systems based on personality traits, *Doctoral dissertation*, *Université Montpellier II-Sciences et Techniques du Languedoc*, 2008.

81. Pan, Z.; Lu, Y.; Wang, B.; and Chau, P.Y. Who do you think you are? Common and differential effects of social self-identity on social media usage. *Journal of Management Information Systems*, *34*, 1, (2017), 71–101.

82. Peer, E.; Vosgerau, J.; and Acquisti, A. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*, 4, (2014), 1023–1031.

83. Pennebaker, J.W.; Francis, M.E.; and Booth, R.J. *Linguistic Inquiry and Word Count: LIWC* 2001. Mahwah, NJ: Lawrence Erlbaum Associates, 2001, p. 71.

84. Pierre, M. The personality of the retail store. *Harvard Business Review*, (1958), 47-55.

85. Plummer, J.T. *Brand Personality: A Strategic Concept for Multinational Advertising*. New York: Young and Rubicam, 1985.

86. Qiu, L.; and Kumar, S. Understanding voluntary knowledge provision and content contribution through a social-media-based prediction market: A field experiment. *Information Systems Research*, *28*, 3, (2017), 529–546.

87. Qiu, L.; Tang, Q.; and Whinston, A.B. Two formulas for success in social media: Learning and network effects. *Journal of Management Information Systems*, *32*, 4, (2015), 78–108.

88. Reed II, A.; Forehand, M.R.; Puntoni, S.; and Warlop, L. Identity-based consumer behavior. *International Journal of Research in Marketing*, *29*, 4, (2012), 310–321.

89. Shami, N.S.; Muller, M.; Pal, A.; Masli, M.; and Geyer, W. Inferring employee engagement from social media. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul, Korea, 2015.

90. Shi, Z.; Lee, G.M.; and Whinston, A.B. Toward a better measure of business proximity: Topic modeling for industry intelligence. *MIS quarterly*, *40*, 4(2016),1035–1056.

91. Shi, Z.; Rui, H.; and Whinston, A.B. Content sharing in a social broadcasting environment: Evidence from twitter. *Available at SSRN 2341243*, (2014).

92. Sibona, C.; and Choi, J.H. Factors affecting end-user satisfaction on Facebook. In *Sixth International AAAI Conference on Weblogs and Social Media*, Dublin, Ireland, 2012.

93. Sirgy, M.J. Self-concept in consumer behavior: A critical review. *Journal of consumer research*, *9*, 3, (1982), 287–300.

94. Steenkamp, J.B.E.; Van Trijp, H.C.; and Berge, J.M.T. Perceptual mapping based on idiosyncratic sets of attributes. *Journal of Marketing Research*, *31*, 1, (1994), 15–27.

95. Steenkamp, J.B.; and Van Trijp, H. Attribute elicitation in marketing research: A comparison of three procedures. *Marketing Letters*, *8*, 2, (1997), 153–165.

96. Stieglitz, S.; and Dang-Xuan, L. Emotions and information diffusion in social media-sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, 29, 4, (2013), 217–248.

97. Sung, Y.; and Tinkham, S.F. Brand personality structures in the United States and Korea: common and culture-specific factors. *Journal of Consumer Psychology*, *15*, 4, (2005), 334–350.

98. Susarla, A.; Oh, J.H.; and Tan, Y. Social networks and the diffusion of user-generated content: Evidence from YouTube. *Information Systems Research*, *23*, 1, (2012), 23–41.

99. Tang, Q.; Gu, B.; and Whinston, A.B. Content contribution for revenue sharing and reputation in social media: A dynamic structural model. *Journal of Management Information Systems*, *29*, 2, (2012), 41–76.

100. Tausczik, Y.R.; and Pennebaker, J.W. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*, 1, (2010), 24–54.

101. Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*, 1, (1996), 267–288.

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

103. Van Rekom, J.; Jacobs, G.; and Verlegh, P.W. Measuring and managing the essence of a brand personality. *Marketing Letters*, *17*, 3, (2006), 181–192.

104. Visser, P.S.; Krosnick, J.A.; and Simmons, J.P. Distinguishing the cognitive and behavioral consequences of attitude importance and certainty: A new approach to testing the common-factor hypothesis. *Journal of Experimental Social Psychology*, *39*, 2, (2003), 118–141.

105. Ward, S. Consumer socialization. *Journal of Consumer Research*, *1*, 2, (1974), 1–14.

106. Wasserman, L. *All of Statistics: A Concise Course in Statistical Inference*. Berlin/ Heidelberg, Germany: Springer Science and Business Media, 2013.

107. Wixom, B.H.; and Todd, P.A. A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research*, *16*, 1, (2005), 85–102.

108. Wixom, B.; Ariyachandra, T.; Douglas, D.E.; Goul, M.; Gupta, B.; Iyer, L.S.; and Turetken, O. The current state of business intelligence in academia: The arrival of big data. *CAIS*, *34*, 1, (2014), 1–13.

109. Xie, K.; and Lee, Y.J. Social media and brand purchase: Quantifying the effects of exposures to earned and owned social media activities in a two-stage decision making model. *Journal of Management Information Systems*, *32*, 2(2015), 204-238.

110. Yang, M.; Ren, Y.C.; and Adomavicius, G. Understanding user-generated content and customer engagement on Facebook business pages. *Information Systems Research*, (2019), (Forthcoming).

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

112. Yin, D.; Bond, S.; and Zhang, H. Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly*, *38*, 2, (2014), 539–560.

113. Zhang, H.; Kim, G.; and Xing, E.P. Dynamic topic modeling for monitoring market competition from online text and image data. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Sydney, NSW, Australia, ACM, 2015, 1425–1434.

114. Zhang, K.; and Moe, W.W. *Measuring Brand Favorability Using Large-Scale Social Media Data*, 2017. Online version only, URL is https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3165530,accessed on July 1, 2019

115. Zhang, K.; Bhattacharyya, S.; and Ram, S. Large-scale network analysis for online social brand advertising. *MIS Quarterly*, *40*, 4, (2016), 849-868.

116. Zou, H.; and Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society*, *67*, 2,(2005), 301-320.

117. Aaker, J.L. Dimensions of brand personality. *Journal of Marketing Research*, (1997), 347-356.

118. Nallapati, R., Zhai, F., and Zhou, B. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *Thirty-First AAAI Conference on Artificial Intelligence*, (2017).