A theoretical framework to identify authentic online reviews

Banerjee, Snehasish; Chua, Alton Yeow Kuan

2014

Banerjee, S., & Chua, A. Y. K. (2014). A theoretical framework to identify authentic online reviews. Online information review, 38(5), 634-649.

https://hdl.handle.net/10356/105158

https://doi.org/10.1108/OIR-02-2014-0047

A Theoretical Framework to Identify Authentic Online Reviews

Snehasish Banerjee, Alton Y.K. Chua

Wee Kim Wee School of Communication and Information,

Nanyang Technological University, Singapore, Singapore

Abstract

Purpose – This paper investigates the extent to which textual characteristics of online reviews help identify authentic entries from manipulative ones across positive and negative comments.

Design/Methodology/Approach – A theoretical framework is proposed to identify authentic online reviews from manipulative ones based on three textual characteristics, namely, comprehensibility, informativeness and writing style. The framework is tested using two publicly available datasets, one comprising positive reviews to hype own offerings, and the other including negative reviews to slander competing offerings. Logistic regression is used for analysis.

Findings – The three textual characteristics offered useful insights to identify authentic online reviews from manipulative ones. In particular, the differences between authentic and manipulative reviews in terms of comprehensibility and informativeness were more conspicuous for negative entries. On the other hand, the differences between authentic and manipulative reviews in terms of writing style were more conspicuous for positive entries.

Research limitations/implications – The findings of this paper are somewhat constrained by the scope of the datasets used for analysis.

Originality/value – The paper represents one of the earliest attempts to develop a theoretical framework to identify authentic online reviews. Prior research has shed light on ways to classify reviews as authentic or manipulative. However, literature on specific differences between the two in terms of textual characteristics is relatively limited. Moreover, by suggesting differences between authentic and manipulative reviews across positive and negative comments, the findings offer nuanced insights into a research area that is growing in importance.

Keywords User-generated content, Online reviews, Theoretical framework, Comprehensibility, Informativeness, Writing Style

Paper type Research paper

Introduction

Social media has been growing in popularity and acceptance. As it grows into an integral part of the Web-blitzed society, users continue to hop on the social media bandwagon. For example, more than 1.4 billion users used social media in 2012, a significant 19% rise from a year earlier (eMarketer, 2013). Such expansion in the user base of social media spawns an ever growing volume of user-generated content, a common form of which includes online reviews (henceforth, reviews). Reviews are user-generated evaluation of products or services posted in review websites such as Tripadvisor.com and Epinions.com to share post-purchase experiences with the online communities. Since reviews are perceived as being posted free of commercial interests, they are often considered more authentic than marketer-generated information (Otterbacher, 2009).

However, authenticity of reviews is not guaranteed. Users' preference for reviews has fuelled a new spamming boom, called spam 2.0 (Hayati and Potdar, 2009). Deemed as being more sophisticated than traditional spam activities (Jindal and Liu, 2008), spam 2.0 involves infiltrating social media platforms with inappropriate content. In the context of review websites, spam 2.0 entails malicious propagation of reviews that appear authentic but are meant to distort the reputation of products and services. Some organizations surreptitiously post not only positive manipulative reviews to hype their own offerings but also negative manipulative reviews to slander offerings of their rivals. For instance, publishers could post positive manipulative reviews to boost sales of their own books, and negative manipulative reviews to tarnish the reputation of competing ones (Harmon, 2004). Similar problems of review manipulation are also reported against several hospitality establishments across the world (Keates, 2007). Therefore, it is challenging to distinguish between authentic and manipulative reviews (Chiou *et al.*, 2013; Martin and Camarero, 2009).

Prior research suggests that authentic reviews can be identified on the basis of various textual characteristics. For example, authentic reviews could be generally terse and easy to read whereas manipulative entries tend to be verbose and ambiguous (Anderson and Simester, 2013; Daft and Lengel, 1984; Vartapetiance and Gillam, 2012). While authentic reviews may contain informative content, manipulative entries may lean towards being imaginative (Ott *et al.*, 2011; Rayson *et al.*, 2001). Moreover, their differences could be discerned from the use of affective cues, perceptual words, tense and punctuations (Yoo and Gretzel, 2009; Zhou *et al.*, 2004). Amidst such ongoing scholarly efforts however, there is a lack of an overarching understanding of the nuances between authentic and manipulative reviews. Furthermore, the extent to which textual characteristics of authentic and manipulative reviews vary across positive and negative comments is largely unknown.

Given that posting manipulative reviews represents a viable business malpractice, they could be written with guile to be passed off as authentic. This calls for the development of a theoretical framework to identify authentic reviews among the plethora of positive and negative entries. Hence, the objective of this paper is three-fold. First, it aims to develop a theoretical framework, encompassing comprehensibility, informativeness and writing style, to help identify authentic reviews. Second, it attempts to validate the framework empirically using a combination of two publicly available datasets, one comprising positive reviews meant to hype own offerings (Ott *et al.*, 2011), and the other including negative reviews intended to slander competing offerings (Ott *et al.*, 2013). Both datasets comprise equal number of authentic and manipulative reviews. Third, it seeks to disinter ways in which differences between authentic and manipulative reviews vary across positive and negative comments. Specifically, the datasets used for analysis include reviews distributed across 20 popular hotels in Chicago.

This paper is potentially significant on two counts. First, it seems that extant literature lacks a comprehensive understanding of textual differences between authentic and manipulative reviews. Hence, this paper represents a modest attempt to develop a theoretical framework to identify authentic reviews. Second, by testing the framework across positive as well as negative comments, the paper has the potential to shed light on ways to distinguish between authentic and manipulative reviews.

Literature Review

In one of the earliest works, Jindal and Liu (2008) identified authentic from manipulative reviews using several review-centric, reviewer-centric and product-centric features. They suggested that manipulative reviews could be mostly duplicates of existing reviews. Yoo and Gretzel (2009) indicated that authentic reviews differed from manipulative reviews with respect to textual characteristics. Specifically, the former could be more readable and use fewer affective cues. In contrast, Harris (2012) found that authentic reviews were less readable than manipulative ones. In another study, Wu *et al.* (2010a) argued that extreme positive or negative reviews are more likely to be manipulative than those with mixed opinions. Other studies attempted to identify manipulative reviews by merging positive or negative extremeness of reviews with textual characteristics such as length of reviews or average words per sentence (Anderson and Simester, 2013; Wu *et al.*, 2010b). Research has also indicated that authentic and manipulative reviews could be classified using psycholinguistic dimensions (Ott *et al.*, 2011; 2013).

Despite such ongoing efforts, extant literature lacks an overarching theoretical framework to identify authentic reviews. Moreover, specific differences between authentic and manipulative reviews in terms of textual characteristics are largely unknown. Hence, this paper proposes a theoretical framework to identify authentic reviews based on three textual characteristics, namely, comprehensibility, informativeness and writing style. These textual characteristics are described as follows.

Review Comprehensibility

Users who post authentic reviews (henceforth, reviewers) as well as those who contribute manipulative reviews (henceforth, spammers) would want to make their entries comprehensible. However, since authentic and manipulative reviews are articulated to serve different purposes, their comprehensibility could be different (Vartapetiance and Gillam, 2012). For the purpose of this paper, comprehensibility of a given review refers to the extent to which it is of appropriate length and readability. Reviews should be of optimum length. While overly short reviews might be too sketchy to comprehend, overly detailed ones could be too intimidating for a large section of the online community (Otterbacher, 2009). Moreover, reviews should be easily readable because plain and simple comments are crucial to reach a wide audience (Ghose and Ipeirotis, 2011).

With respect to length, authentic entries were found to be terser than manipulative ones in settings such as mock theft experiments (Burgoon *et al.*, 2003) and financial statements (Humphreys *et al.*, 2011). Unlike authentic reviews, manipulative reviews perhaps contain more explanations to appear convincing (Anderson and Simester, 2013; Yoo and Gretzel, 2009). However, in an attempt to blur such differences, manipulative reviews could also be rendered terse. Hence, it is interesting to study differences between authentic and manipulative reviews in terms of length.

With respect to readability, there exist two opposing perspectives on ways authentic and manipulative reviews may differ. The first perspective holds that manipulative reviews could be more readable than authentic ones. This is because writing manipulative content is more cognitively demanding than articulating authentic experiences (Newman *et al.*, 2003). People performing a writing task with a high cognitive load tend to write more lucid language than those doing the task with a low cognitive load (Burgoon and Qin, 2006). This in turn may render manipulative reviews more readable than authentic reviews. The second perspective suggests that reviews that are overly simplistic could be perceived as being less credible by the online community (Ghose and Ipeirotis, 2011). Hence, unlike reviewers, spammers might deliberately use complex and sophisticated language to make manipulative reviews appear credible. Therefore, authentic reviews could be more readable than manipulative ones. Interestingly, Yoo and Gretzel (2009) found that authentic reviews were more readable vis-à-vis manipulative ones. The converse was indicated by Harris (2012). Given the inconsistent findings, it can be useful to study differences in readability between authentic and manipulative reviews across positive as well as negative comments.

Review Informativeness

Posting manipulative reviews requires articulating events that did not occur in reality, but in a convincing manner (Newman *et al.*, 2003). Texts written based on real experiences could be largely different from accounts hinged on imagined experiences (Vrij *et al.*, 2000). Generally, authentic reviews based on real experiences could be informative whereas manipulative reviews imaginative (Ott *et al.*, 2011).

However, such differences might be inconspicuous for two reasons. First, the informativeness of authentic reviews cannot be taken for granted. All authentic reviews need not necessarily be written in the most informative ways. Second, spammers could write manipulative reviews in a manner to deliberately make them as informative as authentic ones. Hence, shedding light on the informativeness of authentic and manipulative reviews across positive and negative comments can uncover interesting insights. Specifically, informative texts tend to differ from those that are imaginative in their distribution of part-of-speech (POS) tags (Rayson *et al.*, 2001). The higher the informativeness of a given text, the greater is the proportion of nouns, adjectives, articles, prepositions and the lesser is the proportion of verbs, adverbs, pronouns (Nakamura, 1991; Rayson *et al.*, 2001).

Among pronouns, personal pronouns need special attention in the context of authentic and manipulative reviews. On one hand, extant literature expects spammers to feel guilty and avoid statements of ownership, thereby using fewer personal pronouns (Newman *et al.*, 2003). On the other hand, Yoo and Gretzel (2009) found manipulative reviews to be richly laden with personal pronouns. Given that spammers increasingly appear to make manipulative reviews informative, it is interesting to study the extent to which informativeness can help identify authentic reviews across positive and negative comments.

Review Writing Style

Writing style refers to the ways specific types of words are used to convey opinions in reviews. For the purpose of this paper, writing style of authentic and manipulative reviews is conceived in terms of the use of affective cues, perceptual words, tense and punctuations.

First, given that authentic reviews could be written without any specific agenda, they could be mild in using affective cues. However, to create a more lasting impact, manipulative reviews could be richly embellished with affective cues (Maurer and Schaich, 2011). Second, authentic reviews written after real experiences could be richer in perceptual words vis-à-vis manipulative ones (Vrij *et al.*, 2000). This is because experiences with hotels are largely affected through sensory perceptions (Lin, 2004; Schiffman, 2001). This could be reflected through increased use of perceptual words in authentic reviews. Third, authentic reviews might contain more past tense compared with manipulative ones. Given that positive (negative) reviews could favorably (adversely) impact future sales and revenues of a given hotel (Dellarocas, 2003; Duan *et al.*, 2008), manipulative reviews might be articulated not only to describe past experiences but also to express present or future perceptions on the hotel. Therefore, nuances in the use of tense might shed light on identifying authentic reviews. Finally, punctuations could often be more telling than words to distinguish between authentic and manipulative reviews. In particular, manipulative reviews might contain more question marks and exclamation points than authentic reviews as a part of rhetorical strategies (Kim *et al.*, 2006; Zhou *et al.*, 2004). Table I summarizes the proposed theoretical framework to identify authentic reviews based on the three textual characteristics, namely, comprehensibility, informativeness and writing style.

Table I. Theoretical framework to identify authentic reviews

	Textual sub-dimensions	**Possible differences between authentic and manipulative reviews**	**References**
Comprehensibility	Length	Authentic reviews could be terser than manipulative ones.	Burgoon *et al.* (2003) Humphreys *et al.* (2011) Yoo and Gretzel (2009)
	Readability	Research on readability of authentic and manipulative reviews has yielded inconsistent findings.	Ghose and Ipeirotis (2011) Harris (2012) Zhou (2005)
Informativeness	POS tags	Authentic reviews could contain more nouns, adjectives, articles and prepositions as well as fewer verbs, adverbs and pronouns than manipulative ones.	Nakamura (1991) Rayson *et al.* (2001) Vrij *et al.* (2000)
	Personal pronouns	Research on the use of personal pronouns in authentic and manipulative reviews is inconclusive.	Newman *et al.* (2003) Yoo and Gretzel (2009)
Writing style	Affective cues	Authentic reviews could contain fewer affective cues than manipulative ones.	Maurer and Schaich (2011)
	Perceptual words	Authentic reviews could contain more perceptual words than manipulative ones.	Schiffman (2001) Vrij *et al.* (2000)
	Tense	Authentic reviews could contain more past tense than manipulative ones.	Dellarocas (2003) Duan *et al.* (2008)
	Punctuations	Authentic reviews could contain fewer question marks and exclamation points than manipulative ones.	Kim *et al.* (2006) Zhou *et al.* (2004)

Methodology

Dataset

Research on authentic and manipulative reviews is often hindered by the lack of ground truth (Wu *et al.*, 2010b). This has often led scholars to employ heuristic approaches to label reviews as authentic and manipulative. For instance, Jindal and Liu (2008) deemed duplicate or near duplicate reviews as manipulative ignoring that duplications might be coincidental. Although such approaches could occasionally be intuitive, they lack a compelling thrust.

Given the importance of ground truth, this paper uses a combination of two publicly available secondary datasets (Ott *et al.*, 2011; 2013). The combined dataset includes 1,600 reviews equally distributed across 20 popular hotels in Chicago. In particular, the first dataset comprises 800 positive reviews, of which, 400 are authentic and the rest are manipulative. The second dataset consists of 800 negative reviews, of which, 400 are authentic and the remainder are manipulative.

Operationalization

With respect to comprehensibility, length was operationalized in terms of number of words in a given review (Yoo and Gretzel, 2009). For readability, five metrics used in prior research were identified (Ghose and Ipeirotis, 2011; Korfiatis *et al.*, 2012). These include Gunning-Fog Index (FOG), Coleman-Liau Index (CLI), Automated-Readability Index (ARI), Flesch-Kincaid Grade Level (FKG) and Simple Measure of Gobbledygook (SMOG). Specifically, FOG is based on average sentence length and proportion of complex words, CLI and ARI rely on average word length and sentence length, while FKG and SMOG depend on average sentence length and syllables per word. A lower value for the metrics suggests a more readable review. Among these, SMOG was excluded because it requires texts of at least 30 sentences (Ayello, 1993). About 98% of all reviews in the dataset failed to meet this criterion. Hence, readability was operationalized in terms of four metrics, namely, FOG, CLI, ARI and FKG.

With respect to informativeness, eight POS tags that could differ between authentic and manipulative texts were identified. These include nouns, adjectives, articles, prepositions, verbs, adverbs, pronouns and personal pronouns.

With respect to writing style, affective cues were operationalized based on the proportion of both positive emotion words and negative emotion words. This was necessary because the dataset comprised both positive and negative reviews. Perceptual words were operationalized based on the proportion of words related to the senses of sight, hearing and feeling. Tense was operationalized as the proportion of past, present and future tense. For punctuations, question marks and exclamation points were considered. These were measured using the Linguistic Inquiry and Word Count (LIWC) algorithm (Pennebaker *et al.*, 2007), which is widely used for text analysis purposes such as detection of sentiment (Thelwall *et al.*, 2010; Paltoglou *et al.*, 2013) and deception (Newman *et al.*, 2003; Ott *et al.*, 2011).

Data Analysis

Data analysis includes 23 predictor variables: length of reviews in words, the four readability metrics, the eight POS tags, the two affective cues, the three indicators of perceptual words, the three types of tense and the two types of punctuations. On the other hand, the outcome variable comprises review authenticity. It was dummy-coded such that 1 (0) indicates authentic (manipulative) reviews.

There are several statistical and machine-learning approaches to address problems with a binary outcome variable. However, for datasets with sample size of 300 or more, both perform equally well (Otterbacher, 2013; Stamatatos *et al.*, 2000). Since this paper aims to identify authentic reviews based on textual characteristics rather than specifically contributing to machine learning research, the statistical approach of logistic regression was preferred for analysis.

Logistic regression converts the outcome variable into its logit equivalent and employs maximum likelihood estimation. The model can be approximated as follows:

$$\ln[P(y_i=1)/P(y_i=0)] = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n$$

where x_1 to x_n are the n predictors, and y_i is the outcome. Given a vector of n predictors describing a review, logistic regression uses the predicted likelihood to assign the vector to one of the two groups, either authentic (1) or manipulative (0), based on a threshold of 0.5. If the predicted likelihood for a review to be authentic is greater than 0.5, the review is labeled as authentic, else manipulative.

Logistic regression is however sensitive to multicollinearity. In particular, a correlation greater than 0.80 between any two predictors is problematic (Licht, 1995).) One of the predictor of a highly correlated pair can be eliminated if it is theoretically substantive (O'Brien, 2007). Based on these insights, the model was checked for multicollinearity.

Thereafter, the performance of the model was examined using Omnibus test. Pseudo-R^2 measures such as Cox and Snell R^2 (CS-R^2) and Nagelkerke R^2 (N-R^2) were checked to measure the strength of association between the predictors and the outcome. Deviance statistic defined as negative-two-log-likelihood (-2LL) was also reported. Classification accuracy of the model was reported in terms of the fraction of accurately predicted authentic reviews (APAR) and the fraction of accurately predicted manipulative reviews (APMR).

When the likelihood ratio test for a given predictor was found to be statistically significant, the estimated odds ratio (Exp(β)) was checked to identify its relationship with review authenticity. All logistic regression analyses were repeated thrice to identify authentic reviews from manipulative ones in (1) the aggregated dataset of 1,600 reviews comprising both positive and negative comments (henceforth, aggregated dataset), (2) the subset comprising only 800 positive reviews (henceforth, positive dataset), as well as (3) the subset comprising only 800 negative reviews (henceforth, negative dataset).

Results

To check for multicollinearity, the correlations among all pairs of the 23 predictors in the aggregated dataset were examined. Conceivably, the four readability metrics were strongly correlated. It would have been theoretically substantive to retain any one metric for analysis. However, Ghose and Ipeirotis (2011) advocated the use of multiple metrics to avoid idiosyncratic uniqueness specific to one. Since the correlation between FOG and CLI was less than 0.80 ($r = 0.59$, $p < 0.001$), both were included for analysis. However, ARI and FKG were dropped. Among all pairs of the revised set of 21 predictors, correlations were less than 0.80 indicating that the model was free from multicollinearity. The positive and the negative datasets with the 21 predictors were also free from the problem.

Result of the Omnibus test indicated an acceptable performance of the model for all the three datasets. Based on pseudo-R^2 and deviance statistic, the positive dataset had the highest model fitness, followed by the negative dataset and the aggregated dataset. The positive dataset also had the highest accuracy. In particular, the model accurately predicted 294 of the 400 positive authentic reviews, and 304 of the 400 positive manipulative reviews. The overall model accuracy for the dataset was 74.75%. On the other hand, the overall model accuracy of the negative and the aggregated datasets were 71.25% and 70.56% respectively. The performance of the logistic regression model across the three datasets is summarized in Table II. The odds ratios of all the textual sub-dimensions across the datasets are summarized in Table III.

Table II. Summary of the logistic regression model performance

Datasets	**Omnibus test (df = 21)**	**CS-R^2**	**N-R^2**	**-2LL**	**APAR**	**APMR**
Aggregated	$\chi^2 = 354.59$; $p < 0.001$	0.199	0.265	1863.48	554/800	575/800
Positive	$\chi^2 = 269.63$; $p < 0.001$	0.286	0.381	839.41	294/400	304/400
Negative	$\chi^2 = 190.15$; $p < 0.001$	0.212	0.282	918.88	274/400	296/400

Review Comprehensibility

With respect to comprehensibility, all textual sub-dimensions emerged as significant predictors of review authenticity in the aggregated dataset as follows: review length in words ($\text{Exp}(\beta) = 1.002$, $p < 0.01$), FOG ($\text{Exp}(\beta) = 0.944$, $p < 0.05$) and CLI ($\text{Exp}(\beta) = 0.731$, $p < 0.001$). Given that length was positively associated with review authenticity, authentic reviews emerged as being more verbose compared with manipulative reviews. Based on readability, the negative associations of the outcome with both FOG and CLI indicate that authentic reviews scored lower than manipulative reviews in terms of these metrics, and hence were more readable.

In the positive dataset, only CLI was negatively associated with review authenticity ($\text{Exp}(\beta) = 0.705$, $p < 0.001$). Authentic reviews scored significantly lower than manipulative reviews in the positive dataset, and hence were more readable.

In the negative dataset, review authenticity was significantly predicted by review length in words (Exp(β) = 1.002, $p < 0.05$) and CLI (Exp(β) = 0.757, $p < 0.01$). This indicates that authentic reviews were more verbose and readable compared with manipulative reviews.

Review Informativeness

With respect to informativeness, six POS tags, namely, articles (Exp(β) = 0.901, $p < 0.001$), prepositions (Exp(β) = 0.933, $p < 0.01$), verbs (Exp(β) = 0.813, $p < 0.01$), adverbs (Exp(β) = 0.941, $p < 0.05$), pronouns (Exp(β) = 0.929, $p < 0.05$), and personal pronouns (Exp(β) = 0.844, $p < 0.001$), were negatively associated with review authenticity in the aggregated dataset. These were used less in authentic reviews vis-à-vis manipulative ones. The proportion of nouns was however positively related to review authenticity (Exp(β) = 1.041, $p < 0.05$). Authentic reviews contained more nouns compared with manipulative reviews.

In the positive dataset, only articles (Exp(β) = 0.898, $p < 0.01$), pronouns (Exp(β) = 0.909, $p < 0.05$) and personal pronouns (Exp(β) = 0.793, $p < 0.001$) significantly predicted review authenticity. The negative associations indicate that authentic reviews contained fewer articles, pronouns and personal pronouns than manipulative reviews.

In the negative dataset, five POS tags emerged as significant predictors, namely, nouns (Exp(β) = 1.088, $p < 0.01$), articles (Exp(β) = 0.914, $p < 0.05$), prepositions (Exp(β) = 0.901, $p < 0.01$), verbs (Exp(β) = 0.695, $p < 0.001$), and personal pronouns (Exp(β) = 0.881, $p < 0.05$). Authentic reviews contained more nouns but fewer articles, prepositions, verbs and personal pronouns compared with manipulative reviews.

Review Writing Style

With respect to writing style, visual cues (Exp(β) = 0.759, $p < 0.001$), feeling cues (Exp(β) = 0.718, $p < 0.001$) and exclamation points (Exp(β) = 0.875, $p < 0.01$) were negatively associated with review authenticity in the aggregated dataset. On the other hand, aural cues (Exp(β) = 1.541, $p < 0.001$), present tense (Exp(β) = 1.153, $p < 0.05$) and question marks (Exp(β) = 1.763, $p < 0.05$) were positively associated with the outcome. Authentic reviews contained fewer visual cues, feeling cues and exclamation points but more aural cues, present tense and question marks compared with manipulative reviews.

In the positive dataset, seven textual sub-dimensions of writing style were significantly associated with review authenticity. These include positive emotion words (Exp(β) = 0.928, $p < 0.05$), negative emotion words (Exp(β) = 1.767, $p < 0.001$), visual cues (Exp(β) = 0.784, $p < 0.01$), aural cues (Exp(β) = 2.372, $p < 0.001$), feeling cues (Exp(β) = 0.713, $p < 0.01$), future tense (Exp(β) = 0.675, $p < 0.05$) and exclamation points (Exp(β) = 0.803, $p < 0.001$). Authentic reviews contained fewer positive emotion words, visual cues, feeling cues, future tense and exclamation points but more negative emotion words and aural cues compared with manipulative reviews.

In the negative dataset, five textual sub-dimensions significantly predicted review authenticity. These include negative emotion words (Exp(β) = 0.858, $p < 0.05$), visual cues (Exp(β) = 0.723, $p < 0.01$), aural cues (Exp(β) = 1.241, $p < 0.05$), feeling cues (Exp(β) = 0.758, $p < 0.05$) and present tense (Exp(β) = 1.386, $p < 0.01$). Authentic reviews contained fewer negative emotion words, visual cues and feeling words but more aural cues and present tense compared with manipulative reviews.

Table III. Odds ratios of the textual sub-dimensions

Textual characteristics	Textual sub-dimensions	Datasets		
		Aggregated	**Positive**	**Negative**
Comprehensibility	Words	1.002^{**}	1.001	1.002^{*}
	FOG	0.944^{*}	0.951	0.930
	CLI	0.731^{***}	0.705^{***}	0.757^{**}
Informativeness	Nouns	1.041^{*}	1.011	1.088^{**}
	Adjectives	1.003	1.043	0.957
	Articles	0.901^{***}	0.898^{**}	0.914^{*}
	Prepositions	0.933^{**}	0.964	0.901^{**}
	Verbs	0.813^{**}	0.959	0.695^{***}
	Adverbs	0.941^{*}	0.938	0.947
	Pronouns	0.929^{*}	0.909^{*}	0.963
	Personal pronouns	0.844^{***}	0.793^{***}	0.881^{*}
Writing style	Positive emotion words	0.994	0.928^{*}	1.090
	Negative emotion words	0.988	1.767^{***}	0.858^{*}
	Visual cues	0.759^{***}	0.784^{**}	0.723^{**}
	Aural cues	1.541^{***}	2.372^{***}	1.241^{*}
	Feeling cues	0.718^{***}	0.713^{**}	0.758^{*}
	Past tense	1.145	1.030	1.242
	Present tense	1.153^{*}	0.937	1.386^{**}
	Future tense	0.935	0.675^{*}	1.223
	Question marks	1.763^{*}	1.326	1.745
	Exclamation points	0.875^{**}	0.803^{***}	1.027

Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Discussion

Three interesting findings were gleaned from this paper. First, with respect to comprehensibility, authentic reviews were more verbose and readable vis-à-vis manipulative reviews. Prior studies found authentic comments terser than manipulative ones in mock theft experiments (Burgoon *et al.*, 2003) and financial statements (Humphreys *et al.*, 2011). In contrast, authentic reviews emerged as being more verbose than manipulative reviews. However, the finding that authentic reviews were more readable than manipulative reviews was consistent with prior studies (Daft and Lengel, 1984; Yoo and Gretzel, 2009). For example, a poorly readable manipulative review contained a long sentence of 95 words as follows: "*If you want the downtown experience of a lifetime, with historical living that will bring you back to Chicago in the early 1900's…you are sure to have your excitement whims met, and walk away with the memory of a lifetime.*"

The differences between authentic and manipulative reviews in terms of comprehensibility were more conspicuous for negative reviews compared with positive ones. For the negative dataset, length and readability in terms of CLI could significantly predict review authenticity. On the other hand, for the positive dataset, only CLI was a significant predictor.

Contrary to extant literature (Burgoon *et al.*, 2003; Humphreys *et al.*, 2011, Yoo and Gretzel, 2009), the findings suggest that length may not be a significant proxy to identify positive authentic reviews. However, negative authentic reviews could be more verbose than negative manipulative reviews. In terms of readability, CLI emerged as a better metric than FOG to identify authentic reviews from manipulative ones.

Second, with respect to informativeness, authentic reviews were generally more informative compared with manipulative reviews. Literature on text informativeness suggests that informative texts contain more nouns, adjectives, articles and prepositions, but fewer verbs, adverbs and pronouns compared with imaginative texts (Nakamura, 1991; Rayson *et al.*, 2001). The findings indicated that authentic reviews contained more nouns, as well as fewer verbs, adverbs and pronouns compared with manipulative entries. Surprisingly, the former included fewer articles and prepositions, while the use of adjectives was not a significant predictor. The presence of fewer personal pronouns in authentic reviews vis-à-vis manipulative ones is consistent with that of Yoo and Gretzel (2009).

The differences between authentic and manipulative reviews in terms of informativeness were more conspicuous for negative reviews compared with positive ones. For the negative dataset, review authenticity could be predicted by five POS tags, namely, nouns, articles, prepositions, verbs and personal pronouns. On the other hand, for the positive dataset, only three POS tags could help identify authentic reviews. Those include articles, pronouns and personal pronouns.

This suggests that differences between authentic and manipulative reviews in terms of informativeness are perhaps more easily blurred for positive reviews compared with negative ones. The dominance of personal pronouns in both positive and negative manipulative reviews connotes the lack of guilt among spammers. For instance, a manipulative review indicated, "*I was on a business trip…and I had the unfortunate luck…from the first moment I walked in my experience was poor… I had to resolve an issue…I understand…I will hopefully not be staying there again...*" Even though prior research expects spammers to use less personal pronouns to dissociate themselves from their manipulative content (Vartapetiance and Gillam, 2012; Vrij *et al.*, 2000), such a trend was inconspicuous.

Third, with respect to writing style, the use of perceptual words, tense and punctuations could offer useful cues to identify authentic reviews. Vrij *et al.* (2000) suggested that authentic reviews written after real experiences could be richer in perceptual words compared with manipulative ones. However, the findings indicated that authentic reviews contained more aural cues but fewer visual and feeling cues compared with manipulative entries. Perhaps, manipulative reviews were deliberately rendered rich in visual and feeling cues to appeal to sensory perceptions (Lin, 2004; Schiffman, 2001). In terms of tense, although authentic reviews could be written to share past experiences with hotels, it was surprising that the use of past tense did not significantly predict review authenticity. Instead, increased use of present tense was significantly associated with review authenticity. With respect to punctuations, prior studies revealed that manipulative reviews might contain more question marks and exclamation points than authentic reviews (Kim *et al.*, 2006; Zhou *et al.*, 2004). To augment extant literature, the findings indicated that authentic reviews contained more question marks and fewer exclamation points vis-à-vis manipulative ones.

The differences between authentic and manipulative reviews in terms of writing style were more conspicuous for positive reviews compared with negative ones. For the positive dataset, review authenticity could be predicted by seven indicators, namely, positive emotion words, negative emotion words, visual cues, aural cues, feeling cues, future tense and exclamation points. On the other hand, for the negative dataset, authentic reviews could be identified by only five indicators, namely, negative emotion words, visual cues, aural cues, feeling cues and present tense.

The presence of fewer positive emotion words in positive authentic reviews and fewer negative emotion words in negative authentic reviews complies with the argument that manipulative content could exaggerate the use of affective cues (Maurer and Schaich, 2011; Yoo and Gretzel, 2009). Interestingly, positive authentic reviews used more negative emotion words compared with positive manipulative reviews. For example, a positive authentic review mentioned, "*The room was not huge but there was plenty of room to move around…The bathroom was small but well appointed…*" Perhaps, positive authentic reviews maintained a reasonable tone with occasional use of negative sentiments. On the other hand, positive manipulative reviews were probably written solely to applaud, and hence, negative cues were rare. In terms of perceptual words, the findings augment studies such as Vrij *et al.* (2000) by highlighting that authentic reviews contained more aural cues but fewer visual and feeling cues compared with manipulative entries across both positive and negative comments. In terms of tense, future tense had a negative relationship with review authenticity for positive reviews while present tense showed a positive relationship with review authenticity for negative reviews. Although positive authentic reviews used fewer exclamation points compared with positive manipulative reviews as suggested in literature (Kim *et al.*, 2006), such a pattern was missing in the negative dataset. The findings in light of the proposed theoretical framework are summarized in Table IV.

Conclusion

This paper has developed a theoretical framework to identify authentic reviews from manipulative ones based on three textual characteristics, namely, comprehensibility, informativeness and writing style. The framework has been tested by drawing data from two publicly available datasets, one comprising positive reviews, and the other including negative reviews. Results indicate that the three textual characteristics offer useful insights to identify authentic reviews.

Table IV. Summary of the findings in light of the theoretical framework

	Textual sub-dimensions	**Differences between authentic and manipulative reviews**
Comprehensibility	Length	Authentic reviews were more verbose than manipulative ones. In particular, the difference was conspicuous in the negative dataset.
	Readability	Authentic reviews were more readable than manipulative ones. Specifically, CLI emerged as a better metric than FOG to identify authentic reviews across the aggregated, the positive as well as the negative datasets.
Informativeness	POS tags	Authentic reviews contained more nouns but fewer articles, prepositions, verbs, adverbs and pronouns than manipulative ones. In particular, the differences were more conspicuous in the negative dataset.
	Personal pronouns	Authentic reviews contained fewer personal pronouns than manipulative ones across the aggregated, the positive as well as the negative datasets.
Writing style	Affective cues	Although there were no differences in the aggregated dataset, authentic reviews contained fewer positive emotion words in the positive dataset. Likewise, authentic reviews contained fewer negative emotion words in the negative dataset.
	Perceptual words	Authentic reviews contained more aural cues but fewer visual and feeling cues compared with manipulative ones across the aggregated, the positive as well as the negative datasets.
	Tense	Authentic reviews contained more present tense than manipulative ones. In the positive dataset, authentic reviews contained fewer future tense than manipulative ones. In the negative dataset, authentic reviews contained more present tense than manipulative ones.
	Punctuations	Authentic reviews contained fewer exclamation points but more question marks than manipulative ones. In particular, the differences were more conspicuous in the positive dataset.

On the theoretical front, this paper is significant on two counts. First, it represents one of the earliest attempts to develop a theoretical framework to identify authentic reviews. Prior research has shed light on ways to classify reviews as authentic or manipulative. However, literature on specific differences between the two in terms of textual characteristics is relatively limited. Second, by suggesting differences between authentic and manipulative reviews across positive and negative comments, the findings offer nuanced insights into a research area that is growing in importance.

On the practical front, this paper offers implications for moderators and users of review websites. Guided by the findings, moderators can design systems to recommend reviews that are likely to be authentic, and flag those that are perhaps manipulative. Most recommendation systems work based on users' preferences (Pazzani and Billsus, 2007) or activity patterns (Taraghi *et al.*, 2013). However, they are not always tailored to filter out suspicious entries. Such an additional functionality to state-of-the-art review recommendation systems can help identify organizations that maliciously hype their own offerings, as well as those whose offerings have been slandered by their potential rivals. This in turn can help users identify reviews that are likely to be authentic. If users are able to discern between authentic and manipulative reviews, they can make better informed purchase decisions.

However, the paper is constrained by two limitations. First, the two datasets used for analysis comprised reviews for popular hotels in Chicago. Hence, caution should be exercised in generalizing the findings of this paper to reviews for less popular hotels in other geographical regions. Second, reviews often violate grammar rules by ignoring terminal punctuations or using emoticons instead of words (Petz *et al.*, 2012; 2013). Such grammatical violations were not studied.

Nevertheless, this paper serves as a springboard for further exploration on at least three disparate research strands. First, scholars interested to study cyberpsychology may want to investigate reasons for which reviewers might use more question marks but fewer visual and feeling cues in authentic reviews compared with spammers in manipulative reviews. The richness of negative emotion words in positive authentic reviews is also an idiosyncrasy, which warrants further inquiry. Second, the findings open a few research trajectories for linguists to explore. For example, the results indicated that authentic reviews, which were supposedly informative, did not comply with all POS tags as indicated in extant literature. Hence, it might be interesting to unearth if the meaning of text informativeness differs between online and offline contexts. Third, for scholars interested to study online deception, the textual characteristics of comprehensibility, informativeness and writing style could be extrapolated to identify authentic from manipulative content in social media platforms such as blogs, discussion forums or dating websites. Such studies could help verify the generalizability of the framework.

References

Anderson, E. and Simester, D. (2013), “Deceptive reviews: The influential tail”, available at: http://www.iammodern.com/webimages/business/DeceptiveReviews_Study.pdf (accessed 1 August 2013).

Ayello, E. A. (1993), “A critique of the AHCPR’s ‘Preventing pressure ulcers - A patient’s guide’ as a written instructional tool”, *Advances in Skin and Wound Care*, Vol. 6 No. 3, pp. 44-51.

Burgoon, J.K., Blair, J.P., Qin, T. and Nunamaker Jr, J.F. (2003), “Detecting deception through linguistic analysis”, in *Intelligence and Security Informatics*, Springer, Berlin, pp. 91-101.

Burgoon, J.K. and Qin, T. (2006), “The dynamic nature of deceptive verbal communication”, *Journal of Language and Social Psychology*, Vol. 25 No. 1, pp. 76-96.

Chiou, J.S., Hsu, A.C.F., and Hsieh, C.H. (2013), “How negative online information affects consumers’ brand evaluation: The moderating effects of brand attachment and source credibility”, *Online Information Review*, Vol. 37 No. 6, pp. 910-926.

Daft, R.L. and Lengel, R.H. (1984), “Information richness: A new approach to managerial behavior and organizational design”, in Cummings, L.L. and Staw, B.M. (Eds.), *Research in Organizational Behaviour*, 6, JAI, Homewood, pp. 191-233.

Dellarocas, C. (2003), “The digitization of word of mouth: Promise and challenges of online feedback mechanisms”, *Marketing Science*, Vol. 49 No. 10, pp. 1407-1424.

Duan, W., Gu, B. and Whinston, A.B. (2008), “Do online reviews matter? An empirical investigation of panel data”, *Decision Support Systems*, Vol. 45 No. 4, pp. 1007-1016.

eMarketer. (2013), “Social Media”, available at https://www.emarketer.com/Coverage/Social Media.aspx (accessed 15 January 2013).

Ghose, A. and Ipeirotis, P.G. (2011), “Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics”, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 23 No. 10, pp. 1498-1512.

Harmon, A. (2004), “Amazon glitch unmasks war of reviewers”, *The New York Times*, 14 February, available at http://www.nytimes.com/2004/02/14/us/ amazon-glitch-unmasks-war-of-reviewers.html?pagewanted=allandsrc=pm/ (accessed 26 April 2013).

Harris, C. (2012), “Detecting deceptive opinion spam using human computation”, in *AAAI Workshop on Artificial Intelligence*, AAAI Press, California, pp. 87-93.

Hayati, P. and Potdar, V. (2009), “Toward spam 2.0: An evaluation of web 2.0 anti-spam methods”, in *IEEE International Conference on Industrial Informatics*, IEEE, pp. 875-880.

Humpherys, S.L., Moffitt, K.C., Burns, M.B., Burgoon, J.K. and Felix, W.F. (2011), “Identification of fraudulent financial statements using linguistic credibility analysis”, *Decision Support Systems*, Vol. 50 No. 3, pp. 585-594.

Jindal, N. and Liu, B. (2008), “Opinion spam and analysis”, in *Proceedings of the International Conference on Web search and Web Data Mining*, ACM, New York, pp. 219-230.

Keates, N. (2007), “Deconstructing TripAdvisor”, *The Wall Street Journal*, 1 June, available at http://online.wsj.com/article/SB118065569116920710.html (accessed 25 April 2013).

Kim, S.M., Pantel, P., Chklovski, T. and Pennacchiotti, M. (2006), “Automatically assessing review helpfulness”, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, ACL, Stroudsburg, pp. 423-430.

Korfiatis, N., García-Bariocanal, E. and Sánchez-Alonso, S. (2012), “Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content”, *Electronic Commerce Research and Applications*, Vol. 11 No. 3, pp. 205-217.

Licht, M.H. (1995), “Multiple regression and correlation”, in *Reading and understanding multivariate statistics*, American Psychological Association, Washington, pp. 19-64.

Lin, I.Y. (2004), “Evaluating a servicescape: The effect of cognition and emotion”, *International Journal of Hospitality Management*, Vol. 23 No. 2, pp. 163-178.

Martin, S.S. and Camarero, C. (2009), “How perceived risk affects online buying”, *Online Information Review*, Vol. 33 No. 4, pp. 629-654.

Maurer, C. and Schaich, S. (2011), “Online customer reviews used as complaint management tool”, in *Information and Communication Technologies in Tourism*, Springer, Vienna, pp. 499-511.

Nakamura, J. (1991), “The relationship among genres in the LOB corpus based upon the distribution of grammatical tags”, *JACET Bulletin*, Vol. 22, pp. 44-74.

Newman, M.L., Pennebaker, J.W., Berry, D.S. and Richards, J.M. (2003), “Lying words: Predicting deception from linguistic styles”, *Personality and Social Psychology Bulletin*, Vol. 29 No. 5, pp. 665-675.

O’Brien, R.M. (2007), “A caution regarding rules of thumb for variance inflation factors”, *Quality & Quantity*, Vol. 41 No. 5, pp. 673-690.

Ott, M., Choi, Y., Cardie, C. and Hancock, J.T. (2011), “Finding deceptive opinion spam by any stretch of the imagination”, in *Proceedings of the Association for Computational Linguistics*, 2011, ACL, Stroudsburg, pp. 309-319.

Ott, M., Cardie, C. and Hancock, J.T. (2013), “Negative deceptive opinion spam”, in *North American Chapter of the Association for Computational Linguistics*, 2013, ACL, Stroudsburg, pp. 497-501.

Otterbacher, J. (2009), “'Helpfulness' in online communities: a measure of message quality”, in *Proceedings of the Conference on Human Factors in Computing Systems*, ACM, New York, pp. 955-964.

Otterbacher, J. (2013), “Gender, writing and ranking in review forums: a case study of the IMDb”, *Knowledge and Information Systems*, Vol. 35 No. 3, pp. 645-664.

Paltoglou, G., Theunis, M., Kappas, A. and Thelwall, M. (2013), “Predicting emotional responses to long informal text”, *IEEE Transactions on Affective Computing*, Vol. 4 No. 1, pp. 106-115.

Pazzani, M. J. and Billsus, D. (2007), “Content-based recommendation systems”, in Brusilovsky, P., Kobsa, A. and Nejdl, W. (Eds.), *The Adaptive Web*, Springer, Berlin, pp. 325-341.

Pennebaker, J.W., Chung, C.K., Ireland, M., Gonzales, A. and Booth, R.J. (2007), “The development and psychometric properties of LIWC 2007”, Austin, TX: LIWC.net.

Petz, G., Karpowicz, M., Fürschuß, H., Auinger, A., Stříteský, V. and Holzinger, A. (2013), “Opinion mining on the web 2.0 – Characteristics of user generated content and their impacts”, in Holzinger, A. and Pasi, G. (Eds.), Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data, *Lecture Notes in Computer Science*, Vol. 7947, Springer, Berlin, pp. 35-46.

Petz, G., Karpowicz, M., Fürschuß, H., Auinger, A., Winkler, S., Schaller, S. and Holzinger, A. (2012), “On text preprocessing for opinion mining outside of laboratory environments”, in Huang, R., Ghorbani, A., Pasi, G., Yamaguchi, T., Yen, N. and Jin, B. (Eds.), Active Media Technology, *Lecture Notes in Computer Science*, Vol. 7669, Springer, Berlin, pp. 618-629.

Rayson, P., Wilson, A. and Leech, G. (2001), “Grammatical word class variation within the British National Corpus sampler”, *Language and Computers*, Vol. 36 No. 1, pp. 295-306.

Schiffman, H.R. (2001), *Sensation and Perception*, 5th Edition, Wiley, New York.

Stamatatos, E., Fakotakis, N., and Kokkinakis, G. (2000), “Automatic text categorization in terms of genre and author”, *Computational Linguistics*, Vol. 26 No. 4, pp. 471-495.

Taraghi, B., Grossegger, M., Ebner, M. and Holzinger, A. (2013), “Web analytics of user path tracing and a novel algorithm for generating recommendations in Open Journal Systems”, *Online Information Review*, Vol. 37 No. 5, pp. 672-691.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D. and Kappas, A. (2010), “Sentiment strength detection in short informal text”, *Journal of the American Society for Information Science and Technology*, Vol. 61 No. 12, pp. 2544-2558.

Vartapetiance, A. and Gillam, L. (2012), “‘I don't know where he is not’: Does deception research yet offer a basis for deception detectives?”, in *Workshop on Computational Approaches to Deception Detection*, 2012, ACL, Stroudsburg, pp. 5-14.

Vrij, A., Edward, K., Roberts, K.P. and Bull, R. (2000), “Detecting deceit via analysis of verbal and nonverbal behavior”, *Journal of Nonverbal Behavior*, Vol. 24 No. 4, pp. 239-264.

Wu, G., Greene, D., Smyth, B. and Cunningham, P. (2010a), “Distortion as a validation criterion in the identification of suspicious reviews”, in *Workshop on Social Media Analytics*, ACM, New York, pp. 10-13.

Wu, G., Greene, D. and Cunningham, P. (2010b), “Merging multiple criteria to identify suspicious reviews”, in *Proceedings of the Conference on Recommender Systems*, ACM, New York, pp. 241-244.

Yoo, K.H. and Gretzel, U. (2009), “Comparison of deceptive and truthful travel reviews”, in *Information and Communication Technologies in Tourism*, Springer, Vienna, pp. 37-47.

Zhou, L., Burgoon, J.K., Twitchell, D.P., Qin, T. and Nunamaker Jr, J.F. (2004), “A comparison of classification methods for predicting deception in computer-mediated communication”, *Journal of Management Information Systems*, Vol. 20 No. 4, pp. 139-166.