Using Supervised Learning to Classify Authentic and Fake Online Reviews

Snehasish Banerjee
Wee Kim Wee School of Communication & Information, Nanyang Technological University, Singapore.
snehasis002@ntu.edu.sg

Alton Y. K. Chua
Wee Kim Wee School of Communication & Information, Nanyang Technological University, Singapore.
altonchua@ntu.edu.sg

Jung-Jae Kim
School of Computer Engineering, Nanyang Technological University, Singapore.
jungjae.kim@ntu.edu.sg

ABSTRACT

Before making a purchase, users are increasingly inclined to browse online reviews that are posted to share post-purchase experiences of products and services. However, not all reviews are necessarily authentic. Some entries could be fake yet written to appear authentic. Conceivably, authentic and fake reviews are not easy to differentiate. Hence, this paper uses supervised learning algorithms to analyze the extent to which authentic and fake reviews could be distinguished based on four linguistic clues, namely, understandability, level of details, writing style, and cognition indicators. The model performance was compared with two baselines. The results were generally promising.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing – *linguistic Processing*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *information filtering*.

General Terms

Algorithms, Management, Measurement, Reliability, Verification.

Keywords

Internet shopping; authentic online reviews; fake online reviews; linguistic clues; supervised learning.

1. INTRODUCTION

Advancement in web technologies has led to the growth of Internet shopping that enables users make online purchases [16]. Concurrently, proliferation of user-generated content such as online reviews allows users to easily access others' post-purchase experiences before making purchase decisions [8]. Users are increasingly inclined to harness reviews that are heralded to have been posted without any commercial and marketing interests.

However, not all reviews on the Internet express authentic post-purchase experiences. Some entries could be fake and written based on imagination. For example, businesses could post fake reviews to laud their products and services, as well as to criticize offerings of their competitors [18]. Besides, users could post fake reviews to gain status in the community, or simply for fun [1]. For the purpose of this paper, authentic reviews refer to entries written with post-purchase experiences. On the other hand, fake reviews refer to entries articulated based on imagination without any post-purchase experience. In particular, hotel reviews are considered in this paper given their growing popularity [31].

Since fake reviews are deliberately written to appear authentic, it is challenging for users to differentiate between the two. If users are not able to discern review authenticity, they could be misled into making sub-par purchase decisions. This in turn could jeopardize the existence of Internet shopping in the long run. Therefore, it is pertinent to devise automated strategies to distinguish between authentic and fake reviews.

Although fake reviews are not easy to identify, the ways in which they are written could offer clues to differentiate them from authentic entries. For example, authentic and fake reviews could be different in terms of their understandability. Authentic reviews based on post-purchase experiences could be more understandable than fake entries, which are hinged on imagination [31]. They could differ from each other based on the level of details. Authentic reviews could be more detailed than fake ones [14]. Writing style of authentic and fake reviews could also be different. The former could be simple while the latter might be exaggerated [33]. Moreover, they could vary in terms of cognition indicators. Since writing authentic reviews is cognitively less challenging than articulating fake entries [20], the former could contain fewer indicators of cognition [23]. Moreover, prior research suggests the possibility of using such linguistic clues in supervised learning algorithms to classify authentic and fake reviews [4, 22, 33]. Supervised learning refers to machine learning techniques in which models are trained with data containing class labels. After the model is supervised to learn, it is tested on unknown data for performance evaluation [29].

IMCOM '15, January 08 - 10 2015, BALI, Indonesia
Copyright 2015 ACM 978-1-4503-3377-1/15/01…$15.00
http://dx.doi.org/10.1145/2701126.2701130

Inspired by such state-of-the-art scholarly investigation, the goal of this paper is to analyze the extent to which authentic and fake reviews are distinguishable using supervised learning based on four linguistic clues, namely, understandability, level of details, writing style, and cognition indicators. Specifically, the research goal entails two objectives. The first is to use several supervised learning algorithms to classify authentic and fake reviews based on the linguistic clues. In particular, the following algorithms will be used: Logistic Regression, Decision Tree, Neural Network, JRip, Naïve Bayes, Random Forest, Support Vector Machine, and Voting. The second objective is to compare the classification performance of the proposed approach across all the algorithms against baselines from the literature. Specifically, two baselines were used based on [31] and [33]. The findings of this paper can serve as a significant dovetailing effort to extant literature. It can also shed light on ways fake reviews in review websites could be flagged off to allow for a healthier Internet shopping experience for users.

2. LITERATURE REVIEW

This paper proposes that authentic and fake reviews could be distinguished in terms of four linguistic clues. These include understandability, level of details, writing style, and cognition indicators. Understandability refers to the extent to which a review is clear and easy to comprehend. Authentic reviews could be more understandable than fake reviews. They could contain plain and simple arguments describing post-purchase experiences. On the other hand, fake reviews could be complex and sophisticated [31]. This is because reviews that are too lucid and simplistic are often perceived as being less credible by majority of the online community [12]. Therefore, fake reviews could be deliberately written to make ostentation of cultivated linguistic skills as a means to impress others [9]. Understandability of reviews could be ascertained based on their readability, use of familiar words, as well as surface-level characteristics such as length of words and sentences [6, 7, 12].

Level of details refers to the extent to which a review is rich in objective information. Authentic reviews based on real experiences tend to be richer in details compared with fake ones that are hinged on imagination. After all, writing fake reviews entail describing events that did not occur in reality, as well as expressing attitudes that were non-existent [20]. As a result, fake reviews could be often found wanting in terms of the level of details. Thus, while authentic reviews could include substantial objective information, fake reviews could be rich in vague and non-content words without adequate details [14, 30]. Level of details in reviews could be ascertained based on their informativeness, perceptual details, contextual details, lexical diversity, and the use of function words [4, 22, 31].

Writing style refers to the ways specific types of words are used to put forth opinions in a review convincingly. The differences between authentic and fake reviews in terms of writing style could stem from the extent of exaggeration as well as rhetorical strategies used in the entries [1, 31]. On the one hand, authentic reviews could resemble innocuous opinion sharing entries that are written without any intention to prove a point. On the other hand, fake reviews could resemble attention grabbing entries that are deliberately written to sound convincing [33]. Writing style of reviews could be ascertained based on the use of emotions, tenses, and cues of emphases [2, 28, 31].

Cognition indicators refer to linguistic clues that might leak out due to negligence in writing fake reviews. Individuals engaged in fake behavior generally get aroused both physiologically and psychologically, which is often difficult to be masked [34]. Despite attempting to conceal, the arousal often leak out in the form of linguistic clues for deception detection [30]. For example, writing fake reviews is considered cognitively demanding [20]. Given its challenges, fake reviews could contain more indicators of cognition than authentic entries [23]. Cognition indicators in reviews could be ascertained based on the use of discrepancy words such as “should” and “may”, fillers such as “you know” and “like”, tentative words such as “perhaps” and “guess”, causal words such as “because” and “hence”, insight words such as “think” and “consider”, motion words such as “arrive” and “go”, as well as exclusion words such as “without” and “except” [5, 20, 23].

3. METHODOLOGY

3.1 Data Collection

For the purpose of this paper, a gold standard dataset was created. It comprised a corpus of 900 authentic reviews, and a corpus of 900 fake reviews for 15 popular hotels in Asia. The former was collected from authenticated review websites whereas the latter was obtained from participants in a research setting.

Authentic reviews were collected from three authenticated review websites, namely, Agoda.com, Expedia.com and Hotels.com. These review websites allow reviews to be posted for a given hotel only after a stay in the hotel. In other words, it is infeasible to create fictitious accounts and post fake reviews in these websites. This assures that all reviews obtained from the three websites are authentic. Besides, the three websites are largely comparable in terms of affordances. All of these seek reviews as combinations of titles and descriptions.

Collecting authentic reviews from the three websites involved two steps. First, a set of 15 hotels that attract large volume of authentic reviews were identified. All the identified hotels had attracted about more than 1,000 reviews in Agoda.com, Expedia.com and Hotels.com collectively. This allowed for a wide scope of data collection.

Second, a total of 60 reviews were randomly collected for each of the 15 hotels (15 hotels x 60 reviews = 900). Reviews were admitted into the dataset only if they were in English, contained meaningful titles, and meaningful descriptions of at least 150 characters [22]. The set of 60 reviews collected for every hotel included 20 positive, 20 moderate, and 20 negative entries. Polarity of reviews was determined based on their ratings [11]. These two steps yielded 900 authentic reviews. Specifically, it included 300 positive, 300 moderate, and 300 negative entries.

To ensure comparability with the authentic reviews, 60 fake reviews were collected for each of the 15 hotels. These entries were solicited from participants through email instructions. Informed by studies such as [22] and [31], participants were asked to imagine as if they work for the marketing department of a hotel. Their boss had asked them to write some realistic fake reviews in English for hotels. Each review had to contain a meaningful title, and a meaningful description of at least 150 characters.

Participants were recruited for the study based on three criteria. First, they had to be in the age group of 21 to 45 years. After all, most reviews are posted by users in this age group [17]. Second, their educational profile had to be minimally undergraduate students. This is because reviews are generally posted by educated users [27]. Third, they had to be familiar with the use of reviews in review websites.

Collecting fake reviews from participants involved two steps. First, the invitation to participate in the study was disseminated using snowballing. The researchers' personal contacts were leveraged through social networking sites and word-of-mouth. When a participant volunteered to participate in the study, they were sent the detailed email instructions to write fake reviews. Participants were asked not to submit entries if they had stayed in a hotel earlier. This assures that the fake reviews obtained from participants were not unduly biased.

Second, when a participant sent the fake reviews, they were examined if they could be admitted into the dataset. If a review was in English, contained meaningful title, as well as a meaningful description of at least 150 characters long, it was included into the dataset. Otherwise, it was ignored. These two steps were iterated for the 15 hotels concurrently to yield 900 fake reviews. Specifically, it included 300 positive, 300 moderate, and 300 negative reviews, submitted by a total of 284 participants.

To sum up, the gold standard dataset contained a total of 1,800 reviews (900 authentic + 900 fake) uniformly distributed across the 15 identified hotels. For each hotel, there were 60 authentic reviews (20 positive + 20 moderate + 20 negative) and 60 fake reviews (20 positive + 20 moderate + 20 negative).

3.2 Operationalization of Linguistic Clues

As indicated earlier, this paper attempts to distinguish authentic and fake reviews based on four linguistic clues, namely, understandability, level of details, writing style, and cognition indicators. Understandability was operationalized as readability, word familiarity, and surface-level characteristics. Readability of texts was measured using six indicators, namely, Flesch-Kincaid Grade Level, Gunning-Fog Index, Automated-Readability Index, Coleman-Liau Index, Lasbarhets Index, and Rate Index. Lower values of the indicators suggest more readable reviews [12]. The six values of a given review were averaged to create a composite variable, which is henceforth referred as the mean readability index. Word familiarity was measured as the proportion of words that are easily recognized. For this purpose, words in reviews were compared against the Dale-Chall lexicon of 3,000 familiar words [7]. Structural features were calculated as the number of characters per word, number of words, fraction of words containing 10 or more characters (henceforth, long words), and number of words per sentence in reviews [6].

Level of details was operationalized as informativeness, perceptual details, contextual details, lexical diversity and function words. Informativeness was measured by examining the use of part-of-speech (POS) in reviews. Specifically, eight POS tags were considered, namely, nouns, adjectives, prepositions, articles, conjunctions, verbs, adverbs, and pronouns. The first four are higher in informative texts, while the remaining POS tags could be fewer [4, 25]. Among pronouns, the use of both first person singular and first person plural words was taken into account [14]. Perceptual details included proportion of visual, aural and feeling words, while contextual details comprised temporal and spatial references used in reviews [4]. Lexical diversity was measured based on type-to-token ratio, while function words included non-content words that reduce the level of details in reviews [4, 22, 31].

Writing style was operationalized based on the use of emotions, tenses and emphases. The use of emotions was measured in terms of reviews' emotiveness, as well as the use of positive and negative emotion words [20]. Tenses were measured as the fraction of past, present and future tense words used in reviews. Since reviews are known to influence present image and future revenues of businesses [4], fake reviews could contain fewer past tense but more present and future tense to influence the present and future reputation of hotels [28]. Use of emphases were measured based on the proportion of firm words such as "always" and "never", upper case characters, brand references (i.e., references to the hotel names), as well as use of punctuations such as ellipses "…", question marks "?", and exclamation points "!". The presence of such rhetorical cues is known to connote exaggeration [2, 28, 31].

Cognition indicators were operationalized based on the use of discrepancy, fillers, as well as tentative, causal, insight, motion and exclusion words. In particular, discrepancy, fillers and tentative words in reviews indicate uncertainty, which are often more substantial in fake reviews compared with authentic ones [34]. Fake reviews could also use more causal, insight and motion words, but fewer exclusion words than authentic entries [5, 20, 28].

As shown in Table 1, there were a total of 43 variables as features. These were measured separately for titles and descriptions of all reviews in the dataset. However, mean readability index (feature #1), number of words per sentence (feature #6), and ellipses (feature #33) were not calculated for review titles. The first two rely on the number of sentences in text. Titles of reviews do not necessarily contain full sentences. Besides, ellipses in review titles did not occur at all in the dataset. Thus, the linguistic clues were represented by a set of 83 features: 40 for review titles (43 - 3), and all the 43 for review descriptions.

To measure these 83 features, the following were utilized: Linguistic Inquiry and Word Count Algorithm [24], Stanford Parser's POS tagger [19], and some customized Java programs. The obtained feature matrix (1800 reviews x 83 features) was used as input for data analysis.

3.3 Data Analysis

Data were analyzed using supervised learning, which includes machine learning algorithms that use labeled data for training and testing [29]. Related studies have used various supervised learning algorithms. For instance, logistic regression (LogReg), C4.5 decision tree (C4.5), and back-propagation neural network (BPN) were used in [33]. Algorithms such as BPN and support vector machine with polynomial kernel (SVMp) were used in [32]. JRip, C4.5 and Naïve Bayes (NB) were used in [21]. Again, random forest classifier (RF) was used in [12], while support vector machine with linear kernel (SVM_L) was utilized in [22]. More recently, [2] used support vector machine with radial basis function kernel (SVM_{RBF}) and C4.5 for analysis. Since no single

Table 1. Operationalized features corresponding to the four linguistic clues

Linguistic Clues	Operationalized Features
Understandability	(1) Mean readability index, (2) Familiar words, (3) Characters per word, (4) Words, (5) Long words, (6) Words per sentence
Level of details	(7) Nouns, (8) Adjectives, (9) Prepositions, (10) Articles, (11) Conjunctions, (12) Verbs, (13) Adverbs, (14) Pronouns, (15) 1st person singular words, (16) 1st person plural words, (17) Visual words, (18) Aural words, (19) Feeling words, (20) Spatial words, (21) Temporal words, (22) Lexical diversity, (23) Function words
Writing style	(24) Emotiveness, (25) Positive emotion words, (26) Negative emotion words, (27) Past tense, (28) Present tense, (29) Future tense, (30) Firm words, (31) Upper case characters, (32) Brand References, (33) Ellipses, (34) Exclamation points, (35) Question marks, (36) All punctuations
Cognition indicators	(37) Discrepancy, (38) Fillers, (39) Tentative words, (40) Causal words, (41) Insight words, (42) Motion words, (43) Exclusion words

Table 2. Summary of the data analysis

Analysis	Approach	Precision	Recall	Accuracy	F_1-measure	AUC
LogReg	**Proposed**	**0.728**	**0.691**	**71.67 %**	**0.709**	**0.789**
	Baseline 1 [31]	0.639	0.603	63.17 %	0.620	0.668
	Baseline 2 [33]	0.656	0.637	64.61 %	0.646	0.700
C4.5	**Proposed**	**0.698**	**0.694**	**69.72 %**	**0.696**	**0.691**
	Baseline 1 [31]	0.620	0.538	60.44 %	0.576	0.607
	Baseline 2 [33]	0.589	0.550	58.33 %	0.569	0.574
BPN	**Proposed**	**0.667**	**0.676**	**66.89 %**	**0.671**	**0.725**
	Baseline 1 [31]	0.624	0.547	60.89 %	0.583	0.640
	Baseline 2 [33]	0.603	0.590	60.11 %	0.596	0.632
JRip	**Proposed**	**0.723**	**0.694**	**71.44 %**	**0.708**	**0.747**
	Baseline 1 [31]	0.614	0.590	61.00 %	0.602	0.627
	Baseline 2 [33]	0.614	0.590	60.94 %	0.602	0.632
NB	**Proposed**	**0.739**	**0.456**	**64.72 %**	**0.564**	**0.748**
	Baseline 1 [31]	0.665	0.489	62.11 %	0.564	0.645
	Baseline 2 [33]	0.592	0.689	60.67 %	0.637	0.667
RF	**Proposed**	**0.748**	**0.619**	**70.50 %**	**0.677**	**0.781**
	Baseline 1 [31]	0.612	0.522	59.94 %	0.563	0.637
	Baseline 2 [33]	0.648	0.534	62.22 %	0.586	0.655
SVM_L	**Proposed**	**0.700**	**0.696**	**69.89 %**	**0.698**	**0.700**
	Baseline 1 [31]	0.656	0.556	63.22 %	0.602	0.632
	Baseline 2 [33]	0.654	0.634	64.94 %	0.644	0.649
SVM_P	**Proposed**	**0.707**	**0.680**	**69.89 %**	**0.693**	**0.700**
	Baseline 1 [31]	0.672	0.580	64.83 %	0.623	0.648
	Baseline 2 [33]	0.653	0.639	64.94 %	0.646	0.649
SVM_{RBF}	**Proposed**	**0.685**	**0.671**	**68.11 %**	**0.678**	**0.681**
	Baseline 1 [31]	0.658	0.176	54.22 %	0.278	0.542
	Baseline 2 [33]	0.658	0.476	61.44 %	0.552	0.614
Voting	**Proposed**	**0.755**	**0.711**	**74.00 %**	**0.732**	**0.815**
	Baseline 1 [31]	0.682	0.556	64.83 %	0.612	0.674
	Baseline 2 [33]	0.664	0.624	65.39 %	0.643	0.700

supervised learning algorithm has been consistently superior in performance, all of these were chosen to distinguish between authentic and fake reviews. Besides, another technique was used to involve a voting among the other algorithms with average probabilities. Thus, a total of 10 supervised learning algorithms were used for analysis as follows: (1) LogReg, (2) C4.5, (3) BPN, (4) JRip, (5) NB, (6) RF, (7) SVM_L, (8) SVM_P, (9) SVM_{RBF}, and (10) Voting. These were implemented using Weka by setting all parameters to their default values [13].

The performance of the proposed approach to distinguish between authentic and fake reviews was compared against two baselines, [31] and [33]. Both the studies have greatly informed this paper. Hence, these facilitate ascertaining if the four linguistic clues proposed in this paper outperform extant approaches.

In particular, [31] suggested that descriptions of authentic and fake reviews differ in terms of the following features: (1) length of reviews in words, (2) characters per word, (3) lexical diversity, (4) first person singular words, (5) first person plural words, (6) brand references, (7) positive emotion words, and (8) negative emotion words. This set of features comprises baseline 1.

Moreover, based on [33], it appears that descriptions of authentic and fake reviews differ in terms of the following features: (1) verbs, (2) adverbs, (3) adjectives, (4) characters per word, (5) words per sentence, (6) all punctuations, (7) modal verbs, (8) first person singular words, (9) first person plural words, (10) lexical diversity, (11) emotiveness, (12) function words, (13) spatial words, (14) temporal words, (15) visual words, (16) aural words, (17) feeling words, (18) positive emotion words, and (19) negative emotion words. This set of features constitutes baseline 2.

4. RESULTS

The proposed approach, baseline 1, and baseline 2 were analyzed using the 10 selected supervised learning algorithms through five-fold cross-validation. For testing, if a fake review was correctly classified, it is termed as a true positive. If incorrectly classified, it is called a false negative. Likewise, if an authentic review was correctly classified, it is termed as a true negative. If incorrectly classified, it is called a false positive. Based on these definitions, performance was evaluated using five metrics, namely, precision, recall, accuracy, F1-measure, and area under the receiver operating characteristic curve (AUC).

Results of the analysis are summarized in Table 2. The proposed approach showed promising results. It outperformed the baselines using most supervised learning algorithms across all the five performance metrics. In particular, voting emerged as the best algorithm to distinguish between authentic and fake reviews. Using voting among the other nine algorithms, the proposed approach could correctly identify 692 of the 900 authentic reviews, and 640 of the 900 fake entries.

However, it should be acknowledged that using NB, the proposed approach did not outperform the baselines in terms of recall and F1-measure. A possible reason is that NB is known to perform poorly if the attribute independence assumption is violated [26].

5. DISCUSSION

Three key findings could be gleaned from this paper. First, understandability, level of details, writing style, and cognition indicators offer useful linguistic clues to distinguish between authentic and fake reviews. Prior studies indicated that authentic and fake reviews could be distinguished based on the ways they are written [2, 12, 14, 23, 28, 30, 34]. The results of the proposed approach to distinguish between the two in terms of the four linguistic clues therefore comply with such studies. The proposed approach was also found to outperform extant approaches such as [31] and [33]. Therefore, the set of features highlighted earlier in Table 1 appears to be more comprehensive than previous studies in distinguishing between authentic and fake reviews.

Second, even though prior studies have often attempted to distinguish between authentic and fake reviews using a huge array of features, this paper demonstrates that it is possible to develop more parsimonious supervised learning models. Studies such as [22] used all possible unigrams, bigrams and trigrams to distinguish between authentic and fake reviews. Although such approaches ensure very good performance, the findings could often be merely due to chance. To aggravate the problem, such approaches are computationally intensive. This in turn hampers storage and network resources for both training and testing of classifiers [10]. Thus, in order to distinguish between authentic and fake reviews, this paper presents a set of features that are generally more parsimonious than extant studies such as [22].

Third, titles of reviews are as useful as their descriptions to distinguish between authentic and fake reviews. Most studies on authentic and fake reviews have thus far used features based on descriptions of reviews [4, 22, 31]. This is surprising given that popular review websites such as Amazon.com, IMDb.com and TripAdvisor.com require users to post entries comprising titles as well as descriptions. In fact, titles often command greater attention from users than descriptions of reviews [3]. It is therefore conceivable that differences in titles could also offer clues to identify authentic and fake reviews. In this vein, it was found that features such as the use of exclamation points, nouns and articles in review titles had relatively high information gains to classify authentic and fake reviews (0.09, 0.03, and 0.05 respectively). Hence, future studies should take into consideration nuances in review titles in order to distinguish between authentic and fake reviews.

6. CONCLUSION

With the growth of Internet shopping, users are inclined to browse reviews before making a purchase. However, not all reviews are necessarily authentic. Some entries could be fake yet written to appear authentic. Hence, this paper used 10 supervised learning algorithms to analyze the extent to which authentic and fake reviews could be distinguished based on four linguistic clues, namely, understandability, level of details, writing style, and cognition indicators. The results were generally promising. Using most algorithms, the proposed approach outperformed the two chosen baselines [31, 33].

This paper is significant on two counts. First, it shows that authentic and fake reviews could be distinguished based on their understandability, level of details, writing style, and cognition indicators. These linguistic clues are operationalized as a set of features that is more comprehensive than studies such as [31], but more parsimonious than studies such as [22]. Furthermore, it demonstrates that authentic and fake reviews differ based on these features not only in terms of review descriptions, but also in terms of review titles. Second, this paper contributes to the machine learning research by demonstrating that problems that require addressing through supervised learning could be tackled by testing multiple classification algorithms. Most related studies thus far had confined their analyses to few algorithms [2, 12, 21, 22, 33]. Moreover, meta-classification algorithms such as voting, which emerged as the best performing algorithm in this paper, has not been widely used in related studies on text analysis. Nonetheless, it has been used in domains such as pattern recognition [15]. Hence, this paper serves as a call for the concurrent use of multiple supervised learning algorithms including voting in text mining research.

A few shortcomings of this paper should be acknowledged. For one, the scope of the dataset was limited to hotel reviews. Future research should consider examining the four linguistic clues to distinguish between authentic and fake reviews posted in evaluation of other types of products and services. Moreover, even though this paper explored a set of features to distinguish between authentic and fake reviews, it was beyond its scope to develop any new classification methods. This further offers a potential avenue for future research. Another possibility is to propose novel semi-supervised learning algorithms. Such studies could build on the findings of this paper to eventually devise automated algorithms in order to distinguish authentic from fake reviews on the Internet. By triggering such scholarly discussion, this paper hopes to contribute to a better Internet shopping experience – free from the threat of fake reviews – for users in the long run.

7. REFERENCES

[1] Anderson, E., and Simester, D. 2013. Deceptive reviews: The influential tail. Working paper, Sloan School of Management, MIT, Cambridge, http://www.poshseven.com/webimages/business/DeceptiveReviews_Study.pdf

[2] Afroz, S., Brennan, M., and Greenstadt, R. 2012. Detecting hoaxes, frauds, and deception in writing style online. In *IEEE Symposium on Security and Privacy*, 461-475. DOI = http://dx.doi.org/10.1109/SP.2012.34

[3] Ascaniis, S. D., and Gretzel, U. 2012. What's in a travel review title?. In *Proceedings of the Information and Communication Technologies in Tourism Conference*, Springer, Vienna, 494-505. DOI = http://dx.doi.org/10.1007/978-3-7091-1142-0_43

[4] Banerjee, S., and Chua, A. Y. K. 2014. A linguistic framework to distinguish between genuine and deceptive online reviews. In *Proceedings of the International Conference on Internet Computing and Web Services*, IAENG, Hong Kong, 501-506, http://www.iaeng.org/publication/IMECS2014/IMECS2014_pp501-506.pdf

[5] Boals, A., and Klein, K. 2005. Word use in emotional narratives about failed romantic relationships and subsequent mental health. *Journal of Language and Social Psychology*, 24, 252-268. DOI = http://dx.doi.org/10.1177/0261927X05278386

[6] Cao, Q., Duan, W., and Gan, Q. 2011. Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems*, 50, 511-521. DOI = http://dx.doi.org/10.1016/j.dss.2010.11.009

[7] Chall, J. S., and Dale, E. 1995. *Readability Revisited: The New Dale-Chall Readability Formula*. Brookline, Cambridge.

[8] Chua, A. Y. K., and Banerjee, S. 2014. Developing a theory of diagnosticity for online reviews. In *Proceedings of the International Conference on Internet Computing and Web Services*, IAENG, Hong Kong, 477-482, http://www.iaeng.org/publication/IMECS2014/IMECS2014_pp477-482.pdf

[9] DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. 2003. Cues to deception. *Psychological Bulletin*, 129, 74-118. DOI = http://dx.doi.org/10.1037/0033-2909.129.1.74

[10] Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. *Journal of Machine Learning Research*, 3, 1289-1305.

[11] Gerdes Jr., J., Stringam, B. B., and Brookshire, R. G. 2008. An integrative approach to assess qualitative and quantitative consumer feedback. *Electronic Commerce Research*, 8, 217-234. DOI = http://dx.doi.org/10.1007/s10660-008-9022-0

[12] 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*, 23, 1498-1512. DOI = http://dx.doi.org/10.1109/TKDE.2010.188

[13] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. 2009. The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, 11, 10-18. DOI = http://dx.doi.org/10.1145/1656274.1656278

[14] Hancock, J. T., Curry, L. E., Goorha, S., and Woodworth, M. 2008. On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, 45, 1-23. DOI = http://dx.doi.org/10.1080/01638530701739181

[15] Ho, T. K., Hull, J. J., and Srihari, S. N. 1994. Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16, 66-75. DOI = http://dx.doi.org/10.1109/34.273716

[16] Huang, C. E., Chen, P. K., and Chen, C. C. 2014. A study on the correlation between the customers' perceived risks and online shopping tendencies. In *Advances in Computer Science and its Applications*, H. Y. Jeong, M. S. Obaidat, N. Y. Yen, and J. J. Park, Eds. Springer, Berlin, 1265-1271. DOI = http://dx.doi.org/10.1007/978-3-642-41674-3_175

[17] Ip, C., Lee, H. A., and Law, R. 2012. Profiling the users of travel websites for planning and online experience sharing.

Journal of Hospitality & Tourism Research, 36, 418-426. DOI = http://dx.doi.org/10.1177/1096348010388663

[18] Kerr, D. 2013. Samsung probed for allegedly bashing rival HTC online. *CNET News*, April 15, 2013, http://news.cnet.com/8301-1035_3-57579749-94/samsung-probed-for-allegedly-bashing-rival-htc-online/

[19] Klein, D., and Manning, C. D. 2003. Accurate unlexicalized parsing. In *Annual Meeting Association for Computational Linguistics*, 423-430. DOI = http://dx.doi.org/10.3115/1075096.1075150

[20] 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*, 29, 665-675. DOI = http://dx.doi.org/10.1177/0146167203029005010

[21] O'Mahony, M. P., and Smyth, B. 2009. Learning to recommend helpful hotel reviews. In *Proceedings of the ACM Conference on Recommender systems*, ACM, New York, 305-308. DOI = http://dx.doi.org/10.1145/1639714.1639774

[22] Ott, M., Choi, Y., Cardie, C., and Hancock, J. T. 2011. Finding deceptive opinion spam by any stretch of the imagination. In *Annual Meeting of the Association for Computational Linguistics*, 309-319.

[23] Pasupathi, M. 2007. Telling and the remembered self: Linguistic differences in memories for previously disclosed and previously undisclosed events. *Memory*, 15, 258-270. DOI = http://dx.doi.org/10.1080/09658210701256456

[24] Pennebaker, J. W., Booth, R. J., and Francis, M. E. 2007. Linguistic Inquiry and Word Count: LIWC [Computer software]. Austin, TX, LIWC.net.

[25] Rayson, P., Wilson, A., and Leech, G. 2001. Grammatical word class variation within the British National Corpus sampler. *Language and Computers*, 36, 295-306.

[26] Rennie, J. D., Shih, L., Teevan, J., and Karger, D. R. 2003. Tackling the poor assumptions of naive bayes text classifiers. In *Proceedings of the International Conference on Machine Learning*, 616-623, http://www.aaai.org/Papers/ICML/2003/ICML03-081.pdf

[27] Rong, J., Vu, H. Q., Law, R., and Li, G. 2012. A behavioral analysis of web sharers and browsers in Hong Kong using targeted association rule mining. *Tourism Management*, 33, 731-740. DOI = http://dx.doi.org/10.1016/j.tourman.2011.08.006

[28] Tausczik, Y. R., and Pennebaker, J. W. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29, 24-54. DOI = http://dx.doi.org/10.1177/0261927X09351676

[29] Thakar, U., Tewari, V., and Rajan, S. 2010. A higher accuracy classifier based on semi-supervised learning. In *IEEE International Conference on Computational Intelligence and Communication Networks*, 665-668. DOI = http://dx.doi.org/10.1109/CICN.2010.137

[30] Vrij, A., Edward, K., Roberts, K. P., and Bull, R. 2000. Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, 24, 239-263. DOI = http://dx.doi.org/10.1023/A:1006610329284

[31] Yoo, K. H., and Gretzel, U. 2009. Comparison of deceptive and truthful travel reviews. In *Proceedings of the Information and Communication Technologies in Tourism Conference*, Springer, Vienna, 37-47. DOI = http://dx.doi.org/10.1007/978-3-211-93971-0_4

[32] Zheng, R., Li, J., Chen, H., and Huang, Z. 2006. A framework for authorship identification of online messages: Writing-style features and classification techniques. *Journal of the American Society for Information Science and Technology*, 57, 378-393. DOI = http://dx.doi.org/10.1002/asi.20316

[33] 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*, 20, 139-165.

[34] Zuckerman, M., DePaulo, B. M., and Rosenthal, R. 1981. Verbal and nonverbal communication of deception. In *Advances in Experimental Social Psychology*, L. Berkowitz, Ed., Academic Press, New York, 1-59.