Distinguishing between authentic and fictitious user-generated hotel reviews

Banerjee, Snehasish; Chua, Alton Y. K.; Jung-Jae Kim

2015

Banerjee, S., Chua, A. Y. K., & Jung-Jae Kim. (2015). Distinguishing between authentic and fictitious user-generated hotel reviews. 2015 6th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 1-7.

https://hdl.handle.net/10356/82626

https://doi.org/10.1109/ICCCNT.2015.7395179

Distinguishing between Authentic and Fictitious User-generated Hotel Reviews

Snehasish Banerjee, Alton Y. K. Chua
Wee Kim Wee School of Communication & Information
Nanyang Technological University
Singapore
{snehasis002, altonchua}@ntu.edu.sg

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

***Abstract*—The objective of this paper is to distinguish between authentic and fictitious user-generated hotel reviews. To achieve this objective, it adopts a two-step approach. The first seeks to classify authentic and fictitious reviews by leveraging on their possible textual differences. The second step attempts to identify the textual traits that are unique to authentic and fictitious reviews. For the purpose of this paper, a ground truth dataset of 1,800 reviews, uniformly divided between authentic and fictitious, was created. With respect to the first step, authentic and fictitious reviews were classified by using four forms of textual differences: understandability, level of details, writing style, and cognition indicators. Classification was performed using voting by average probability among logistic regression, C4.5, Support Vector Machine, JRip, and Random Forest classifiers. Using five-fold cross-validation, the proposed approach was found to outperform two existing baselines. Furthermore, with respect to the second step, the textual traits unique to authentic and fictitious reviews were identified using Information Gain, and Chi-squared feature selection techniques. A sequential forward feature selection approach was further adopted to identify the top five features that aid the classification of authentic and fictitious reviews. These include the use of nouns, articles, function words, punctuations, and in particular, exclamation points in reviews. The implications of the results are discussed.**

Keywords—text analysis, machine learning, data mining, classification algorithms

I. Introduction

The growing popularity of social media allows anyone with Internet access to easily create user-generated content on the web. A key form of user-generated content includes hotel reviews. Commonly posted in websites such as TripAdvisor.com, hotel reviews allow users to share post-trip experiences of staying in a hotel with other potential travellers in the online community.

Travellers increasingly rely on such user-generated hotel reviews (henceforth, reviews) to book their accommodation. They assume reviews to be more authentic than content produced by advertising agencies that could present hotels by concealing negative aspects [1]. In contrast, reviews written by fellow travellers are perceived as honest disclosures of hotels' strengths and weaknesses.

Such a perception is however questionable because not all reviews could be blindly trusted as authentic accounts of post-trip experiences. Some biased reviews could be maliciously written by marketing professionals to distort travellers' perceptions about hotels [2]. Yet other misleading reviews could be benignly written by users simply to poke fun and pass time [3]. Since such fictitious entries are intentionally written to be camouflaged as authentic, it is challenging for travellers to distinguish between the two. For the purpose of this paper, authentic reviews refer to those written after a stay in a given hotel. In contrast, fictitious reviews refer to those written out of imagination without staying in the specific hotel.

Although authentic and fictitious reviews are generally difficult to be distinguished, their textual characteristics could set them apart [4, 5, 6]. Yet, a comprehensive understanding of textual differences between authentic and fictitious reviews is missing in extant literature. Hence, this paper aims to distinguish between authentic and fictitious reviews using a two-step approach. The first seeks to classify authentic and fictitious reviews by leveraging on their possible textual differences. To expand the boundaries of the literature, this paper strives to identify textual differences more comprehensively compared with prior studies such as [5]. Thereafter, to delve deeper, the second step attempts to identify the textual traits that are unique to authentic and fictitious reviews. The findings might help various stake-holders associated with review websites in sifting grain from chaff among the plethora of posted reviews.

The remainder of the paper is organized as follows. The next section explains the contributions of the current paper in relation to prior studies. This is followed by the methods of data collection, measurement and analysis. The results pertaining to the two-fold objectives are presented next, followed by the discussion and the conclusion.

II. Relation to Prior Studies

To classify authentic and fictitious reviews, scholars in recent years have often relied on their underlying textual differences. For instance, studies such as [5] showed that authentic and fictitious reviews differ in terms of textual characteristics such as number of characters per word, and the use of pronouns as well as sentiments. Studies such as [6] suggested additional textual characteristics such as the use of various part-of-speech (POS) tags, lexical diversity, function words and punctuations. More recently, studies such as [4] used numerous textual characteristics as well as n-grams. In particular, a set of some 81 textual characteristics featured in the Linguistic Inquiry and Word Count (LIWC) dictionary was found to perform well in classifying authentic and fictitious reviews [7]. Among n-grams, bigrams were found extremely useful to classify the two [4].

Despite such scholarly efforts, extant literature lacks a comprehensive understanding of textual differences between authentic and fictitious reviews. This is because on the one hand, studies such as [5] investigated differences between the two across a limited set of textual characteristics. On the other hand, studies such as [4] relied on an innumerable set of textual characteristics, some of which might lack robust theoretical reasoning, thus standing the chance to influence classification performance fortuitously.

This paper contributes to the literature by differing from the prior studies in at least two ways. First, it identifies a set of textual characteristics, more comprehensive than studies such as [5] but more parsimonious than studies such as [4], in order to classify authentic and fictitious reviews. Specifically, it proposes that authentic and fictitious reviews could be classified based on four broad textual characteristics: understandability, level of details, writing style, and cognition indicators. For one, authentic reviews based on post-trip experiences could be more understandable than fictitious reviews, which are written based on imagination [8]. Given the real experiences of staying in a hotel, the former could be more detailed than the latter [9]. Writing style of authentic reviews could be less exaggerated compared with that of fictitious entries [10]. Furthermore, given that writing authentic reviews is cognitively easier than articulating fictitious entries [8], the two could differ in terms of cognition indicators [11].

Second, this paper relies on both titles and descriptions of reviews to classify authentic and fictitious entries. Thus far, investigation of authentic and fictitious reviews has been largely confined to descriptions of reviews even though most popular review websites such as TripAdvisor.com and Expedia.com require users to post entries comprising both titles and descriptions. In fact, titles often command greater attention than descriptions of reviews [12]. It is hence conceivable that the textual characteristics of titles might also offer hints to predict if reviews are authentic or fictitious. Therefore, this paper leverages on the four textual characteristics, namely, understandability, level of details, writing style, and cognition indicators, of titles as well as descriptions to classify authentic and fictitious reviews. Thereafter, it shows that the proposed approach offers promising classification performance compared with two existing baselines, namely, LIWC and bigrams. Finally, using feature selection techniques, the paper further identifies textual traits that are unique to titles and descriptions of authentic as well as fictitious reviews.

III. METHODS

A. *Data Collection*

The ground truth created for this paper comprised 900 authentic reviews and 900 fictitious reviews. Specifically, the authentic reviews were collected from three authenticated 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. This might assure that all reviews are authentic. In contrast, the fictitious reviews were solicited from participants in a research setting. They had no experience of staying in the hotels for which they were asked to submit entries. This ensures the validity of the ground truth.

For the purpose of the data collection, a set of 15 hotels in Asia that attract large volumes of reviews in Agoda.com, Expedia.com and Hotels.com were identified. All the identified hotels had cumulatively attracted more than 1,000 reviews across the three websites. This offered a large initial pool of reviews to admit into the dataset.

For each of the 15 hotels, a total of 60 authentic reviews were randomly collected (15 x 60 = 900 authentic reviews). Reviews were admitted into the dataset only if they were in English, contained meaningful titles, and meaningful descriptions of at least 150 characters. Meaningfulness of titles and descriptions were inspected manually to ensure that there were no linguistically nonsensical comments. The final 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 star ratings [13]. This yielded 900 authentic reviews (300 positive + 300 moderate + 300 negative).

Likewise, for each of the 15 hotels, a total of 60 fictitious reviews were solicited from about 300 participants through email instructions. They were recruited based on convenient sampling. Care was taken to ensure that their profile was similar to those who are likely to write reviews on the Internet. Specifically, all participants were in the age group between 21 to 45 years with an educational profile of minimally undergraduate students. They were also familiar with the use of reviews, and had travel experience within the last year. For the purpose of snowballing, these 300 participants were further requested to disseminate the study invitation to their contacts, who would fit the above eligibility criteria.

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. Similar methodology has been followed in prior works such as [5]. For the purpose of this study, some participants were asked to submit positive reviews. Others were required to submit moderate reviews, while the rest needed to submit negative reviews. Each review had to contain a meaningful title, and a meaningful description of at least 150 characters. However, participants were asked not to submit entries if they had stayed in a hotel earlier. Finally, 900 fictitious reviews (300 positive + 300 moderate + 300 negative) submitted by some 284 participants were admitted into the dataset.

B. Features Measuring Textual Characteristics

Understandability – conceived as readability, word familiarity, and surface-level characteristics – was measured as follows: Readability was calculated using the average of six indicators, namely, Flesch-Kincaid Grade Level, Gunning-Fog Index, Automated-Readability Index, Coleman-Liau Index, Lasbarhets Index, and Rate Index. Lower value of the average would indicate more readable reviews and vice-versa [14]. To calculate the fraction of familiar words in reviews, every word was compared against Dale-Chall's list of 3,000 familiar words [15]. Surface-level characteristics included 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 [16].

Level of details entailed informativeness, perceptual and contextual details, lexical diversity, and function words. Informativeness was measured based on the use of eight POS tags, 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 [17]. Among pronouns, the use of both first person singular and first person plural words was taken into account separately [9]. Perceptual details included proportion of visual, aural and feeling words, while contextual details comprised temporal and spatial words. Lexical diversity was measured based on type-to-token ratio, while function words was calculated by examining the use of non-content words that reduce the level of details [9], [18].

Writing style – conceived as the use of emotions, tenses, and emphases – was measured as follows: The use of emotions was calculated in terms of reviews' emotiveness, as well as the use of positive and negative emotion words [8]. Tenses were examined as the fraction of past, present and future tense words used in reviews. The use of emphases was measured based on the proportion of firm words such as "always" and "never", upper case characters, references to the hotel brand names, as well as punctuations such as ellipses, exclamation points, and question marks. After all, such rhetorical strategies are known to express hyperboles and emphases [5, 18].

Cognition indicators were measured as the fraction of discrepancy, fillers, as well as tentative, causal, insight, motion and exclusion words used in reviews. Except exclusion words, the other word categories could be fewer in authentic reviews compared with fictitious entries. Such differences generally stem from the higher cognition load associated with writing fictitious reviews vis-à-vis authentic ones [8], [18, 19].

There were collectively a total of 43 classification features (Table I), which were calculated using Stanford Parser's POS tagger, LIWC, and some customized Java programs. In particular, each feature was measured separately for titles and descriptions of all reviews in the dataset. However, average readability (feature #1), number of words per sentence (feature #6), and ellipses (feature #33) were not calculated for titles. The first two rely on the number of sentences in text. However, review titles do not necessarily contain sentences. Besides, ellipses in review titles did not occur at all in the dataset. Thus, reviews' textual characteristics were measured using a set of 83 features: 40 for review titles, and all the 43 for review descriptions.

By measuring the variables corresponding to review titles, a reviews-by-title features matrix (1800 x 40) was created. Likewise, by measuring the variables corresponding to review descriptions, a reviews-by-description features matrix (1800 x 43) was created. The two matrices were horizontally concatenated to form the reviews-by-features matrix (1800 reviews x 83 variables), which was used for subsequent analysis.

TABLE I. FEATURES FOR THE TEXTUAL CHARACTERISTICS

Characteristics	Operationalized Features
Understandability	(1) Average readability, (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, (16) 1st person plural, (17) Visual, (18) Aural, (19) Feeling, (20) Temporal, (21) Spatial, (22) Type-to-token ratio, (23) Function words
Writing style	(24) Emotiveness, (25) Positive emotions, (26) Negative emotions, (27) Past, (28) Present, (29) Future, (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, (40) Causal, (41) Insight, (42) Motion, (43) Exclusion words

C. Analytical Approaches

As indicated earlier, this paper uses a two-step approach to distinguish between authentic and fictitious reviews. The first seeks to classify authentic and fictitious reviews by leveraging on their textual differences, while the second attempts to identify the textual traits unique to authentic and fictitious reviews. For the first step, authentic and fictitious reviews were classified using voting by average probability among logistic regression, C4.5, Support Vector Machine, JRip, and Random Forest classifiers. These five classification techniques were involved in the voting because they have been used in similar studies [4, 6, 14, 20]. However, they have seldom been combined using voting hitherto. Classification results were assessed using five-fold cross-validation. Performance was evaluated using six metrics: precision, recall, specificity, accuracy, F_1-measure, and AUC.

For the second step, the textual characteristics unique to authentic and fictitious reviews were identified based on two feature selection techniques: Information Gain (IG), and Chi-squared (χ^2). After feature selection, the classification using voting was repeated only with variables having non-zero IG and χ^2 values. Finally, among the features with relatively high IG and χ^2 values, a sequential forward feature selection was performed to identify the top five features to classify authentic and fictitious reviews [21].

Besides, the classification performance of the proposed approach was compared with that of two baselines. The first involved features corresponding to review descriptions for the 81 textual characteristics offered by the LIWC dictionary [7]. The second baseline involved bigrams as features. After lowercasing and Porter stemming, length normalized bigrams were computed for all review descriptions using the WEKA toolkit [22]. In order to enhance computational efficiency, bigrams were retained only if they occurred at least three times in authentic and fictitious reviews. A total of 3,878 such unique bigrams were found.

These baselines were chosen due to two reasons. First, both are known to perform well in distinguishing between authentic and fictitious reviews [4, 23, 24]. Thus, these baselines facilitate ascertaining if the proposed approach outperforms existing solutions. Second, both the baselines relied only on review descriptions. If the proposed approach outperforms the baselines, it would demonstrate the relevance of review titles to distinguish between authentic and fictitious reviews.

IV. RESULTS

With respect to the first step of analysis using classification, the textual characteristics of understandability, level of details, writing style and cognition indicators together could accurately identify 714 of the 900 fictitious reviews. On the other hand, 677 of the 900 authentic reviews were accurately classified, thus yielding an accuracy of 77.28%. This suggests that the proposed approach offers reasonable classification performance. To draw greater insights, it was necessary to identify the textual traits that are useful to distinguish between authentic and fictitious reviews. This called for the second step of the analytical approach. With respect to the second step using feature selection, 41 of the 83 features were found to have non-zero IG and χ^2 values (Table II). These represent the textual traits that are unique for authentic and fictitious reviews.

The classification performance of the proposed approach superseded that of the baselines (Table III), both with and without feature selection. It is noteworthy that the 41 feature-selected textual characteristics of the proposed approach demonstrate superior performance over the bigrams-baseline, which included as many as 3,878 features for classification.

To delve deeper, 10 features with relatively high IG as well as high χ^2 values were admitted for sequential forward feature selection (cf. Table II). The ten admitted features are as follows: (1) nouns in titles [higher among authentic reviews]; (2) articles in titles [higher among fictitious reviews]; (3) function words in titles [higher among fictitious reviews]; (4) exclamation points in titles [higher among fictitious reviews]; (5) all punctuations in titles [higher among fictitious reviews]; (6) pronouns in descriptions [higher among fictitious reviews]; (7) first person singular words in descriptions [higher among fictitious reviews]; (8) past tense in descriptions [higher among fictitious reviews]; (9) firm words in descriptions [higher among fictitious reviews]; and (10) upper case characters in descriptions [higher among fictitious reviews]. An example of an authentic review title that was rich in nouns pointed, "*…location…thumbs up…value for money…*" Fictitious reviews rich in punctuations, pronouns as well as past tense used titles such as "*Overpriced!!!!!*" and descriptions such as "*…we were delighted…I was surprised…*"

Conceivably, the features with high IG and χ^2 values are considered superior over those with low values. However, IG and χ^2 rank features based on the data characteristics only, independently from the classifier. Hence, to corroborate these rankings, feature selection was further conducted on the 10 features with relatively high IG as well as high χ^2 values by taking the classifier into account through what is known as a wrapper approach [25]. Specifically, a sequential forward feature selection was adopted to identify the top five features from among the 10 with relatively high IG and χ^2 values [21]. It started with an empty set. In every iteration, it collected all possible feature subsets containing the best feature combination obtained in the previous iteration plus one additional feature. The classification model was trained and evaluated using voting based on each subset. The subset that yielded the best classification accuracy was selected. The process was iterated five times to identify the top five features among those ten.

Using the sequential forward feature selection, the following features were identified in its five iterations respectively: (1) all punctuations in titles, (2) exclamation points in titles, (3) articles in titles, (4) nouns in titles, and (5) function words in descriptions. In other words, these represent the top five features that help to classify authentic and fictitious reviews. These five features together yielded an accuracy of 69.94%. Titles of fictitious reviews contained more punctuations such as exclamation points, more articles but fewer nouns compared with titles of authentic entries. Descriptions of fictitious reviews were richer in function words compared with those of authentic ones. Interestingly, four of the top five features are associated with review titles. It appears that review titles could be more eloquent than descriptions in offering clues to distinguish between authentic and fictitious reviews.

V. DISCUSSION AND CONCLUSION

Two main findings are gleaned from the results. First, authentic and fictitious reviews differ substantially in terms of the textual characteristics of understandability, level of details, writing style, and cognition indicators. Multiple features corresponding to each of the four textual characteristics were found to play a significant role in the classification of authentic and fictitious reviews (cf. Table II). This finding complies with several prior studies over the years that suggested the presence of potential textual differences between authentic and fictitious information [4, 5, 6, 26, 27, 28]. In a recent literature survey paper along this research theme, it is no wonder why textual content of reviews was identified as "*the first thing to be considered in spam detection practice*" [29, p. 3635]. Even though it is commonly believed that spammers are increasingly getting efficient in blurring the lines between authentic and fictitious reviews, the differences are still substantial enough to set the two apart. This is an encouraging sign for research on review spam because it highlights that the community of spammers is yet to catch up with the scholarly community. Nonetheless, it is important for future studies to keep on uncovering even more textual differences between authentic and fictitious reviews so that spammers are never able to fully blur the underlying textual nuances.

Second, titles of reviews offer useful clues to distinguish between authentic and fictitious entries. As indicated earlier, four of the top five features that helped to classify authentic and fictitious reviews measured textual properties of review titles. Specifically, they included the use of punctuations such as exclamation points, as well as part-of-speech tags such as articles and nouns. Moreover, using both titles and descriptions to classify authentic and fictitious reviews, the approach proposed in this paper managed to outperform two existing baselines, which relied on only review descriptions. The need to analyze reviews both in terms of titles and descriptions could be justified through the lens of the relevance theory [30, 31], which suggests that individuals interpret information through a tension between retrieving contextual details and using cognitive efforts. Especially when information is presented in two parts: titles and descriptions, the former serves as relevance optimizers for the latter [32]. In other words, titles often dictate the extent to which individuals pay attention to review descriptions by playing a role similar to headlines in newspapers or taglines in advertisements [12, 31, 32]. Hence, it offers adequate motivation for spammers to manipulate not only descriptions but also titles of fictitious reviews. This finding suggests that future research on review spam detection should leverage on the textual characteristics of both review titles and review descriptions for comprehensive investigation.

TABLE II. FEATURE SELECTION ACROSS TITLES AND DESCRIPTIONS OF REVIEWS

Characteristics	**Features**	**Review Titles**		**Review Descriptions**	
		IG	χ^2	**IG**	χ^2
Understandability	Average readability[a]	*NA*		0.009	18.407
	Familiar words	0	0	0	0
	Characters per word	0	0	0.012	27.883
	Words	0.020	50.578	0	0
	Long words	0.004	9.203	0.010	25.233
	Words per sentence[a]	*NA*		0.013	27.920
Level of details	Nouns	0.028	67.980	0.019	48.134
	Adjectives	0	0	0	0
	Prepositions	0	0	0	0
	Articles	0.051	103.332	0.014	35.736
	Conjunctions	0.008	20.847	0	0
	Verbs	0	0	0.008	21.191
	Adverbs	0	0	0	0
	Pronouns	0.009	22.361	0.026	64.405
	1st person singular	0.014	32.466	0.035	86.309
	1st person plural	0	0	0	0
	Visual	0	0	0	0
	Aural	0	0	0	0
	Feeling	0	0	0	0
	Temporal words	0	0	0	0
	Spatial words	0.016	39.583	0.013	32.189
	Type-to-token ratio	0	0	0	0
	Function words	0.046	104.965	0.016	38.755
Writing style	Emotiveness	0.020	44.565	0.010	24.032
	Positive emotions	0.015	37.267	0	0
	Negative emotions	0.015	37.143	0.010	24.224
	Past	0	0	0.036	81.973
	Present	0	0	0	0
	Future	0.007	15.373	0.005	12.027
	Firm words	0.010	25.186	0.021	52.074
	Upper case characters	0.020	50.000	0.068	151.280
	Brand References	0.003	6.897	0.009	22.636
	Ellipses[a]	*NA*		0	0
	Exclamation points	0.093	200.547	0.008	19.112
	Question marks	0.005	9.045	0	0
	All punctuations	0.107	234.224	0	0
Cognition indicators	Discrepancy	0.008	19.663	0	0
	Fillers	0.004	7.027	0	0
	Tentative words	0	0	0	0
	Causal words	0	0	0	0
	Insight words	0	0	0	0
	Motion words	0	0	0	0
	Exclusion words	0	0	0.012	30.291

[a] Features computed only for review descriptions. The corresponding IG and χ^2 values for review titles are hence denoted as *Not Applicable (NA)*.

TABLE III. CLASSIFICATION PERFORMANCE

Approach	Precision	Recall	Specificity	Accuracy	F_1-measure	AUC
Proposed	0.784	0.752	0.793	77.28 %	0.768	0.853
Proposed (feature selected)	0.785	0.746	0.796	77.06 %	0.765	0.847
LIWC baseline	0.713	0.704	0.717	71.06 %	0.708	0.767
Bigrams baseline	0.773	0.714	0.790	75.22 %	0.742	0.839

Despite the insights offered in this paper, it is constrained by two limitations. First, its scope is trained on authentic and fictitious reviews in English language for hotels, thereby hindering the scope for generalizability. For the purpose of this paper, English was chosen because it is the most popular language used on the Internet, supported by some 536.6 million (26.8 %) users [33]. Moreover, reviews for hotels were considered as the test case for investigation because they are not only growing in popularity [34, 35], but also represent a viable dataset for analysis [4, 5]. Nonetheless, investigation of non-English reviews or entries for other products and services could be a possible extension of this work.

Second, the focus of this research was to distinguish between authentic and fictitious reviews rather than truthful and deceitful entries. For the purpose of this paper, authentic reviews were defined as those contributed in authenticated review websites such as Agoda.com, Expedia.com and Hotels.com by users who had visited a given hotel. On the other hand, fictitious reviews refer to those written by participants in a research setting without the experience of staying in the hotel. Hence, authentic reviews in the context of this paper should not be misinterpreted as truthful. Likewise, fictitious reviews are not necessarily deceitful. New methodological approaches are needed to understand nuances between authentic and truthful reviews, as well as between what is fictitious and deceitful.

Additional research is needed to address these limitations. Furthermore, this paper calls for future studies to rely on semi-supervised learning algorithms to distinguish between authentic and fictitious reviews [36, 37, 38]. This is especially important given the difficulties associated with obtaining annotated datasets in this field. More scholarly efforts are needed to further improve the classification performance. Moreover, future studies need to offer probabilistic estimates of the likely distributions for authentic and fictitious reviews in the real world. The dataset in this paper used the proportion of 50% authentic and 50% fictitious, which is certainly informed by prior research [4, 39]. Nonetheless, if the proportions are unbalanced in real world settings, the extent to which the present approach of analysis might work remains to be seen. To mitigate the problems posed by overly unbalanced datasets, subsequent studies might experiment with resampling approaches such as over-sampling the minority class, or under-sampling the majority class [40, 41, 42]. Besides, future research also needs to investigate the extent to which Internet users are able to discern the authenticity of reviews. By offering these directions for interested scholars, this paper hopes to further expand the boundaries of this research arena.

REFERENCES

[1] A. Papathanassis, and F. Knolle, “Exploring the adoption and processing of online holiday reviews: A grounded theory approach,” Tourism Management, vol. 32, pp. 215-224, April 2011.

[2] E. Allen, “‘Dear staff. We could do with some positive comments’: Hotel boss is caught telling his workers to post fake reviews on TripAdvisor,” MailOnline (2012, October 9). Retrieved February 27, 2013 from http://www.dailymail.co.uk/news/article-2214974/Hotel-boss-caught-telling-workers-post-fake-reviews-TripAdvisor.html

[3] F. Bronner, and R. de Hoog, “Vacationers and eWOM: Who posts, and why, where, and what?,” Journal of Travel Research, vol. 50, pp. 15-26, January 2011.

[4] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, “Finding deceptive opinion spam by any stretch of the imagination,” Proceedings of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 309-319, 2011, Stroudsburg, PA: ACL.

[5] K. H. Yoo, and U. Gretzel, “Comparison of deceptive and truthful travel reviews,” In: Information and Communication Technologies in Tourism, pp. 37-47, 2009, Vienna: Springer.

[6] L. Zhou, J. K. Burgoon, D. P. Twitchell, T. Qin, and J. F. Nunamaker Jr, “A comparison of classification methods for predicting deception in computer-mediated communication,” Journal of Management Information Systems, vol. 20, pp. 139-165, Spring 2004.

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

[8] M. L. Newman, J. W. Pennebaker, D. S. Berry, and J. M. Richards, “Lying words: Predicting deception from linguistic styles,” Personality and Social Psychology Bulletin, vol. 29, pp. 665-675, May 2003.

[9] J. T. Hancock, L. E. Curry, S. Goorha, and M. Woodworth, “On lying and being lied to: A linguistic analysis of deception in computer-mediated communication,” Discourse Processes, vol. 45, pp. 1-23, March 2008.

[10] E. Anderson, and D. Simester, “Deceptive reviews: The influential tail,” Working paper, Sloan School of Management, Cambridge, MA, 2013. Retrieved March 3, 2014 from http://www.iammodern.com/webimages/business/DeceptiveReviews_Study.pdf

[11] M. Pasupathi, “Telling and the remembered self: Linguistic differences in memories for previously disclosed and previously undisclosed events,” Memory, vol. 15, pp. 258-270, April 2007.

[12] S. D. Ascaniis, and U. Gretzel, “What’s in a travel review title?,” In: Information and Communication Technologies in Tourism, pp. 494-505, 2012, Vienna: Springer.

[13] J. Gerdes Jr., B. B. Stringam, and R. G. Brookshire, “An integrative approach to assess qualitative and quantitative consumer feedback,” Electronic Commerce Research, vol. 8, pp. 217-234, December 2008.

[14] A. Ghose, and P. G. Ipeirotis, “Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics,” IEEE Transactions on Knowledge and Data Engineering, vol. 23, pp. 1498-1512, August 2011.

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

[16] Q. Cao, W. Duan, and Q. Gan, “Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach,” Decision Support Systems, vol. 50, pp. 511-521, January 2011.

[17] P. Rayson, A. Wilson, and G. Leech, “Grammatical word class variation within the British National Corpus sampler,” Language and Computers, vol. 36, pp. 295-306, November 2001.

[18] Y. R. Tausczik, and J. W. Pennebaker, “The psychological meaning of words: LIWC and computerized text analysis methods,” Journal of Language and Social Psychology, vol. 29, pp. 24-54, March 2010.

[19] A. Boals, and K. Klein, “Word use in emotional narratives about failed romantic relationships and subsequent mental health,” Journal of Language and Social Psychology, vol. 24, pp. 252-268, September 2005.

[20] A. Pal, “Design and development of a sentiment analysis system using supervised learning,” Master’s thesis, Jadavpur University, India (2014). Retrieved February 2015 from http://dspace.jdvu.ac.in/handle/123456789/29136

[21] D. Kollmar, and D. H. Hellmann, “Feature selection for a real-world learning task,” In: Machine Learning and Data Mining in Pattern Recognition, pp. 157-172, 2001, Springer Berlin Heidelberg.

[22] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software: An update,” ACM SIGKDD Explorations Newsletter, vol. 11, pp. 10-18, June 2009.

[23] N. Jindal, and B. Liu, “Opinion spam and analysis,” Proceedings of the International Conference on Web search and Web Data Mining, pp. 219-230, 2008, New York: ACM.

[24] R. Mihalcea, and C. Strapparava, “The lie detector: Explorations in the automatic recognition of deceptive language,” Proceedings of the International Joint Conference on Natural Language Processing, pp. 309-312, 2009, Stroudsburg, PA: ACL.

[25] S. Shilbayeh, and S. Vadera, “Feature selection in meta-learning framework,” IEEE Science and Information Conference, pp. 269-275, 2014.

[26] M. K. Johnson, and C. L. Raye, “Reality monitoring,” Psychological Review, vol. 88, pp. 67-85, January 1981.

[27] U. Undeutsch, Forensische psychologie [Forensic psychology]. Gottingen, Germany: Verlag fur Psychologie, 1967.

[28] A. Vrij, K. Edward, K. P. Roberts, and R. Bull, “Detecting deceit via analysis of verbal and nonverbal behavior,” Journal of Nonverbal Behavior, vol. 24, pp. 239-263, December 2000.

[29] A. Heydari, M. ali Tavakoli, N. Salim, and Z. Heydari, “Detection of review spam: A survey,” Expert Systems with Applications, vol. 42, pp. 3634-3642, May 2015.

[30] D. Sperber, and D. Wilson, Relevance: Communication and Cognition. Oxford: Blackwell, 1986.

[31] D. Sperber, and D. Wilson, “Précis of relevance: Communication and cognition,” Behavioral and Brain Sciences, vol. 10, pp. 697-710, December 1987.

[32] D. Dor, “On newspaper headlines as relevance optimizers,” Journal of Pragmatics, vol. 35, pp. 695-721, May 2003.

[33] Internet World Stats, “Internet world users by language: Top 10 languages,” 2013. Retrieved February 28, 2013 from http://www.internetworldstats.com/stats7.htm

[34] H. Chipkin, “Consumer trends 2012: Review websites,” Travel Weekly (2012, July 31). Retrieved October 23, 2013 from http://www.travelweekly.com/Travel-News/Online-Travel/Consumer-Trends-2012-Review-websites

[35] O. Kwon, and Y. Sung, “Shifting selves and product reviews: How the effects of product reviews vary depending on the self-views and self-regulatory goals of consumers,” International Journal of Electronic Commerce, vol. 17, pp. 59-82, Fall 2012.

[36] X. Zhu, J. Lafferty, and Z. Ghahramani, “Combining active learning and semi-supervised learning using gaussian fields and harmonic functions,” Workshop on the Continuum from Labeled to Unlabeled Ddata in Machine Learning and Data Mining, pp. 58-65, 2003.

[37] V. Sindhwani, and S. S. Keerthi, “Large scale semi-supervised linear SVMs,” Proceedings of the Annual International Conference on Research and Development in Information Retrieval, pp. 477-484, 2006, New York: ACM.

[38] D. H. Fusilier, R. G. Cabrera, M. M. Gomez, and P. Rosso, “Using PU-learning to detect deceptive opinion spam,” Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 38-45, 2013, Stroudsburg, PA: ACL.

[39] C. G. Harris, “Detecting deceptive opinion spam using human computation,” Workshop on Artificial Intelligence, pp. 87-93, 2012, California: AAAI.

[40] E. M. Karabulut, and T. Ibrikci, “Effective automated prediction of vertebral column pathologies based on logistic model tree with SMOTE preprocessing,” Journal of Medical Systems, vol. 38, pp. 1-9, April 2014.

[41] T. Maciejewski, and J. Stefanowski, “Local neighbourhood extension of SMOTE for mining imbalanced data,” IEEE Symposium on Computational Intelligence and Data Mining, pp. 104-111, 2011.

[42] H. A. Najada, and X. Zhu, “iSRD: Spam review detection with imbalanced data distributions,” IEEE International Conference on Information Reuse and Integration, pp. 553-560, 2014.