Inconsistency Investigation between Online Review Content and Ratings

Completed Research

Guohou Shan
University of Maryland, Baltimore County
gshan2@umbc.edu

Dongsong Zhang
University of Maryland, Baltimore County
zhangd@umbc.edu

Lina Zhou
University of Maryland, Baltimore County
zhoul@umbc.edu

Lingge Suo
University of Maryland, Baltimore County
lingge1@umbc.edu

Jaewan Lim
University of Maryland, Baltimore County
jae12@umbc.edu

Chunming Shi
Wilfrid Laurier University
cshi@wlu.ca

Abstract

Despite the tremendous role of online consumer reviews (OCRs) in facilitating consumer purchase decision making, the potential inconsistency between product ratings and review content could cause the uncertainty and confusions of prospect consumers toward a product. This research is aimed to investigate such inconsistency so as to better assist potential consumers with making purchase decisions. First, this study extracted a reviewer's sentiments from review text via sentiment analysis. Then, it examined the correlation and inconsistency between product ratings and review sentiments via Pearson correlation coefficients (PCC) and box plots. Next, we compared such inconsistency patterns between fake and authentic reviews. Based on an analysis of 24,539 Yelp reviews, we find that although the ratings and sentiments are highly correlated, the inconsistency between the two is more salient in fake reviews than in authentic reviews. The comparison also reveals different inconsistency patterns between the two types of reviews.

Keywords

Inconsistency, sentiments, Pearson correlation coefficient, box plot.

Introduction

Online consumer reviews (OCRs), including product/service ratings and textual comments, play a more and more important role in guiding potential consumers to make informed purchase decisions. A product rating reflects consumers' general attitudes and feelings towards certain products they bought or services they experienced, which could provide a general buying advice for potential buyers. Review content could provide specific product buying and/or using experiences. Previous research also found that OCRs could greatly affect consumers' purchase intention (Park et al. 2007). Particularly, positive reviews with high product ratings can attract consumers to buy, while negative reviews with low product ratings may dampen consumers' buying probability (Zhang et al. 2016). The information contained in OCRs can help potential customers reduce product uncertainty and make more reasonable purchase decisions (Baek et al. 2012). However, an OCR may not help consumers if there exists inconsistency between the rating and sentiments of the same OCR calculated based on review content, which is the focus of this study.

Normally, consumers tend to see the product rating first, then look into review content when they want to find more details about the assessed products to make a better purchasing decision (Baek et al. 2012). However, sometimes a numeric rating and the associated textual review content may be significantly inconsistent (Islam 2014). For example, consumers may utilize high or low ratings to attract other potential consumers to read their reviews and then display ads for totally different products with inconsistent attitudes and content. Thus, it would be interesting to explore the patterns of inconsistency between product ratings and sentiments expressed in review content. In addition, given the rise of fake OCRs (Zhang et al. 2016), a comparison of the inconsistency patterns between fake and authentic reviews would provide insights into the quality of OCRs and improve automated detection of fake OCRs.

This study makes the following research contributions: (1) it empirically examines the correlation and inconsistency between product ratings and review content sentiments; (2) it identifies specific product ratings where inconsistency is most likely to occur; and (3) it innovatively compares the inconsistency patterns of fake reviews against those of authentic reviews, which could help identify quality reviews.

The rest of this paper is organized as follows. First, we will introduce related work, including sentiment analysis methods. Then, the proposed methods will be presented. Next, we will present data analysis and results, and finally discuss research contributions, practical implications, and limitations part of this research.

Related Work

Research of online reviews

A variety of researches have been explored on investigating online product ratings and review content. Based on their different research focus, those researches can be classified into two broad categories: investigating the effects/applications of online reviews and looking into the characteristics of online reviews.

Researches on studying effects of online reviews focus on investigating the applications of online reviews. The applications of online reviews include spam reviews detection, book sales prediction, consumer purchase intentions, etc. (Chong et al. 2017; Fan et al. 2017; Hu et al. 2014; Islam 2014; Lee and Shin 2014; Salehan and Kim 2016; Sharma and Lin 2013; Tan et al. 2018). For example, Hu et al. (2014) investigated the effect of product ratings and review sentiments on Amazon book sales. In their research, they developed a multiple equation models to examine the inter-relationships between ratings, sentiments, and sales by using reviews of more than 4,000 Amazon books and found that ratings had indirect impact on sales while sentiments had a direct influence on book sales. Lee et al. (2014) investigated the influence of review qualities on consumer purchase intention. In their research, they did a web-based experiment with 201 participants to examine how the quality of online product reviews affected consumers' purchase decision and how such effects varied depending on the product type. They found that positive reviews with high quality would lead to a strong purchase intention and review quality had a negative direct effect on the purchase intention for the experience good, with no corresponding effect for the search good.

Researches on investigating online review characteristics mainly focus on analyzing helpful review characteristics, predicting review helpfulness, analyzing facts influencing review helpfulness values, etc. (Askalidis et al. 2017; De Pelsmacker et al. 2018; Fang et al. 2016; Ghose and Ipeirotis 2007; Liu and Park 2015). For instance, Liu et al. (2015) utilized travel product websites to investigate the characteristics of a helpful review. In their research, they collected 5,090 online reviews from Yelp and utilized TOBIT regression model to analyze the relationship of review helpfulness with reviewer's information, expertise, number of friends, fans, word count of reviews and so on. They found that combination of both reviewer's characteristics and review content positively affected the perceived usefulness of reviews. Fang et al. (2016) implemented an empirical analysis to analyze the facts influencing review helpfulness values of online tourism reviews. In their research, they utilized online review data crawled from TripAdvisor and conducted a two-level empirical analysis to explore factors affecting the value of review helpfulness while applied a Tobit regression model at the reviewer level to investigate the effects of reviewer characteristics. Their empirical analysis results indicated that both text readability and reviewer characteristics affected the perceived review helpfulness values.

However, there is a lack of research on empirically studying the correlation and inconsistency between product ratings and content sentiments of the same reviews, and the inconsistency patterns among authentic and fake reviews.

Sentiment Analysis of OCRs

Sentiment analysis (SA) is the method of extracting sentiments from text (Liu 2012). It has been commonly applied to social media content, online website reviews, such as OCRs and online discussion forums, to discover people's attitude, emotions, and feelings toward an entity (Gangadharbatla 2008). From a technical point of view, recent sentiment analysis methods can be classified into four categories, including lexicon-based, machine learning, statistical, and rule-based approaches (Collomb et al. 2014).

A lexicon-based approach to SA (e.g., Taboada et al. 2011) utilizes sentiment terms in dictionaries to calculate text polarity and sentiment value. Machine learning based methods often adopt supervised learning algorithms for SA by training a model using a known dataset (Zou et al. 2015). Statistical models represent each review as a mixture of latent aspects and ratings. This method tries to cluster head terms into aspects and sentiments into ratings. It is assumed that the latent aspects and ratings can be represented by multinomial distributions (Collomb et al. 2014). The rule based approach looks for opinion words in a text based on predefined rules and then classifies them based on the number of positive and negative words (Gilbert 2014). The specific comparisons of these four approaches are shown in Table 1 (Devika et al. 2016).

Approaches	Advantages	Disadvantages
Lexicon Based	Labelled data and procedure of learning is not required	Requires pre-defined lexicons
Machine learning	Dictionary is not necessary; High accuracy of classification	Domain specific; Labeled data is hard to get.
Statistical models	Easy to implement	No linguistic insights
Rule based	High performance at review and sentence level (91% and 86% accuracy respectively)	Efficiency and accuracy depend on the defining rules

Table 1 Comparisons of the Four Approaches

Due to the existence of various online review related lexicons and the disadvantages of other three approaches, we adopt the lexicon-based SA method in this research.

The Proposed Method

The primary goal of this study is to investigate the correlation between product ratings and sentiments of review content, and compare their inconsistency patterns between fake and authentic reviews. The proposed approach is shown in Figure 1.

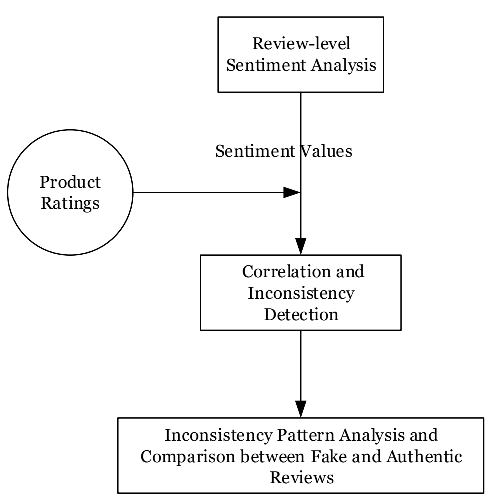

Figure 1 Process of the Proposed Approach

First, this research performs review-level sentiment analysis to attain reviews' sentiment value, then analyzes whether product ratings and review sentiments are correlated and inconsistent. Next, we analyze and compare specific inconsistency patterns among fake and authentic reviews.

Review-level Sentiment Analysis

Normally, reviews express opinions that involve sentiments (Gonçalves et al. 2013). There are three levels of sentiment analysis: document (i.e., review) level, sentence level, and aspect level (Liu 2012). This research mainly focuses on the review level sentiment analysis. The process of extracting reviewers' sentiments from their reviews mainly consists of three steps: tokenizing the textual content of a review and tagging the terms in the review to identify sentiment terms; calculating the sentiment polarity of sentiment terms based on lexicons; and aggregating the sentiments of the identified sentiment terms to form the sentiment of the entire review (Pang and Lee 2008). In this paper, we adopt the lexical based method to discover sentiment within reviews. There are more than 50 different lexicons, such as LIWC (Tausczik and Pennebaker 2010), POMS-ex (Bollen et al. 2011) and SentiWordNet3.0 (Baccianella et al. 2010). Considering its high accuracy (Ribeiro et al. 2016), we adopt SentiWordNet3.0 for SA in this research. The specific process of review sentiment discovery is shown in Figure 2.

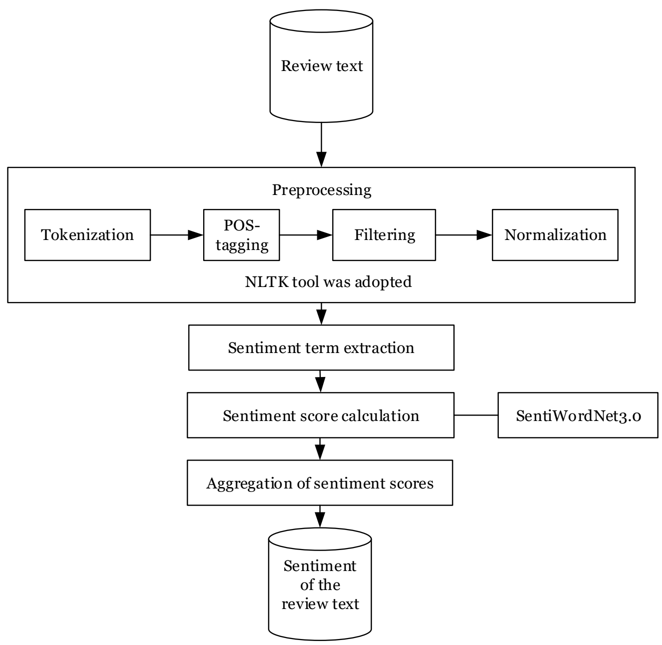

Figure 2 Review-level Sentiment Analysis based on SentiWordNet3.0

Because SentiWordNet3.0 can't directly work on multi-sentence reviews (Kreutzer and Witte 2013), this research preprocessed review text with the following four steps before using SentiWordNet3.0: (1) Tokenization. Tokenization transformed review text into tokens, making it convenient for POS-tagging; (2) POS-tagging, which assigned parts of speech to each word in the review text; (3) Term filtering, which reduced the number of terms being processed to nouns, adjectives, verbs, adverbs; and (4) Normalization, which was used to stem and lemmatize the remaining terms in a review.

After preprocessing, we extracted sentiment terms based on their POS in a review and utilized SentiWordNet3.0 to calculate the sentiment score of each sentiment term. Finally, the sentiment scores of individual terms in the review text were aggregated to attain the final sentiment score of an entire review.

Product ratings and review sentiment correlation and inconsistency detection

After deriving review sentiments, we first explored the possible correlation between product ratings and review sentiments through PCC (Pearson Correlation Coefficient (Benesty et al. 2009)). PCC is calculated as follows:

$$\begin{cases} cov(X,Y) = \dfrac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{n-1} & (1) \\ p_{X,Y} = corr(X,Y) = \dfrac{cov(X,Y)}{\sigma_X \sigma_Y} = \dfrac{E[(X-u_X)(Y-u_Y)]}{\sigma_X \sigma_Y} & (2) \end{cases}$$

Where $cov(X, Y)$ represents the covariance of two variables (X and Y), which are product ratings and review sentiments. $P_{X,Y}$ denotes their PCC. X_i *and* Y_i represent the i^{th} examples of X *and* Y, respectively, and $\overline{X}$ and $\overline{Y}$ denote the average values of X *and* Y, respectively. n represents the total number of examples. u_X and u_Y represent the expected average values of X and Y, respectively, while σ_X and σ_Y are the standard deviations of X and Y, respectively. The value of a PCC is in the range of [-1, 1], with -1 indicating that the two variables are totally negatively correlated, 1 meaning that the two variables are totally positively correlated, and 0 indicating no relations (Nagelkerke 1991).

The consistency of two things means that one is uniform or retaining form with the other (Complete 2009). In this paper, we regard consistency of product ratings and sentiments as a uniform/normal distribution of review sentiments in product ratings. In this case, the inconsistency between sentiments and ratings, if exists, could be regarded as an anomaly distribution of review sentiments in product ratings.

Inconsistency pattern analysis

To examine the potential inconsistency patterns between a product rating and its review sentiments, we analyzed average review length (ARL), average number of nouns (AN), average number of verbs (AVN), average number of adjectives (AAN) and average number of adverbs (ARN) of each review. In the term filtering stage of review-level sentiment analysis, nouns/adjectives/verbs/adverbs of a review were filtered, which may affect the sentiment term extraction and the final sentiment score of the review. Besides, review length may have influence on review sentiments (Hong and Fang 2015). In this case, we analyzed the specific ARL, AN, AVN, AAN and ARN of reviews and combined these patterns with product ratings to analyze their inconsistency patterns in detail.

Analysis and Result

OCR Dataset

Based on the focus of this research, we collected 24,539 restaurant reviews from Yelp.com, including reviewer id, review ratings, review content, and "recommended" or "not recommended" labels (Luca 2016). Among them, 11,641 reviews were labelled as "recommended" and 12,898 were labelled as "not recommended" by the fake review detection system of Yelp, which could be regarded as authentic reviews and fake reviews, respectively (Sussman et al. 2014).

Pearson correlation coefficient analysis and comparison

By following the process shown in Figure 2, we calculated the review sentiment scores for each rating. Examples of the first 10 reviews are presented in Table 2.

Reviewer ID	Star Rating	Review Sentiment Score	Label
zpvnHZ	1	-0.28645	Y
YrWoZw	4	0.362747	N
5AjacS	1	-0.10958	Y
Z2pnai	1	0.101393	Y
v8JUPx	4	0.224799	N

t_GPDT	5	0.303504	Y
wlm-Ug	2	-0.08785	N
RNyOJT	5	0.121646	Y
LXGuXk	4	0.655798	N
5VkVBk	3	0.178926	Y

Table 2 Review Sentiments of the First 10 Reviews

To protect user privacy, only the first 6 characters of the original 22-digit user ID were extracted as the reviewer ID (see Table 1). In the "Label" column, 'Y' (or 'N') indicates that this review is a fake (or authentic) review.

We used PCC to analyze correlations between ratings and review sentiments. The values of PCC between ratings and review sentiments of fake and authentic reviews are 0.374 (p<.01) and 0.398 (p<.01), respectively. It shows that review ratings and sentiments are both significantly positively correlated at 0.01 of both fake and authentic reviews.

Inconsistency between review ratings and sentiments

We used box plot by SPSS 18.0 to analyze whether inconsistency exists between ratings and review sentiments. The results are shown in Figures 3 and 4.

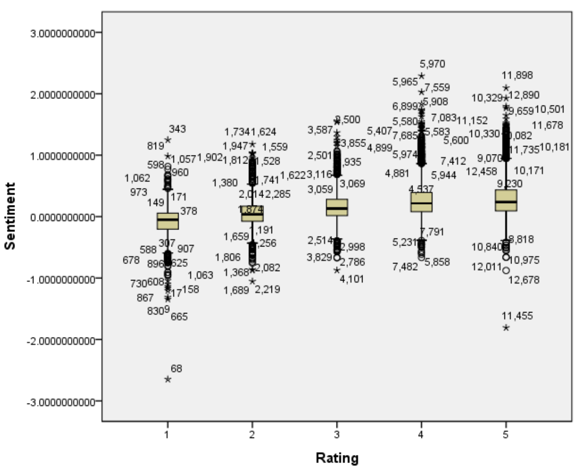

Figure 3 Box plot of fake reviews

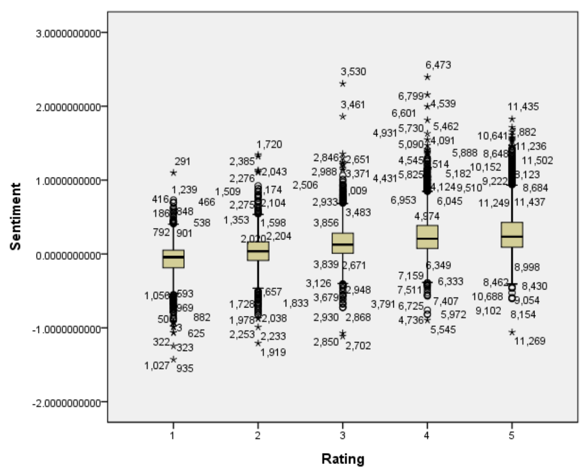

Figure 4 Box plot of authentic reviews

Figure 3 shows the box plot of review sentiment distribution on each product rating. The number at each point indicates the review index in the dataset. The point type 'o' means "mild outlier" and '∗' means "extreme outlier", which indicate abnormal points in SPSS 18.0. It could be seen in Figure 3 that all review ratings have abormal points. It means sentiments are abnormally distributed, which indicates that the inconsistency exists in each rating of fake reviews. Specially, the extreme outliers tend to be more abnormally distributed under the minimum normal range in fake reviews with low product ratings (i.e., 1 and 2), but more abnormally distributed above the maximum normal range in high ratings (i.e., 4 and 5).

As shown in Figure 4, the inconsistency distribution of review sentiments in each review rating for authentic reviews is also salient. There are many outliers of review sentiments of reviews with each rating, which shows that inconsistency does occur in authentic reviews. At the same time, low rating authentic reviews tend to have more extreme outliers ('∗') under minimum normal range, while high rating reviews have a tendency of having more extreme outliers above maximum normal range.

According to Figures 3 and 4, it could be concluded that the inconsistency between restaurant ratings and review sentiments exists in both fake and authentic reviews. We further examine whether there are significant differences in their inconsistency patterns between the two groups of reviews.

	Fake reviews					Authentic reviews				
Ratings	1	2	3	4	5	1	2	3	4	5
Sample size	1106	1208	1880	4300	4404	1318	1173	1575	3544	4031
Minimum	-2.648	-1.053	-.873	-.665	-1.808	-1.427	-1.208	-1.115	-.897	-1.060
Maximum	1.249	1.178	1.557	2.289	2.096	1.099	1.343	2.305	2.395	1.827
Mean	-.083	.051	.158	.265	.291	-.076	.037	.158	.255	.284
SD^2	.070	.060	.060	.079	.085	.063	.065	.070	.075	.081

Table 3 Distributions of Fake and Authentic Reviews cross Ratings

Table 3 reveals that fake reviews with restaurant ratings of 1, 4, and 5 have higher sentiment standard deviations (SD), which are 0.70, 0.79, and 0.85 respectively, than authentic reviews, which are .063, 0.75 and 0.81, respectively. This shows that fake reviews have higher inconsistency between ratings and sentiments than authentic reviews with a rating of 1, 4, or 5. This is in contrast to reviews with a rating of 2 or 3, suggesting that fake reviewers provide more consistent sentiments in reviews with these two ratings.

Product ratings and review content inconsistency patterns

In order to understand consumer rating-giving behavior and distinguish fake reviews from authentic reviews better, we analyzed and compared review content characteristics: average review length (ARL), average number of nouns (AN), average number of verbs (AVN), average number of adjectives (AAN) and average number of adverbs (ARN) of fake reviews versus authentic reviews. By applying the python NLTK tool, the distributions of textual characteristics of reviews with different ratings are presented in Table 4.

	Authentic reviews					Fake reviews				
Rating	1	2	3	4	5	1	2	3	4	5
ARL	554.80	577.17	512.41	492.59	493.81	529.88	502.41	460.00	441.01	430.20
AN	24.05	25.19	23.02	22.89	22.62	23.32	21.88	20.65	20.30	19.85
AVN	20.78	20.63	17.52	16.09	16.43	19.74	18.07	15.82	14.59	14.32
AAN	7.29	9.11	9.01	9.08	8.49	7.01	8.33	8.46	8.31	7.62
ARN	8.39	9.08	7.94	7.21	6.88	7.92	8.32	7.29	6.59	6.23

Table 4 Distributions of Textual Characteristics of Authentic and Fake Reviews

As shown in Table 4, authentic reviews with a rating of 2 have the longest review length and include most nouns, adjectives and adverbs; those with a rating of 1 use the most verbs. On the other hand, fake reviews with a rating of 1 are the longest, use the most nouns and verbs, but the least adjectives, and those with a rating of 3 have the most adjective words. This shows that the inconsistency between review ratings and review content really exists. Table 4 also shows that the average length of authentic reviews with any rating is larger than that of fake reviews. This may indicate that fake reviewers tend to spend less time and write less review content than authentic reviewers. At the same time, authentic reviews include more adjectives, adverbs, nouns, verbs than fake reviews. In order to validate our research findings about different inconsistency patterns of fake and authentic reviews, we did the following paired sample T test.

	Difference in pairs					T	df	Sig. (two-sides)
	Mean	SD	SE Mean	95%confidence interval				
ARL	53.45600	18.54643	8.29422	30.42756	76.48444	6.445	4	.003
AN	2.35400	.97215	.43476	1.14691	3.56109	5.414	4	.006
AVN	1.78200	.58088	.25978	1.06074	2.50326	6.860	4	.002
AAN	.65000	.23801	.10644	.35447	.94553	6.107	4	.004
ARN	.04600	1.05881	.47351	-1.26869	1.36069	.097	4	.927

Table 5 T-test Comparison of Authentic and Fake Review Patterns

Tables 5 shows that the differences of inconsistency patterns (ARL, AN, AVN, AAN) are significant between authentic and fake reviews while the ARN inconsistency pattern difference is not significant.

Discussion

Review ratings and textual content play a more and more important role in helping consumers make informed purchase decisions (Park et al. 2007). Given the large volume of OCRs, assisting consumers to identify quality and consistent reviews could help save consumers' time while making a decision. This research analyzes the relationship and inconsistency between review sentiments and product ratings, and compares inconsistency patterns among fake and authentic reviews.

This study makes several research contributions. First, this research reveals that star ratings and review sentiments are positively correlated based on Pearson Correlation Coefficient.

Second, the inconsistency between ratings and review sentiments really exists in reviews. We used box plots to visualize abnormal distribution of review sentiments for every rating to statistically verify the existence of inconsistency. Then, this paper dug in deeper into textual content of reviews and revealed that review length, review adjective words, noun words, verb words and adverbs were all inconsistently distributed among the five star ratings.

Third, the inconsistency patterns between ratings and review content are different in authentic reviews versus fake reviews. This research found that for authentic reviews, ratings and review sentiments have higher positive correlations than those for fake reviews. At the same time, fake reviews with a rating of 1 or 5 tend to have much fewer abnormal extreme outliers than authentic reviews, but authentic reviews with a rating of 3 tend to have much more abnormal extreme outliers than fake reviews. Besides, fake reviews and authentic reviews also have significantly different inconsistency patterns of ARL, AN, AVN, AAN between ratings and textual content characteristics of reviews. This research finding may be useful to improve fake review detection accuracy by adding more features related to significantly different textual content characteristics of fake and authentic reviews.

This study has some limitations that offer potential opportunities for future research. In our study, we only used data collected from Yelp.com. Reviews on this platform are mainly about local businesses, as well as the online reservation service (i.e., Yelp Reservations) (Luca 2016). However, there are many other OCR platforms, such as Amazon.com, that provide OCRs on a large number of diverse products or services. It would be interesting to examine if the findings of this study will still hold true with OCRs on other platforms. Similarly, we only used OCRs of restaurants, which can be viewed as a type of experience goods. It would be worth investigating if those inconsistency patterns also exist in reviews of search goods.

ACKNOWLEDGMENT

This research was partially supported by the National Science Foundation under Grants SES-1527684 and CNS-1704800. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the above funding agency.

REFERENCES

Askalidis, G., Kim, S. J., and Malthouse, E. C. 2017. "Understanding and Overcoming Biases in Online Review Systems," *Decision Support Systems* (97), pp. 23-30.

Baccianella, S., Esuli, A., and Sebastiani, F. 2010. "Sentiwordnet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," *LREC*, pp. 2200-2204.

Bollen, J., Mao, H., and Pepe, A. 2011. "Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena," *Icwsm* (11), pp. 450-453.

Chong, A. Y. L., Ch'ng, E., Liu, M. J., and Li, B. 2017. "Predicting Consumer Product Demands Via Big Data: The Roles of Online Promotional Marketing and Online Reviews," *International Journal of Production Research* (55:17), pp. 5142-5156.

Collomb, A., Costea, C., Joyeux, D., Hasan, O., and Brunie, L. 2014. "A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation," *Rapport de recherche RR-LIRIS-2014-002*).

Complete, C. E. D. "Unabridged.(2009)," *Bank of English. Glasgow: HarperCollinsPublishers*).

De Pelsmacker, P., Dens, N., and Kolomiiets, A. 2018. "The Impact of Text Valence, Star Rating and Rated Usefulness in Online Reviews," *International Journal of Advertising*), pp. 1-20.

Devika, M., Sunitha, C., and Ganesh, A. 2016. "Sentiment Analysis: A Comparative Study on Different Approaches," *Procedia Computer Science* (87), pp. 44-49.

Fan, Z.-P., Che, Y.-J., and Chen, Z.-Y. 2017. "Product Sales Forecasting Using Online Reviews and Historical Sales Data: A Method Combining the Bass Model and Sentiment Analysis," *Journal of Business Research* (74), pp. 90-100.

Fang, B., Ye, Q., Kucukusta, D., and Law, R. 2016. "Analysis of the Perceived Value of Online Tourism Reviews: Influence of Readability and Reviewer Characteristics," *Tourism Management* (52), pp. 498-506.

Gangadharbatla, H. 2008. "Facebook Me: Collective Self-Esteem, Need to Belong, and Internet Self-Efficacy as Predictors of the Igeneration's Attitudes toward Social Networking Sites," *Journal of interactive advertising* (8:2), pp. 5-15.

Ghose, A., and Ipeirotis, P. G. 2007. "Designing Novel Review Ranking Systems: Predicting the Usefulness and Impact of Reviews," *Proceedings of the ninth international conference on Electronic commerce*: ACM, pp. 303-310.

Gilbert, C. H. E. 2014. "Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf*.

Gonçalves, P., Araújo, M., Benevenuto, F., and Cha, M. 2013. "Comparing and Combining Sentiment Analysis Methods," *Proceedings of the first ACM conference on Online social networks*: ACM, pp. 27-38.

Hong, J., and Fang, M. 2015. "Sentiment Analysis with Deeply Learned Distributed Representations of Variable Length Texts," Technical report, Stanford University.

Hu, N., Koh, N. S., and Reddy, S. K. 2014. "Ratings Lead You to the Product, Reviews Help You Clinch It? The Mediating Role of Online Review Sentiments on Product Sales," *Decision support systems* (57), pp. 42-53.

Islam, M. R. 2014. "Numeric Rating of Apps on Google Play Store by Sentiment Analysis on User Reviews," *Electrical Engineering and Information & Communication Technology (ICEEICT), 2014 International Conference on*: IEEE, pp. 1-4.

Kreutzer, J., and Witte, N. 2013. "Opinion Mining Using Sentiwordnet," *Uppsala University*).

Lee, E.-J., and Shin, S. Y. 2014. "When Do Consumers Buy Online Product Reviews? Effects of Review Quality, Product Type, and Reviewer's Photo," *Computers in Human Behavior* (31), pp. 356-366.

Liu, B. 2012. "Sentiment Analysis and Opinion Mining," *Synthesis lectures on human language technologies* (5:1), pp. 1-167.

Liu, Z., and Park, S. 2015. "What Makes a Useful Online Review? Implication for Travel Product Websites," *Tourism Management* (47), pp. 140-151.

Luca, M. 2016. "Reviews, Reputation, and Revenue: The Case of Yelp. Com,").

Nagelkerke, N. J. 1991. "A Note on a General Definition of the Coefficient of Determination," *Biometrika* (78:3), pp. 691-692.

Pang, B., and Lee, L. 2008. "Opinion Mining and Sentiment Analysis," *Foundations and Trends® in Information Retrieval* (2:1–2), pp. 1-135.

Park, D.-H., Lee, J., and Han, I. 2007. "The Effect of on-Line Consumer Reviews on Consumer Purchasing Intention: The Moderating Role of Involvement," *International journal of electronic commerce* (11:4), pp. 125-148.

Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., and Benevenuto, F. 2016. "Sentibench-a Benchmark Comparison of State-of-the-Practice Sentiment Analysis Methods," *EPJ Data Science* (5:1), p. 23.

Salehan, M., and Kim, D. J. 2016. "Predicting the Performance of Online Consumer Reviews: A Sentiment Mining Approach to Big Data Analytics," *Decision Support Systems* (81), pp. 30-40.

Sharma, K., and Lin, K.-I. 2013. "Review Spam Detector with Rating Consistency Check," *Proceedings of the 51st ACM southeast conference*: ACM, p. 34.

Sussman, S., Garcia, R., Cruz, T. B., Baezconde-Garbanati, L., Pentz, M. A., and Unger, J. B. 2014. "Consumers' Perceptions of Vape Shops in Southern California: An Analysis of Online Yelp Reviews," *Tobacco induced diseases* (12:1), p. 22.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. 2011. "Lexicon-Based Methods for Sentiment Analysis," *Computational linguistics* (37:2), pp. 267-307.

Tan, H., Lv, X., Liu, X., and Gursoy, D. 2018. "Evaluation Nudge: Effect of Evaluation Mode of Online Customer Reviews on Consumers' Preferences," *Tourism Management* (65), pp. 29-40.

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* (29:1), pp. 24-54.

Zou, H., Tang, X., Xie, B., and Liu, B. 2015. "Sentiment Classification Using Machine Learning Techniques with Syntax Features," *Computational Science and Computational Intelligence (CSCI), 2015 International Conference on*: IEEE, pp. 175-179.