Gaming The System: An Insight Into Online Fake Reviews

Paresh Sharma

Contents

1. Abstract

Online customer reviews have played an increasingly important role in the purchasing decision of consumers buying goods and services. Consumers are more inclined to purchase highly rated products; consequently, businesses fiercely compete to obtain positive reviews for their products. Unfortunately, some businesses compete by employing unethical tactics such as writing fake reviews to promote their products and demote competitors. This results in asymmetric information in the marketplace which has led to poor purchasing decisions and legitimate sellers out of pocket. 'Gaming the system' in this manner has become increasingly problematic over time and many believe more needs to be done to tackle the problem.

To investigate this problem, a dataset of 800 deceptive and 800 truthful hotel reviews is used to explore the key linguistic differences between real and fake hotel reviews. Additionally, several supervised machine learning models will be deployed to try and automatically detect fake reviews. The findings suggest some structural and keyword differences between real and fake reviews. Furthermore, the Naïve Bayes classifier using 1:2 n-grams provides the best classification performance with an 89% accuracy.

2. Introduction

Since the inception of the internet, it has become increasingly common for individuals to read and share their own product experiences online. A plethora of studies have found that 90+% of shoppers read online reviews before making a purchase [1, 2, 3]. These reviews act as a trusted source of information when comparing products to buy. Although personal recommendations from friends and family tend to be the most trusted source of information, a study by BrightLocal found that 85% of individuals trusted reviews as much as personal recommendations [3]. Furthermore, 73% of customers valued a written review over an overall star rating [2]. Along with the quality of the review, the volume of reviews a product receives plays a role in a consumer's decision. Given two products with similar ratings, consumers are more likely to purchase the product with more reviews [4]. Moreover, the conversion rate for a product with five reviews is 270% greater than the purchase likelihood of a product with no reviews and this increases to 380% for higher-priced products [5].

Strong reviews are vital to a business's sales success, with one economic consultancy estimating reviews boost sales by 18% [6]. Naturally, the positive correlation between strong reviews and sales incentivises businesses to provide good products and services. If they do not, businesses could face reputational damage from customers submitting negative reviews, especially given that negative reviews tend to impact sales more than positive reviews as they discourage future customers from consuming products and services [7]. Thus, customer reviews can help to regulate businesses and provide an effective and efficient mechanism to overcome the information asymmetry between online buyers and sellers.

Unfortunately, some companies chose to 'game the system' by writing and getting others to write fake reviews to promote their products and demote competitors. Fake reviews are reviews that are not a person’s honest and impartial opinion or do not reflect their genuine experience of a product [8]. Individuals writing fake reviews are called opinion spammers and their activities are called opinion spamming. It is hard to determine the extent of opinion spamming online, but some research has suggested it could be as high as 67% of online reviews[1] [9].

An interesting example of where several strategies have been employed to manipulate the marketplace is the case of Amazon. One method used by opinion spammers on Amazon is to ask genuine buyers to pay for an item and leave a positive review, but then reimburse them the cost of the item. There are multiple groups[2] on Facebook which facilitate this by allowing sellers to offer their products for free or at discounted rates, in exchange for reviews. The consumer group, Which?, found that collectively, just seven of these groups had more than 87,000 members, and advertise hundreds of products each day [8]. Another tactic used by sellers is to make fake orders and leave positive reviews with fake Amazon ‘zombie’ accounts. These accounts often emulate ‘real’ customer browsing behaviour to not arouse suspicion and send products to real addresses (after compensating the homeowner). The final method discussed is where sellers have bought their own and competitor data from Amazon employees who steal the information from Amazon’s databases and resell the data. This data can be used in several ways including contacting customers to remove/change negative reviews in exchange for a financial incentive and running advertising campaigns to target specific customers directly through email or indirectly through a Facebook remarking campaign [10].

As a result, opinion spamming online has led to asymmetric information in the marketplace which has culminated in poor purchasing decisions and legitimate sellers out of pocket. Despite businesses and consumers being held accountable for posting fake reviews under the 2008 Consumer Protection from Unfair Trading Regulations[3] [11], ‘gaming the system’ in this manner has become increasingly problematic over time and many believe more needs to be done by companies, policymakers, and regulators to tackle the problem.

3. Literature review

Spam has historically been examined in the context of emails [12] and actions intended to influence the page ranking from search engines [13, 14]. However, more recently researchers have extended their scope in this field to also look at opinion spamming. Research has tended to focus on identifying the key linguistic differences between real and fake reviews along with attempting to automatically detect fake reviews with supervised machine learning models.

Yoo and Gretzel, 2009, observe the linguistic differences between 42 deceptive and 40 truthful reviews. The results show that deceptive reviews tended to be more complex and more likely to include self-references. However, no significant difference was found between the length of the review and the number of unique words used [15].

---

[1] An article by The Washington Post suggests 67% of the reviews for the Amazon product ‘testosterone booster’ were found to be questionable. Furthermore, Bluetooth speakers, weight-loss pills and Bluetooth headphones all scored between 50-60% [9].

[2] Groups include Amazon Deals Group and Amazon UK Reviewers.

[3] This is a UK law and laws regarding deception may vary by country.

A larger study was conducted by Jindal and Liu, 2008, who investigated opinion spamming in a dataset consisting of 5.8 million Amazon reviews from 2.14 million reviewers. They discovered many duplicate reviews written by the same reviewers on the different products or by different reviews on the same products. Using these findings, they attempted to manually label a dataset which categorised 470 fake reviews (although the authors concede the difficulty and potential inaccuracy of this task given that some fake reviews are very similar to real reviews). A logistic regression model was then trained on the data and produced an Area Under the ROC Curve (AUC) score of 78% [16].

Ott et al., 2011, find that humans can only correctly identify fake positive reviews with 61.9%[4] accuracy and suggest this could have negatively impacted previous studies. To overcome this problem, they build a new gold-standard dataset using Chicago hotel reviews. The data consists of 400 truthful positive reviews from tripadvisor.com and 400 fake positive reviews written by users following a careful design procedure on Amazon Mechanical Turk (AMT)[5]. The data is trained using unigrams, bigrams, trigrams, and features from the Linguistic Inquiry and Word Count (LIWC) tool for a Support Vector Machine (SVM) model. Bigrams and LIWC perform best with an accuracy of 89.8% [17]. This accuracy is boosted to 91.2% by Feng et al., 2012, using deep syntax rule-based features and unigrams [18].

In 2013, Ott et al., build on their original work by exploring 400 real negative hotel reviews from Chicago and 400 fake negative hotel reviews created on AMT as their gold-standard data. The authors find that human deception detection is better for negative reviews than positive reviews (65% to 61.9%), but the best detection performance is still achieved through automated classifiers which achieved 86% accuracy [19].

4. Research objective

This paper will combine the gold-standard hotel review data from Ott et al., 2011 & 2013, to further explore the key linguistic differences between real and fake positive and negative reviews. Additionally, several supervised machine learning models will be deployed to try and automatically detect fake reviews when both positive and negative reviews are present in the data. This study hopes to provide the end user with a better understanding of the linguistic differences between real and fake reviews along with a classification system for automatically detecting fake reviews.

5. Corpus description

The corpus consists of 1600 truthful and deceptive hotel reviews split by sentiment from 20 of the most popular Chicago hotels and is taken from the author’s personal website [20]. The corpus can be broken down as 20 reviews for the 20 Chicago hotels, split by sentiment:

- 400 truthful positive reviews from TripAdvisor [17]
- 400 deceptive positive reviews from AMT [17]
- 400 truthful negative reviews from Expedia, Hotels.com, Orbitz, Priceline, TripAdvisor and Yelp [19]

[4] Three humans were tested, and this was the score of best performing human.
[5] AMT is an online crowdsourcing website used to employee humans to manually label data.

- 400 deceptive negative reviews from AMT [19]

The author tried to maintain a consistent pre-processing procedure across the data, however, there are minor differences. The specifics are outlined in [17, 19].

The hotels used in the study are:

- affinia: Affinia Chicago (now MileNorth, A Chicago Hotel)
- allegro: Hotel Allegro Chicago - a Kimpton Hotel
- amalfi: Amalfi Hotel Chicago
- ambassador: Ambassador East Hotel (now PUBLIC Chicago)
- conrad: Conrad Chicago
- fairmont: Fairmont Chicago Millennium Park
- hardrock: Hard Rock Hotel Chicago
- hilton: Hilton Chicago
- homewood: Homewood Suites by Hilton Chicago Downtown
- hyatt: Hyatt Regency Chicago
- intercontinental: InterContinental Chicago
- james: James Chicago
- knickerbocker: Millennium Knickerbocker Hotel Chicago
- monaco: Hotel Monaco Chicago - a Kimpton Hotel
- omni: Omni Chicago Hotel
- palmer: The Palmer House Hilton
- sheraton: Sheraton Chicago Hotel and Towers
- sofitel: Sofitel Chicago Water Tower
- swissotel: Swissotel Chicago
- talbott: The Talbott Hotel

The data is stored in a CSV (Comma Separated Value) file, this allows data to be stored in a table structured format.

deceptive	hotel	polarity	source	text
truthful	conrad	positive	TripAdvisor	We stayed for a one night getaway with family on a thursday. Triple AAA rate of 173 was a steal. 7th floor room complete with 44in plasma TV bose stereo, voss and evian water, and gorgeous bathroom(no tub but was fine for us) Concierge was very helpful. You cannot beat this location... Only flaw was breakfast was pricey and service was very very slow(2hours for four kids and four adults on a friday morning) even though there were only two other tables in the restaurant. Food was very good so it was worth the wait. I would return in a heartbeat. A gem in chicago...
deceptive	affinia	positive	MTurk	Although much too overpriced in my opinion, the hotel is spotless. The staff was very courteous. And the spa service ? Was a God send ! In a relatively flexible location for traveling for sight seeing so I didnt spend major bucks trying to get around the city ! LOVE IT ! Going back for my anniversary

Table 1: An example of the dataset.

6. Exploratory analysis

6.1. Exploring the data

To provide some context before exploring fake reviews in the dataset, we look at the top words used in the reviews, the semantic similarity between each hotel's reviews and the key topics discussed in the reviews.

What are the most used words in the data?

From the word cloud below, as expected, words related to hotels are the most frequently used.

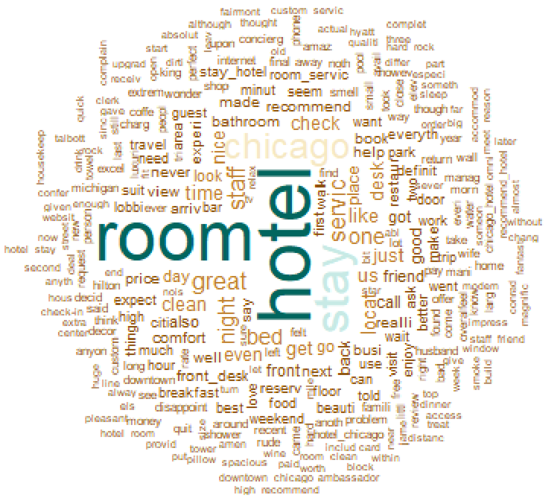

Figure 1: The top 200 words used in the hotel reviews.

How similar are reviews split by hotel?

To observe how similar reviews are by hotel, we construct a hierarchical clustering model using complete linkage after pre-processing the text data. Hierarchical clustering is an algorithm which aims to group objects into clusters. Complete linkage is one of several methods for calculating a hierarchical cluster and works by using the distance furthest away from two elements (one in each cluster) when calculating the distance between clusters.

Figure 2 alludes to 4 clusters among the hotel reviews. However, if we look at the dissimilarity score (height) the clusters do not differ greatly and many fall into one cluster suggesting that the reviews for all hotels are similar despite minor differences. Unfortunately, we are not able to identify why this is the case, which is a major disadvantage of unsupervised approaches. Furthermore, it can be difficult to know which linkage method provides accurate results, as there are no explicit performance metrics. However, intuitively, this model makes sense as one would expect reviews from hotels in the same area to be similar.

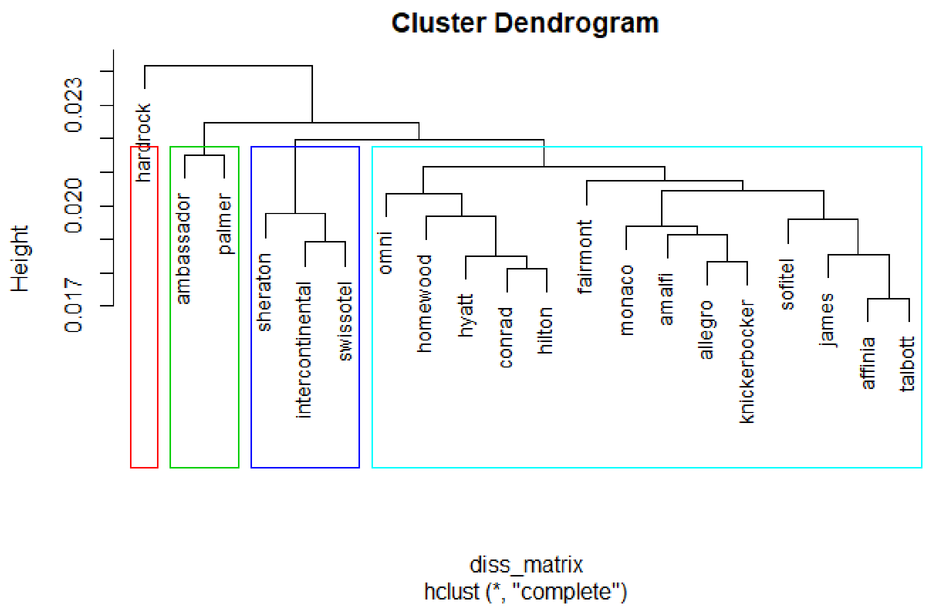

Figure 2: Hierarchical cluster for hotels, grouped into 4 classes.

What are the key topics discussed in the reviews?

Topic modelling, an unsupervised learning approach, can be used to automatically identify topics present in a text corpus. There are many techniques to obtain topics from documents, but Latent Dirichlet Allocation (LDA) developed by Blei et al., 2003 [21] is one of the most popular.

In the LDA model, documents are viewed as a mixture of topics that are present in the corpus and topics are viewed as a distribution over words. The results from an LDA model iterated 1000 times with 5 topics is shown in Table 2. There does not appear to be a clear distinction between topics, rather reviewers appear to discuss hotels more generally. This is further confirmed when looking at the gamma of the LDA model, a parameter which measures the probability of a document falling in each class, as for most documents each topic scores a gamma of 10-40%. This holds true when using a different number of pre-selected topics and on real and fake reviews separately.

	Words									
Topic 1	hotel	chicago	stay	place	recommend	suit	busi	visit	definit	trip
Topic 2	room	bed	night	bathroom	floor	get	use	day	two	lobbi
Topic 3	great	locat	staff	room	hotel	comfort	view	nice	restaur	walk
Topic 4	room	us	desk	check	call	front	one	arriv	get	front_desk
Topic 5	hotel	servic	stay	room	like	look	even	price	time	expect

Table 2: LDA word topics for k = 5.

6.2 Exploring the key linguistic differences between real and fake reviews

Now that we have a basic understanding of the data, the key linguistic differences between real and fake reviews will now be discussed using 3 methods:

i) Basic statistics

In Table 3 structural and word complexity features are taken for different subsets of the original data. The table indicates that the average number of tokens, types and sentences is greater for negative than positive reviews regardless of whether the review is truthful or deceptive. Furthermore, these features are greater for truthful reviews then deceptive reviews. The average use of punctuation follows a similar trend, although deceptive positive reviews contain a lot less punctuation than the other review classes observed. There was not a significant difference between the lexical diversity and readability scores for each review class suggesting that the complexity of the language used does not differ by sentiment or deception.

Statistic	Truthful positive	Deceptive positive	Truthful negative	Deceptive negative	Truthful	Deceptive
Average no. of tokens	140.65	128.46	204.08	196.98	172.37	162.72
Average no. of types	89.52	81.81	118.71	114.51	104.11	98.16
Average no. of sentences	8.52	7.55	11.07	10.34	9.79	8.95
Average use of punctuation	17.64	12.73	25.44	19.41	21.54	16.07
Average Lexical diversity score (TTR[6])	0.7	0.69	0.65	0.64	0.68	0.66
Readability score (Flesch-Kincaid[7])	8.28	8.2	7.94	8.25	8.11	8.22

Table 3: Summary statistics.

ii) Keyness analysis

Next, to identify the most prevalent keywords in deceptive and truthful reviews we used keyness analysis with the chi-squared measure. From figure 3 we can see that keywords for deceptive reviews include less descriptive words like ‘luxury’, ‘hotel’ and ‘chicago’ along with many first-person pronouns whereas the keywords for the truthful reviews are more descriptive e.g. ‘night’, ‘bathroom’ and ‘floor’. One explanation for deceptive reviews including less descriptive words and first-person pronouns is that the authors of the reviews have not visited the hotel and thus find it difficult to comment on anything specific like the truthful reviewers.

---

[6] Type-Token Ratio, where TTR = V/N where V = total types and N = total tokens.
[7] Flesch-Kincaid Readability Score (Flesch and Kincaid, 1975): 0.39*ASL+11.8*(NSy/Nw)-15.59 where ASL = average sentence length, NSy = number of syllables and Nw = number of words.

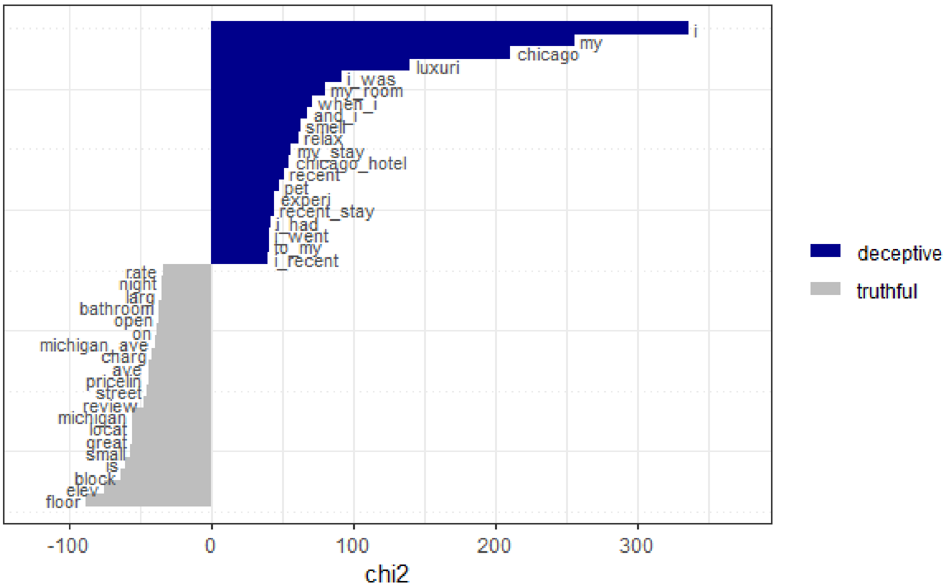

Figure 3: Keyness analysis truthful vs deceptive reviews.

iii) Dictionary

Finally, we used the publicly available Moral Foundations Dictionary (MFD) [22] to assess the difference between truthful and deceptive reviews. This dictionary looks at proportions of virtue and vice for each foundation – care, fairness, loyalty, authority, and sanctity, for a given document. In the analysis, a weighted data frame matrix (DFM) is used to obtain virtue and vice scores for each foundation for a given set of documents. Virtue is subtracted from vice in the table below to compute an overall score for each group of documents.

From Table 4, we can see that truthful and deceptive reviews produce similar scores for each moral foundation. This also holds when broken down by sentiment e.g. truthful positive scores 0.443 for ‘Care’ and deceptive positive scores 0.471. This implies that the deceptive reviews do a good job of mirroring the moral foundations in truthful reviews. Moreover, positive reviews score higher for ‘Care’, ‘Loyalty’ and ‘Sanctity’, but there is no significant difference between sentiment for the other foundations. This is expected given these three morals are highly positive morals, whereas ‘Fairness’ and ‘Authority’ are more sentiment neutral morals.

Moral foundations	Truthful positive (%)	Deceptive positive (%)	Truthful negative (%)	Deceptive negative (%)	Truthful (%)	Deceptive (%)
Care	0.443	0.471	0.178	0.135	0.286	0.267
Fairness	0.003	0.012	0.004	0.008	0.004	0.009
Loyalty	0.146	0.253	0.122	0.151	0.131	0.191
Authority	0.111	0.158	0.219	0.142	0.175	0.148
Sanctity	0.398	0.504	0.011	0.112	0.169	0.267

Table 4: Summary of scores for each moral foundation.

7. Automatically detecting fake reviews

After exploring the linguistic differences in the data, we attempted to create a system for automatically detecting fake reviews. Firstly, we used a keywords in context (KWIC) model on the keywords found in the keyness analysis conducted in section 6. However, it is not easy to identify fake reviews from human observation. Thus, we employed several supervised machine learning algorithms to try and automatically detect fake reviews in the dataset.

The dataset includes 800 truthful reviews and 800 deceptive reviews. As the two classes are already equal we did not need to rebalance the dataset[8]. However, before implementing the models, the data was pre-processed by removing numbers, URLs, and symbols along with lowercasing and stemming all tokens. Punctuation and stop words were kept as they were likely to contain useful features for fake review identification.

The pre-processed tokens were used to create a DFM[9] which was used to train and test different machine learning models for different n-grams. A range performance metrics were computed, however, this study used accuracy to determine the best model, which is in line with previous research. The results are outlined in Table 5.

N-grams	Model	Accuracy	Precision	Recall	F1	AUC
1	Naïve Bayes	88%	86	89	87	88
1:2		**89%**	**85**	**96**	**90**	**89**
1:3		87%	82	96	88	87
1	Ridge	86%	84	89	86	88
1:2		86%	80	95	87	89
1:3		83%	75	97	84	87
1	Lasso	87%	86	87	86	88
1:2		83%	82	83	83	89
1:3		83%	83	81	82	87
1	Elastic net[10]	86%	84	88	86	88
1:2		84%	84	85	84	89
1:3		84%	83	83	83	87
1	Support Vector Machines (linear kernel)	86%	85	86	85	86
1:2		87%	88	86	87	87
1:3		88%	89	86	87	88
1	Random Forests	81%	78	85	81	82
1:2		85%	82	89	85	87
1:3		84%	83	85	84	83
1	Extreme Gradient Boosting	84%	82	87	84	91
1:2		85%	84	87	85	92
1:3		85%	83	87	85	93

Table 5: Performance metrics for supervised machine learning classifiers.

[8] If the data was not balanced, we would need to rebalance it using an under or over sampling technique as some algorithms e.g. Random Forests, do not work with imbalanced classes.

[9] Term frequency – inverse document frequency was also applied, but later removed after noticing it worsened model performance.

[10] Note, Elastic Net, SVM and Extreme Gradient Boosting underwent parameter tuning for model selection.

The Naïve Bayes (NB) model with 1:2 n-grams and Laplace smoothing performed best with 89% accuracy. The model also performed relatively well for the other performance metrics. The NB model is one of the simplest algorithms and can be represented as the following:

$$\boldsymbol{\hat{Y} = argmax_y \, P(y) \prod_{i=1}^{n} P(w_i \,|y)}$$

where $\hat{Y}$ = predicted class, $P(y)$ = prior probability of the class, $\prod_{i=1}^{n} P(w_i \,|y)$ = the product of the probability of a word given the class, for all words.

Table 6 presents the features most predictive of a deceptive and truthful review. Like keyness analysis words like 'luxury' and 'chicago' are highly predictive of deceptive reviews and words like 'pricelin' and 'night' are highly predictive of truthful reviews.

Deception class		**Truthful class**	
Features	**Probability**	**Features**	**Probability**
of_cigarett	0.9183	pricelin	0.9681
even_have	0.9246	/_night	0.9556
park_hotel	0.9246	the_confer	0.9490
chicago_area	0.9273	._love	0.9400
sheraton_chicago	0.9299	was_larg	0.9400
was_cold	0.9299	four_season	0.9400
monaco_chicago	0.9299	hancock	0.9400
'_luxuri	0.9299	:_-	0.9400
relax_.	0.9346	thru	0.9363
i_enjoy	0.9346	._*	0.9363
to_relax	0.9346	&	0.9331
pamper	0.9346	hotwir	0.9320
allegro_chicago	0.9346	great_rate	0.9320
._onc	0.9367	a_corner	0.9320
hotel_chicago	0.9440	subway	0.9320
jame_chicago	0.9455	millenium_park	0.9320
the_millennium	0.9592	the_minibar	0.9320
millennium_knickerbock	0.9623	minut_walk	0.9320
intercontinent_chicago	0.9623	avenu_.	0.9272
chicago_millennium	0.9650	travelzoo	0.9272

Table 6: Naïve Bayes top 20 important predictive features for each class.

8. Conclusion

We conclude that an automated classifier for detecting fake reviews using both positive and negative reviews can significantly outperform human classification (89% to approximately 50-65% [19]). We also find that fake hotel reviews use less descriptive words, first-person pronouns and frequently mention the name of the hotel – possibly to highlight their own credibility. On the other hand, real reviews use specific words detailing their experience at the hotels.

Future work may wish to consider some of the following challenges:

- Many of the characteristics indicating a fake review in this paper may be overcome by well-trained fake review writers, thus resulting in opinion spam harder to detect.
- The mental state of AMT users may not be the same as the state of those writing fake reviews. More research could be conducted to obtain an improved gold-standard dataset.
- This research uses a relatively small sample of reviews, is it possible to accurately classify reviews with big data?
- What further measures could businesses, policymakers and regulators do to help tackle this problem?

9. References

1) Spiegel Research Center. (2016). *How Online Reviews Influence Sales*. [online]. Available at: (https://spiegel.medill.northwestern.edu/_pdf/Spiegel_Online%20Review_eBook_Jun2017_FINAL.pdf [Accesses 7th April 2019]
2) Fan & Fuel. (2017). *No online customer reviews means BIG problems in 2017*. [online]. Available at: https://fanandfuel.com/no-online-customer-reviews-means-big-problems-2017/ [Accessed 7th April 2019].
3) Bright Local. (2018). *Local Customer Review Survey.* [online]. Available at: https://www.brightlocal.com/learn/local-consumer-review-survey/ [Accessed 7th April 2019].
4) Powell et at., 2017. *The Love of Large Numbers: A Popularity Bias in Consumer Choice.* [online]. Available at: https://journals.sagepub.com/doi/full/10.1177/0956797617711291 [Accessed 7th April 2019].
5) Askalidis et al., 2016. *The Value of Online Customer Reviews*. [online]. Available at: https://dl.acm.org/citation.cfm?id=2959181 [Accessed 8th April 2019].
6) Econsultancy. (2012). *Ecommerce consumer reviews: why you need them and how to use them*. [online]. Available at: https://econsultancy.com/ecommerce-consumer-reviews-why-you-need-them-and-how-to-use-them/ [Accessed 8th April 2019].
7) The Australian. (2012). *The economics of fake reviews*. [online]. Available at: (https://www.theaustralian.com.au/business/business-spectator/the-economics-of-fake-reviews/news-story/4053c6ac551e1f73f1544520206db43c [Accessed 8th April 2019].
8) Which? (2018). *The facts about fake reviews*. [online]. Available at: https://www.which.co.uk/news/2018/10/the-facts-about-fake-reviews/ [Accessed 8th 2019].
9) The Washington Post. (2018). *How merchants use Facebook to flood Amazon with fake reviews*. [online]. Available at: https://www.washingtonpost.com/business/economy/how-merchants-secretly-use-facebook-to-flood-amazon-with-fake-reviews/2018/04/23/5dad1e30-4392-11e8-8569-26fda6b404c7_story.html?noredirect=on&utm_term=.e8d0d3a23737 [Accessed 8th April 2019].
10) Ecomcrew. (2018). *How Chinese sellers are manipulating Amazon*. [online]. Available at: https://www.ecomcrew.com/chinese-sellers-manipulating-amazon/ [Accessed 9th April 2019]
11) Gov. (2019). *Consumer Protection from Unfair Trading Regulations*. [online]. Available at: https://www.gov.uk/government/publications/consumer-protection-from-unfair-trading-regulations-traders [Accessed 10th April].
12) Drucker, H., 1999. *Support vector machines for spam categorization*. [online]. Available at: http://www.site.uottawa.ca/~diana/csi5180/paper-itnn_1999_09_1048.pdf [Accessed 10th April].
13) Gyongyi, Z. Garcia-Molina, H., 2005. *Web Spam Taxonomy*. [online]. Available at: http://ilpubs.stanford.edu:8090/771/1/2005-9.pdf [Accessed 11th April].
14) Ntoulas, A., 2006. *Detecting spam web pages through content analysis*. [online]. Available at: https://dl.acm.org/citation.cfm?id=1135794 [Accessed 11th April 2019].
15) Yoo, K. Gretzel, U., 2009. *Comparison of Deceptive and Truthful Travel Reviews*. [online]. Available at: https://www.researchgate.net/publication/221357520_Comparison_of_Deceptive_and_Truthful_Travel_Reviews [Accessed 12th April 2019].

16) Jindal, N. Liu, B., 2008. *Opinion Spam and Analysis*. [online]. Available at: https://www.cs.uic.edu/~liub/FBS/opinion-spam-WSDM-08.pdf [Accessed 12th April].
17) Ott, M. et al., 2011. *Finding Deceptive Opinion Spam by Stretch of the Imagination*. [online]. Available at: https://www.aclweb.org/anthology/P11-1032 [Accessed 12th April].
18) Banerjee, R. Yejin, C., 2012. *Syntactic Stylometry for Deception Detection*. [online]. Available at: https://www.researchgate.net/publication/233844582_Syntactic_Stylometry_for_Deception_Detection [Accessed 12th April 2019].
19) Ott, M. et al., 2013. *Negative Deceptive Opinion Spam*. [online]. Available at: https://www.cs.cornell.edu/home/cardie/papers/NAACL13-Negative.pdf [Accessed 12th April 2019].
20) Myle Ott. (2019). *Deceptive Opinion Spam Corpus v1.4*. [online]. Available at: https://myleott.com/op-spam [Accessed 12th April].
21) Blei, D. M., Ng, A. Y., and Jordan, M., I., 2003. *Latent Dirichlet Allocation*. [online]. Available at: http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf [Accessed 12th April 2019].
22) Jeremy Frimer. (2017). *Moral Foundations Dictionaries for Linguistic Analyses, 2.0*. [online]. Available at: http://www.jeremyfrimer.com/uploads/2/1/2/7/21278832/summary.pdf [Accessed 12th April 2019].