

Sentiment and Emotion Based Text Representation for Fake Reviews Detection

Alimuddin Melleng
Queen's University Belfast
amelleng01@qub.ac.uk

Anna-Jurek Loughrey
Queen's University Belfast
a.jurek@qub.ac.uk

Deepak P
Queen's University Belfast
deepaksp@acm.org

Abstract

Fake reviews are increasingly prevalent across the Internet. They can be unethical and harmful. They can affect businesses and mislead customers. As opinions on the Web are increasingly relied on, the detection of fake reviews has become more critical. In this study we explore the effectiveness of sentiment and emotions based representations for the task of building machine learning models for fake reviews detection. The experiment performed with three real-world datasets demonstrate that improved data representation can be achieved by combining sentiment and emotion extraction methods, as well as by performing sentiment and emotion analysis on a part-by-part basis by segmenting the reviews.

1 Introduction

The Internet has evolved into a content creation platform where people express their opinions and experiences. Online reviews written by users have significant impact on customers and companies. Potential customers often consult reviews before making a purchase. Reviews help potential customers to gain insights from other people's experiences, particularly in making choices on purchasing products or services. At the same time, companies need reviews on their products or services in order to get feedback and maintain good reputation. However, not all reviews available in the Internet are genuine. Profusion of reviews of questionable quality increase concerns about their trustworthiness. Moreover, users with malintent often post fake reviews (FR) to mislead customers by promoting or demoting products or target stores. Authors of FR can sway customer

choices towards companies with which they are associated, or against competitors making fake reviews a lucrative business. There has been an increase in FR profusion lately. According to the report of the Harvard Business School (Luca and Zervas, 2016) the percentage of fake reviews on YELP¹ increased from 5 % in 2006 to 20% in 2013. This makes FR detection an important challenge to be addressed.

FR were firstly categorized by Jindal et al. (2008) into three groups: (1) Untruthful opinions: mislead readers by giving positive reviews to promote or demote target object, (2) Reviews on brands only: the reviewer focus on the brands, producers or sellers of a product or service without commenting on the product or service, (3) Non-reviews: the reviews are irrelevant to the product and do not contain opinions but advertisements or questions. The first category is the most challenging type to detect, and that is the focus of our paper. Given the large numbers of reviews posted daily, automatic methods would be preferred over manual ones as illustrated in (Ott et al., 2011). Recent years have witnessed an increased impetus on machine learning methods for data-driven FR detection (Mukherjee et al., 2013; Ott et al., 2011; Rout et al., 2017)

The performance of machine learning models for detecting FR is heavily influenced by the data representation (or features) in their application (Bengio et al., 2013). Text analytics has conventionally focused on domains such as labelling news stories or grouping disease reports based on severity where the human authors of text documents are largely passive to the usage of downstream analytics. FR mitigation methods, on the other hand, are in direct conflict with the intents of FR peddlers, generating interest-

¹<https://www.yelp.co.uk/>

ing gamification dynamics. This makes it important for data-driven FR solutions to rely on more generic or higher-level data representations rather than simple lexical ones based on words, phrases and sentences. This is because FR filters using higher-level generic features may naturally be expected to be more robust and resistant to simple workarounds by FR authors such as word and phrase replacements. Further, higher-level features may have limited volatility across domains; thus, FR detection methods based on them may be more transferable across domains.

In this paper, we evaluate the effectiveness of emotion and sentiment based representations for the task of building machine learning models for FR detection. In particular, we illustrate that improved data representations can be achieved by leveraging a plurality of emotion and sentiment extraction methods, as well as by estimating emotions and sentiments on a part-by-part basis by segmenting the reviews. We illustrate the improved effectiveness of multiple emotion and sentiment features as well as review-segmented features by evaluating over real-world datasets.

2 Related Work

Representation learning focuses on developing a more instructive feature set for training a classification model that helps to boost the FR detection process (Li et al., 2017; Yilmaz and Durahim, 2018). Within past research, diverse features selection methods have been employed to detect FR. These may be divided into two classes: review-centric and reviewer-centric features. Reviewer-centric features are related to the reviewer's behaviour (Fontanarava et al., 2017) rather than the review itself. Those features include textual features, rating features, and temporal features. Review-centric features are derived from the content of a review. Commonly used review-centric features include Bag-of-words, TF-IDF (Term-frequency inverse-document-frequency), POS (part of speech) tags, word n-grams (Ahmed et al., 2018), and word embedding vectors (e.g. Word2vec, Doc2vec) (Krishnamurthy et al., 2018; Yilmaz and Durahim, 2018). A recent study by Jia et al. (2018) explored the application of linguistic features to distinguish between fake and non-fake reviews. They used Yelp filter dataset in their study and applied Term Frequency, Word2vec, and Latent Topic Distribution for data representation.

They trained three machine learning models i.e. SVM, Logistic Regression, and Multi-layer Perceptron and found that LDA+Logistic Regression and LDA+Multi-layer Perceptron performed better with 81.3% of accuracy.

With representations being only a means to enable better FR identification, it is useful to briefly outline the classification techniques that have been employed for FR detection. Ott et al., (2011) used word n-gram features in combination with a SVM classifier. Banerjee and Chua (2014) employed a Logistic Regression classifier over POS tags and writing style features (e.g., tense of words) for FR detection. Algur et al., (2010) explored a similarity-oriented method for FR detection over domain-specific product features.

As mentioned earlier, our representations are centred on emotion and sentiment based features. There has been very little prior work on using such features for FR detection. An early work in sentiment analysis for FR detection was conducted by Peng and Zhong (2014), whereas (K et al., 2019) explore utility of emotions in health fake news detection. Peng and Zhong (2014) chose SentiWordNet and MPQA lexicons and analysed sentiment on review and product features. In our experiment, we used IBM, Afinn, SenticNet, and Biu Liu lexicons. To our knowledge, this is the first study detecting FR by means of combination emotion and sentiment analysis. Taking cue from the previous work of FR detection, we use Random Forest classifier, in our experiments.

3 Methodology

In this section we describe our proposed approach to online FR detection using emotion and sentiment based text representation.

3.1 Emotion and Sentiment Analysis

For the purpose of sentiment and emotion analysis, we apply three different sentiment lexicons and one emotion analysis API.

- **IBM Watson Natural Language Understanding.** Natural Language Understanding (NLU)² is a collection of APIs that offer text analysis through natural language processing. One of the feature of IBM Watson NLU is emotion analysis. The API takes a text as an input and returns the category which the

²<https://www.ibm.com/services/natural-language-understanding/>

text belongs to, stored in a list variable: $\langle \text{KeyValuePair} \langle \text{String}, \text{Double} \rangle \rangle$ e.g. "emotion" : {"sadness":0.336228}. Each item in the list contains the category (emotion) name and the categorization score. IBM Watson NLU can detect five emotions: anger, disgust, fear, joy, and sadness. For example, for an input 'I love apples! I don't like oranges', the NLU API returns (sadness: 0.32665, joy: 0.563273, fear: 0.033387, disgust: 0.022637, anger: 0.041796).

- **SenticNet lexicon.** SenticNet³ performs tasks such as polarity detection and emotion recognition. Instead of merely relying on word co-occurrence frequencies, it leverages semantics and linguistics. This lexicon contains a list of words with their polarity and intensity values. The intensity is a float number between -1 and +1. For example, according to the SenticNet lexicon 'abandoned' is a negative word with intensity of -0.85. Each word in the lexicon is assigned with only one polarity and intensity value.
- **AFINN lexicon.** AFINN⁴ lexicon is a list of English terms rated with valence on a scale -5 (negative) and +5 (positive). This lexicon has been manually labelled by Finn Årup Nielsen (2011). AFINN provides two versions of lexicon: the newest version AFINN-111 with 2477 words and phrases and AFINN-96 with 1468 unique words and phrases on 1480 lines. Our experiment use AFINN-111 as it is the most up-to-date version.
- **Biu Liu lexicon.** Biu Liu⁵ lexicon consists of 6789 words including 2006 positive and 4783 negative words (Hu and Liu, 2004). This lexicon does not provide any sentiment scores and only provides positive/negative labels.

3.2 Representation Learning

In this work we explore whether sentiment and emotions extracted from a review can be used to train machine learning models for distinguishing between fake and non-fake reviews. We perform the sentiment/emotion analysis with different levels of granularity on a part-by-part basis by segmenting the reviews.

³<https://sentic.net/>

⁴<https://pypi.org/project/afinn/>

⁵<http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar>

3.2.1 Sentiment Based Representation

The process of constructing sentiment based representation of a review is presented in Algorithm 1. We first split a review into P segments, each one containing the same number of sentences. For example, if $P=4$, then we split a review into 4 segments. For each segment we identify all positive and all negative words using the lexicons. In the next step, all positive sentiment values and all negative sentiment values within the segment are accumulated together. In the case of AFINN and SenticNet, all positive and negative values are summed in each segment. For Biu Liu lexicon, all positive and all negative words are counted. Following this, the segment is represented by a two dimensional vector $[pos(s_i), neg(s_i)]$, where $pos(s_i)$ and $neg(s_i)$ represent the accumulated/counted positive and negative sentiment values. Finally, all P vectors (one generated for each segment) are concatenated. The concatenated vector is returned as the sentiment representation of the entire review. The process looks the same for all sentiment lexicons.

Algorithm 1 Sentiment Based Representation

Input: Review R , number of segments P , sentiment lexicon L

Output: Sentiment representation of R

- 1: Split R into P equal segments s_1, \dots, s_P
 - 2: **for all** s_1, \dots, s_P **do**
 - 3: Tokenise s_i into set of words W
 - 4: Retrieve sentiment values for all words in W using L
 - 5: Accumulate all positive sentiment values in W as $pos(s_i)$
 - 6: Accumulate all negative sentiment values in W as $neg(s_i)$
 - 7: $v_i = [pos(s_i), neg(s_i)]$
 - 8: **end for**
 - 9: $v(R) := [v_1, \dots, v_P]$
 - 10: **return** $v(R)$
-

3.2.2 Emotion Based Representation

The process of generating emotion based representation is presented in Algorithm 2. As in the case of the sentiment based representation, a review is first divided in P segments. All sentences in each segment is then passed to the IBM Watson API. As the output we obtain vector with the five emotions' scores. Finally, the emotion vectors obtained for all the segments are concatenated. The

output vector is returned as the emotion representation of the entire review.

Algorithm 2 Emotion based Representation

Input: Review R , number of segments P , emotion lexicon L

Output: Emotion representation of R

- 1: Split R into P equal segments s_1, \dots, s_P
 - 2: **for all** s_1, \dots, s_P **do**
 - 3: Get vector v_i with emotions scores from L
 - 4: **end for**
 - 5: $v(R) := [v_1, \dots, v_P]$
 - 6: **return** $v(R)$
-

3.2.3 Multi-Segment Based Representation

The process of multi-segment representation learning is presented in Algorithm 3. With this technique, the sentiment/emotion based representation is first generated for different numbers of segments $1 \dots P$. Following this, all vectors obtained for $p = 1 \dots P$ are concatenated to form the final representation. In this way, the output vector contains more granular information on the distribution of sentiment or emotions within a review.

Algorithm 3 Multi-Segment Representation

Input: Review R , maximum number of segments P , lexicon L

Output: Vector representation of R

- 1: **for all** $p \in 1 \dots P$ **do**
 - 2: Obtain $v_p(R)$ calling Algorithm 1 or 2 and passing R , p and L as parameters
 - 3: **end for**
 - 4: $v(R) := [v_1(R), \dots, v_P(R)]$
 - 5: **return** $v(R)$
-

3.2.4 Combined Sentiment and Emotion Based Representation

The last representation type that we explore is the combined sentiment and emotion based representation. The process is presented in Algorithm 4. First, a review is divided into P segments. The representation of each segment is generated by concatenation of sentiment and emotion representations obtained with Algorithms 1 and 2 respectively. Finally, representations of all segments are merged together.

Algorithm 4 Combined sentiment and Emotion Based Representation

Input: Review R , number of segments P , sentiment lexicon L_s , emotion lexicon L_e

Output: Vector representation of R

- 1: Split R into P equal segments s_1, \dots, s_P
 - 2: **for all** s_1, \dots, s_P **do**
 - 3: Get sentiment representation $V_s(R)$ applying Algorithm 1 with R , P and L_s
 - 4: Get emotion representation $V_e(R)$ applying Algorithm 2 with R , P and L_e
 - 5: $v_i = [V_s(R), V_e(R)]$
 - 6: **end for**
 - 7: $v(R) := [v_1, \dots, v_P]$
 - 8: **return** $v(R)$
-

4 Experimental Results and Discussion

In this section we present the experimental evaluation of the proposed four different sentiment/emotion based representations. Each of the representations are separately used to build a machine learning model for FR detection. We conducted an extensive set of experiments in order to answer the following key questions:

- Do sentiment/emotion based representations help in FR detection?
- Which of the proposed representations is the most effective for FR detection?
- Can higher sentiment/emotion granularity level improve the data representation?

4.1 Experimental Setup

Datasets. We collected our datasets from two different sources. We used gold standard spam review dataset from Ott et al., (2011), and Yelp dataset from Rayana and Akoglu (2015). The Ott dataset contains reviews about hotels. Yelp Zip and Yelp NYC are extracted from Yelp filtered dataset. Yelp NYC is a collection of reviews from restaurants located in New York City (NYC) while Yelp Zip is a collection of restaurant's reviews in zip code area in NY State. Each of the datasets contains true labels of the reviews, i.e. fake or non-fake label assigned to each review. Table 1 shows the size and class distribution for each of the datasets. In our experiment, we only consider reviews that contain more than 10 sentences. We presume that proposed representation learning techniques would not be effective for short reviews

since sparse text does not allow to identify emotions and sentiments well. Table 2 demonstrates the statistics of the datasets after filtering.

Dataset	Non-fake	Fake
YELP ZIP	528019	80439
YELP NYC	322097	36860
Ott	800	800

Table 1: Statistics of the datasets

Dataset	Non-fake	Fake
YELP ZIP	170261	15108
YELP NYC	105080	6185
Ott	340	270

Table 2: Statistics of the datasets after filtering

Learning. As the machine learning algorithm we used Random Forest (RF) given that it was reported as one of the most effective in FR detection (Chowdhary and Pandit, 2018; Saumya and Singh, 2018; Viviani and Pasi, 2017). However, any other learning algorithm can be applied instead. We set $n_estimator=100$ and $random_state=42$ for the RF parameter. All the experiments are performed with 5-fold cross-validation and the prediction performance is evaluated with application of F-measure. Given the very high class imbalance in the Yelp NYC and Yelp Zip, we randomly select number of non-fake reviews equal to the number of FR in order to balance the training data.

4.2 Sentiment and Emotion Granularity

In this section we investigate what level of granularity in terms of sentiment and emotion is the most representative for FR detection. Tables 3-5 demonstrate the F-measure obtained by RF with each of the datasets and sentiment and emotion based representations for reviews. For the parameter P we used values from 1 to 4. For each table, the first row represents results obtained by RF applied with the emotion based representation obtained with the IBM Watson API. The three bottom columns contain results obtained for the sentiment based representation generated with each of the three sentiment lexicons. Each column refers to a different value of parameter $P = 1 \times 4$. The last column presents results obtained for multi-segment based representation.

We can observe from the tables that in the majority of cases, the higher the granularity (P) the

better the prediction performance. It can also be noted that the multi-segment based representation tends to perform better than when a single segmentation is applied. The only exception is the Biu Liu lexicon, which for Yelp, Zip, and Ott obtained the best results for $P = 1$.

Lexicon	P=1	P=2	P=3	P=4	P1-4
IBM	0.570	0.584	0.589	0.584	0.597
SenticNet	0.506	0.510	0.522	0.523	0.524
Biu Liu	0.574	0.540	0.547	0.558	0.557
AFINN	0.550	0.542	0.549	0.555	0.563

Table 3: RF’s F-measure over Yelp ZIP dataset.

Lexicon	P=1	P=2	P=3	P=4	P1-4
IBM	0.554	0.569	0.578	0.569	0.584
SenticNet	0.511	0.520	0.523	0.525	0.526
Biu Liu	0.546	0.523	0.543	0.543	0.555
AFINN	0.524	0.529	0.541	0.544	0.557

Table 4: RF’s F-measure over Yelp NYC dataset.

Lexicon	P=1	P=2	P=3	P=4	P1-4
IBM	0.620	0.605	0.543	0.533	0.590
SenticNet	0.533	0.529	0.483	0.525	0.570
Biu Liu	0.618	0.561	0.592	0.576	0.600
AFINN	0.523	0.560	0.580	0.545	0.600

Table 5: RF’s F-measure over Ott dataset.

4.3 Sentiment vs. Emotion

In this section we compare the results obtained by RF applied with the sentiment and the emotion based representations of data. We can see from Tables 3-5 that IBM emotion lexicon obtained the best performance in comparison to the three sentiment lexicons in the Yelp Zip and NYC datasets. This may be considered unsurprising since emotions provide more fine grained information for the classifiers to work with. For the Ott dataset, Biu Liu and AFINN lexicon obtained better results for some of the greater values of P .

In order to perform better comparison between the sentiment and emotion based representations we calculated average of the results obtained for each of the granularity levels: $P1, P2, P3, P4, P1 - 4$. The results are demonstrated in Figure 1. We can observe from the graphs that the IBM emotion lexicon performs significantly better than any of the other sentiment lexicons apart from the Ott dataset where it is outperformed by the Biu Liu lexicon.

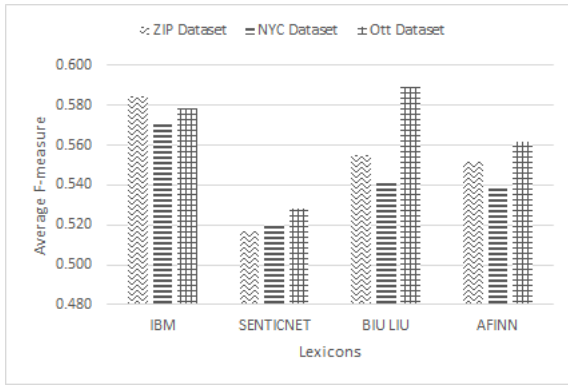


Figure 1: Average F-measure obtained for all values of P

4.4 Combined Sentiment and Emotion Based Representation

Table 6 demonstrates results obtained for the combined sentiment and emotion based representation generated according to the Algorithm 4. We can observe that for Zip and NYC datasets the best results were obtained when multi-segment based representation was applied. With Ott the best performance was obtained for $P1$.

Dataset	P1	P2	P3	P4	P1-4
ZIP	0.589	0.596	0.599	0.599	0.602
NYC	0.580	0.580	0.589	0.584	0.588
Ott	0.653	0.624	0.606	0.604	0.640

Table 6: F-measure obtained with combined sentiment and emotion representation learning.

In Figures 2-4 we compare the performance of the combined sentiment and emotion based representation with the emotion based representation, which so far obtained the most promising results. We can observe that the combined approach obtained better results in each case, with the difference in F-measure being quite significant for Zip and NYC datasets. This demonstrate that improved data representation can be achieved by applying combination of different emotion and sentiment extraction methods.

5 Conclusions and Future Work

In this paper, we analyzed the effectiveness of emotion and sentiment based representations estimated over varying text granularities, for the task of fake review classification. Through an empirical study across three real-world datasets, we find consistent evidence that combinations of emotions and sentiments work better than either of

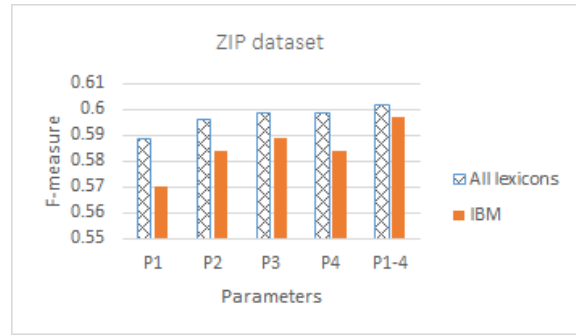


Figure 2: Combined sentiment-emotion vs. emotion representation learning for Zip

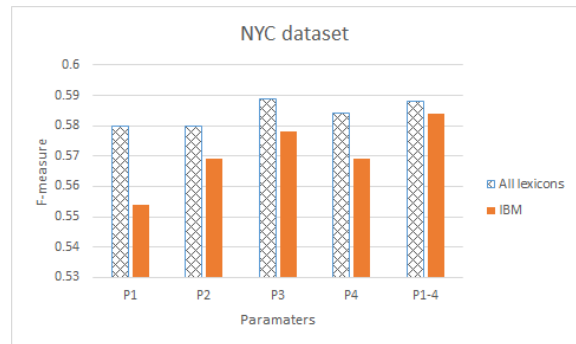


Figure 3: Combined sentiment-emotion vs. emotion representation learning for NYC

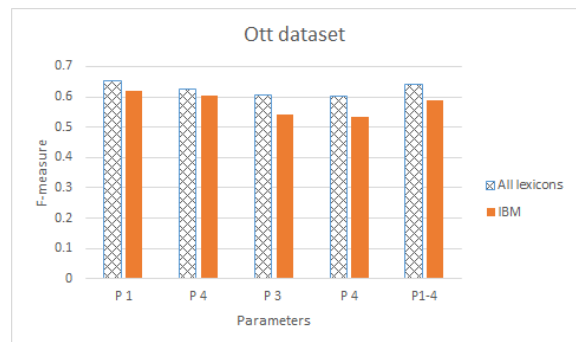


Figure 4: Combined sentiment-emotion vs. emotion representation learning for Ott

them separately. Further, we observe that combining emotion and sentiment representations obtained across different text granularities yields better accuracies over the restaurant review datasets. As future work, we plan to carry on research on cross domain between different datasets. We also want to observe how sentiment and emotion work on neural network model such as CNN and LSTM using generic as well as custom-built lexicons (Bandhakavi et al., 2017).

Acknowledgments

The authors thank LPDP (Indonesia Endowment Fund for Education) for funding this research.

References

- Ahmed, Hadeer, et al. "Detecting Opinion Spams and Fake News Using Text Classification." *Security and Privacy*, vol. 1, no. 1, 2017, p. e9, doi:10.1002/spy2.9.
- Algur, Siddu P., et al. "Conceptual Level Similarity Measure Based Review Spam Detection." *Proceedings of the 2010 International Conference on Signal and Image Processing, ICSIP 2010*, IEEE, 2010, pp. 416–23, doi:10.1109/ICSIP.2010.5697509.
- Anoop, K., et al. "Emotion Cognizance Improves Fake News Identification." *ArXiv Preprint ArXiv:1906.10365*, 2019.
- Bandhakavi, Anil, et al. "Lexicon Generation for Emotion Detection from Text." *IEEE Intelligent Systems*, vol. 32, no. 1, IEEE, 2017, pp. 102–08, doi:10.1109/MIS.2017.22.
- Banerjee, Snehasish, and Alton Y. K. Chua. "Applauses in Hotel Reviews: Genuine or Deceptive?" *Proceedings of 2014 Science and Information Conference, SAI 2014*, The Science and Information (SAI) Organization, 2014, pp. 938–42, doi:10.1109/SAI.2014.6918299.
- Bengio, Yoshua, et al. "Representation Learning: A Review and New Perspectives." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, IEEE, 2013, pp. 1798–828.
- Chowdhary, Neha S., and Anala A. Pandit. "Fake Review Detection Using Classification." *International Journal of Computer Applications*, vol. 180, no. 50, 2018, pp. 16–21.
- Fontanarava, Julien, et al. "Feature Analysis for Fake Review Detection through Supervised Classification." *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE, 2017, pp. 658–66.
- Hu, Minqing, and Bing Liu. "Mining and Summarizing Customer Reviews." *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 50, no. 08, 2004, pp. 50-4466-50-4466, doi:10.5860/choice.50-4466.
- Jia, Shaohua, et al. "Fake Reviews Detection Based on LDA." *2018 4th International Conference on Information Management, ICIM 2018*, IEEE, 2018, pp. 280–83, doi:10.1109/INFOMAN.2018.8392850.
- Jindal, Nitin, and Bing Liu. "Opinion Spam and Analysis." *Proceedings of the 2008 International Conference on Web Search and Data Mining*, ACM, 2008, pp. 219–30.
- Krishnamurthy, Gangeshwar, et al. "A Deep Learning Approach for Multimodal Deception Detection." *ArXiv Preprint ArXiv:1803.00344*, 2018.
- Kumar, Abhinav, et al. "Spotting Opinion Spammers Using Behavioral Footprints." *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2013, doi:10.1145/2487575.2487580.
- Li, Luyang, et al. "Document Representation and Feature Combination for Deceptive Spam Review Detection." *Neurocomputing*, vol. 254, Elsevier B.V., 2017, pp. 1339–51, doi:10.1016/j.neucom.2016.10.080.
- Luca, Michael, and Georgios Zervas. "Fake It till You Make It: Reputation." *Competition, and Yelp Review Fraud., SSRN Electronic Journal*, 2016.
- Nielsen, Finn Årup. "A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs." *CEUR Workshop Proceedings*, vol. 718, 2011, pp. 93–98, doi:10.1016/j.knosys.2015.06.015.
- Ott, Myle, et al. "Finding Deceptive Opinion Spam by Any Stretch of the Imagination." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, Association for Computational Linguistics, 2011, pp. 309–19.
- Peng, Qingxi, and Ming Zhong. "Detecting Spam Review through Sentiment Analysis." Peng, Q. and Zhong, M. (2014) 'Detecting Spam Review through Sentiment Analysis', *Journal of Software*. Doi: 10.4304/Jsw.9.8.2065-2072." *Journal of Software*, 2014, doi:10.4304/jsw.9.8.2065-2072.
- Rayana, Shebuti. "Collective Opinion Spam Detection: Bridging Review Networks and Metadata." *Proceedings of the 21th ACM SIGKDD*, 2015.
- Rout, Jitendra Kumar, et al. "Deceptive Review Detection Using Labeled and Unlabeled Data." *Multimedia Tools and Applications*, vol. 76, no. 3, Multimedia Tools and Applications, 2017, pp. 3187–211, doi:10.1007/s11042-016-3819-y.

- Saumya, Sunil, and Jyoti Prakash Singh. "Detection of Spam Reviews: A Sentiment Analysis Approach." *CSI Transactions on ICT*, vol. 6, no. 2, Springer, 2018, pp. 137–48.
- Viviani, Marco, and Gabriella Pasi. "Quantifier Guided Aggregation for the Veracity Assessment of Online Reviews." *International Journal of Intelligent Systems*, vol. 32, no. 5, Wiley Online Library, 2017, pp. 481–501.
- Yilmaz, Cennet Merve, and Ahmet Onur Durahim. "SPR2EP: A Semi-Supervised Spam Review Detection Framework." *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2018*, 2018, pp. 306–13, doi:10.1109/ASONAM.2018.8508314.