The Language of Deception

Subjects: Language & Linguistics

Contributor: Alibek Jakupov, Julien Longhi, Besma Zeddini

Digital forensic investigations are becoming increasingly crucial in criminal investigations and civil litigations, especially in cases of corporate espionage and intellectual property theft as more communication occurs online via e-mail and social media. Deceptive opinion spam analysis is an emerging field of research that aims to detect and identify fraudulent reviews, comments, and other forms of deceptive online content.

Keywords: digital investigation ; NLP-based forensics ; deceptive opinion spam ; feature engineering ; stylometry

1. Introduction

Digital communication mediums like emails and social networks are crucial tools for sharing information and communication, but they can also be misused for criminal and political purposes. A notable instance of this misuse was the spread of false information during the U.S. election. Lazer et al. highlighted that “misinformation has become viral on social media” (Lazer et al. 2018). They underscored the importance for researchers and other relevant parties to encourage cross-disciplinary studies aimed at curbing the propagation of misinformation and addressing the root issues it exposes. Reports and worries have also arisen about terrorists and other criminal groups taking advantage of social media to promote their unlawful endeavors, such as setting up discrete communication pathways to share information (Goodman 2018). Therefore, it is not unexpected that government bodies are closely scrutinizing these platforms or communication paths. Most existing studies focus on creating a map of individual relationships within a communication network. The primary goal in these methods is to pinpoint the closest associates of a known target. These methods aim to enhance precision, recall, and/or the F1 score, often overlooking the significance of the content within conversations or messages. As a result, these methods can be highly specific (tailored for particular outcomes), may lack accuracy, and may not be ideal for digital investigations (Keatinge and Keen 2020). For example, in the tragic incident at the Gilroy Garlic Festival, the shooter had reportedly expressed his anger on his Facebook page before the incident. This post, however, did not attract the attention of pertinent parties until after the tragedy. This lack of attention is not surprising, given that the shooter was not a recognized threat on the social network, and his post might not have been given high priority using traditional methods (Sun et al. 2021).

The example mentioned above demonstrates how written information can be employed to influence public opinion and impact the outcome of important events. There is a field within Natural Language Processing (NLP) that concentrates on scrutinizing on a similar phenomenon, called Deceptive Opinion Spam. Therefore, certain findings within this field could significantly enhance our comprehension of forensic linguistic analysis. Opinion Spam refers to reviews that are inappropriate or fraudulent, which can take on various forms such as self-promotion of an unrelated website or blog, or deliberate review fraud that could lead to monetary gain (Ott et al. 2011). Organizations have a strong incentive to detect and eliminate Opinion Spam via automation. This is because the primary concern with Opinion Spam is its influence on customer perception, particularly with regards to reviews that inaccurately praise substandard products or criticize superior ones (Vogler and Pearl 2020). Compared to other NLP tasks like sentiment analysis or intent detection, there has been relatively little research on using text classification approaches to detect Opinion Spam (Barsever et al. 2020). One can easily identify certain types of opinion spam, such as promotional content, inquiries, or other forms of non-opinionated text (Jindal and Liu 2008). The described situations can be classified as Disruptive Opinion Spam, characterized by irrelevant comments that are easily recognizable by the audience and pose a minimal threat, as individuals are empowered to disregard them if they so choose (Ott et al. 2011). When it comes to Deceptive Opinion Spam, which involves more nuanced forms of fake content, the task of identifying it is not as simple; the reason being that these statements are intentionally constructed to seem authentic and mislead the assessor (Ott et al. 2011). Deceptive Opinion Spam is a type of fraudulent behavior where a malicious user creates fictitious reviews, either positive or negative, with the intention of either boosting or damaging the reputation of a business or enterprise (Barsever et al. 2020). Thus, the deliberate intention to deceive readers in certain statements makes it challenging for human reviewers to accurately identify such deceptive texts, resulting in a success rate that is not significantly better than chance (Vogler and Pearl

2020). Consequently, discoveries in Deceptive Opinion Spam could prove valuable for designing digital investigation techniques for studying different communication channels, such as social networks. In contrast to traditional methods, the strategy that incorporates NLP techniques, particularly those used for Deceptive Opinion Spam analysis, places emphasis on both the interaction among individuals and the substance of the communication which may significantly improve the investigation process (Sun et al. 2021).

The problem is commonly addressed as a task of classifying text. Text classification systems typically consist of two key elements: a module for vectorization and a classifier. The vectorization module is tasked with creating features from a provided text sequence, while the classifier assigns category labels to the sequence using a set of matching features. These features are usually categorized into lexical and syntactic groups. Lexical features may include metrics such as total words or characters per word, as well as the frequency of long and unique words. On the other hand, syntactic features primarily consist of the frequency of function words or word groups, such as bag-of-words (BOW), n-grams, or Parts-Of-Speech (POS) tagging (Brown et al. 1992). In addition to vocabulary and sentence structure aspects, there are also methods known as lexicon containment techniques. These techniques symbolize the presence of a term from the lexicon in a text as a binary value, with positive indicating its existence and negative denoting its absence (Marin et al. 2014). The lexicons for such kind of features are constructed by a human expert (Pennebaker et al. 2001; Wilson et al. 2005) or generated automatically (Marin et al. 2010). Several approaches suggest integrating the text's morphological relationships and reliant linguistic components as input vectors for the classification algorithm (Brun and Hagege 2013). In addition to this, there are semantic vector space models which serve to characterize each word via a real-valued vector, determined using the distance or angle between pairs of word vectors (Sebastiani 2002). In the field of automatic fraudulent text detection, various approaches have been applied, mostly relying on linguistic features, such as n-grams (Fornaciari and Poesio 2013; Mihalcea and Strapparava 2009; Ott et al. 2011), discourse structure (Rubin and Vashchilko 2012; Santos and Li 2009), semantically related keyword lists (Burgoon et al. 2003; Pérez-Rosas et al. 2015), measures of syntactic complexity (Pérez-Rosas et al. 2015), stylometric features (Burgoon et al. 2003), psychologically motivated keyword lists (Almela et al. 2015), and parts of speech (Fornaciari and Poesio 2014; Li et al. 2014).

These vectorization strategies are typically utilized to examine the significance of the features, which helps to highlight recurring patterns in the framework of fraudulent statements that are less prevalent in truthful texts. Although this technique shows some effectiveness, it has significant drawbacks due to the difficulty in controlling the quality of the training set. For example, while many of the classification algorithms, trained using this method, show acceptable performance within their specific fields, they struggle to generalize effectively across different domains, thereby lacking resilience in adapting to domain changes. (Krüger et al. 2017). As an illustration, a mere alteration in the polarity of fraudulent hotel evaluations (that is, training the model on positive reviews while testing it on negative ones) has the potential to significantly reduce the F score (Ott et al. 2013). This observation holds when the training and the testing dataset originate from different domains (Mihalcea and Strapparava 2009). Additionally, specific categorization models that rely on semantic vector space models could be significantly influenced by social or personal biases embedded in the training data. This can lead the algorithm to make incorrect deductions. (Papakyriakopoulos et al. 2020). Furthermore, certain studies suggest that deceptive statements differ from truthful ones more in terms of their sentiment then other linguistic features (Newman et al. 2003). According to certain cases, the deceivers display a more positive affect in order to mislead the audience (Zhou et al. 2004), whereas certain instances demonstrate that deception is characterized by more words reflecting negative emotion (Newman et al. 2003).

Based on the evidence mentioned above, it can be inferred that feature extraction methodologies utilized in classical NLP tasks exhibit limited reliability when applied to forensic investigations. This is primarily due to their strong association with particular lexical elements (like n-grams and specific keywords) or linguistically abstract components that may not be directly influenced by the style of verbal deception (such as specific parts of speech, stylometric features, and syntactic rules) (Vogler and Pearl 2020). From this point of view, it is more favorable to develop a novel set of features based on domain-independent approaches like sentiment analysis or stylometric features, as it offers superior generalization capabilities and independence from the training dataset domain.

2. Deep Learning Methods in Social Networks

The idea of employing machine learning and deep learning methods to identify dubious activities in social networks has garnered general attention. For instance, Bindu et al. introduced an unsupervised learning method that can automatically spot unusual users in a static social network, albeit assuming that the network's structure does not change dynamically Bindu et al. (2017). Hassanpour et al. applied deep convolutional neural networks for images and long short-term memory (LSTM) to pull out predictive characteristics from Instagram's textual data, showing the capability to pinpoint potential substance use risk behaviors, aiding in risk evaluation and strategy formulation (Hassanpour et al. 2019). Tsikerdekis

used machine learning to spot fraudulent accounts trying to enter an online sub-community for prevention purposes (Tsikerdekis 2016). Ruan et al. also used machine learning to detect hijacked accounts based on their online social behaviors (Ruan et al. 2015). Fazil and Abulaish suggested a mixed method to detect automated spammers on Twitter, using machine learning to examine related aspects like community-based features (e.g., metadata, content, and interaction-based features) (Fazil and Abulaish 2018). Cresci et al. employed machine learning to spot spammers using digital DNA technology, with the social fingerprinting technique designed to distinguish between spam bots and genuine accounts in both supervised and unsupervised manners (Cresci et al. 2017). Other applications focused on urban crime perception utilizing the convolutional neural network as their learning preference (Fu et al. 2018; Shams et al. 2018).

Certain studies showed the potential of focusing purely on textual data, especially in the context of social network analysis (Ala'M et al. 2017). One example of this application was in 2013, when Keretna et al. used a text mining tool, Stanford POS tagger, to pull out features from Twitter posts that could indicate a user's specific writing style (Keretna et al. 2013). These features were then used in the creation of a learning module. Similarly, Lau et al. used both NLP and machine learning techniques to analyze Twitter data. They found that the Latent Dirichlet Allocation (LDA) and Support Vector Machine (SVM) methods yielded the best results in terms of the Area Under the ROC Curve (AUC) (Lau et al. 2014). In addition, Egele et al. developed a system to identify compromised social network accounts by analyzing message content and other associated features (Egele et al. 2015). Anwar and Abulaish introduced a unified social graph text mining framework for identifying digital evidence from chat logs based on user interaction and conversation data (Anwar and Abulaish 2014). Wang et al. treated each HTTP flow produced by mobile applications as text and used NLP to extract text-level features. These features were then used to create an effective malware detection model for Android viruses (Wang et al. 2017). Al-Zaidya et al. designed a method to efficiently find relevant information within large amounts of unstructured text data, visualizing criminal networks from documents found on a suspect's computer (Al-Zaidy et al. 2012). Lastly, Louis and Engelbrecht applied unsupervised information extraction techniques to analyze text data and uncover evidence, a method that could potentially find evidence overlooked by a simple keyword search (Louis and Engelbrecht 2011).

Li et al. applied their findings to detect fraudulent hotel reviews, using the Ott Deceptive Opinion spam corpus, and obtained a score of 81.8% by capturing the overall dissimilarities between truthful and deceptive texts (Li et al. 2014). The researchers expanded upon the Sparse Additive Generative Model (SAGE), which is a Bayesian generative model that combines both topic models and generalized additive models, and this resulted in the creation of multifaceted latent variable models via the summation of component vectors. Since most studies in this area focus on recognizing deceitful patterns instead of teaching a solitary dependable classifier, the primary difficulty of the research was to establish which characteristics have the most significant impact on each classification of a misleading review. Additionally, it was crucial to assess how these characteristics affect the ultimate judgment when they are paired with other attributes. SAGE is a suitable solution for meeting these requirements because it has an additive nature, which allows it to handle domain-specific attributes in cross-domain scenarios more effectively than other classifiers that may struggle with this task. The authors discovered that the BOW method was not as strong as LIWC and POS, which were modeled using SAGE. As a result, they formulated a general principle for identifying deceptive opinion spam using these domain-independent features. Moreover, unlike the creator of the corpus (Ott et al. 2011), they identified the lack of spatial information in hotel reviews as a potential indicator for identifying fraudulent patterns, of which the author's findings suggest that this methodology may not be universally appropriate since certain deceptive reviews could be authored by experts in the field. Although the research found that the domain-independent features were effective in identifying fake reviews with above-chance accuracy, it has also been shown that the sparsity of these features makes it difficult to utilize non-local discourse structures (Ren and Ji 2017); thus, the trained model may not be able to grasp the complete semantic meaning of a document.

(Ren and Ji 2017) built upon earlier work by introducing a three-stage system. In the first stage, they utilized a convolutional neural network to generate sentence representations from word representations. This was performed by employing convolutional action, which is commonly used to synthesize lexical n-gram information. To accomplish this step, they employed three convolutional filters. These filters are effective at capturing the contextual meaning of n-grams, including unigrams, bigrams, and trigrams. This approach has previously proven successful for tasks such as sentiment classification. (Wilson et al. 2005). Subsequently, they created a model of the semantic and discourse relations of these sentence vectors to build a document representation using a two-way gated recurrent neural network. These document vectors are ultimately utilized as characteristics to train a classification system. The authors achieved an 85.7% accuracy on the dataset created by Li et al. and showed that neural networks can be utilized to obtain ongoing document representations for the improved understanding of semantic features. The primary objective of this research was to practically show the superior efficacy of neural features compared to conventional discrete feature (like n-grams, POS, LIWC, etc.) due to their stronger generalization. Nevertheless, the authors' further tests showed that by combining

discrete and neural characteristics, the total precision can be enhanced. Therefore, discrete features, such as the combination of sentiments or the use of non-functional words, continue to be a valuable reservoir of statistical and semantic data.

(Vogler and Pearl 2020) conducted a study investigating the use of particular details in identifying disinformation, both within a single area and across various areas. Their research focused on several linguistic aspects, including n-grams, POS, syntactic complexity metrics, syntactic configurations, lists of semantically connected keywords, stylometric properties, keyword lists inspired by psychology, discourse configurations, and named entities. However, they found these features to be insufficiently robust and adaptable, especially in cases where the area may substantially differ. This is mainly because most of these aspects heavily rely on specific lexical elements like n-grams or distinct keyword lists. Despite the presence of complex linguistic aspects such as stylometric features, POS, or syntactic rules, the researchers consider these to be of lesser importance because they do not stem from the psychological basis of verbal deceit. In their research, they saw deceit as a product of the imagination. Consequently, in addition to examining linguistic methods, they also explored approaches influenced by psychological elements, like information management theory (Burgoon et al. 1996), information manipulation theory (McCornack 1992), and reality monitoring and criteria-based statement analysis (Vogler and Pearl 2020). Since more abstract linguistic cues motivated by psychology may have wider applicability across various domains (Kleinberg et al. 2018), the authors find it beneficial to use these indicators grounded in psychological theories of human deception. They also lean on the research conducted by Krüger et al. which focuses on identifying subjectivity in news articles and proposes that linguistically abstract characteristics could potentially be more robust when used on texts from different fields (Krüger et al. 2017). For their experiment, Vogler and Pearl employed three different datasets for the purpose of training and evaluation, accommodating shifts in the domain, ranging from relatively subtle to considerably extensive: the Ott Deceptive Opinion Spam Corpus (Ott et al. 2011), essays on emotionally charged topics (Mihalcea and Strapparava 2009), and personal interview questions (Burgoon et al. 1996). The linguistically defined specific detail features the authors constructed for this research proved to be successful, particularly when there were notable differences in the domains used for training and testing. These elements were rooted in proper nouns, adjective phrases, modifiers in prepositional phrases, exact numeral terms, and noun modifiers appearing as successive sequences. The characteristics were derived from appropriate names, descriptive phrase clusters, prepositional phrase changes, precise numerical terms, and noun modifiers that showed up as successive sequences. Each attribute is depicted as the total normalized number and the average normalized weight. The highest F score they managed to obtain was 0.91 for instances where content remained consistent, and an F score of 0.64 for instances where there was a significant domain transition. This suggests that the linguistically determined specific detail attributes display a broader range of application. Even though the classifier trained with these features showed fewer false negatives, it struggled to accurately categorize truthful texts. The experimental results clearly indicate that a combination of n-gram and language-specific detail features tends to be more dependable only when a false positive carries a higher cost than a false negative. It is worth noting that features based on n-grams might have a superior ability for semantic expansion when they are built on distributed meaning representations like GloVe and ELMo.

References

1. Lazer, David M. J., Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, and et al. 2018. The science of fake news. Science 359: 1094–96.
2. Goodman, Anka Elisabeth Jayne. 2018. When you give a terrorist a twitter: Holding social media companies liable for their support of terrorism. Pepperdine Law Review 46: 147.
3. Keatinge, Tom, and Florence Keen. 2020. Social media and (counter) terrorist finance: A fund-raising and disruption tool. In Islamic State's Online Activity and Responses. London: Routledge, pp. 178–205.
4. Sun, Dongming, Xiaolu Zhang, Kim-Kwang Raymond Choo, Liang Hu, and Feng Wang. 2021. Nlp-based digital forensic investigation platform for online communications. Computers & Security 104: 102210.
5. Ott, Myle, Yejin Choi, Claire Cardie, and Jeffrey T. Hancock. 2011. Finding deceptive opinion spam by any stretch of the imagination. arXiv arXiv:1107.4557.
6. Vogler, Nikolai, and Lisa Pearl. 2020. Using linguistically defined specific details to detect deception across domains. Natural Language Engineering 26: 349–73.
7. Barsever, Dan, Sameer Singh, and Emre Neftci. 2020. Building a better lie detector with bert: The difference between truth and lies. Paper presented at 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, July 19–24; pp. 1–7.

8. Jindal, Nitin, and Bing Liu. 2008. Opinion spam and analysis. Paper presented at 2008 International Conference on Web Search and Data Mining, Palo Alto, CA, USA, February 11–12; pp. 219–30.

9. Brown, Peter F., Vincent J. Della Pietra, Peter V. Desouza, Jennifer C. Lai, and Robert L. Mercer. 1992. Class-based n-gram models of natural language. Computational Linguistics 18: 467–80.

10. Marin, Alex, Roman Holenstein, Ruhi Sarikaya, and Mari Ostendorf. 2014. Learning phrase patterns for text classification using a knowledge graph and unlabeled data. Paper presented at Fifteenth Annual Conference of the International Speech Communication Association, Singapore, September 14–18.

11. Pennebaker, James W., Martha E. Francis, and Roger J. Booth. 2001. Linguistic Inquiry and Word Count: Liwc 2001. Mahway: Lawrence Erlbaum Associates, vol. 71.

12. Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. Paper presented at Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Vancouver, BC, Canada, October 6–8; pp. 347–54.

13. Marin, Alex, Mari Ostendorf, Bin Zhang, Jonathan T. Morgan, Meghan Oxley, Mark Zachry, and Emily M. Bender. 2010. Detecting authority bids in online discussions. Paper presented at 2010 IEEE Spoken Language Technology Workshop, Berkeley, CA, USA, December 12–15; pp. 49–54.

14. Brun, Caroline, and Caroline Hagege. 2013. Suggestion mining: Detecting suggestions for improvement in users' comments. Research in Computing Science 70: 5379–62.

15. Sebastiani, Fabrizio. 2002. Machine learning in automated text categorization. ACM Computing Surveys (CSUR) 34: 1–47.

16. Fornaciari, Tommaso, and Massimo Poesio. 2013. Automatic deception detection in italian court cases. Artificial Intelligence and Law 21: 303–40.

17. Mihalcea, Rada, and Carlo Strapparava. 2009. The lie detector: Explorations in the automatic recognition of deceptive language. Paper presented at ACL-IJCNLP 2009 Conference Short Papers, Singapore, August 4; pp. 309–12.

18. Rubin, Victoria L., and Tatiana Vashchilko. 2012. Identification of truth and deception in text: Application of vector space model to rhetorical structure theory. Paper presented at Workshop on Computational Approaches to Deception Detection, Avignon, France, April 23; pp. 97–106.

19. Santos, Eugene, and Deqing Li. 2009. On deception detection in multiagent systems. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 40: 224–35.

20. Burgoon, Judee K., J. Pete Blair, Tiantian Qin, and Jay F. Nunamaker. 2003. Detecting deception through linguistic analysis. Paper presented at Intelligence and Security Informatics: First NSF/NIJ Symposium, ISI 2003, Tucson, AZ, USA, June 2–3; Proceedings 1. Berlin/Heidelberg: Springer, pp. 91–101.

21. Pérez-Rosas, Verónica, Mohamed Abouelenien, Rada Mihalcea, and Mihai Burzo. 2015. Deception detection using real-life trial data. Paper presented at 2015 ACM on International Conference on Multimodal Interaction, Seattle, WA, USA, November 9–13; pp. 59–66.

22. Almela, Ángela, Gema Alcaraz-Mármol, and Pascual Cantos. 2015. Analysing deception in a psychopath's speech: A quantitative approach. DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada 31: 559–72.

23. Fornaciari, Tommaso, and Massimo Poesio. 2014. Identifying fake amazon reviews as learning from crowds. Paper presented at 14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg, Sweden, April 26–30; Toronto: Association for Computational Linguistics, pp. 279–87.

24. Li, Jiwei, Myle Ott, Claire Cardie, and Eduard Hovy. 2014. Towards a general rule for identifying deceptive opinion spam. Paper presented at 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, MD, USA, June 22–27; pp. 1566–76.

25. Krüger, Katarina R., Anna Lukowiak, Jonathan Sonntag, Saskia Warzecha, and Manfred Stede. 2017. Classifying news versus opinions in newspapers: Linguistic features for domain independence. Natural Language Engineering 23: 687–707.

26. Ott, Myle, Claire Cardie, and Jeffrey T. Hancock. 2013. Negative deceptive opinion spam. Paper presented at 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Atlanta, Georgia, June 9–14; pp. 497–501.

27. Papakyriakopoulos, Orestis, Simon Hegelich, Juan Carlos Medina Serrano, and Fabienne Marco. 2020. Bias in word embeddings. Paper presented at 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, January 27–30; pp. 446–57.

28. Newman, Matthew L., James W. Pennebaker, Diane S. Berry, and Jane M. Richards. 2003. Lying words: Predicting deception from linguistic styles. Personality and Social Psychology Bulletin 29: 665–75.

29. Zhou, Lina, Judee K. Burgoon, Douglas P. Twitchell, Tiantian Qin, and Jay F. Nunamaker Jr. 2004. A comparison of classification methods for predicting deception in computer-mediated communication. Journal of Management Information Systems 20: 139–66.

30. Bindu, P. V., P. Santhi Thilagam, and Deepesh Ahuja. 2017. Discovering suspicious behavior in multilayer social networks. Computers in Human Behavior 73: 568–82.

31. Hassanpour, Saeed, Naofumi Tomita, Timothy DeLise, Benjamin Crosier, and Lisa A. Marsch. 2019. Identifying substance use risk based on deep neural networks and instagram social media data. Neuropsychopharmacology 44: 487–94.

32. Tsikerdekis, Michail. 2016. Identity deception prevention using common contribution network data. IEEE Transactions on Information Forensics and Security 12: 188–99.

33. Ruan, Xin, Zhenyu Wu, Haining Wang, and Sushil Jajodia. 2015. Profiling online social behaviors for compromised account detection. IEEE Transactions on Information Forensics and Security 11: 176–87.

34. Fazil, Mohd, and Muhammad Abulaish. 2018. A hybrid approach for detecting automated spammers in twitter. IEEE Transactions on Information Forensics and Security 13: 2707–19.

35. Cresci, Stefano, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi. 2017. Social fingerprinting: Detection of spambot groups through dna-inspired behavioral modeling. IEEE Transactions on Dependable and Secure Computing 15: 561–76.

36. Fu, Kaiqun, Zhiqian Chen, and Chang-Tien Lu. 2018. Streetnet: Preference learning with convolutional neural network on urban crime perception. Paper presented at 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle, WA, USA, November 6–9; pp. 269–78.

37. Shams, Shayan, Sayan Goswami, Kisung Lee, Seungwon Yang, and Seung-Jong Park. 2018. Towards distributed cyberinfrastructure for smart cities using big data and deep learning technologies. Paper presented at 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS), Vienna, Austria, July 2–6; pp. 1276–83.

38. Ala'M, Al-Zoubi, Ja'far Alqatawna, and Hossam Paris. 2017. Spam profile detection in social networks based on public features. Paper presented at 2017 8th International Conference on information and Communication Systems (ICICS), Irbid, Jordan, April 4–6; pp. 130–35.

39. Keretna, Sara, Ahmad Hossny, and Doug Creighton. 2013. Recognising user identity in twitter social networks via text mining. Paper presented at 2013 IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, October 13–16; pp. 3079–82.

40. Lau, Raymond Y. K., Yunqing Xia, and Yunming Ye. 2014. A probabilistic generative model for mining cybercriminal networks from online social media. IEEE Computational Intelligence Magazine 9: 31–43.

41. Egele, Manuel, Gianluca Stringhini, Christopher Kruegel, and Giovanni Vigna. 2015. Towards detecting compromised accounts on social networks. IEEE Transactions on Dependable and Secure Computing 14: 447–60.

42. Anwar, Tarique, and Muhammad Abulaish. 2014. A social graph based text mining framework for chat log investigation. Digital Investigation 11: 349–62.

43. Wang, Shanshan, Qiben Yan, Zhenxiang Chen, Bo Yang, Chuan Zhao, and Mauro Conti. 2017. Detecting android malware leveraging text semantics of network flows. IEEE Transactions on Information Forensics and Security 13: 1096–1109.

44. Al-Zaidy, Rabeah, Benjamin C. M. Fung, Amr M. Youssef, and Francis Fortin. 2012. Mining criminal networks from unstructured text documents. Digital Investigation 8: 147–60.

45. Louis, A. L., and Andries P. Engelbrecht. 2011. Unsupervised discovery of relations for analysis of textual data. Digital Investigation 7: 154–71.

46. Ren, Yafeng, and Donghong Ji. 2017. Neural networks for deceptive opinion spam detection: An empirical study. Information Sciences 385: 213–24.

47. Burgoon, Judee K., David B. Buller, Laura K. Guerrero, Walid A. Afifi, and Clyde M. Feldman. 1996. Interpersonal deception: Xii. information management dimensions underlying deceptive and truthful messages. Communications Monographs 63: 50–69.

48. McCornack, Steven A. 1992. Information manipulation theory. Communications Monographs 59: 1–16.

49. Kleinberg, Bennett, Maximilian Mozes, Arnoud Arntz, and Bruno Verschuere. 2018. Using named entities for computer-automated verbal deception detection. Journal of Forensic Sciences 63: 714–23.