A Framework for Fake Review Detection: Issues and Challenges

Jitendra kumar Rout
KIIT Deemed to be University
Bhubaneswar, INDIA
Email: jitu2rout@gmail.com

Amiya Kumar Dash
KIIT Deemed to be University
Bhubaneswar, INDIA
Email: dash.amiyakumar@gmail.com

Niranjan Kumar Ray
KIIT Deemed to be University
Bhubaneswar, INDIA
Email: rayniranjan@gmail.com

Abstract—With evolution of 4G and availability of wireless Internet access at a much higher speed, makes it possible for developing countries to use E-commerce for getting product and services. As each and every company now-a-days have presence in online mode of marketing, getting the right product as well as service is also tough. This leads to the importance of online reviews on the Internet. For taking purchase as well as financial decisions, every individual have to depend on online reviews. Online reviews given by users regarding a particular product or service may not be always genuine. Some companies as well as individuals trick the reviews to promote a specific product or brand and demote its competitors. A little has been done in past to address this issue and still companies as well as researches are trying to get-rid of this. In this work, we have tried our best to summarize the overall issues as well as challenges for detection of fake reviews as well as fake reviewers. Finally, a framework has been proposed to deal with fake reviews.

Index Terms—Fake Reviews; Opinion Spammers; Machine Learning; Review Spammer; E-commerce.

I. Introduction

Growth of Internet technology leads to the growth of e-commerce as well as associated review sites. Gradually people used to prefer online mode of marketing. Availability of millions of products and services on e-commerce sites makes it difficult to search for the best suitable product according to requirement. The best way to overcome this is to follow the opinion of others who have already tried the product. A large number of review sites are available indifferent domains shown in the Table I. Apart from many independent review sites few companies have their own review system. The huge amount of user generated content provided by these sites provides a lot of invaluable information regarding various products and services. Nowadays an average customer used to read reviews of others before taking any financial or purchase decisions. In other words, these opinions influence perspective customer to take purchase decisions or reverse the purchase decision. While online reviews are helpful but blind trust may lead to dangerous consequences for both the seller as well as buyer. Unfortunately, driven by the desire of profit or promotion, reviews may also be faked. Business owners might give incentives to write favorable reviews about their product and negative reviews about their competitors. These fake reviews are called review spam and person who write it are called review spammers. A review spam may also be referred as opinion spams, fake reviews, deceptive reviews, fraudulent reviews, or non-genuine reviews etc.

TABLE I
Different Review sites for products and business

Domain	Review Sites
Electronics product	Cnet.com, Digital Trends
General product	ConsumerSearch.com, ConsumerReports.org, Mouthshut, Buzzillions, ReviewCentre.com
Consumer Products	Amazon, Youtube, Which? Choice, Trustpilot, TestFreaks
Household goods and appliances	Viewpoints, Good Housekeeping
Photography	DPReview
Companies and Management	Glassdoor
Bike	RoadBikeReview
Food and Restaurant	Zomato, America's Test Kitchen, OpenTable
Consumer focused reviews	TrustedReviews, Epinions, ConsumerAffairs,Yelp
Travel and Hotel	TripAdvisor
Health and well-being	ConsumerHealthDigest
Business	Yelp, Google My Business, Yahoo! Local Listings, Foursquare, Angie's List (Local Business),
Home services	Home Advisor

A lot of work has been done on spam detection such as Email Spam [1], SMS Spam [2], Web Spam [3], Social Media Spam [4], Search Engine Spam [5], video Spam [6, 7] etc. Opinion Spams however bit different and are basically found in product review sites. The intention is to give positive reviews regarding a specific product for profit and promotion and give unjustified negative ones to demote the competing brand or product. The opinion spam problem is first formulated by Jindal and Liu [8, 9] in 2007 in the context of product reviews and hence believed to be the first documented research in this domain. There has been a considerable growth of deceptive reviews overtime, starting from individual spammers to group spammers both are on rise. Unless detected or removed, it will damage the e-commerce business to a great extent and the social media which is believed to be a trusted source of public opinion may lose its luster. However, over the past few years, both industry as research community have made significant contribution in combating opinion spamming. Despite that, the problem is still huge and needs significant research advancement. Some of the issues and challenges that need to be addressed are as follows:

- Opinion Spams are highly implicit and pretend to be honest opinions voiced by actual users or customers.

978-1-7281-0259-7/18/$31.00 ©2018 IEEE
DOI 10.1109/ICIT.2018.00014

It is logically impossible to recognize fake reviews by simply reading it.

- Availability of labeled data sets (Most of the user generated contents are unlabeled).
- Appropriate feature selection.
- Early detection (i.e. how to detect the reviews immediately after it is being posted?)
- Cross-site verification (compare reviews of the same product across multiple sites).
- How to deal with the enormous amount of data generated and gather from review as well as e-commerce sites.

The rest of this paper has been organized as follows. Novel contributions of the paper were discussed in Section II. Section III presents a review of the related works done in the field of review spam detection. A framework has been proposed to deal with the issues in Section IV. We present and discuss our findings in Section V. The paper has been concluded and future scopes have been indicated in Section VI.

II. Contributions of the work

The contribution of the work are as follows

- Issues and challenges were outlined from literature.
- A framework has been proposed to deal with the issues.
- Possible features were explored and discussed.
- With an available dataset [10, 11], the experiments were carried out and results were compiled.
- Possible future directions were discussed.

III. Related Work

The opinion spam problem was first formulated and documented by Jindal et al. [8, 12] in the context of online reviews. The research is broadly classified into two types: Spam(fake) review detection and review spammer detection. In this work, we solely focus on the spam review detection. Spam review detection can be categorized into three types based on the types of data to be used i.e. Supervised learning (based on labeled data), Unsupervised learning (based on unlabeled data) and Semi-supervised learning (based on both labeled and unlabeled data). Basically, fake review detection is a classification problem [8–11, 13–17], but the key difficulties lies in availability of labeled datasets. The process of labeling is also a time consuming and tedious process followed by the most difficult task of validation. So, few authors have tried unsupervised learning approach [18, 19] which gives a broad view of categories of the reviews. However, to derive a clear conclusion with the huge amount of unlabeled data and with few labeled ones, semi-supervised model [20–24] was preferred. Afterwards, few works have been proposed to find the fake reviews from the graphs on relationship between reviews, reviewers and store information [25–27].

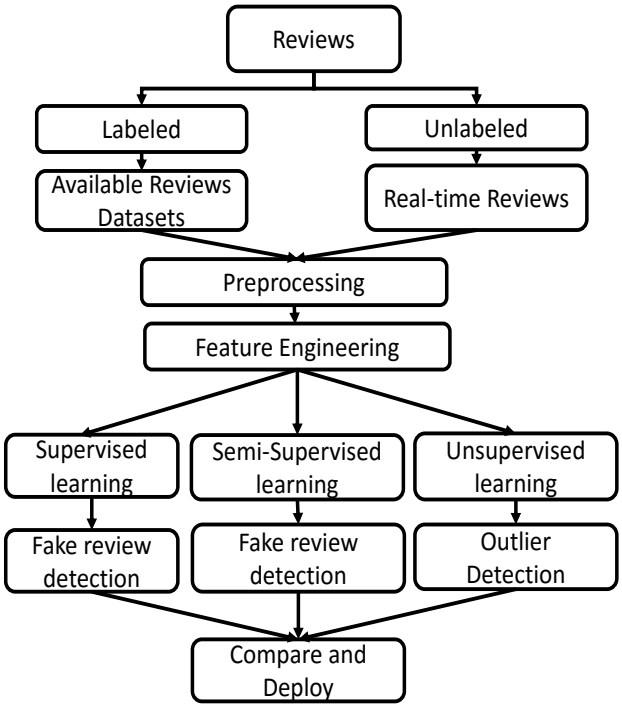

Fig. 1. Proposed framework for fake review detection

IV. Proposed Framework

Along with the proposed framework, this section describes the types of data used for deception detection and feature engineering. Even if the dataset we have used has 1600 instances (which is claimed to be the only available gold standard dataset in the domain) but in reality very few labeled data is available and most of the reviews produced are unlabeled. So, semi-supervised which uses minimum labeled data to provide label for huge unlabeled data is the most appropriate one for deception detection. The proposed Framework is shown in the Figure 1.

A. Types of Data

The available data for fake review detection can be categorized into three types:

- Review Content
- Meta Data about each review
- Product Information

Review Content is the actual content of a review including the review title. From the content various linguistic features can be extracted such as word, POS tags, n-grams and different, semantic and stylometric/syntactic features. Metadata (information about the reviews rather than content of review) about reviews includes reviewer-id, review-id, ratings, time/date of the review, helpfulness score, no of reviews written by a reviewer etc. Product information includes product description, brand, sales information etc.

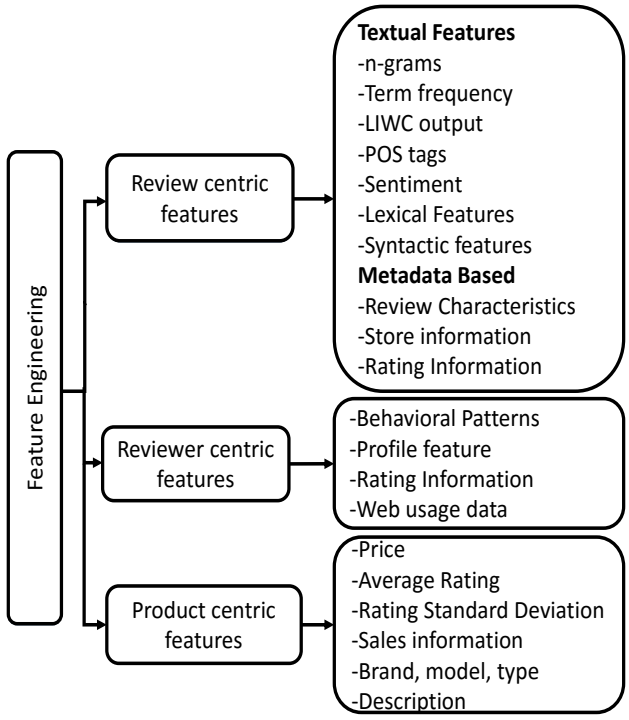

Fig. 2. Common features used in the domain

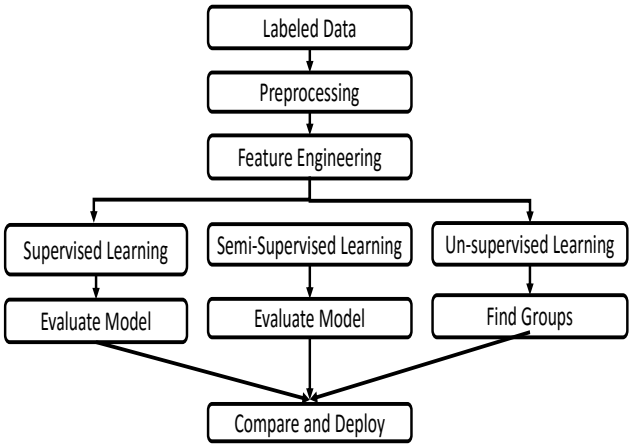

Fig. 3. Framework experimented with the Available Dataset

TABLE II
Results obtained using Supervised Learning

Supervised Learning		
Features Used	Learner	Accuracy
n-gram, POS tags, Opinion Polarity,	SVM	88.67
	Naïve Bayes	90.19
Lexical and Syntactic features	Decision Tree	90.02

B. Feature Engineering

Feature Engineering is the selection or construction of features from the available data. Commonly used features which are used in the domain of spam review detection can be categorized into three types:

- Review Centric Features
- Reviewer Centric Features
- Product Centric Features

Figure 2 describes the different types of features used in this domain.

C. Dataset Description

In this work, the 'gold standard' dataset by Ott et al. [10, 11] is used for evaluation of the proposed framework. The dataset comprises 1600 reviews on 20 Chicago-based hotels. The corpus contains 80 reviews for each hotel containing 40 spam and 40 non-spam reviews i.e. 800 deceptive and 800 genuine reviews. Each reviews consist of a unique ID, Hotel name, content of review, polarity of the review, binary label for depicting whether a review is a spam or not. For experimental purpose we have considered 400 from each category. The Framework shown in Figure 3 is designed to deal with both labeled data (available dataset) as well as unlabeled data (possibly real time reviews). It is derived version of the first framework is for experimental verification and validation. The data set used [10, 11] is assumed to unlabeled in case of unsupervised learning (after removing label) and for semi supervised learning dataset is partitioned into subset with size in the ratio 80:20 and (0.2*80) % is assumed to be labeled and rest as unlabeled data.

V. Results and Discussion

For unsupervised learning which is basically used to find out outliers in case of spam detection, is used in this case to check the existence of two natural clusters (i.e. Deceptive and Genuine). K-mean clustering is used for the purpose. For supervised and semi-supervised leaning methods the results were shown in the Table II and Table III respectively. In case of supervised learning we got an accuracy of 90.19% while in case of semi-supervised learning it is 83.70%.

TABLE III
Results obtained using Supervised Learning

Semi-Supervised Learning			
Features Used	Algorithms Used	Learner	Accuracy
n-gram, POS tags, Opinion Polarity, LIWC output	Co-Training	k-NN	74.66
		Logistic Regression	50.94
		Random Forest	71.89
	PU Learning	k-NN	82.49
		Logistic Regression	83.7
		Random Forest	60.17
	EM Algorithm	k-NN	83.27
		Logistic Regression	82.76
		Random Forest	70.5

VI. Conclusion & Future Work

With the growth of usage of social media/ review sites for purchase/business decisions, opinion spamming also gaining its pace and fake reviewers are also becoming more careful for review writings. Major online sites (Like Amazon, yelp etc.) have developed their own mechanism to deal with fake reviews but still a lot has to be done with time. However, it is too hard to make a system completely free of opinion spamming. Our proposed framework can be used to deal with both labelled and unlabeled data but for dealing with huge amount of data Big Data analysis techniques need to be integrated with it. From the experiments for the framework, we have obtained an accuracy of 90.19% for Supervised learning and 83.70% for Semi-Supervised Learning. Some of the possible future works are: i. More research needs to be done in other forms of social networks like discussion forums (quora, stack exchange etc.), blogs, microblogs, Facebook, Youtube. Each of these forums have its own review types and different features needs to be explored for deceptive detection. ii. Multiple site comparison needs to be done for the same product and brand including same reviewer (if there). iii. Early detection (immediately after its posting) of the fake reviews needs to be done for minimizing the damage to business and company. iv. As huge amount of data (data streams) are being generated by the review sites as well as other social forums, Big Data techniques need to be explored.

References

[1] I. Fette, N. Sadeh, and A. Tomasic, "Learning to detect phishing emails," in Proceedings of the 16th International Conference on World Wide Web, 2007, pp. 649–656, DOI:10.1145/1242572.1242660.

[2] A. Karami and L. Zhou, "Improving static sms spam detection by using new content-based features," in Proceedings of the 20th Americas Conference on Information Systems, 2014.

[3] J. Karimpour, A. A. Noroozi, and S. Alizadeh, "Web spam detection by learning from small labeled samples," International Journal of Computer Applications, vol. 50, no. 21, 2012.

[4] X. Zheng, Z. Zeng, Z. Chen, Y. Yu, and C. Rong, "Detecting spammers on social networks," Neurocomputing, vol. 159, pp. 27–34, 2015.

[5] L. Becchetti, C. Castillo, D. Donato, R. Baeza-Yates, and S. Leonardi, "Link analysis for web spam detection," ACM Transactions on the Web (TWEB), vol. 2, p. 2, 2008.

[6] A. Hess and J. Klaue, "A video-spam detection approach for unprotected multimedia flows based on active networks," in Euromicro Conference, 2004. Proceedings. 30th, 2004, pp. 461–465.

[7] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida, C. Zhang, and K. Ross, "Identifying video spammers in online social networks," in Proceedings of the 4th international workshop on Adversarial information retrieval on the web, 2008, pp. 45–52.

[8] N. Jindal and B. Liu, "Analyzing and detecting review spam," in Proceedings of the Seventh IEEE International Conference on Data Mining, 2007, pp. 547–552, DOI: 10.1109/ICDM.2007.68.

[9] ——, "Review spam detection," in Proceedings of the 16th international conference on World Wide Web, 2007, pp. 1189–1190, DOI: 10.1145/1242572.1242759.

[10] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, 2011, pp. 309–319.

[11] M. Ott, C. Cardie, and J. T. Hancock, "Negative deceptive opinion spam," in Proceedings of North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2013, pp. 497–501.

[12] N. Jindal and B. Liu, "Opinion spam and analysis," in Proceedings of the 2008 International Conference on Web Search and Data Mining, 2008, pp. 219–230, DOI: 10.1145/1341531.1341560.

[13] C. Lai, K. Xu, R. Y. Lau, Y. Li, and L. Jing, "Toward a language modeling approach for consumer review spam detection," in Proceedings of IEEE 7th International Conference on e-Business Engineering, 2010, pp. 1–8, DOI: 10.1109/ICEBE.2010.47.

[14] S. P. Algur, A. P. Patil, P. Hiremath, and S. Shivashan, "Conceptual level similarity measure based review spam detection," in International Conference on Signal and Image Processing, 2010, pp. 416–423, DOI: 10.1109/ICSIP.2010.5697509.

[15] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, "Spotting opinion spammers using behavioral footprints," in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, 2013, pp. 632–640, DOI: 10.1145/2487575.2487580.

[16] S. Shojaee, M. A. A. Murad, A. Bin Azman, N. M. Sharef, and S. Nadali, "Detecting deceptive reviews using lexical and syntactic features," in Proceedings of 13th International Conference on Intelligent Systems Design and Applications, 2013, pp. 53–58, DOI:10.1109/ISDA.2013.6920707.

[17] J. K. Rout, S. Singh, S. K. Jena, and S. Bakshi, "Deceptive review detection using labeled and unlabeled data," Multimedia Tools and Applications, pp. 1–25, 2016, DOI:10.1007/s11042-016-3819-y.

[18] G. Wu, D. Greene, B. Smyth, and P. Cunningham, "Distortion as a validation criterion in the identification of suspicious reviews," in Proceedings of the 1st Workshop on Social Media Analytics, 2010, pp. 10–13, DOI:10.1145/1964858.1964860.

[19] R. Y. Lau, S. Liao, R. C. W. Kwok, K. Xu, Y. Xia, and Y. Li, "Text mining and probabilistic language modeling for online review spam detecting," ACM Transactions on Management Information Systems, vol. 2, no. 4, pp. 1–30, 2011, DOI:10.1145/2070710.2070716.

[20] J. K. Rout, A. Dalmia, K.-K. R. Choo, S. Bakshi, and S. K. Jena, "Revisiting semi-supervised learning for online deceptive review detection," IEEE Access, pp. 1–1, 2017, DOI:10.1109/ACCESS.2017.2655032.

[21] D. Hernández, R. Guzmán, M. Móntes y Gomez, and P. Rosso, "Using PU-learning to detect deceptive opinion spam," in Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2013, pp. 38–45.

[22] W. Zhang, C. Bu, T. Yoshida, and S. Zhang, "CoSpa: A co-training approach for spam review identification with support vector machine," Information, vol. 7, no. 1, p. 12, 2016, DOI:10.3390/info7010012.

[23] H. Li, B. Liu, A. Mukherjee, and J. Shao, "Spotting fake reviews using positive-unlabeled learning," Computación y Sistemas, vol. 18, no. 3, pp. 467–475, 2014, DOI:10.13053/CyS-18-3-2035.

[24] Y. Ren, D. Ji, and H. Zhang, "Positive unlabeled learning for deceptive reviews detection." in EMNLP, 2014, pp. 488–498.

[25] G. Wang, S. Xie, B. Liu, and S. Y. Philip, "Review graph based online store review spammer detection," in Proceedings of the 11th IEEE International Conference on Data Mining, 2011, pp. 1242–1247, DOI:10.1109/ICDM.2011.124.

[26] G. Wang, S. Xie, B. Liu, and P. S. Yu, "Identify online store review spammers via social review graph," ACM Transactions on Intelligent Systems and Technology, pp. 61:1–61:21, 2012, DOI:10.1145/2337542.2337546.

[27] D. Liang, X. Liu, and H. Shen, "Detecting spam reviewers by combing reviewer feature and relationship," in International Conference on Informative and Cybernetics for Computational Social Systems (ICCSS), 2014, pp. 102–107.