Online Fake Review Detection Techniques : A Systematic Review

Akanksha Singh
Shri Shankaracharya group of Institution
Dept. of Computer Science & Engineering
Bhilai, Chhattisgarh, India
Akanksha13008@gmail.com

Prof. Dr. Siddharth Choubey
Shri Shankaracharya Group of Institution
Dept. of Computer Science & Engineering
Bhilai, Chhattisgarh

Prof. Dr. Abha Choubey
Shri Shankaracharya Group of Institution
Dept. of Computer Science & Engineering
Bhilai, Chhattisgarh, India
Abha.is.shukla@gmail.com

Abstract

A fantastic source of collecting the reviews on particular item is different web based shopping sites where individuals share their reviews on items and their shopping experiences. Individuals may get through the wrong conclusions known as review spam. Hence, for this it is fundamental to distinguish it by a few means. In this paper, an overview is done on the different methods acquainted with distinguish the review spam with their outcomes.

Keywords— *Review spam, un-truthful reviews (type-1), non-reviews (type-3).*

I. INTRODUCTION

"What other individuals think" has dependably been an essential snippet of data for the greater part of us amid the basic leadership process. By and by, web clients post their reviews and suppositions on sites, online journals, web-based social networking like Facebook and Twitter, web-gatherings, e-commerce sites, RSS channels and so on. These conclusions are useful for both business associations and people. This reviewing frameworks supports a few individuals to enter their phony review to advance a few items or stigmatize some others [1]. Those individuals are called conclusion spammers and their activities are called feeling spamming. The primary explanation behind this activity is to make more benefit by composing dishonest reviews and beguiling evaluations .So to make items and administrations trustable, these faked conclusions must be distinguished and evacuated. Detection procedures are utilized to discover fake reviews, fake reviewers and fake reviewer group [2].

II. SPAM DETECTION APPROACHES

Depending upon the approach used for review spam detection, it can be classified as:

A. Review centric approach

Fake Review Spam are categories mainly into three different types as given below:

- Type I: Un-Truthful Review – the reviews composed are not founded on honest to goodness encounters of reviewers of utilizing the items or administrations. The review posted are ordered into two - hyper spam (contain undeserving positive opinions about some objective elements so as to advance the substances) and defaming spam (contains false negative conclusions about some different elements keeping in mind the end goal to harm their reputations) [1].
- Type II: Review on brand only - remarks just on the brands or the producers of the items. Some might be authentic, however are considered as spam as they are not focused at particular items [1].
- Type III: Non-Reviews – it isn't a review, however has ads and other unimportant reviews containing no feelings (e.g., inquiries, answers, and irregular content). A spammer may work independently, or purposely or unwittingly work as a member from a group [1].

Based on these types different techniques are used to detect different review spams.

B. Reviewer centric approach

Methods involved in this section identify various characteristics behaviors so as to detect the spammers.

This review paper trade sees about different strategies adopted for detecting fake review and opinion spammers. Whatever is left of the paper talks about the kinds of reviews spam and exactness level accomplished utilizing innovation.

III. TECHNIQUES OF REVIEW SPAM DETECTION

A. Language model and feature selection

This method identifies Type-I and Type-II fake review spam. Type I utilizes n-gram model to identify copies by checking whether

- There is same audit for same item from various client ids.
- There is same review for various items from same and distinctive client ids
- There is a same client id for various items.

Type II utilizes feature selection to identify mark spam. The copies are seen with semantic language model. This model has more productivity and viability for discovering copies. This investigation comes about organization to enhance their business execution and clients to purchase items.

B. Modified method of iterative computation framework

This method is utilized for distinguishing Type-I, Type-II and Type-III of review spam.

Type I and II are fathomed by the ICF++ method-experiences a few phases like Pos labeling, production of exchange record, FP development, extremity age and count of assertion values.

Type– III utilized IR-based assessment technique. The circumstances an analyst gives inverse conclusion to different audits of same item is checked to recognize a review spam.

IV. LITERATURE SURVEY

Nitin Jindal [1], focused on review spam and spam detection. Three main types of spam were identified. Detection of such spam is done first by detecting duplicate reviews. We then detected type 2 and type 3 spam reviews by using supervised learning with manually labeled training examples. Results showed that the logistic regression model is highly effective. However, to detect type 1 spam reviews, the story is quite different because it is very hard to manually label training examples for type 1 spam. We presented an approach to use three kinds of duplicates, which are very likely to be spam, as positive training examples to build a classification model. The results are promising.

Nitin Jindal [2], focused on s importance of reviews also gives good incentive for spam, which contains false positive or malicious negative opinions. Author makes an attempt to study review spam and spam detection. To the best of our knowledge, there is still no reported study on this problem.

Siddu P. Algur [3], focused on a novel and effective technique for detecting the trustworthiness of customer reviews for a particular product based on the features of the product being commented by the reviewers. Spam reviews are been categorized as duplicate and near duplicate reviews and non-spam reviews as partially related and unique reviews. Results demonstrate the effectiveness of the proposed technique in detecting spam and non-spam reviews. The efficiency of the task of web based customer review spam detection can be enhanced by identifying and eliminating duplicate and near duplicate spam reviews, thereby providing a summary of the trusted reviews for customers to make buying decisions.

C.L. Lai [4], focused on the development of a novel computational methodology to combat online review spam. Our experimental results confirm that the KL divergence and the probabilistic language modeling based computational model is effective for the detection of untruthful reviews. Empowered by the proposed computational methods, our empirical study found that around 2% of the consumer reviews posted to a large e-Commerce site is spam.

RAYMOND Y. K. LAU [5], focused on the proposed models outperform other well-known baseline models in detecting fake reviews. To the best of our knowledge, the work discussed in this article represents the first successful attempt to apply text mining methods and semantic language models to the detection of fake consumer reviews. A managerial implication of our

research is that firms can apply our design artifacts to monitor online consumer reviews to develop effective marketing or product design strategies based on genuine consumer feedback posted to the Internet.

[6], with the fast growing and importance of inline reviews, malicious users start to abuse the online review websites and deliberately post low quality or even fraudulent reviews, which are typically referred to as “Spam detection”. There are many existing studies on review spam detection are based on classification models. Features such as the numerous verbs used in the reviews are commonly used to construct the spam review classification model. In this paper, author focuses on various sorts of linguistic features and evaluate their performance on detecting spam reviews.

[7], Opinionated online networking, for example, item surveys are presently broadly utilized by people and associations for their decision making. Be that as it may, because of the reason of benefit or distinction, individuals attempt to amusement the framework by assessment spamming (e.g., composing counterfeit audits) to advance or downgrade some objective items. For surveys to reflect real client encounters and sentiments, such spam audits ought to be distinguished. Earlier takes a shot at sentiment spam concentrated on distinguishing counterfeit surveys and individual phony analysts. In any case, a fake review group (a gathering of commentators who work cooperatively to compose counterfeit audits) is considerably additionally harming as they can take add up to control of the conclusion on the objective item because of its size. This paper examines spam detection in the shared setting, i.e., to find fake reviewer group. The proposed strategy first uses a visit item set mining technique to locate an arrangement of competitor gatherings. It at that point utilizes a few behavioral models got from the plot wonder among fake reviewer and connection models in light of the connections among gatherings, single reviewer, and items they explored to recognize fake reviewer group. Moreover, we additionally manufactured a named dataset of fake reviewer group. Despite the fact that naming individual phony surveys and commentators is hard, shockingly naming fake reviewer groups is much less demanding. We additionally take note of that the proposed strategy leaves from the conventional directed learning approach for spam discovery in view of the inborn idea of our concern which makes the great managed learning approach less viable. Test comes about demonstrate that the proposed technique beats various solid baselines including the best in class managed characterization, relapse, and figuring out how to rank calculations.

[8], Buyers progressively rate, survey and research items on the web (Jansen, 2010; Litvin et al., 2008). Thusly, sites containing buyer surveys are getting to be focuses of supposition spam. While late work has concentrated essentially on physically identifiable occurrences of supposition spam, in this work we think about tricky feeling spam—imaginary assessments that have been purposely composed to sound genuine. Incorporating work from brain research and computational semantics, we create and contrast three methodologies with identifying misleading conclusion spam, and eventually build up a classifier that is about 90% exact on our best quality level supposition spam dataset. In view of highlight investigation of our educated models, we furthermore make a few hypothetical commitments, including uncovering a connection between misleading assessments and creative composition.

[9], Fake review detection has been studied by researchers for several years. However, so far all reported studies are based on English reviews. This paper reports a study of detecting fake reviews in Chinese. Our review dataset is from the Chinese review hosting site Dianping, which has built a fake review detection system. They are confident that their algorithm has a very high precision, but they don't know the recall. This means that all fake reviews detected by the system are almost certainly fake but the remaining reviews may not be all genuine. This paper first reports a supervised learning study of two classes, fake and unknown. However, since the unknown set may contain many fake reviews, it is more appropriate to treat it as an unlabeled set. This calls for the model of learning from positive and unlabeled examples (or PU-learning). Experimental results show that PU learning not only outperforms supervised learning significantly, but also detects a large number of potentially fake reviews hidden in the unlabeled set that Dianping fails to detect.

[10], In this paper, author propose a novel review spam detection method. Detecting review spams is a difficult task, particularly because even human beings are not always able to reliably determine which reviews are spams.

TABLE I. *Comparisons of various techniques and method used in present system*

Ref. No.	Method Used	Data Source	Approach	Strength	Limitation
[1]	Logistic Regression	Reviews downloaded from Amazon.com	Had check for SVM and Naïve Bayes additionally and discovered LR(Logistic Regression) better	Experimental outcome showed that the logistic regression model is highly effective	Additional performance measures are there.
[2]	Proposed a Review centric supervised machine learning technique	Reviews downloaded from Amazon.com	Proposed work influences an endeavor to consider the review of spam and spam detection. To the best of our insight, there is still no announced examination on this issue.	Duplicate detection and classification to detect review spam.	Improvement is required for the accuracy and detecting more sophisticated spam reviews
[3]	based on conceptual level similarity	Reviews of format pros and cons	Proposed work exhibits the viability of the proposed method in detecting spam and non-spam reviews. The productivity of the undertaking of online client review spam detection can be upgraded by distinguishing and dispensing with copy and close copy spam reviews, along these lines giving an outline of the put stock in reviews for clients to settle on purchasing choices.	Outcome shows that there are huge numbers of duplicate spam reviews detected using the conceptual level similarity measure	-
[4]	Language model	Reviews downloaded from Amazon.com	Results affirm that the KL divergence and the probabilistic dialect displaying based computational model is viable for the detection of untruthful reviews. Enabled by the proposed computational strategies, our experimental examination found that around 2% of the customer reviews presented on a huge online business webpage is spam.	This results confirm that the KL divergence and the probabilistic language modeling based computational model is effective for the detection of untruthful review	Additional performance measures are there.
[5]	Semantic Language Model	Reviews downloaded from Amazon.com	The proposed SVM computational model is more successful in detecting non-reviews than other regulated machine learning models. The proposed computational models accomplish a genuine positive rate of over 95%in fake review detection. Empowered by the design artifacts, an observational investigation of the reliability of online	The outcome confirm that semantic language modeling and a text mining-based computational model are effective for the detection of untruthful reviews, even if spammers exercise obfuscation strategies	Improvement is required for the accuracy and detecting more sophisticated spam reviews

			customer reviews is then performed.		
[6]	Linguistic	a spam review benchmark dataset	In this paper, author concentrated on various sorts of etymological features and assess the execution on detecting spam reviews	The experimental outcome showed that by incorporating many linguistic features of reviews, the detection performance of spam reviews can be greatly improved, comparing with the state-of-the-art methods.	Not experimented for opinions coming from other domains such as product reviews.
[7]	frequent item set mining	Author built a labeled dataset using expert human judges.	The proposed strategy initially utilized regular item set mining to locate an arrangement of applicant gatherings, from which a marked arrangement of spammer groups was delivered.	Experiments shows multiple strong baselines including the state-of-the-art supervised classification, regression, and learning to rank algorithms.	Additional performance measures are there.
[8]	n-gram–based text categorization	Author built a labeled dataset using expert human judges.	Author develop and contrast three methodologies for detecting deceptive sentiment spam	Develop a classifier that is nearly 90% accurate on our gold-standard opinion spam dataset.	Need more work on both negative opinions, as well as opinions coming from other domains
[9]	Kernel Density Estimation (KDE) technique	Amazon.com reviews	Author experiment comes about utilizing Amazon.com reviews from the product area demonstrated that the proposed strategy is viable, which not just exhibited its adequacy unbiasedly in view of administered learning (or order),	Experimental outcome show that the proposed method outperforms strong baselines	Need to propose a novel evaluation method to evaluate the detected spammers automatically
[10]	Author design a multi-scale anomaly detection algorithm on multidimensional time series based on curve fitting	Amazon.com reviews	Author propose a hierarchical algorithm to vigorously recognize the time windows where such attacks are probably going to have happened.	Experimental outcome show that the proposed algorithm is effective in detecting singleton review spams.	-

V. CONCLUSION

This paper fundamentally considers the issue of singleton review spam detection, which is both testing and imperative to settle. This paper indicates diverse methodologies for review spam detection. All approach has some preferred standpoint and in addition some disadvantage. Fundamental aim is to accurately classify the review as a spam or not.

REFERENCES

[1] Nitin Jindal and Bing Liu, "Analyzing and Detecting Review Spam", Seventh IEEE International Conference on Data Mining 2007.

[2] SNEHAL DIXIT & A.J.AGRAWAL, "REVIEW SPAM DETECTION", International Journal of Computational Linguistics and Natural Language Processing Vol 2 Issue 6 June 2013 ISSN 2279 –0756.

[3] Siddu P. Algur, Amit P.Patil, P.S Hiremath, S. Shivashankar, "Conceptual level Similarity Measure based Review Spam Detection", 2010 IEEE.

[4] C.L. Lai, K.Q. Xu, Raymond Y.K. Lau, Y. li, L. Jing "Toward A Language Modeling Approach for Consumer Review Spam Detection", International Conference on E-Business Engineering 2010.

[5] RAYMOND Y. K. LAU, S. Y. LIAO, RON CHIWAI KWOK, KAIQUAN XU, YUNQING XIA, YUEFENG LI, "Text Mining and Probabilistic Language Modeling for Online Review Spam Detection", ACM Trans. Manag. Inform. Syst. 2, 4, Article 25, December 2011.

[6] Amir Karami and Bin Zhou, "Online Review Spam Detection by New Linguistic Features", ieee 2014.

[7] Mukherjee, A., Liu, B., & Glance, N., "Spotting fake reviewer groups in consumer reviews", in Proceedings of the 21st international conference on world wide web (pp. 191–200), 2012.

[8] Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). 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 (pp. 309–319).

[9] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Exploiting burstiness in reviews for review spammer detection." in ICWSM. Citeseer, 2013.

[10] S. Xie, G. Wang, S. Lin, and P. S. Yu, "Review spam detection via temporal pattern discovery," in Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '12. New York, NY, USA: ACM, 2012, pp. 823–831.