



International Journal of Intellectual Advancements and Research in Engineering Computations

A survey on online spam reviews and techniques for identification

Abarna B¹, Dinesh Kumar R T²

¹PG Scholar, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tiruchengode, Namakkal, TamilNadu, India.

²Assistant Professor, Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tiruchengode, Namakkal, TamilNadu, India.

ABSTRACT

Online products reviews plays a major role in purchase decision of consumers. The product companies are now-a-days spamming the reviews. They create a bunch of account to manipulate the comments. Here a simple survey for spam reviews, reviewers and identification techniques are detected. Here gives the detail description about spammers, consumer characteristics, businesses characteristics and techniques. To know the techniques which are widely used in this survey. Techniques like PU, yelp, machine learning method are explained.

Index Terms: Spam, Spammers, PU, Machine learning, Yelp, Reviews, Graph method.

INTRODUCTION

In e-commerce sites, on-line reviews become additive and extra-important because shoppers' area unit buying options are powerfully influenced by these reviews. Thanks to cash incentives, try to use information and game systems and shoppers by posting ratings and reviews in favor of pushing sales across multiple counterfeits or even selling their competitors. These imposters, also called *Review Spammers* or *Opinion Spammers*, become more and more damager as they could be organized by crowdsourcing tasks. As there are lots of accounts, the organized spammers, called *Spammer Group*, could take total control of the reaction on their target products with little irregular actions. Although many efforts have been done for review spam and individual spammer detection, limited attention has been received at the spammer group detection. Generally, as there are usually no labeled instances (groups), most obtainable work locate spammer group candidates first, and then use unsupervised ranking methods to identify real spammer groups from these

candidates. Nevertheless, according to the research in, we could easily label some groups yourself to obtain some labeled instances (i.e., labeled spammer groups or non-spam groups). It is noticeable that combining these labeled instances and other unlabeled groups will considerably improve the accuracy of spammer group detection.

Our main contributions are summarized as follows

A proposed method PSGD, a partially supervised learning model to detect review spammer groups. Specifically, we only label some spammer groups as positive instances and learn a classifier from the positive and unlabeled instances. To the best of our knowledge, this is the first time PU-Learning is applied to spammer group detection.

Here they design a reliable negative set (*RN*) extraction algorithm which defines a feature strength function to measure the discriminative power of group features, and then iteratively removes instances containing high discriminative

Author for correspondence:

Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women, Tiruchengode, Namakkal, TamilNadu, India.

features from the unlabeled instances set to obtain *RN*. By combining the positive instances and the extracted negative instances, the PU-Learning problem can be converted into the well-known semi-supervised learning problem, thus many mature methods such as Naive Bayesian model and EM algorithm can be applied to construct the classifier.

In this paper they conducted an extensive experiments on a real-life dataset collected from Amazon.cn. We propose two new group features and verify their effect for improving the performance of detection. Given the overall performance of PSGD, we also analyze the impact of the weighting factor of unlabeled data and evaluate the effectiveness of our proposed *RN* extraction algorithm. The experimental results demonstrate that PSGD can effectively detect spammer groups and outperforms the state-of-the-art spammer group detection methods.

Positive Unlabeled Learning

To overcome the deceptive reviews a semi-supervised model, called mixing population and individual property PU (positive unlabeled) learning (MPIPUL), is proposed. Firstly, few dependable negative examples are documented from the unlabeled dataset. Secondly, few representative examples of positive and negative generated examples based on LDA (Latent Dirichlet Allocation). Thirdly, for the residual unlabeled examples (we call them spy examples), which cannot be explicitly recognized as positive and negative, two similarity weights are assigned, by which the probability of a spy example belonging to the positive class and the negative class are displayed. Finally, spy examples and their similarity weights are incorporated into SVM

(Support Vector Machine) to build an accurate classifier. An experiment on gold-standard dataset states the usefulness of MPIPUL which outperforms the present baselines.

This paper makes the following contributions

- For the first time, PU learning is defined in the atmosphere of identifying deceptive reviews.
- A novel PU learning is planned based on LDA and SVM.
- Experimental outcome reveals that our anticipated technique outperforms the present baselines.

Detection of review spam

Feedback processing technologies and methods are collected and set up through a number of analytics for consumer reviews and help traders and individuals. Four useful opinion-mining tasks for customers and vendors are the following:

1. Sentiment categorization that determines whether a review is positive, negative or neutral.
2. Featured base-opinion mining that discovers features or aspects of a reviewed article with the goal of gaining the opinion of a reviewer about that particular aspect.
3. Comparative sentence and relation result that compares one article with one or more other similar articles.
4. Opinion searches that facilitate users in search of impression on any particular article.

By capturing burst patterns as spam attacks and work reviews have fallen within the pattern is that the most effective technique to notice spam reviews. Moreover, in terms of clues to notice spam reviews, linguistic and cognitive psychology variations of real and deceitful reviews have a major influence on the detection of spam reviews.

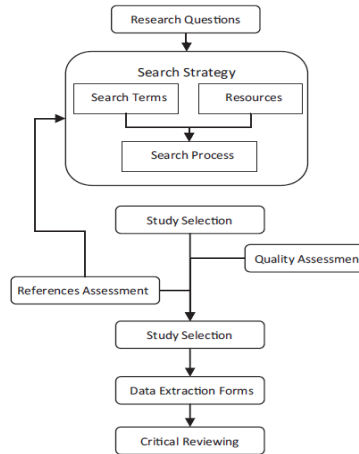


Fig1 review Process

Temporal dynamics of opinion spamming

This paper performed thoroughly analyses on the temporal dynamics of opinion spamming. It used a large-set of reviews from Yelp restaurants and its filtered reviews to characterize the approach opinion spamming operate in a very industrial setting. Vector automobile regression could be a model model wont to capture the linear interdependencies among multiple statistic. Temporal Dynamics models generalize the univariate autoregressive model by leaving over one evolving variable. Victimization time-series analyses here, it showed that there exist 3 dominant spamming policies: early, mid, and late across varied eating house. Our analyses showed that the deception rating time-series for every eating house had statistically vital correlations with the dynamics of truthful ratings time-series indicating that spam injection could probably be coordinated by the restaurants/spammers to counter the impact of unfavorable ratings over time. Causative time-series analysis of deceptive like rating time-series as response with totally different covariates time-series established the presence to 2 further trends of spam injection: buffered and reduced spamming. The covariate time-series beside varied alternative options were then wont to predict future deceptive ratings, long term/imminent future quality and rating of a eating house within the presence of deception victimization vector automobile regression. The framework additional allowed USA to indirectly

validate Yelp's filter that was shown to be affordable.

NetSpam

Based on a metapath concept also as a replacement graph-based method to label reviews counting on a rank-based labeling approach. The performance of the proposed framework is evaluated by using two real-world labeled datasets of Yelp and Amazon websites. Our observations show that calculated weights by using this metapath concept are often very effective in identifying spam reviews and results in better performance. Additionally, we found that even without a plaything, NetSpam can calculate the importance of every feature and it yields better performance within the features' addition process, and performs better than previous works, with only little number of features. Moreover, after defining four main categories for features our observations show that the reviews behavioral category performs better than other categories, in terms of AP, AUC also as within the calculated weights. The results also confirm that using different supervisions, almost like the semi-supervised method, has no noticeable effect on determining most of the weighted features, even as indifferent datasets.

Spotting Fake Reviews

This paper reports a study of detecting fake reviews in Chinese. Here 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 involves the model of learning from positive and unlabeled examples (or PU-learning). A simple PU learning framework called PU-LEA that iteratively removes positive training data from unlabeled data. However, they presume a ongoing but gradual reduction of the negative instances over iterations which unfortunately isn't always true.

Uncovering Crowdsourced Manipulation

This paper tackles the unseen challenge of crowdsourced in online reviews through a three-part effort: (A) first, we propose to target a seed

collection of deceptive reviewers who have invented a completely unique sampling method for finding products and listing them. (B) Second, we backed up a Markov Random Domain (between two reviewers) and pair energies (single reviewers) where we define the individual's energies to enhance this basic set of deceptive critics with a critic-critical graph synthesis approach. (C) Finally, we use the framework to characterize the results of this probabilistic model as a classification of crowdsourced criticism. Our classification approach using reviewer set results as a feature is substantially implemented by a classification approach to reviewer set results.

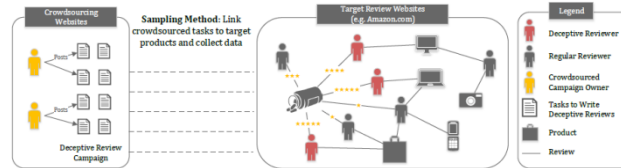


Fig2 Overall Sampling Framework

Collective Spam Detection

A new holistic approach called SpEagle that which ties together relational data with metadata. It considers the user-review-product graph to formulate the matter as a network-based classification task, during which users are labeled as a spammer or benign, reviews as fake or genuine, and products as target or non-target. Especially, it uses the metadata to style and extracts indicative features of spam which are converted into a spam score to be used as a part of class priors. It works during a completely unsupervised fashion. However, it's amenable to simply leverage label information. As such, we introduce a semi-supervised version of our method called SpEagle+, which accepts as input labels for a (small) set of user, review, and/or product nodes within the graphed not that they are not perfect.

Yelp Fake Review

There are two major approaches to filtering: supervised and unsupervised learning. In terms of features used, there are also roughly two types: linguistic features and behavioral features. In this work, we will take a supervised approach as we

can make use of Yelp's filtered reviews for training. To expose the precise psycholinguistic difference between AMT reviews and Yelp reviews (crowdsourced vs. commercial fake reviews) yelp filtering technique is used.

Exploiting Burstiness

Markov Random Field (MRF), and use the Loopy Belief Propagation (LBP) method to infer whether a reviewer may be a spammer or not within the graph. We also propose several features and use feature induced message passing within the LBP framework for network inference. The key characteristic of the approach is that the features utilized in detecting spammers are entirely different from the features utilized in classification (i.e., there's no feature overlap). KDE is closely associated with histograms, but are often endowed with properties like smoothness and continuity, which are desirable properties for review burst detection during a product.

Learning to Identify Review Spam

Here a proposal of a machine learning method to spot review spams were introduced. First analyze the effect of varied features in spam identification and also observe that the review

spammer consistently writes spam. This provides another view to spot review spam: we will identify if the author of the review is spammer. supported this observation, we offer a two view semi-supervised method, co-training, to take advantage of the massive amount of unlabeled data. The two-view co-training algorithms with the assistance of semi-supervised method are able to do better results than the single-view algorithm.

Detecting Review Spammer Groups

Each group member isn't required to review every target product. So to seek out loose scammers within the group scammers bipartite graph projection is employed. We propose a group of group spam indicators to live the spamicity of a loose spammer group, and style a completely unique algorithm to spot highly suspicious loose spammer groups during a divide and conquer manner. By exploiting effective group spam indicators to gauge the spamicity of detected groups, a divide and conquer algorithm is meant to efficiently detect and rank loose spammer groups with high precision and recall.

Data Stream Classification

Here proposed a completely unique PU learning technique LELC (PU Learning by Extracting Likely positive and negative micro-clusters) for document classification. LELC only

requires little set of positive examples and a group of unlabeled examples which is definitely obtainable within the data stream environment to create accurate classifiers. LELC can automatically extract high-quality positive and negative micro-clusters from data streams, the restrictions related to the first positive set P, like its limited size, doesn't have an excellent impact on our algorithm. Augmented by the top quality likely positive set LP and certain negative set LN that resulted, our LELC algorithm is thus ready to build a strong classifier for data stream classification.

Impact of Online Consumer Reviews on Sales

Here proposed a conceptual framework and hypothesize that product- and consumer-specific characteristics affect consumers' reliance on online consumer reviews and thus are important factors governing the efficacy of online reviews. consumers commonly seek quality information when purchasing new products. With the Internet's growing popularity, online consumer reviews became a crucial resource for consumers seeking to get product quality. Our study suggests that niche producers and producers that sell mostly through online channels should be more concerned about online consumer reviews and manipulations of online review systems because online reviews could significantly affect their sales.

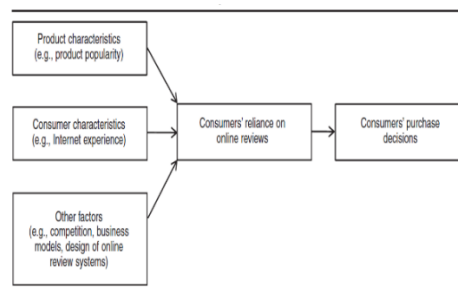


Fig3 Conceptual Framework

Graph based Spammer Detection

Heterogeneous review graph is used here to capture the relationships among reviewers, reviews and stores that the reviewers have reviewed. We explore how interactions between nodes in this graph can reveal the cause of spam and propose an iterative model to identify suspicious reviewers.

This is the first time such intricate relationships have been identified for review spam detection. We also develop an effective computation method to quantify the trustiness of reviewers, the honesty of reviews, and the reliability of stores.

CONCLUSION

In this paper, we survived various papers from 2010 to 2017 to give detailed information about spam reviews, different kinds of spam reviewers and different techniques used to identify them. Also here explains importance of online reviews for the consumers and businesses and their characteristics. the techniques like PU, yelp, Markov Random Field (MRF), the Loopy Belief

Propagation (LBP), Machine Learning Method, graph method etc., are used to identify spammers like group of spammer, loosely spammers from groups, business enemies, lone spammers or authors/ producers themselves. from the above survey PU and yelp techniques are widely used to identify spammers. These techniques also worked more efficiently than other techniques. Thus from my survey result PU and Yelp techniques are the bests approach to identify the spam reviewers.

REFERENCES

- [1]. Y. Ren, D. Ji, and H. Zhang, "Positive unlabeled learning for deceptive reviews detection," in *Proc. EMNLP*, 2014, 488_498.
- [2]. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, "Detection of review spam: A survey," *Expert Syst. Appl.*, 42(7), 2015, 634_3642.
- [3]. K. C. Santosh and A. Mukherjee, "On the temporal dynamics of opinion spamming: Case studies on yelp," in *Proc. 25th Int. Conf. WorldWideWeb*, 2016, 369_379.
- [4]. S. Shehnepoor, M. Salehi, R. Farahbakhsh, and N. Crespi, "NetSpam: A network-based spam detection framework for reviews in online social media," *IEEE Trans. Inf. Forensics Security*, 12(7), 2017, 1585_1595.
- [5]. H. Li, B. Liu, A. Mukherjee, and J. Shao, "Spotting fake reviews using positive-unlabeled learning," *Comput. Sistemas*, 18(3), 2014, 467_475.
- [6]. Fayazi, K. Lee, J. Caverlee, and A. Squicciarini, "Uncovering crowdsourced manipulation of online reviews," in *Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2015, 233_242.
- [7]. S. Rayana and L. Akoglu, "Collective opinion spam detection: Bridging review networks and metadata," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, 985_994.
- [8]. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What yelp fake review _lter might be doing?" in *Proc. 7th Int. AAAI Conf. Weblogs Soc. Media*, 2013, 409_418.
- [9]. G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Exploiting burstiness in reviews for review spammer detection," in *Proc. ICWSM*, vol. 13. 2013, 175_184.
- [10]. F. Li, M. Huang, Y. Yang, and X. Zhu, "Learning to identify review spam," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, , 22(3), 2011, 219_230.
- [11]. M. Ott, C. Cardie, and J. Hancock, "Estimating the prevalence of deception in online review communities," in *Proc. 21st Int. Conf. World Wide Web*, 2012, 201_210.
- [12]. Z. Wang, T. Hou, D. Song, Z. Li, and T. Kong, "Detecting review spammer groups via bipartite graph projection," *Comput. J.*, 59(6), 2016, 861_874.
- [13]. X.-L. Li, P. S. Yu, B. Liu, and S.-K. Ng, "Positive unlabeled learning for data stream classification," in *Proc. SIAM Int. Conf. Data Mining*, 2009, 259_270
- [14]. F. Zhu and X. Zhang, "Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics," *J. Market- ing*, 74(2), 2010, 133_148.
- [15]. G. Wang, S. Xie, B. Liu, and P. S. Yu, "Review graph based online vstore review spammer detection," in *Proc. IEEE 11th Int. Conf. Data Mining (ICDM)*, 2011, 1242_1247.