International Journal of Current Engineering and Technology
©2021 INPRESSCO®, All Rights Reserved

E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
Available at http://inpressco.com/category/ijcet

Research Article

Spam Detection Framework for Product Reviews using Machine Learning

Ms. Deepika V. Vachane and Prof. Gopal D. Upadhye

Department of Computer Engineering JSPM's Rajarshi Shahu College of Engineering Pune, India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

Generally the people trust on product on the basis of that product reviews and rating. People can take off a survey give a chance to spammers to compose spam surveys about goods and services for various benefits. Recognizing these fake reviewers and the spam content is a big debated issue of research and despite of the fact that an various number research has been done already. Up till now the procedures set hardly differentiate spam reviews, and no one show the significance of every property type. In this investigation, a structure, named NetSpam, which uses spam highlights for demonstrating review data sets as heterogeneous information networks to design spam identification method into a group of issue in this networks. Utilizing the significance of spam features help us to acquire good outcomes regarding different metrics on review data sets. The contribution work is when user search query it will display all n-no of products as well as recommendation of the product.

Keywords: *Social Media, Social Network, Spammer, Spam Review, Fake Review, Heterogeneous Information Networks.*

Introduction

Public social network portals play an vital role in the spread of information. Today a lot of people rely on the written reviews of other users in the selection of products and services. Additionally written reviews help service providers to improve the quality of their products and services. The reviews therefore play an important role in success of a business. While positive reviews can provide boost to a business, negative reviews can highly affect credibility and cause economic losses. Since anyone can leave comments as review, provides a tempting opportunity for spammers to write spam reviews which mislead users' choices. A lot of techniques have been used to identify spam reviews. Graph based algorithms are also used to identify spammers. However many aspects are still unsolved. The general concept of the NetSpam framework is to create a set of audit data retrieved as HIN (Heterogeneous Information Network) and transform the problem of spam detection into a classification issue. In particular, convert the product review data set as a HIN where the reviews are linked through different characteristics. Then a weighting algorithm is used to calculate the importance of each characteristic. These weights are used to calculate the last labels for reviews that use unsupervised and semi-supervised procedures.

NetSpam is able to find features' importance relying on metapath denition and based on values calculated for each review. NetSpam improves the accuracy and reduces time complexity. It highly depends to the number of features used to identify spam reviews. Thus using features with more weights will resulted in detecting spam reviews easier with lesser time complexity.

Literature Survey

The pair wise features are first explicitly utilized to detect group colluders in online product review spam campaigns, which can reveal collisions in spam campaigns from a more fine-grained perspective. A novel detecting framework [1] named Fraud Informer is proposed to cooperate with the pair wise features which are intuitive and unsupervised. Benefits are: Pair wise features can be more robust model for correlating colluders to manipulate perceived reputations of the targets for their best interests to rank all the reviewers in the website globally so that top-ranked ones are more likely to be colluders. it is difficult problem to automate.

In [2] paper, They proposes to build a network of reviewers appearing in different bursts and model reviewers and their co-occurrence in bursts as a Markov Random Field (MRF) and apply the Loopy Belief Propagation (LBP) method to induce whether a reviewer is a spammer or not in the graph. A novel assessment method to evaluate the detected spammers automatically using supervised classification of their reviews. Positive points are: High accuracy, the proposed method is effective. To detect review spammers in review bursts. To detect spammers automatically. Drawback is: a generic framework is not used for detect spammers.

In [3] paper, The challenges are: The detection of fraudulent behaviors, determining the trustworthiness of review sites, since some may have strategies that enable misbehavior, and creating effective review aggregation solutions. The TrueView score, in three different variants, as a proof of concept that the synthesis of multi-site views can provide important and usable information to the end user. Positive points are : develop novel features capable of finding cross-site discrepancies effectively, a hotel identity-matching method with 93% accuracy. Enable the site owner to detect misbehaving hotels. Enable the end user to trusted reviews. The drawback is difficult problem to automate.

In [4] paper, To describes unsupervised anomaly detection techniques over user behavior to distinguish probably bad behavior from normal behavior. To discover diverse attacker schemes fake, compromised, and colluding Facebook identities with no a priori labeling while maintaining low false positive rates. Anomaly detection technique to forcefully identify anomalous likes on Facebook ads. Achieves a detection rate of over 66% (covering more than 94% of misbehavior) with less than 0.3% false positives. The attacker is trying to drain the budget of some advertiser by clicking on ads of that advertiser.

In [5] paper, A grouped classification algorithm called Multi-typed Heterogeneous Collective Classification (MHCC) and then extends it to Collective Positive and Unlabeled learning (CPU).The proposed models can markedly increase the F1 scores of strong baselines in both PU and non-PU learning environment. Positive points are: Proposed models can markedly increase the F1 scores of strong baselines in both PU and non-PU learning settings. Models only use language self-contained features; they can be smoothly generalized to other languages. It detects a huge number of potential fake reviews hidden in the unlabeled set. Fake reviews hiding in the unlabeled reviews that Dianping's algorithm did not capture. The ad-hoc labels of users and IPs used in MHCC may not be very specific as they are computed from labels of neighboring reviews.

In [6] paper, To elaborates two distinct methods of reducing feature subset size in the review spam domain. The methods involves filter-based feature rankers and word frequency based feature selection. Lead points are: The first method is to simply select the words which appear most often in the text. Second method can use filter based feature rankers to rank the features and then select the top ranked features. Drawback is : There is no fix size to fits all the approach.

In [7] paper, To provide an efficient and effective method to identify review spammers by incorporating social relations based on two assumptions that people are more likely to consider reviews from those connected with them as trustworthy, and review spammers are less likely to maintain a large relationship network with normal users. Powers are: The proposed trust-based prediction achieves a higher accuracy than standard CF method. To overcome the sparsity problem and compute the overall trustworthiness score for every user in the system, which is used as the spamicity indicator.The drawback is to Review dataset is required.

In [8] paper, They proposes the detect fake reviews for a product by using the text and rating property from a review. In short, the proposed system (ICF++) will measure the honesty value of a review, the trustiness value of the reviewers and the reliability value of a product. Powers are: Accuracy is better than ICF method. Precision is maximizing. Drawbacks are : the Process need to be optimized.

In [9] paper , To provides an overview of existing challenges in a range of problem domains associated with online social networks that can be addressed using anomaly detection. It provides an overview of existing techniques for anomaly detection, and the manner in which these have been applied to social network analysis. Positive points are: The Detection of anomalies used to identify illegal activities. Drawbacks are: Need to improve the use of anomaly detection techniques in SNA.

In [10] paper, They proposes a new holistic approach called SpEagle that utilizes clues from all metadata (text, timestamp, and rating) as well as relational data (network), and harness them collectively under a unified system to spot suspicious users and reviews, as well as products targeted by spam. SpEagle employs a review-network-based classification task which accepts prior knowledge on the class distribution of the nodes, estimated from metadata. Positive points are: It enables seamless integration of labeled data when available. It is extremely efficient.

In [11] this paper a novel framework, named NetSpam, which utilize spam feature for modeling review datasets as heterogeneous information networks to plan spam detection procedure into arrangement problem in such networks. Using the importance of spam features obtain well again results in terms of Special metrics experiment real-world review datasets from Yelp and Amazon websites.

In [12] this paper, the Convolution Neural Network(CNN) and Particle Swarm Optimization (PSO), those two approaches use for recognition of the isolated handwritten digit.Customized PSO is used to reduce the overall computation time of the proposed system.

In [13] this paper mangoes are graded in four types like Green Mango, Yellow Mango and Red Mango which are based on machine learning method. This system considers RGB values size and shape of mangoes. Following analysis is used to obtain good probability. This helps to train system to identify appropriate maturity of mangoes. This research is conducted on two machine learning method i.e. Naive Byes and SVM(Support Vector Machine).

Proposed Methodology

A new proposed framework consists in representing a set of reviews data provided as HIN (Heterogeneous Information Network) and solving the issue of spam detection in a problem of HIN classification. In particular, to show the reviews data set as a HIN where the reviews are linked through different types of nodes (such as functionality and users). Then a weighting algorithm is used to calculate the importance (or weight) of each function. These weights are used to calculate the latest review labels using supervised and unsupervised procedures. Based on our observations, defining two views for features (review-user and behavioral-linguistic), the classified features as review behavioral have more weights and yield better performance on spotting spam reviews in both semi-supervised and unsupervised approaches. The feature weights can be added or removed for labeling and hence time complexity can be scaled for a specific level of accuracy. Categorizing features in four major categories (review-behavioral, user-behavioral, review-linguistic, user-linguistic), helps us to understand how much each category of features is contributed to spam detection.

A. Architecture

The Fig.1 shows the proposed system architecture.

1) NetSpam framework that is a novel network based approach which models review networks as heterogeneous information networks.

2) A new weighting method for spam features is proposed to determine the relative importance of each feature and shows how effective each of features are in identifying spams from normal reviews.

3) NetSpam framework improves the accuracy against the state-of-the art in points of time complexity, which extremely depends to the number of features utilized to detect a spam review.

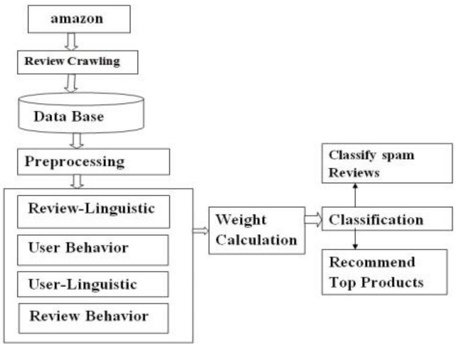

Fig. 1. Proposed System Architecture

The general concept of our proposed framework is to model a given review dataset as a Heterogeneous Information Network and to map the problem of spam detection into a HIN classification problem. In particular, model review dataset as in which reviews are connected through different node types. The fig. 2 shows the flowchart of NetSpam framework.

Advantages of Proposed System:

1) It identifies spam and spammers as well as different type of analysis on this topic.

2) Written reviews also help service providers to enhance the quality of their products and services.

3) It identifies the spam user using positive and negative reviews in online social media.

4) This framework displays only trusted reviews to the users.

B. Algorithms

1. Sentiment Analysis Algorithm:

- Input: Text File(comment or review) T, The sentiment lexicon L.
- Output:S_mt = {P,Ng $andN$} and strength S where P:
- Positive, Ng: Negative, N: Neutral
- Initialization: SumPos = SumNeg =0, where,
- SumPos: accumulates the polarity of positive tokens tismt in T,
- SumNeg: accumulates the polarity of negative tokens ti-smt in T,

```
Begin
    For each $t_i \in T$ do
        Search for $t_i$ in L
        If $t_i \in Pos - list$ then
            $SumPos \leftarrow SumPos + ti - smt$
        Else $ift_i \in Pos - list$ then
            $SumNeg \leftarrow SumNeg + ti - smt$
        End If
    End For
    If $SumPos > |SumNeg|$ then
        Smt = P
        S=SumPos/(SumPos+SumNeg)
    Else If $SumPos < |SumNeg|$ then
        Smt = Ng
        S=SumNeg/(SumPos+SumNeg)
    Else
        Smt = N
        S=SumPos/(SumPos+SumNeg)
    End IfEnd
```

2. Latent Semantic Analysis Algorithm

1) Step 1: Documents should be prepared in the following way:

   - Exclude trivial words as well as low-frequency terms.
   - Conflate terms with techniques like stemming or lemmatization.

2) Step 2: A term-frequency matrix (A) must be created that includes the occurrences of each term in each document.

3) Step 3: Singular Value Decomposition (SVD):

- Extract least-square principal components for two sets of variables: set of terms and set of documents. • SVD products include the term eigenvectors U, the document eigenvectors V , and the diagonal matrix of singular values ^{P}.

4) Step 4: From these, factor loadings can be produced for terms U^{P} and documents V^{P}

3. Netspam Algorithm

```
• Input:review–dataset,spam-feature-list,pre–labeled– reviews
• Output: features importance (W), spamicity probability (Pr)
• Process:
• Step 1: u, v: review, $y_u$: spamicity probability of review u
• Step 2: $f(x_{lu})$: initial probability of review u being spam • Step 3: $P_l$ : metapath based on feature l, L: features number
• Step 4: n: number of reviews connected to a review
• Step 5: $m_u^{P_l}$ : the level of spam certainty
• Step 6: $m_{u,v}^{P_l}$ : the metapath value
• Step 7: Prior Knowledge
• Step 8: if semi-supervised mode
• Step 9:     if $u \in pre - labeled - reviews$
• Step 10:    $y_u = label(u)$
• Step 11:    else
• Step 12:    $y_u = 0$
• Step 13: else unsupervised mode
• Step 14:    $y_u = \frac{1}{L} \sum_{l=1}^{L} f(x_{l_u})$
• Step 15: Network Schema Definition
• Step 16: schema = defining schema based on spamfeature-list
• Step 17: Metapath Definition and Creation
• Step 18: for $p_l \in schema$
• Step 19:    for $u, v \in review - dataset$
• Step 20:        $m_u^{p_l} = \frac{|s \times f(x_{l_u})|}{s}$
• Step 21:        $m_v^{p_l} = \frac{|s \times f(x_{l_v})|}{s}$
• Step 22:    if $m^{p}{}_u{}^{l} = m^{p}{}_v{}^{l}$
• Step 23:    $mppu,vl = mpul$
• Step 24:    else
• Step 25:    $mp^{p}{}_{u,v}{}^{l} = 0$
• Step 26: Classification - Weight Calculation
• Step 27: for $p_l \in schemes$
• Step 28:    $W_{p_l} = \frac{\sum_{r=1}^{n} \sum_{s=1}^{n} mp_{r,s}^{p_l} \times y_r \times y_s}{\sum_{r=1}^{n} \sum_{s=1}^{n} mp_{r,s}^{p_l}}$
• Step 29: Classification - Labeling
• Step 30: for $u,v \in review - dataset$
• Step 31:    $Pr_{u,v} = 1 - \prod_{p_l=1} 1 - mp_{u,v}^{p_l} \times W_{p_l L}$
• Step 32:    $Pr_u = avg(Pr_{u,1}, Pr_{u,2}, \ldots, Pr_{u,n})$
• Step 33: return (W,Pr)
```

Result And Discussions

Table II Classification results of NetSpam Framework for product reviews

Reviews	Count
Spam	257
Non-Spam	301

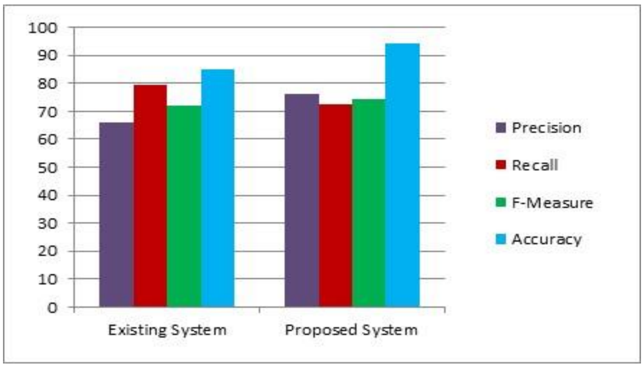

Fig. 2. Performance Analysis between existing and proposed system

The proposed NetSpam framework time complexity is $O(e^2n)$. The netspam framework accuracy is 94.06% which is better than SPaglePlus Algorithm accuracy is 85.14% on using product dataset.

Conclusion

In this proposed system investigation presents a novel spam detection system in particular NetSpam framework for product reviews based on Sentiment analysis (SA) and latent semantic analysis (LSA). This paper has used SA and LSA with netspam algorithm for spam detection. Our perceptions appear that computed weights by utilizing this metapath idea can be exceptionally successful in distinguishing spam reviews and prompt a superior performance. Additionally, NetSpam can compute the significance of each element furthermore, it yields better execution in the highlights' expansion process, and performs superior to anything past works, with just a modest number of highlights. In addition, in the wake of characterizing four primary classifications for highlights our perceptions demonstrate that the reviews behavioral classification. LSA is used in the proposed system to reduce similar comments and try to improve spam detection accuracy. The outcomes about additionally affirm that utilizing distinctive supervisions, comparative to the semi-supervised technique, have no observable impact on deciding the vast majority of the weighted highlights, similarly as in various datasets.

References

[1]. Ch. Xu , J. Zhang," Combating product review spam campaigns via multiple heterogeneous pairwise features", In SIAM InternationalConference on Data Mining, 2014.

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

[3]. A. j. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos, ”Trueview: Harnessing the power of multiple review sites”, In ACM WWW, 2015.

[4]. B. Viswanath, M. Ahmad Bashir, M. Crovella, S. Guah, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, ”Towards detecting anomalous user behavior in online social networks”, In USENIX, 2014.

[5]. H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao , ”fake reviews via collective PU learning”, In ICDM, 2014.

[6]. M. Crawford, T. M. Khoshgoftaar, and J. D. Prusa,” Reducing Feature set Explosion to Faciliate Real-World Review Sapm Detection”, In Proceeding of 29th International Florida Artificial Intelligence Research Society Conference, 2016.

[7]. H. Xue, F. Li, H. Seo, and R. Pluretti,” Trust-Aware Review Spam Detection”, IEEE Trustcom/ISPA.,2015

[8]. E. D. Wahyuni , A. Djunaidy,” Fake Review Detection From a Product Review Using Modified Method of Iterative Computation Framework”, In Proceeding MATEC Web of Conferences, 2016.

[9]. R. Hassanzadeh ,”Anomaly Detection in Online Social Networks: Using Datamining Techniques and Fuzzy Logic”, Queensland University of Technology, Nov, 2014.

[10]. R. Shebuti , L. Akoglu,” Collective opinion spam detection: bridging review networks and metadata”, In ACM KDD, 2015.

[11]. Saeedreza Shehnepoor, Mostafa Salehi*, Reza Farahbakhsh, Noel Crespi, ”Netspam : a network-based spam detection framework for reviews in online social media”, IEEE conference paper, 2017.

[12]. G.D.Upadhye,P.Barhate,“Classifying Handwritten Digit Recognition Using CNN and PSO”,IJRTE,ISSN: 2277-3878, Volume-8 Issue-2, July 2019.

[13]. G.D.Upadhye,D.Pise,“Grading of Harvested Mangoes Quality and Maturity Based on Machine Learning Techniques”,IEEE International conference on smart city and Emerging Technology,2018.