International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

An Efficient spam detection system for online reviews using advance data mining algorithms

Mrs. Sonali Ratnakar Nawale and Prof. S. M. Aher

Computer Engineering Vishwabharati Academy's College of Engineering Ahmednagar, India

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

Abstract

Major Society of people using internet trust the contents of net. The liability that anyone can take off a survey give a brilliant chance to spammers to compose spam surveys about hotels and services for various interests. Recognizing these spammers and the spam content is a widely debated issue of research and in spite of the fact that an impressive number of studies have been done as of late towards this end, yet so far the procedures set forth still scarcely distinguish spam reviews, and none of them demonstrate the significance of each extracted feature type. In this application, use a novel structure, named NetSpam, which proposes spam features for demonstrating hotel review datasets as heterogeneous information networks to design spam review detection method into a classification issue in such networks. Utilizing the significance of spam features helps us to acquire better outcomes regarding different metrics on review datasets. The outcomes represent that NetSpam results with the previous methods and encompassed by four categories of features; involving review-behavioral, user-behavioral, review linguistic, user-linguistic, the first type of features performs better than the other categories. The contribution work is when user will search query it will display all top hotels as well as there is recommendation of the hotel by using user's point of interest.

Keywords: *Social Media, Spammer, Spam Review, Heterogeneous Information Networks, Sentiment Analysis, Content Similarity, Netspam*

Introduction

Social network portals play an important role in the propagation 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[2][6]. 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 based on linguistic patterns, behavioral patterns. 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[1]. In particular, convert the hotel/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[7][8].

NetSpam is able to find features' importance relying on metapath definition 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[1]. Thus using features with more weights will resulted in detecting spam reviews easier with lesser time complexity[10]. Motivation:

- Identifying the spam user using positive and negative reviews in online social media.
- Display only trusted reviews at the users' side.
- When user search query, it will show top-k hotel and recommends one of the hotel using user's point of interest.

Literature Survey

The pair wise features are first explicitly utilized to detect group colluders in online product review spam campaigns, which can reveal collusions 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. Advantages 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. Disadvantage is difficult problem to automate. The paper [2] 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. Advantages are: High accuracy, the proposed method is effective. To detect review spammers in review bursts. To detect spammers automatically. Disadvantage 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. Advantages 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. Disadvantage is difficult problem to automate. In [4] paper describes unsupervised anomaly detection techniques over user behavior to distinguish probably bad behavior from normal behavior. To find 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 Multityped 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. Advantages are: Proposed models can markedly increase the F1 scores of strong baselines in both PU and nonPU 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. The paper [6] 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. Advantages 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. Disadvantages are: There is not a one size fits all approach that is always better.

In [7] paper, providing 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. Advantages 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. Disadvantages are: Review dataset required. The paper [8] proposes to 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. Advantages are: Accuracy is better than ICF method. Precision is maximizing. Disadvantages are: Process need to be optimized.

The paper [9] 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. Advantages are: Detection of anomalies used to identify illegal activities. Disadvantages are: Need to improve the use of anomaly detection techniques in SNA. The paper [10] 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. Advantages are: It enables seamless integration of labeled data when available. It is extremely efficient.

Online Social Media websites play a main role in information propagation which is considered as an important source for producers in their advertising operations as well as for customers in selecting products and services. People mostly believe on the

written reviews in their decisionmaking processes, and positive/negative reviews encouraging/discouraging them in their selection of products and services. These reviews eventually be an important factor in success of a business while positive reviews can bring benefits for a company, negative reviews can potentially impact credibility and cause economic losses. The fact that anyone with any identity can leave comments as reviews provides a tempting opportunity for spammers to write fake reviews designed to mislead users' opinion. These misleading reviews are then multiplied by the sharing function of social media and propagation over the web. The reviews written to change users' perception of how good a product or a service are considered as spam, and are often written in exchange for money.

Disadvantages:

1) There is no information filtering concept in social network.
2) People believe on the written reviews in their decisionmaking processes, and positive/negative reviews encouraging/discouraging them in their selection of products and services.
3) Anyone can access application through registration and gives feedbacks as reviews for spammers to misguide other user's opinion.
4) Less accuracy.
5) More time complexity.

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.

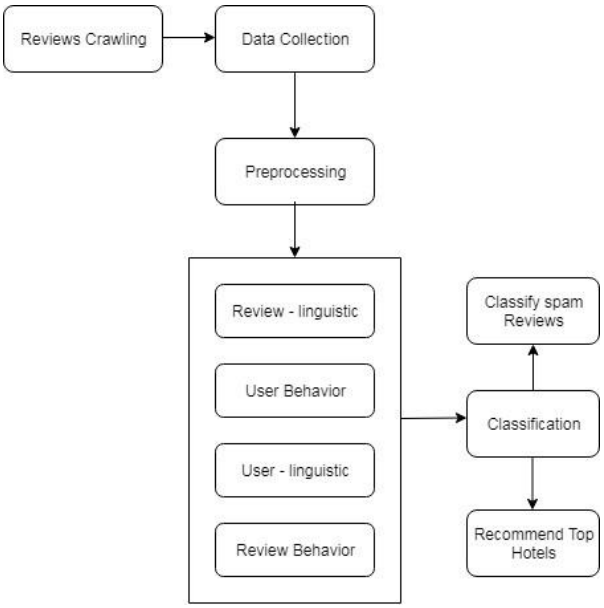

Fig. 1. Proposed System Architecture

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.

A weighting algorithm is then employed to calculate each feature's importance. These weights are used to calculate the very last labels for reviews using both unsupervised and supervised procedures. Based on the observations defining two views for features.

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 ti-smt in
T,
SumNeg: accumulates the polarity of negative tokens ti-smt in
T,
Begin

1. For each $t_i \in T$ do
2. Search for t_i in L
3. If $t_i \in Pos - list$ then
4. $SumPos \leftarrow SumPos + ti - smt$
5. Else $if t_i \in Pos - list$ then
6. $SumNeg \leftarrow SumNeg + ti - smt$
7. End If
8. End For
9. If *SumPos* > |*SumNeg*| then
10. Smt = P
11. S=SumPos/(SumPos+SumNeg)
12. Else If *SumPos* < |*SumNeg*| then
13. Smt = Ng
14. S=SumNeg/(SumPos+SumNeg)
15. Else
16. Smt = N
17. S=SumPos/(SumPos+SumNeg)
18. End If

End
2. Latent Semantic Analysis Algorithm 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.

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

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 spam-feature-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: $W_{p_l} = \frac{\sum_{r=1}^{n}\sum_{s=1}^{n} mp_{r,s} \times y_r \times y_s}{\sum_{r=1}^{n}\sum_{s=1}^{n} mp_{r,s}^{p_l}}$ ^{p_l}
Step 29: Classification - Labeling
Step 30: for $u,v \in review - dataset$
Step 31: $Pr_{u,v} = 1 - \prod_{p_l=1} 1 - mp_{u,v}^{p_l} \times W_{p_lL}$
Step 32: $Pr_u = avg(Pr_{u,1}, Pr_{u,2}, ..., Pr_{u,n})$
Step 33: return (W,Pr)

C. Spam Features

- User-Behavioral (UB) based features: Burstiness: Spammers, usually write their spam reviews in short period of time for two reasons: first, because they want to impact readers and other users, and second because they are temporal users, they have to write as much as reviews they can in short time.

$$x_{BST}(i) = \begin{cases} 0 & (L_i - F_i) \notin (0, \tau) \\ 1 - \frac{L_t - F_t}{\tau} & (L_i - F_i) \in (0, \tau) \end{cases} \quad (1)$$

Where,
$L_i - F_i$, describes days between last and first review for τ = 28.
Users with calculated value greater than 0.5 take value 1 and others take 0.

- User-Linguistic (UL) based features: Average Content Similarity, Maximum Content Similarity: Spammers, often write their reviews with same template and they prefer not to waste their time to write an original review. In result, they have similar reviews. Users have close calculated values take same values (in [0; 1]).
- Review-Behavioral (RB) based features: Early Time Frame: Spammers try to write their reviews a.s.a.p., in order to keep their review in the top reviews which other users visit them sooner.

$$x_{ETF}(i) = \begin{cases} 0 & (L_i - F_i) \notin (0, \delta) \\ 1 - \frac{L_t - F_t}{\delta} & (L_i - F_i) \in (0, \delta) \end{cases} \quad (2)$$

Where,
$L_i - F_i$ denotes days specified written review and first written review for a specific business. We have also δ = 7. Users with calculated value greater than 0.5 takes value 1 and others take 0.
Rate Deviation using threshold: Spammers, also tend to promote businesses they have contract with, so they rate these businesses with high scores. In result, there is high diversity in their given scores to different businesses which is the reason they have high variance and deviation.

$$x_{DEV}(i) = \{ \quad 1 - \frac{r_{ij} - avg_{e \in E * j} r(e)}{4} > \beta_1 0$$

Otherwise
(3) Where,
β_1 is some threshold determined by recursive minimal entropy partitioning. Reviews are close to each other based on their calculated value, take same values (in [0;
1)).
• Review-Linguistic (RL) based features: Number of first Person Pronouns, Ratio of Exclamation Sentences containing '!': First, studies show that spammers use second personal pronouns again and again as compared to first personal pronouns. In addition, spammers put "!" in your prayers as much as possible to increase the impression on users and highlight their reviews among others. The reviews are close to each other according to their calculated value, taking the same values (in [0; 1]).

Result and Discussions

Experimental evaluation outcomes shows the Tripadvisor API uses for hotel review dataset with higher percentage of spam reviews have better performance because when fraction of spam reviews increases, probability for a review to be a spam review increases and as a result more spam reviews will be labeled as spam reviews. The results of the dataset show all the four behavioral features are ranked as first features in the final overall weights. The Fig.3 graph shows the NetSpam framework features for the dataset have more weights and features for Review-based dataset stand in the second position. Third position belongs to User-based dataset and finally Itembased dataset has the minimum weights (for at least the four features with most weights).

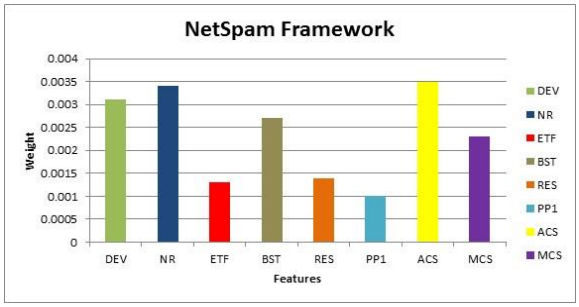

Fig. 2. Feature weights for NetSpam Framework

Table I Weights of all features

Features	Weight
DEV	0.0031
NR	0.0034
ETF	0.0013
BST	0.0027
RES	0.0014
PP1	0.001
ACS	0.0035
MCS	0.0023

Table II Classification results of NetSpam Framework for hotel reviews

Reviews	Count
Spam	257
Non-Spam	301

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 TripAdvisor hotel dataset.

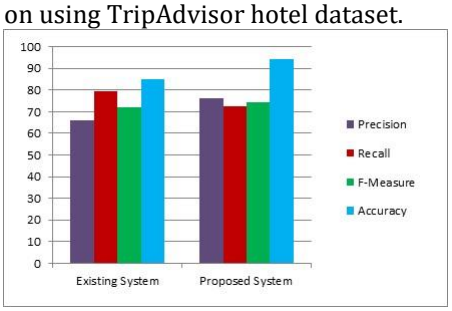

Fig. 3. Performance Analysis between existing and proposed system

Conclusion

This paper presents a novel spam detection system in particular NetSpam in view of a metapath idea and another graph based strategy to name reviews depending on a rankbased naming methodology. The execution of the proposed structure is assessed by utilizing review datasets. The perceptions demonstrate that ascertained weights by utilizing this metapath idea can be exceptionally powerful in recognizing spam surveys and prompts a superior execution. Furthermore, found that even without a prepare set, NetSpam can figure the significance of each element and it yields better execution in the highlights' expansion procedure, and performs superior to anything past works, with just few highlights. In addition, in the wake of characterizing four fundamental classifications for highlights our perceptions demonstrate that the review behavioral classification performs superior to anything different classifications, regarding AP, AUC and in the ascertained weights. The outcomes likewise affirm that

utilizing diverse supervisions, like the semi-administered strategy, have no detectable impact on deciding the vast majority of the weighted highlights, similarly as in various datasets. Contribution part in this project, for user when searches query as location he will get the top-k hotel lists as well as one recommendation of hotel by using personalized recommendation algorithm with the help of user's point of interests.

References

[1] Ch. Xu and J. Zhang. Combating product review spam campaigns via multiple heterogeneous pairwise features. In SIAM International Conference 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. Spotting 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 and 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 and L. Akoglu. Collective opinion spam detection: bridging review networks and metadata. In ACM KDD, 2015.