International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 4 | Issue 2 | ISSN : 2456-3307

An Efficient Network-Based Spam Detection Structure for Reviews In Online Social Media

G. Prashanti[*1], **Tiruveedhula Priyanka**[2]
[1*]prashantiguttikonda77@gmail.com,priyaind3@gmail.com[2]

ABSTRACT

Todays, a major part of everyone trusts on content in social media like opinions and feedbacks of a topic or a product. The liability that anyone can take off a survey give a brilliant chance to spammers to compose spam surveys about products and services for various interests. Recognizing these spammers and the spam content is a wildly debated issue of research and in spite of the fact that an impressive number of studies have been done as of late toward 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 investigation, we propose a novel structure, named NetSpam, which uses spam highlights for demonstrating review datasets as heterogeneous information networks to design spam detection method into a classification issue in such networks. Utilizing the significance of spam features help us to acquire better outcomes regarding different metrics on review datasets. The outcomes demonstrate that NetSpam results the existing methods and among four categories of features; including 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 search query it will display all top-k products as well as recommendation of the product.

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

I. INTRODUCTION

Online Social Media portals play an influential role in information propagation which is considered as an important source for producers in their advertising campaigns as well as for customers in selecting products and services. In the past years, people rely a lot on the written reviews in their decision-making processes, and positive/negative reviews encouraging/ discouraging them in their selection of products and services. In addition, written reviews also help service providers to enhance the quality of their products and services. These reviews thus have become 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 review, 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.

Online Social Media portals play an influential role in information propagation which is considered as an impor-tant source for producers in their advertising campaigns as well as for customers in selecting products and services. In the past years, people rely a lot on the written reviews in their decision-making processes, and positive/negative reviews encouraging/discouraging them in their selection of products and services. In addition, written reviews also help service providers to enhance the quality of their products and services. These reviews thus have become 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 review, provides a tempting opportunity for spammers to write fake reviews designed to mislead users' opinion. These misleading reviews are

CSEIT1833622 | Received : 01 April 2018 | Accepted : 10 April 2018 | March-April-2018 [(4) 2 : 101-107]

then multiplied by the sharing function of social media and prop-agation over the web. The reviews written to change users' perception of how good a product or a service are considered as spam [11], and are often written in exchange for money.

S.R. Shehnepoor is with the University of Tehran, Tehran, Iran. M. Salehi (*corresponding author) is with the University of Tehran, Tehran, Iran. R. Farahbakhsh is with the Institut Mines-Telecom, Telecom SudParis, Paris, France. N. Crespi is with the Institut Mines-Telecom, Telecom Sud-Paris, Paris, France. emails: fshehnepoor@ut.ac.ir, mostafa salehi@ut.ac.ir, reza.farahbakhsh@it-sudparis.eu, noel.crespi@institut-telecom.fr.g As shown in [1], 20% of the reviews in the Yelp website are actually spam reviews.

On the other hand, a considerable amount of literature has been published on the techniques used to identify spam and spammers as well as different type of analysis on this topic [30], [31]. These techniques can be classified into different categories; some using linguistic patterns in text [2], [3], [4], which are mostly based on bigram, and unigram, others are based on behavioral patterns that rely on features extracted from patterns in users' behavior which are mostly metadata-based [34], [6], [7], [8], [9], and even some techniques using graphs and graph-based algorithms and classifiers [10], [11], [12].

Despite this great deal of efforts, many aspects have been missed or remained unsolved. One of them is a classifier that can calculate feature weights that show each feature's level of importance in determining spam reviews. The general concept of our proposed framework is to model a given review dataset as a Heterogeneous Information Network (HIN) [19] and to map the problem of spam detection into a HIN classification problem. In particular, we model review dataset as a HIN in which reviews are connected through different node types (such as features and users). A weighting algorithm is then employed to calculate each feature's importance (or weight). These weights are utilized to calculate the final labels for reviews using both unsupervised and supervised approaches.

To evaluate the proposed solution, we used two sample review datasets from Yelp and Amazon websites. 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 unsu-pervised approaches. In addition, we demonstrate that using different supervisions such as 1%, 2.5% and 5% or using an unsupervised approach, make no noticeable variation on the performance of our approach. We observed that feature weights can be added or removed for labeling and hence time complexity can be scaled for a specific level of accuracy. As the result of this weighting step, we can use fewer features with more weights to obtain better accuracy with less time complexity. In addition, 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.

In summary, our main contributions are as follows:

(i) We propose NetSpam framework that is a novel network-based approach which models review networks as hetero-geneous information networks. The classification step uses different metapath types which are innovative in the spam detection domain.

- A new weighting method for spam features is pro-posed to determine the relative importance of each feature and shows how effective each of features are in identifying spams from normal reviews. Previous works [12], [20] also aimed to address the importance of features mainly in term of obtained accuracy, but not as a build-in function in their framework (i.e., their approach is dependent to ground truth for determining each feature importance). As we explain in our unsupervised approach, NetSpam is able to find features importance even without ground truth, and only by relying on metapath definition and based on values calculated for each review.

- NetSpam improves the accuracy compared to the state-of-the art in terms of time complexity, which highly depends to the number of features used to identify a spam review; hence, using features with more weights will resulted in detecting fake reviews easier with less time complexity.

II. LITERATURE SURVEY

The whole literature review is focused on the following literary works being done by an array of scholars and researchers from the field of Review Spam Detection. The following papers are selected for review keeping in mind the traditional and conventional approaches of Spam detection along with the emerging techniques.

LITERATURES REVIEWED

Spam Filtering by Semantics-based Text Classification 2016

This paper, we described a novel and efficient Chinese spam filtering approach based on semantic information delivered in the body text of emails. The fundamental step is the extracting of semantic information from texts, which will be treated as feature terms for classification later. The extraction of semantic information of text was achieved by attaching semantic annotations on the words and sentences of it. We get these feature terms through attaching annotations on text layer-by-layer, then these terms are used to build up the decision tree and selected by pruning. The method of adding annotations on text is usually applied to the pre-processing of text in natural language processing. The application of text classification in semantic extraction and feature selection is limited because of the low training speed.

Trust-Aware Review Spam Detection 2015

The focus is on the problem of detecting review spammers using contextual social relationships that are available in several online review systems. We first present a trust-based rating predication algorithm using local proximity derived from social relationships, such as friendships and Complements relationships, using the random walk with restart. Results show a strong correlation between social relationships and the computed trustworthiness scores.
Model works under the assumption that review spammers tend to be socially inactive. Many of them would be isolated or barely connected with other users in the system. Our prediction model only aggregates the ratings from trusted users, which potentially filters out the influence of spammers. Experiments on the collected Yelp dataset show that the proposed trust based prediction achieves a higher accuracy than standard CF method.

Spam Mails Filtering Using Different Classifiers with Feature Selection and Reduction Techniques 2015

Work proposes a methodology to detect an email as spam or legitimate mail on the basis of text categorization. Various techniques for pretreatment of email format are applied such as applying stop words removing, stemming, feature reduction and feature selection techniques to fetch the keywords from all the attributes and finally using different classifiers to segregate mail as spam or ham. The papers have used PCA (Principal Component Analysis) and CFS (Correlation Feature Selection) technique for feature reduction. Methodology is totally based on data mining approach for classifying ham and spam emails from large text and text embedded image datasets. Time taken to build model is less by using CFS comparatively PCA applied on different classifiers which are Naive Bayesian, SVM, Random Forest, Bayes Net. Using CFS saves a lot of time for classifiers to build than using PCA. PCA and CFS reduce the attributes without loss their value. Logistic Model Tree (LMT) classifier produce accurate results comparative to others but takes a lot of computational time. Future researches need to consideration on co-evolutionary problem of the spam filtering at server level, because while the spam filter tries to develop its prediction capacity, the spammer try to develop their spam messages in order to overreach the classifiers.

Opinion Spam Detection Using Feature Selection 2014

Rinki Patel and Priyank Thakkar modeled the problem as the classification problem and Naïve Bayes (NB) classifier and Least Squares Support Vector Machine (LS-SVM) are used on three different representations (Boolean, bag-of-words and term frequency–inverse document frequency (TF-IDF)) of the opinions.In this paper experiments are carried out on widely used gold-standard dataset. The paper focuses on modelling deceptive spam detection task as classification problem with deceptive and truthful as two classes. Experiments are carried out with unigram, bigram and bigram plus sequence of words approaches. Learning a classifier using appropriate number of features improves the

accuracy. Detection of spam review using adjective, noun and verbs is not possible.

SMS Classification Based on Naïve Bayes Classifier and Apriori Algorithm Frequent Itemset 2014

The paper proposes a hybrid system of SMS classification to detect spam or ham, using Naïve Bayes classifier and Apriori algorithm. Work done here not only considered each and every word as independent and mutually exclusive but also frequent words as a single, independent and mutually exclusive. Training the system for the first time requires little bit more time than Naïve Bayes Classifier. The main contribution of this paper is better accuracy.

III. EXISTING SYSTEM

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 decision-making processes, and positive/negative reviews encouraging/discouraging them in their selection of products and services. These reviews thus have become 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:

- There is no information filtering concept in online social network.
- People believe on the written reviews in their decision-making processes, and positive/negative reviews encouraging/discouraging them in their selection of products and services.
- Anyone create registration and gives comments as reviews for spammers to write fake reviews designed to misguide users' opinion.
- Less accuracy.
- More time complexity.

IV. PROPOSED SYSTEM APPROACH

The proposed framework is to model a given review dataset as a Heterogeneous Information Network (HIN) and to map the problem of spam detection into a HIN classification problem. In particular, we model review dataset as a HIN in which reviews are connected through different node types (such as features and users). A weighting algorithm is then employed to calculate each feature's importance (or weight). These weights are utilized to calculate the final labels for reviews using both unsupervised and supervised approaches. 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.

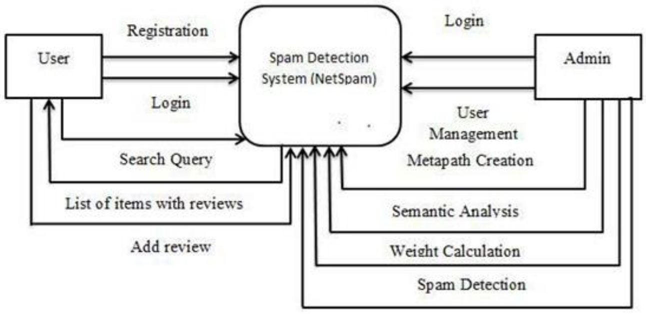
1. NetSpam framework that is a novel network based approach which models review networks as heterogeneous information networks.
2. A new weighting method for spam features is proposed to determine the relative importance of each feature and shows how effective each of features are in identifying spams from normal reviews.
3. NetSpam improves the accuracy compared to the state-of-the art in terms of time complexity, which highly depends to the number of features used to identify a spam review.

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, we model review dataset as in which reviews are connected through different node types.

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

Advantages:

1. To identify 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. To identify the spam user using positive and negative reviews in online social media.
4. To display only trusted reviews to the users.



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 fort= 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:

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

Where,

- $L_i - F_i$ describes days between last and first review for t= 28.

Users with calculated value greater than 0.5 take value 1 and others take 0. denotes days specified written review and first written review for a specific business. We have also =

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.

$$x_{DEV}(i) = \begin{cases} 0 & Otherwise \\ 1 - \frac{r_{ij} - avg_{e \in E_{*j}} r(e)}{4} > \beta_1 \end{cases} \qquad (3)$$

Where, 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 much more than first personal pronouns. In addition, spammers put '!' in their sentences as much as they can to increase impression on users and highlight their reviews among other ones. Reviews are close to each other based on their calculated value, take same values (in [0; 1]).

V. CONCLUSION

This paper presents a unique spam detection system particularly NetSpam in sight of a metapath plan and another graph based mostly strategy to name reviews counting on a rank-based naming methodology. The execution of the projected structure is assessed by utilizing review datasets. Our perceptions demonstrate that discovered weights by utilizing this metapath plan is exceptionally powerful in recognizing spam surveys and prompts a superior execution. moreover, we tend to found that even while not a prepare set, NetSpam will figure the importance of every component and it yields higher execution within the highlights' growth procedure, and performs superior to something past works, with simply few highlights. additionally, within the wake of characterizing four basic classifications for highlights our perceptions demonstrate that the review activity classification performs superior to something completely different classifications, concerning AP, terrorist group and within the discovered weights. The outcomes likewise affirm that utilizing diverse supervisions, just like the semi-administered strategy, haven't any detectable impact on deciding the overwhelming majority of the weighted highlights, equally as in varied datasets. Contribution half during this project, for user once searches question he can get the top-k edifice lists in addition mutually recommendation edifice by mistreatment customized recommendation rule.

VI. REFERENCES

[1]. J. Donfro, A whopping 20 % of yelp reviews are fake. http://www.businessinsider.com/20-percent-of-yelp-reviews-fake-2013-9. Accessed: 2015-07-30.

[2]. M. Ott, C. Cardie, and J. T. Hancock. Estimating the prevalence of deception in online review communities. In ACM WWW, 2012.

[3]. M. Ott, Y. Choi, C. Cardie, and J. T. Hancock. Finding deceptive opinion spam by any stretch of the imagination.In ACL, 2011.

[4]. Ch. Xu and J. Zhang. Combating product review spam campaigns via multiple heterogeneous pairwise features. In SIAM International Confer-ence on Data Mining, 2014.

[5]. N. Jindal and B. Liu. Opinion spam and analysis. In WSDM, 2008.

[6]. F. Li, M. Huang, Y. Yang, and X. Zhu. Learning to identify review spam. Proceedings of the 22nd International Joint Conference on Artificial Intelligence; IJCAI, 2011.

[7]. G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh. Ex-ploiting burstiness in reviews for review spammer detection. In ICWSM, 2013.

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

[9]. 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.

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

[11]. L. Akoglu, R. Chandy, and C. Faloutsos. Opinion fraud detection in online reviews bynetwork effects. In ICWSM, 2013.

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

[13]. S. Feng, R. Banerjee and Y. Choi. Syntactic stylometry for deception detection. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers; ACL, 2012.

[14]. N. Jindal, B. Liu, and E.-P. Lim. Finding unusual review patterns using unexpected rules. In ACM CIKM, 2012.

[15]. E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw. Detecting product review spammers using rating behaviors. In ACM CIKM, 2010.

[16]. A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh. Spotting opinion spammers using behavioral footprints. In ACM KDD, 2013.

[17]. S. Xie, G. Wang, S. Lin, and P. S. Yu. Review spam detection via temporal pattern discovery. In ACM KDD, 2012.

[18]. G. Wang, S. Xie, B. Liu, and P. S. Yu. Review graph based online store review spammer detection. IEEE ICDM, 2011.

[19]. Y. Sun and J. Han. Mining Heterogeneous Information Networks; Principles and Methodologies, In ICCCE, 2012.

[20]. A. Mukerjee, V. Venkataraman, B. Liu, and N. Glance. What Yelp Fake Review Filter Might Be Doing?, In ICWSM, 2013.

[21]. S. Feng, L. Xing, A. Gogar, and Y. Choi. Distributional footprints of deceptive product reviews. In ICWSM, 2012.

[22]. Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. In VLDB, 2011.

[23]. Y. Sun and J. Han. Rankclus: integrating clustering with ranking for heterogeneous information network analysis. In Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology, 2009.

[24]. C. Luo, R. Guan, Z. Wang, and C. Lin. HetPathMine: A Novel Transduc-tive Classification Algorithm on Heterogeneous Information Networks. In ECIR, 2014.

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

[26]. M. Luca and G. Zervas. Fake It Till You Make It: Reputation, Compe-tition, and Yelp Review Fraud., SSRN Electronic Journal, 2016.

[27]. 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.

[28]. 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.

[29]. A. Mukherjee, B. Liu, and N. Glance. Spotting Fake Reviewer Groups in Consumer Reviews. In ACM WWW, 2012.

[30]. A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari. Detection of review spam: A survey. Expert Systems with Applicants, Elsevier, 2014.

[31]. M. Crawford, T. D. Khoshgoftar, J. N. Prusa, A. Al. Ritcher, and H. Najada. Survey of Review Spam Detection Using Machine Learning Techniques. Journal of Big Data. 2015.

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

[33]. C. L. Lai, K. Q. Xu, R. Lau, Y. Li, and L. Jing. Toward a Language Modeling Approach for Consumer Review Spam Detection. In Proceed-ings of the 7th international conference on e-Business Engineering. 2011.

[34]. N. Jindal and B. Liu. Opinion Spam and Analysis. In WSDM, 2008.

[35]. S. Mukherjee, S. Dutta, and G. Weikum. Credible Review Detection with Limited Information using Consistency Features, In book: Machine Learning and Knowledge Discovery in Databases, 2016.

[36]. K. Weise. A Lie Detector Test for Online Reviewers. http://bloom.bg/1KAxzhK. Accessed: 2016-12-16.

[37]. M. Salehi, R. Sharma, M. Marzolla, M. Magnani, P. Siyari, and D. Mon-tesi. Spreading processes in multilayer networks. In IEEE Transactions on Network Science and Engineering. 2(2):65–83, 2015.

AUTHOR DETAILAS

G.PRASHANTI is working an assistant professor in VIGNAN'S LARA INSTITUTE OF TECHNOLOGY & SCIENCE. Vadlamudi-522213 Guntur Dist.She has Experience in the teaching field For 7 years and her interested in research areas are network security Steganography and data mining.

TIRUVEEDHULA PRIYANKA she is Currently pursuing MCA in MCA Department,Vignan's Lara Institute Of Technology&Science,Vadlamudi, Guntur(Dt), Andhra Pradesh, India. she received Bachelor of science from KRISHNA UNIVERSITY