International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

Spam Detection using Sentiment Analysis of Text

Vignesh N[1], Thanigaivel L[2], Vimalraj B.[3], Rajkamal J[4], AP V. Balamurugan[5]
[1, 2, 3, 4, 5]*Department of Computer Science and Engineering, S.A. Engineering College*

Abstract: Nowadays, a big part of people rely on available content in social media in their decisions (e.g. reviews and feedback on a topic or product). The possibility that anybody can leave a review provides a golden opportunity for spammers to write spam reviews about products and services for different interests. Identifying these spammers and the spam content is a hot topic of research and although a considerable number of studies have been done recently toward this end, but so far the methodologies put forth still barely detect spam reviews, and none of them show the importance of each extracted feature type. In this study, we propose a novel framework, named NetSpam, which utilizes spam features for modeling review datasets as heterogeneous information networks to map spam detection procedure into a classification problem in such 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 or negative . 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 tempting opportunity for spammers to write fake reviews designed to mislead users'.

II. LITERATURE SURVEY

A. Estimating The Prevalence Of Deception In Online Review Communities

Consumers' purchase decisions are increasingly influenced by user-generated online reviews. Accordingly, there has been growing concern about the potential for posting deceptive opinion spam---fictitious reviews that have been deliberately written to sound authentic, to deceive the reader. But while this practice has received considerable public attention and concern, relatively little is known about the actual prevalence, or rate, of deception in online review communities, and less still about the factors that influence it. We propose a generative model of deception which, in conjunction with a deception classifier, we use to explore the prevalence of deception in six popular online review communities: Expedia, Hotels.com, Orbitz, Priceline, TripAdvisor, and Yelp.

B. Finding Deceptive Opinion Spam By Any Stretch Of The Imagination

1) Authors: M. Ott, Y. Choi

Consumers increasingly rate, review and research products online (Jansen, 2010;Litvin et al., 2008). Consequently, websites containing consumer reviews are becoming targets of opinion spam. While recent work has focused primarily on manually identifiable instances of opinion spam, in this work we study deceptive opinion spam---fictitious opinions that have been deliberately written to sound authentic. Integrating work from psychology and computational linguistics, we develop and compare three approaches to detecting deceptive opinion spam, and ultimately develop a classifier that is nearly 90% accurate on our gold-standard opinion spam dataset. Based on feature analysis of our learned models, we additionally make several theoretical contributions, including revealing a relationship between deceptive opinions and imaginative writing.

C. Learning to identify review spam. Proceedings of the 22nd international joint conference on artificial intelligence

1) Authors: F. Li, M. Huang

In the past few years, sentiment analysis and opinion mining becomes a popular and important task. These studies all assume that their opinion resources are real and trustful. However, they may encounter the faked opinion or opinion spam problem. In this paper, we study this issue in the context of our product review mining system. On product review site, people may write faked reviews, called review spam, to promote their products, or defame their competitors' products. It is important to identify and filter out the review spam. Previous work only focuses on some heuristic rules, such as helpfulness voting, or rating deviation, which limits the performance of this task.In this paper, we exploit machine learning methods to identify review spam. Toward the end, we manually build a spam collection from our crawled reviews. We first analyze the effect of various features in spam identification. We also observe that the review spammer consistently writes spam. This provides us another view to identify review spam: we can identify if

©IJRASET: All Rights are Reserved

the author of the review is spammer. Based on this observation, we provide a twoview semi-supervised method, co-training, to exploit the large amount of unlabeled data. The experiment results show that our proposed method is effective. Our designed machine learning methods achieve significant improvements in comparison to the heuristic baselines.

D. Faloutsos. True view: harnessing the power of multiple review sites

1) Authors: A. j. Minnich, N. Chavoshi

Online reviews on products and services can be very useful for customers, but they need to be protected from manipulation. So far, most studies have focused on analyzing online reviews from a single hosting site. How could one leverage information from multiple review hosting sites? This is the key question in our work. In response, we develop a systematic methodology to merge, compare, and evaluate reviews from multiple hosting sites. We focus on hotel reviews and use more than 15 million reviews from more than 3.5 million users spanning three prominent travel sites. Our work consists of three thrusts: (a) we develop novel features capable of identifying cross-site discrepancies effectively, (b) we conduct arguably the first extensive study of cross-site variations using real data, and develop a hotel identity-matching method with 93% accuracy, (c) we introduce the TrueView score, as a proof of concept that cross-site analysis can better inform the end user. Our results show that: (1) we detect 7 times more suspicious hotels by using multiple sites compared to using the three sites in isolation, and (2) we find that 20% of all hotels appearing in all three sites seem to have low trustworthiness score. Our work is an early effort that explores the advantages and the challenges in using multiple reviewing sites towards more informed decision making.

E. Spotting Fake Reviews Via Collective PU Learning

1) Authors: H. Li, Z. Chen

Online reviews have become an increasingly important resource for decision making and product designing. But reviews systems are often targeted by opinion spamming. Although fake review detection has been studied by researchers for years using supervised learning, ground truth of large scale datasets is still unavailable and most of the existing approaches of supervised learning are based on pseudo fake reviews rather than real fake reviews. Working with Dianping, the largest Chinese review hosting site, we present the first reported work on fake review detection in Chinese with filtered reviews from Dianping's fake review detection system. Dumplings algorithm has a very high precision, but the recall is hard to know. This means that all fake reviews detected by the system are almost certainly vf fake, but the remaining reviews (unknown set) may not be all genuine. Since the unknown set may contain many fake reviews, it is more appropriate to treat it as an unlabeled set. This calls for the model of learning from positive and unlabeled examples (PU learning). By leveraging the intricate dependencies among the reviews, users and IP addresses, we first propose a collective classification algorithm called Multi-typed Heterogeneous Collective Classification (MHCC) and then extend it to Collective Positive and Unlabeled learning (CPU). Our experiments are conducted on real-life reviews of 500 restaurants in Shanghai, China. Results show that our proposed models can markedly improve the F1 scores of strong baselines in both PU and non-PU learning settings. Since our models only use language independent features, they can be easily generalized to other languages.

F. An Approach of Two-Way Spam Detection Based on Boosting Pages Analysis

1) Authors: Chakrit Likitkhajorn

Web spam is an attempt to increase rank of inappropriate web pages. Link spam is one of web spam technique which aims for increasing rank by creating artificial popularity of the page by increasing in-links on the page. This task required creating a lot of boosting pages which are a page spammer created for boosting the rank of the spam page. In this paper, we proposed a technique to detect web spam based on finding these boosting pages instead of web spam page itself. We start from a small set of spam seed pages to find boosting pages. Then web spam pages will be detected using boosting pages. Moreover, we evaluate normal pages to increase the preciseness of our algorithm. Experimental results shows that most of spam pages can be retrieved. Moreover, the accuracy of our algorithm can be increased by normal page analysis

G. A Novel Approach Toward Spam Detection Based on Iterative Patterns

1) Authors: Mohammad Razmara

Spamming is becoming a major threat that negatively impacts the usability of e-mail. Although lots of techniques have been proposed for detecting and blocking spam messages, Spammers still spread spam e-mails for different purposes such as advertising, phishing, adult and other purposes and there is not any complete solution for this problem. In this work we present a novel solution

©IJRASET: All Rights are Reserved

toward spam filtering by using a new set of features for classification models. These features are the sequential unique and closed patterns which are extracted from the content of messages. After applying a term selection method, we show that these features have good performance in classifying spam messages from legitimate messages. The achieved results on 6 different datasets show the effectiveness of our proposed method compared to close similar methods. We outperform the accuracy near +2% compared to related state of arts. In addition our method is resilient against injecting irrelevant and bothersome words.

H. Sentiment Analysis in A Cross-Media Analysis Framework

1) Author: Yonas Woldemariam

This paper introduces the implementation and integration of a sentiment analysis pipeline into the ongoingopen source cross-media analysis framework. The pipeline includes the following components; chat room cleaner, NLP and sentiment analyzer. Before the integration, we also compare two broad categories of sentiment analysis methods,
namey lexicon-based and machine learning approaches. We mainly focus on finding out which method is appropriate to detect sentiments from forum discussion posts. In order to conduct our experiments, we use the apache-hadoop framework with its lexicon-based sentiment prediction algorithm and Stanford coreNLP library with the Recursive Neural Tensor Network (RNTN) model. The lexicon- based uses sentiment dictionary containing words annotated with sentiment labels and other basic lexical features, and the later one is trained on Sentiment Treebank with 215,154 phrases, labeled using Amazon Turk. Our overall performance evaluation shows that RNTN outperforms the lexicon-based by 9.88% accuracy on variable length positive, negative, and neutral comments. How- ever, the lexicon-based shows better performance on classifying positive comments. We also found out that the F1-score values of the Lexicon-based is greater by 0.16 from the RNTN.

I. Spam Mail Detection through Data Mining Techniques

1) Author: Shubhi Shrivastava

In todays electronic world a huge part of communication, both professional and private, takes place in the form of electronic mails or emails. However, due to advertising agencies and social networking websites most of the emails circulated contain unwanted information which is not relevant to the user. Spam emails are a type of electronic mail where the user receives unsolicited messages via email. Spam emails cause inconvenience and financial loss to the recipients so there is a need to filter them and separate them from the legitimate emails. Many algorithms and filters have been developed to detect the spam emails but spammers continuously evolve and sophisticate their spamming techniques due to which the existing filters are becoming less effective. The method proposed in this paper involves creating a spam filter using binary and continuous probability distributions. The algorithms implemented in building the classifier model are Naive Bayes and Decision Trees. The effect of overfitting on the performance and accuracy of decision trees is analyzed. Finally, the better classifier model is identified based on its accuracy to correctly classify spam and non-spam emails.

J. Support Vector Machines, Import Vector Machines And Releance Vector Machines For Hyperspectral Classification-Acomparison

1) Author: Andreas Ch. Braun, Uwe Weidner, Stefan Hinz

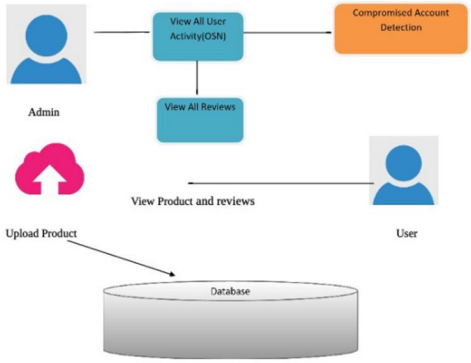
Support Vector Machines (SVM) have gained increasing attention due to their classification accuracy, robustness and indifference towards the input data type. Thus, they are widely used in the remote sensing community – and especially among researchers working on hyperspectral datasets. However, since their first publication a lot of enhancements and adaptations have been proposed, many of which aim at introducing probability distributions and the Bayes theorem to SVM. Within this paper, we present a classification result of a HyMap dataset using two of the proposed enhancements – Import Vector Machines and Relevance Vector Machines – and compare them to the Support Vector Machine.

III. PROPOSED SYSTEM

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.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.

©IJRASET: All Rights are Reserved

Architecture Diagram



IV. CONCLUSION

This study introduces a novel spam detection framework namely NetSpam based on a metapath concept as well as a new graph-based method to label reviews relying 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 can be very effective in identifying spam reviews and leads to a better performance. In addition, we found that even without a train set, NetSpam can calculate the importance of each feature and it yields better performance in the features' addition process, and performs better than previous works, with only a small number of features.

Moreover, after defining four main categories for features our observations show that the reviewsbehavioral category performs better than other categories, in terms of AP, AUC as well as in the calculated weights.

The results also confirm that using different supervisions, similar to the semi-supervised method, have no noticeable effect on determining most of the weighted features, just as in different datasets. For future work, metapath concept can be applied to other problems in this field.

For example, similar framework can be used to find spammer communities. For finding community, reviews can be connected through group spammer features (such as the proposed feature in [29]) and reviews with highest similarity based on metapth concept are known as communities.

In addition, utilizing the product features is an interesting future work on this study as we used features more related to spotting spammers and spam reviews. Moreover, while single networks has received considerable attention from various disciplines for over a decade, information diffusion and content sharing in multilayer networks is still a young research.

Addressing the problem of spam detection in such networks can be considered as a new research line in this field.

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 Conference 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. Exploiting 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.

©IJRASET: All Rights are Reserved