ISSN 2395-1621

A Survey on Machine Learning Methods for Online Deceptive Review Detection

[#1]Priyanka M. Shinde, [#2]Prof. H.P.Channe

[1]shindep156@gmail.com
[2]hpchanne@pict.edu

[#1]Student, Computer Department,
[#2]Prof., Computer Department

S.P.P.U University
P.I.C.T., Pune, India

ABSTRACT

Online reviews are often the primary factor in a customer's decision to purchase a product or service, and are a valuable source of information that can be used to determine public opinion on these products or services Now-a-days not only users but organizers also consider opinions as their main impact factor in manufacturing product or service. Opinion of user is very important for business. Because of opinion of actual user further consumers should think to use that resource. In Business, opinion review has great impact to economical bottom line. Reliance on online reviews gives rise to the potential concern that wrongdoers may create false reviews to artificially promote or devalue products and services. This practice is known as Opinion (Review) Spam, where spammers manipulate and poison reviews (i.e., making fake, untruthful, or deceptive reviews) for profit or gain. Since not all online reviews are truthful and trustworthy, it is important to develop techniques for detecting review spam. By extracting meaningful features from the text, it is possible to conduct review spam detection using various machine learning techniques. In our survey we are going to study the different semi-supervised learning algorithms for finding out these spam reviews and ensemble method for multiclass and multilabel text categorization.

Keywords— Online review spam, semi-supervised learning, unlabeled reviews, PU learning, Co-training, EM algorithm

ARTICLE INFO

Article History

Received: 15th March 2018

Received in revised form :

15th March 2018

Accepted: 17th March 2018

Published online :

21st March 2018

I. INTRODUCTION

As with more consumers using online opinion reviews to inform their service decision making, opinion reviews have an economical impact on the bottom line of businesses. Opinion spamming is becoming more sophisticated and, in some cases, organized, due to the potential to profit from such activities. For example, some businesses reportedly recruited online users such as professional fake review writers to post fake opinions. These opinions can be used to market and promote a particular business, spread rumors and damage the reputation of a competing business, or influence online users opinions and views about a particular topic.[1]

Machine learning attempts to tell how to automatically find a good predictor based on past experiences. we're going to look at some of the significant results from machine learning. One goal is to learn some of the techniques of machine learning, but also, just as significant we are going to get a glimpse of the research front and the sort of approaches researchers have take toward how to increase the classification accuracy in online deceptive review detection by solving the problem of multiclass and multilabel text categorization using ensemble method

Machine learning has three approaches the tradition one is supervised and unsupervised. Supervised learning has been traditionally used to detect fake reviews, supervised learning approaches suffer from several limitations. For example, unless one can be assured of the ``quality'' of the reviews used in the training dataset, we will have a garbage-in-garbage-out situation. In addition, the amount of labelled data points used to train the classifier can be difficult to obtain and update, given the dynamic nature of online reviews[2]

Some limitations in supervised learning methods could be addressed using automatic labeling, a process

known as semi-supervised learning. Semi-supevised learning is the third approach after the previous two.In the latter, a large number of unlabeled data points are used, instead of labelled data points. As such, labelled data points can be sparsely present and using those points, labels of the unknown instances are automatically generated first, which can then be used to train a classifier and evaluate a given review[3]

In other domains, it has been found that using unlabeled data in conjunction with a small amount of labeled data can considerably improve learner accuracy compared to completely supervised methods. a two-view semi-supervised method for review spam detection was created by employing the framework of a co-training algorithm to make use of the large amount of unlabeled reviews available. The co-training algorithm is a bootstrapping method that uses a set of labeled data to incrementally apply labels to unlabeled data. It trains 2 classifiers on 2 distinct sets of features and adds the instances most confidently labelled by each classifier to the training set. [2][3]This effectively allows large datasets to be generated and used for classification, reducing the demand to manually produce labelled training instances. A modified version of the co-training algorithm that only adds instances that were assigned the same label by both classifiers was also proposed. Their dataset was generated with the assistance of students who manually labelled 6000 reviews collected from Epinions.com, 1394 of which were labelled as review spam. Four groups of review centric features were created: content, sentiment, product and metadata. Another two groups of reviewer centric features were created: profile and behavioural.[12]

In order to use the two-view method for adding unlabeled instances to the training set, classifiers were trained on each set of features such as one with review. centric features and another with reviewer centric ones. Note that these 2 classifiers are only used to add instances to the labeled data and the final classifier is trained using all available features, both review centric and reviewer centric. Experiments were conducted using Naïve Bayes, [2]Logistic Regression and SVM [6]with 10-fold cross validation, and it was found that Naïve Bayes was the best performer, so all additional work was performed with Naïve Bayes. They observed that using the co-training semi-supervised method, they were able to obtain an F-Score of .609, which was higher than the 0.583 they obtained when not including any unlabeled data. Further, it was observed that by using their co-training with agreement modification, they were able to raise this value to 0.631. While these F-Scores appear low, it is hard to compare them with the performance from other studies as they used their own dataset. The results do seem to indicate that this type of semi-supervised learning may indeed help in the area of review spam detection and demands further study with additional datasets.

Ensemble learning algorithms train multiple classifiers and then combine their predictions. Since the generalization ability of an ensemble classifier can be much better than a single learner, the algorithms and applications of ensemble learning have been widely studied in recent years. In many successful applications, ensemble learning classifiers usually achieve the best performance in the literature[16]

In this paper, we are going to study several semi-supervised learning approaches to improve the classication, as well as we will identify that how the new dimensions in the feature vector that are Parts-of-Speech features, Linguistic and Word Count features and Sentimental Content features will help us to obtain better results. We further study and survey that whether these techniques help us in our proposed method of text categorization for multiclass and multilabel problem using ensemble method.

II. POPULAR ML ALGORITHMS

a. Naïve Bayes Classifier: It is a supervised classification method developed using Bayes' Theorem of conditional probability with a 'Naive' assumption that every pair of feature is mutually independent. That is, in simpler words, presence of a feature is not effected by presence of another by any means. Irrespective of this over-simplified assumption, NB classifiers performed quite well in many practical situations, like in text classification and spam detection. Only a small amount of training data is need to estimate certain parameters. Beside, NB classifiers have considerably outperformed even highly advanced classification techniques

b. Support Vector Machine: SVM, another supervised classification algorithm proposed by Vapnik in 1960s have recently attracted an major attention of researchers. The simple geometrical explanation of this approach involves determining an optimal separating plane or hyper plane that separates the two classes or clusters of data points justly and is equidistant from both of them. SVM was defined at first for linear distribution of data points. Later, the kernel function was introduced to tackle nonlinear datas as well.[7][15]

c. Decision Tree: A classification tree, popularly known as decision tree is one of the most successful supervised learning algorithm. It constructs a graph or tree that employs branching technique to demonstrate every probable result of a decision. In a decision tree representation, every internal node tests a feature, each branch corresponds to outcome of the parent node and every leaf finally assigns the class label. To classify an instance, a top-down approach is applied starting at the root of the tree. For a certain feature or node, the branch concurring to the value of the data point for that attribute is considered till a leaf is reached or a label is decided[9][15]

According to the previous system the comparative results of these ML algorithms is as follws. Dataset here is The raw tweets were taken from Sentiment140 data set. Then those are pre-processed and labeled using a python program. Each of these classifier were exposed to same data. Same algorithm of feature selection, dimensionality reduction and k-fold validation were employed in each cases. The algorithms were compared based on the training time, prediction time and accuracy of the prediction is as follows in Table1.[15]

Algorithm	Training time(sec.)	Prediction time(sec)	Accuracy
Naive bayes	2.708	0.328	0.692
SVM	6.485	2.054	0.6565
Decision Tree	454.609	0.063	0.69

Table1: Comparison between ML algorithms

III. RELATED WORK

Opinions from social media are increasingly used by individuals and organizations for making purchase decisions and making choices at elections and for marketing and product design. Positive opinions often mean profits and fames for businesses and individuals, which, unfortunately, give strong incentives for people to game the system by posting fake opinions or reviews to promote or to discredit some target products, services, organizations, individuals, and even ideas without disclosing their true intentions, or the person or organization that they are secretly working for. Such individuals are called opinion spammers and their activities are called opinion spamming,[12] The key challenge of opinion spam detection is that unlike other forms of spam, it is very hard, if not impossible, to recognize fake opinions by manually reading them, which makes it difficult to find opinion spam data to help design and evaluate detection algorithms. For other forms of spam, one can recognize them fairly easily[1][12]
Three main types of data have been used for review spam detection.

Review content : The actual text content of each review. From the content, we can extract linguistic features such as word and POS n-grams and other syntactic and semantic clues for deceptions and lies. However, linguistic features may not be enough because one can fairly easily craft a fake review that is just like a genuine one. For example, one can write a fake positive review for a bad restaurant based on his true experience in a good restaurant.

Meta-data about the review : The data such as the star rating given to each review, user-id of the reviewer, the time when the review was posted, the time taken to write the review, the host IP address and MAC address of the reviewer 's computer, the geo-location of the reviewer, and the sequence of clicks at the review site. From such data, we can mine many types of abnormal behavioural patterns of reviewers and their reviews. For example, from review ratings, we may find that a reviewer wrote only positive reviews for a brand and only negative reviews for a competing brand. Along a similar line, if multiple user-ids from the same computer posted a number of positive reviews about a product, these reviews are suspicious. Also, if the positive reviews for a hotel are all from the nearby area of the hotel, they are clearly not trustworthy.

Product information : Information about the entity being reviewed, e.g., the product description and sales volume /rank. For example, a product is not selling well but has many positive reviews, which is hard to believe.[10][11]

Authors used review and reviewer features to design a two-view semi-supervised method, by employing the framework of co-training algorithm to detect spam reviews. In this approach, the co-training algorithm uses the large amount of unlabeled examples to train the algorithm. PU-learning reportedly achieves an F-measure of 83.7% with Näive Bayes, using only 100 positive examples. While this is better than the findings reported by author, where 6000 labelled instances and co-training were used, it is difficult to make a conclusive statement as both approaches use different datasets[1]

Positive Unlabeled (PU) learning is another semi-supervised learning approach , which can be used to build an accurate classifier even without having labelled negative training examples. Several PU learning techniques have been applied successfully in document classification with promising results. Hernández et al. first used this technique to detect review spam.[2] This technique is applied to this algorithm for deceptive review detection using half the datasets used here. Although [2] achieved an F-Score of 0.837 using just 100 positive instances for training, the results published did not disclose the accuracy or feature characteristics of their methods, which made it difficult to compare performance. We remark that both datasets also have different sentimental polarities .Their assumption regarding continual reining of negative instances over iterations will not always hold in practice, as pointed out by Li et al. [3]. then showed that PU-LEA identified much fewer positive examples from the unlabeled set. In addition, the authors attempted to detected review spams in Chinese language reviews of restaurants from Dianping.com. In their approach, LP (Labelled Positive) was used, also considering the fact that unknown set is really an unlabeled set rather than the non-fake review set. According to the authors, PU learning not only outperforms SVM but also detects a large number of potentially fake reviews hidden in the unlabeled set. The authors used publicly available PU learning system.[3]

Today's search engines are seriously threatened by malicious web spam that attempts to subvert the unbiased searching and ranking services provided by the engines. Search engines are today combating web spam with a variety of ad hoc, often proprietary techniques. We believe that our work is a first attempt at formalizing the problem and at introducing a comprehensive solution to assist in the detection of web spam. Our experimental results show that we can effectively identify a significant number of strongly reputable (non-spam) pages. In a search engine, TrustRank can be used either separately to filter the index, or in combination with PageRank and other metrics to rank search results. TrustRank and two baseline strategies, and evaluate their performance using our sample X: The baseline precision score for the sample X is 613/(613 + 135) = 0.82. (using TrustRank algorithm here the two baseline are ignorant trust and PageRank)[4] Challenge for the future research are Instead of selecting the entire seed set at once, one could think of an iterative process: after the oracle has evaluated some pages, we could reconsider what pages it should evaluate next, based on the previous outcome. Such issues are a challenge for future research[4].

The survey gives various aspects of content-based spam on the web using a real-world data set from the MSN

Search crawler. author have presented a number of heuristic methods for detecting content based spam. Some of their spam detection methods are more effective than others, however when used in isolation their methods may not identify all of the spam pages. For this reason, they combined their spam-detection methods to create a highly accurate C4.5 classifier. their classifier can correctly identify 86.2% of all spam pages, while flagging very few legitimate pages as spam. [5] Challenges for future research are Some of the methods for spam detection presented in this paper may be easily fooled by spammers. For example, their is a section in paper may be fooled by adding frequent words to otherwise meaningless content. Although we believe that it will be relatively hard for a spam web page to fool all of their techniques, we may see some degradation of the classification performance over time. To accommodate for this we plan to study how we can use natural language techniques to recognize artificially generated text. Additionally, the heuristic methods that we presented in this paper may very well serve as part of a "multi-layered" spam detection system. In the first layer, we can apply the computationally cheap methods presented in this paper to capture most of the spam. After that, we can apply more computationally expensive techniques or link analysis to capture the remaining spam pages. Therefore, we plan to investigate how we can build and evaluate such a layered spam-detection system[5]

Shojaee et al. [11] proposed a novel method for detecting review spam by using Stylometric (Lexical and Syntactic) features. (For further details on Stylometric featurers). The features in this work are categorized as either lexical features or syntactic features. Lexical features are character/word based features, while syntactic features represent the writing style of the reviewers at the sentence level, such as occurrences of punctuations or function words. In this work they built SVM and Naïve Bayes classifiers on the dataset created by Ott et al. [3] using a hybrid set of both the lexical and syntactic features and compared this with using either lexical or syntactic features alone. Using 10-fold cross validation, they observed that the hybrid feature set using the SVM learner achieved the highest performance, an F-measure of 84 %. Additionally, SVM outperformed Naïve Bayes for all sets of features. A potential concern of this study is that the model was trained and evaluated on synthetic fake reviews. Due to this, it is possible that the classifier performance measured is a poor indication of real world performance, Also there is no comparison evaluation to determine if using these Stylometric features in addition to n-gram features enhances classification performance.[17]

The study describes the use of support vector machines (SVM's) in classifying e-mail as spam or nonspam by comparing it to three other classification algorithms: Ripper, Rocchio, and boosting decision trees. These four algorithms were tested on two different data sets: one data set where the number of feature were constrained to the 1000 best features and another data set where the dimensionality was over 7000. SVM's performed best when using binary features. For both data sets, boosting trees and SVM's had acceptable test performance in terms of accuracy and speed. However, SVM's had significantly less training time[6] Challenges for future research here, it is challenging to identify fake opinions, as one may need to also understand the context of the postings in order to determine whether the particular opinion is deceptive[6]

Author proposes that using profile compatibility to differentiate genuine and fake product reviews. For each product, a collective profile is derived from a separate collection of reviews. Such a profile contains a number of aspects of the product, together with their descriptions. For a given unseen review about the same product, we build a test profile using the same approach. We then perform a bidirectional alignment between the test and the collective profile, to compute a list of aspect-wise compatible features. We adopt Ott et al. op spam v1.3 dataset for identifying truthful vs. deceptive reviews. We extend the recently proposed N- GRAM+SYN model of Feng et al. by incorporating profile compatibility features, showing such an addition significantly improves upon their state-ofart classification performance. [7] In the methodology author for a fair comparison consider using unigrams, bigrams, and the union of both, and choose the best combination with deep syntax features as our baseline system Challenges for future research here, Another particularly interesting direction is to In future explore on how their C+N-GRAM+SYN model performs on identifying fake negative reviews, as recently released by author, rather than the positive reviews used in this work. While negative opinion spam is more hazardous to a brand's fame compared to positive ones, and thus identifying fake negative reviews might be more crucial, one potential difficulty of our approach is that genuine extremely negative reviews written by renowned reviewers are much more sparse than extremely positive ones, especially for famous products, such as the most popular Chicago hotels in the op spam v1.3 dataset.[7]

Experiment is conducted using a large number of reviewers and reviews of manufactured products from Amazon.com . A user study is used to verify whether the ranking produced by our algorithm confirms to people's perceptions of spammer groups. Frequent pattern mining and ranking: The number of candidate groups mined in step 1 was 2,273 with the minimum support of 3. Each group consists of at least two reviewers. SVM rank was then applied to produce the final ranking of the candidate spammer groups. This ranked result was employed in user study. User agreement study: This was conducted using three (3) independent human judges. the following three types of groups for the user agreement study: top 100 groups, middle 100 groups and bottom 100 groups. The Cohen's Kappa scores are all above' 0.8 'which indicates almost perfect agreements[8]

IV. CONCLUSION

The fore most target of ML researchers is to design more efficient (in terms of both time and space)and practical general purpose learning methods that can perform better over a widespread domain. In the context of ML, the efficiency with which a method utilises data resources that is also an important performance paradigm along with time and space complexity.

As the influence of online opinion and reviews on users increasing day by day so,the capability to detect deceptive

online reviews is much necessary than before. In this paper, we study different ML algorithms and observe that which semi-supervised approach is more accurate in results by comparing with several algorithms .we also studied that extracting the features from text such as Parts-of-Speech features, Linguistic and Word Count features and Sentimental Content features give better results in online deceptive review detection. The study shows that these algorithms also used to solve the multiclass and multilabel problem by introducing twin SVM in text categorization.
.

ACKNOWLEDGMENT

Prof. H.P.Channe for giving me all the help and guidance I needed. I am really grateful to them for their kind support throughout the analysis and design phase. Their valuable criticism and suggestions were very helpful. I extend my thank to my dearest parents who always motivate me for doing something different.

REFERENCES

[1] J.Rout, A.Dalmia,"Revisiting Semi-Supervised Learning for Online Deceptive Review Detection," Ministry of Electronics and Information Technology (MeitY) India,2017

[2] F. Li, M. Huang, Y. Yang, and X. Zhu, ``Learning to identify review spam," in Proc. 22nd Int. Joint Conf. Artif. Intell. (IJAI), 2011, pp. 24882493

[3] D. Hernández, R. Guzmán, M. Móntes-y-Gomez, and P. Rosso, ``Using PU-learning to detect deceptive opinion spam," in Proc. 4th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Anal., 2013,pp. 3845.

[4] H. Li, B. Liu, A. Mukherjee, and J. Shao, ``Spotting fake reviews using positive-unlabeled learning," Comput. Sistemas, vol. 18, no. 3, pp. 467475, 2014, doi: 10.13053/CyS-18-3-2035

[5] Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen, ``Combating Web spam with trustrank," in Proc. 13th Int. Conf. Very Large Data Bases (VLDB), vol. 30. 2004, pp. 576587

[6] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly, ``Detecting spam Web pages through content analysis," in Proc. 15th Int. Conf. World Wide Web (WWW), 2006, pp. 8392, doi: 10.1145/1135777.1135794

[7] H. Drucker, D. Wu, and V. N. Vapnik, ``Support vector machines for spam categorization," IEEE Trans. Neural Netw., vol. 10, no. 5, pp. 10481054, Sep. 1999, doi: 10.1109/72.788645

[8] V. W. Feng and G. Hirst, ``Detecting deceptive opinions with prole compatibility," in Proc. 6th Int. Joint Conf. Natural Lang. Process. (IJCNLP), 2013, pp. 338346.

[9] A. Mukherjee, B. Liu, J. Wang, N. Glance, and N. Jindal, ``Detecting group review spam," in Proc. 20th Int. Conf. Companion World Wide Web, 2011, pp. 9394, doi: 10.1145/1963192.1963240

[10] Jindal, Nitin, Bing Liu, and Ee-Peng Lim. Finding unusual review Patterns using unexpected rules. In Proceedings of acm international conference on information and Knowledge management (ciKm-2010). 2010. doi:10.1145/1871437.1871669

[11] Jindal, Nitin and Bing Liu," review spam detection," In Proceedings of WWW (Poster paper). 2007

[12] Liu, Jing jing, Yunbo Cao, Chin-Yew Lin, Yalou Huang, and Ming Zhou. ,"Low-quality product review detection in opinion summarization," In Proceedings of the Joint conference on empirical methods in natural Language Processing and computational natural Language Learning (emnLP-conLL-2007). 2007.

[13] Bing Liu ,"Sentiment Analysis and Opinion Mining," ISBN: 9781608458851

[14]. A. Blum and T. Mitchell, ``Combining labeled and unlabeled data with co-training," in Proc. 11th Annu. Conf. Comput. Learn. Theory, 1998, pp. 92100, doi: 10.1145/279943.279962

[15] Kajaree Das, Rabi Narayan Behera, ``A Survey on Machine Learning: Concept, Algorithms and Applications," in IJIRCCE Vol. 5, Issue 2, February 2017

[16] B. Zhang ,X. Xu and J. Su," An Ensemble Method for Multi-class and Multi-label Text Categorization," ACM Computing Surveys,2015, vol.10., no.3. , pp. 15-45

[17] N.Cristianini and J. Shawe-Taylor. An Introduction to Suppor t Vector Machines. Cambridge University Press,2000

[18] Adankon, M. And Cheriet, M. Genetic Algorithm Based training for semi-supervised SVM, Neural Computing And applications,volume19(8), 1197-1206,2010

[19] Dixit S, Agrawal AJ (2013) Survey on review spam detection. Int J Comput Commun Technol ISSN (PRINT) 4:0975–7449

