Survey on Deceptive Reviews Detection using Text Mining

Priyanka Patani[1] Pinal Patel[2]
[1,2]Department of Computer Engineering
[1,2]GEC Gandhinagar, India

Abstract— As eCommerce and online service providing business is boosting up with huge availability of internet to end users. User are preferring to avail online services and searching for online purchasing, reviews and opinions of previous users are becoming important for any type of business. Online reviews are not only important for businesses those provide online services but also for the businesses those are not providing online services. Due to high importance of opinions from various users there is one another issue occurred in this era that is 'Fake Reviews' or 'Deceptive Reviews'. To find out reviews published by real users or the users those have really obtained the services from the provide has become the important research topic. As a part of our research work we have studied so many technique that can try to identify the deceptive reviews. During the paper we have discussed many methods and approaches those are useful for detecting deceptive reviews.

Keywords: PU Learning, KNN, Deceptive Reviews

I. Introduction

As Deceptive Opinions are not real experience, they do not present real facts of the Product or any Services. Deceptive Opinions are written by Human or by machine generated to show more positive impact or more negativity to any products or any services. With the high usage of E-Commerce, customer chooses online purchase and Online Booking of various services, they do not have aware real facts of the products or services, and they are dependent on various Reviews written by previous customers. Their decision is based on Reviews and Ratings of the Service or Product on Website or Application.

Opinion spamming are being part of business and those are becoming sophisticated during the time of online digital marketing. In some cases those are intensively organized, because of the potential to hidden benefit from such activities. For example, many business houses reportedly assigned work for online users (e.g. professional fake review writers) to post fake reviews for their own services or for the competitors. These reviews can be used for marketing purpose and promote a particular activity or services, spread rumours and damage the reputation of a another related business, or influence online users' real opinions and views about a particular topic (e.g. during elections) [1].

There are main two basic methods for detecting the deceptive spam reviews. One of them is Supervised Learning and another one is Semi Supervised learning. In most cases supervised learning method can be used for detecting review spam by considering it as the classification problem. It classifies the detection problem into two classes (Binary Classes): spam reviews and non-spam reviews. As per our research, the first researchers to have studied this type of method on deceptive opinion spam were Jindal et al. [2]. Main issue with the same system is to find accurately labelled datasets of review spam, due to this problem the use of supervised learning is not always applicable. Unsupervised learning provides a solution for this, as it doesn't require labelled data.

A novel unsupervised text mining model (Unsupervised Learning Model) was developed and aggregated into a semantic language model for detecting fake reviews or opinions by Raymond et al. [1] and it was compared against supervised learning methods. In other domains then Text Mining, there is a fact has been found that when unlabelled data is used with a small amount of labelled data, this method can considerably increase accuracy of learner with comparison of entirely supervised methods [7]. In an another study conducted by Li et al. [9], a two-view semi-supervised algorithm for "review spam detection" was developed by providing the framework of a co-training algorithm to make use of the larger numbers of unlabelled opinions available.

By Blum and Mitchell [3] initially developed the co-training algorithm with a special method. Their method used a set of labelled opinion reviews to apply correct labels to unlabelled opinion data in incrementally way. Their method trains two classifiers on 2 different sets of features (selected from all features) and adds the instances from the dataset those are most confidently labelled by each classifier used to the training set. This special method effectively allows huge datasets to be generated and used for classification problem of opinion spam. It is also reducing the manual work to manually produce labelled training instances. A newer updated version of the co-training classification algorithm that only adds instances (records) that were assigned the same result by both classifiers used in co-training was also proposed. They generated dataset with the help of students who manually labelled more than six thousand reviews gathered from the website Epinions.com, 1394 of the same dataset were labelled as opinion spam. They generated 4 groups of review centric features: content, sentiment, product and metadata. Another 2 groups of reviewer those are based on reviewer centric features were created: profile based and behavioural based.

II. Related work

In the paper, "Revisiting Semi-Supervised Learning for Online Deceptive Review Detection" by "JITENDRA KUMAR ROUT1, ANMOL DALMIA1, KIM-KWANG RAYMOND CHOO2, (Senior Member, IEEE), SAMBIT BAKSHI1, (Member, IEEE), AND SANJAY KUMAR JENA1, (Senior Member, IEEE)" published in 2017 at IEEE access, they explain how semi-supervised learning methods can be used to detect spam reviews, prior to demonstrating its utility using a data set of hotel reviews.

The following feature points are chosen to be extracted and used for the experiments from the dataset:

- Sentiment Polarity
- Parts of Speech (POS) tags
- Linguistic Inquiry and Word Count (LIWC)

– Bigram frequency counts

For their future research along this direction includes implementing and evaluating the proposed approach in the real-world, for example, using the approach on data collected from various websites in real-time. Also, minimal meta-data are considered in this work during classification. Future investigation may include a better integrating of minimal meta-data. Details steps for their flow is like.

Algorithm 1 Co-Training Algorithm

INPUT: Labeled instance set L, and unlabeled instance set U.

OUTPUT: Deployable classifier, C.

Create set of unlabeled examples, U', by randomly sampling u examples from U;
for each feature vector x in $L \cup U$ **do**
  partition x to tuple of views, (x_1, x_2);
end for
for k iterations **do**
  $h_1 \leftarrow train(x_1)\ \forall (x_1, x_2) \in L$;
  $h_2 \leftarrow train(x_2)\ \forall (x_1, x_2) \in L$;
  Let h_1 label p positive and n negative examples from U';
  Let h_2 label p positive and n negative examples from U';
  Add labeled examples to L;
  Randomly sample $2(p+n)$ examples from U to U';
end for

As they have partitioned their dataset in various blocks and evaluated the result for the same partitions.

A. *Result Obtained from Their Research Work:*

Partition	Learner	Accuracy	Precision	Recall	F-Score
75-25	k-NN	0.7626	0.9150	0.7011	0.7939
	Logistic Regression	0.5025	0.9950	0.5012	0.6667
	Random Forest	0.6075	0.7800	0.5799	0.6652
	Stochastic Gradient Descent	0.5075	0.9900	0.5038	0.6678
80-20	k-NN	0.7469	0.8063	0.7207	0.7612
	Logistic Regression	0.5094	1.0	0.5047	0.6702
	Random Forest	0.7281	0.9	0.6699	0.7680
	Stochastic Gradient Descent	0.5031	1.0	0.5016	0.6681
90-10	k-NN	0.7353	0.8750	0.6863	0.7692
	Logistic Regression	0.5125	1.0	0.5063	0.6723
	Random Forest	0.6813	0.8000	0.6465	0.7175
	Stochastic Gradient Descent	0.6750	0.9500	0.6129	0.7451

Table 1: Result of PU-KNN

In another research "Detecting Spammer Groups from Product Reviews: A Partially Supervised Learning Model" researchers "LU ZHANG, ZHIANG WU, (MEMBER, IEEE), AND JIE CAO" used a novel approach. They propose a partially supervised learning model (PSGD) to detect spammer groups. By labeling some spammer groups as positive instances, PSGD applies Positive Unlabeled Learning (PU-Learning) to study a classifier as spammer group detector from positive instances (labeled spammer groups) and unlabeled instances (unlabeled groups). They extract reliable negative set in terms of the positive instances and the distinctive features by combining the positive instances, extracted negative instances and unlabeled instances.

They convert the PU-Learning problem into the well-known semi-supervised learning problem, and then use Naive Bayesian model and EM algorithm to train a classifier for spammer group detection. Below is the model of PSDG.

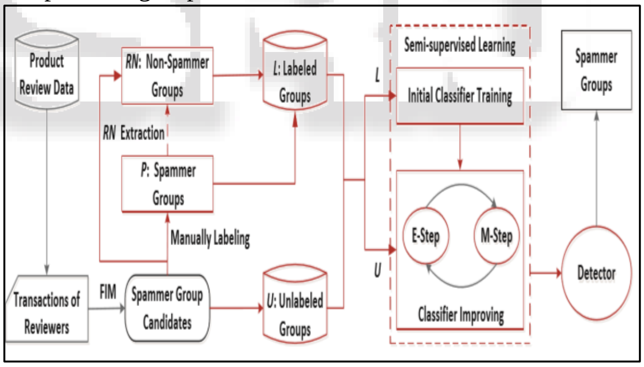

Fig. 1: Model for PSDG

They have used amzon china ecommerce dataset that contains 21,51,963 reviews from various users and for various products. They considered the product those have been reviewed for atleast 15 times. They have generated various feature sets and based on ration of positive instanance.

Feature Set	Precision	Recall	F-Measure
Set 1	0.8	0.86	0.82
Set 2	0.86	0.88	0.84
Set 3	0.91	0.93	0.92

Table 2: Result for PSDG

In research paper "A NOVEL APPROACH FOR POSITIVE AND UNLABELED LEARNING BY LABEL PROPAGATION" published by "Shuangxun Ma, Ruisheng Zhang" label propagation algorithm is applied to train the final classifier. But before applying their label propagation algorithm, they have also enlarged positive set of instances

by extracting reliable positive instances. They have used KATZ index for measuring similarity between reviews.
They have used balance-scale and ecoli dataset from UCI repository.

Their best performance has been presented in below table with various iteration m.
Precision values are present here.

Dataset	M=1	M=2	M=3
Balance-scale	95.12%	97.39%	95.50%
Ecoli	87.75%	84.08%	85.00%

Table 3: Result for Label Propagation

In research work "hPSD: A Hybrid PU-Learning-Based Spammer Detection Model for Product Reviews" by "ZhiangWu , Member, IEEE, Jie Cao, Yaqiong Wang, Youquan Wang , Lu Zhang , and Junjie Wu" they proposed a hybrid semi supervised learning model for spammer detection to leverage both the users' characteristics and the user–product relations. They have injected positive opinions by various iterations. They have used movie dataset and used shilling injection method for their evaluation.

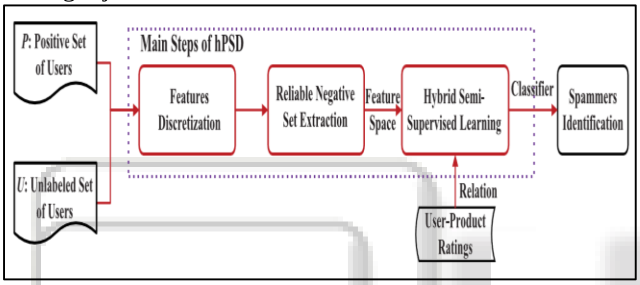

Fig. 2: Model for hPSD

They have also applied their approach to amazon dataset and the following table shows the result obtained from their method.

Dataset	Precision	Recall	F-Measure
Amazon	0.81	0.97	0.9

Table 4: Result for hPSD

In another research work "PUED: A Social Spammer Detection Method Based on PU Learning and Ensemble Learning" by "Yuqi Song, Min Gao (B), Junliang Yu , Wentao Li , Lulan Yu , and Xinyu Xiao" proposed a novel method only relying on normal users to detect spammers exactly. They evaluated two steps: 1) pick out reliable spammers from unlabelled dataset using voting classifier and 2) to train the random forest classifier.

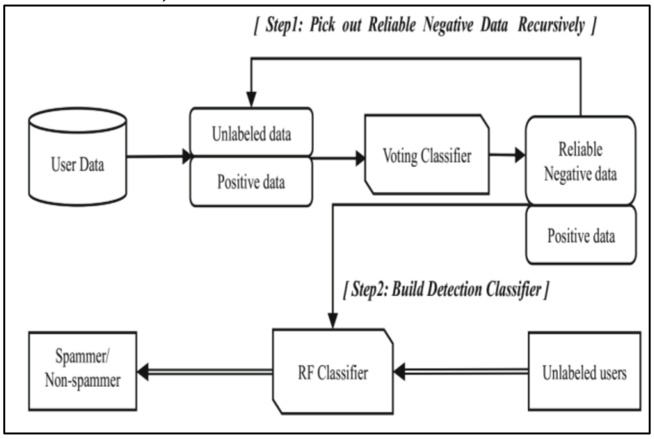

Fig. 3: Model for PUED

Below is the result for PUED methods.

Dataset	Precision	Recall	F-measure
Twitter	0.876	0.792	0.826
YouTube	0.862	0.668	0.76

Table 5: Result for PUED

III. Conclusion

As we have reviewed many papers and after research we have founded that deceptive review detection system is a data oriented problem and selection of features may play a crucial role for the problem. We have founded the PU Learning wih KNN give better result for the dataset of Hotel Services. Other research works reviewed by us also important with type of services and based on dataset they have used. For future work we plan to use more meta features and weighted KNN for more accurate classification work.

References

[1] J. K. Rout, A. Dalmia, K. R. Choo, S. Bakshi and S. K. Jena, "Revisiting Semi-Supervised Learning for Online Deceptive Review Detection," in IEEE Access, vol. 5, pp. 1319-1327, 2017. doi: 10.1109/ACCESS.2017.2655032

[2] L. Zhang, Z. Wu and J. Cao, "Detecting Spammer Groups From Product Reviews: A Partially Supervised Learning Model," in IEEE Access, vol. 6, pp. 2559-2568, 2018. doi: 10.1109/ACCESS.2017.2784370

[3] Shuangxun Ma and Ruisheng Zhang, "PU-LP: A novel approach for positive and unlabeled learning by label propagation," 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Hong Kong, 2017, pp. 537-542. doi: 10.1109/ICMEW.2017.8026296

[4] Z. Wu, J. Cao, Y. Wang, Y. Wang, L. Zhang and J. Wu, "hPSD: A Hybrid PU-Learning-Based Spammer Detection Model for Product Reviews," in IEEE Transactions on Cybernetics. doi: 10.1109/TCYB.2018.2877161

[5] Song, Yuqi&Gao, Min & Yu, Junliang& Li, Wentao& Yu, Lulan& Xiao, Xinyu. (2018). PUED: A Social Spammer Detection Method Based on PU Learning and Ensemble Learning: 13th International Conference, CollaborateCom 2017, Edinburgh, UK, December 11–13, 2017, Proceedings. 10.1007/978-3-030-00916-8_14

[6] Donato Hernández Fusilier, Manuel Montes-y-Gómez, Paolo Rosso, Rafael Guzmán Cabrera, Detecting positive and negative deceptive opinions using PU-learning, Information Processing &Management, Volume 51, Issue 4, 2015, Pages 433-443,ISSN 0306-4573, https://doi.org/10.1016/j.ipm.2014.11.001.

[7] Visani, Chirag& Jadeja, Navjyotsinh & Modi, Manali. (2017). A Study on Different Machine Learning Techniques for Spam Review Detection. 10.1109/ICECDS.2017.8389522.

[8] Taeho Jo, String Vector Based KNN for Text Categorization, ISBN 978-89-968650-9-4, ICACT201719 ~ 22, 2017

[9] Fang Lu and QingyuanBai, "A refined weighted K-Nearest Neighbors algorithm for text categorization," 2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering, Hangzhou, 2010, pp. 326-330. doi:10.1109/ISKE.2010.5680854

[10] Bruno Trstenjaka*,SasaMikacb, DzenanaDonkoc, "KNN with TF-IDF Based Framework for Text Categorization" doi: 10.1016/j.proeng.2014.03.129