International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 5, Issue 04, April-2019

Semi Supervised PU learning method for deceptive detection using MKNN

Priyanka Patani[1], Pinal Patel[2]

[1]*Computer Department, GEC Gandhinagar,*
[2] *Computer Department, GEC Gandhinagar,*

Abstract— Currently, in the era of E-Commerce a huge number of user opinion, user reviews and feedbacks are posted on the portal by thousands of users for any services and products. Reviews posted by the users are important source of information both end of the business. Mostly in online selling customers are rely on reviews before ordering the services online. Unfortunately, there is an problem of deceptive opinions, that is, not by real users. Deceptive reviews are aimed to improve the ratings of low quality products or services (Positive Reviews) or they are aimed to downgrade the high quality services of other businesses (negative reviews). For our research in this research paper we have focused on the identification of every type of deceptive reviews, either it is* positive or it is negative*. Because of the scarcity of samples of deceptive reviews, we propose to present the problem of the "detection of deceptive opinions employing PU-learning". PU-learning (Positive Unlabelled Learning) is a semi-supervised technique that is used to build a binary classifier. Mostly it builds classifier on the basis of positive (i.e., deceptive opinions) and unlabelled examples only. Concretely, we propose a novel approach for detection of deceptive spam using PU Learning and Modified KNN algorithm.

Keywords— PU Learning, KNN, MKNN, Spam Detection, Classification

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.

In this paper we have applied PU Learning method with Modified KNN Algorithm and that is discussed in our next section. As per our knowledge, the term PU Learning was described in research paper ECML-2005 paper[8]. It stands for positive and unlabelled learning. This method is also called learning from positive and unlabelled records. Their first paper regarding PU learning was published in ICML-2002. They focused on text classification in the same paper. This paper is organized in total VI sections. Section II contains related work, Section III contains the work detail for which we have got idea for improvement. Section IV contains our proposed methodology, Section V contains the result analysis and last Section VI has conclusion and future work.

II. RELATED WORK

In the paper: "Revisiting Semi-Supervised Learning for Online Deceptive Review Detection" [9] by "Jitendra Kumar Rout, Anmol Dalmia, Kim-Kwang Raymond Choo , Sambit Bakshi And Sanjay Kumar Jena", they explained the use of semi-supervised learning methods to find out spam opinions from fake users. They have applied their model on a data set of hotel users reviews.

They have used the following feature points for the testing and training from the dataset:

- Sentiment Polarity of Reviews
- Parts of Speech (POS) tags detection from the Reviews
- Linguistic Inquiry and Word Count (LIWC) to the Reviews
- Bigram frequency counts from the reviews of the dataset

For their future research they have mentioned that as their model is not tested over real word live reviews, further work may include implementation of system and evaluate their proposed model in the real-world, for that reviews to be collected directly from websites. They also mentioned that minimal meta-data are considered in their work during classification, They have mentioned better integrating of minimal meta-data in their future expansion section.

Lu Zhang, Zhiang Wu And Jie Cao in their research 'Detecting Spammer Groups from Product Reviews: A Partially Supervised Learning Model' [10] proposed a partially supervised learning model (PSGD) to find out spammer groups. They applied method of labelling some spammer groups as positive instances, In research work PSGD they applied Positive Unlabelled Learning (PU-Learning) for study a classifier as spammer group detector by using positive instances (labelled spammer groups) and unlabeled instances (unlabeled groups). They extract reliable negative set from the dataset in terms of the positive instances and the features by combining the positive instances of the dataset, extracted negative instances and unlabeled instances. They converted the PU-Learning problem into the semi-supervised learning problem, and then applied well known Naive Bayesian model and EM algorithm for training a classifier for spammer group detection. Their experiments on real-life Amazon.cn dataset they have founded that the their method PSGD is effective and better performer among spammer group detection methods. Shuangxun Ma, Ruisheng Zhang have worked on A Novel Approach For Positive And Unlabeled Learning By Label Propagation [11]. In their research work they proposed PU-LP, a graph-based PU learning algorithm. This method is based on similar label assumption. They have measured similarities between examples based on Katz index. The positive set is increased by extracting reliable positive instances from unlabeled set, and then the reliable negative example set is extracted. They also applied label propagation algorithm to train their final classifier.

III. EXISTING WORK

As mentioned in sector II, in research paper[1] they have used PU Learning method with k-NN as co-training algorithm. They have used 'goldstandard' dataset by Ottetal.[14],[50] for their evaluations. The dataset they used contains total 1600 text reviews from different 20 hotels situated in the Chicago area of United States of America, among them 800 reviews are deceptive reviews and 800 real user reviews. In the same dataset, 400 reviews are written with a negative sentimental polarity and 400 reviews with a positive sentimental polarity. They have obtained the same reviews from various sources. They have generated deceptive opinions using Amazon Mechanical Turk (AMT) for their evaluation purpose and the rest of reviews were obtained from so many online reviewing online services platform like Yelp, TripAdvisor.com, Expedia, and Hotels.com. For the research, the dataset was partitioned with fixed partition method of the One thousand six hundred instances in the corpus, two sets of instances were generated, the training data set and the testing data set. The partition ration of the corpus used in their research is 75% Testing and 25% Training Dataset, 80% Testing Data and 20% training data and 90% of whole dataset to testing and remaining10% to training by using the 4-fold, 5-fold and 10-fold partitioning schemes, respectively. They also used random sampling to choose the example in the set.

For their research work they applied four variations of classifiers. They used k-Nearest Neighbour classifier (k-NN), Logistic Regression classifier, RandomForest classifier and the Stochastic Gradient Descent classifier. For the k-NN

algorithm, the value of 'k' was set as 4. Also, for the Random Forest classifier, 100 worker instances were used for testing the result.

Algorithm 1 Co-Training Algorithm

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

OUTPUT: Deployable classifier, C.

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

After the evaluation, they have obtained best score 76.50% of accuracy and an F-Score of 0.775. The dataset was divided in a 75:25 partition ration for training and test dataset in the evaluation for which they got better result. From the training dataset, 20% of the instances were chosen as labelled and all others are as unlabelled. The k-NN classifier was used for the final evaluations, and the results are presented in Table 1.

TABLE 1
RESULT FOR EXISTING SYSTEM

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

IV. PROPOSED METHODOLOGY

Steps for Proposed Method

Step 1: Extract Reviews (TripAdvisor,GOIBIBO)

Step 2: PreProcess Reviews

Step 3: Extract Meta Data

Step 4: Weight the Reviews based on Meta Data

Step 5: Train with MKNN

Step 6: Apply PU Learning Approch

Step 7: Testing with Output Classifier Model

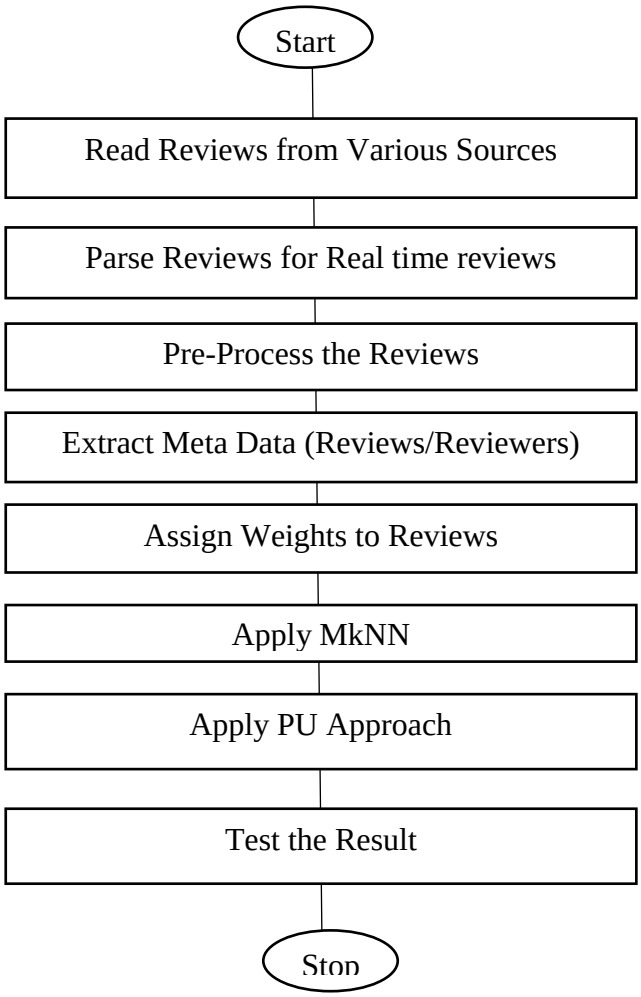


Figure 1
Proposed Flow Diagram

During the pre process step we have removed stop works and duplicate reviews from the dataset. For weight calculated we have extracted various services from the reviews. And applied a following formula to assign an weight.

Formula for calculating weight is like

W(r) = Sum(Total postive(i)* URF(i) + Total Neg (i)* URF(i))

Where URF is User Review for feature i

By using this formula we tried to assign weight based on overall opinion of particular feature. In Table 1 we have provided sample information related to various services extracted from the reviews and negative or positive impact of the particular service.

Table 2

Weight For Various Services In Reviews

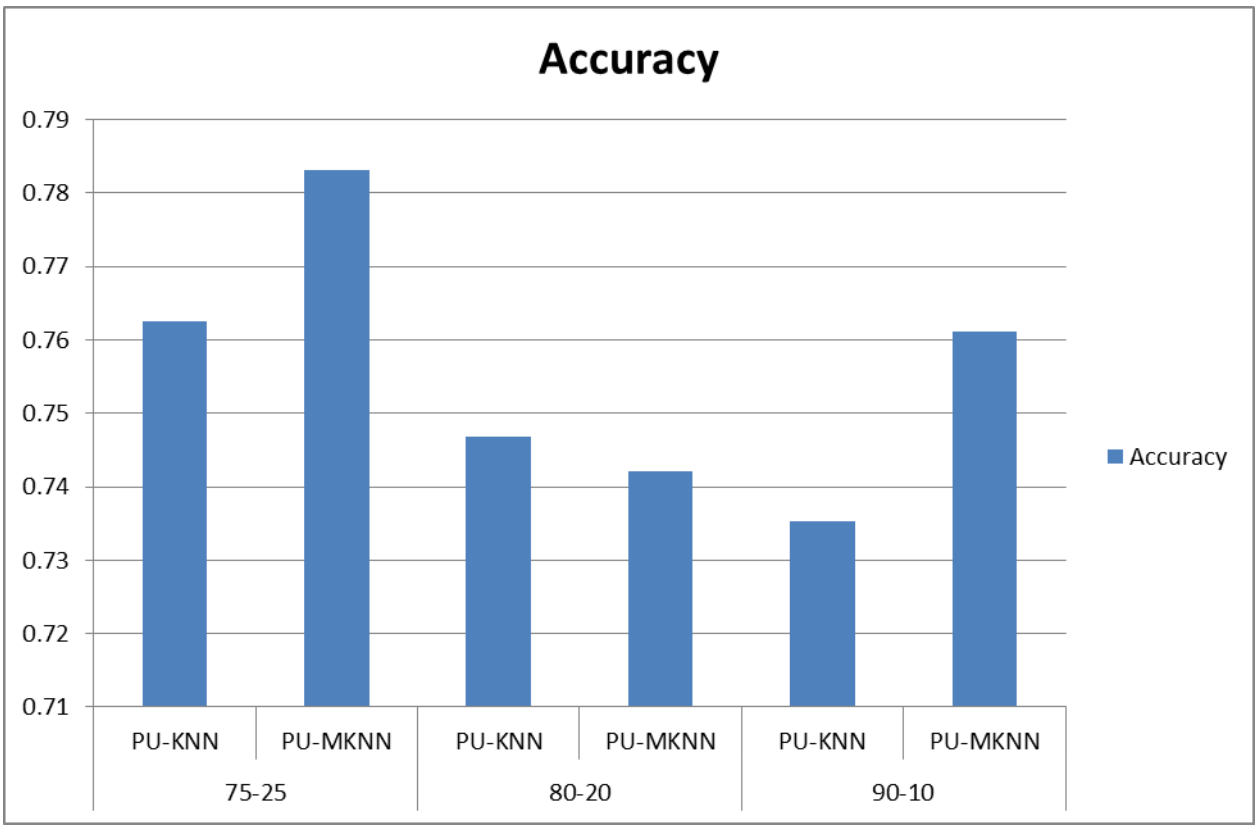
Feature	Pos	Neg	WeightP	WeightN
Wifi/ Internet Service	1	0	75	0
Location of the Hotel	0	0	0	0
Food	0	1	0	71.42857
Cost	1	0	50	0
Room Service	1	0	46.9697	0
			34.39394	**14.28571**

V. RESULT ANALYSIS

We have applied our model to various datasets as mentioned in Section IV. We have founded some significant improvement in accuracy. Table 3 contains the summary of the result over different datasets.

Dataset	Method	Accuracy
75-25	PU-KNN	0.7626
	PU-MKNN	**0.7832**
80-20	PU-KNN	0.7469
	PU-MKNN	0.7421
90-10	PU-KNN	0.7353
	PU-MKNN	**0.7612**

Graphs for Accuracy



VI. CONCLUSION AND FUTURE WORK

As too much work has been done with detection of Deceptive Reviews in various online services like Hotel or any E-Commerce related products, but still there is a large scope to work with the same domain (Deceptive Reviews Detection). As we have extracted some extra details from reviews and after applying modified k-NN with PU Learning approach to our work we have obtained more accuracy to the existing work. We have identified various deceptive reviews using some meta features. Still there is a scope to get better result by modifying weight calculation and combination of various classification algorithms.

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, ICACT2017 19 ~ 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