Opinion Spam Detection using Feature Selection

Conference Paper · November 2014

CITATIONS
7

READS
380

1 author:

Rinki Patel
Nirma University
2 PUBLICATIONS **21** CITATIONS

SEE PROFILE

OPINION SPAM DETECTION USING FEATURE SELECTION

Ms. Rinki Patel[1]

Department of Computer Science & Engineering
Institute of Technology, Nirma University
Ahmedabad, Gujarat, India
1rinkiparikh.patel@nirmauni.ac.in

Prof. Priyank Thakkar[2]

Department of Computer Science & Engineering
Institute of Technology, Nirma University
Ahmedabad, Gujarat, India
2priyankbthakkar@gmail.com

***Abstract*— In modern times, it has become very essential for e-commerce businesses to empower their end customers to write feedback or reviews about the products or services that they have utilized. There is also growing fad of appraising online services and products that the customer have experienced. Such reviews provide vital sources of information on these products or services. This information is utilized by the future potential customers before deciding on purchase of new products or services. These opinions or reviews are also exploited by marketers to find out the drawbacks of their own products or services and alternatively to find the vital information related to their competitor's products or services. This in turn allows to identify weaknesses or strengths of products. Unfortunately, this significant usefulness of opinions has also raised the problem for spam, which contains forged positive or spiteful negative opinions.**

This paper focuses on the detection of deceptive opinion spam. A recently proposed opinion spam detection method which is based on *n*-gram techniques is extended by means of feature selection and different representation of the opinions. The problem is modelled 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. All the experiments are carried out on widely used gold-standard dataset.

***Keywords*—** Opinion Spam Detection, Text Classification, Feature Selection

I. INTRODUCTION

In modern times, people use web for everything, they use web to solve their questions, to find solutions of unsolved problems, to know about not so known products or services etc. They also use web, to know opinions of others before finalizing their decision on purchase of a new product or service. Positive reviews about products generally results in a purchase of a product and vice-versa. This reveals that opinions influence decision making of individuals and organizations. However, the significant influence of opinions in decision making has also encouraged spammer and is also the reason behind the increasing number of opinion spams.

Positive opinions can result in significant financial gains and/or fame for business, organizations and individuals. Whereas negative opinions on some entities can damage their reputations. Deceptive opinions/fictitious reviews are purposefully written to sound authentic and victimize readers. The task of deceptive opinion spam detection can be modelled as binary classification problem with two classes, deceptive and truthful.

Many of the previous studies on detecting deceptive opinions were based on methods that seek for duplicate reviews [2]. Some other researchers have also used meta-information such as the IP address of the reviewer or the average rating of the product, rather than the actual content of the review [4].

A gold standard dataset consisting of 800 truthful and 800 deceptive hotel reviews was released by authors in [3] and [5]. However studies prior to this were not having access to any standard dataset and therefore their evaluations were based on some ad hoc procedures.
The paper focuses on modelling deceptive spam detection task as binary classification problem with deceptive and truthful as two classes. Impact of representation of the opinions in terms of different information retrieval models and feature selection is studied. Experiments are carried out on gold-standard dataset with naïve-Bayes and LS-SVM classifiers.

II. RELATED WORK

Text content, behavioral analysis and supervised methods are used by many researchers to address the problem opinion spam detection. Jindal and Liu had first attempted the study of spam detection and had given two methods for spam detection based on duplicate detection and spam classification [1]. Jindal and Liu, in another study, identified opinion spam by detecting exact text duplicates in an Amazon.com dataset. They found out three types of duplicate positive reviews that were used as a spam: (1) duplicates from different user id on the same product (2) duplicates from the same user id on different products and (3) duplicates from different user id on different products [2].

Authors in [3] proposed *n*-gram text categorization techniques to detect negative deceptive opinion spam with performance

far surpassing that of human judges. Similar techniques for detecting positive deceptive opinion spam are proposed in [5].

Some studies that tried to trick better features to improve classifier performance used sentiment scores, product brand, and reviewer's profile attributes to train classifiers [6]. Score computation based on behavioral heuristics, such as rating deviation is proposed in [7].The study reported in [8] focused on finding fraudulent reviewer groups by using frequent item set mining.

In [9], different stylistic, syntactical and lexical features describing opinions were identified. They used support vector machine to learn a classifier based on these features of the opinions.

III. PRE-PROCESSING

Gold-standard English dataset assembled by the authors in [5] is used in this study. The dataset consist of 800 positive and 800 negative reviews. Out of 800 positive reviews, 400 reviews are truthful while remaining reviews are deceptive. Same is the case with negative opinions. All the characters are converted to lower case and stop words are removed from each of the reviews. Each of the review is then defined as vector in multidimensional Euclidean space. The axes of this multidimensional Euclidean space are terms appearing in opinion collection. Three different representations namely Boolean, bag-of-words and TFIDF of these vectors are exercised in this study. In Boolean representation, these vectors are Boolean vectors and each element of the vector represents absence or presence of the corresponding term in the corresponding opinion. Each element of the vector, in case of bag-of-words representation is a natural number indicating how many times the corresponding term has appeared in the corresponding review. These vectors are real-valued vectors when represented in terms of TFIDF values of the terms.TF, IDF and TFIDF are defined in equations (1), (2) and (3) respectively [11].

$$TF = \log(1 + f_{ij}) \quad \dots (1)$$

$$IDF = f_{ij} * \log\left(\frac{\text{number of opinions}}{\text{number of opinions with word i}}\right) \quad \dots (2)$$

$$TFIDF = TF * IDF \quad \dots (3)$$

Where, f_{ij} is the frequency of word i in opinion j.

In addition to this, sequence of words approaches, such as, unigram, bigram and bigram+ are used for each of the representations. The total number of features exhibited by unigram, bigram and bigram+ profiles of these opinions are 9378, 82093 and 92054 respectively.

IV. PROPOSED APPROACH

The task of detecting deceptive opinion is modelled as binary classification problem in this study. The two classes considered are deceptive and truthful. Two popular machine learning techniques namely naïve-Bayes and LS-SVM are used to learn the classifier.

A. Naïve-Bayes Classifier

Class conditional independence is the naïve assumption of the naïve-Bayes classifier. Bayesian classifier predicts the probability of data belonging to a particular class given test data. It uses concept of Bayes' theorem to predict probability. A way of calculating the posterior probability P(C|X), from P(C), P(X|C) and P(X) is provided by Bayes' theorem. It states that

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad \dots (4)$$

Here, P(C|X) is the posterior probability which tells us the probability of hypothesis C being true given that event X has occurred. The probability of belonging to class deceptive/truthful is the hypothesis C and event X is the test data. A conditional probability of occurrence of event X given hypothesis C is true is defined as P(X|C). Training data is used to estimate it. The working of naive Bayesian classifier is summarized as follows.

Assume that, m classes C_1, C_2,...,C_m and event of occurrence of test data, X, is given. The test data is classified into a class with highest probability by Bayes' theorem (Equation (4)).

$$P(Ci|X) = \frac{P(X|Ci)P(Ci)}{P(X)} \quad \dots (5)$$

Given data sets with many attributes (A_1, A_2,...,A_n), computation of $P(X|C_j)$ is extremely computationally expensive. The naive assumption of class conditional independence is made in order to reduce computation in evaluating $P(X|C_j)$. This presumes that given the class label of the tuple (i.e. that there are no dependence relationships among the attributes), the values of the attributes are conditionally independent of one another, Therefore,

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i) \quad \dots (6)$$

$$= P(x_1|C_i) \times P(x_2|C_i) \times \dots \times P(x_n|C_i)$$

Here, value of attribute Ak for tuple X is denoted by x_k. Computation of $P(x_k|Ci)$ depends on whether it is categorical or continuous. $P(x_k|Ci)$ is the number of observations of class C_i in training set having the value x_k for A_k, divided by the number of observations of class C_i in the training set, if A_k is categorical. Gaussian distribution is fitted to the data, if A_k is

continuous-valued, and the value of $P(x_k|C_i)$ is calculated based on Equation (7).

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

So that,

$$P(x_k|C_i) = f\left(x, \mu_{C_i}, \mu_{C_i}\right) \quad \ldots(7)$$

Here, the mean (i.e., average) and standard deviation of the values of attribute A_k for training tuples of class C_i are denoted as μ_{C_i} and and σ_{C_i} respectively. In order to estimate, $P(x_k|Ci)$, these two quantities are then plugged into Equation (7) together with x_k,

Bayesian model based on term counts classifies the test data as follows.

Assume that there are m terms $t_1, t_2, \ldots, t_m$ (corresponding to the attributes in the opinion description) and n opinions $o_1, o_2, \ldots, o_n$ from class C. Assume that n_{ij} denotes the number of times that term t_i occurs in document d_j and $P(t_i|C)$ denotes the probability with which term t_i occurs in all documents from class C. The latter is estimated with the number of times that t_i occurs in all opinions from class C over the total number of terms in the opinions from class C.

$$P(t_i|C) = \frac{\sum_{j=1}^{n} n_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} n_{ij}} \qquad (8)$$

The multinominal distribution defines the probability of opinion o_j given class C as

$$P(o_j|C) = \left(\sum_{i=1}^{m} n_{ij}\right)! \prod_{i=1}^{m} \frac{P(t_i|C)^{n_{ij}}}{n_{ij}!} \qquad (9)$$

In order to predict the class label of X, $P(X|C_i)P(C_i)$ is evaluated for each class C_i. The class label of observation X is predicted as class C_i, if and only if

$$P(X|Ci)P(Ci) > P(X|Cj)P(Cj) \text{ for } 1 \le j \le m; j \ne i \quad (10)$$

B. Least Square – Support Vector Machin

Given a set of training examples which are linearly separable, $T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, in LS-SVM, learning is to solve the following constrained minimization problem,

$$\text{Minimize: } \frac{< w \cdot w >}{2} + \frac{C}{2}\sum_{i=1}^{n} \xi_i^2 \qquad (11)$$

$$\text{Subject to: } y_i(< w \cdot x_i > + b) = 1 - \xi_i, \quad i = 1, 2, \ldots, n \qquad (12)$$

Where $C \ge 0$ is a user defined parameter and ξ_i is a slack variable.

Solving the constrained minimization problem defined in Equations (11) and (12), solutions for w and b are obtained, which in turn give the maximum margin hyperplane $< w \cdot x_i > + b = 0$ with the margin $\frac{2}{\|w\|}$.

C. Feature Selection

As reported in the previous section, each of the opinion is represented by large number of features. All the features may not be important and useful in distinguishing whether the opinion is truthful or deceptive. Information Gain is used in this paper to measure the importance of the features. Details can be found in [12].

V. EXPRIMENTAL EVALUATION

This section discusses about evaluation measure, dataset and results.

A. Evaluation Measure

The assessment of the usefulness of the proposed method is carried out by way of the f-measure. F-measure is the harmonic mean of Precision (P) and Recall (R) values. It is more intuitive than the arithmetic mean when computing a mean of ratios. Computation of f-measure requires estimating Precision and Recall which are evaluated from True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). These parameters are defined in Equations (13), (14), (15) and (16).

$$\text{Precision}_{\text{positive}} = \frac{TP}{TP + FP} \quad \ldots(13)$$

$$\text{Precision}_{\text{negative}} = \frac{TN}{TN + FN} \quad \ldots(14)$$

$$\text{Recall}_{\text{positive}} = \frac{TP}{TP + FN} \quad \ldots(15)$$

$$\text{Recall}_{\text{negative}} = \frac{TN}{TN + FP} \quad \ldots(16)$$

Precision is the weighted average of precision positive and negative while Recall is the weighted average of recall positive and negative. Accuracy and f-measure are estimated using Equation (17) and (18) respectively.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad \ldots(13)$$

$$F - \text{measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \ldots(13)$$

B. Dataset

Gold-standard English dataset assembled by the authors in [5] is used in this study. The dataset consist of 800 positive and 800 negative reviews. Out of 800 positive reviews, 400 reviews are truthful while remaining reviews are deceptive. Same is the case with negative opinions

C. Exprerimental Results

Tables I, II and III show results for unigram, bigram and bigram+ sequence of words approaches respectively when all the features are used to learn the classifiers. For each of the sequence of words approaches, experiments are carried out for Boolean, bag-of-words and TFIDF representations. It can be seen that, in case of naïve-Bayes classifier, bag-of-words representation performs the best for all the sequence of words approaches. Impact of feature selection is depicted in Figures 1, 2 and 3.

Table I Unigram approach (total 9378 attributes)

Classifier	Representations								
	Boolean			Word Count			TFIDF		
	P	R	F	P	R	F	P	R	F
Naïve Bayes	86.34	85.12	85.73	86.0	85.88	**85.95**	74.81	71.9	73.36
SVM (LS)	79.11	78.81	78.96	78.5	78.25	78.41	79.60	79.1	79.36

Table II Bigram approach (total 82093 attributes)

Classifier	Representations								
	Boolean			Word Count			TFIDF		
	P	R	F	P	R	F	P	R	F
Naïve Bayes	82.0	74.88	78.29	85.1	82.3	**83.75**	76.69	76.05	76.37
SVM (LS)	82.7	78.75	80.68	83.4	79.6	81.51	82.87	79	80.89

Table III Bigram Plus approach (total 92054 attributes)

Classifier	Representations								
	Boolean			Word Count			TFIDF		
	P	R	F	P	R	F	P	R	F
Naïve Bayes	83.75	78.88	81.2	88.85	87.75	**88.29**	78.7	78.45	78.58
SVM (LS)	85.72	83.31	84.5	84.79	82.31	83.53	84.6	82.12	83.34

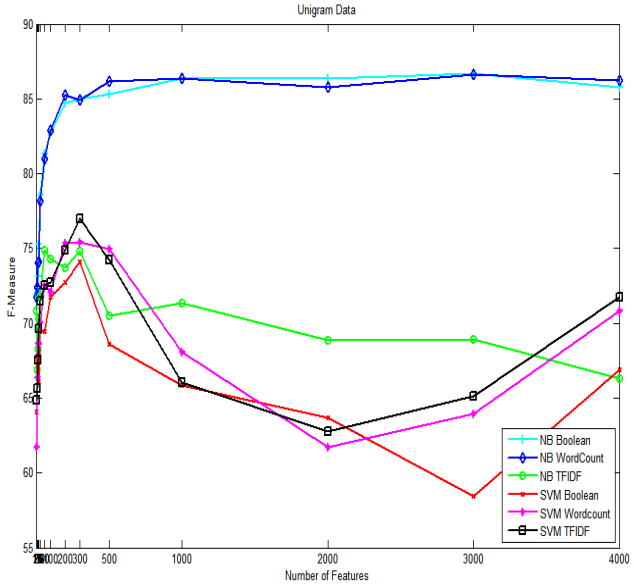

Fig. 1 Impact of feature selection - Unigram approach

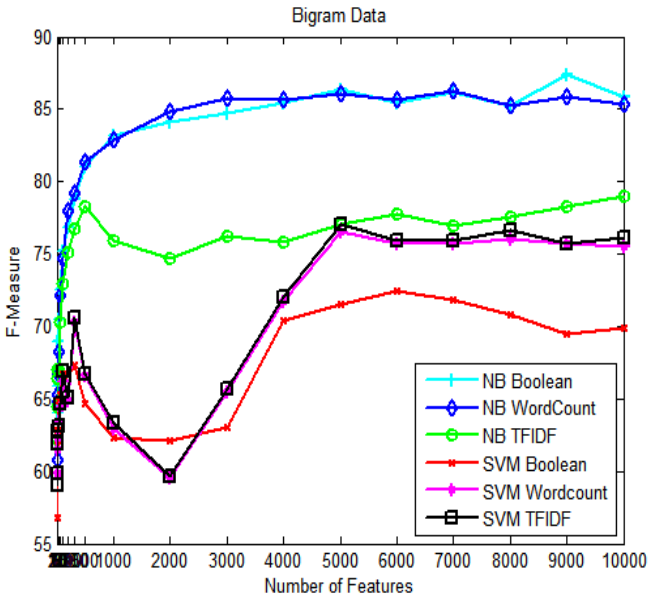

Fig. 2 Impact of feature selection - Bigram approach

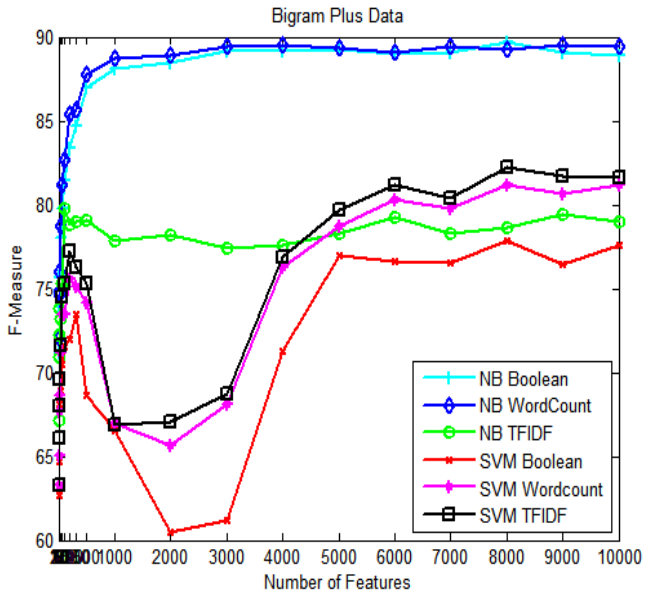

Fig. 3 Impact of feature selection - Bigram Plus approach

VI. CONCLUSIONS

The task of opinion spam detection is focused in this paper. The problem of opinion spam detection is modelled as the classification problem. Experiments are carried out with unigram, bigram and bigram+ sequence of words approaches.. For each of these approaches, opinions are modelled as Boolean, bag-of-words and TFIDF vectors. Naïve-Bayes and LS-SVM are used as the classification techniques. It is evident from the results that naïve-Bayes classifier with bag-of words representation of the opinions performs the best. Impact of feature selection is also studied in the paper. It is apparent form the result that learning a classifier using appropriate number of features improves the accuracy.

VII. FUTURE WORK

The use of more than certain percentage of adjectives, adverb, and missing certain key facts about products or services can be applied as feature to achieve more accuracy in detection of opinion spam. Also plan is to do improvement in detection of spam review using POS tagging for adjective, noun and verbs etc.

REFERENCES

[1] Nitin Jindal, Bing Liu ,"Review Spam Detection" , ACM Proceedings of the 16th international conference on World Wide Web, pp-1189-1190,2007.

[2] Nitin Jindal, Bing Liu," Opinion Spam and Analysis",ACM Proceedings of the international conference on Web search and web data mining, pp.219-229, 2008.

[3] MyleOtt, Yejin Choi, Claire Cardie, Jeffrey T. Hancock, "Finding deceptive opinion spam by any stretch of imagination" , ACM Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1 , pp.309-319,2011.

[4] SihongXie, Guan Wang, Shuyang Lin, Philip S. Yu "Review spam detection via time series pattern discovery", ACM Proceedings of the 21st international conference companion on World Wide Web,pp.635-636,2012.

[5] Ott, Myle, Claire Cardie, and Jeffrey T. Hancock. "Negative deceptive opinion spam." Proceedings of NAACL-HLT. 2013.

[6] Li, F.; Huang, M.; Yang, Y.; and Zhu, X. 2011. Learning to Identify Review Spam. In IJCAI.

[7] Lim, E.-P.; Nguyen, V.-A.; Jindal, N.; Liu, B.; and Lauw, H. W. 2010. Detecting product review spammers using rating behaviors.In CIKM, 939–948.

[8] Mukherjee, A.; Liu, B.; and Glance, N. S. 2012. Spotting fake reviewer groups in consumer revies. In WWW.

[9] Raymond Y. K. Lau, S. Y. Liao, Ron Chi-Wai Kwok, Kaiquan Xu, Yunqing Xia, Yuefeng Li," Text mining and probabilistic modeling for online review spam detection" ACM Transactions on Management Information Systems (TMIS) , Volume 2 Issue 4,Article 25,2011.

[10] B.Azhagusundari, Antony SelvadossThanamani "Feature Selection based on Information Gain", ISSN: 2278-3075,Volume-2, Issue-2, January 2013. In IJITEE

[11] Hall, Mark and Frank, Eibe and Holmes, Geoffrey and Pfahringer, Bernhard and Reutemann, Peter and Witten, Ian H, "The WEKA data mining software: an update", ACM SIGKDD Exploration Newsletter, 11(1), 10-18, 2009.

[12] Jiawei Han, Micheline Kamber, and Jian Pei, "Data Mining: Concepts and Techniques", Morgan Kaufmann, 2006.