Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 113 (2017) 273–279

Procedia Computer Science

www.elsevier.com/locate/procedia

The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017)

The Impact of applying Different Preprocessing Steps on Review Spam Detection

Wael Etaiwi* and Ghazi Naymat

Princess Sumaya University for Technology, Amman. Jordan

Abstract

Online reviews become a valuable source of information that indicate the overall opinion about products and services, which affect customer's decision to purchase a product or service. Since not all online reviews and comments are truthful, it is important to detect fake and poison reviews. Many machine learning techniques could be applied to detect spam reviews by extracting a useful features from review's text using Natural Language Processing (NLP). Many types of features could be used in this manor such as linguistic features, Word Count, n-gram feature sets and number of pronouns. In order to extract such features, many types of preprocessing steps could be performed before applying the classification method, this steps may include POS tagging, n-gram term frequencies, stemming, stop word and punctuation marks filtering, etc. this preprocessing steps may affect the overall accuracy of the review spam detection task. In this research, we will investigate the effects of preprocessing steps on the accuracy of reviews spam detection. Different machine learning algorithms will be applied such as Support Victor Machine (SVM) and Naïve Bayes (NB), and a labeled dataset of Hotels reviews will be analyze and process. The efficiency will be evaluated according to many evaluation measures such as: precision, recall and accuracy.

Peer-review under responsibility of the Conference Program Chairs.

Keywords: spam reviews, preprocessing, Bag-of-Words, feature selection, machine learning.

* Corresponding author. Tel.: +962795744288.
E-mail address: w.etaiwi@psut.edu.jo

Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2017.08.368

1. Introduction

Business owners might push reviewers to write good reviews about their products or services, or push them to write bad reviews about their competitor's products or services. These fake reviews are considered as spam reviews, and can have a great impact on the online marketplace. According to[1], spam review detection is a challenging task as there are no clear indications in the review text that the review is trustful or fake, in addition to the spam review text that usually looks perfectly normal. Therefore, Human efforts are required in order to distinguish good review among fake review manually, which is not practical and time consuming task. Some websites allow users to label reviews as helpful or not, which is still requires humans to judge reviews. Spam reviews were categorized into three main types, as proposed by Dixit et al.[2]: (1) Untruthful Reviews, which undermine the integrity of the online review system, (2) Reviews on Brands – where the reviews are only concerned with the brand or the seller of the product and fail to review the product itself, and (3) Non-Reviews – where reviews contain either unrelated text or advertisements. Feature extraction is the process of finding and constructing features from data or text. In our problem, features were categorized into two main types: features of reviews and features of reviewers. Information contained in the review itself are review's features, while features extracted from the reviewers and related to his personality and reviewing behavior are reviewer's features. Most of spam detection methods proposed in this topic used a combination of multiple features from the same type or from the two different types. Using a combination of features from two different types to train the classification (detection) method yield to better results as demonstrated by Jindal et. al.[3, 4]. Bag-of-words, POS tags, word count and term frequencies are some examples of review's features. While number of reviews, average review length, and percentage of reviewer's reviews are some examples of reviewer's features. Many approaches used to detect online spam reviews such as supervised learning and unsupervised learning techniques. In the supervised learning approach, the reviews are classified into two main categories (spam and not-spam) based on labeled training dataset that contains a labeled reviews. Many common algorithms used in this track such as: SVM, NB and decision tree. On the other hand, in order to overcome the problem of finding a labeled dataset to train the detection model, the unsupervised learning approaches used to detect spam reviews without the need for labeled dataset. Some similarity measure used in this approach such as cosine similarity measure. 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[5] such as using a nomination technique to list spam candidates, and complete the detection task by human experts.

In order to detect spam reviews from a given text, many standard Natural Language Processing (NLP) preprocessing steps could be applied in order to prepare the original reviews to be analysed. These prepossessing steps may affects the overall performance of the detection algorithm. A study of the effects of applying many different combination of standard NLP preprocessing steps on many different classification algorithms is proposed in this paper. A balanced dataset of hotels reviews used to evaluate the performance of the many different classification algorithms after applying preprocessing steps in case of spam reviews detection. The organization of this paper is as follow: Section two discusses the related works. Experiments discussed in section three. The last section presents the experimental results, discussions and conclusions.

2. Related Works

Ott et al.[6, 7] developed and compared three different approaches to detect deceptive opinion spam: standard text classification, psycholinguistic deception detection, and a problem of genre identification. Many features were used, such as: the frequency distribution of part-of-speech (POS) tags in a text, Linguistic Inquiry and Word Count (LIWC) features, and n-gram feature sets. NB and SVM algorithms were used to train the proposed model on truthful positive reviews for hotels found on TripAdvisor, data gathered for the same hotels using Amazon Mechanical Turk (AMT). The authors evaluated the detection approaches using a 5-fold nested Cross Validation (CV) procedure. The experimental results showed that the automatic classifiers outperformed human judges. Karam[8] used new features to detect spam reviews, he used many different types of linguistic features such as the number of pronouns, psychological features such as the affective processes, current concerns such as degree of leisure, spoken features such as degree of assent, and punctuation such as number of colons. The author evaluated the spam classification performance by considering more than 40 different classification algorithms on a spam review benchmark dataset, and the classification evaluation results overcame the others methods with more than 93%