Expert Systems With Applications 58 (2016) 83–92

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Detection of fake opinions using time series

Atefeh Heydari*, Mohammadali Tavakoli, Naomie Salim

Faculty of Computer Science and Information System, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia

ARTICLE INFO

Article history:
Received 2 September 2015
Revised 29 January 2016
Accepted 6 March 2016
Available online 24 March 2016

Keywords:
Review spam
Spam detection
Opinion spam
Fake reviews

ABSTRACT

Today's e-commerce is highly depended on increasingly growing online customers' reviews posted in opinion sharing websites. This fact, unfortunately, has tempted spammers to target opinion sharing websites in order to promote and demote products. To date, different types of opinion spam detection methods have been proposed in order to provide reliable resources for customers, manufacturers and researchers. However, supervised approaches suffer from imbalance data due to scarcity of spam reviews in datasets, rating deviation based filtering systems are easily cheated by smart spammers, and content based methods are very expensive and majority of them have not been tested on real data hitherto.

The aim of this paper is to propose a robust review spam detection system wherein the rating deviation, content based factors and activeness of reviewers are employed efficiently. To overcome the aforementioned drawbacks, all these factors are synthetically investigated in suspicious time intervals captured from time series of reviews by a pattern recognition technique. The proposed method could be a great asset in online spam filtering systems and could be used in data mining and knowledge discovery tasks as a standalone system to purify product review datasets. These systems can reap benefit from our method in terms of time efficiency and high accuracy. Empirical analyses on real dataset show that the proposed approach is able to successfully detect spam reviews. Comparison with two of the current common methods, indicates that our method is able to achieve higher detection accuracy (F-Score: 0.86) while removing the need for having specific fields of Meta data and reducing heavy computation required for investigation purposes.

1. Introduction

With the development of internet, people became more confident to explain their thoughts on websites and share them with millions of people (Heydari, Tavakoli, Ismail, & Salim, 2016). Web 2.0 slowly changed different aspects of people living. For instance, by creating online groceries, a huge number of daily trades are virtualized.

Nowadays people are more dependent to the internet for purchasing products and services. Long time ago, when they wanted to purchase a product, the best method was asking other customers who have purchased it before and know about the quality of that product very well to ensure that they will have a successful transaction.

Similarly, now they can visit customer reviews about various products or services that they tend to purchase via opinion sharing websites. Hence they can easily trade off the pros and cons of a specific good. The increasingly propensity of people to use online opinion sharing websites has created a challenging situation for manufacturers, business holders and stores (Peñalver-Martinez et al., 2014). Hence dishonest producers who tend to control and optimize the customers' opinions flow on their products and brand attempt to publish fake reviews among review websites. Sometimes they hire individual or in some cases groups of spammers to create not only glamorized positive reviews on their products but also harmful negative reviews on competitors'. These types of non-truthful reviews motivate customers to find their products the best option to purchase among similar products offered by other brands.

Fake opinions are extremely harmful not only for potential customers but also for business holders. Therefore, opinion mining techniques are assisting business to analyze posted customers' opinions on offered products to detect and filter spam reviews and proffer truthful reviews to purchasers (Savage, Zhang, Yu, Chou, & Wang, 2015). However research in this area is not adequate and many critical problems related to spam detection are not solved yet.

A bunch of previous approaches relied on content based factors to detect spam reviews (see Section 2). In an approach proposed by Lim, Nguyen, Jindal, Liu, and Lauw, (2010), for example, the

* Corresponding author. Tel.: +989212378119/+60197623624.
E-mail addresses: hatefeh2@live.utm.my, a_tav_ir@yahoo.com (A. Heydari), tmohammadali2@live.utm.my (M. Tavakoli), naomie@utm.my (N. Salim).
http://dx.doi.org/10.1016/j.eswa.2016.03.020
0957-4174/

product features mentioned in a review are compared with other reviews to identify duplicate reviews and filter them as spam. Although they are applicable on any type of reviews, content based methods naturally need expensive computations.

Other approaches focused on rating behaviors (Algur, Patil, Hiremath, & Shivashan, 2010) or/and other available Meta data of reviews (see Section 2). Majority of these methods require certain features that are included in a few number of datasets. However, most of these features, such as rating, author's ID, and helpfulness, could be manipulated perfectly by spammers to appear as real opinion.

Our approach differs significantly from former studies in several manners. Firstly, our method narrows down the selection of candidates for textual similarity investigation by constructing time series of reviews for each product and capturing only suspicious time intervals. This novation removes the need of expensive comparisons. Secondly, spam reviews generated by spammers who try to mislead customers without exhibiting any anomalous rating behavior are easily detectable by our method. This is because our approach does not merely focus on rating behaviors for detection of spam reviews. Finally, there are a few number of widely available fields of Meta data required in our method making it comprehensively applicable on different review websites and datasets.

The contribution of this paper is (1) To testing the suitability of using time series analysis approaches accompanied with a synthetic spam scoring system for detection of spam reviews and, consequently, developing a robust review spam detection system, (2) To reduce the need for expensive computations of detection phase by narrowing down the selection of samples.

The remainder of the paper is structured as follows: the next section discusses related work. Section 3 discusses the proposed method for detection of spam reviews. Experimental analysis are presented in Section 4. Finally, Section 5 presents our conclusions and future work.

2. Related work

In comparison with other types of spam such as e-mail spam (Wu, Feng, Wang, & Liang, 2015), web spam (Fdez-Glez et al., 2015), and SMS spam (Ahmed et al. 2015), detection of spam reviews is very nontrivial because manual evaluation of reviews and distinguishing fake reviews from real opinions is almost impossible (Jindal et al. 2008). Hence, state-of-the art methods in detecting various types of spam are not applicable in this domain. Accordingly, detection of spam reviews could be considered as one of the sophisticated problems in Natural Language Processing domain. A comprehensive review of state-of-the-art approaches in detection of spam reviews is provided in our previous research (Heydari, ali Tavakoli, Salim, & Heydari, 2015). These approaches can be broken down into the three categories of detecting group of spammers, detecting spammers, and detecting spam reviews:

(A) Detection techniques for group spammers

Some of the spam attacks are organized by group of spammers and a part of spam detection approaches have focused on detecting group of spammers, though the number of these approaches is limited. Mukherjee, Liu, and Glance, (2012),and Mukherjee, Liu, Wang, Glance, and Jindal (2011) defined diverse group spam indicators to detect group spammers such as rating deviation of members of a group of spammers, content similarity between group members, and number of products for which the group is creating spam reviews. By the construction of a relational model, the authors used the relationship between groups, individuals and products to score candidate groups. Similar relational models and features were used latter in Kolhe, Joshi, Jadhav, and Abhang (2014). Although both of the studies considered textual similarity of reviews as a spam sign, Ye and Akoglu, (2015) only used a graph-based measure to find statistical distortions caused by spamming activities and cluster the groups of spammers.

(B) Spammers detection techniques

Graph-based approaches consisting of graphs with review, reviewer, and store nodes (Akoglu, Chandy, & Faloutsos, 2013; Fayazbakhsh & Sinha, 2012; Wang, Xie, Liu, & Yu, 2011) focused mainly on using rating behaviors of reviewers to detect spammers. Rating deviation was one of the main features (Lim et al., 2010; Mukherjee et al., 2013; Sharma et al. 2013) or the only feature (Akoglu et al., 2013; Aye & Oo, 2014; Jindal, Liu, & Lim, 2010; Savage et al., 2015; Xue et al., 2015) used in detection of opinion spammers.

One of the indicators of the quality and fame of a product is its rank obtained from reviews. Thus, distortion of product's rank is one of spammers' main targets. However, with observation of online reviews, it could be seen that in some spam reviews, given rate is incompatible with the content. It shows that spammers are conscious about filtering technologies and try to pass through rating deviation-based filtering systems. They rate a product moderately, while trying to mislead customers by their words. We alleviate this problem in our approach by taking into account the context of reviews and activeness of a reviewer in every captured suspicious time interval. With this method, spam reviews of not only rating deviators, but also smarter spammers would be captured.

With the goal of detecting singleton spammers, the study carried out by Xie, Wang, Lin, and Yu (2012) was focused on reviewers' behaviors. A singleton reviewer is a reviewer who has written only one review. The authors assumed that reviewers' behaviors can be divided into two phases: arrival phase: when a customer purchase a product or a spammer get hired, and writing phase: when they start developing reviews. The authors analyzed spammers and customers' behaviors in normal arrival, promotion arrival and spam attack arrival. Accordingly, they found that spammers start writing phase immediately after arrival but customers have delay for receiving product and testing it. Therefore, the authors focused on nonstandard patterns in arrival phase to do their task. Consequently, in another method proposed in Fei et al. (2013) posting time of reviewers were used to detect spammers. The authors generated 5 new spammer behavioral features as indicators to be used in review spammer detection. Their method reveals more accurate results comparing to Xie et al., (2012). However, one of these 5 features is 'Ratio of Amazon verified purchase', a rarely available feature, which possibility of using this feature in any detection technique optimizes the accuracy of the method profoundly.

In order to develop a comprehensive detection system, utilized features should be general to enable the proposed system to work in disparate circumstances and on different datasets. The Amazon verified purchase sign indicates that the reviewer has really purchased the target product and the probability of being a spammer for him/her is almost zero. Although the work of Fei et al., (2013) was successful in detecting spammers, it could be used in a limited number of datasets. In contrast, our system uses effective features acquired from processing abundantly available data that could be found in almost all of the product review datasets. Moreover, our scope is to detect all types of spam reviews, including singleton, multiplton, advertisements, random texts, rank promoters, and fake reviews.