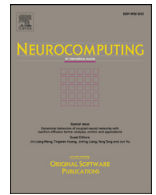




Contents lists available at ScienceDirect

## Neurocomputing

journal homepage: [www.elsevier.com/locate/neucom](http://www.elsevier.com/locate/neucom)

## Towards automatic filtering of fake reviews

Emerson F. Cardoso, Renato M. Silva, Tiago A. Almeida\*

Department of Computer Science, Federal University of São Carlos (UFSCar), Sorocaba, São Paulo, Brazil

## ARTICLE INFO

## Article history:

Received 24 November 2017

Revised 21 March 2018

Accepted 30 April 2018

Available online xxx

Communicated by Dr Fenza Giuseppe

## Keywords:

Fake review

Spam review

Natural language processing

Text categorization

Machine learning

## ABSTRACT

Online opinions significantly influence consumer purchase decisions. Unfortunately, this has led to a dramatic increase of fake (or spam) reviews that can damage the reputation of brands and artificially manipulate users' perceptions about products and companies. Despite the efforts of several studies on fake review detection, important questions still remain open. For instance, there is no consensus if the performance of the classification methods is affected when they are used in real-world scenarios that require online learning. Moreover, it is also not known if the performance of the methods decreases due to the time-ordered nature of the reviews. To answer these and other important open questions, this work presents a comprehensive analysis of content-based classification methods for fake review detection. The experiments were performed in multiple settings, employing different types of learning and datasets. A careful analysis of the results provided sufficient evidence to respond appropriately to the open questions, which can be used as a baseline for future studies.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

Expressing opinions is an inherent characteristic of humans. Nowadays, with the popularization of the Internet, opinions are increasingly easier to reach more and more people in different locations around the world. In fact, websites such as TripAdvisor and Yelp are used to share reviews mainly about services, places, and establishments. In the same way, e-commerce platforms, such as Amazon, also allow their users to post reviews about products and services. These online systems have popularized the sharing of reviews and increased people's trust on online opinions. They have also raised competitiveness and, as a consequence, some companies have unfortunately hired people to write fake reviews promoting their products and services or defaming their competitors.

Fake reviews are also known as deceptive opinions, spam opinions, or spam reviews, while their authors are called spammers. They can cause financial loss for product manufacturers and service providers because their brand reputation can be damaged by negative fake reviews. Moreover, companies can also lose customers when fake reviews promote their competitors.

Some examples of spam reviews are shown in Fig. 1. These opinions were written by a group of people in order to elevate the popularity of three specific products. They published 5-star reviews to the same products but on different dates.

Some recent cases of fraudulent use of reviews became public in the news. For example, a chef published fake negative reviews about rival restaurants on TripAdvisor, getting fired after his boss discovered the fraud on social media<sup>1</sup>. In another case, Samsung was fined for hiring spammers to post negative fake reviews about HTC smartphones<sup>2</sup>.

The spread of spam reviews is a serious problem on the Internet and they already represent a considerable volume of existing online reviews. Recently, Luca and Zervas [2] estimated that 16% of Yelp restaurant reviews are spam.

In order to reduce this problem, some social networks allow users to report suspicious reviews that might be spam. However, humans are rarely able to detect spam review accurately, since they are written to look authentic [3,4]. To illustrate the difficulty in discerning spam reviews, Ott et al. [5] offered the two real examples presented below. Just one is legitimate.

*I have stayed at many hotels traveling for both business and pleasure and I can honestly say that The James is tops. The service at the hotel is first class. The rooms are modern and very comfortable. The location is perfect within walking dis-*

\* Corresponding author.

E-mail addresses: [mersu898@gmail.com](mailto:mersu898@gmail.com) (E.F. Cardoso), [renatoms@dt.fee.unicamp.br](mailto:renatoms@dt.fee.unicamp.br) (R.M. Silva), [talmeida@ufscar.br](mailto:talmeida@ufscar.br) (T.A. Almeida).<sup>1</sup> Chef sacked after putting negative reviews about rivals on TripAdvisor. Available at <https://goo.gl/QcR3EZ>, accessed on Oct 16, 2017.<sup>2</sup> Samsung Fined For Paying People to Criticize HTC's Products. Available at <https://goo.gl/tmFwYk>, accessed on Oct 16, 2017.

<p>1 of 1 people found the following review helpful:            ★★★★★ <b>Practically FREE music</b>, December 4, 2004            This review is from: <a href="#">Audio Xtract (CD-ROM)</a>            I can't believe for \$10 (after rebate) I got a program that gets me free unlimited music. I was hoping it did half what was ....</p>	<p>2 of 2 people found the following review helpful:            ★★★★★ <b>Like a tape recorder...</b>, December 8, 2004            This review is from: <a href="#">Audio Xtract (CD-ROM)</a>            This software really rocks. I can set the program to record music all day long and just let it go. I come home and my ....</p>
<p>3 of 8 people found the following review helpful:            ★★★★★ <b>Yes – it really works</b>, December 4, 2004            This review is from: <a href="#">Audio Xtract Pro (CD-ROM)</a>            See my review for Audio Xtract - this PRO is even better. This is the solution I've been looking for. After buying iTunes, ....</p>	<p>3 of 10 people found the following review helpful:            ★★★★★ <b>This is even better than...</b>, December 8, 2004            This review is from: <a href="#">Audio Xtract Pro (CD-ROM)</a>            Let me tell you, this has to be one of the coolest products ever on the market. Record 8 internet radio stations at once, ....</p>
<p>5 of 5 people found the following review helpful:            ★★★★★ <b>My kids love it</b>, December 4, 2004            This review is from: <a href="#">Pond Aquarium 3D Deluxe Edition</a>            This was a bargain at \$20 - better than the other ones that have no above water scenes. My kids get a kick out of the ....</p>	<p>5 of 5 people found the following review helpful:            ★★★★★ <b>For the price you....</b>, December 8, 2004            This review is from: <a href="#">Pond Aquarium 3D Deluxe Edition</a>            This is one of the coolest screensavers I have ever seen, the fish move realistically, the environments look real, and the ....</p>

Fig. 1. Examples of spam reviews collected from Amazon. Source: Mukherjee et al. [1] (adapted).

*tance to all of the great sights and restaurants. Highly recommend to both business travellers and couples.*

*My husband and I stayed at the James Chicago Hotel for our anniversary. This place is fantastic! We knew as soon as we arrived we made the right choice! The rooms are BEAUTIFUL and the staff very attentive and wonderful!! The area of the hotel is great, since I love to shop I couldn't ask for more!! We will definitely be back to Chicago and we will for sure be back to the James Chicago.*

Several approaches for spam review detection are found in literature and most are based on supervised learning [6]. Despite the existing studies, a number of important questions still remain open. Specifically, there is no consensus if the text categorization methods are affected by:

- (i) the changes in the characteristics of reviews over time;
- (ii) the polarity of the reviews (compliments vs complaints);
- (iii) the use of real-world vs artificial reviews to train and evaluate the classifiers;
- (iv) processing reviews of various type of services or products at the same time;
- (v) scenarios that naturally require online learning.

Furthermore, there is no consensus on which of the evaluated content-based classification approaches is the best available choice to be used as a baseline for further comparisons.

To fill these important gaps, we conducted a comprehensive comparison of benchmark machine learning methods applied for content-based spam review detection. The experiments were performed in multiple settings and using different types of learning and datasets in order to answer the open questions and offer results for future baseline comparisons.

The remainder of this paper is organized as follows. Section 2 presents related work and open questions in spam review filtering. The experimental settings are detailed in Section 3. Section 4 presents all the results and analysis. Finally, we give our main conclusions and suggestions for future research in Section 5.

## 2. Related work and open questions

In this section, we briefly introduce several related work on spam detection and spam review detection. We also present open questions in spam review filtering.

### 2.1. Spam detection

In the past decades, machine learning methods have been applied to a wide range of problems, such as facial recognition [7], font recognition [8], speech recognition [9], diagnosis of diseases [10], and fraud detection [11]. In the last years, machine learning has also been explored to combat spam, a problem that is spreading to various online applications.

Spam detection has been extensively studied in several types of media, such as email [12], webpage [13–15], blogs [16], microblogs [17,18], SMS [19,20], and YouTube [20]. Many traditional machine learning-based methods have been employed, such as support vector machines (SVM) [14,16,17,19–21], naïve Bayes [14,16,20,22], decision trees (DT) [14,16,20,22–24], and k-nearest neighbors (KNN) [14,16,20]. In general, spam filtering approaches are based on the textual content and spam detection is seen as a binary text categorization problem where the categories are spam or ham (non-spam) [19,20,23,25].

Spam review is a different kind of spam. While spam on email, Twitter, blogs, and other media can be easily identified by an experienced user, spam on review imposes extra challenges, since even an experienced user is rarely able to detect it. As a consequence, few labeled datasets are available to train the classification methods [5]. Moreover, there is no consensus if the methods currently used to identify spam are effective to filter spam review.

### 2.2. Spam review detection

The problem of spam review detection has been studied more exhaustively over the last years. In general, the existing researches can be divided based on the data and features they analyze, as presented below [6,26].

1. *Content-based spam filtering*: studies that have proposed approaches based on textual content of the reviews [3,5,23,25,27–32].
2. *Behavior-based spam filtering*: studies that used information to improve identifying atypical behaviors of reviewers, such as the posting time and geolocation [25,30,33–35].
3. *Spam detection based on information about the product*: studies that have analyzed the information about each product, such as the sales volume, price, product description, and star rating of reviews [30,36].
4. *Spammer groups detection*: studies that focused on detecting groups of spammers [1,37].

The first studies analyzed millions of reviews about electronics collected from the Amazon website [27–29]. Jindal and Liu [27] found the following categories of spam: (i) reviews that analyze only specific brands, (ii) non-opinion reviews with unrelated

Download English Version:

<https://daneshyari.com/en/article/6863687>

Download Persian Version:

<https://daneshyari.com/article/6863687>

[Daneshyari.com](https://daneshyari.com)