International Journal of E-Business Research
Volume 14 • Issue 1 • January-March 2018

Review Spam Detection by Highlighting Potential Spammers and Diminishing Their Effect

Fatemeh Keshavarz, University of Calgary, Canada
Ayeshaa Abdul Waheed, University of Calgary, Canada
Btissam Rachdi, University of Calgary, Canada
Reda Alhajj, University of Calgary, Canada

ABSTRACT

Nowadays, millions of products and services are available to the public online. Therefore, searching for the best products which meets individuals' expectations would be difficult due to the existence of too many alternative choices. One of the most reliable approaches to choose a product or service is to exploit the experience of people who have already tried them, and are expected to have reported their almost honest opinions about them. A reviewing system is a place where individuals share their experience on products and services. Individuals may read and/or write their reviews which may be neutral and professional or biased. Moreover, companies utilize reviewing systems to apply opinion mining techniques in order to improve their goods or services and may be to watch their competitors. However, the popularity of reviewing systems ignites this motivation for some people to try to influence viewers by entering their fake reviews to promote some products or defame some others. These spam reviews should be detected and eliminated to prevent misleading potential customers and unethically affect the market. Opinion mining should be adapted to locate and eliminate potential spam reviews. In this paper, some review spam detection approaches have been proposed and examined over a sample dataset. The proposed approaches consider patterns that existed in trends of reviews, as well as reviewers' behavior. The approaches depend on various strategies such as observing abnormal trends, detecting uncommon or suspicious behaviors, investigating group activities, among others. The reported test results revealed some promising outcome.

KEYWORDS

Fake Reviews, Opinion Mining, Opinion Spam, Review Analysis, Review Spam, Text Analysis

INTRODUCTION

Almost immediately after the advent of the Internet and the Web technology, individuals as well as public and private organizations became interested in marketing their products and services on the Web. Since then, the utilization of the Web for selling products became rapidly pervasive such that nowadays it constitutes a considerable proportion of the Internet usage. Along with advertising on the Web and in addition to the independent sites that allow for product evaluation, vendors have given their audience the authority to evaluate and comment on the merchandise. Having this feature available and in most cases added to retailing websites, may be beneficial to both parties, i.e., producers and

DOI: 10.4018/IJEBR.2018010104

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

consumers. Companies can analyze the provided reviews to learn the strengths and weaknesses of their goods and services from consumers' points of view. They can further apply their conclusions in future to achieve higher levels of satisfaction from consumers, as well as to attract more people towards buying their merchandise. Moreover, they can investigate purchasers' reviews of products of competitor companies to become aware of the general appetite of consumers: their preferences, priorities, etc. Considering the admirable features of competing products in purchasers' opinions, producers can improve their own products by adding or magnifying those features. Further, by becoming aware of discreditable characteristics of products from rival companies, manufacturers will make sure not to strengthen those characteristics in their production or maybe to consider deficiencies before they are realized by customers. On the other hand, since the ability to review products online is made available for people, they can easily post their comments on their desired products at any time, from any part of the world, and at their own convenience. Consequently, a large number of opinions with a significant variety of ranking may be encountered. This also gives at almost no cost the opportunity to competitors to watch how their products have been received by consumers. Possessing sentiments of previous consumers of different interests and expectations, individuals can distinguish their most preferred goods or services faster and straighter than the time when such an integrated collection of opinions was not available. Therefore, presently, individuals rely extensively on reviews available online. This means they decide whether to buy products or not by analyzing existing opinions on those products. In fact, when a potential customer gets a positive overall impression of a product by considering its present sentiments, it is highly probable that he/she will act more in favor of purchasing the product. Normally if the percentage of positive opinions is considerable, it is likely that the overall impression will be highly positive. Likewise, if the overall impression is negative, it is less imaginable that vendees buy the product. Again, the overall negative impression can be the result of a great proportion of negative sentiments. The results of a survey conducted in early 2012 indicate that 51% of the customers have used Internet more than 6 times during a year while 72% of them have the same trust in online reviews as they have in personal recommendations (Anderson, 2012).

The systematic approach introduced to mine these vast collection of opinions is known as opinion mining aka sentiment analysis. There are two types of opinions to be identified and detected by opinion mining systems; direct opinions which are the feelings of the individuals about various single products, and comparative opinions in which the opinion holders compare two or more products according to their interests and priorities (Liu, 2010). In the case of direct opinions, opinion mining usually classifies the documents each containing a person's opinion into either positive or negative classes. Some sentiment analysts have tried to detect the orientations of opinions of the given comments by utilizing supervised (Chesley, Vincent, Xu, & Srihari, 2006; Pang, Lee, & Vaithyanathan, 2002; Yu & Hatzivassiloglou, 2003) or unsupervised learning (Dave, Lawrence, & Pennock, 2003; Turney, 2002). However even in the texts provided by opinion holders, not all the sentences contain opinions. Therefore, the sentiment analysis system needs to mine the opinionated sentences out of the given text. Opinionated sentences are the ones consisting of explicit or implicit opinions (Liu, 2010). As a result, some other researchers consider the problem of opinion mining at the sentence level, i.e., they examine each sentence first to assess if it is subjective (Riloff & Wiebe, 2003; Wiebe, Bruce, & O'Hara, 1999), and then to detect its polarity (Gamon, Aue, Corston-Oliver, & Ringger, 2005; Wilson, Wiebe, & Hoffmann, 2005; Yu & Hatzivassiloglou, 2003), even though, the sentence-level opinion mining has its own drawbacks. In the case of comparative opinions, the approach and the intention are different. The mining system needs first to detect the comparative sentences which usually contain some equative (e.g. as as), comparative, or superlative adjectives or adverbs (Jindal & Liu, 2006a). Then, it would be able to distinguish the preferred item among all (Ganapathibhotla & Liu, 2008; Jindal & Liu, 2006b).

The substantial importance of online reviews for a vast range of its users from companies to individual purchasers gives this stimulus to organizations to manipulate the overall polarity of the opinions for products. For example, a vendor might put effort to influence the potential customers

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/review-spam-detection-by-highlighting-potential-spammers-and-diminishing-their-effect/193030

Related Content

Reasons for Non-Engagement in Online Shopping: Evidence from the Philippines
Rex P. Bringula (2016). *International Journal of E-Business Research (pp. 17-30).*
www.irma-international.org/article/reasons-for-non-engagement-in-online-shopping/152316

Customer Decision Making in Web Services
Zhaohao Sun, Ping Zhangand Dong Dong (2012). *Handbook of Research on E-Business Standards and Protocols: Documents, Data and Advanced Web Technologies (pp. 210-232).*
www.irma-international.org/chapter/customer-decision-making-web-services/63472

Sourcing and Outsourcing Arithmetic
Tapen Sinhaand K. Subhadra (2009). *Electronic Business: Concepts, Methodologies, Tools, and Applications (pp. 1723-1738).*
www.irma-international.org/chapter/sourcing-outsourcing-arithmetic/9376

Adoption of Near Field Communication (NFC) for Mobile Payments in the UAE: A Merchants' Perspective
Mohanad Halawehand Hashem Al Qaisi (2016). *International Journal of E-Business Research (pp. 38-56).*
www.irma-international.org/article/adoption-of-near-field-communication-nfc-for-mobile-payments-in-the-uae/163362

Factors Affecting Continuance Intention of M-Government: An Empirical Study
Thamer Alshammari (2023). *International Journal of E-Business Research (pp. 1-23).*
www.irma-international.org/article/factors-affecting-continuance-intention-of-m-government/326550