INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Impact Factor: 8.165

Efficient Spam Reviews Detection System for Real-time Amazon Product Reviews using Machine Learning: A Review

Tanaya Prashant Korhalkar, Sarthak Santosh Bhandare, Swamini Prashant Khardekar,

Prof. Suparna Shirish Naik

Department of Computer Technology, Bharati Vidyapeeth's Jawaharlal Nehru Institute of Technology (Poly), Pune, Maharashtra, India

Department of Computer Technology, Bharati Vidyapeeth's Jawaharlal Nehru Institute of Technology (Poly), Pune, Maharashtra, India

Department of Computer Technology, Bharati Vidyapeeth's Jawaharlal Nehru Institute of Technology (Poly), Pune, Maharashtra, India

Professor, Department of Computer Technology, Bharati Vidyapeeth's Jawaharlal Nehru Institute of Technology (Poly), Pune, Maharashtra, India

ABSTRACT: generally the people trust on product on the basis of that product reviews and rating. Reviews can affect an organization or profile of a brand. The corporation has to assess market reactions towards its goods. However, it is not straightforward to track and organize popular reviews. Many public views are hard to manually process in social media. A methodology is then required to categories positive or negative public assessments automatically. Online feedback will provide customers with an insight into the consistency, efficiency and advice of the product; this provides prospective buyers with a better understanding of the product. One such unrealized opportunity is the usability of web assessments from suppliers in order to fulfil client requirements by evaluating beneficial feedback. Good and negative reviews play a major role in assessing customer needs and in quicker collection of product input from consumers. Sentiment Analysis is a computer study that extracts contextual data from the text. In this study a vast number of online mobile telephone ratings are analyzed. We classify the text as positive and negative, but we also included feelings of frustration, expectation, disgust, apprehension, happiness, regret, surprise and confidence. This delimited grouping of feedback helps to holistically assess the product, allowing buyers to decide better.

KEYWORDS: Machine Learning, Social Media, Text Mining, Text Classification, Sentiment Analysis, Online Reviews.

I. INTRODUCTION

1.1. Overview

Many businesses and software sectors store their data in Social networking creation provides the customer with the ability to share his or her views. That means the organization can't monitor the contents of the virtual universe now. Complaints on social media are submitted by customers who are not pleased with a company's services or goods. On the other hand, consumers are still optimistic about a commodity in social media. This view could affect other potential clients, including positive or negative ones. Potential consumers can find out about a certain product before deciding to purchase goods.

An appraisal of the sentiment is expected to immediately decide whether the feeling is negative or positive. Feeling analyses are a subset of text mining that focuses on the text of a person's feelings, mood, and attitude. The fundamental theory of sentiment analysis consists of categorizing the polarity of texts and determining whether they are positive or negative. Sentiment analyses are commonly used for rapid social network growth. For different places, public opinion is becoming critical. There have been some difficulties in collecting public examinations.

Many product evaluation pages have recently been published on the Internet. It invites scientists to carry out a consumer review sentiment analysis. On product evaluations, customer opinion was evaluated in this paper.

1.2. Open Issues

A lot of work has been done in this field thanks to its extensive use and applications. This section mentions some of the approaches that have been implemented to achieve the same purpose. These works are mainly differentiated from the techniques for sentiment analysis systems.

II. RELATED WORK

The pairwise features are first explicitly utilized to detect group colluders in online product review spam campaigns, which can reveal collusions in spam campaigns from a more fine-grained perspective.

In [1] paper, Spam campaigns spotted in popular product review websites (e.g., amazon. com) have attracted mounting attention from both industry and academia, where a group of online posters is hired to collaboratively craft deceptive reviews for some target products. The goal is to manipulate the perceived reputations of the targets for their best interests.

In [2] paper, Online product reviews have become an important source of user opinions.
Due to profit or fame, imposters have been writing deceptive or fake reviews to promote and/or demote some target products or services. Such imposters are called review spammers. In the past few years, several approaches have been proposed to deal with the problem. This work, take a different approach, which exploits the burrstones nature of reviews to identify review spammers.

In [3] paper, Online reviews on products and services can be very useful for customers, but they need to be protected from manipulation. So far, most studies have focused on analyzing online reviews from a single hosting site. How could one leverage information from multiple reviews hosting sites? This is the key question in our work. In response, develop a systematic methodology to merge, compare, and evaluate reviews from multiple hosting sites. focus on hotel reviews and use more than 15 million reviews from more than3.5million users spanning three prominent travel sites.

In [4] paper, Users increasingly rely on crowd-sourced information, such as reviews on Yelp and Amazon, and liked posts and ads on Facebook. This has the lento market for black hat promotion techniques via fake (e.g., Sybil) and compromised accounts, and collusion networks. Existing approaches to detect such behavior relies mostly on supervised (or semi-supervised) learning over known (or hypothesized) attacks. They are unable to detect attacks missed by the operator while labeling, or when the attacker changes strategy.

In [5] paper, Online reviews have become an increasingly important resource for decision-making and product designing. But review systems are often targeted by opinion spamming. Although fake review detection has been studied by researchers for years using supervised learning, the ground truth of large-scale datasets is still unavailable and most of the existing approaches of supervised learning are based on pseudo-fake reviews rather than real fake reviews. Working with Dianping1, the largest Chinese review hosting site, present the first reported work on fake review detection in Chinese with filtered reviews from Damping's fake review detection system.

In [6] paper, Online reviews are quickly becoming one of the most important sources of information for consumers on various products and services. With their increased importance, there exists an increased opportunity for spammers or unethical business owners to create false reviews in order to artificially promote their goods and services or smear those of their competitors. In response to this growing problem, there have been many studies on the most effective ways of detecting review spam using various machine learning algorithms. One common thread in most of these studies is the conversion of reviews to word vectors, which can potentially result in hundreds of thousands of features.

In [7] paper, it providing an efficient and effective method to identify review spammers by incorporating social relations based on two assumptions that people are more likely to consider reviews from those connected with them as trustworthy, and review spammers are less likely to maintain a large relationship network with normal users. The contributions of this paper are two-fold: (1) elaborate how social relationships can be incorporated into review rating prediction and propose a trust based rating prediction model using proximity as trust weight; and (2) design a trust-aware detection model based on rating variance which iteratively calculates user-specific overall trustworthiness scores as the indicator for spam city.

In [8] paper, to detect fake reviews for a product by using the text and rating property from a review. In short, the proposed system (ICF++) will measure the honesty value of a review, the trustiness value of the reviewers and the reliability value of a product. The honesty value of a review will be measured by utilizing the text mining and opinion mining techniques. The result from the experiment shows that the proposed system has a better accuracy compared with the result from iterative computation framework (ICF) method.

In the [9] paper, Online Social Networks (OSNs), which capture the structure and dynamics of person-to-person and person-to-technology interaction, are being used for various purposes such as business, education, telemarketing, medicine, and entertainment. This technology also opens the door to unlawful activities. Detecting anomalies, in this new perspective of social life that articulates and reflects the off-line relationships, is an important factor as they could be a sign of a significant problem or carrying useful information for the analyzer.

In [10] paper, they propose a new holistic approach called SpEagle that utilizes clues from all metadata (text, timestamp, and rating) as well as relational data (network), and harness them collectively under a unified system to spot suspicious users and reviews, as well as products targeted by spam. SpEagle employs a review-network-based classification task which accepts prior knowledge on the class distribution of the nodes, estimated from metadata. Positive points are: It enables seamless integration of labeled data when available. It is extremely efficient.

In [11] this paper, mangoes are graded in four types   like Green Mango, Yellow Mango and Red Mango which are based on machine learning method. This system considers RGB values size and shape of mangoes. Following analysis is used to obtain good probability. This helps to train system to identify appropriate maturity of mangoes. This research is conducted on two machine learning method i.e. Naive Byes and SVM (Support Vector Machine).

III. PROPOSED SYSTEM

A new proposed framework consists in representing a set of review data provided as FIN (Fake   Information Networks) and solving the issue of spam detection in a problem of FIN classification. In particular, to show the reviews data set as a FIN where the reviews are linked through different types of nodes (such as functionality and users). Then a weighting algorithm is used to calculate the importance (or weight) of each function. These weights are used to calculate the latest review labels using supervised and unsupervised procedures. Based on our observations, defining two views for features (review-user and behavioral-linguistic), the classified features as review behavioral have more weight and yield better performance on spotting spam reviews in both semi-supervised and unsupervised approaches. The feature weights can be added or removed for labeling and hence time complexity can be scaled for a specific level of accuracy. Categorizing features into four major categories (review-behavioral, user-behavioral, review-linguistic, user-linguistic), helps us to understand how much each category of features contributes to spam detection.

IV. CONCLUSION

Sentiment Analysis is a case study that looks at the feeling, mood, entropy or feelings of people. This paper addresses a basic issue of the study of feelings and the classification of feelings of polarity. Data was compiled from online product reviews of Amazon.com. A method known as the categorization of emotion polarity and POS along with thorough explanations of each phase was proposed. These measures include pre-processing, pre-filtering, partitioning, data consistency. Functionality that include machine learning expertise. Much work has been done in opinion mining and consumer evaluation in the form of a study of documents, sentences, and features. Opinion Mining can become a most interesting field of study for potential preferences by using a number of found function expressions derived from the reviews. More novel and successful approaches need to be invented to address the existing difficulties of mining opinion and sentiment analysis.

REFERENCES

[1]. Ch. Xu and J. Zhang," Combating product review spam campaigns via multiple heterogeneous pairwise features", In SIAM International Conference on Data Mining, 2014.

[2]. G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Exploiting bustiness in reviews for review spammer detection", In ICWSM, 2013.

[3]. A. j. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos, "True view: Harnessing the power of multiple review sites",In ACM WWW, 2015.

[4]. B. Viswanath, M. Ahmad Bashir, M. Crovella, S. Guah, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, "Towards detecting anomalous user behavior in online social networks", In USENIX, 2014.

[5]. H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao," Spotting fake reviews via collective PU learning", In ICDM, 2014.

[6]. M. Crawford, T. M. Khoshgoftaar, and J. D. Prusa," Reducing Feature Set Explosion to Faciliate Real-World Review Sapm Detection", In Proceeding of 29th International Florida Artificial Intelligence Research Society Conference, 2016.

[7]. H. Xue, F. Li, H. Seo, and R. Pluretti," Trust-Aware Review Spam Detection",IEEE Trustcom/ISPA.,2015.

[8]. E. D. Wahyuni , A. Djunaidy," Fake Review Detection From a Product Review Using Modified Method of Iterative Computation Framework", In Proceeding MATEC Web of Conferences, 2016.

[9]. R. Hassanzadeh," Anomaly Detection in Online Social Networks: Using Datamining Techniques and Fuzzy Logic", Queensland University of Technology, Nov, 2014.

[10] R. Shebuti, L. Akoglu," Collective opinion spam detection: bridging review networks and metadata", In ACM KDD, 2015.

[11] G.D. Upadhye, D.Pise,"Grading of Harvested Mangoes Quality and Maturity Based on Machine Learning Techniques",IEEE International conference on smart city and Emerging Technology,2018.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING