IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016

FakeTech: Identifying fake reviews using Collective-Positive Unlabeled Learning

Savita K Shetty[1], Kanchana S Kokatanur[2]

Dept. of Information Science and Engineering, MSRIT, Bangalore, Karnataka, India[1,2]

Abstract: The advancement of technology with the Internet has generated plentiful of user-generated data. This content is used to give knowledgeable information using different data mining techniques. Among various types of generated data reviews about product, business or services are becoming more important. Now a days online review is often the primary factor and a valuable source of information in aiding customer's purchase or service decisions. The vitality of the peer reviews has attracted spammers to induct fake and unrealistic reviews. Some online review systems are facilitating interactions between customers to improve its utility and experiences by expressing product or service opinions. Due to the large public opinion generated, directly or in-directly affecting the marketing of the products or service has incepted the manufacturer's interest on online reviews. Observing the reliability of customers on reviews some vendors are trying to Fake It! thus misleading the customers. Despite aware of manipulated reviews, customer is unable to distinguish the fake once from the genuine review which necessitates building a system that filters reviews. In this paper, we approach a dual layer classification based on two -level filtering method. The intent is achieved by splitting into two levels, at first by using metadata followed by review content analysis In the first level of classification, we will consider the metadata parameters (IP address, time)to decide the truthfulness of the review. Next, Auto learning system is built which learns from past history of the user. The real reviews classified may still contain some suspicious reviews which calls second level of classification technique using review content features and reviewer centric features to detect review spam. In both the levels auto learning system is built which learns from past history of the user in the system which reduces the computational time when new data is fed into the system. A comparative study is carried out where our built model showed high performance than other techniques.

Keywords: primary factor and a valuable source, reliability of customers.

INTRODUCTION

Huge amount of data is being generated and made available in information Industry. This raw data is of no use until it is converted into some meaningful information to extract useful knowledge using data mining techniques. Data mining is a process that takes data as input and outputs knowledge [11]. The process involves analyzing data from different perspective and transforming the data into useful information – that can be used in many applications such as Market analysis, Production Control, Fraud Detection, Science Exploration, etc. The proposed Fake Review Detection Technique is designed and developed to identify fake online reviews which help both vendors and customers in their business and purchase decision.

Traditionally, humans are always influenced by oral communication as a means of passing information to their peers as well as successive generations. According to the survey, word of mouth is a primary factor in marketing products or services. Ever since e-commerce trade has come into light it is intensely being used by customers or vendors to purchase or sell utilities. When it comes to e-trade customer reviews and opinions expressed in public domains have replaced the means of passing information thus aiding customer a more convenient way for making his decisions. Having found the richness of the reviews vendors, online retailers and service providers have come up with feedback forms for customers who want to express their exceptionally good or never forgettable bad experiences for the products or services bought. Many people blindly rely on reviews before placing their orders becoming prey for fraudsters. Furthermore manufacturers may fake reviews by providing incentives to whoever writes good reviews about their products or services, or might pay someone to write fake reviews about their competitor's merchandise.

Thus vitality of the information has given a room for manipulating the reviews in their own interest and ultimately scapegoats in all these spamming activities are none other than customers. Spammers are imposters of their own opinions in favors of their incentive providers either by promoting their own goods or demoting their competitor's goods or targeted products. The authenticity of these spammer's comments are very hard to be distinguished by just reading it manually because these comments are appealing and tends to be genuine. In this paper, we propose a two level classification method called "Fake Review Detection method" to detect fake reviews. The method is built based on considering metadata parameters, review and reviewer centric features.

RELATED WORK

The study of spam detection is spread across different parameters each work shows unique ways of utilizing

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016

parameters such as metadata (IP address, time, browser ID), review content, product ID and so on. Each approach filters and refines reviews to the utmost realistic ones. In [1], Nitin Jindal & Bing Liu using shingle method proposed Review Spam Discovery Component in which duplicate reviews were concentrated followed by classification of spam or non-spam. Duplicate detection used Shingle method where near similar posts were considered as a spam with similarity score greater than threshold assumed (score>0.9). To further more classify spam or non-spam 2-class classification model is built using machine learning. The study done by [2] Nitin Jindal & Bing Liu in 2008 introduced new technique for opinion spam and analysis. In 1997, shingle schema was presented by A.z. BRODER for assessing closeness and regulation in the two reports by enrolling similarity score. By then that thinking was related on two sentences to check closeness between them. First the duplicate review were removed and then using supervised learning reviews brand and non-reviews were classified. In 2010[3], Ee-Peng Lim, Viet-A Nguyen, Nitin Jindal, Bing Liu , Hady W. Lauw concentrated on the behavioral approach to detect the review spammers who attempt to manipulate review ratings on particular products or product companies. They derived aggregated scoring methods to rank reviewers and according to the measure they displayed spamming behaviors.

In their study the importance was given to several trademark practices of review spammers and models these practices to perceive the spammers. In year 2011[4] Wang, Guan, Sihong Xie, Bing Liu, and Philip S. Yu proposed review graph concept to capture relationships between reviewers all reviews, and stores where in reviewers have reviewed as a heterogeneous graph. Their work concentrated on how interactions between nodes present in the network graph can be used to reveal the source of spam [9]. In 2012[5], Arjun Mukherjee, Bing Liu and Natalie Glance contributed to the area of fake review social occasions using behavioral model and Frequent item set mining system. In 2013[6], Arjun Mukharjee, Abhinav Kumar, Bing Liu helped in the area of group review spam. They built a principled methodology to capture escapade viewed reviewing practices to identify thought spammers (fake observers) in an unsupervised Bayesian finding construction. Fangtao Li et al., 2011[7] worked on recognizing review spam using regulated learning techniques and break down the impact of different mechanisms in study spam unmistakable affirmation.

METHODOLOGIES

1. Metadata based classification

This type of detection method uses metadata (networking parameters) such as IP addresses [10] and review time as the base for the review classification. The reviews coming from same IP address within a specified time window crosses the threshold value then that review is labeled as Fake. The threshold can be changed accordingly so as system can work with different types of data set.Auto learning system introduced will learn from past activity of the user and is used when new dataset arrives to decide the authenticity of the user.

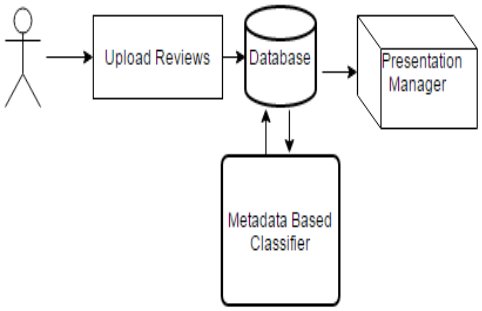

Figure 1: Architecture of Metadata Classification method

Advantages

- Reviews generated by automated systems are easily caught uniquely featured (IP address).
- Repetition of classification techniques helps to detect fake reviews proactively by auto learning.

Disadvantages

- Residues of the fake reviews can be found as system doesn't dig into the content of review.

2. Review and Reviewer Centric Classification

Second type of classification is carried out based on the content of the reviews and the reviewer. Review written by the user, gives a lot of information regarding review being spam/fake.

This study is carried out by carrying out feature engineering for fake review detection. The features considered in this method are (8): Maximum number of reviews, percentage of positive review, maximum content similarity. It is observed that spammers post more than 5 reviews on any particular product.

Keeping this into consideration, we have review count system which will calculate the number of reviews given by user. It is observed that spammers write more positive reviews about any particular service/product. So the reviews with high positive percentage will be considered as un-trustworthy reviews.

Similarity between reviews also gives strong indication about review being fake. So the review content similarity is calculated by using n-gram technique (12).Bi-gram and tri-grams are considered to check the review similarity. Bi-gram feature finds out if particular bigram word is present in a review.

Similarly tri-gram feature finds out if particular word trigram is present in a review and using both, bi-gram and tri-gram similarity is calculated. System calculates total contribution of all three features to identify truthfulness of the review. Threshold value is maintained crossing which will be considered as spam/untruthful review.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016

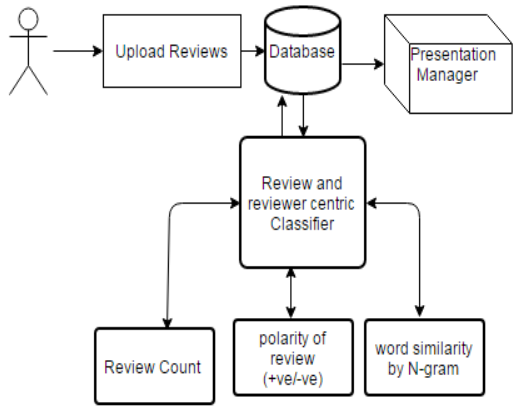

Figure 2: Architecture of review and reviewer centric method

Advantages

- Vulnerability of fake reviews is handled more effectively by taking rich information from the review content.

Disadvantages

- Ignoring networking parameters may confine the strengths of classification techniques.

3. Fake Review Detection Method

This method is based on the idea of above explained techniques. Initially, the review classification is carried out based on the networking parameters. Since real reviews classified may still contain some suspicious reviews which calls second level of classification that uses review and reviewer centric classification method (Type2).

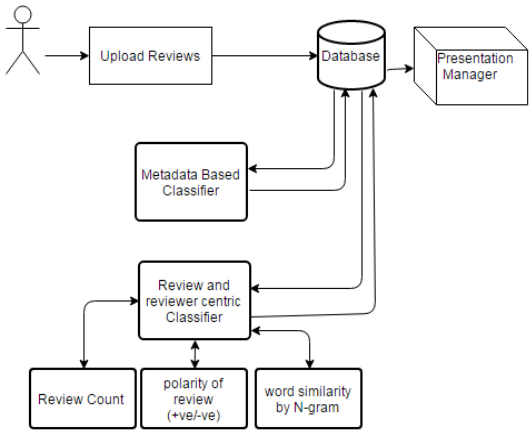

Figure 3: Overall system Architecture

Advantages

- The robustness of the system is enhanced by the combination of networking parameters and along with their contents.
- The classification of fake reviews at two levels yields a better promising result.

Implementation

The structure of dataset considered for review classification has the following components <Sl.no, user-name, review, IP-address, Date, time, Product-name>

Classification algorithm for Fake Review Detection method:

Level 1:

Input: Let R = R_1,R_2,R_3,....,Rnbe the set of reviews given by different users.

Output: Review Ri€ Table I –fake or Table II − real.

Step 1: Search repository for the presence of IP address and user id of input dataset in trained data.

Step 2: Maintain two separate tables If(R_1(ip) == train(ip)) Insert in Table I.

Step 3: Repeat step 1 and step 2 for each review.

Step 4: For remaining reviews,calculate IP count within specified time interval.

If(count(ip) >= IP_Time_Threshold

Insert into Table I.

Else

Insert into Table II.

Step 5: Repeat step 4 for each input review.

Step 6: STOP.

Level2:

Input: Table II reviews represented as R = R_1,R_2,R_3,....,Rn with unlabeled reviews.

Output: Review Ri €{ fake –review or real – review}

Step 1: Calculate count percentage for user.

Step 2: Calculate positive percentage of review for user.

Step 3: Calculate content similarity of a text.

Step 4: Sum up the percentage from Step 1, 2 3.

If total percentage threshold

Assign label as fake

else

Assign label as real

Step 5: Repeat steps 1 - 4 for all users

Step 6: STOP.

EXPERIMENTAL SETUP

Dataset creation for the system is manually carried out based on the considered components.Data base pooling is used so as to run all methods independent of each other. Dataset was fed to all the three methods at different levels of classification yielding different sets of data. The results of the classifiers were stored in separately maintained database. Dataset and results are stored in the native database. Dynamic graphs are generated using jfree chart. Below figure 4 shows the comparative results of three methods.

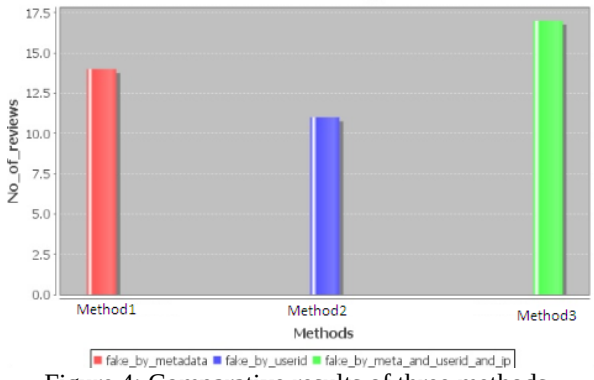

Figure 4: Comparative results of three methods.

CONCLUSION AND FUTURE WORK

Fake review detection is carried out in two different levels taking networking parameters in first level and review content along with the reviewer at second level. At first level will remove suspicious user based on the IP address and review time. Classification method is extended to one more level so as to remove any residual fake reviews left by first method. Combination of both the technique provides very promising results in identifying fake reviews.

In the future work the recommender system can be developed for the products/services which get more positive/real reviews. The system can keep track of number of positive reviews the product/service has got over a time and suggests user based on their search request.

ACKNOWLEDGEMENT

I would like to thank **Dr. Vijaya Kumar B P**, Head of Department of Information Science Engineering, MSRIT and **Mrs. Savitha K. Shetty**, Assistant Professor, MSRIT for their valuable time.

REFERENCES

[1] N. Jindal and B. Liu, "Review Spam Detection", in Proceedings of WWW-2007 (poster paper), May 2007.

[2] Jindal, N. and Liu, B. Opinion spam and analysis. Proceedings of the 2008 WSDM, 2008, pp. 219-229.

[3] E. P. Lim, V. A. Nguyen, N. Jindal, B. Liu and H. Lauw, "Detecting Product Review Spammers using Rating Behaviors," in Proceedingsof the 19th ACM International Conference on Information andKnowledge Management (CIKM-2010, full paper), Oct 2010.

[4] G. Wang, S. Xie, B. Liu, P. S. Yu, "Identify Online Store Review Spammers via Social Review Graph," ACM Transactions onIntelligent Systems and Technology, accepted for publication, 2011.

[5] A. Mukherjee, B. Liu, and N. Glance, "Spotting Fake Reviewer Groups in Consumer Reviews," International World Wide WebConference (WWW-12), April 2012.

[6] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, "Spotting Opinion Spammers using Behavioral Footprints," in Proceedings of SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-13), Aug 2013.

[7] Li, Fangtao, M. Huang, Y. Yang, and X. Zhu, "Learning to Identify Review Spam," in Proceedings of the International Joint Conferenceon Artificial Intelligence (IJCAI-11), 2011.

[8] S. Dixit and A. J. Agrawal, "Survey on review spam detection," International Journal of Computer & Communication Technology, Volume-4, 2013.

[9] G. Wang, S. Xie, B. Liu, P. S. Yu, "Review Graph based Online Store Review Spammer Detection," ICDM-11, 2011.

[10] Li, Huayi and Chen, Zhiyuan and Liu, Bing and Wei, Xiaokai and Shao, Jidong,"Spotting fake reviews via collective positive-unlabeled learning" in 2014 IEEE International Conference on Data Mining , 2014.

[11] Gary M. Weiss, Brian D. Davison, "Data Mining", To appear in the Handbook of Technology Management, H. Bidgoli (Ed.), John Wiley and Sons, 2010.

[12] http://www.text-analytics101.com/2014/11/what-are-n-grams.html, last accessed on 21 june,2016.