International Journal of Computer and Communication Technology

Volume 7 | Issue 1 Article 9

January 2016

SURVEY ON REVIEW SPAM DETECTION

SNEHAL DIXIT
Department of Computer Science, Shri Ramdeobaba College of Engineering & Management, Nagpur, India, snehalsayali.dixit@gmail.com

A.J. AGRAWAL
Department of Computer Science, Shri Ramdeobaba College of Engineering & Management, Nagpur, India, avinashjagrawal@gmail.com

SURVEY ON REVIEW SPAM DETECTION

SNEHAL DIXIT[1] & A.J.AGRAWAL[2]

[1,2]Department of Computer Science, Shri Ramdeobaba College of Engineering & Management, Nagpur, India
E-mail: snehalsayali.dixit@gmail.com, avinashjagrawal@gmail.com

Abstract- The proliferation of E-commerce sites has made web an excellent source of gathering customer reviews about products; as there is no quality control anyone one can write anything which leads to review spam. This paper previews and reviews the substantial research on Review Spam detection technique. Further it provides state of art depicting some previous attempt to study review spam detection.

Keywords- *Natural language processing, reviews centric spam detection, reviewer centric spam detection*

I. INTRODUCTION

Today due to the popularity of Ecommerce sites it became a target for spammers apart from well-known email and web spam. Review spam refers to the fraud spam written by spammer to hype the product features or defame them. Most of the E-commerce sites provide review section for users so that they can post reviews of products at merchant site and express their views. Such content contributed by web is called as user-generated content. This content forms valuable information for merchants other customers, product manufacturers. Though these reviews are important source of information there is no quality control on this user generated data, anyone can write anything on web which leads to many low quality reviews still worse review spam which mislead customers affecting their buying decisions. Though this is the case in past few years there is growing interest in mining opinion from these reviews by academicians and industries; but not much reported study regarding important issue related to trustworthiness of online reviews. Depending upon the approach used for spam detection it can be classified as:

A. (Review centric approach)
Techniques included in this section depend upon the content of review Fake Reviews are classified for first time by [5] in three categories:

- Type 1 (untruthful opinions): Those that deliberately mislead readers or opinion mining systems by giving undeserving positive reviews to some target objects in order to promote the objects (which we call hyper spam) and/or by giving unjust or malicious negative reviews to some other objects in order to damage their reputation (which we call defaming spam).
- Type 2 (reviews on brands only): Those that do not comment on the products in reviews specifically for the products but only the brands, the manufacturers or the sellers of the products. Although they may be useful, we consider them as spam because they are not targeted at the specific products and are often biased.
- Type 3 (non-reviews): Those that are non-reviews, which have two main sub-types: (1) advertisements and (2) other irrelevant reviews containing no opinions (e.g., questions, answers, and random texts).

Based on these types different techniques are used to detect different review spams.

B. (Reviewer centric approach)
Techniques included in this section identify several characteristics behaviors so as to detect the spammers.

In view of above consideration, this paper defines this paper previews and reviews the substantial research on Review Spam detection technique. Further it provides state of art depicting some previous attempt to study review spam detection. Remainder of this paper is organized as follows. Section 2 covers survey of different Review spam detection techniques. Section 3 provides comparison among methods discussed in section 2. Section 4 concludes the paper.

II. RELATED WORK

Moreover work has been done in detecting two types of spam which are web spam and email spam. Web spam refers to the action of misleading search engines to rank some web pages higher than they deserve [2,3].Web spam can be classified as content spam (adding irrelevant word to the document to rank it high) and link spam (spam on hyperlink [1]).review spam is similar to that of web spam in some respect but hyper links which are sparsely used in reviews and adding irrelevant words to web page also doesn't help much in review spam; it's make it different from web spam.

Another type of spam is email spam which may be defined as "Email spam is unsolicited, unwanted email that was sent indiscriminately, directly or indirectly, by a sender having no current relationship with the user" [4]. Spam emails generally contents advertisements which are very rarely used in review

spam and if used can be detected easily by customers which makes it less harmful.[5] has defined problem of review spam as "to classify review into two categories as spam and non spam". In the direction of solving this problem some attempts have been made ;this section is further discuss two different approaches for review spam detection as mention in earlier section .

A. Review centric spam detection

Different types [5] of spam require different treatment for detecting them. Type 2 and 3 are easily recognizable manually, so for automatic detection machine learning approach can be used by using labeled data as its input. Most difficult is to detect Type 1 spam (untruthful spams) as they are difficult to label manually. We can classify them using the concept of duplicate or near duplicate reviews.

Duplicate Review can be defined as exactly similar reviews while near duplicate review refer to partially similar reviews (similarity percentage under consideration varies).This section further previews existing research detecting all such type of reviews concentrating on the content of review as main source of spam. Each paper preview is described as proposed method in paper followed by evaluation method used for evaluating the proposed system.

The task of detecting fake reviews and reviewers was first proposed by Nitin and Liu in [1], which they call opinion spam detection. This paper proposed a supervised machine learning method for detecting TYPE1, 2, 3 spams.

Proposed Method

Method is divided into three steps 1)to detect type 2,3 spam using supervised machine learning 2)identifying duplicates and near duplicates 3)identify type 1 spam. For type 2, 3 spam detection logistic regression was used. Large set of features were defined which were grouped under three categories

1. Review centric features
2. Reviewer centric features
3. Product centric features

Duplicate and near duplicate reviews were detected using shingle method which use 2-gram based review content comparison.

Finally type 1 spam was detected using duplicate and near duplicate review as positive sample and unique review as negative sample for supervised machine learning which also use logistic regression model. This model also identifies outlier reviews to great extent.

Evaluation Method

For evaluation purpose in [1] AUC (Area under ROC curve) is employed .Also lift curve are used to visualize the performance of TYPE1 spam detecting logistic regression model to predict outlier reviews.

A state of art method proposed by [6] is based on conceptual level similarity. It mainly concentrates on different review format used on web which is mentioned in [6].

- Format1: pros and cons

-pros and cons are separately mentioned by the reviewer.

- Format2: pros, cons and detailed review

-along with pros and cons detailed review is asked to the reviewer.

- Format3: free format

-there is no separation of pros and cons in the review.

In [6] they have used product features that have been commented by the reviewers in their reviews. Different review format require different spam detection techniques. Type 1and 2 did not need special extraction of features while in type 3 format features have to be identified first.

Proposed Method

This method makes use of duplicate and near duplicate reviews considering them as spam while partially relate and unique reviews as non spam.

It has three steps

1. Feature extraction-It involves feature extraction from reviews and storing them in feature database. Sample feature extracted stored in database is shown below:

f1	f2	f3	f4	.....	fn
price	lcd	zoom	speed	.....	Size

2. Feature matrix construction-features extracted in step 1 are used to construct feature matrix. Sample matrix s as shown below

Feature matrix	Price	lcd	zoom	speed	size	Total
Review No	f1	f2	f3	f4	fn	
R1	0	1	1	0	0	2
:	:	:	:	:	:	:
Rm	1	0	0	1	1	3

3. Matching feature calculation between reviews-By calculating similarity score of different review pairs they are categorised as spam (duplicate/ near duplicate) or non-spam (partially related /unique) based on threshold value T.

sim (Ri,Rk) =NC –DH(Ri,Rk)

where NC=total number of feature in each review Ri

DH(Ri,Rk)=Hamming distance between review vector Ri and Rk

Evaluation Method

For evaluation purpose confusion matrix is created for pros and con s separately.

Similar to the method proposed by [6] was proposed by [7] but with some refinements in the method .Main idea of this paper was also resemblance calculation of reviews based on their features.

Proposed Method

In this paper a novel technique named as shingling technique is used for detecting spam reviews based on the product features that have been commented in reviews. Steps involved in spam detection are

1. Review pre-processing
2. Feature extraction
3. Shingle's creation
4. Resemblance ratio calculation of the created shingles between the reviews.

Evaluation Method

For evaluation purpose same method was used as [6], confusion matrix is created for pros and con s separately.

Integrating work from psychology and computational linguistics a new method was proposed by [8] of finding deceptive opinion spam. In this paper three automated approaches were used to detect deceptive opinion spam trained on the dataset (with gold standard deceptive opinions) which was specially developed for the technique used.

Proposed method

Feature used for three automated approach used for deception detection as described in [8] are outlined here.

1. Genre identification

Work in computational linguistics has shown that the frequency distribution of part-of-speech POS) tags in a text is often dependent on the genre of the text (Biber et al., 1999; Rayson et al., 2001). In this approach feature is constructed for each review based on the frequencies of each POS tag for testing relationship this feature and truthful and deceptive reviews.

2. Psycholinguistic deception detection

The Linguistic Inquiry and Word Count (LIWC) software output is used to derive features. One feature for each of the 80 LIWC dimensions is created, which can be summarized broadly under the following four categories:

i. Linguistic processes: Functional aspects of text
ii. Psychological processes: Includes all social, emotional, cognitive, perceptual and biological processes, as well as anything related to time or space.
iii. Personal concerns: Any references to work, leisure, money, religion, etc.
iv. Spoken categories: Primarily filler and agreement words.

3. Text categorization

Text categorization approach use to model both content and context with n-gram features. Following three n-gram feature sets were considered, with the corresponding features lowercased and unstemmed: UNIGRAMS, BIGRAMS+, TRIGRAMS+, where the superscript + indicates that the feature set subsumes the preceding feature set.

Features from the three approaches just introduced are used to train Naive Bayes and support Vector Machine classifiers.

Evaluation Method

Three Meta judges were appointed to annote the dataset samples as deceptive (imaginative) or informative. This result is compared with the automated approach.

A state of art method was introduced in [9] which include supervised machine learning as well as two view co-training algorithm is used for semi supervised machine learning.

Proposed method

With labeled review spam dataset a supervised method is designed to identify review spam. Naive Bayesis used as classifier with basic assumption that features are conditionally independent given the reviews category. A co-training algorithm was given with two views of feature set (review and reviewer based) for semi-supervised machine learning which are outlined below:

Review based features:

1. content feature
2. sentiment features
3. product features
4. meta-data features

Reviewer based features:

1. profile features
2. Behavior features

Co-training Algorithm

Require: two views of feature sets for each review: review features

Fr and reviewer features Fu; a small set of labeled reviews L; a

large set of unlabeled reviews U.

Ensure: Loop for I iterations

1: Learn the first view classifier Cr from L based on review features Fr;

2: Use Cr to label reviews from U based on Fr;

3: Choose p positive and n negative most confidently predicted reviews Treview from U.

4: Learn the second view classifier Cu from L based on reviewer features Fu;
5: Use Cu to label reviews from U based on reviewer features Fu;
6: Choose p positive and n negative most confidently predicted reviews T'reviewer from U.
7: Extract the reviews T'review authored by T'reviewer
8: Move Reviews Treview U T'review from U to L with their predicted labels.
Evaluation Method

A10 fold cross validation was conducted by randomly splitting data set into ten folds ,where nine folds are selected for training and tenth fold is selected for test.

[10] Introduces a method of spam detection identical to [1] but revised feature set which improve accuracy to 88.3%.

B. Reviewer centric spam detection
Although multiple reviews posted by a same reviewer seem suspicious it is not always the case that they are spam, they may be the result of multiple purchasing experience or may be the improvement in the same review. So it became necessary to take into consideration reviewer behavior while detecting review spam. This section discusses two papers related to this approach of spam detection.

[11] has introduced a user centric and user behavior driven approach for review spam detection .A user centric approach is preferred over review centric approach as gathering behavioral evidence of spammers is easier than that of spam reviews[11].this paper basically dealt with four different spamming model:
Target Based

1. Targeting Product(TP)
2. Targeting Group(TG)

Deviation Based

1. General rating Deviation(GD)
2. Early rating Deviation(ED)

Proposed method

Data is preprocessed using 4 preprocessing steps listed below before use for spam detection.

1. Removal of anonymous users
2. Removal of duplicate products
3. Removal of inactive users and unpopular products
4. Resolution of brand name synonyms

After preprocessing spam detection is done for three spamming behaviors involving targeted products and product groups and derives their respective spam scores for each reviewer representing the ex-tent to which he practices the behaviors.
Evaluation Method
Review spam detection software is used to facilitate manual evaluation.
In [12] a novel method is used which make use of a heterogeneous graph to detect the relationship between reviewer, review and store. This relationship is used to identify trustiness of reviewers, honesty of reviews and reliability of reviews.
Proposed Method

[12] had proposed an iterative algorithm whose inputs are set of stores, review and reviewers producing set of reliability ,honesty ,and trustiness as output.

Evaluation method
In this paper IR-based evaluation strategy is used.

III.COMPARATIVE ANALYSIS

In section II we had discuss various methods to detect review spam. There comparative analysis based on accuracy to determine review spam is given in table 1.

IV. CONCLUSIONS

In this work we surveyed existing techniques and algorithms created for Review centric and Reviewer centric spam detection. To draw a general picture of the review spam detection, we first provide proposed work in each paper. We also presented a brief overview of evaluation method used to determine accuracy.

At last we had provided a comparative study about different spam detection techniques depending upon their accuracy.

Tabe1: comparative analysis

Sr no	Method	Precision
Review centric methods		
1	Opinion Spam and Analysis[1]	85%
2	Conceptual level Similarity Measure Based Review Spam Detection[6]	43.64%
3	Spam Detection of customer Reviews from Web Pages[7]	75.04%
4	Deceptive Opinion Spam by Any Stretch of Imagination[8]	83.3%
5	A Method for sorting out the Spam from Chinese Product Reviews.[10]	88.3%
Reviewer centric methods		
6	Detecting product review spammers using rating	78%

	behaviors[11]	
7	Review Graph based Online Store Review Spammer Detection[12]	49%

REFERENCES

[1] Nitin Jindal and Bing Liu .Opinion Spam and Analysis Department of Computer Science University of Illinois at Chicago 851 South Morgan Street, Chicago, IL 60607-7053 Copyright 2008

[2] Z. Gyongyi & H. Garcia-Molina. Web Spam Taxonomy.Technical Report, Stanford University, 2004.

[3] Nikita Spirin and Jiawei Han. Survey on Web Spam Detection: Principles and Algorithms Department of Computer Science University of Illinois at Urbana-Champaign Urbana, IL 61801, USA

[4] Gordon Cormack and Thomas Lynam. Spam corpus creation for TREC. In Proceedings of Second Conference on Email and Anti-Spam,CEAS'2005, 2005.

[5] N. Jindal and B. Liu. Analyzing and Detecting Review Spam. ICDM2007.

[6] Siddu P.Algur Amit P.Patil P.S Hiremath S.Shivashankar .Conceptual level Simiarity Measure Based Review Spam Detection

[7] Siddu P.Algur Amit P.Patil P.S Hiremath S.Shivashankar .Spam Detection of customer Reviews from Web Pages.

[8] Myle Ott , Yejin choi ,Claire Cardie , Jeffrey T.Hancock.Finding Deceptive Opinion Spam by Any Stretch of Imagination.

[9] Fangtao Li,Minilie Huang,Yi Yang ,Xiaoyan Zhu.Learning to Identify Spam.Proceedings of the twenty second International Joint Conference on Artificial Intiligence.

[10] Lijia Liu,Yu Wang .A Method for sorting out the Spam from Chinese Product Reviews.2012 IEEE

[11] E.P. Lim, V.A. Nguyen, N. Jindal, B. Liu, and H.W. Lauw. Detecting product review spammers using rating behaviors. In CIKM, 2010.

[12] Guan Wang, Sihong Xie, Bing Liu, Philip S. Yu.Review Graph based Online Store Review Spammer Detection.11th IEEE International Conference on Data Mining 2011.