International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 04, April-2015

A Survey on Review Spam Detection techniques

Vyas Krishna Maheshchandra
Post graduate student
Department of Computer Engineering
Dharamsinh Desai University
Nadiad, Gujarat, India

Prof. Ankit P. Vaishnav
Department of Computer Engineering
Dharamsinh Desai University
Nadiad, Gujarat, India

Abstract: **An excellent source of gathering the reviews on particular product is different shopping sites where people share their reviews on products as well as their shopping experience. Sometimes people may come across the wrong opinions called as review spam. So, for this it necessary to detect it by some means. In this paper a survey is done on the different techniques introduced to detect the Review spam with their results. And a brief comparison is done.**

Keywords: Review spam, un-truthful reviews (type-1), non-reviews (type-3)

I. INTRODUCTION

Now a day there is no quality control for social media sites and one is having freedom to share on media which leads to review spam. And it is a need to recognize review spam because most of the users tale their decision based on the reviews. This situation mainly arises for different shopping sites or the sites or hotels also. Different techniques are introduced and used for detecting review spam.

[1] has given main 3 types of review spam, which are

a. Un-truthful review (False opinions) which is divided in two category. Positive Spam review(Undeserving opinion to promote product) Negative Spam Review(negative opinion to damage reputation
b. Reviews on brand only(reviews on some particular brands)
c. Non-reviews (contain no reviews) which is divided in Advertisements, Question or answers, Comment on other reviews or any Random text.

In this survey paper different techniques used to detect these type of spam are discussed.

The rest of the article is organized as follows. Section II discuss about related work on different types of spam. Section III gives related work. Section IV discusses about analysis, and the last section V is about conclusion.

II OTHER TYPES OF SPAM

Most the work is done to detect Email spam and Web spam.

In day to day life many times we need to access email accounts. And every day we receives many mails about advertisements, or some form of survey which can be fake and harmful software also. Mostly used technique for email spam is White-list and Black-list method which is based on IP Address in which a set is created which defines mail from which IP (black-list) is for spam and mail from which IP (white-list)is not spam. Other method which are widely used are using KNN algorithm [2] for images in mail spam using an OCR (Optical character reorganization) [3].

A technique which gives higher ranking to the pages which is more than actual is called web spamming. Spammer uses different tricks to increase the rankings of their websites in the search results. Mainly used techniques for this type of spam is Content spam and link spam. To identify web spam Linguistic features have also been examined [4].

III RELATED WORK

A. Analyzing and Detecting Review Spam

[5] has introduced main 3 types of review spam which is discussed in above Introduction.

In this work first they have detected near-duplicates reviews by observing whether

a. there is same review for same product from different user ids
b. there is same review for different products from same user id or
c. there is same review for different product from different user id

Their outline of work was to detect duplicates and near duplicates by using 2-gram method, and then to detect type-2 and type-3 spam using supervised learning methods and detect type-1 review based on above types of duplicates and other information. Duplicated reviews identified based on similarity score of 2-grams of the reviews on different products.

Type-2 and 3 were identified by manually labeling reviews by their classes and then given it to logistic regression. For that they identify total 36 type of features which are product, review and reviewer centric.

And for experimental results they had used the measure which is AUC (Area under ROC curve) and lift curve.

B. Review Spam Detection

[6] This paper presents a review based supervised machine learning approach to detect untruthful review (type 1 spam) it also previews some previous attempt to study review spam detection. Depending upon the approach used for spam detection it can be classified as:

A. (Review centric approach)
B. (Reviewer centric approach)
In this work main modules are
 Customer Reviews
 Review pre-processing
 Stop-word-removal
 Detecting Duplicate and near duplicate
 Type 1 Spam(Un-truth full spam review) [1] Classification Technique

Here for type 1 classification technique they used total 12 features extracted from reviews and give labels to each.

For evaluation they have compared accuracy of four machine learning methods Gaussian, naïve bayes, Decision Tree, Multinomial naïve bayes, Logistic Regression. MProducts which is category if products reviews were used for evaluation .And they have shown that Logistic Regression and Gaussian have higher accuracy as compared to Decision tree and Multinomial naïve bayes.

C. Conceptual level Similarity Measure based Review Spam Detection

[7] has detected mainly type 1 spam reviews based on the similarity measure. The format of reviews they had used is pros and cons. According to them the review is not a spam in following two conditions

1. If the number of matched features is below some specified threshold i.e. partially related reviews
2. If no one features are matched between two review i.e. Unique Review

It has three steps
1. Feature extraction-It involves feature extraction from reviews and storing them in feature database
2. Feature matrix construction-features extracted in step 1 are used to construct feature matrix.
3. Matching feature calculation between reviews-By calculating similarity score of

Different review pairs they are categorized as spam (duplicate/ near duplicate) or non-spam (partially related /unique) based on threshold value T. For evaluation purpose confusion matrix is created for pros and con s separately and compared human annotated result with automated result.

D. Toward A Language Modeling Approach for Consumer Review Spam Detection

This paper [8] is to show the trustworthiness of reviews by detecting the review spam. Their experimental result shows that the KL divergence and the probabilistic language model is effective for the detection of untruthful reviews.

In their work they have used

- The pre processing techniques like POS (Part of speech Tagging), stop-word-removal, stemming on the data crawled from web.
- And they have developed their POS tagger based on the word-net lexicon and the publically available Word-Net API. And used the unsupervised probabilistic language model (for untruthful review detection which is type 1 review spam), and a supervised classifier (for non-review detection which is type 3 spam).
- For non-review spam detection they identify features which were used in detecting web spam [4] Which are Syntactical, Lexical and Stylistic features. For classification task they have used SVM (Support Vector Machine) and Logistic Regression.
- For un-truth full type of reviews they build the computational model using KL (Kullback-Leibler) divergence which is a well-known measure commonly used to estimate the distance between two probably distributions. And compare result with Vector Space model and Logistic regression.

E. Text Mining and Probabilistic Language Modeling for Online review Spam Detection

[9]Has detected type 1 and type 3 spam reviews. Main focus on type 1 spam review.

In this study they have detected the fake reviews and the final decision was on the Visitors that the review is fake or not. In their work they have divided their work in following modules.
Module 1: In this the user selects the detection scope.
Module 2: If reviews are not available locally then use API (Application Programming Interface) to retrieve reviews.
Module 3: traditional document preprocessing procedures, which are stop-word removal, Part-of-Speech (POS) tagging, and stemming were applied on data.
Module 4: after the reviews were preprocessed, the high-order concept association mining module was invoked to extract the prominent concepts and their high-order associations for each product category. These association relationships were used to bootstrap the performance
Module 5: non review detection is performed by a supervised SVM classifier.
Module 6: untruthful review here type-1 spam review detection is carried out by an unsupervised probabilistic language model.

For the non-review spam detection they have used SVM (Support Vector Machine) and LR to classify the reviews. For that they have used the features same as in web spam detection technique [4] for SVM[8].

And for un-truth full reviews they developed their model and used different techniques.

The results: they have used the methods for un-truth full reviews are SVM, VS (Vector Space),I-match, LM(unigram Language Model), SLM(Semantic Language Model). The result shows that SLM gives the highest result and SVM gives poor result.

And for non-review based spam they used KNN (Nearest neighbor classifier), LR(Logistic Regression), and SVM (Support Vector Machine). Results shows that SVM is performing well and it has the highest result among them.

For validation and performance checking they adopted following technique

TABLE 1 CONFUSION MATRIX

	Human Classification		
		Spam	*Ham*
System's Classification	*Spam*	a	b
	Ham	c	d

With reference to the above confusion matrix depicted in Table the various effectiveness measures can be defined by

$hm = b/(b+ d)$

$sm = c/(a + c)$

$lam = \text{logit}^{-1}(\text{logit}(hm) + \text{logit}(sm)/2$

Here a, b, c, and d are the number of reviews falling into each category. The ham misclassification rate (hm) is the ratio of all ham misclassified as spam, and the spam misclassification rate (sm) is the ratio of all spam misclassified as ham. All measures are according to the TREC Spam Track [10]. Where lam us the logistic average misclassification rate . As hm, sm, and lam are measures of failure rather than effectiveness. A small ratio means good performance.

IV. ANALYSIS

The comparison and analysis of all the techniques used in previously discussed paper is shown as below.

Table 2 ANALYSES OF TECHNIQUES

Sr no.	Reference Paper	Dataset Used	Type of Review Spam	Method	Result/Conclusion	Remark
A.	Analyzing and Detecting Review Spam	Reviews downloaded from Amazon.com	Type 1 Type 2 Type 3	Logistic Regression	AUC(Area Under ROC Curve): Type-2→98.5% Type-3→99.0%	Had check for SVM and Naïve Bayes also and found LR(Logistic Regression) better. Additional performance measures are there.
B.	Review Spam Detection	Reviews downloaded from Amazon.com	Type 1	Proposed a Review centric supervised machine learning technique	Accuracy: GaussianNB~90% Decision Tree ~65% MultinominalNB ~60% Logistic Regression~92%	Has divided result based on percentage of training data is used.
C.	Conceptual level Similarity Measure based Review Spam Detection	Reviews of format pros and cons	Type 1	based on conceptual level similarity	-	Compared the automated results of proposed system with human annotated results: Results comparison with human perception makes an unrealistic approach of detecting spam reviews
D.	Toward A Language Modeling Approach for Consumer Review Spam Detection	Reviews downloaded from Amazon.com	Type 1 Type 3	Language model	For Untruthful reviews(type-1): True Positive ratio KL-96.38% VS-92.62% LR-17.15% For Non-reviews(Type-3): SVM-92.86% LR-85.17%	Additional performance measures are there. These are shown in above section in paper E.
E.	Text Mining and Probabilistic Language Modeling for Online review Spam Detection	Reviews downloaded from Amazon.com	Type 1 Type 3	Semantic Language Model	For Untruthful reviews: True Positive ratio SLM-97.77% LM-95.88% I-match-95.92% VS-94.52% SVM-56.53% For Non-reviews: SVM-95.06% LR-94.19% KNN-92.05%	Additional performance measures are there. These are shown in above section in paper E.

V CONCLUSION

This paper shows various approaches for review spam detection. All approach has some advantage and some disadvantage. Main aim is to correctly identify the review as a spam or not.

REFERENCES

[1] Nitin Jindal and Bing Liu. "Opinion Spam and Analysis." Proceedings of First ACM International Conference on Web Search and Data Mining (WSDM-2008), Feb 11-12, 2008, Stanford University, California, USA

[2] Loredana Firte Camelia Lemnaru Rodica Potolea *Spam Detection Filter using KNN Algorithm and Resampling* 2010 IEEE

[3] Peng Wan, Minoru Uehara Spam Detection Using Sobel Operators and OCR 2012 26th International Conference on Advanced Information Networking and Applications Workshops

[4] Jakub Piskorski, Marcin Sydow, Dawid Weiss Exploring Linguistic Features for Web Spam Detection:A Preliminary Study ACM 200x

[5] Nitin Jindal and Bing Liu Analyzing and Detecting Review Spam Seventh IEEE International Conference on Data Mining 2007

[6] SNEHAL DIXIT & A.J.AGRAWAL REVIEW SPAM DETECTION International Journal of Computational Linguistics and Natural Language Processing Vol 2 Issue 6 June 2013 ISSN 2279 – 0756

[7] Siddu P. Algur, Amit P.Patil, P.S Hiremath, S. Shivashankar Conceptual level Similarity Measure based Review Spam Detection 2010 IEEE

[8] C.L. Lai, K.Q. Xu, Raymond Y.K. Lau, Y. li, L. Jing Toward A Language Modeling Approach for Consumer Review Spam Detection International Conference on E-Business Engineering 2010

[9] RAYMOND Y. K. LAU, S. Y. LIAO, RON CHI-WAI KWOK, KAIQUAN XU, YUNQING XIA, YUEFENG LI Text Mining and Probabilistic Language Modeling for Online Review Spam Detection ACM Trans. Manag. Inform. Syst. 2, 4, Article 25 (December 2011)

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