SPAM REVIEW DETECTION USING NATURAL LANGUAGE PROCESSING TECHNIQUES

Chaitanya Kale
Department of Information Technology, College Of Engineering, Kopargaon, India
Dadasaheb Jadhav
Department of Information Technology, College Of Engineering, Kopargaon, India
Tushar Pawar
Department of Information Technology, College Of Engineering, Kopargaon, India

ABSTRACT

In recent year, online reviews have become the most important resource of customer opinion. Existing research has been focused on extraction, classification and summarization of opinion from reviews in websites, forums and blogs. Nowadays consumer can obtain information for products and service from online review resources, which can help them make decision. The social tools provided by the content sharing applications allow online user to interact, to express their opinions and to read opinions from other users. But the spammers provide comments which are written intentionally to mislead users by redirecting them to web sites to increase their rating and to promote products less known on the market. Reading spam comments is a bad experience and a waste of time for most of the online users but can also be harming and cause damage to the reader. Several researchers in this field focused on only spam or non-spam comments. But, our goal is to detect comments which are likely to represent spam considering some indicators like a discontinuous flow of text, inadequate and vulgar language or not related to the specific context will helps in giving correct feedback of various customers reviews about given product. Mainly we have observed that previous work is focused on extraction, classification and summarization of opinion and checking of spam and non-spam. But, proposed system aims to find irregular or discontinuous text flow, vulgar language or not related to specific context and check similarity between the comments.

Keywords: Spam; Machine learning; Topic extraction; Post-comment similarity; Feature vector; Opinion sentiment; Review inter-relationship; Group spammers.

INTRODUCTION

The Web is the greatest repository of digital information and communication platform ever invented. People around the world widely use it to interact with each other as well as to express opinion and feelings on different issues and topics. With the increasing availability of online review sites and blogs, customers depend on online review to make their purchase decision and business to respond promptly to their client's expectations. Detecting opinion spam is a very challenging problem since opinion expressed in the web are typically short text, written by unknown people using different style and for different task, particularly because human being are not always able to reliable determine which review are spams.

The proposed work presents a method for detecting whether review is spam or non-spam. These methods essentially rely on duplicate reviews only suited for some special spamming activities. To capture inter-relationship among reviewers, reviews and stores and proposed a novel review graph based on spam reviewer detection approach. The main goal is to detect comments which are likely to represent spam considering some indicators that is a discontinuous text flow, inadequate and vulgar language or not related to a specific context by using machine learning approaches.

LITERATURE SURVEY

Siddu P. Algur [2] in this paper, make an attempt to detect review is spam or non-spam. The trustworthiness of the review is assessed as a spam or non-spam with includes both duplicate and non-duplicate review. Author proposed novel and effective technique namely conceptual level similarity measure used for detection of spam review based on the product features that have been commented in the review. The efficiency of the task of web base customer review spam detection can be enhanced by identifying and eliminating duplicate spam reviews, thereby providing summary of trusted review for customer to make buying decision.

D. Liang, H. Shen. [3] In this paper, author constructed a novel multi-edge graph model in which each node represents a reviewer and each edge represents an inter-relationship between reviewers on one special product. Combing with the features based on reviewers' unreliability score, author proposed an unsupervised iterative computation framework. It is the first algorithm to consider both of the reviewers' features and their inter-relationships, and places emphasis on detecting the spammers who always work together. Experimental results show that the method is effective in detecting spam reviewers with a satisfied precision.

A. Gupta, R. Kaushal [4] OSNs (Online Social Network) have become a new medium for dissemination of information, at the same time; they are also fast becoming a playground for the spread of misinformation, fake news, rumors, unsolicited messages, etc. Spammers, out of malicious intent, post either unwanted (or irrelevant) information or spread misinformation on OSN platforms. As part of work, author proposed mechanisms to detect such users (Spammers) in Twitter social network (a popular OSN). This work is based on a number of features at tweet-level and user-level like Followers/Follows, URLs, Spam Words, Replies and Hash Tags.

H. A. Najada, X Zhu. [5] In this research, author's main aim is to distinguish between spam and non-spam reviews by using supervised classification methods. One of the most effective ways to distinguish spam and non-spam reviews is by using machine learning techniques, which has proved its ability in such problems especially when dealing with text and natural language.

X. Yang. [6] In this paper, author proposed an iterative computation framework to detect spam reviews based on coherent examination. They first define some reviews' coherent metrics to analyze review coherence in the granularity of sentence. In this part author have provided several metrics to measure the coherence of a review based on the flow smoothness information between sentences: Word Transition Probability – Conditional Probability, Word Concurrence Probability – Join Probability.

R. Patel, P. Thakkar. [7] This paper focuses on the detection of deceptive opinion

spam. A recently proposed opinion spam detection method which is based on *n*-gram techniques is extended by means of feature selection and different representation of the opinions. The problem is modelled as the classification problem and Naïve Bayes (NB) classifier and Least Squares Support Vector Machine (LS-SVM) are used on three different representations (Boolean, bag-of-words and term frequency–inverse document frequency (TF-IDF)) of the opinions.

M. L. Ramprasad, M. G. Amudha. [8]In the existing work related to the you Tube the user have to search for the videos which takes a long time and irrelevant videos not related to the user search is displayed. The aim of the proposed work is that, User-generated content (UGC) video systems by definition heavily depend on the input of their community of users and their social interactions for video diffusion and opinion sharing. As a UGC system can achieve a larger audience through improved connectivity, findings motivate to propose a mean to enhance the users' connectivity by taking benefit of friend recommendation and spammer detection of the online videos.

H. Li, Z. Chen, B. Liu, X. Wei, J. Shao,[9]This paper reports the work on fake review detection in Chinese with filtered reviews from Dianping's fake review detection system. Dianping's algorithm has a very high precision, but the recall is hard to know. This means that all fake reviews detected by the system are almost certainly fake but the remaining reviews (unknown set) may not be all genuine. They first proposed a collective classification algorithm called Multi-typed Heterogeneous Collective Classification (MHCC) and then extend it to Collective Positive and Unlabeled learning (CPU). Experiments are conducted on real-life reviews of 500 restaurants in Shanghai, China. Results shows that proposed models can markedly improve the F1 scores of strong baselines in both PU and non-PU learning settings.

Y. Lin, H. Wu, J. Zhang, X. Wang, A. Zhou. [10]In this paper, authors have explored the issue on fake review detection in review sequence, which is crucial for implementing online anti-opinion spam. Authors have analyzed the characteristics of fake reviews firstly. Based on review contents and reviewer behaviors, six time sensitive features are proposed to highlight the fake reviews. The experimental result shows that these methods can identify the fake reviews orderly with high precision and recall.

J. Z. Wang, Z. Yan, L. T. Yang, B. X. Huang,[11] Existing research has been focused on extraction, classification and summarization of opinions from reviews in news websites, forums and blogs. Important issue that has not been well studied is the degree of relevance between a review and its corresponding article. In this paper, author proposed a notion of ''Review Pertinence'' to study the degree of this relevance. Unlike usual methods, they measure the pertinence of review by considering not only the similarity between a review and its corresponding article, but also the correlation among reviews

PROPOSED SYSTEM

In our research show that it is possible to detect spam comments with the proper selection of features which capture different characteristics of legitimate comments in order to differentiate them from spam comments. In our experiment we consider as spam the following type of documents associated with a review by using some indicators: (i) incoherent comments with increased number of punctuation marks, new lines, stop words, non ASCII characters and white spaces, (ii) inadequate comments which contain offensive words and (iii) coherent comments which do not provide relevant content to a specific topic.

Our experiment we makes use of natural language processing techniques in order to identify the relevant features of spam comments. We propose a supervised learning approach and experiment different sets of features to correctly classify comments as spam or not.

SYSTEM ARCHITECTURE

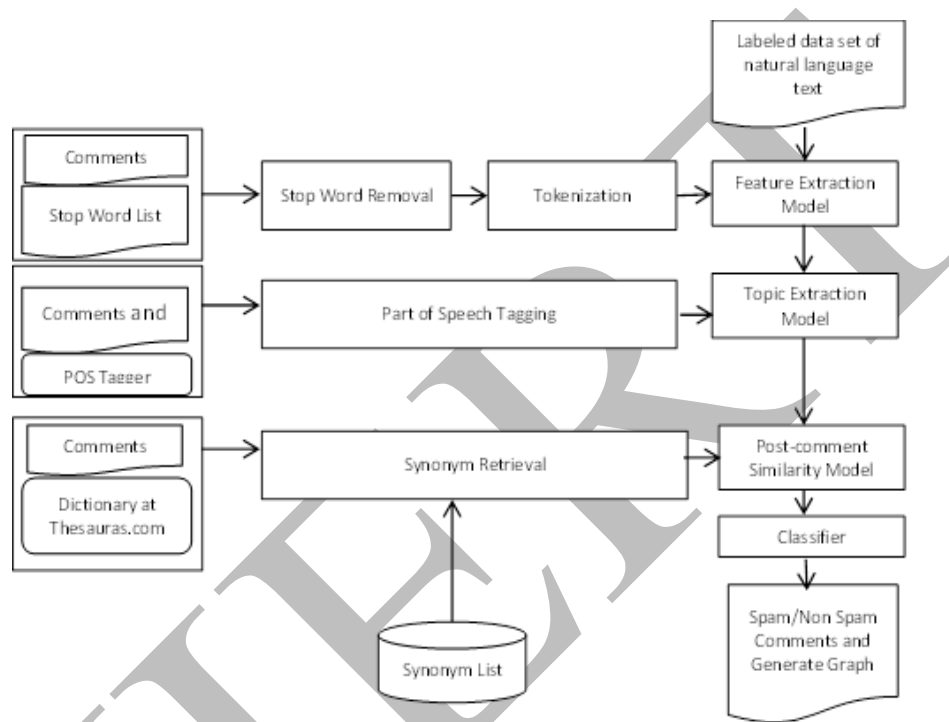As shown in Figure 1, spam review detection system using NLP techniques:

Fig.1. Spam Review Detection System

The architecture of proposed system is composed of three main modules:

- Feature extraction module.
- Post-comment similarity module
- Topic extraction module.

FEATURE EXTRACTION MODULE

In feature extraction module the proposed system eliminate comments which are discontinuous and contain vulgar expressions. The identification of these types of comments relies on the identification of countable features: links, white spaces, sentences, punctuation marks, word duplication, stop words, non ASCII characters, new line, and capital letter.

We are implement feature extraction module based on some characteristics of spam comments:

a) Number of links in the given comments.
b) Number of white spaces in the given comments.
c) Number of sentences in the given comment.
d) Number of punctuation marks in the given comment.
e) Comment Word duplication.
f) Comment Stop words ratio.
g) Number of Non ASCII Characters in the given comment.

POST-COMMENT SIMILARITY

This module detects whether the comment and post contain similar topic based on the following similarity metric: the normalized value of the sum of the frequencies of occurrences of each word and its synonyms from the comments in the post. The post-comment similarity module which connect to online directory to retrieve all the synonyms for the word in the comment. After the synonyms retrieval process the post-comment similarity formula is applied and the post-comment similarity degree is calculated.

TOPIC EXTRACTION MODULE

The topic extraction module is designed to determine if there are common topic between a comment and a post and to find out if the contents of the comment are related to the content of the related post. The proposed system considers two basic types of topics – bigrams and uni-grams which are extracted using combination of shallow natural language processing technique. To identify uni-gram topic system extracts a collection of candidate nouns. To create set of bigrams topic system extract all bigrams from both the comments and the post which conform to one of two basic part-of-speech co-location patterns.

EXPECTED RESULT

We present Spam Review Detection System in that we are using Natural Language Processing Techniques. We can find comment is spam or non-spam by using some indicators such as: irregular or discontinuous text flow, vulgar language or not related to specific context and check similarity between the comments. At the end we can able to count total review spam and non-spam count and we can generate graphical representation.

CONCLUSION

Thus, we are done the research of existing system we have observed that previous work is focused on extraction, classification and summarization of opinion and checking of spam and non-spam. But, proposed system aims to find irregular or discontinuous text flow, vulgar language or not related to specific context and check similarity between the comments.

REFERENCES

[1] C. Radulescu, M. Dinsoreanu, R. Potolea, "Identification of spam comments using Natural Language Processing Techniques", *Technical University of Cluj-Napoca, Romania,* 2014, pp. 29-35.

[2] S. P. Algur, A. P. Patil, P. S. Hiremath, S. Shivashankar, "Conceptual level Similarity Measure based Review Spam Detection*", Dept. of P.G. Studies and Research in Computer Science, Gulbarga University, Gulbarga, Karnataka, India,* 2010, pp. 416-423.

[3] D. Liang, H. Shen, "Detecting spam review by combing reviewer feature and relationship", *2014 International Conference on Informative and Cybernetics for Computational Social Systems (ICCSS),* pp. 102-107.

[4] A. Gupta, R. Kaushal,**"**Improving Spam Detection in Online Social Networks", *Department of Information Technology Indira Gandhi Delhi Technical University for Women Kashmere Gate, Delhi,* pp. 1-6.

[5] H. A. Najada, X Zhu, **"**iSRD: Spam Review Detection with Imbalanced Data Distribution", *Dept. of Computer & Electrical Engineering and Computer Science, Florida Atlantic University Boca Raton 33431 FL, USA,* pp. 553-560.

[6] X. Yang, "One Methodology for Spam Review Detection Based on Review Coherence Metrics", *2015 International Conference on Intelligent Computing and Internet of Things (IC1T),* pp. 99-102.

[7] R. Patel, P. Thakkar, "Opinion Spam Detection Using Feature Selection", *2014 Sixth International Conference on Computational Intelligence and Communication Networks,* pp. 560-564.

[8] M. L. Ramprasad, M. G. Amudha, "Spammer detection and tagging based user generated video search system –a Survey", *ICICES2014 - S.A.Engineering College, Chennai, Tamil Nadu, India,* pp. 6-14.

[9] H. Li, Z. Chen, B. Liu, X. Wei, J. Shao, "Spotting Fake Reviews via Collective Positive-Unlabeled Learning", *2014 IEEE International Conference on Data Mining,* pp. 899-904.

[10] Y. Lin, H. Wu, J. Zhang, X. Wang, A. Zhou, "Towards Online Anti-Opinion Spam: Spotting Fake Reviews from the Review Sequence", *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014),* pp. 261-264.

[11] J. Z. Wang, Z. Yan, L. T. Yang, B. X. Huang, "An approach to rank reviews by fusing and mining opinions based on review pertinence", *Non-Traditional Security Center, Huazhong University of Science and Technology, China, Information Fusion 23 (2015),* pp. 3-15.