

Research Article

## Classification and Blocking of Spam Users based on Review Using Expected Maximization Algorithm

Hema Dewangan<sup>1</sup>, Prof. Om Prakash Dewangan<sup>2</sup>, Prof. Toran Verma<sup>3</sup>

<sup>1-3</sup> Dept. of Computer Science and Engineering , Rungta College of Engineering and Technology, Bhilai, Chhattisgarh, India

### \*Corresponding author

Hema Dewangan

Email: [dewangan.hema1990@gmail.com](mailto:dewangan.hema1990@gmail.com)

**Abstract:** An excellent source of collecting the reviews on specific product is various online shopping sites where people share their reviews on products and their shopping experience. People may come through the wrong opinions known as review spam. Therefore, for this it is essential to detect it by some means. In this paper, presents methods for detection of spam users using feature extraction and discretization, in combination with EM algorithm. Our framework can detect multiple spammers by knowing only small set of spammer sets. Proposed method effectively selects relevant features and builds features set to identify the spammers. In this paper, we have blocked the users with fake id or who are predicted as spammer.

**Keywords:** Review spam, un-truthful reviews, opinion spam, rating spam.

### INTRODUCTION

At the present, there is no quality control for social networking sites and one has having freedom to share their reviews on social networking sites which helps to lead the review spam. And it is a requirement to recognize review spam because most of the users make their decision based on the reviews. This condition mainly arises for various online shopping sites or the sites or hotels also. Various techniques are introduced and used for detecting review spam [6-7].

Many copies of similar content is being distributed among multiple users in the internet. These messages are of no use because of duplication. The importance of message is reduced due to the duplication of many similar messages. Most of the spams are present on the rating websites such as e-commerce, movie rating, product rating etc. People misuse platform and try to capture audience attention by providing many fake reviews about the product. This fake review definitely tries to deceive users by their fancy fake comments.

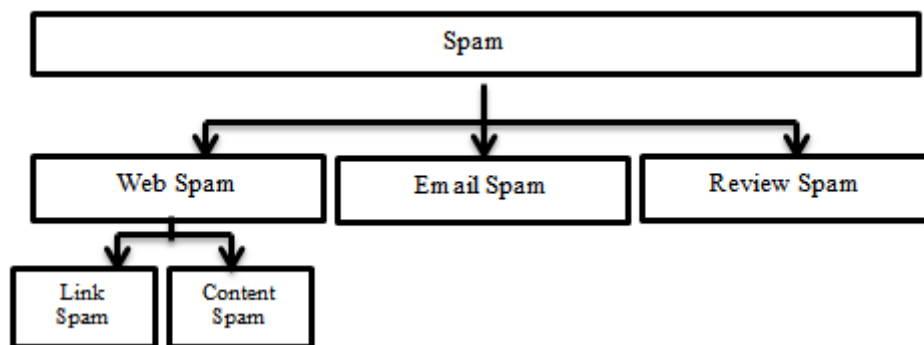


Fig.1. Various types of spam

### Web Spam

The main purpose of web spam is to make people follow some untrusted link. This link will steal all the user information and send data to the hacker server [10]. Spammer tries to manipulate web pages to steal user private information. They try to hide some scripts in a web page and when user clicks on that page, the suspicious link got activated and sends all private information like. Cookies, history to the spammer server.

### Email Spam

A substitute kind of spam is email spam, which is furthermore not the same as review spam. Email spam (furthermore called rubbish messages) incorporates getting the uninvited business plugs [11]. Usually spammer tries to send bulk emails to unknown or random users of their interest as a target. Some suspicious scripts are hidden in the email. Whenever user clicks on the particular link, the scripts activated and steal private information.

### Review Spam

Spam Reviews refers to "illegal" activities that try to mislead readers by giving false negative opinions to some other entities in order to damage their reputations [12]. If the product review is good or satisfactory then user will buy the product. So, to sold product on e-commerce websites people who sell product often do spamming on reviews to attack more number of customers.

### Other Types of Spam

Rating spam generally used to deceive people by providing higher rating for any product and service. Example is Flipkart, where fake users review product and provide higher rating [14].

Various authors have proposed various methods and techniques in order to detect fake reviews and reviewer. For detection behavioral based review detection are proposed by many of authors, which uses probability of content in the review. Removal of spams from any website is very challenging tasks [15, 16].

Most the work is done to detect Email spam and Web spam. In today's 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 or harmful software also. Mostly used technique for email spam is Whitelist and Blacklist method which is based on IP Address in which a set is created which defines mail from which IP (blacklist) is for spam and mail from which IP (whitelist) is not spam. Another technique which are widely used are using KNN algorithm for images in mail spam using an OCR (Optical character reorganization) [17].

### LITERATURE SURVEY

Nitin Jindal [1], focused on review spam and spam detection. Three main types of spam were identified. Detection of such spam is done first by detecting duplicate reviews. We then detected type 2 and type 3 spam reviews by using supervised learning with manually labeled training examples. Results showed that the logistic regression model is highly effective. However, to detect type 1 spam reviews, the story is quite different because it is very hard to manually label training examples for type 1 spam. We presented an approach to use three kinds of duplicates, which are very likely to be spam, as positive training examples to build a classification model. The results are promising.

Nitin Jindal [2], focused on s importance of reviews also gives good incentive for spam, which contains false positive or malicious negative opinions. Author makes an attempt to study review spam and spam detection. To the best of our knowledge, there is still no reported study on this problem.

Siddu P. Algur [3], focused on a novel and effective technique for detecting the trustworthiness of customer reviews for a particular product based on the features of the product being commented by the reviewers. Spam reviews are been categorized as duplicate and near duplicate reviews and non-spam reviews as partially related and unique reviews. Results demonstrate the effectiveness of the proposed technique in detecting spam and non-spam reviews. The efficiency of the task of web based customer review spam detection can be enhanced by identifying and eliminating duplicate and near duplicate spam reviews, thereby providing a summary of the trusted reviews for customers to make buying decisions.

C.L. Lai [4], focused on the development of a novel computational methodology to combat online review spam. Our experimental results confirm that the KL divergence and the probabilistic language modeling based computational model is effective for the detection of untruthful reviews. Empowered by the proposed computational methods, our empirical study found that around 2% of the consumer reviews posted to a large e-Commerce site is spam.

Raymond Y. K. Lau [5], focused on the proposed models outperform other well-known baseline models in detecting fake reviews. To the best of our knowledge, the work discussed in this article represents the first successful attempt to apply text mining methods and semantic language models to the detection of fake consumer reviews. A managerial implication of our research is that firms can apply our design artifacts to monitor online consumer reviews to develop effective

marketing or product design strategies based on genuine consumer feedback posted to the Internet.

Nitin Jindal et.al. [6], it is presently a typical practice for web-based business Web locales to empower their clients to compose surveys of items that they have bought. Such audits give profitable wellsprings of data on these items. They are utilized by potential clients to discover feelings of existing clients before choosing to buy an item. They are additionally utilized by item makers to recognize issues of their items and to discover aggressive knowledge data about their rivals.

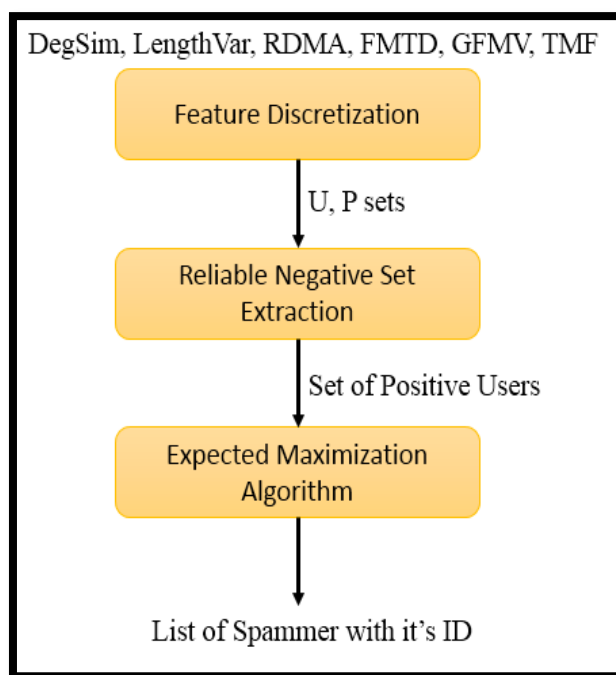
**Methodology**

In this section, we introduce the proposed framework for classification of spammers and non-

spammers. Some of the basic notations are presented below:

- n → number of users
- U → unlabeled set of users
- I → set of m products
- R (i, j) is the binary relation, where user i has review product j is described in the form of binary (yes or no)
- P → set of positive spammers (identified by some manual techniques)

From the given P, R, U, our framework identifies the spam users from U (set of uses).



**Fig. 2. Show the proposed system architecture**

**Feature Discretization**

In our proposed system, the dataset which we have considered is amazon review dataset. Amazon review feature dataset contains only numeric dataset. Modeling of all the features directly is not recommended for spammer detection. Hence discretization is required. We discretize feature f into set of v categories. The categories are:

- i. DegSim
- ii. LengthVar
- iii. RDMA
- iv. FMTD
- v. GFMV
- vi. TMF

These all are the features through which our algorithm runs and produce output.

**Reliable Negative Set Extraction**

This step aims to pick out extremely small set of instances from the set of users U that are different from instances in P (which is positive labeled set). Using below equation we can single out instances.

$$D_f = n_P(f) \log \frac{|P| + |U|}{n_P(f) + n_U(f)} = a \log \frac{n}{a + b},$$

Where, a and b are number of instances in P and U sets respectively. D<sub>f</sub> is feature discriminative for class of positive labeled sets.

### Expected Maximization Algorithm

To classify the users from being spam or non-spam category, EM algorithm is applied over U, P, sets. EM is based on maximization of conditional likelihood.

EM is based on the maximum posteriori algorithm, estimates of the attributes in statistical models. EM algorithm efficiently classifies the spammers with their ID's.

```
{
  "reviewerID": "196",
  "asin": "242",
  "overall": 3,
  "reviewTime": "881250949"
}
{
  "reviewerID": "186",
  "asin": "302",
  "overall": 3,
  "reviewTime": "891717742"
```

Fig. 3. Shows the snapshot of Amazon Review Dataset

### RESULTS

We have performed experiment by taking Amazon review dataset. It is of numerical type only. Snapshot of dataset is presented in fig.3. The experiment is conducted to predict the amount of spam and non-spam users present in any particular website which floods spam messages while reviewing product.

The Fig. 4. Presents the output of blocked users with their ID which are predicted to be a spammer. Our algorithm efficiently takes the set of unlabeled users list and their reviews and classifies them in the spammer and non-spammer category. Fig. 5. Shows the number of spam and non-spam users from the set of unlabeled set U.

```
User with ID - 1 is Blocked
User with ID - 6 is Blocked
User with ID - 7 is Blocked
User with ID - 10 is Blocked
User with ID - 11 is Blocked
User with ID - 13 is Blocked
User with ID - 18 is Blocked
User with ID - 40 is Blocked
User with ID - 41 is Blocked
User with ID - 54 is Blocked
User with ID - 56 is Blocked
User with ID - 57 is Blocked
User with ID - 58 is Blocked
User with ID - 60 is Blocked
User with ID - 74 is Blocked
User with ID - 77 is Blocked
User with ID - 80 is Blocked
User with ID - 81 is Blocked
User with ID - 83 is Blocked
```

Fig. 4. Shows the snapshot of blocked users

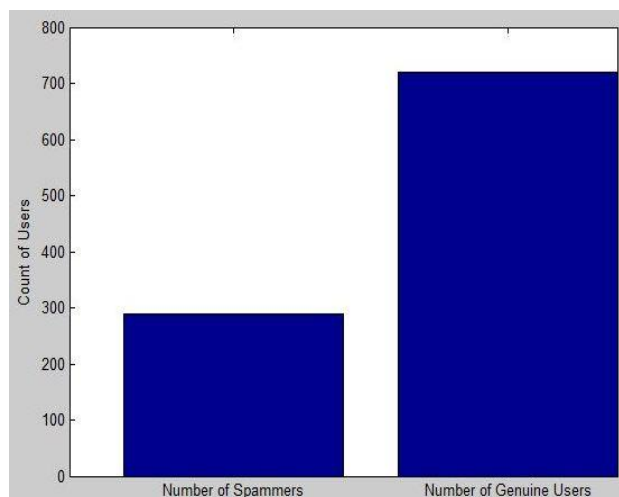


Fig. 5. Shows the snapshot of number of spammer and non-spammers from the U set

## CONCLUSION

This paper basically provides the solution for the problem of singleton review spam detection, which is both challenging and important to solve. This paper presents feature extraction and selection method through which the relevant features are selected used for classification of spam reviews. The experiment is conducted on the MATLAB software. The Maximum Expectation algorithm is considered and evaluated the performance of our framework. Our framework blocked 300 spam users out of 1000+ users. The rest 700 users are detected as genuine user's shows in fig.5.

## REFERENCES

- Jindal N, Liu B. Analyzing and detecting review spam. InData Mining, 2007. ICDM 2007. Seventh IEEE International Conference on 2007 Oct 28 (pp. 547-552). IEEE.
- Jindal N, Liu B. Review spam detection. InProceedings of the 16th international conference on World Wide Web 2007 May 8 (pp. 1189-1190). ACM.
- Algur SP, Patil AP, Hiremath PS, Shivashankar S. Conceptual level similarity measure based review spam detection. InSignal and Image Processing (ICSIP), 2010 International Conference on 2010 Dec 15 (pp. 416-423). IEEE.
- Lai CL, Xu KQ, Lau RY, Li Y, Jing L. Toward a language modeling approach for consumer review spam detection. Ine-Business Engineering (ICEBE), 2010 IEEE 7th International Conference on 2010 Nov 10 (pp. 1-8). IEEE.
- Lau RY, Liao SY, Kwok RC, Xu K, Xia Y, Li Y. Text mining and probabilistic language modeling for online review spam detection. ACM Transactions on Management Information Systems (TMIS). 2011 Dec 1;2(4):25.
- Jindal N, Liu B. Review spam detection. In Proceedings of the 16th International Conference on World Wide Web, 2007;1189–1190.
- Karami A, Zhou B. Online review spam detection by new linguistic features. iConference 2015 Proceedings. 2015 Mar 15.
- Mukherjee A, Liu B, Glance N. Spotting fake reviewer groups in consumer reviews. InProceedings of the 21st international conference on World Wide Web 2012 Apr 16 (pp. 191-200). ACM.
- Ott M, Choi Y, Cardie C, Hancock JT. Finding deceptive opinion spam by any stretch of the imagination. InProceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1 2011 Jun 19 (pp. 309-319). Association for Computational Linguistics.
- Fei G, Mukherjee A, Liu B, Hsu M, Castellanos M, Ghosh R. Exploiting Burstiness in Reviews for Review Spammer Detection. ICWSM. 2013 Jul 8;13:175-84.
- Xie S, Wang G, Lin S, Yu PS. Review spam detection via temporal pattern discovery. InProceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining 2012 Aug 12 (pp. 823-831). ACM.
- Wang G, Xie S, Liu B, Yu PS. Identify online store review spammers via social review graph. ACM Transactions on Intelligent Systems and Technology (TIST). 2012 Sep 1;3(4):61.
- Hu M, Liu B. Mining and summarizing customer reviews. InProceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining 2004 Aug 22 (pp. 168-177). ACM.
- Sahu S, Dongre B, Vadhvani R. Web Spam Detection Using Different. IJSCE, 2011; 1(3).
- Liu Y, Jin J, Ji P, Harding JA, Fung RY. Identifying helpful online reviews: a product

- designer's perspective. *Computer-Aided Design*. 2013 Feb 28;45(2):180-94.
16. Lai CL, Xu KQ, Lau RY, Li Y, Song D. High-order concept associations mining and inferential language modeling for online review spam detection. In *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on* 2010 Dec 13 (pp. 1120-1127). IEEE.
  17. Zheleva E, Kolcz A, Getoor L. Trusting spam reporters: A reporter-based reputation system for email filtering. *ACM Transactions on Information Systems (TOIS)*. 2008 Dec 1;27(1):3.
  18. Martinez-Romo J, Araujo L. Web spam identification through language model analysis. In *Proceedings of the 5th international workshop on adversarial information retrieval on the web 2009* Apr 21 (pp. 21-28). ACM.
  19. Jindal N, Liu B. Review spam detection. In *Proceedings of the 16th International Conference on World Wide Web, 2007*; 1189–1190.
  20. Cormack GV, Gómez Hidalgo JM, Sáenz EP. Spam filtering for short messages. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management 2007* Nov 6 (pp. 313-320). ACM.
  21. Cormack GV, Lynam TR. Online supervised spam filter evaluation. *ACM Transactions on Information Systems (TOIS)*. 2007 Jul 1;25(3):11.