International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X; pISSN:2321-2241 Volume: 14; Issue: 8; August- 2025

NET SPAM: A NETWORK-BASED SPAM DETECTION FRAMEWORK FOR REVIEWS IN ONLINE SOCIAL MEDIA

Karuna S H
PG,Student
Dept. of MCA
The Oxford College of Engineering,
Bommanahalli, Bengaluru- 560068
karunash037@gmail.com

Mridula Shukla
Assistant Professor
Dept. of MCA
The Oxford College of Engineering,
Bommanahalli, Bengaluru- 560068
mridulatewari@theoxford.edu

ABSTRACT

Ratings of online enterprises are an essential component as regards making decisions of action in e-commerce and the spent social media longitudinal. They also find themselves falling into the grips of the spammers who post shallow or even false opinions despite them being able to be used by the customers to make their judgments on the products and services. The spam reviews divert trust, hurt an honest business and defraud purchasers. Existing detection mechanisms are more likely to adopt text analysis or naive behavioral tests and will lack any counter to intelligent spammers, who train behaviors that are similar to those of legitimate humans. The research paper presents as a solution to the problem the idea of applicability to Net-Spam, a model of network based spam detection whose reviews, users, and features are modeled as a heterogeneous information network being linked to each other. In contrast to the traditional frameworks, Net Spam adopts the restoratively weight process to limit emphasis on the strongest indicators of the spam, such as the patters of user and review behavior. Most accurate and most efficient Spam configures between Yelp and Amazon, and the Amazon and Yelp. Another facility that was presented in the framework is flexibility, in the forms of supervised, semi supervised, and unsupervised types of learning.

KEYWORDS: *Online enterprises, Intelligent spammers, Legitimate humans, Heterogeneous Information Network.*

INTRODUCTION

Internet review has become one of the most trusted form of information by customers that are carrying out a purchase decision in the modern era of digital media. When purchasing new product with Amazon, where to eat in a certain place, people rely on the opinions of others. Hotel or accommodation to stay or restaurant or diner to eat in. This system results in the

online sites susceptible to abuse, although it provides greater authority to the consumers. Bad actors like spammers and malicious users can abuse such review systems with false and/or misleading reviews in order to: To make higher sales of bad quality commodity or to undermine the reputation of the contending firm. This kind of deceiving activity reduces the trustworthiness of online market places besides misleading buyers.

The common approaches to spam detection might be based on text level analysis or very basic behavioral checks which would tend to fail in some instances against clever spammers who attempt to pass themselves off as real users. These kinds of models rarely possess the characteristics that make some characteristics better at identifying spam than others. In order to solve this problem, the model named Net Spam displays a net-based type, attributes trait-weighing approach, in order to establish the leading indications first. The concept in reading user actions and what they also state in reviews in mixed info webs is attributed to using cell phone webs in linking mixed info webs. There is greater power, size, and bend in obtaining spam reviews as there are many online locations connected to spam posts.

LITERATURE SURVEY

Spam reviews identification is the research question that is active and the outcome is a taking several directions. Previous studies have concentrated mainly on the linguistic component of analysis such that the spam reviews would be described here as a diversionary style of writing, sub strain in expressive or homogenous proportion of text. The methods though useful to a certain degree can be circumnavigated when the spammers go to the efforts of exactly what ordinary use of language is. To sidestep these limitations, the researchers turned to behavioural models, and they used the frequencies of review, variation in ratings, and time of post. There were always such types of solutions which could not identify planned. The more recently, network spam runs have occurred but they are necessitated by mechanisms that histories user activities, are representatives of users, products and capture reviews in graph structures.

Such models help derive relationships and connections that is eluded by textual- or behavioral approaches. Nevertheless, not all of them distinguish between the various features, and this lowers their efficiency. The Net Spam framework is based on these principles because it couples heterogeneous information networks with feature-weighting scheme. It is superior when it comes to determining the efficiency of detection since it produces high weights on highly discriminative values of features. Such harmonious combination provides Net Spam with greater versatility and consistency than previous models because this embraces deficiencies in linguistic and behavioral approaches.

EXISTING WORK

Spam reviews identification is the research question that is active and the outcome is a taking several directions. Previous studies have concentrated mainly on the linguistic component of analysis such that the spam reviews would be described here as a diversionary style of writing, sub strain in expressive or homogenous proportion of text. The methods though useful to a certain degree can be circumnavigated when the spammers go to the efforts of exactly what ordinary use of language is. To sidestep these limitations, the researchers turned to behavioural models, and they used the frequencies of review, variation in ratings, and time of post. There were always such types of solutions which could not identify planned. The more recently, network spam runs have occurred but they are necessitated by mechanisms that histories user activities, are representatives of users, products and capture reviews in graph structures.

PROPOSED SYSTEM

The novel system (Net Spam) would be better than the allowances of the older systems. The paper contributes to the area of networks since it also provides a network based approach on spam detection. As opposed to solely considering the reviews, it includes the summation of the relationships between users, reviews, products and Heterogeneous Information Network (HIN). Such representation also discovers the presupposed correlations which remain hidden and may pass unnoticed, namely, within the literal discourse a qualitative approach to behaviour modeling.

The procedure starts by gathering and pre-cleaning reviews offered on the Internet. Then features are pulled out in four areas: review-behavioral (e.g. frequency of posts, average ratings deviation), user-behavioral (e.g. account activity, variety of reviews), review-linguistic (e.g. sentiment, exaggerations), and user-linguistic (e.g. writing style consistency). These are characteristics that constitute the building blocks of the HIN. One of the most important innovations of Net Spam is its feature-weighting mechanism, that places more weight (importance) on attributes contributing most toward correct detection. Net Spam uses only the most discriminative signals and unlike the traditional models, Net Spam assigns unequal weights to individual features. The system is flexible and can work in supervised, semi-supervised, and unsupervised mode, therefore, is suitable to be applied in practice to use-cases when the labeled data is limited. Such portability will make the Net Spam to have a wide applicability in numerous platforms in online settings.

METHODOLOGY

Net Spam methodology tries to achieve the systematic detection and classification of spam reviews in an accurate and scalable manner. This is done in an orderly manner through these steps: data collection,

preprocessing, feature extraction, network modeling, feature weighting and classification. The initial studies focused more on text-based approaches in which frequencies of words, sentiment strength and style are investigated. These methods helped to determine articles that are written in a bad or exaggerated manner, but they could not be applied when spammers used the natural and heterogeneous expressions. The behavioral-based approaches were than used and centered on the presence of reviews, bursts, and discrepancies in ratings. Whereas these were used to detect abnormality, they in most cases required a lot of metadata and could not stand a strategic poster in the center over time, particularly spammers. Furthermore, a second type of models (network-based) emerged to model the links between user-cum-product-and-reviews. Such models recognized unseen groups of spammers but all features received equal weight and they were not as accurate. An analysis of frequency and diversity of reviews is one of the user- behavioral features. Review-Linguistic features revolve around sentiment polarity, exaggerations, and stylistic indications, whereas user-linguistic feature identifies the consistency of writing in more than one post. The features are then embedded on a Heterogeneous Information Network (HIN) wherein users, reviews, and products are framed as nodes, and their linkages are depicted as edges. By using this structure one can have the system detect hidden patterns and relationships. The mechanism of feature weighting is essential since it ranks the most discriminative features with increased weights, and classifications can focus on what is significant with reduced willingness to adhere to noise. Lastly, categorization is done in supervised, semi-supervised or unsupervised environments subject to the presence of labeled data. The approach makes Net Spam flexible, efficient, and applicable in the various real life settings.

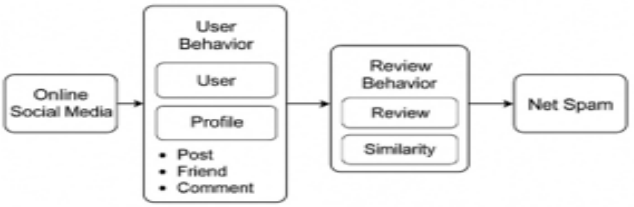

Fig 1. Architecture of Net Spam Framework

Task	Task Name	Status
1	Requirement Analysis & Feasibility Study	Done
2	Design of Net Spam System Architecture & Context Diagram	Done
3	Implementation of Feature Extraction & HIN Modeling	Done
4	Integration of Feature Weighting & Classifier Module	Done
5	Final Deployment & Documentation	Done

Table 1. Net spam development tasks.

EXPERIMENTAL RESULTS

In order to compare the effectiveness of the anticipated Net Spam structure, tests were performed on benchmark levels associated with Amazon and Yelp where both involve a combination of authentic reviews with spam-labeled reviews. It was to measure literacy, performance and success of the feature-weighting mechanism against standard models. The results showed that Net Spam was always superior to the present text based and behavioural models suggestions. Among its key contributions was the finding out that review-behavioral characteristics, such as bursts in posting periods were done. The deviations of the rating in relation to the average rating, proved to be the most powerful factors influencing survey of importance is also such user-behavioral features, in particular, when it is essential to make considerations in the case of their necessity. The estimation of the patterns of the account activity. The linguistic characteristics had a marginal role but it was helpful when combined with others characteristics. The other striking note was that the framework performed reasonably with young labeled data. Net Spam could work well in a semi-supervised scenario (labels were only given to 5 percent of the reviews), producing accuracy levels near those of fully supervised models. This shows its flexibility to real-world platforms where labeled data is considered small. In sum, the experiments showed that Net Spam outperformed on accuracy, time complexity, and robust classifications and thus is a viable and scale-up solution to detecting spam in online reviews.

Fig 2. A Social Media Marketing Interface.

Fig 3. Admin Login for reviewing Spam Reviews.

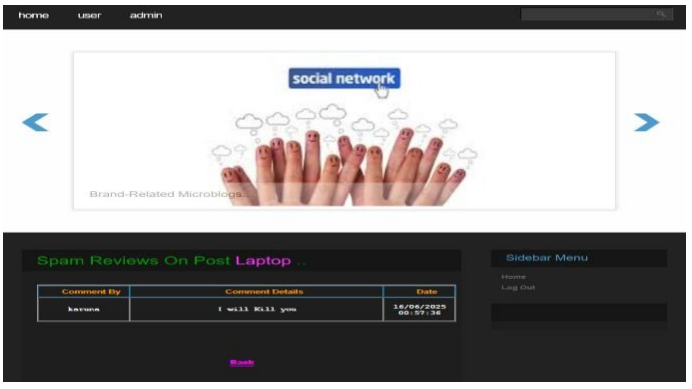

Fig 4. Categorized spam reviews on various online products

CONCLUSION

The paper has introduced Net Spam, a network-based system to find spam reviews in online social media websites. Compared with the previous models based on a homogeneous information network, in which all features are given equal attention, the proposed approach adds a feature-weighting scheme on a heterogeneous information network, which helps distinguish more discriminative signals. The accuracy of the review-behavioral and user-behavioral survey experiments on Amazon and Yelp data sets proved the same properties, however, this time it was the ability to distinguish the real reviews and the ones with spam. The flexibility of the framework is inherent, and it excels where all supervised and semi-supervised and, in addition to the supervised, the unsupervised and un-supervised modality are concerned making it highly adaptive in a large sphere in the real world. Besides the increased precision, it is possible to notice that the interpretability is improved in that visualisation functions are required to allow the administrators to differentiate either suspicious accounts and surveying trends. In short, by ensuring that online platforms like Net Spam are reliable, scalable and the mechanisms to detect spam are customizable, credence of such platforms is being stimulated. Bings can be in the form of on-line malicious statements, cross platform abuse anomaly analysis and This is the detection of spammers in the community.

REFERENCES

[1] N. Jindal and B. Liu, “Opinion spam and analysis,” in Proc. 1st Int. Conf. Web Search Data Mining (WSDM), 2008, pp. 219–230.

[2] M. Ott, C. Cardie, and J. T. Hancock, “Finding deceptive opinion spam by any stretch of the imagination,” in Proc. 49th Annu. Meeting Assoc. Comput. Linguistics (ACL), 2011, pp. 309–319.

[3] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, “Exploiting burstiness in reviews for review spammer detection,” in Proc. 7th Int. AAAI Conf. Web Social Media (ICWSM), 2013, pp. 175–184.

[4] L. Akoglu, R. Chandy, and C. Faloutsos, “Opinion fraud detection in online reviews by network effects,” in Proc. 7th Int. AAAI Conf. Web Social Media (ICWSM), 2013, pp. 1–10.

[5] G. Wang, S. Xie, B. Liu, and P. S. Yu, “Review graph based online store review spammer detection,” in Proc. IEEE 11th Int. Conf. Data Mining (ICDM), 2011, pp. 1242–1247.

[6] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, “What Yelp fake review filter might be doing?,” in Proc. 7th Int. AAAI Conf. Web Social Media (ICWSM), 2013, pp. 40