



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 4, April 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

A Survey on Product Recommendation System using Sentiment Analysis and Semantic Analysis based on Machine Learning

Amol Pawale¹, Rohit Patil², Om Chavan³, Ajit Ravan⁴, Ayesha Sayyad⁵

B.E. Student Department of Information and Technology, Trinity College of Engineering and Research,
Pune, India^{1,2,3,4}

Assistant Professor, Department of Information and Technology, Trinity College of Engineering and Research,
Pune, India⁵

ABSTRACT: People often rely on a product on the basis of its reviews and ratings. People can remove the update, allowing spam senders to do spam studies about the assets, in addition to directing different benefits. Identifying these fake reviewers and spam content is a major research issue, despite the way much research has been done. Up till now, the ways set hardly differentiate spam reviews, and no one shows the significance of every property type. In this study, a framework, developed by NetSpam, uses spam features to display updated data sets as separate databases to design spam detection into a problem group for these networks. Using the criticality of spam features helps us obtain good outcomes regarding different metrics on review data sets. When a client asks a research question, it will display all the n-no of items, not just a suggestion of an item.

KEYWORDS: product, review, sentiment analysis, machine learning.

I. INTRODUCTION

Nowadays, the world is seeing more support for the electronic trade, where online reviews are taking on a significant role. The customer is currently involved in the pursuit of product and service reviews when deciding what to buy or where to get it. Spam analysts have used this lucky opportunity to write risky reviews to damage reputable stores or use fraudulent polls to lure customers into low-quality items. This is often viewed as a spam review. These spam updates represent a real risk to web-based businesses, with people, organizations, links, and organisations losing all their valuable resources. Customer feedback also plays an important role in determining purchasing options. A big part of the customer these days is posting their experiences on online journals, site reviews, web-based business sites, and personal contact sites. The above data is consumed by businesses or corporate organizations, as they are committed to assessing customer perceptions about their assets, management, and support. As people buy things after reading surveys, the kind of reviews the object gets worries sellers. This means that a positive survey on an item will find deals and a negative one will reduce them.

The Spam Recovery Framework describes a method that uses the concept of various data systems in classifying spam surveys. The system is built with research databases and is managed to define metadata-based metadata with regard to updates. Spam highlights are defined in the framework. The general significance of each element is resolved by utilising a weighting strategy. The element esteems are calculated using a likelihood-based methodology based on total dispersion work. Slightly controlled reading is also used, and naming is done by taking common objects that may be present.

In this work, a distributed approach is proposed to improve the effectiveness of spam detection using emotional analysis and semantic analysis in the Amazon product review database. As a result, the set of updated data falls under a lot of information, and as a result, it will be difficult to process surveys and separate spam updates into large survey databases that use spam detection.

II. RELATED WORK

Shigang Hu and others [1] This research work augments heuristic-driven user interest profiling with reviewer credibility analysis and fine-grained feature sentiment analysis to devise a robust recommendation methodology. The proposed Integrity, Interest, and Advanced Model (CISER) model consists of five modules, which are a candidate feature release feature, reviewer loyalty analysis, user interest mining, candidate feature allocation, and a

recommendation module. The Update corpus is provided as an input to the CISER model. The learner feature output module uses context and emotional confidence to extract key features. To make our model stronger in reviews and against counterfeit and inappropriate reviewers, reviewing the reviewer's credibility provides a way to link expertise, trust, and influence scores with reviewers to gauge their opinions in terms of their credibility. Consumer mining module users use the beauty of review writing as an interest mining heuristic pattern. The candidate feature sentiment assignment module scores candidate features present in the review based on their fast text sentiment polarity.

Dematis, E. Karapistoli and others [2] suggest a method that combines content and usage information to obtain forged product reviews. The proposed model uses both product reviews and the behavioural features of reviewers linked to specific spam indicators. In this paper, the detection of fine-grained explosive patterns is used to better evaluate updates produced in "suspicious" periods. The reviewer's previous review history is also used to determine the "author's" reputation of the reviewer as an indicator of the level of integrity of their recent updates.

S. Zhou and others [3] adopted a comprehensive data analysis approach to investigate the impact of online customer reviews on customer performance and product performance. The authors developed a semantic keyword similarity method based on a single value to measure customer intelligence using large customer review documents and product release notes. Using a mobile app data set with over 3 million online reviews, our in-depth research finds that the review volume has a close relationship with customer technology. In addition, customer speed has a curvilinear relationship with product performance. This research contributes to the design literature by demonstrating the influence of firms' ability to use online customer reviews and their impact on product performance. It also helps to consolidate the contradictions found in the literature relating to the relationship between the three structures.

H. Li and others [4] found that postal reviewers also follow an interesting distribution pattern, which has never been reported before. That is, shipping costs are bimodal. Multiple spam senders also often share reviews of the same set of products over a short period of time, which we call combusting, and the author finds some interesting patterns in the short-term power of individual reviewers and their behaviour in concert with other reviewers. Encouraged by these findings, the authors first proposed a Markov model with a dual-coded mode so that the spam model would use only the update times for each reviewer. The authors then expanded on the Coupled Hidden Markov Model to capture both the reviewer's post-behaviour and cohesive signals.

M. Hazim and others [5] Aimed to improve spam detection in the marketplace for mobile applications, this study recommends the use of features based on modified development methods such as Extreme Gradient Boost (XGBoost) and General Enhanced Depression Model (GBM) for testing multi-language data sets (i.e., English and Malay). From the tests conducted, it was found that XGBoost is best suited for finding spam ideas in the English database, while GBM Gaussian is best suited for the Malay database. Comparative analysis also shows that the implementation of the features based on the proposed statistics achieved an acquisition accuracy level of 87.43 percent in the English database and 86.13 percent in the Malay database.

N. Kumar and others [6] a novel guided teaching method to increase the chances of finding confusion by analysing a few aspects of users' behaviour and identifying their behaviour as a group in a cohesive manner. Clearly, the user characteristics of the author's model and their interactions, such as consistent distribution and multidisciplinary diversity, They then compile this stream using a few supervised reading strategies, such as retrieval, vector support, and neighbours near k that provide solid meta separators. The authors conduct a detailed examination of the methods and developed the technical details. This approach is of interest to online business forums because it can help reduce false reviews and increase consumer confidence in the reliability of their online knowledge. This research contributes to the literature by integrating feature distribution features into machine learning strategies, which can improve the performance of artificial retrieval detectors on digital platforms.

N. Hussain and others [7] identified different performance metrics that are often used to test the accuracy of models for receiving spam reviews. Finally, this activity presents a general discussion of the different methods of extracting features from the updated data sets, the proposed tax for spam review acquisition methods, test ratings, and publicly available data sets. Future research posts and guidelines on the field of spam review findings are also presented. This study has shown that the success factors of any way to get spam reviews are interdependent. Feature releases depend on the update database, and the accuracy of the spam detection review process depends on the selection of the engineering method feature. Therefore, in order to successfully use the spam review acquisition model and achieve better accuracy, these factors need to be considered in conjunction with each other.

C. Pandey and D.S. Rajpoot [8] Today, online reviews play an important role in customer decisions. From buying a shirt to an e-commerce store to a restaurant, online reviews have become a staple of choice. However, people are always busy as they do not have time to pay attention to the internal details of products and services, so the reliance on online reviews has been raised. Because of their reliance on online reviews, some individuals and organisations are arrogantly creating spam updates to promote or discredit a person, product or organization. Therefore, it is impossible to pinpoint whether the review is spam or ham in the flesh, and it is impossible to distinguish all reviews by hand. Therefore, the spiral cuckoo search based clustering method has been introduced to obtain spam reviews. The proposed method uses cuckoo search power and Fermat spiral to solve the problem of cuckoo search method integration. The effectiveness of the proposed method was tested on four spam databases and one Twitter spam set.

R. Narayan and others [9] Today, with the growing popularity of the internet, online marketing will become very popular. This is because many products and services are readily available online. Therefore, reviews of all of these products and services are very important for customers and organizations. Unfortunately, driven by the will of profit or promotion, fraudsters are often used to produce false reviews. These false reviews, written by fraudsters, prevent customers and organisations from reaching real conclusions about the products. This false review or spam review should be detected and eliminated in order to prevent potential fraudulent customers. In this paper, we used a supervised reading method to detect spam reviews. The proposed work uses a set of different features and emotional points to construct models, and their performance was assessed using different categories.

R. Barbado, O. Araque, and C. A. Iglesias [10] propose a feature framework to detect false reviews tested in the consumer electronics domain. Contributions are fourfold: (i) the creation of a database for classifying false reviews in the consumer electronics domain in four different cities based on scratching techniques, (ii) a description of the framework for the discovery of false reviews, and (iv) the assessment and analysis of the results of each unread town.

R. Ghai and others [11] whose reference shows that a proposed review process is proposed. Some restrictions have been raised for review assistance. These parameters indicate the difference between one update and another, thus increasing the chances of spam. The method presented classifies a review as useful or unhelpful depending on the school given the update.

Problem Statement:

To develop a product recommendation system based on real reviews and identify spammers and spam content using reviews, and to perform four categories of features, including review-behavioural, user-behavioural, review-linguistic, and user-linguistic features for showing only trusted reviews to the user.

Motivation:

1. People rely on written reviews in their decision-making processes, and positive or negative reviews encouraging or discouraging them in their selection of products and services.
2. Anyone can access the application through registration and give feedback as reviews for spammers to guide other user's opinion.

Objectives:

- Spam and spam detection and a different type of analysis of this topic.
- Written reviews also help service providers improve the quality of their products and services.
- Identify a spam user by using positive and negative reviews on online forums.
- Only show trusted updates to users.

Existing Result

- Approaches to the state of the art to generate recommendations positive evaluations are often based on sentiment analysis using data mining.
- However, they suffer from low accuracy.

Proposed System:

- The proposed new framework consists of representing a review data set of product recommendations provided as HIN and resolving the issue of spam detection in the HIN classification problem.

- In particular, to demonstrate their database ideas such as HIN, where reviews are linked to different types of nodes (such as working with users). Then a measuring algorithm is used to calculate the value (or weight) of each task.
- These weights are used to calculate the latest review labels using both monitored and unsupervised processes.
- Based on our observations, to describe the two aspects of the review (user-review and language-behaviour), features that are classified as “behavioural reviews” are more weighty and produce better performance in detecting spam reviews in both unchecked and unchecked methods.
- The feature weights can be added or subtracted for labelling, and that is why complexity can be measured with a certain degree of accuracy.
- Dividing features into four main categories (behaviour-review, user behaviour, language-review, and language usage) helps us to understand how each category contributes to spam detection.

System Architecture:

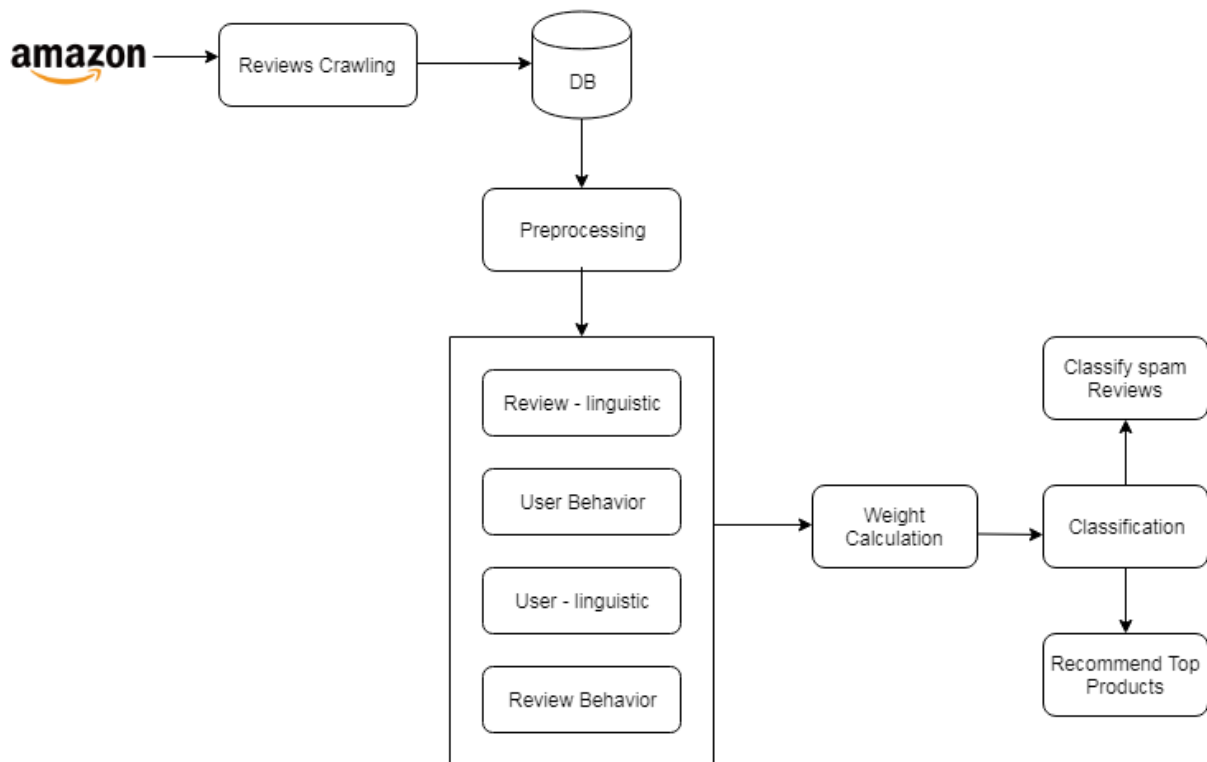


Fig. System Architecture

III. CONCLUSION

This proposed system investigation presents a novel spam detection system, in particular the NetSpam framework for product reviews based on sentiment analysis (SA) and latent semantic analysis (LSA). This paper uses SA and LSA with the NetSpam algorithm for spam detection. Our ideas seem to suggest that computer-generated weights using this metapath concept can be very effective in distinguishing spam reviews and reporting higher performance. Additionally, NetSpam can last longer, produce better performance in the process of maximising highlights, and perform higher performance than ever before, with just a small number of great features. In addition, in the wake of defining four primary classifications, highlighting our ideas shows that a review of ethical categories is warranted. LSA is used in the proposed system to reduce similar comments and try to improve spam detection accuracy. The results also show that using different supervisions, as opposed to the semi-supervised technique, has no discernible effect on deciding the vast majority of the weighted highlights, as in other datasets.

REFERENCES

1. Shigang Hu, Akshi Kumar, Fad-Al-Turjman, Shivam Gupta, Simran Seth, Shubham, "Reviewer Credibility and Sentiment Analysis Based User Profile Modelling for Online Product Recommendation," vol. 8, pp. 26172–26186, Jan. 2020.
2. Dematis, E. Karapistoli, and A. Vakali, "Fake review detection via exploitation of spam indicators and reviewer behaviour characteristics," in Proc. Int. Conf. Current Trends Theory Pract. Inform. Cham, Switzerland: Edizioni Della Normale, 2019, pp. 581–595.
3. S. Zhou, Z. Qiao, Q. Du, G. A. Wang, W. Fan, and X. Yan, "Measuring customer agility from online reviews using big data text analytics," J. Manage. Inf. Syst., vol. 35, no. 2, pp. 510–539, Apr. 2018.
4. H. Li, G. Fei, S. Wang, B. Liu, W. Shao, A. Mukherjee, and J. Shao, "Bimodal distribution and co-bursting in review spam detection," in Proceedings of the 26th International World Wide Web Conference (WWW), 2019, pp. 1063–1072.
5. M. Hazim, N. B. Anuar, M. F. A. Razak, and N. A. Abdullah, "Detecting opinion spam using a supervised boosting approach," PLoS ONE, vol. 13, no. 6, 2019, e0198884.
6. N. Kumar, D. Venugopal, L. Qiu, and S. Kumar, "Detecting review manipulation on online platforms using hierarchical supervised learning," J. Manage. Inf. Syst., vol. 35, no. 1, Jan. 2018, pp. 350–380.
7. N. Hussain, H. Turab Mirza, G. Rasool, I. Hussain, and M. Kaleem, "Spam review detection techniques: A systematic literature review," Appl. Sci., vol. 9, no. 5, p. 987, 2019.
8. C. Pandey and D. S. Rajpoot, "Spam review detection using spiral cuckoo search clustering method," Evol. Intell., vol. 12, no. 2, pp. 147–164, Jun. 2019.
9. R. Narayan, J. K. Rout, and S. K. Jena, "Review spam detection using opinion mining," in Progress in Intelligent Computing Techniques: Theory, Practice, and Applications. Singapore: Springer, 2018, pp. 273–279.
10. R. Barbado, O. Araque, and C. A. Iglesias, "A framework for fake review detection in online consumer electronics retailers," Inf. Process. Manage., vol. 56, no. 4, pp. 1234–1244, Jul. 2019.
11. R. Ghai, S. Kumar, and A. C. Pandey, "Spam detection using rating and review processing method," in Smart Innovations in Communication and Computational Sciences. Singapore: Springer, 2019, pp. 189–198.
12. M. Xu and S. Liu, "Semantic-enhanced and context-aware hybrid collaborative filtering for event recommendation in event-based social networks," IEEE Access, vol. 7, pp. 17493–17502, 2019.
13. A. Da'u and N. Salim, "Sentiment-Aware Deep Recommender System With Neural Attention Networks," IEEE Access, vol. 7, pp. 45472–45484, 2019.
14. S. Reddy, S. Nalluri, S. Kuniseti, S. Ashok, and B. Venkatesh, "Contentbased movie recommendation system using genre correlation," in Smart Intelligent Computing and Applications, S. C. Satapathy, V. Bhateja, and S. Das, Eds. Singapore: Springer, 2019, pp. 391–397.
15. V. Srivastava and A. D. Kalro, "Enhancing the helpfulness of online consumer reviews: The role of latent (content) factors," J. Interact. Marketing, vol. 48, pp. 33–50, Nov. 2019.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details