



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 5, May 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

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

Amol Pawale, Ajit Ravan, Om Chavan , Rohit Patil, Prof.Ayesha Sayyad

Department of Information Technology College Name Trinity College of Engineering and Research, Pune, India

ABSTRACT: People often rely on products based on reviews and ratings. People can remove updates, which allows spammers to conduct property spam investigations and receive a variety of benefits. Despite a lot of research being done, identifying these fake reviewers and spam content still remains an important research challenge. Traditionally, spam reviews have been difficult to distinguish, and no one has proven the importance of each type of attribute. In this study, we use the Net Spam framework to represent the updated dataset as a separate database to predict spam detection with problem groups in these networks. You can use the spam severity feature to get good results from multiple survey dataset metrics. When a client submits an investigation request, the system displays all of them, not just the suggestion. Keywords: products, reviews, sentiment analysis, machine learning.

I. INTRODUCTION

Global support for e-commerce is increasing as online reviews play an increasingly important role. When determining what to buy or where to buy today, customers check for product and service reviews. Spam analysers use this opportunity to write risky reviews that harm reputable stores or use fake surveys to entice shoppers to buy low-quality items. This is often considered a spam review. Spam updates pose a huge threat to web companies as individuals, organizations, links, and organizations lose all their valuable assets. Customer feedback is very important when making an alternative purchase decision. Most customer experiences today are documented in online diaries, site reviews, business websites, and personal contact sites. Businesses and corporate organizations use the above information to evaluate customers' experiences with assets, management and support. Product reviews worry sellers because people read reviews and buy things. In other words, if the product reviews are positive, offers are found, and if the reviews are bad, deals are reduced. The

Spam Recovery Framework describes how to classify a survey as spam using the concepts of various data systems. The system consists of a research database supported to define metadata based on metadata about changes. The main points of spam are defined by the framework. To evaluate the overall importance of each factor, use a weighting approach. Item scores are determined based on the total variance task using a likelihood-based methodology. A slightly controlled read is also used, and naming is done using common objects that may exist.

This article proposes a distributed strategy for improving spam detection performance in the Amazon Product Review Database using sentiment and semantic analysis.

II. RELATED WORK

Shigang Hu and others [1] This study combines heuristic-driven user interest profiling with reviewer credibility analysis and fine-grained feature sentiment analysis to produce a robust recommendation way. 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. Update corpus is provided with input into the CISER model. The learner feature output module extracts key features based on context and emotional confidence. To make our model stronger in reviews and against counterfeit and inappropriate reviewers, reviewing a reviewer's credibility allows you to link their knowledge, trust, and impact scores with their opinions to measure their credibility. The attractiveness of review writing is used as an interest mining heuristic pattern by users of the consumer mining module. 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] propose a method for obtaining fabricated product reviews that incorporates content and use data. The proposed model uses both product reviews and the behavioural features of reviewers linked to specific spam indicators. The detection of fine-grained explosive patterns is employed in this paper to improve the

evaluation of updates delivered during "suspect" times. The "author's" character of the critic is also utilised to judge the quantum of integrity of their current updates grounded on the critic's former review history.

S. Zhou and colleagues [3] used a rigorous data analytic technique to study the impact of online customer evaluations on customer and product performance. Using huge customer review documents and product release notes, the authors devised a semantic keyword similarity method based on a single value to quantify customer intellect. 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, client 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 aids in the consolidation of the literature's discrepancies regarding the link between the three structures.

Postal reviewers, according to H. Li and others [4], follow an intriguing distribution pattern that has never been recorded before. To put it another way, 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 extended the Coupled Hidden Markov Model to include the reviewer's post-behaviour as well as coherent signals.

M. Hazim and co-workers [5] set out to improve spam detection in mobile apps. This research suggests that traits 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). According to the results of the experiments, XGBoost is best suited for identifying spam ideas in the English database, while GBM Gaussian is best suited for the Malay database. A comparative analysis also reveals that the features based on the provided statistics were implemented with a high level of acquisition accuracy of 87.43% in the English database and 86.13% 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 study contributes to the literature by incorporating feature distribution capabilities into machine learning strategies, which can improve the performance of artificial search detectors on digital platforms.

N. Hussain et al. and others [7] identified various performance metrics frequently used to test the accuracy of spam feedback models. Finally, this walkthrough provides a general discussion of different methods for extracting features from updated datasets, proposed taxes on how to gather spam feedback, test scores, and public datasets. Future research publications and proposals based on the spam scan results are also under discussion. This study showed that the success factors of how to receive spam reviews are interdependent. Feature releases depend on the update database, and the accuracy of the spam detection testing process depends on the engineering method feature selection. Consequently, to successfully apply the spam feedback model and increase its accuracy, these factors must be tested together.

C. Pandey and D.S. Rajpoot [8] Today, online reviews play an important role in customer decision making. From shirt shopping to online stores to restaurants, online reviews have become a staple of choice. However, people have become more reliant on online reviews because they are usually busy and do not have time to pay attention to complex items and services. Because of our reliance on online reviews, some people and organizations arrogantly create spam updates to promote or discredit people, products, or organizations. As a result, it is impossible to determine whether a review is spam or ham, and it is also impossible to manually distinguish all reviews. The proposed method uses the cuckoo search ability and Fermat's spiral to solve the cuckoo search method integration problem. The proposed method was tested on four spam databases and one spam collection on Twitter.

R. Narayan and others. [9] With the development of the Internet today, Internet marketing will become popular. This is because many products and services are readily available online. As a result, feedback from consumers and organizations on all these products and services is critical. Unfortunately, scammers for profit or promotion are often used to write bad reviews. These false reviews written by scammers prevent customers and organizations from drawing real conclusions about the product. These false or spam reviews should be identified and removed to avoid fraudulent customers. In this article, we used a controlled-read method to detect spam reviews. The suggested work constructs models utilising a variety of characteristics and emotive points, and their performance is evaluated using various categories.

R. Barbados, O. Araque and C.A. Iglesias [10] proposes a functional framework for false positive detection tested in the field of consumer electronics. The contribution consists of four parts: (i) creating a database to classify false consumer electronics reviews in 4 different cities based on scratch method, (ii) explaining false reviews detection based, (iv) evaluating and analysing each result. Unread city.

R. Ghaian and others. [11], that reference indicates that the proposed review process was proposed. Some restrictions have been introduced to aid validation. This value shows inconsistencies between one update and another, increasing the likelihood of spam. The presented method categorizes reviews as useful or not useful depending on the school that received the update.

Proposed System:

- The proposed new structure consists of presenting a survey data set with product recommendations presented as HINs and solving the spam detection problem in HIN classification questions.
- In particular, it shows an idea of a database such as a HIN whose outlines are linked to different types of nodes (e.g users and tasks). It then uses a measurement algorithm to calculate the value (or weight) of each task.
- These weights are used in both supervised and unsupervised algorithms to generate up-to-date view representations.
- According to our findings, the feature marked as “behavioural review” is to detect spam reviews using both untested and unvalidated methods when describing two elements of a review (user feedback and verbal behaviour). more weights and better performance.
- Complexity can be measured with some precision because feature weights can be added or subtracted for labelling purposes.
- Divide the features into four main categories (behavioural checks, user actions, language checks and language usage) to help you understand how each category contributes to spam detection.

System Architecture:

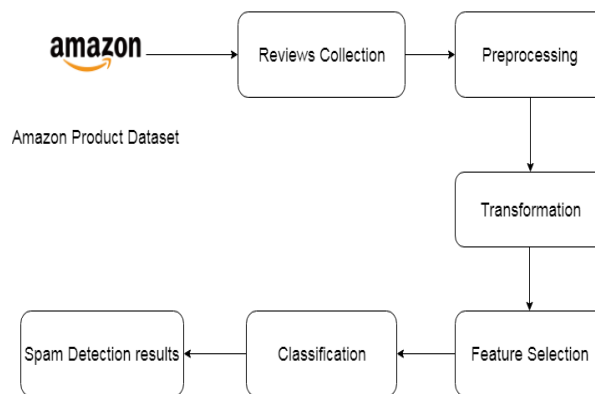


Fig. System Architecture

Algorithm:

Algorithm I: Spam review detection using behavioral features method

Input: review R . $T=0.5, 0.55, 0.6$

//threshold value for labelling the review

Output: Spam or Not-Spam

for each review R_1 in review dataset do

// behavior features ($F_1, F_2, F_3, \dots, F_{13}$)

for each behavior feature F_i calculate normalize value do

// variable V is calculating normalize value of F

V_i = calculate normalize value F_i

Sum += V_i

```

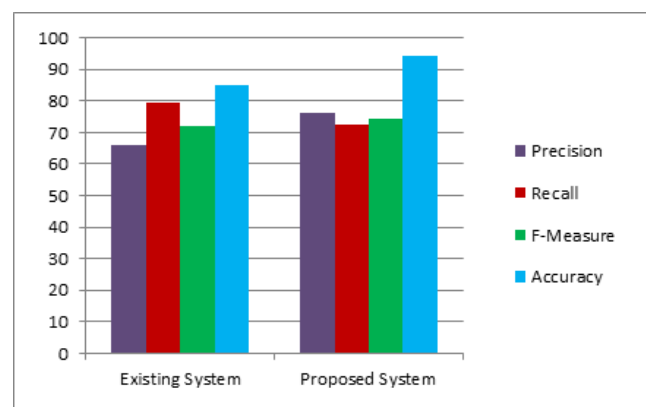
end for
// calculating average score
Average Score=Sum/13
for each value Vi do
// calculating drop score
DropScore = (Sum - V)/12
if Average Score - DropScore>=0.05 then
    assign weight Wi 2
    Total Weight += 2
else
    assign weight Wi 1
    Total Weight +=1
end if
end for
for each value Vi do
// calculating total spam score
Score += W; * Vi
end for
Spam Score = Score / Total Weight
if Spam Score > then
    label Ri ← Spam
else
    label Ri ← Not-Spam
end if
end for

```

III. RESULT AND DISCUSSION

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and Jdk 1.8. The application is web application used tool for design code in Eclipse and execute on Tomcat server.

Fig 2.Performance Analysis between existing and proposed system



The proposed SpamDup framework time complexity is $O(e^2 n)$. The SpamDup framework accuracy is 94.06\% which is better than SPaglePlus Algorithm accuracy is 85.14\% on using product dataset.

IV. 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 computergenerated weights using this meta path 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. The proposed system uses LSA to reduce the number of these comments and increase the accuracy of spam detection. The results also show that, in contrast to the partial observation method, the use of another observation method does not have as appreciable effect on the decision of the majority of weighted emphasis as it does with the 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. 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