IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

Fraud Apps Detection using Sentiment Analysis and Spam Filtering

Priyanka Rekhawar[1], Ketki Shinde[2], Sakshi Shinde[3], Prajkta Shelke[4], Prof. S. P. Sneha Vanjari[5]
Students, Department of Information Technology[1,2,3,4]
Professor, Department of Information Technology[5]
Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Abstract: *In the mobile app industry, ranking fraud is the practise of engaging in dishonest or deceitful behaviour with the intention of artificially boosting an App's position on a popularity list. In fact, ranking fraud by app developers is becoming more and more common. These practises include inflating their apps' sales or uploading fake app reviews. Although the significance of preventing ranking fraud has long been understood, little knowledge and research have been done in this field. In order to do this, we present a comprehensive analysis of fraud app detection using sentiment analysis and spam filtering in this study and suggest a system for detecting it in mobile apps. We specifically suggest mining the active times, or leading sessions, of mobile Apps to precisely locate the ranking scam in the first place.*

Keywords: Mobile Apps, Fraud Detection, Rating and Review, sentiment analysis, spam filtering

I. INTRODUCTION

Over the past few years, the number of smartphone apps has increased at an astounding rate. For instance, the Apple App Store and Google Play each had more than 1.6 million Apps available as of the end of April 2013. Many App shops created daily App leaderboards, which show the chart rankings of most popular Apps, to encourage the creation of mobile Apps. The App leader board is undoubtedly one of the most crucial tools for promoting mobile apps. As a result, in order to have their apps rated as highly as possible in such App leader boards, App developers frequently investigate various strategies, such as advertising campaigns, to promote their apps. However, as a recent trend, unethical App developers turn to various fraudulent techniques to purposefully raise their Apps and ultimately influence the chart positions on an App store rather than depending on conventional marketing strategies. This is typically accomplished by deploying "bot farms" or "human water armies" to quickly inflate the number of App downloads, ratings, and reviews.

II. RELATED WORK

Detailed The pair wise features are first explicitly utilized to detect group colluders in online product review spam campaigns, which can reveal collusions in spam campaigns from a more fine-grained perspective. In [1] paper, Spam campaigns spotted in popular product review websites (e.g., amazon. com) have attracted mounting attention from both industry and academia, where a group of online posters are hired to collaboratively craft deceptive reviews for some target products. The goal is to manipulate perceived reputations of the targets for their best interests.

In [2] paper, Online product reviews have become an important source of user opinions. Due to profit or fame, imposters have been writing deceptive or fake reviews to promote and/or to demote some target products or services. Such imposters are called review spammers. In the past few years, several approaches have been proposed to deal with the problem. In this work, take a different approach, which exploits the burrstones nature of reviews to identify review spammers.

In [3] paper, Online reviews on products and services can be very useful for customers, but they need to be protected from manipulation. So far, most studies have focused on analyzing online reviews from a single hosting site. How could one leverage information from multiple review hosting sites? This is the key question in our work. In response, develop a systematic methodology to merge, compare, and evaluate reviews from multiple hosting sites. focus on hotel reviews and use more than 15million reviews from more than3.5million users spanning three prominent travel sites.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

In [4] paper, Users increasingly rely on crowd sourced information, such as reviews on Yelp and Amazon, and liked post sand ads on Facebook. This has lento market for black hat promotion techniques via fake (e.g., Sybil) and compromised accounts, and collusion networks. Existing approaches to detect such behavior relies mostly on supervised (or semi-supervised) learning over known (or hypothesized) attacks. They are unable to detect attacks missed by the operator while labeling, or when the attacker changes strategy.

In [5] paper, Online reviews have become an increasingly important resource for decision making and product designing. But reviews systems are often targeted by opinion spamming. Although fake review detection has been studied by researchers for years using supervised learning, ground truth of large scale datasets is still unavailable and most of existing approaches of supervised learning are based on pseudo fake reviews rather than real fake reviews. Working with Dianping1, the largest Chinese review hosting site, present the first reported work on fake review detection in Chinese with filtered reviews from Damping's fake review detection system.

In [6] paper, Online reviews are quickly becoming one of the most important sources of information for consumers on various products and services. With their increased importance, there exists an increased opportunity for spammers or unethical business owners to create false reviews in order to artificially promote their goods and services or smear those of their competitors. In response to this growing problem, there have been many studies on the most effective ways of detecting review spam using various machine learning algorithms. One common thread in most of these studies is the conversion of reviews to word vectors, which can potentially result in hundreds of thousands of features.

In [7] paper, it providing an efficient and effective method to identify review spammers by incorporating social relations based on two assumptions that people are more likely to consider reviews from those connected with them as trustworthy, and review spammers are less likely to maintain a large relationship network with normal users. The contributions of this paper are two-fold: (1) elaborate how social relationships can be incorporated into review rating prediction and propose a trust based rating prediction model using proximity as trust weight; and (2) design a trust-aware detection model based on rating variance which iteratively calculates user-specific overall trustworthiness scores as the indicator for spam city.

In [8] paper, to detect fake reviews for a product by using the text and rating property from a review. In short, the proposed system (ICF++) will measure the honesty value of a review, the trustiness value of the reviewers and the reliability value of a product. The honesty value of a review will be measured by utilizing the text mining and opinion mining techniques. The result from the experiment shows that the proposed system has a better accuracy compared with the result from iterative computation framework (ICF) method.

In [9] paper, Online Social Networks (OSNs), which captures the structure and dynamics of person-to-person and person-to-technology interaction, is being used for various purposes such as business, education, telemarketing, medical, entertainment. This technology also opens the door for unlawful activities. Detecting anomalies, in this new perspective of social life that articulates and reflects the off-line relationships, is an important factor as they could be a sign of a significant problem or carrying useful information for the analyzer.

In [10] paper, they propose a new holistic approach called SpEagle that utilizes clues from all metadata (text, timestamp, and rating) as well as relational data (network), and harness them collectively under a unified system to spot suspicious users and reviews, as well as products targeted by spam. SpEagle employs a review-network-based classification task which accepts prior knowledge on the class distribution of the nodes, estimated from metadata. Positive points are: It enables seamless integration of labeled data when available. It is extremely efficient.

In [11] this paper, mangoes are graded in four types like Green Mango, Yellow Mango and Red Mango which are based on machine learning method. This system considers RGB values size and shape of mangoes. Following analysis is used to obtain good probability. This helps to train system to identify appropriate maturity of mangoes. This research is conducted on two machine learning method i.e. Naive Byes and SVM (Support Vector Machine).

III. PROPOSED SYSTEM

Lot of work has been done in this field because of its extensive usage and applications. In this section, some of the approaches which have been implemented to achieve the same purpose are mentioned. These works are majorly differentiated by the algorithm for fraud app detection.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

3.1 System Architecture

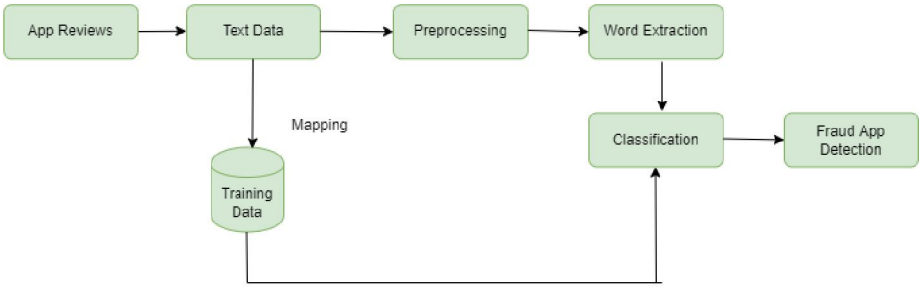

Fig. System Architecture

3.2 Algorithm

Naïve Bayes:

Step 1: Convert the data set into a frequency.

Step 2: Create Likelihood table by finding the probabilities like Overcast probability = 0.29 and probability of playing is 0.64.

Step 3: Now, use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

For example:

Problem: Players will play if weather is sunny. Is this statement is correct?

We can solve it using above discussed method of posterior probability.

P(Yes | Sunny) = P(Sunny | Yes) * P(Yes) / P (Sunny)

Here we have P (Sunny |Yes) = 3/9 = 0.33, P(Sunny) = 5/14 = 0.36, P(Yes)= 9/14 = 0.64

Now, P (Yes | Sunny) = 0.33 * 0.64 / 0.36 = 0.60, which has higher probability.

Naive Bayes uses a similar method to predict the probability of different class based on various attributes. This algorithm is mostly used in text classification and with problems

4) Performance Evaluation: The evaluation can be done based on following factors:

i) Performance matrices such as TPR FPR Precision Recall etc.

ii) Impact of Different Sampling method

iii) Investigation of time related data

3.3 Spam Detection

Algorithm 1: Spam review detection using behavioral features method

Input: review R_i , τ = 0.5,0.55,0.6 //threshold value for labelling the review
Output: Spam or Not-Spam

```
for each review Ri in review dataset do
    // behavior features (F1, F2, F3, ..., F13)
    for each behavior feature Fi calculate normalize value do
        // variable V is calculating normalize value of F
        Vi = calculate normalize value Fi
        Sum += Vi
    end for
    // calculating average score
        Average Score = Sum / 13
    for each value Vi do
        // calculating drop score
        DropScore = (Sum - Vi) / 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 += Wi * Vi
    end for
    Spam Score = Score / Total Weight
    if Spam Score >= τ then
        label Ri ◄— Spam
    else
        label Ri ◄— Not-Spam
    end if
end for
```

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

IV. RESULTS AND DISSCUSSION

Experimental evaluation is done to compare the proposed system with the existing system for evaluating the performance. The simulation platform used is built using Java framework (version jdk 8) on Windows platform. The system does not require any specific hardware to run; any standard machine is capable of running the application.

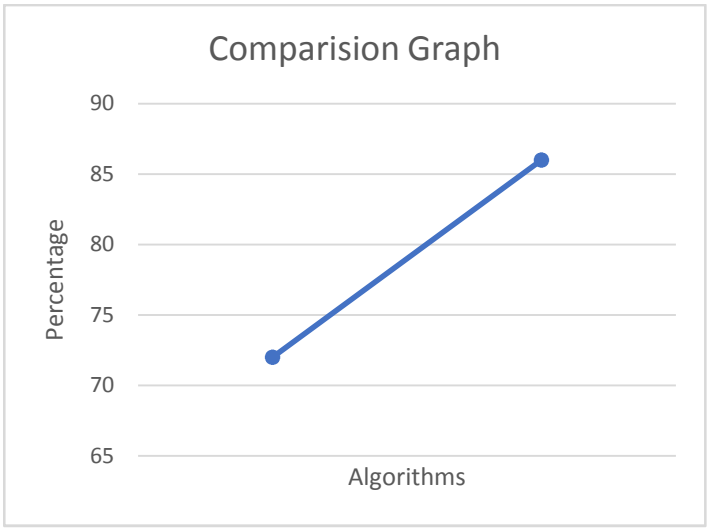

Figure 2. comparison Graph

Sr. No.	Existing System	Proposed System
Algorithm	Binary Support vector machine	Naïve Bayes
Precision	60.2%	65.4 %
Recall	85.5%	89.7%
Accuracy	Prediction Accuracy:72%	Prediction Accuracy:86%

V. CONCLUSION

In this study, we created a fraud app detection system for mobile apps. In more detail, we first demonstrated how leading sessions were the source of ranking fraud and offered a technique for mining leading sessions from each App's historical ranking records using sentiment analysis and spam detection. Then, for detecting fraud apps, we identified evidence based on review.

REFERENCES

[1]. Ch. Xu and J. Zhang," Combating product review spam campaigns via multiple heterogeneous pairwise features", In SIAM International Conference on Data Mining, 2014.
[2]. G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Exploiting bustiness in reviews for review spammer detection", In ICWSM, 2013.
[3]. A. j. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos, "True view: Harnessing the power of multiple review sites",In ACM WWW, 2015.
[4]. B. Viswanath, M. Ahmad Bashir, M. Crovella, S. Guah, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, "Towards detecting anomalous user behavior in online social networks", In USENIX, 2014.
[5]. H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao," Spotting fakereviews via collective PU learning", In ICDM, 2014.
[6]. M. Crawford, T. M. Khoshgoftaar, and J. D. Prusa," Reducing Feature Set Explosion to Faciliate Real-World Review Sapm Detection", In Proceeding of 29th International Florida Artificial Intelligence Research Society Conference, 2016.
[7]. H. Xue, F. Li, H. Seo, and R. Pluretti," Trust-Aware Review Spam Detection",IEEE Trustcom/ISPA.,2015.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

[8]. E. D. Wahyuni , A. Djunaidy," Fake Review Detection From a ProductReview Using Modified Method of Iterative Computation Framework", In Proceeding MATEC Web of Conferences, 2016.
[9]. R. Hassanzadeh," Anomaly Detection in Online Social Networks: Using Datamining Techniques and Fuzzy Logic", Queensland University of Technology, Nov, 2014.
[10] R. Shebuti, L. Akoglu," Collective opinion spam detection: bridging review networks and metadata", In ACM KDD, 2015.
[11] G.D. Upadhye, D.Pise,"Grading of Harvested Mangoes Quality and Maturity Based on Machine Learning Techniques",IEEE International conference on smart city and Emerging Technology,2018.