

Discovery of Ranking Fraud for Mobile Apps

Sakshi Shetty¹, Harsh Gathe², Sheetal Bhosale³, Nitin Nautiyal⁴, Prof. Arpita Raut⁵

^{1,2,3,4}Department of Information Technology Engineering, Vidyalkar Institute of Technology, Mumbai

⁵Asst.Professor, Department of Information Technology Engineering, Vidyalkar Institute of Technology, Mumbai

Abstract- Deceptive activities in the app are done in order to appear on the top ranking list of apps. Nowadays, many App developers use shady means such as increasing their app sales or posting phony app ratings to commit ranking fraud. Preventing the ranking fraud has nowadays become an important criteria but there is less research done in this particular area. We provide a view of fraud ranking in mobile apps and propose a fraud app detection system. In this system, initially main process is based upon data preprocessing. In which various methods like stop word filtering, stemming, removal of URLs, non-dictionary word removal and also POStagging (Parts Of Speech tagging) are included. There are three types of evidences, 1) ranking based evidences 2) rating based evidences and 3) review based evidences. We propose an evidence aggregation method to integrate all the evidences for fraud detection. Finally, we show the apps which are fraud or genuine based on the polarity of the reviews.

Key Words: Fraud, POStagging, Data Preprocessing, Evidence aggregation method

1. INTRODUCTION

The mobile apps are growing rapidly and everyday many apps are getting launched as well as many apps are getting closed. There are many fraudulent apps which can completely damage mobile phones or can gain access to device maliciously. There are many apps supporting many operating system such as android and iOS. As apps are increasing day by day, many new apps are launched everyday so it gets difficult for the users to select the best app, so many App stores developers are launching daily App leader boards, which shows the ranking of various popular Apps. The App leader board is one of the most important ways for detecting whether that app is legitimate or not. A higher rank on the leader board usually shows that a large number of downloads have taken place and that app developer is getting millions of dollars in revenue. Many App developers try to find out various ways like advertising to promote their Apps so that their Apps is ranked as high as possible, on the leader board. Instead of applying traditional marketing techniques, false App developers have started fraudulent apps to increase their position that is getting on top of the leaderboard of apps and ultimately manipulating the rankings at the store.

Malicious manipulation of user generated data is done to impudent the usage patterns of the data. Review manipulation is one of the most important viewpoint of ranking fraud for mobile apps. Specifically, before downloading or purchasing a new mobile App, users often read its historical reviews and an App that contains more positive reviews and ratings may attract more number of users to download it. Therefore Review evidence can be extracted. Once an App is published, it can be rated by anyone whoever downloads it. User rating is also one of the most important characteristic feature of an App advertisement. Higher rating may attract more number of users to download it and can also be ranked higher in the popularity list of apps. A higher rank usually leads to more

number of downloads and million dollars in the revenue. Thus, rating manipulation is an important viewpoint of ranking fraud. So, it is important to develop a system for fraud app detection. We propose a simple yet effective system to classify the apps based on predefined set of features.

2. LITERATURE SURVEY

[1] Hengshu Zhu, Hui Xiong, Yong Ge, and Enhong Chen proposed a system for Discovery of Ranking Fraud for mobile apps. They did two main steps for mining leading sessions. Firstly, they discovered leading events from the App's historical ranking records. Then they merged adjacent leading events for constructing leading sessions. After this they collected all types of evidences and performed evidence aggregation. Finally, they used gradient based approach with exponentiated updating so as to learn the weight of evidences.

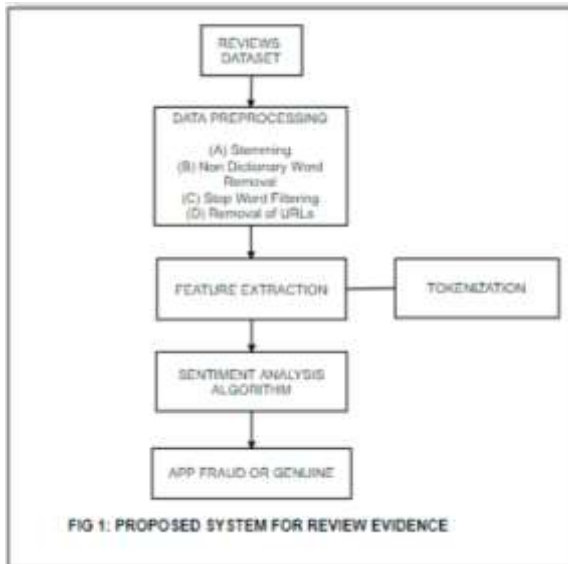
[2] Prof. Mariappan.R, Ms. Deivanai.A, Ms. B. Hema, Ms. Thamizharasi.V have studied and researched about ranking, rating and review using statistical hypotheses test. They proposed an unsupervised approach based on fraud similarity to combine these evidences.

[3] Monika Pandey, Prof. Tripti Sharma proposed a Fraud App Detection system using Fuzzy Logic model based on Sentiment Analysis. Initially, they did data pre-processing such as stemming, removal of URLs etc. Then, they collected nouns, verbs and adjective to get sentiment score. Finally they used fuzzy classification algorithm so as to detect the app as genuine or fraud.

[4] Ms. Manasi Mhatre, Ms. Surabhi Mhatre, Ms. Dhikshashri Dhemre, Prof. Saroja.T.V proposed the system for Detection of Ranking Fraud in mobile applications. Firstly, they collected all evidences such as ranking, rating and reviews.

Then, they performed Naïve Bayes Classification algorithm using predefined set of features.

3. PROPOSED SYSTEM



STEPWISE PROCEDURE:

1)Reviews Data Collection.

2)Data Preprocessing.

a) Stop word filtering- The stop word doesn't convey any sentiment, eg: is, what, when, where etc. This will be used to remove stop words from review.

b) Stemming-It will be used to remove suffix and prefix from word eg :ies ,ing ,tion ,etc.

c) Non Dictionary word will be removed as it can lead to wrong sentiment.

d) URLs will also be removed because it does not contain any sentiment.

3)Feature Extraction: In this phase, the 3 dictionary are used for sentiment score calculation. The nouns, verbs, adjective are collected and tokenize in order to get sentiment score.

4)POS tagging: POS tagging (Parts Of Speech tagging) is a supervised learning solution that uses features like the previous word, next word. It basically extracts the noun, verbs, adjective etc.

5)Sentiment Analysis: This technique takes a string and returns the sentiment rating for positive, negative and neutral based on sentiment polarity. For this purpose, we tend to use classifiers in Sentiment Analysis to see the subjective value of text document i.e. positive, negative or neutral.

4. METHODOLOGY

This system contains 2 major modules which is given as follows:

1. Admin

a. Manage Apps:

i. The system allows admin to add new apps with details of the app (name, image, URL, description, category etc.)

ii. This system also helps to view the app.

b. Detection of Fraud apps:

i. Admin will view the ratings and reviews given by the user.

ii. Admin will then analyze the ratings and reviews which will then help him/her to decide whether the app is fraud or genuine.

c. Manage Users:

i. The admin can view the details of the user who has installed the app.

ii. He can also add new user.

d. Manage App Provider:

i. The admin can add new app providers.

ii. He can also view the details of the App Provider.

2. User

a. Registration / Login:

i. To view the mobile apps, user has to first create an account by filling the registration form.

ii. After successfully creating an account, user can now login into the system using login credentials such as username and password.

b. View App Details:

i. The user can view the application's details, ratings and comments.

ii. The user can also download the app after viewing the app details.

c. Give Reviews:

i. After using the application, user can give their feedback based on the experience while using that app.

Proposed Algorithm:

- 1) Create an input file "sample-input.txt" containing 1 or more sentences to check the polarities.
- 2) The file can be read with the help of a file reader object.
- 3) Each sentence or text is given a token with the help of POSTagger.
- 4) POS tagger assigns a tag (JJ or JJS) to each token.
- 5) Check the tag of each token if the tag is "JJ" or "JJS" (i.e. the tagged token is an adjective/opinion word) then pass this word in SentiWordnet to check the sentiment score as well as polarity of that particular word.
- 6) SentiWordNet will return the sentiment type of that word (eg: positive, weak_positive, strong_positive, negative, strong_negative, neutral etc.) based on the sentiment score.
- 7) Count the no of positive(pos_count) and no of negative (neg_count) adjectives for each sentence and the ratings given by the user and remove aggregate for each app

[estimable(J,3)] "may be computed or estimated" with the help of adjective estimable, that has an Obj score of 1:0 (and Pos and Neg scores of 0.0), while the synset [estimable(J,1)] has a Pos score of 0:75, a Neg score of 0:0, and an Obj score of 0:25.

SentiWordNet scores are calculated as positive and negative terms and are used to determine sentiment orientation by assigning the document to the class with the highest score.

Score	Positive	Negative	Objective
	Adjectives		
0.0	65.77%	62.81%	0.08%
0.125	12.12%	7.32%	2.14%
0.25	8.81%	8.68%	7.42%
0.375	4.85%	5.19%	11.73%
0.5	3.74%	5.63%	9.50%
0.625	2.94%	5.53%	7.65%
0.75	1.28%	3.72%	9.21%
0.875	0.47%	1.07%	7.57%
1.0	0.03%	0.04%	44.71%

5. CONCLUSION

We proposed a fraud app detection system for mobile Apps. We showed that the system provided a method for mining leading sessions for each App from its historical ranking records. Then, on the basis of the historical ranking records, we identified ranking, rating and review based evidences, for detecting the fraud apps. Moreover, we proposed an evidence based aggregation method to combine all types of evidences for evaluating the ranking of leading sessions for mobile Apps. Finally, we proposed a system with extensive experiments on reviews data. In the future, we plan to study more efficient and effective fraud evidences and analyse the rating, review and rankings. Moreover, we will extend our fraud app detection approach with other mobile App, for enhancing the user experience.

6. FUTURE SCOPE

In future, we plan to study more efficient and effective fraud evidences and analyse the rating, review and rankings. Also, we will study more fraud evidence techniques. Moreover, we will extend our fraud app detection approach with other mobile App, for enhancing the user experience. We plan to study more efficient and effective fraud evidences and analyse the rating, review and rankings.

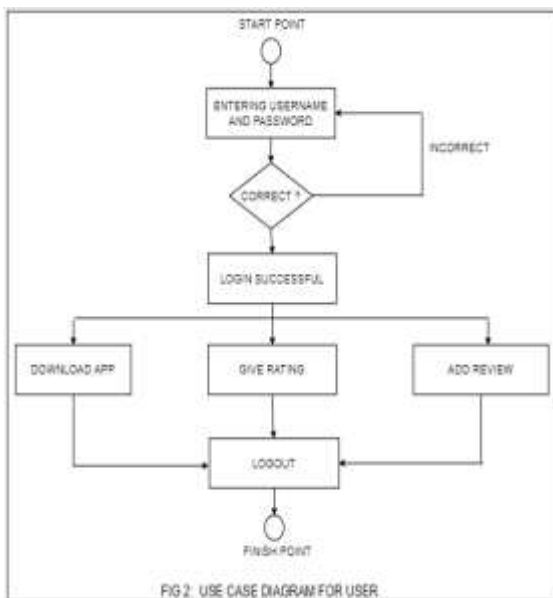


FIG 2. USE CASE DIAGRAM FOR USER

SentiWordNet

SentiWordNet is the result of all the synsets of WordNet according to the sentiment such as "positivity", "negativity", and "neutrality". Each synset is associated to three numerical scores Pos(s), Neg(s), and Obj(s) which indicate how positive, negative, and "objective" (i.e., neutral) the terms contained in the synset are. For example, in SENTIWORDNET 1.0 the synset

7. REFERENCES

- [1] Hengshu Zhu, Hui Xiong, Yong Ge, and Enhong Chen "Discovery of Ranking Fraud for Mobile Apps", IEEE Transactions On Knowledge And Data Engineering, Vol. 27, No. 1, January 2015
- [2] Prof.Mariappan.R*, Ms.Deivanai.A, Ms.B.Hema, Ms.Thamizharasi.V, "Investgation of Ranking Rating and Review using Statistical Hypotheses Tests", International Journal Of Engineering Sciences & Research Technology, April, 2015
- [3] Monika Pandey, Prof. Tripti Sharma, "Fraud App Detection using Fuzzy Logic Model Based on Sentiment of Reviews", International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 09 | Sep 2018
- [4] Ms.Manasi Mhatre, Ms. Surabhi Mhatre, Ms. Dhikshashri Dhemre, Prof.Saroja.T.V, "Detection of Ranking Fraud in Mobile Applications", International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 03 | Mar-2018