Detection of fraud in ranking in mobile applications using Data Mining

Mohit Bangroo[1], Govind Nair[2], Bhushan Dhodi[3]

Abstract- *Rating fraud in the mobile App market refers to fraudulent or deceptive activities which have a purpose of bumping up the Apps in the popularity list. Indeed, it becomes more and more frequent for App developers to use shady means, such as inflating their Apps' sales or posting phony App ratings, to commit ranking fraud. To this end, in this paper, we provide a holistic view of ranking fraud and propose a ranking fraud detection system for mobile App. Furthermore, we investigate three types an evidences, ranking based evidences, rating based evidences and review based evidences, by modeling Apps' ranking, rating and review behaviors. In addition, we propose an optimization based aggregation method to integrate Al the evidences for fraud detection. In the experiments, we validate the effectiveness of the proposal system, and show the scalability of the detection algorithm as well as some regularity of ranking fraud activities*

Keywords- ranking fraud, genuine, comment

I. INTRODUCTION

Most of us use android and IOS Mobiles these days and also uses the play store or app store capability normally. These app stores helps user to download any application of his/her choice anytime. For the convenience of users many app stores demonstrate chart rankings of popular apps. Higher ranking on the charts usually leads to increase in download of the application and in turn generate revenue for the developer.

Both the stores provide great number of application but unluckily few of those applications are fraud. Such application does damage to phone and also may be data thefts. Hence, such applications must be marked, so that they will be identifiable for store users. So we are proposing a web application which will process the information, comments and the review of the application. So it will be easier to decide which application is fraud or not. Multiple applications can be processed at a time with the web application. Also User cannot always get correct or true reviews about the product on internet. So rating/comments will be judged by the admin and it would be easy for admin to predict the application as Genuine or Fraud.

II. RELATED WORK

The previous research papers that are related o the fraud detection of mobile application are divided into different sections, they are:

- **Web Ranking Spam Detection:**

The web ranking spam refers to any intentional actions which bring to selected webpage an inexcusable auspicious relevant importance.

- **Online Review Spam:**

Online reviews are often the primary factor in a customer's decision to purchase a product or service, and are a valuable source of information that can be used to determine public opinion on these products or services.

III. REVIEW BASED EVIDENCE USING SENTIMENTANALYSIS

Sentiment Analysis refers to the practice of applying Natural Language Processing (NLP) and Text Analysis techniques to identify and extract subjective information from a piece of text. A person's opinion or feelings are for the most part subjective and not facts. We can check for user's sentimental comments on multiple applications. The reviews may be fake or genuine. But after comparing reviews of admin as well as user's, we can get more clear idea.

Hence, we can get higher probability of getting real reviews. So we are proposing a system to develop a web application that will take reviews from registered users for single product, and analyze them for positive negative rating. For every user reviews and comments will be fetched separately and analyzed for positive negative rating.The user comment is a long string. The easiest way to generate features from text is to split the text up into words. Each word in a review will then be a feature that we can then work with. In order to do this, we'll split the reviews based on whitespace. We'll then count up how many times each word occurs in the negative reviews, and how many times each word occurs in the positive reviews. This will allow us to eventually compute the probabilities of a new review belonging to each class. In Review Based Evidences, besides ratings, most of the App

stores also allow users to write some textual comments as App reviews. Such reviews can reflect the personal perceptions and usage experiences of existing users for particular mobile Apps. Indeed, review manipulation is one of the mast important perspectives of App ranking fraud.

IV. RATING BASED EVIDENCE USING NAÏVE BAYES

Bayesian classifiers are statistical classifiers. They can predict class membership probabilities, such as the probability that a given sample belongs to a particular class. Bayesian classifier is based on Bayes' theorem. Naive Bayesian classifiers assume that the effect of an attribute value on a given classis independent of the values of the other attributes. This assumption is called class conditional independence. It is made to simplify the computation involved and, in this sense, is considered "naïve".

According to Bayes' theorem, the probability that we want to
compute P(H|X) can be expressed in terms of probabilities P(H),

P(X|H), and P(X) as

$$P(H|X) = \frac{P(X|H)\ P(H)}{P(X)}$$

$$P(c_i / \mathbf{x}) \quad \frac{P(\mathbf{x}/c_i)P(c_i)}{P(\mathbf{x})} \quad P(\mathbf{x}/c_i)P(c_i)$$
$$\text{for } i \quad 1, 2, \quad , L$$

In the system from the given set of comments, the classifier will predict which of the classes have the higher probability of posterior. For eg from a given sample, Y is predicted to belong to class Ci, if and only if

P(Ci|Y)>P(Cj|Y) for1 _ j _ m, j!= i.

Gaussian distribution with a mean μ and standard deviation _ defined by

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp -\frac{(x-\mu)^2}{2\sigma^2},$$
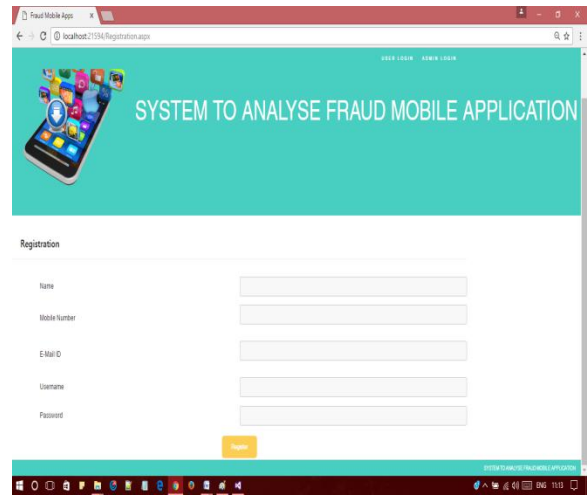
V. SCREENSHOTS
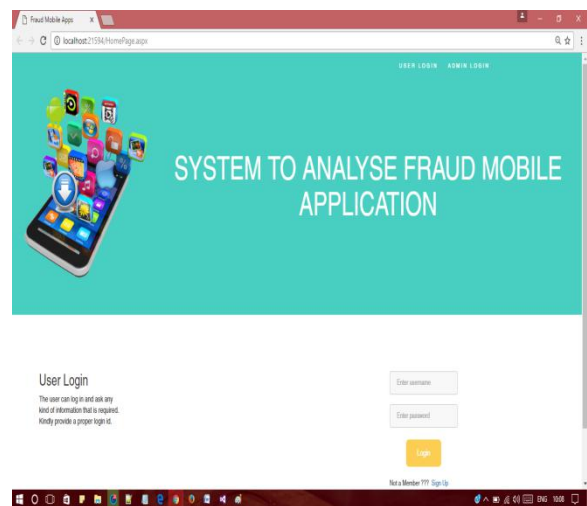

Figure 1. User Login

Figure 1. User Login

VI. FUTURE SCOPE

The proposed framework is scalable and can be extended with other domain generated evidences for ranking fraud detection. Experimental results show the effectiveness of the proposed system, the scalability of the detection algorithm as well as some regularity of ranking fraud activities. To the best of our knowledge, there is no existing benchmark to decide which leading sessions or Apps really contain ranking fraud.

VII. CONCLUSION

We propose a methodology to evaluate the security of Android mobile apps based on sentimental rating, commenting and data mining. In this project we propose the system to detect fraud apps for android and IOS mobiles. In this system we describe such mechanism which will help to detect fraud apps. This system is based on sentiment analysis and data mining. We are proposing a system that helps us to

determine an application's credibility bytaking comments from users on an application which will help admin to analyze them for positive and negative rating.

REFERENCES

[1] D. F. Gleich and L.-h. Lim. Rank aggregation via nuclear norm minimization. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '11, pages 60–68, 2011

[2] Klementiev, D. Roth, and K. Small. An unsupervised learning algorithm for rank aggregation. In Proceedings of the 18th European conference on Machine Learning, ECML '07, pages 616–623, 2007.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan. Lantentdirichlet allocation. Journal of Machine Learning Research, pages 993–1022, 2003.

[4] G. Heinrich. Parameter stimaion for text analysis.Technicalreport,University of Lipzig, 2008.

[5] B. Yan and G. Chen.Appjoy: personalized mobile application discovery. In Proceedings of the 9th international conference on Mobile systems, applications, and services, MobiSys '11, pages 113–126, 2011

[6] Jiawei Han and MichelineKamber, Data Mining: Concepts andTechniques, Elsevier 2006, ISBN 1558609016.

[7] http://en.wikipedia.org/wiki/cohen's kappa.

[8] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw,"Detecting product review spammers using rating behaviors," inProc. 19thACMInt. Conf. Inform. Knowl.Manage., 2010, pp. 939–948