Findings on detection of fraud smartphone apps by ranking using mining leading sessions

Prof A Sai Hanuman¹, P. Sujan Mohan² and Kanegonda Ravi Chythanya³

¹,²Gokaraju Rangaraju Institute of Engineering and Technology, Bachupally, Hyderabad -500090
³Department of CSE, S R University, Warangal, Telangana, India - 506371

E-mail: a_saihanuman@hotmail.com

Abstract. The world as we know now is full of mobile users everywhere and the majority leads with smartphone users which are filled with android and iOS applications. This has produced huge development in smartphone technology at a lightning rate because the apps have become popular and a common concept. For this we are dwelling into data mining in which the concept of mining data from a specific application is tough and pivotal job. These are the methodologies we are using in this paper. We are combining these two methods of mining data and ranking fraud in app store which will be a formidable task. According to report at google play store and apple store there are 2.4 million apps as of May 2019 so that shows quick surge in the past few years. Furthermore, above 500k independent app developers competing for all the end users. Primarily finding fraud apps is done by scanning the top apps ranked apps which will be around 30-40 of them but they are on top for short period of time i.e. few days for verifying their authenticity. However, this can’t be applied when dealing with thousands of apps for each day. Hence, we have to go for an extensive approach in which we can apply a methodology to validate each app. As the importance of mining is established so we need to mine the critical data of app which includes reviews, ratings and rankings of an app that get integrated into an algorithm that can determine the fraud. So, in this paper of discovering the ranking fraud of smartphone app, we need to develop a precise and faultless methodology that gives result of trusted apps.

1. Introduction
In the last few years, a drastic number of all-purpose apps deployed is at a booming figure. The evolution in the smartphone industry has made the smartphone apps very more popular and familiar idea. Main challenges in popular operating system called android are to detect and remove malicious apps. Most of the Application stores launched daily App leaderboard to rank most popular Apps which inspired the development of smartphone Apps. Apps which are on the top list of the leaderboard in turn lead to a very big spike in downloads and crores of rupees in profits. Thus, App developers are computing with different methods like advertising campaign to support and boost Apps to move up the ranks in app store leaderboards. Usually in the market dishonest App programmers for Apps enhancement treacherous methods to consciously enhance apps and its performance level on the leaderboard table in the play store. Ranking fraud has become chief facet in the smartphone app market more so as large quantity of mobile apps are produced. So, a unique answer has to be emerged to identify fraud apps.

The business market has boomed to come up with a greater number of marketing strategies to spike their sales and revenue because of inventing all-around apps for daily uses. The usage of
smartphone mobiles has increased so directly proportional smartphone apps of android and iOS have also risen in greater number. Amongst all this technology the competition in market is cut throat. The fundamental factors of an app are end users’ ratings, reviews and rankings that are used by developers to improve their analytics and faults to satisfy their customers. These factors are taken into consideration by us too after app is downloaded.

In this study of all the papers in this field we noticed a repeat pattern of manipulative methods done manually to forge the performance of app on the chart rankings and in reviews also. Few companies have gone lengths to get false reviews and mendacious methods to pay for teams who give fraud details to the app and submit fraud performance analytics. This has become a new trend among many business marketing of apps. Hence now there are committed teams of people who are doing this deceitful around the clock at all stages of app performance after its deployment. So, to be safe from this wrongfully rated apps we need to know the apps authenticity before installing it that can be achieved with an automated technique. Therefore, an automation system is the solution to verify authentic reviews, ratings and also trustworthy rankings of the app leaderboard.

2. Literature survey

In paper [2], author has proposed a pioneering methodology to calculate the aggregation of rank with matrices which provide precise data and avoid fuzzy data. Proposed methodology solution deals a well-structured matrix-completion problemsubduing spaces of a skew-symmetric matrix. The author has gone all out in introducing advance techniques to this field and in applying them to the algorithm for obtaining matrix completion over skew-symmetric matrices. This enabled to handle the skew-symmetric data for matrix completion by just improving the existing algorithm.

The author in paper [3] came up with a modern proposal to deal with the rank aggregation problem to regenerate a linear combination of ranking function which scans through the authenticity related documents. To overcome the problem author presented an unsupervised learning algorithm called ULARA which upon execution gives linear combination of individual ranking functions as established in the fundamentals of ranking. This proposed algorithm productive success is analysed by two experimental settings that exploits two functions which are synthetic data that computes the performance with spearman’s rank corelation coefficient and a precision filled performance of information recovery data fusion task.

In paper [4], the author has recommended a more mathematical and algorithmic framework to aggregate ranks without man management control. Author designed an EM-based algorithm that demonstrates efficiency for unvarying decomposable distance functions. So proposed framework is used for types of partial rankings but also for instances in which ranking data is of different type.

The author in paper [5] has focused on spammers who have been giving unauthentic reviews on apps and that too asgroups. They have been doing these mistrustful reviews and ratings in a targeted way on the products. So the author has proposed a novel scheme to detect these certain behaviours which are done repeatedly by the spammers on numerous products and have had significant impact on their performance in the market. After analysis the proposed ranking and supervised methods have proven to be potent in identifying spammers.

The study done in paper [6] also deals with online spammers who are attacking with their mistrustful opinions and to attack this the author has proposed a fundamental method called Author Spamicity Model (ASM). Proposed scheme is unprecedented because this is an existing system that uses heuristics or ad-hoc in identifying the spam reviews. The solution mentioned by the author is still not in action by any of the existing methods. The author has mentioned using unsupervised Bayesian inference framework because of which the outcome of both assessments indicate that the proposed model is strong competitor and successful.

A semi-supervised learning system which simplifies both unlabelled and labelled user profiles for multi-class modelling is the proposed scheme in paper [7]. The author has focused on detecting hybrid shilling attack detection and proposed trailblazing scheme called HySAD. The author has put this to test with real world case study on amazon.com with this HySAD method and found intriguing
attacker characteristics which enabled in enhancing the performance of a filtering recommender system.

The problem of unsupervised web spam detection has been studied in paper [8]. The author went after finding the probability of a web page spam and for that initiated the notion of spamicity. Author proposes the use of established methods link spam and term spam detection which need not go through training data set and are cost friendly. Upon execution with real time data set that proposed technique proved to be proficient to capture spam pages.

Table 1. Literature survey

| Sr. no. | Paper Title                                                                 | Proposed                                                                                                                                                                                                 | Advantages                                                                                                                                  | Principal Method                                                                 |
|---------|------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| 1.      | Spotting opinion spammers using behavioral footprints [6]                    | An unsupervised Bayesian inference framework is used to catch sight of opinion spammers. So, a fundamental system of Author Spamicity Model (ASM) is proposed by the author. | The outcome displays exceptional values which are better than the existing methods. The analysis metrics expresses efficiency and is a strong competitor. | Heuristics or ad-hoc label based unsupervised Bayesian inference framework       |
| 2.      | Detecting product review spammers using rating behaviors [5]                 | Author has focused on review spammers who as groups are impacting a product with dishonest reviews. For this the author proposed a novel technique to detect them.     | After executing the process of proposed ranking and supervised methods showed potent success of identifying scammers.                          | The techniques of scoring to compute the degree of spam for each reviewer.         |
| 3.      | Rank aggregation via nuclear norm minimization [2]                          | The technique of rank aggregation is introduced by the author. He proposed to use matrix completion that ignores incomplete and noise data.            | The space of skew-symmetric matrices is subdued by structured matrix completion in the proposed technique                                 | Matrix operations                                                               |
| 4.      | Unsupervised rank aggregation with distance-based models [4]                | Here to achieve automation in aggregating rankings, the author proposed a fundamental mathematical and algorithmic framework.                          | The proposed method easily more efficient compared to the existing system.                                                                  | Augmented permutation and anorthodox decomposable distance function               |
5. An unsupervised learning algorithm for rank aggregation [3]

To deal with the rank aggregation problem the author with improving issues proposes a linear combination of ranking functions to scan the authenticity.

The Ad hoc retrieval systems does data fusion task which values are used as metrics in determining the efficiency of the proposed method.

Linear Combination of Ranking functions

3. Proposed system

The existing method incorporates the results of different evidences to verify the authenticity of an app. This ranking fraud detection method is important job in this field. We again merge it with a recommendation system which will take in the ranking spots gained after execution then result it in a suggested system which will suggest authentic and appropriate apps.

![Evidence aggregation](image)

**Figure 1.** Evidence aggregation

4. Conclusion and future scope

In this study we have showed the overall to identifying fraud of smartphone apps. The prime attributes, the advantages and drawbacks of each system are outlined. The number of all-purpose apps havevery quickly grown over the last few years. End users are attracted by this fake review detection. As popularity rises, smartphones are major target for harmful applications. So, identifying and getting rid of mischievous apps from app stores is the main probe. Main aim of these fraud activities is to get the fraud smartphone apps into the trending list [12-14].

Thus, there is need to have novel system to effectively analyze fraud apps. In proposed system, system performance can be enhanced by adding the recommendation based on the ranking. In Future a study can also be done to make prime analysis of social media campaigning and come up with a detection system for it. Campaigning on social media may be used to gain a particular benefit that may be of business, politics or something else.

5. References

[1] Hengshu Zhu, Hui Xiong, Yong Ge, and Enhong Chen 2003 “Discovery of ranking fraud for smartphone apps” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING 27(1)1289-1305.
[2] D F Gleich and Lh Lim 2011 17th ACM SIGKDD International Conference on Knowledge Discovery and Datamining (KDD) 60–68.
[3] YT Yu and MF Lau 2005 Journal of Systems and Software in Press.
[4] A Klementiev, D Roth, and K Small 2007 The 18th European Conference on Machine Learning (ECML ’07) 616–23.
[5] A Klementiev, D Roth, and K Small 2008 The 25th International Conference on Machine Learning (ICML) 472–79.
[6] EP Lim, VA Nguyen, N Jindal, B Liu, and H W Lauw 2010 The 19th ACM International Conference on Information and Knowledge Management (CIKM ’10) 939–48.
[7] P PramodKumar, S Naresh Kumar and Ch Sandeep, 2019 Journal of Mechanics of Continua and Mathematical Sciences 15(6) 201-12.
[8] A Mukherjee, A Kumar, B Liu, J Wang, M Hsu, M Castellanos, and R Ghosh 2013 The 19th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD ’13).
[9] Ravi Kumar R, Babu Reddy M & Praveen P 2019, International Journal of Scientific and Technology Research 8(12) 2071-74.
[10] H Zhu, H Cao, E Chen, H Xiong, and J Tian 2012 In Proceedings of the 21st ACM international conference on Information and knowledge management (CIKM’12) 1617–21.
[11] Kumar BV, Chanti Y, Yamsani N, Aluvala S and Bhaskar B 2019 International Journal of Recent Technology and Engineering (IJRTE) 8(2S3) 1121-29.
[12] Mahender K, Kumar TA and Ramesh KS 2019 Simple transmit diversity techniques for wireless communications Advances in Intelligent Systems and Computing 669 329-342 10.1007/978-981-10-8968-8_28
[13] Rajasri I, Guptha AVSSKS and Rao YVD 2016 Generation of Egts: Hamming Number Approach Procedia Engineering 144 537-542 10.1016/j.proeng.2016.05.039
[14] Seena Naik K and Sudarshan E 2019 Smart healthcare monitoring system using raspberry Pi on IoT platform ARPN Journal of Engineering and Applied Sciences 14(4) 872-876.