Fuzzy Systems and Soft Computing
ISSN : 1819-4362

UNVEILING RANKING FRAUD IN MOBILE APPS: STRATEGIES FOR DETECTION AND MITIGATION

#1UPPARAPU SRINIVAS,
#2PULKAM NAVEEN SAI,
#3PEDDI KISHOR, *Associate Professor,*
Department of Computer Science and Engineering,
SREE CHAITANYA INSTITUTE OF TECHNOLOGICAL SCIENCES, KARIMNAGAR, TS.

ABSTRACT : The competition for visibility in app stores has intensified due to the proliferation of mobile applications; therefore, app ranking is an essential determinant of success. Nevertheless, the surge in popularity of applications has concurrently facilitated the emergence of deceitful methodologies that seek to manipulate rankings in order to gain an unwarranted edge. This article investigates the occurrence of ranking fraud in mobile applications and suggests methods to identify and mitigate it. Through the examination of diverse data points including user reviews, download patterns, and engagement metrics, we are able to detect dubious activities that are suggestive of fraudulent rankings. By harnessing the power of machine learning and statistical analysis methodologies, we fabricate resilient detection algorithms that possess the ability to accurately identify fraudulent activities. In addition, we deliberate on proactive strategies that developers and app store operators can adopt to proactively thwart and prevent fraudulent rankings, thereby cultivating an app ecosystem that is equitable and transparent. Our objective with these endeavors is to maintain the credibility of mobile application rankings and guarantee an equitable environment for every developer.

***KEYWORDS*:** Mobile Applications, fraud detection, evidences, Historical Record

1. INTRODUCTION

Smartphone owners resort to apps for all manner of purposes. Numerous app developers can create apps with a variety of features. It's not uncommon for programmers to break the law in pursuit of improved system performance. As a result of these unlawful activities, the use of apps for mobile devices is on the rise. Applicants for application writing positions are becoming increasingly dishonest. Poor mobile app reviewers are identified in this research. This research proposes the R3-RFD approach for detecting score fraud by analyzing the most frequent mobile app sessions. The provided language is also used accurately to compute ratings. Falsely boosting the rating of an application through comments is illegal. When evaluating whether or not to allow an application's data through, two different limitations are considered. Each subscriber can only submit one review per app per month. The stipulation in question is crucial. Each MAC address is limited to five daily logins with the second option.

Purpose

This project aims to investigate rating fraud and provide a trustworthy method for detecting suspicious behavior in app store rankings. When people engage in dishonest or fraudulent practices to make an app appear more popular than it actually is, this is known as a "ranking scam" in the mobile app industry. By keeping an eye on peak usage periods—especially during leadership sessions—the proposed method successfully identifies instances of phony ranking in mobile apps. Rank data, rating data, and review data are the three types of evidence available.

2. LITERATURE SURVEY

D. M. Blei, A. Y. Ng, and M. I. Jordan, Latent Dirichlet allocation :

Blei, Ng, and Jordan pioneered the Latent Dirichlet Allocation (LDA) model, a probabilistic generative framework designed specifically for the analysis of discrete data sets, such as textual content. The model is held together by a hierarchical framework and a three-stage Bayesian model. Each member of the group is represented by a simple conglomeration of concepts at each tier. It is demonstrated that there is an endless number of distinct permutations of a finite set of subject probabilities that constitute each and every topic. With the help of text modeling, you can generate plausible topics that adequately characterize the content of a document. This manual will show you how to estimate things using a number of different approaches. The book also demonstrates how to use the expectation-maximization (EM) technique to determine empirical Bayes values. An LSI model that assesses the likelihood of each word's meaning is used to demonstrate the outcomes in the fields of document modeling, text categorization, and joint filtering.

Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou, A Taxi Driving Fraud Detection Systemices in City Taxis :

As demonstrated by Ge, Xiong, Liu, and Zhou (year), advancements in GPS tracking technology have made it simpler to install tracking devices in vehicles, allowing for the rapid collection of a large number of GPS tracks. When police use GPS tracking, they are more likely to apprehend dishonest taxi drivers. This demonstrates a reliable strategy for locating dishonest taxi drivers by the accurate recognition of dishonest behavior. Here, evidence of either the trip's route or distance can be used. The third function is a product of the first two, thanks to the Dempster-Shafer theory. At first, a massive trove of automobile GPS data is mined for key locations. After that, it's shown how to retrieve trip-plan details without providing any inputs. In addition, waymarking has been installed to denote the path that serves as a link between the various locations. The purpose of these markers is to uncover proof of driving distances and facilitate the construction of a model illustrating the distribution of driving distances. A method has been demonstrated that can detect taxi fraud utilizing large amounts of GPS data from legitimate cabs.

T. L. Griffiths and M. Steyvers, Rank Aggregation Via Nuclear Norm Minimization :

T. L. Griffiths and M. Steyvers introduce rank aggregation, which is intrinsically linked to the use of skew-symmetric matrices. Our novel method of ranking uses cutting-edge work in matrix completion theory to improve upon previous approaches. This concept originated from the rake aggregation technique, which provides a comprehensive view of a partially filled skew-symmetric matrix. The matrix completion technique can be used to determine individual ranks when working with skewed-symmetric data. The data is ranked based on a single set of criteria, and both comparisons are performed simultaneously. The proposed method fares well in the face of noisy or missing data due to its emphasis on matrix completion approaches.

A. Klementiev, D. Roth, And K. Small, An Unsupervised Learning Algorithm for Rank Aggregation :

Knowledge learning, data mining, and natural language processing are all topics that Klementiev, Roth, and Small have covered in their respective research articles. They discuss how case grading is crucial in many scenarios in these domains but is not provided by categorization techniques. Rank aggregation describes the process by which ratings from many systems are combined into a single set of criteria. The findings reveal ULARA, a novel method of unsupervised learning that makes extensive use of a linear array of scoring functions. These additions were developed to facilitate consensus amongst evaluators.

A. Klementiev, D. Roth, And K. Small, Unsupervised Rank Aggregation with Distance-Based Models [8]:

The difficulty of merging a set of ranks, which typically requires putting together specifically prepared ranked data, is addressed by the approach described by Klementiev, Roth, and Small (year). Rank grouping makes use of a wide variety of heuristic and supervised learning approaches. These techniques require an in-depth understanding of the issue at hand and access to data that has been manually sorted. An approach is proposed that would eliminate this issue by automatically calculating aggregate rankings. This strategy is employed when numerous approaches exist for compiling a top-k set.

3. EXISTING SYSTEM

The number of apps available for smartphones has increased dramatically over the past few years. By the end of April 2013, more than 1.6 million apps were available in both the Apple App Store and Google Play. Many app stores feature daily app leaderboards that rank apps by the number of downloads they've received in an effort to promote app discovery and growth. Getting your mobile apps included on the App top board is a certain strategy to increase downloads. If you move up the rankings, your sales and income are likely to increase by the millions. Promoting their apps is only one strategy developers use to boost downloads and rankings in the app stores.

Drawbacks of Existing System

How to detect rate fraud in mobile apps is an area of study that is just getting started. There is no way to manually manipulate the rankings of all the available mobile apps.

Proposed System

This study recommends developing a system that can identify fraud as a means of preventing the manipulation of mobile app rankings. Several key issues have been raised as having been overlooked here. Rating theft is uncommon during the application procedure. This highlights the need of accurately establishing the time of the theft. The problem at hand is discovering anomalies in mobile apps on a regional rather than worldwide scale. The enormous number of mobile apps has also been cited as a cause for concern, as it complicates efforts to root out instances of review fraud. Therefore, it is crucial to develop a system that can detect fraudulent ratings independently of normal data and is applicable in any context. Due to the dynamic nature of the music industry, spotting and proving blatant incidents of ranking fraud can be challenging. Due to this issue, we are investigating possible tells within mobile apps that lie covertly.

Finding fraudulent ranking manipulations in critical app sessions is the focus of mobile app ranking fraud detection. By analyzing review data from the past, we make it simple to locate the most crucial sessions for each given application. After analyzing rating trends throughout multiple crucial sessions, researchers discovered that fraudulent apps perform differently from their real-world counterparts. This investigation into mobile app fraud uses historical ratings data. In this article, we'll go over three simple strategies for locating app-based evidence of false conduct.

4. PROBLEM STATEMENT

Numerous app shops regularly update their tree-like rankings of the day's top downloaded apps. A leaderboard is required to encourage software downloads among the general public. Because there are so many copies floating around, the quality of the original application may suffer. Recent research has demonstrated the compatibility between software releases. We must correct this serious problem. There is a direct correlation between the amount of downloads and the app's rating, which the author uses to their financial advantage. In this scenario, phony applications could be submitted. The fact that the user searches for more, similar fraudulent apps is further evidence that they are ill-informed. None of the accepted or real review, rating, or ranking ratios are widely known.

5. SYSTEM MODEL

More than 400 thousand programs may be found in app shops like the Apple App Store and the Google Android Market. This demonstrates the company's rapid expansion over the past few years. Users seeking out novel and efficient application models face new obstacles in light of the explosion of mobile app development. This is why savvy marketers are turning to user feedback in the form of app ratings and comments. In this analysis, the popular sessions were identified using the powerful R3-RFD technique. According to the findings, programs deviate from their regular behavior by displaying distinct ranking trends across sessions. These ratings provide sufficient evidence to conclude that mobile app fraud does occur. Real-world evidence-based methods for detecting these fraudulent actions have been developed in an effort to mitigate their negative impacts. Fake mobile apps can easily go undetected in the app stores' rating systems. The history of the app's ratings and reviews revealed multiple instances of fraud and peculiar tendencies. In addition, we introduce a novel unsupervised data gathering approach for verifying the integrity of existing connections. Long-term testing of the proposed strategy using data collected from the Google Play store. Real-world evidence supporting the effectiveness of the proposed approach is presented in Figure X.

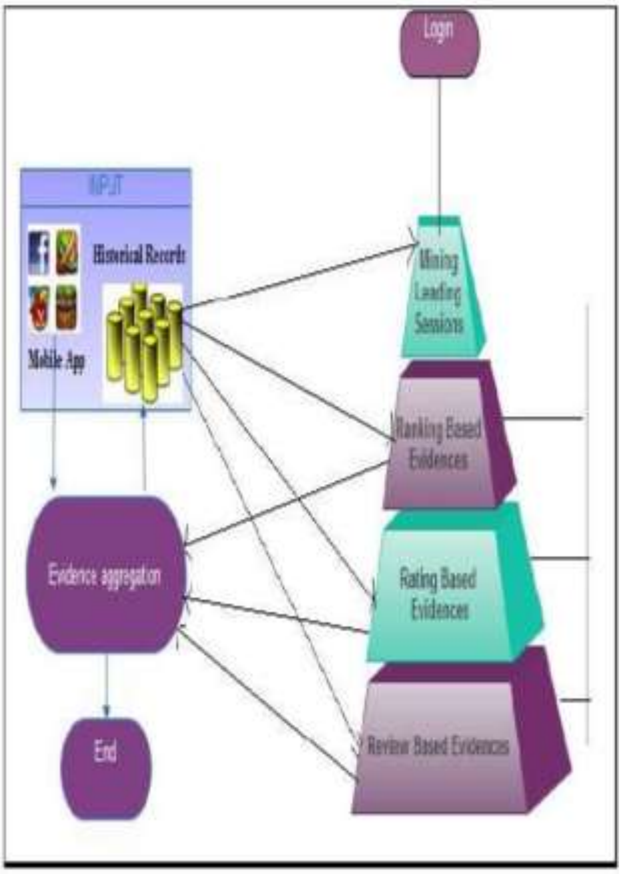

Fig. 1: It's fascinating to see the system model.

6. EXPERIMENTAL RESULTS

Experimental Data

The information was gathered from the Google Play Store's Top Free 300 and Top Paid 300 app lists. Each collection includes a daily list of the 300 best mobile apps (both paid and unpaid). User ratings and comments are included in all of the datasets, so keep that in mind.

Mining Leading Sessions

The minimum allowed value for K* is 7, and its maximum value is 300.

The process of identifying pivotal moments is split into two distinct phases. Examining the historical App ratings is a great way to spot outliers. The success of an introductory course depends on the clarity of its first tasks. Using pseudocode, Algorithm 1 demonstrates how to determine which program sessions occur most frequently

.

Algorithm 1 Mining Leading Sessions

Input 1: a's historical ranking records R_a;
Input 2: the ranking threshold K^*;
Input 2: the merging threshold ϕ;
Output: the set of a's leading sessions S_a;
Initialization: $S_a = \emptyset$;

```
E_s = ∅; e = ∅; s = ∅; t^e_start = 0;
for each i ∈ [1, |R_a|] do
    if r^a_i ≤ K* and t^e_start == 0 then
        t^e_start = t_i;
    else if r^a_i > K* and t^e_start ≠ 0 then
        //found one event;
        t^e_end = t_{i-1}; e =< t^e_start, t^e_end >;
        if E_s == ∅ then
            E_s ∪ = e; t^s_start = t^e_start; t^s_end = t^e_end;
        else if (t^e_start − t^s_end) < φ then
            E_s ∪ = e; t^s_end = t^e_end;
        else then
            //found one session;
            s =< t^s_start, t^s_end, E_s >;
            S_a ∪ = s; s = ∅ is a new session;
            E_s = {e}; t^s_start = t^e_start; t^s_end = t^e_end;
        t^e_start = 0; e = ∅ is a new leading event;
return S_a
```

7. CONCLUSION AND FUTURE WORK

This research demonstrates a novel approach to discovering instances of location-based blackmail in mobile apps. Using each app's historical ranking data, this study outlines how driving sessions were acquired and demonstrates how location deception could occur during these sessions. At the present time, it is believed that verifications based on positions, ratings, and surveys can aid in detecting incidents of dishonesty in positions. This study also reveals a novel model that improves the system as a whole by providing all the data required to verify the validity of driving sessions transmitted via mobile apps. This is a novel approach since it allows direct comparison of all data to empirical evidence. This facilitates the integration of further geographic data in the search for instances of positional deception. Real data from the Apple App Store's downloads have been used to test the proposed strategy extensively. Initial findings indicated that the proposed approach was effective. The next step is to examine and verify the genuine relationship between rankings, surveys, and scores via the lens of plausible lying. Improvements to the user experience can be made across the board by integrating geolocation editing with existing mobile app functions, such as app recommendations.

REFERENCES

1. D. M. Blei, A. Y. Ng, and M. I. Jordan, Latent Dirichlet allocation, J. Mach. Learn. Res., pp. 993–1022, 2003.
2. Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou, A taxi driving fraud detection system, in Proc. IEEE 11th Int. Conf. Data Mining, 2011, pp. 181– 190. D. F.
3. Gleich and L.-h. Lim, Rank aggregation via nuclear norm minimization, in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2011, pp. 60–68.
4. T. L. Griffiths and M. Steyvers, Finding scientific topics, Proc. Nat. Acad. Sci. USA, vol. 101, pp. 5228–5235, 2004.
5. G. Heinrich, Parameter estimation for text analysis, Univ. Leipzig, Leipzig, Germany, Tech. Rep., http://faculty.cs.byu.edu/~ringger/CS601R/papers/Heinrich- GibbsLDA.pdf, 2008.
6. N. Jindal and B. Liu, Opinion spam and analysis, in Proc. Int.Conf. Web Search Data Mining, 2008, pp. 219–230.