

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

Mobile App Ranking, Rating and Fraud Detection

Vishakha Kothari, Prof. Prabhudev Irabashetti

ME, Dept. of Computer Engineering, VACOE, Ahmednagar, India

Asst. Professor, Dept. of Computer Engineering, VACOE, Ahmednagar, India

ABSTRACT: Positioning coercion in the versatile App business segment implies fake or deluding practices which have an inspiration driving thumping up the Apps in the notoriety list. Without a doubt, it ends up being more progressive for App creators to use shady means, for instance, swelling their Apps' business or posting fake App evaluations, to submit situating coercion. While the importance of turning away situating coercion has been extensively seen, there is confined perception and examination here.

There are such a variety of uses accessible on web in view of that client cannot generally get right or genuine audits about the item on web. The essential point of this anticipate is to upgrade the counteractive action of positioning cheats in portable applications utilizing mining and NLP procedures. In the current framework the main occasion and driving session of an application is distinguished from the gathered chronicled records. At that point three unique sorts of confirmations are gathered from the client criticisms in particular positioning based proof, rating based confirmation and survey based confirmation. These three proofs are accumulated by utilizing proof conglomeration strategy.

The cell phone industry has become colossally as of late. The two driving stages, Google Android and Apple iOS, every component commercial centers offering a huge number of programming applications, or applications. The incomprehensible determination has encouraged a developing industry, with new business and income models rising. As an application engineer, fundamental measurements and information for one's applications are accessible by means of the commercial center, additionally by means of outsider information sources. The primary object of this proposal venture is to research whether discoveries from application measurements can give understanding in how to outline more fruitful applications, that pull in more downloads and/or more income

KEYWORDS: Mobile Apps, ranking fraud detection, evidence aggregation, historical ranking records, rating and review.

I. INTRODUCTION

Presently a day, the vast majority of us uses android Mobile furthermore utilizes the play store ability ordinarily. Play store give extraordinary number of utilization yet unfortunately few of those applications are extortion. Such applications harm to telephone furthermore might be information robberies. Henceforth such applications must be checked, with the goal that they will be identifiable for play store clients. Positioning coercion in the convenient App business part suggests fake or deluding practices which have an inspiration driving thumping up the Apps in the famous rundown. So we are proposing an application which will prepare positioning based confirmation, rating based proof and survey based confirmation with stop words evacuation, NLP and mining strategies. So it will be less demanding to choose which application is extortion or not. There are more than 1.6 million Apps at Apple's App store and Google Play. Rather than relying upon customary showcasing arrangements, shady App designers resort to some false intends to deliberately help up their Apps and at last impact the graph rankings on an App store. This is executed by utilizing Bot-homesteads or human water armed forces to expand the App downloads, appraisals and audits in a brief timeframe.

While the significance of avoiding positioning extortion has been broadly perceived, there is constrained comprehension and examination here. To this end, in this paper, we give an all-encompassing perspective of positioning misrepresentation and propose a positioning extortion identification framework for portable applications. We give an all-encompassing perspective of positioning extortion and propose a positioning misrepresentation

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

recognition framework for portable Apps. In particular, we first propose to precisely find the positioning misrepresentation by mining the dynamic time frames, to be specific driving sessions, of portable Apps. Such driving sessions can be utilized for recognizing the nearby irregularity rather than worldwide abnormality of App ranking.

Background

Official versatile application measurements are accessible from the commercial centers, for example, Google Play and Apple App Store. Notwithstanding, these are regularly deficiently nitty gritty for an application distributor's needs. The interval arrangement of outsider information devices that accumulate and order extra information are frequently in view of the reason that application distributors will impart their official commercial center qualifications to this outsider, so that their product may follow up for the application distributor's benefit. These outsider devices are by and large electronic apparatuses that limit access to a portion of the crude information. Furthermore, while these outsider apparatuses might be helpful for creating basic outlines and reports, they are regularly not reasonable for spotting patterns. Also there are outsider apparatuses that require alterations of the application itself, for instance to implant particular libraries for following purposes. The ideal application insights arrangement would be in-house with a specific end goal to dodge reliance on outside administrations that could possibly exist later on or whose expense may increment. Putting away all application measurements in a neighborhood database would likewise be favorable position over an outsider site where the information is perused just for the application engineer. A solitary neighborhood information source offers better conceivable outcomes for factual investigation and guarantees traceability of the information. The size and framework prerequisites ought not be significant, either. Typically, all information is likewise anonymized for individual trustworthiness reasons. What is reasonably exceptionally fascinating is to consolidate a hefty portion of the current application measurements devices into a complete arrangement covering all angles from demographics and gadget dissemination, to number of downloads and rating rendition following, market positioning in view of class and area, in-application utilization designs, application survey examination, and even programmed treatment of bugs and programming issues. Obviously, for relative correlation of one's own applications versus rivalry some solid outsider information sources should in any case be utilized. In any case, having admittance to information for every one of these ranges could possibly be a capable apparatus. Giving chose client's access to it would be much all the more intense. Controlling the database additionally makes the likelihood for imparting this application information to others. This sharing may be finished by utilizing a particular API or by permitting read-just access to some bit of or the whole database by others. Another probability is RSS channels. Case in point, an application distributed firm may impart measurements for App A to Company B over a web administration which guarantees that exclusive Company B can get to the information about App A. Organization B could install or advance distribute this information all alone, subject to their concurrence with the application distributed firm who give this information. The measurements stage could give semi-computerized report era, which at present requires extensive manual push to order the insights acquired from a few distinct sources.

Google Play Store Parser:

Take out the information from Google play store. Provide the URL path of apps, single or multiple path. Extracts following property of android application

- Ranking
- App Category
- App Version
- Size
- Installs
- Review
- Rating

We are proposing system with 2 enhancements. Firstly, we are using Senti-word dictionary to identify the exact reviews scores. Secondly, the fake feedbacks by a same person for pushing up that app on the leader board are restricted. Two different constraints are considered for accepting the feedback given to an application. The first

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

constraint is that an app can be rated only once from a user login. And the second is implemented with the limited number of user login logged per day into system.

So we have proposed an optimization based aggregation method to integrate all the evidences for evaluating the credibility of leading sessions from mobile Apps. A unique perspective of this approach is that all the evidences can be modeled by statistical hypothesis tests, thus it is easy to be extended with other evidences from domain knowledge to detect ranking fraud.

Business models

Notwithstanding the conventional deals strategy for offering one's application available to be purchased at a specific coincidental cost or for an occasional expense new plans of action have risen in the application exchange. The selection of these different plans of action is developing with time. These new plans of action are portrayed in the rest of the section.

I. One-off payment

This is the most evident application plan of action. Build up an application, set a cost, and offer it to clients. There is just a one-time cost for boundless application use. Income is part 70-30 between the engineer and the commercial center/stage proprietor. One issue with this installment model is that the gigantic choice of applications (700 000+ in both prevailing commercial centers) has pushed costs down to levels that make it difficult to be productive exclusively based upon the income stream from new downloads. The normal offering cost in Apple's App Store is ~US\$2.15 [3]. Be that as it may, one-off installments are well known among the top amusement distributors, (for example, Electronic Arts and Gameloft), where 35% of their incomes in 2012 originated from irregular expenses [3]. Actually, the rundown of the main 100 earning applications on App Store is regularly 75% amusements. I trust that the explanation behind this is these all around perceived distributors have increased adequate purchaser trust so that the in advance price tag is not saw as a hazardous venture. Obviously, setting a coincidental cost for an application or amusement does not keep the distributor from likewise actualizing a plan for in-application buys.

II. Subscriptions

Applications that give access to something – an administration or highlight – are frequently reasonable for the membership installment model. A case is music spilling administrations, with applications that are allowed to download, yet require month to month restoration of the record for an expense. Online multiplayer amusements may likewise fit this profile. A few daily papers and magazines likewise offer their premium material (not accessible on their open sites) through application channels, where access requires a dynamic membership. These memberships might be naturally recharged or delayed at altered time interims.

III. Free or ad-sponsored

The free application plan of action for the motivations behind this postulation incorporates both totally free applications that might be distributed as advancement or by novices, and applications that are allowed to download however get wage from inserted notices (promotions). Such advertisement segments are offered by Google through its Admob administration, Apple through its iPad program, and by others, for example, Crisp Wireless stage, utilizing a substantial assortment of portable promotion sorts.

IV. Freemium

A variation of free is the plan of action freemium, a wit mix of free and premium begat by JaridLukin. The thought is to discharge two renditions of the application: a lightweight free form of the application that makes enthusiasm for the paid calling (expert) adaptation. The model looks like the shareware or exhibit idea broadly utilized as a part of the PC world. Gartner's exploration report "Market Trends: Mobile App Stores, Worldwide, 2012" [5] ventures that 89% of application downloads are free, and that this rate will keep on rising for a considerable length of time to come. See Table 4 for subtle elements.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

V. *In-app purchases*

The freemium line of thinking has prompted a developing pattern of in-application buys, where the application is generally a shell and its genuine substance is obtained independently. For instance, amusements could be given only a couple levels and offer an in-application store that offers extra substance. Information from the Apple App Store propose that this plan of action really makes the real piece of application income today

II. LITERATURE REVIEW: [PAPER 1]

A mobile developer intentionally or accidentally keeps the in-app advertising control near to where the user must touch, or scroll on usage of smart phone. With the given micro screen real-estate, the user will be lead to mistap while working on the mobile app.

AKlementiev, D. Roth, and K. Small, produces a model which has to integrate the set of rankings often deals with aggregating and it only comes up when a certain ranked data is developed. Even though the various heuristic and supervised learning approaches to rank aggregation, a requirement of domain knowledge and supervised ranked data exists. Therefore, to solve this issue, a framework is proposed for learning aggregate rankings without supervision. This framework is instantiated for the cases of permutations and combinations of top-k lists [4].

D. M. Blei, A. Y. Ng, and M. I. Jordan, introduces a unique model called as Dirichlet allocation (LDA) a generative probabilistic model for collections of discrete data such as text amount. Basically it is a three level hierarchical Bayesian model in which each element of a group is demonstrated as a finite mixture over a fundamental set of topics. Each topic is demonstrated as an infinite mixture over fundamental set of topic probabilities. With the reference of text modeling, the topic probabilities provide an open representation of a document. An efficient approximation inference technique is presented based on various methods and an EM algorithm for empirical Bayes parameter estimation is also presented. The results are reported in document modeling, text classification and collaborative filtering, which compares to a collection of unigrams and probabilistic LSI model [6].

Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou, illustrated that growth in the field of GPS tracking technology have allowed the users to install GPS tracking devices in taxis to gather huge amount of GPS traces under some time period. These traces by GPS offered an unparalleled opportunity to uncover taxi driving fraud traces and then fraud detection system is proposed which is able to identify taxi driving fraud. First, two sort of function are uncovered here i.e. travel route evidence and driving distance evidence. Even a third function is developed to combine the previous functions based on Dempster-Shafer theory. First identification of interesting locations is done from tremendous amount of taxi GPS logs and then parameter free method is proposed to extract the travel route evidences. Secondly, concept of route mark is developed to illustrate the driving path between locations and based on those mark, specific model is characterized for the distribution of driving distance and discover the driving distance evidences. And finally, taxi driving fraud detection system with large scale real world taxi GPS logs [7].

T. L. Griffiths and M. Steyvers, introduces the process of rank aggregation which is interweave with the structure of skew-symmetric matrices. Recent development in the principles of matrix completion matrices is been applied and this idea gives rise to a new method for ranking a set of items. The core of this idea deals with the raking aggregation method which intimately describes a partially filled skew-symmetric matrix. The algorithm for matrix completion is raised to hold skew-symmetric data and use that to extract ranks for each item. This algorithm applies same strategy for both pairwise comparisons as well as for rating data. It becomes robust to noise and incomplete data as it is based on matrix completion [8].

A. Klementiev, D. Roth, and K. Small, describes the field of information retrieval, data mining, and natural language, many applications needs a ranking of instances which is not present in classification. Furthermore, a rank aggregation is a result of aggregating the results of the established ranking models into a formalism and then result represents a novel unsupervised learning algorithm (ULARA) which gives a linear combination of individual ranking functions. These functions were developed based on the axiom of rewarding ordering agreement between the rankers [9].

Click-spam detection [17] : Many researchers mainly focused on (on early generation) bots to detect the click-spam. When Sbotminer looked for the anomalies in query distribution, he found search engine bots. Millie and

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

Gillberg reported the unusual collusion in users' associated with different publishers that may be pointing to bot behavior. User-Driven Access Control, Bluff ads and Premium Clicks, aim to authenticate user presence to mitigate click-spam. More general approach is proposed by Vachadev, to target every kind of click-spam including bot and non-bot mechanisms.

Since, in the literature, while there are some related work, such as web ranking spam detection, online review spam detection and mobile App recommendation, the problem of detecting ranking fraud for mobile Apps is still under-explored.

Generally speaking, the related works of this study can be grouped into three categories. The first category is about web ranking spam detection. The second category is focused on detecting online review spam. Finally, the third category includes the studies on mobile App recommendation.

Also we can easily conclude from existing system that some of the existing approaches can be used for anomaly detection from historical rating and review records, they are not able to extract fraud evidences for a given time period (i.e., leading session). These cannot able to detect ranking fraud happened in Apps' historical leading sessions. There is no existing benchmark to decide which leading sessions or Apps really contain ranking fraud.

III. PROPOSED SYSTEM

In recent years, mobile app has been growing tremendously while boosting more than 400,000 applications like Apple app store and Google Android market. This tremendous growth of mobile App has made it difficult to user for finding unique and trusted patterns of Application in crowded App stores. Thus to solve this important issue, existing marketing executives precisely use the App download history and ratings by the users to recommend the mobile applications which is totally trusted. Identifying ranking fraud is actually to identify ranking fraud of mobile apps within such leading sessions.

In this paper, an useful algorithm is used to discover the leading sessions based on the historical records and with the help of analysis of those records, it is proved that deceptive apps usually have different ranking patterns in each leading sessions as compared to the normal apps. Therefore it is illustrated from those ranking records that some fraud is taking place in mobile app market and to restrict those frauds, three main evidences are developed to detect such fraud. As only ranking based evidences does not seem to be much sufficient to detect the fraud of mobile app, based on apps rating and review history some fraud evidences were discovered which showed anomaly patterns by those history. Specifically, an unsupervised evidence aggregation method is also proposed here to integrate all such types of evidences for evaluating the trustworthiness of leading sessions.

And finally, the proposed system is estimated with real world app data gathered from the Apple's app store for time consuming period, i.e., more than two years. The results of these experiments showed an effectiveness of proposed approach in fig 1.

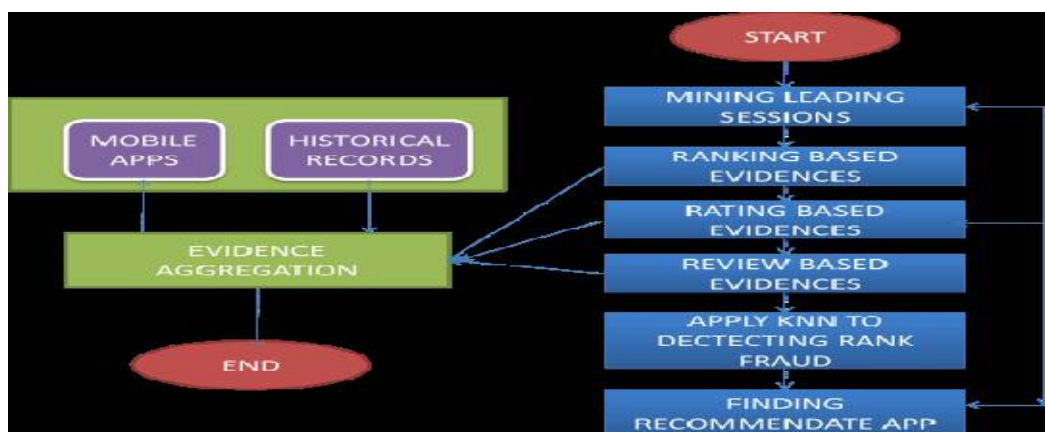


Fig 1: Architecture of Evidence Aggregation based Ranking Fraud Detection.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

We are first proposed a simple yet effective algorithm to identify the leading sessions of each App based on its historical ranking records. Then, with the analysis of Apps' ranking behaviors, we find that the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps. Thus, we characterize some fraud evidences from Apps' historical ranking records, and develop three functions to extract such ranking based fraud evidences.

Following details block diagram will gives the flow of data.

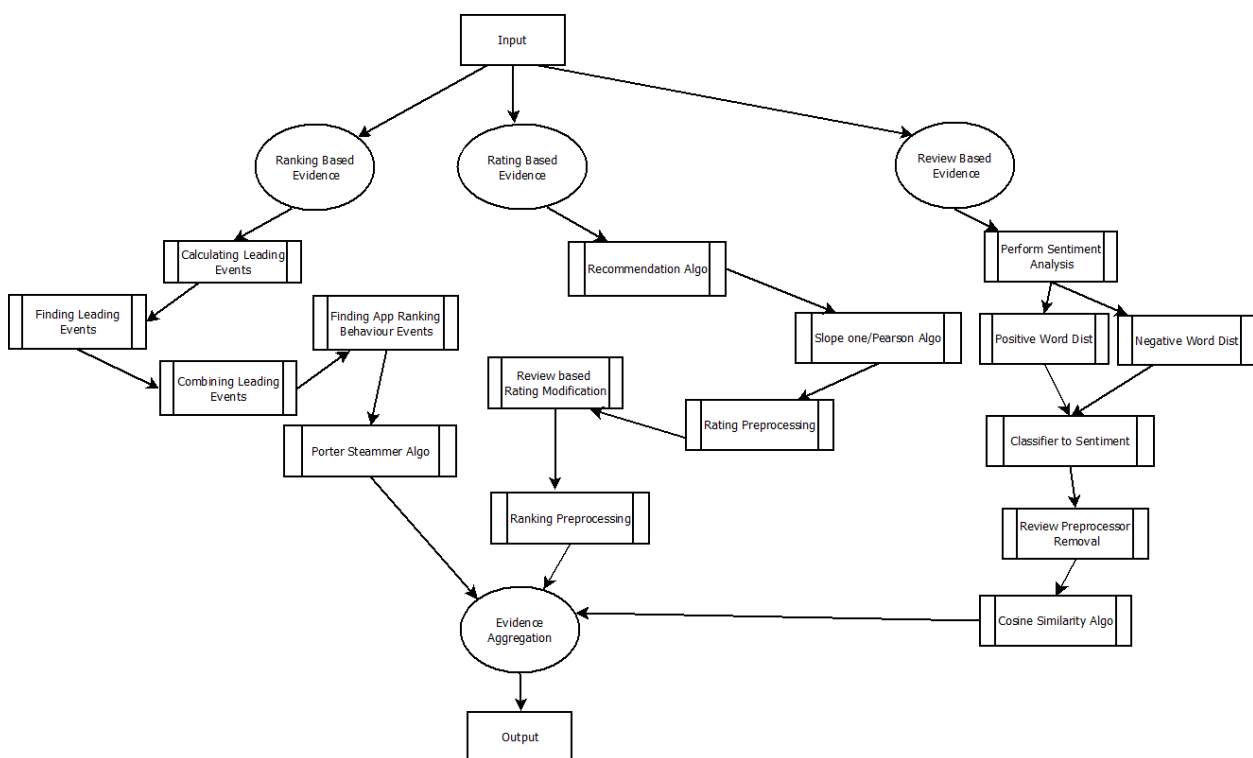


Fig 2: Block diagram of Evidence Aggregation based Ranking Fraud Detection

Basically, mining leading sessions has two types of steps concerning with mobile fraud apps. Firstly, from the Appshistorical ranking records, discovery of leading events is done and then secondly merging of adjacent leading events is done which appeared for constructing leading sessions. Certainly, some specific algorithm is demonstrated from the pseudo code of mining sessions of given mobile App and that algorithm is able to identify the certain leading events and sessions by scanning historical records one by one.

We further propose two types of fraud evidences based on Apps' rating and review history, which reflect some anomaly patterns from Apps' historical rating and review records. In Ranking Based Evidences, by analyzing the Apps' historical ranking records, we observe that Apps' ranking behaviors in a leading event always satisfy a specific ranking pattern, which consists of three different ranking phases, namely, rising phase, maintaining phase and recession phase.

In Rating Based Evidences, specifically, after an App has been published, it can be rated by any user who downloaded it. Indeed, user rating is one of the most important features of App advertisement. An App which has higher rating may attract more users to download and can also be ranked higher in the leaderboard. Thus, rating manipulation is also an important perspective of ranking fraud. The rating based confirmation is utilized to rate by any client who downloaded it.

The KNN calculation is utilized to enhance effectiveness and precision of the application. These all proofs are consolidated for recognizing the extortion applications. 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

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

perceptions and usage experiences of existing users for particular mobile Apps. Indeed, review manipulation is one of the most important perspective of App ranking fraud.

- Evidence Aggregation:** To the best of our knowledge, there is no existing benchmark to decide which leading sessions or Apps really contain ranking fraud. Thus, we develop four intuitive baselines and invite five human evaluators to validate the effectiveness of our approach Evidence Aggregation based Ranking Fraud Detection (EA-RFD). Audit based confirmation is utilized to check the surveys of the application. We developed an unsupervised evidence-aggregation method to integrate these three types of evidences for evaluating the credibility of leading sessions from mobile Apps. Figure 1 shows the framework of our ranking fraud detection system for mobile Apps.

1. Advantages of Proposed System:

- 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.

IV. RESULT ANALYSIS

Sentiment Analysis Precision:

Below result analysis shows the sentiment analysis precision among all existing technology. Result shows the precision of positive sentiment and precision of negative sentiment of all existing technology System will give more than 85% precision values

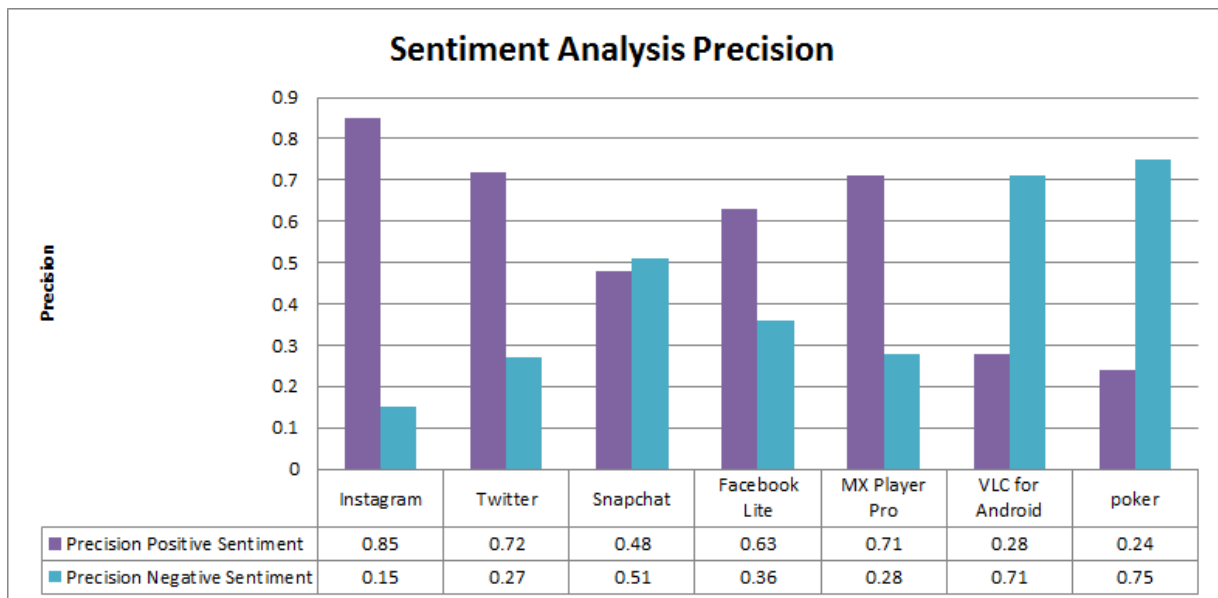


Figure: Sentiment Analysis Precision

Sentiment Analysis on Extracted Review:

Below result analysis shows theSentiment analysis on extracted review among all existing technology. The results shows the negative sentiment, positive sentiment and the total sentiment of all existing technology System will give more than 85% precision values

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

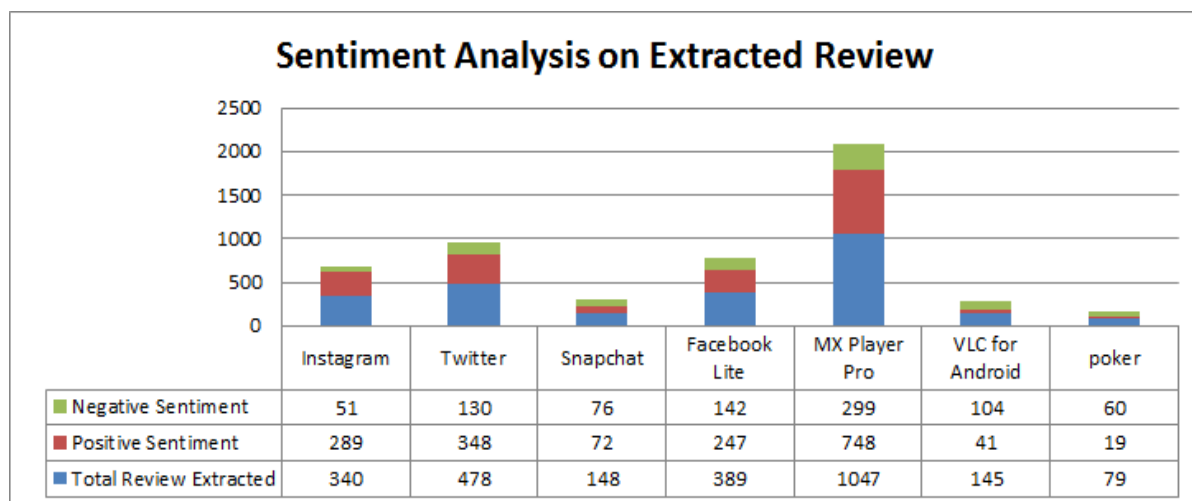


Figure: Sentiment Analysis on Extracted Review

V. CONCLUSION

This paper introduces a system which is built up and it is actually a positioning extortion discovery framework for mobile Apps. In particular, initially it is demonstrated that positioning misrepresentation happened in driving sessions and gave a system to digging driving sessions for each App from its chronicled positioning records.

Complaints of an original version of application provider can be undertaken by using Mining Leading Session algorithm. The duplicate version is identified by the admin by means of Historical Records. The admin will also see the date of publication of the apps. When the apps is detected as fraudulently published by the admin then the respective app will be blocked. The user can give the feedback at only once.

Sentiword dictionary is used for finding the exact reviews. The admin can block the fake application. The Review or Rating or Ranking given by users are Correctly Calculated. Hence, a new user who wants to download an app for some purpose can get clear view about the available applications.

REFERENCES

- [1] Discovery of Ranking fraud for mobile apps. HengshuZhu, HuiXiong, Seniorsmembers, IEEE, YongGe, and EnhongChen, Seniorsmember, IEEE, IEEE transactions on knowledge and data engineering, vol .27, No.1, January 2015.
- [2] Detecting product review spammers using rating behaviors. E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw. In Proceedings of the 19th ACM international conference on Information and knowledge management.
- [3] Supervised rank aggregation. Y.-T. Liu, T.-Y. Liu, T. Qin, Z.-M. Ma, and H. Li In Proceedings of the 16th international conference on World Wide Web.
- [4] An unsupervised learning algorithm for rank aggregation, A. Klementiev, D. Roth, and K. Small In Proceedings of the 18th European conference on Machine Learning, ECML '07, pages 616–623, 2007.
- [5] An unsupervised learning algorithm for rank aggregation, A. Klementiev, D. Roth, and K. Small In Proceedings of the 18th European conference on Machine Learning, ECML '07, pages 616–623, 2007.
- [6] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., pp. 993–1022, 2003.
- [7] 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.
- [8] T. L. Griffiths and M. Steyvers, "Finding scientific topics," Proc. Nat. Acad. Sci. USA, vol. 101, pp. 5228–5235, 2004.
- [9] A. Klementiev, D. Roth, and K. Small, "An unsupervised learning algorithm for rank aggregation," in Proc. 18th Eur. Conf. Mach. Learn., 2007, pp. 616–623.
- [10] Getjar mobile application recommendations with very sparse datasets. K. Shi and K. Ali. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '12, pages 204–212, 2012.
- [11] Ranking fraud Mining personal context-aware preferences for mobile users. H. Zhu, E. Chen, K. Yu, H. Cao, H. Xiong, and J. Tian. In Data Mining (ICDM), 2012 IEEE 12th International Conference on, pages 1212–1217, 2012.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

- [12] detection for mobile apps H. Zhu, H. Xiong, Y. Ge, and E. Chen. A holistic view. In Proceedings of the 22nd ACM international conference on Information and knowledge management, CIKM '13, 2013.
- [13] Exploiting enriched contextual information for mobile app classification, H. Zhu, H. Cao, E. Chen, H. Xiong, and J. Tian. In Proceedings of the 21st ACM international conference on Information and knowledge management, CIKM '12, pages 1617–1621, 2012.
- [14] spammers using behavioral Footprints A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '13, 2013.
- [15] Detecting product review spammers using rating behaviors. E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw In Proceedings of the 19th ACM international conference on Information and knowledge management, CIKM '10, pages 939–948, 2010.
- [16] “Detection of Ranking Fraud for Mobile App and Prevention from User’s Recommendation” IJIRCE, DOI: 10.15680/IJIRCE.2015.0309112.
- [17] Dr.D.Vasumati, M. Sree Vani, Dr.R.Bhramaramba, and O.YaswanthBabu “Data Mining Approach to Filter Click-spam in Mobile Ad Networks” ICCDMME'2015 April 20-21, 2015 Bangkok.