

MobSafe: Forensic Analysis For Android Applications And Detection Of Fraud Apps Using CloudStack And Data Mining

Patil Rohini, Kale Pallavi, Jathade Pournima, Kudale Kucheta, Prof.Pankaj Agarkar.
Dr. D. Y. Patil School Of Engineering, Lohegaon, Pune
Computer Engineering Department

Abstract— Nowadays there are so many applications available on internet because of that user can not always get correct or true reviews about the product on internet. So we can check for more than 2 sites, for reviews of same product. The reviews may be fake on individual sites. But after comparing reviews from 2 sites 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 android application that will take reviews from two different websites for single product, and analyze them with NLP for positive negative rating. In this , User will give 2 different URLs from 2 different sites for same product to system as input. For every URL Reviews and comments will be fetched separately and analyzed with NLP for positive negative rating .Then their rating will be combined together with average to give final rating for the product. In this paper we propose the system to develop a android app which help to detect fraud apps using cloudstack and data mining. To develop propose system we use two methods natural language processing and K-means algorithm.

Keywords- Mobile Apps, ranking fraud detection, evidence aggregation, historical ranking records, rating and review, K-means, natural language processing (NLP).

I. INTRODUCTION

Most of us use android Mobile these days and also uses the play store capability normally. Play store provide great number of application but unluckily few of those applications are fraud. Such applications dose damage to phone and also may be data thefts. Hence such applications must be marked, so that they will be identifiable for play store users. So we are proposing a web application which will process the information , *comments and thee reviews of the application with natural language processing to give results in the form of graph*. So it will be easier to decide which application is fraud or not. Multiple application can be processed at a time with the web application. Also User can not always get correct or true reviews about the product on internet. So we can check for more than 2 sites, for reviews of same product. Hence we can get higher probability of getting real reviews.

II. Literature survey

- [1] Hengshu Zhu, Hui Xiong, Senior Member, IEEE, Yong Ge, and Enhong Chen, Senior Member, IEEE “Discovery of Ranking Fraud for Mobile Apps” IEEE Transactions On Knowledge And Data Engineering, Vol. 27, No. 1, January 2015.
This paper proposed a ranking fraud detection system for android mobile apps. In this paper particularly, authors showed that ranking fraud happened in primary sessions for each app from its past ranking records. Then, they identified ranking based, rating based and review based evidences for finding ranking fraud. Additionally, authors proposed an optimization based aggregation method to combine all the evidences for evaluating the reliability of leading sessions from mobile apps.
- [2] Pranjali Deshmukh, Pankaj Agarkar “Mobile Application For Malware Detection” International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 02 Issue: 02 | May-2015 www.irjet.net
In this paper author proposed methods for evaluation of analysis and design pattern of android apps based on cloud computing and data mining. In this paper authors developed mechanism ASEF and SAAF for android apps to achive security. In this authors describe a methodology that performs apps security and provide user friendly interface on a mobile phone.[2]
- [3] Anuja A. Kadam ,Pushpanjali M. Chouragade “A Review Paper on: Malicious Application Detection in Android System” International Journal of Computer Applications (0975 – 8887) National Conference on Recent Trends in Computer Science & Engineering (MEDHA 2015).
In this paper, authors provide a methodical study on the different techniques of malicious application detection in android mobiles. The investigation of permission-induced risk in Android apps on a large-scale in three levels[3]. First upon rank all the individual permissions with respect to their probable risk with different methods. Then, categorize subsets of risk permissions. Then using several algorithms detect the malapps based on the identified subsets of risky permissions.

[4] Muneer Ahmad Dar & Javed Parvez, University of Kashmir, India “Evaluating Smartphone Application Security: A Case Study on Android” Global Journal of Computer Science and Technology Network, Web & Security Volume 13 Issue 12 Version 1.0 Year 2013 Type: Double Blind Peer Reviewed International Research Journal Publisher: Global Journals Inc. (USA) Online ISSN: 0975-4172 & Print ISSN: 0975-4350.
In this paper authors describes the limitations in the current Android security model in detail[4]. In this paper authors gives the detail description of security requirements which need to be consider at the time of designing security mechanism for smartphones.

III . RELATED WORK

1. CLOUD COMPUTING

Cloud computing is the use of hardware and software computing property via internet.



Figure1. Cloud computing process

Load balancing technique is the process of improving presentation in cloud computing. In load balancing distribution of loads in the center of the processor. As the significance of or use of web raise every day, so there is need to increase the necessities of load balancing.

IV. SYSTEM ARCHITECTURE

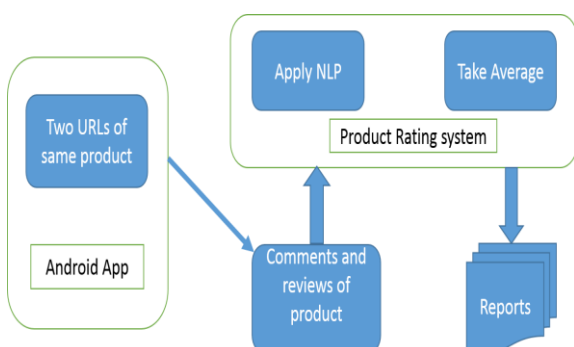


Figure2. System architecture

Natural Language Processing (NLP):

Natural language processing (NLP) is a ground of computer science, artificial intelligence, and computational linguistics afraid with the communications between computers and human (natural) languages. As such, NLP is connected to the area of human-computer communication. Many challenges in NLP involve natural language thoughtful, that is, enabling computers to derive meaning from human or natural language contribution, and others involve natural language generation.

a) SENTIMENT ANALYSIS:

Extract one-sided information usually from a set of papers often using reviews to determine split about specific substance. It is especially useful for identify trends of public view in the social media for the purpose of selling.

b) TOKENIZATION:

Transforming a stream of typeset into a stream of dispensation units called tokens.

c) STOP WORDS FILTERING:

Consists in eliminate stop-words.

d) STEMMING:

The process of dropping each word to its stem or origin form by removing its suffix.

e) STEM FILTERING:

Consists in dropping the number of each amount.

K-MEANS:

We can get safety by using data mining concepts also k-means and clustering algorithm. Usability and viability is increased day by day of mobile apps ,for the purpose , increasing security of mobile apps explain the k-means technique against the hateful activity of android apps. Most of tread names in android:

1. Andriod.PDaspy
3. Andriod.Obad
4. DriodDreamtrojen
5. Android .lucky cat

COMPONENTS:

a) *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

b) *WEB APPLICATION:* GUI Design, Servlets, JSP

c) *GRAPH :*

R is a language and environment for arithmetic computing and graphics. R provides a wide multiplicity of arithmetical and graphical technique, and is highly extensible. R language is used to generate the analysis graphs, like histogram, bar graph, pie chart.

d) *WEB SERVICE :*

Web service is a method of communiqué between two electronic devices over a system. It is a software function provide at a network address over the Web with the examine always on as in the idea of utility computing. The W3C defines a Web service generally as, a software system measured to support interoperable machine-to-machine interaction over a system.

V. MATHEMATICAL MODEL

Set Theory Analysis:

a. Let 'S' be the | Fraud application detection as the final set

$$S = \{ \dots \dots \dots \}$$

b. Identify the inputs as N, Q, C, K, Z, V, U

$$S = \{N, Q, C, K, Z, V, U, \dots\}$$

N = {N1, N2, N3, ...| 'N' gives the application name/ product name}

Q= {Q1,Q2, Q3, ... | 'Q' gives the reviews related application/ product}

C= {I1, I2, ... |'C' gives the comments related application/ product}

K= {P1, P2, ... |'K' gives the rank of application/ product}

Z={D1, D2, D3, D4, ...| 'Z' gives size of application}

V={D1, D2, D3, D4, ...| 'V' gives the version of application}

U={D1, D2, D3, D4, ...| 'U' gives the URL for the application/ product}

c. Identify the outputs as O

S = {N, Q, C, K, Z, V, U, R, ...

N = {N1, N2, N3, ...| 'N' gives the application name/ product name}

R= {R1, R2 ...| 'R' is the Report for the selected apps/ product}

d. Identify the functions as 'F'

S = {N, Q, C, K, Z, V, U, R, F...}

F = {F1(), F2(), F3(), F4()}

F1(N) :: Select Application

F2 (U) :: Pass URL

F3 (Q, C, K,Z, V,) :: Apply NLP

F4 (U) :: Report generation

VI. ANALYSIS

The model of discovery of fraud for mobile apps was proposed by Hengshu Zhu, Hui Xiong, Senior Member, IEEE, Yong Ge, and Enhong Chen [1].

The experimental result[1] are shown in figure 3 and figure 4.

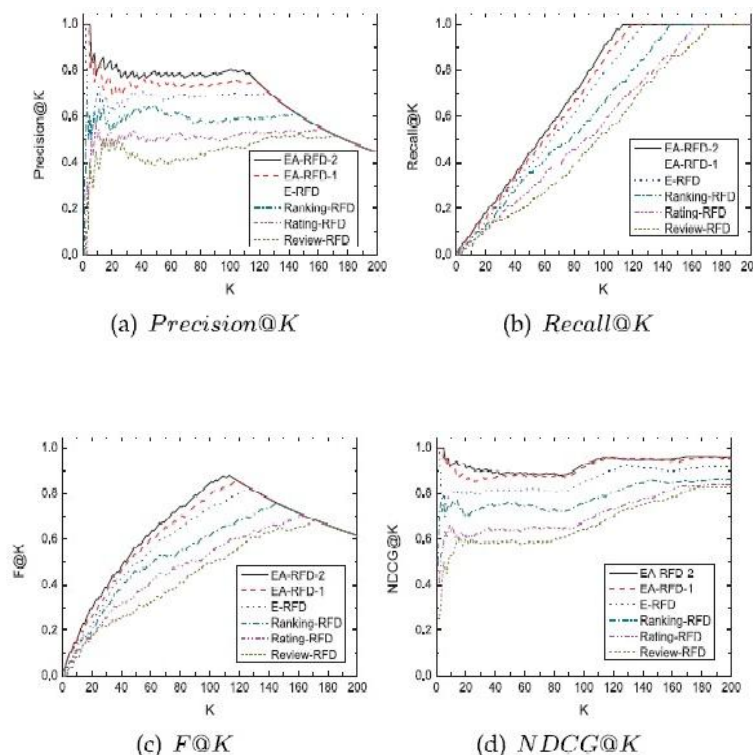


Figure 3. The overall performance of each detection

approach in Top Free 300 data set[1].

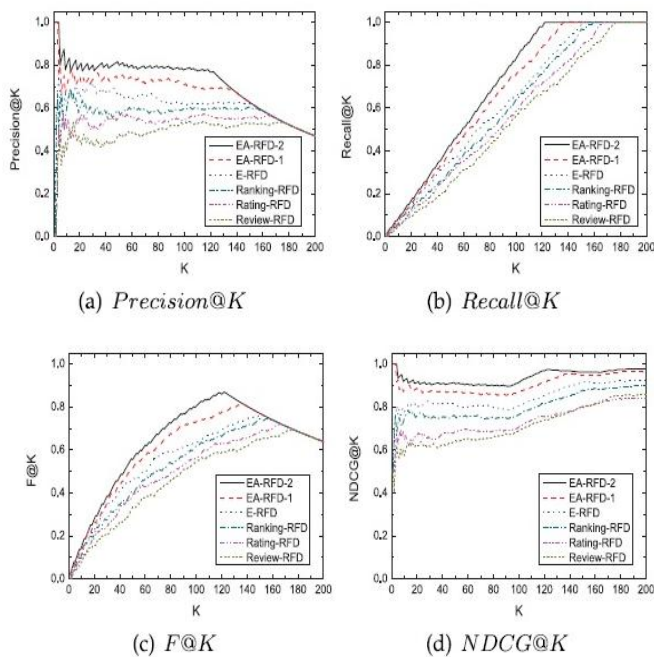


Figure 4. The overall performance of each detection approach in Top Paid 300 data set[1].

In this analysis it describe the overall performances of fraud apps. In this analysis, approach is describe on the basis of different estimation metrics they are as follows Precision@K, Recall@K, F@K and NDCG@K. Here K is set to 200 as maximum. Requirements for this experiment are 2.8 GHZ2 quad-core CPU, 4G main memory PC[1].

Figure 3 and 4 shows performance result of 2 different data sets with each detection approach, the evaluation result of 2 data sets are consistent.

This result shows the usefulness of data aggregation based on frame work which detect ranking fraud apps.

VII. PROJECT SCOPE

This application will be used to detect fraud product selling online or fraud application on mobile. It will help to take reviews from two different websites for single product, and analyze them with NLP for positive negative rating. The same application will be used to detect fraud application and the mobile security as well.

VIII. CONCLUSION

We propose a methodology to evaluate the security of Android mobile apps based on cloud computing and data mining. In this paper we propose the system to detect fraud apps for android mobile. In this system we describe a android app which help to detect fraud apps. This system is based on cloudstack and data mining. To develop propose system we use two methods first one is natural language processing (NLP) and second is K-means Algorithm. We are proposing

a system to develop a android app that will take reviews from two different websites for single product, and analyze them with NLP for positive negative rating. In this system k-means is used to group the applications and then find out the fraud application.

ACKNOWLEDGMENT

We would like to take this opportunity to express our profound gratitude and deep regard to our project guide "Prof. Pankaj Agarkar", for his guidance, valuable feedback and for constant encouragement for the project. Working under him was extremely knowledgeable experience for us.

REFERENCES

- [1] Hengshu Zhu, Hui Xiong, Senior Member, IEEE, Yong Ge, and Enhong Chen, Senior Member, IEEE "Discovery of Ranking Fraud for Mobile Apps" IEEE Transactions On Knowledge And Data Engineering, Vol. 27, No. 1, January 2015
- [2] Pranjali Deshmukh, Pankaj Agarkar "Mobile Application For Malware Detection" International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 02 Issue: 02 | May-2015 www.irjet.net
- [3] Anuja A. Kadam, Pushpanjali M. Chouragade "A Review Paper on: Malicious Application Detection in Android System" International Journal of Computer Applications (0975 - 8887) National Conference on Recent Trends in Computer Science & Engineering (MEDHA 2015)
- [4] Muneer Ahmad Dar & Javed Parvez, University of Kashmir, India "Evaluating Smartphone Application Security: A Case Study on Android" Global Journal of Computer Science and Technology Network, Web & Security Volume 13 Issue 12 Version 1.0 Year 2013 Type: Double Blind Peer Reviewed International Research Journal Publisher: Global Journals Inc. (USA) Online ISSN: 0975-4172 & Print ISSN: 0975-4350.
- [5] (2012). [Online]. Available: <http://www.lextek.com/manuals/onix/index.html>
- [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] D. F. 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.
- [9] T. L. Griffiths and M. Steyvers, "Finding scientific topics," Proc. Nat. Acad. Sci. USA, vol. 101, pp. 5228–5235, 2004.
- [10] A. Klementiev, D. Roth, and K. Small, "Unsupervised rank aggregation with distance-based models," in Proc. 25th Int. Conf. Mach. Learn., 2008, pp. 472–479.