Anomaly Detection based on Review Burstness and Ranking Fraud Discovery

Anpu Alexander
PG Student
Dept. of Computer Science
TKM Institute of Technology Kollam

Rahila N. A.
Assistant Professor
Dept. of Computer Science
TKM Institute of Technology Kollam

P. Mohamed Shameem, PhD
Head of Department
Dept. of Computer Science
TKM Institute of Technology Kollam

ABSTRACT
Nowadays everyone is using smart phone. Many applications are in smart phone. To download an application user visit App store such as Google play store, Apple play store etc, then he or she is able to see the different application lists. User has no awareness about the application. So user looks at the list and download the application from App Store based on the mobile app rank. App developers use different ways to promote their Apps in order to get top position in App store for example, high rating and good reviews are given about the mobile app i.e. there is fraud behavior occur it. To detect fraud behavior first identify the active periods of mobile app, namely leading session of mobile apps. In the existing system the leading event and leading session of an app identified from the collected historical records. Then ranking based evidence, rating based evidence and review based evidence were collected from the historical records. These evidence score value is used to detect fraud behavior occur in the mobile app. In proposed system from the reviews of mobile app it identifies if it is a fake review or not.

Keywords
Aggregation, Leading session, SVM.

1. INTRODUCTION
Data mining is defined as extracting information from huge set of data. The extracted information can be used for any of the following applications. Market Analysis, Fraud Detection, Customer Retention, Production Control and Science Exploration. Due to faster development in the mobile technology mobile apps are growing on a very large scale. Different app stores launched their leader board to display the chart ranking of most populated apps. Leader board is the way to promote mobile app in the market. A high ranked app gets a large number of downloads and million dollars in revenue.

To promote their apps in top position the app developers use different ways. High rating and good review are given about the mobile app. This study focuses on an integrated approach, for various evidences, to find mobile app ranking fraud. Some challenges are faced to find out fraud. First is what time the fraud is happening. It means exact time of fraud occur. Secondly there is number of apps present in market so it is impossible to physically mark ranking fraud for every app, so it's difficult to distinguish fraud without utilizing any essential data. Mobile apps are not ranked high in the leader board, but in a few events. In this way, the fundamental task is to recognize ranking fraud of mobile apps in leading sessions. Initially propose an efficient algorithm to identify the main sessions of every app depends on its previous ranking records. From the review based evidence similarity of the review is measured. It not focus the review is fake or authentic.

This paper focuses on the review authenticity of the mobile app. Not all reviews are necessarily authentic. Some reviews are fake but it likes to be authentic. So, authentic and fake reviews are not easy to differentiate. Hence, this paper uses supervised learning algorithms to analyze authentic and fake reviews. It is based on linguistic clues, namely, understandability, level of details, writing style, and cognition indicators.

The rest of the paper is marshalled as 2.Preliminaries, 3. Related works, 4.System model, 5.Proposed system, 6. Experiment and result, 7.Conclusion, and 8.Future work.

2. PRELIMINARIES
In this section first introduce some preliminaries, and then show how to mine leading sessions for mobile apps from their historical ranking records.

2.1 Leading Event
Given a ranking threshold K ∈ [1, K], a leading event e of app a contains a time range $T_e = [t_{e\,start}, t_{e\,end}]$ and corresponding rankings of a, which satisfies $R_{a\,start} \leq K^* < R_{a\,start-1}$ and $R_{a\,end} \leq K^* < R_{a\,end+1}$. Moreover,$\forall t_k \in (t_{e\,start} t_{e\,end})$, $R_{a\,k} \leq K^*$.

2.2 Leading Session
A leading session s of app a contains a time range $T_s = [t_{s\,start}, t_{s\,end}]$ and n adjacent leading events $\{e_1, .., e_n\}$, which satisfies $t_{s\,start} = t_{e1\,start}, t_{s\,end} = t_{en\,end}$ and there is no other leading session s^* that makes $T_s \subset T_s$ Meanwhile, $\forall_i \in [1, n)$, so $(t_{ei+1\,start} - t_{ei\,end}) < \emptyset$, where $\emptyset$ is a predefined time threshold for merging leading events.

3. RELATED WORK
The related works can be grouped into three classes. Web ranking spam detection is the first class. The Web ranking spam means that any deliberate actions which bring to selected Web pages an unjustifiable favourable relevance or importance. It introduces the concept of spamicity, to measure how likely a page is spam. Spamicity is more flexible and user controllable measure than the traditional supervised classification methods. They propose efficient online link spam and term spam detection methods using spamicity. These methods do not need training and also cost effective. A real data set is used to evaluate the effectiveness and the efficiency. With the increase in the number of web apps to detect the fraudulent apps, then a simple and effective algorithm which identifies the leading sessions of each app based on its historical ranking of records is introduced [1]. By analysing the ranking behaviours of apps, the fraudulent apps often have different ranking patterns in each leading session compared with normal app is discovered. So some fraud evidences from app's historical ranking records are identified

and develop three functions to obtain such ranking based fraud evidences.

4. SYSTEM MODEL

The mobile industry is growing rapidly, subsequently the number of mobile apps is also increasing. As there are many apps available, users are confused in downloading the apps for their use. They check the daily app leader boards for selecting an app. But some fraud occurs in the leader board in order to get revenue. So detect such fraud apps, a system is developed based on evidences.

The system model is shown in Fig 1. First a simple effective algorithm is used to identify the leading sessions of each app based on its historical records. Fraud signature values was aggregated from the rank, rate, review behavior and then find whether fraud occur or not. The main modules are:

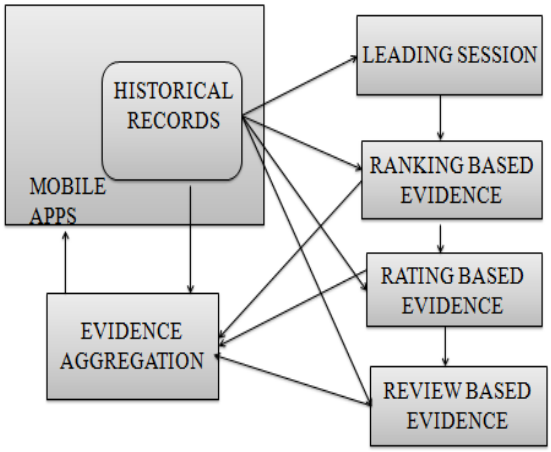

Fig 1: System model for ranking fraud detection

4.1 Mining Leading Session

This is the first step in the proposed scheme. The leading session find out from the historical record of mobile app. The user rated the mobile app in order to enter the popularity list. The leading session can find base on the threshold value. The main aim of the leading session is to find the fraud ratings.

4.2 Ranking Based Evidence

The ranking based evidence is composed by three different ranking phases. i.e. rising phase, maintaining phase and recession phase. In the leader board, every new app is rated. The highly rated apps are ranked to the first place is called rising phase. The same place is occupied for several periods of time are known as maintaining phase. The same app rank decreases over certain period of time are known as recession phase.

4.3 Rating Based Evidence

The ranking based evidence is not enough to detect the fraud apps. So studied the rating based evidences of the mobile app. The marketing services offer limited discount that mostly affects the outcomes of the rating based evidences. It is used for extracting the rating records from the historical records.

4.4 Review Based Evidence

Mobile apps are allowed to write review about the mobile apps. It tells some experiences of the user. This analysis is helpful to detect the fraud applications.

4.5 Evidence Aggregation

The final score is used to find whether there is fraud occur or not. The final score is the total score value of all evidence scores.

5. PROPOSED SYSTEM

Reviews are not authentic. Some reviews are fake but it likes to be authentic. So, authentic and fake reviews are not easy to differentiate. Hence, this paper uses supervised learning algorithms to analyze authentic and fake reviews. It based on four linguistic clues, namely, understandability, level of details, writing style, and cognition indicators [5], [6].

Understandability means that which a review is more understandable to user. Authentic reviews contain plain and simple arguments for describing post-purchase experiences. Understandability was performed as surface-level characteristics. Structural features were calculated as the number of characters per word, number of words, and fraction of words containing 7 or more characters is called long word [4].

Level of details means which review contains objective information. Authentic reviews based on real experiences fake reviews are based on imagination. Level of details contains informativeness, perceptual details, contextual details, and the use of function words. Informativeness was measured by examining the use of part-of-speech (POS) in reviews. The eight POS tags are nouns, adjectives, prepositions, articles, conjunctions, verbs, adverbs, and pronouns [2], [8], [11]. With the help of NLP tool, taggers take input as file containing review text and annotate each word with corresponding tags. Perceptual details contain visual, aural and feeling words. In NLP terms, visual, auditory, kinesthetic and auditory digital words are called predicates. The predicates that a person uses will provide you with an indication of the person's preferred representational system. Visual predicate for thing you see. For example see, look. Auditory predicates for thing you hear. For example tell, sound. Kinesthetic predicates things you feel. For example feel, unfeeling. Function words are words that have little lexical meaning or have ambiguous meaning, but instead serve to express grammatical relationships with other words within a sentence. Function words included non-content words that reduce the level of details in reviews [2], [7].

Writing style of reviews based on the use of emotions, tenses, punctuations, uppercase character. The Tenses were measured as the fraction of past, present and future tense words used in reviews. Fake reviews could contain less past tenses but more present [10]. Emphases were measured based on the proportion upper case characters, as well as use of punctuations such as ellipses "…", question marks "?", and exclamation points "!".

Cognition indicators in reviews based on the use of fillers and motion words. A filler word is an apparently meaningless word. Some of the common filler words in English are um, uh, er, ah, like, okay, right, and you know. Motion words such as "arrive" and "go". Fake reviews could also use more motion words, but fewer exclusion words than authentic entries [3], [9].

The Data were analyzed using supervised learning, which includes machine learning algorithms that use labeled data for training and testing.

SVM is the best classifier that predicts the test data according to the training set available. Authenticity is labelled based on

the various linguistic features arrived from review text analysis. The feature parameter is fed to an SVM kernel to classify into various groups. Linear kernel SVM with random forest classifier is used to conduct supervised learning.

6. EXPERIMENT AND RESULT

In this section evaluate the performance of burstness detection based ranking fraud discovery.

6.1 Experimental Data

An experiment is conducted in pc based environment with 2.7 GHZ intel processor. The data was collected from various mobile application that have feedbacks and ratings for a period starting from Jan 2016 to Dec 2016.The data set is well analyzed to remove discontinuous data and incomplete session.

From the preliminary investigation it self can see the distribution of app rating are not even. Hence mining of session and events were conducted. Evidence were calculated and aggregated besides the rating data. Review based evidence was also incorporated. These all from a baseline method for ranking fraud detection. Table 1 shows the evidence history generated for different apps in different periods.

Table 1. Evidence history for different apps for different period

Evidence	Period 1	Period 2	Period 3	Period 4	Period 5
Evidence1	0.57	0.73	0.6	0.7	0.72
Evidence2	0.67	0.68	0.52	0.6	0.71
Evidence3	0.89	0.79	0.72	0.6	0.87

Enhancing the base line method the concept of linguistic clues added. It generated 4 main characteristics. These characters give the guidelines for assigning a review as fake or authentic.

Stanford NLP library is used to extract NLP tags that representing the parts of speech property. Specifically 8 tags are considered for the generation of level of details.

Review emotiveness is also measured by comparing DECHAL chart lexicon which contains recognizable word forms, cognition data were measured using filters as well as tentative casual and exclusion words.

As shown in Table 2 a total of 13 features are extracted. The obtained variable feature metric is used for data analysis.

Table 2. Feature Observations- Average Feature value of Reviews

Apps	Level of details	Cognition indicators	Understa-ndability	Writing style
App1	0.85	0.6	0.725	0.75
App2	0.83	0.7	0.765	0.7
App3	0.9	0.85	0.875	0.77

6.2 Data Analysis

The ranking fraud detection process generated 6 evidences. The variation of evidence for different app is shown in Fig 2. The significance of linguistic feature is analysed and truth label is assigned as per human judgement and various trails. LIB SVM is the tool used for supervised learning.

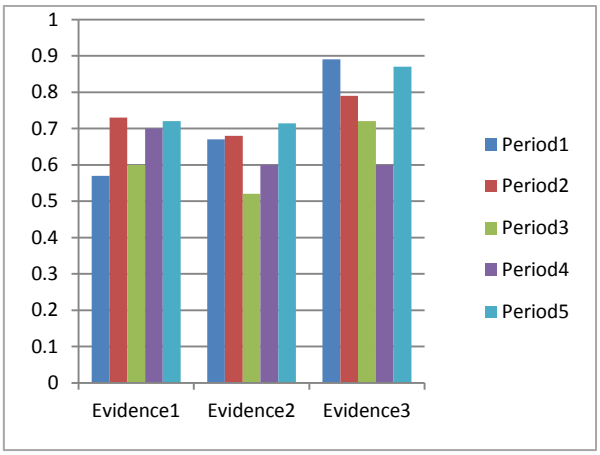

Fig 2: The variation of evidence for all different app

The input reviews are trained to predict the authenticity using SVM classifiers. The results obtained with various app are analysed. The precision and accuracy of the classification are calculated. It is found that the classification gave better result in terms of accuracy, precision. A graphical analysis of the accuracy of SVM process under different data set and different review set given in Fig 3 based on Table 3 values.

Table 3. Overall accuracy under different data set

App name	Accuracy	F-Ratio
App1	0.725	0.75
App2	0.765	0.7
App3	0.875	0.77

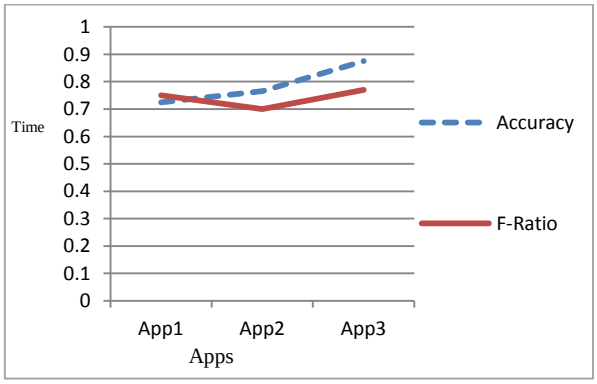

Fig 3: Accuracy of SVM process under different data set and different review set

Table 4 shows the Precision of classifier based authentic review detection. Table 5 shows the Recall of classifier based authentic review detection. The Precision and Recall of classifier is also analysed and given on Fig 4, Fig 5.

Table 4. Precision of classifier based authentic review detection

Trial	No of reviews	Precision RF	Precision ranking with RF
1	8	0.6	0.9
2	6	05	0.8
3	7	0.5	0.74
4	8	0.4	0.7

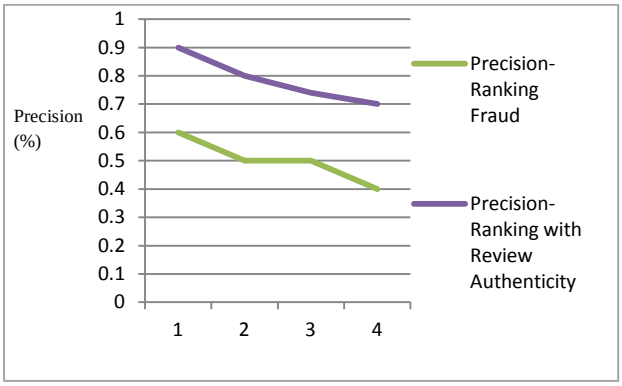

Fig 4: Precision of classifier based authentic review detection

Table 5. Recall of classifier based authentic review detection

Trial	No of reviews	Precision RF	Precision ranking with RF
1	8	0.86	0.9
2	6	0.75	0.78
3	7	0.65	0.74
4	8	0.45	0.67

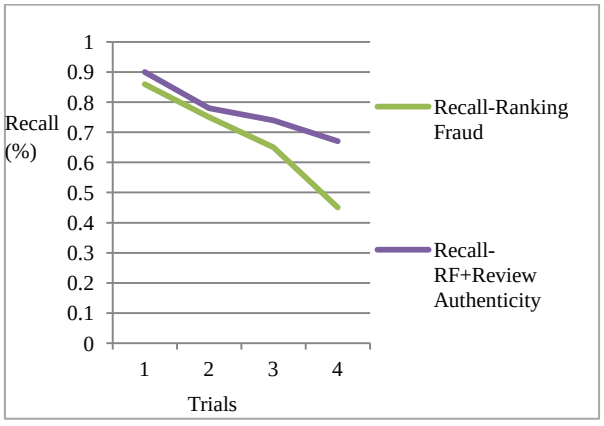

Fig 5: Recall of classifier based authentic review detection

The fraud detection in ranking and fake detection in reviews will generate a clear result for decision making purpose for end users. The system is also favourable in time complexity and memory.

7. CONCLUSION

The project has succeeded in mining the fraudulent reviews and rankings in review data sets. Most popular data sets contain the anomaly in ratings session and event session. Also the burstness of reviews were detected and reported. The experiment could reveal the category of review, whether fake or not. The linguistic analysis is approved by human judgment and machine accuracy. The consumer aspect of this project is very important that consumers and genuine users rely on this fraud and burstness detection strategy. The idea could be further applied on other various text mining application.

8. FUTURE SCOPE

The objective of this work is satisfied with the discovery of fraud and burstness of reviews in a mobile app scenario still there are lot of improvements needed for better performance and accuracy. Rather than SVM some modern ANN based tool can be used for review prediction. Also the strategy of finding evidences can be modified according to apps business cycle.

9. REFERENCES

[1] Hengshu Zhu, Young Ge, and Enhong Chen. 2015. Discovery of Ranking Fraud.

[2] Banerjee, S., and Chua, A. Y. K. 2014. A linguistic framework to distinguish between genuine and deceptive online reviews.

[3] Boals, A, and Klein. 2005. Word use in emotional narratives about failed romantic relationships and subsequent mental health.

[4] Cao, Q. Duan, and Gan, Q. 2011. Exploring determinants of voting for the "helpfulness" of online user reviews.

[5] Banerjee, Snehasish, Alton YK Chua, and Jung-Jae Kim. 2015. "Using supervised learning to classify authentic and fake online reviews."

[6] Hancock, J. T, Curry, L. E., Goorha, S., and Woodworth, M. 2008. On lying and being lied to: A linguistic analysis of deception in computer-mediated communication.

[7] Ott, M., Choi, Y., Cardie, C., and Hancock, J. T. 2011. Finding deceptive opinion spam by any stretch of the imagination. In Annual Meeting of the Association for Computational Linguistic.

[8] Rayson, P., Wilson, A., and Leech, G. 2001. Grammatical word class variation within the British National Corpus.

[9] Tausczik, Y. R., and Pennebaker, J. W. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology, 29, 24-54.

[10] Yoo, K. H., and Gretzel, U. 2009. Comparison of deceptive and truthful travel reviews.

[11] Zuckerman, M., De Paulo, B. M., and Rosenthal, R. 1981. Verbal and nonverbal communication of deception. In Advances in Experimental Social Psychology, L. Berkowitz, Ed., Academic Press, New York.