ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

Recognizing the Ranking Scam Occurred in Mobile Apps

Vaishali Date[1], Dipali Dongare[2], Pooja Jadhav[3], Tejal Wayal[4], Asmita Mali[5]

B.E Students, Dept. of IT, Dr. D. Y. Patil Institute of Engineering and Technology, Pimpri, Savitribai Phule Pune University, Pune, India.

Assistant Professor, Dept. of IT, Dr. D. Y. Patil Institute of Engineering and Technology, Pimpri, Savitribai Phule Pune University, Pune, India.

ABSTRACT: Ranking scam in mobile application market refers to deceptive activities done by someone which have a purpose of raising the rank of application in leaderboard. It has becomes more frequent for App developers to use such means as inflating their Apps' sales or posting fraud ratings, to commit ranking scam. While the importance of preventing ranking fraud has been widely recognized, there has been limited research in this area. This paper may contribute the methods to enhance Scam i.e. fraud detection to improve the ranking of mobile applications. In this paper, we propose a system which discovers the presence of ranking scam and had tried to minimize the fraud in ranking of mobile applications. We are going to list the applications by finding the active period of the application named as leading session. We recognize ranking based evidences, ratingbased evidences and review based evidences. Using these three evidences we calculate aggregation of them and performed hypothesis test on it to get actual rank of mobile application and authentication of user is done for further access which avoids the ranking scam further. We hadretrieved our system with genuine data collected form application play store for long period of time.

KEYWORDS: Mobile Applications, ranking scam, ranking, rating, review, evidence aggregation, authentication of user.

I. INTRODUCTION

Now a day's everyone is using smart phones. There is need of various apps to be installed on smart phone for much purpose. To download application customer needs to visit play store, for instance, Google Play Store, Apples store etc. Right when customer visit play store then he can see the application records. This once-over depends on the reason of progression or advancement. Customer doesn't think about the application (i.e. which applications are useful or useless). So customer looks at and downloads the applications. In any case, as a less than dependable rule it happens that the downloaded application won't work or not profitable. That infers it is distortion in convenient application list. To avoid this, we make system in which we list the applications. The quantity of versatile Apps has developed at an amazing rate in the course of recent years Mobile Apps number has grown very fast as of the end of April 2013; there were more than 1.6 million Apps at Apples store and Google Play store. To countenance the development of mobile Apps, many Apps stores launched daily leader boards, which convey the chart rankings of most well-known Apps. To be sure, the App leader board is the most important way for promoting Apps.

A favorable rank on the leader board leads to a huge number of downloads and million dollars in revenue. In this manner, App designers have a tendency to investigate and explore various ways to raise their Apps in order to have their Apps ranked as high as possible. Instead of relying on traditional marketing solutions, customary promoting arrangements Disreputed App developers resort to some false intends to intentionally help to boost their Apps and eventually manipulate the chart rankings on store. This is generally actualized by utilizing alleged called as "boot farms" and "human water armies" to increase the App downloads in a very small time. For eg, an article from Venture Beat reported that, when an Application was promoted with the help of ranking manipulation, it should be propelled from 1,800 to the top 25 in Apple's top free leader board and more than 50,000 to 100,000 new users could be acquired

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

within a couple of days. Indeed, such positioning extortion raises awesome worries to the versatile App industry.Apple has recommended of cracking down the App developers who commit ranking scam in the Apple's App store.

Subsequently, it is not adequate to just utilize positioning based confirmations, we advance propose two sorts of misrepresentation proofs in view of Apps' evaluating and survey history, which mirror some irregularity designs from Apps' verifiable rating and audit records. What's more, we add to an unsupervised confirmation collection technique to incorporate these three sorts of proofs for assessing the validity of driving sessions from versatile Apps. Fig. 1 demonstrates the structure of our positioning misrepresentation recognition framework for versatile Apps. It is important that every one of the proofs are removed by displaying Apps' positioning, rating and audit practices through factual theories tests. The proposed system is adaptable and can be reached out with other area created confirmations for positioning misrepresentation identification. At last, we assess the proposed framework with genuine App information gathered from the Google App store. Exploratory results demonstrate the viability of the proposed framework, the versatility of the discovery calculation and in addition some consistency of positioning extortion exercises.In the literature, some work is done related to web ranking spam, online review Spam detection and recommendation of mobile Application, the problem of detecting ranking scam is still under explore, To fill this we propose a system which discover the presence of ranking scam in mobile applications.

II. RELATED WORK

The related work of this study can be grouped into three categories. The first is about web ranking spam detection [1], the second is focused on online review Spam detection [2], and finally, the third includes the studies on mobile App recommendation [3].

In [1] we continue our inquiry of "web spam" introduction of artificially-created pages into the web in order influence the solutions from search engines, to drive traffic to certain pages for profit. This paper considers some prior-un described techniques for automatically detecting spam pages, examinee the effectiveness of these techniques in isolation and using grouping algorithms for aggregation. In [2] it aims to detect users producing spam reviews. Different characteristic behaviors of review spammers were identified and model was done for detecting the spammers. In particular, we seek to model the behaviors spammers may target specific product groups in order to maximize their impact and they tend to deviate from the other reviewer in their ratings of products. We propose achieved method to measure the degree of spam for each reviewer and to take affect them on an Amazon review dataset. In [3] we illustrate how to extract personal context-aware preference from the context-rich device logs for building personalized context-aware warned systems. First learn general context-aware preferences from the context logs of most of users. Then, the preference user can be represented as a distribution of these general context-aware preferences. Two approaches are there for mining general context-aware preferences based on two distinct assumptions, namely, context independent and context dependent. Finally, spacious experiments on a real-world data set that show the approaches are effectively and perform baselines with respect to mining personal context-aware preferences for mobile users. In [4] GPS tracking technology has enabled us to install GPS tracking devices in city taxis to collect a huge amount of GPS traces under operational time limitation. In thissystem, we first provide functions to find two evidences: travel route evidence and driving distance evidence and a third function is designed to combine the two evidences based on dempster-Shafer theory. Then, we propose a parameter-free method and we introduce route mark to connote a typical driving path from an interesting site to another one. Based on route mark, we use statistical model to characterize the division of driving distance and identify the driving distance evidences. We evaluate the taxi driving fraud detection system. We uncover some orderly of driving fraud activities and investigate the impetus of drivers to commit a driving fraud by analyze the produced taxi fraud data .In[5] we study issue in the context of product reviews, that are opinion rich and are vastly used by consumers and manufacturers. In the past two years, several startup companies also not disappeared which aggregate opinions from product reviews. We will see that opinion spam is as it is different from Web spam and email spam, and thus requires different quest techniques. On performing the analysis of 5.8 million reviews and 2.14 million reviewers from amazons, we see that opinion spam in reviews is large spread. This paper analyzes such spam activities and presents some techniques to detect them. In [6] many applications in information recompense, data mining, and related fields require a ranking of instances with respect to specified criteria as opposed to a classification. Also, for such problems, established multiple ranking models have been studied and it is desirable to collect their results into a joint ranking, formalism denoted as rank aggregation. This work presents an unsupervised learning algorithm for rank

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

aggregation (ULARA) which returns a linear cohesion of the separate ranking functions based on the principle of rewarding ordering agreement during the rankers. To present ULARA, we exhibition its effectiveness on a data fusion task across ad hoc retrieval systems.

II. EXISTING APPROACH

In the writing, while there are some related work, for example, web positioning spam location [1], [8], [9], online audit spam discovery [2], [11], [12], and versatile App proposal [14], [13], [10], [3], the issue of identifying positioning misrepresentation for portable Apps is still under-investigated. To fill this vital void, in this paper, we propose to add to a positioning misrepresentation identification framework for portable Apps. Along this line, we distinguish a few essential difficulties. To begin with, positioning misrepresentation does not generally happen in the entire life cycle of an App, so we have to distinguish the time when extortion happens. Such test can be viewed as recognizing the neighborhood inconsistency rather than worldwide oddity of portable Apps. Second, because of the immense number of portable Apps, it is hard to physically name positioning extortion for each App, so it is imperative to have an adaptable approach to naturally distinguish positioning misrepresentation without utilizing any benchmark data. At long last, because of the dynamic way of outline rankings, it is difficult to recognize and affirm the proofs connected to positioning extortion, which inspires us to find some verifiable misrepresentation examples of versatile Apps as confirmations.

III. PROPOSED APPROACH

Indeed, our careful observation reveals that mobile Apps are not always ranked high in the leaderboard, but only in some leading events, which form different leading sessions. In other words, ranking fraud usually happens in these leading sessions. Therefore, detecting ranking fraud of mobile Apps is actually to detect ranking fraud within leading sessions of mobile Apps.

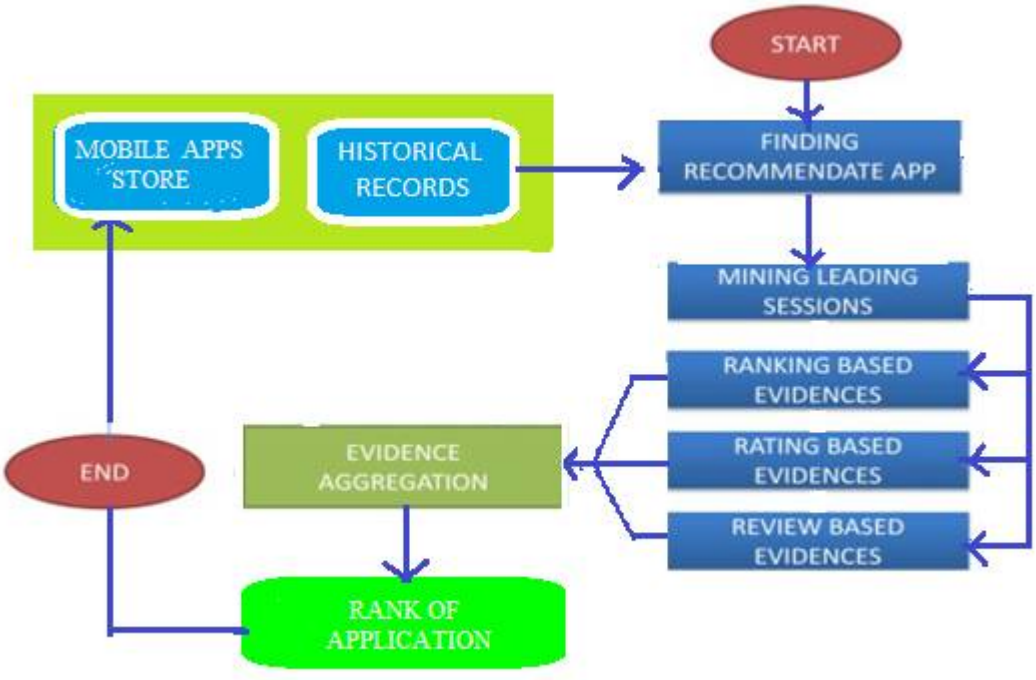


Figure 1: The framework of our ranking fraud detection system for mobile Apps.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

Specifically, we first propose 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. Nonetheless, the ranking based evidences can be affected by App developers' reputation and some legitimate marketing campaigns, such as "limited-time discount". As a result, it is not sufficient to only use ranking based evidences. Therefore, 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 addition, we develop an unsupervised evidence-aggregation method to integrate these three types of evidences for evaluating the credibility of leading sessions from mobile Apps. Fig. 1 shows the framework of our ranking fraud detection system for mobile Apps.

3.1 IDENTIFYING LEADING SESSIONS FOR MOBILE APPS

The App leaderboard demonstrates top K popular Apps with respect to different categories, such as "Top Free Apps" and "Top Paid Apps". Moreover, the leaderboard is usually updated periodically (e.g., daily). Therefore, each mobile App a has many historical ranking records which can be denoted as a time series, $R_a = \{r^a_1, \ldots, r^a_i, \ldots, r^a_n\}$, where $r^a_i \in \{1, \ldots, K, +\infty\}$ is the ranking of a at time stamp t_i; $+\infty$ means a is not ranked in the top K list; n denotes the number of all ranking records. Note that, the smaller value r^a_ihas, the higher ranking position the App obtains.

By analyzing the historical ranking records of mobile Apps, we observe that Apps are not always ranked high in the leaderboard, but only in some leading events. For example, Fig. 2a shows an example of leading events of a mobile App. Formally; we define a leading event as follows:

Note that we apply a ranking threshold* which is usually smaller than K here because K may be very big (e.g., more than 1,000), and the ranking records beyond K* (e.g., 300) are not very useful for detecting the ranking manipulations.

Furthermore, we also find that some Apps have several adjacent leading events which are close to each other and form a leading session. For example, Fig. 2b shows an example of adjacent leading events of a given mobile App, which form two leading sessions. Particularly, a leading event which does not have other nearby neighbors can also be treated as a special

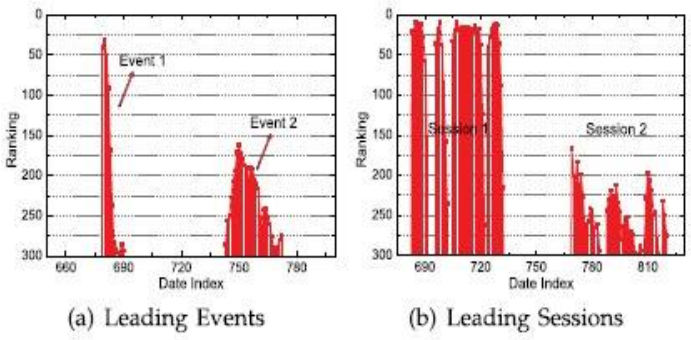


Figure 2: (a) Example of leading events; (b) Example of leading sessions of mobile Apps.

The formal definition of leading session is as follows intuitively, the leading sessions of a mobile App represent its periods of popularity, so the ranking manipulation will only take place in these leading sessions. Therefore, the problem of detecting ranking fraud is to detect fraudulent leading sessions. Along this line, the first task is how to mine the leading sessions of a mobile App from its historical ranking records.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

3.2 EXTRACTING EVIDENCES FOR RANKING FRAUD DETECTION

In this section, we study how to extract and combine fraud evidences for ranking fraud detection.

3.2.1 Ranking Based Evidences

By investigating the Apps' chronicled positioning records, we watch that Apps' positioning practices in a main occasion dependably fulfill a particular positioning example, which comprises of three diverse positioning stages, to be specific, rising stage, keeping up stage and retreat stage. In particular, in every driving occasion, an App's positioning first increments to a top position in the leaderboard (i.e., rising stage), then keeps such top position for a period (i.e., looking after stage), lastly diminishes till the end of the occasion (i.e., retreat stage). Fig. 3 demonstrates a case of various positioning periods of a main occasion. For sure, such a positioning example demonstrates an essential comprehension of driving occasion. In the accompanying, we formally characterize the three positioning periods of a main occasion.

3.2.2 Rating Based Evidences

The positioning based proofs are helpful for positioning misrepresentation identification. Be that as it may, some of the time, it is not adequate to just utilize positioning based proofs. For instance, some Apps made by the celebrated designers.

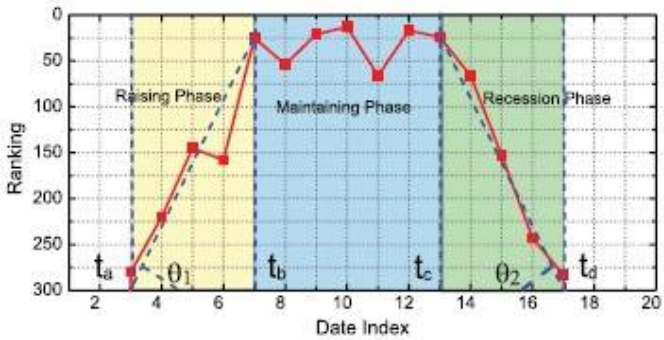
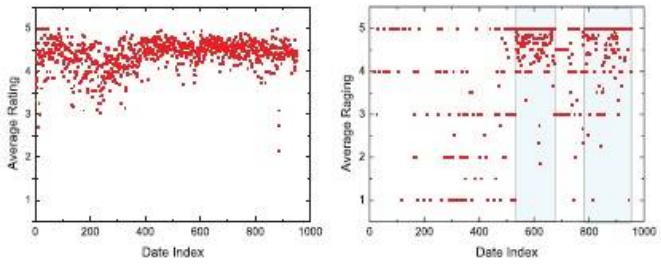


Figure 3: An example of different ranking phases of a leading event.



(a) Example 1 (b) Example 2

Figure 4: Two real-world examples of the distribution of Apps' daily average ratings.

Estimations of u1 because of the engineers' believability and the "informal" promoting impact. Additionally, a percentage of the lawful promoting administrations, for example, "restricted time markdown", may likewise bring about critical positioning based proofs. To explain this issue, we likewise concentrate how to concentrate misrepresentation confirmations from Apps' verifiable rating records. In particular, after an App has been distributed, it can be appraised by any client who downloaded it. Without a doubt, client rating is a standout amongst the most essential components of App promotion. An App which has higher rating may draw in more clients to download and can likewise be positioned higher in the leaderboard. Therefore, appraising control is likewise a vital point of view of positioning extortion. Naturally, if an App has positioning extortion in a main session s, the evaluations amid the time of s may have oddity designs contrasted and its authentic appraisals, which can be utilized for building rating based proofs. For instance, Figs. 4a and 4b demonstrate the appropriations of the day by day normal rating of a well-known App "WhatsApp" and a suspicious App found by our methodology, individually. We can watch that an ordinary App

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

dependably gets comparable normal rating every day, while a deceitful App may get moderately higher normal evaluations in some time periods (e.g., driving sessions) than different times.

3.2.3 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 perceptions and usage experiences of existing users for particular mobile Apps. Indeed, review manipulation is one of the most important perspectives of App ranking fraud. Specifically, before downloading or purchasing a new mobile App, users often first read its historical reviews to ease their decision making, and a mobile App contains more positive reviews may attract more users to download. Therefore, imposters often post fake reviews in the leading sessions of a specific App in order to inflate the App download, and thus propel the App's ranking position in the leaderboard.

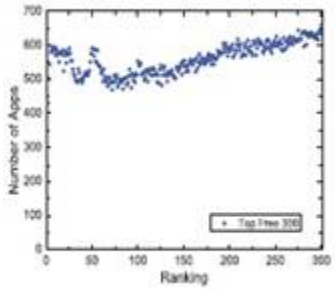


Figure 5: The distribution of the number of Apps w.r.t different rankings.

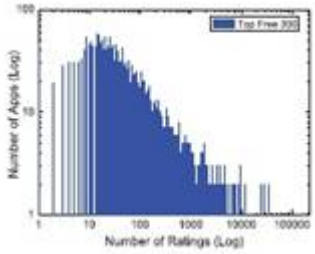


Figure 6: The distribution of the number of Apps w.r.t different numbers of ratings.

Although some previous works on review spam detection have been reported in recent years [14], [19], [21], the problem of detecting the local anomaly of reviews in the leading sessions and capturing them as evidences for ranking fraud detection are still under-explored.

To this end, here we propose two fraud evidences based on Apps' review behaviors in leading sessions for detecting ranking fraud.

3.2.4 Evidence Aggregation

After extracting three types of fraud evidences, the next challenge is how to combine them for ranking fraud detection. Indeed, there are many ranking and evidence aggregation methods in the literature, such as permutation based models [16], [17], score based models [15], [21] and Dempster Shafer rules [4], [20]. However, some of these methods focus on learning a global ranking for all candidates. This is not proper for detecting ranking fraud for new Apps. Other methods are based on supervised learning techniques, which depend on the labeled training data and are hard to be exploited. Instead, we propose an unsupervised approach based on fraud similarity to combine these evidences.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

V. EXPERIMENTAL RESULTS

In this section, we evaluate the performances of ranking scam detection using real-world App data.

3.3 The Experimental Data

The experimental data sets were collected from the Top Free leaderboards of Google App Store. The data sets contain the daily chart rankings. Moreover, each data set also contains the user ratings and review information. Thedistributions of the number of Apps with respect to different rankings are collectin data sets. There are number of Apps with low rankings and some have high rankings. The distribution of the number of Apps with respect to different number of ratings is collect in data sets. The distribution of App ratings is not even, which indicates that only a small percentage of Apps are very popular.

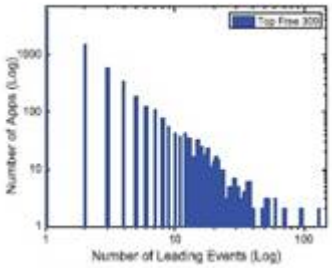


Figure 7: The distribution of the number of Apps w.r.t different numbers of leading events.

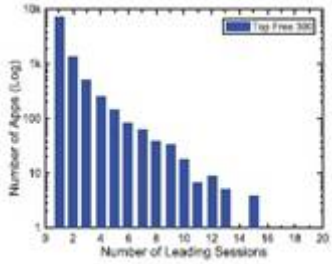


Figure 8: The distribution of the number of Apps w.r.t different number of leading sessions.

3.4 Mining Leading Sessions

Here, we demonstrate the results of mining leading sessions in data sets. Specifically, in Algorithm 1, we set the ranking threshold k* and threshold ϕ. This denotes two adjacent leading events can be segmented into the same leading session if they occur within one week of each other. The distributions of the number of Apps with respect to different numbers of contained leading events and leading sessions in both data sets.We can see that only a few Apps have many leading events and leading sessions. The distribution of the number of leading sessions with respect to different numbers of contained leading events in both data sets. We can find only a few leading sessions contain many leading events.

VI. FUTURE WORK

The limitation can be overcome very easily. One of the available methods is to have a password mechanism. In password mechanism the user will be provided with a unique code or password. When network moderator has a doubt of misuse of a particular mobile the user can be asked for password confirmation.

VII. CONCLUSION

In this paper, we added to a positioning misrepresentation identification framework for portable Apps. In particular, we initially demonstrated that positioning extortion happened in driving sessions and gave a strategy to digging driving

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

sessions for each App from its authentic positioning records. At that point, we distinguished positioning based confirmations, rating based proofs and audit based confirmations for identifying positioning misrepresentation. In addition, we proposed a streamlining based accumulation strategy to incorporate every one of the confirmations for assessing the validity of driving sessions from versatile Apps. A novel point of view of this methodology is that every one of the confirmations can be displayed by measurable theory tests, subsequently it is anything but difficult to be stretched out with different proofs from area learning to recognize positioning extortion. At long last, we accept the proposed framework with broad tests on certifiable App information gathered from the Google's App store. Exploratory results demonstrated the adequacy of the proposed approach.

REFERENCES

[1] A.Ntoulas, M. Najork, M. Manasse, and D. Fetterly, "Detecting spam web pages through content analysis," in Proc. 15th Int. Conf.World Wide Web, 2006, pp. 83–92.
[2] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in Proc. 19thACMInt. Conf. Inform. Knowl. Manage, 2010, pp. 939–948.
[3] H. Zhu, E. Chen, K. Yu, H. Cao, H. Xiong, and J. Tian, "Mining personal context-aware preferences for mobile users," in Proc. IEEE 12th Int. Conf. Data Mining, 2012, pp. 1212–1217.
[4] 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.
[5] Jindal and B. Liu, "Opinion spam and analysis," in Proc. Int.Conf. Web Search Data Mining, 2008, pp. 219–230.
[6] Klementiev, D. Roth, and K. Small, "An unsupervised learning algorithm for rank aggregation," in Proc. 18th Eur. Conf. Mach.Learn., 2007, pp. 616–623.
[7] Hengshu Zhu, HuiXiong, Yong Ge, and EnhongChen,"Discovery of Ranking Fraud for Mobile Apps" in Proc IEEE Jan 2015.
[8] N.Spirin and J. Han, "Survey on web spam detection: Principles and algorithms," SIGKDD Explor. Newslett., vol. 13, no. 2, pp. 50–64, May 2012.
[9] B. Zhou, J. Pei, and Z. Tang, "A spamicity approach to web spam detection," in Proc. SIAM Int. Conf. Data Mining, 2008, pp. 277–288.
[10] H. Zhu, H. Cao, E. Chen, H. Xiong, and J. Tian, "Exploiting enriched contextual information for mobile app classification," in Proc. 21stACMInt. Conf. Inform. Knowl. Manage., 2012, pp. 1617–1621.
[11] Z. Wu, J. Wu, J. Cao, and D. Tao, "HySAD: A semi-supervised hybrid shilling attack detector for trustworthy product recommendation," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012, pp. 985–993.
[12] S. Xie, G. Wang, S. Lin, and P. S. Yu, "Review spam detection via temporal pattern discovery," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012, pp. 823–831.
[13] B. Yan and G. Chen, "AppJoy: Personalized mobile applicationdiscovery," in Proc. 9th Int. Conf. Mobile Syst., Appl., Serv., 2011, pp. 113–126.
[14] K. Shi and K. Ali, "Getjar mobile application recommendations with very sparse datasets," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012, pp. 204–212.
[15] 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.
[16] 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.
[17] A. Klementiev, D. Roth, K. Small, and I. Titov, "Unsupervised rank aggregation with domain-specific expertise," in Proc. 21st Int. Joint Conf. Artif. Intell., 2009, pp. 1101–1106.
[18] N. Jindal and B. Liu, "Opinion spam and analysis," in Proc. Int. Conf. Web Search Data Mining, 2008, pp. 219–230.
[19] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, "Spotting opinion spammers using behavioral footprints," in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013, pp. 632–640.
[20] G. Shafer, A Mathematical Theory of Evidence. Princeton, NJ, USA: Princeton Univ. Press, 1976.
[21] M. N. Volkovs and R. S. Zemel, "A flexible generative model for preference aggregation," in Proc. 21st Int. Conf. World Wide Web, 2012, pp. 479–488.