www.ijecs.in

International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issues 6 June 2016, Page No. 16830-16835

Efficient Retrival Of Mobile Apps Using EIRQ

***R.Vinodharasi*[1], *P.Ramadoss*[2]**

[1]M.E. Student, Department of CSE, Parisutham Institute of Technology and Science, Tamil Nadu, India [2]Asst. Professor, Department of CSE, Parisutham Institute of Technology and Science, Tamil Nadu, India
[1]rvinodharasi@gmail.com,[2]vp.ramadoss@gmail.com

***Abstract*— Mining the needed data based on our application was the crucial activity in the computerized environment.For that mining techniques was introduced.This project used to extract the mobile apps.The Ranking fraud in the mobile App market refers to fraudulent or deceptive activities which have a purpose of bumping up the Apps in the popularity list. Indeed, it becomes more and more frequent for App developers to use shady means, such as inflating their Apps' sales or posting phony App ratings, to commit ranking fraud. Here first propose to accurately locate the ranking fraud by mining the active periods, namely leading sessions, of mobile Apps. Furthermore, we investigate three types of evidences, i.e., ranking based evidences, rating based evidences and review based evidences, by modeling Apps' ranking, rating and review behaviors through statistical mining based hypotheses tests. In addition, In this project an optimization based application used to integrate all the evidences for fraud detection based on EIRQ (efficient information retrieval for ranked query) algorithm. Finally, evaluate the proposed system with real-world App data collected from the iOS App Store for a long time period. Experiment was need to be done for validate the effectiveness of the proposed system, and show the scalability of the detection algorithm as well as some regularity of ranking fraud activities.**

Keywords- *Mobile Apps, ranking fraud detection, evidence aggregation, historical ranking records, rating and review.*

1.INTRODUCTION

Data Mining ,it is an Extraction of concealed, prescient data from huge databases .It is likewise called as Knowledge Discovery from Databases (KDD).It perform an Identification and assessment of shrouded examples in database. It is capable innovation with incredible potential to help associations to find and produce data from their information stockrooms. Information mining instruments anticipate future patterns and practices. It help association to settle on proactive learning driven choices, they get ready databases for distinguishing shrouded designs furthermore robotizes the identification of applicable examples in a database, utilizing characterized methodologies and calculations to investigate present and authentic information that can then be examined to anticipate future patterns. Since information mining apparatuses anticipate future patterns.

To mine such kind of information there are number of information mining devices are accessible. As an aftereffect of this, it has turned out to be fairly troublesome for an obscure client to choose the most ideal information digging device for his work. This paper shows a review of information mining with the strides incorporated into mining information and the diverse information mining strategies and it likewise gives the per user the correlations investigation of different openly accessible information mining apparatuses, for example, WEKA instrument, Rapid Miner device and Net Tool Spider for web mining accessible today with their own particular qualities and shortcomings.

Information mining alludes to extricating or "mining" learning from a lot of information. It is additionally called Knowledge-Discovery in Databases (KDD) or Knowledge-Discovery and Data Mining, is the procedure of consequently looking expansive volumes of information for examples, for example, affiliation rules. It applies numerous more established computational systems from measurements, data recovery, machine learning and example acknowledgment. Taking after are the information mining steps.

Information Cleaning:In the initial step, information that contain undermined or purge records are evacuated. Information Integration:In request to continue with information mining, information should be gathered and integratedinto a solitary designed structure. Be that as it may, diverse wellsprings of information as a rule don't give uniform structures and translations of information; subsequently mix into a solitary organization needs to take place.DataSelection:Not the greater part of the information gathered are required however. Information choice takes into account picking just such information that are pertinent to the undertaking to be performed.

Information Transformation:The information that have passed the cleaning step are still not prepared for information miningpurposes, for despite everything they should be changed into arrangement acknowledged by the information mining algorithm.DataMining:In this stride, different calculations might be connected on the information keeping in mind the end goal to discoverpotential learning covered up inside of the information. Design Evaluation:The significance of results gave by information mining should be assessed, for not the majority of the discoveries might be of enthusiasm to the request. Repetitive examples are hence uprooted. Learning Presentation:Results that seem, by all accounts, to be the most vital experience change and perception with a specific end goal to be introduced in the most justifiable structure.

The fascinating, helpful (possibly valuable and already obscure principles and examples) data can be separated from these huge data vaults. Specialists regard Data mining as the fundamental procedure of Knowledge Discovery in Database (KDD). It is otherwise called extraction of data, information/design examination, information prehistoric studies, information digging, data collecting and business knowledge. Visit thing set mining prompts the disclosure of affiliations and connections among things in huge value-based or social datasets. The conventional calculations for mining affiliation rules based on parallel characteristicsdatabases. An efficient algorithm should reduce the I/O operation of the process of mining by means of decreasing the times of database searching.

2 OVERVIEW

2.1PROBLEM STATEMENT

Application engineers resort to some fake intends to intentionally help their Apps and in the long run control the outline rankings on an App store. This is generally executed by utilizing purported bot ranches utilizing copy surveys. Extortion recognition incorporates observing of the spending conduct of clients with a specific end goal to determination, discovery, or evasion of undesirable conduct. As utilization of cellular telephone turns into the most winning method of installment for both online and additionally customary buy, extortion relate with it are likewise quickening. Misrepresentation location is worried with not just catching the false application like surveys like copy ID.

Conceivable access to this individual data by unapproved parties puts clients at danger, and this is not where the dangers end. These gadgets incorporate numerous sensors and are about dependably with us, giving profound experiences into our advanced lives as well as our physical lives. The GPS unit can tell precisely where you are, while the receiver can record sound, and the camera can record pictures. Furthermore, cell phones are regularly connected specifically to some money related dangers, by means of SMS messages, telephone calls, and information arranges, which can affect a client's month to month bill, or progressively, as a way to validate to a bank or straightforwardly connection to a budgetary record through a 'computerized wallet'. This entrance implies that any application (or application) that is permitted to keep running on the gadgets possibly can take advantage of specific parts of the data.

As of late brilliant cell phones have gotten to be pervasive. More than 50 percent of every cell telephone are currently smartphones,and this measurement does not represent different gadgets, for example, tablet PCs that are running comparable versatile working frameworks. By, more than 400 million Android gadgets were enacted in 2012 alone. Android gadgets have across the board appropriation for both individual and business use. From kids to the elderly, beginners to master , and in various societies around the globe, there is a fluctuated client base for cell phones.

The pervasive use of these cell phones postures new protection and security dangers. Our whole advanced lives are regularly put away on the gadgets, which contain contact records, email messages, passwords, and access to documents put away privately.furthermore, in the cloud. Conceivable access to this individual data by unapproved parties puts clients at danger, and this is not where the dangers end. These gadgets incorporate numerous sensors and are almost dependably with us, giving profound bits of knowledge into our advanced lives as well as our physical lives. The GPS unit can tell precisely where you are, while the amplifier can record sound, and the camera can record pictures.

Also, cell phones are frequently connected straightforwardly such real-time ranking frauds also can bedetected by the proposed approach of fraud detection system. to some money related dangers, by means of SMS messages, telephone calls, and information arranges, which can affect a client's month to month bill, or progressively, as a way to validate to a bank or specifically connection to a monetary record through a 'computerized wallet'. This entrance implies that any application (or application) that is permitted to keep running on the gadgets conceivably can take advantage of certain

parts of the data. In the kindhearted case the entrance is performed to give helpful functionalities, however in different situations it might be utilized to gather a lot of individual data and even as a way to have some unfriendly effect on a client. Moreover, the line in the middle of considerate and vindictive is regularly fluffy, with numerous applications falling into a hazy area where they might be excessively intrusive however not altogether malevolent.

Prescribing the most exceedingly terrible application to the client is not a decent process.For maintaining a strategic distance from that need to create one procedure for finding a decent versatile apps.Ranking misrepresentation in the portable App market alludes to fake or beguiling exercises which have a reason for knocking up the Apps in the ubiquity list. Without a doubt, it turns out to be progressively and morefrequent for App designers to utilize shady means, for example, expanding their Apps' deals or posting imposter App evaluations, to submit positioning misrepresentation.

2.2EXISTING SYSTEM

Generally speaking, the related works of this study can be grouped into three categories. Mobile App ranking spam detection. Specifically, the web ranking spam refers to any deliberate actions which bring to selected webpages an unjustifiable favorable relevance or importance. Spam detection is mainly based on the analysis of ranking principles of search engines, such as PageRank and query risk score information. This is different from ranking fraud detection for mobile Apps.

Detecting online review spam. Specifically, they solved this problem by detecting the co-anomaly patterns in multiple review based time series. However, to the best of our knowledge, none of previous works has studied the problem of ranking fraud detection for mobile Apps.

The main drawback of the existing system was malware maybe attacked during the downloading of app.Second,The app may be downloaded based on the fake references.Another one was No alerted occurring during download malware file.

3 PROPOSED SYSTEM

The instant download information of each mobile App is often not available for analysis. The App developers themselves are also reluctant to release their download information for various reasons. Therefore, in this paper, we mainly focus on ranking, rating and review records for ranking fraud detection using EIRQ (efficient information retrieval for ranked query) algorithm. Sometimes we need to detect such ranking fraud from Apps' current ranking observations.

Actually, given the current ranking now of an App, we can detect ranking fraud for it in two different cases.The algorithm EIRQ which was used to develop this project. This project also helpful in finding and block the datas.

This is the main adavantage of implementing the malware when application was downloaded. This project was also helpful in choosing the best app for oueneed.The additional advantage of this was providing the security for using the application

3.1DEVELOPER UPLOADING APP TO SERVER

A server is both a running occasion of some product equipped for tolerating demands from customers, and the PC such a server keeps running on. In today's innovation parcel of portable application for informing, scanning, altering and so forth yet this application might be made by the application designer and transferred by the server. In server part application arrives taking into account the classifications. It might be with unique example rights or with copy malware application.

The application engineer need to transfer the application to the server.The server need to keep up the procces of the designer that is the perspectives of the application likewise the downloading process.The engineer was check with the application advance intermittently.

3.2APPLICATION REVIEW BY DEVELOPER ARMIES

First, the download information is an important signature for detecting ranking fraud, since ranking manipulation is to use so-called "bot farms" or "human armies" to inflate the App downloads and ratings in a very short time. However, the instant download information of each mobile App is often not available for analysis. Every application has some historical data due to the based on the reviews and response of the users. The review may be uploaded by the users or developer by fake ID. The App developers themselves are also reluctant to release their download information for various reasons for introduce the applications. Therefore, in this paper, we mainly focus on extracting evidences from Apps' historical ranking, rating and review records for reach the application to people usage priority increasing.

4.ARCHITECTURE AND METHODOLOGY

4.1EIRQ-Privacy Scheme

The working process of EIRQ-Privacy is similar to the main differences lie in the Matrix Construct and File Filter algorithms. Intuitively, EIRQ-Privacy adopts one buffer, with different mapping times for files of different ranks. Let i denote the mapping times for a Rank-I query, and let l be the highest rank of queries that choose the i-th keyword in the dictionary. The mask matrix M is a d-row and m-column matrix, where d is the number of keywords in the dictionary, and m ¼ max.

The Matrix Construct algorithm constructs M in the following way: for the i-th row of M that corresponds to Dic½i, the ADL sets M½i; 1; . . .;M½i; l to 1, and M½i; l þ 1; . . . ; M½i;m to 0, and then encrypts each element under its public key. Note that for a row that corresponds to a Rank-l keyword, the ADL sets the first l elements, rather than random l elements, to 1.

The reason is to ensure that, given any Rank-l file, when wemultiply the rows that correspond to file keywords together in a element-by-element way, the resulting row contains l elements whose values are larger than.

4.2ARCHITECTURE

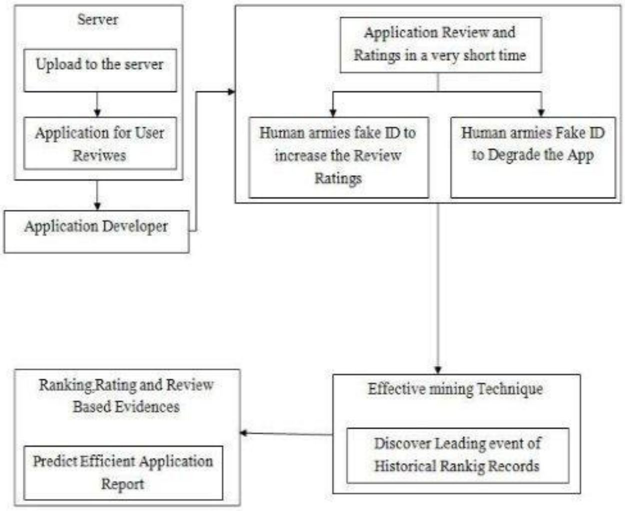
Server
Upload to the server
Application for User Reviwes
Application Developer
Application Review and Ratings in a very short time
Human armies fake ID to increase the Review Ratings
Human armies Fake ID to Degrade the App
Ranking,Rating and Review Based Evidences
Predict Efficient Application Report
Effective mining Technique
Discover Leading event of Historical Rankig Records

Fig1:System Architecture

5 EXPERIMENT RESULT

5.1 THE EXPERIMENTAL DATA

The test information sets were gathered from the "Best Free 300" and "Top Paid 300" leaderboards of Apple's Application Store (U.S.) from February 2, 2010 to September 17, 2012. The information sets contain the every day diagram rankings1 of top 300 free Apps and main 300 paid Apps, individually. Besides, every information set additionally contains the client appraisals and audit data.

Demonstrate the appropriations of the quantity of Apps concerning diverse rankings in these information sets. In the figures, we can see that the quantity of Apps with low rankings is more than that of Apps with high rankings. Besides, the rivalry between free Apps is more than that between paid Applications, particularly in high rankings (e.g., main 25 demonstrate the circulation of the quantity of Apps with deference to various number of evaluations in these information sets. In the figures, we can see that the circulation of App evaluations is not, which demonstrates that just a little rate of Apps are exceptionally well known.

5.2 HUMAN JUDGEMENT BASED EVALUATION

To the best of our insight, there is no current benchmark to choose which driving sessions or Apps truly contain positioning misrepresentation. Therefore, we create four instinctive baselines and welcome five human evaluators to accept the adequacy of our methodology Evidence Aggregation based Ranking Fraud Detection (EA-RFD). Especially, we mean our methodology with score based total (i.e., Principle 1) as EA-RFD-1, and our methodology with rank based accumulation (i.e., Principle 2) as EA-RFD-2, individually.

5.3 BASELINES

The first baseline Ranking-RFD stands for ranking evidence based ranking fraud detection, which estimates ranking fraud for each leading session by only using ranking based evidences (i.e., C1 to C3). These three evidences are integrated by our aggregation approach.

The second baseline Rating-RFD stands for Rating evidence based ranking fraud detection, which estimates the ranking fraud for each leading session by only using rating based evidences (i.e., C4 and C5). These two evidences are integrated by our aggregation approach.effectiveness of different kinds of evidences, and our preliminary experiments validated that baselines with Principle 2 always outperform baselines with Principle 1. The last baseline E-RFD stands for evidence based ranking fraud detection, which estimates the ranking fraud for each leading session by ranking, rating and review based evidences without evidence aggregation. Specifically, it ranks leading sessions by Equation (18), where each wi is set to be 1=7 equally. This baseline is used for evaluating the effectiveness of our ranking aggregation method. Note that, according to Definition 3, we need to define some ranking ranges before extracting ranking based evidences for EA-RFD-1, EA-RFD-2, Rank-RFD and E-RFD. In our experiments, we segment the rankings into five different ranges, i.e., ½1; 10_, ½11; 25_, ½26; 50_, ½51; 100_, ½101; 300_, which are commonly used in App leaderboards. Furthermore, we use the LDA model to extract review topics as introduced in Section 3.3. Particularly, we first normalize each review by the Stop-Words Remover [6] and the Porter Stemmer [7]. Then, the number of latent topic Kz is set to 20 according to the perplexity based estimation approach.

5.4 PERFORMANCE

In this area, we show the general exhibitions of every positioning extortion location approach concerning differentb assessment measurements, i.e., Precision@K, Recall@K, F@k, and NDCG@K. Especially, here we set the most extreme K to be 200, and all examinations are led on a 2.8 GHZ2 quad-center CPU, 4G primary memory PC. Figs. 12 and 13 demonstrate the assessment execution of every identification approach in two information sets. From these figures we can watch that the assessment results in two information sets are steady. In reality, by breaking down the assessment results, we can acquire a few shrewd perceptions. In particular, to start with, we find that our methodology, i.e., EA-RFD-2/EA-RFD-1, reliably outflanks different baselines and the upgrades are more critical for littler K (e.g., K < 100). This outcome plainly accepts the adequacy of our confirmation conglomeration based system for identifying positioning extortion. Second, EA-RFD-2 beats EA-RFD-1 sightly as far as all assessment measurements, which demonstrates that rank based total (i.e., Principle 2) is more successful than score based accumulation (i.e., Principle 1) for coordinating extortion confirmations. Third, our methodology reliably outflanks E-RFD, which accepts the viability of confirmation aggradation for distinguishing positioning extortion. Fourth, E-RFD have preferred discovery execution over Ranking-RFD, Rating-RFD and Review-RFD.

This shows utilizing three sorts of confirmations is more powerful than just utilizing one kind of proofs, regardless of the fact that without confirmation collection. At long last, by looking at Ranking-RFD, Rating-RFD and Review-RFD, we can watch that the positioning based confirmations are more powerful than rating and audit based proofs. It is on the grounds that rating and audit controls are just supplementary to positioning control. Especially, we watch that Review-RFD will most likely be unable to prompt the great execution as far as all assessment measurements on the two information sets. A conceivable explanation for this marvel is that audit control (i.e., fake-positive surveys) does not specifically influence the graph positioning of Apps, but rather might build the likelihood of blowing up App downloads and appraisals.

In this way, the audit control does not as a matter of course result in positioning misrepresentation because of the obscure positioning standards in the App Store. In any case, the proposed audit based confirmations can be useful as supplementary for positioning extortion identification. Really, in our preparatory examinations, we found that the audit based proofs could simply enhance the identification exhibitions while being utilized together with different confirmations. This unmistakably accepts the adequacy of the survey based confirmations. To encourage approve the trial results, we additionally lead a progression of matched T-test of 0.95 certainty level which demonstrate that the upgrades of our methodology, i.e., EA-RFD-2/EA-RFD-1, on all assessment measurements with various K contrasted with different baselines are all measurably huge.

6 RELATED WORK

Generally speaking, the related works of this study can be grouped into three categories. The first category is about web ranking spam detection. Specifically, the web ranking spam refers to any deliberate actions which bring to selected webpages an unjustifiable favorable relevance or importance [30].

For example, Ntoulas et al. [22] have studied various aspects of content-based spam on the web and presented a number of heuristic methods for detecting content based spam. Zhou et al. [30] have studied the problem of unsupervised web ranking spam detection. Specifically, they proposed an efficient online link spam and term spam detection methods using spamicity.

Recently, Spirin and Han [25] have reported a survey on web spam detection, which comprehensively introduces the principles and algorithms in the literature. Indeed, the work of web ranking spam detection is mainly based on the analysis of ranking principles of search engines, such as PageRank and query term frequency. This is different from ranking fraud detection for mobile Apps. The second category is focused on detecting online review spam. For example, Lim et al. [19] have identified several representative behaviors of review spammers and model these behaviors to detect the spammers. Wu et al. [27] have studied the problem of detecting hybrid shilling attacks on rating data. The proposed approach is based on the semisupervised learning and can be used for trustworthy product recommendation. Xie et al. [28] have studied the problem of singleton review spam detection. Specifically, they solved this problem by detecting the co-anomaly patterns in multiple review based time series. Although some of above 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

Finally, the third category includes the studies on mobile App recommendation. For example, Yan and Chen [29] developed a mobile App recommender system, named Appjoy, which is based on user's App usage records to build a preference matrix instead of using explicit user ratings. Also, to solve the sparsity problem of App usage records, Shi and Ali [24] studied several recommendation models and proposed a content based collaborative filtering model, named Eigenapp, for recommending Apps in their website Getjar. In addition, some researchers studied the problem of exploiting enriched contextual information for mobile App recommendation. For example, Zhu et al. [32] proposed a uniform framework for personalized context-aware recommendation, which can integrate both context independency and dependency assumptions.

However, to the best of our knowledge, none of previous works has studied the problem of ranking fraud detection for mobile Apps.

7 CONCLUSIONS AND FUTURE WORK

In this paper, we developed a ranking fraud detection system for mobile Apps. Specifically, we first showed that ranking fraud happened in leading sessions and provided a method for mining leading sessions for each App from its historical ranking records. Then, we identified ranking based evidences, rating based evidences and review based evidences for detecting ranking fraud. Moreover, we proposed an optimization based aggregation method to integrate all the evidences for evaluating the credibility of leading sessions from mobile Apps. An 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. Finally, we validate the proposed system with extensive experiments on real-world App data collected from the Apple's App store. Experimental results showed the effectiveness of the proposed approach.

In the future, we plan to study more effective fraud evidences and analyze the latent relationship among rating, review and rankings. Moreover, we will extend our ranking fraud detection approach with other mobile App related services, such as mobile Apps recommendation, for enhancing user experience.

REFERENCES

[1] (2014). [Online]. Available: http://en.wikipedia.org/wiki/ cohen's_kappa

[2] (2014). [Online]. Available: http://en.wikipedia.org/wiki/ information_retrieval

[3] (2012). [Online]. Available: https://developer.apple.com/news/ index.php?id=02062012a

[4] (2012). [Online]. Available: http://venturebeat.com/2012/07/03/ apples-crackdown-on-app-ranking-manipulation/

[5] (2012). [Online]. Available: http://www.ibtimes.com/applethreatens- crackdown-biggest-app-store-ranking-fra ud-406764

[6] (2012). [Online]. Available: http://www.lextek.com/manuals/ onix/index.html

[7] (2012). [Online]. Available: http://www.ling.gu.se/lager/ mogul/porter-stemmer.

[8] L. Azzopardi, M. Girolami, and K. V. Risjbergen,

"Investigating the relationship between language model perplexity and ir precision- recall measures," in Proc. 26th Int. Conf. Res. Develop. Inform. Retrieval, 2003, pp. 369–370.

[9] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., pp. 993–
1022, 2003.

[10] 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.

[11] 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.

T. L. Griffiths and M. Steyvers, "Finding scientific topics," Proc. Nat. Acad. Sci. USA, vol. 101, pp.
5228–5235, 2004.

[12] G. Heinrich, Parameter estimation for text analysis, "

Univ. Leipzig, Leipzig, Germany, Tech. Rep., http://faculty.cs.byu.edu/~ringger/ CS601R/papers/Heinrich-GibbsLDA.pdf, 2008.

[13] N. Jindal and B. Liu, "Opinion spam and analysis," in Proc. Int. Conf. Web Search Data Mining, 2008, pp. 219–230.

[14] J. Kivinen and M. K. Warmuth, "Additive versus exponentiated gradient updates for linear prediction," in Proc. 27th Annu. ACM Symp. Theory Comput., 1995, pp. 209–218.

[15] Klementiev, D. Roth, and K. Small, "An unsupervised learning algorithm for rank aggregation," in Proc. 18th Eur. Conf. Mach. Learn., 2007, pp. 616–623.

[16] 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] 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] 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.

[20] Y.-T. Liu, T.-Y. Liu, T. Qin, Z.-M. Ma, and H. Li,
"Supervised rank aggregation," in Proc. 16th Int.
Conf. World Wide Web, 2007, pp. 481–490.

[21] 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.

[22] 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.

[23] G. Shafer, A Mathematical Theory of Evidence. Princeton, NJ, USA: Princeton Univ. Press, 1976.

[24] 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.

[25] 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.

[26] 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.