International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No. (01), January 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

LEVERAGING CROWD SOURCING FOR EFFICIENT MALICIOUS USERS DETECTION IN LARGE-SCALE SOCIAL NETWORKS

R.SHANTHI BRABHA[1], M.CHANDRALEKHA[2]

[1]Assistant Professor, Department of Computer science,

Sri Adi Chunchanagiri Women's College, (India)

[2] Research Scholar, Department of Computer science,

Sri Adi Chunchanagiri Women's College, (India)

ABSTRACT

Crowd sourcing is increasingly being used as a means to tackle problems requiring human intelligence. With the ever growing worker base that aims to complete micro tasks on crowd sourcing platforms in exchange for financial gains, there is a need for stringent mechanisms to prevent exploitation of deployed tasks. Quality control mechanisms need to accommodate a diverse pool of workers, exhibiting a wide range of behavior. A pivotal step towards fraud-proof task model is understanding the behaviors patterns of microtask workers. In this paper, we predict the prevalent malicious activity on crowd sourcing platforms and study the behavior exhibited by trustworthy and untrustworthy workers, particularly on crowd sourced surveys. Based on our analysis of the typical malicious activity, we define and identify different types of workers in the crowd, propose a method to measure malicious activity.

Keywords: Crowd sourcing, malicious user's detection, large scale networks.

I.INTRODUCTION

Crowd sourcing has gained quick popularity, because of the data-intensive nature of emerging tasks, requiring validation, evaluation and annotation of large volumes of data. While developing a sound definition of crowd sourcing, Estelles and Guevara suggest that micro tasks are of variable complexity and modularity, and entail mutual benefit to the worker and the requester. Gathering small contributions through such micro tasks facilitates the accomplishment of work that is not easily automatable, through rather minor contributions of each individual worker. With the universality of the internet, it became possible to distribute tasks at global scales, leading to the recent success of crowd sourcing, being later defined as an online, distributed problem-solving and production model. In the recent past, there has been a considerable amount of work towards developing appropriate platforms and suggesting frameworks for efficient crowd sourcing an increasing large amount of research communities benefit from using crowd sourcing platforms in order to either gather distributed and unbiased data, to validate results, evaluate aspects, or to build ground truths. While the demand for using crowd sourcing to solve several problems is on an upward climb, there are some obstacles that hinder requesters from attaining reliable, transparent, and non-skewed results. Herein, a primary nuisance is introduced through malicious workers, understood by as workers with ulterior motives, who either simply sabotage a task or try to quickly attain task completion for monetary gains.

Gold-standards are the typically adopted solution to improve task performance. In general practice, gold-standards are questions where answers are known apriori to the task administrators. Thus, if a worker fails to provide the correct answer for a particular question, he is automatically flagged as an untrustworthy worker. However, with the success of crowd sourcing market, we believe that malicious activities and adversarial approaches will also become more advanced and popular, overcoming common gold standards. Quality control mechanisms should thereby account for a diverse pool of workers that exhibit a wide range of behavioral patterns. Methods have been considered and used in order to tackle poor worker performance in the past. However, there is a need to understand the behavior of these workers and the kinds of malicious activity they bring about in crowdsourcing platforms. In this paper, we present our work towards analyzing the behavior of malicious microtask workers, and reflect on guidelines to overcome such workers in the context of online surveys.

II.LITERATURE REVIEW

Behrend et al. showed the suitability of crowdsourcing asan alternative data source for organizational psychology research [11]. Kittur et al. promoted the suitability of crowdsourcing user studies, while cautioning that special attention should be given to the task formulation [12]. Even though these works outline shortcomings of using crowdsourcing, they do not consider the impact of malicious activity that can emerge in differing ways. In our work, we show that varying types of malicious activity is prevalent in crowdsourced surveys, and propose measures to curtail such behavior. Marshall et al. profiled Turkers who take surveys, and examined the characteristics of surveys that may determine the data reliability [13].

Related to their work, we adopt the approach of collecting data through crowdsourced surveys in order to draw meaningful insights. Our analysis quantitatively and qualitatively extends their work, and additionally provides a sustainable classification of malicious workers that sets precedents for an extension to different categories of micro tasks. Through their work, Ipeirotis et al. provoked the need for techniques that can accurately estimate the quality of workers, allowing for the rejection or blocking of low-performing workers and spammers [5]. The authors presented algorithms that improve the existing techniques to enable the separation of bias and error rate of the worker. Baba et al. released on their study of methods to automatically detect improper tasks on crowdsourcing platforms [14]. The authors reflected on the importance of controlling the quality of tasks in crowdsourcing marketplaces. Complementing these existing works, our work propels the consideration of both aspects (task design as well as worker behaviour), for effective crowdsourcing. Dow et al. introduced a feedback system for improving the quality of work in the crowd [15]. Oleson et al. present a method to achieve quality control for crowdsourcing, by providing training feedback to workers while relying on programmatic creation of gold data [8]. But for gold-based quality assurance, task administrators need to understand the behavior of malicious workers and anticipate the likely types of worker errors with respect to different types of tasks. Understanding the behavior of workers, is therefore an important objective of this paper. In the realm of studying the reliability and performance of crowd workers with respect to the incentives offered, Mason et al.

investigated the relationship between financial incentives and the performance of the workers [16]. They found that higher monetary incentives increase the quantity of workers but not the quality of work. A large part of their results align with our findings presented in the following sections.

III.WORKER TRAITS, TASKS DESIGN AND METRICS

Researchers in the field have approved the importance and need for techniques to deal with inattentive workers, scammers, incompetent and malicious workers. Ross et al. studied the demographics and usage behaviors characterizing workers on Amazon's Mechanical Turk[17]. Kazai et al. defined types of developers in the crowd by type-casting developers as either sloppy, spammer, incompetent, competent, or diligent [18]. By doing so, the authors expect their insights to help in designing tasks and attracting the best developer to a task. While the authors use developer-performance in order to define these types, we delve into the behavioral patterns of developers.

IV.CONCLUSION

The ubiquity of the internet, allows distributing crowd sourcing tasks that require human intelligence at an increasingly large scale. This field has been gaining rapid popularity, not least because of the data-intensive nature of emerging tasks, requiring validation, evaluation and annotation of large volumes of data. Although certain tasks require human intelligence, humans can exhibit maliciousness that can disrupt accurate and efficient utilization of crowdsourcing platforms. In our work, we aim to understand this phenomenon.

REFERENCES

[1.] E.Estelles-Arolas and F.Gonz alez-Ladr on-de Guevara, "Towards an integrated crowdsourcing definition," *Journal of Information science, vol. 38, no. 2, pp. 189–200, 2012.*

[2.] D.C.Brabham, "Crowdsourcing as a Model for Problem Solving," *Convergence: The International Journal of Research into New Media Technologies, vol. 14, no. 1, pp. 75–90, Feb. 2008.*

[3.] K.Ntalianis, N.Tsapatsoulis, A.Doulamis, and N.Matsatsinis, "Automatic annotation of image databases based on implicit crowdsourcing, visual concept modeling and evolution," *Multimedia Tools and Applications, vol. 69, no. 2, pp. 397–421, 2014.*

[4.] G.Kazai, J.Kamps, and N. Milic-Frayling, "An analysis of human factors and label accuracy in crowdsourcing relevance judgments," Inf. Retr.vol. 16, no. 2, pp. 138–178, 2013.

[5.] P. G. Ipeirotis, F. Provost, and J. Wang, "Quality management on amazon mechanical turk," in Proceedings of the ACM SIGKDD workshop on human computation. ACM, 2010, pp. 64–67.

[6.] 6.C.Eickhoff and A.de Vries, "How crowdsourcable is your task," in Proceedings of the workshop on crowdsourcing for search and data mining (CSDM) at the fourth ACM international conference on web search and data mining (WSDM), 2011, pp. 11–14.

[7.] 7. R.Gennaro, C.Gentry, and B.Parno, "Non-interactive verifiable computing: Outsourcing computation to untrusted workers," in Advances in Cryptology–CRYPTO 2010. Springer, 2010, pp. 465–482.

IJARSE
ISSN: 2319-8354

[8.] 8. D.Oleson, A.Sorokin, G. P.Laughlin, V.Hester, J.Le, and L.Biewald, "Programmatic gold: Targeted and scalable quality assurance in crowdsourcing." *Human computation, vol. 11, p. 11, 2011.*

[9.] 9. D.E.Difallah, G.Demartini, and P.Cudre-Mauroux, "Mechanical cheat: Spamming schemes and adversarial techniques on crowdsourcing platforms." *In CrowdSearch, 2012, pp. 26–30.*

[10.] 10. S.Dow, A.Kulkarni, S.Klemmer, and B.Hartmann, "Shepherding the crowd yields better work," *In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. ACM, 2012, pp. 1013–1022.*

[11.] 11. T.S.Behrend, D.J.Sharek, A.W.Meade, and E.N.Wiebe, "The viability of crowdsourcing for survey research," *Behavior research methods, vol. 43, no. 3, pp.800–813, 2011.*

[12.] 12. A.Kittur, E.H.Chi, and B.Suh, "Crowdsourcing user studies with mechanical turk," *In Proceedings of the SIGCHI conference on human factors in computing systems. ACM, 2008, pp. 453–456.*

[13.] 13. C.C.Marshall and F.M.Shipman, "Experiences surveying the crowd: Reflections on methods, participation, and reliability", *In Proceedings of the 5th Annual ACM Web Science Conference, ser. WebSci 13.New York, NY, USA: ACM, 2013, pp. 234–243.*

[14.] 14. Y.Baba, H.Kashima, K.Kinoshita, G.Yamaguchi, and Y.Akiyoshi, "Leveraging crowdsourcing to detect improper tasks in crowdsourcing marketplaces*," In Twenty-Fifth IAAI Conference, 2013*.

[15.] 15. S.Dow, A.Kulkarni, B.Bunge, T.Nguyen, S.Klemmer, and B.Hartmann, "Shepherding the crowd: managing and providing feedback to crowd workers," *In CHI'11 Extended Abstracts on Human Factors in Computing Systems. ACM, 2011, pp. 1669–1674.*

[16.] 16. W.Mason and D. J.Watts, "Financial incentives and the performance of crowds," *ACM SigKDD Explorations Newsletter, vol. 11, no. 2, pp. 100–108, 2010.*

[17.] 17. J.Ross, L.Irani, M.Silberman, A.Zaldivar, and B.Tomlinson, "Who are the crowdworkers shifting demographics in mechanical turk," *In CHI'10 Extended Abstracts on Human Factors in Computing Systems. ACM, 2010, pp. 2863–2872.*

[18.] 18. G.Kazai, J.Kamps, and N.Milic-Frayling, "Worker types and personality traits in crowdsourcing relevance labels," *In Proceedings of the 20th ACM international conference on Information and knowledge management.ACM, 2011, pp. 1941–1944.*