Time-controlled incentive federated crowdsourcing

Jing Zhang[a,*], Xiaoqian Jiang[b]

[a]School of Cyber Science and Engineering, Southeast University, No. 2 SEU Road, Nanjing 211189, China
[b]School of Cyber Science and Engineering, Southeast University, No. 2 SEU Road, Nanjing 211189, China

Abstract

Keywords:

1. Introduction

Crowdsourcing is widely used to solve the imbalance problems with supervised-learning data sets (e.g., barriers to expertise, regional restriction (Ye et al., 2018; Sigurdsson et al., 2016; Amgad et al., 2022)). There are crowdsourcing platforms such as MTurk, Zooniverse, and Datatang that enable efficient interaction between task publishers and workers to facilitate the execution of crowdsourced tasks. However, research has shown that some labelling tasks that involve sensitive personal information such as physical traits, personality bias, and traces of life may expose workers' privacy (Xia & McKernan, 2020). People's increasing awareness of privacy security makes privacy research in crowdsourcing more and more necessary. State-of-the-art mechanisms to prevent privacy disclosure all come at the expense of the accuracy of crowdsourcing data(e.g.cloaking (Pournajaf et al., 2014; Ren et al., 2022) or inaccuracy(e.g. obfuscation like local differential privacy (Wang et al., 2018; Wei et al., 2019)))(Wang et al., 2020a). Notwithstanding, these mechanisms have to undermine the quality of crowdsourcing, because the original information uploaded by the worker needs to be blurred.

For the above problem, the federated learning (FL) (McMahan et al., 2017) framework is introduced. FL allows multiple clients to collaborate on training shared models by iteratively aggregating model updates without exposing the raw data (Wang et al., 2020b; Gao et al., 2022). In traditional crowdsourcing, the platform centrally processes the data collected by workers (Yu et al., 2020; Tu et al., 2020; Wu et al., 2021; Zhang et al., 2022). However, Centralized platforms are prone to privacy breaches. Fortunately, we can introduce the framework of FL to develop mobile crowdsourcing. As a distributed learning framework, FL allows crowdsourcing workers to process data locally before uploading it. Federate crowdsourcing has been approved well (Li et al., 2020b; Ciftler et al., 2020; Zhang et al., 2021).

*Corresponding author
Email addresses: jingz@seu.edu.cn (Jing Zhang), 220224926@seu.edu.cn (Xiaoqian Jiang)

FL can effectively mitigate privacy breaches in crowdsourcing, but FL requires clients to contribute their data, computing, and communications resources (Zhan et al., 2021a). Without enough motivation, the client will not actively participate in FL. Furthermore, Although FL allows clients to process data locally, malicious third parties can recover part of a participant's data from a server's shared data update (Lyu et al., 2020; Suri et al., 2022). Such potential privacy concern makes even less active participants in FL (Mothukuri et al., 2021). Unless there's enough compensation that clients are willing to take those risks and contribute their resources. Moreover, the clients in FL have absolute control over their own devices and data(Liu et al., 2022). In other words, only the owners of the clients can decide when, where and how to participate in FL (Li et al., 2020a). Therefore, Sufficient incentive can improve the performance of the model by encouraging clients to respond optimally. To sum up, Incentive mechanism is integral to crowdsourcing and FL.

Current research on incentive mechanism focuses on FL, which is designed around the driving factors of clients' contribution, reputation, and resource allocation (Zhan et al., 2020, 2021b; Li et al., 2023). The problem they address is how to measure the contribution of each client and how to attract and retain more clients. However, those incentive mechanisms designed for FL do not work well in federated crowdsourcing. The reasons are as follows:(1) In federated crowdsourcing, clients need to manually label crowdsourcing tasks, so the incentive mechanism also needs to mobilize workers' interest and enthusiasm for crowdsourcing tasks. (2) Due to the heterogeneity of client devices (e.g., computing and communication resources) and individual workers (e.g., knowledge reserve and professional quality), time control should be considered in federated crowdsourcing. (3) Federated crowdsourcing is not done as quickly as possible. It needs to both assess the quality of submitted data to prevent malicious workers from submitting low-quality data for quick rewards, and respond to client delays in updating the model due to emergencies(e.g., the instability of workers, devices, and network). (4) In federal crowdsourcing, platforms need to recruit and retain high-quality workers, workers (clients) need to be paid fairly on time, and task publishers (servers) need to pay as little as possible to maximize their own benefits. The incentive mechanism of federated crowdsourcing must meet the requirements of all three parties at the same time.

To meet above challenges, we put forward Time-controlled incentive federated crowdsourcing (TiFedCrowd) which inspires the client to complete data collection and local model training within the given time to optimize the global model. TiFedCrowd modeled the above issue as a two-stage Stackelberg game (Li & Sethi, 2017) with the time limit for analysis and discussion. In the second stage, TiFedCrowd allots rewards based on the local accuracy level of the client. At the same time, the costs paid by workers to complete federated crowdsourcing tasks are considered, mainly computing and communication costs thereby motivating workers to complete crowdsourcing tasks actively and efficiently. In the first stage, TiFedCrowd maximizes the net utility of the server, which is the total utility of global model minus the total incentive costs delivered to clients. Meanwhile, the maximum and minimum completion times are specified so that only models uploaded within the given time will be accepted. Finally, we deduce the Nash equilibrium in the Stackelberg game. Figure 1 shows the working flow diagram of the eFedCrowd.The main contributions of this paper are as follows:

- We propose the Time-controlled incentive federated crowdsourcing (TiFedCrowd), which protects the privacy of participants while encouraging more workers to complete crowdsourcing tasks with high quality and efficiency. A time threshold is set to preliminarily screen the quality of submitted data, which is applicable to both immediate and non-instant crowdsourcing.

- The Nash equilibrium of the stackelberg game is derived, and we prove that the server and clients utility maximization is globally optimal.
- TiFedCrowd allocates rewards according to contributions, having strong interpretability, which is conducive to attracting and retaining high-quality crowdsourcing workers, thus maintaining the fairness and accountability of the federated crowdsourcing market.

References

Amgad, M., Atteya, L. A., Hussein, H., Mohammed, K. H., Hafiz, E., Elsebaie, M. A., Alhusseiny, A. M., AlMoslemany, M. A., Elmatboly, A. M., Pappalardo, P. A. et al. (2022). Nucls: A scalable crowdsourcing approach and dataset for nucleus classification and segmentation in breast cancer. GigaScience, 11.

Ciftler, B. S., Albaseer, A., Lasla, N., & Abdallah, M. (2020). Federated learning for rss fingerprint-based localization: A privacy-preserving crowdsourcing method. In 2020 International Wireless Communications and Mobile Computing (IWCMC) (pp. 2112–2117). IEEE.

Gao, D., Yao, X., & Yang, Q. (2022). A survey on heterogeneous federated learning. arXiv preprint arXiv:2210.04505, .

Li, B., Shi, Y., Kong, Q., Du, Q., & Lu, R. (2023). Incentive-based federated learning for digital twin driven industrial mobile crowdsensing. IEEE Internet of Things Journal, .

Li, L., Fan, Y., Tse, M., & Lin, K.-Y. (2020a). A review of applications in federated learning. Computers & Industrial Engineering, 149, 106854.

Li, T., & Sethi, S. P. (2017). A review of dynamic stackelberg game models. Discrete & Continuous Dynamical Systems-B, 22, 125.

Li, Z., Liu, J., Hao, J., Wang, H., & Xian, M. (2020b). Crowdsfl: A secure crowd computing framework based on blockchain and federated learning. Electronics, 9, 773.

Liu, J., Huang, J., Zhou, Y., Li, X., Ji, S., Xiong, H., & Dou, D. (2022). From distributed machine learning to federated learning: A survey. Knowledge and Information Systems, 64, 885–917.

Lyu, L., Yu, H., Zhao, J., & Yang, Q. (2020). Threats to federated learning. Federated Learning: Privacy and Incentive, (pp. 3–16).

McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Artificial intelligence and statistics (pp. 1273–1282). PMLR.

Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. Future Generation Computer Systems, 115, 619–640.

Pournajaf, L., Xiong, L., Sunderam, V., & Goryczka, S. (2014). Spatial task assignment for crowd sensing with cloaked locations. In 2014 IEEE 15th International Conference on Mobile Data Management (pp. 73–82). IEEE volume 1.

Ren, Y., Li, X., Miao, Y., Luo, B., Weng, J., Choo, K.-K. R., & Deng, R. H. (2022). Towards privacy-preserving spatial distribution crowdsensing: A game theoretic approach. IEEE Transactions on Information Forensics and Security, 17, 804–818.

Sigurdsson, G. A., Varol, G., Wang, X., Farhadi, A., Laptev, I., & Gupta, A. (2016). Hollywood in homes: Crowdsourcing data collection for activity understanding. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14 (pp. 510–526). Springer.

Suri, A., Kanani, P., Marathe, V. J., & Peterson, D. W. (2022). Subject membership inference attacks in federated learning. arXiv preprint arXiv:2206.03317, .

Tu, J., Yu, G., Domeniconi, C., Wang, J., Xiao, G., & Guo, M. (2020). Multi-label crowd consensus via joint matrix factorization. Knowledge and Information Systems, 62, 1341–1369.

Wang, L., Qin, G., Yang, D., Han, X., & Ma, X. (2018). Geographic differential privacy for mobile crowd coverage maximization. In Proceedings of the AAAI Conference on Artificial Intelligence. volume 32.

Wang, L., Yu, H., & Han, X. (2020a). Federated crowdsensing: framework and challenges. arXiv preprint arXiv:2011.03208, .

Wang, L., Yu, H., & Han, X. (2020b). Federated crowdsensing: framework and challenges. arXiv preprint arXiv:2011.03208, .

Wei, J., Lin, Y., Yao, X., & Zhang, J. (2019). Differential privacy-based location protection in spatial crowdsourcing. IEEE Transactions on Services Computing, 15, 45–58.

Wu, M., Li, Q., Bilal, M., Xu, X., Zhang, J., & Hou, J. (2021). Multi-label active learning from crowds for secure iiot. Ad Hoc Networks, 121, 102594.

Xia, H., & McKernan, B. (2020). Privacy in crowdsourcing: a review of the threats and challenges. Computer Supported Cooperative Work (CSCW), 29, 263–301.

Ye, C., Coco, J., Epishova, A., Hajaj, C., Bogardus, H., Novak, L., Denny, J., Vorobeychik, Y., Lasko, T., Malin, B. et al. (2018). A crowdsourcing framework for medical data sets. AMIA Summits on Translational Science Proceedings, 2018, 273.

Yu, G., Tu, J., Wang, J., Domeniconi, C., & Zhang, X. (2020). Active multilabel crowd consensus. IEEE Transactions on Neural Networks and Learning Systems, 32, 1448–1459.

Zhan, Y., Li, P., Guo, S., & Qu, Z. (2021a). Incentive mechanism design for federated learning: Challenges and opportunities. IEEE Network, 35, 310–317.

Zhan, Y., Li, P., Qu, Z., Zeng, D., & Guo, S. (2020). A learning-based incentive mechanism for federated learning. IEEE Internet of Things Journal, 7, 6360–6368.

Zhan, Y., Zhang, J., Hong, Z., Wu, L., Li, P., & Guo, S. (2021b). A survey of incentive mechanism design for federated learning. IEEE Transactions on Emerging Topics in Computing, 10, 1035–1044.

Zhang, C., Guo, Y., Jia, X., Wang, C., & Du, H. (2021). Enabling proxy-free privacy-preserving and federated crowdsourcing by using blockchain. IEEE Internet of Things Journal, 8, 6624–6636.

Zhang, J., Wu, M., Zhou, C., & Sheng, V. S. (2022). Active crowdsourcing for multilabel annotation. IEEE Transactions on Neural Networks and Learning Systems, .