Merging Collaborative Learning and Blockchain: Privacy in Context

Sandi Rahmadika and Kyung-Hyune Rhee*
Department of ITConvergence of Information Security, Graduate School
Pukyong National University, Republic of Korea
*Department of ITConvergence and Application Engineering
Pukyong National University, Republic of Korea

Abstract

The emergence of collaborative learning to the public is to tackle the user's privacy issue in centralized learning by bringing the AI models to the data source or client device for training. Collaborative learning employs computing and storage resources on the client's device. Thus, it is privacy preserved by design. In harmony, blockchain is also prominent since it does not require an intermediary to process a transaction. However, these approaches are not yet fully ripe to be implemented in the real world, especially for the complex system (several challenges need to be addressed). In this work, we present the performance of collaborative learning and potential use case of blockchain. Further, we discuss privacy issues in the system.

1. Introduction

Collaborative learning and blockchain technology are widely discussed recently. These technologies count on the top of the decentralized form as a part of the distributed system. Both can be interpreted as an intersection of on-device AI, decentralized ledger, and edge computing. The main objective of this approach is to cover the weaknesses in the centralized architecture.

Collaborative learning is a breakthrough in the machine learning. It turns the centralized raw data into a decentralized form. The raw data owned by clients are never leaving the devices [1]. Thus, the issues of privacy in centralized learning can be tackled.

Blockchain appears with the merits of offering solutions that are faced in the centralized model. The core idea behind it is a chain-shaped data structured as known as a chain of blocks [2]. Due to the merits, blockchain can be used for many purposes such as storing data, managing data, and incentive mechanisms [3]. Therefore, the application of the blockchain has reached many aspects such as finance, healthcare, supply chain, and to name a few.

In this work, we briefly show the performance of collaborative learning where the users train the model on the devices using their dataset (built-in training). Once the training is completed, the user sends the upgraded model back to the global or aggregation server. This sort of activity is carried out continuously for as long as necessary to improve the AI model.

The users are incentivized since they provide the valid upgraded model and train the model using their resources. The rewards are propagated by relying on the blockchain technology. Eventually, blockchain with its merits offerred can be used in the collaborative learning as an incentive mechanism. However, this paper only presents the initial approach of collaborative learning and incentive mechanism. Further research is a necessity, especially related to potential attacks.

2. Collaborative Intelligence

One of the most prominent benefits of utilizing blockchain technology is not involving a middleman nor intermediaries to manage the transactions [4]. This feature is also useful for collaborative intelligence where data is not concentrated. The data is scattered among the user devices in the same application. In the collaborative learning algorithm, there is a set of users with a distinct local dataset.

$$f(w) = \frac{1}{n}\sum_{i=1}^{n} f_i(w) \qquad (1)$$

$$\sum_{n=1}^{n} \frac{c_n}{c} w_{t+1}^{n} \qquad (2)$$

Intuitively, $fi(w)=(xi,yi,w)$ can be defined as a loss of the prediction on example (xi,yi) which is managed by model parameters w as shown in (1). It also can be interpreted as IoT devices send gradients or parameters Δw1+ Δw2+Δw3+⋯+Δwn to the cloud server, which is partitioned homogenously [5]. Finally, the aggregated server computes the model received and applies it to the new parameters as defined in (2).

3. Blockchain Incentive Structure

A blockchain-based loyalty rewards system can reduce management costs with smart contracts that are secure, trackable, transparent, and lower costs. A smart contract can be utilized further such as merging with AI as shown in Figure 1. This combination can provide a more resilient and efficient path for a decentralized system. By implementing this, the reliance on the middleman to take care of the transaction can be eliminated, so that many benefits are obtained.

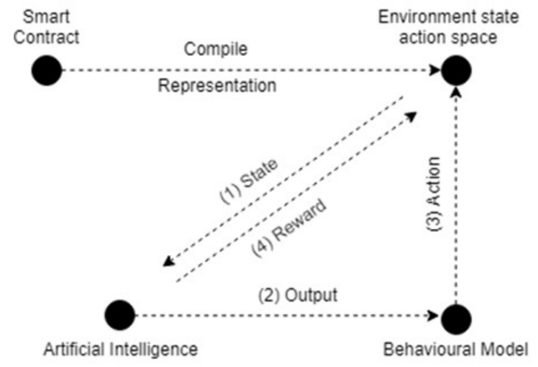

Fig. 1. The basic high-level structure of the AI smart contract scheme.

Figure 1 presents a smart contract that handles the training of AI agents with an interpretable result. As the transition model towards the blockchain economy space, this could become a critical tool in offering safer decentralized application (DApps) [6].

4. Results and Discussion

For the presentation and implementation, we use five devices acting as clients in the collaborative system. Afterward, the clients download the first global model from the aggregated server. The devices train the model using their dataset. When training is completed, the client sends the updated model back to the aggregation server.

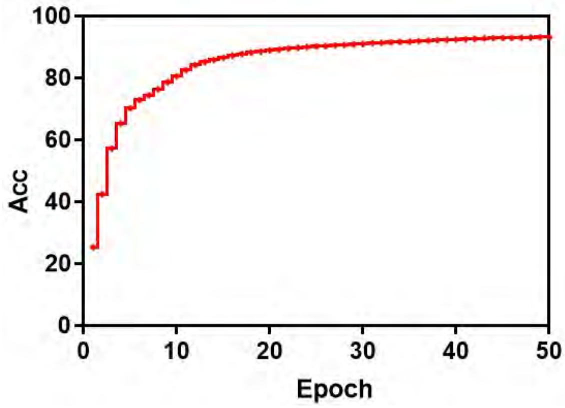

Fig. 2. Average accuracy of training data

We use a standard convolutional neural network (CNN) as the model installed within devices. The model trains the letter and number images.

Table 1. Inference attacks for different fractions

Property / % parameters update	10%	50%	100%
Top region (Antwerpen)	0.84	0.86	0.93
Gender	0.90	0.91	0.93
Veracity	0.94	0.99	0.99

Based on the implementation, the accuracy increased over time (see Fig 2). The average loss is around 0.23 and the accuracy reaches 93%. Nevertheless, we find that the number of devices affects the training time. The more devices within the collaborative system, the slower the training time it gets. When the transaction is conducted and verified by the aggregated server, then the rewards are delivered to the clients by using smart contracts. Nevertheless, in this paper, we do not elaborate on the smart contract in detail.

Even though the combination of decentralized learning and blockchain brings a lot of benefits, but it is also still vulnerable to attacks such as inference attacks as shown in Table 1 [7]. Roughly speaking, in a certain way, the malicious can leak the training dataset of the clients. Thus, the main objective of decentralized learning has been disrupted by malicious clients. The bright side is the performance of the malicious decreases with the increasing number of honest clients (a large number of clients) in the same decentralized network. In short, the attempt from malicious is negligible.

5. Conclusion

The collaboration of decentralized learning and blockchain brings a lot of merits that overcome many issues in the centralized system. We presented the model as well as the performance which needs to be developed further in all aspects. Even though this combination provides goodness in the real world, the privacy of the users is still an issue. The malicious with a certain condition can gather the training dataset of the users. For future work, we plan to analyze the data protection policies with the appropriate incentives by leveraging blockchain.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2018R1D1A1B07048944) and partially was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2020-2015-0-00403) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)

[References]

[1] Weng, Jiasi, et al. "Deepchain: Auditable and privacy-preserving deep learning with blockchain-based incentive." IEEE Transactions on Dependable and Secure Computing (2019).

[2] S. Wu, Y. Chen, Q. Wang, M. Li, C. Wang, and X. Luo, "CReam: A Smart Contract Enabled Collusion-Resistant e-Auction," IEEE Transactions on Information Forensics and Security, 2018.

[3] Wang, Jingzhong, et al. "A blockchain based privacy-preserving incentive mechanism in crowdsensing applications." IEEE Access 6 (2018): 17545-17556.

[4] G. A. Montes and B. Goertzel, "Distributed, decentralized, and democratized artificial intelligence," Technological Forecasting and Social Change. 2019.

[5] Jakub Konecny, H Brendan McMahan, Daniel Ramage, and Peter Richtarik. "Federated optimization: Distributed machine learning for on-device intelligence", arXiv preprint arXiv:1610.02527, 2016.

[6] "Incentivai." [Online]. Available: https://incentivai .com/product/. [Accessed: 12-Oct-2019].

[7] Melis, Luca, et al. "Exploiting unintended feature leakage in collaborative learning." 2019 IEEE Symposium on Security and Privacy (SP). IEEE, 2019.