Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/18/20230953

Overcoming data silos in healthcare: The potential of blockchain and federated learning on the Hedera platform

Jiashao He

School of Software, South China Normal University, Guangzhou, 528225, China

20202005399@m.scnu.edu.cn

Abstract. The lack of access to extensive and varied datasets remains one of the major issues facing the field of machine learning, despite recent advancements. This is especially true in the healthcare sector, where it can be challenging to gather and use patient data for research because it is frequently compartmentalized across many healthcare providers. By enabling secure and privacy-preserving access to distributed data, blockchain technology, and federated learning have the potential to overcome these difficulties. In this article, we'll look at how federated learning and blockchain are used in the healthcare industry and talk about their benefits and drawbacks. We will also examine the Hedera platform, which makes use of blockchain technology and a new algorithm called Gossip Degree to provide a revolutionary method of federated learning. We will also go over the potential effects of federated learning on the healthcare sector and what it means for future research.

Keywords: federated learning, blockchain, Hedera.

1. Introduction

Due to privacy and security issues, new technologies like Federated Learning (FL) and Hedera Hashgraph are becoming more and more crucial in today's digital environment [1-3]. FL is a sophisticated Machine Learning (ML) technique that makes it possible to create models while maintaining the security and privacy of user data. It permits data to stay on user devices rather than being sent to a centralized server. This approach is perfect for a variety of industries, such as healthcare and medical, insurance and finance, robots and autonomous systems, virtual reality, and more. A decentralized and distributed ledger technology called Hedera Hashgraph allows for secure communication and transactions while protecting user privacy [4-6]. It is an appropriate platform for creating dependable, scalable, and latency-sensitive decentralized applications.

The paper is structured in the manner described below. To solve the issue of accessing and using patient data for research while respecting data privacy, we sketch a realistic scenario of leveraging the blockchain in federated learning in Section 2. We also explain the idea and potential of federated learning [7]. The current state and detrimental effects of data silos on the healthcare sector are discussed in Section 3, along with the difficulties that various healthcare companies face when trying to use data mining and machine learning. The Hedera platform is also introduced as a potential remedy for federated learning. In this section, we also highlight the limitations and shortcomings of existing federated learning algorithms. In Section 4, based on the limitations discussed in Section 3, the paper offers

DOI: 10.54254/2755-2721/18/20230953

recommendations and future prospects for federated learning. And in Section 5, we make a conclusion of this paper.

2. Case analysis

In the healthcare industry, patient data is often dispersed across multiple healthcare providers, leading to challenges in accessing and using the data for research purposes. However, blockchain technology combined with federated learning can potentially solve this issue.

For instance, let's consider a scenario where a research team intends to train a machine-learning model to identify risk factors for a particular disease. Instead of collecting patient data in a centralized location, which could violate regulations like HIPAA and raise privacy concerns, the researchers can use a blockchain network to access and exchange data securely [8-9]. Each healthcare provider maintains control over their data and can authorize its usage in the model training process. The blockchain offers a secure and transparent way to manage data access and sharing, while federated learning enables the model to be trained on distributed data without jeopardizing data privacy [1].

Federated learning can address this situation's significant difficulty of healthcare data silos. Researchers can access a larger dataset while maintaining the security and privacy of specific patient data by using blockchain and federated learning. In turn, this may result in more thorough and accurate models that benefit patients by improving disease diagnosis and therapy.

In conclusion, the application of blockchain technology and federated learning in the healthcare industry can aid in overcoming the problem of data silos and encourage a secure and cooperative method of data management.

3. Analysis

The current state of data silos poses significant challenges to data mining and machine learning, particularly in the healthcare industry. However, the Hedera platform offers a promising solution to this problem through its innovative approach to federated learning.

Federated learning is a distributed learning technique that allows multiple entities to learn a shared model without sharing their data. Instead of centralizing the data in a single location, federated learning decentralizes the data and moves the learning process to where the data resides. This approach offers several advantages, including preserving data privacy and reducing communication costs.

The Hedera platform employs a novel approach to federated learning by leveraging blockchain technology [10]. This technology provides an immutable and decentralized ledger that ensures the integrity of the training process and the privacy of the data. Specifically, the platform preserves data privacy by utilizing the blockchain's distributed architecture to keep sensitive data encrypted and decentralized. This prevents unauthorized access and tampering by malicious actors, thereby ensuring the confidentiality and integrity of the training data.

In addition to preserving data privacy, the Hedera platform also reduces communication costs associated with traditional federated learning approaches. By leveraging blockchain technology, the platform can minimize the need for data transfers between participants. This is because the blockchain stores the model parameters, which means that each participant can update their local model without the need to share raw data. As a result, communication costs are reduced, making federated learning more efficient and cost-effective.

Furthermore, the Hedera platform introduces a new algorithm called GossipGrad, which enhances the efficiency and convergence speed of federated learning. GossipGrad is a gossip-based algorithm that allows nodes to communicate with each other in a peer-to-peer fashion, without the need for a central coordinator. This approach reduces the communication overhead and improves the scalability of federated learning, making it suitable for large-scale distributed systems [2].

Despite the advantages of federated learning, existing algorithms still face several challenges, such as slow convergence and scalability issues. The Hedera platform's innovative approach to federated learning, combined with the GossipGrad algorithm, presents a viable solution to these challenges. By leveraging blockchain technology and gossip-based communication, the Hedera platform offers a more

DOI: 10.54254/2755-2721/18/20230953

efficient and scalable approach to federated learning, which could revolutionize the way we perform data mining and machine learning in the healthcare industry and beyond.

4. Discussion

Moreover, there are a number of difficulties with federated learning implementation, including security, communication overhead, and convergence speed, which may restrict its usefulness in practical applications. However, the Hederaplatform provides a number of benefits that potentially address these issues, improving the prospects for federated learning applications.

Firstly, to secure data privacy and integrity, the Hedera platform first uses cutting-edge security technologies including blockchain and cryptography. These innovations guarantee the security of federated learning by preventing data leakage, manipulation, and theft. In order to keep and verify all data records during the data exchange process and stop hostile nodes from assaulting the system, the platform uses blockchain technology. Additionally, it uses secure multi-party computing and encryption cryptographic techniques to safeguard data privacy.

Secondly, the Hedera platform uses advanced algorithms such as GossipGrad to improve the convergence speed and efficiency of federated learning. The GossipGrad algorithm, based on message passing and local computation, reduces the communication and computation overhead among nodes, improving the speed and efficiency of federated learning. Compared to traditional federated learning algorithms, the GossipGrad algorithm has a faster convergence speed and higher accuracy. Additionally, it can dynamically adjust node selection and message passing frequency to better adapt to different application scenarios.

Furthermore, the Hederaplatform is appropriate for a variety of federated learning application situations due to its great scalability and flexibility. Users may conveniently use federated learning for model training and prediction thanks to its openness and simplicity. The platform's distributed computing and communication technologies, which support thousands of nodes for federated learning, contribute to its scalability and flexibility. Moreover, users can create and implement federated learning models using a variety of computer languages and frameworks thanks to their openness and simplicity. Furthermore, the platform offers a comprehensive array of APIs and SDKs that let users do model training and prediction in various data formats and computing environments.

Overall, the Hederaplatform offers a secure, effective, flexible, and user-friendly federated learning platform that enables users to utilize scattered data resources for model training and prediction while maintaining data privacy. Federated learning has a wide range of potential applications in a number of industries, including healthcare, finance, and intelligent transportation. Examples of such applications include disease prediction, fraud detection, and traffic forecasting. As a result, federated learning can become much more widely used and valuable by being implemented on the Hederaplatform, which could lead to it playing a bigger part in the future of machine learning and data analysis.

5. Conclusion

Federated learning is a promising technique for performing machine learning on distributed data without compromising data privacy. The application of blockchain technology to improve the security and privacy of federated learning in the healthcare industry has been covered in this article. We also introduced the Hedera platform, which provides a revolutionary method for federated learning and makes use of blockchain and gossip-based algorithms to boost its effectiveness, scalability, and security.

Data protection, integrity, and openness are just a few benefits of using blockchain technology in federated learning. The blockchain's distributed architecture makes a guarantee that data is encrypted and decentralized, preventing tampering and illegal access.

Additionally, a public and immutable record provided by the blockchain enables stakeholders to monitor and audit the training process, ensuring its fairness and integrity.

Finally, the GossipGrad algorithm, which enhances the efficiency and speed of federated learning, is also a feature of the Hedera platform. This technique uses a gossip-based method of communication,

DOI: 10.54254/2755-2721/18/20230953

which lowers communication and computation overhead, increasing the effectiveness and scalability of federated learning.

In conclusion, federated learning and blockchain technology give a fresh and efficient method for data mining and machine learning in the healthcare sector and beyond. The Hedera platform provides a safe, effective, versatile, and user-friendly platform for federated learning, offering a promising solution to implementation issues. The adoption of blockchain technology in federated learning is projected to increase due to the growing requirement for data privacy and security in machine learning.

References

[1] Zhang, Xiuxian, et al. "Hashgraph Based Federated Learning for Secure Data Sharing." Wireless and Satellite Systems: 11th EAI International Conference, WiSATS 2020, Nanjing, China, September 17-18, 2020, Proceedings, Part II. Springer International Publishing, 2021.

[2] Hu, Chenghao, Jingyan Jiang, and Zhi Wang. "Decentralized federated learning: A segmented gossip approach." arXiv preprint arXiv:1908.07782 (2019).

[3] Monrat A A, Schelén O, Andersson K. A survey of blockchain from the perspectives of applications, challenges, and opportunities. IEEE Access, 2019, 7: 117134-117151.

[4] Zhou Q, Huang H, Zheng Z, et al. Solutions to scalability of blockchain: A survey. Ieee Access, 2020, 8: 16440-16455.

[5] Berdik D, Otoum S, Schmidt N, et al. A survey on blockchain for information systems management and security. Information Processing & Management, 2021, 58(1): 102397.

[6] Lu Y. The blockchain: State-of-the-art and research challenges. Journal of Industrial Information Integration, 2019, 15: 80-90.

[7] Kholidy H A, Kamaludeen R. An Innovative Hashgraph-based Federated Learning Approach for Multi Domain 5G Network Protection//2022 IEEE Future Networks World Forum (FNWF). IEEE, 2022: 139-146.

[8] Zhang X, Zhao L, Li J, et al. Hashgraph Based Federated Learning for Secure Data Sharing//Wireless and Satellite Systems: 11th EAI International Conference, WiSATS 2020, Nanjing, China, September 17-18, 2020, Proceedings, Part II. Springer International Publishing, 2021: 556-565.

[9] Abid Haleem, Mohd Javaid, Ravi Pratap Singh, Rajiv Suman, Shanay Rab, Blockchain technology applications in healthcare: An overview, International Journal of Intelligent Networks, Volume 2, 2021, Pages 130-139, ISSN 2666-6030.

[10] Shah K, Kanani S, Patel S, et al. Blockchain‐based object detection scheme using federated learning. Security and Privacy, 2023, 6(1): e276.