



Federated Learning for 6G Edge Intelligence: Concepts, Challenges and Solutions

Han Wang^{1,2,4,5(✉)}, Jianhe Hu^{3,4,5}, Chunxiao Xing^{1,2}, and Liang-Jie Zhang^{4,5}

¹ Research Institute of Information Technology, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China

² Department of Computer Science and Technology, Institute of Internet Industry, Tsinghua University, Beijing 100084, China

³ Department of Economics, College of Letters and Science, University of California Davis, Davis, CA 95616, USA

⁴ National Engineering Research Center for Supporting Software of Enterprise Internet Services, Shenzhen 518057, China

⁵ Kingdee Research, Kingdee International Software Group Company Limited, Shenzhen 518057, China

Abstract. Along with continuous evolution, the future 6G network will become a converged “Cloud-Edge-Terminal” ecosystem which can carry various crucial AI applications on edge computing units, formulating an ubiquitous “Edge Intelligence” paradigm to enable differentiated service innovations and empower intelligent transformation of vertical industries. However, due to issues of data security, user privacy, wireless network transmission capability and etc., it is not feasible for conventional machine learning methods to build AI models by directly collecting massive distributed edge data together, and hence resulting a large number of “isolated data islands” in the edge units. In order to break the data sharing barrier and drive cross-edge data cooperation, this paper studies a federated learning based AI model training method by which sensitive raw data can be maintained and protected in its original edge units. Based on the general scheme, some challenging problems are discussed to implement this new paradigm in practical scenarios, and the corresponding promising solutions and key techniques are proposed to inspire further researches.

Keywords: 6G · Federated learning · Edge intelligence

1 Introduction

Inspired by the idea of cloud computing, the brand new “Service Based Architecture (SBA)” [1] is formulated to support the long-term evolution of the cellular network, from the current 5G to the future 6G and beyond era, so as to meet the critical requirements of various applications. In order to match the whole

This work is supported by National Key R&D Program of China (2018YFB1402701).

© Springer Nature Switzerland AG 2022

Y. Pan et al. (Eds.): AIMS 2021, LNCS 12987, pp. 99–112, 2022.

https://doi.org/10.1007/978-3-030-96033-9_9

new architecture, the Mobile Network Operators (MNO) upgrade their existing telecom equipment rooms into cloud computing data centers, sink the user plane function to the network edges or even user premises, and build the Multi-access Edge Computing (MEC) infrastructures [2–5]. By deploying applications on the MEC hosts, service providers can offer high-quality cloud services with high bandwidth, low latency and massive connections to various kinds of users. In recent years, the rapid development of Artificial Intelligence (AI) technologies, represented by Machine Learning (ML) and Deep Learning (DL), has produced many major breakthroughs in the fields of computer vision, speech recognition and natural language processing, forming a complete ecological chain from smart terminals to cloud platforms and then to application services. By converging technologies of 5G/6G, MEC and AI under the SBA framework, building network edge based ubiquitous intelligence applications will enable differentiated service innovations and empower intelligent transformation of vertical industries. Currently, many cutting-edge 6G researches, represented by the largest international cooperation research project “6G Flagship” [6], have identified “Edge Intelligence” as an important goal in the evolution roadmap from “Cloud Native 5G” to “Intelligence Native 6G” [7–13].

Conventional AI training process is usually concentrated in the cloud computing center, where sufficient and high-quality supply of big data guarantees the quality of model training. However, the paradigm of “5G/6G + MEC + AI” is mainly oriented to vertical industries (e.g. industrial internet, internet of vehicles, and internet of things), where on-premise data in the enterprise’s private MEC is exclusive to its owner, and high-value data in the MNO’s public MEC cannot simply be opened to the third party due to property rights and security issues as well. In addition, the massive amount of data generated by a large number of smart terminals also cannot be shared due to network transmission capacity and privacy protection. The above restrictions have formed a large number of “isolated data islands” in various edge locations such as MECs and smart terminals, coupled with increasingly stringent data protection regulations worldwide, and hence it has become difficult to complete model training by directly aggregating edge data to the central cloud.

Recently, Federated Learning (FL) is proposed to address the dilemma of “isolated data islands” [14–16]. FL is a distributed ML framework in which each participant does not need to share its raw data in the process of model training, and completes the training procedures jointly by transferring encrypted local model parameters to build a shared global ML model. By using FL to build edge intelligence, the global model can be fully trained under the condition that raw data maintains in its origins. This distributed federated training framework requires only a small number of model parameters to be transferred, instead of aggregating all the raw data from the edge to the central cloud, which can save a lot of network bandwidth and transmission time, as well as save a lot of storage and computing resources in the central cloud, and also fully protect user privacy and data security.

The FL based AI framework is essential to the evolution from “5G Edge Computing” to “6G Edge Intelligence”, which will drive cross-field enterprise-level

data cooperation, generate new industry and business models based on collaboration, promote innovation development without excessive costs of technology upgrades [17–26]. The paradigm of FL based edge intelligence will enable the 6G system to become an ubiquitous smart information infrastructure that connects everything and empowers industries, supporting the overall development of the entire social economy, and hence has important theoretical significance and practical application value.

This paper discusses some potential challenges and promising solutions to implement federated learning based edge intelligence in the future 6G era. The rest of this paper is organized as follows. Firstly, we introduce the preliminaries of edge computing and federated learning, and propose the paradigm of FL based edge intelligence in Sect. 2. Some potential challenges of conducting the proposed paradigm are discussed in Sect. 3. Then, we propose some promising solutions to defuse the above challenges in Sect. 4. At last, the paper is concluded in Sect. 5.

2 Federated Learning Based Edge Intelligence

Modern communication systems are evolving rapidly, from conventional “terrestrial dumb data transmission channel” to future “space-air-ground-sea integrated smart information grid”, during which edge units (e.g., smart terminals and edge computing centers) are playing a more and more important role of producing, caching and processing data. However, most raw data in edge units are so sensitive that they cannot be shared directly resulting isolated data islands. Hence, how to utilize distributed data to construct edge based intelligent applications becomes a tough problem.

Federated learning is one of the most promising techniques to formulate the ubiquitous “Edge Intelligence” paradigm for the future 6G and beyond era. In this section, we introduce the preliminaries of future converged network architecture, discuss the origins of isolated data islands in edge units, and then propose the federated learning based AI model training method.

2.1 Service Based Architecture and Edge Computing

The specifications of SBA [1], defined by the 3rd Generation Partnership Project (3GPP), leverage the service based interactions between different network functions, aligning system operations with network function virtualization and software defined networking to meet the critical requirements in the 5G and beyond era. The similar characteristics above are shared by the MEC specifications [3], which are defined by the European Telecommunications Standards Institute (ETSI). The 3GPP specifications define the enablers for edge computing, allowing that SBA and MEC interact collaboratively in traffic routing and policy control. Integrating the above MEC with SBA, taking the current 5G MEC system for example in Fig. 1, can formulate a powerful edge computing environment providing high-quality cloud services with high bandwidth, low latency and massive connections to various kinds of users.

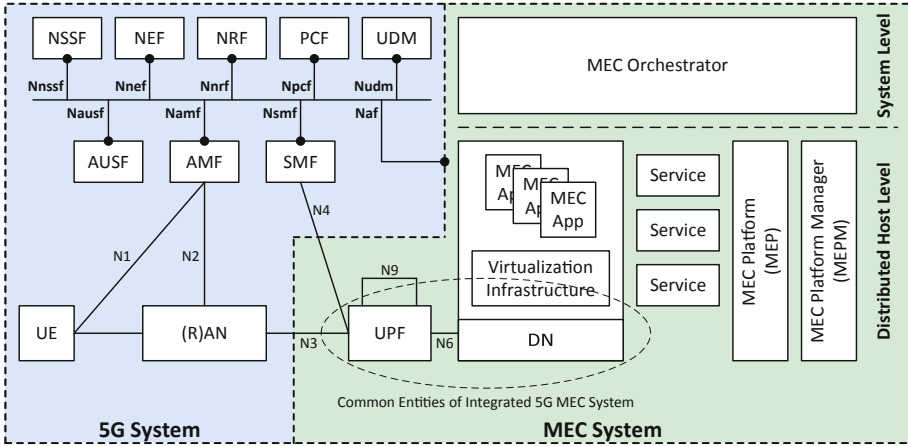


Fig. 1. Typical architecture of MEC integrated SBA (3GPP 5G) system

In the current 5G era, the resources of network and computing are “integrated” together in the paradigm of “5G + MEC”. Along with continuous evolution, the future 6G and beyond systems will bring more resources of network, computing and devices together to formulate a “converged” ecosystem, where network resources contain not only high-speed wired fiber links, but also terrestrial and satellite wireless links, while computing resources are ubiquitously distributed in various kinds of clouds, edges and terminals which are abstracted in Fig. 2 and detailed as follows.

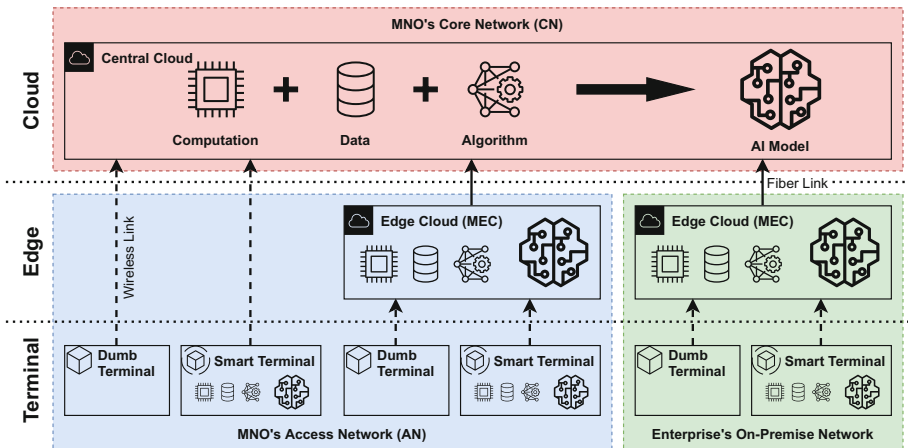


Fig. 2. Computing resources in 6G and beyond systems

- **Cloud** refers to the large scale cloud computing center located in the MNO’s Core Network (CN) and connected with edges or terminals via fibers or wireless links, and hence is specifically identified as the “Central Cloud”.
- **Edge** refers to a large number of small and medium scale cloud computing centers located at the network edge and providing MEC services, which are widely distributed in the MNO’s Access Network (AN) and also can be deployed down to the enterprise premise on demand, and hence is specifically identified as the “Edge Cloud”.
- **Terminal** refers to all kinds of user terminal devices, including smart phones, intelligent industrial robots, intelligent connected vehicles and other “Smart Terminals” with rich computing resources, as well as feature phones, cameras, sensors and other “Dumb Terminals” which has only the function of sending and receiving data.

In Fig. 2, the “Central Cloud” and “Edge Cloud” are MNO-managed telecom-grade cloud computing centers with complete and abundant computing resources and high-speed and reliable network connections, which are high-quality platforms for deploying AI services. In recent years, the techniques of smart terminal evolves rapidly so that capabilities of computing, storage and network have increased significantly. Along with the corresponding flourished AI ecology, both of complex inferring and simple training can be achieved on a single smart terminal. As the data producer and privacy concern, “Smart Terminals” are suitable and necessary to participate the process of training AI models directly. 6G Flagship and other related studies have already regarded “Edge Cloud” and “Smart Terminals” together as “Edge Unit” to host AI applications.

2.2 Isolated Data Islands Dilemma in Edge Units

Building a deep learning based AI system contains two parts: training and inference. The training process requires sufficient data input and intensive computation supply, with the continuous increase of diverse data, and then continuous training is required to improve the model accuracy. On the contrary, the inference process uses the trained model to identify new data in a single step, and hence requires very little computation so that most of the current terminals can achieve such simple inference tasks. Therefore, we mainly focus on the difficult process of AI model training.

The conventional AI training process is usually done centrally in cloud computing centers. The training platform is deployed on the “Central Cloud” in Fig. 2, and all kinds of terminals upload their own raw data directly (or indirectly via “Edge Cloud”) to the central cloud, and the central cloud invests a large amount of computation under the control of algorithms to complete the model training process. However, this kind of centralized data aggregation and processing for model training is facing increasing challenges:

- **Transmission**
The amount of data generated by various terminals is getting larger and

larger, and the continuous uploading of massive multimedia data to the cloud leads to many problems such as high load overheating and frequency reduction of terminals, rapid consumption of battery power, surge in traffic costs and congestion of transmission networks, which seriously affects the normal use of terminals and stable operation of networks.

– Security

Along with the rapid development of various vertical industries, the converged system carries more and more enterprise applications, and the massive data accumulated inside various industrial terminals and edge clouds are highly sensitive with many crucial issues such as data property rights, commercial value, personal and enterprise privacy, application system operation security, etc. Moreover, domestic and international data protection regulations are also becoming more and more strict, resulting the raw data can no longer be shared directly.

The above two aspects together lead to a large number of “isolated data islands” at the edge units, which has become a critical dilemma for training large-scale AI models in the converged architecture of the future 6G and beyond era.

2.3 Federated Learning for Distributed Model Training

Federated Learning (FL) is a recent addition to the distributed machine learning approaches, which aims at training AI models across multiple local datasets, contained in decentralized edge units holding local data samples, without aggregating or exchanging their raw data, thus addressing critical issues such as privacy, security and access rights to heterogeneous data. The approach of FL based AI model training is an effective solution to the above mentioned critical dilemma of “isolated data islands” in edge units.

The FL approach is different from both techniques of traditional centralized learning and classical distributed machine learning, where the former requires that all data samples should be uploaded to a centralized cloud server while the latter assumes that all data samples are identically distributed with same dimension. The general FL procedures are designed as iterative steps: training local model with local data samples, and exchanging parameters (e.g., weights in a Deep Neural Network (DNN)) among temporarily trained local models to generate global model. A centralized server can be used as a reference clock to manage the iterative steps of the FL algorithm, while a peer-to-peer scheme without center is also feasible for performing the FL training process.

Specifically, from top to bottom in Fig. 2, “Central Cloud”, “Edge Cloud” and “Smart Terminal” are regarded as three-level computing units of which AI training platforms are deployed with the same initial DNN model. Federated learning is performed between the adjacent upper-to-lower (i.e., central cloud \leftrightarrow edge clouds, edge cloud \leftrightarrow smart terminals, and central cloud \leftrightarrow smart terminals) computing units. As illustrated in Fig. 3, the procedures of FL based model training under the framework of Fig. 2 is divided into multiple rounds, each consisting of the following four steps:

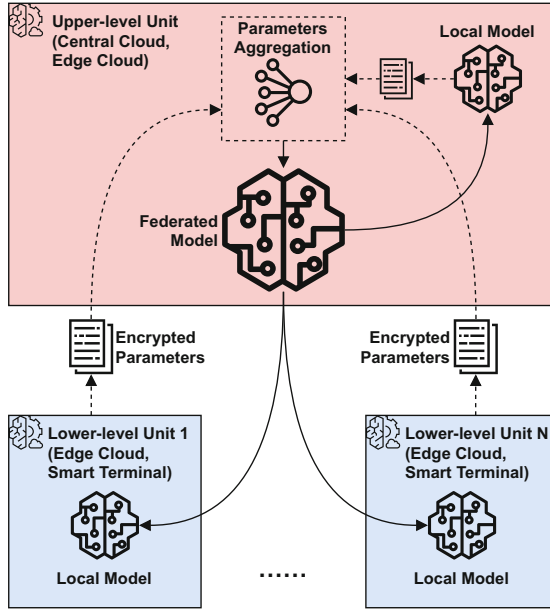


Fig. 3. Procedures of federated learning based model training

– **Local Training**

All of the computing units calculate gradients or parameters locally, and then the lower-level units forward their trained model parameters to the corresponding upper-level unit.

– **Model Aggregating**

The upper-level unit performs secure aggregation (e.g., homomorphic encryption) of the uploaded parameters from all lower-level units without learning any local information.

– **Parameter Broadcasting**

The upper-level unit broadcasts the aggregated parameters to all of the lower-level units.

– **Model Updating**

All lower-level units update their respective models with the received aggregated parameters, and then examine the performances of updated models.

After several local training and update exchanges between the upper-level unit and the corresponding lower-level units, it is possible to achieve a global optimal learning model. It is worth noting that, according to the source and feature of datasets, “horizontal” or “vertical” federating scheme can be chosen to perform model parameters aggregation to achieve “cross-sample collaborative modeling in the same industry” or “cross-feature collaborative modeling among different industries” to better meet the needs of practical applications.

3 Challenging Problems

From the current 5G era to the future, more and more industrial applications are characterized by latency stringency and demand, hence the latency induced by communicating and executing AI models in the central cloud may violate these requirements. Empowered by the above FL scheme, training AI model at the edge units ensures network scalability by distributing the procedures iteratively from centralized architectures in the remote central cloud to various edge units located closer to the users. This allows faster response to user requests since computation, data aggregation, and analysis are handled within user proximity. Moreover, it provides latency improvements for real-time applications as AI models are executed near the user.

Therefore, the above FL based distributed AI model training scheme is suitable for building the ubiquitous “Edge Intelligence” for the critical requirements of future 6G and beyond era. However, there are some practical challenging problems in heterogeneous modeling, efficiency improvement and security reinforcement that should to be discussed in detail as follows.

3.1 Heterogeneous Modeling

The performances of computing units in Fig. 2 vary significantly, from stable and efficient cloud computing centers to resource-constrained smart terminals, forming a heterogeneous model training system. In the standard FL training process described in Sect. 2.3, the upper-level unit can only aggregate parameters to renew the federated model after receiving all of the lower-level parameters before executing the subsequent steps. However, in a practical system, smart terminals may often fail to upload model parameters in time or even lose connections with their upper-level unit, due to unstable network connection, overheating frequency reduction, battery volume exhaustion, etc., which will result a halt of the whole process of model training. Since there are massive smart terminals participating the process of federated model training, the above halt phenomenon is almost bound to occur unavoidably.

In order to solve the above problems, it is necessary to study the federated modeling method applicable to heterogeneous edge units, so as to reduce the impact of terminal anomalies to ensure smooth and efficient in process of FL based training.

3.2 Efficiency Improvement

Currently, AI techniques are mainly used for processing multimedia contents, such as image recognition or video analysis, which requires high-dimensional DNN models with massive parameters to describe complex multimedia data. Although the FL based distributed training scheme does not require uploading raw multimedia data, the frequent exchanges of massive model parameters between upper and lower levels will also put certain stress on the network transmission capability. Edge cloud is connected to central cloud through high-stable

high-speed fiber network, and hence the parameters are transferred smoothly and fluidly. However, smart terminals are usually connected to edge or central cloud through wireless links, which are susceptible to various factors and poor network stability, and hence the frequent exchange of massive parameters may encounter transmission bottlenecks so as to affect the overall training efficiency.

In order to solve the above problems, it is necessary to study the efficient training methods for high-dimensional models, so as to improve the efficiency of model parameter exchange to ensure a high-efficient federated training process.

3.3 Security Reinforcement

As described in Sect. 2.3, the upper-level unit plays an important role in aggregating lower-level parameters and renew the federated model, which is the “critical core” in the framework of federated training. Although the upper-level unit is served by central or edge cloud with complete security measures, such a training system with center may still be subject to security threats from its inside. It has been shown that the original raw data can be deduced out with high probability by tracking gradient changing process during the model training. If the “critical core” is invaded by an attacker, the malicious program can collect all parameters during the entire process of model training, and may recover the original raw data through decryption, deduction and other technical methods, which will pose a great threat to user privacy, business interests, and even system operation security.

In order to solve the above problems, it is necessary to study the decentralized security reinforcement methods and build a safe and reliable federated training mechanism to ensure the operation security of the edge intelligence systems.

4 Promising Solutions

The above mentioned problems involve three aspects of federated modeling, i.e., heterogeneity, efficiency and security, which are practical challenges in the future 6G and beyond era. In order to inspire further researches, we will propose some promising ideas to solve the above challenges, which are inducted as a logical diagram in Fig. 4 and discussed in detail as follows.

4.1 Federated Modeling for Heterogeneous Edge Units

The standard FL training process requires all of the lower-level parameters collected together before aggregating uniformly, which is a “synchronous” strategy of parameter aggregation. However, in the heterogeneous model training system described in Sect. 3.1, due to the limitations of wireless networks and smart terminals, it is not guaranteed that all smart terminals can upload their model parameters in time, and some ones may even be temporarily or permanently disconnected, which will cause the entire training process to stagnate. Therefore, in the practical heterogeneous system, we need to propose an “asynchronous”

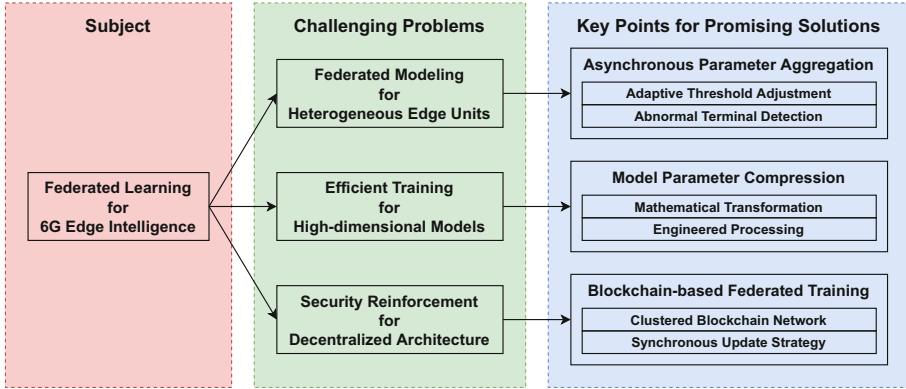


Fig. 4. Promising solutions to federated learning based edge intelligence

parameter aggregating strategy to reduce the impact of terminal and network anomalies to ensure efficient and smooth during the training process.

Specifically, considering the limited capacities of wireless network and terminals, we only require smart terminals to “do their best” to upload parameters. When the number or proportion of received parameters reaches a certain “threshold”, the upper-level unit instantly start to perform parameter aggregation to build the federated model. Few terminals, which cannot upload in time and join the current round of parameter aggregation, can participate in the next aggregating round after their parameters have been uploaded completely. It is worth noting that how to determine an appropriate threshold is an important issue. If the threshold is set too low, the parameters will be aggregated so early that the new federated model improves little due to insufficient parameter collection, and hence more training rounds are needed to get a better modeling quality. On the contrary, if the threshold is set too high, the waiting time will be too long to pull down the training efficiency. Therefore, the threshold should be adjusted adaptively and dynamically according to different datasets, models and training stages. Moreover, during multiple rounds of parameter aggregation, if a terminal consistently fails to complete uploading tasks, the upper-level unit should identify it as a suspected abnormal terminal. Once detected and confirmed, it will be removed from the lower-level units list to avoid lowering the proportion of valid units and affecting the model training efficiency.

The above proposed “Asynchronous Parameter Aggregation” strategy can greatly reduce the waiting time of upper-level units and improve the entire efficiency of federated model training. The corresponding key points are summarized in Fig. 4 and explained as follows:

- **Adaptive Threshold Adjustment**

By tracking the real-time convergence state during the process of federated model training, the threshold can be determined adaptively to balance the training speed and accuracy.

– **Abnormal Terminal Detection**

The management entities (e.g., Access and mobility Management Function (AMF) in Fig. 1) on control plane can be utilized to detect working status of suspected abnormal terminals, which will be removed to avoid low efficiency once be confirmed as anomaly.

The proposed adaptive threshold adjustment algorithm and the abnormal terminal detection method together can formulate an adaptive asynchronous parameter aggregation mechanism to ensure an efficient and smooth process of federated modeling training.

4.2 Efficient Training for High-Dimensional Models

The FL based distributed training requires frequent exchange of model parameters between adjacent upper-to-lower computing units, and hence the communication efficiency between upper and lower levels is crucial to the overall efficiency of federated modeling. Under the condition of constant network transmission capacity, improving the communication efficiency can be considered from a combination of two aspects: on the one hand, reducing the round number of parameter exchange, which is the main issue discussed in the previous section of asynchronous parameter aggregation; on the other hand, reducing the data volume in every exchange to save transmission time, which is the main topic of this section.

High-dimensional DNN models can better describe complex multimedia data, but also bring a huge amount of model parameters which put greater pressure on the processes of homomorphic encryption, wireless transmission and secure aggregation during federated training. However, it has been shown that not every value of the massive parameters in the high-dimensional model plays an important role which has a large amount of redundancy, and additionally the model parameters and the gradient changes in the training process often have certain structural and sparse features, and thus providing basic premises to compress the model parameters and reduce the amount of data.

Specifically, the high-dimensional “Model Parameter Compression” can be achieved by performing two types of methods, “Mathematical Transformation” and “Engineered Processing”, which are summarized in Fig. 4 and explained as follows:

– **Mathematical Transformation**

According to structural and sparse features, some mathematical transformation methods, such as singular value decomposition, Huffman coding, principal component analysis and compressive sensing, can be used to indirectly reduce data volume to compress the high-dimensional model. This kind of methods utilizes the inherent structure and sparsity of the original raw data to compress model parameters, usually with advantages of high compression rate and lossless, but suffering the drawback of computational intense.

– **Engineered Processing**

Many conventional engineering methods, such as quantification, dimensionality reduction, pruning, truncation and precision reduction, can be used to directly reduce the amount of model parameters, which is simple and efficient to implement, but usually with a defect of lossy compression.

After data processing, the amount of model parameters will be significantly reduced, which can consequently improve the communication efficiency during the federated training process. It is worth noting that most commonly used methods are precision-impaired which may degrade the modeling quality due to the cumulative effects of continuous iterations between upper and lower units, while a few lossless methods are computational complex which may also increase burden on smart terminals. Therefore, it is necessary to choose appropriate compression methods, according to features of actual business data and model structures, balancing computation-communication-storage overheads and modeling accuracy, so as to perform high-dimensional model training more efficiently.

4.3 Security Reinforcement for Decentralized Architecture

In the process of federated training, the upper-level unit plays an important role as the “critical core”, which may lead to a serious systematic security risk once be invaded. Therefore, we need to strengthen the system architecture with the idea of “decentralization” to ensure the secure operation of edge intelligence systems.

In recent years, the rapidly developing technology of “Blockchain”, with anonymous, immutable and distributed features, has been widely used to provide a reliable secure system among multiple untrustworthy participants. By transforming the conventional centralized network structure to distributed, the blockchain is essentially a distributed ledger which can ensure the data security by various cryptographic techniques and guarantee the data reliability among multiple untrustworthy distributed participants through consensus mechanisms, smart contracts, etc. In the federated training process, the upper and lower units naturally formulate as a distributed architecture, and hence all of the computing units can join together in a blockchain network and share equivalent rights. Each computing unit updates the global record in the whole blockchain network when interacting model parameters, so that each on-chain unit records the whole changing process of model parameters and thus the “critical core” is decentralized from a single upper-level unit to the whole network. Under the constraints of consensus mechanism and smart contracts, illegal operations initiated by malicious nodes at any network location will be discovered in time, and hence the security risks in the federated training process can be fundamentally solved.

It is worth noting that latency will be introduced during the process of updating blockchain. The larger the network scales, the greater the latency introduces, and excessive scale of the blockchain will suffers too large additional delay to operates normally. Therefore, in order to maintain high-efficiency, the “Blockchain-based Federated Training” should be operated as a “Clustered

Blockchain Network” with “Synchronous Update Strategy”, of which key points are summarized in Fig. 4 and explained as follows:

– **Clustered Blockchain Network**

According to the physical and logical architecture of the practical federated training system, various of engineering methods, such as slicing, sub-chaining and multi-channelization, can be utilized to split a large-scale blockchain into clustered sub parts to avoid long-latency and low-efficiency.

– **Synchronous Update Strategy**

The intra-cluster (i.e., within a particular slice, sub-chain or channel) data are synchronously updated to promote efficiency, while the updating strategy of inter-cluster (i.e., between slices, sub-chains and channel) data maintains asynchronous to avoid big changes of the FL training process.

The above proposed clustered blockchain network structure and its corresponding synchronous update strategy can effectively balance efficiency and security to guarantee stable operation of the blockchain-based federated training system.

5 Conclusion

In this paper, we studied a federated learning based distributed training method, which is dedicated to solving the isolated data islands and breaking the data sharing barrier, to facilitate building ubiquitous edge-based AI applications in the future 6G and beyond era. We first proposed the general scheme of federated training, and then discussed three challenging aspects of heterogeneous modeling, efficiency improvement and security reinforcement in practical scenarios. Based on the discussed problems, we proposed some promising ideas and corresponding key technical points to inspire further researches.

References

1. 3GPP TS 23.501 V16.8.0: Technical Specification Group Services and System Aspects; System architecture for the 5G System (5GS); Stage 2 (Release 16). 3GPP Technical Specification (2021)
2. ETSI MEC: <https://www.etsi.org/technologies/multi-access-edge-computing>
3. ETSI GS MEC 003 V2.1.1: Multi-access Edge Computing (MEC); Framework and Reference Architecture. ETSI Group Specification (2019)
4. Kekki, S., et al.: MEC in 5G networks. ETSI White Paper, pp. 1–28 (2018)
5. ETSI GR MEC 031 V2.1.1: Multi-access Edge Computing (MEC); MEC 5G Integration. ETSI Group Report (2020)
6. 6G Flagship Homepage. <https://www.6gflagship.com/>
7. 6G Flagship: Key Drivers and Research Challenges for 6G Ubiquitous Wireless Intelligence. <https://www.6gchannel.com/items/key-drivers-and-research-challenges-for-6g-ubiquitous-wireless-intelligence/>
8. Peltonen, E., et al.: 6G white paper on edge intelligence. 6G Research Visions (8) (2020). <https://www.6gchannel.com/items/6g-white-paper-edge-intelligence/>

9. Samsung Research: 6G: The Next Hyper-Connected Experience for All. Samsung Research Report (2020)
10. IMT-2030 (6G) Promotion Group: White Paper on 6G Vision and Candidate Technologies. IMT-2030 (6G) Promotion Group White Paper (2021)
11. 6GANA: From Cloud AI to Network AI: A View from 6GANA. 6GANA Report (2021)
12. NTT DOCOMO: White Paper: 5G Evolution and 6G (Version 3.0). NTT DOCOMO White Paper (2021)
13. 5G PPP: AI and ML - Enablers for Beyond 5G Networks (Version 1.0). 5G PPP White Paper (2021)
14. Konečný, J., McMahan, H.B., Yu, F.X., Richtarik, P., Suresh, A.T., Bacon, D.: Federated learning: strategies for improving communication efficiency. In: NIPS Workshop on Private Multi-Party Machine Learning (2016)
15. McMahan, H.B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) (2017)
16. Yang, Q., Liu, Y., Chen, T., Tong, Y.: Federated machine learning: concept and applications. *ACM Trans. Intell. Syst. Technol.* **10**(2), 1–19 (2019)
17. You, X., et al.: Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts. *Sci. China Inf. Sci.* **64**(1), 110301 (2020). <https://doi.org/10.1007/s11432-020-2955-6>
18. Wang, X., Han, Y., Wang, C., Zhao, Q., Chen, X., Chen, M.: In-edge AI: intelligent mobile edge computing, caching and communication by federated learning. *IEEE Netw.* **33**(5), 156–165 (2019)
19. Liu, Y., Yuan, X., Xiong, Z., Kang, J., Wang, X., Niyato, D.: Federated learning for 6G communications: challenges, methods, and future directions. *China Commun.* **17**(9), 105 (2020)
20. Wang, X., Han, Y., Leung, V.C.M., Niyato, D., Yan, X., Chen, X.: Convergence of edge computing and deep learning: a comprehensive survey. *IEEE Commun. Surv. Tutor.* **22**(2), 869–904 (2020)
21. Niknam, S., Dhillon, H.S., Reed, J.H.: Federated learning for wireless communications: motivation, opportunities, and challenges. *IEEE Commun. Mag.* **58**(6), 46–51 (2020)
22. Lim, W.Y.B., et al.: Federated learning in mobile edge networks: a comprehensive survey. *IEEE Commun. Surv. Tutor.* **22**(3), 2031–2063 (2020)
23. Feng, C., Zhao, Z., Wang, Y., Quek, T.Q.S., Peng, M.: On the design of federated learning in the mobile edge computing systems. *IEEE Trans. Commun.* **69**(9), 5902–5916 (2021)
24. Lu, Y., Huang, X., Zhang, K., Maharjan, S., Zhang, Y.: Blockchain and federated learning for 5G beyond. *IEEE Netw.* **35**(1), 219–225 (2021)
25. Yu, R., Li, P.: Toward resource-efficient federated learning in mobile edge computing. *IEEE Netw.* **35**(1), 148–155 (2021)
26. Zhou, X., Liang, W., She, J., Yan, Z., Wang, K.I.K.: Two-layer federated learning with heterogeneous model aggregation for 6G supported internet of vehicles. *IEEE Trans. Veh. Technol.* **70**(6), 5308–5317 (2021)