Data Collaboration Model for Nuclear Power Business

Hui Zhang[1](✉), Siqi Jin[1], Wei Dong[2], Menghao Han[3], and Zhenyan Ji[3]

[1] Shenzhen Zhongguanghe Engineering Design Co., Ltd., Shenzhen 518172, China
zhanghui560203@163.com, jinsiqi@cgnpc.com.cn

[2] Institute of Software, Chinese Academy of Sciences, Beijing 100190, China
dongwei@iscas.ac.cn

[3] Department of Software, Beijing Jiaotong University, Beijing 100044, China
{21121731,zhyji}@bjtu.edu.cn

Abstract. To unify different departments, business systems and terminal protocols in nuclear power business, data collaboration technology and distributed data storage are introduced to realize intelligent collaboration between cloud servers and edge nodes with the help of edge-cloud collaboration, which can break through the bottleneck of multi-source heterogeneous data access. We propose a secure and efficient federated learning scheme for sharing private energy data in smart grids through edge-cloud collaboration to cope with unpredictable communication delays and potential security and efficiency issues. This model sinks the cloud computing technology to the edge side, realizes the cloud-side data collaboration model, promotes the transformation of nuclear power business to a unified and intelligent advanced model. This can effectively improve the working efficiency of nuclear power business, and ensure the security of data in the collaborative mode.

Keywords: deep learning · data collaboration · blockchain · distributed storage

1 Introduction

With the deepening of informatization in various industries, huge industry data has been accumulated, and safe and efficient data access has become a major concern for industries. The application of new technologies such as artificial intelligence, cloud computing, big data, and the Internet of Things has driven the production fast and intelligent in various industries [1, 2]. The nuclear power industry produces a large amount of data every day and has an urgent need for dataization and intelligence. Currently in the large scale domestic nuclear power business, data management is self-contained between different power plants, departments and companies of a power corporation conglomerate. A unified management and application model is lacking. The multi-base management model requires information resources to be shared on the company side and the plant side. The unified management and operation of data, models and applications between the company side and the plant side needs to be established to improve the data resource

G. Guan et al. (Eds.): ICBDSS 2022, AHCS 8, pp. 175–180, 2023.
https://doi.org/10.2991/978-94-6463-064-0_20

circulation efficiency between the company side and the plant side. To solve this problem, this paper proposes a safe and efficient federated learning scheme that utilizes distributed data storage to share private energy data in smart grids through edge-cloud collaboration to cope with unpredictable communication delays and potential safety and efficiency issues.

2 Related Work

The introduction of edge computing into the industrial Internet can meet the requirements of real-time transmission and high reliability of industrial data. The in-depth development and wide application of edge intelligence technology based on the Internet of Things have led to research on heterogeneous data collaboration under the edge cloud collaboration architecture. Zhang L et al. [1] proposed a secure and efficient data storage and sharing scheme for blockchain-based mobile edge computing. By implementing data agent signatures on strong computing nodes, massive data is guaranteed to be complete and real during uploading to the cloud server. The key shares of the signed private key is stored in different blocks of the blockchain to improve fault tolerance. Su Z et al. [3] proposed a safe and efficient joint learning AIoT (Artificial Intelligence of Things) solution, which can effectively motivate energy data owners to share high-quality local model updates and improve communication efficiency through edge cloud collaboration. Guo J et al. [4] established the Intelligent Trust Collaboration Network System to collect data through collaboration with Mobile Vehicles and Unmanned Aerial Vehicles in the Internet of Things, effectively solving the problem of lacking of security considerations in AI-based data collection methods. Bogdanova A et al. [5] experimentally demonstrated that federated learning can be successfully applied in multiparty collaborative environments by using federated learning for medical fields that require the combination of large amounts of pre-collected patient data. It not only reaches the baseline of the centralized model, but also has a regularization effect on the training process.

In order to fuse the nuclear power heterogeneous data and fully mine the data business value, we have endeavored to research on data collaboration and constructed a data collaboration model for the fusion, collaboration and sharing of massive heterogeneous data in the nuclear power industrial Internet platform.

3 Data Collaboration of Nuclear Power Business

Nowadays the companies, the departments, and the nuclear power plants of a power corporation conglomerate have their own data storage and data management and a unified data management hasn't been established. It's difficult to analyze the heterogeneous data and extract useful information effectively. To solve the problem, data collaboration technology is introduced to fulfill the equipment object model delivery from the company side to the plant side, the time series data collection and reporting, the data resource packaging/unpacking of collaborative tasks, the sending and receiving of data packets, the control of data quality, and the security and stability of data transmission, as shown in Fig. 1.

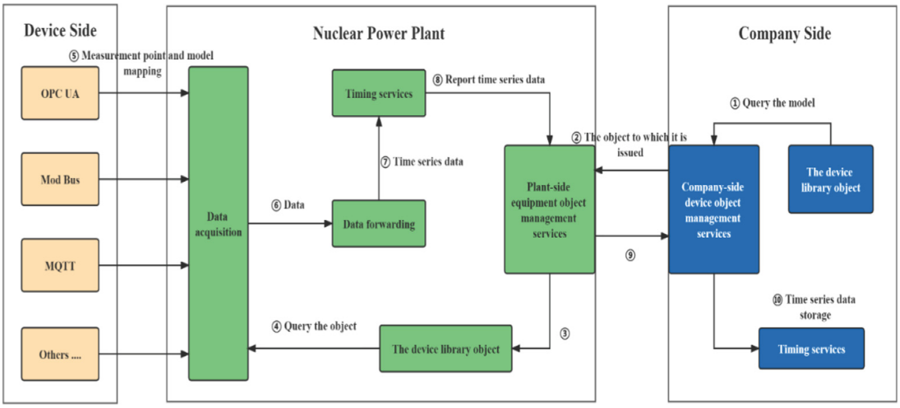

Fig. 1. Nuclear power data collaboration

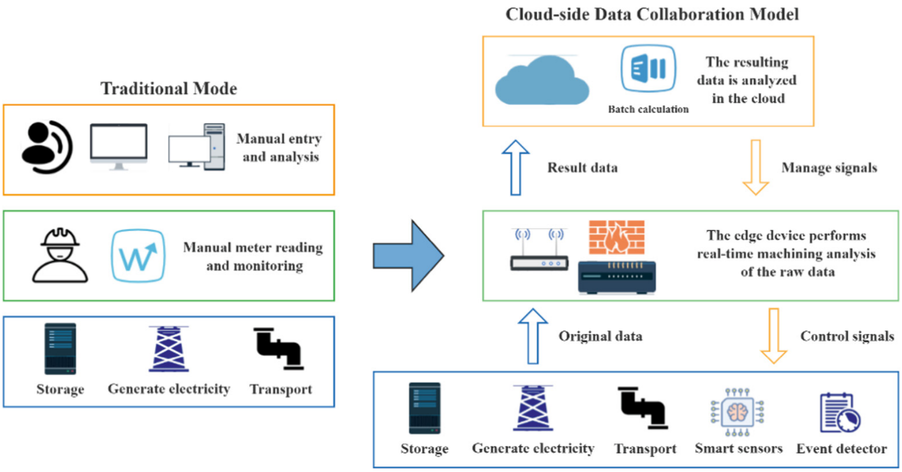

Fig. 2. Comparison of old and new modes

As shown in Fig. 2. Federated learning is adopted to protect sensitive data on plant sides and increase the efficiency of data processing. The details are as follows:

Step 1: The edge nodes collect data from corresponding sensors according to tasks assigned by the cloud, and use ADC rules to filter out noisy data. The valid data is then input into the task handler on the edge node for corresponding business data processing, and finally the task results are stored in a unified edge database to be uploaded to the cloud side.

Step 2: Cloud servers collect data from the edge nodes, and summarize the data between the companies, the departments, and the nuclear power plants of a power corporation conglomerate.

Step 3: The DAE-LSTM model is deployed to analyze the data stored in the cloud database, generate a task completion list, and decide whether and when to send the aggregated results to edge nodes to update the programs and parameters on the edge nodes.

Algorithm 1: DAE-LSTM

Input: Origin data collected to the cloud server
Output: DAE-LSTM prediction model generated by multiple rounds of training;

```
Data preprocessing: remove outliers, normalize, etc;
Inject Gaussian noise into preprocessed data;
Unsupervised pre-training: DAE-LSTM for feature extraction;
i ← 0;
while i less than IterNum do
    Supervised training: Classify data using a softmax classifier;
    Cross-validation: Compare and update hyperparameters;
    i ← i + 1;
Take the optimal parameters to generate the model weight file;
```

Step 4: Task procedures and parameter configurations of edge nodes are adjusted dynamically by analyzing the task completion degree list generated by the DAE-LSTM model. Finally, the adjustment results are transferred back to the corresponding edge node to make corresponding adjustments.

Through data collaboration technology, it effectively reduces the company's overall operation and management costs, shortens the implementation cycle of various projects, reduces the waste of people and property.

Algorithm 2: Cloud-edge Collaboration Model. m : donates the total number of edge nodes; B : donates the local minibatch size; E : donates the number of local epochs;

```
initialize the weight parameter;
for each round t = 1,2,... do
    S ← a collection of m edge nodes ;
    for each edge node k ∈ S do
        the kth edge node calls edge_update to perform the training;
    processes the currently collected weight parameters w_t;
Function edge_update(k, ω):
    split the data on the edge node into batches of size B;
    for each local epoch i from 1 to E do
        for batch b ∈ B do
            Update the weight parameters using the SGD;
    return w to server;
```

3.1 Resource Distribution

Resource distribution is responsible for packaging the collaborative data resources. The packaging abides by the company's unified format specifications. The resource distribution function packages the local tools' data model, historical dataset, real-time data, algorithm model, etc. The packages are distributed to the specified nodes. During data packaging, both manual packaging and automatic packaging based on customized rules are supported, the functional association between data package and management collaboration tasks is established. In the end, the resources are distributed to different target nodes.

3.2 Resource Reception

Data demanders receives packaged data transmitted from data owners through the resource receiving port. When the resource receiving port receives the packaged data, first stores the data in the form of a packet and provides API access.

3.3 Data Synchronization Policies

According to the association between tasks, users can set and adjust the task execution order for collaborative tasks initiated by themselves, or tasks performed by themselves but initiated by other users to ensure the execution order of the tasks.

3.4 Uni-/Bi-directional Data Management and Strict Data Permission Management

On the company side, data access control of all data assets on the edge side is designed as follows.

(1) Authorization object: the company only authorizes the administrator of the edge platform. The user authorization in the edge platform is performed by the corresponding edge platform administrator. In other words, a two-level authorization is designed.
(2) Authorized content: The administrator can authorize the relevant authorization according to the content of standard directory classification. He can specify a node or all the nodes that are aggregated in this directory, so as to facilitate the data access control within the node.
(3) Users can access specific business data only when they have both relevant services access permission and data access permissions simultaneously.
(4) Provide an integrated interface for task collaboration. Relevant permissions are obtained automatically after the permission application is approved, and manually permission configuration is not needed.

4 Conclusion

This paper studies the data management problem in the edge-cloud collaboration system. A data collaboration architecture for nuclear power business is constructed to meet the urgent requirements of the industry. In the future, the data collaboration model will be deployed to various nuclear power business to solve various data processing issues such as intelligent monitoring, intelligent decision-making, and etc.

Acknowledgements. This study is supported by the National Key Research and Development Plan Project (No. 2020YFB1711700).

References

1. Zhang L, Peng M, Wang W, et al. Secure and efficient data storage and sharing scheme for blockchain-based mobile-edge computing[J]. Transactions on Emerging Telecommunications Technologies, 2021, 32(10): e4315.
2. Hong W, You J, Zhang H. A Survey on Blockchain Application: a Data Collaboration Perspective[C]//2021 16th International Conference on Computer Science & Education (ICCSE). IEEE, 2021: 125–132.
3. Su Z, Wang Y, Luan T H, et al. Secure and Efficient Federated Learning for Smart Grid with Edge-Cloud Collaboration[J]. IEEE Transactions on Industrial Informatics, 2021, 18(2): 1333-1344.
4. Guo J, Liu A, Ota K, et al. ITCN: an intelligent trust collaboration network system in IoT[J]. IEEE transactions on network science and engineering, 2021, 9(1): 203-218.
5. Bogdanova A, Attoh-Okine N, Sakurai T. Risk and advantages of federated learning for health care data collaboration[J]. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 2020, 6(3): 04020031.
6. Imakura A, Sakurai T. Data collaboration analysis framework using centralization of individual intermediate representations for distributed data sets[J]. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 2020, 6(2): 04020018.