Hitachi Review

October 2024 Technical Information

Federated Learning toward Network-distributed AI

#Generative AI #IoT/Data Utilization #Research & Development

Author

Takahito Tanimura, Ph.D.
Edge Intelligence Research Department, Digital Platform Innovation Center, Research & Development Group, Hitachi, Ltd.
Current work and research: Research and development of networks and distributed AI.
Society memberships: A senior member of the Institute of Electronics, Information and Communication Engineers (IEICE) and a member of the Physical Society of Japan (JPS).

Shinji Tarumi
Healthcare IT Research Department, Healthcare Innovation Center, Research & Development Group, Hitachi, Ltd.
Current work and research: Research and development of healthcare data analysis and distributed AI.
Society memberships: The Japanese Society for Artificial Intelligence (JSAI).

Yuichi Kitagawa
Edge Intelligence Research Department, Digital Platform Innovation Center, Research & Development Group, Hitachi, Ltd.
Current work and research: Research and development of distributed AI.
Society memberships: IEICE.

Yuya Isoda
Data Management Research Department, Digital Service Platform Innovation Center, Research & Development Group, Hitachi, Ltd.
Current work and research: Research and development of data management and application management.
Society memberships: The Information Processing Society of Japan (IPSJ).

Wakako Nakano
Edge Intelligence Research Department, Digital Platform Innovation Center, Research & Development Group, Hitachi, Ltd.
Current work and research: Research and development of computer vision and distributed AI.
Society memberships: IEICE and IPSJ.

Masayuki Takase
Edge Intelligence Research Department, Digital Platform Innovation Center, Research & Development Group, Hitachi, Ltd.
Current work and research: Research and development of networks and distributed AI.
Society memberships: IEICE.

Highlight

Large-scale data acquisition is a critical issue for the development of industry-specific AI models, including deep learning and generative AI, trained with domain-specific adaptation. To address this issue, federated learning, a technique that enables the use of highly confidential data distributed across different organizations without compromising its confidentiality, has been proposed and developed.

This article describes how federated learning works and Hitachi's activities in this area. We also discuss its potential for use in generative AI, a field that has attracted considerable attention in recent years.

1. Introduction

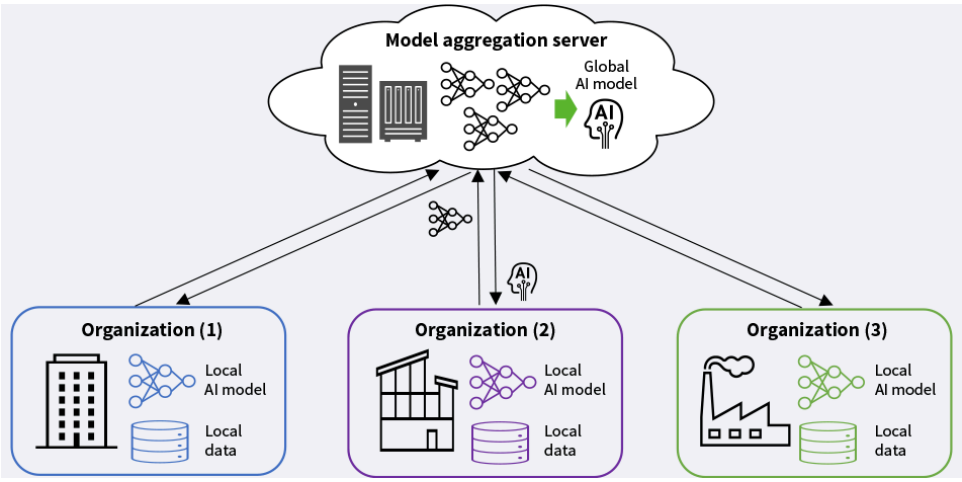
The rapid development of artificial intelligence (AI) today relies heavily on large amounts of data for training. The acquisition of large volumes of high-quality domain- and industry-specific data is a critical step in the development of advanced AI services that target specific domains or industries.

To address the challenges associated with acquiring this data, federated learning 1) provides a technique for exploiting data that is distributed across multiple organizations. Traditional techniques have worked by consolidating data in a single location, a practice that raises potential issues of confidentiality and privacy. With federated learning, distributed data can be used to train AI without compromising data privacy. This article describes the basics of how federated learning works, the challenges it faces, and considers its potential for use in generative AI.

2. Overview of Federated Learning

Federated learning is a technique for distributed AI training that allows an AI model to learn using all the data that is distributed across multiple locations, without physically collecting the data at one location. Figure 1 shows a basic overview of how federated learning works.

Figure 1—How Federated Learning Work



AI: artificial intelligence

Federated learning can train an AI model collaboratively across multiple organizations using training data held separately in each organization.

Federated learning trains an AI model through the following steps.

(1) Each participant trains its own AI model (local AI model) using its own training data (local data).

(2) Each participant sends its trained local AI model to the model aggregation server instead of sending its training data.

(3) The model aggregation server aggregates these local AI models to generate an AI model (global AI model) that inherits the features of all the models received from the participants.

(4) The model aggregation server sends the global AI model to each location, and at each location, the participants retrain the global AI model.

(5) Steps (1) to (4) are repeated to keep the model updated.

An essential point to note about this technique is that it generates a global AI model from all the local data of the participants without sending any of this data to the server. A global AI model generated by the federated learning technique should outperform an individually trained model using only the limited data of each individual participant.

3. Use Cases for Federated Learning

Federated learning is particularly useful for groups of organizations that face common challenges and want to work together to solve them. Federated learning can be used in areas where participants can benefit from sharing knowledge and insights.

Healthcare is a typical area for federated learning: Federated learning allows a group of healthcare organizations, such as hospitals, to work together to build advanced diagnostic models without having to share patient medical records, which are sensitive and private. The use of federated learning provides access to a much larger amount of data than could be obtained by any one medical institution, while still protecting patient privacy.

The One Hitachi initiative focused on the property insurance industry, where sales are made through agencies. It gained extensive knowledge by using federated learning to virtually integrate documents and data stored in different agencies.

Elsewhere, federated learning has the potential to be used in a wide range of areas, such as the joint development of fraud detection models in the financial industry and failure risk diagnosis solutions for the manufacturing industry.

4. Technical Challenges of Federated Learning and Hitachi's Technology

4.1 Imbalance in Data Held by Different Organizations and How This Is Solved

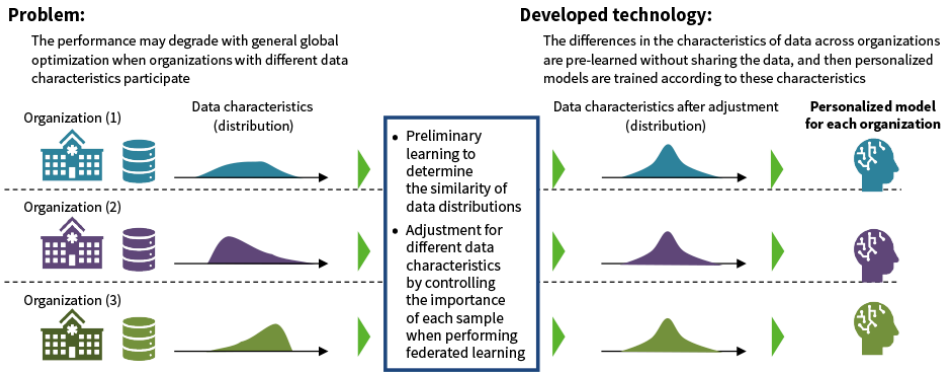
One of the major challenges in federated learning is how to address data imbalances. It is well-known that the performance of models built using federated learning suffers from bias when the data held by the different participants vary in quantity and quality. To address this problem, Hitachi has developed a personalized federated learning technique [2] that takes account of the characteristics of participants' data when building models. In contrast to conventional federated learning, which trains a single global model, this technique is characterized by its ability to generate personalized models for each participant.

In more detail, this technique first estimates the similarity of the data held by each participant, without requiring any of the participants to disclose their data to one another directly. It then uses the estimated similarity to perform federated learning while flexibly adjusting the importance, i.e., the contribution to learning, of each data sample (see Figure 2). This method can create personalized models for each participant by prioritizing learning on data with characteristics similar to that of their own.

Using a publicly available medical dataset collected from multiple hospitals, the performance of survival prediction models created using personalized federated learning was evaluated. The results indicated that the models outperformed the model created using conventional federated learning. Its performance also compared favorably to the ideal scenario in which machine learning was performed on the consolidated set of all hospital data.

The use of this technique ensures that all participants in federated learning can reliably be provided with high-performance models. This makes it suitable for a wide variety of applications where data imbalances are a problem.

Figure 2—Personalized Federated Learning



Whereas conventional federated learning creates a single global model, user-specific personalized models can be created for each participant.

4.2 Increased Network Usage and How This Is Solved

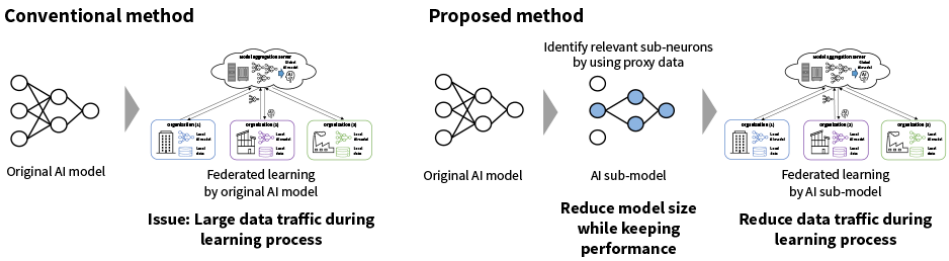
One of the technical issues with federated learning is that network traffic increases significantly during the training process as each participant sends model information to the aggregation server and the aggregation server sends model information back to the participants. The issue becomes more serious when both the number of participants and the size of the model are large. To solve this problem, Hitachi has developed a communication-efficient federated learning technique that does not place a heavy load on the network.

This technique works by extracting and transmitting smaller AI sub-models derived from the original AI model using the lottery ticket hypothesis* and proxy data sets [3]. Data-driven techniques for extracting an AI sub-model from an original AI model are difficult to use in practice because the model aggregation server used in federated learning is not allowed to store local confidential data. Hitachi has solved this problem by extracting AI sub-models using proxy data that can be synthesized at the model aggregation server. This solution provides results comparable to those using the actual confidential data (see Figure 3) with reduced communication traffic.

By reducing traffic during the federated learning process with this technique, there is an opportunity to use federated learning in narrow-band networks, including mobile and IoT networks.

* The hypothesis that a dense deep neural network model contains sub-models, called "lottery tickets," that perform similarly to the original model but with fewer parameters.

Figure 3—Diagram of Communication-efficient Federated Learning



The volume of data traffic during the federated learning process is reduced by using AI sub-models for the exchange of model information across the network.

5. Future of Federated Learning with Generative AI

5.1 Use of Federated Learning to Customize Generative AI

Recent advances in generative AI, especially generative AI that is customized by internal data sets and knowledge, could be an essential part of improving workflows in organizations. The most common method of customizing generative AI is to fine-tune an existing foundation model using data that different users have for their particular purposes.

If a sufficiently large and high-quality dataset is available for this fine-tuning, it is possible to obtain a useful custom generative AI model that is suitable for a more specialized objective. In other cases, however, fine-tuning is sometimes difficult due to a lack of training data. The training data used for this fine-tuning is sometimes confidential to the organization and therefore cannot be easily shared with other entities.

By using federated learning in generative AI fine-tuning, it is possible to address this issue on the training data while still keeping it private. This collaborative fine-tuning through federated learning can provide a way for generative AI models to have not only generic knowledge, but also domain-specific expertise based on the experience of the organizations, while keeping all training data confidential.

5.2 Use of Federated Learning to Extend Generative AI Capability

Federated learning can be used to extend the capabilities of generative AI and/or generative AI-based agents, which are autonomous agents augmented by generative AI, to solve problems. In practice, generative AI is augmented with predictive AI models that can be trained by federated learning. The federated-learned predictive AI models, which are customized by confidential domain-specific datasets as described in the previous section, can be used as its utility tools for the generative AI model. The two types of models can be kept separate but connected, with the generative AI as a human interface, and the predictive AI as its tool.

One potential application is in financial advisory services. Such services could combine flexible customer interaction with domain-specific analysis capabilities by using an interactive agent based on generative AI to handle communication with the customer and to invoke advanced financial prediction models built using federated learning when they are needed for specific financial predictions or analysis.

Other possibilities for using this type of integration could include presenting the output of federated learning models in a natural language format, or incorporating predictions made by federated learning models into the generative AI inference process. Such an approach has the potential to build highly reliable AI systems that are easier for users to understand.

6. Conclusions

This article provided an overview, the challenges faced, and the solutions found for federated learning. As an extension of federated learning in the future, it also discussed the use of federated learning with generative AI.

REFERENCES

1) H. B. McMahan et al., “Communication-Efficient Learning of Deep Networks from Decentralized Data,” Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) 2017, pp. 1273-1282 (2017)

2)

S. Tarumi et al., “Personalized Federated Learning for Institutional Prediction Model using Electronic Health Records: A Covariate Adjustment Approach,” 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 1-4 (Jul. 2023)

3) T. Tanimura et al., “Compressing Model before Federated Learning by Transferrable Surrogate Lottery Ticket,” 2023 IEEE 20th Consumer Communications & Networking Conference (CCNC), pp. 620-623 (Jan. 2023)

© Hitachi, Ltd. 1994, 2024. All rights reserved.

Hitachi Review

Hitachi Review is a technical medium that reports on Hitachi's use of innovation to address the challenges facing society.

The *Hitachi Review* website contains technical papers written by Hitachi engineers and researchers, special articles such as discussions or interviews, and back numbers.

Hitachi Hyoron (Japanese) website	*Hitachi Review* (English) website
https://www.hitachihyoron.com/jp/	https://www.hitachihyoron.com/rev/

Hitachi Review Newsletter

Hitachi Review newsletter delivers the latest information about Hitachi Review when new articles are released.