Babylonian Journal of Internet of Things
Vol.2023, **pp**. 1–7
DOI: https://doi.org/10.58496/BJIoT/2023/001; ISSN: 3006-1083
https://mesopotamian.press/journals/index.php/BJIoT

Research Article

Federated Learning in IoT: A Survey on Distributed Decision Making

Rasha Talal Hameed[1], Omar Abdulwahabe Mohamad[1,*]

[1] *Computer Science, College of Education, Al-Iraqia University, Iraq.*

ARTICLE INFO

Article History

Received 17 Nov 2022
Accepted 04 Jan 2023
Published 10 Jan 2023

Keywords

Federated Learning

Internet of Things (IoT)

Distributed Decision Making

Decentralized Machine Learning

IoT Applications

ABSTRACT

The proliferation of Internet of Things (IoT) devices has ushered in an era of unprecedented data generation, necessitating innovative approaches to extract valuable insights while respecting privacy and resource constraints. Federated Learning (FL) has emerged as a promising paradigm, enabling decentralized model training across a network of edge devices. This paper presents a comprehensive survey on the application of Federated Learning in IoT, with a specific focus on the challenges and solutions related to distributed decision making. The survey begins by elucidating the foundational concepts of Federated Learning and IoT, highlighting their convergence and the potential benefits of leveraging FL in decentralized IoT ecosystems. The paper explores the diverse applications of FL in various IoT domains, including smart cities, healthcare, industrial IoT, and smart grid systems. It investigates how FL can address the challenges posed by the distributed nature of IoT data, such as data heterogeneity, privacy concerns, and communication constraints. A significant portion of the survey is dedicated to examining the methodologies and algorithms employed in federated learning for distributed decision making in IoT. This encompasses a discussion on federated optimization techniques, communication-efficient algorithms, and privacy-preserving mechanisms. The survey also delves into the role of edge computing in facilitating efficient FL in IoT, considering the resource constraints inherent in edge devices. Furthermore, the paper reviews the current state-of-the-art federated learning frameworks and platforms tailored for IoT environments. It evaluates their capabilities in handling real-world challenges and providing scalable solutions for distributed decision making. The survey concludes by identifying open research challenges and potential avenues for future developments in federated learning for IoT, emphasizing the need for novel algorithms, robust security measures, and standardized frameworks. In summary, this paper offers a comprehensive overview of Federated Learning in the context of IoT, shedding light on its potential, challenges, and solutions with a specific emphasis on the critical aspect of distributed decision making. The insights provided aim to guide researchers, practitioners, and policymakers in navigating the complex landscape of FL in IoT and fostering advancements in this rapidly evolving field.

1. INTRODUCTION

The proliferation of smart devices and Internet-of-Things (IoT)[1] networks has led to more distributed computing environments where data is generated and stored across countless nodes.[2] This decentralized nature of IoT ecosystems poses notable challenges for applying data-driven machine learning models traditionally dependent on accessing large, centralized datasets. Federated learning has emerged as a distributed collaborative framework that enables training machine learning algorithms across decentralized edge devices or servers while keeping data localized and private. This approach is especially compelling to explore for IoT networks, with vast amounts of sensitive user data generated across smart devices and sensors.

In federated learning applied to IoT systems, a shared global model can be cooperatively trained from localized data samples on devices without the need to exchange or pool raw data from users.[3] This preserves privacy while allowing collective learning from statistical aggregations across the network. The trained global model is then disseminated to devices to utilize for productive decision making tailored to new localized data.

*Corresponding author. Email: engit2020@gmail.com

While showing promise, federated learning opens stimulating research questions around efficiency, robustness, and security when put into practice - especially for real-world IoT deployments at population-scale. This survey reviews federated learning developments for IoT contexts with a lens towards distributed decision making in particular. We summarize the current landscape of algorithms, applications, and open challenges for effectively coordinating intelligent decisions across decentralized IoT devices and users via federated learning's distributed model training approach.

The organization of this paper is as follows. First, we outline the fundamentals of federated learning and its distinguishing components. We then detail popular applications of federated learning specifically using IoT network data across domains like smart homes, smart cities, and healthcare. Focusing on decision making, we next categorize different consensus algorithms and other distributed techniques investigated in the literature to coordinate decisions across IoT devices informed by shared models from federated learning. Finally, we suggest promising directions and open problems to advance the convergence of federated learning and decentralized decision making tailored to practical IoT networks with potential real-world societal impact.

2. BACKGROUND

2.1 Federated Learning:

Federated Learning (FL)[4] represents a paradigm shift in the field of machine learning, particularly in the context of distributed and decentralized data environments. It addresses the challenges posed by the increasing ubiquity of connected devices, such as smartphones, Internet of Things (IoT)[5] devices, and edge computing nodes. The traditional model of centralized machine learning, where data is collected and processed in a central server, faces limitations in scenarios where data privacy, communication bandwidth, and real-time processing are critical concerns. The fundamental premise of Federated Learning is to bring the machine learning model to the data, rather than centralizing the data for model training. This decentralized approach is especially pertinent in settings where data is sensitive, massive, or distributed across a network of devices with limited connectivity. FL enables collaborative model training across a multitude of devices without the need to transfer raw data to a central server, thereby addressing privacy concerns and alleviating the communication burden associated with transmitting large datasets.

Decentralized Model Training: In FL, model training occurs locally on individual devices, and only model updates or gradients are shared with a central server or among neighboring devices. This decentralized nature reduces the need for extensive data transfer and minimizes the risk associated with exposing raw, sensitive data.

Privacy Preservation: Privacy is a critical consideration in FL. By keeping data local and transmitting only model updates, FL ensures that sensitive information remains on the device. This is particularly important in applications such as healthcare, finance, and personal devices where user privacy is paramount.

Communication Efficiency: FL optimizes communication by transmitting only essential model updates. This is crucial for resource-constrained devices, as it reduces the need for high-bandwidth communication and minimizes latency, making FL suitable for real-time and low-latency applications.

Adaptability to Edge Computing: The rise of edge computing, where data processing occurs closer to the data source, aligns well with the principles of FL. Edge devices, such as sensors and IoT devices, can actively participate in model training without relying heavily on centralized cloud infrastructure, leading to faster decision-making and reduced dependence on external servers.

Applications of Federated Learning span various domains, including healthcare, finance, smart cities, and industrial IoT. For instance, FL can be employed in predictive maintenance of industrial machinery, collaborative learning in educational apps, or improving personalized healthcare recommendations without compromising patient data.

As Federated Learning continues to evolve, ongoing research efforts are addressing challenges related to model aggregation, security, robustness, and scalability. The versatility of FL makes it a promising solution for the growing demand for intelligent systems in decentralized and privacy-sensitive environments..

2.2 Internet of Things (IoT)

The Internet of Things (IoT)[6] refers to the interconnected network of physical devices embedded with sensors, actuators, software, and other technologies, allowing them to collect and exchange data.[7] These devices, often referred to as "smart" or "connected" devices, communicate with each other and with centralized systems to perform various tasks, monitor environments, and provide valuable insights. The concept of IoT has evolved over the years, driven by advancements in technology and the desire to create more efficient and interconnected systems.

Key Components of IoT:

Sensors and Actuators: IoT devices are equipped with sensors to gather data from the surrounding environment and actuators to perform specific actions based on the collected data. Common sensors include temperature sensors, motion sensors, and cameras, while actuators might include motors, valves, or other mechanisms for control.

Connectivity: IoT[8] devices rely on connectivity to share data. This can be achieved through various communication protocols such as Wi-Fi, Bluetooth, Zigbee, cellular networks, or Low Power Wide Area Networks (LPWAN). The choice of connectivity depends on factors like data transfer speed, range, and power consumption.

Data Processing: IoT devices often process data locally using onboard computing capabilities or send the data to centralized cloud servers for analysis. Edge computing, where data is processed closer to the source, has gained prominence in IoT to reduce latency and enhance real-time decision-making.

Cloud Computing: Cloud platforms play a crucial role in storing and analyzing the vast amounts of data generated by IoT devices. Cloud computing provides the scalability and computational resources needed for handling the diverse and often massive datasets generated by IoT ecosystems.

Security: Given the interconnected nature of IoT, security is a paramount concern. IoT devices may be vulnerable to cyber-attacks, and compromised devices can have serious consequences. Security measures include encryption, authentication protocols, and regular software updates to patch vulnerabilities.

Interoperability: The interoperability of diverse IoT devices and systems is essential for seamless communication and collaboration. Standardization efforts, such as common communication protocols and data formats, are ongoing to ensure compatibility among different IoT devices.

3. SIGNIFICANCE OF FL IN THE CONTEXT OF INTERNET OF THINGS (IOT)

The integration of Federated Learning (FL)[9] within the landscape of the Internet of Things (IoT)[10] holds profound significance, ushering in a paradigm shift in how machine learning models are trained and decision-making is executed within decentralized and interconnected environments. Several factors underscore the critical importance of FL in the context of IoT:

1. Decentralized Data Governance:

IoT ecosystems generate vast amounts of data distributed across diverse devices, often with stringent privacy concerns. FL enables decentralized model training without the need to centralize sensitive data, preserving individual device privacy and adhering to data governance regulations.

2. Preserving Data Locality:

Transmitting large volumes of raw IoT data to centralized servers can be impractical, resource-intensive, and pose security risks. FL brings the model to the data, allowing devices to locally process and contribute insights without compromising data security or burdening communication networks.

3. Mitigating Communication Overhead:

Traditional machine learning models require frequent communication with a central server, which may be inefficient, particularly in low-bandwidth or intermittent connectivity scenarios. FL minimizes communication overhead by transmitting only model updates, making it well-suited for IoT environments where efficient communication is essential.

4. Privacy-Preserving Machine Learning:

Protecting user privacy is a critical concern in IoT, where personal and sensitive data is often involved. FL employs techniques such as federated averaging and homomorphic encryption to ensure privacy, allowing devices to contribute to model training without revealing individual data.

5. Edge Intelligence and Real-time Decision-Making:

Traditional cloud-based approaches may introduce latency in decision-making, impacting real-time applications in IoT. FL is adaptable to edge computing, facilitating localized model updates and enabling faster, context-aware decision-making directly on IoT devices.

4. CHALLENGES IN CENTRALIZED MACHINE LEARNING FOR IOT

The conventional approach of centralized machine learning, where data is collected and processed in a central server, faces several challenges when applied to the dynamic and distributed nature of the Internet of Things (IoT) environments. Recognizing and addressing these challenges is essential for the effective integration of machine learning in IoT scenarios. The primary challenges include:

1. Data Privacy and Security:

Distributed Sensitive Data: IoT devices often generate sensitive data such as personal health information, location data, or industrial process details. Centralized models require data to be transmitted to a central server, raising concerns about data privacy during transit.

Vulnerability to Cyber Attacks: Centralized systems become attractive targets for cyber-attacks. A breach in the central server compromises the entirety of the collected data, posing significant privacy and security risks.

2. Communication Overhead:

Bandwidth Limitations: Transmitting large volumes of data from numerous IoT devices to a central server can strain network bandwidth, leading to increased latency and potential data transmission bottlenecks.

Real-time Processing Challenges: Applications demanding real-time decision-making may be hindered by the time delay caused by data transfer to a centralized location.

3. Scalability Issues:

Growing Number of Devices: The expanding network of IoT devices results in an exponential increase in data volume. Centralized systems may struggle to scale efficiently, leading to performance degradation and resource constraints.

Resource-Intensive Processing: Centralized processing requires significant computational resources, which can become a bottleneck as the number of connected devices rises.

4. Data Heterogeneity:

Diverse Data Sources: IoT ecosystems consist of diverse devices with varying data formats, structures, and semantics. Centralized models may struggle to adapt to the heterogeneity of data sources, leading to suboptimal performance.

5. Reliability and Availability:

Dependency on Central Server: Centralized systems are vulnerable to server failures, network outages, or maintenance issues. Such dependencies can result in disruptions in service availability and reliability.

6. Regulatory Compliance:

Data Localization Requirements: Some jurisdictions impose regulations that necessitate data to be processed and stored within specific geographic boundaries. Centralized models may face challenges in adhering to these regulations.

7. Energy Consumption:

Communication Energy Overhead: Transmitting data from IoT devices to a central server consumes energy, especially in resource-constrained devices. This energy overhead may be significant for battery-powered devices, affecting their operational lifespan.

Addressing these challenges is crucial for ensuring the successful integration of machine learning in IoT environments. Federated Learning (FL) emerges as a promising solution by decentralizing the model training process, mitigating privacy concerns, reducing communication overhead, and enhancing scalability and adaptability to diverse IoT ecosystems.

5. APPLICATIONS OF FEDERATED LEARNING IN IOT

Federated Learning (FL) has gained traction as a transformative approach in the realm of Internet of Things (IoT), offering a decentralized and privacy-preserving alternative to traditional machine learning models. The applications of Federated Learning in IoT span various domains, showcasing its versatility and adaptability to diverse scenarios. Some notable applications include:

1. Smart Cities:

Traffic Management: FL enables traffic prediction and optimization by utilizing data from various sources such as traffic cameras, sensors, and mobile devices. Decentralized decision-making allows for real-time adjustments to traffic flow.

Energy Consumption Optimization: FL can be applied to optimize energy consumption in smart grids, ensuring efficient energy distribution and reducing wastage.

2. Healthcare:

Remote Patient Monitoring: FL facilitates the development of personalized healthcare models without compromising patient privacy. Healthcare data from wearable devices, sensors, and electronic health records can be collectively used to enhance predictive models for disease monitoring and early intervention.

Drug Discovery: Collaborative learning through FL can accelerate drug discovery processes by aggregating insights from various research institutions and pharmaceutical companies without sharing sensitive data.

3. Industrial IoT (IIoT):

Predictive Maintenance: FL enables predictive maintenance models by aggregating information from sensors on industrial machinery. This ensures timely identification of potential faults, reducing downtime and maintenance costs.

Quality Control: Decentralized decision-making in FL allows for real-time quality control in manufacturing processes by analyzing data from sensors on the production line.

4. Smart Agriculture:

Precision Farming: FL can aid in precision farming by combining data from various agricultural IoT devices, such as soil sensors, drones, and weather stations. This allows for optimized decision-making regarding irrigation, fertilization, and crop management.

5. Edge Devices and IoT Sensors:

Energy-Efficient Devices: FL can be employed in optimizing the energy consumption of IoT devices by training models locally on edge devices. This reduces the need for frequent communication with central servers and extends the operational life of battery-powered devices.

Anomaly Detection: FL enables collaborative anomaly detection in sensor networks by aggregating insights from multiple sensors. This is particularly valuable in applications like environmental monitoring and security surveillance.

6. Smart Homes:

Personalized Automation: FL allows for the creation of personalized automation models by learning user preferences from various smart home devices. This includes adjusting lighting, temperature, and other parameters based on individual habits.

7. Finance:

Fraud Detection: FL can enhance fraud detection models in the financial sector by collaboratively learning from transaction data across multiple banks without exposing sensitive customer information. Decentralized decision-making in FL can improve risk assessment models by aggregating insights from diverse financial sources, contributing to more accurate predictions.

These applications demonstrate the broad impact of Federated Learning in addressing challenges related to data privacy, communication efficiency, and decentralized decision-making in the context of Internet of Things. As FL technologies continue to advance, the scope for innovative applications in IoT is expected to expand further.

6. OPEN RESEARCH CHALLENGES AND FUTURE DIRECTIONS

As Federated Learning (FL) continues to evolve as a prominent paradigm, particularly within the intricate landscape of the Internet of Things (IoT), several open research challenges beckon the attention of scholars, practitioners, and policymakers. Addressing these challenges is crucial for unlocking the full potential of FL in IoT ecosystems and steering the research community towards meaningful advancements. Furthermore, identifying future directions provides a roadmap for sustained innovation and the integration of FL into a wide array of applications. The following outlines key open research challenges and potential avenues for future exploration:

Developing robust model aggregation techniques capable of accommodating dynamic and heterogeneous IoT environments remains a challenge. Adapting to changing device participation and data distributions poses inherent complexities.

Adversarial Threats: Mitigating adversarial attacks on FL models in decentralized IoT settings is a critical challenge. Developing robust defenses against privacy breaches and model poisoning attacks is imperative for widespread adoption.

Reducing Communication Overhead: Optimizing communication protocols and minimizing the volume of exchanged data while preserving model accuracy is an ongoing challenge. Efficient strategies for handling intermittent connectivity and low-bandwidth scenarios are essential.

Large-Scale Deployments: Scaling FL frameworks to accommodate a massive number of IoT devices and ensuring efficient coordination among them present challenges. Scalability becomes crucial as the number of connected devices in IoT ecosystems continues to grow.

Optimal Edge-Cloud Balance: Determining the optimal balance between edge computing and centralized cloud processing in FL-IoT settings is a challenge. Striking the right balance ensures effective utilization of resources while meeting real-time processing requirements.

Framework Interoperability: The absence of standardized protocols and frameworks for FL in IoT hampers interoperability. Establishing common standards is essential for fostering collaboration among diverse devices and platforms.

Minimizing Device Energy Consumption: FL on resource-constrained edge devices demands energy-efficient algorithms. Balancing model accuracy with energy consumption is critical, especially in IoT scenarios reliant on battery-powered devices.

Integrating FL with Other Techniques: Exploring hybrid models that combine FL with other machine learning techniques, such as transfer learning or reinforcement learning, to enhance model performance and adaptability in dynamic IoT environments.

Interpretable FL Models: Investigating methods to make FL models more interpretable and transparent, addressing the "black-box" nature of some machine learning models in decentralized settings.

7. CONCLUSION

In conclusion, this survey has delved into the dynamic intersection of Federated Learning (FL) and the Internet of Things (IoT), illuminating the significance of decentralized decision-making in IoT environments. The exploration of applications, challenges, methodologies, and future directions underscores the transformative potential of FL in addressing the complexities inherent to the vast and diverse IoT landscape.Federated Learning emerges as a compelling solution to the challenges posed by centralized machine learning in IoT, offering a paradigm that aligns with the principles of data privacy, communication efficiency, and adaptability to edge computing. The applications of FL in various domains, including smart cities, healthcare, and industrial IoT, showcase its versatility and potential to reshape how insights are gleaned from distributed IoT data sources. The challenges identified, spanning data privacy, security, communication overhead, and scalability, underscore the need for continued research and innovation. Addressing these challenges will be pivotal in ensuring the seamless integration of FL in IoT ecosystems, fostering trust among users, and promoting the adoption of decentralized decision-making models. Looking ahead, the outlined future directions provide a roadmap for researchers and practitioners to explore hybrid models, enhance explainability, promote cross-domain collaboration, and navigate ethical considerations. These avenues pave the way for the evolution of FL applications in IoT beyond theoretical frameworks, fostering real-world implementations that contribute to the advancement of both fields. As Federated Learning in IoT matures, collaboration among academia, industry, and policymakers becomes increasingly vital. Standardization efforts, interdisciplinary research, and the establishment of regulatory frameworks will play crucial roles in shaping the ethical, legal, and practical aspects of FL deployment in IoT ecosystems.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

No funding is provided and no financial support is received to carry out the research presented in this paper.

Acknowledgment

The author would like to express gratitude to the institution for their invaluable support throughout this research project.

References

[1] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. V. Poor, "Federated learning for internet of things: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1622-1658, 2021.

[2] A. Imteaj, U. Thakker, S. Wang, J. Li, and M. H. Amini, "A survey on federated learning for resource-constrained IoT devices," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 1-24, 2021.

[3] A. Imteaj and M. H. Amini, "Distributed sensing using smart end-user devices: Pathway to federated learning for autonomous IoT," in *2019 International conference on computational science and computational intelligence (CSCI)*, 2019, pp. 1156-1161: IEEE.

[4] L. Li, Y. Fan, M. Tse, and K.-Y. Lin, "A review of applications in federated learning," *Computers & Industrial Engineering*, vol. 149, p. 106854, 2020.

[5] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," *Knowledge-Based Systems*, vol. 216, p. 106775, 2021.

[6] S. Madakam, V. Lake, V. Lake, and V. Lake, "Internet of Things (IoT): A literature review," *Journal of Computer and Communications*, vol. 3, no. 05, p. 164, 2015.

[7] C. Srinivasan, B. Rajesh, P. Saikalyan, K. Premsagar, and E. S. Yadav, "A review on the different types of Internet of Things (IoT)," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 11, no. 1, pp. 154-158, 2019.

[8] K. Ashton, "That ‘internet of things’ thing," *RFID journal*, vol. 22, no. 7, pp. 97-114, 2009.

[9] L. Farhan, S. T. Shukur, A. E. Alissa, M. Alrweg, U. Raza, and R. Kharel, "A survey on the challenges and opportunities of the Internet of Things (IoT)," in *2017 Eleventh International Conference on Sensing Technology (ICST)*, 2017, pp. 1-5: IEEE.

[10] N. Hossein Motlagh, M. Mohammadrezaei, J. Hunt, and B. Zakeri, "Internet of Things (IoT) and the energy sector," *Energies*, vol. 13, no. 2, p. 494, 2020.