SATBAYEV UNIVERSITY

Computing & Engineering

Volume 1 (2023), Issue 3, 25-28

https://doi.org/10.51301/ce.2023.i3.05

Applications in federated machine learning

G. Bektemyssova, G. Bakirova*, G. Shaikemelev

International Information Technology University, Almaty, Kazakhstan

**Corresponding author: g.bakirova@iitu.edu.kz*

Abstract. In our paper we figured out, that federated learning (FL) is a deep learning technique used in various industries, including medicine, agriculture, vehicles, retail and finance. It offers privacy, data ownership, localized model training, bandwidth efficiency, real-time learning, scalability and resilience to device failures. In medicine FL can improve patient's representation, drug development, medical image analysis, sickness diagnosis and individualized treatment planning. In agriculture, FL can improve crop irrigation, fertilization, harvesting and monitoring animal health. In retail, FL can analyze customer behavior data, preserving privacy. As we understand, federated learning divides model training among local data sources using sensors like GPS, microphones, and cameras. But learning models can be hacked by various threats, including data poisoning attacks.

Keywords: *federated learning, poisoning attack, decentralized, centralized, cross silo.*

1. Introduction

Global scientific interest in federated learning has increased dramatically in recent years. Finding ways to improve Federated Learning approach performance in the three key areas of resource allocation, communication cost efficiency, and security and privacy—all of which are essential for the technology's practical application—took up a sizable amount of the research [1].

Federated learning's fundamental concept is to divide up model training among several local data sources. Despite their rapid sequence of changes, many devices are equipped with strong sensors including GPS, microphones, and cameras. It implies that they have an incredible amount of access to data, the majority of which is naturally confidential. The goal of federated learning is to train machine learning algorithms across many decentralized edge devices that store and exchange local data samples. During the federated learning process, the training data is retained locally with each member. This approach allows for the sharing of each member's training data while simultaneously guaranteeing the privacy of each member.

The FL architecture can be broadly divided in different categories based on Figure 1.

1) Scenario: cross device, cross silo. There are many customers in cross-device FL scenarios, only a tiny portion can be trained. A lack of resources and communication barriers lead to a higher number of customers quitting during training. Conversely, cross-silo FL incorporates many companies or data centers, where the primary issues are in computing overhead and communication.

2) Methods are divided into federated transfer learning, vertical and horizontal, when datasets have various sample spaces but the same feature space.

3) Communication architecture. There are three different kinds of communication architectures: decentralized, hierarchical and centralized. Clients are connected to a worldwide server using the cloud-based architecture of centralized FL. Being an edge-based architecture, hierarchical FL serves as a mediator between clients and servers. Decentralized FL enables clients to share local model updates and aggregate at any client's end while cooperatively training models without requiring a connection to any cloud or edge server [2].

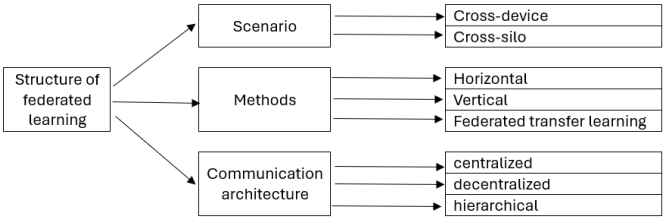

Figure 1. Structure of FL

2. Materials and methods

FL has becoming more well-known as a cooperative method for creating machine learning models, and it has been used in many different fields. Early adopters have seen its possibilities and used it in practical situations [3]. Examples of applications in federated machine learning, you can see on Figure 2.

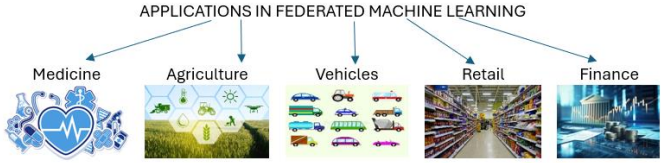

Figure 2. Applications in federated machine learning

Applications in federated machine learning:

1. Medicine. While protecting patient privacy, FL provides a useful technique for training machine learning mod-

els in the medicine sector. Enhancing patient representation, it may be applied to drug development, medical image analysis, sickness diagnosis, and individualized treatment planning.

2. Agriculture. FL has potential applications in agriculture, providing insights on crop irrigation, fertilization, and harvesting. It can also monitor animal health. In agriculture, FL can combine data from weather forecasts, soil and satellite imagery for accurate predictions. Industrial automation involves training models on data from factory sensors, enabling more accurate predictions on agricultural outcomes. FL is a technology that trains models on decentralized data, ensuring privacy and promoting collaboration among farmers, researchers and organizations. It can transform industries, including agriculture, by avoiding centralization, preserving data privacy, facilitating collaborative model training, tailoring models to local needs, aiding in disease and pest detection, and aiding in crop improvement. FL also aids in crop rotation, fertilization, and irrigation decisions, boosting productivity and identifying genetic markers for improved crop varieties.

Federated learning proposes a revolutionary strategy for combating crop diseases in global agriculture. In addition to promoting sustainable agriculture and global food security, this method, which protects the privacy and security of data, has demonstrated encouraging results in the detection and classification of illnesses. The federated learning model should be improved in the future, the investigation of sophisticated privacy-preserving techniques, and the expansion of applications to other crop and disease categories [4].

3. Vehicles. Autonomous vehicles can share training models among themselves, allowing them to learn from each other's experiences while maintaining privacy. For instance, a vehicle can update its model based on new road conditions or obstacles [5].

Also, tire grip, suspension responsiveness, brake efficiency, and vehicle performance are all strongly impacted by the kind and quality of the road surface. Inadequate road conditions can shorten a vehicle's lifespan, increase wear and tear, and raise maintenance expenses. By being aware of these variables, drivers may modify their driving techniques, maximize fuel efficiency, and lower emissions. A broader spectrum of road users can benefit from the accurate road surface categorization that a FL method provides across a large geographic region [6].

Smart driving analysis tracks driver behavior by capturing vehicle data and detecting smartphones. Machine learning algorithms, rules, and model building are examples of classification techniques. Neural network algorithms, mobile phone sensors, and vehicle-mounted cameras are some of the real-time and non-real-time analytic techniques employed. Surveys are used to obtain categorization in non-real time [7].

On-board sensors in cars with sophisticated safety measures and self-driving capabilities produce enormous volumes of data. In order to interpret and evaluate this data, machine learning algorithms have been devised, which have the benefit of minimal computing complexity and the capacity to extrapolate new characteristics. The application of machine learning (ML) in vehicular networks is now centered on centralized learning (CL), which involves training a potent learning algorithm—typically a neural network—on a sizable dataset [8].

Vehicular networks (VNs) are key components of future Intelligent Transportation Systems (ITS), which have the potential to improve efficiency and safety. Better assistance for linked cars is provided by 5G technology, and VN dynamics may be captured by machine learning (ML). Exciting research problems arise from federated learning (FL), which provides a networked machine learning platform while maintaining privacy [9].

4. Retail. Retail in FL is a tool that enables retailers to analyze customer behavior data from multiple locations, preserving customer privacy, thereby improving inventory management, understanding customer preferences, and overall shopping experience.

FL offers numerous benefits, but collecting sensitive data like face expressions poses privacy risks. Emotion recognition is crucial for social communication and understanding customer behavior. Model can combine these two models, allowing for decentralization and data safety.

Federated learning and emotion recognition offer advantages, but privacy concerns arise. A hybrid Split Federated Learning approach provides data security and decentralization, achieving high accuracy rates in the Emotion Classifier on various datasets [10].

5. Finance. The financial sector is utilizing AI to provide personalized services, but faces data privacy challenges due to data dispersion. Federated Learning (FL) offers a solution, but improper use can compromise stakeholder interests. Responsible and Effective Federated Learning (RE-FL) explores in finance, identifying six dimensions: accountability, controllability, fairness, privacy, security, and effectiveness [11].

Other way of usage of FL you can see in the article [12] presents an approach for unsecured loan risk assessment for decentralized finance (DeFi) lending networks that protects privacy. For unsecured loans, it employs federated learning techniques to precisely assess the likelihood of borrower default. The system offers a safe and effective solution for DeFi platforms as it is based on a trusted execution environment (TEE) with program-level isolation.

3. Results and discussion

We described the areas of application of FIs, wanted to present what risks exist and what measures can be taken, even with its encouraging outcomes, federated learning is susceptible to attacks from fraudulent users and criminals when combined with practical applications. Despite its advantages, before it is used in practice, thorough testing is required to guarantee its dependability.

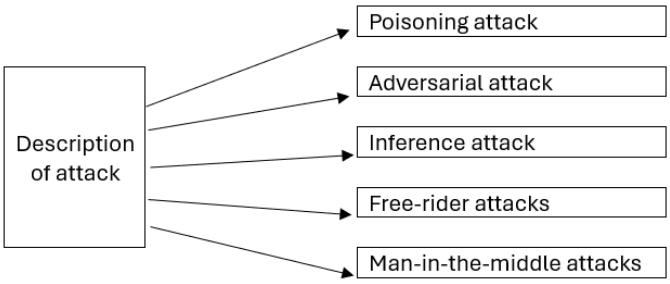

Figure 3. Description of attack

3.1. Attack description

1. Poisoning attack. A malicious individual can affect the prediction of a machine learning model by attacking the

training dataset using poisoning. A federated learning model's validity or dependability may be compromised by data poisoning, in which a hacker taints the training set with erroneous labels or biased data. By changing training data samples and/or model updates poorly, they aim to corrupt models, turning a good model into a bad one.

Defense from poisoning attack: prior to training the model, data security and reliability may be guaranteed, which helps prevent data and model poisoning assaults in federated learning. In order to safeguard data integrity throughout training, stability is essential. The behaviors of the poisoned party may be misinterpreted as typical user behavior, rendering identification unfeasible. Through data and parameter adjustments, the user may improve the model's attack.

2. Adversarial attack. Adversarial attacks modify input data to produce conclusions that aren't totally accurate. Confrontation training and data augmentation are two popular strategies used to counter these threats. Confrontation training fortifies the model's robustness by combining adversarial and real datasets, but it exposes it to trained adversarial samples. Data augmentation randomizes the original data in order to increase the model's capacity for generalization.

3. Inference attack. By enabling local data training, federated learning reduces the amount of indirect access to local knowledge. However, any privacy leaks jeopardize both safety and privacy. Differential privacy provides statistical defenses against adversaries' knowledge, whereas forward and reverse inference assaults take use of model secrecy. For safe storage, secret sharing systems disperse sensitive data, and homomorphic encryption is a well-liked and safe technique. Inference attacks can be thwarted by hybrid protection techniques including homomorphic encryption, differential privacy, and secret sharing protocols. These techniques guard against malicious usage of shared parameters and guarantee data security.

Determine illicit data about the FL process (participants, data, characteristics, etc.) by analyzing disclosed participant information. Then, utilize that knowledge to create an attack against the FL [13].

4. Free-rider attack. An opportunistic client that dissimilates fair participation in the FL training to obtain the global model parameters, without actually performing the local training.

An unscrupulous customer that dissimilates equitable participation in the FL training without carrying out the local training in order to get the global model parameters.

5. Man-in-the-middle attack. The transferred model updates may be intercepted by an external attacker, who may then replace them with malicious ones in order to steal, alter, or redirect the updates to a different location [14].

The security flaws of FL enabled systems are figure out with an emphasis on three possible adversaries: clients, aggregator servers and outsiders. Clients have complete control over local model updates, hyper-parameter adjustments, and the training process itself. Aggregator servers have the ability to inspect model changes, deduce personal data, and launch injection attacks. Theft of model updates and private data recovery is possible. By protecting communication lines from outside threats, security concerns are mitigated for the FL system. A secure aggregation FL approach can be complicated by outsiders and FL actors working together. On the other hand, attacks against data, algorithms, or federation can be targeted or untargeted in Federated Learning systems. Attacks fall into three categories: concentrating on algorithms, threatening to disrupt the global model, and misbehaving subtasks [14].

4. Conclusions

In conclusion, federated learning is employed in a number of sectors, including banking, retail, healthcare, agriculture and transport. FL may enhance personalized treatment planning, medical image analysis and patient representation. FL can enhance agricultural irrigation, fertilization, harvesting and animal health monitoring in agriculture. FL may evaluate consumer activity data in retail while protecting privacy. But, federated learning models can be compromised by various threats, including data poisoning, adversarial attacks, inference attacks, free-rider attacks, and man-in-the-middle attacks. Poisoning attacks involve malicious individuals attacking the training dataset, causing the model's validity or dependability to be compromised. To protect against these attacks, data security and reliability are guaranteed before training the model. Adversarial attacks modify input data, leading to inaccurate conclusions. Confrontation training and data augmentation are popular strategies to counter these threats. Inference attacks can be thwarted by hybrid protection techniques like homomorphic encryption, differential privacy, and secret sharing protocols. Free-rider attacks involve opportunistic clients dissimulating fair participation in training to obtain global model parameters. Man-in-the-middle attacks intercept transferred model updates, allowing external attackers to steal, alter, or redirect them.

References

[1] El-Sayed, T., Moustafa, A., Elrashidy, M. & El-Sayed, A. (2023). Optimizing Federated Learning Approach: Literature Survey and Open Points. *2023 3rd International Conference on Electronic Engineering (ICEEM), 1-5*

[2] Panigrahi, M., Bharti, S. & Sharma, A. (2023). Federated Learning for Beginners: Types, Simulation Environments, and Open Challenges. *2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications (IC2E3), 1-6*

[3] Reddy, G.P. & Pavan Kumar, Y.V. (2023). A Beginner's Guide to Federated Learning. *2023 Intelligent Methods, Systems, and Applications (IMSA), 557-562*

[4] Muhammad, B.L. & Kusharki, M.B. (2023). Federated Learning for Collaborative Crop Disease Monitoring in Wheat Production. *2023 2nd International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS), 1*, 1-5

[5] Le, D., Dao, M., Tran, A., Nguyen, T. & Le-Thi, H. (2023). Federated Learning in Smart Agriculture: An Overview. *2023 15th International Conference on Knowledge and Systems Engineering (KSE), 1-4*

[6] Vondikakis, I.V., Panagiotopoulos, I.E. & Dimitrakopoulos, G.J. (2023). An adaptive federated learning framework for intelligent road surface classification. *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), 4121-4126*

[7] Dikbıyık, D. & Alagöz, F. (2023). Driving Behavior Classification Using Smartphone Sensor Data. *2023 8th International Conference on Computer Science and Engineering (UBMK), 370-375*

[8] Elbir, A.M., Soner, B., Coleri, S., Gündüz, D. & Bennis, M. (2020). Federated Learning in Vehicular Networks. *2022 IEEE International Mediterranean Conference on Communications and Networking (MeditCom), 72-77*

[9] Tan, K., Bremner, D.J., Kernec, J.L. & Imran, M.A. (2020). Federated Machine Learning in Vehicular Networks: A summary

of Recent Applications. *2020 International Conference on UK-China Emerging Technologies (UCET), 1-4*
[10] Waref, D. & Salem, M.E. (2022). Split Federated Learning for Emotion Detection. *2022 4th Novel Intelligent and Leading Emerging Sciences Conference (NILES), 112-115*
[11] Shi, Y. Song, H., & Xu, J. (2023). Responsible and Effective Federated Learning in Financial Services: A Comprehensive Survey. *2023 62nd IEEE Conference on Decision and Control (CDC), 4229-4236*
[12] Mao, Q., Wan, S., Hu, D., Yan, J., Hu, J. & Yang, X. (2023). Leveraging Federated Learning for Unsecured Loan Risk Assessment on Decentralized Finance Lending Platforms. *2023 IEEE International Conference on Data Mining Workshops (ICDMW), 663-670*
[13] Tyagi, S., Rajput, I.S. & Pandey, R. (2023). Federated learning: Applications, Security hazards and Defense measures. *2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT), 477-482*
[14] Arbaoui, M., Brahmia, M. & Rahmoun, A. (2022). Towards secure and reliable aggregation for Federated Learning protocols in healthcare applications. *2022 Ninth International Conference on Software Defined Systems (SDS), 1-3*

Федерациялық машина оқытудағы қолданбалар

Г. Бектемысова, Г. Бакирова*, Г. Шайкемелов
Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан
**Корреспонденция үшін автор: g.bakirova@iitu.edu.kz*

Аңдатпа. Мақалада Федеративті оқыту (FL) әртүрлі салаларда, соның ішінде медицина, ауыл шаруашылығы, көлік, бөлшек сауда және қаржы салаларында қолданылатын терең оқыту әдісі екені анықталды. Ол құпиялылықты, деректерді иеленуді, локализацияланған модельді оқытуды, өткізу қабілеттілігінің тиімділігін, нақты уақыттағы оқытуды, масштабтауды және құрылғы ақауларына төзімділікті қамтамасыз етеді. Медицинада FL пациенттің түсінігін, дәріні әзірлеуді, медициналық кескінді талдауды, ауруды диагностикалауды және жеке емдеуді жоспарлауды жақсарта алады. Ауыл шаруашылығында FL суаруды, тыңайтқышты, егін жинауды және жануарлардың денсаулығын бақылауды жақсарта алады. Бөлшек саудада FL құпиялылықты сақтай отырып, тұтынушы әрекеті деректерін талдай алады. Біз түсінетініміздей, федеративті оқыту GPS, микрофондар және камералар сияқты сенсорларды пайдалана отырып, жергілікті деректер көздері бойынша модельдік оқытуды бөліседі. Дегенмен, оқу үлгілері әртүрлі қауіптермен, соның ішінде деректерді улану шабуылдарымен бұзылуы мүмкін.

Негізгі сөздер: *федеративті оқыту, уытты шабуыл, орталықтандырылмаған оқыту, орталықтандырылған оқыту, кросс-сило.*

Приложения в федеративном машинном обучении

Г. Бектемысова, Г. Бакирова*, Г. Шайкемелов
Международный университет информационных технологий, Алматы, Казахстан
**Автор для корреспонденции: g.bakirova@iitu.edu.kz*

Аннотация. В статье определено, что федеративное обучение (FL) - это метод глубокого обучения, используемый в различных отраслях, включая медицину, сельское хозяйство, транспортные средства, розничную торговлю и финансы. Он обеспечивает конфиденциальность, владение данными, локализованное обучение моделей, эффективность использования полосы пропускания, обучение в реальном времени, масштабируемость и устойчивость к сбоям устройств. В медицине FL может улучшить представление о пациенте, разработку лекарств, анализ медицинских изображений, диагностику заболеваний и индивидуальное планирование лечения. В сельском хозяйстве FL может улучшить орошение, удобрение, сбор урожая и мониторинг здоровья животных. В розничной торговле FL может анализировать данные о поведении покупателей, сохраняя конфиденциальность. Как мы понимаем, федеративное обучение распределяет обучение моделей между локальными источниками данных, используя такие датчики, как GPS, микрофоны и камеры. Однако модели обучения могут быть взломаны различными угрозами, включая атаки с отравлением данных.

Ключевые слова: *федеративное обучение, отравляющая атака, децентрализованное обучение, централизованное обучение, кросс-сило.*

Received: 24 May 2023
Accepted: 15 September 2023
Available online: 30 September 2023