TAQSIMLANGAN MA'LUMOTLAR ASOSIDA MAXFIYLIKNI SAQLOVCHI FEDERATIV MASHINAVIY O'QITISH ALGORITMLARINI ISHLAB CHIQISH

Matkarimov Otabek Oybekovich
Toshkent shahridagi Belarus–O'zbekiston qo'shma tarmoqlararo amaliy texnik kvalifikatsiyalar instituti
E-mail: matkarimov.otabek95@gmail.com
https://doi.org/10.5281/zenodo.21906587

Annotatsiya. Mazkur maqolada taqsimlangan ma'lumotlar asosida maxfiylikni saqlovchi federativ mashinaviy o'qitish algoritmlarini ishlab chiqish masalalari tadqiq etilgan. Federativ mashinaviy o'qitish texnologiyasining asosiy afzalligi ma'lumotlarni markaziy serverga uzatmasdan, ularni mahalliy qurilmalarda qayta ishlash orqali maxfiylik va ma'lumotlar xavfsizligini ta'minlash imkoniyatidan iborat. Tadqiqotda taqsimlangan ma'lumotlar muhitida model parametrlarini almashish, lokal modellarni birlashtirish, kommunikatsiya samaradorligini oshirish hamda maxfiylikni ta'minlash mexanizmlari tahlil qilingan.
Kalit so'zlar: federativ mashinaviy o'qitish, taqsimlangan ma'lumotlar, ma'lumotlar maxfiyligi, axborot xavfsizligi, lokal model, global model, federativ algoritm, mashinaviy o'qitish, differensial maxfiylik, xavfsiz agregatsiya.

Аннотация. В статье исследуются вопросы разработки алгоритмов федеративного машинного обучения с сохранением конфиденциальности на основе распределённых данных. Основным преимуществом технологии федеративного машинного обучения является возможность обработки данных непосредственно на локальных устройствах без их передачи на централизованный сервер, что способствует обеспечению конфиденциальности и безопасности информации. В исследовании проанализированы механизмы обмена параметрами моделей, агрегации локальных моделей, повышения эффективности коммуникации и защиты конфиденциальных данных в распределённой среде. Особое внимание уделено проблемам децентрализованного характера данных, различий вычислительных ресурсов устройств и коммуникационных затрат в процессе федеративного обучения.
Ключевые слова: федеративное машинное обучение, распределённые данные, конфиденциальность данных, информационная безопасность, локальная модель, глобальная модель, федеративный алгоритм, машинное обучение, дифференциальная конфиденциальность, безопасная агрегация.

Abstract. This article investigates the development of privacy-preserving federated machine learning algorithms based on distributed data. The main advantage of federated machine learning is its ability to process data locally without transferring raw data to a centralized server, thereby improving data privacy and security. The study analyzes mechanisms for model parameter exchange, local model aggregation, communication efficiency, and privacy protection in distributed environments. Particular attention is paid to the challenges associated with decentralized data, differences in computational resources among participating devices, and communication costs during the federated learning process. Algorithmic approaches aimed at improving training efficiency and reducing the risks of confidential information disclosure are also considered.

Keywords: federated machine learning, distributed data, data privacy, information security, local model, global model, federated algorithm, machine learning, differential privacy, secure aggregation.

Kirish. Bugungi kunda raqamli texnologiyalarning jadal rivojlanishi va sun'iy intellektning turli sohalarga keng joriy etilishi katta hajmdagi ma'lumotlardan samarali foydalanish zaruratini yuzaga keltirmoqda. Mashinaviy o'qitish algoritmlari tibbiyot, moliya, ta'lim, sanoat, transport, telekommunikatsiya va boshqa ko'plab yo'nalishlarda murakkab masalalarni avtomatlashtirish hamda prognozlash imkonini bermoqda. Shu bilan birga, mashinaviy o'qitish modellarini o'qitish uchun zarur bo'lgan ma'lumotlarning markazlashtirilgan holda yig'ilishi shaxsiy va maxfiy axborotning himoyasi bilan bog'liq muammolarni kuchaytirmoqda. Ayniqsa, shaxsiy ma'lumotlar, moliyaviy axborot, tibbiy ma'lumotlar va korporativ ma'lumotlarning yagona markaziy serverda saqlanishi ularning ruxsatsiz foydalanilishi yoki kiberhujumlarga uchrash xavfini oshiradi.

An'anaviy mashinaviy o'qitish yondashuvida turli manbalardan olingan ma'lumotlar markaziy server yoki yagona ma'lumotlar bazasiga to'planadi va ushbu ma'lumotlar asosida model o'qitiladi. Mazkur yondashuv hisoblash jarayonini nisbatan soddalashtirsa-da, taqsimlangan ma'lumotlarning maxfiyligi va xavfsizligini ta'minlashda qator cheklovlarga ega. Ma'lumotlarni markazlashtirish nafaqat maxfiylik bilan bog'liq xavflarni, balki katta hajmdagi ma'lumotlarni uzatish, saqlash va qayta ishlash bilan bog'liq texnik va iqtisodiy xarajatlarni ham oshiradi. Shu sababli ma'lumotlarni markazlashtirmagan holda mashinaviy o'qitish modellarini shakllantirishga imkon beruvchi yangi texnologik yondashuvlarni ishlab chiqish dolzarb ilmiy-amaliy vazifalardan biri hisoblanadi.

Adabiyotlar tahlili. Mazkur yo'nalishdagi asosiy ilmiy tadqiqotlar quyidagicha tahlil qilindi. McMahan va hammualliflar tomonidan amalga oshirilgan tadqiqotda federativ mashinaviy o'qitishning asosiy konsepsiyasi ishlab chiqilib, ma'lumotlarni mobil qurilmalarda saqlagan holda umumiy mashinaviy o'qitish modelini shakllantirish yondashuvi taklif etilgan [1]. Konečný va hammualliflar federativ o'qitish tizimlarida kommunikatsiya samaradorligini oshirish masalasiga alohida e'tibor qaratgan [2]. Tadqiqotda lokal qurilmalardan markaziy serverga uzatiladigan model yangilanishlari hajmini kamaytirishga qaratilgan structured updates va sketched updates kabi yondashuvlar taklif etilgan. Mualliflar ma'lumotlar qurilmalarda taqsimlangan holatda saqlanishi sababli kommunikatsiya kanallarining samaradorligi butun tizim unumdorligiga bevosita ta'sir qilishini asoslaganlar. Bonawitz va hammualliflar tomonidan ishlab chiqilgan tadqiqotda federativ mashinaviy o'qitish jarayonida Secure Aggregation, ya'ni xavfsiz agregatsiya mexanizmi taklif etilgan [3]. Geyer, Klein va Nabi tomonidan olib borilgan tadqiqotda federativ mashinaviy o'qitishda differensial maxfiylik mexanizmlaridan foydalanish masalasi o'rganilgan [4]. Li va hammualliflar tomonidan ishlab chiqilgan FedProx algoritmi federativ tarmoqlardagi tizimli va statistik geterogenlik muammolarini hal etishga qaratilgan [5]. Tadqiqotda federativ tizimdagi qurilmalar hisoblash quvvati, aloqa tezligi va mavjud ma'lumotlar hajmi bo'yicha bir-biridan farq qilishi, shuningdek, ma'lumotlar bir xil taqsimlanmasligi mumkinligi qayd etilgan. Zhang va hammualliflarning tadqiqotida federativ mashinaviy o'qitishning xavfsizlik muammolari tizimli ravishda tahlil qilingan [6]. Mualliflar federativ o'qitish xom ma'lumotlarni markaziy serverga yig'masligi sababli maxfiylikni ta'minlash imkoniyatiga ega ekanligini qayd etish bilan birga, taqsimlangan arxitektura yangi turdagi xavfsizlik tahdidlarini ham yuzaga keltirishini ko'rsatganlar. Federativ mashinaviy o'qitish bo'yicha olib borilgan tadqiqotlar umumlashtirilganda, Kairouz va hammualliflar federativ o'qitishning mavjud holati, asosiy muammolari va istiqboldagi yo'nalishlarini keng qamrovda tahlil qilganlar [7].

Materiallar va usullar. Tadqiqotda taqsimlangan ma’lumotlar muhitida maxfiylikni saqlagan holda mashinaviy o‘qitish modelini shakllantirish uchun federativ mashinaviy o‘qitish yondashuvidan foydalanildi. Tadqiqot jarayonida lokal qurilmalarda modelni o‘qitish, model parametrlarini markaziy serverga uzatish va ularni xavfsiz agregatsiya qilish bosqichlari amalga oshirildi. Asosiy algoritmik yondashuv sifatida FedAvg tamoyili, maxfiylikni kuchaytirish uchun esa differensial maxfiylik va xavfsiz agregatsiya mexanizmlaridan foydalanish ko‘zda tutildi.

1-jadval. Tadqiqotda foydalanilgan asosiy komponentlar

Komponent	Vazifasi
Taqsimlangan ma’lumotlar	Lokal qurilmalarda saqlanadi
Lokal model	Har bir ishtirokchida o‘qitiladi
FedAvg	Lokal modellarni birlashtiradi
Differensial maxfiylik	Maxfiy ma’lumotlarni himoyalaydi
Secure Aggregation	Model yangilanishlarini xavfsiz agregatsiya qiladi

Tadqiqotda federativ o‘qitish samaradorligini baholash uchun model aniqligi, yo‘qotish funksiyasi, kommunikatsiya xarajati va maxfiylik darajasi asosiy mezonlar sifatida tanlandi.

2-jadval. Tadqiqotda qo‘llanilgan baholash mezonlari

Mezon	Baholash maqsadi
Model aniqligi	Modelning to‘g‘ri ishlash darajasini aniqlash
Loss qiymati	O‘qitish samaradorligini baholash
Kommunikatsiya xarajati	Ma’lumot almashinuvi samaradorligini aniqlash
Maxfiylik darajasi	Foydalanuvchi ma’lumotlarini himoyalashni baholash

Tadqiqot natijalarini baholashda an’anaviy markazlashtirilgan mashinaviy o‘qitish va federativ mashinaviy o‘qitish yondashuvlari o‘rtasidagi farqlar solishtirildi. Shu orqali taklif etilayotgan maxfiylikni saqlovchi algoritmning aniqlik, xavfsizlik va kommunikatsiya samaradorligi bo‘yicha afzalliklari aniqlanadi.

Tadqiqot muhokamasi. Tadqiqot davomida taqsimlangan ma’lumotlar sharoitida federativ mashinaviy o‘qitishning samaradorligi, maxfiylikni ta’minlash imkoniyatlari va modelni o‘qitish jarayonining barqarorligi tahlil qilindi. Lokal ma’lumotlarni markaziy serverga uzatmasdan modelni o‘qitish an’anaviy markazlashtirilgan yondashuvga nisbatan ma’lumotlar xavfsizligini oshirish imkonini beradi. Bunda har bir ishtirokchi o‘z ma’lumotlari asosida lokal modelni shakllantiradi va faqat model yangilanishlarini uzatadi.

Tadqiqotda FedAvg algoritmining lokal modellarni yagona global modelga birlashtirish imkoniyati asosiy mexanizm sifatida qaraldi. Biroq ma’lumotlarning turli ishtirokchilar o‘rtasida bir xil taqsimlanmaganligi model aniqligi va konvergentsiya tezligiga ta’sir qilishi mumkin. Shu sababli federativ o‘qitish jarayonida lokal ma’lumotlarning hajmi va taqsimlanish xususiyatlarini hisobga olish muhim hisoblanadi.

Tahlil natijalari shuni ko‘rsatadiki, federativ mashinaviy o‘qitishda maxfiylikni oshirish bilan model aniqligi o‘rtasida muayyan muvozanat mavjud. Masalan, differensial maxfiylik uchun qo‘shiladigan shovqin miqdorining ortishi maxfiylikni kuchaytirishi mumkin, biroq haddan tashqari shovqin modelning o‘rganish qobiliyatini pasaytiradi. Shu sababli maxfiylik parametrlari modelning aniqligi va konvergentsiyasiga ta’sirini hisobga olgan holda tanlanishi lozim.

Xulosa. Tadqiqot natijalari taqsimlangan ma’lumotlar asosida maxfiylikni saqlovchi federativ mashinaviy o‘qitish algoritmlaridan foydalanish zamonaviy axborot tizimlarida ma’lumotlar xavfsizligini ta’minlashning istiqbolli yo‘nalishlaridan biri ekanligini ko‘rsatdi. Federativ yondashuvda ma’lumotlarning lokal qurilmalarda saqlanishi va markaziy serverga faqat model yangilanishlarining uzatilishi maxfiy ma’lumotlarning bevosita tarqalish xavfini kamaytiradi.

Tadqiqot davomida FedAvg algoritmi, differensial maxfiylik va xavfsiz agregatsiya mexanizmlarining o‘zaro uyg‘unligi tahlil qilindi. Ushbu mexanizmlardan kompleks foydalanish modelni markazlashtirilmagan sharoitda o‘qitish, foydalanuvchi ma’lumotlarini himoyalash hamda model yangilanishlarining xavfsizligini oshirish imkonini beradi. Shu bilan birga, maxfiylik darajasini oshirish jarayonida model aniqligi, konvergentsiya tezligi va kommunikatsiya xarajatlari o‘rtasidagi muvozanatni ta’minlash muhimligi aniqlandi.

Foydalanilgan adabiyotlar ro’yxati

1. McMahan B., Moore E., Ramage D., Hampson S., Agüera y Arcas B. Communication-Efficient Learning of Deep Networks from Decentralized Data // Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS). 2017. Vol. 54. P. 1273–1282.
2. Konečný J., McMahan H.B., Yu F.X., Richtárik P., Suresh A.T., Bacon D. Federated Learning: Strategies for Improving Communication Efficiency. arXiv preprint arXiv:1610.05492, 2016.
3. Bonawitz K., Ivanov V., Kreuter B., Marcedone A., McMahan H.B., Patel S., Ramage D., Segal A., Seth K. Practical Secure Aggregation for Privacy-Preserving Machine Learning // Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. 2017. P. 1175–1191.
4. Geyer R.C., Klein T., Nabi M. Differentially Private Federated Learning: A Client Level Perspective. arXiv preprint arXiv:1712.07557, 2017.
5. Li T., Sahu A.K., Zaheer M., Sanjabi M., Talwalkar A., Smith V. Federated Optimization in Heterogeneous Networks // Proceedings of Machine Learning and Systems. 2020. Vol. 2. P. 429–450.
6. Zhang K., Song X., Zhang C., Yu S. Challenges and Future Directions of Secure Federated Learning: A Survey // Frontiers of Computer Science. 2022. Vol. 16. Article 165817.
7. Kairouz P., McMahan H.B., Avent B., Bellet A., Bennis M., Bhagoji A.N., et al. Advances and Open Problems in Federated Learning // Foundations and Trends in Machine Learning. 2021. Vol. 14, No. 1–2. P. 1–210.