



МОДЕЛЬ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ В ІОТ

М. Климаш, О. Гордійчук-Бублівська, Б. Коваль

Національний університет “Львівська політехніка”, вул. С. Бандери, 12, Львів, 79013, Україна

Відповідальний за рукопис: О. Гордійчук-Бублівська (e-mail: olena.v.hordiichuk-bublivska@lpnu.ua).

(Подано 24 грудня 2021)

В роботі подано огляд методів інтелектуальної обробки даних у системах промислового Інтернету речей. Наведено порівняння методів аналізу великих даних у промислових системах зі значним навантаженням. Запропоновано використовувати для опрацювання даних методи розподіленого машинного навчання. Розроблено програмну модель для аналізу даних різних обсягів. Проаналізовано основні підходи до організації машинного навчання: федеративне і нерозподілене навчання. Експериментально доведено ефективність використання федеративного машинного навчання, оскільки воно забезпечує вищу точність оброблення даних, навіть у разі збільшення їх обсягів. Визначено, що нерозподілене машинне навчання працює швидше, отже, може використовуватися в системах, пріоритетом для яких є менший час обробки даних. Такий підхід відкриває можливості створення адаптивної моделі системи промислового Інтернету речей, що здатна самонавчатися та коригувати власну інфраструктуру залежно від зміни параметрів.

Ключові слова: *ІоТ; Big Data; федеративне машинне навчання.*

УДК 621.126

1. Вступ

У сучасних промислових системах постійно виникають нові виклики, оскільки через використання технології промислового Інтернету речей (Industrial Internet of Things, ІоТ) потрібно постійно збирати та опрацьовувати дані із величезної кількості пристроїв і датчиків. Для ефективного опрацювання великих масивів інформації застосовують методи машинного і глибокого навчання, а також штучний інтелект [1].

Для архітектури розподілених промислових систем складно адаптувати алгоритми машинного навчання. Оскільки для проведення навчання всі дані треба зібрати з пристроїв, надіслати центральному серверу для оброблення, а потім результати повернути назад, такий підхід спричиняє значне навантаження на канали передавання даних та ставить під загрозу цілісність і конфіденційність інформації.

Для вирішення проблем із опрацюванням великих обсягів даних у системах промислового Інтернету речей використовують методи розподіленого машинного навчання. Всі пристрої в такому випадку навчаються на власних локальних наборах даних та надсилають центральному пристрою тільки їх результати [2].

2. Принципи функціонування систем промислового Інтернету речей

Щоб забезпечити організовану роботу на підприємстві, використовують найсучасніші методи і засоби оброблення даних. Необхідно постійно здійснювати моніторинг стану пристроїв та компонент мережі для того, щоб завчасно визначити можливі збої та несправності.

Інтернет речей – це сукупність пристроїв, які можуть самостійно здійснювати комунікацію між собою, збирати та аналізувати дані, надсилати їх для оброблення на сервер. Така архітектура системи дає змогу обійтися без втручання людини в процес її функціонування. Натомість відносна автономність пристроїв означає, що:

- пристрої самі збирають дані, тобто оснащені давачами, сенсорами;
- всі дані потім необхідно опрацювати;
- оскільки обсяг інформації з усіх пристроїв дуже великий, його складно опрацювати на одному пристрої;
- щоб опрацювати дані на кінцевих пристроях Інтернету речей, потрібно використовувати спеціальні розподілені технології.

Отже, оброблення та зберігання великих даних є важливою проблемою великомасштабних інфокомунікаційних систем. Для ефективнішого оброблення інформації використовують розподілені системи. В них дані розподіляються між декількома обчислювальними пристроями, що зменшує навантаження на них та прискорює їх роботу.

Якщо Інтернет речей охоплює здебільшого побутові пристрої, наприклад, холодильники, годинники, навіть цілі будинки, то промисловий Інтернет речей використовується для покращення роботи промислових об'єктів. На великих підприємствах часто працює безліч пристроїв, різних за характеристиками, властивостями та функціоналом. Вони розташовані на великій території, тому їх важко постійно моніторити. До того ж є підприємства різного спрямування. Якщо, для прикладу, на заводі, який виробляє харчову продукцію, виникне збій у роботі одного з механізмів, це може призвести до зміни технології виробництва, погіршення якості продукції. Потім доволі важко простежити виникнення певної помилки, виправити її та унеможливити повторення.

На підприємствах, що працюють з небезпечними речовинами або матеріалами, збої у роботі можуть спричинити аварійну ситуацію, завдати значних економічних, екологічних та соціальних збитків. Тому необхідно постійно аналізувати стан системи. Спеціальні давачі, встановлені на пристроях, дають змогу вимірювати їх параметри, зокрема температуру, точність роботи, спеціальні характеристики.

Оскільки всі дані передаються для подальшого оброблення на центральний сервер, за значного навантаження можлива ситуація, коли не встигають обробити всі інформацію, вчасно відреагувати та врегулювати ситуацію. Також часто пристрої територіально розподілені на значній відстані. Це також унеможлиблює використання жорсткого централізованого управління системою промислового Інтернету речей [3].

3. Аналіз особливостей федеративного машинного навчання

Розподілені технології оброблення даних дають змогу істотно пришвидшити обчислення великих даних та зменшити навантаження на окремі пристрої. В розподілених системах всі пристрої опрацьовують частини одного великого завдання, а потім надсилають центральному пристрою дані для об'єднання та отримання остаточного результату. Для ефективної комунікації та синхронізації використовують спеціальні протоколи передавання даних.

Розподілені обчислення дають можливість значно гнучкіше організувати обчислення в інформаційній системі. Зокрема, якщо на вхід розподіленої системи надійде великий обсяг даних, що потребують обробки, для обчислень виділяється певна кількість пристроїв. В той час інші ресурси системи можуть використовуватися для вирішення інших завдань. Також часто в інфокомунікаційних системах їх елементи розташовані на дуже великій відстані один від одного. У такому разі для організації роботи та управління такими системами можуть використовуватися хмарні сервери або бази даних. В автоматичному режимі конфігурація системи обчислень може змінюватися залежно від поставлених завдань та вимог до функціонування.

Перед початком роботи блок вхідних даних або задача надходить на керуючий пристрій. Там здійснюється розподіл даних на менші частини відповідно до обчислювальних пристроїв чи процесів у системі. Очевидною перевагою розподілених обчислень є підвищення надійності та відмовостійкості. У системі, де всі дані обробляє тільки один пристрій, вихід його з ладу призводить до глобального збою у роботі та повної втрати даних. Проте застосування архітектури розподілених обчислень в промисловому Інтернеті речей супроводжується певними складностями.

Розподілені системи складаються із самостійних обчислювальних пристроїв, здатних обробляти задачі. Тому, використовуючи давачі як елементи розподіленої системи, складно забезпечити високу ефективність. Отже, необхідні ефективніші підходи до обчислення даних.

Розподілене машинне навчання. Для того, щоб знаходити певні шаблони і закономірності в великому обсязі даних, передбачати майбутні імовірні події, використовують машинне навчання та штучний інтелект. Спеціальні алгоритми класифікують та структурують дані, відкидають надлишковість, що зменшує обсяг інформації. Важлива також можливість самонавчання на основі початкових даних чи тих, які набулі під час навчання.

Процес машинного навчання складається із кількох етапів. Спочатку встановлюються певні стандартні значення, які повинні бути отримані під час обчислень. Потім задають певні характеристики і параметри та визначають результати роботи. Порівнюючи результат обчислень з очікуваними даними, можна виявити закономірності між параметрами та отриманими значеннями. В наступному раунді параметри обчислень коригують до досягнення необхідного результату (або близького до нього з певною точністю) [4].

Розподілене обчислення може відбуватися за допомогою вертикальної та горизонтальної фрагментації. За горизонтальної фрагментації кожна частина даних, що потребують обчислень, міститься на окремому вузлі в розподіленій системі. У разі застосування вертикальної фрагментації множина всіх даних розподіляється на підмножини, що опрацьовуються декількома пристроями.

Для того, щоб здійснювати машинне навчання в розподілених системах, необхідно збирати дані з різних пристроїв у мережі. Потім на центральному пристрої виконується алгоритм машинного навчання і його результати розсилають назад усім пристроям. Такий підхід характеризується певними недоліками:

- збільшується навантаження на канали передавання даних;
- оскільки передаються приватні дані з кінцевих пристроїв, є ризик їх перехоплення.

Також для виконання алгоритмів машинного навчання у розподілених системах необхідно забезпечити відповідність програмного та апаратного забезпечення поставленим завданням, використовувати спеціальні протоколи комунікації та синхронізації, додатково шифрувати дані.

Щоб уникнути цих недоліків, запроваджено використання федеративного машинного навчання. За такого підходу всі дані опрацьовуються на локальних пристроях, а потім тільки результати навчання надсилають центральному пристрою для оновлення глобальної моделі. Відтак центральний пристрій розсилає усім пристроям у мережі дані про оновлену модель.

Федеративне машинне навчання. Федеративне навчання є підвидом розподіленого машинного навчання. Навчання відбувається безпосередньо на пристроях, наприклад, смартфонах чи елементах системи “Розумний будинок” або “Розумне місто”. Всі дані з пристроїв не потрібно пересилати для опрацювання, що підвищує надійність системи та інформації в ній.

Під час федеративного машинного навчання дані на окремих пристроях використовуються для обчислення локальних навчальних моделей. Після того, як процес навчання успішно завершився, тільки його результати, що можуть впливати на навчання глобальної моделі, пересилаються центральному пристрою у системі. Коли отримано дані зі всіх локальних навчальних моделей, здійснюється модифікація глобальної моделі згідно із отриманими результатами. Результати після завершення обчислення глобальної моделі надсилаються усім пристроям у мережі.

Отже, порівняно зі звичайним розподіленим машинним навчанням, за якого дані пересилаються для обчислень на окремий пристрій, федеративне навчання пропонує новий спосіб

локального навчання. Такий підхід зменшує ризик того, що під час передавання інформація спотвориться або загубиться. Оскільки у роботі моделі федеративного навчання беруть участь малопотужні пристрої, зменшується енергоспоживання у мережі, а також тривалість затримок під час пересилання даних через їх локалізацію в одному місці [5].

Важливою перевагою федеративного машинного навчання є можливість адаптувати модель відповідно до потреб та особливостей конкретного користувача. Оскільки в обчисленнях беруть участь не потужні сервери, а, здебільшого, телефони користувачів, розумні гаджети, то вони опрацьовують персональні дані, що дає змогу вирішувати конкретні вузькопрофільні проблеми.

У разі використання певного програмного забезпечення, наприклад такого, що уможлиблює контроль за фізичною активністю, станом здоров'я, надає персональні рекомендації під час прослуховування музики тощо, важливо якнайшвидше виявити закономірності в поведінці користувача та послуги, необхідні йому. Це дає змогу швидко й ефективно обробляти дані. Використовуючи федеративне машинне навчання, можна навчатися на локальних даних та не пересилати іншим пристроям персональну інформацію. Проте, коли глобальна модель машинного навчання оновиться, використавши локальні результати, її розсилають назад усім пристроям. Тому на основі отриманих даних персональні пристрої користувача можуть виконувати точніші обчислення.

Приклад реалізації архітектури промислового Інтернету речей із використанням технологій федеративного навчання подано на рис. 1.

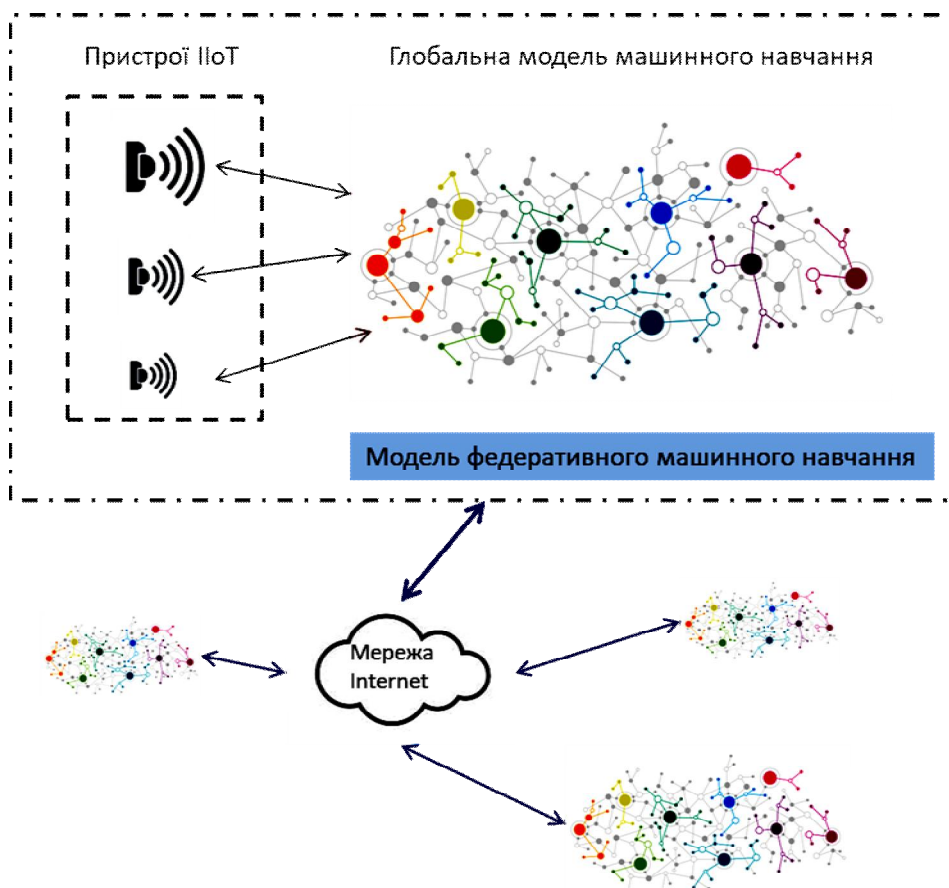


Рис. 1. Федеративне навчання у системі ІІоТ

Наприклад, коли користувач інформаційної системи на власному пристрої користується програмою з інтелектуальною клавіатурою введення даних, відбувається навчання на основі введеної інформації. Потім, вводячи перші букви, користувач отримує перелік слів, які починаються на ці літери та, найімовірніше, зацікавлять його, оскільки доволі подібні до введених

раніше. Проте процес машинного навчання на даних тільки одного користувача доволі довгий, до того ж не завжди трапляються усі варіанти подій, що зменшує точність роботи алгоритму рекомендацій.

Використовуючи федеративне машинне навчання, не потрібно передавати іншим пристроям всі повідомлення із пристрою для їх аналізу, оскільки це ставить під загрозу їх конфіденційність. На локальному вузлі відбувається обчислення алгоритму машинного навчання та тільки його результатами він обмінюється з усіма іншими учасниками в мережі. Також пристрій отримує результати навчання інших користувачів та на їх основі має більше даних для власного навчання [6].

За допомогою федеративного навчання можна також здійснювати управління конфігурацією системи. Кожна локальна мережа, опрацювавши та проаналізувавши результати, надсилає їх у систему управління конфігурацією. Там приймається рішення про вимкнення можливих аварійних вузлів, додавання нових пристроїв, зміну топології мережі тощо (рис. 2).

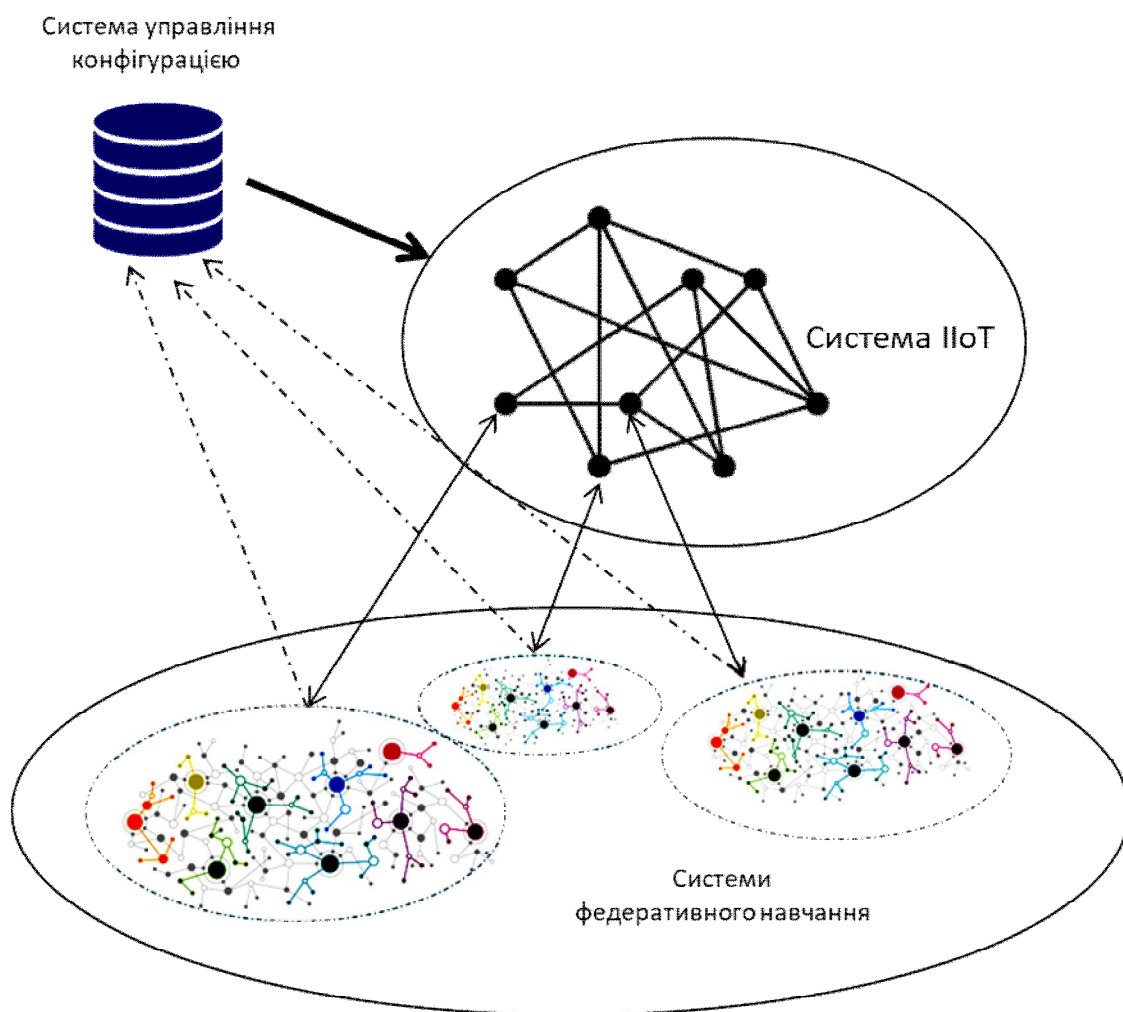


Рис. 2. Система управління конфігурацією системи ІіоТ із використанням федеративного машинного навчання

Недоліками федеративного навчання є його особливості – використання персональних пристроїв, які можуть вийти з ладу. Також переважно для зв'язку в мережі застосовують безпроводні технології, наприклад Wi-Fi. Якщо з'єднання із мережею буде розірване, пристрій не зможе обмінюватися даними з глобальною моделлю, що впливає на точність проведення машинного навчання.

4. Дослідження ефективності оброблення великих даних за допомогою федеративного навчання

Проте переваги федеративного навчання все ж дають змогу використовувати його у великомасштабних інфокомунікаційних системах. Це:

- зменшення трафіку в мережі;
- конфіденційність персональних даних;
- гнучкість організації обчислень;
- можливість швидко змінювати конфігурацію системи;
- підвищення ефективності обчислень.

Для систем Інтернету речей та промислового Інтернету речей ці переваги критично важливі. Оскільки обсяг даних, що потребують оброблення, є значним, доцільніше зберігати їх на власних пристроях. Також, оскільки в Інтернеті речей пристрої володіють важливими персональними даними, саме федеративне навчання здатне забезпечити їх конфіденційність під час оброблення [7]. Модель використання технологій інтелектуального аналізу даних в інформаційно-комунікаційній промисловій системі наведено на рис. 3.

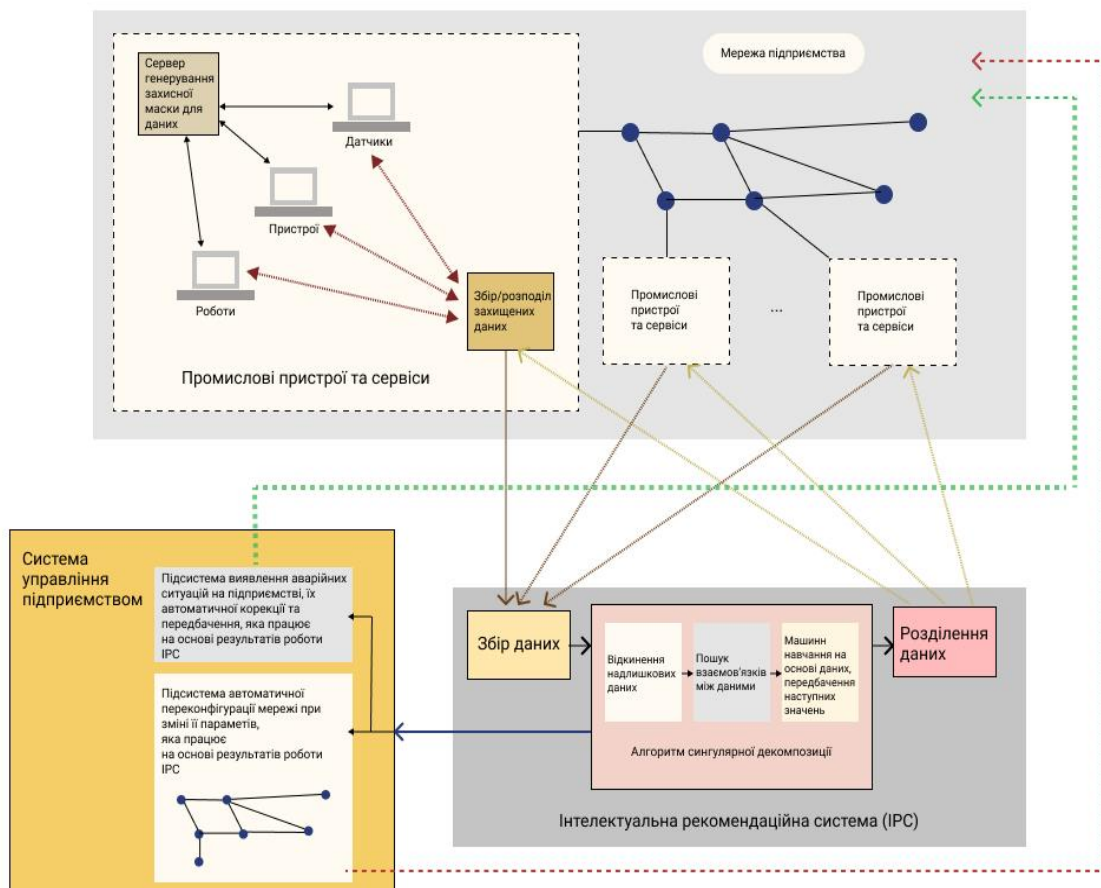


Рис. 3. Модель використання технологій інтелектуального аналізу даних в інформаційно-комунікаційній промисловій системі

Як бачимо на рис. 3, використовуючи в промисловій системі методи розподіленого машинного навчання, інтелектуального аналізу даних, зокрема рекомендаційних систем, можна забезпечити збирання та оброблення інформації з різних пристроїв. На основі отриманих даних

визначають характеристики стану кожного пристрою та мережі загалом. У разі необхідності здійснюється корегування функціонала компонентів системи, а також зміна її конфігурації.

У роботі створено програмну модель оброблення даних мовою Python. Визначено тривалості оброблення різних обсягів даних із використанням нерозподіленого і федеративного машинного навчання. Експеримент виконано на основі алгоритму визначення і підрахунку спеціальних символів у тексті (рис. 4).

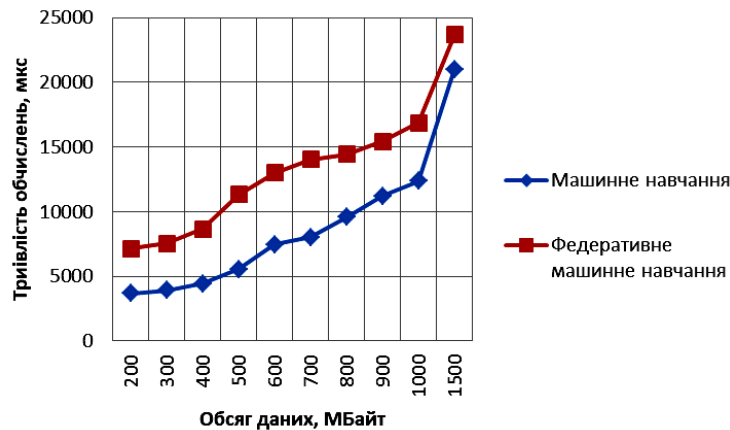


Рис. 4. Порівняння тривалості обчислень за допомогою алгоритму звичайного і федеративного машинного навчання

Як бачимо з рис. 4, для алгоритму федеративного машинного навчання тривалість обчислень є більшою, оскільки складність обчислень зростає. Натомість федеративне навчання дає змогу підвищити точність роботи алгоритму машинного навчання під час оброблення даних, тобто зменшити коефіцієнт помилкового прийняття рішень, що забезпечує перспективу подальшого його вивчення і використання (рис. 5).

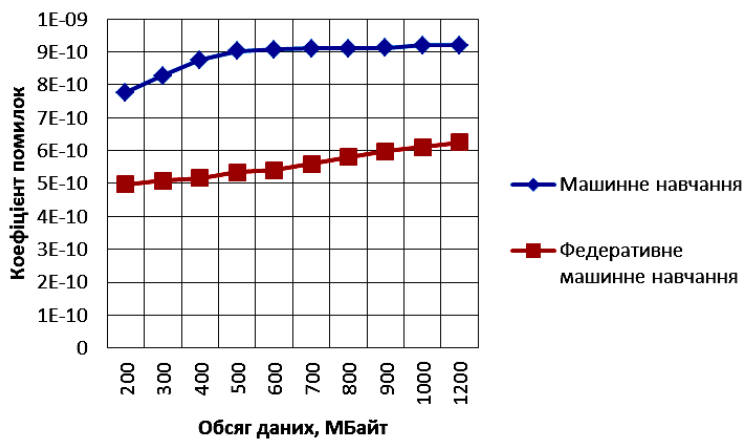


Рис. 5. Порівняння точності обчислень за допомогою алгоритму звичайного і федеративного машинного навчання

Результати досліджень показують переваги використання методів розподіленого машинного навчання у великомасштабних системах промислового Інтернету речей, якщо необхідно забезпечити високу точність та надійність обчислень. Запропонований метод дає змогу створити інтелектуальну модель управління промисловими системами, що на основі зібраних з різних пристроїв та підсистем даних самонавчається для автоматичного усунення неполадок та зміни конфігурації.

Висновки

Системи промислового Інтернету речей нині використовуються для автоматизації виробництва, визначення і передбачення можливих аварійних ситуацій, що дає змогу істотно підвищити продуктивність роботи промислових об'єктів. Для оброблення і аналізу великих даних у великомасштабних системах доцільно використовувати методи машинного навчання. Через високе навантаження на системи промислового Інтернету речей використовується федеративне машинне навчання, в якому, на противагу традиційному, навчання відбувається не на центральному вузлі, а на локальних пристроях. Такий підхід підвищує конфіденційність оброблення даних, проте потребує додаткових витрат на проектування та експлуатацію.

Досліджено використання федеративного машинного навчання для проведення аналізу даних в системах промислового Інтернету речей. Визначено, що у разі оброблення великих даних такий підхід забезпечує вищу точність виконання алгоритмів, навіть у випадку збільшення обсягу інформації. Підвищується також конфіденційність даних порівняно із використанням традиційного машинного навчання, оскільки вся інформація опрацьовується на локальних пристроях. Отримані переваги забезпечують перспективу застосування методів федеративного машинного навчання у великомасштабних промислових системах. На їх основі можна створити інтелектуальну модель аналізу даних, що здатна не тільки швидко та ефективно виконувати опрацювання даних, але й самонавчатися та автоматизувати функції управління і конфігурації.

Список використаних джерел

- [1] Hwaiyu, G. (2017), "The Industrial Internet of Things (IIoT)" *Internet of Things and Data Analytics Handbook*, Wiley, pp. 41–81. DOI: 10.1002/9781119173601.ch3.
- [2] Abuhasel, K. and Khan, M. (2020), "A Secure Industrial Internet of Things (IIoT) Framework for Resource Management in Smart Manufacturing", in *IEEE Access*, Vol. 8, pp. 117354–117364.
- [3] Paul, D., Chakraborty, T., Datta, S. and Paul, D. (2018), "IoT and Machine Learning Based Prediction of Smart Building Indoor Temperature", *4th International Conference on Computer and Information Sciences (ICCOINS)*, pp. 1–6. DOI: 10.1109/ICCOINS.2018.8510597.
- [4] Khan, L., Saad, W., Han, Z., Hossain, E. and Hong, C. (2021), "Federated Learning for Internet of Things: Recent Advances, Taxonomy, and Open Challenges", in *IEEE Communications Surveys & Tutorials*, Vol. 23, No. 3, pp. 1759–1799.
- [5] Yang, K., Shi, K., Zhou, Y., Yang, Z., Fu, L. and Chen, W. (2020), "Federated Machine Learning for Intelligent IoT via Reconfigurable Intelligent Surface", in *IEEE Network*, Vol. 34, No. 5, pp. 16–22.
- [6] Nguyen, D., Ding, M., Pathirana, P., Seneviratne, A. and Vincent Poor, H. (2021), "Federated Learning for Internet of Things: A Comprehensive Survey", in *IEEE Communications Surveys & Tutorials*, Vol. 23, No. 3, pp. 1622–1658.
- [7] Elayan, H., Alogaily, M. and Guizani, M. (2021), "Deep Federated Learning for IoT-based Decentralized Healthcare Systems", *2021 International Wireless Communications and Mobile Computing (IWCMC)*, pp. 105–109. DOI: 10.1109/IWCMC51323.2021.9498820.

INTELLECTUAL DATA ANALYSIS MODEL IN IIOT

M. Klymash, O. Hordiichuk-Bublivska, B. Koval

Lviv Polytechnic National University, 12, S. Bandery Str., Lviv, 79013, Ukraine

An overview of intelligent data processing methods in the systems of the Industrial Internet of Things is presented in this paper. A comparison of Big Data analysis methods in industrial systems with a significant load is provided. The methods of distributed machine learning for data processing are offered. A software model for data analysis of different volumes is developed in the work. The analysis of the basic approaches to the organization of machine learning is carried out: federal and undistributed. The effectiveness of the use of federal machine learning was experimentally proven, as it provides higher accuracy of data processing, even when increasing their volume. It is determined that unallocated machine learning works faster, so it can be used in systems where fast data processing is a priority. This approach opens up the possibility of creating an adaptive model of the Industrial Internet of Things system that can self-learn and adjust its infrastructure depending on changing parameters.

Key words: *IIoT; Big Data; Federated Learning.*