



ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ ВИКОРИСТАННЯ РОЗПОДІЛЕНОЇ МАТРИЧНОЇ ФАКТОРИЗАЦІЇ В ПРОМИСЛОВИХ СИСТЕМАХ

О. Гордійчук-Бублівська ^[ORCID:0000-0002-6439-549X]

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 26 липня 2023)

У роботі проаналізовано основні особливості оброблення великих обсягів даних в промислових smart grid системах. Визначено переваги розподілених обчислень для ефективнішого аналізу інформації. Досліджено рекомендаційні алгоритми, що дають змогу обробляти великі дані швидше та точніше. Запропоновано метод розподіленої матричної факторизації для надання рекомендацій користувачам smart grid систем, що передбачає обмін публічними даними між пристроями, із обробленням приватних локально. Визначено переваги запропонованої розподіленої рекомендаційної моделі, зокрема гнучкість у разі зміни параметрів, вищу точність обчислень внаслідок взаємообміну результатами між вузлами, високу швидкість оброблення даних, а також масштабованість.

Ключові слова: машинне навчання; Funk SVD; big data; розподілені обчислення.

1. Вступ

Завдяки упровадженню сучасних технологій змін зазнало багато галузей. Безліч позитивних змін, таких як розширення спектра послуг, що надають інформаційні системи, або підвищення продуктивності їх роботи свідчать про необхідність подальшого вдосконалення методів опрацювання даних. Складні інформаційні системи характеризуються розподіленою архітектурою. Такі системи складаються з великої кількості віддалених кінцевих пристроїв, серверів, баз даних, шлюзів, що разом формують єдину інформаційну структуру. Обмін даними між різними вузлами забезпечує спільне виконання поставлених завдань. Прикладами комплексних розподілених систем є smart city, промисловий Інтернет речей, smart grid, інтелектуальні системи різного призначення, що містять безліч компонентів. Системи, керовані програмними алгоритмами, можуть виконувати багато завдань самостійно, реагувати на події та приймати рішення [1].

Дані, зібрані з різноманітних джерел, у масштабних інформаційних системах називають великими, оскільки вони характеризуються певними особливостями. По-перше, інформація подана в різних форматах, наприклад, текстові повідомлення, відео, зображення тощо. Тому необхідне оброблення даних та подання у зручнішому для подальшого аналізу вигляді. Для обробки великих даних доцільно використовувати методи штучного інтелекту, машинного та глибокого навчання. Завдяки такому підходу можна швидко знаходити складні взаємозв'язки між прихованими характеристиками даних, аналізувати попередні події та передбачати майбутні. Завдяки цьому можна не тільки усувати загрози, а й запобігати їм [2].

У роботі досліджено методи оброблення даних у промислових системах. Розглянуто принципи функціонування систем smart grid для ефективного та гнучкого розподілу електроенергії в промисловості. Проаналізовано роботу алгоритму Funk SVD для оброблення великих даних і обчислення рекомендацій користувачам промислових систем. Визначено переваги Funk SVD та запропоновано його розподілену модифікацію, що дає змогу збирати та аналізувати дані з різних пристроїв системи smart grid. Для цього визначаються їх приватну та публічну частини. Приватні дані користувача залишаються на пристрої, а публічні використовують для обміну результатами формування рекомендацій з іншими вузлами. Досліджено ефективність роботи запропонованого алгоритму та визначено точність, гнучкість та швидкість під час оброблення даних.

2. Аналіз та постановка задачі

Оскільки доводиться постійно обробляти велику кількість даних, інформаційні системи стикаються з необхідністю постійного вдосконалення. Розподілена архітектура системи передбачає використання множини пристроїв для оброблення і зберігання даних, а також кінцевих користувачьких терміналів. Взаємодія між пристроями, які територіально розподілені та відрізняються за функціоналом та характеристиками, є комплексним завданням, для вирішення якого використовують спеціальні алгоритми та протоколи оброблення даних. Також важливим етапом оброблення великих даних є їх оптимізація, що передбачає виявлення найважливішої інформації, подання її у зрозумілому вигляді.

У розподілених інформаційних системах оптимізація даних ускладнена з таких причин:

- необхідність забезпечення розподіленого доступу до даних, що розташовані на різних пристроях. У разі використання неефективного методу доступу до даних тривалість обчислень значно зростає, а користувачі отримують неякісні послуги;
- важливість синхронізації оброблення даних. У разі невчасного оновлення станів усіх вузлів користувачі можуть одержати застарілу або ж помилкову інформацію;
- безпека приватних даних під час зберігання та передавання по каналах зв'язку. Розподілені системи повинні бути захищені від хакерських атак, несанкціонованого доступу та змін важливої інформації;
- відмовостійкість розподіленої системи є гарантією обслуговування користувачів з належною якістю. Забезпечується використанням резервних вузлів оброблення даних;
- якщо зростає кількість користувачів та створюваний ними трафік, важливо забезпечити масштабованість розподіленої системи. Зокрема, використання хмарних обчислювальних ресурсів дає змогу швидко та гнучко підлаштовувати папараметри системи залежно від викликів.

Перехід до методів інтелектуального опрацювання даних дає змогу ефективніше здійснювати оптимізацію та подальшу їх обробку [3]. Завданнями систем оброблення даних часто є аналіз текстів, зображень, визначення взаємозв'язків між категоріями товарів. Також обробляються аудіо- та відеофайли, дані з соціальних мереж та вебсторінок. Для ефективної подальшої обробки даних доцільно використовувати методи попередньої оптимізації. Інформація перевіряється на важливість для результатів обчислень. Оскільки часто дані, зібрані з різних джерел, містять багато надлишковості, доцільно відкидати частину з них, залежно від вимог до точності результату. Одним із потужних інструментів опрацювання великих масивів даних є рекомендаційні системи (recommender systems, RS). Такі системи виконують аналіз взаємозв'язків між користувачами та товарами або послугами. Потім формуються рекомендації про товари, імовірно, цікаві для кожного користувача окремо чи групи загалом. У такий спосіб можна значно покращити якість обслуговування клієнтів комерційних систем. Також рекомендаційні системи використовують для виявлення аномалій у роботі складних структур, наприклад, несправності обладнання на підприємстві, для інтелектуальної фільтрації контенту в соцмережах [4,5].

Переваги використання RS – удосконалення користувацького досвіду взаємодії з цифровими платформами, визначення найбільш та найменш популярних товарів, покращення показників продажів. Основою роботи рекомендаційних систем є збирання даних про користувачів та їх взаємодію з товарами. Для цього використовують різноманітні оцінки, які користувачі надавали запропонованим їм сервісам, наприклад, рейтинги товарів, статистику продажів, перегляд реклами чи пошукові запити. Отримані дані аналізують, визначаючи особливості користувачів, які замовляють або купують відповідну продукцію.

3. Особливості обробки даних у розподілених системах

Упровадження новітніх технологій в сучасні інформаційні системи сприяє їх ефективній роботі. Машинне навчання (machine learning, ML) дає змогу істотно прискорити опрацювання великих даних та вирішувати складні завдання швидше та з високою точністю. У випадку, коли традиційні обчислювальні методи вже не можуть проаналізувати всю необхідну інформацію, алгоритми ML працюють гнучкіше, легко масштабуються та автоматизують громіздкі обчислення. Для аналізу даних, розташованих на різних пристроях, доцільно використовувати розподілені алгоритми машинного навчання [6–8].

Машинне навчання дає змогу виявити складні взаємозалежності у великих обсягах даних: можна виявити шаблони поведінки різних процесів та передбачити з високою точністю майбутні події. Оброблення даних за допомогою машинного навчання в розподілених системах містить декілька важливих етапів, зокрема збирання та попередню оптимізацію даних, розподілене їх зберігання, об'єднання отриманих на різних вузлах результатів у глобальну модель та їх обмін між користувачами, оптимізацію результатів навчання.

Smart grid (SG) – інтелектуальні промислові системи, спрямовані на гнучке та ефективне розподілення електроенергії між користувачами. Завдяки використанню алгоритмів управління та аналізу даних з багатьох кінцевих пристроїв визначають оптимальне значення електроенергії для користувачів залежно від споживання в різні пори року, місяця та дня. Перевагами систем SG є їх гнучкість у розподілі ресурсів, масштабованість та надійність. Надання послуг користувачам за результатом аналізу їхніх потреб та попередніх дій сприяє поліпшенню якості обслуговування. У разі виявлення зростання чи зменшення споживання енергії система smart grid корегує рівень надання послуг, оптимізуючи навантаження [9–10].

Рекомендаційні системи використовують у smart grid для регулювання розподілу електроенергії залежно від потреб конкретних кінцевих користувачів [11, 12]. Завданням рекомендаційних систем є надання чітких і персоналізованих рекомендацій про моделі енергооптимізації на основі аналізу зібраних даних про активність кінцевих, розподільчих та керуючих пристроїв. Розумні електромережі обробляють величезні обсяги даних з різних джерел (пристрої IoT, давачі). Рекомендаційні алгоритми можуть аналізувати ці дані для оптимізації моделей використання енергії, балансування та розподілу навантаження, допомагаючи мінімізувати втрати енергії та зменшити загальне її споживання. Також рекомендаційні системи можуть передбачати періоди пікового навантаження і забезпечувати стабільне та ефективне енергопостачання. У разі виявлення збоїв у роботі, несправності обладнання чи зміни навантаження на систему smart grid рекомендаційні системи сприяють вирішенню проблеми, підбираючи оптимальне рішення. Чим більше даних аналізують такі рекомендаційні системи, тим точніші й коректніші результати їх роботи.

Основні особливості реалізації рекомендаційної системи для smart grid такі:

- оптимізація затрат електроенергії. Постійний моніторинг рівня споживання енергії, стану навколишнього середовища, промислової інфраструктури дають змогу надавати рекомендації для належного обслуговування, зменшення ризиків виходу з ладу системи в години пікового навантаження;
- гнучке обслуговування запитів користувачів, визначення найвідповідніших тарифів для них, пакетів послуг, програм з енергозбереження тощо;

- використання відновлюваних джерел енергії. Рекомендаційна система аналізує виготовлення та використання відновлюваної енергії та пропонує користувачам оптимальні для них способи її використання;

- надійність та відмовостійкість. Наявність у системі smart grid великої кількості кінцевих пристроїв та сенсорів дає можливість своєчасно отримувати сигнали про аварійні чи небезпечні ситуації. Рекомендаційні системи сприяють визначенню параметрів роботи мережі, запобігаючи виникненню таких ситуацій;

- планування та організованість роботи. Завдяки комплексному аналізу процесів роботи smart grid систем визначають загальні шаблони їх функціонування, що потім застосовують для досягнення визначених цілей.

Рекомендаційні системи в smart grid використовують методи машинного навчання та математичної статистики для виявлення закономірностей у роботі окремих підсистем та системи в цілому. Для оброблення великих даних в smart grid системах застосовують різноманітні алгоритми статистичного аналізу даних, машинного навчання. Також відзначають матричну факторизацію, тобто перетворення початкових матриць даних для зручнішого аналізу. Зокрема, ефективно застосовують алгоритм сингулярної декомпозиції даних (Singular Value Decomposition, SVD) [14], що дає змогу перетворити громіздкі матриці даних на менші, відкинувши надлишковість. Це можливо завдяки перетворенню початкової матриці даних на добуток трьох підматриць, в яких потім визначають найважливіші елементи та видаляють інші. Також SVD дає змогу знаходити відсутні значення у даних, доповнюючи зібрані значення, тим самим ефективніше аналізуючи їх та формуючи рекомендації користувачам (рис. 1).

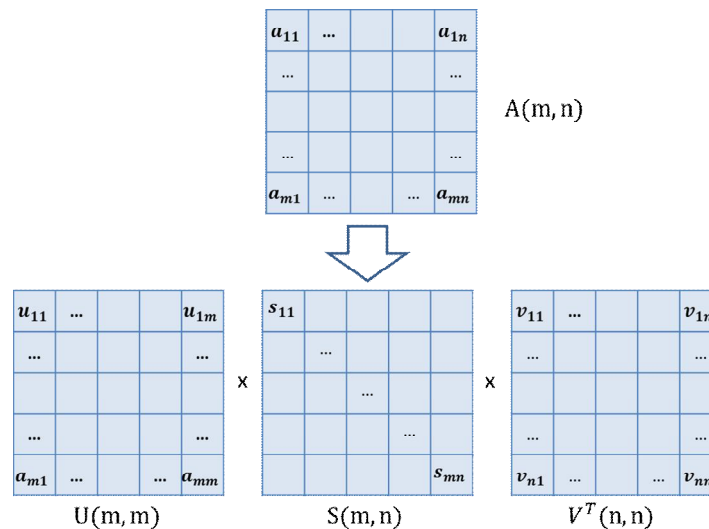


Рис. 1. SVD декомпозиція даних

На вхід рекомендаційної системи надходить матриця A , що містить m рядків та n стовпців. Кожен її елемент $a_{i,j}$ визначає оцінку, яку користувач i поставив товару j . Після SVD розкладу формуються матриці $U(m, n)$, що містять дані про користувачів з початкової матриці; $V(n, n)$, яка містить дані про товари; та $S(m, n)$, що є діагональною матрицею. Матриця $S(m, n)$ містить дані на головній діагоналі за зменшенням, ці дані є коефіцієнтами взаємозв'язку між користувачами і товарами. Якщо початкова матриця A містить порожню клітинку $a_{i,j}$, її значення можна отримати, обчисливши скалярний добуток рядка i матриці U та стовпця j матриці V^T . Так можна визначити, якою буде імовірна оцінка користувача i щодо товару j , та заповнити порожні клітинки початкової матриці даних.

Однак такий підхід не завжди оптимальний для реальних промислових систем, оскільки зібрані в них матриці дані розріджені. Щоб опрацювати їх алгоритмом SVD, потрібні великі затрати обчислювальних потужностей. Для вирішення цієї проблеми використовують алгоритм Funk SVD [15, 16], що одразу визначає, скільки елементів з початкової матриці залишити, не опрацюючи зайві. Завдяки гнучкій технології обчислень Funk SVD зручний для роботи в рекомендаційних системах.

Втім, у промислових системах також виникають труднощі під час використання наведених алгоритмів. Для спільного обчислення необхідно збирати дані з різних кінцевих пристроїв та передавати центральному серверу. Це створює навантаження на канали передавання, ставить нові вимоги щодо надійності даних і уповільнює обчислення. Натомість, не всі кінцеві пристрої мають достатні набори даних для того, щоб самостійно формувати рекомендації. Для вирішення цієї проблеми пропонуємо вдосконалений розподілений Funk SVD, що уможливило спільне обчислення рекомендацій для різних пристроїв, зберігаючи надійність приватних даних. Такий алгоритм працює у декілька етапів:

1. Здійснюється Funk SVD декомпозиція локальних даних A_{node} на кожному вузлі:

$$A_{node} = U_{local} * S_{local} * V_{local}^T, \quad (1)$$

де U_{local} – матриця, що містить приховані характеристики про користувачів; V_{local}^T – матриця прихованих характеристик про товари; S_{local} – діагональна матриця, що містить значення рівня залежностей між користувачами та послугами.

2. U_{local} та S_{local} надсилаються центральному пристрою від кожного користувача.

3. Центральний обчислювальний пристрій визначає глобальні U_{global} та S_{global} у системі з n вузлів:

$$U_{global} = \text{mean}(U_{local_1}, \dots, U_{local_n}), \quad (2)$$

$$S_{global} = \text{mean}(S_{local_1}, \dots, S_{local_n}), \quad (3)$$

де $\text{mean}(\)$ – функція визначення середнього значення для кожного елемента матриць.

4. U_{global} та S_{global} надсилають назад кінцевим пристроям.

5. Кінцеві пристрої можуть обчислити власне V_{global}^T :

$$V_{global}^T = \frac{U_{global}^T * A_{node}}{S_{global}}. \quad (4)$$

Роботу модифікованого Funk SVD у розподіленій промисловій системі зображено на рис. 2.

На рис. 2 зображено процес обчислень для одного центрального пристрою та відповідних йому кінцевих пристроїв. Для вибору даних від кінцевих пристроїв за ними повинен бути закріпленний набір міток, що допомагають серверу встановити ефективну комунікацію. Таких міток може бути дуже багато, залежно від особливостей роботи промислових систем. У цій роботі визначено декілька найважливіших:

- унікальний ідентифікатор у мережі (Device ID), що дає змогу встановити, з якого вузла надійшла інформація, чи він аутентифікований та авторизований;
- обчислювальна потужність, яка може бути різних рівнів (low-power, medium-power, high-power), та визначати, чи пристрій може брати участь в обробленні даних на належному рівні. Якщо потужність менша за встановлене порогове значення, його участь в обчисленнях відхиляють;
- інформація про відстань до сервера, що визначає розподіл за територіальною ознакою вузлів за певним центральним пристроєм. У дослідженні обчислено за алгоритмом K-Nearest Neighbors (KNN).

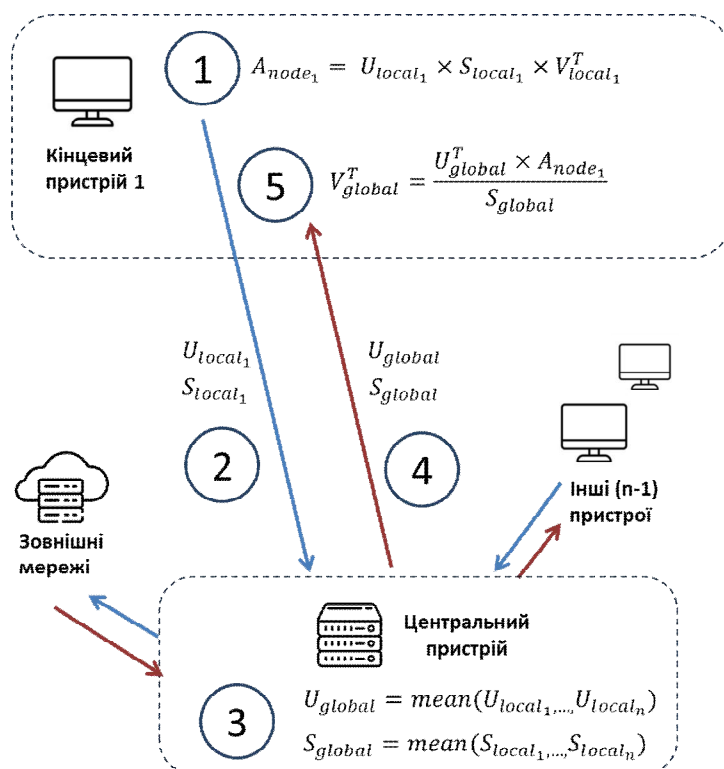


Рис. 2. Розподілений Funk SVD алгоритм

Наведені параметри важливі для організування розподілених обчислень, дають змогу регулювати кількість вузлів, що беруть участь в обробленні даних. Отже, сервер може відхиляти запити та дані від невідомих пристроїв, або тих, які не мають ресурсів, необхідних для обчислень. Кількість підмереж у реальних промислових системах може змінюватися залежно від потужності пристроїв, відстані між ними, сумісності тощо. Центральним обчислювальним пристроєм може також бути шлюз, що володіє достатніми ресурсами оброблення даних від вузлів. Дані користувача A_{node} залишаються на кінцевому пристрої, їх складно відновити на основі тільки U_{local} та S_{local} .

У роботі дослідження зосереджено на підвищенні точності обчислень рекомендацій у рекомендаційній системі. Завдяки запропонованій модифікації Funk SVD вузли обмінюються інформацією про рекомендації, пришвидшуючи навчання рекомендаційної системи. Замість того, щоб кожен вузол самостійно визначав рекомендації з нуля, він може використовувати частково вивчені моделі з інших вузлів, зменшуючи загальний час навчання. Нижче детальніше розглянуто дослідження та моделювання роботи різних алгоритмів.

4. Дослідження та моделювання розподіленого рекомендаційного алгоритму

Для дослідження ефективності запропонованого розподіленого Funk SVD алгоритму його порівняно з SVD та звичайним Funk SVD. Для моделювання використано мову програмування Python та вбудовані бібліотеки, зокрема NumPy. В моделюванні використано чотири обчислювальні процеси та однакові матриці даних розмірністю 1500 рядків та стовпців для всіх експериментів. Для дослідження взято дані з різним ступенем розрідженості, тобто заповненості елементів матриць. Для цього з початкових даних видалили частину значень, яку досліджувані алгоритми мали відновити.

На рис. 3 відображено порівняння точності рекомендацій (що відповідає ступеневі подібності початкової та після відновлення матриць) для різних методів. Для зручності результат подано в відсотках.

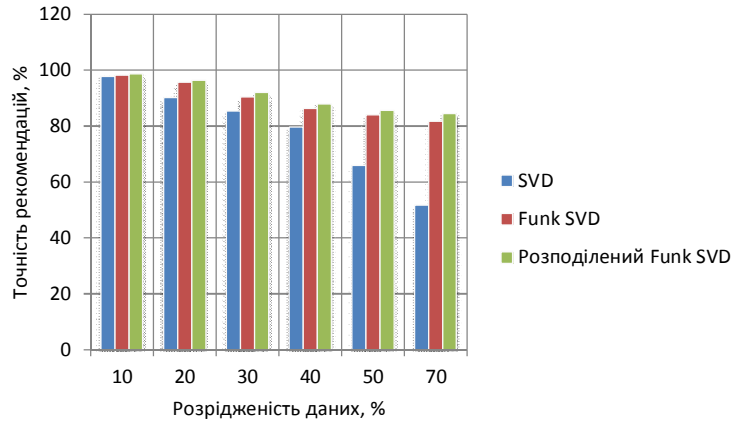


Рис. 3. Порівняння точності реконструкції матриці даних для SVD, Funk SVD та розподіленого Funk SVD алгоритмів залежно від ступеня розрідження даних

Точність обчислень визначається середнім відхиленням між значеннями елемента початкової матриці даних та тим, який відновлено внаслідок розкладу різними алгоритмами. Як зазначено вище, порожній елемент початкової матриці можна отримати, скалярно помноживши рядок і стовпець матриць U та V з відповідними індексами. Можна обчислити значення середньої квадратичної помилки MSE (Mean Squared Error) для початкової A_{node} та відновленої $A_{predicted}$ матриць:

$$MSE = \frac{1}{n} \sum_{i,j=1}^n (A_{i,j_{node}} - A_{i,j_{predicted}})^2, \tag{5}$$

де $A_{i,j_{node}}$ та $A_{i,j_{predicted}}$ є елементами матриць A_{node} та $A_{predicted}$ відповідно;

n – розмірність матриць (кількість рядків і стовпців, стало число);

i та j – індекси елемента матриці, тобто номери стовпця та рядка, на перетині яких він розташований.

Згідно із рис. 3, алгоритм розподіленого Funk SVD демонструє найвищу точність рекомендацій. Це пояснюється обміном даними між вузлами. Отримані від центрального пристрою U_{global} та S_{global} локальний вузол використовує для відновлення значень, відсутніх у власних початкових даних, тобто $A_{predicted}$:

$$A_{predicted} = U_{global} * S_{global} * V_{global}^T \dots \tag{6}$$

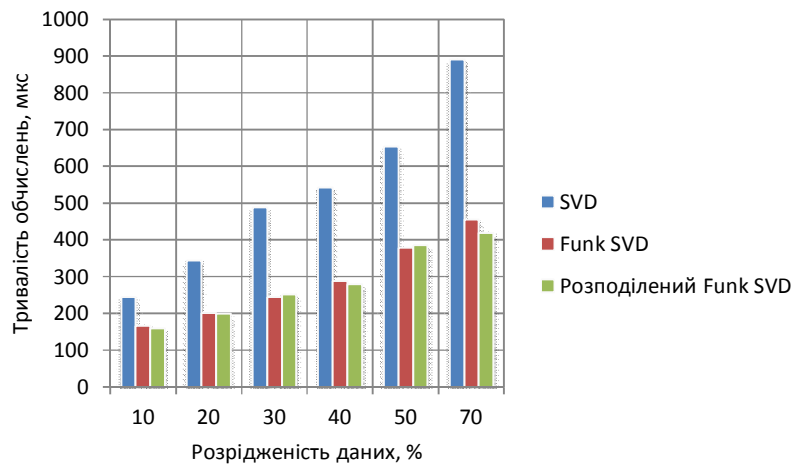


Рис. 4. Порівняння швидкості обчислення даних для SVD, Funk SVD та розподіленого Funk SVD алгоритмів залежно від ступеня розрідження даних

Враховуючи особливості експерименту, в промислових системах результати обчислень можуть різнитися залежно від кількості пристроїв, властивостей даних. На рис. 4 також продемонстровано залежності часу виконання алгоритмів від ступеня розрідження даних, що обробляються.

На рис. 4 видно, що алгоритм SVD працює найповільніше, а звичайний та модифікований Funk SVD швидше.

Коротко підсумуємо: результати досліджень, зображені на рис. 3, підтверджують вищу точність обчислень рекомендацій модифікованим розподіленим Funk SVD алгоритмом. Запропонований алгоритм гнучкий у разі використання у розподілених промислових системах, дає змогу обробляти користувацькі дані, не пересилаючи їх між вузлами в повному обсязі, а тільки частину Funk SVD розкладу. Швидкість обчислень, згідно із рис. 4, залишається доволі високою, що сприяє наданню своєчасних рекомендацій. Запропонована модифікація демонструє власні переваги та перспективи подальших досліджень у галузі розподіленої матричної факторизації.

Висновки

Розроблення нових методів та засобів оброблення даних сприяє підвищенню ефективності роботи інформаційних систем. Використовувати велику кількість кінцевих пристроїв проблематично через обсяги трафіку, які вони створюють. Також необхідно забезпечувати високу якість оброблення запитів користувачів, надійність і конфіденційність даних. Рекомендаційні системи дають змогу виявляти залежності між подіями, користувачами і товарами, тому вони є важливим інструментом для оброблення інформації.

У роботі проаналізовано особливості оброблення великих даних в smart grid системах. Визначено важливість використання методів машинного навчання для підвищення ефективності роботи таких систем. Досліджено функціонування алгоритмів SVD та Funk SVD під час оброблення розріджених даних великих обсягів, отриманих від великої кількості користувачів. Запропоновано модифікований розподілений Funk SVD метод, що уможливорює розподілене виконання алгоритму Fed SVD у промислових smart grid системах. Алгоритм, на відміну від звичайного, дає змогу обмінюватися результатами обчислень рекомендацій між вузлами, причому не потрібно пересилати приватні дані користувача. Такий підхід сприяє використанню методів матричної факторизації в розподілених системах. Результати дослідження модифікованого Funk SVD продемонстрували вищу точність обчислень, порівняно з подібними алгоритмами, надійність, а також доволі велику швидкість.

Список використаних джерел

- [1] M. Li, H. Wang and J. Li, "Mining conditional functional dependency rules on big data", in *Big Data Mining and Analytics*, Vol. 3, No. 1, pp. 68–84, March 2020. DOI: 10.26599/BDMA.2019.9020019.
- [2] A. Cuzzocrea, "Big Data Lakes: Models, Frameworks, and Techniques", 2021 *IEEE International Conference on Big Data and Smart Computing (BigComp)*, Jeju Island, Korea (South), 2021, pp. 1–4. DOI: 10.1109/BigComp51126.2021.00010.
- [3] F. Peng, H. Wang, L. Zhuang, M. Wang and C. Yang, "Methods of enterprise electronic file content information mining under big data environment", 2020 *International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*, Bangkok, Thailand, 2020, pp. 5–8. DOI: 10.1109/ICBASE51474.2020.00008.
- [4] M. P. Maharani, P. Tobianto Daely, J. M. Lee and D.-S. Kim, "Attack Detection in Fog Layer for IIoT Based on Machine Learning Approach", 2020 *International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju, Korea (South), 2020, pp. 1880–1882. DOI: 10.1109/ICTC49870.2020.9289380.
- [5] M. Klymash, O. Hordiichuk-Bublivska, M. Kyryk, L. Fabri and H. Kopets, "Big Data Analysis in IIoT Systems Using the Federated Machine Learning Method", 2022 *IEEE 16th International Conference on Advanced*

- Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, Lviv-Slavske, Ukraine, 2022, pp. 248–252. DOI: 10.1109/TCSET55632.2022.9766908.
- [6] S. K. Kishore, G. Vasukidevi, E. P. C. Prasad, T. R. Patnala, V. P. Reddy and P. B. Chanda, “A Real-Time Machine learning based cloud computing Architecture for Smart Manufacturing”, 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2022, pp. 562–565. DOI: 10.1109/ICAAIC53929.2022.9792860.
- [7] M. D. Choudhry, J. S. B. Rose and S. M. P., “Machine Learning Frameworks for Industrial Internet of Things (IIoT): A Comprehensive Analysis”, 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Trichy, India, 2022, pp. 1–6. DOI: 10.1109/ICEEICT53079.2022.9768630.
- [8] B. Walek and P. Fajmon, “A Recommender System for Recommending Suitable Products in E-shop Using Explanations”, 2022 3rd International Conference on Artificial Intelligence, Robotics and Control (AIRC), Cairo, Egypt, 2022, pp. 16–20. DOI: 10.1109/AIRC56195.2022.9836983.
- [9] R. Sharma, S. Rani and S. Tanwar, “Machine Learning Algorithms for building Recommender Systems”, 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 785–790. DOI: 10.1109/ICCS45141.2019.9065538.
- [10] B. P. Kulkarni, S. Sai Krishna, K. Meenakshi, P. Kora and K. Swaraja, “Performance Analysis of Optimization Algorithms GA, PSO, and ABC based on DWT-SVD watermarking in OpenCV Python Environment”, 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1–5. DOI: 10.1109/INCET49848.2020.9154134.
- [11] R. K. Patel, A. Kumari, S. Tanwar, W. -C. Hong and R. Sharma, “AI-Empowered Recommender System for Renewable Energy Harvesting in Smart Grid System”, in *IEEE Access*, 2022, Vol. 10, pp. 24316–24326. DOI: 10.1109/ACCESS.2022.3152528.
- [12] Yu Jun, Olena Hordiichuk-Bublivska, Yan Lingyu, Marian Kyryk, Mykola Beshley, Hu Jiwei, “Big Data Analysis in Smart Grid Systems”, 18th IMEKO TC10 Conference “Measurement for Diagnostics, Optimisation and Control to Support Sustainability and Resilience” Warsaw, Poland, September 26–27, 2022.
- [13] N. Zhi-hong and Z. Fei, “Research on Semi-supervised Recommendation Algorithm Based on Hybrid Model”, 2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), Taiyuan, China, 2020, pp. 344–348. DOI: 10.1109/MLBDBI51377.2020.00073.
- [14] Y. Jaradat, M. Masoud, I. Jannoud, A. Manasrah and M. Alia, “A Tutorial on Singular Value Decomposition with Applications on Image Compression and Dimensionality Reduction”, 2021 International Conference on Information Technology (ICIT), Amman, Jordan, 2021, pp. 769–772. DOI: 10.1109/ICIT52682.2021.9491732.
- [15] M. Klymash, M. Kyryk, Y. Pyrih, O. Hordiichuk-Bublivska and T. Andrukhiv, “Model of Large Sparse Datasets Processing Efficiency in IIOT”, 2023 17th International Conference on the Experience of Designing and Application of CAD Systems (CADSM), Jaroslaw, Poland, 2023, pp. 45–49. DOI: 10.1109/CADSM58174.2023.10076508.
- [16] G. Xin, J. Qin, X. Song and J. Zheng, “Dual Auto-Encoder Based Rating Prediction Recommendation Algorithm”, in *IEEE Access*, 2022, Vol. 10, pp. 97289–97297. DOI: 10.1109/ACCESS.2022.3205610.

INVESTIGATION OF DISTRIBUTED MATRIX FACTORISATION EFFICIENCY IN THE INDUSTRIAL SYSTEMS

Olena Hordiichuk-Bublivska

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

The processing of big data is an exceedingly urgent challenge in the functioning of modern information systems. The latest information technologies must be employed to collect, store, and analyze vast amounts of information. Intelligent data processing systems were implemented in numerous fields, particularly in the industry. Smart industrial systems also utilize data from various devices, enabling automated management processes and network component analysis. A prime example of an intelligent industrial system is the smart grid, which efficiently distributes electricity to users by considering demand, network parameters, load, etc.

Processing large amounts of information necessitates the use of machine learning methods and mathematical data analysis. Matrix factorization serves as an exemplary technique for transforming information into a more convenient form for further processing, establishing relationships between elements, and optimizing outcomes. In particular, the SVD (Singular Value Decomposition) and Funk-SVD algorithms are employed to address big data processing challenges, and they were discussed in this work. The key features of processing large data volumes in industrial smart grid systems were analyzed in the paper. The advantages of distributed computing for more efficient information analysis were identified. The recommendation algorithms that enable faster and more accurate processing of extensive data were explored in the study. Specifically, the SVD and Funk-SVD algorithms, used in recommendation systems for large data processing, were examined. A method of distributed matrix factorization to provide recommendations to smart grid system users was proposed in the paper. This approach involves the exchange of public data between devices and the local processing of private data. The advantages of this distributed model include flexibility in adjusting parameters, improved calculation accuracy through result exchange between nodes, high data processing speed, and scalability were identified. The conclusion that the proposed method can be effectively used in recommendation systems within the smart grid context, enhancing automated management processes and resource distribution was exclaimed.

Key words: *machine learning; Funk SVD; big data; distributed computing.*