

**О. В. Гордійчук-Бублівська, Л. П. Фабрі**

Національний університет "Львівська політехніка", м. Львів, Україна

МАТРИЧНА ФАКТОРИЗАЦІЯ ВЕЛИКИХ ДАНИХ У ПРОМИСЛОВИХ СИСТЕМАХ

Створення нових технологій та їх впровадження в різні сфери зумовило потребу оброблення та зберігання великих обсягів даних. В промислових системах модернізація означає використання великої кількості розумних пристроїв, що виконують спеціалізовані функції, а зібрані дані використовують для управління системою та автоматизації виробничих процесів. В роботі визначено основні характеристики розумних промислових систем. Проаналізовано концепцію Промислового Інтернету речей (англ. *Industrial Internet of Things*, IIoT) та актуальність проблеми модернізації висобництва. Досліджено проблеми оброблення великих даних в системах Промислового Інтернету речей. Розглянуто використання рекомендаційних систем для швидкого знаходження взаємозв'язків між користувачами та послугами виробництва. Проаналізовано використання алгоритмів матричної факторизації SVD (англ. *Singular-Value Decomposition*) та FunkSVD для оброблення розріджених матриць даних. Проведено моделювання роботи запропонованих алгоритмів і визначено переваги FunkSVD для роботи з розрідженими даними. Встановлено, що алгоритм FunkSVD опрацьовує дані за меншу тривалість часу, ніж SVD, але це не впливає на точність результату. З'ясовано, що SVD також є складнішим у реалізації та вимагає більше обчислювальних ресурсів. Удосконалено алгоритм FunkSVD для покращення ефективності оброблення великих масивів інформації так, що використовують менше даних для формування рекомендацій. Встановлено, що модифікований метод працює швидше за звичайний, проте зберігає високу точність обчислень, що є важливим для роботи в рекомендаційних системах. Виявлено можливість надавати рекомендації користувачам промислових систем за короткий поміжок часу, в такий спосіб покращуючи їх актуальність. Запропоновано продовжувати дослідження для знаходження оптимальних параметрів алгоритму FunkSVD.

Ключові слова: Промисловий Інтернет речей; сингулярне подання матриці; розріджені дані; рекомендаційні системи.

Вступ / Introduction

Модернізація промислових систем дала можливість змінити вигляд сучасного виробництва, пропонувати нові послуги користувачам, підвищувати продуктивність виробництва. Натомість, використання великої кількості розумних остаточних пристроїв вимагає постійного та швидкого аналізу даних. Розподілена система промислових систем і залучення хмарних ресурсів передбачає використання програмних алгоритмів управління ресурсами, згідно з актуальними вимогами до продуктивності виробництва. Оскільки розумні промислові системи розглядають як єдину структуру, що займається як виробленням продукції, так і реалізації її користувачам, постійно проводять аналіз ефективності на підставі відгуків користувачів, як явних, так і прихованих. До явних відгуків можна віднести такі, що користувачі самі залишили. Неявні відгуки визначають на підставі взаємодії користувачів і товарів. Наприклад, кількість покупок, пошукових запитів, вподобань у соціальних мережах тощо. Всі ці дані мають значення для визначення зацікавленості користувачів в наявних товарах та виробництва нових [7], [6]. Інформація, отримана від користувачів, є величезного обсягу, а також не вся вона є важливою для подальшого аналізу. Тому використовують спеціальні рекомендаційні системи, що дають змогу впорядкувати дані про користувачів і товари, визначити взаємозв'язки між ними і відкинути надлишковість. В роботі розглянуто алгоритми для реалізації роботи рекомендаційних систем SVD (англ. *Singular-Value Decomposition*) та FunkSVD (алгоритм, популяри-

зований Саймоном Функом). Проведено порівняння ефективності роботи наведених алгоритмів для роботи з великими масивами даних. Запропоновано також вдосконалити FunkSVD для формування рекомендацій з використанням меншої кількості даних.

Об'єкт дослідження – оброблення великих обсягів даних у рекомендаційних системах.

Предмет дослідження – матрична факторизація даних з використанням алгоритму FunkSVD.

Мета роботи – підвищення ефективності оброблення даних в рекомендаційних системах.

Для досягнення зазначеної мети визначено такі основні завдання дослідження:

- проаналізувати наявні методи матричної факторизації SVD та FunkSVD;
- вдосконалити метод FunkSVD для визначення рекомендацій користувачам на підставі меншої кількості даних;
- за результатами роботи вдосконаленого методу FunkSVD встановити оптимальні параметри роботи рекомендаційних систем.

Аналіз останніх досліджень та публікацій. Проблема цифровізації промислового виробництва та оброблення великих даних є зараз актуальною. V. C. Jadała та ін. [7] розглядають наводять переваги Промислового Інтернету речей та визначають потребу його інтегрування з хмарними технологіями для більш гнучкого управління. X. Nan [6] в своїй роботі розглядає вирішення проблеми оптимізації транспортних перевезень в промислових системах за допомогою Інтернету речей (англ. *Internet of Things*, IIoT) та хмарних обчислень. N.

Lin та ін. досліджують аналіз показників роботи системи Інтернету речей для ефективної маршрутизації транспортних засобів роботі [9]. В роботі [16] P. Zhang та ін. використовують федеративне та глибоке машинне навчання для оброблення даних з різних джерел. При цьому забезпечується надійність, конфіденційність та ефективність навчання. A. Simeone, Y. Zeng та ін. [13] розглядають проблему автоматизації виробництва і пропонують використання рекомендаційних систем для визначення потреб користувачів. D. Petrik та ін. пропонують створювати цифрові додатки в сфері IoT на підставі аналізу попередньо зібраних даних [11]. Y. Qiu, X. Zhu та J. Lu [12] досліджують методи оптимізації даних при визначенні фізичних навантажень для користувачів фітнес додатків. S. M. Kasongo [8] пропонує систему виявлення вторгнень в IoT на підставі таких алгоритмів, як генетичний та дерево прийняття рішень. В роботі [10] S. Mantravadi та ін. розглядають проблеми використання великої кількості остаточних пристроїв в IoT та формують рекомендації щодо покращення рівня надійності і конфіденційності. I.T. Christou та ін. пропонують алгоритми машинного навчання, що аналізують кількісні властивості даних у промислових системах [2]. В роботі [4] H. Gao та ін. використовують колаборативну фільтрацію даних в системах IoT, а F. Sun та X. Li [14] досліджують алгоритм Сингулярної Декомпозиції Даних для покращення ефективності функціонування систем астрономічної навігації. В роботі [15] J. Wang та ін. пропонують вдосконалення алгоритмів кластеризації, використовуючи сингулярну декомпозицію даних. K. El Handri та A. Idrissi досліджують метод багатокритеріального прийняття рішень на підставі сингулярної декомпозиції в роботі [3]. S. Guo та C. Li в роботі [5] пропонують гібридну рекомендаційну систему, використовуючи вдосконалений алгоритм FunkSVD. В роботі [1] K. Virul модифікує FunkSVD, використовуючи для надання рекомендацій дані не про всіх користувачів, а тільки одного. При цьому дані про товари обробляють в повному обсязі. Такий підхід зумовлений тим, що кількість користувачів промислової системи збільшуються значно швидше, ніж товарів. Метод демонструє високу точність оброблення даних і надання рекомендацій. В даній роботі запропоновано використання даних від різної кількості користувачів для досягнення оптимальних показників функціонування рекомендаційної системи.

Методи оброблення та оптимізації великих даних. Завдяки вдосконаленню інформаційних технологій змінюються також різні сфери. Використання різноманітних розумних пристроїв дає змогу значно спростити виконання різноманітних операцій та оброблення великих даних. В зв'язку з пандемічною ситуацією, особливо важливою задачею є забезпечення безпеки для користувачів і персоналу інформаційних систем. Тому делегування функцій людини програмним і апаратним засобам дає змогу організувати дистанційну роботу персоналу. Промисловий Інтернет Речей (англ. *Industrial Internet of Things*, IoT) є сучасною концепцією роботи промислових систем, що передбачає децентралізацію виробництва і залучення остаточних пристроїв до збирання даних і управління виробництвом. Вбудовані в комп'ютерну мережу датчики та сенсори визначають різноманітні параметри виробництва: температуру нав-

колишнього середовища, точність роботи пристроїв, ефективність виконання задач тощо [9]. Концепція Промислового Інтернету речей стосується не тільки вдосконалення процесу виробництва продукції. Змінюється також і основні мета роботи промислових систем. Отже, розумне промислове виробництво не тільки спрощує процес виготовлення товарів, а й дає змогу оптимізувати складні обчислення [11], [12], [13], [16].

Рекомендаційні системи зазвичай використовують для пошуку взаємозв'язків між покупцями та товарами. Оскільки часто немає інформації про рівень зацікавленості певного товару конкретним покупцем, його можна обрахувати, беручи до уваги рейтинги схожих товарів чи покупців. В системах Промислового Інтернету речей рекомендаційні системи можуть використовуватися для визначення найбільш затребуваних товарів, необхідних затрат на виробництво, а також для встановлення зв'язків між різними факторами, що впливають на роботу системи. Інформація з різних джерел збирається у вигляді матриць відповідностей. Оскільки часто матриці про взаємозв'язок між певними факторами є розріджені, тобто окремі клітинки є порожніми, доцільно їх ущільнювати перед подальшою обробкою [2], [8], [10]. На рис. 1 наведена архітектура промислового виробництва з використанням рекомендаційної системи.



Рис. 1. Архітектура промислового виробництва з використанням рекомендаційної системи / Architecture of industrial production using a recommendation system

Зібрані дані про користувачів і товари формують спеціальні таблиці рекомендацій. В таких таблицях, наприклад, по вертикальній осі розташовують дані про користувачів, а по горизонтальній – про товари. На перетині цих координат розміщені рейтинги для конкретного товару, який виробляється промисловою системою, від користувача. Часто таблиці рекомендацій не повністю заповнені, оскільки не всім товарам користувач виставив оцінку. Для надання рекомендацій про нові товари слід аналізувати поведінку споживача щодо інших товарів і підбирати ті, що схожі з попередніми за певним набором ознак. Також можна визначати необхідні рекомендації на підставі подібності користувачів між собою. До прикладу, споживачі одного віку і професії можуть мати схожі вподобання щодо визначених товарів. Отже, вирахувавши подібність попередніх вподобань, або знайшовши схожість користувачів, можна заповнити порожні клітинки в таблиці рекомендацій.

Найчастіше в рекомендаційних системах застосовуються методи колаборативної фільтрації, що виділяють групи людей зі схожими ознаками та надають їм подібні рекомендації. Оскільки таблиці рекомендацій не завжди є повністю заповнені (тобто розріджені), а також не всі елементи потрібні для подальшого обчислення рекомендацій, виникає проблема нераціонального використання ресурсів. Визначення найважливіших для аналізу та прийняття рішень даних і відкинення зайвих необхідне для ефективної роботи промислових систем, особливо при обробленні великих обсягів інформації [4]. Для підвищення ефективності оброблення даних, що зібрані з різних пристроїв, необхідно використовувати ефективні математичні і статистичні методи. Одним з ефективних є метод сингулярної декомпозиції даних SVD (англ. *Singular Value Decomposition*). За допомогою методу SVD можна розкласти матрицю даних N на добуток трьох під матриць:

$$N = U \times \Sigma \times V^T, \quad (1)$$

де N – матриця розмірності (m, n) , $U(m, m)$, $\Sigma(m, n)$, $V(n, n)$. Матриця Σ є діагональною, отже, елементи розміщені тільки на головній діагоналі в порядку спадання значень. Ці елементи називаються сингулярними значеннями або числами матриці, а їх кількість s дає змогу визначити точність подальших обчислень взаємозв'язків між даними. Якщо в матриці Σ обрати s найбільших значень, в матрицях U та V також можна використовувати тільки s рядків. Після їх перемноження отримаємо матрицю N_1 , яка володітиме найважливішими статистичними властивостями початкової, але буде меншого обсягу. Отже, можна пришвидшити процес оброблення великих даних і вибору найважливіших [14], [15]. SVD є прикладом матричної факторизації, тобто приведення матриці даних до меншої розмірності зі збереженням основних властивостей.

Алгоритм SVD є універсальним і часто застосовують для роботи з великими даними, при вирішенні задач пошуку закономірностей, передбачень тощо. В системах Промислового Інтернету речей також часто необхідно впорядкувати інформацію, зібрану з різних пристроїв. Проте, виникає проблема розрідженості даних, коли багато значень в таблиці є незаповненими. Наприклад, при зборі даних про вподобання користувачів, тільки частина їх збирається для кожного товару. Під час формування таблиць, що відображають взаємозв'язки між процесами в системі, незаповнені клітинки не допомагають їх подальшому обробленню, проте займають обчислювальні ресурси.

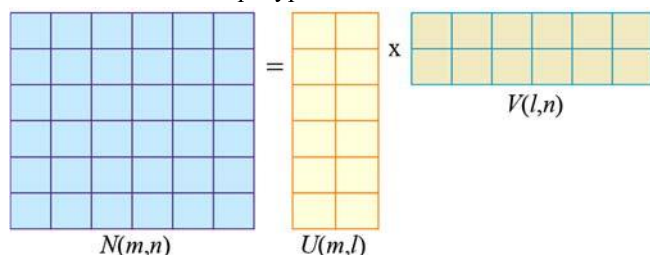


Рис. 2. Розклад даних за допомогою алгоритму FunkSVD / Data decomposition using the FunkSVD algorithm

SVD проводить розклад початкової матриці даних на добуток трьох, що є аналогічного розміру. В подальшому можна визначити найважливіші дані та відкинути надлишкові, але тільки після сингулярного розкладу.

Тому, для вирішення проблем роботи з розрідженими даними, використовують алгоритм FunkSVD, що теж є прикладом матричної факторизації [3], [5]. FunkSVD проводить розклад матриці $N(m, n)$ на дві менші $U(m, l)$ та $V(l, n)$, де l – точність оброблення даних (рис. 2).

Результати дослідження та їх обговорення / Research results and their discussion

Проведемо дослідження ефективності матричної факторизації даних. Дані містять оцінки товарів користувачами, обчислення виконано з використанням мови програмування Python та бібліотеки Numpy і TensorFlow. На рис. 3 зображено залежність часу виконання обчислення рекомендацій для даних алгоритмом SVD за різної кількості сингулярних чисел.

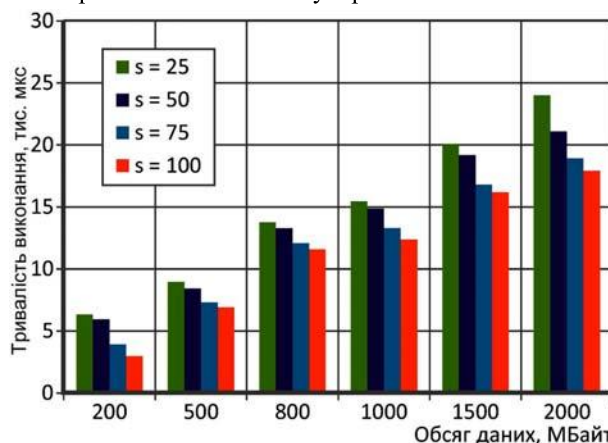


Рис. 3. Залежність тривалості обчислення даних алгоритмом SVD за різної кількості сингулярних чисел s / Dependence of the data calculation time by the SVD algorithm for different singular numbers s

Завдяки корекції кількості сингулярних чисел можна досягти оптимізації часу виконання алгоритму SVD залежно від вимог до точності обчислень. Оскільки при відкиданні певної кількості даних частина інформації не використовують, це може дещо вплинути на точність надання рекомендацій. Залежність ефективності надання рекомендацій від кількості сингулярних чисел приведено на рис. 4.

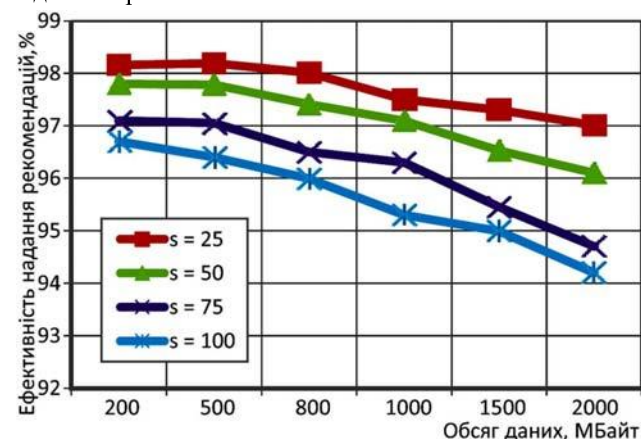


Рис. 4. Залежність ефективності надання рекомендацій алгоритмом SVD від кількості сингулярних чисел s / Dependence of recommendations effectiveness by the SVD algorithm on the singular numbers s

Ефективність надання рекомендацій обчислюють як середню різницю між відомими та передбачуваними даними, подають у відсотках (100 % означає абсолютно

точне надання рекомендацій). Зазвичай, такого точного результату неможливо досягти навіть при використанні усіх даних. Тому за допомогою визначення найбільш важливих і відкинення частини сингулярних чисел пришвидшується процес обчислення великих обсягів даних, зберігаючи достатньо високу точність надання рекомендацій.

Порівняємо швидкість опрацювання даних. Залежність тривалості обчислень для алгоритмів SVD та FunkSVD наведена на рис. 5.

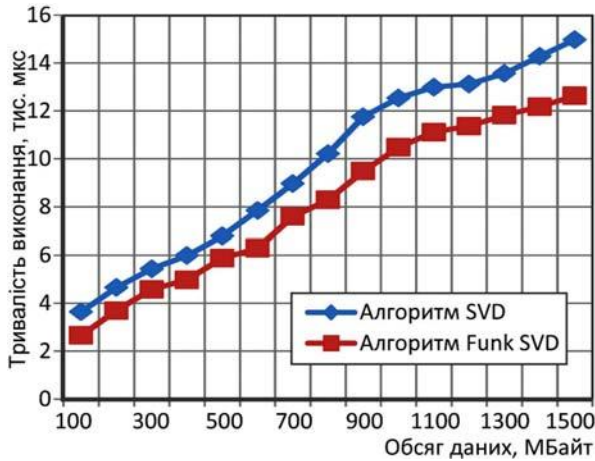


Рис. 5. Залежність тривалості обчислень для алгоритмів SVD та FunkSVD / Calculation duration dependence for SVD and FunkSVD algorithms

Як бачимо з рис. 5, FunkSVD проводить обчислення швидше ніж звичайний SVD. Це дає змогу більш ефективно обробляти великі дані, а також оптимально використовувати ресурси пам'яті. Залежність ефективності надання рекомендацій для алгоритмів SVD та FunkSVD наведена на рис. 6.

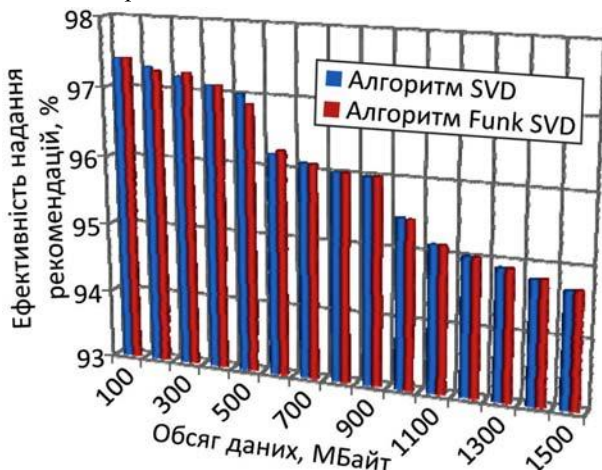


Рис. 6. Залежність ефективності надання рекомендацій для алгоритмів SVD та FunkSVD / Dependence of recommendation efficiency for SVD and FunkSVD algorithms

З рис. 6 зрозуміло, що алгоритми SVD та FunkSVD показують майже однакові показники ефективності надання рекомендацій. Розглянемо детальніше роботу алгоритму FunkSVD. Для проведення розкладу матриці генеруються випадкові матриці U та V , а потім ітеративно підбирається значення, які відповідають початковій матриці.

Рекомендації відповідно до алгоритму FunkSVD надаються, обчислюючи добуток рядка та стовпця, що відповідають індексу шуканого елемента в матриці:

$$r_{i,j} = \sum_{\lambda \in I} (u_{i,\lambda} \times v_{\lambda,j}) \quad (2)$$

В роботі запропоновано не використовувати повністю рядок з користувачами, оскільки при обчисленні великих даних він є значного розміру. Натомість рядок товарів частіше менший, тому можна його повністю обчислювати [1]. Для дослідження обрано обчислення рекомендацій для розріджених матриць даних, при використанні даних від різної кількості користувачів, а не від усіх. На рис. 7 зображено порівняння тривалості обчислень.

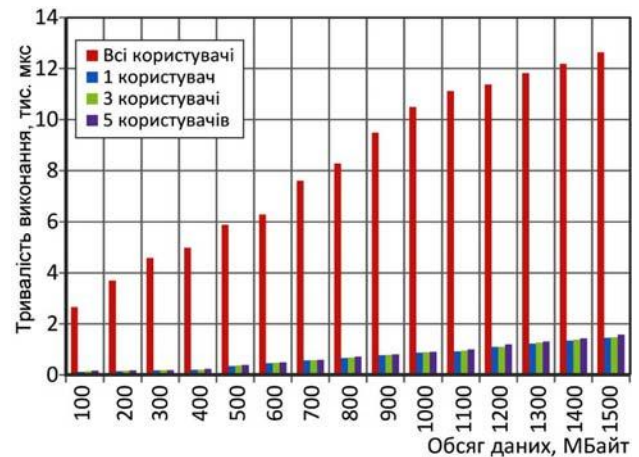


Рис. 7. Залежність тривалості обчислень для алгоритму FunkSVD від кількості користувачів, що враховуються при наданні рекомендацій / Dependence of the calculation time for the FunkSVD algorithm on the number of users used for providing recommendations

Дослідимо ефективність рекомендацій, які надає алгоритм. Залежності для модифікованого алгоритму FunkSVD приведені на рис. 8.

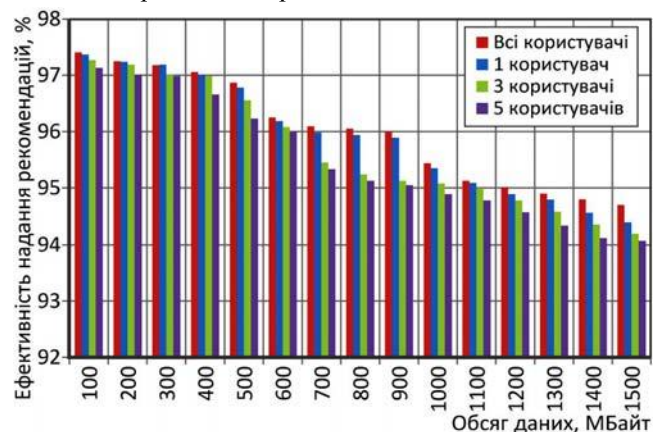


Рис. 8. Залежність ефективності надання рекомендацій для алгоритму FunkSVD від кількості користувачів, що враховуються при наданні рекомендацій / Dependence of the recommendations effectiveness for the FunkSVD algorithm on the number of users used for providing recommendations

Як бачимо з рис. 7 і рис. 8, зменшення кількості проаналізованих користувачів під час формування рекомендацій не впливає надто сильно на точність результату, проте може значно скоротити тривалість обчислень. Такий підхід дає змогу ефективніше обробляти та аналізувати великі і розріджені матриці даних, що сприяє роботі рекомендаційних систем. Завдяки точному і швидкому аналізу даних можна забезпечувати роботу інтелектуальних промислових систем та автоматизувати значну частину виробничих процесів.

Обговорення результатів дослідження. Згідно з проведеними дослідженнями, оброблення великих обсягів даних у промислових системах є актуальною проблемою. Розглянуті методи SVD і FunkSVD дають змогу визначати взаємозв'язки між даними. Оскільки метод FunkSVD більш ефективний для роботи з розрідженими матрицями даних, доцільно його використовувати в промислових системах для оброблення інформації, отриманої від різних джерел. Запропонована модифікація методу FunkSVD спрощує його роботу, проте зберігає точність обчислень.

Отримані результати дослідження дають змогу сформулювати наукову новизну та практичну значущість проведеного дослідження.

Наукова новизна отриманих результатів дослідження – удосконалено метод FunkSVD для матричної факторизації даних, який використовує менше інформації від користувачів для надання рекомендацій щодо цікавих товарів чи послуг.

Практична значущість результатів дослідження – отримані результати можна використати для подальших досліджень та покращення ефективності роботи рекомендаційних систем при обробленні розріджених матриць даних. Тривалість обчислень модифікованого FunkSVD скорочується, а точність результатів залишається достатньо високою. Запропонований метод можна застосовувати для надання своєчасних і актуальних рекомендацій користувачам промислових систем.

Висновок / Conclusions

Промисловість змінюється та модернізується, використовуючи сучасні методи оброблення та аналізу даних. В роботі розглянуто особливості функціонування інтелектуального промислового виробництва. Визначено важливість оброблення великих масивів даних за допомогою рекомендаційних систем. Проаналізовано ефективність методів SVD і FunkSVD для оброблення великих даних. Проведено моделювання роботи наведених методів і визначено, що метод FunkSVD працює швидше, ніж SVD, а також вимагає менше затрат обчислювальних ресурсів, при цьому точність надання рекомендацій залишається високою. Також запропоновано модифікувати алгоритм Fed SVD для роботи з великими матрицями даних і використовувати тільки частину даних про користувачів для формування рекомендацій. Отримані результати показують, що модифікований метод пришвидшує обчислення та зберігає достатньо високу точність. Тому доцільно продовжувати дослідження даного алгоритму для покращення ефективності роботи рекомендаційних систем в системах промислового виробництва.

References

- [1] Birul, K. (2016). A novel latent factor model for recommender system," in *JISTEM – Journal of Information Systems and Technology Management*, 13(3), 497–514. <https://doi.org/10.4301/S1807-17752016000300008>
- [2] Christou, I. T., Kefalakis, N., Zalonis, A., & Soldatos, J. (2020). Predictive and Explainable Machine Learning for Industrial Internet of Things Applications, 16th International Conference on Distributed Computing in Sensor Systems (DCOSS), 213–218. <https://doi.org/10.1109/DCOSS49796.2020.00043>
- [3] El Handri, K., & Idrissi, A. (2021). Parallelization of Top_k Algorithm Through a New Hybrid Recommendation System for Big Data in Spark Cloud Computing Framework, *IEEE Systems Journal*, 5(4), 4876–4886. <https://doi.org/10.1109/JSYST.2020.3019368>
- [4] Gao, H., Qin, X., Barroso, R. J. D., Hussain, W., Xu, Y., & Yin, Y. (2022). Collaborative Learning-Based Industrial IoT API Recommendation for Software-Defined Devices: The Implicit Knowledge Discovery Perspective, *IEEE Transactions on Emerging Topics in Computational Intelligence*, 6(1), 66–76. <https://doi.org/10.1109/TETCI.2020.3023155>
- [5] Guo, S., & Li, C. (2020). Hybrid Recommendation Algorithm based on User Behavior, *IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, 2242–2246. <https://doi.org/10.1109/ITA-IC49862.2020.9339083>
- [6] Han, X. (2022). Design and Implementation of Intelligent Logistics Equipment Scheduling Platform based on Internet of Things and Cloud Computing, *IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI)*, 979–986. <https://doi.org/10.1109/ICET-CI55101.2022.9832062>
- [7] Jadala, V. C., Pasupuletti, S. K., Raju, S. H., Kavitha, S., Sai Bhaba, Ch. M H, & Sreedhar, B. (2021). Need of Internet of Things, Industrial IoT, Industry 4.0 and Integration of Cloud for Industrial Revolution, *Innovations in Power and Advanced Computing Technologies (IPACT)*, 1–5. <https://doi.org/10.1109/i-PACT52855.2021.9696696>
- [8] Kasongo, S. M. (2021). An Advanced Intrusion Detection System for IIoT Based on GA and Tree Based Algorithms, *IEEE Access*, 9, 113199–113212. <https://doi.org/10.1109/ACCESS.2021.3104113>
- [9] Lin, N., Shi, Y., Zhang, T., & Wang, X. (2019). An Effective Order-Aware Hybrid Genetic Algorithm for Capacitated Vehicle Routing Problems in Internet of Things, *IEEE Access*, 7, 86102–86114. <https://doi.org/10.1109/ACCESS.2019.2925831>
- [10] Mantravadi, S., Schnyder, R., Møller, C., & Brunoe, T. D. (2020). Securing IT/OT Links for Low Power IIoT Devices: Design Considerations for Industry 4.0, *IEEE Access*, 8, 200305–200321. <https://doi.org/10.1109/ACCESS.2020.3035963>
- [11] Petrik, D., Schönhofen, F., & Herzwurm, G. (2022). Understanding the Design of App Stores in the IIoT, *IEEE/ACM International Workshop on Software-Intensive Business (IW-SIB)*, 43–50.
- [12] Qiu, Y., Zhu, X., & Lu, J. (2021). Fitness Monitoring System Based on Internet of Things and Big Data Analysis, *IEEE Access*, 9, 8054–8068. <https://doi.org/10.1109/ACCESS.2021.3049522>
- [13] Simeone, A., Zeng, Y., & Caggiano, A. (2021). Intelligent decision-making support system for manufacturing solution recommendation in a cloud framework, *International Journal of Advanced Manufacturing Technology*, 112. <https://doi.org/10.1007/s00170-020-06389-1>
- [14] Sun, F., & Li, X. (2021). Star Chart Recognition Algorithm Based on Singular Value Decomposition, *IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 124–129. <https://doi.org/10.1109/IAEAC50856.2021.9391032>
- [15] Wang, J., Wang, K., Jia, R., & Chen, X. (2020). Research on Load Clustering Based on Singular Value Decomposition and K-means Clustering Algorithm, *Asia Energy and Electrical Engineering Symposium (AEEES)*, 831–835. <https://doi.org/10.1109/AEEES48850.2020.9121555>
- [16] Zhang, P., Wang, C., Jiang, C., & Han, Z. (2021). Deep Reinforcement Learning Assisted Federated Learning Algorithm for Data Management of IIoT, *IEEE Transactions on Industrial Informatics*, 17(12), 8475–8484. <https://doi.org/10.1109/TII.2021.3064351>

MATRIX FACTORIZATION OF BIG DATA IN THE INDUSTRIAL SYSTEMS

The creation of new technologies and their implementation in various fields necessitated Big Data processing and storage. In industrial systems, modernization means the use of a large number of smart devices that perform specialized functions. Data from such devices are used to control the system and automate production processes. A change in the parameters of individual components of the manufacturing system may indicate the need to adjust the global management strategy. The intelligent industrial systems main characteristics were defined in the paper. The Industrial Internet of Things concept and the relevance of the modernization problem for manufacturing were analyzed. The problems of processing Big Data in Industrial Internet of Things systems were examined in the paper. The use of recommendation systems for quickly finding relationships between users and production services was considered. The advantages of Big Data analysis by recommendation systems, which have a favourable effect on industrial enterprise efficiency were given. The use of SVD and FunkSVD matrix factorization algorithms for processing sparse data matrices was analyzed. The possibility of optimizing arrays of information, choosing the most important, and rejecting redundancy with the help of the above algorithms was determined. The proposed algorithms were simulated. The advantages of FunkSVD for working with sparse data were assigned. It was found that the FunkSVD algorithm processes the data in a shorter time than SVD, but this does not affect the accuracy of the result. The SVD is also more difficult to implement and it requires more computing resources was established. It has been shown that FunkSVD uses a lot of data to determine the relationships between it and make recommendations about the products most likely to be of interest to users. To increase the efficiency of processing large sets of information the FunkSVD algorithm was improved in such a way that it uses fewer data to generate recommendations. Based on the results of the research, the modified method works faster than the non-modified one but retains high calculation accuracy, which is important for work in recommender systems. The possibility of providing recommendations to users of industrial systems in a shorter period, thus improving their relevance, was revealed. It was proposed to continue research for finding the optimal parameters of the FunkSVD algorithm for Big Data processing.

Keywords: Industrial Internet of Things; Singular-Value Decomposition (SVD); sparse data; recommendation systems.

Інформація про авторів:

Гордійчук-Бублівська Олена Василівна, аспірантка, кафедра автоматизованих систем управління.

Email: olena.v.hordiichuk-bublivska@lpnu.ua

Фабрі Людвіг Павлович, канд. техн. наук, доцент, кафедра автоматизованих систем управління. **Email:** liudvih.p.fabri@lpnu.ua

Цитування за ДСТУ: Гордійчук-Бублівська О. В., Фабрі Л. П. Матрична факторизація великих даних у промислових системах. *Український журнал інформаційних технологій*. 2022, т. 4, № 2. С. 68–73.

Citation APA: Hordiichuk-Bublivska, O. V., & Fabri, L. P. (2022). Matrix factorization of big data in the industrial systems. *Ukrainian Journal of Information Technology*, 4(2), 68–73. <https://doi.org/10.23939/ujit2022.02.068>