

УДК 519.21 + 681.3

## КОМП'ЮТЕРНЕ ОБЧИСЛЕННЯ ПСЕВДО-ЗВОРОТНИХ МАТРИЦЬ ТА ЙОГО ЗАСТОСУВАННЯ ДО ЗАДАЧ ДОСЛІДЖЕННЯ СТАТИСТИЧНИХ ДАНИХ

Студ. О.Л. Василенко

Науковий керівник проф. С.М. Краснитський  
Київський національний університет технологій та дизайну

**Мета роботи** – розробка програмного забезпечення для ефективного обчислення псевдо-зворотних матриць в рамках сучасних підходів до комп'ютерної обробки даних. Одержання можливості використання отриманих результатів в задачах досліджень статистичних даних, зокрема при виконанні операцій з виродженими або погано зумовленими матрицями в лінійному регресійному аналізі при побудови моделі за допомогою методу найменших квадратів.

**Завдання роботи** полягає в розробці програмного продукту для ефективного (з точки зору кількості обчислювального часу) обчислення псевдо-зворотних матриць та їх застосування для рішення проблеми методу найменших квадратів.

**Об'єкт та предмет дослідження.** Об'єктом дослідження є методи обчислення псевдо-зворотних матриць та їх застосування до задач дослідження даних. Предметом дослідження є показники ефективності методів псевдо обернення за допомогою комп'ютерних систем.

**Методи та засоби дослідження.** При дослідженні підходів до обчислення псевдо-зворотних матриць було застосовано такі підходи:

- теоретичні: метод оцінки складності алгоритму по кількості потрібних операцій та порівняння алгоритмів за асимптотичною оцінкою швидкості росту складності обчислень.

- експериментальні: використання утиліти Java Monitoring and Management Console для спостереження за поведінкою роботи програмного забезпечення. Використання обчислювальних ресурсів таких як пам'ять, процесорний час та кількості задіяних потоків виконання байт-коду.

Для програмної реалізації використана інтегроване середовище розробки IntelliJ IDEA 2018.

**Наукова новизна та практичне значення отриманих результатів.** Сучасні наукові дослідження характеризуються збільшеними об'ємами статистичних даних, тому загальнодоступні інструменти не завжди справляються з поставленою задачею. Поява таких систем аналізу даних як “Big Data” та “Data mining” потребують більш ефективного підходу до обчислень. Проведені дослідження дозволяють удосконалити процес комп'ютерної обробки статистичних даних за рахунок зменшення затрат часу на обрахування псевдо-зворотної матриці, що є суттєвим при збільшенні об'єму вхідної матриці.

**Результати дослідження.** Псевдо-зворотні матриці використовуються в одному з найпопулярніших статистичних методів – методом найменших квадратів для пошуку приближеного розв'язку СЛАР. Сучасна література зазвичай пропонує різні підходи для обчислення псевдо-оберненої матриці. Приміром в [1],[2] запропоновані наступні методи: Рангова декомпозиція; алгоритм Грама — Шмідта; трансформації Хаусхолдера; Поворот Гівенса; SVD – розклад; обчислення за допомогою мінорів; розклад Холецкого. QR - декомпозиція може бути виконана за алгоритмом Грама — Шмідта, але він потребує обчислення матричних квадратних коренів, що при обчисленні за допомогою комп'ютерів приводить до погіршення якості результату. Модифікований

алгоритм Грама — Шмідта позбавлений необхідності обчислення квадратного кореня, що призводить до більшої точності роботи алгоритму.

Серед варіацій QR-розкладу найбільш широко використовуваним є трансформація (перетворення) Хаусхолдера, яка має вигляд:

$$H = I - \frac{1}{y}uu', \text{ де } y = \frac{1}{2}\|u\|^2.$$

Тут  $H$  – симетрична та ортогональна матриця, ключова властивість якої полягає в наступному: нехай  $x$  та  $y$  — вектори однакової довжини, тоді матриця Хаусхолдера  $H$  відповідає рівностям  $u = x - y$ ,  $Hx = y$ . Трансформацію Хаусхолдера можна використовувати для обчислення QR - матриці розміром  $m \times n$  з матриці  $A$  з  $m \geq n$ . Цей метод має більшу числову стійкість ніж метод Грама-Шмідта.

Більшість пакетів програмного забезпечення використовують метод сингулярного розкладу матриці, який є більш точним, але потребує більше операцій обчислення з плаваючою комою.

Якщо порівнювати QR - розклад та сингулярний розклад, то QR - метод потребує менше обчислювальних операцій, але надає гіршу точність. Використання чисел з подвійною точністю (округлення виконується до  $10^{-16}$ ) може збільшити точність розрахунку при QR - розкладі.

Популярні пакети програмного забезпечення для статистичних розрахунків (Statistica, Mathcad, Matlab) виконують обчислення на комп'ютері, де вони встановлені. Це приводить до наступних наслідків:

1) Персональний комп'ютер повинен мати достатні характеристики для обробки великих об'ємів даних.

2) Потрібно купувати для кожного комп'ютера окрему ліцензію для програмного пакету.

Виходячи з попередніх двох пунктів, легко зробити висновок – це не раціональне використання коштів, оскільки всі обчислення дешевше робити на одному, але потужному сервері, який має потужний процесор та й отримувати доступ до нього з комп'ютерів-клієнтів за допомогою технології WEB.

**Висновки.** Розробка програмного забезпечення, що імплементує вище описані підходи, дозволить оптимізувати процес статистичного аналізу при застосуванні лінійного регресійного аналізу шляхом зменшення накладних витрат на обчислення псевдо-зворотних матриць. Використання віддаленого обчислення за допомогою клієнт-серверної взаємодії дозволяє ефективно використовувати обчислювальні комп'ютерні системи групою користувачів, отримуючи доступ, за допомогою звичайного браузера.

**Ключові слова:** псевдо-зворотня матриця, метод найменших квадратів, QR-розклад, сингулярний розклад матриці.

#### ЛІТЕРАТУРА

1. A. D. Snider, E. B. Saff Fundamentals of Matrix Analysis with Applications John Wiley & Sons, Inc., Hoboken, 2015
2. Уон Фоккинк Распределенные алгоритмы. Интуитивный подход - Издательский дом «Питер» 2017
3. Одерски М., Спун Л., Веннерс Б. Scala. Профессиональное программирование - Издательский дом «Питер» 2018