

Дослідження методів, моделей та інформаційних технологій редукції даних

Васильєва Л.В., Житченко А.С.

Донбаська державна машинобудівна академія, м. Краматорськ

На даний момент існує досить велика кількість різноманітних програмних засобів, які можуть виконати це завдання, проте порівняльний огляд найпоширеніших показав, що модуль редукції даних є або додатковою можливістю (наприклад, в Excel), або входить в загальний пакет математичних функцій в складі спеціалізованого програмного забезпечення (Mathlab, STATISTICA, SPSS). Вузько направлені програмні засоби для обробки результатів чисельного моделювання є розробками підприємств і не поширюються на продаж або представляють собою деякі бібліотеки функцій. Таким чином, дослідження теми даної роботи, а також власна розробка програмного комплексу для вирішення задачі редукції даних є актуальною [1–3].

Мета дослідження – аналіз і підвищення ефективності роботи з алгоритмами редукції даних. Об'єкт дослідження – моделі та інформаційні технології редукції даних. Предмет дослідження – результати роботи алгоритмів редукції даних, їх порівняльний аналіз і характеристика побудованих моделей. Наукова новизна – удосконалення якості вихідних даних, модифікації візуалізації та оптимізації швидкості і зручності роботи методів редукції багатовимірних даних.

Метод головних компонент є одним з найбільш популярним і зручним засобом факторного аналізу і візуалізації багатовимірних даних. Саме тому його було обрано для реалізації у програмному засобі. Використовуючи технологію Web Workers API всі операції завантаження і розрахунків виконуються у фоновому процесі паралельно з роботою браузера. Таким чином можливо підвищити швидкість і уникнути «замороження» сторінки веб-браузеру у момент виконання важких операцій. Клас методу головних компонент «РСА» містить необхідний функціонал для обробки багатовимірних даних, але переважною частиною

даного алгоритму є візуалізація [4]. Побудування графічної частини розроблено за допомогою бібліотеки D3.js і React.js. Приклад побудування графіку типу «Biplot» представлено на рис. 1.

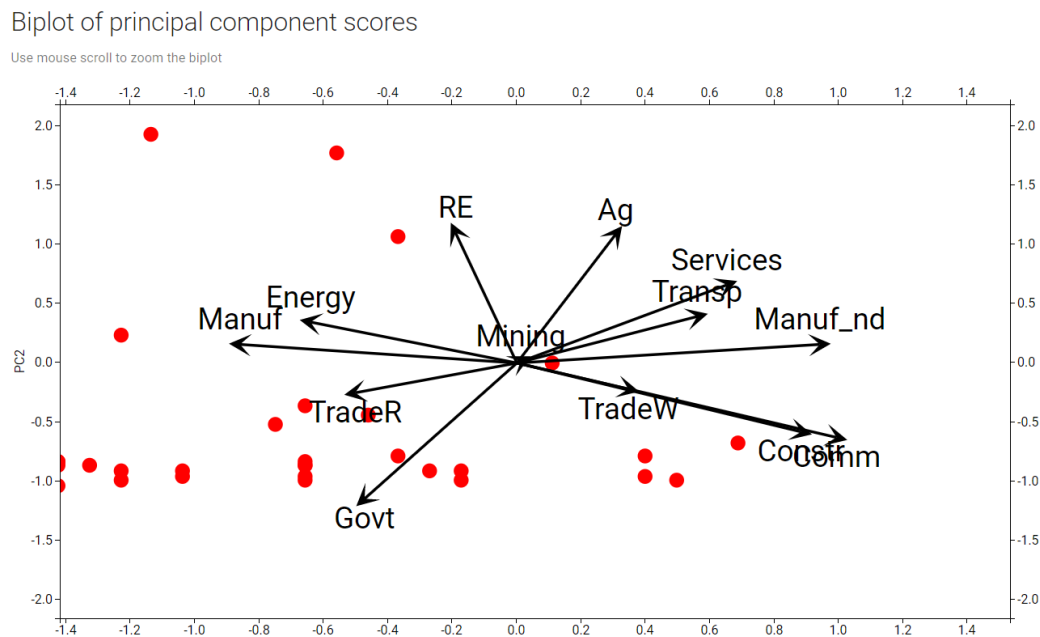


Рисунок 1 – Приклад побудованого графіку типу «Biplot» з масштабуванням.

Аналогічно до методу головних компонент розрахунки алгоритму карт Кохонена виконуються за допомогою Web Workers API. Алгоритм також працює з розробленим раніше методом головних компонент, використовуючи власні вектори з найбільшим значенням для визначення напрямку навчання, що дозволяє значно прискорити даний процес. Набір даних не потрібно стандартизувати, це буде зроблено автоматично алгоритмом класу карти Кохонена шляхом масштабування кожної функції до діапазону [0,1].

Література

1. Wang L., Wang G., Alexander C. A. *Big data and visualization: methods, challenges and technology progress* // *Digital Technologies*. – 2015. – Т. 1. – №. 1. – С. 33-38.
2. Hurter C. *Image-based visualization: Interactive multidimensional data exploration* // *Synthesis Lectures on Visualization*. – 2015. – Т. 3. – №. 2. – С. 1–127.
3. Васильєва Л.В. *Методика розв'язання задачі групування багатомірних об'єктів за допомогою кластерного аналізу* // *Фізико-математична освіта*. – 2017. – Випуск 3 (13). – С. 31-34.
4. <https://quernest.github.io/reduction/#/>