

## **ВИКОРИСТАННЯ WEB-ТЕХНОЛОГІЙ ДЛЯ РЕАЛІЗАЦІЇ МЕТОДІВ РЕДУКЦІЇ І ДИНАМІЧНОЇ ВІЗУАЛІЗАЦІЇ БАГАТОВИМІРНИХ ДАНИХ**

Основною ідеєю методів редукції є відображення багатовимірних даних у простір малої вимірності шляхом розв'язання задачі візуалізації багатовимірних даних, яке зберігає їх структуру. Це дозволяє виявити велику кількість нових закономірностей та визначити в подальшому методи аналізу цих даних. Термін редукція даних зазвичай відноситься до аналітичних методів (багатовимірні методи аналізу: факторний аналіз, багатовимірне шкалювання, кластерний аналіз, канонічна кореляція або нейронні мережі), які можуть пояснити мінливість вихідних багатовимірних даних [1]. Візуалізація – це важливий підхід, що допомагає отримати повне уявлення про дані і виявити значення даних [2].

Метою цієї роботи є дослідження методів редукції і динамічної візуалізації багатовимірних даних та розробка WEB-додатку, що їх реалізує. Була використана D3.js – data-driven JavaScript бібліотека для маніпулювання DOM елементами. D3.js дозволяє візуалізувати дані, використовуючи HTML, SVG та CSS. Акцент на веб стандарти дозволяє користуватися всіма можливостями сучасних браузерів без прив'язки до фреймворків, поєднуючи компоненти візуалізації та data-driven підхід до DOM маніпуляцій.

Розроблений WEB-додаток був використаний на наборі реальних даних [3], отриманих національним центром біотехнологічної інформації США (NCBI) і представляє собою таблицю порівняння нуклеотидів вірусу на збіг в ланцюгу DNA. Таблиця у форматі .csv і містить 262 рядки і 12 колонок. Кожен з рядків у таблиці представляє опис однієї мутації вірусу. Дослідження кожної характеристики певної мутації є дуже трудомісткою задачею і може не виправдати затраченого часу і очікувань.

Використовуючи алгоритми редукції даних ми можемо значно скоротити розрахункові витрати при обробці, виділити найважливішу інформацію і виконати проектування вибірки на двовимірний простір, що

---

<sup>1</sup> К.т.н., доц., доцент кафедри комп'ютерних інформаційних технологій, Донбаська державна машинобудівна академія (м. Краматорськ)

<sup>2</sup> магістр з комп'ютерних наук, Донбаська державна машинобудівна академія (м. Краматорськ)

дозволяє графічно представити дані. Було досліджено набір даних, використовуючи метод головних компонент і самоорганізаційні карти Кохонена. Метод головних компонент є одним з популярних засобів факторного аналізу і візуалізації багатовимірних даних. Використовуючи технологію Web Workers API, всі операції завантаження і розрахунків виконуються у фоновому процесі паралельно з роботою браузера. Клас методу головних компонент «РСА» містить необхідний функціонал для обробки багатовимірних даних, але переважною частиною даного алгоритму є візуалізація. Побудування графічної частини розроблено за допомогою бібліотеки D3.js і React.js. 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 напряму 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].

За результатами дослідження було зроблено візуалізацію у двовимірному просторі і ідентифіковано 5 основних типів мутації в коронавірусі, проведено порівняльний аналіз результатів роботи алгоритмів, які збігаються з результатами дослідженням BioTech [3].

Використані алгоритми редукції даних показали себе ефективними і мають достатню інформативність в обох випадках і добре підходять для вирішення багатьох проблем розмірності.

Подальші наукові пошуки можуть бути спрямовані на удосконалення якості вихідних даних, модифікації візуалізації та оптимізації швидкості і зручності роботи методів редукції багатовимірних даних.

#### ПЕРЕЛІК ПОСИЛАНЬ

1. 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.

2. 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.

3. James C. Ingram, «SARS CORONAVIRUS ACCESSION» Kaggle, doi: 10.34740/KAGGLE/DSV/953866.