

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ РАДИОФИЗИКИ И КОМПЬЮТЕРНЫХ
ТЕХНОЛОГИЙ**

Кафедра системного анализа и компьютерного моделирования

**Снижение размерности и классификация в экспериментах
секвенирования одиночных клеток**

Аннотация к дипломной работе

Дунского Никиты Дмитриевича
студента 4 курса, специальность
«компьютерная безопасность»

Научный руководитель:
доктор физико-математических наук
профессор Апанасович В.В.

Допущена к защите

«___» _____ 2022 г.

Зав. кафедрой системного анализа и
компьютерного моделирования,
кандидат физико-математических наук,
доцент В.В. Скакун

Минск, 2022

РЕФЕРАТ

Дипломная работа, 54 страницы, 22 рисунка, 18 источников, 5 приложений.

СЕКВЕНИРОВАНИЕ РНК ОДИНОЧНЫХ КЛЕТОК, МЕТОД ГЛАВНЫХ КОМПОНЕТ, МЕТОД Т-РАСПРЕДЕЛЕННЫХ СТОХАСТИЧЕСКИХ СОСЕДЕЙ, МЕТОД РАВНОМЕРНОГО ПРИБЛИЖЕНИЯ И ПРОЕКЦИИ, МЕТОД МНОГОМЕРНОГО ШКАЛИРОВАНИЯ.

Объект исследования – секвенирование молекул РНК одиночных клеток.

Цель работы – разработка и сравнительный анализ алгоритмов снижения размерности для классификации клеток на данных секвенирования одиночных клеток.

В результате выполнения работы реализованы алгоритмы предсказания снижения размерности для данных секвенирования одиночных клеток, а также алгоритмы классификации. Проведен сравнительный анализ методов. Проведено исследование экспериментальных данных лучшей комбинацией методов. Реализован алгоритм отбора признаков для повышения точности. Программная реализация выполнена в среде программирования Spyder на языке программирования Python.

РЭФЕРАТ

Дыпломная праца, 54 старонкі, 22 малюнка, 18 крыніц, 5 прыкладанняў.

СЕКВЕНІРАВАННЕ РНК АДЗІНАЧНЫХ КЛЕТАК, МЕТАД
ГАЛОЎНЫХ КАМПАНАТ, МЕТАД Т-РАЗМЕРКАВАННЫХ
СТАХАСТЫЧНЫХ СУСЕДЗЯЎ, МЕТАД РАЎНАМЕРНАГА НАБЛІЖЭННЯ І
ПРАЕКЦЫІ, МЕТАД ШМАТМЕРНАГА ШКАЛІРАВАННЯ.

Аб'ект даследавання – секвеніраванне малекул РНК адзіночных клетак.

Мэта працы – распрацоўка і параўнальны аналіз алгарытмаў зніжэння памернасці для класіфікацыі клетак на дадзеных секвеніравання адзіночных клетак.

У выніку выканання працы рэалізаваны алгарытмы прадказання зніжэння памернасці для дадзеных секвеніравання адзіночных клетак, а таксама алгарытмы класіфікацыі. Праведзены параўнальны аналіз метадаў. Праведзена даследаванне эксперыментальных даных лепшай камбінацыяй метадаў. Рэалізаваны алгарытм адбору прыкмет для павышэння дакладнасці. Праграмная рэалізацыя выканана ў асяроддзі праграмавання Spyder на мове праграмавання Python.

ABSTRACT

Graduate work consists of 54 pages. It contains 18 images, 5 sources.

SINGLE CELL RNA SEQUENCING, PRINCIPAL COMPONENT ANALYSIS METHOD, T-DISTRIBUTED STOCHASTIC NEIGHBOR EMBEDDING METHOD, UNIFORM MANIFOLD APPROXIMATION AND PROJECTION METHOD, MULTIDIMENSIONAL SCALING METHOD.

Object of research – sequencing of single cell RNA molecules.

Objective – development and comparative analysis of dimensionality reduction algorithms for cell classification based on single cell sequencing data.

As a result of the work, algorithms for predicting dimension reduction for single cell sequencing data, as well as classification algorithms, have been implemented. A comparative analysis of the methods was carried out. The study of experimental data by the best combination of methods was carried out. Implemented a feature selection algorithm to improve accuracy. The software implementation is performed in the Spyder programming environment in the Python programming language.