

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра дифференциальных уравнений и системного анализа

Аннотация к дипломной работе

**РАЗРАБОТКА АЛГОРИТМА КЛАСТЕРИЗАЦИИ НА ОСНОВЕ
МЕТОДОВ КОМБИНАТОРНОЙ ОПТИМИЗАЦИИ**

Меркушев Марк Ильич

Научный руководитель:
кандидат физ.-мат. наук,
О. А. Лаврова

2024

В дипломной работе 26 страниц, 11 рисунков, 16 источников, 1 приложение.

МАШИННОЕ ОБУЧЕНИЕ, КОМБИНАТОРНАЯ ОПТИМИЗАЦИЯ, ТЕОРИЯ ГРАФОВ, АЛГОРИТМ ГОЙМАНСА-УИЛЬЯМСОНА, АЛГОРИТМЫ КЛАСТЕРИЗАЦИИ, WOLFRAM MATHEMATICA, PYTHON

Объектом исследования является алгоритм кластеризации, разработанный на основе аппроксимационного алгоритма Гойманса-Уильямсона поиска максимального разреза графа.

Цель работы: разработка и реализация алгоритма в среде Wolfram Mathematica и на Python, тестирование его на реальном примере.

Для достижения поставленной цели были использованы следующие методы и инструменты: метод внутренней точки, метод сферической релаксации, метод k-средних, метод главных компонент, Wolfram Mathematica, Python.

В дипломной работе достигнуты следующие результаты:

- 1) Реализован алгоритм кластеризации на основе алгоритма Гойманса-Уильямсона в среде Wolfram Mathematica и на Python. Предложено применение алгоритма в задачах бинарной кластеризации. Приведена возможная реализация разбиения данных на два и более кластера посредством рекурсивной кластеризации.
- 2) Алгоритм протестирован на примере набора данных “Титаник”. На основе тестовых вычислений алгоритм сравнивается с базовым алгоритмом кластеризации k-Means. Продемонстрирована сопоставимость разработанного алгоритма с базовым по точности. Проанализированы причины расхождения времени работы алгоритмов.

Дипломная работа является теоретико-практической. Ее результаты могут быть использованы в сфере анализа данных.

Дипломная работа является завершенной, поставленные задачи решены в полной мере, присутствует возможность дальнейшего развития исследований.

Дипломная работа выполнена автором самостоятельно.

Thesis project is presented in the form of an explanatory note of 26 pages, 11 figures, 16 references, 1 application.

MACHINE LEARNING, COMBINATORIAL OPTIMIZATION, GRAPH THEORY, GOEMANS-WILLIAMSON ALGORITHM, CLUSTERING ALGORITHMS, WOLFRAM MATHEMATICA, PYTHON

The research object of this thesis project is a clustering algorithm developed based on the Goemans-Williamson approximation algorithm for finding the maximum cut of a graph.

The purpose of this work is to develop and implement the algorithm in Wolfram Mathematica and Python environments, and to test it on a real example.

The following methods and tools were used to achieve the goal: interior-point method, spherical relaxation method, k-means method, principal component analysis, Wolfram Mathematica, Python.

The main results of the thesis project are as follows:

- 1) The clustering algorithm based on the Goemans-Williamson algorithm has been implemented in the Wolfram Mathematica environment and in Python. An application of the algorithm in binary clustering tasks is proposed. Possible implementation of data partitioning into two and more clusters through recursive clustering is presented.
- 2) The algorithm was tested on the example of the “Titanic” dataset. Based on test calculations, the algorithm is compared with the basic clustering algorithm k-Means. The comparability of the developed algorithm with the basic one in terms of accuracy has been demonstrated. Reasons for the discrepancy in the algorithms’ running time are analyzed.

This thesis project is a theoretical and practical one. Its results can be used in the field of data analysis.

The thesis project is complete, all tasks have been successfully done, there is a possibility for further research and development.