ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»

Отчёт по лабораторной работе № 4

«Кластеризация массивов»

Выполнил работу

Кулигин Евгений

Академическая группа C3100

Принято

Ментор, Вершинин Владислав Константинович

Санкт-Петербург

2024

1. Введение

Цель работы: разработка алгоритма кластеризации массива с использованием полного перебора всех возможных разбиений.

Задачи

1.Изучить основные принципы кластеризации массивов.

2.Реализовать алгоритм кластеризации массива с минимизацией метрики разброса элементов.

3.Провести тестирование алгоритма на массивах различных размеров и оценить производительность.

4.Сравнить результаты тестирования с теоретическими оценками.

1. Теоретическая подготовка

Кластеризация — это процесс разделения массива на группы (кластеры), элементы которых схожи между собой.

Используемый алгоритм

1. Генерация масок

Все возможные разбиения массива генерируются в виде масок. Маска представляет собой набор чисел, где каждое число указывает, к какому кластеру принадлежит соответствующий элемент массива.

2. Метрика кластера

Для каждого кластера вычисляется метрика. Метрика – это сумма модулей разности элемента кластера и его среднего арифметического

3. Общая метрика

Общая метрика вычисляется как сумма метрик всех кластеров.

1. Реализация

Этапы выполнения

1. Генерация всех масок  
Маски создаются как числа в системе счисления с основанием n\_clusters

2. Распределение элементов по кластерам  
Элементы массива распределяются по кластерам согласно текущей маске.

3. Вычисление метрики  
Метрика рассчитывается для каждого разбиения, и сохраняется лучшее разбиение.

4. Контроль времени выполнения  
Если алгоритм превышает заданный лимит времени, выполнение прерывается.

Использование библиотек

#include <iostream> – Для ввода и вывода данных.

#include <vector> – Для работы с динамическими массивами.

#include <cmath> – Для выполнения математических операций.

#include <limits> – Для получения предельных значений типов данных.

#include <chrono> – Для измерения времени выполнения программы.

#include <cassert> – Для проверки корректности программы.

#include <stdexcept> – Для обработки ошибок и исключений.

1. Экспериментальная часть

Теоретическая асимптотика: O(4^N) \* (N^2), память: 4^N \* N^2 + 4^N \* N + 32N + 32 байт. (Подсчет приведен в приложении). Теоретическая сложность задачи составляет O(4^N) \* (N^2). Для тестирования алгоритма была собрана статистика, приведенная в таблице №1

Таблица №1 – Подсчет сложности реализованного алгоритма

|  |  |  |
| --- | --- | --- |
| Входной набор элементов | Время, мс | Теоретическая сложность |
|  |  |  |
| 1 | 0 | 0,0004 |
| 3 | 0 | 0,0576 |
| 5 | 1 | 2,56 |
| 6 | 8 | 14,7456 |
| 7 | 35 | 80,2816 |
| 8 | 144 | 419,4304 |
| 9 | 691 | 2123,366 |
| 10 | 2872 | 10485,76 |
| 11 | 11261 | 50751,08 |

Рисунок 1 – График O-нотации для выполненной программы.

5. Заключение

В ходе выполнения работы мною был реализован алгоритм кластеризации массива на основе полного перебора всех возможных комбинаций. Цель работы была достигнута путём тестирования алгоритма на массивах различного размера. Полученные результаты подтвердили корректность работы алгоритма, а также совпали с теоретическими оценками сложности как по времени, так и по памяти.

В качестве дальнейших исследований можно предложить оптимизацию алгоритма для снижения экспоненциального роста использования памяти и времени. Это может включать использование динамического программирования.

ПРИЛОЖЕНИЕ А

lab-4.cpp

1. Измерение времени выполнения программы, в зависимости от количества элементов массива.

Изображение выглядит как снимок экрана, текст

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

1. Проведение тестов Cassert

Изображение выглядит как текст, снимок экрана, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание