

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
“ВЫСШАЯ ШКОЛА ЭКОНОМИКИ”

# ОТЧЁТ ПО ПРАКТИКЕ

ИССЛЕДОВАНИЕ ВОЗМОЖНОГО СОКРАЩЕНИЯ ПЕРЕБОРА  
ПРИ ВЫБОРЕ ПАРАМЕТРОВ  $p, \beta$   
ДЛЯ АЛГОРИТМА  $A - Ward_{p\beta}$

**Студент:**  
Еремейкин П.А.  
группа  
мНoД16\_ТМСС

**Руководитель:**  
профессор  
Миркин Б.Г.

Москва 2018

# Содержание

<b>1</b>	<b>Основные положения</b>	<b>3</b>
<b>2</b>	<b>Алгоритм <math>A - Ward_{p\beta}</math></b>	<b>4</b>
2.1	Постановка задачи. Кластеризация . . . . .	4
<b>3</b>	<b>Методика эксперимента</b>	<b>5</b>
<b>4</b>	<b>Экспериментальное обеспечение</b>	<b>5</b>
<b>5</b>	<b>Результаты</b>	<b>5</b>
<b>6</b>	<b>Выводы</b>	<b>5</b>

## 1 Основные положения

Исследование выполняется в рамках развития пакета программ СИК (Система Интеллектуальной Кластеризации), который был разработан в ходе курсового проекта “Алгоритмы интеллектуализации метода k-средних”. Этот пакет предназначен для применения современных интеллектуальных методов при решении задач кластеризации.

В состав пакета входят методы иерархического кластер-анализа: метод аномальных кластеров, алгоритмы  $Ward$ ,  $A - Ward$ ,  $A - Ward_{p\beta}$  а также дивизивные методы.

С технической точки зрения СИК представляет собой набор Python модулей, объединённых в единую программу при помощи графического пользовательского интерфейса.

В рамках данной практики рассматривается проблема выбора параметров для алгоритма  $A - Ward_{p\beta}$ . Этот алгоритм представляет собой модифицированную версию иерархического алгоритма  $A - Ward$  и вводит два параметра:  $p$  и  $\beta$ . Оптимальные значения параметров зависят от конкретной задачи и данных, к которым применяется алгоритм. На настоящее время не существует рекомендаций по эффективному выбору этих параметров, а единственный обоснованный метод — перебор всех возможных значений с последующей оценкой результата для каждой пары  $(p_i, \beta_i)$  по эмпирической характеристике. Такой подход требует большого времени вычисления, что во многих случаях делает его неприменимым на практике.

Для решения задачи выбора параметров  $p, \beta$  в условиях ограниченного времени была выдвинута гипотеза о возможном сокращении перебора. Согласно этой гипотезе, результаты выбора оптимальных значений по всем доступным объектам и по сокращённой выборке из этих объектов различаются не существенно. Цель данной работы состоит в экспериментальной проверке приведённой гипотезы, оценке различных стратегий формирования сокращённой выборки, их характеристик относительно качества результата и затрачиваемого времени.

## 2 Алгоритм $A - Ward_{p\beta}$

### 2.1 Постановка задачи. Кластеризация

Алгоритм  $A - Ward_{p\beta}$  предназначен для решения задачи кластеризации, то есть выделения из таблиц наблюдения множеств (кластеров) таким образом, чтобы сходные объекты попадали в один и тот же кластер, а несходные — в разные кластеры [?].

### **3 Методика эксперимента**

### **4 Экспериментальное обеспечение**

### **5 Результаты**

### **6 Выводы**