

НОВЫЕ АЛГОРИТМЫ ИЕРАХИЧЕСКОЙ КЛАСТЕРИЗАЦИИ В СРАВНЕНИИ С КЛАССИЧЕСКИМИ МЕТОДАМИ ИЕРАХИЧЕСКОЙ КЛАСТЕРИЗАЦИИ

Валиева З. И., магистр

Миссаров М.Д., д.ф.-м.н., профессор

КФУ, г. Казань, Россия

Аннотация. Предложены два новых метода аггломеративной иерархической кластеризации, основанные на новых способах расчета расстояния между кластерами. Качество кластеризации оценивается с использованием

3

внутренних
показателей
качества:
коэффициента
силуэта,
индекса
Данна
и
индекса
Дэвиса-Болдуина.

Эксперименты

с
использованием
метода
Монте

-
Карло
проводились
для
разного
количества
точек,
разного
количества
кластеров
и
разных
размерностей
данных.

Новые
алгоритмы в большинстве экспериментов
показывают
средние
характеристики
показателей
качества

,
превышающие
средние
характеристики
этих
показателей
для
классических
алгоритмов

монотонной
иерархической
кластеризации.

Ключевые слова: иерархическая кластеризация, расстояния между кластерами, индексы качества, метод Монте-Карло.

Кластерный анализ относится к области машинного обучения без учителя. Методы кластерного анализа используются в экономике, финансах, социологии, медицине, биоинформатике и других областях [1]- [3].

...

полный текст во вложении