

0.1. Рылов С.А. Структуры данных для сеточной кластеризации мультиспектральных изображений

Алгоритмы кластеризации (разбиение данных на непересекающиеся подмножества по некоторой мере схожести) удобно использовать при решении многих прикладных задач. Сеточные алгоритмы кластеризации позволяют обрабатывать большие объёмы данных и при этом способны выделять кластеры сложной, заранее неизвестной формы [1]. В связи с этим сеточные алгоритмы хорошо подходят для сегментации мультиспектральных спутниковых изображений. Однако на практике встает проблема применения сеточных алгоритмов кластеризации к данным высокой размерности, возникающая из-за экспоненциальной зависимости объема сеточной структуры от размерности пространства признаков [2]. В докладе предлагается новая структура данных для хранения многомерной сеточной структуры, позволяющая существенно сократить зависимость объема занимаемой памяти от размерности с помощью перехода к хранению информации только о непустых ячейках сеточной структуры. Кроме того, были реализованы многомерные сеточные структуры на основе использования хеш-таблиц `HashMap` из `Java.Util` и `LongLongHashMap` из библиотеки `Eclipse Collections`. Все разработанные структуры данных были внедрены в сеточный алгоритм кластеризации НСА [3].

Проведенные экспериментальные исследования реализованных подходов на мультиспектральных спутниковых изображениях показывают, что они позволяют проводить сеточную кластеризацию данных высокой размерности (от 5 до 10) при адекватных затратах памяти. При этом предложенная структура данных демонстрирует наилучшие результаты. Можно отметить, что, начиная с размерности 9, хорошо наблюдается эффект «проклятия размерности». Таким образом, проведенное исследование позволяет расширить возможности применения сеточных алгоритмов кластеризации к данным высокой размерности.

Список литературы

- [1] Сидорова В.С. Гистограммный иерархический алгоритм и понижение размерности пространства спектральных признаков // Журнал Сибирского федерального университета. Техника и технологии. 2017. Т. 10. № 6. С. 714–722.
- [2] Пестунов И.А., Рылов С.А., Бериков В.Б. Иерархические алгоритмы кластеризации для сегментации мультиспектральных изображений // Автометрия. 2015. Т. 51. № 4. С. 12–22.
- [3] RYLOV S. A., PESTUNOV I. A. Fast hierarchical clustering of multispectral images and its implementation on NVIDIA GPU // Journal of Physics: Conference Series. 2018. Vol. 1096. P. 012039.