

Быстрый состоятельный сеточный алгоритм кластеризации

Тарасенко Антон Сергеевич

Институт математики им. С. Л. Соболева СО РАН

tarasenko@math.nsc.ru

Соавторы: Бериков В. Б., Пестунов И. А., Рузанкин П. С., Рылов С. А.

Секция: Теория вероятностей

Предлагается быстрый и состоятельный сеточный алгоритм, который оценивает количество кластеров для наблюдений в R^d и строит их приближения. Временная сложность алгоритма может быть сведена к линейной без потери свойства состоятельности.

Несмотря на то, что сеточные алгоритмы демонстрируют впечатляющую производительность, обеспечивая эффективную обработку больших наборов данных, их эвристическая природа часто оставляет место для неопределенности относительно достоверности их результатов. Теоретическая состоятельность, однако, обозначает способность алгоритма, при определенных условиях, давать корректные оценки как количества кластеров, так и их состава. Помимо теоретического доказательства состоятельности, мы проводим численные симуляции и тесты на реальных наборах данных, чтобы сравнить производительность нового алгоритма с устоявшимися сеточными методами.

Работа выполнена при поддержке Математического Центра в Академгородке, соглашение с Министерством науки и высшего образования Российской Федерации №075-15-2022-281.