

ОЦЕНКА КАЧЕСТВА КЛАСТЕРИЗАЦИИ С ПОМОЩЬЮ ГРАФОВЫХ АЛГОРИТМОВ

Козин И.А., Черних Е.П.

*Национальный технический университет
«Харьковский политехнический институт»,
г. Харьков*

В наше время рост объемов видеоданных с различными свойствами в различных проблемно-ориентированных областях приводит к необходимости анализа существующих и создания новых быстродействующих и надежных средств кластеризации. Оценка качества очень важна для всего процесса кластеризации, так как без неё полученная структура кластеров не может быть сочтена достаточно достоверной для того, чтобы делать из неё определённые выводы или проводить дальнейший анализ на её основе. Поэтому оценка качества кластеризации является актуальной областью исследования.

Решение задачи кластеризации принципиально неоднозначно. Существует несколько причин:

- 1) нет однозначно наилучшего критерия качества кластеризации;
- 2) число кластеров, как правило, неизвестно заранее и устанавливается в соответствии с некоторым субъективным критерием.

Для определения оптимального числа кластеров в работе были рассмотрены алгоритм связанных компонентов и алгоритм послойной кластеризации. При использовании алгоритма связанных компонентов на гистограмме не было точно выраженной кластерной структуры данных. Было большое количество пиков, что затруднило определение внутркластерных и межкластерных расстояний. При таком условии управлять кластерами при помощи порога затруднительно. С помощью послойной кластеризации были выбраны связанные компоненты на разных уровнях и задана нужная глубина получаемых кластеров, что, в свою очередь, ускорило работу кластеров.

Для достижения наилучшего результата необходимо экспериментировать с выбором мер расстояний, а иногда даже менять алгоритм.

Достоинствами графовых алгоритмов кластеризации являются наглядность, относительная простота реализации и возможность внесения различных усовершенствований, основанных на геометрических соображениях.

При использовании любого алгоритма важно понимать его достоинства и недостатки, учитывать природу данных, с которыми он лучше работает.

Литература:

1. Айвазян С.А. Прикладная статистика: классификация и снижение размерности // С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин. – М.: 1989. – 224 с.
2. Шлезингер М. Десять лекций по статистическому и структурному распознаванию // М. Шлезингер, В. Главач. – Киев: Наукова думка, 2004. – 343 с.