
Р. Э. Здор

(ГГУ им. Ф. Скорины, Гомель)

СРАВНЕНИЕ ВЫЧИСЛИТЕЛЬНОЙ СЛОЖНОСТИ АЛГОРИТМОВ КЛАСТЕРИЗАЦИИ ДАННЫХ

Кластеризация (или кластерный анализ) – это задача разбиения множества объектов на группы, называемые кластерами. Внутри каждой группы должны оказаться “похожие” объекты, а объекты разных группы должны быть как можно более отличны.

Алгоритмы иерархической кластеризации. Алгоритм в начале работы помещает каждый объект в отдельный кластер, а затем объединяет кластеры во все более крупные, пока все объекты выборки не будут содержаться в одном кластере. Таким образом, строится система вложенных разбиений. Результаты таких алгоритмов обычно представляются в виде дерева – дендрограммы.

Алгоритм выделения связанных компонент. В алгоритме выделения связанных компонент задается входной параметр R и в графе удаляются все ребра, для которых “расстояния” больше R . Соединенными остаются только наиболее близкие пары объектов. Смысл алгоритма заключается в том, чтобы подобрать такое значение R , лежащее в диапазоне всех “расстояний”, при котором граф “развалится” на несколько связанных компонент. Полученные компоненты и есть кластеры.

Послойная кластеризация. Алгоритм послойной кластеризации основан на выделении связанных компонент графа на некотором уровне расстояний между объектами (вершинами). Уровень расстояния задается порогом расстояния s . Алгоритм послойной кластеризации формирует последовательность подграфов графа G , которые отражают иерархические связи между кластерами.

Таблица 1 – Вычислительная сложность алгоритмов

Алгоритм кластеризации	Вычислительная сложность
Иерархический	n^2
к-средних	nkl , где k – число кластеров, l – число итераций
с-средних	
Выделения связанных компонент	$n^2 \log n$
Послойная кластеризация	$\max(n, m)$, где $m < n(n-1)/2$

ЛИТЕРАТУРА

1. Воронцов, К. В. Алгоритмы кластеризации и многомерного шкалирования / К. В. Воронцов. – СПб.: Питер, 2007. – 752 с.

2. Курс лекций «Data Mining» / Интернет-университет информационных технологий [Электронный ресурс]. – 2005. – Режим доступа: www.intuit.ru/departament/database/datamining/. – Дата доступа: 01.02.2015.