

АГЛОМЕРАЦИОННЫЕ АЛГОРИТМЫ

Методы иерархической кластеризации основываются на двух идеях: *агломерации*, т.е. последовательного объединения индивидуальных объектов или их групп во всё более крупные подмножества, или обратном по смыслу процессе, который начинается с корня и на каждом шаге делит образующие группы по степени их гетерогенности. В обоих случаях результат работы алгоритма представляет собой древовидную структуру [1].

Агломерационные алгоритмы (АА), требуют выявления различий между кластерами. Наиболее распространенными являются максимальная или полная связь, в которой различие между двумя кластерами является максимумом всех пар различий между парами точек в разных кластерах, минимальная или единичная связь или ближайший сосед, в котором различие между двумя кластерами является минимальным для всех этих пар различий, и средняя связь, в которой различие между двумя кластерами является средним или подходящим средневзвешенным значением для всех этих пар различий.

АА начинаются с набора одноэлементных кластеров, состоящих из всех объектов. Совершая агломерации пары кластеров минимальной разности, чтобы получить новый кластер, удалив два кластера из дальнейшего рассмотрения, следует повторять этот этап агломерации, пока не будет получен единственный кластер, содержащий все наблюдения. Множество кластеров, полученных по пути, образует иерархическую кластеризацию.

Иерархическая кластеризация особенно хорошо работает с небольшими наборами данных. А поскольку они не только должны определять наилучший способ объединения кластеров на каждой итерации, но и когда кластеризация завершена, становятся дороже с точки зрения вычисления, поскольку учитывается больше точек данных. Алгоритмы разделения более эффективны, т.к. на каждом временном шаге алгоритму нужно только разбить каждый кластер на два таким образом, чтобы он удовлетворял некоторым критериям.

Литература

1 Нейский, И.М. Классификация и сравнение методов кластеризации / А. В. Егоров, Н. И. Куприянов. – Мн., 2011. – 10 с.