

КЛАСТЕРИЗАЦИЯ ОБЪЕКТОВ

Фищукова Н.В.

Национальный технический университет

«Харьковский политехнический институт», г. Харьков

Как известно, задача группирования (кластеризация) состоит в разделении совокупности объектов на группы таким образом, чтобы объекты каждой из образованных при этом групп были «близки» друг к другу, а объекты, принадлежащие разным группам, возможно более далеки. Перечислим основные, наиболее часто используемые на практике варианты построения алгоритмов кластеризации.

Выбор центров группирования. Центры группирования выбираются произвольно или так, чтобы минимальное расстояние в группе точек, задающих центры группирования, было максимально возможным; строится гиперсфера минимального радиуса, поглощающая все точки, соответствующие исходной совокупности объектов. Центры группирования выбираются на этой гиперсфере таким образом, чтобы минимальное расстояние между ними было максимальным.

Процедура подключения точек к группам. Точка присоединяется к группе, расстояние до центра группирования которой является минимальным или расстояние до центра тяжести которой минимально; для каждой точки вычисляются расстояния до центров тяжести кластеров, выбирается минимальное из них и первой присоединяется та точка, для которой выбранное минимальное расстояние является наименьшим; точка присоединяется к тому из кластеров, среднее расстояние до которого минимально.

Коррекция положения центров группирования. После того, как все точки распределены, вычисляется критерий качества кластеризации, представляющий собой дробь, числитель которой характеризует степень компактности полученных кластеров, а знаменатель определяет расстояние между центрами тяжести двух ближайших кластеров. Наилучшему варианту группирования будет соответствовать минимальное значение критерия.

По результатам полученного распределения осуществляется коррекция исходного положения центров группирования. Основным вариантом проведения коррекции состоит в выборе в качестве новых центров группирования центры тяжести образованных кластеров.

Для оценки эффективности возможных вариантов кластеризации построена имитационная модель. В докладе приводятся результаты сравнения алгоритмов кластеризации и делаются рекомендации.