

АНАЛИЗ КЛАСТЕРНЫЙ (автоматическая классификация, таксономия, распознавание образов) - совокупность многомерных статистических методов, предназначенных для исследования структуры некоторой совокупности объектов, переменных или других единиц анализа.

Анализ структуры объектов, т.е. разделение их на классы (кластеры - см.), производится на основе матрицы расстояний - квадратной таблицы, в которой представлены расстояния между всеми возможными парами объектов в многомерном пространстве переменных. Выбор мер расстояния зависит от типа измерительных шкал; наиболее просто он определяется, если все признаки измерены с помощью однотипных шкал - количественных, порядковых или дихотомических. Для некоррелированных количественных и порядковых переменных обычно используется расстояние Евклида, для коррелированных - расстояние Махалобиса; для дихотомических шкал - расстояние Хемминга (city-block). Иногда вместо мер расстояния используются меры сходства или подобия объектов.

Алгоритмы А.К. можно разделить на иерархические, неиерархические и методы классификации "с обучением". Иерархические методы предполагают последовательное объединение объектов в кластеры по степени их близости друг к другу или, напротив, последовательное разбиение совокупности объектов на все более мелкие кластеры. В этом случае кластерное решение представляет собой иерархическую структуру вложенных друг в друга кластеров.

Неиерархические методы позволяют находить и идентифицировать "сгущения" объектов в пространстве переменных. Кластеризация "с обучением" предполагает, что количество классов известно заранее, и имеется обучающая выборка - набор объектов, для которых известно, к каким классам они принадлежат. Остальные объекты классифицируются по степени их близости к объектам из выборки обучающей (см.).

Результаты А.К. чаще всего представляются графически, в виде дендрограммы ("дерева"), показывающей порядок объединения объектов в кластеры. Интерпретация кластерной структуры, которая во многих случаях начинается с определения числа кластеров, является творческой задачей.

Для того, чтобы она могла быть эффективно решена, исследователь должен располагать достаточной информацией о кластеризуемых объектах. При кластеризации "с обучением" результаты могут быть представлены в виде списков объектов, отнесенных к каждому классу.

Основными преимуществами А.К. являются отсутствие ограничений на распределение

АНАЛИЗ КЛАСТЕРНЫЙ

Автор: словарь
02.09.2007 03:05 -

переменных, используемых в анализе; возможность классификации (кластеризации) даже в тех случаях, когда нет никакой априорной информации о количестве и характере классов; универсальность (кластерный анализ может применяться не только к совокупностям объектов, но также к наборам переменных или любых других единиц анализа).

О.В. Терещенко