



Методология информационно-аналитической работы

Практическая работа №2

Виды анализа, которые входят в сферу аналитики

Доклад на тему:
“Кластерный анализ”

Студенты группы ИВМО-05-22:
Калашников Григорий
Григорьев Валерий



Определение

Кластерный анализ - многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы.

Задача кластеризации относится к статистической обработке, а также к широкому классу задач самообучения.

Входные данные в случае кластерного анализа являются объекты со своими признаками, характеризующие эти объекты.

По этим же признакам происходит анализ и сравнение.

Задачи кластерного анализа



Кластерный анализ выполняет следующие основные задачи:

- Разработка типологии или классификации.
- Исследование полезных концептуальных схем группирования объектов.
- Порождение гипотез на основе исследования данных.
- Проверка гипотез или исследования для определения, действительно ли типы (группы), выделенные тем или иным способом, присутствуют в имеющихся данных.

Независимо от предмета изучения применение кластерного анализа предполагает следующие этапы:

- Отбор выборки для кластеризации. Подразумевается, что имеет смысл кластеризовать только количественные данные.
- Определение множества переменных, по которым будут оцениваться объекты в выборке, то есть признаковового пространства.
- Вычисление значений той или иной меры сходства (или различия) между объектами.
- Применение метода кластерного анализа для создания групп сходных объектов.
- Проверка достоверности результатов кластерного решения.

Сфера применения кластерного анализа



-Биология

-Информатика

-Социология

-Аналитика

-Анализ финансового рынка

-Маркетинг и др.

Методы кластеризации



1. Вероятностный подход. Предполагается, что каждый рассматриваемый объект относится к одному из классов.
2. Подходы на основе систем искусственного интеллекта
3. Логический подход. Построение дендрограммы осуществляется с помощью дерева решений.
4. Теоретико-графовый подход.

Этапы кластеризации





Пример работы кластерного анализа

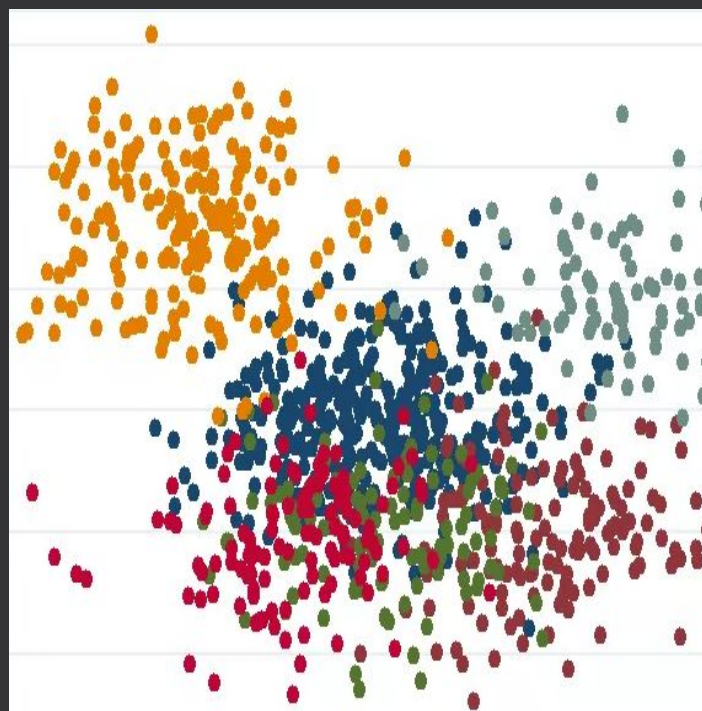


рис. 1 до процедуры кластерного анализа

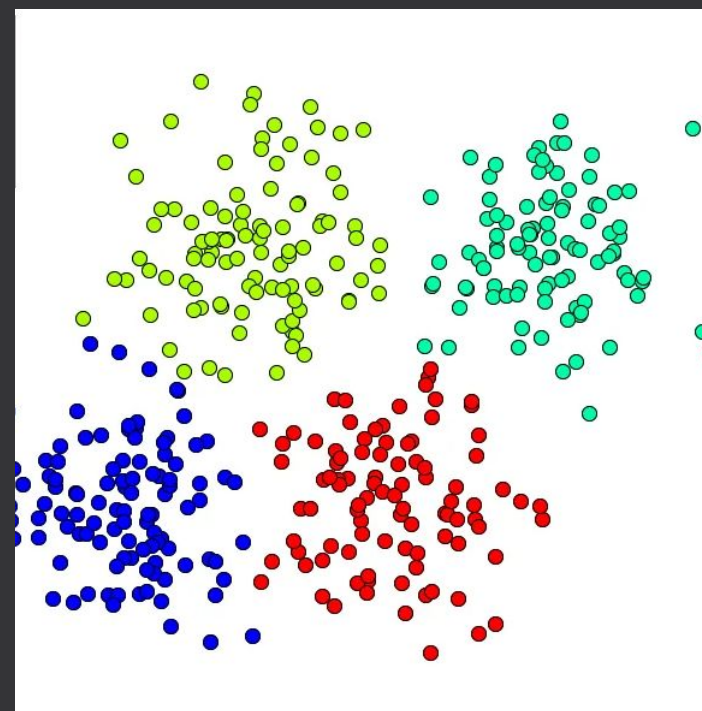


рис. 2 после процедуры кластерного анализа

Цели кластерного анализа



1. Понимание данных путём выявления кластерной структуры.
 - Разбиение выборки на группы схожих объектов позволяет упростить дальнейшую обработку данных и принятия решений, применяя к каждому кластеру свой метод анализа.
2. Сжатие данных.
 - Если исходная выборка избыточно большая, то можно сократить её, оставив по одному наиболее типичному представителю от каждого кластера.
3. Обнаружение новизны.
 - Выделяются нетипичные объекты, которые не удаётся присоединить ни к одному из кластеров.

Выводы

Кластерный анализ или кластеризация - это универсальный, точный аналитический метод анализа данных, задача, которого группировка набора объектов таким образом, чтобы объекты в одной группе (называемой кластером) были более похожи друг на друга, чем объекты в других группах (кластерах). Это основная задача исследовательского анализа данных и распространенный метод статистического анализа данных, используемый во многих областях, включая распознавание образов, анализ изображений, поиск информации, биоинформатику, сжатие данных, компьютерную графику и машинное обучение.



Спасибо за внимание