

А. О. Сахарчук

(ГрГУ им. Я. Купалы, Гродно)

О РАЗРАБОТКЕ СЕРВИСОВ АНАЛИЗА МНОГОМЕРНЫХ ДАННЫХ НА ОСНОВЕ АЛГОРИТМОВ КЛАСТЕРИЗАЦИИ

Под кластерным анализом понимается статистическая функция, работающая с многомерными данными, выполняющая их сбор по некоторым параметрам и упорядочивающая объекты в сравнительно однородные группы. Задача кластеризации относится к статистической обработке данных, а также к классу задач обучения без учителя.

Спектр применений кластерного анализа очень широк: его используют в медицине, маркетинге, социологии и других дисциплинах. Существует около 100 разных алгоритмов кластеризации. Наиболее часто используемые – кластеризация методом k -средних (либо его модификацией: s -средних). Прелесть данного метода – он работает даже тогда, когда данных мало и не выполняются требования нормальности распределений случайных величин и другие требования классических методов статистического анализа.

Чтобы определять «похожесть» объектов нужно составить вектор характеристик для каждого объекта (некий набор числовых значений, например: количество подписчиков аккаунта и количество лайков на ресурсе). Затем необходимо провести нормализацию (Все значения приводятся к некоторому диапазону, например, $[-1, -1]$ или $[0, 1]$). Наконец, для каждой пары объектов измеряется «расстояние» между ними – степень похожести. Существует множество метрик, например: Евклидово расстояние, Манхэттенское расстояние и другие. Реализация алгоритма нечёткой кластеризации (с-средних) состоит из следующих шагов:

- Выбрать начальное нечеткое разбиение n объектов на k кластеров путем выбора матрицы принадлежности U размера $n \times k$.
- Используя матрицу U , найти значение критерия нечеткой ошибки (Примечание: формула нахождения критерия может быть различной, в зависимости от поставленной задачи).
- Перегруппировать объекты с целью уменьшения этого значения критерия нечеткой ошибки.
- Возвращаться в п. 2 до тех пор, пока изменения матрицы U не станут незначительными.

Литература

1 Хайкин, С. Нейронные сети: полный курс / С. Хайкин. – 2-е издание.: пер. с англ. – М. : Издательский дом «Вильямс»: Н.Н. Куссуль, А.Ю. Шелестова, 2006. – 1104 с.