

Задание по теме 4

Склифасовский Денис

Задание:

Составьте краткое описание перечисленных методов кластеризации. Для каждого метода укажите:

1. Четкая или нечеткая кластеризация. В четкой кластеризации каждый элемент входит только в один кластер, в нечеткой может входить в несколько кластеров.
2. Исчерпывающая или неисчерпывающая кластеризация. В исчерпывающей кластеризации каждый объект должен принадлежать хотя бы одному кластеру, в неисчерпывающей могут оставаться "выбросы", не входящие ни в один кластер.
3. Тип алгоритма.

Например:

Алгоритм k-средних относится к итерационным алгоритмам, выполняет исчерпывающую четкую кластеризацию.

Алгоритмы, которые нужно описать:

- k-средних
- c-средних
- expectation-maximisation
- иерархическая кластеризация
- DBSCAN
- самоорганизующиеся карты Кохонена

Решение:

1. k-средних
 - a. Кластеризация: четкая исчерпывающая
 - b. Тип алгоритма: итерационный
 - c. Алгоритм k-средних относится к итерационным алгоритмам, выполняет исчерпывающую четкую кластеризацию. Каждый элемент принадлежит только одному кластеру.
2. c-средних
 - a. Кластеризация: нечеткая исчерпывающая
 - b. Тип алгоритма: итерационный
 - c. Алгоритм c-средних является итерационным методом кластеризации, но в отличие от k-средних он позволяет элементам принадлежать к нескольким кластерам с различной степенью принадлежности.
3. expectation-maximisation
 - a. Кластеризация: нечеткая исчерпывающая
 - b. Тип алгоритма: итерационный

- c. Алгоритм expectation-maximisation выполняет нечеткую исчерпывающую кластеризацию, где каждый элемент имеет вероятностную принадлежность к каждому кластеру.
- 4. иерархическая кластеризация
 - a. Кластеризация: четкая исчерпывающая
 - b. Тип алгоритма: иерархический
 - c. Иерархическая кластеризация строит иерархию кластеров путем последовательного объединения или разделения кластеров. В иерархической кластеризации каждый элемент принадлежит только одному кластеру на каждом уровне иерархии.
- 5. DBSCAN
 - a. Кластеризация: четкая неисчерпывающая
 - b. Тип алгоритма: плотностный (для кластеризации пространственных данных с присутствием шума)
 - c. DBSCAN является алгоритмом кластеризации, основанным на плотности данных. Он определяет кластеры как плотно связанные области в пространстве данных. DBSCAN может также обнаруживать выбросы, элементы, которые не принадлежат ни одному кластеру.
- 6. самоорганизующиеся карты Кохонена
 - a. Кластеризация: нечеткая исчерпывающая
 - b. Тип алгоритма: нейронные сети
 - c. Самоорганизующиеся карты Кохонена (Self-Organizing Maps, SOM) представляют собой тип нейронных сетей, применяемых для кластеризации и визуализации данных. Каждый нейрон представляет прототип кластера. Каждый элемент данных может быть связан с несколькими ближайшими нейронами, представляя нечеткую принадлежность к кластерам.