



K中心聚类

杜逆索



K-中心点聚类算法

1.简介

K中心点聚类算法重复迭代，直到每个代表对象都成为它的簇的实际中心点，或最靠中心的对象。

聚类结果的质量用代价函数来评估，该函数用来度量对象与其簇的代表对象之间的平均相异度。



K-中心点聚类算法

2. K-中心点聚类算法原理

K-中心点聚类算法的基本思想为：选用簇中位置最中心的对象，试图对 n 个对象给出 k 个划分，代表对象也被称为是中心点，其他对象则被称为非代表对象。



K-中心点聚类算法

2. K-中心点聚类算法原理

在K-中心点聚类算法中需要计算所有非选中对象与选中对象之间的相异度作为分组的依据。一般情况下，数据对象为数值型，选用曼哈顿距离：

$$d(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \cdots + |x_{in} - x_{jn}|$$

此处： $i = (x_{i1}, x_{i2}, \cdots, x_{in})$ 和 $j = (x_{j1}, x_{j2}, \cdots, x_{jn})$ 是两个n维的数据对象。



K-中心点聚类算法

2. K-中心点聚类算法原理

K-中心点聚类算法描述:

输入: 簇的数目 k 和包含 n 个对象的数据库

输出: k 个簇, 使得所有对象与其最近中心点的相异度总和最小

- (1) 任意选择 k 个对象作为初始的簇中心点,
- (2) Repeat
- (3) 指派每个剩余对象给离他最近的中心点所表示的簇
- (4) Repeat
- (5) 选择一个未被选择的中心点
- (6) Repeat



K-中心点聚类算法

2. K-中心点聚类算法原理

K-中心点聚类算法描述:

- (7) 选择一个未被选择过的非中心点对象
- (8) 计算用代替的总代价并记录在S中
- (9) Until 所有非中心点都被选择过
- (10) Until 所有的中心点都被选择过
- (11) If 在S中的所有非中心点代替所有中心点后的计算出总代价有小于0的存在, then找出S中的用非中心点替代中心点后代价最小的一个, 并用该非中心点替代对应的中心点, 形成一个新的k个中心点的集合;
- (12) Until 没有再发生簇的重新分配, 即所有的S都大于0.



K-中心点聚类算法

3. K-中心点聚类算法特点及应用

3.1 K-中心点聚类算法特点

优势:

- 对噪声点/孤立点不敏感，具有较强的数据鲁棒性；
- 聚类结果与数据对象点输入顺序无关；
- 聚类结果具有数据对象平移和正交变换的不变性等。

缺点:

- 在于聚类过程的高耗时性。



K-中心点聚类算法

3. K-中心点聚类算法特点及应用

3.2 K-中心点聚类算法应用

- (1) K-中心点算法在暂住人口分析中的应用
- (2) K-中心点算法在软件测试中的应用



K-中心点聚类算法

4. 小结

本章详细地介绍了K-中心点算法的基本概念、基本原理,并介绍了该算法的特点和存在的缺陷,最后介绍了K-中心点算法的应用,从中可以看出K-中心点算法的应用非常广泛。