

聚类分析方法有什么好处

通常，我们在研究与处理事物时，经常需要将事物进行分类，例如地质勘探中根据物探、化探的指标将样本进行分类；古生物研究中根据挖掘出的骨骼形状和尺寸将它们分类；大坝监控中由于所得的观测数据量十分庞大，有时亦需将它们分类归并，获得其典型代表再进行分析等，对事物进行分类，进而归纳并发现其规律已成为人们认识世界、改造世界的一种重要方法。

由于对象的复杂性,仅凭经验和专业知识有时不能确切地分类,随着多元统计技术的发展和计算机技术的普及,利用数学方法进行更科学的分类不仅非常必要而且完全可能。

近些年来,数值分类学逐渐形成了一个新的分支,称为聚类分析,聚类分析适用于很多不同类型的数据集合,很多研究领域,如工程、生物、医药、语言、人类学、心理学和市场学等,都对聚类技术的发展和起到了推动作用。



什么是聚类分析？

聚类分析也称群分析或点群分析,它是研究多要素事物分类问题的数量方法,是一种新兴的多元统计方法,是当代分类学与多元分析的结合。其基本原理是,根据样本自身的属性,用数学方法按照某种相似性或差异性指标,定量地确定样本之间的亲疏关系,并按这种亲疏关系程度对样本进行聚类。

聚类分析是将分类对象置于一个多维空间中,按照它们空间关系的亲疏程度进行分类。

通俗的讲，聚类分析就是根据事物彼此不同的属性进行辨认，将具有相似属性的事物聚为一类，使得同一类的事物具有高度的相似性。

聚类分析方法，是定量地研究地理事物分类问题和地理分区问题的重要方法，常见的聚类分析方法有系统聚类法、动态聚类法和模糊聚类法等。

聚类分析方法有什么好处

聚类分析：将个体（样品）或者对象（变量）按相似程度（距离远近）划分类别，使得同一类中的元素之间的相似性比其他类的元素的相似性更强。目的在于使类间元素的同质性最大化和类与类间元素的异质性最大化。其主要依据是聚到同一个数据集中的样本应该彼此相似，而属于不同组的样本应该足够不相似。

常用聚类方法：系统聚类法，K-均值法，模糊聚类法，有序样品的聚类，分解法，加入法。

注意事项：

1. 系统聚类法可对变量或者记录进行分类，K-均值法只能对记录进行分类；
2. K-均值法要求分析人员事先知道样品分为多少类；
3. 对变量的多元正态性，方差齐性等要求较高。

应用领域：细分市场，消费行为划分，设计抽样方案等

优点：聚类分析模型的优点就是直观，结论形式简明。

缺点：在样本量较大时，要获得聚类结论有一定困难。由于相似系数是根据被试的反映来建立反映被试间内在联系的指标，而实践中有时尽管从被试反映所得出的数据中发现他们之间有紧密的关系，但事物之间却无任何内在联系，此时，如果根据距离或相似系数得出聚类分析的结果，显然是不适当的，但是，聚类分析模型本身却无法识别这类错误。

聚类分析是一种探索性的分析，在分类的过程中，人们不必事先给出一个分类的标准，聚类分析能够从样本数据出发，自动进行分类，聚类分析所使用方法的的不同，常常会得到不同的结论，不同研究者对于同一组数据进行聚类分析，所得到的聚类数未必一致。



聚类分析的意义是什么

聚类分析指将物理或抽象对象的集合分组为由类似的对象组成的多个类的分析过程。它是一种重要的人类行为。

聚类分析的目标就是在相似的基础上收集数据来分类。聚类源于很多领域，包括数学，计算机科学，统计学，生物学和经济学。在不同的应用领域，很多聚类技术都得到了发展，这些技术方法被用作描述数据，衡量不同数据源间的相似性，以及把数据源分类到不同的簇中。

商业：聚类分析被用来发现不同的客户群，并且通过购买模式刻画不同的客户群的特征。聚类分析是细分市场的有效工具，同时也可用于研究消费者行为，寻找新的潜在市场、选择实验的市场，并作为多元分析的预处理。

生物：聚类分析被用来动植物分类和对基因进行分类，获取对种群固有结构的认识

地理：聚类能够帮助在地球中被观察的数据库商趋于的相似性

保险行业: 聚类分析通过一个高的平均消费来鉴定汽车保险单持有者的分组, 同时根据住宅类型, 价值, 地理位置来鉴定一个城市的房产分组

因特网：聚类分析被用来在网上进行文档归类来修复信息

电子商务:聚类分析在电子商务中网站建设数据挖掘中也是很重要的一个方面,通过分组聚类出具有相似浏览行为的客户,并分析客户的共同特征,可以更好的帮助电子商务的用户了解自己的客户,向客户提供更合适的服务。