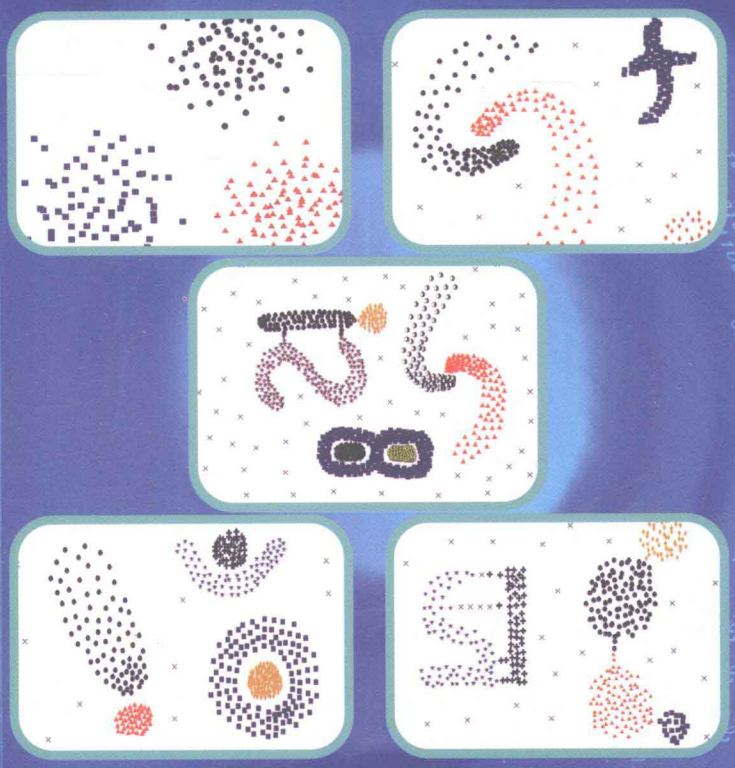
地球观测与导航技术丛书

空间聚类分析及应用

邓 敏　刘启亮　李光强　黄健柏 著



科学出版社

地球观测与导航技术丛书

空间聚类分析及应用

邓　敏　刘启亮　李光强　黄健柏　著

科学出版社

北京

内 容 简 介

空间聚类分析是空间数据挖掘与知识发现的主要手段之一，已广泛应用于地理学、地质学、气象学、地图学、天文学及公共卫生等诸多领域。本书系统阐述了空间聚类分析的理论框架，并对当前国内外空间聚类分析领域研究的主要内容与进展进行了介绍。书中首先阐述了空间聚类分析研究的重要意义，明确了空间聚类分析研究中的基本问题，建立了空间聚类分析的理论框架，并据此对空间聚类分析的各个主要研究内容分别进行阐述，主要包括空间数据清理与聚类趋势分析、空间相似性度量、空间点实体聚类算法、空间面实体与动态轨迹聚类算法及空间聚类有效性评价方法等内容，同时介绍了空间聚类分析方法在地震模式分析、气象、环境、社会经济等领域的具体应用实例。

本书可供地理、地质、测绘、计算机、环境等相关领域的科研人员与研究生阅读参考。

图书在版编目(CIP)数据

空间聚类分析及应用 / 邓敏等著. —北京：科学出版社，2011.10
（地球观测与导航技术丛书）

ISBN 978-7-03-032533-4

Ⅰ.①空… Ⅱ.①邓… Ⅲ.①地理信息系统：空间信息系统－数据采集 Ⅳ.①P208

中国版本图书馆 CIP 数据核字(2011)第 206611 号

责任编辑：孙 芳 / 责任校对：鲁 素
责任印制：赵 博 / 封面设计：鑫联必升

科学出版社出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

新蕾印刷厂印刷

科学出版社发行 各地新华书店经销

*

2011 年 10 月第 一 版 开本：787×1092 1/16
2011 年 10 月第一次印刷 印张：11 3/4
印数：1—2 000 字数：261 000

定价：50.00 元

（如有印装质量问题，我社负责调换）

《地球观测与导航技术丛书》编委会

《地球观测与导航技术丛书》出版说明

地球空间信息科学与生物科学和纳米技术三者被认为是当今世界上最重要、发展最快的三大领域。地球观测与导航技术是获得地球空间信息的重要手段，而与之相关的理论与技术是地球空间信息科学的基础。

随着遥感、地理信息、导航定位等空间技术的快速发展和航天、通信和信息科学的有力支撑，地球观测与导航技术相关领域的研究在国家科研中的地位不断提高。我国科技发展中长期规划将高分辨率对地观测系统与新一代卫星导航定位系统列入国家重大专项；国家有关部门高度重视这一领域的发展，国家发展和改革委员会设立产业化专项支持卫星导航产业的发展；工业与信息化部和科学技术部也启动了多个项目支持技术标准化和产业示范；国家高技术研究发展计划（863 计划）将早期的信息获取与处理技术（308、103）主题，首次设立为“地球观测与导航技术”领域。

目前，“十一五”计划正在积极向前推进，“地球观测与导航技术领域”作为 863 计划领域的第一个五年计划也将进入科研成果的收获期。在这种情况下，把地球观测与导航技术领域相关的创新成果编著成书，集中发布，以整体面貌推出，当具有重要意义。它既能展示 973 和 863 主题的丰硕成果，又能促进领域内相关成果传播和交流，并指导未来学科的发展，同时也对地球观测与导航技术领域在我国科学界中地位的提升具有重要的促进作用。

为了适应中国地球观测与导航技术领域的发展，科学出版社依托有关的知名专家支持，凭借科学出版社在学术出版界的品牌启动了《地球观测与导航技术丛书》。

丛书中每一本书的选择标准要求作者具有深厚的科学研究功底、实践经验，主持或参加 863 计划地球观测与导航技术领域的项目、973 相关项目以及其他国家重大相关项目，或者所著图书为其在已有科研或教学成果的基础上高水平的原创性总结，或者是相关领域国外经典专著的翻译。

我们相信，通过丛书编委会和全国地球观测与导航技术领域专家、科学出版社的通力合作，将会有一大批反映我国地球观测与导航技术领域最新研究成果和实践水平的著作面世，成为我国地球空间信息科学中的一个亮点，以推动我国地球空间信息科学的健康和快速发展！

李德仁

2009 年 10 月

前　言

聚类分析是人类认识客观事物最朴素、最常用的手段之一。早在几千年前的《易经》、《战国策》等经典著作中，就已经出现了聚类分析的思想。聚类分析作为一个专门的研究领域提出至今仅有七十年左右，但其发展速度确实极为惊人，几乎遍及了所有数据处理的领域。伴随着地理信息技术的提出与发展及对地观测能力的迅速提升，1994 年，李德仁院士在国际上首先提出了从空间数据库中挖掘知识的思想，即空间数据挖掘与知识发现。自此，聚类分析在空间信息领域得到了空前的发展，空间聚类分析应运而生并已经成为空间数据挖掘与知识发现的主要技术手段之一。

自 20 世纪 90 年代，第一个专门用于空间聚类分析的算法——CLARANS 被提出以来，空间聚类分析已成为地理信息科学与计算机科学领域共同关注的热点研究问题之一。目前，国内外学者已经在空间聚类分析领域取得了可喜的研究成果，其应用领域也正逐步扩大，但需要注意的是，空间数据的复杂特性（如空间数据的几何形态特征、空间关系、空间数据的相关性、异质性及多尺度特性）导致空间聚类分析研究要比传统的聚类分析研究更加复杂。因此，空间聚类分析并不是传统的聚类分析技术在空间信息领域的简单套用，而需要开展专门的研究。目前，国内外已出版多部有关聚类分析的研究著作，但还没有专门针对空间聚类分析的学术著作出现。为此，结合作者多年来在空间数据挖掘领域的研究心得与成果，尝试撰写一本专门介绍空间聚类分析的学术著作。本书旨在明确空间聚类分析的基本问题与理论框架，同时对现有国内外学者的研究成果进行梳理，并介绍作者在空间聚类分析领域的一些研究成果与见解，希望能够促进空间聚类分析研究的深入与应用。

本书基于空间数据的基本特征、性质及聚类分析的内涵，首先提出了以“空间数据清理与聚类趋势分析-空间聚类算法设计-空间聚类有效性评价”为核心的空间聚类分析研究框架，进而针对具体内容分别进行阐述，同时对空间聚类分析所涉及的专门技术（如空间相似度量）进行了介绍。在内容设置上，为使读者对空间聚类分析的研究领域有较为全面的了解，本书专门回顾了国内外学者针对该领域的各个研究内容所取得的代表性成果，在此基础上，介绍了作者针对新的应用需求所发展的空间聚类分析算法与软件成果。本书从数据和应用两个角度分门别类地对空间聚类算法进行阐述，其中，依据数据的特征介绍了点、线（如动态轨迹）、面的空间聚类算法；并从应用的角度介绍了二维空间实体空间聚类、顾及空间障碍的空间聚类及顾及专题属性的空间聚类；还介绍了大量的算法实例及在地震、气象、环境、社会经济等领域的应用实例，以及自主开发的一款空间聚类分析软件系统 EasyCluster。

本书出版受到了国家 863 计划项目（2009AA12Z206）、教育部新世纪人才支持计划（NCET-10-0831）、中南大学前沿研究计划（2010QYZD002）及中南大学升华育英计划优秀人才资助项目的联合资助。感谢石岩、孙前虎、彭东亮、林雪梅等硕士研究生为本书所

作的部分算法实现与软件开发工作。感谢梅小明老师、赵玲老师、刘慧敏老师、赵彬彬博士、陈杰博士在撰写过程中给予的有益建议。感谢英国伦敦大学程涛教授、王佳璆博士对本书出版的帮助。

本书的出版也得到了中南大学各级领导的关心与支持。感谢中南大学人事处唐忠阳副处长、科技部吴厚平副部长、李启厚副处长、研究生院刘少军副院长、陈立章主任给予的鼓励与帮助！感谢中南大学图书馆馆长、地球科学与信息物理学院副院长朱建军教授、地球科学与信息物理学院副院长柳建新教授、刘兴权教授、邹峥嵘教授等学院领导在本书撰写过程中给予的指导、关心和支持！

空间聚类分析研究方兴未艾，本书的出版仅能起到抛砖引玉的作用。虽然本书撰写过程中力求尽善尽美，但限于作者的学识与经验，不妥之处在所难免，敬请读者批评指正。

作　者

2011 年 6 月

目　录

第1章 绪 论

1.1 空间聚类分析的产生

从人类诞生之日起，其认识、适应、改造自然的步伐就从未停止。聚类与分类是人类认识自然最基本的、最有效的技能之一，在人类社会的发展历程中发挥了重要的作用(Everitt et al.，2001；Anderberg，1973)。当人们试图去认识一类新事物或新现象时，往往会首先采用一定的特征去描述它们，然后根据一定的准则去跟其他已知的事物或现象进行比较(Xu et al.，2009)。例如，自然界大致可以分为动物、植物和矿物三界，动物界下设有纲、目、科、属、种等5个级别，我们通过这样的分类可以很容易发现其共性特征，如鱼类在水中生存，鸟类能够飞翔，同时也可以通过这种分类去推测具体动物的生活习性，如当我们看到喜鹊落在屋顶时，可以自然联想到它在天空中飞行的情景，而不需要看到它在飞行。

历史上，我国劳动人民最早将聚类分析的思想应用到实践中，五千多年前的《易经》就已经提出了“物以类聚，人以群分”的认识思想。公元前4世纪，齐国著名谋士东莱人(今山东省龙口市)淳于髡进一步阐述了这一观点，并通过实例说明了聚类思想的重要价值。在长期的生产实践中，聚类分析的思想一直是以一种经验的形式出现在人类的日常生活中。例如，水果根据颜色、大小分成不同的种类，并且通常可以卖出不同的价钱。聚类分析真正得到空前的发展还要得益于数学方法的引入。1939年，Tryon首次采用聚类分析的思想从相关矩阵中提取互相关的组，标志着聚类分析学科的正式提出。与传统的分类学不同，聚类分析是采用数学工具来研究类的划分及各类间的异同，而传统的分类则多是借助经验或专业知识。在随后的几十年里，聚类分析技术得到了飞速发展，一些沿用至今的经典算法被相继提出，如50～60年代提出的K-means聚类算法在2006年IEEE国际数据挖掘大会组织的评选中被评为十大最具影响力的数据挖掘算法之一(Wu et al.，2008)。聚类分析在70年代初首先由数学地质学家引入我国，中国科学院方开泰等(1982)首先开展了较为系统的研究，并于80年代初编纂出版了我国第一本系统介绍聚类分析技术的学术著作《聚类分析》。

20世纪80年代，伴随着数据库技术与数据采集技术的突破性进展，数据的爆炸性增长，使人们面临了所谓“数据丰富，但信息贫乏”的困境，人们迫切需要强有力的工具从存储在大型数据库中的海量数据获取有用的信息或知识(Han et al.，2005；Tan et al.，2005)。为走出这种困境，数据挖掘技术应运而生。1989年，在美国底特律市召开的第11届国际人工智能学术会议首次提出了“从数据库中发现知识(knowledge discovery in databases，KDD)”的概念。数据挖掘即从数据集中识别出有效的、新颖的、潜在有用的、最终可理解模式的信息，其主要手段包括聚类分析、关联规则挖掘、分类与预测、异常探测

及演变分析等(Han et al.,2005;Fayyad et al.,1996)。聚类分析既可以作为一种独立的数据挖掘工具,又可以与其他数据挖掘方法结合使用,挖掘更深层次的知识,提高数据挖掘效率,其已成为数据挖掘研究中的热点课题。

实际上,人们日常生活所接触和利用的现实世界数据中,大约有80%与地理位置、属性及其空间分布有关(李德仁等,2000)。伴随着计算机、网络、全球定位系统(global positioning system,GPS)、遥感(remote sensing,RS)及地理信息系统(geographical information system,GIS)等技术的迅猛发展和应用,空间数据的数量、复杂性及传输速度都在快速增长,其膨胀速度也极大超过了常规事务型数据(李德仁等,2006),尤其是近十年来对地观测技术与空间数据基础建设的突破性进展,空间数据的爆炸性增长已成为数据处理与分析领域的一个重要特征。空间数据除了具有属性特征外,还具有空间位置特征、空间关系特征和时间特征(王家耀,2001)。因此,与传统的事务性数据相比,空间数据更为复杂,具体表现为空间数据的海量性、空间属性间的非线性关系、尺度特性、模糊性、高维特性及数据的缺值(裴韬等,2001)。传统的空间分析主要是针对空间实体及其属性特征,采用统计分析的手段进行分析,其仅仅考虑了与样本性相关的统计量,没有顾及这些样本在地理空间的分布特征和相互间的位置关系,从而并不非常适用于处理空间相关的数据(Koperski,1999;郭仁忠,1997)。因此,虽然以GIS数据库为主体的空间数据库得到了极大发展,但由于缺乏高效、精确、科学的手段分析这些数据,使得空间数据同样面临了数据丰富而分析不足的尴尬局面,造成了空间数据的极大浪费(马荣华等,2007)。

鉴于数据挖掘技术在事务性数据库中的成功应用,从空间数据中挖掘知识已引起了国内外学者的广泛关注。1994年,李德仁院士在加拿大渥太华举行的GIS国际学术会议上首次提出了从GIS数据库中发现知识的概念,并系统地分析了空间数据挖掘的特点和方法,这标志着空间数据挖掘理论的正式提出(Li et al.,1994)。国际上,Han、Miller、Ester及Shekhar等学者也对空间数据挖掘开展了较早且持续的研究(Miller et al.,2009;Shekhar et al.,2005,2003;Ester et al.,2000,1997;Han et al.,1997)。1994年,Ng和Han提出了一个专门针对空间数据库的聚类算法——CLARANS(clustering large applications based upon RANdomized search),这标志着空间数据挖掘研究的正式兴起。空间数据挖掘是从空间数据库中提取隐含的、用户感兴趣的空间和非空间的模式、普遍特征、规则和知识的过程,已成为数据挖掘领域的一个崭新分支(邸凯昌,2000)。空间聚类、空间关联规则挖掘、空间分类、空间演变与预测、空间异常探测等构成了空间数据挖掘的主要研究内容。因此,空间聚类是空间数据挖掘的一个主要研究方向,也是传统聚类分析技术在空间数据库中的进一步应用,尤其是从空间数据中挖掘聚集模式符合人类对客观世界的认知方式,可以显著提高人们对空间数据的分析和认识水平。

1.2 空间聚类分析的研究概况与基本问题

1.2.1 空间聚类分析的研究概况

空间聚类分析既可以发现隐含在海量数据中的聚类规则,又可以与其他的空间数据挖掘方法结合,挖掘更深层次的知识,提高空间数据数据挖掘的效率和质量(杨春成,

2004)。空间实体的自然聚集现象经常反映一定的规律或趋势。《战国策·齐策三》中,淳于髡在解释"物以类聚,人以群分"时用到这样一个例子:"人们要寻找柴胡、桔梗这类药材,如果到水泽洼地去找,恐怕永远也找不到;要是到梁文山的背面去找,那就可以成车地找到。"柴胡、桔梗这种聚集现象实际上反映了周围环境的特殊性,继而对于人们认识、利用这种特性提供了重要的指示作用。1854 年,琼·斯诺博士采用空间聚集分析的手段发现伦敦霍乱病起源的案例堪称空间聚类分析最早的成功应用(Miller et al.,2009),当琼·斯诺博士将霍乱病死者居住位置标注在一张 1∶6500 比例尺的城区地图上后(图 1.1),发现死者大多集中在一口名为"布洛多斯托"的水井附近(图 1.1 中圆圈处),当关闭这口井后,新的霍乱病例也就没有再出现。

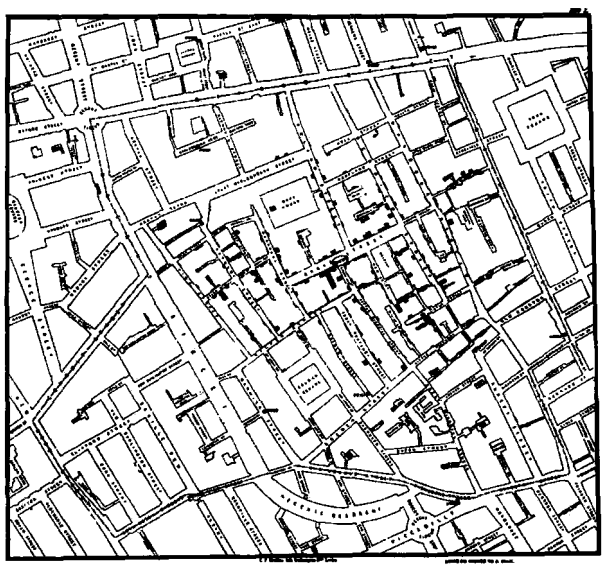


图 1.1　1854 年伦敦霍乱病空间分布图

空间聚类分析的一个重要作用在于能够发现空间实体自然的空间集聚模式,对于揭示空间实体的分布规律、提取空间实体的群体空间结构特征、预测空间实体的发展变化趋势具有重要的作用。进一步,结合空间实体的非空间属性在空间上的分布与差异,对于解释复杂的地理现象具有重要的意义。在城市规划领域,空间聚类在公共设施选址中具有明显的优势,并且已经得到了成功的应用(Liao et al.,2008;毛政元等,2004)。在制图综合领域,空间聚类已被广泛应用于点群特征简化、点群空间特征提取、建筑物聚合操作及等高线简化(武芳等,2008;Qi et al.,2008;Li,2007;郭庆胜等,2007;卢林等,2005)。在地震分析领域,空间聚类在提取地震空间分布特征及地质构造方面也体现出了独特的优势(Pei et al.,2009,2006;Xu et al.,1998)。在地价评估领域,空间聚类技术已被成功用于地价的分级(焦利民等,2009;邓羽等,2009)。在图像处理领域,空间聚类方法同样成功应用于遥感影像分类、分割研究中(秦昆等,2008;骆剑承等,1999;Sander et al.,1998)。在全球气候变化研究领域,借助空间聚类手段发现对陆地气候具有显著影响的极地、海洋大气压力模式、海表气温分布对于理解全球气候具有重要的价值(Birant et al.,2007;Tan et al.,2005)。在公共安全领域,犯罪热点分析是空间聚类分析对社会安全的又一个贡献,

可以有力地帮助警察对地方治安维护做出决策(Estivill-Castro et al.,2002)。近年来,空间动态轨迹聚类已成为空间聚类技术的一个新的应用,借助空间聚类技术可以发现热带风暴等空间轨迹数据的空间分布模式,这对于理解局部气候变化具有重要的意义(Lee et al.,2007)。

空间聚类不仅可以单独作为一种数据分析的手段,而且还可以作为其他空间数据挖掘方法的重要基础。例如,采用空间聚类分区预处理后构建神经网络的精度比全局构建的神经网络具有更强的预测能力(李光强等,2009;王海起等,2008)。空间聚类结果可以作为空间关联分析的输入来挖掘空间关联规则(Malerba et al.,2002;Koperski et al.,1995)。例如,采用空间聚类分析获取居民区边界,再分析聚类边界与高尔夫球场的邻接关系,继而预测房屋价格走向(Knorr et al.,1996)。空间聚类分析也可以用来指导空间分类,建立分类模型,再对遥感影像进行分类,其效率和精度将大大提高(Cihlar et al.,2003;Faber,1994)。空间聚类也是挖掘空间异常的一种有力手段(邓敏等,2010;李光强等,2008;林甲祥等,2008)。

此外,空间聚类算法在大多数情况下可以直接或稍加修改后进行传统事物型数据的聚类分析,在智能计算、机器学习、模式识别、生物学、心理学、信息检索、经济学等领域进行应用。鉴于聚类分析的巨大应用潜力,当前针对聚类分析的研究已经进入了高速发展的时期。据统计,在过去 10 年间,全球有超过 12000 篇期刊或会议论文在题目、摘要或关键词中包含聚类分析的字眼(Xu et al.,2009),同时,这些论文涉及的范围也是极其广泛的,囊括了超过 200 个主要学科及 3000 多种杂志。从 1996 年到 2006 年的 10 年间,与聚类分析有关的论文几乎呈指数增长(图 1.2)。此外,目前有近 80 种主要期刊在刊登关于聚类分析及空间聚类的文章,近 50 个国际会议接收聚类分析有关的论文(Gan et al.,2007)。作为聚类分析研究的一个重要分支,空间聚类分析已成为地球信息科学与计算机科学领域共同关注的热点。

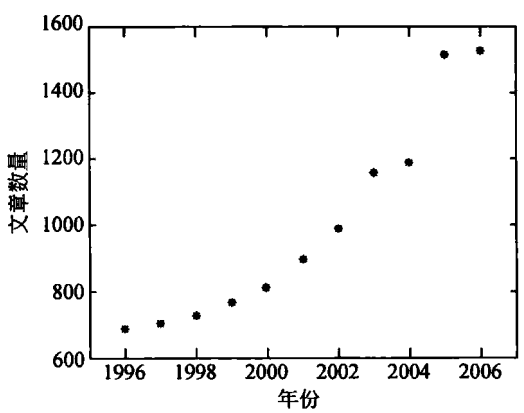


图 1.2 1996～2006 年每年发表的与聚类分析有关的科技论文(Xu et al.,2009)

1.2.2 空间聚类分析的定义

空间聚类分析是传统聚类分析研究的一个延伸与发展,因此,有必要回顾聚类分析的

相关定义。目前，尚缺乏一个正式的、公认的聚类分析定义，下面选取了4种较有代表性的定义：

(1) 聚类分析旨在将一组实体依据一定的相似性度量准则划分成一系列较为均匀的子类，同一类中实体间的相似度要尽可能大于不同类间的实体(Bacher et al.，1981)。

(2) 聚类分析旨在发现 K 个簇或者一种包含 K 个簇的划分方式，相同类中的实体互相相似，而不同类中实体是不相似的(Bacher，1996)。

(3) 聚类分析的目的在于将一个有限的、未标记的数据集分解成一系列有限的、自然的潜在数据结构，而不需要提供一个由同概率分布获得的未观测样本的精确分类(Baraldi et al.，2002；Cherkassky et al.，1998)。

(4) 将物理或抽象对象的集合划分成相似对象的过程称为聚类(Han et al.，2005)。

首先要明确一个非常容易混淆的概念，即聚类与分类，Cherkassky 等(1998)对聚类分析的定义很好地回答了这个问题。分类系统分为监督分类与非监督分类两种，两者的根本区别在于是否将实体分配到一个预先设计好的分类系统中去。聚类分析不需要预设的分类系统，因此属于非监督分类；而我们通常所熟悉的分类则属于监督分类的范畴。综合上述定义，可以给出一个更为完整的空间聚类分析定义，即空间聚类旨在将一组具有相关性的空间实体依据一定的相似性度量准则划分成一系列由若干空间实体构成的、具有一定意义的空间簇，同一空间簇中实体尽可能相似，不同空间簇内的实体尽可能相异。可见，空间聚类分析的定义并不是传统聚类分析定义的简单套用，两者之间具有明显的差别，其集中体现在实体的定义、相似性定义及类的定义三个方面，下面将分别进行比较分析。

在传统的聚类分析中，对象的概念有多种不同的说法，如数据、元组、记录、观测资料、项目等，通常是对应了数据库中一条包含多个属性的记录，习惯性地抽象为空间中的一个点(Gan et al.，2007)，当属性维数超过3后，这种抽象就失去了物理意义，因此，聚类结果也很难可视化。然而，在地理空间中，空间对象(亦称实体)总是有明确的物理意义的，并具有一定的空间位置，通常采用特定的坐标表示，同时也具有几何特征，即大小、形状、分布等。属性特征是附加于空间实体之上的，一般称之为专题属性。因此，不管属性的维数多高，空间实体本身总是可以在地理空间中唯一表示，故空间聚类的结果总是可以进行可视化表达。同时需要指出的是，不是所有的空间数据库都可以运用空间聚类方法，空间实体进行空间聚类的先决条件是它们之间存在一定的相关性。空间聚类必须在满足地理学第一定律(Tobler，1970)的前提下才能进行，即空间实体之间具有一定的依赖关系。相似性的定义在聚类分析中起关键作用，在传统的聚类分析中多采用各种距离、相关系数等度量实体间的相似性；而在空间聚类中，相似性的定义包含了两方面的意义：一种是属性上的相似，这与传统的聚类分析类似；另一种是空间关系上的相似，即要求空间实体在位置上接近或相邻，这是传统的聚类分析所不考虑的。簇也称为组、类，当前还没有一个公认的定义，但从根本上簇的定义是基于相似性的。定义簇的主要原则是要求同一个簇中的实体要尽可能相似，而不同簇中的实体要存在较大差异，内部均匀性和外部分离性是簇的主要特征(Everitt et al.，2001；Gordon，1999；Hansen et al.，1997)。空间聚类中，簇的定义的一个重要特点是要顾及空间关系与空间自相关，即实体间必须满足直接或间接的邻近关系，同时还要求簇内实体要满足空间自相关的条件，对空间不相关的实体进行聚类是

没有意义的。

可以将空间聚类形式化描述为：令 $S=\{S_1,\cdots,S_i,\cdots,S_n\}$ 表示一组具有空间相关性的空间实体集合；$S_i=\{s_{i1},\cdots,s_{ij},\cdots,s_{in}\}$ 表示空间实体的特征向量；s_{ij} 表示空间实体 i 的一维属性；空间聚类获得 K 个空间簇，$S=C_1\cup C_2\cup\cdots C_i\cdots\cup C_k$，$C_i=\{S_{i1},\cdots,S_{ij},\cdots,S_{iz}\}$；Similar($S_{mi}$，$S_{nj}$)表示第 m 个空间簇中第 i 个实体与第 n 个空间簇中第 j 个实体的相似度。因此，对于空间聚类结果 $C_1,\cdots,C_i,\cdots,C_k$，需满足下列条件：

(1) $\bigcup_{i=1}^{k}C_i=S$。

(2) 对于 $\forall C_m,C_n\subseteq S,m\neq n$，需要同时满足：

① $C_m\cap C_n=\varnothing$(仅针对硬聚类)；

② $\mathrm{MAX}_{\forall s_{mi}\in C_m,\forall s_{nj}\in C_n}(\mathrm{Similar}(S_{mi},S_{nj}))<\mathrm{MIN}_{\forall s_{mx},s_{my}\in C_m}(\mathrm{Similar}(S_{mx},S_{my}))$。

根据空间实体特征向量 S_i 的特点，又可以将空间聚类区分为以下三种类型：

(1) S_i 仅包含了空间位置属性。

(2) S_i 既包含空间位置属性，又包含专题属性。

(3) S_i 仅包括专题属性。

第(1)种类型的空间聚类分析可以用来发现空间实体的空间分布模式与规律；第(2)种类型综合考虑了空间位置与专题属性特征的双重意义，可以用于发现更深层次的地学规律；第(3)种类型仅考虑了专题属性的差异，需要结合空间实体的空间分布进行分析，在很大程度上退化为传统的聚类分析手段，在本书中将不做详细讨论。因此，本书将重点研究前两种类型的空间聚类分析方法。

1.2.3 空间聚类分析的基本框架

根据空间聚类的定义，一个完整的空间聚类分析过程包括以下 6 个部分：空间数据清理、空间聚类趋势分析、属性提取与相似性度量、空间聚类算法选择与设计、空间聚类有效性评价、空间聚类结果解释与应用。具体流程如图 1.3 所示。

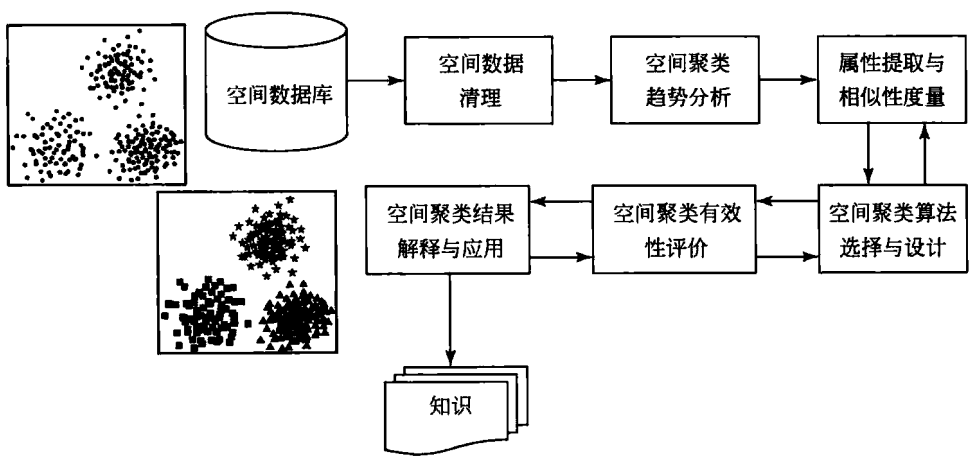


图 1.3 空间聚类的基本流程