土地空间信息技术

目录

**[第四章](#bookmark76)****[土地空间大数据与决策分析](#bookmark76)****[156](#bookmark76)**

[4.2 土地空间大数据技术方法 158](#bookmark81)

[4.3.1 数据描述与探测方法 159](#bookmark82)

[4.3.2 相关性分析方法 160](#bookmark83)

[4.3.3 空间格局分析方法 160](#bookmark84)

**4.2** **土地空间大数据技术方法**

通过整合多源异构数据，运用先进的数据分析方法，如数据描述与探测、相关性分析、 空间格局、时序分析、时空融合、预测与分类等（如图 4-1 所示），揭示土地规划、土地利用、土地保护的模式和趋势，发挥在土地适宜性评价、城市发展模拟、生态环境评估等方面的重要作用。基于这些分析结果构建决策模型，为土地开发、利用、规划、整治和保护业务提供科学依据。其中， 数据描述与探测类方法用于概括和展示时空数据的基本特征及潜在模

式，如描述性统计分析和可视化分析；相关性分析类方法旨在研究时空数据中变量间的关联 关系，包括线性或非线性相关关系，通过相关系数等手段确定其关联程度和方向；空间格局 分析类方法着重于剖析空间数据的分布格局，空间插值、空间自相关分析以及空间聚类分析 等方法可揭示空间上的聚集、分散或分类情况；时序分析类方法专注于时间序列数据的动态 变化，涵盖趋势分析、季节性分析和周期性分析，用于识别时间上的增减趋势、季节性波动 以及周期性规律；时空融合分析类方法综合考虑空间和时间信息，如时空插值、时空自相关 分析和时空聚类分析，以更好地把握时空数据的整体分布规律和相关性；预测与分类方法则 用于对时空数据进行分类和预测，如决策树算法、支持向量机、神经网络、聚类算法以及关 联规则挖掘等，可依据历史数据对未来情况进行推断和分类。

描述性统计分析

数据描述与探测

可视化分析

线性相关关系

相关性分析

非线性相关关系

空间插值

空间自相关分析

土 地 大 数 据 分 析 方 法

空间格局

空间聚类分析

趋势分析

季节性分析

时序分析

周期性分析

时空插值

时空自相关分析

时空融合

时空聚类分析

决策树算法

支持向量机

神经网络

预测与分类

聚类算法

关联规则挖掘

图 4-1 土地空间大数据分析方法

**4.3.1** **数据描述与探测方法**

查找位置用于确定符合多种指定标准的区域。通过按顺序排列的要素（称为轨迹），确 定哪些要素是感兴趣的事件。它由启用时间的点图层组成，以查找对象在指定距离和持续时 间内停留的位置，确定数据在空间模式方面是否存在任何统计显著性的聚类。OD 流（Origin- Destination Flow）分析方法使用点的启用了时间的图层，该图层在时间上可表示时刻。选 择标识字段以及要进行 OD 流分析的时间段，对输入数据进行该时段的OD 流分析。

地理大数据异常探测旨在发现非观测误差导致的、偏离整体或局部分布特征的少部分地 理实体。传统的空间数据异常探测主要是针对时空位置信息与专题属性信息的异常数据识 别。空间大数据异常探测将位空间邻近性度量的概念扩展至“动态流空间”与“多维场景空 间”，从地理流过程相似性、地理场景综合相似性等多视角刻画地理实体在演化过程中的突 变程度。在时空数据中，异常模式通常代表着事物发展的某种特殊规律，在现实生活中更能 引起人们的兴趣。当前空间大数据异常检测方法主要可以分为基于图形、基于距离、基于密

度、基于图论、基于智能学习、基于空间聚类等方法（邓敏等, 2016）。

**4.3.2** **相关性分析方法**

空间大数据关联关系挖掘旨在定量或定性地发现地理实体间的联系。定量挖掘方面，回 归分析是其中的代表性方法，近年来学者们在如何建模时空相关性、异质性与变量间非线性 关系等方面取得了一些重要进展，例如：地理加权回归、时空回归克里金等。

其中，地理加权回归（GWR）模型引入对不同区域的影响进行估计，能够反映参数在不 同空间位置的变化而变化，其结果更符合客观实际。GWR 扩展了传统的回归框架，在全局回 归模型的基础上进行局部的参数估计，模型结构如下：

yi = β(ui, vi ) + ∑kβk (ui, vi )xik + εi （4-1）

式中，(ui, vi )是第i个样本空间单元的地理中心坐标，βk (ui, vi )是连续函数βk (u, v)在i样 本空间单元的值。

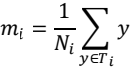
定性挖掘方面，关联规则挖掘是其中的代表性方法，旨在发现在相邻时空区域内频繁出 现的地理对象关系。当前地理大数据关联规则挖掘多基于邻近性与频繁性思路发现连续时空 区域内的关联关系，常用的算法可以大致分为：基于空间拓扑关系的挖掘算法、基于空间距 离关系的挖掘算法、基于空间方向关系的挖掘算法三类（陆新慧等, 2013）。

**4.3.3** **空间格局分析方法**

空间大数据聚类分析旨在依据地理实体间的相似性，将地理实体划分为一系列内部相似、 外部相异的簇结构。空间聚类是描述地理现象空间依赖性的重要手段。空间依赖性是地理现 象的内蕴特征，其表现为邻近空间数据间通常表现出较强的相似性，对于研究地理现象的产 生机理、分布规律及发展变化趋势具有重要的价值。空间聚类可以单独作为一种空间依赖性 的探索性分析工具。例如，在公共卫生领域，空间聚类已经被广泛用于疾病爆发模式识别； 在犯罪学领域，空间分析已经成为犯罪热点探测的主要工具；在地质学领域，空间聚类已经 被成功应用于地震分布规律的探测。

（1）划分聚类方法

划分聚类方法的核心步骤在于通过不断优化基于簇内距离定义的目标函数，将空间数据 划分为给定数目的空间簇，常用算法包括：K-means、ISODATA、FCM 等。其中，K-means 算 法在各个领域得到了很好地应用，其算法基础是误差平方和准则，若Ni 是第 *i*聚类Ti 中的样 本数目，mi 是Ti 中样本的均值，即

 （4-2）

把Ti 的各样本 *y* 与均值mi 的误差平方和对所有类相加为

Je = yi IIy - mi II2 （4-3）

Je 就称为误差平方和聚类准则，使Je 极小的聚类即误差平方和准则下的最优结果。

K-means 算法具体步骤如下：

① 选 *K*个初始聚类中心，z1 (k) ，z2 (k),…, zk (k)，其中 *k*为迭代运算的次序号，

初始时 *K=1*。

② 逐个将模式样本 *x*按最小距离准则分配给 K 个聚类中心中的某一个zj (k)。

1) j = argi min IIx - zi (k)II2 （4-4）

③ 计算新的聚类中心，zi (k + 1), 1 ≤ j ≤ K，

2）  （4-5）

其中，Nj 为第 *j*个聚类域sj 中所包含的样本个数。

④ 若zj (k + 1) = zj (k), 1 ≤ j ≤ K，则迭代结束，否则K = K + 1，返回①。

（2）层次聚类算法

层次聚类算法的核心内容在于定义簇间距离，不断凝聚或分裂获得指定数目的空间簇， 常用算法包括：最近距离、Delaunay 三角网、最小方差距离等。其中，Delaunay 三角网的 应用领域非常广泛，根据其三角网构建过程的不同，生成 Delaunay 三角网的算法可以分为 逐点插入法、三角网生长法和分治法三种方法。逐点插入法的基本步骤如下：

① 构建初始大三角形；

② 随机排列点集 P 中的所有点 P0 ， P1，···, Pn-1；

③ 对点集 P 中的点 Pr ，按以下步骤执行插入操作：定位包含点 Pr 的三角形 PiPjPk；

如果 Pr 在三角形 *PiPjPk* 内部，则连接点 Pr 和三角形 PiPjPk 的各个顶点，将其剖分成三个子三 角形；如果 Pr 在三角形 PiPjPk 的边 PiPj 上，则连接 Pr 与共边 PiPj 的两个三角形的第三个点， 剖分形成四个子三角形；通过边交换方法规格化三角剖分；移除包含大三角形任意顶点的所 有三角形。

（3）密度聚类算法

密度聚类算法的核心在于发现空间实体聚集的高密度区域，常用算法包括： DBSCAN、 OPTICS、STSNN 等。其中，DBSCAN 方法克服了早期方法难以准确描述空间簇形态的问题，实 现了较好的效果。

目前，空间聚类已经成为空间统计、空间数据挖掘、图像模式识别等多个领域研究的核 心研究内容，并已经取得了一些代表性研究成果。然而，由于空间聚类的尺度依赖特征建模 能力的不足，导致聚类结果存在“无中生有”、“无从理解”的困境尚待解决。