

引用格式: 邓敏, 蔡建南, 杨文涛, 等. 多模态地理大数据时空分析方法[J]. 地球信息科学学报, 2020, 22(1): 41-56. [ Deng M, Cai J N, Yang W T, et al. Spatio-temporal analysis methods for multi-modal geographic big data[J]. Journal of Geo-information Science, 2020, 22(1): 41-56. ] DOI: 10.12082/dqxxkx.2020.190491

# 多模态地理大数据时空分析方法

邓 敏<sup>1</sup>, 蔡建南<sup>1\*</sup>, 杨文涛<sup>2</sup>, 唐建波<sup>1</sup>, 杨学习<sup>1</sup>, 刘启亮<sup>1</sup>, 石 岩<sup>1</sup>

1. 中南大学地理信息系, 长沙 410083; 2. 湖南科技大学地理空间信息技术国家地方联合工程实验室, 湘潭 411100

## Spatio-temporal Analysis Methods for Multi-modal Geographic Big Data

DENG Min<sup>1</sup>, CAI Jiannan<sup>1\*</sup>, YANG Wentao<sup>2</sup>, TANG Jianbo<sup>1</sup>, YANG Xuexi<sup>1</sup>, LIU Qiliang<sup>1</sup>, SHI Yan<sup>1</sup>

1. Department of Geo-information, Central South University, Changsha 410083, China; 2. National-Local Joint Engineering Laboratory of Geospatial Information Technology, Hunan University of Science and Technology, Xiangtan 411100, China

**Abstract:** Multi-modal spatio-temporal analysis is aimed at discovering valuable knowledge about the spatio-temporal distributions, associations and revolutions underlying the multi-modal geographic big data. It is a core task of the pan-spatial information system, and is expected to facilitate the study of relationship between human and space. With emerging opportunities and challenges in an era of geographic big data, we systematically summarized four main methods for spatial-temporal analysis based on previous study, including spatio-temporal cluster analysis, spatio-temporal outlier detection, spatio-temporal association mining and spatio-temporal prediction. We discussed the challenges when applying the four methods in multi-scale modeling, multi-view fusion, multi-characteristic cognition, and multi-characteristic expression for spatial-temporal analysis. First, two types of scales (including data scale and analysis scale) are of great importance in the spatio-temporal clustering task. Given the data scale, the best analysis scale for detecting spatio-temporal clusters can be determined using a permutation test method by evaluating the significance of clusters. Second, in the spatio-temporal outlier detection method, the cross-outliers in the context of two types of points are known as the abnormal associations between different types of points and the validity of cross-outliers is assessed through significance tests under the null hypothesis of complete spatial randomness. Third, in the spatio-temporal association mining method, the multi-modal distribution characteristics of each feature quantitatively described in the observed dataset are employed to construct the null hypothesis that the spatio-temporal distributions of different features are independent of each other, and then the evaluation of spatio-temporal associations is modeled as a significance test problem under the null hypothesis of independence. Finally, in the spatio-temporal prediction model, the effects of multiple characteristics of spatio-temporal data (e.g., spatio-temporal auto-correlation and heterogeneity) on the prediction results are fully considered using a space-time support vector regression model. These methods can reveal the geographic knowledge in a more comprehensive, objective, and accurate way, and play a key role in supporting the smart city applications, such as meteorological and environmental monitoring, public safety

收稿日期: 2019-09-04; 修回日期: 2019-11-25.

基金项目: 国家自然科学基金重点项目(41730105); 国家重点研发计划项目(2016YFB0502303)。[ **Foundation items:** National Natural Science Foundation of China, No.41730105; National Key Research and Development Foundation of China, No.2016YFB0502303. ]

作者简介: 邓 敏(1974—), 男, 江西临川人, 教授, 博士生导师, 从事时空数据挖掘与信息服务研究。E-mail: dengmin@csu.edu.cn

\*通讯作者: 蔡建南(1992—), 男, 江苏靖江人, 博士, 从事时空关联模式挖掘与时空统计方法研究。E-mail: jiannan.cai@csu.edu.cn

management, and urban facility planning. For example, the spatio-temporal clustering method can be used to identify the meteorological division, the spatio-temporal outliers can contribute to the detection of the abnormal distribution of urban facilities, the spatio-temporal association mining method can help discover and understand the relationship among different types of crimes, and the spatio-temporal prediction method can be employed to predict the concentration of air pollutants.

**Key words:** pan-spatial information system; geographic big data; multi-model characteristics; spatio-temporal analysis; spatio-temporal clustering; spatio-temporal outlier detection; spatio-temporal association mining; spatio-temporal prediction

**\*Corresponding author:** CAI Jiannan, E-mail: jiannan.cai@csu.edu.cn

**摘要:**多模态地理大数据时空分析旨在融合地理大数据的多模态信息发现有价值的时空分布规律、异常表现、关联模式与变化趋势,是全空间信息系统的核心研究内容,并有望成为推进地理学人地关系研究的重要突破口。为应对地理大数据时代的新机遇与挑战,本文围绕4类核心的时空分析方法(时空聚类分析、时空异常分析、时空关联分析与时空预测分析),系统归纳了国内外研究现状,探讨了时空分析中多尺度建模、多视角协同、多特征认知与多特性表达的研究难点。进而,介绍了多模态地理大数据时空聚类、异常、关联与预测分析模型,更加全面、客观、精准地认知与理解时空大数据中潜在的地理知识,并且能够在气象环境监测、公共安全管理、城市设施规划等多个应用领域发挥关键作用。

**关键词:**全空间信息系统;地理大数据;多模态特征;时空分析;时空聚类分析;时空异常分析;时空关联分析;时空预测分析

## 1 引言

空间分析旨在借助计算机与数学手段量化分析地理数据的空间关系与模式,是地理信息系统区别于一般管理信息系统的特有功能<sup>[1]</sup>。20世纪80年代时态GIS<sup>[2]</sup>的提出进一步推动了时空分析的研究,通过在空间分析中结合地理数据的时间属性,发现地理数据中潜在的、有价值的时空模式、关联关系及其随时间的演变趋势<sup>[3]</sup>,对于理解复杂的地理学规律具有重要的科学价值。近年来,随着对地观测传感器技术(如卫星遥感、激光雷达、航测航拍等)以及人类活动感知手段(如社交签到、导航出行、手机定位等)日新月异的飞速发展,地理数据已经明显呈现出大数据的典型“5V”特征(即Volume、Velocity、Variety、Value和Veracity<sup>[4]</sup>),使得人们能够更加精细、全面、多样、实时地刻画涵盖自然与人文地理要素的地球表层系统<sup>[5]</sup>。地理大数据与其他大数据的主要区别在于其特有的时空属性,并且时空属性信息的时空分析方法受到了国内外学者的广泛关注<sup>[6-8]</sup>,有望成为地理大数据时代揭示人地关系的重要突破口<sup>[9]</sup>。

新一代地理信息系统的时空分析与传统时空分析的典型区别在于其分析对象的多模态特点,即多粒度、多类型、多参考系、多元关联、多维动态和多能自主<sup>[10]</sup>。多模态地理大数据除了一般大数据普遍存在的“5V”表象特征之外,亦具有“5度”内在

特征<sup>[11]</sup>,即时空粒度细、时空广度宽、时空密度大、时空偏度重与时空精度低。地理大数据的前3“度”特征,使得时空分析由粗粒度、单尺度、低密度的单模态小数据分析,向细粒度、多尺度、高密度的多模态大数据分析发展成为了可能<sup>[12-13]</sup>。然而,由于地理大数据获取过程的有偏性和自发性,面向单模态小数据的传统采样与误差理论将难以在多模态地理大数据时空分析模型中对后2“度”特征进行有效控制<sup>[14]</sup>。面对这些机遇与挑战,如何扬长避短,合理利用地理大数据,对于提升多模态地理大数据时空分析方法的性能至关重要。为此,本文首先对地理数据时空分析的当前研究进行回顾与总结,剖析地理大数据时代下时空分析面临的挑战,介绍多模态地理大数据时空分析的关键技术,展望未来研究重要的发展方向。

## 2 地理数据时空分析的现状与挑战

新时期地理学的核心使命在于探究自然与人文要素的时空分异规律及其相互作用关系,进而预测地球表层系统的时间演变过程<sup>[5]</sup>。为了全力完成新时期所赋予的核心使命,迫切需要4类时空分析的理论、技术与方法支撑,包括时空聚类分析、时空异常分析、时空关联分析与时空预测分析。时空聚类分析与时空异常分析分别用于揭示地理实体时空分布模式中存在的聚集规律与异常表现,时空关

联分析意在刻画不同地理实体间的时空交互关系,而时空预测分析旨在建模地理实体时空属性未来发展变化的演变趋势<sup>[15-17]</sup>。虽然每类时空分析方法面向不同知识类型,但在具体地学应用中,可以通过不同知识之间的彼此增益,为解决综合性地理问题提供系统方案。例如,在空气污染时空分析案例中,首先借助时空聚类与异常分析方法提取空气污染物浓度的时空分区与突变结构,为构建局部预测模型提供基础<sup>[18]</sup>;进而,基于时空关联分析方法深入挖掘空气污染物浓度影响因子的作用机制,辅助预测模型的协变量选择<sup>[19-20]</sup>;最后,在空气污染浓度的时空分布与关联知识的基础之上,建立多变量自适应协同的局部时空预测模型,以实现空气污染浓度的高精度预测<sup>[21]</sup>。下面对4类核心时空分析方法的地理内涵、发展脉络与存在问题进行回顾与总结。

## 2.1 时空聚类分析

时空聚类分析旨在发现地理空间中各类实体的时空分布格局与规律,对于揭示地理实体或地理现象的本质特征、相互依赖关系和演化趋势具有重要的指导意义。根据聚类方法是否顾及非空间专题属性,可以将当前时空聚类分析方法分为两大类:①时空位置聚类,用于发现事物在空间位置上毗邻、时间上邻近发生的时空分布格局与规律。根据所采用的主要思想和工具的不同,时空位置聚类可进一步细分为5类<sup>[22]</sup>:基于划分的方法、基于层次的方法<sup>[23-24]</sup>、基于密度的方法<sup>[25-26]</sup>、基于模型的方法<sup>[27-28]</sup>和混合策略的方法<sup>[29-31]</sup>;②顾及非空间专题属性的时空聚类,旨在发现具有相似专题属性的地理实体或现象(如监测站点的空气质量、气温等)在时空域上聚集分布的特征,需要同时满足空间/时间毗邻和专题属性相似的双重约束<sup>[32]</sup>。依据属性结合策略的不同,可进一步分为时空一属性耦合的方法和时空一属性分治的方法。时空一属性耦合的方法将地理实体的空间位置、时间和非空间属性视为附加在地理实体上的多维属性,每一个地理实体由一个多维向量进行表示,定义一种混合的距离或相似性度量进行聚类<sup>[33]</sup>。时空一属性分治方法分别从空间和非空间属性2个方面施加层次约束进行聚类操作,以获得综合多维特征相似性的聚类结果<sup>[34-37]</sup>。

面对泛在化、动态性的地理时空大数据,传统“静态”数据的时空聚类算法面临着巨大挑战。近年来,国内外学者针对时空数据的动态性(即“流数据”)聚类分析亦开展了深入研究,取得了丰硕的研

究成果<sup>[38-39]</sup>,其中具有代表性的是针对移动对象的时空聚类分析,所提取的聚集模式主要分为4类,即Convoy模式、Swarm模式、Platoon模式和Moving cluster模式<sup>[40]</sup>。Convoy模式又称为护航模式,指多个对象一起移动一定连续时间的群体运动模式,如执行护航任务的飞机编队机群;Swarm模式也被称为蜂群模式,指一定规模的群体对象在多个时间段(不要求连续)一起移动的群体运动模式,如蜂群的移动模式;Platoon模式是一种时间约束介于Convoy模式和Swarm模式之间的一种群体运行模式,群体对象可以在某个或几个时间段分离,然后又汇合在一起连续移动一定时间段;Moving cluster模式不同于前几个模式,其不要求群体内的对象不变,即允许群体对象一起运动过程中有新对象加入群体以及对象退出群体,如动物迁徙。

## 2.2 时空异常分析

时空异常分析旨在从海量时空数据库中挖掘得到偏离整体或局部分布特征的少部分实体,为深入剖析地理现象或地理过程的特殊分布状况、变化或潜在发展规律提供重要的理论依据和实践指导<sup>[41-43]</sup>。时空异常分析可视为事务型异常探测<sup>[44]</sup>在空间域和时空域的扩展与延伸,进一步顾及了时空数据的相关性、异质性、非平稳性等特征<sup>[45-46]</sup>。当前研究主要集中于如何在异常分析模型中纳入位置—时间—属性的约束。根据位置、时间、属性三者之间的组合关系,时空异常分析大致分为:①基于位置的空间异常探测,提取空间位置具有明显偏离全局或局部分布特征的小部分空间实体集合<sup>[47-49]</sup>;②基于位置—属性的空间异常探测,提取与空间邻近实体属性相比具有明显差异,而与整体相比差异可能不明显的空间实体集合<sup>[50-54]</sup>;③基于位置—时间的时空轨迹异常探测,主要研究时空轨迹的形状异常和分布异常<sup>[55-56]</sup>;④基于位置—时间—属性的时空序列异常探测,提取属性与时空域相比具有显著差异的实体集合<sup>[46,57-61]</sup>。

近年来,国内外地理信息科学领域的学者们大致从3个方面持续开展深入研究:①“时空”维,从时间—空间信息耦合的视角研究空间异常<sup>[62]</sup>→时空异常<sup>[63]</sup>→实时异常<sup>[64]</sup>,如极端气候事件的空间分布、时空分布及实时分布;②“属性”维,从地理实体属性维度的视角研究一元异常<sup>[65]</sup>→二元异常<sup>[66-67]</sup>→多元异常<sup>[68]</sup>,如单一类型犯罪事件的分布异常、两种类型犯罪事件间的交叉异常,以及多种类型犯罪事件相互作用异常;③“分析”维,从分析任务的视角

研究异常探测<sup>[69]</sup>→异常关联<sup>[70]</sup>→异常演化<sup>[71]</sup>,如典型空气污染事件的异常探测、造成污染事件因素的关联分析、以及污染随时间的演化与推断。通过对各维进行抽取和组合可以探索地理时空异常的各类形态以及所蕴含的多层次内涵。

### 2.3 时空关联分析

时空关联分析旨在从时空数据集中识别不同类别地理实体间频繁存在特定时空关联关系(如时空邻近关系)的实体类别集合<sup>[3]</sup>,对于理解地理实体间的时空交互具有重要的科学意义。不同于事务型数据,连续的时/空域内不存在离散的时空事务,因而需要对经典的事务型关联规则挖掘(如Apriori算法<sup>[72]</sup>)拓展到时空数据关联关系挖掘。早期研究主要集中于时/空事务的定义与构建,如格网划分方法<sup>[73]</sup>、参考实体划分方法<sup>[74]</sup>、Voronoi图划分方法<sup>[75]</sup>和路网划分方法<sup>[76]</sup>等。对连续时/空域的离散划分会割裂邻近时空位置上数据间潜在的时空关系,为此,Shekhar等<sup>[77-78]</sup>首次提出了非事务化模型,用于发现连续空间域内频繁同现于邻近空间位置的地理实体集合(称空间同位模式)。遵循该模型的基本思想,地理信息科学领域的学者们主要在3个方面开展深入研究:①空间域向时空域的拓展,借助时空分治<sup>[79-80]</sup>或时空耦合<sup>[81-82]</sup>策略在空间同位模式挖掘模型中纳入时间因子,用于发现频繁同时或依次出现于邻近时空位置的地理实体集合,如深夜发生于酒吧附近的群发案件;②欧氏空间向网络空间的拓展,以网络空间最短路径距离定义地理实体间的邻近关系<sup>[83]</sup>,发现地理实体在网络约束下的频繁同现规律,如邻近街道上的合作商铺;③全局模型向局部模型的拓展,顾及地理数据的异质性,通过区域划分<sup>[84-85]</sup>或聚类分析<sup>[86-87]</sup>思想识别地理实体频繁满足空间邻近关系的空间关联区域,如占据特定空间的生物共生群落。

### 2.4 时空预测分析

时空预测是通过构建反映时空变量间关系的模型对地理事件或现象未知的空间属性值或专题属性值进行估计。时空变量间关系建模是时空预测分析的关键内容,有效地表达时空变量间关系所具有时空依赖性与非平稳性(异质性)是时空预测建模的主要特征。依据分析理论的差异可将时空预测模型分为时空统计模型与机器学习模型。其中,时空统计模型通过统计推断来刻画变量间的关

系,如处理(时)空间依赖性的地统计学模型<sup>[88]</sup>、时空自回归移动平均模型<sup>[89]</sup>等,以及表达(时)空间非平稳性的地理加权回归<sup>[90]</sup>、地理时空加权回归模型<sup>[91]</sup>等。与时空统计模型相比,机器学习模型能够自适应地对任何复杂非线性关系进行建模,近年来被广泛地应用于时空预测分析中。然而,机器学习模型通常要求样本满足独立同分布条件,违背了时空数据的时空依赖性与非平稳性<sup>[92]</sup>,如何建立适用于时空数据的机器学习模型是当前研究热点,融合时空自相关的神经网络<sup>[93]</sup>、地理加权极限学习机<sup>[94]</sup>、以及时空残差神经网络<sup>[95]</sup>等模型的提出则极大地丰富了时空预测分析理论与方法体系。

### 2.5 时空分析的难点问题

为适应于地理大数据多粒度、多类型、多元关联和多维动态等多模态特点<sup>[10]</sup>,地理数据时空分析现有研究仍存在一些亟需解决的难点问题,主要包括:

(1)时空聚集模式的尺度依赖性建模。时空聚集模式是地理对象在一定空间和时间内聚集分布的特征形式。挖掘时空聚集模式的本质目的是认识地理数据中所蕴含的地理事物的时空分布规律。地理数据的自相关、异质和多尺度特性是认识和理解地理事物时空分布规律的基础。当前时空聚集模式挖掘的框架强调从“数据→算法→模式”的过程,在实际应用中甚至将聚集模式挖掘视为一个“黑箱”过程,而忽略了对地理数据的自相关、异质和多尺度特性的探索与分析,亦缺少对所挖掘模式显著性的评价,导致挖掘结果的可解释性差,甚至发现一些“非自然存在”的虚假模式。时空聚集模式通常具有显著的尺度依赖性。在不同尺度下,聚集模式具有不同的表现形式,如图1所示。现有时空聚类方法较少地顾及尺度对挖掘结果的影响,能够挖掘数据中多尺度聚集模式的时空聚类方法研究还相对初步。

(2)时空异常模式的多视角协同探测。现有的时空异常分析模型大多基于单模态特征探测时空异常模式,缺乏对地理时空数据多模态特征间关联关系的客观描述及异常表现形式的深刻认知,难以揭示多视角下时空异常模式的深刻内涵,进而可能导致异常探测结果的片面性,甚至出现错误。如图2(a)和(b)所示,分别描述了两类空间点事件A和B,图2(c)描述了两类点事件的联合空间分布,

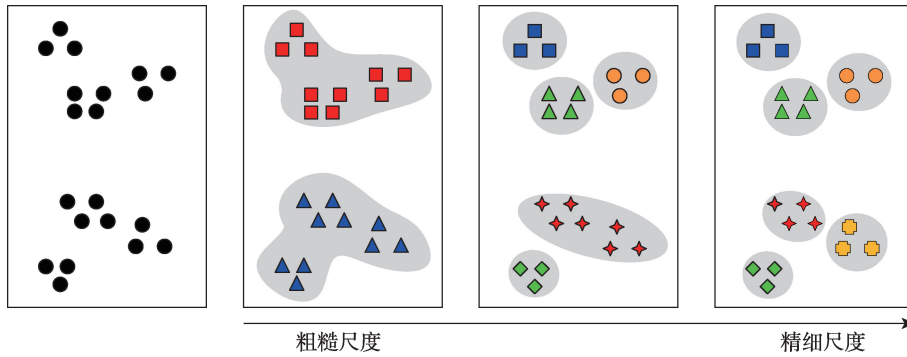


图1 时空聚集模式的尺度依赖性

Fig. 1 Dependence of spatio-temporal clustering on scale

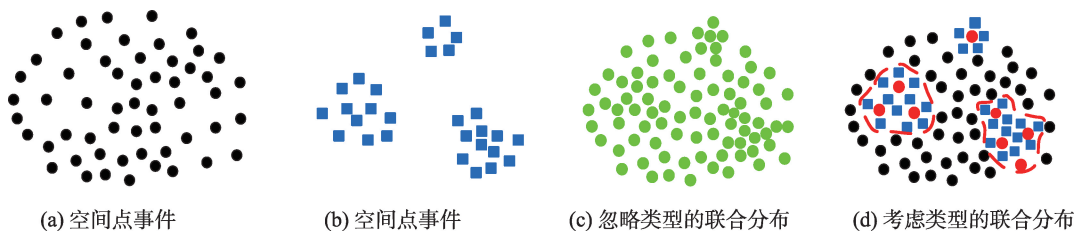


图2 2种类型空间点事件的联合分布示例

Fig. 2 Spatial distributions of two different types of events

基于现有异常的定义,则不存在异常点。若考虑事件的类型时,则可发现某些区域内A类型点与其空间邻域内其他A类型点相比,其周边聚集了较多的B类型点,从而说明这些A类型点在局部范围与B类型点表现为交叉关联异常强烈,称为空间交叉异常。显然,忽略空间点事件类型将无法识别任何异常模式。因此,需要在时空异常分析模型中融入地理数据多模态特征间的关联关系,通过多视角协同探测,以克服现有单模态特征分析结果的片面性。

(3)时空关联模式的多特征可信认知。现有时空分析模型大多基于人为主观设置的频繁度阈值筛选频繁的时空关联模式,缺乏对地理数据多模态分布特征的深度认知,难以揭示地理数据内蕴特征下的客观关联关系,进而可能导致分析结果不可信。以时空同位模式为例,图3(a)中2个随机分布的时空实体在特定分析尺度下亦可能普遍存在随机的时空同位关系,若频繁度阈值设置不合理,现有方法可能会将该实体集合识别为频繁同位模式;同理,图3(b)中随机分布的时空实体与聚集分布的时空实体之间亦可能出现虚假的交互结构,进而导致结果的误判。显然,这2种情况随机产生的频繁模式均为无效的关联模式。因此,需要在时空关联分析模型中融入关于地理数据多模态分布

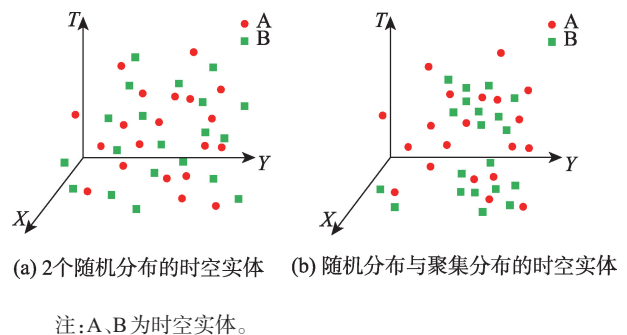


图3 随机交互结构导致的虚假时空同位模式

Fig. 3 Spurious spatial co-location patterns caused by the random interactions

特征的空间认知知识,以提升时空关联模式挖掘结果的可信度。

(4)融合时空数据多重特性的时空预测。时空数据往往不只表现为单一时空特性,可能同时存在时空依赖与时空非平稳等多重特性。例如,在人口时空分布预测中,一方面分析单元中的人口数量受相邻时空单元中人口数量影响,另一方面单元间人口流动机制随时空发生改变;在空气污染浓度时空预测中,相邻时空位置的污染物浓度存在相互依赖的同时,污染源分布、地形与气象等因素的空间异质性会导致污染物浓度与其它变量间关系的时空

变异。因此,在时空预测分析中需要融合时空数据的多重特性,建立更为可靠的时空预测方法是当前亟需解决问题之一。

### 3 多模态地理大数据时空分析的关键技术

针对上述时空分析的难点问题,下面分别介绍多模态地理大数据时空分析的4项关键技术,包括:

- ① 尺度驱动的时空聚类分析;
- ② 融合多视角关联知识的时空异常分析;
- ③ 顾及多模态分布特征的时空关联分析;
- ④ 融合时空数据多重特性的时空预测分析。

#### 3.1 尺度驱动的时空聚类分析

地理现象及其空间分布模式都是在特定尺度下的观测结果<sup>[96]</sup>。在尺度问题的研究中,尺度的描述以及地学分析模型中尺度的参数化是需要重点研究的关键问题<sup>[97]</sup>。从广义的角度来讲,尺度表示描述事物或现象在地理空间中的分布特征(如位置、状态)时所采用的空间或时间单位,同时尺度也可以表示地理现象或过程在空间/时间上的分布范围和发生的频率,以及人们在分析事物或现象的格局、过程时所采用的窗口大小<sup>[98]</sup>。根据对人类视觉多尺度认知过程的分析 and 国内外学者对多尺度时空聚类分析的研究进展,可以将多尺度时空聚类分析涉及的尺度概念主要分为2类:① 数据尺度(或称为观测尺度),表示地理空间数据的采样粒度,如不同分辨率传感器获取的遥感影像具有

不同的数据尺度;② 分析尺度,是指对地理空间数据进行后加工处理、分析、决策、推理时所采用的尺度,如确定某一地理实体的空间邻近实体时所采用的距离阈值。

在人类视觉多尺度认知或聚类过程中,数据尺度主要受观测距离的控制,改变观测距离相当于改变了传感器的分辨率,由此传递给大脑不同观测尺度的地理空间数据进行分析和处理;分析尺度则主要受到视网膜中感受野尺寸的控制,研究表明感受野对信息的转换作用可以近似为一种核函数对信息的处理,而感受野尺寸则相当于核函数的带宽作用,带宽越大对信息的模糊作用越强,不同尺寸的感受野(或分析尺度)获得的图像上发现的对象空间分布模式也不相同<sup>[99]</sup>。如图4所示,显示了人类视觉认知多尺度感知的生理学过程示意图<sup>[100]</sup>。将地理空间数据的比例尺(数据尺度)与观察距离相关联,可以建立起多尺度聚集模式与数据尺度间的关联关系。如图5所示,在不同数据尺度下,通过人类视觉认知到的时空数据中存在的显著聚集模式会随着数据尺度的改变而发生变化。

多尺度时空聚类分析中除了尺度依赖性建模这一关键问题外,还需要研究在一定数据尺度下,聚类模式本源尺度(即最佳分析尺度)的选择问题。最佳分析尺度的选择主要依赖于人类视觉对所观察到的对象聚集信号的响应强度,即聚集模式的显著性。为了从统计意义上对时空聚集模式的显著性进行推断,介绍一种顾及非空间专题属性的时空聚集模式统计挖掘方法<sup>[102]</sup>。该方法首先以时空簇的可靠性或显著性的判别研究为切入点,在时

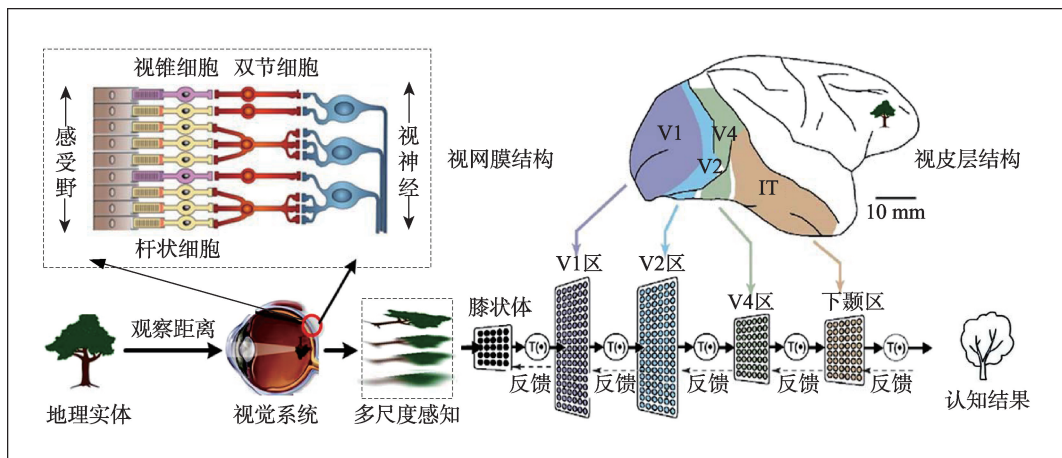


图4 人类视觉认知的生理结构示意图<sup>[100]</sup>

Fig. 4 Illustration of human visual system

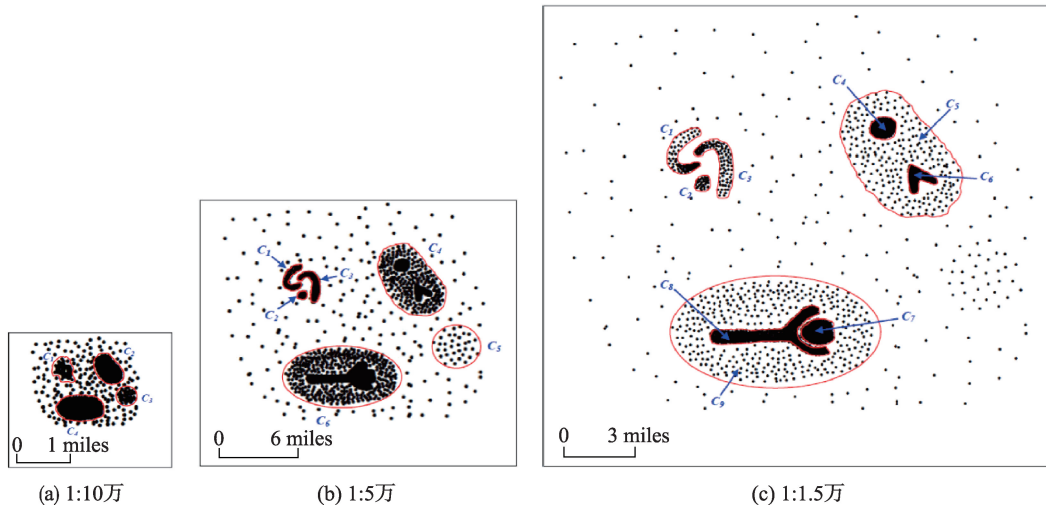


图5 时空聚集模式与数据尺度间关联关系<sup>[101]</sup>

Fig. 5 Relationship between spatio-temporal clusters and data scale

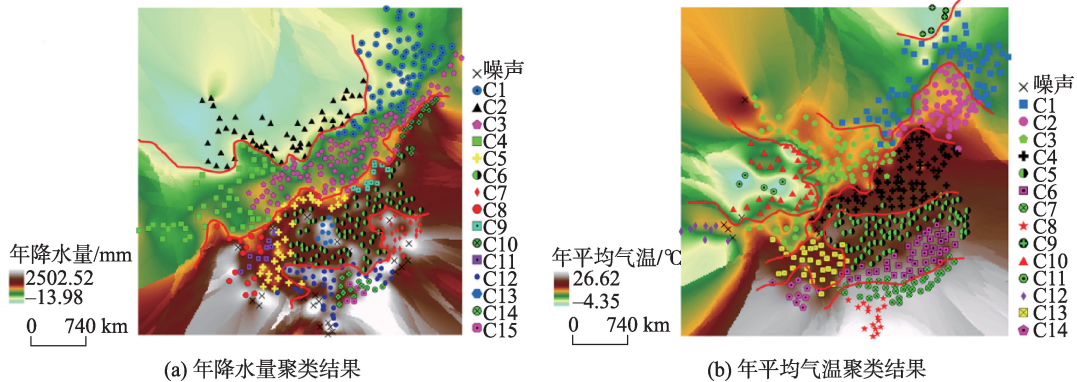
空自相关显著性检验研究的基础上,通过时空随机重排检验的策略,构建适用于时空聚集模式显著性检验的统计判别方法,并与时空层次聚类算法相融合。如图6所示,采用该方法探测我国降水与气温时空分布模式,将有助于气象学研究领域分析我国降水与气温的时空分异性和变化规律。图6(a)中降水数据被划分为15个聚类(即C1-C15),较好地反映了我国降水的空间分异性;图6(b)中的年平均气温聚类结果(C1-C14)也与我国气温分布的等值线(图中红色线)相吻合。

### 3.2 融合多视角关联知识的时空异常分析

针对单模态特征下时空异常分析结果的片面性问题,介绍一种融合多视角关联知识的时空对象交叉异常统计推断方法。相比于现有方法,该方法

在异常分布模式分析过程中通过融合多模态关联知识,建立关联知识引导下的异常统计推断,实现对时空对象异常分布模式的多视角关联透视,克服了异常模式主观判别的难题。

为明确分析目的,首先需要理解地理学视角下的时空交叉异常的表现形式。该方法将时空交叉异常解释为显著偏离主导分布格局的交叉关联关系的实体集合。如图7所示,包含了多类型的空间实体集合,以实体类型a为分析对象,事件a→b交叉关联主导下的事件a→c集合为交叉异常,如区域2和3;事件a→b交叉关联主导下的局部交叉关联强度明显偏强的实体集合为交叉异常,如区域4;去除交叉异常,事件a中呈现局部聚集实体集合为交叉异常,如区域5。在此认知基础上,时空交叉异常



注:图中标记的红色线由北向南分别代表了我国年降水量400 mm,800 mm和1600 mm等值线。

图6 时空聚类统计挖掘方法对气象监测数据的聚类分区结果<sup>[102]</sup>

Fig. 6 Meteorological division identified by the statistical method for spatio-temporal clustering

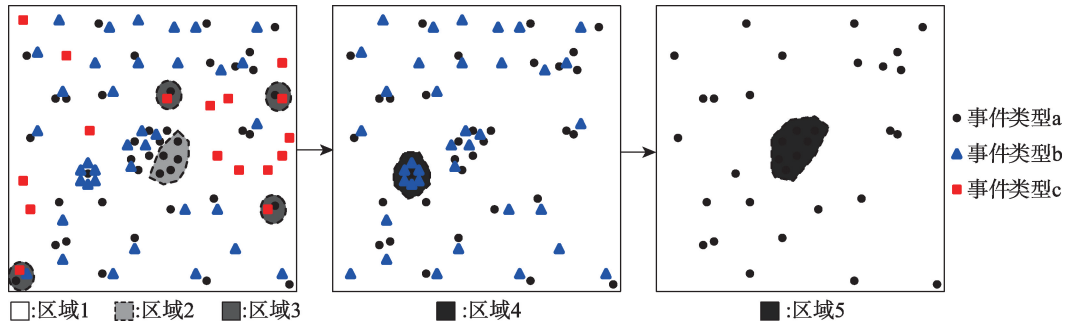


图7 时空交叉异常分布示例

Fig. 7 Illustration of spatio-temporal cross outliers

探测可以建模为不同类型实体之间交叉关联关系强度的显著性检验问题。

考虑一个由目标类型实体和参考类型实体组成的数据集,首先对指定的目标类型实体,采用交叉K-函数统计检验其与参考类型实体之间的空间依赖显著性,具体地,对任一目标类型实体,综合考虑与其空间邻近的参考类型实体数目及其之间的空间距离,度量目标-参考类型点之间的空间交叉关联强度。在此基础上,构建目标类型实体的空间邻域,进行目标类型实体交叉异常的统计推断。针对具有显著空间依赖的目标-参考类型实体数据集,借助图论与点模式分析介绍2种研究策略,即基于层次约束TIN的方法<sup>[103]</sup>和基于假设检验的方法<sup>[104]</sup>。

(1)基于层次约束TIN的方法是将不同类别实体之间的交叉关联关系建模为图,通过约束图形结构提取异常。首先进行空间交叉邻近关系构建,对数据集中的所有空间实体构建Delaunay三角网(TIN),在此基础上,通过对TIN边施加空间多层次

约束准确获取两类实体之间的空间交叉邻近关系,根据目标类型实体与邻近参考类型实体数目及其之间的空间距离计算空间交叉关联强度。针对目标类型实体,采用层次约束TIN构建空间邻域,将相邻目标类型实体之间的交叉关联强度差异转化为连接边的权重,采用类似的策略对边施加层次约束,从而获得一系列TIN的子图,根据子图所包含的目标类型实体的数目进行交叉异常的统计推断。

(2)基于假设检验的方法是将交叉异常判别建模为邻近目标类别实体之间交叉关联强度的显著性检验问题。首先,采用层次约束TIN获取目标类型实体的空间邻域,如图8(a)中 $P_i$ 的空间邻域 $SN(P_i)=\{P_1, P_2, P_3, P_4, P_5, P_6, P_7\}$ ,并根据给定参考半径 $r$ 范围内参考类型实体的分布特征计算交叉关联强度。针对目标类型实体及其空间邻域实体集合,构建给定参考半径范围下的支撑域,如图8(a)所示。在此基础上,零假设 $H_0$ 认为目标类型实体与其空间邻域实体的交叉关联强度没有明显差异,即参考类型实体在支撑域范围满足完全空间随机分布

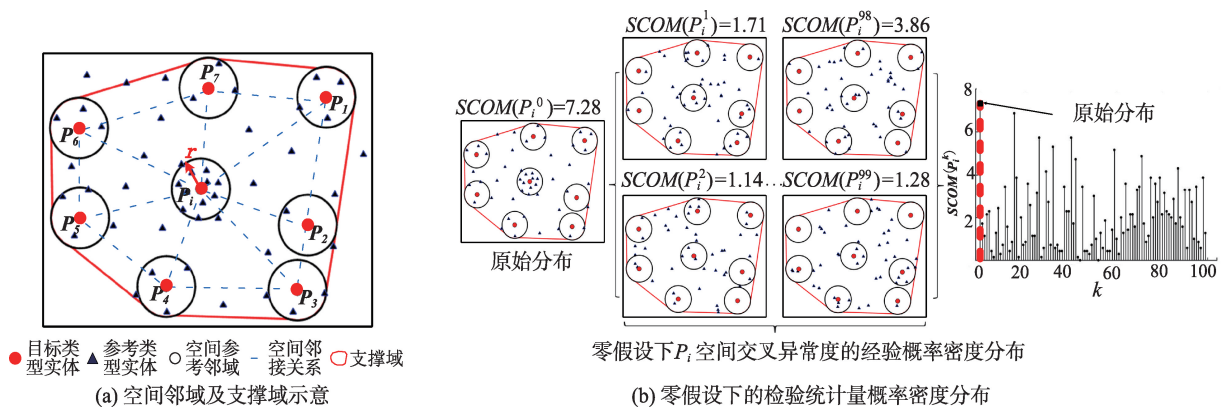


图8 基于假设检验的异常分布模式统计推断

Fig. 8 Significance tests on the cross outliers



(CSR, Complete Spatial Randomness)。因此,采用蒙特卡洛随机模拟在支撑域范围内生成大量随机数据集,选取交叉关联强度差异度量指标(如空间交叉异常度 SCOM<sup>[104]</sup>)作为统计推断的检验统计量,由大量随机数据集估算检验统计量在零假设下的经验概率密度分布,对交叉异常的显著性进行统计判别,如图8(b)中,目标类型实体 $P_i$ 在原始分布下的空间交叉异常度为7.28,大于 $m=99$ 次随机数据集下的空间交叉异常度,则其 $p\text{-value}(P_i^0)=1/(99+1)=0.01$ ,在给定的显著性水平 $\alpha$ (通常为0.05)条件下拒绝零假设 $H_0$ ,则 $P_i$ 为显著的空间交叉异常。

融合多视角关联知识的时空对象异常分布模式统计推断方法已成功应用于不同类型犯罪事件之间、城市金融设施与抢劫犯罪事件间的时空交叉异常分布模式探测<sup>[103-104]</sup>,可以为进一步实现城市突发事件的联合预警提供技术支持。

### 3.3 顾及多模态分布特征的时空关联分析

针对时空关联分析结果有效性难以评价的问题,本文以时空同位模式为例,介绍一种地理实体时空关联模式的统计分析方法。相比于现有方法,该方法以地理实体多模态分布特征知识为引导,可以有效排除随机交互结构对结果的干扰,实现对多实体关联关系的客观认知。

为明确分析目的,首先需要理解时空同位模式形成背后的地理学内涵。时空同位模式可解释为单类实体时空分布结构相互诱导的产物<sup>[105]</sup>,如图9

所示,若某类地理实体或潜在实体的分布结构(如空间聚集结构)会诱导其它多类地理实体在相近位置呈现相似分布结构<sup>[106]</sup>,则这些地理实体会形成具有地理内涵的时空同位模式。在该认知基础之上,时空同位模式挖掘可以进一步建模为多类地理实体分布结构间诱导关系的显著性检验问题<sup>[107]</sup>。

针对由 $k$ 类地理实体组成的 $k$ 元候选同位模式 $CP=\{f_1, \dots, f_k\}$ ,零假设 $H_0$ 描述为 $k$ 类实体 $f_1, \dots, f_k$ 的分布结构间不存在诱导关系,即分布相互独立;备择假设 $H_1$ 则描述为 $k$ 类实体间存在诱导性分布结构,即分布相互依赖。为构建服从零假设 $H_0$ 的零模型,需要在保持 $k$ 类实体分布独立的同时,维持各类实体在观测数据集中的分布结构。由于每类地理实体的潜在分布过程是预先未知的,因此在缺乏充分先验知识的情况下难以选择合理的分布模型(如泊松簇过程<sup>[108]</sup>)对观测数据进行模拟<sup>[109]</sup>。如图10所示,首先选取多个互补模态的分布特征统计量,

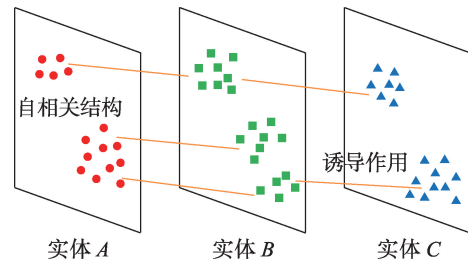


图9 不同地理实体空间自相关结构间的诱导作用  
Fig. 9 Induced spatial auto-correlations between different features

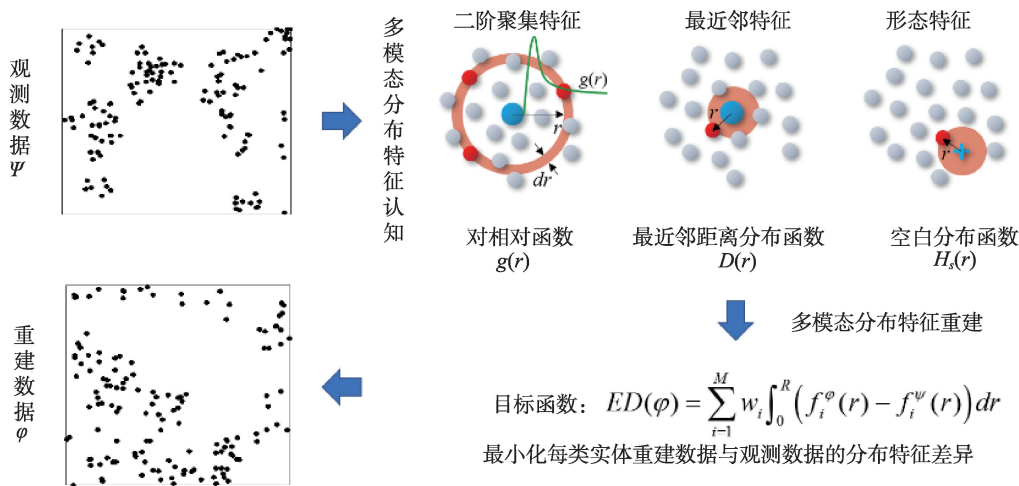


图10 基于多模态分布特征的模式重建

Fig. 10 Pattern reconstruction based on multi-modal summary characteristics

对每类地理实体的分布结构进行多视角定量认知,进而借助模式重建策略<sup>[110]</sup>对观测数据的单元分布特征进行非参数拟合,独立生成具有相似单元分布特征的多元重建数据集。在此基础上,选取多实体时空交互的兴趣度量指标(如参与指数<sup>[78]</sup>)作为统计推断的检验统计量,由大量重建数据集估计检验统计量在零假设下的经验分布,从而检验候选模式中各实体分布特征间诱导关系的显著性(即 $p$ 值),识别统计显著的时空同位模式。如图11所示,时空同位模式统计分析模型能够有效识别由诱导性时空分布结构生成的显著模式,并且能够有效剔除随机发生的虚假模式。

在实际应用时,时空关联模式统计挖掘模型需要考虑地理数据的时空维度、度量空间和研究范围,进而设计具体的实现方法。例如,针对城市环境中沿道路网络分布的地理实体(如城市设施),需要在分布特征认知、重建数据生成和检验统计量构建等3个重要环节中同时建模网络约束特征<sup>[105]</sup>。针对不同应用问题,已发展了空间数据、时空数据和网络空间数据的全局或局部关联模式统计分析的具体方法<sup>[105,107,109,111]</sup>,并成功运用于湿地植物物种共生关系分析、城市设施兴趣点空间集群结构提取以及犯罪案件诱发机制识别,可以为改善生态环境、优化城市结构、维护公共安全等智慧城市应用领域提供重要的决策支持。

### 3.4 融合时空数据多重特性的时空预测分析

时空数据的特性将对时空预测分析具有重要影响,也是时空预测建模需要考虑的主要因素<sup>[112]</sup>。本文以空气污染物浓度预测应用为例,介绍一种兼顾时空依赖性与空间异质性的时空支持向量回归预测方法,该方法可以抽象为融合时空数据多重特

性的时空预测分析通用框架,进而扩展到不同的应用领域。

大气运动直接影响空气污染物扩散与迁移,该物理过程在空间的连续性决定了风向路径上越邻近位置污染物浓度具有越高的相关度,因此,为了描述污染浓度数据的时空依赖特征,采用高斯矢量函数构建时空自相关变量。如图12(a)所示,假设风向为东北(NE),则 $p_0$ 位置的污染浓度变化与 $p_1, p_2, p_3$ 以及 $p_4$ 位置污染源有关,关系的强弱则取决于 $p_0$ 与其它位置的空间关系,通常表现为随角度( $\theta$ )与距离( $d$ )的增加而呈现衰减特征,据此可定义高斯矢量权重 $w_{ij}$ ,并通过加权平均可构建 $p_0$ 位置的污染浓度时空自相关变量。

污染源、地形地貌等因素的空间分异,导致不同区域中污染物浓度与关联因素/协变量的关系不尽相同,即表现出所谓的空间非平稳性或异质性。在地理时空数据分析中,空间异质可分为局部异质与分层异质<sup>[113]</sup>,两者分别强调空间分异的位置与区域依赖,其中局部异质可以借助于局部统计量度量,如LISA指数,而空间分层异质可利用 $q$ 统计量或者空间聚类分析技术进行分析。如图12(b)所示,针对污染物浓度与关联变量在较大尺度研究区中具有空间分层异质,利用空间聚类分析算法探测污染物浓度数据中的空间聚集模式,实现均质或近似均质区域的划分。进而,以不同区域为分析单元,构建参数随空间区域变化的局部支持向量回归模型,其中模型的输入包括协变量与相应的时空自相关变量,从而实现时空数据多重特性表达的时空预测建模。由于时空数据的多重特性在不同的应用场景表现形式各异,如何探测与表达应用场景中时空特性具体形式,并将其与模型进行有效融合是时

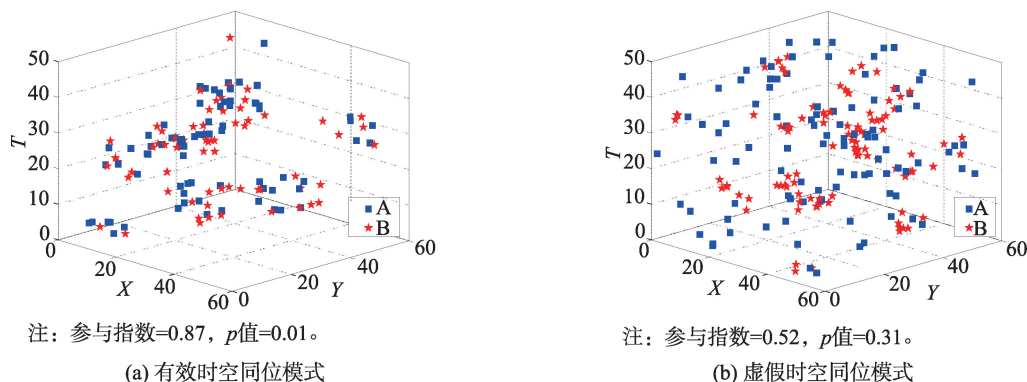


图11 时空同位模式统计分析模型的判别结果

Fig. 11 Spatio-temporal co-location patterns determined by the statistical method

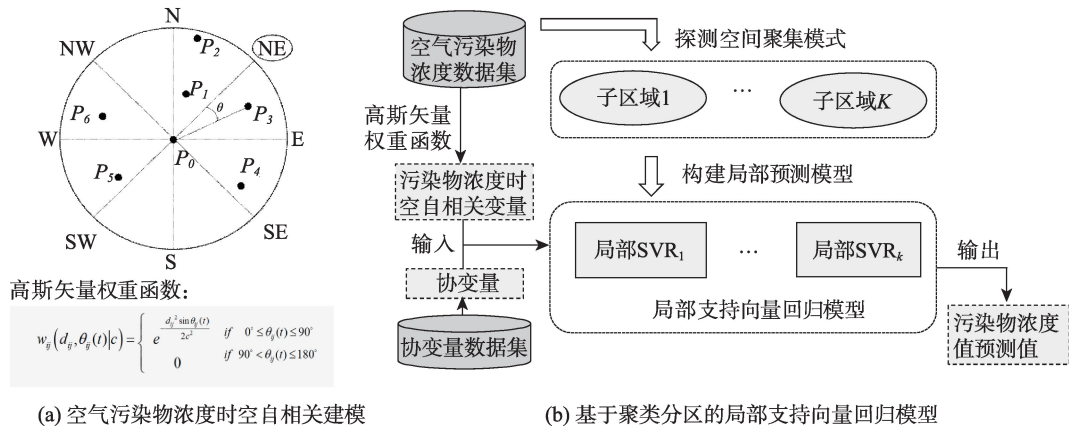


图12 兼顾时空依赖性与空间异质性的时空支持向量回归预测方法

Fig. 12 Space-time support vector regression model considering both auto-correlation and heterogeneity

空预测分析中重点关注的内容。

### 4 总结与展望

本文系统回顾了4类主要时空分析方法(时空聚类分析、时空异常分析、时空关联分析与时空预测分析)的发展脉络与研究现状,探讨了地理大数据时代下时空分析潜在的难点问题。针对这些问题,给出了时空分析模型中多尺度、多视角、多特征等多模态信息的建模方法,并针对不同分析任务介绍了多模态地理大数据时空分析模型,能够为全空间信息系统的研发及其在智慧城市中的应用提供有力的技术支撑。

为应对地理大数据时代带来的新机遇与挑战,多模态地理大数据时空分析的研究仍任重而道远,未来需要重点开展以下4个方向的研究工作:

(1)地理大数据的“5V”与“5度”特征对时空分析方法的计算性能提出了严峻挑战,如何在并行计算等高性能计算模型融入地理数据的特性(自相关性、异质性与尺度依赖性),开展面向地理大数据的实时或近实时的时空分析方法研究<sup>[114]</sup>;

(2)人类行为大数据多以数据流(如人口流动)的形式存在,不同于对地观测大数据所聚焦的位置数据,流数据能够表达不同地理位置或区域间的交互信息<sup>[115]</sup>,需要进一步认知流数据空间内时空模式的新型表现形式与地理内涵,并发展相应的分析方法;

(3)地理大数据的多模态信息使得复杂时空对象的刻画更加精细<sup>[116]</sup>,面向点线面实体的时空分析方法难以发挥多模态信息的价值<sup>[112]</sup>,为此,需要研

究融合多模态信息的复杂对象时空分析方法;

(4)时空对象从诞生、发展、变化到消亡的全生命周期特征表达在地理大数据时代已成为可能<sup>[13]</sup>,如何实现地理大数据多模态特征及其交互关系的全链建模将更加困难,建立面向全周期过程的动态时空分析模型将是未来值得关注的研究热点。

#### 参考文献(References):

[1] Goodchild M F. Geographical information science[J]. International Journal of Geographical Information Systems, 1992,6(1):31-45.

[2] Langran G, Chrisman N R. A framework for temporal geographic information[J]. Cartographica: The International Journal for Geographic Information and Geovisualization, 1988,25(3):1-14.

[3] Shekhar S, Jiang Z, Ali R, et al. Spatiotemporal data mining: A computational perspective[J]. ISPRS International Journal of Geo-information, 2015,4(4):2306-2338.

[4] Marr B. Big data: Using smart big data, analytics and metrics to make better decisions and improve performance [M]. Chichester, UK: John Wiley & Sons, 2015.

[5] 傅伯杰,冷疏影,宋长青.新时期地理学的特征与任务[J]. 地理科学,2015,35(8):939-945. [Fu B J, Leng S Y, Song C Q. The characteristics and tasks of geography in the new era[J]. Scientia Geographica Sinica, 2015,35(8):939-945.]

[6] Vatsavai R R, Ganguly A, Chandola V, et al. Spatiotemporal data mining in the era of big spatial data: Algorithms and applications[C]// Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data. ACM, 2012:1-10.

[7] Li S, Dragicevic S, Castro F A, et al. Geospatial big data

- handling theory and methods: A review and research challenges[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016,115:119-133.
- [8] 朱庆,付萧.多模态时空大数据可视分析方法综述[J].测绘学报,2017,46(10):1672-1677. [Zhu Q, Fu X. The review of visual analysis methods of multi-modal spatio-temporal big data[J]. *Acta Geodaetica et Cartographica Sinica*, 2017,46(10):1672-1677. ]
- [9] 杨振山,龙瀛, DOUAY Nicolas.大数据对人文—经济地理学研究的促进与局限[J].地理科学进展,2015,34(4):410-417. [Yang Z S, Long Y, DOUAY Nicolas. Opportunities and limitations of big data applications to human and economic geography: The state of the art[J]. *Progress in Geography*, 2015,34(4):410-417. ]
- [10] 华一新,周成虎.面向全空间信息系统的多粒度时空对象数据模型描述框架[J].地球信息科学学报,2017,19(9):1142-1149. [Hua Y X, Zhou C H. Description frame of data model of multi-granularity spatio-temporal object for pan-spatial information system[J]. *Journal of Geo-information Science*, 2017,19(9):1142-1149. ]
- [11] 裴韬,刘亚溪,郭思慧,等.地理大数据挖掘的本质[J].地理学报,2019,74(3):586-598. [Pei T, Liu Y X, Guo S H, et al. Principle of big geodata mining[J]. *Acta Geographica Sinica*, 2019,74(3):586-598. ]
- [12] 周成虎.全空间地理信息系统展望[J].地理科学进展,2015,34(2):129-131. [Zhou C H. Prospects on pan-spatial information system[J]. *Progress in Geography*, 2015,34(2):129-131. ]
- [13] 华一新.全空间信息系统的核心问题和关键技术[J].测绘科学技术学报,2016,33(4):331-335. [Hua Y X. The core problems and key technologies of pan-spatial information system[J]. *Journal of Geomatics Science and Technology*, 2016,33(4):331-335. ]
- [14] 宋长青.地理学研究范式的思考[J].地理科学进展,2016,35(1):1-3. [Song C Q. On paradigms of geographical research[J]. *Progress in Geography*, 2016,35(1):1-3. ]
- [15] 李德仁,王树良,史文中,等.论空间数据挖掘和知识发现[J].武汉大学学报·信息科学版,2001,26(6):491-499. [Li D R, Wang S L, Shi W Z, et al. On spatial data mining and knowledge discovery (SDMKD) [J]. *Geomatics and Information Science of Wuhan University*, 2001,26(6):491-499. ]
- [16] 李德仁,王树良,李德毅,等.论空间数据挖掘和知识发现的理论与方法[J].武汉大学学报·信息科学版,2002,27(3):221-233. [Li D R, Wang S L, Li D Y, et al. Theories and technologies of spatial data mining and knowledge discovery[J]. *Geomatics and Information Science of Wuhan University*, 2002,27(3):221-233. ]
- [17] 王劲峰,葛咏,李连发,等.地理学时空数据分析方法[J].地理学报,2014,69(9):1326-1345. [Wang J F, Ge Y, Li L F, et al. Spatiotemporal data analysis in geography[J]. *Acta Geographica Sinica*, 2014,69(9):1326-1345. ]
- [18] Deng M, Fan Z, Liu Q, et al. A hybrid method for interpolating missing data in heterogeneous spatio-temporal datasets[J]. *ISPRS International Journal of Geo-information*, 2016,5(2):13.
- [19] 何占军,邓敏,蔡建南,等.顾及背景知识的多事件序列关联规则挖掘方法[J].武汉大学学报·信息科学版,2018,43(5):766-772. [He Z J, Deng M, Cai J N, et al. A context-based association rules mining method for multiple event sequences[J]. *Geomatics and Information Science of Wuhan University*, 2018,43(5):766-772. ]
- [20] He Z, Deng M, Cai J, et al. Mining spatiotemporal association patterns from complex geographic phenomena[J]. *International Journal of Geographical Information Science*, 2019:1-26. [https://doi.org/10.1080/1365\\_8816.2019.1566549](https://doi.org/10.1080/1365_8816.2019.1566549)
- [21] Deng M, Yang W, Liu Q, et al. Heterogeneous space-time artificial neural networks for space-time series prediction[J]. *Transactions in GIS*, 2018,22(1):183-201.
- [22] 邓敏,刘启亮,李光强,等.空间聚类分析及应用[M].北京:科学出版社,2011. [Deng M, Liu Q L, Li G Q, et al. *Spatial clustering methods and applications*[M]. Beijing: Science Press, 2011. ]
- [23] Guo D. Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP)[J]. *International Journal of Geographical Information Science*, 2008,22(7):801-823.
- [24] Liu Q, Liu W, Tang J, et al. Permutation-test-based clustering method for detection of dynamic patterns in spatio-temporal datasets[J]. *Computers, Environment and Urban Systems*, 2019,75:204-216.
- [25] Birant D, Kut A. ST-DBSCAN: An algorithm for clustering spatial-temporal data[J]. *Data & Knowledge Engineering*, 2007,60(1):208-221.
- [26] Pei T, Zhou C, Zhu A-X, et al. Windowed nearest neighbour method for mining spatio-temporal clusters in the presence of noise[J]. *International Journal of Geographical Information Science*, 2010,24(6):925-948.
- [27] Kulldorff M, Heffernan R, Hartman J, et al. A space-time permutation scan statistic for disease outbreak detection [J]. *PLOS Medicine*, 2005,2(3):216-224.
- [28] Bação F, Lobo V, Painho M. The self-organizing map, the Geo-SOM, and relevant variants for geosciences[J]. *Computers & Geosciences*, 2005,31(2):155-163.
- [29] 汪中,刘贵全,陈恩红.一种优化初始中心点的K-means

- 算法[J].模式识别与人工智能,2009,22(2):299-304. [ Wang Z, Liu G Q, Chen E H. A k-means algorithm based on optimized initial center points[J]. Pattern Recognition and Artificial Intelligence, 2009,22(2):299-304. ]
- [30] Bai L, Cheng X, Liang J, et al. Fast density clustering strategies based on the k-means algorithm[J]. Pattern Recognition, 2017,71(1):375-386.
- [31] Lin C R, Chen M S. Combining partitional and hierarchical algorithms for robust and efficient data clustering with cohesion self-merging[J]. IEEE Transactions on Knowledge and Data Engineering, 2005,17(2):145-159.
- [32] 焦利民,张欣,毛立凡.自组织双重空间聚类算法的城市扩张结构分析应用[J].地球信息科学学报,2015,17(6):638-643. [ Jiao L M, Zhang X, Mao L F. Self-organizing dual spatial clustering algorithm and its application in the analysis of urban sprawl structure[J]. Journal of Geo-information Science, 2015,17(6):638-643. ]
- [33] Joshi D, Samal A, Soh L K. Spatio-temporal polygonal clustering with space and time as first-class citizens[J]. Geo-informatica, 2012,17(2):387-412.
- [34] Liu Q, Deng M, Shi Y, et al. A density-based spatial clustering algorithm considering both spatial proximity and attribute similarity[J]. Computers & Geosciences, 2012, 46:296-309.
- [35] 石岩,刘启亮,邓敏,等.融合图论与密度思想的混合空间聚类方法[J].武汉大学学报·信息科学版,2012,37(11):1276-1280. [ Shi Y, Liu Q L, Deng M, et al. A hybrid spatial clustering method based on graph theory and spatial density[J]. Geomatics and Information Science of Wuhan University, 2012,37(11):1276-1280. ]
- [36] Han J, Zhu L, Kullendorff M, et al. Using gini coefficient to determining optimal cluster reporting sizes for spatial scan statistics[J]. International Journal of Health Geographics, 2016,15(1):27.
- [37] Lin C R, Liu K H, Chen M S. Dual clustering: Integrating data clustering over optimization and constraint domains [J]. IEEE Transactions on Knowledge and Data Engineering, 2005,17(5):628-637.
- [38] 吉根林,孙鸿艳,赵斌.时空轨迹群体运动模式挖掘研究进展[J].南京航空航天大学学报,2016,48(5):615-624. [ Ji G L, Sun H Y, Zhao B. Research progress in group moving patterns mining of spatio-temporal trajectories[J]. Journal of Nanjing University of Aeronautics & Aeronautics, 2016,48(5):615-624. ]
- [39] 龚玺,裴韬,孙嘉,等.时空轨迹聚类方法研究进展[J].地理科学进展,2011,30(5):522-534. [ Gong X, Pei T, Sun J, et al. Review of the research progresses in trajectory clustering methods[J]. Progress in Geography, 2011,30(5):522-534. ]
- [40] 刘文凯,唐建波,蔡建南,等.面向城市交通应用的移动对象聚类算法比较研究[J].地理与地理信息科学,2016,32(6):69-74. [ Liu W K, Tang J B, Cai J N, et al. Algorithm comparison for clustering of moving objects in traffic system[J]. Geography and Geo-information Science, 2016,32(6):69-74. ]
- [41] Han J, Kamber M, Pei J. Data mining: Concepts and techniques, third edition[M]. Morgan Kaufman, San Francisco, 2012.
- [42] 邓敏,石岩,龚健雅,等.时空异常探测方法研究综述[J].地理与地理信息科学,2016,32(6):43-50. [ Deng M, Shi Y, Gong J Y, et al. A summary of spatio-temporal outlier detection[J]. Geography and Geo-Information Science, 2016,32(6):43-50. ]
- [43] Campos G O, Zimek A, Jörg Sander, et al. On the evaluation of unsupervised outlier detection: Measures, datasets, and an empirical study[J]. Data Mining and Knowledge Discovery, 2016,30(4):891-927.
- [44] Hawkins D. Identification of outliers[M]. London: Chapman and Hall, 1980.
- [45] Shekhar S, Lu C T, Zhang P. A unified approach to detecting spatial outliers[J]. Geo-informatica, 2003,7(2):139-166.
- [46] Cheng T, Li Z. A multiscale approach for spatio-temporal outlier detection[J]. Transactions in GIS, 2006,10(2):253-263.
- [47] Knorr E M, Ng R T. Algorithms for mining distance-based outliers in large datasets[C]// Proceedings of the International Conference on Very Large Data Bases, 1998, 98:392-403.
- [48] Breunig M M, Kriegel H P, Ng R T, et al. LOF: Identifying density based local outliers[C]// Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 2000,29(2):93-104.
- [49] Shi Y, Deng M, Yang X, et al. Adaptive detection of spatial point event outliers using multilevel constrained delaunay triangulation[J]. Computers, Environment and Urban Systems, 2016,59:164-183.
- [50] Chawla S, Sun P. SLOM: A new measure for local spatial outliers[J]. Knowledge & Information Systems, 2006,9(4):412-429.
- [51] Chen D, Lu C T, Kou Y, et al. On detecting spatial outliers [J]. Geo-informatica, 2008,12(4):455-475.
- [52] 邓敏,刘启亮,李光强.采用聚类技术探测空间异常[J].遥感学报,2010,14(5):944-958. [ Deng M, Liu Q L, Li G Q. Spatial outlier detection method based on spatial clustering[J]. Journal of Remote Sensing, 2010,14(5):944-958. ]
- [53] Lu C T, Raimundo F, Liu X, et al. A graph-based

- approach to detect abnormal spatial points and regions[J]. *International Journal on Artificial Intelligence Tools*, 2011,20(4):721-751.
- [54] Shi Y, Deng M, Yang X, et al. A spatial anomaly points and regions detection method using multi- constrained graphs and local density[J]. *Transactions in GIS*, 2017,21(2):376-405.
- [55] Lee J G, Han J, Li X. Trajectory outlier detection: A partition- and- detect framework[C]// *Proceedings of the IEEE International Conference on Data Engineering*, 2008:140-149.
- [56] 毛嘉莉,金澈清,章志刚,等.轨迹大数据异常检测:研究进展及系统框架[J]. *软件学报*,2017,28(1):17-34. [ Mao J L, Jin C Q, Zhang Z G, et al. Anomaly detection for trajectory big data: Advancements and framework[J]. *Journal of Software*, 2017,28(1):17-34. ]
- [57] Birant D, Kut A. Spatio-temporal outlier detection in large databases[J]. *Journal of Computing and Information Technology*, 2006,14(4):291-297.
- [58] Wu E, Liu W, Chawla S. Spatio-temporal outlier detection in precipitation data[C]// *Proceedings of the International Workshop on Knowledge Discovery from Sensor Data*. Springer, Berlin, Heidelberg, 2008:115-133.
- [59] 刘启亮,邓敏,王佳璆,等.时空一体化框架下时空异常探测[J]. *遥感学报*,2011,15(3):457-474. [ Liu Q L, Deng M, Wang J Q, et al. Spatio-temporal outliers detection within the space-time framework[J]. *Journal of Remote Sensing*, 2011,15(3):457-474. ]
- [60] George G, Yorgos N. A fuzzy index for detecting spatio-temporal outliers[J]. *Geo-informatica*, 2012,16:597-619.
- [61] Telang A, Deepak P, Joshi S, et al. Detecting localized homogeneous anomalies over spatio-temporal data[J]. *Data Mining & Knowledge Discovery*, 2014,28(5- 6):1480-1502.
- [62] Ernst M, Haesbroeck G. Comparison of local outlier detection techniques in spatial multivariate data[J]. *Data Mining & Knowledge Discovery*, 2017,31(2):1-29.
- [63] Gupta M, Gao J, Aggarwal C C, et al. Outlier detection for temporal data: A survey[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2014,26(9):2250-2267.
- [64] Shi Y, Deng M, Yang X, et al. Detecting anomalies in spatio-temporal flow data by constructing dynamic neighbourhoods[J]. *Computers, Environment and Urban Systems*, 2018,67:80-96.
- [65] Schubert E, Zimek A, Kriegel H P. Local outlier detection reconsidered: a generalized view on locality with applications to spatial, video, and network outlier detection[J]. *Data Mining & Knowledge Discovery*, 2014,28(1):190-237.
- [66] Liu X, Chen F, Lu C T. On detecting spatial categorical outliers[J]. *GeoInformatica*, 2014,18(3):501-536.
- [67] Papadimitriou S, Faloutsos C. Cross-outlier detection[C]// *Proceedings of the International Symposium on Spatial and Temporal Databases*. Springer, Berlin, Heidelberg, 2003:199-213.
- [68] Janeja V P, Palanisamy R. Multi-domain anomaly detection in spatial datasets[J]. *Knowledge & Information Systems*, 2013,36(3):749-788.
- [69] Singh A K, Lalitha S. A novel spatial outlier detection technique[J]. *Communication in Statistics- Theory and Methods*, 2018,47(1):247-257.
- [70] Zheng Y, Zhang H, Yu Y. Detecting collective anomalies from multiple spatio-temporal datasets across different domains[C]// *Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2015:1-10.
- [71] Palacio S M. Abnormal pattern prediction: Detecting fraudulent insurance property claims with semi-supervised machine-learning[J]. *Data Science Journal*, 2019,18(1):1-15.
- [72] Agrawal R, Srikant R. Fast algorithms for mining association rules[C]//*Proceedings of the 20<sup>th</sup> International Conference on Very Large Data Bases*. Santiago, Chile, 1994, 1215:487-499.
- [73] Shekhar S, Huang Y. Co-location rules mining: A summary of results[C]// *Proceedings of the Spatio-Temporal Symposium on Databases*, 2001.
- [74] Koperski K, Han J. Discovery of spatial association rules in geographic information databases[C]// *Proceedings of the International Symposium on Spatial Databases*. Springer Berlin Heidelberg, 1995:47-66.
- [75] Li G Q, Deng M, Zhu J J. Spatial association rules mining methods based on voronoi diagram[J]. *Geomatics and Information Science of Wuhan University*, 2008,33(12): 1242-1245.
- [76] 田晶,王一恒,颜芬,等.一种网络空间现象同位模式挖掘的新方法[J]. *武汉大学学报·信息科学版*,2015,40(5): 652-660. [ Tian J, Wang Y H, Yan F, et al. A new method for mining co-location patterns between network spatial phenomena[J]. *Geomatics and Information Science of Wuhan University*, 2015,40(5):652-660. ]
- [77] Shekhar S, Huang Y. Discovering spatial co-location patterns: A summary of results[C]// *Proceedings of the International Symposium on Spatial and Temporal Databases*. Springer, Berlin, Heidelberg, 2001:236-256.
- [78] Huang Y, Shekhar S, Xiong H. Discovering colocation patterns from spatial data sets: A general approach[J].

- IEEE Transactions on Knowledge & Data Engineering, 2004,16(12):1472-1485.
- [79] Celik M, Shekhar S, Rogers J P, et al. Mixed-drove spatio-temporal co-occurrence pattern mining: A summary of results[C]//Proceedings of the Sixth International Conference on Data Mining (ICDM 2006). IEEE, 2006:119-128.
- [80] Celik M, Shekhar S, Rogers J P, et al. Mining at most top-k% mixed-drove spatio-temporal co-occurrence patterns: a summary of results[C]// Proceedings of the 2007 IEEE 23<sup>rd</sup> International Conference on Data Engineering Workshop. IEEE, 2007:565-574.
- [81] Huang Y, Zhang L, Zhang P. A framework for mining sequential patterns from spatio-temporal event data sets[J]. IEEE Transactions on Knowledge and Data Engineering, 2008,20(4):433-448.
- [82] Mohan P, Shekhar S, Shine J A, et al. Cascading spatio-temporal pattern discovery[J]. IEEE Transactions on Knowledge and Data Engineering, 2012,24(11):1977-1992.
- [83] Yu W, Ai T, He Y, et al. Spatial co-location pattern mining of facility points-of-interest improved by network neighborhood and distance decay effects[J]. International Journal of Geographical Information Science, 2016,31(2):280-296.
- [84] Celik M, Kang J M, Shekhar S. Zonal co-location pattern discovery with dynamic parameters[C]// Proceedings of the Seventh International Conference on Data Mining (ICDM 2007). IEEE, 2007:433-438.
- [85] Qian F, Chiew K, He Q, et al. Mining regional co-location patterns with kNNG[J]. Journal of Intelligent Information Systems, 2014,42(3):485-505.
- [86] Eick C F, Parmar R, Ding W, et al. Finding regional co-location patterns for sets of continuous variables in spatial datasets[C]//Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, 2008:30.
- [87] Mohan P, Shekhar S, Shine J A, et al. A neighborhood graph based approach to regional co-location pattern discovery: a summary of results[C]// Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2011: 122-132.
- [88] Cressie N A, Wikle C. Statistics for spatio-temporal data [J]. Chance, 2011,27(2):64.
- [89] Pfeifer P E, Deutch S J. A three-stage iterative procedure for space-time modeling[J]. Technometrics, 1980,22(1): 35-47.
- [90] Brunson C, Fotheringham A S, Charlton M E. Geographically weighted regression: A method for exploring spatial nonstationarity[J]. Geographical Analysis, 1996,28(4): 281-298.
- [91] Huang B, Wu B, Barry M. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices[J]. International Journal of Geographical Information Science, 2010,24(3):383-401.
- [92] Deng M, Yang W, Liu Q L, et al. Heterogeneous space-time artificial neural networks for space-time series prediction[J]. Transactions in GIS, 2018,22(1):183-201.
- [93] Cheng T, Wang J. Accommodating spatial associations in DRNN for space-time analysis[J]. Computers, Environment and Urban Systems, 2009,33(6):409-418.
- [94] Deng M, Yang W, Liu Q. Geographically weighted extreme learning machine: A method for space-time prediction[J]. Geographical Analysis, 2017,49(4):433-450.
- [95] Zhang J, Zheng Y, Qi D, et al. Predicting citywide crowd flows using deep spatio-temporal residual networks[J]. Artificial Intelligence, 2018,259:147-166.
- [96] 李小文,曹春香,张颢.尺度问题研究进展[J].遥感学报, 2009,13(s1):12-20. [ Li X W, Cao C X, Zhang H. Progress in scale research[J]. Journal of Remote Sensing, 2009,13(s1):12-20. ]
- [97] Goodchild M F. Scale in GIS: An overview[J]. Geomorphology, 2011,130(1):5-9.
- [98] 孙庆先,李茂堂,路京选,等.地理空间数据的尺度问题及其研究进展[J].地理与地理信息科学,2007,23(4):53-56. [ Sun Q X, Li M T, Lu J X, et al. Scale issue and its research progress of geospatial data[J]. Geography and Geo-Information Science, 2007,23(4):53-56. ]
- [99] Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002,24(5):603-619.
- [100] 李志林,刘启亮,唐建波.尺度驱动的空间聚类理论[J].测绘学报,2017,46(10):1534-1548. [ Li Z L, Liu Q L, Tang J B. Towards a scale-driven theory for spatial clustering[J]. Acta Geodaetica et Cartographica Sinica, 2017, 46(10):1534-1548. ]
- [101] Liu Q, Li Z, Deng M, et al. Modeling the effect of scale on clustering of spatial points[J]. Computers, Environment and Urban Systems, 2015,52:81-92.
- [102] 唐建波,刘启亮,邓敏,等.空间层次聚类显著性判别的重排检验方法[J].测绘学报,2016,45(2):233-240. [ Tang J B, Liu Q L, Deng M, et al. A permutation test for identifying significant clusters in spatial dataset[J]. Acta Geodaetica et Cartographica Sinica, 2016,45(2):233-240. ]
- [103] Shi Y, Gong J, Deng M, et al. A graph-based approach

- for detecting spatial cross-outliers from two types of spatial point events[J]. *Computers, Environment and Urban Systems*, 2018,72:88-103.
- [104] Deng M, Yang X, Shi Y, et al. A non-parametric statistical test method to detect significant cross-outliers in spatial points[J]. *Transactions in GIS*, 2018,22(6):1462-1483.
- [105] Cai J, Deng M, Liu Q, et al. Nonparametric significance test for discovery of network-constrained spatial co-location patterns[J]. *Geographical Analysis*, 2019,51(1):3-22.
- [106] Fortin M J, Dale M R T. *Spatial analysis: a guide for ecologists*[M]. Cambridge University Press, 2005.
- [107] Cai J, Deng M, Liu Q, et al. A statistical method for detecting spatiotemporal co-occurrence patterns[J]. *International Journal of Geographical Information Science*, 2019, 33(5):967-990.
- [108] Neyman J, Scott E L. Statistical approach to problems of cosmology[J]. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1958,20(1):1-29.
- [109] Deng M, He Z, Liu Q, et al. Multi-scale approach to mining significant spatial co- location patterns[J]. *Transactions in GIS*, 2017,21(5):1023-1039.
- [110] Wiegand T, He F, Hubbell S P. A systematic comparison of summary characteristics for quantifying point patterns in ecology[J]. *Ecography*, 2013,36(1):92-103.
- [111] Cai J, Liu Q, Deng M, et al. Adaptive detection of statistically significant regional spatial co-location patterns[J]. *Computers, Environment and Urban Systems*, 2018,68: 53-63.
- [112] Yang W T, Deng M, Xu F, et al. Prediction of hourly PM<sub>2.5</sub> using a space-time support vector regression model[J]. *Atmospheric Environment*, 2018,181:12-19.
- [113] Wang J F, Zhang T L, Fu B J. A measure of spatial stratified heterogeneity[J]. *Ecological Indicators*, 2016,67:250-256.
- [114] Prasad S K, Aghajarian D, McDermott M, et al. Parallel processing over spatial-temporal datasets from geo, bio, climate and social science communities: a research roadmap[C]// *Proceedings of the 2017 IEEE International Congress on Big Data (Bigdata Congress)*. IEEE, 2017: 232-250.
- [115] 孙中伟,路紫.流空间基本性质的地理学透视[J].*地理与地理信息科学*,2005,21(1):109-112. [ Sun Z W, Lu Z. A geographical perspective to the elementary nature of space of flows[J]. *Geography and Geo-Information Science*, 2005,21(1):109-112. ]
- [116] 程昌秀,史培军,宋长青,等.地理大数据为地理复杂性研究提供新机遇[J].*地理学报*,2018,73(8):1397-1406. [ Cheng C X , Shi P J, Song C Q , et al. Geographic big-data: A new opportunity for geography complexity study [J]. *Acta Geographica Sinica*, 2018,73(8):1397-1406. ]