

基于大数据的城市人地关系分析与应用计算平台 ——2018年中国地理学会地理大数据计算环境“优秀实用案例”

薛冰^{1,2*}, 李京忠^{2,3}, 肖骁^{1,2}, 谢潇^{1,2}, 庞敏^{1,2,4},
姜璐^{1,5}, 逯承鹏^{1,2}, 任婉侠^{1,2}

1. 中国科学院污染生态与环境工程重点实验室(沈阳应用生态研究所), 沈阳 110016; 2. 辽宁省环境计算与可持续发展重点实验室, 沈阳 110016; 3. 许昌学院城乡规划与园林学院, 许昌 461000; 4. 西北师范大学地理与环境科学学院, 兰州 730070; 5. 兰州大学资源环境学院, 兰州 730000

摘要: 以兴趣点(Points of Interest, POI)为主要代表的地理空间大数据对于人地关系研究具有重要革新意义。研发大数据多维语义关联模型、语义关联元数据索引及组织技术以及大数据可视化算法等关键技术体系,同时以基于电子地图POI、网络文本大数据的城市商住空间格局及关联关系计算,运用空间核密度估计等GIS技术分析城市人地系统要素的空间聚集形态及相关性特征,并揭示要素之间的关联机理建立数据计算平台。本平台为利用大数据进行人-地关系研究提供计算环境,为创新人地关系的大数据分析方法与理论体系,促进科学知识发现(人地关系)与决策应用服务(政策调控)提供实验基地。

关键词: 人地关系; 空间格局; POI; 空间分析; 沈阳市

DOI: 10.3974/geodp.2018.03.06

1 前言

全面推进大数据发展和应用,加快建设数据强国,已经成为我国的国家战略^[1]。大数据应用正快速发展为对数量巨大、来源分散、格式多样的数据进行采集、存储和关联分析,从中发现新知识、创造新价值、提升新能力的新一代信息技术和服务业态^[2]。地学大数据是生态文明建设、生态环境保护与区域可持续发展工作的关键技术支撑内容之一^[3-4],也是支撑智慧城市建设与相关地理信息产业发展的核心科技要素^[5-8]。随着空间信息技术、计算机技术和网络服务技术的快速发展,使地学大数据逐步从简单空间坐标信息发展为一类包含丰富环境语义内涵的新型地理空间数据类型^[9-11],成为当前国际地理空间信息和环境领域的新兴产业增长点和研究热点^[12-14]。地学大数据赋予数据空间特征,体现出区域人地系统因素中的广度、宽度及深度以及相关信息的挖掘和分析,在公共环境政策评价、环境监测分析、区域可持续发展等热点应用中扮演着越来越重要的角色^[15-16]。地学大数据的研究

收稿日期: 2018-07-20; 修订日期: 2018-09-20; 出版日期: 2018-09-25

基金项目: 国家自然科学基金(41471116, 41701142, 41701466); 沈阳市科技局重点科技研发计划(17-117-6-00)及双百工程(Z17-7-030); 中国科学院青年创新促进会(会员号: 2016181)

*通讯作者: 薛冰 D-1830-2009, 中国科学院沈阳应用生态研究所, xuebing@iae.ac.cn

作者ID: 薛冰 D-1830-2009, 李京忠 S-3218-2018, 肖骁 S-3189-2018, 谢潇 S-3278-2018, 庞敏 S-3205-2018, 姜璐 S-9924-2018, 逯承鹏 S-3201-2018, 任婉侠 S-3252-2018

引用格式: 薛冰, 李京忠, 肖骁等. 基于大数据的城市人地关系分析与应用计算平台——2018年中国地理学会地理大数据计算环境“优秀实用案例”[J]. 全球变化数据学报, 2018, 2(3): 290-294. DOI: 10.3974/geodp.2018.03.06.

也是构建虚拟生态系统的基础理论体系，特别是在数据支撑体系和空间分析应用方面具有重要的实际意义。以兴趣点（Points of Interest, POI）为主要代表的地理空间大数据对于人地关系研究具有重要革新意义，且主要体现在其具有的人类经济活动及范围的表征能力^[17]。

“基于大数据的城市人地关系分析与应用计算平台”是基于 POI 大数据的城市经济地理要素空间格局及关联关系为实证研究案例，期望实现如下科学目标：第一，在理论研究上，明晰大数据用于人地关系研究的现状及未来潜力研究，并为今后利用大数据进行人地关系研究提供借鉴；第二，为创新应用地学大数据开展城市人地要素布局及要素之间的内在关联机制研究提供参考，同时也为优化区域人地系统及政策调控提供决策参考；第三，发展应用包含多维语义关联建模技术等在内的关键技术体系及可视化技术，有效解决海量数据的重叠掩盖缺点，为探寻城市人地系统宏观和微观多尺度层面的分布规律提供数据支撑和决策支持。

2 数据源

以 2017 年为基准年，采集包括沈阳市全域范围内经济地理全类型要素 POI 数据 43.37 万条（18 个大类）。其中，购物服务类 POI 信息最多，占总数的 31% 以上；餐饮服务、生活服务、公司企业 POI 信息量占比总数的 40%；其余类型依次为医疗保健服务、政府机构及社会团体、科教文化服务、交通设施服务、汽车服务等。POI 数据采用 GCJ-02 火星坐标系。此外，基于网络文本抓取分析获取平均房价数据和住宅年代等非结构化数据 1,279 条。

3 技术路线与计算方法

将沈阳市 POI 数据按照国家行业标准进行分类，利用核密度估计等空间统计分析方法，提取城市空间结构与各行业聚集特征信息，并开展两类空间经济要素之间的关联特征挖掘（图 1）。以零售业与住宅为例，在挖掘二者空间关联关系基础上，揭示住宅价格空间异质性及影响机制（图 2）。计算方法包括：

（1）核密度估计（Kernel Density Estimation, KDE）方法：由 Rosenblatt 和 Emanuel Parzen^[18] 最先提出，基本原理是以每个样点 $i(x, y)$ 为中心，通过核函数计算出每个样点在指定半径范围内对各个栅格单元中心点的密度贡献值，搜索半径范围的栅格单元中心点距离样点越近，其样点的密度贡献值越大。本案例将零售业 POI 核密度分布图与行政区划图叠加，识别出零售商店核密度值显著高于周围栅格的中心区域，将其定义为“零售商业中心”。

（2）空间自相关分析方法（局域 Getis-Ord G_i^* 指数）：空间自相关分析方法由 Ord 和 Getis 提出^[19]，是一种基于距离权矩阵的局部空间自相关指标，能探测高值聚集和低值聚集。高值聚集区称为热点，低值聚集区称为冷点。本案例利用该方法探究零售业热点街区空间分布。

（3）相关性分析方法：通常用来判断两变量之间是否存在相互关系，并解释相互关系的密切程度。相关系数则是衡量两变量之间相关性的基本指标。本案例通过计算沈阳市零售业与住宅核密度的相关系数来衡量零售业与住宅空间的空间关联特征。

（4）缓冲区分析方法：以 POI 数据为基础，自动建立周围一定宽度范围内的缓冲面，然后与目标图层进行叠加，进行分析得到所需结果。本案例将 3 km 作为沿道路网步行到购

物中心的最大距离，在“3 km”沿线距离范围内的区域视为购物中心的服务区。

(5) 地理加权回归模型 (Geographically Weighted Regression, GWR): 是由英国学者 Fotheringham 在研究空间非平稳性时提出的新方法^[20], 它的主要优势在于把空间权重矩阵应用到线性回归模型中, 通过引入地理坐标位置对传统的回归模型进行了扩展, 具有灵活的局域调整优化权重功能, 其模型系数能较好揭示地理要素的空间非均质性。本案例运用 GWR 模型探究各因素对房价的影响因子以及各因素回归系数的空间异质性。

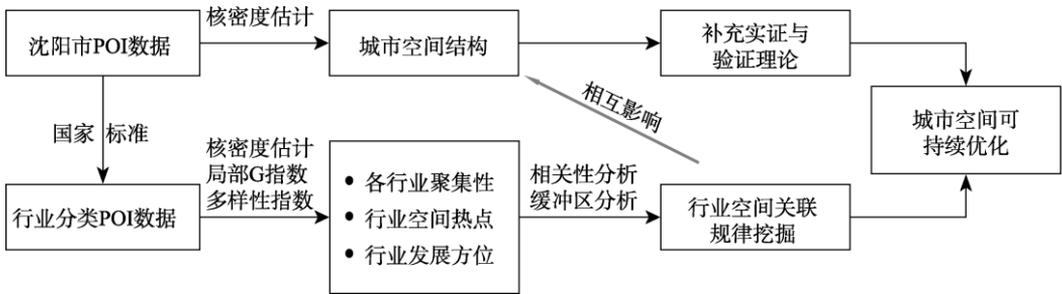


图1 基于POI的城市人地关系分析与应用

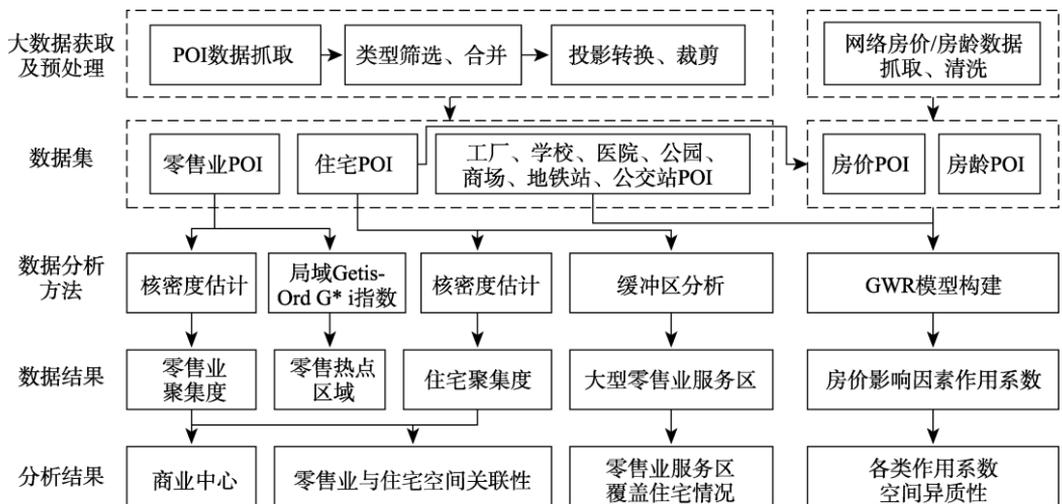


图2 城市经济地理要素空间关联规律挖掘分析框架图(以零售业、住宅为例)

4 主要应用及其成果

本平台的建设主要是为沈阳老工业基地城市经济活动和人文活动分析提供计算平台, 从而有助于从全局范围内揭示城市经济要素的空间集聚形态及关联性。沈阳市零售商业中心在空间分布上形成块状聚集、多中心发展的格局; 各零售业态发展不均衡, 商场商厦过于集中于核心商区, 超市业态的聚集区域分布广泛, 便利店的分布与整个零售业态的分布特征相似; 验证了中心地理论中的中心地等级划分思想以及零售业区位选择的聚集理论。零售业的区位分布与人口的分布密度规律一致, 验证了冯·杜能提出的竞价租金理论。房龄、住宅密度、公共交通、公共配套设施等对房价有显著的提升作用, 而工业企业等对房价有一定的抑制作用, 证实了亨利·乔治定理, 即公共物品投资空间差异是导致城市内部

住宅地价差异的原因之一，并可以利用这一理论指导老工业区的可持续空间重构。

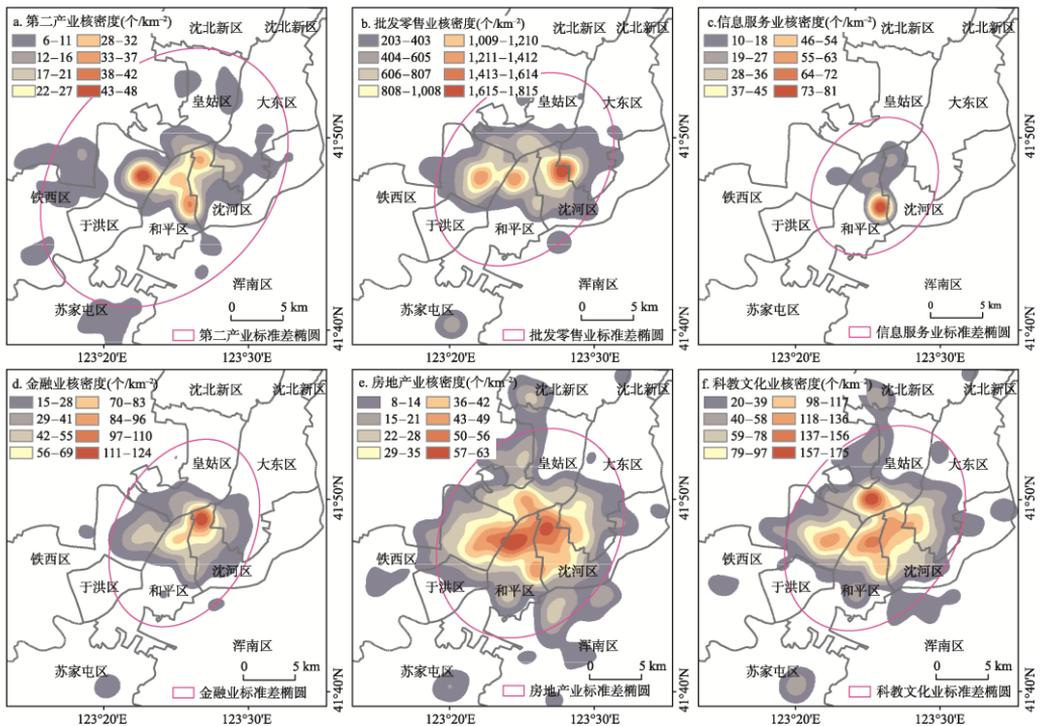


图 3 沈阳市分行业经济地理要素格局分布图

在实践上，从老工业基地全面振兴建设着手，结合工业-资源型城市发展实际，构建区域人地系统多尺度表征体系并建模解析其耦合演化过程，在数据源上采用大数据与传统数据相结合，在方法学集成上采用模块分析与体系集成相结合，在路线设计上采用方法研究与实证研究相结合，这充分反映了生态文明建设过程中城市人地系统及其要素多尺度嵌套和复合性，突破了传统的较为侧重于单一主体的研究手段，有助于丰富可持续发展科学体系。例如，研究结果表明沈阳市各行业空间错位特征较为显著，且在外围城区形成若干散状聚集点，行业密度呈现圈层扩散特征（图 3）。此外，从微观尺度上开展城市住宅与零售业的空间关联量化研究（图 4），可为政府、规划部门以及购房者提供更加完整的决策参考。同时还尝试把以政策导向为主的科学机理研究与地方政策决策服务相结合，以弥补科学与政策之间的距离。

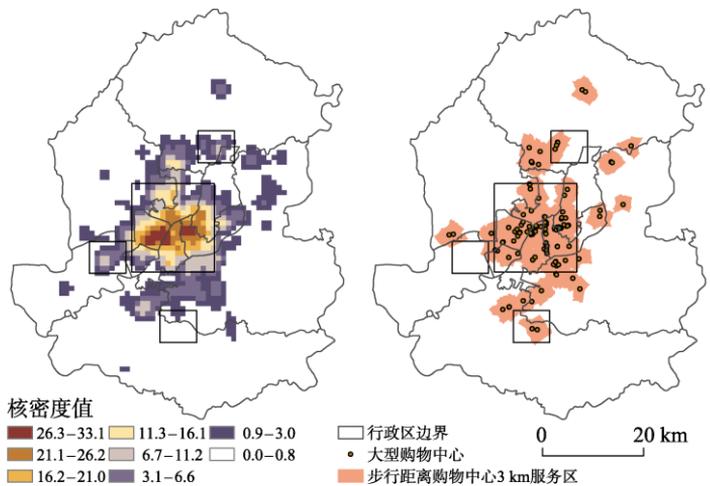


图 4 沈阳市住宅核密度与大型购物服务区对比图

5 结论

采用数据与计算相结合的方法开展大数据及其在人地分析中的应用,有助于分析大数据应用于人地关系研究的潜力、明晰基于大数据的人地关系研究的水平,为今后利用大数据进行人-地关系研究提供借鉴和启示。“基于大数据的城市人地关系分析与应用计算平台”将大数据与GIS相结合,对零售业与住宅空间集聚特征及热点分布开展量化关联表达和分析,并探究老工业区房价驱动因素影响效应的空间异质特征,可帮助政府和相关部门了解房产市场的供需关系、监测老工业区搬迁改造后区域的发展成效、促进老工业城市空间的重构和可持续发展。

参考文献

- [1] 人民网-人民日报. 习近平在中国共产党第十九次全国代表大会上的报告[OL]. <http://cpc.people.com.cn/n1/2017/1028/c64094-29613660.html>, 2017-10-28.
- [2] 李国杰, 程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. 中国科学院院刊, 2012, 27(6): 647-657.
- [3] 吴志峰, 柴彦威, 党安荣等. 地理学碰上“大数据”: 热反应与冷思考[J]. 地理研究, 2015, 34(12): 2207-2221.
- [4] 李德仁. 展望大数据时代的地球空间信息学[J]. 测绘学报, 2016, 45(4): 379-384.
- [5] Ivan, I., Singleton, A., Horák, J., et al. The Rise of Big Spatial Data [M]. Switzerland: Springer International Publishing, 2016.
- [6] 谢潇. 语义感知的地理视频大数据自适应关联组织方法[J]. 测绘学报, 2016, 45(10): 1260.
- [7] 周顺平, 徐枫. 大数据环境下地理信息产业发展的几点思考[J]. 地理信息世界, 2014, 21(1): 45-50.
- [8] 刘瑜, 詹朝晖, 朱逸等. 集成多源地理大数据感知城市空间分异格局[J]. 武汉大学学报(信息科学版), 2018, 43(3): 327-335.
- [9] McKenzie, G., Janowicz, K., Gao, S., et al. How where is when? On the regional variability and resolution of geosocial temporal signatures for points of interest [J]. *Computers Environment and Urban Systems*, 2015, 54: 336-346.
- [10] Gao, S., Janowicz, K., Couclelis, H. Extracting urban functional regions from points of interest and human activities on location-based social networks [J]. *Transactions in GIS*, 2017, 21(3): 446-467.
- [11] Kunze, C., Hecht, R. Semantic enrichment of building data with volunteered geographic information to improve mappings of dwelling units and population [J]. *Computers Environment and Urban Systems*, 2015, 53: 4-18.
- [12] Becker, R. A., Caceres, R., Hanson, K., et al. A tale of one city: using cellular network data for urban planning [J]. *IEEE Pervasive Computing*, 2011, 10(4): 18-26.
- [13] Sagl, G., Resch, B., Hawelka, B., et al. From social sensor data to collective human behaviour patterns: Analysing and visualising spatiotemporal dynamics in urban environments [C]. *Proceedings of the GI-Forum 2012: Geovisualization, Society and Learning*, 2012: 54-63.
- [14] 李德毅. 大数据时代的位置服务[J]. 测绘科学, 2014, 39(8): 3-6.
- [15] 甄峰, 秦萧, 席广亮. 信息时代的地理学与人文地理学创新[J]. 地理科学, 2016, 35(1): 11-18.
- [16] 龙瀛, 刘伦. 新数据环境下定量城市研究的四个变革[J]. 国际城市规划, 2017, 31(1): 64-73.
- [17] 薛冰, 肖骁, 李京忠等. 基于 POI 大数据的城市零售业空间热点分析: 以辽宁省沈阳市为例[J]. 经济地理, 2018, 38(5): 36-43.
- [18] 王法辉. 基于 GIS 的数量方法与应用[M]. 北京: 商务印书馆, 2009.
- [19] Ord, J. K., Getis, A. Local spatial autocorrelation statistics: distributional issues and an application [J]. *Geographical Analysis*, 1995, 27(4): 286-306.
- [20] Fotheringham, A. S., Brunson, C., Charlton, M. Geographically weighted regression: the analysis of spatially varying relationships [J]. *American Journal of Agricultural Economics*, 2004, 86(2): 554-556.