

基于模糊理论的局部空间同位模式挖掘方法

段韶鹏¹ 娄文璐² 张金鹏³ 张 占⁴

DUAN Shaopeng LOU Wenlu ZHANG Jinpeng ZHANG Zhan

摘要

地理空间具有显著的异质性，不同区域的地理属性分布不均衡，导致全局分析方法难以精准定位具有相似地理属性的局部区域，挖掘结果的准确性不足。为此，文章提出了一种基于模糊理论的局部空间同位模式挖掘方法。首先，采用多层次融合整合地理属性和特征，形成全面的地理属性空间表达。通过参与指数评估空间同位模式普遍性，筛选候选局部同位模式集合。然后，针对该集合，运用模糊理论处理地理空间异质性和不确定性，构建模糊上下近似描述地理属性，提取相似地理属性的局部感兴趣区域。利用图结构模型挖掘局部空间同位模式，将区域视为节点，通过边和权重体现相似性或空间关联，构建邻接矩阵。最后，采用谱聚类技术分析图结构模型，观察聚类簇特性，深入了解局部空间同位模式分布规律和特征，从而实现挖掘。实验结果表明，所提方法准确率为0.945，决定系数为0.932，相较于对照组方法显著提升，证明了其可行性和可靠性。

关键词

模糊理论；同位模式；地理属性；图结构模型；谱聚类技术；模式挖掘

doi: 10.3969/j.issn.1672-9528.2025.07.040

0 引言

在空间数据分析领域，局部空间同位模式的挖掘对于理解地理现象的空间分布和相互关联具有重要意义。随着空间数据的快速增长，如何高效地挖掘这些模式成为研究热点。

1. 河南开放大学 河南郑州 450008
 2. 云南财经大学商学院 云南昆明 650221
 3. 云南财经大学信息学院 云南昆明 650221
 4. 长沙民政职业技术学院软件学院 湖南长沙 410004
- [基金项目] 2025年河南省科技攻关项目(252102210149)；云南省基础研究计划面上项目(202501AT070455)；云南省教育厅科学研究基金项目青年人才基础研究专项(2024J0643)；湖南省自然科学基金资助项目(2024JJ8025)

然而，现有的同位模式挖掘方法在处理复杂空间数据时，往往存在效率低下的问题。刘宇情等人^[1]提出了基于网格空间团的多级同位模式挖掘方法，通过划分网格并构建空间团来挖掘同位模式。在地理空间异质性显著的情况下，固定的网格划分难以适应不同区域的地理属性分布不均衡，导致密集区域出现过噪声数据，影响挖掘的准确性和全面性。王靖涵等人^[2]则采用了基于图结构的局部空间同位模式挖掘方法，利用图模型表示空间关系，并通过图算法进行挖掘。在定位具有相似地理属性的局部区域时，该方法依赖于节点或子图的相似性度量。由于相似性度量的不精确，导致局部区域的相似性被错误评估。吴静等人^[3]从全局视角筛选不频繁同位模式为候选，通过粗糙集处理其实例位置特点，对局部自然分

2024, 2 (2):134-149.

- [6] 翟社平, 霍媛媛, 杨锐, 等. 基于一致性哈希和随机选取的PBFT算法改进[J]. 计算机工程与应用, 2024, 60(12):294-302.
- [7] CASTRO M, LISKOV B. Practical byzantine fault tolerance [C]//OSDI '99: Proceedings of the Third Symposium on Operating Systems Design and Implementation. New York: ACM, 1999: 173-186.
- [8] 宋宇哲, 郑广海, 张鑫. RG-BFT: 基于随机分组的拜占庭容错算法[J]. 计算机工程与设计, 2024, 45 (6): 1661-1667.
- [9] 陈珩, 黄世成, 郑明辉. 一种双主节点的PBFT共识算法[J].

湖北民族大学学报(自然科学版), 2024, 42 (2): 185-190.

【作者简介】

周岩龙(2001—), 男, 河北廊坊人, 硕士研究生, 研究方向: 区块链共识算法、大数据计算, email: 1164460643@qq.com.

宋淑彩(1970—), 女, 河北张家口人, 博士, 研究生导师、教授, 研究方向: 大数据技术、计算机视觉, email: ssc1330@hebiace.edu.cn.

(收稿日期: 2025-02-04 修回日期: 2025-07-11)

布区域进行探测，并评估其频繁出现的程度，进而生成局部空间的同位分布模式。由于地理属性的异质性，全局筛选和粗糙集处理无法适应局部区域特征，导致同位模式挖掘的遗漏或误判，影响结果的全面性和准确性。蒋希文^[4]采用新密度度量函数，设计并行区域同位模式挖掘算法。由于地理属性分布不均衡，固定的区域划分策略难以适应不同区域特征，导致密集区域噪声数据过多，而稀疏区域遗漏重要信息，影响挖掘的准确性和全面性。针对上述问题，本文提出基于模糊理论的局部空间同位模式挖掘方法，旨在提升挖掘效率和准确性，为空间数据分析提供新视角。

1 基于模糊理论的局部空间同位模式挖掘方法

1.1 基于多层次融合的候选局部同位模式筛选

空间同位模式指的是一种地理现象，即一组地理事件倾向于在相邻区域内频繁共同出现并相互关联。由于这种地理现象具有复杂性和多样性，单一层次的表达方式无法充分捕捉地理空间的全部复杂性，导致模式挖掘结果的不准确^[5]。因此，为更全面地反映各区域的地理特征，本研究采用多层次融合的方式来表达地理属性空间。通过多层次融合，能够将不同层次的地理属性和特征进行整合，形成一个更全面、更准确的地理属性空间表达，为局部空间同位模式的挖掘提供更丰富、更准确的数据基础。

针对空间特征集 $F = \{f_1, f_2, \dots, f_n\}$ ，为评估其 K 阶候选空间同位模式中各个元候选模式在全局范围内的普遍性，采用参与指数作为评估标准。计算公式为：

$$\begin{aligned} PR(P, f_i) &= \frac{|\pi_{f_i}(\text{table_instance}(k))|}{|\text{table_instance}(\{f_i\})|} \\ PI(P) &= \min_{i=1}^k \{PR(P, f_i)\} \end{aligned} \quad (1)$$

式中： π 表示去除重复项关系的投影操作； $PRF = (P, f_i)$ 表示高阶同位模式 $P = \{f_1, f_2, \dots, f_k\}$ 中某个空间特征的参与率； table_instance 表示实例集合，即空间同位模式的实例。将未满足预设标准的空间同位模式归类为候选的局部同位模式集合。由于空间中存在的异质性，各种空间同位模式的分布呈现出差异性，这些模式的实例更倾向于在地理属性相近的区域内共同出现^[6]。为更全面地分析这些候选局部同位模式，针对候选集中的每一个空间同位模式，将每个热点区域视为一个实体，其多层次的地理特征构成了该实体的特征集合 A ；这些实体共同构成了一个集合 U ，其特征信息则通过多层次融合的方式整合成了一个更全面、更准确的模糊数据集，为局部空间同位模式的挖掘提供了更丰富、更准确的数据基础。

1.2 基于模糊理论的局部感兴趣区域提取

地理现象表现出的空间异质性源自空间特征分布的不均衡，这使得在识别同位模式的分布区域时，仅仅依靠确定的地理属性或规则不足以准确捕捉局部区域的特性。模糊理论因其能够处理不确定性和模糊性，特别适用于描述和分析这种复杂多变的地理空间现象^[7]。为了科学识别同位模式分布区域，减少人为干扰，本研究采用模糊理论提取局部感兴趣区域。通过灵活处理地理属性相似性和差异性，准确定位相似地理属性的局部区域，作为同位模式分布的重要候选。

为准确描述和分析同位模式分布区域，本研究利用选定的属性集 B 构建对象 X 的模糊上下近似，从而对 X 进行描述。模糊上下近似的构造可以用公式表示为：

$$\begin{aligned} \underline{B}X &= \{x | [x]_B \subseteq X\} \\ \overline{B}X &= \{x | [x]_B \cap X \neq \emptyset\} \end{aligned} \quad (2)$$

式中： B 表示整体的属性集； U 表示对象集； X 表示对象集，即地理空间中的各个区域； $[x]$ 表示所有与对象 X 在属性集 B 上满足模糊相似关系的元素所构成的一个模糊集合。

在本文中，地理空间被视为对象集合 U ，而整合后的地理属性数据则构成了属性集合 A 。局部同位模式的热点区域被视为对象 X ，其地理属性数据则组成了相应的属性集合 $[x]_B$ 。

针对每一个局部候选同位模式，本研究均采用模糊上下近似方法来抽取其热点区域属性集合，并将这些抽取出的属性集合全部作为样本数据用于训练模糊规则^[8]。

上述方法能挂钩有效提取出具有相似地理属性的局部区域，为局部空间同位模式的深入分析与应用提供有力支持。

1.3 图结构模型与谱聚类结合的局部空间同位模式挖掘

针对已识别的局部感兴趣区域，传统方法难以全面表达区域关联性，导致模式挖掘不完善。本文提出用图结构模型挖掘局部空间同位模式的新方法。图结构模型能直观展示区域间空间关系和相似性，通过节点和边体现关联程度，形成邻接矩阵。在此基础上，采用谱聚类技术，计算度矩阵和拉普拉斯矩阵，分解特征值构成新特征空间，通过聚类簇观察特性，揭示同位模式分布规律。

首先，构建一个图结构模型，其中每个局部感兴趣区域被视作图中的一个节点，节点集合表示为 $D = \{d_1, d_2, \dots, d_n\}$ ， n 为局部感兴趣区域的总数。节点间的相似性或空间关联通过边来体现，边的权重 ω_{ij} 用以衡量区域 d_i 与 d_j 之间的关联紧密程度^[9]。权重计算公式为：

$$\omega_{ij} = \alpha \cdot \sin_{\text{geo}}(d_i, d_j) + \beta \cdot \sin_{\text{elec}}(d_i, d_j) \quad (3)$$

式中： α 、 β 表示权重调整系数； $\sin_{\text{geo}}(d_i, d_j)$ 、 $\sin_{\text{elec}}(d_i, d_j)$ 分别表示区域 d_i 与 d_j 的地理位置的相似度函数。

构建完图结构模型后，本文采用谱聚类技术进行局部空

间同位模式的挖掘^[10]。根据权重 ω_{ij} ，可以进一步计算出节点的度矩阵 Z ，其表达式为：

$$Z = \sum_{j=1}^n \omega_{ij} \cdot Z_{ij} \quad (4)$$

式中： Z_{ij} 表示局部感兴趣区域节点 d_i 的度，即与节点 d_i 相连的所有边的权重总和。

在上述的基础上，构造出拉普拉斯矩阵 P ，其计算公式为：

$$P = Z - W \quad (5)$$

对拉普拉斯矩阵 L 分解特征值，选前 k 个最小非零特征值对应的向量，构成新特征空间。

在这个新的特征空间中，对各个局部感兴趣区域节点进行聚类分析，得到 k 个聚类簇。通过观察聚类簇内节点连接强度、形态等特性，可深入了解不同簇内局部空间同位模式分布规律和特征，实现挖掘。

2 实验结果与讨论

2.1 实验说明

为全面验证基于模糊理论的局部空间同位模式挖掘方法的有效性与优越性，本研究设计并实施了一项细致的仿真对比实验。实验中，选择了两种具有代表性的局部空间同位模式挖掘方法作为对照组：一种是基于网格空间团的多级同位模式挖掘方法（对照组1），该方法通过划分网格并构建空间团来挖掘同位模式，在处理某些特定类型的空间数据时具有一定优势；另一种是基于粗糙集的局部同位模式挖掘算法（对照组2），该算法利用粗糙集理论处理空间数据的不确定性和模糊性，适用于处理含有噪声和不确定性的数据集。

本实验的主要目的是，在相同的数据集和实验环境下，对各方法在处理复杂空间数据时挖掘局部同位模式的能力、准确性以及效率进行全面、客观的比较和分析。通过对比实验结果，可以直观地评估基于模糊理论的局部空间同位模式挖掘方法的性能表现，进一步验证其可行性。

2.2 搭建实验环境

本次实验的仿真环境搭建完成，具体配置参数详见表1。

表1 实验环境配置表

序号	配置名称	具体配置
1	处理器	Intel Core i7-10700K
2	内存	32 GB
3	显卡	NVIDIA GeForce RTX 3080
4	编程语言环境	Python 3.8
5	深度学习框架	TensorFlow 2.5
6	机器学习库	scikit-learn 0.24
7	数据处理库	Pandas 1.2.4
8	数值计算库	NumPy 1.19.5

根据表1所示的配置，搭建了用于基于模糊理论的局部空间同位模式挖掘方法的实验环境。

2.3 实验数据集

实验采用了某领域提供的真实空间数据作为数据集，包含了1500个局部空间区域的同位模式样本数据。这些数据集被划分为6个不同的空间行为模式，每个模式下的训练样本和测试样本数量分布如表2所示。

表2 实验数据分布表

标签	空间行为模式	训练样本数	测试样本数
1	城区居住模式	170	40
2	商业区活动模式	190	45
3	工业区生产模式	210	50
4	公园休闲模式	205	55
5	教育机构学习模式	225	60
6	交通枢纽流通模式	235	65

随后，将基于模糊理论的局部空间同位模式挖掘方法以及对照组中的两种方法，分别应用于该实验数据集，进行空间行为模式的挖掘与预测工作。最后，对实验结果进行了对比分析。

2.4 实验指标

在本次实验中，为直观且客观地评估所提方法与对比方法的性能，本文分别采用准确率 A 与决定系数 R^2 作为挖掘结果的评价指标，具体定义为：

$$A = \frac{1}{N_0} \sum_{i=1}^{N_0} \gamma_i 1(y'_i = y_i)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N_0} (Y'_i - Y_i)^2}{\sum_{i=1}^{N_0} (\bar{Y} - Y_i)^2} \quad (6)$$

式中： A 表示局部空间同位模式挖掘结果的准确率； N_0 表示空间单元的总数； γ_i 表示第 i 个空间单元中同位模式在模糊隶属度函数下的概率分布； y'_i 、 y_i 分别表示第 i 个空间单元的同位模式挖掘结果与实际观测结果； R^2 表示衡量局部空间同位模式挖掘结果的拟合优度， R^2 值越大，表明挖掘结果与实际数据的拟合效果越佳，即挖掘精度越高； Y'_i 、 Y_i 分别表示第 i 个空间单元的同位模式预测值与实际观测值； \bar{Y} 表示所有空间单元同位模式的平均值。

通过上述指标，可以量化评估基于模糊理论的局部空间同位模式挖掘方法的有效性及准确性。

2.5 实验结果分析

在完成局部空间同位模式挖掘后，采用上述公式分别计算了3种方法所得挖掘结果的准确率以及模式识别效果的决

定系数，作为本次实验的结果。局部空间同位模式挖掘结果的准确率对比结果如图 1 所示。

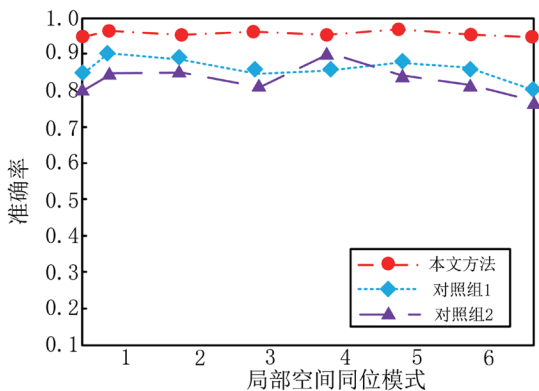


图 1 局部空间同位模式挖掘结果

如图 1 所示，所提方法局部空间同位模式挖掘准确率达 0.945，较对照组两种方法分别提升 0.090 和 0.115。这主要得益于图结构模型与谱聚类技术的结合，能更精确挖掘同位模式。图结构模型直观表达区域关联性，谱聚类技术将相似区域聚类，使结果更具可解释性。传统方法难以表达区域关联性，而此结合能全面揭示复杂关联，提升挖掘全面性和深度。

在局部空间同位模式挖掘中，决定系数用于评估挖掘结果与实际数据之间的拟合程度，从而证明挖掘效果的优劣。不同方法的决定系数对比结果如图 2 所示。

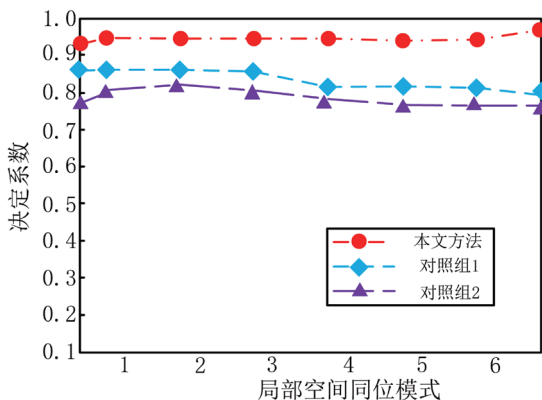


图 2 电力用户行为模式挖掘结果

图 2 展示局部空间同位模式识别效果的决定系数，实验结果显示，本文方法决定系数达 0.932，较对照组分别提升 0.100 和 0.140。这一提升归因于图结构模型与谱聚类技术的结合，能更精确挖掘同位模式。通过观察聚类簇内节点连接强度、形态等特征，可揭示同位模式内在规律，为模式识别提供精确依据。

3 结语

本文提出基于模糊理论的局部空间同位模式挖掘方法，通过多层次融合表达地理属性空间，利用模糊理论提取局部

感兴趣区域，并构建图结构模型挖掘模式。该方法能揭示同位模式内在规律，为空间数据分析提供新视角。虽具理论和应用优势，但存在不足，如模糊规则构建和参数选择受主观影响，图结构模型在大规模数据处理中面临复杂度与计算效率挑战。未来研究将优化模糊规则自动化生成，增强模型鲁棒性和适应性，提升方法在大规模空间数据中的实用性和可扩展性。

参考文献：

- [1] 刘宇情, 王丽珍, 杨培忠, 等. 基于网格空间团的多级同位模式挖掘方法 [J]. 浙江大学学报 (工学版), 2024, 58(5): 918-930.
- [2] 王靖涵, 艾廷华, 吴昊, 等. 基于图结构的局部空间同位模式挖掘 [J]. 测绘学报, 2024, 53(4): 724-735.
- [3] 吴静, 傅俊杰, 程朋根. 基于粗糙集的局部同位模式挖掘算法 [J]. 测绘通报, 2022(10): 80-85.
- [4] 蒋希文, 王丽珍, Vanha TRAN. 基于模糊密度峰值聚类的区域同位模式并行挖掘算法 [J]. 中国科学: 信息科学, 2023, 53(7): 1281-1298.
- [5] 张华. 图数据库 Neo4j 下的 POI 同位特征提取 [J]. 测绘技术装备, 2024, 26(3): 88-94.
- [6] 宋宇轩. 基于自适应神经模糊理论的医疗数据安全评估模型设计 [J]. 微型电脑应用, 2024, 40(4): 202-204.
- [7] 陈景琪, 丁凌. 基于模式挖掘的智能楼宇能源消耗量预测方法 [J]. 电子设计工程, 2022, 30(4): 134-137.
- [8] 刘逸, 陈奎, 刘子惠, 等. 基于同位模式的休闲产业空间集聚特征研究 [J]. 旅游学刊, 2022, 37(2): 94-104.
- [9] 秦晓磊. 基于模糊粗糙集的东北地区河流水位流量关系单值化研究 [J]. 水利技术监督, 2025(3): 117-119.
- [10] 杨秋菊. 基于模糊理论的多源异构传感器数据融合模型 [J]. 吉林大学学报 (工学版), 2024, 54(10): 3058-3063.

【作者简介】

段韶鹏 (1991—), 男, 河南郑州人, 硕士, 讲师, 研究方向: 大数据、人工智能。

姜文璐 (1996—), 女, 河南周口人, 博士, 助教, 研究方向: 信息安全。

张金鹏 (1989—), 男, 河南郑州人, 博士, 讲师, 研究方向: 空间数据挖掘。

张占 (1990—), 男, 湖南张家界人, 博士, 讲师, 研究方向: 大数据分析。

(收稿日期: 2025-03-21 修回日期: 2025-07-07)