基于图神经网络聚类的露天矿区山洪风险等级区划研究*

潘臻[1,2],杨浩翔[3],王向鹏[4],李伟[5],翟东豪[6]

(1.成都高新减灾研究所, 四川 成都 610095;
2.地震预警与多灾种预警应用信息技术四川省重点实验室, 四川 成都 611731;
3.成都市美幻科技有限公司, 四川 成都 610041;
4.成都理工大学 地球物理学院, 四川 成都 610059;
5.甘肃煤田地质局一四五队, 甘肃 张掖市 734000;
6.成都理工大学 计算机与网络安全学院, 四川 成都 610059)

摘要:针对现有的露天矿区山洪风险等级区划方法存在的空间关系忽视问题,提出了一种基于图神经网络(GNN)聚类的露天矿区山洪风险等级区划方法。研究区域选取白云鄂博矿区,矿区山洪风险等级区划方法主要有三个步骤:(1)选取该矿区遥感图像划分栅格单元并将每个栅格作为图的节点,节点特征主要包括高程、坡度、坡向、降雨量等;(2)利用余弦相似度法计算环境特征相似性并将相似节点的空间关系表示为边,构建邻接矩阵;(3)构建的图和对应的邻接矩阵被输入到网络模型,采用图采样和聚合(GraphSAGE)与图注意力机制(GAT)组成的编码器架构,使用KL散度损失函数优化模型,最终实现白云鄂博矿区山洪风险等级的聚类预测。试验结果表明,模型的聚类准确率达到87.5%,与传统聚类算法k-means、层次聚类和随机森林相比分别提升了11.2, 9.4, 5.3个百分点,验证了该方法在露天矿区的山洪风险等级区划中的可靠性。研究成果可为露天矿区的山洪预防提供参考。

关键词:露天矿区;山洪风险等级;聚类分析;图神经网络

中图分类号:TD745 **文献标识码**:A **文章编号**:1005-2763(2025)12-0144-09

Research on Regional Classification of Mountain Flood Risk Based on Graph Neural Network Clustering in the Open-Pit Mining Area

PAN Zhen[1,2], *YANG Haoxiang*[3], *WANG Xiangpeng*[4], *LI Wei*[5], *ZHAI Donghao*[6]

(1.Institute of Care-Life (ICL) of Chengdu High-Tech Zone, Chengdu, Sichuan 610095, China;
2.Multi-Hazard Early Warning Key Laboratory of Sichuan Province, Chengdu, Sichuan 611731, China;
3.Chengdu Meihuan Technology Co., Ltd, Chengdu, Sichuan 610041, China;
4.College of Geophysical, Chengdu University of Technology, Chengdu, Sichuan 610059, China;
5.145 Brigade Gansu Bureau of Coal Geology, Zhangye, Gansu 734000, China;
6.College of Computer Science and Cyber Security, Chengdu University of Technology, Chengdu, Sichuan 610059, China)

Abstract: A method based on graph neural network (GNN) clustering was proposed for the regional classification of mountain flood risk in open-pit mining areas, so as to address the issue of spatial relationships neglected in existing classification methods. The research area selects the Baiyunebo mining area, and the classification method of mountain flood risk in mining area mainly consists of three steps. Firstly, the remote sensing images of the mining area were selected to divide into grid units and each grid was used as a node of the map. The node features mainly included elevation, slope, aspect, rainfall and precipitation. Secondly, the cosine similarity method was used to calculate the similarity of environmental features and represent the spatial relationship of similar nodes as edges, constructing an adjacency matrix. Thirdly, the constructed graph

* **收稿日期**:2024-12-23
基金项目:四川省科技创新人才项目(2023JDRC0012);四川省重点实验室开放基金项目(2024KFKT003)
作者简介:潘臻(1981—),男,四川内江人,博士,讲师,主要研究方向为自然灾害监测预。E-mail:182319702@qq.com
通信作者:翟东豪(2002—),男,河南洛阳人,硕士研究生,主要研究方向为图神经网络、人工智能与地学大数据。E-mail:1102428556@qq.com

and corresponding adjacency matrix were input into the network model, which adopted an encoder architecture consisting of graph sample and aggregate (GraphSAGE) and graph attention network (GAT). The KL divergence loss function was used to optimize the model, ultimately achieving clustering prediction of the classification of mountain flood risk in the Bayyunebo mining area. The experimental results show that the clustering accuracy of the model reaches 87.5%, which is 11.2, 9.4 and 5.3 percentage points higher than traditional clustering algorithms *k*-means, hierarchical clustering, and random forest, respectively. The research result can verify the ability of this method to classify mountain flood risk in open-pit mining areas and can provide a reference for the prevention of mountain flood in open-pit mining areas.

Key words: Open-pit mining area, Classification of mountain flood risk, Cluster analysis, Graph neural network

0 引言

露天矿山在矿产资源开采中发挥着重要作用。开采和加工过程中,露天矿山会对周边环境产生许多影响[1],导致矿区易受自然灾害的威胁,尤其是山洪灾害。山洪产生的原因复杂多样[2],在雨季,边坡滑坡、崩塌等极易诱发泥石流灾害,它不仅威胁矿区设施和人员安全,还会导致严重的水土流失、生态退化和水源污染,进一步加剧矿区及其周边环境的破坏。现有的山洪评估方法多侧重于水文分析、经验判断或统计模型分析。牛全福等[3]通过引入山洪潜在指数,构建了一个多层次评价体系,用于山洪风险等级评价;韩俊太等[4]提出了基于动态临界雨量的山洪灾害分级预警方法,延长了灾害的预见期;ZHAI Xiaoyan 等[5]将水文建模方法和聚类分析结合来评估整个山洪过程。然而,这些方法仍然难以揭示露天矿区不同地理区域间的复杂关系与多重影响因素的交互作用。

图神经网络(Graph Neural Network, GNN)是一种专门处理图结构数据的神经网络[6],因能够有效捕捉拓扑结构并对非线性关系进行表达而受到关注[7]。近年来,GNN 被广泛引入计算机视觉(Computer Vision, CV)模型,其不仅提升了 CV 相关任务的性能表现,而且为 CV 模型提供了更多可解释的分解[8]。基于 GNN 的聚类方法能充分挖掘节点更深层次的潜在特征表示,从而有效提升聚类的准确性和效果。图卷积网络(Graph Convolutional Network, GCN)作为 GNN 的一种重要变体,通过对图的卷积操作将邻域节点的信息汇总到目标节点,实现对空间结构信息的有效学习[9]。而图注意力网络(Graph Attention Network, GAT)允许模型根据节点间的权重差异灵活调整信息传播的强度[10],特别适合处理复杂多变的环境因素,弥补传统 GCN 在处理多因素交互时的局限性。图采样和聚合(Graph Sample and Aggregate, GraphSAGE)通过采样并聚合邻居节点和顶点的特征来获得顶点的图嵌入,该方法能够有效处理大规模图数据[11],避免邻接矩阵的稠密计算。另外,遥感(Remote Sensing)为山洪灾害的监测和评估提供了更为丰富的高分辨率数据[12],其中包含的地形变化、植被覆盖、水体分布等精细地表特征,能够为露天矿区山洪灾害风险评估提供精细的地理背景信息。

白云鄂博矿区拥有丰富的矿产资源,但地形变化、植被破坏和降水集中等因素可能加剧该地区的山洪风险。本文选取白云鄂博矿区为研究区域,进行露天矿区山洪风险等级区划研究。在构图过程中,通过地理信息系统(Geographic Information System,GIS)将遥感图像栅格化,并将每个栅格单元作为图的节点,节点的属性特征包括高程、坡度、坡向、降雨量等地理信息,旨在充分利用露天矿区的特殊环境因素,增强特征表达;在构建边时,借鉴 ZHANG Qing 等[13]使用的余弦相似度方法计算环境相似性,并将相似节点之间的空间关系表示为边,据此生成邻接矩阵,根据节点的相似度定义边的权重。该方法不仅有助于模型捕捉方向性特征,尤其是坡向、坡度和降雨量等属性,同时也降低了图的规模。本研究将 GraphSAGE 模块和 GAT 模块融入编码器,以 KL 散度作为损失函数优化聚类结果,将研究区域划分为 5 个风险等级:低、较低、中、较高、高。研究主要作出了以下贡献。

(1) 将 GNN 应用于露天矿区的安全防治。通过对矿区的地理特征和环境数据进行建模,捕捉不同区域之间的相互关系,有效解决了矿山环境中复杂的地形、气候和人为因素对山洪等级预测的影响。

(2) 提出了一种基于 GNN 聚类分析的方法。

首先根据地理信息构建出露天矿区的图结构,并将GraphSAGE和GAT融入图神经聚类模型,实现对露天矿区山洪风险等级的精准聚类。

1 风险等级区划指标

研究表明,区域下垫面条件直接影响地表径流汇集过程,相对稳定的下垫面条件是山洪形成的孕灾环境,在山洪汇流过程中起着关键作用[14]。此外,地形和地貌也为滑坡和泥石流等次生灾害的发生提供了物质条件。坡度陡峭、地形深度切割的区域,地表汇流速度较快、径流损失较少,易发生山洪。因此,在山洪风险评估中,必须重视地形起伏和地貌类型等下垫面因素的影响。

露天矿区的开采会引发地质环境问题[15]。开采改变了原有地形,矿坑的深度和高程差使水流容易积聚或加剧径流。此外,开采过程清除了大量植被,导致土壤抗侵蚀能力下降,水土流失加剧,径流增加,最终引发山洪。而且露天矿山可能位于降雨频发区域,降水集中且强度大时会使水位迅速上升,水流汇聚,若排水系统设计不合理,矿区积水无法有效排出,暴雨后形成洪水区,进一步加剧山洪灾害的发生风险。

根据以上诱因,本文重新梳理了露天矿区山洪风险指标,选取高程、高程差、坡度、坡向、坡度差作为孕灾环境指标,选取年降雨天数以及年降雨量均值作为致灾指标,使用归一化植被指数(NDVI)作为地表状态指标。

2 研究区域与数据

2.1 研究区域描述

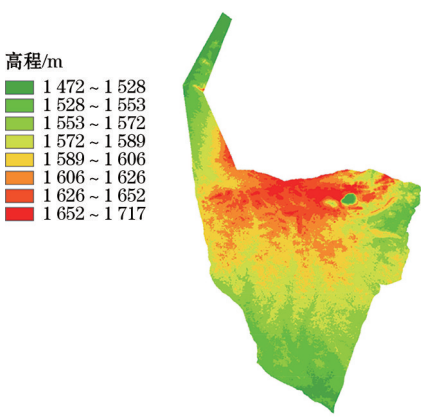
白云鄂博矿区位于内蒙古自治区中部,是全球重要的稀土资源产地之一,主要包含两个露天矿区(矿区1和矿区2)。该矿区的主要矿产包括稀土、铁矿、铜矿、钼矿等。矿区坐落于内蒙古巴彦淖尔市白云鄂博镇,周围为典型的草原生态环境,且零星分布着一些水域。气候方面,白云鄂博矿区属于温带大陆性气候,夏季干热,冬季寒冷且干燥,年降水量较少,集中在夏季,年降水量为200~350 mm,气温波动较大,夏季温度可超过30 ℃,而冬季最低气温可降至-30 ℃左右。该矿区地质构造较为复杂,主要由花岗岩、石英岩和片麻岩等构成,地质条件稳定,具有较高的矿产资源潜力,以稀土矿、铁矿、铜矿、钼矿为主。白云鄂博稀土矿是全球最大的稀土矿之一,轻稀土元素如铈、镧的储量丰富。此外,铁矿和铜矿的储量也非常可观,矿区内的钼矿资源同样具有重要的经济价值。

尽管矿区拥有丰富的矿产资源,但其开采活动对环境造成了明显影响。当地的草原生态系统受采矿和加工过程的影响而遭到破坏,水源也面临污染风险。同时,矿区的地形变化、植被破坏以及降水集中等因素共同加剧了该地区的山洪灾害风险。因此,有必要结合区域地理和水文特征进行全面分析,以制定切实有效的防灾减灾措施。

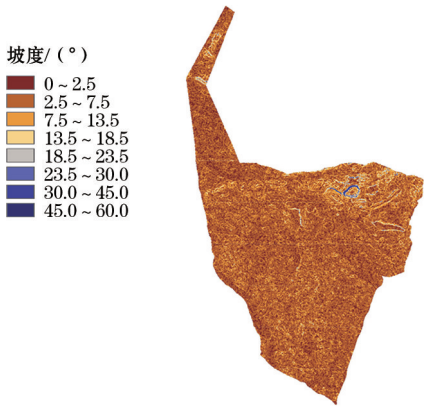
本文研究数据源于地理空间数据云和国家遥感数据与应用服务平台上的高分二号卫星图像。地理数据包括栅格化的高程信息、坡度等信息,降雨数据来自中央气象局,包括年降雨天数及年降雨量均值等。

2.2 数据预处理

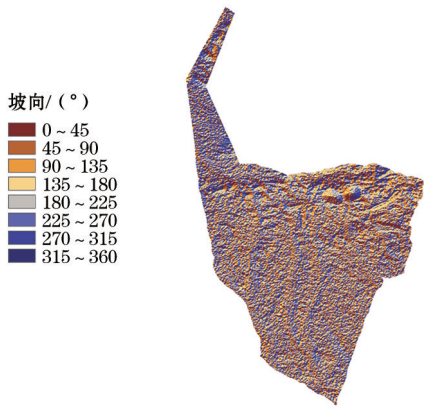
对数字高程模型(Digital Elevation Model,DEM)数据进行预处理,确保后续分析的准确性和有效性。首先,对获取的DEM数据进行拼接、裁剪和投影,以行政区作为边界分割出白云鄂博矿区并生成该区域的地形分析图像,如图1所示。将不同来源的气象数据整合为统一格式,确保高程、坡度等栅格数据具有一致的空间分辨率和坐标系统。然后,对地理数据进行清洗,去除缺失值和异常值,并使用归一化方法[16]对提取的特征数据进行标准化处理,以消除数据源和量纲的差异。



(a) 高程



(b) 坡度



(c) 坡向

图 1 地形分析图像

Fig.1 Images of terrain analysis

3 研究方法

3.1 图构建

3.1.1 节点定义

利用 GIS 计算各地理节点的属性特征。对于节点 v_i,构建其特征向量 $\boldsymbol{X}_i=(x_1, x_2, \cdots, x_n)$,其中,$n$ 为特征的维度,x 为特征值。

常见的节点权重赋值方法有:基于网格的特征赋值和基于地理特征赋值。由于矿区地理环境复杂且具有一定的规律性,如坡度是影响露天矿山安全性的重要因素,凹地或坡度较小的区域可能面临积水。因此,本文基于地理特征赋值,采用加权和归一化处理进行权重优化。

对于节点 v_i,其归一化权重定义为:

$$\omega=\alpha_1\frac{x_1}{\max(x_1)}+\cdots+\alpha_n\frac{x_n}{\max(x_n)} \tag{1}$$

式中:ω 为节点的权重;$\frac{x}{\max(x)}$ 为归一化后的特征;α 为特征的权重系数,可根据实际来调整。设定 $n=5$,对应具体的 5 个特征:坡度、坡向、高程、降雨量、NDVI。对于的 α 设定,考虑到坡度对山洪形成的影响较大,可以设置较大的权重系数;而土地覆盖类型的指标影响较小,则可以设置较小的权重系数。表 1 为某节点的权重计算。

表 1 节点权重计算

Table 1 Node weight calculations

特征	原始值	最大值	归一化值	权重系数	加权值
坡度	33.6°	47.4°	0.71	0.3	0.21
坡向	110.3°	178.6°	0.62	0.2	0.12
高程	1 141.1 m	1 783.9 m	0.64	0.1	0.06
降雨量	218.7 mm	397.3 mm	0.55	0.3	0.17
NDVI	0.1	1	0.10	0.1	0.03
总权值					0.59

3.1.2 边定义

通过计算邻域节点的相似性确定节点之间的边。在网格结构中,节点的邻居通常由其空间位置决定,使用余弦相似度方法计算两个节点间的相似度,每个节点都与其相似度最大的节点构成边。对于节点 v_i 和 v_j,通过余弦相似度方法计算邻域节点的相似度 $\mathrm{Sim}(v_i, v_j)$,当相似度值取最大时,则认为这两个节点之间存在边,最终构建出图 $\boldsymbol{G}=(\boldsymbol{V}, \boldsymbol{E})$,以及邻接矩阵 $\boldsymbol{A}$。余弦相似度方法计算过程见式(2)、式(3)。

$$\cos\theta_{\boldsymbol{X}_i\boldsymbol{X}_j}=\frac{\boldsymbol{X}_i\cdot\boldsymbol{X}_j}{|\boldsymbol{X}_i|\cdot|\boldsymbol{X}_j|} \tag{2}$$

$$\mathrm{Sim}(v_i, v_j)=1-\cos\theta_{\boldsymbol{X}_i\boldsymbol{X}_j} \tag{3}$$

式中,$\cos\theta_{\boldsymbol{X}_i\boldsymbol{X}_j}$ 为两特征向量的夹角的余弦值。

3.2 基于 GNN 的聚类模型

基于 GNN 的聚类模型采用了 GraphSAGE 和 GAT 的图编码器(Encoder)架构。在输入端,图通过节点的地理特征以及基于环境相似度计算的邻接矩阵构建。GraphSAGE 模块通过采样并聚合邻居节点的特征来生成节点表示,这使得模型可以有效地捕捉局部结构信息。GAT 具有注意力机制,可赋予不同邻居节点不同的权重,从而在信息聚合过程中自动聚焦对当前节点最有意义的邻居节点。全连接层通过上述两个模块的结合,能够在全局和局部尺度上精确地学习节点表示。在训练过程中,模型使用 KL 散度作为损失函数进行优化,目标是最小化预测结果与真实分布之间的差异。模型构建和分析流程如图 2 所示。

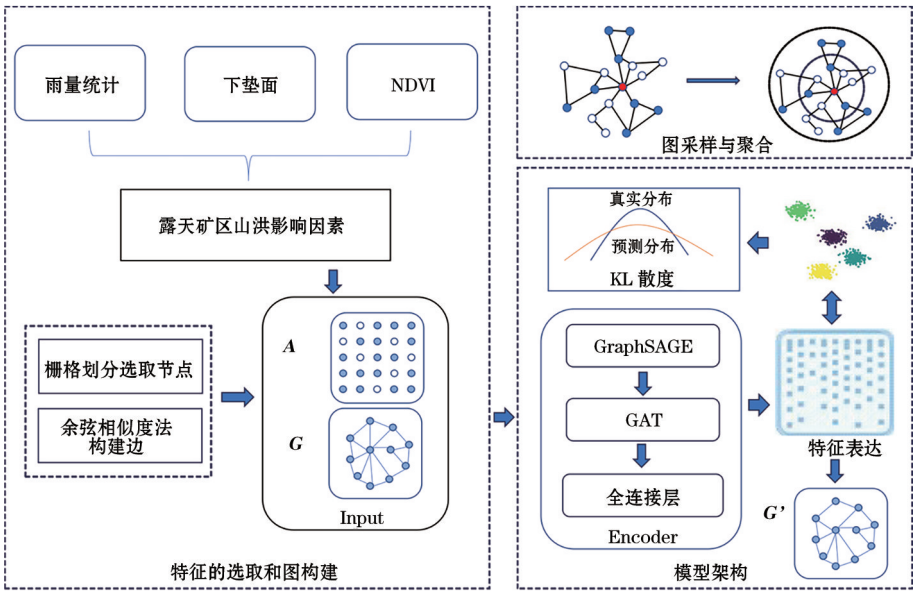


图 2 模型构建和分析流程

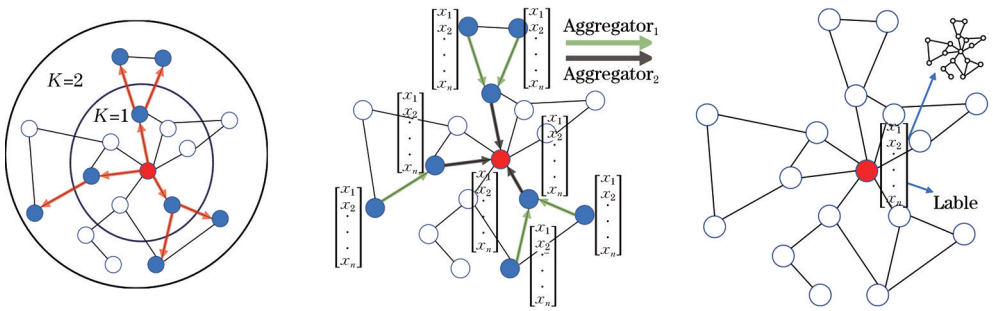
Fig.2 Construction and analysis process of the model

3.2.1 GraphSAGE 模块

GraphSAGE 模块由三层 GCN 层、两个全连接层和一个 Dropout 层组成,并采用 GELU 作为非线性激活函数。与卷积神经网络(Convelutional Neural Networks, CNN)不同,GCN 更注重图的采样和信息的聚合。GELU 函数的引入增强了模型的非线性表达能力,使其能够学习复杂的网络关系。全连接层用于整合特征并将嵌入维度映射到输出维度,作为 GAT 模块的输入。GraphSAGE 模块的主要目标是从每个节点的 K 跳邻域中进行采样,并通过聚合函数(Aggregator)聚合邻居节点信息,以生成节点的图嵌入表示(Label)。该表示利用聚合后的信息预测节点的聚类类别,同时输出节点 v_i 的特征 h_{v_i},见式(4)。GraphSAGE 过程如图 3 所示。

$$h_{v_i}=\mathrm{Aggregator}(\mathrm{GELU}(\boldsymbol{A}\boldsymbol{X}_i\boldsymbol{W})) \quad (4)$$

式中,$\boldsymbol{W}$ 为卷积层的权重矩阵。



(a) 采样邻居 (b) 聚合特征 (c) 嵌入生成

图 3 GraphSAGE 过程

Fig.3 The process of GraphSAGE

3.2.2 GAT 模块

模型进一步引入了多头注意力机制,构成 GAT 模块对潜在特征进行重构。多头注意力机制即通过调用 T 组相互独立的注意力机制,将各自生成的节点表示进行拼接,形成最终的节点特征表达:

$$h'_{v_i}=\mathrm{Concat}(h_{v_i}^{(1)},h_{v_i}^{(2)},\cdots,h_{v_i}^{(T)})\cdot\boldsymbol{W} \quad (5)$$

式中:h'_{vi} 表示拼接后节点 v_i 的特征表达;$h_{vi}^{(T)}$ 表示第 T 组注意力头的输出;Concat 函数将所有注意力头结果进行拼接,并通过线性变换矩阵得到最终的节点特征表达。特征表达方式可有效结合图的结构和属性信息,提升特征的表达深度。多头注意力机制如图 4 所示。

模型通过注意力系数度量节点间的邻域关系,同时引入相邻节点间的属性特征,将邻居关系与属性特征相结合。该联合特征经过激活函数的处理后生成输出,可以实现同时捕捉节点结构和属性的特

征表达效果。最终的输出即为每个节点在不同聚类中心的分布,可用于后续的聚类分析。

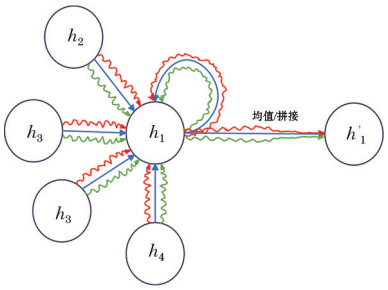


图 4 多头注意力机制

Fig.4 Multi-head attention mechanism

3.2.3 KL 散度损失

为确保隐藏表示的分布与先验分布相匹配,以便在潜在空间中形成清晰的聚类结构,模型采用 KL 散度作为聚类损失。将节点与聚类中心之间的相似性分布表示为 Q,利用其二次方构建 P 分布函数作为目标分布。聚类误差 L 由 P 与 Q 之间的 KL 散度定义:

$$L=\mathrm{KL}(P\,||\,Q) \tag{6}$$

通过最小化 KL 散度,模型可以在潜在空间中形成更加连贯的分布,从而提高聚类的效果。

3.3 最佳聚类数量选择

本文通过计算聚类质量指数(CQI,I_{CQ})[17] 来确定最佳聚类方案。CQI 用于综合评估每个节点的属性特征差异,以确保同一类别的空间单元之间的差异最小化,从而找到最合适的聚类数量。CQI 值越高,说明同类别节点之间的属性特征差异越大。此外,随着聚类数量的增加,CQI 值也会相应增大。因此,本研究将 CQI 值最小的聚类结果选为最佳聚类方案,见式(7)。计算了 4 种方法在不同聚类数量下的 CQI 值,如图 5 所示。

$$I_{\mathrm{CQ}}=-\frac{1}{M}\sum_{r=1}^{a}n_r\sum_{j=1}^{b}C_{rj}+\delta \tag{7}$$

式中:M 为节点总数;a 为聚类的类别数;b 为属性因子的总数;n_r 为类型为 r 的节点数量;C_{rj} 为类别 r 中第 j 属性因子的变异系数;δ 为常数,用于对类别数量进行尺度调整。

由图 5 可知,随着聚类数量的增加,4 种方法的 CQI 值均呈现出先下降后上升的趋势。在聚类数量较少的情况下,增加聚类数量有助于提升类别内部属性特征的相似性。然而,当聚类数量进一步增加,作为惩罚项的影响逐渐显现,导致 CQI 值开始回升。具体来看,GNN 聚类算法和 k-means 方法在聚类数为 5 时,CQI 值最低;层次聚类(Ward)方法在聚类数为 3 时达到最小 CQI 值;而随机森林(Random Forest,RF)方法则在聚类数为 4 时取得最低 CQI 值。

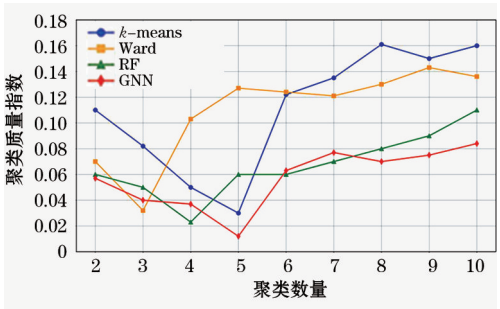


图 5 不同聚类数量下的 CQI 值

Fig.5 CQI values of different numbers of clusters

3.4 模型训练设置

为确保模型的有效性,超参数设置见表 2。训练过程采用 Adam 优化器进行参数自适应性并更新参数,实现动态学习率衰减以提高收敛速度。训练中配置验证集监控机制,当性能不再提升时及时停止训练,防止过拟合。采用 mini-batch 方法,充分利用 GPU 资源加速训练过程。

表 2 超参数设置

Table 2 Hyperparameters setting

参数	设置
模型层数	2 层,3 层,4 层
嵌入维度	64,128,256
注意力头数	2,4,6
激活函数	GeLU
学习率	0.001
正则化	L2

3.5 风险等级划分结果

采矿活动对山洪风险的空间分布具有较大的影响。白云鄂博矿区山洪风险等级区划结果如图 6 所示。

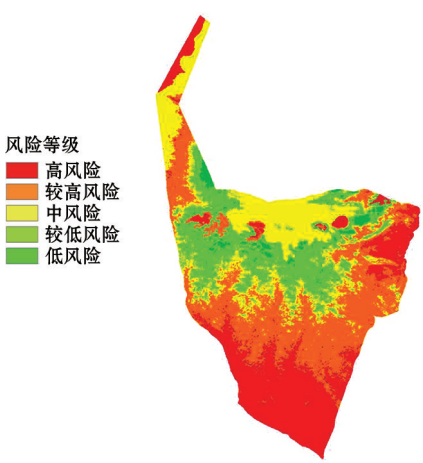


图 6 山洪风险等级区划结果

Fig.6 Results of regional classfication of mountain flood risk

基于GNN的聚类算法综合地形特征等多因素的分析结果,将研究区域的山洪风险划分为以下5个等级。

(1) 高风险区。主要分布在陡坡、主沟谷及矿区中心地带。该区域坡度较大,叠加地表开挖、植被破坏与废石堆积等因素,导致地表径流强度显著增大。

(2) 较高风险区。集中于次级沟谷及受采矿影响的坡地。该区域排水系统受损,且土壤结构松散共同加剧了该区域的水土流失程度。

(3) 中风险区。位于坡度适中的过渡地带,受采矿干扰较轻。部分区域植被有所恢复,对径流强度起到一定的减缓作用。

(4) 较低风险区。多见于缓坡和山间盆地,受采矿活动影响较小,自然排水条件良好,山洪风险相对较低。

(5) 低风险区。主要分布于平坦地带,远离采矿区域,植被覆盖度高,排水通畅,基本无山洪风险。

4 试验与分析

4.1 对比试验

在聚类分析中,为了全面评估聚类效果,常用的指标包括聚类准确率、聚类紧密度、聚类质量指数。聚类准确率通过准确率衡量聚类结果与真实标签的一致性,准确度越高,说明聚类效果越好。聚类紧密度评估同一聚类内样本之间的相似度,通过计算样本到聚类中心的距离,距离越小,聚类越紧密,效果越好。通过综合分析上述指标,可以全面评估聚类效果,为数据分析和决策提供依据。为对比图神经网络聚类方法与传统聚类方法的效果,本研究选择了三种经典聚类方法[18](*k*-means、Ward和RF)进行对比试验。

不同聚类算法的对比试验结果见表3。可以看出,本文提出的GNN聚类方法在山洪综合风险等级区划任务中表现出显著优势。相较于传统的*k*-means、Ward和RF聚类方法,GNN在聚类准确率上分别提升了11.2,9.4和5.3个百分点,同时该方法凭借其对节点属性与空间结构关系的深度捕捉,取得了最高的聚类紧密度(0.72)和最低的CQI值(0.012)。尽管*k*-means与Ward等传统方法在部分场景中表现良好,但在处理高维复杂数据时存在局限性,尤其对山洪数据中强烈的空间关联性难以实现有效捕捉,导致在准确性与空间连贯性等方面存在不足。RF虽在捕捉非线性特征方面有所提升,但在空间结构信息处理上仍存在不足。试验结果验证了图神经网络在处理强关联性数据方面具有独特优势。

表3 不同聚类算法对比试验结果

Table 3 Results of comparative experiments among different clustering algorithms

聚类算法	聚类准确率/%	轮廓系数	聚类紧密度	CQI值
k-means	76.3	0.60	0.55	0.029
Ward	78.1	0.65	0.60	0.032
RF	82.2	0.70	0.65	0.020
GNN	87.5	0.73	0.72	0.012

4.2 消融试验

为了深入理解模型中GraphSAGE模块、GAT模块以及KL散度损失对整体性能的贡献,并验证该模型的鲁棒性和稳定性,本文设计了一系列消融试验,结果见表4。

表4 消融试验结果

Table 4 Results of ablation experiments

模型	聚类准确率/%	轮廓系数	聚类紧密度	CQI值
GAT+KL	84.0	0.68	0.67	0.028
GraphSAGE+KL	83.1	0.61	0.70	0.020
GraphSAGE+GAT	85.3	0.71	0.65	0.019
GraphSAGE+GAT+KL	87.5	0.73	0.72	0.012

消融试验结果表明,GraphSAGE、GAT与KL散度损失均对模型性能具有重要贡献。当分别移除GraphSAGE、GAT或KL散度模块时,聚类准确率分别下降至84.0%、83.1%、85.3%;而完整的基于GNN聚类模型(GraphSAGE+GAT+KL)聚类准确率达到最优值,为87.5%。这一结果表明:GraphSAGE模块增强了模型对图结构信息的捕捉能力,GAT模块提高了对节点间关联关系的敏感度,而KL散度损失则有效优化了潜在空间中的聚类分布。各模块的协同作用显著提升了模型整体的聚类效果,凸显了基于GNN聚类模型在处理复杂结构化数据中的优势。

5 结论

本文提出了一种基于图神经网络(GNN)的露天矿区山洪风险等级区划聚类方法,并将其应用于白云鄂博矿区,实现了对该矿区的山洪风险等级划分,得到以下主要结论。

(1) GNN是处理具有强空间关联性地理数据的有效工具。模型基于GNN的空间依赖关系进行

建模,能够有效捕捉地理空间数据的内在关联。与传统聚类方法 k-means、Ward 和 RF 相比,模型的聚类准确率分别提升了 11.2,9.4 和 5.3 个百分点,表明本文的模型在山洪灾害风险等级预测中的效果优于传统聚类方法。

(2) 通过集成 GraphSAGE、图注意力机制与 KL 散度损失的模型框架,多模块协同的基于 GNN 聚类模型在图结构信息的特征捕捉能力,对节点间关联关系的敏感度和聚类分布优化等方面具有显著优势。

(3) 本研究所提出的方法为露天矿区山洪风险防控提供了新思路,成果可用于指导矿区监测布设与重点防控区域识别,对矿山安全管理和灾害预警具有一定启示意义。然而,当前方法构建的图规模较大,会导致节点采样和聚合的复杂度增加,从而增加时间和空间开销。未来的研究仍需关注大规模图优化策略。

参考文献(References):

[1] 郑凯.浅谈露天矿山开采过程中的地质灾害防治[J].建筑技术研究,2019(12):87-88.
ZHENG Kai. Abrief discussion on the prevention and control of geological disasters in the open-pit mining process [J]. Building Technology Research,2019(12):87-88.

[2] RIJAL M, LUO P P, MISHRA B K, et al. Global systematical and comprehensive overview of mountainous flood risk under climate change and human activities[J]. Science of The Total Environment,2024,941:173672.

[3] 牛全福,熊超,雷姣姣,等.基于 FFPI 模型的甘肃陇南山区山洪灾害风险评价[J].自然灾害学报,2023,32(4):36-47.
NIU Quanfu, XIONG Chao, LEI Jiaojiao, et al. Risk assessment of mountain flood disasters in the Longnan mountain area of Gansu Province based on the FFPI model [J]. Journal of Natural Disasters,2023,32(4):36-47.

[4] 韩俊太,王政荣,杨雨亭.基于动态临界雨量的小流域山洪灾害分级预警[J].水力发电学报,2022,41(9):67-76.
HAN Juntai, WANG Zhengrong, YANG Yuting. Flash flood grading and warning based on dynamic rainfall thresholds[J]. Journal of Hydroelectric Engineering,2022,41(9):67-76.

[5] ZHAI Xiaoyan, ZHANG Yongyong, ZHANG Yongqiang, et al. Simulating flash flood hydrographs and behavior metrics across China: Implications for flash flood management[J]. Science of The Total Environment,2021,763:142977.

[6] 吴博,梁循,张树森,等.图神经网络前沿进展与应用[J].计算机学报,2022,45(1):35-68.
WU Bo, LIANG Xun, ZHANG Shusen, et al. Advances and applications in graph neural network [J]. Chinese Journal of Computers,2022,45(1):35-68.

[7] WANG Xianmin, DU Aiheng, HU Fengchang, et al. Landslide susceptibility evaluation based on active deformation and graph convolutional network algorithm[J]. Frontiers in Earth Science,2023,11:1132722.

[8] 吴凌飞.图神经网络:基础、前沿与应用[M].北京:人民邮电出版社,2022:346.
WU Lingfei. Graph neural networks: basics, frontiers, and applications[M]. Beijing: Posts & Telecom Press,2022:346.

[9] 徐冰冰,岑科廷,黄俊杰,等.图卷积神经网络综述[J].计算机学报,2020,43(5):755-780.
XU Bingbing, CEN Keting, HUANG Junjie, et al. A survey on graph convolutional neural networks[J]. Chinese Journal of Computers,2020,43(5):755-780.

[10] 李刚,杜亚波,杨庆贺,等.基于多尺度特征融合和注意力机制的矿区道路障碍检测[J].中国安全科学学报,2024,34(9):87-98.
LI Gang, DU Yabo, YANG Qinghe, et al. Obstacle detection on mining roads based on multi-scale feature fusion and attention mechanism[J]. China Safety Science Journal,2024,34(9):87-98.

[11] LIU Tao, JIANG Aimin, ZHOU Jia, et al. GraphSAGE-based dynamic spatial-temporal graph convolutional network for traffic prediction[J]. IEEE Transactions on Intelligent Transportation Systems,2023,24(10):11210-11224.

[12] 黄登冕,张聪,姚晓军,等.矿山环境遥感监测研究进展[J].遥感技术与应用,2022,37(5):1043-1055.
HUANG Dengmian, ZHANG Cong, YAO Xiaojun, et al. Research progress of mine environment remote sensing monitoring[J].Remote Sensing Technology and Application,2022,37(5):1043-1055.

[13] ZHANG Qing, HE Yi, ZHANG Lifeng, et al. A landslide susceptibility assessment method considering the similarity of geographic environments based on graph neural network [J]. Gondwana Research,2024,132:323-342.

[14] 张若婧,陈跃红,张晓祥,等.基于参数最优地理探测器的江西省山洪灾害时空格局与驱动力研究[J].地理与地理信息科学,2021,37(4):72-80.
ZHANG Ruojing, CHEN Yuehong, ZHANG Xiaoxiang, et al. Spatial-temporal pattern and driving factors of flash flood disasters in Jiangxi Province analyzed by optimal parameters-based geographical detector[J]. Geography and Geo-information Science,2021,37(4):72-80.

[15] 刘恒材,黄河,帅博文,等.露天煤矿地质灾害防治研究进展[J].矿山工程,2020,8(3):311-316.
LIU Hengcai, HUANG He, SHUAI Bowen, et al. Research progress on geological disaster prevention and control in open-pit coal mines [J]. Mining Engineering,2020,8(3):311-316.

[16] MAZZIOTTA M, PARETO A. Normalization methods for spatio-temporal analysis of environmental performance: revisiting the min-max method[J]. Environmetrics,2022,33(5):e2730.

[17] 毛祺,彭建,刘焱序,等.耦合 SOFM 与 SVM 的生态功能分区方法:以鄂尔多斯市为例[J].地理学报,2019,74(3):460-474.
MAO Qi, PENG Jian, LIU Yanxu, et al. An ecological function zoning approach coupling SOFM and SVM: a case study in Ordos[J]. Acta Geographica Sinica,2019,74(3):460-474.

[18] 徐金霞,郭海燕,邓国卫,等.山洪灾害临界雨量确定方法对比研究:以雅安名山河流域为例[J].自然灾害学,2024,33(3):89-99.
XU Jinxia, GUO Haiyan, DENG Guowei, et al. Comparative study on rainfall threshold methods of flash flood disaster: take Mingshan River of Ya'an as an example [J]. Journal of Natural Disasters,2024,33(3):89-99.