

基于SVM的地下采煤区沉陷灾害发育敏感性分区研究

薛永安 邹友峰 张文志 张明媚 柳广春 宋明伟

引用本文:

薛永安, 邹友峰, 张文志, 等. 基于SVM的地下采煤区沉陷灾害发育敏感性分区研究[J]. 煤田地质与勘探, 2022, 50(10): 108–118.

XUE Yong'an, ZOU Youfeng, ZHANG Wenzhi, et al. SVM-based sensitivity zoning of subsidence disaster development in the underground coal mining areas[J]. *Coal Geology & Exploration*, 2022, 50(10): 108–118.

在线阅读 View online: <https://dx.doi.org/10.12363/issn.1001-1986.22.01.0104>

您可能感兴趣的其他文章

Articles you may be interested in

贵州省开阳县斜坡地质灾害孕灾因子敏感性分析

Sensitivity analysis of disaster-pregnant environmental factors for slope geological hazards in Kaiyang County, Guizhou Province
煤田地质与勘探. 2020, 48(4): 190–198 <https://doi.org/10.3969/j.issn.1001-1986.2020.04.026>

基于VMD和SVM的煤厚预测方法研究

Research on the coal thickness prediction method based on VMD and SVM
煤田地质与勘探. 2021, 49(6): 243–250 <https://doi.org/10.3969/j.issn.1001-1986.2021.06.029>

厚黄土与煤炭采空区地质灾害易发性评价

Geological hazard susceptibility assessment in thick loess and gob
煤田地质与勘探. 2017, 45(5): 112–120 <https://doi.org/10.3969/j.issn.1001-1986.2017.05.020>

无人机在矿区表土特征及地质灾害监测中的应用

Application of unmanned aerial vehicle in surface soil characterization and geological disaster monitoring in mining areas
煤田地质与勘探. 2021, 49(6): 200–211 <https://doi.org/10.3969/j.issn.1001-1986.2021.06.024>

基于加权确定性系数法的地质灾害易发性分析

Analysis of geological hazard susceptibility based on the weighted certainty factor method
煤田地质与勘探. 2018, 46(6): 108–114 <https://doi.org/10.3969/j.issn.1001-1986.2018.06.015>

煤矿地区的滑坡灾害及其防治技术

Landslide disaster and its prevention and control technology in coal mining area
煤田地质与勘探. 2018, 46(2): 1–7 <https://doi.org/10.3969/j.issn.1001-1986.2018.02.001>



移动阅读

薛永安, 邹友峰, 张文志, 等. 基于 SVM 的地下采煤区沉陷灾害发育敏感性分区研究[J]. 煤田地质与勘探, 2022, 50(10): 108–118. doi: 10.12363/issn.1001-1986.22.01.0104
XUE Yong'an, ZOU Youfeng, ZHANG Wenzhi, et al. SVM-based sensitivity zoning of subsidence disaster development in the underground coal mining areas[J]. Coal Geology & Exploration, 2022, 50(10): 108–118. doi: 10.12363/issn.1001-1986.22.01.0104

基于 SVM 的地下采煤区沉陷灾害发育敏感性分区研究

薛永安^{1,2}, 邹友峰^{1,3}, 张文志^{1,3}, 张明媚^{4,5}, 柳广春^{1,3}, 宋明伟^{1,3}

(1. 河南理工大学 测绘与国土信息工程学院, 河南 焦作 454000; 2. 太原理工大学 矿业工程学院, 山西 太原 030024; 3. 河南省采空区场地生态修复与建设技术工程研究中心, 河南 焦作 454000; 4. 山西能源学院 地质与测绘工程系, 山西 晋中 030600; 5. 中国科学院地理科学与资源研究所 中国科学院陆地表层格局与模拟重点实验室, 北京 100101)

摘要: 地下采煤区沉陷灾害发育重点区预测目前尚无固定程式, 且敏感区预测结果存在不确定性较大的问题。以山西省太原市西山地区沉陷灾害为研究对象, 分别以 2012 年和 2014 年核查编录的沉陷灾害数据为建模数据和验证数据, 以高程、坡度、坡向、地势起伏度、地面曲率、地层岩组、地质构造为敏感性评价因子, 综合运用 GIS 空间分析、统计分析和支持向量机(SVM) 等方法, 构建了 4 种核函数 SVM 沉陷灾害敏感性分区预测模型, 分别从模型的评价因子权重、模型优选、敏感性分区预测结果、预测精度和模型适用性进行了分析。结果表明: 多项式核函数 SVM 模型(PL-SVM) 的训练精度(受试者特征曲线下面积 AUC=0.854)与验证精度(AUC=0.755)均较高, 模型预测能力良好, 是 4 种模型中表现最好的模型, 所划分敏感性分区结果合理, 极高与高敏感区以较小面积分布较多沉陷灾害点, 而低敏感区则以较大面积分布极少沉陷灾害点。PL-SVM 模型预测的太原西山地区沉陷灾害发育极高、高、中和低敏感区的面积占比分别为: 20.19%、17.43%、21.18%、41.20%, 频率比值与敏感性等级之间呈良好的正相关, 符合线性函数关系。PL-SVM 模型敏感性评价结果可靠, 适用性好, 对地下采煤区沉陷灾害发育特征研究及灾害普查重点区预判具有参考意义。

关 键 词: 地下采煤区; 沉陷灾害; 敏感性分区; 评价因子; 支持向量机; 预测模型; 地质灾害

中图分类号: TD325; P208; X4 **文献标志码:** A **文章编号:** 1001-1986(2022)10-0108-11

SVM-based sensitivity zoning of subsidence disaster development in the underground coal mining areas

XUE Yong'an^{1,2}, ZOU Youfeng^{1,3}, ZHANG Wenzhi^{1,3}, ZHANG Mingmei^{4,5}, LIU Guangchun^{1,3}, SONG Mingwei^{1,3}

(1. School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China; 2. College of Mining Engineering, Taiyuan University of Technology, Taiyuan 030024, China; 3. Henan Engineering Research Center for Ecological Restoration and Construction Technology of Goaf Sites, Jiaozuo 454000, China; 4. Department of Geological and Surveying Engineering, Shanxi Institute of Energy, Jinzhong 030600, China; 5. Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China)

Abstract: At present, there is no fixed way to predict the sensitive zones of subsidence disaster development in underground coal mining areas, and the prediction result of sensitive zones has a great uncertainty. Herein, the subsidence disaster in Xishan area of Taiyuan City, Shanxi Province was taken as the research object. Totally 4 types of kernel SVM based prediction model for sensitivity zoning of subsidence disaster were constructed with the methods of GIS spatial analysis, statistical analysis and Support Vector Machine (SVM) in combination, taking the subsidence disaster data checked and recorded in 2012 and 2014 as the modeling and verification data respectively, as well as the elevation, slope

收稿日期: 2022-02-26; 修回日期: 2022-06-27

基金项目: 国家自然科学基金山西煤基低碳联合基金重点项目(U1810203); 山西省高等学校科技创新计划项目(2021L596); 河南省高校基本科研业务费专项资金项目(NSFRF220432)

第一作者: 薛永安, 1980 年生, 男, 陕西凤翔人, 博士研究生, 讲师, 从事 3S 技术融合应用、精密工程变形监测与预测研究.

E-mail: xueyongan@tyut.edu.cn

gradient, slope aspect, topographic relief, surface curvature, stratigraphic strata and geological structure as the sensitivity assessment factors. Meanwhile, analysis was performed on the weight of assessment factors, the model optimization, the prediction results of sensitivity zoning, the prediction accuracy, and the applicability of models respectively. The results show that the polynomial kernel-SVM model (PL-SVM) has relatively high training accuracy (with the area under the receiver characteristic curve of AUC=0.854) and validation accuracy (AUC=0.755), as well as good prediction capability. Thus, it has the best performance among the 4 types of models, and the sensitivity zoning is reasonable, with more points of subsidence disaster distributed in a small area of the very-high and high sensitive zones, while few points of subsidence disaster distributed in a large area of the low sensitive zones. As predicted by the PL-SVM model, the area proportion of very-high, high, moderate and low sensitive zones of subsidence disaster in Taiyuan Xishan area is 20.19%, 17.43%, 21.18% and 41.20%, respectively. Besides, the frequency ratio and the sensitivity grade are in good positive correlation, showing a linear functional relation. The sensitivity assessment result based on PL-SVM model is reliable and has good applicability, which has reference significance to the study on the development characteristics of subsidence disasters in underground coal mining areas and the prediction of key areas in disaster survey.

Keywords: underground coal mining area; subsidence disaster; sensitivity zoning; assessment factor; support vector machine; prediction model; geological disaster

地下矿产采出后,开采区周围岩土体的原始应力平衡状态被破坏,岩土体上出现位移与变形^[1],这是一个十分复杂的过程,是采矿、地质等许多因素及多场耦合综合影响的结果,在地表上表现为塌陷坑与地裂缝^[2],属于矿山地质灾害的范畴,具有渐进性特点,本文称为沉陷灾害。沉陷灾害是地下采煤扰动导致地表快速或缓慢沉降的一种地质灾害,在矿区分布广泛,既造成大量土地资源损毁,又严重破坏当地的地形地貌景观,还会严重影响周边植物生长^[2],也是诱发斜坡地质灾害和影响区域生态地质环境的重要因素。因此,开展地下采煤区沉陷灾害发育敏感性分区研究对国土综合整治与生态修复、下伏老采空区重大工程项目设计与稳定性监测、地质灾害防治与预警等至关重要,同时对地下采煤区生态环境修复战略^[3]具有助力作用。

灾害敏感性是指在特殊地形或某些因素作用下发生灾害的可能性^[4],灾害敏感性评价则是针对某一地区现存或潜在灾害空间分布概率的定性或定量分析。目前,地质灾害敏感性评价工作已经取得众多研究成果^[5],针对地表沉陷灾害敏感性评价,主要集中在开采地下水引起的地表沉降^[6-8]和废弃矿区地表沉降的灾害敏感性评价研究^[9-11],提出了众多评价模型并被应用和开展对比分析^[12-14],为沉陷灾害发育敏感性分区研究提供了评价模型选择参考。敏感性评价模型一般可分为定性模型和定量模型,随着研究的不断深入,更多评价模型被引入该领域^[15-16],为灾害风险评估与定量评价提供了良好的模型基础。目前,常用的评价模型主要包括:频率比(Frequency Ratio, FR)^[17]、确定性系数(Certainty Factor, CF)^[18]、支持向量机(Support Vector Machine, SVM)^[19]、随机森林(Random Forest, RF)^[7-8]和自适应神经模糊推理系统(Adaptive-Network-Based Fuzzy Inference System, ANFIS)^[20]等。在模型

应用中,确定地表沉陷影响因子是沉陷灾害发育敏感性分区评价的前提,如有研究者认为覆岩岩性、地表坡度、采空区范围、岩体力学性质、地下水等是诱发地表沉陷的主要影响因素^[21-22],沉陷灾害敏感性评价则是在上述基础上所做的进一步研究工作^[23]。评价因子的可靠性选择,评价模型的适用性对比,为开展地下采煤区沉陷灾害发育敏感性分区研究提供了丰富的理论与实践参考^[21-23]。

山西省太原市西山地区地下采煤强度高,沉陷灾害及次生地质灾害分布广泛,给人民生活安全及经济发展带来较大影响。文献[24]开展了崩塌、滑坡、不稳定斜坡和地面塌陷4种地质灾害与发育特征因子之间关联性的空间统计分析,为太原西山地区沉陷灾害发育敏感性评价提供了特征因子选择参考。文献[25]针对斜坡地质灾害敏感性评价中地势起伏度提取最佳尺度开展了研究,为太原西山地区地质灾害敏感性评价时地势起伏度因子可靠提取提供了依据。

针对地下采煤区沉陷灾害发育敏感性分区,笔者选择在小样本情况下应用广泛的SVM模型为敏感性评价模型,分别以不同年度核查编录的沉陷灾害作为模型建立与预测能力验证数据,开展太原市西山前山区沉陷灾害发育敏感性分区研究,以期为地下采煤区沉陷灾害及次生地质灾害防治提供科学的方法与数据支持。

1 研究区与数据源

1.1 研究区概况

太原市西山前山区位于山西省太原市西部山区(图1),北以尖草坪区、万柏林区和晋源区行政界线为边界,南东以山前坡脚线为界,总面积441.063 km²。区内采煤活动历史悠久,已近百年,煤矿开采主要为地下开采

方式,既有国有大矿,包括西铭矿、杜儿坪矿、官地矿、白家庄矿和西峪煤矿,还有分布广泛且数量庞大的乡办、村办、联营小型煤矿,逐渐在空间上形成了大矿与小矿既有邻近、又有重叠的地下采煤区,地表沉陷情况极其复杂,沉陷灾害及次生地质灾害分布广泛,部分地裂缝发育于山顶基岩或山体出露岩层(图 2)。

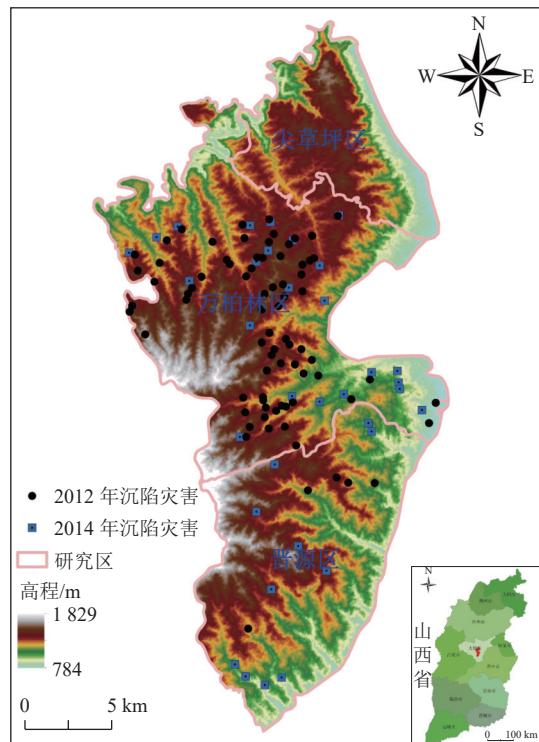


图 1 太原市西山前山区地理位置

Fig.1 Geographical map of Xishan area of Taiyuan City



图 2 太原市西山前山区沉陷灾害

Fig.2 Subsidence disaster in Xishan area of Taiyuan City

1.2 数据源

1) 沉陷灾害数据

收集到研究区 2012 年、2014 年核查编录的地面塌陷与地裂缝灾害共 113 处,提取地裂缝的中心位置,以点形式与地面塌陷构建沉陷灾害点图层,其空间分布如图 1 所示。其中,2012 年 79 处,2014 年 34 处,数据来源于山西省太原市规划和自然资源局。

2) DEM 数据

在地理空间数据云网站下载了研究区的 ASTER GDEM V2 数据,并重采样为 30 m 分辨率的 DEM 数

据,作为研究区的数字地貌因子提取基础数据,高程中误差为 $\pm 22.723 \text{ m}$ ^[25]。

3) 地质数据

研究区 1 : 200 000 数字地质图来源于山西省太原市规划和自然资源局。

2 研究方法

沉陷灾害发育敏感性分区步骤如下:

(1) 收集研究区实际核查编录的 2012 年和 2014 年的沉陷灾害空间分布数据,以 2012 年发育的沉陷灾害为训练样本建立模型,以 2014 年发育的沉陷灾害为验证样本评价模型预测精度。

(2) 从地形地貌和地质因素两个方面出发建立评价因子组合。

(3) 以评价因子提取结果输入基于机器学习思想的 SVM 模型开展沉陷灾害发育敏感性评价,以受试者特征曲线(Receiver Operating Characteristic, ROC)下面积(Area Under the ROC Curve, AUC)作为定量评价指标对分区结果进行精度验证。

(4) 以自然间断点法进行等级划分并制作研究区沉陷灾害发育敏感性分区图。

总体技术路线如图 3 所示。

2.1 评价因子

针对地下采煤活动对矿区微地貌、土地利用等带来的扰动影响,文献 [26-27] 选择研究区中西部的杜儿坪煤矿南三盘区为对象,其中,文献 [26] 采用 1 : 10 000 地形图所生产的 3 期 DEM 数据分别代表矿区采前、采中和采后 3 个时期的地形表面,开展了数字地貌时空演变特征分析,结果表明坡度与地势起伏度统计结果随时间推进而变化明显,坡向也有一定的影响。因此,DEM 数据的获取时间对数字地貌因子分析结果是否可靠具有重要约束,灾害敏感性预测研究中应尽可能采用与沉陷灾害发育时间相匹配的 DEM 数据进行数字地貌因子提取。文献 [27] 则以采前、采中和采后的 5 期遥感影像为数据源,采用最小距离分类法进行了矿区土地利用信息提取与演变分析,结果表明,土地利用类型转移过程和方向与煤炭开采扰动过程相吻合。因此,遥感影像时相不同,所提取的土地利用信息则不同,以特定年份的遥感影像所提取的土地利用数据和植被覆盖数据与历史编录的沉陷灾害数据进行相关分析,其结果不确定性极大。

沉陷灾害影响因子众多,在开采沉陷预计中多采用地形地貌、地质因素、采煤方法、煤层厚度、埋深、工作面尺寸等因素^[1,28],专家学者基于不同目的建立了不同的影响因子序列。受地下采煤相关数据获取困难、

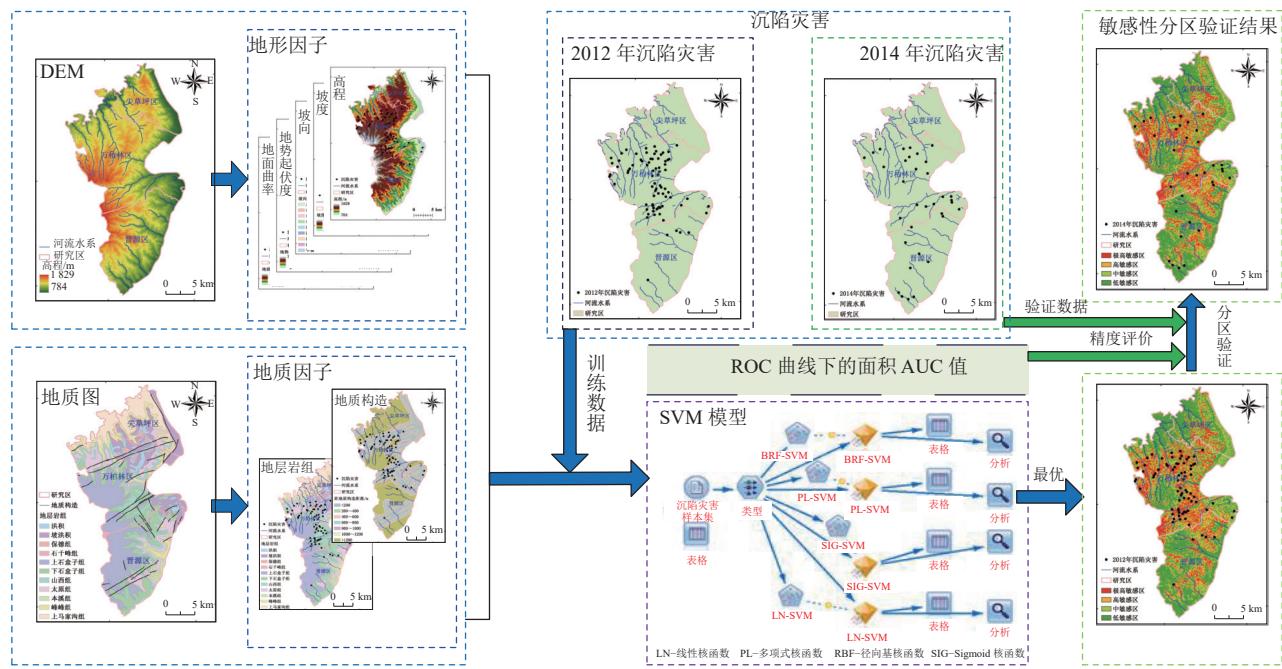


图 3 总体技术路线
Fig.3 Technology roadmap for the study

定量表述不足制约,并考虑方法的通用性,本文在现有研究的基础上^[24-25],综合考虑地下采煤区数字地貌特征和土地利用时空演变的影响^[26-27],从地形地貌与地

质因素出发,选取高程、坡度、坡向、地势起伏度、地面曲率、地层岩组、地质构造共 7 个因子作为沉陷灾害发育敏感性评价因子组合(表 1)。

表 1 研究区沉陷灾害敏感性评价因子
Table 1 Assessment factors of subsidence disaster sensitivity in the study area

影响因素	评价因子	数据分类	孕灾特征	备注
地形地貌	高程/m	784~1 829	影响植被类型、植被覆盖度	连续值
	坡度/(°)	0~66.20	坡度大易引起滑移破坏	连续值
	坡向	平面; 北; 东北; 东; 东南; 南; 西南; 西; 西北	影响光照、风的干湿、降雨、植被覆盖和土壤潮湿程度等	分类值
	地势起伏度/m	9~343	地形起伏大, 易引起滑移破坏	连续值, 9×9 网格提取 ^[25]
地质因素	地面曲率	-17.529~17.072	定量表征地形表面某点的扭曲变化程度	连续值
	距地质构造距离/m	<200; ≥200~400; ≥400~600; ≥600~800; ≥800~1 000; ≥1 000~1 200; ≥1 200	地质构造附近易形成地裂缝	分类值, 距构造距离
	地层岩组	洪积; 坡洪积; 保德组; 石千峰组; 上石盒子组三段; 上石盒子组二段; 上石盒子组一段; 下石盒子组; 山西组; 太原组; 本溪组; 峰峰组; 上马家沟组	覆岩性质对地表破坏形式具有重要影响	分类值, 组

以 ASTER GDEM V2 数据在 ArcGIS 软件中提取高程、坡度、坡向、地势起伏度和地面曲率, 基于 1:200 000 地质图提取地层岩组与地质构造信息, 利用 ArcGIS 软件制作研究区 2012 年沉陷灾害发育敏感性评价因子图(图 4)。

经图 4 统计, 沉陷灾害在各行政区之间的空间分布规模差异性较大, 2012 年的平均发育密度为 0.179 处/km², 万柏林区的发育密度最高, 达到 0.336 处/km², 高于研

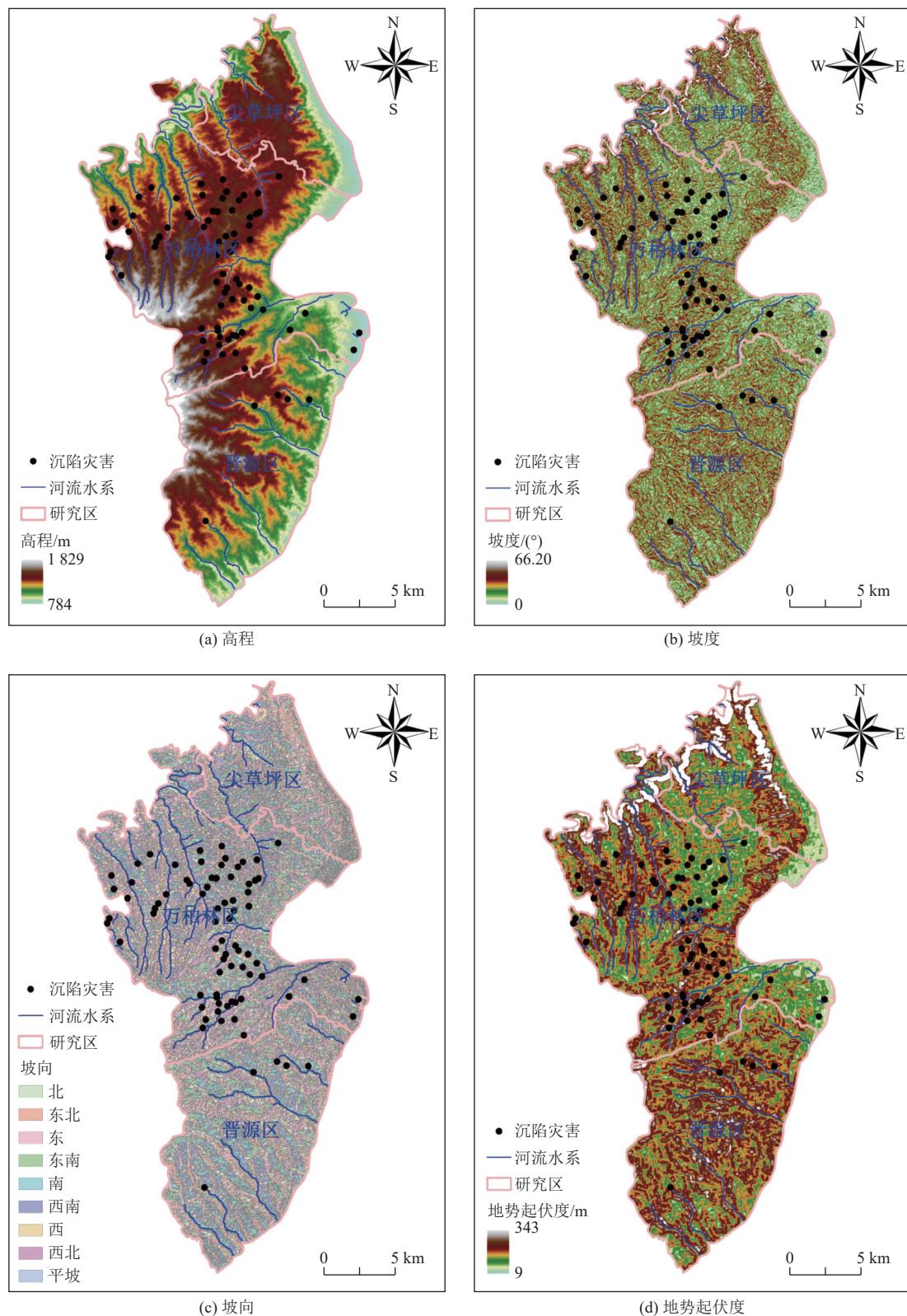
究区平均水平。其次是晋源区, 为 0.034 处/km², 发育密度相对较低。

2.2 支持向量机(SVM)

支持向量机是一种基于最大间隔的线性判别分类方法, 其功能强大, 可以最大化模型的预测准确度, 被认为是目前针对小样本统计估计和预测学习的最佳理论, 最早被用于研究小样本情况下的机器学习^[29], 在地质灾害敏感性评价研究中应用较多^[13,30]。

SVM 通过引入核函数有效地解决了非线性分类问题,使敏感性评价中的非线性分类计算复杂度不再取决于空间维数,而是取决于用于建模的沉陷灾害样本数量,尤其是其中支持向量的灾害点数量。显然,核函数是 SVM 克服维数灾难实现沉陷灾害敏感性评价有效分类的关键。SVM 常用核函数(表 2)主要有线性核函数(Linear Kernel, LN)、多项式核函数(Polynomial Kernel, PL)、径向基核函数(Radial Basis Function, RBF)和 Sigmoid 核函数(Sigmoid Kernel, SIG)^[13]。

目前,地质灾害敏感性评价中径向基核函数应用较多^[30],但研究区内沉陷灾害样本数量较少,应对比 4 种核函数 SVM 模型的预测精度,选择最优核函数建



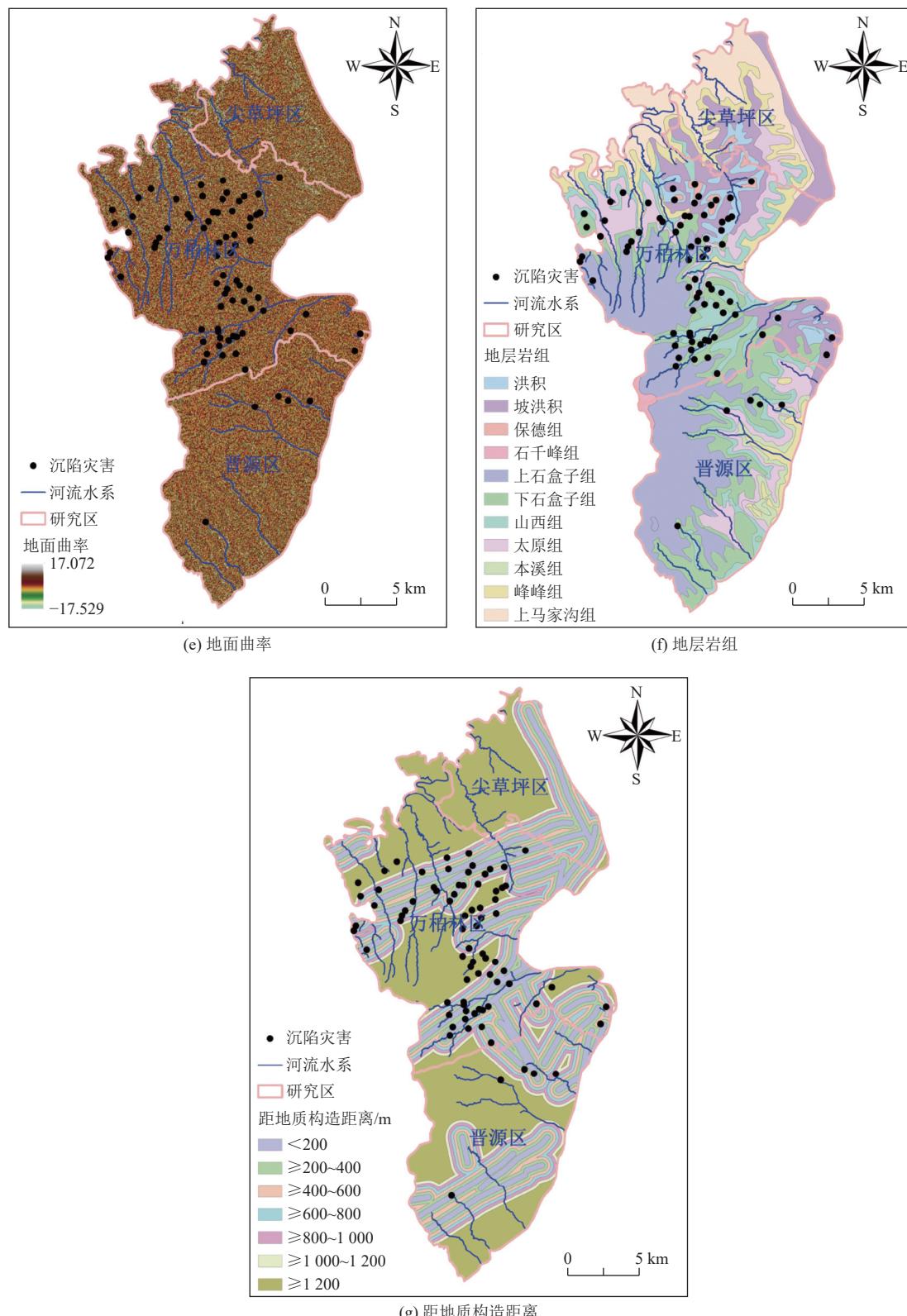


图4 太原市西山前山区沉陷灾害评价因子

Fig.4 Assessment factors of the subsidence disaster in Xishan area of Taiyuan City

立 SVM 预测模型。

沉陷灾害敏感性分区 SVM 模型预测步骤如下:

(1) 建立研究区沉陷灾害数据集, 包括训练样本集和验证样本集。

(2) 提取沉陷灾害点在评价因子图中的值, 作为属性值记录下来, 包括: 高程、坡度、坡向、地势起伏度、地面曲率、地层岩组和距地质构造的距离。

(3) 建立 4 种 SVM 模型, 包括 LN-SVM 模型、PL -

表 2 SVM 常用核函数^[31]Table 2 Common kernel functions of SVM^[31]

核函数类型	核函数	解决问题
线性核函数	$k(x_i, x_j) = x_i^T x_j$	线性
多项式核函数	$k(x_i, x_j) = (\gamma x_i^T x_j + C)^d$	偏线性
径向基核函数	$k(x_i, x_j) = \exp(-\gamma x_i - x_j^2)$	偏非线性
Sigmoid核函数	$k(x_i, x_j) = \tanh((\gamma x_i^T x_j + C))$	非线性

SVM 模型、RBF-SVM 模型和 SIG-SVM 模型。

(4) 采用训练样本对 4 种 SVM 模型进行训练, 统计评价因子在每一种模型中的权重排序, 分析模型预测正确率与错误率, 避免出现过拟合和欠拟合。

(5) 选择预测正确率最高的 SVM 模型作为沉陷灾害敏感性分区预测模型。

(6) 采用验证样本评价最优 SVM 模型的预测精度, 对比训练样本的预测精度, 分析预测结果的合理性。

2.3 精度评价

敏感性评价是一个二值分类问题, 一般将样本分为正类和负类, 现有研究中通常采用真正类率和真负类率构建受试者特征曲线(ROC), 利用 ROC 曲线下的面积(AUC 值)作为敏感性分区预测精度的度量^[13,31], AUC 值的范围为 [0,1], 值越大越好, 表明模型预测能力越强, 灾害敏感性分区预测结果精度越高。

根据 AUC 值的大小可以划分敏感性分区预测模型的精度等级(表 3)^[13]。

表 3 敏感性预测模型精度等级^[13]Table 3 Accuracy levels of sensitivity prediction model^[13]

AUC	0~0.6	>0.6~0.7	>0.7~0.8	>0.8~0.9	>0.9~1.0
精度等级	差	一般	好	非常好	极好

对于二值分类问题, 假设 C_1 为正类, C_2 为负类, 对应的 2×2 混淆矩阵见表 4。其中, TP 表示分类方法准确预测为 C_1 的点的数目; FP 表示分类方法预测为 C_1 , 但实际属于 C_2 的点的数目; FN 表示分类方法预测为 C_2 , 但实际属于 C_1 的点的数目; TN 表示分类方法准确预测为 C_2 的点的数目。

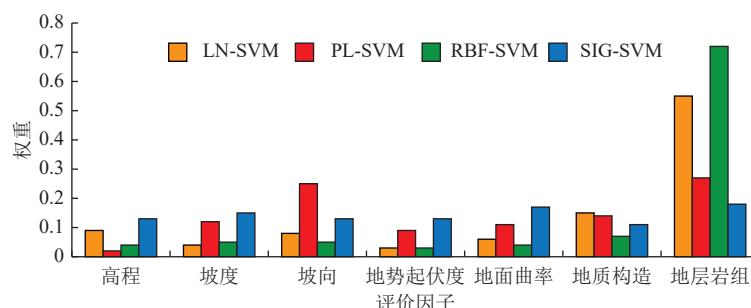


图 5 研究区不同核函数 SVM 模型沉陷灾害敏感性评价因子权重对比

Fig.5 Comparison of the weight of subsidence disaster sensitivity evaluation factors in different kernel function SVM models in the study area

表 4 二值分类混淆矩阵

Table 4 Confusion matrix of binary classification

预测类	真实类	
	正类(C_1)	负类(C_2)
正类(C_1)	真正类 (TP)	假正类 (FP)
负类(C_2)	假负类 (FN)	真负类 (TN)

敏感性 S_p , 也称为真正类率, 是 C_1 中所有点被正确预测的比例, 计算公式^[32]如下:

$$S_p = \frac{TP}{TP+FN} \quad (1)$$

特异性(X_p), 也称为真负类率, 计算公式^[33]如下:

$$X_p = \frac{TN}{FP+TN} \quad (2)$$

AUC 值计算公式^[33]如下:

$$AUC = \sum_{p=1}^n (X_{p+1} - X_p) \times (S_{p+1} - S_p - S_p/2) \quad (3)$$

3 沉陷灾害发育敏感性分区结果与讨论

3.1 结果

本文按照图 3 的技术流程开展研究区沉陷灾害发育敏感性分区预测。

训练样本和验证样本分别随机生成同等数量的非沉陷灾害样本, 建立由 79(沉陷灾害)+79(非沉陷灾害)个样本点组成的训练样本数据集, 共有训练样本点 158 个。建立由 34(沉陷灾害)+34(非沉陷灾害)个样本点组成的验证样本数据集, 共有验证样本点 68 个。其中, 沉陷灾害样本点与非沉陷灾害样本点之间的距离为 500 m。

3.1.1 评价因子权重

利用 SPSS 软件统计 4 种常用核函数 SVM 模型评价因子的重要性排序, 结果对比如图 5 所示。

图 5 显示, 4 种模型中评价因子的权重顺序并不相同, 其权重从高到低的排列顺序分别为:

(1) LN-SVM 模型: 地层岩组, 地质构造, 高程, 坡向, 地面曲率, 坡度, 地势起伏度。

(2) PL-SVM模型: 地层岩组, 坡向, 地质构造, 坡度, 地面曲率, 地势起伏度, 高程。

(3) RBF-SVM模型: 地层岩组, 地质构造, 坡向, 坡度, 地面曲率, 高程, 地势起伏度。

(4) SIG-SVM模型: 地层岩组, 地面曲率, 坡度, 地势起伏度, 坡向, 高程, 地质构造。

3.1.2 SVM模型优选

4种核函数SVM模型预测正确率在SPSS软件中的统计结果见表5。

由表5可以看出, 4种核函数SVM模型的预测正确率由高到低排列为: PL-SVM模型、LN-SVM模型、RBF-SVM模型、SIG-SVM模型, 其中, PL-SVM模型表现最优。

3.1.3 评价结果

通过因子权重分析及预测正确率对比, PL-SVM模型的因子重要性排序更合理, 预测正确率最高, 明显

表5 研究区不同核函数SVM模型预测正确率
Table 5 Prediction accuracy of SVM models with different kernel functions in the study area

模型	预测正确率/%	错误率/%
LN-SVM	75.32	24.68
PL-SVM	92.41	7.59
RBF-SVM	72.15	27.85
SIG-SVM	62.66	37.34

注: 正确率=(样本正确数/样本数)×100%; 错误率=1-正确率。

优于其他3种模型。因此, 本文采用PL-SVM模型进行研究区沉陷灾害发育敏感性分区预测, 并在ArcGIS中采用自然间断点法将研究区划分为4个敏感性等级, 分别为: 极高敏感区、高敏感区、中等敏感区和低敏感区, 得到研究区沉陷灾害敏感性分区结果(图6a), 验证结果如图6b所示, 对训练样本和验证样本的分区统计结果见表6。

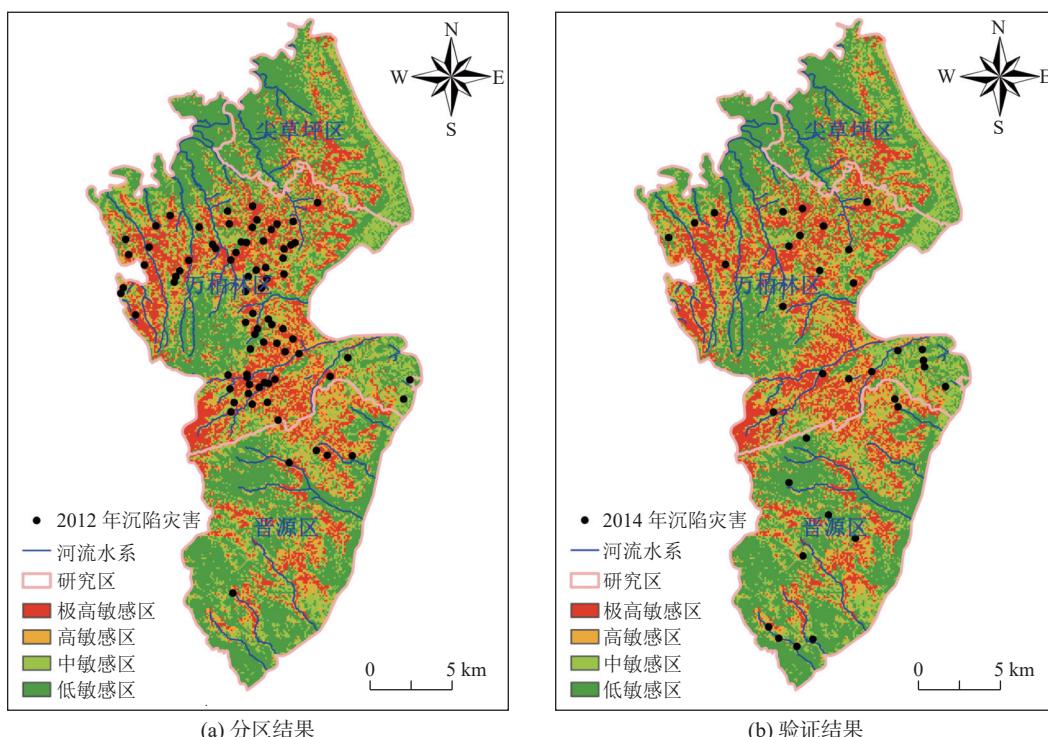


图6 研究区沉陷灾害敏感性分区与验证
Fig.6 Sensitivity zoning and verification of subsidence disaster in the study area

表6 研究区沉陷灾害敏感性分区统计
Table 6 Statistics on sensitivity zoning of subsidence disaster in the study area

敏感性等级	分区栅格数	栅格比例/%	训练样本数	灾害点比例/%	验证样本数	灾害点比例/%	训练样本频率比	验证样本频率比
极高	8 899	20.19	39	49.37	13	38.24	2.45	1.89
高	7 680	17.43	23	29.11	9	26.47	1.67	1.52
中	9 334	21.18	14	17.72	7	20.59	0.84	0.97
低	18 160	41.20	3	3.80	5	14.71	0.09	0.36

由表 6 可以看出, PL-SVM 模型通过训练样本和验证样本分别统计得到的频率比值(频率比=灾害点比例/分区栅格比例)均随敏感性等级升高而递增, 经对比其趋势均符合线性函数, 拟合方程分别为:

(1) 训练样本频率比趋势: $y = -0.791x + 3.24, R^2 = 0.9997$ 。

(2) 验证样本频率比趋势: $y = -0.514x + 2.47, R^2 = 0.9887$ 。

其中, y 为频率比值; x 为敏感性等级, 极高为 1, 高为 2, 中为 3, 低为 4。

可以看出, 2 个拟合方程的决定系数 R^2 的值均接近于 1, 表明 PL-SVM 模型通过训练样本和验证样本所计算的频率比值与敏感性等级之间呈良好的正相关, 均以更小的极高与高敏感性分区面积分布了更多的沉陷灾害。训练样本频率比的趋势较验证样本频率比的趋势相关性更好, 表明 2012 年与 2014 年发育的沉陷灾害在研究区内的空间分布略有不同。

3.1.4 精度评价

利用式(3)分别计算训练样本与验证样本在 PL-SVM 模型敏感性分区预测结果中的 AUC 值, 其中, 以 2014 年沉陷灾害作为验证样本的 PL-SVM 预测模型的 AUC 值为 0.755, 低于以 2012 年沉陷灾害作为训练样本的 PL-SVM 预测模型的 AUC 值(0.854), 表明 PL-SVM 模型在验证样本集的预测精度略低于训练样本集, 符合预测模型在测试数据集上的精度一般低于训练数据集精度的规律。但 2 种模型预测精度均较高, 训练模型精度等级为非常好, 预测模型精度等级为好。预测精度表明, 模型在训练样本与验证样本上既没有过拟合, 也没有欠拟合, 结果较为合理。

3.2 讨论

3.2.1 评价因子权重

由图 5 可以看出, 地层岩组在 4 种模型中均是最重要的评价因子, 表明沉陷灾害发育与地层岩组分布关系密切, 是地面塌陷与地裂缝发育的最重要影响因素。尽管在 SIG-SVM 模型中地质构造表现为影响最弱的因子, 但从 4 种模型中前 3 位因子出现的频率可以看出, 地质构造对沉陷灾害发育的影响仅次于地层岩组, 表明在构造断裂附近, 岩层破碎, 地下采煤扰动再次打破断裂附近的岩体力学平衡, 诱发地表塌陷或开裂。

由图 5 可知, 地势起伏度因子在 LN-SVM 和 RBF-SVM 模型中均是沉陷灾害发育影响最弱的因子, 在 PL-SVM 模型中为次最弱因子, 表明区域范围内高程最大值与最小值之间的差值大小对沉陷灾害发育影响较小。一般情况下, 当沉陷灾害处于沟谷边缘时极易

诱发崩塌与滑坡灾害, 而沉陷灾害仅包括地面塌陷与地裂缝, 因此, 地势起伏度因子在敏感性评价中贡献较小。

由图 5 同时可以看出, LN-SVM 和 RBF-SVM 模型中地层岩组的权重远超其他因子, SIG-SVM 模型中各因子的权重相对平衡, 地层岩组略高于其他因子, 而 PL-SVM 模型中地层岩组与坡向因子的权重明显高于其他因子, 在评价因子组合分析中, 既没有过分依赖某一个因子, 也没有分散因子权重, 其合理性高于其他 3 种核函数模型。

3.2.2 预测模型优选

SVM 预测模型的性能优劣主要取决于核函数的选择, 本文分别对比了 LN-SVM 模型、PL-SVM 模型、RBF-SVM 模型和 SIG-SVM 模型在研究区沉陷灾害敏感性分区中的预测精度, 预测正确率分别为 75.32%、92.41%、72.15%、62.66%, 其中, PL-SVM 模型预测正确率明显高于其他 3 种核函数 SVM 模型, 分别比 LN-SVM 模型、RBF-SVM 模型和 SIG-SVM 模型的预测正确率高 22.69%、28.08% 和 47.48%。SIG-SVM 模型的预测正确率是 4 种模型中最低的, 分别比 LN-SVM 模型、PL-SVM 模型和 RBF-SVM 模型的预测正确率低 16.81%、32.19% 和 13.15%。

由图 5 可知, PL-SVM 模型的评价因子权重排序合理性最好, 而 SIG-SVM 模型则最差, 与预测正确率完全一致, 选择 PL-SVM 模型作为研究区沉陷灾害发育敏感性分区预测模型是可靠的。

3.2.3 沉陷灾害敏感性分区预测结果

敏感性评价研究中通常采用总样本的 70% 或 80% 作为训练点建立预测模型, 剩余 30% 或 20% 作为验证点验证预测模型精度^[30-31]。本文以研究区 2012 年和 2014 年实际发育的沉陷灾害分别作为训练样本和验证样本, 基于 PL-SVM 模型进行沉陷灾害发育敏感性分区预测。结果表明, 2012 年沉陷灾害发育在极高、高、中、低 4 个等级的数量百分比分别为 49.37%、29.11%、17.72%、3.80%, 极高和高敏感区分布了共 78.48% 的沉陷灾害。2014 年沉陷灾害发育在极高、高、中、低 4 个等级的数量百分比分别为 38.24%、26.47%、20.59%、14.71%, 极高和高敏感区分布了共 64.71% 的沉陷灾害。

由图 6 可知, 2012 年沉陷灾害在研究区内有明显的聚集特征, 2014 年则演变为随机分布, 部分沉陷灾害出现在低敏感区, 导致敏感性分区预测结果对验证样本(AUC=0.755)的精度略低于训练样本(AUC=0.854), 但图 6b 可以看出, 模型的预测结果良好, 2014 年新发育的沉陷灾害大部分位于极高和高敏感区, 为沉陷灾

害核查提供了重点区预测, 可极大减少野外调查的盲目性与野外工作量。

3.2.4 综合分析

目前, 地下采煤区地质灾害敏感性评价研究中一般将多年灾害点混合作为样本进行模型训练, 但忽视了受采煤扰动影响的矿区地貌特征和土地利用时空演变^[26-27], 采用特定年份的灾害数据进行建模又需要面对样本数量过少对敏感性评价结果所带来的不确定性影响。今后应就此问题继续展开研究, 利用遥感解译手段增加特定年份灾害样本数量, 提升区域灾害点小样本情况下沉陷灾害发育敏感性分区预测精度。

研究区含煤地层出露, 极高和高敏感区主要分布于王封煤业、西铭煤矿、杜儿坪煤矿、西峪煤矿、白家庄煤矿等多个大型煤矿区域, 相比较地形地貌与地质因素, 大、小矿地下采煤扰动是研究区沉陷灾害发育的主控因素。本文基于方法适用性考虑, 未将地下采矿扰动影响列为敏感性评价指标, 今后应继续研究如何定量化引入地下采矿相关指标参与沉陷灾害敏感性评价。

太原市西山地区近年所开展的采煤沉陷区治理效果明显, 玉泉山城郊森林公园、长风城郊森林公园、四达沟生态恢复景区等分布于研究区东缘, 生态地质环境恢复较好, 但沉陷灾害极高和高敏感区与植被覆盖区重叠度较高, 对区内地下水保护形成潜在威胁, 应防止水资源破坏导致植被衰退而引起矿区生态环境再次恶化^[28], 也是矿山生态环境治理需要重点关注的区域。同时, 分布于林地内的沉陷灾害点给遥感解译和野外调查带来困难, 灾害数据集完备性受限, 给敏感性分区评价结果的精准性与可靠性带来一定的影响。

4 结论

a. PL-SVM 模型是研究区沉陷灾害敏感性分区预测的最优模型, 分别以 2012 年和 2014 年实际发育沉陷灾害点为训练模型和验证模型, 其训练精度与验证精度均较高, 且训练精度较验证精度高, 对地下采煤区沉陷灾害发育敏感性分区具有良好的适用性和预测性。

b. PL-SVM 模型划分了研究区沉陷灾害极高、高、中和低敏感区 4 个等级的面积比例, 极高与高敏感区主要分布于万柏林区和晋源区边山一带, 与研究区内大矿与小矿的重合带高度吻合。灾害点比例随敏感性等级升高而递增, 与敏感性等级之间呈现良好的线性正相关。分区结果合理, 对研究区沉陷灾害普查与防治重点区确定具有指导价值, 对开展同类研究具有参考意义。

c. 应进一步改进、完善评价因子组合, 并选择更

多评价模型开展对比, 组建更可靠的沉陷灾害发育敏感性预测体系, 同时收集研究区最新的沉陷灾害数据进行模型验证, 为地下采煤区沉陷灾害发育敏感性分区预测提供最适用的结果。

参考文献(References)

- [1] 邹友峰, 邓喀中, 马伟民. 矿山开采沉陷工程[M]. 徐州: 中国矿业大学出版社, 2003.
- [2] 毕银丽, 伍越, 张健, 等. 采用 HYDRUS 模拟采煤沉陷地裂缝区土壤水盐运移规律[J]. 煤炭学报, 2020, 45(1): 360–367.
BI Yinli, WU Yue, ZHANG Jian, et al. Simulation of soil water and salt movement in mining ground fissure zone based on HYDRUS[J]. Journal of China Coal Society, 2020, 45(1): 360–367.
- [3] 彭苏萍, 毕银丽. 黄河流域煤矿区生态环境修复关键技术与战略思考[J]. 煤炭学报, 2020, 45(4): 1211–1221.
PENG Suping, BI Yinli. Strategic consideration and core technology about environmental ecological restoration in coal mine areas in the Yellow River Basin of China[J]. Journal of China Coal Society, 2020, 45(4): 1211–1221.
- [4] POURGHASEMI H R, PRADHAN B, GOKCEOGLU C. Application of fuzzy logic and Analytical Hierarchy Process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran[J]. Natural Hazards, 2012, 63(2): 965–996.
- [5] PHAM B T, SHIRZADI A, BUI D T, et al. A hybrid machine learning ensemble approach based on a radial basis function neural network and rotation forest for landslide susceptibility modeling: A case study in the Himalayan area, India[J]. International Journal of Sediment Research, 2018, 33(2): 157–170.
- [6] ZAMANIRAD M, SARRAF A, SEDGHI H, et al. Modeling the influence of groundwater exploitation on land subsidence susceptibility using machine learning algorithms[J]. Natural Resources Research, 2020, 29: 1127–1141.
- [7] ELMAHDY S I, MOHAMED M M, ALI T A, et al. Land subsidence and sinkholes susceptibility mapping and analysis using random forest and frequency ratio models in Al Ain, UAE[J]. Geocarto International, 2020, 37(1): 315–331.
- [8] EBRAHIMY H, FEIZIZADEH B, SALMANI S, et al. A comparative study of land subsidence susceptibility mapping of Tasuj plane, Iran, using boosted regression tree, random forest and classification and regression tree methods[J]. Environmental Earth Sciences, 2020, 79(10): 223.
- [9] LEE S, HYUN-JOO O, KI-DONG K. Statistical spatial modeling of ground subsidence hazard near an abandoned underground coal mine[J]. Disaster Advances, 2010, 3(1): 11–23.
- [10] OH H, AHN S, CHOI J, et al. Sensitivity analysis for the GIS-based mapping of the ground subsidence hazard near abandoned underground coal mines[J]. Environmental Earth Sciences, 2011, 64(2): 347–358.
- [11] PRAKASH S S, MANISH Y, JYOTI D A, et al. Multivariate statistical approach for assessment of subsidence in Jharia coalfields, India[J]. Arabian Journal of Geosciences, 2017, 10(8): 191.

- [12] OH H, SYIFA M, LEE C, et al. Land subsidence susceptibility mapping using bayesian, functional, and meta-ensemble machine learning models[J]. *Applied Sciences*, 2019, 9(6): 1248.
- [13] MOHAMMADY M, POURGHASEMI H R, AMIRI M. Assessment of land subsidence susceptibility in Semnan plain (Iran): A comparison of support vector machine and weights of evidence data mining algorithms[J]. *Natural Hazards*, 2019, 99(2): 951–971.
- [14] EMAMI S N, YOUSEFI S, POURGHASEMI H R, et al. A comparative study on machine learning modeling for mass movement susceptibility mapping (a case study of Iran)[J]. *Bulletin of Engineering Geology and the Environment*, 2020, 79: 5291–5308.
- [15] ARABAMERI A, SAHA S, ROY J, et al. A novel ensemble computational intelligence approach for the spatial prediction of land subsidence susceptibility[J]. *Science of the Total Environment*, 2020, 726: 138595.
- [16] NADIRI A A, KHATIBI R, KHALIFI P, et al. A study of subsidence hotspots by mapping vulnerability indices through innovative “ALPRIFT” using artificial intelligence at two levels[J]. *Bulletin of Engineering Geology and the Environment*, 2020, 79: 3989–4003.
- [17] BIANCHINI S, SOLARI L, SOLDATO M D, et al. Ground subsidence susceptibility (GSS) mapping in Grosseto plain (Tuscany, Italy) based on satellite InSAR data using frequency ratio and fuzzy logic[J]. *Remote Sensing*, 2019, 11(17): 11.
- [18] 覃乙根, 杨根兰, 谢金, 等. 贵州省开阳县斜坡地质灾害孕灾因子敏感性分析[J]. *煤田地质与勘探*, 2020, 48(4): 190–198.
QIN Yigen, YANG Genlan, XIE Jin, et al. Sensitivity analysis of disaster-pregnant environmental factors for slope geological hazards in Kaiyang County, Guizhou Province[J]. *Coal Geology & Exploration*, 2020, 48(4): 190–198.
- [19] ABDOLLAHI S, POURGHASEMI H R, GHANBARIAN G A, et al. Prioritization of effective factors in the occurrence of land subsidence and its susceptibility mapping using an SVM model and their different kernel functions[J]. *Bulletin of Engineering Geology and the Environment*, 2019, 78(6): 4017–4034.
- [20] GHORBANZADEH O, ROSTAMZADEH H, BLASCHKE T, et al. A new GIS-based data mining technique using an Adaptive Neuro-Fuzzy Inference System (ANFIS) and k-fold cross-validation approach for land subsidence susceptibility mapping[J]. *Natural Hazards*, 2018, 94: 497–517.
- [21] SINGH K B, DHAR B B. Sinkhole subsidence due to mining[J]. *Geotechnical & Geological Engineering*, 1997, 15(4): 327–341.
- [22] SAHU P, LOKHANDE R D. An investigation of sinkhole subsidence and its preventive measures in underground coal mining[J]. *Procedia Earth and Planetary Science*, 2015, 11: 63–75.
- [23] WU Zhiyong, NIU Qinghe, LI Wenping, et al. Ground stability evaluation of a coal-mining area: A case study of Yingshouyingzi mining area, China[J]. *Journal of Geophysics and Engineering*, 2018, 15(5): 2252–2265.
- [24] 张明媚, 葛永慧, 薛永安, 等. 地下采煤区地质灾害发育空间特征及其成因[J]. *太原理工大学学报*, 2019, 50(4): 472–477.
ZHANG Mingmei, GE Yonghui, XUE Yong'an, et al. Spatial characteristics and genesis of geological disasters in underground coal mining area[J]. *Journal of Taiyuan University of Technology*, 2019, 50(4): 472–477.
- [25] 张明媚, 薛永安. 斜坡地质灾害敏感性评价中地势起伏度提取最佳尺度研究[J]. *太原理工大学学报*, 2020, 51(6): 881–888.
ZHANG Mingmei, XUE Yong'an. Optimal scale for extracting relief amplitude in slope geological hazard sensitivity evaluation[J]. *Journal of Taiyuan University of Technology*, 2020, 51(6): 881–888.
- [26] 张明媚, 薛永安, 吕义清, 等. 地下采煤扰动影响区数字地貌时空演变特征[J]. *煤矿安全*, 2019, 50(5): 243–246.
ZHANG Mingmei, XUE Yong'an, LYU Yiqing, et al. Temporal and spatial evolution characteristics of digital geomorphology in affected zone of underground coal mining disturbance[J]. *Safety in Coal Mines*, 2019, 50(5): 243–246.
- [27] 张明媚, 薛永安, 吕义清, 等. 地下采煤扰动影响土地利用时空演变研究[J]. *中国煤炭*, 2019, 45(5): 102–106.
ZHANG Mingmei, XUE Yong'an, LYU Yiqing, et al. Research on temporal-spatial evolution of land utilization affected by underground coal mining disturbance[J]. *China Coal*, 2019, 45(5): 102–106.
- [28] 邹友峰. 高强度开采地表生态环境演变机理与调控[M]. 北京: 科学出版社, 2019.
- [29] GIROSI F. An equivalence between sparse approximation and support vector machines[J]. *Neural Computation*, 1998, 10(6): 1455–1480.
- [30] 徐胜华, 刘纪平, 王想红, 等. 熵指数融入支持向量机的滑坡灾害易发性评价方法: 以陕西省为例[J]. *武汉大学学报·信息科学版*, 2020, 45(8): 1214–1222.
XU Shenghua, LIU Jiping, WANG Xianghong, et al. Landslide susceptibility assessment method incorporating index of entropy based on support vector machine: A case study of Shaanxi Province[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(8): 1214–1222.
- [31] 唐睿旋, 晏鄂川, 唐薇. 基于粗糙集和 BP 神经网络的滑坡易发性评价[J]. *煤田地质与勘探*, 2017, 45(6): 129–138.
TANG Ruixuan, YAN Echuan, TANG Wei. Landslide susceptibility evaluation based on rough set and back-propagation neural network[J]. *Coal Geology & Exploration*, 2017, 45(6): 129–138.
- [32] 古天龙, 李龙, 常亮, 等. 公平机器学习: 概念、分析与设计[J]. *计算机学报*, 2022, 46(5): 1018–1051.
GU Tianlong, LI Long, CHANG Liang, et al. Fair machine learning: Concepts, analysis, and design[J]. *Chinese Journal of Computers*, 2022, 46(5): 1018–1051.
- [33] WANG Yue, WEN Haijia, SUN Deliang, et al. Quantitative assessment of landslide risk based on susceptibility mapping using random forest and geo-detector[J]. *Remote Sensing*, 2021, 13(13): 2625.

(责任编辑 周建军 聂爱兰)