

多源数据支持下的三峡库区 滑坡灾害空间智能预测

武雪玲¹ 任 福² 牛瑞卿¹

(1 中国地质大学(武汉)地球物理与空间信息学院,武汉市鲁磨路 388 号,430074)

(2 武汉大学资源与环境科学学院,武汉市珞喻路 129 号,430079)

摘 要:针对传统空间分析技术不易发掘多源、海量滑坡数据中隐藏的模式、趋势和关系等问题,以三峡库区为研究对象,通过多源数据融合提取滑坡孕灾环境和影响因素信息,进而利用动态构建算法建立 BP 神经网络模型,定量预测滑坡空间易发性,生成滑坡易发性分区图。采用成功率曲线分析预测结果,三层 BP 神经网络的预测精度达到 89.75%,预测结果与野外调查实际情况吻合较好。

关键词:多源数据;三峡库区;滑坡;智能;空间预测

中图分类号:P237.9;P208

三峡水库自 2008-10 首次 172.8 m 试验性高水位运行以来,在库水周期性波动的作用下,库岸岩土体斜坡受到影响^[1-3],威胁三峡大坝的正常运行和库区居民的生命财产安全。因此,三峡库区高水位蓄水后,滑坡稳定性定量分析成为库区地质灾害防治中亟待解决的问题。

国内外学者对滑坡稳定性分析已经开展了一系列相关研究,主要集中在:① 通过多源、海量空间数据融合,挖掘滑坡时空分布特征及其诱发因素信息^[4-5],在空间尺度上覆盖从宏观到微观,在时间尺度上覆盖从间断到连续^[6-9]。② 针对多源、海量的滑坡灾害数据,顾及滑坡灾害的不确定性和非线性特征,采用数据挖掘中的智能算法,定量分析滑坡稳定性^[10-13]。尽管上述研究提高了滑坡稳定性分析的效率,但亟需发展利用多维滑坡信息提取的滑坡稳定性智能分析模型。

本文选择三峡库区秭归至巴东库岸段长江干流岸坡作为研究区,通过多源滑坡空间数据融合提取滑坡孕灾环境和影响因素信息,利用动态构造算法建立 BP 神经网络结构模型,定量预测滑坡稳定性,采用统计分析法绘制滑坡稳定性分区图,并通过成功率曲线评价预测精度。

1 BP 神经网络

滑坡灾害的孕育是一个时间和空间尺度上线性与非线性、均变与突变、确定性与非确定性、规则与随机共存的过程,目前仍然无法精确确定影响滑坡灾害的全部影响因素及其影响程度。BP 神经网络能对网络中各层的权系数进行修正,适用于处理知识背景不清楚、推理规则不明确的复杂、模糊、随机的信息识别问题,能为客观反映滑坡灾害易发程度提供一条有效途径。

BP 算法是前馈多层网络结构,包括输入层、隐含层及输出层。隐含层有一个或者多个,不与外界连接,但它们的能影响输入输出之间的关系,隐含层神经元采用一个单调上升的非线性可连续可微函数——Sigmoid 型激活函数^[14]。相邻层神经元之间用连接权系数相互连接,而同层神经元之间没有连接。本文采用动态构造方法,按照 BP 网络的典型结构,随机生成相关的网络参数。在学习训练过程中,根据网络的输入和训练状态,有针对性地增加新结点,逐渐形成固定的优化结构,即试探性地逐渐增加其他因素指标进

收稿日期:2013-05-26。

项目来源:国家自然科学基金资助项目(41271455/D0108);数字制图与国土信息应用工程国家测绘地理信息局重点实验室开放研究基金资助项目(GCWD201101);地理空间信息工程国家测绘地理信息局重点实验室经费资助项目(201115);中央高校基本科研业务费专项资金资助项目(CUGL120207)。

入训练和分析,从而避免训练数据过多使得网络陷入过度适应的危险。

2 研究区概况及数据源

研究区位于湖北省西部长江三峡境内,地理坐标为东经 $110^{\circ}00' \sim 110^{\circ}18'$,北纬 $30^{\circ}47' \sim 31^{\circ}05'$,总长度 77 km,区域总面积 732 km^2 ,其中,长江水面面积为 38.28 km^2 (135 m 水位线)。图 1 为研究区地理位置示意图,该区属于奉节以东巫山山脉中秭归——巴东段三峡峡谷,相对高程达 $800 \sim 2\,000 \text{ m}$ 。研究区地质条件复杂,且处于暴雨频繁的亚热带气候区,受新构造运动的影响,山原期夷平面快速抬升,河流强烈下切,形成高陡岸坡,卸荷效应显著,滑坡地质灾害广泛发育且频繁发生。

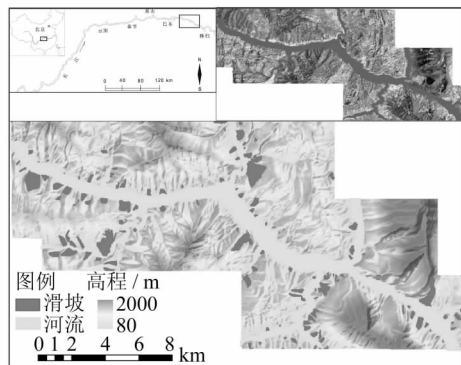


图 1 研究区地理位置示意图

Fig. 1 Location of the Study Area

本文采用的主要数据源包括:① Landsat ETM+ 卫星影像 1 景(2000-05-14,轨道号为 125/39)和我国环境减灾卫星影像 HJ-1A 1 景(2011-04-13,轨道号为 7/76),用于提取土地利用、植被指数等地表覆盖信息;② 1:10 000 地形图,用于

提取地形地貌和水文特征等信息;③ 1:50 000 和 1:200 000 地质图,用于提取断层、斜坡结构和工程岩组等地质信息;④ 其他历史滑坡存档资料和野外调查资料及部分航片等,用于历史滑坡解译和空间定位。研究区内共解译出滑坡 138 处,总面积 13.7 km^2 ,其中,最大滑坡面积为 1.1 km^2 ,最小滑坡面积为 $5\,157 \text{ m}^2$ 。

3 研究方法

3.1 多源空间数据预处理和融合

多源空间数据预处理和融合是提取滑坡易发性评价因子的基础,多源数据主要包括地形地貌和地质数据、遥感数据和野外调查资料。多源数据的类型、结构、坐标系统、投影方式和比例尺(或分辨率)各不相同。因此,数据融合是关键,其主要步骤包括:① 建立统一的坐标系、地图投影,对数据进行校正和配准;② 统一采用栅格单元作为滑坡评价单元,将所有矢量数据转换为栅格格式;③ 通过栅格重采样,将所有栅格单元大小统一为 $28.5 \text{ m} \times 28.5 \text{ m}$;④ 针对提取不同信息的需要,科学合理地选择相关数据,进行滑坡信息提取和分析。

3.2 评价因子选取

根据研究区的具体特点和前人研究成果,基于多源滑坡空间数据,采用三维地形分析、空间分析、水文分析、遥感图像分类、遥感图像信息提取等方法,提取地形地貌、地质、水文和地表覆被类 10 个滑坡易发性评价因子,作为滑坡空间预测模型的输入变量,通过统计分析研究区内历史滑坡与各评价因子之间的相关性,将连续因子离散化为分类变量,评价因子体系见表 1。

表 1 滑坡评价因子体系

Tab. 1 Landslide Assessment Factors

评价因子		划分指标
地形地貌类	高程/m	<180, 180~380, 380~580, 580~780, >780
	坡度	<15°, 15°~25°, 25°~35°, 35°~45°, >45°
	剖面曲率	<-1.5, -1.5~-0.6, -0.6~0.2, 0.2~1, >1
地质类	工程岩组	软岩、软硬相间岩、硬岩
	斜坡结构	顺倾坡、伏倾坡、顺斜坡、横向坡、逆斜坡、逆向坡
	断层平均距离/m	<1 280, 1 280~2 713, 2 713~4 300, 4 300~6 000, >6 000
水文类	TWI	<8, 8~9, 9~10, 10~12, >12
	水系平均距离/m	<600, 600~1 200, 1 200~2 000, 2 000~3 000, >3 000
地表覆被类	土地利用	建筑用地、耕地、林地
	NDVI	<103, 103~137, 137~170, 170~209, >209

为保证各评价因子的相对独立性,采用 Pearson 积矩相关系数统计分析因子之间的相关程

度,统计研究区上述 10 个评价因子的 Pearson 相关系数(见表 2),相关系数的绝对值均小于 0.75。

表 2 各因子 Pearson 相关系数

Tab. 2 Pearson Correlation Coefficient of Factors

相关性	高程	坡度	剖面曲率	工程岩组	斜坡结构	断层	TWI	水系	土地利用	NDVI
坡度	0.197	1.000								
剖面曲率	0.077	0.004	1.000							
工程岩组	0.089	0.016	-0.014	1.000						
斜坡结构	0.126	0.014	-0.006	0.121	1.000					
断层	-0.010	0.052	0.013	0.011	0.010	1.000				
TWI	-0.219	-0.293	0.161	-0.019	-0.031	-0.046	1.000			
水系	0.708	0.048	-0.005	0.054	0.083	0.091	-0.085	1.000		
土地利用	0.195	0.168	-0.007	0.015	0.009	0.036	-0.079	0.141	1.000	
NDVI	0.430	0.300	-0.004	0.025	0.055	-0.010	-0.133	0.291	0.316	1.000

3.3 滑坡空间智能预测

3.3.1 评价因子归一化处理

滑坡评价因子数据类型、取值范围和量纲的不一致,会加大激活函数波动性,降低网络训练速度。因此,采用极差标准化方法对连续型评价因子进行归一化处理。针对离散型因子,根据已知滑坡栅格单元中各因子属性所占比例确定各属性值,再采用极差标准化方法对各属性值进行归一化处理。斜坡结构、工程岩组和土地利用的属性归一化结果见表 3。

表 3 离散因子属性归一化处理

Tab. 3 Normalization of Discrete Factors

类型	滑坡格网	量化	归一化
斜坡结构	飘倾坡	324	0.020
	伏倾坡	2 730	0.168
	顺斜坡	2 806	0.173
	横向坡	6 248	0.383
	逆斜坡	2 507	0.154
	逆向坡	1 664	0.102
工程岩组	软岩	8 323	0.507
	软硬相间岩	7 492	0.458
	硬岩	575	0.035
土地利用	建筑用地	1 062	0.067
	耕地	5 834	0.369
	林地	8 911	0.564

3.3.2 格网模型单元划分

格网单元是栅格 GIS 中常用的基本单元,采用矩阵的形式组织和管理数据,计算机实现起来十分简单、高效。此外,数据被规则划分,便于数据重采样。所以,结合本文研究区地形地质资料的比例尺和卫星遥感影像的空间分辨率,对所有评价因子重采样为 28.5 m×28.5 m,研究区共划分为 275 758 个格网模型单元,其中,历史滑坡单元 16 480 个,稳定区单元 259 278 个。

3.3.3 模型决策表构建

10 个滑坡评价因子作为条件属性,滑坡作为决策属性(1 代表滑坡单元,0 代表非滑坡单元),构成 275 758 行 11 列的二维表格。为反映总体样本特征,保证样本中滑坡与非滑坡比例和总样本集中滑坡与非滑坡比例一致,随机选取 70% 的历史滑坡单元和稳定区单元作为训练样本,其余 30% 作为测试样本。

3.3.4 滑坡空间智能预测

BP 神经网络隐含层节点数目的确定是一个相对复杂的问题,隐含层单元的输入和输出是单调上升的非线性函数,节点数目太多,网络学习时间变长,节点数目太少,降低网络容错性。因此,采用实验测试方法确定最佳隐含层节点数,根据预测精度择优选取。

滑坡稳定性分析的分类误差是一个成本问题,即“危险区”与“稳定区”的错分不等价。因此,定义 A、B 错误,A 错误是指将危险区划分为稳定区,B 错误是指将稳定区划分为危险区。A 错误率=A 错误数/滑坡总数,B 错误率=B 错误数/非滑坡总数。对三层神经网络结构进行测试,结果见表 4。

实验结果表明,随着网络复杂度的增加,测试样本的总错误率保持在 13% 左右。当网络结构为 10×60×1,采用快速训练法,训练函数为变学习动量梯度下降算法,动量参数为 0.9,初始学习率为 0.3,学习率衰减周期为 30,最高学习率为 0.1,最低学习率为 0.01,交叉验证为假,训练周期为 200 时,训练样本和测试样本的 A 错误率达到了最小值,满足总错误率较低且 A 错误率较小的隐层节点选择原则。因此,该网络结构的神经网络模型是文本研究区滑坡易发性预测的最佳选择。

表4 训练样本和测试样本的错误率

Tab. 4 Error Rate of Training Sample and Test Sample

隐层节点	训练样本/%		测试样本/%	
	A 错误	总错误	A 错误	总错误
10	16.50	13.25	16.58	13.56
15	13.65	12.08	15.74	13.50
20	13.59	11.41	14.68	12.37
25	13.55	11.75	15.33	13.00
30	13.42	11.00	15.05	12.29
35	12.23	10.55	13.54	11.80
40	11.38	9.72	13.25	11.23
45	10.84	9.49	12.88	11.17
50	11.61	9.97	14.81	12.02
55	11.23	9.69	13.56	11.47
60	10.47	9.14	12.47	10.94
65	12.43	9.96	15.65	12.00
70	12.37	11.23	13.37	12.06

利用训练好的模型对研究区 275 758 个格网单元进行分类,得到滑坡易发性指数,值越大表示滑坡发生概率越大。为了便于区分滑坡的易发程度,根据直方图的分布特点,采用自然断点方法将滑坡易发性指数划分为 0~0.1、0.1~0.3、0.3~0.5、0.5~0.75 和 0.75~1,分别为稳定区、低易发区、中易发区、高易发区和危险区,结果见图 2。

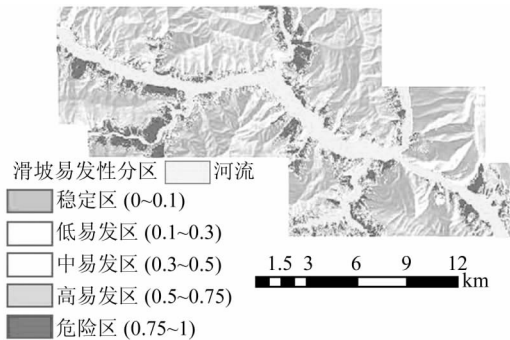


图2 滑坡易发性分区图

Fig. 2 Landslide Susceptibility Zoning Map

从预测结果的总体空间分布形态来看,滑坡灾害呈线状分布,主要分布于库水强影响区和波动区,即长江干流及支流沿岸,高易发区主要分布在青干河左岸、归州河右岸、童庄河右岸、巴东段附近长江右岸和靠近秭归段长江左岸,其余地方为零星点状分布。可见,研究区将进入一个库岸再造、斜坡失稳等滑坡灾害相对多发的不稳定时期。

以滑坡易发性分区为依据,分析每个输入变量对输出结果的敏感程度,得到滑坡易发性评价因子对滑坡灾害易发性影响的重要程度,各评价因子的权重比较见图 3,坡度、高程、TWI、工程岩组、水系距离和断层距离对滑坡的控制和影响较大。滑坡主要分布于 15°~30°的缓倾坡,该坡度

范围的岸坡前缘地形较缓而后缘地形较陡,如果坡体两侧分布有冲沟或走向与坡向相近的断层条件,则容易形成稳定性较低的六面滑坡体。侏罗系、三叠系中统巴东组和志留系等软岩岩组滑坡易发,其中三叠系巴东组地层是调查中库区滑坡发育最多的地层,岩性为灰岩、白云岩、泥灰岩、砂页岩互层等。土地利用、斜坡结构、剖面曲率和 NDVI 对滑坡的影响相对较小。

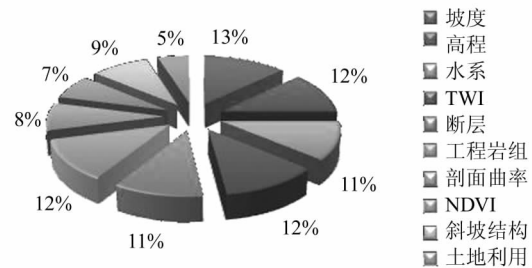


图3 评价因子权重比较

Fig. 3 Weight Comparison of Predictor Factors

4 结果分析与验证

成功率曲线能定量衡量模型预测的准确度,取值范围为[0.5,1],曲线越陡,则模型预测精度越高,曲线下面积为 1 表示一种理想状态,即模型预测结果与滑坡实际空间分布完全一致。隐层节点为 60 的三层 BP 神经网络模型的成功率曲线见图 4,曲线下面积为 0.897 5,即预测的准确率为 89.75%。对于 87%~100% (横坐标 13%处) 的预测指数,对应预测分区中的高易发区和危险区,能预测 51%的滑坡发生在高易发区和危险区;对于 60%~100% (横坐标 40%处) 的预测指数,对应预测分区中的危险区,能预测 75%的滑坡发生在高易发区和危险区。

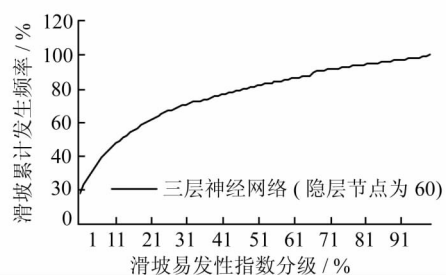


图4 模型成功率曲线

Fig. 4 Success Rate Curve

选择研究区内两个典型检验滑坡对预测结果进行验证(见图 5)。秭归县沙镇溪镇干将坪村,地处长江支流——青干河左(北)岸,预测结果为

危险区,根据实际滑坡调查资料,结合历史滑坡数据,该区为滑坡高发区。现场踏勘发现千将坪特大滑坡发育于该区。此外,位于秭归县龙江区境内长江北岸的区域,预测结果为危险区,根据实际野外调查资料和历史滑坡数据,该区有滑坡灾害发育,根据现场踏勘,该滑坡为新滩滑坡,发育于与两侧古生界岩层走向斜交的堆积层切向斜坡上,历史上发生过多次崩滑变形,属于继发性复活型滑坡。

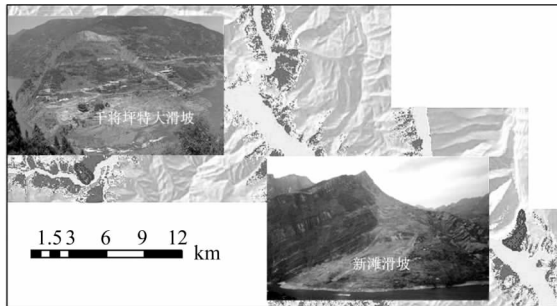


图 5 典型滑坡野外验证图

Fig. 5 Field Validation of Typical Landslides

5 结 语

本文针对三峡库区滑坡灾害受时空多源素影响的特征,提取了滑坡易发性评价因子,利用动态构建方法建立 BP 神经网络模型,定量预测滑坡稳定性,生成滑坡稳定性分区图,并采用成功率曲线和野外调查数据分析、评价预测结果。研究表明,本文提出的方法适合于库区滑坡预测,可为滑坡灾害预警与评估提供重要依据,但目前滑坡监测预警对遥感技术的重视不足,遥感与其他技术手段的协同研究工作缺乏,遥感数据与非遥感数据融合进行滑坡监测预警的研究程度较浅。因此,顾及滑坡系统行为的复杂性和非线性特征,发展多源数据支持下的多维滑坡信息提取模型和利用多种优势互补的智能算法提高滑坡空间预测的准确度是后续工作的方向。

参 考 文 献

- [1] 王尚庆,徐进军,罗勉. 三峡库区白水河滑坡险情预警方法研究[J]. 武汉大学学报·信息科学版, 2009, 34(10): 1 218-1 221
- [2] 郑守仁. 三峡工程试验性蓄水 175 m 水位运行的相关问题[J]. 人民长江, 2010, 41(8): 1-4
- [3] 陈德基,满作武. 三峡工程几个重大地质问题的研究与论证[J]. 中国工程科学, 2011, 13(7): 43-50

- [4] Cojean R, Ca Y J. Analysis and Modeling of Slope Stability in the Three Gorges Dam Reservoir (China): The Case of Huangtupo Landslide[J]. J Mt Sci, 2011(2): 166-175
- [5] 牛瑞卿,韩舸. 利用数据挖掘的滑坡监测数据处理流程[J]. 武汉大学学报·信息科学版, 2012, 37(7): 869-872
- [6] Pradhan B, Chaudhari A, Adinarayana J, et al. Soil Erosion Assessment and Its Correlation with Landslide Events Using Remote Sensing Data and GIS: A Case Study at Penang Island, Malaysia[J]. Environ Monit Assess, 2012, 184: 715-727
- [7] Nandi A, Shakoor A. A GIS-based Landslide Susceptibility Evaluation Using Bivariate and Multivariate Statistical Analyses[J]. Eng Geol, 2010, 110(2): 11-20
- [8] Kawabata D, Bandibas J. Landslide Susceptibility Mapping Using Geological Data, a DEM from ASTER Images and an Artificial Neural Network (ANN)[J]. Geomorphology, 2009, 113: 97-109
- [9] Carrara A. GIS Technology and Models for Assessing Landslide Hazard and Risk[J]. Geomorphology, 2008, 94: 257-260
- [10] Liu J P, Zeng Z P, Liu H Q, et al. A Rough Set Approach to Analyze Factors Affecting Landslide Incidence[J]. Comput Geosci, 2011, 37: 1 311-1 317
- [11] Yilmaz I. Comparison of Landslide Susceptibility Mapping Methodologies for Koyulhisar, Turkey: Conditional Probability, Logistic Regression, Artificial Neural Networks, and Support Vector Machine[J]. Environ Earth Sci, 2010, 61: 821-836
- [12] Gorsevski P V, Piotr J. Discerning Landslide Susceptibility Using Rough Sets[J]. Computers, Environment and Urb, 2008, 32: 53-65
- [13] Ilanloo M. A Comparative Study of Fuzzy Logic Approach for Landslide Susceptibility Mapping Using GIS: An Experience of Karaj Dam Basin in Iran [J]. Procedia Social and Behavioral Sciences, 2011, 19: 668-676
- [14] Kanungo D P, Sarkar S, Sharma S. Combining Neural Network with Fuzzy, Certainty Factor and Likelihood Ratio Concepts for Spatial Prediction of Landslides[J]. Nat Hazards, 2011, 59: 1 491-1 512

第一作者简介:武雪玲,博士,副教授,现主要从事滑坡灾害预测研究。

E-mail:snowforesting@163.com

Spatial Intelligent Prediction of Landslide Hazard Based on Multi-source Data in Three Gorges Reservoir Area

WU Xueling¹ REN Fu² NIU Ruiqing¹

(1 Institute of Geophysics and Geomatics, China University of Geosciences, 388 Lumo Road, Wuhan 430074, China)

(2 School of Resources and Environmental Sciences, Wuhan University, 129 Luoyu Road, 430079 Wuhan, China)

Abstract: Landslide hazard is influenced by many temporal and spatial factors. Traditional spatial analytical techniques cannot easily discover new and unexpected patterns, trends, and relationships that can be hidden deep within very large diverse geographic datasets. Focusing on Three Gorges Reservoir Area, environmental and triggering factors for landslide occurrences were extracted from multi-source data. Then, quantitative landslide susceptibility indices were calculated using the trained three-layered BP neural network, and the landslide susceptibility maps were generated. Finally, success rate curve was used to verify the results of landslide susceptibility mapping, and the results showed the best accuracy of 89.75%. The validation showed sufficient agreement between the prediction results and existing landslide. Therefore, the proposed model is an efficient method for landslide intelligent prediction, and can provide a significant reference for landslide hazard prediction and assessment.

Key words: multi-source data; Three Gorges Reservoir Area; landslide; intelligent; spatial prediction

About the first author: WU Xueling, Ph. D., associate professor, research interest is in landslide hazard prediction.

E-mail: snowforeting@163.com