

doi: 10.12068/j.issn.1005-3026.2020.05.001

基于深度信念网络的滑坡敏感性评价

王卫东<sup>1,2</sup>, 何卓磊<sup>1</sup>, 韩 征<sup>1</sup>, 钱 于<sup>3</sup>

(1. 中南大学 土木工程学院, 湖南 长沙 410075; 2. 中南大学 重载铁路工程结构教育部重点实验室, 湖南 长沙 410075;  
3. 南卡罗来纳大学 土木与环境工程系, 哥伦比亚 29208)

**摘 要:** 滑坡敏感性评价中各致灾因子之间存在复杂非线性关系, 传统的评价模型难以揭示该类复杂关系, 以致评价结果精度受限. 基于文献调查与实地调研, 选取高程、地貌类型、岩性、坡度、与构造线距离、与水系距离和年均降雨量为主要致灾因素, 在地理信息系统(GIS)中建立了基于深度信念网络(DBN)模型的区域滑坡敏感性区划模型, 并以四川区域为例进行了实例分析. 最后通过 ROC 曲线特征将评价结果与逻辑回归(LR)和人工神经网络(BPNN)模型评价结果进行了对比分析, 并探讨了各评价模型对不同致灾因子的响应. 研究表明 DBN 模型具有更高精度以及较低的假阳性率和假阴性率, 更适用于大区域、复杂致灾因子的区划滑坡敏感性评价工作.

**关 键 词:** 地理信息系统; 滑坡敏感性评价; 深度学习; 深度信念网络; ROC 曲线  
**中图分类号:** X 43      **文献标志码:** A      **开放科学(资源服务)标识码(OSID):**  
**文章编号:** 1005-3026(2020)05-0609-07



Landslides Susceptibility Assessment Based on Deep Belief Network

WANG Wei-dong<sup>1,2</sup>, HE Zhuo-lei<sup>1</sup>, HAN Zheng<sup>1</sup>, QIAN Yu<sup>3</sup>  
(1. School of Civil Engineering, Central South University, Changsha 410075, China; 2. The Key Laboratory of Engineering Structures of Heavy Haul Railway, Ministry of Education, Central South University, Changsha 410075, China; 3. Department of Civil and Environmental Engineering, The University of South Carolina, Columbia 29208, America. Corresponding author: HAN Zheng, assistant professor, E-mail: zheng\_han@csu.edu.cn)

**Abstract:** Complex non-linear relationships exist among causing factors in landslides susceptibility assessment. Traditional assessment models is difficult to reveal such complex relationships, and limit the accuracy of assessment results. Based on literature review and field survey, the altitude, landform, lithology, slope, distance to tectonic line, distance to drainage network and annual average rainfall were chosen as the main causing factors. A regional landslides susceptibility mapping model based on deep belief network (DBN) model in geographic information system (GIS) was established, and Sichuan Province was taken as an example. Finally, through the characteristics of ROC curves, the assessment results were compared with logistic regression (LR) and BP neural network (BPNN), and the response of each model to different causing factors was discussed. The results show that the DBN model has high accuracy, low false positive rate and false negative rate, and is suitable for landslides susceptibility assessment in large area with complex causing factors.

**Key words:** geographic information system (GIS); landslides susceptibility assessment; deep learning; deep belief network(DBN); ROC curve

滑坡敏感性是指在综合分析相关地质和地形条件的情况下, 某一区域发生滑坡的可能性. 在确定了致灾因子之后, 评估模型是编制滑坡敏感性区划图的一个关键问题. 过去几十年的研究提供

了多种可选择模型,一般可分为两类,即定性模型和定量模型.定性方法是基于专家经验的知识驱动型方法,如层次分析法<sup>[1-2]</sup>等,具有不确定性和主观性.随着计算机技术、遥感技术以及地理信息系统技术的全面发展,研究人员更倾向于采用定量方法来进行地质灾害区划.定量方法是基于统计学的数据驱动型方法,如逻辑回归(LR)分析法<sup>[3]</sup>等.在机器学习逐渐发展起来后,一些学者开始使用支持向量机法<sup>[4]</sup>等方法来进行滑坡敏感性分析.定量方法具有较好的确定性和客观性.

Polykretis 等<sup>[5]</sup>的研究认识到由于致灾因子之间的复杂非线性关系、缺乏相关数据以及环境动态变化的整合,导致编制滑坡敏感性区划图存在困难.然而迄今为止,大多数的分类和回归学习方法都是浅层结构的算法,不同特征之间的非线性关系没有得到充分反映.深度学习的本质是通过构建具有许多隐藏层和大量训练数据的模型来学习更多特性,实现对上层的特征抽象,一定程度上对传统模型所暴露出的目标函数优化和收敛不足等问题进行了提升与改进.因此,这种方法提高了分类或预测的准确性.本研究首次将深度信念网络(DBN)模型用于滑坡地质灾害评估,并将结果与 LR 和 BPNN 模型进行比较分析.

# 1 滑坡及致灾因子数据集

## 1.1 研究区域

四川省具有明显的西高东低、高低悬殊的地形特点.西部是川西高原,主要为高原和山地;东部是四川盆地,主要为平原和丘陵.四川处在中国 1 级和 2 级阶梯交界处和地震分布带中,省内存

在岷江断裂带、龙门山断裂带等一些大规模的断裂带,构造运动活跃,对滑坡的形成提供了充足的地质条件.由于地貌和地质背景复杂,水文条件丰富,该地区每年都会发生大量滑坡灾害.

## 1.2 历史滑坡数据集

在编制滑坡敏感性区划图的过程中,通过模型对历史滑坡分布数据进行分析可以获取每个因子之间的内在联系,从而最终预测每个点的滑坡发生概率.本文将调研得到的四川省滑坡分布图在 GIS 环境下数字化,制成四川省滑坡历史灾害分布图(图 1).从图 1 可以看出,四川省滑坡历史灾点主要分布在南部、东北部和中部.

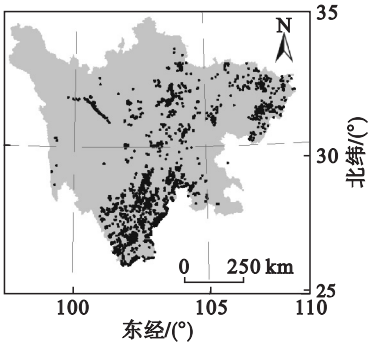


图 1 历史滑坡数据  
Fig. 1 Historical landslides data

## 1.3 致灾因子数据集

滑坡灾害是多因素相互作用的结果.评价指标的选取是否合理对滑坡评价过程和结果有着重要的影响.根据以往相关研究经验<sup>[6-7]</sup>并结合四川省的自然条件情况,本文选取了高程、地貌、岩性、坡度、与构造带距离、与水系距离、年均降雨量等 7 个致灾因子(图 2),每个因子具体的来源和格式见表 1.

表 1 数据来源与格式  
Table 1 Data source and format

致灾因子	来源	格式
高程	中国科学院地理空间数据云 DEM	栅格数据
地貌、与构造带距离、与水系距离	国家基础地理信息系统(NFGIS - M50DB)	ArcGIS 矢量图层
岩性	四川省地质图	ArcGIS 矢量图层
坡度	高程 DEM 处理结果	栅格数据
年均降雨量	中国气象局气象数据中心	文本数据

# 2 方法模型

## 2.1 深度信念网络(DBN)

深度学习的概念来源于对人工神经网络的研究.它将低级特性结合起来形成更抽象的高级表

示(属性类或特性),以发现数据的分布式特性.DBN 是一种高效的无监督学习算法.迄今为止,它已广泛应用于图像分类、语音识别和其他领域<sup>[8-9]</sup>.

DBN 在编制滑坡敏感性区划图中的一个显著优势是,它从可能对滑坡的发生至关重要的因

素中学习特征. 本文的 7 个致灾因子是研究者经常使用的,但这些因素之间的复杂相互作用和内在联系难以解释和量化. 传统的网络训练方法如反向传播 BPNN 和径向基函数 (RBF) 对因素的

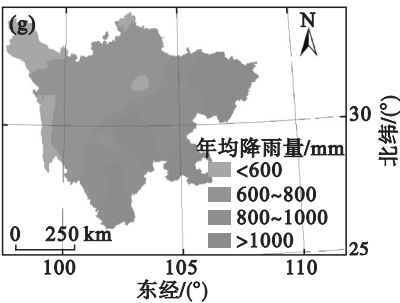
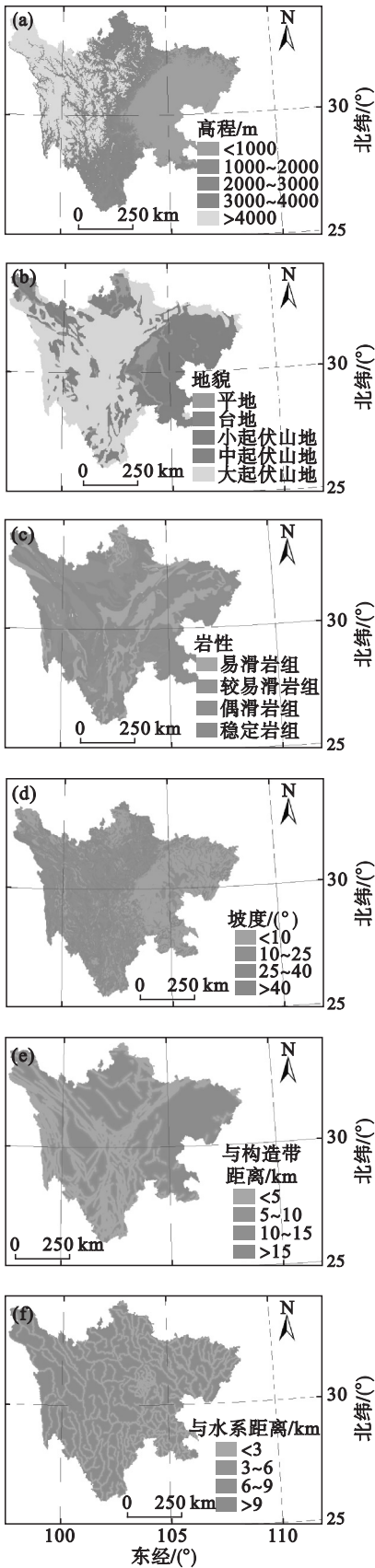


图 2 致灾因子图层  
Fig. 2 Layers of triggering factors

(a)—高程; (b)—地貌; (c)—岩性; (d)—坡度; (e)—与构造带距离; (f)—与水系距离; (g)—一年均降雨量.

原始特征依赖性很强,缺乏揭示因素组合效应的能力. 在这种情况下, DBN 被认为是解决这个问题的较好的方法. 鉴于此,本研究尝试将 DBN 应用于四川省的滑坡敏感性区划图.

DBN 是一个双向深度网络,是由受限玻尔兹曼机器 (RBM) 组成的概率生成模型. RBM 由输入数据层 (可视层  $v$ ) 和隐藏层  $h$  组成,各层神经元之间不存在连接 (图 3).

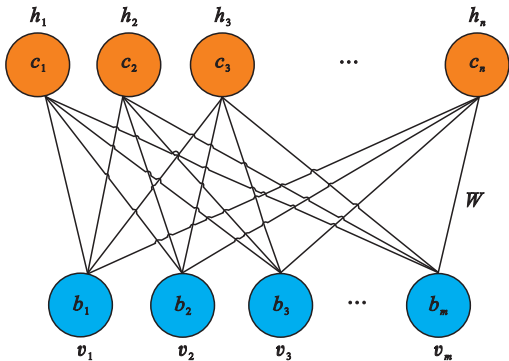


图 3 RBM 网络  
Fig. 3 RBM network

将若干个 RBM 连接起来则构成了一个 DBN,其中,上一个 RBM 的隐层即为下一个 RBM 的显层,上一个 RBM 的输出即为下一个 RBM 的输入. 训练过程中,需要充分训练上一层的 RBM 后才能训练当前层的 RBM,直至最后一层 (图 4).

## 2.2 对比模型

### 2.2.1 LR 分析

LR 模型是分类问题中常用的广义线性回归分析模型. 在编制滑坡敏感性区划图的过程中, LR 分析了致灾因子之间的线性关系,对每个点的滑坡发生率进行分类和预测. 各致灾因子的权重记录在表 2 中. 随后,历史滑坡数据集和所有致灾因子图层叠加在 GIS 平台下的 6 391 个栅格单元中.

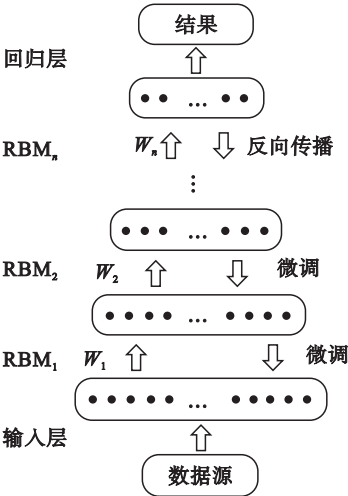


图 4 DBN 网络  
Fig. 4 DBN network

2. 2. 2 BPNN

反向传播神经网络 BPNN 是人工神经网络中应用最广泛、最具代表性的模型. 它是一种由误差反向传播算法训练的多层前馈网络. BPNN 具有大规模并行处理能力、自适应性、自学习能力、强容错性、鲁棒性和分布式信息存储等特点. 本研究采用 Matlab 环境下的人工神经网络工具箱进行建模分析.

3 模型预测结果

3. 1 DBN 模型

本研究选取典型的 4 层 DBN, 包含 3 个 RBM 和一个输出层来对已有数据进行训练. 整个

表 2 致灾因子分类及权重  
Table 2 Classification and weight of the triggering factors

致灾因子	二级因子	二级因子面积 $A_{ij}/\text{km}^2$	因子面积占比 $R_{ij}/\%$	历史滑坡面积 $a_{ij}/\text{km}^2$	滑坡面积占比 $r_{ij}/\%$	$T_{ij} = a_{ij}/R_{ij}$	规范化权重 $I_{ij}$
高程/m	<1 000	139 255. 30	28. 74	34. 37	22. 53	0. 783 9	0. 106 1
	1 000 ~ 2 000	52 426. 67	10. 82	56. 40	36. 96	3. 415 9	0. 462 4
	2 000 ~ 3 000	55 333. 87	11. 42	48. 94	32. 07	2. 808 2	0. 380 1
	3 000 ~ 4 000	106 161. 58	21. 91	11. 84	7. 76	0. 354 2	0. 048 0
	>4 000	131 357. 39	27. 11	1. 03	0. 68	0. 025 1	0. 003 4
地貌	平地	31 010. 23	6. 40	2. 28	1. 50	0. 234 4	0. 044 2
	台地	8 624. 72	1. 78	5. 19	3. 41	1. 915 7	0. 361 0
	小起伏山地	109 214. 15	22. 54	17. 58	11. 54	0. 512 0	0. 096 5
	中起伏山地	86 634. 82	17. 88	42. 53	27. 92	1. 561 5	0. 294 3
	大起伏山地	249 050. 89	51. 40	84. 74	55. 63	1. 082 3	0. 204 0
岩性	易滑岩组	98 118. 30	20. 25	39. 09	25. 66	1. 267 2	0. 338 2
	较易滑岩组	123 217. 20	25. 43	20. 38	13. 38	0. 526 2	0. 140 4
	偶滑岩组	173 851. 09	35. 88	78. 12	51. 29	1. 429 5	0. 381 4
	稳定岩组	89 348. 22	18. 44	14. 73	9. 67	0. 524 4	0. 140 0
坡度/(°)	<10	164 644. 93	33. 98	32. 25	21. 18	0. 623 3	0. 138 3
	10 ~ 25	181 312. 92	37. 42	35. 29	23. 17	0. 618 9	0. 447 7
	25 ~ 40	128 159. 46	26. 45	78. 33	51. 43	1. 944 6	0. 351 7
	>40	10 417. 50	2. 15	5. 65	3. 71	1. 725 0	0. 062 3
与构造带 距离/km	<5	188 521. 62	38. 91	92. 61	60. 80	1. 562 6	0. 413 5
	5 ~ 10	89 144. 13	18. 40	30. 03	19. 72	1. 071 7	0. 283 6
	10 ~ 15	52 257. 75	10. 78	13. 25	8. 70	0. 807 1	0. 213 5
	>15	154 611. 31	31. 91	16. 43	10. 78	0. 337 8	0. 089 4
与水系距 离/km	<3	136 031. 84	28. 08	69. 09	45. 36	1. 615 4	0. 399 4
	3 ~ 6	110 085. 30	22. 72	44. 48	29. 20	1. 285 2	0. 317 8
	6 ~ 9	85 969. 63	17. 74	20. 74	13. 62	0. 767 8	0. 189 9
	>9	152 448. 04	31. 46	18. 01	11. 82	0. 375 7	0. 092 9
年均降雨 量/mm	<600	51 554. 50	10. 64	3. 15	2. 07	0. 194 5	0. 051 7
	600 ~ 800	174 965. 52	36. 11	17. 29	11. 35	0. 314 3	0. 083 6
	800 ~ 1 000	128 062. 55	26. 43	61. 55	40. 41	1. 528 9	0. 406 7
	>1 000	129 952. 24	26. 82	70. 33	46. 17	1. 721 5	0. 457 9

训练过程分为两步:首先是对 RBM 进行预训练, 神经元数量、学习率和批次更新数量等参数需要在训练过程中根据研究需要不断调整;最终确定

RBM 的三层神经元数量分别为 100,50 和 20,训练方法为 SGD,学习率为 1,每批次更新数量为 100,采用适用于概率预测的性能函数 MSE 和激



活函数 Sigmoid. 在对前三层 RBM 训练之后,对整个网络进行训练,输入层节点选择 7 个,代表 7 个致灾因子,最后一层网络节点数为 1,表示单元滑坡敏感预测值;优化函数使用 Adam,损失函数采用 MSE,学习率为 0.01,迭代次数为 1 000 次,以 70% 的数据(4 474 个单元)为训练集,30% 的数据(1 917 个单元)为测试集,整个网络准确率为 83%.

利用训练好的 DBN,对四川省进行滑坡敏感性预测,预测范围为 0~0.990 5. 根据自然断点法将滑坡敏感性分为低 0~0.205 5,中 0.205 5~0.509 3,高 0.509 3~0.760 3 和极高 0.760 3~0.990 5 四个等级,得到四川省滑坡敏感性区划图(图 5a).

3.2 LR 模型

在 LR 分析中,单个历史滑坡被视为因变量. 存在则标记为 1,不存在标记为 0. 数据导出到 SPSS,得到代表各致灾因子之间线性关系的值:

$$z = -5.273 + 1.444I_{\text{高程}} + 2.285I_{\text{地貌}} + 3.705I_{\text{岩性}} + 0.867I_{\text{坡度}} + 2.068I_{\text{构造带}} + 3.151I_{\text{水系}} + 3.569I_{\text{降雨}},$$

(1)

$$P = \frac{1}{1 + e^{-z}}.$$

(2)

式中, $I_{\text{因子}}$ 代表各个致灾因子的权重.

根据式(1)和式(2),推导出四川地区  $P$  的分布范围为 0.019 5~0.799 1. 根据自然断点分类法,将滑坡敏感性区划图分为低(0.019 5~0.160 6),中(0.160 6~0.318 6),高(0.318 6~0.508 9)和极高(0.508 9~0.799 1). 预测结果如图 5b 所示.

3.3 BPNN 模型

选取典型的 3 层 BPNN 对输入数据进行训练. 隐含层和输出层传递函数分别采用 Tansig 和 Logsig,训练函数采用 Traingdx,阈值和权值的学习函数采用 Learngd,性能函数采用 MSE. 输入层节点数设为 7,即本文选取的滑坡一级致灾因子个数;输出层节点数设为 1,表示单元滑坡敏感性值.

利用 BPNN 得到四川省滑坡敏感性分布范围为 0~0.992 5,根据自然断点分类法分为低(0~0.348 7),中(0.348 7~0.536 7),高(0.536 7~0.726 7)和极高(0.726 7~0.992 5). 预测结果如图 5c 所示.

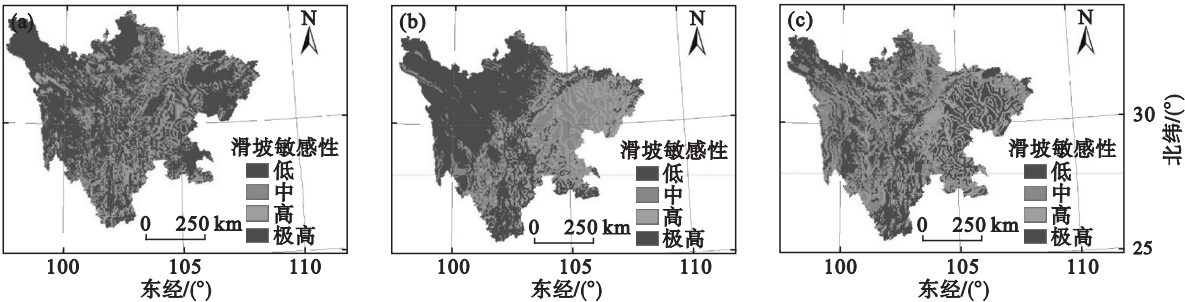


图 5 滑坡敏感性区划结果  
Fig. 5 Results of landslides susceptibility mapping  
(a)—DBN 模型; (b)—LR 模型; (c)—BPNN 模型.

4 结果分析

4.1 ROC 曲线精度评估

ROC 曲线分析是评估模型预测结果准确性的常用方法. 为了进一步讨论 DBN 的特点,利用 ROC 来评估 3 个模型编制滑坡敏感性区划图结果的准确性. 其中,曲线下面积(AUC)是模型精度的最终评价标准. DBN,LR 和 BPNN 的 ROC 曲线如图 6 所示,AUC 值分别为 0.899,0.782 和 0.867. 从这些数据来看,DBN 在 3 个模型中具有最好的预测精度,表明了深层神经网络处理复杂因素特征的优势.

4.2 模型结果对比

图 5 从几个方面揭示了结果的差异. 首先,3 个模型编制的滑坡敏感性区划结果具有相似的特征,即低、中敏感区主要集中在西部,而高、极高敏感区则分布在南部. 其次,3 个模型预测得到不同比例的 4 种敏感性水平. 显然,DBN 和 BPNN 之间的差异并不显著,而 DBN 和 LR 之间的差异主要集中在西北和东部地区. LR 与 DBN 及 BPNN 在四川省西北部和中部地区的区划差异明显地揭示了三者数据挖掘能力的不同,同为非线性处理模型的 DBN 和 BPNN 有更接近的结果,而两者在四川省西北部的区划结果相比于 LR 模型更符合实际情况.

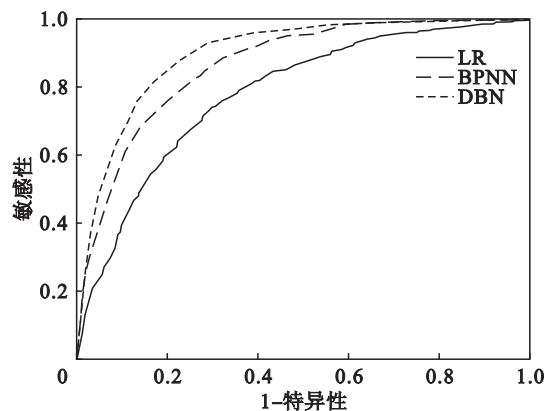


图 6 ROC 曲线分析  
Fig. 6 ROC curves analysis

图 7 和图 8 探讨了 3 个模型结果的差异,最明显的是 3 个模型中 DBN 对敏感性进行了保守的预测,“极高”和“高”敏感性覆盖的面积分别占总面积的 27.57% 和 18.57%。同时,52.72% 的历史滑坡被确定为“极高”敏感性等级,优于 BPNN 的 33.75% 和 LR 的 43.06%。另外,计算了不同

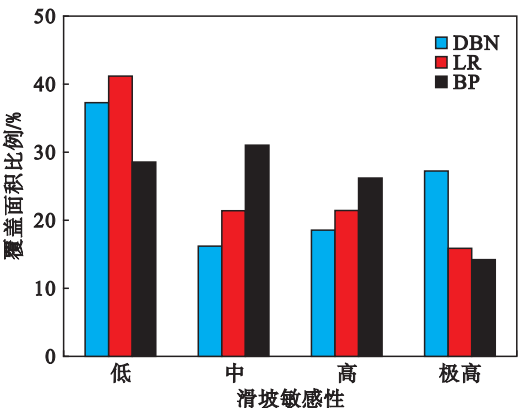


图 7 敏感性等级覆盖面积比例  
Fig. 7 Proportion of coverage area of each landslide susceptibility level

敏感性等级中的历史滑坡数,DBN 预测处于“极高”和“高”敏感性的历史滑坡占总量的 81.04%,为 3 个模型中最好。

与 BPNN 的 29.63% 相比,DBN 中总共 18.96% 的历史滑坡被错误地归类为“低”和“中”敏感性,更好地控制假阴性。

4.3 致灾因子的敏感性

为了探究不同模型中致灾因子的影响,进行了统计试验以探讨预测得到的滑坡敏感性与各级不同致灾因子面积的比例(图 9)。由该图可知,DBN 和 BPNN 都反映了平均年降雨量和与水系距离跟滑坡发生之间的正相关关系,因为预测的“高”和“极高”的敏感滑坡集中在年降雨量较大且与水系距离较近的区域,这与先前研究的结论一致<sup>[10-11]</sup>。如图 9 所示,高程因素也起着至关重要的作用。LR 的结果表明,海拔 1~2 km 的区域覆盖了大部分“极高”的敏感区域,而 DBN 更倾向于小于 1 km 的区域。

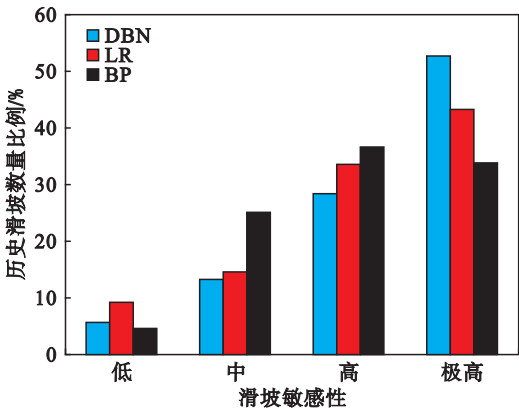
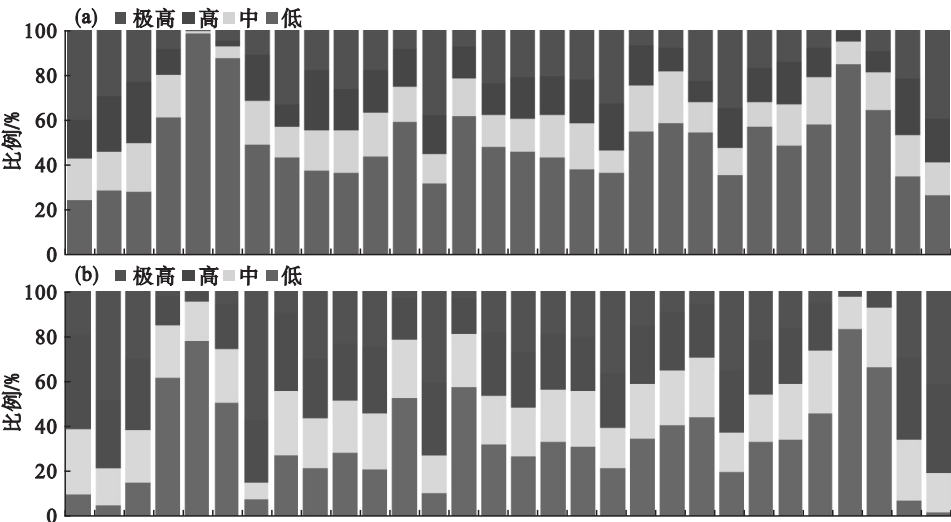


图 8 不同敏感性等级的历史滑坡数量比例  
Fig. 8 Percentage of historical landslides located in different susceptibility areas



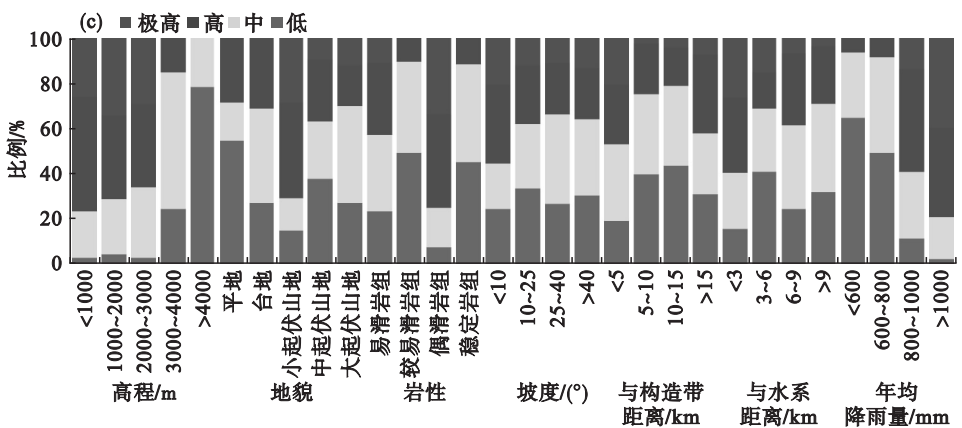


图9 不同滑坡敏感性在不同致灾因子面积中的比重  
Fig. 9 Proportion of landslide susceptibility to the areas of each level of different triggering factors  
(a)—DBN; (b)—LR; (c)—BPNN.

在 3 个模型中,坡度因素对滑坡发生的影响没有如此显著的统计意义,因为滑坡在该因素的不同水平上分布几乎均匀,这与侧重于个别滑坡机制的研究的预期和结论有些差异.这一现象的出现可以解释为四川省内的历史滑坡同时存在土质滑坡和岩质滑坡,而坡度对土质滑坡影响较大,对岩质滑坡影响较小.

总的来说,DBN 和 BPNN 的结果显示出非常相似的分布特征,在高程、与水系距离和年平均降雨量等因素上差异不大,而在剩下的 4 个因素上则存在较大差异.LR 的结果与 DBN 和 BPNN 有显著差异,特别是在地貌因素上,LR 预测大多数高敏感区域将分布在台地中.这一发现虽然是初步的,但仍然有助于理解每个模型对致灾因子的响应.

## 5 结 论

1) 基于 DBN 和 LR,BPNN 模型进行滑坡敏感性评价,结果表明四川西南部和东南部更容易发生滑坡.其中 DBN 模型区划结果显示整个四川省处于“极高”和“高”敏感性的比例为 46.14%,LR 模型为 37.44%,BPNN 模型为 40.35%.

2) DBN 具有较好的假阴性率控制性能,区划结果精度高于 LR 和 BPNN,验证了该方法的准确性.

3) DBN 与 LR,BPNN 模型对致灾因子的响应不同.利用深度学习 DBN 模型能综合反映致灾因子在外在条件下复杂的内在相互关系,尤其是在大范围、大数据量的区域中,可以成为以后地质灾害区划研究的重要手段.

## 参考文献:

[1] Avtar R, Singh C K, Singh G, et al. Landslide susceptibility zonation study using remote sensing and GIS technology in the Ken - Betwa River Link area, India [J]. *Bulletin of Engineering Geology and the Environment*, 2011, 70 (4): 595 - 606.

[2] Abedini M, Tulabi S. Assessing LNRF, FR, and AHP models in landslide susceptibility mapping index; a comparative study of Nojian Watershed in Lorestan Province, Iran [J]. *Environmental Earth Sciences*, 2018, 77 (11): 405 - 417.

[3] Aditian A, Kubota T, Shinohara Y. Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a Tertiary Region of Ambon, Indonesia [J]. *Geomorphology*, 2018, 318: 101 - 111.

[4] Ballabio C, Sterlacchini S. Support vector machines for landslide susceptibility mapping: the Staffora River Basin case study, Italy [J]. *Mathematical Geosciences*, 2012, 44 (1): 47 - 70.

[5] Polykretis C, Chalkias C. Comparison and evaluation of landslide susceptibility maps obtained from weight of evidence, logistic regression, and artificial neural network models [J]. *Natural Hazards*, 2018, 93 (3): 1 - 26.

[6] Kanungo D P, Sarkar S, Sharma S. Combining neural network with fuzzy, certainty factor and likelihood ratio concepts for spatial prediction of landslides [J]. *Natural Hazards*, 2011, 59 (3): 1491 - 1512.

[7] Chen W, Xie X S, Wang J L, et al. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility [J]. *Catena*, 2017, 151: 147 - 160.

[8] Djurdjevic P D, Huber M. Deep belief network for modeling hierarchical reinforcement learning policies [C]//Proceedings of 2013 International Conference on Systems, Man, and Cybernetics. Manchester: IEEE, 2013: 2485 - 2491.

[9] Kim H, Lee S, Lee D, et al. Real-time human pose estimation and gesture recognition from depth images using super pixels and SVM classifier [J]. *Sensors*, 2015, 15: 12410 - 12427.

[10] Cheng D Q, Cui Y F, Su F H, et al. The characteristics of the Mocoa compound disaster event, Colombia [J]. *Landslides*, 2018, 15: 1223 - 1232.

[11] Han Z, Wang W D, Li Y G, et al. An integrated method for rapid estimation of the valley incision by debris flows [J]. *Engineering Geology*, 2018, 232: 34 - 45.