

Construction and Verification of a Susceptibility Evaluation Model for Geological Hazards on Highways

Yishu Zhang

Shanxi Transportation Science Research Institute Group Co., Ltd., Taiyuan, Shanxi, 030006, China

Abstract

Geological hazards on highways have become an important factor restricting China's infrastructure construction, when evaluating and predicting their susceptibility, it is necessary to select appropriate indicators based on the actual engineering conditions and meet the basic requirements of universality, independence, and operability. In response to the complex and ever-changing geological environment of highways, it is necessary to evaluate the susceptibility of geological disasters in order to reduce the impact of frequent geological disasters on highway construction. Based on this, the paper constructs a geological hazard susceptibility evaluation model for highways, and combines engineering examples to randomly select 139 grid points, the ROC method is used to compare and verify the accuracy of geological hazard susceptibility prediction between the random forest model and the support vector machine model. The results indicate that both the random forest model and the support vector machine model can accurately reflect the susceptibility of geological disasters, but the accuracy of the random forest model is higher.

Keywords

highway; geological disasters; susceptibility evaluation

高速公路地质灾害易发性评价模型构建与验证

张一舒

山西交通科学研究院集团有限公司, 中国·山西 太原 030006

摘要

高速公路地质灾害已经成为制约中国基础工程建设的一项重要因素,而在对其进行易发性的评价与预测时,需要结合工程实际条件选择适当的指标,并满足基本的普适性、独立性及可操作性要求。针对高速公路地质环境复杂多变问题,为了减少地质灾害频发对公路建设的影响,有必要对地质灾害易发性进行评价。基于此,论文通过构建高速公路地质灾害易发性评价模型,结合工程实例,随机选取139个栅格点,采用ROC法,对比验证了随机森林模型和支持向量机模型两种模型地质灾害易发性预测精度。结果表明,随机森林模型和支持向量机模型均能精确反映地质灾害的易发性,但随机森林模型的精度更高。

关键词

高速公路; 地质灾害; 易发性评价

1 引言

在确定评价单元时,常用的栅格单元法、斜坡单元法均具有自身独特的应用范围,但前者的精度一般更高。此外,在建立评价模型的同时,也需要对模型的评价精度有充分的了解,并在计算后对模型进行检验,比对模型预测的可靠性。

2 地质灾害评价指标的选取原则及相关性

2.1 地质灾害评价指标的选取原则

地质灾害评价需要结合工程实际情况,充分考虑各项因素的作用,尤其应当关注水文地质、地形地势等的影响^[1]。

因此,根据工程实际情况完善科学、合理的指标体系就显得尤为重要,有助于提升评价的便捷性、准确性^[2]。总体来看,选择指标时应当符合以下基本原则。

①普适性。对于同一项目的不同研究区域而言,其地质、气候等条件不可避免地会存在偏差,因此就需要保证指标具有足够的普适性,能够最大程度地满足工程评价的要求。在实际工程中,一般需要技术人员通过实地野外调查、数据整理分析来具体确定指标,如有必要还应将人文风俗纳入考量范围,确定其中最佳的指标组合。

②独立性。实践表明,地质灾害的萌芽与发展并非仅受单一因素的影响,往往是多重因素复合作用导致的。因此,对已有的指标进行分析、选择时就需要尽可能谨慎,最大程度地降低指标体系中各指标之间的相关性,以此来为后续的量化评价提供足够的合理性支持。在选择时,一般可先处理

【作者简介】张一舒(1990-),女,中国山西大同人,本科,工程师,从事交通运输研究。

高相关性的指标，并剔除其中权重较小的指标，建立起适用于具体工程的指标体系。

③可操作。可操作性集中体现在指标是否可以在 ArcGIS 软件中得到具体应用，并被量化处理。对于评价指标而言，即便其具有足够的合理性、科学性，但如果指标数据的获取难度较大，则难以直接应用到评价工作中去，同时也会间接增加评价工作量。

2.2 地质灾害评价指标的相关性

如上文所述，在构建指标体系时应特别注意指标间的相关性。若不同指标的相关性较高，则在模型建立与分析过程中就很容易导致多重共线性，降低模型分析的精度^[3]。所以，在指标选择初期，就需要对指标间的相关性进行判定，从其中剔除多余指标，改进指标体系的合理性。论文选择 Pearson 相关系数来完成指标相关性分析，并借助相关性系数 R 来量化描述指标的相关程度，R 的取值位于 [-1,1] 的区间内，其评价标准如表 1 所示。

表 1 相关系数表

| | 取值范围 | 相关程度 |
|---|---------|------|
| R | <0.4 | 低度相关 |
| | 0.4—0.7 | 显著相关 |
| | 0.7—1.0 | 高度相关 |

|R| 越接近 1，则表示指标间的相关性越突出；反之，|R| 越接近 0，则标识两指标间的相关性越低。

3 地质灾害评价的方法

论文以中国某高速公路工程建设为例展开。

3.1 评价单元的划分

在使用 ArcGIS 软件对灾害的易发性进行量化评价之前，先需要技术人员合理划分评价单元。具体来看，评价单元的划分一般需要确定各单元的类型、大小及规格。

作为最为常见且效果良好的一类评价单元划分方法，栅格单元法以规则形状的网络为划分依据，将研究区段划分为若干个独立单元^[4]。在栅格单元法下，评价工作能够快速提取出所需的数据信息并对其进行叠加分析，但采用栅格单元法所划分的单元比较规则，因此难以准确反映地形的变化。

选择栅格单元法来划分评价单元，虽然难以准确、详细地描述地势起伏情况，但其评价结果的过渡较为平稳，因此相较斜坡单元法精度更好，应用该方法得到的评价结果也具有更高的参考价值。

在确定单元大小时，论文主要借鉴了已有研究与实践，基于研究区段大小及工程条件来分析，并未形成具体、严格的定量表达。同时，在研究中不断调整单元的大小，通过观察、比对其实际效果，得出最终的单元大小为 30m×30m，并以此将指标图层转换为栅格文件，总计划分了 556941 个栅格。借助 ArcGIS 软件内置的栅格转点功能，即可将各个栅格单元转换为单个的点数据，并在此基础上完成各图层的

处理，以此得到研究区段内全部的评价指标数据。

3.2 研究样本的分类

通过对研究区段内开展实地野外调研，共发现地质灾害点 139 处，为了建立评价模型，则还需要增加 139 处非地质灾害点进行比对。本研究在距灾害点 100m 范围外又随机选取了 139 个栅格点，并将所选取这些测点的数据信息作为非地质灾害点。从 139 处地质灾害点以及 139 处非地质灾害点构成的数据集中，各挑出 70% 的测点数据，作为评价模型的训练集样本。而剩下 30% 的测点数据则归入验证集样本中，来检验模型的泛化能力，评估其是否具有足够的可靠性。

3.3 评价结果

3.3.1 随机森林模型

选用随机森林模型来完成灾害易发性评价，并借助 MATLAB 软件通过有放回地随机抽取 (bootstrap) 从训练样本中随机抽得部分样本组成新的训练集。按以上方式重复 500 次即可得到 500 组新的训练集，基于每组的训练集建立起相应的决策树，并从 10 个特征变量中随机抽取 8 个作为决策树的节点。在按照上述要求完成模型建设后，将相应的参数导入算法中，同时借助训练集建立分类模型。随后即可运行模型对训练集和验证集循环分析，分别得到决策结果为 0 及 1 的决策树数量。

模型运行结果显示，训练集总共 196 个样本全部被预测正确，精度为 100%；验证集总共 82 个样本中有 16 个的预测结果与实际情况不符，其整体精度约为 80%。计算得到的地质灾害的易发性指数如表 2 所示。

采用 ArcGIS 软件的栅格转点功能即可将 dem 栅格文件转换为点文件，并在其中导入 MATLAB 预测所得的易发性指数数据，借助点转栅格功能即可将点文件转回为栅格文件，并形成研究区段的易发性指数分布图。随后，还可利用自然间断点法对易发性指数展开分类，根据其指数的大小将研究区段的评价结果划分为五个不同的定量评价，分别为极高、高、中、低以及极低。

3.3.2 支持向量机模型

若采用支持向量机模型来完成易发性评价，那么技术人员就需要特别慎重地选择核函数及其基本参数，以此来保障评价结果的准确性，在论文中选择径向基函数进行分析。在确定核函数之后，还应分别确定 c、g 的取值。在实际的评价工作中，一般可借助 K-折交叉验证法来确认 c、g 的取值，而在本研究中这一值的区间为 [-3.5, 3]。在将边界条件导入 MATLAB 进行交叉验证分析后，即可得出当 c、g 的取值分别为 1.93 及 -2.16 时，验证精度达到峰值，其最佳精度约为 72%，故将其作为 c、g 最终取值。

采用 ArcGIS 软件得到各评价因子数据后，将数据导入至 MATLAB 中已完成训练的支持向量机分类器中，分析研究区段内各栅格点的地质灾害易发性。计算得到的地质灾害的易发性指数如表 3 所示。

表 2 随机森林模型下得到的评价结果

| 栅格点 | 易发性指数 | 栅格点 | 易发性指数 | 栅格点 | 易发性指数 | 栅格点 | 易发性指数 | 栅格点 | 易发性指数 |
|-----|-------|-----|-------|-----|-------|-----|-------|-----|-------|
| 1 | 0.510 | 15 | 0.500 | 29 | 0.606 | 43 | 0.380 | 57 | 0.282 |
| 2 | 0.462 | 16 | 0.510 | 30 | 0.568 | 44 | 0.496 | 58 | 0.172 |
| 3 | 0.500 | 17 | 0.536 | 31 | 0.544 | 45 | 0.374 | 59 | 0.184 |
| 4 | 0.484 | 18 | 0.544 | 32 | 0.412 | 46 | 0.544 | 60 | 0.232 |
| 5 | 0.484 | 19 | 0.510 | 33 | 0.450 | 47 | 0.544 | 61 | 0.200 |
| 6 | 0.478 | 20 | 0.574 | 34 | 0.450 | 48 | 0.398 | 62 | 0.262 |
| 7 | 0.470 | 21 | 0.594 | 35 | 0.444 | 49 | 0.374 | 63 | 0.356 |
| 8 | 0.442 | 22 | 0.574 | 36 | 0.390 | 50 | 0.514 | 64 | 0.402 |
| 9 | 0.460 | 23 | 0.598 | 37 | 0.534 | 51 | 0.384 | 65 | 0.220 |
| 10 | 0.460 | 24 | 0.594 | 38 | 0.528 | 52 | 0.384 | 66 | 0.408 |
| 11 | 0.512 | 25 | 0.584 | 39 | 0.294 | 53 | 0.376 | 67 | 0.428 |
| 12 | 0.476 | 26 | 0.588 | 40 | 0.304 | 54 | 0.448 | 68 | 0.538 |
| 13 | 0.484 | 27 | 0.572 | 41 | 0.372 | 55 | 0.594 | 69 | 0.504 |
| 14 | 0.482 | 28 | 0.622 | 42 | 0.384 | 56 | 0.254 | 70 | 0.494 |

表 3 支持向量机模型下得到的评价结果

| 栅格点 | 易发性指数 | 栅格点 | 易发性指数 | 栅格点 | 易发性指数 | 栅格点 | 易发性指数 | 栅格点 | 易发性指数 |
|-----|--------|-----|--------|-----|--------|-----|--------|-----|--------|
| 1 | -0.653 | 15 | -0.534 | 29 | -0.581 | 43 | -0.759 | 57 | -0.801 |
| 2 | -0.797 | 16 | -0.534 | 30 | -0.557 | 44 | -0.725 | 58 | -0.814 |
| 3 | -0.694 | 17 | -0.804 | 31 | -0.406 | 45 | -0.771 | 59 | -0.826 |
| 4 | -0.597 | 18 | -0.444 | 32 | -0.450 | 46 | -0.715 | 60 | -0.789 |
| 5 | -0.496 | 19 | -0.607 | 33 | -0.419 | 47 | -0.714 | 61 | -0.785 |
| 6 | -0.532 | 20 | -0.596 | 34 | -0.396 | 48 | -0.732 | 62 | -0.765 |
| 7 | -0.523 | 21 | -0.600 | 35 | -0.357 | 49 | -0.742 | 63 | -0.749 |
| 8 | -0.537 | 22 | -0.553 | 36 | -0.344 | 50 | -0.717 | 64 | -0.691 |
| 9 | -0.531 | 23 | -0.562 | 37 | -0.685 | 51 | -0.746 | 65 | -0.714 |
| 10 | -0.584 | 24 | -0.612 | 38 | -0.552 | 52 | -0.749 | 66 | -0.640 |
| 11 | -0.505 | 25 | -0.582 | 39 | -0.803 | 53 | -0.777 | 67 | -0.614 |
| 12 | -0.500 | 26 | -0.598 | 40 | -0.803 | 54 | -0.801 | 68 | -0.508 |
| 13 | -0.529 | 27 | -0.672 | 41 | -0.767 | 55 | -0.624 | 69 | -0.450 |
| 14 | -0.533 | 28 | -0.605 | 42 | -0.750 | 56 | -0.806 | 70 | -0.435 |

4 两种模型的检验

为了检验以上两种模型的评价、预测精度，本研究采用 ROC 法展开分析。其中，曲线图形的 X 轴（特异性）表征非灾害点数被错误预测的概率，而 Y 轴（敏感性）则表征灾害点被成功预测的概率，曲线与 X 轴所围成的图形面积越大则代表模型的精度越高，一般也将这一面积称为 AUC 值，其介于 0~1 的范围内。采用 ROC 法对以上两种模型的 ROC 曲线进行绘制并研究发现，随机森林模型、支持向量机模型的精度值分别为 0.931、0.874，两者均能比较精确地描述地质灾害的易发性，但前者的精度相对更高。

5 结语

综上所述，由于高速公路建设环境复杂多变，在实际建设过程中会遇到诸多地质灾害问题，对高速公路地质灾害易发性进行评价，能够减少地质灾害对公路建设的影响，避

免发生较大的危害，具有一定的实际意义。笔者主要从地质灾害评价指标的选取原则及其相关性出发，结合工程实践分析了地质灾害易发性评价的基本方法与流程，并对构建的评价模型进行了检验。受地质环境和建设条件的影响，论文的研究成果仅限于本工程区域内的地质灾害易发性评价，其他类似工程可借鉴论文研究思路，需进一步进行研究。

参考文献

- [1] 许玉龙.甘肃省某高速公路工程地质灾害危险性评估及防治措施[J].西部资源,2021(4):59-61.
- [2] 魏敏.仁博高速公路新增崧岭互通立交项目地质灾害危险性评估研究[J].西部资源,2021(2):124-126.
- [3] 甘贤军,支墨墨.典型山区高速公路基础设施灾害应对方法[J].运输经理世界,2020(4):46-48.
- [4] 魏雨.西南山区高速公路沿线地质灾害易发性评价[D].成都:成都理工大学,2020.