

主成分分析法在地质灾害危险性综合评价中的应用

范文^{1,2}, 刘雪梅¹, 高德彬¹, 苏生瑞¹, 常中华¹

(1. 长安大学 地质工程与测绘工程学院 陕西 西安 710054 2. 西安交通大学 建力学院 陕西 西安 710049)

[摘要] 介绍了主成分分析法的分类原理和实施步骤以及多指标体系中指标的选取,针对地质灾害的危险性,建立了评价的多指标体系和分级特征,同时分析了该方法在地质灾害危险性综合评价中的应用方法与过程。并通过对某高速公路沿线地质灾害危险性进行分段综合评价,给出了建设工程可能诱发或加剧的地质灾害和建设工程本身可能遭受的地质灾害危害的程度。证明该方法在地质灾害危险性综合评价中是切实可行的。

[关键词] 主成分分析法 指标体系 地质灾害 综合评价

[中图分类号] P642 [文献标识码] A [文章编号] 1007-9955(2001)04-0053-05

[作者简介] 范文(1967-)男,副教授,博士研究生,主要从事地质工程与岩土工程的教学与科研工作。

综合评价从数学的眼光来看,就是建立一种从高维空间到低维空间的映射,这种映射能保持样本在高维空间的某种“结构”^[1]。综合评价的研究对象通常是自然、社会、经济等领域中的同类事物(横向)或同一事物在不同时期的表现(纵向)。综合评价的依据是指标,而指标按不同的标志可分为实物指标和价值指标、相对指标和绝对指标、单向指标和综合指标等^[2]。在评价问题时,由于指标体系的不足,造成评价结果的差异。根据这一不足,又发展了多指标综合评价方法,即把反映被评价事物的多个指标的信息综合起来,得到综合指标,以此来反映被评价事物的整体情况,并进行横向和纵向的比较。这样既有全面性,又有综合性。近年来,其他领域的相关知识不断渗入多指标综合评价,使得多指标综合评价方法不断丰富。其中主成分分析法是多元统计分析的一种。其实质是对原坐标系进行平移和旋转变换,使得新坐标的原点与数据群点的重心重合,新坐标系的第一轴与数据变异的最大方向对应,新坐标系的第二轴与第一轴标准轴正交,并且对应于数据变异的第二大方向……依次类推。使高维数据系统综合简化。简化原则为数据变异最大方向之原则、最小二乘原则、群点相似性改变最小的原则、对

原变量系统最佳综合表现力原则等^[3]。

1 主成分分析法原理和实施步骤

主成分分析的基本思想是一种把原来各个指标化为可数几个互不相关(或相互独立)的综合指标的一种方法,可以达到数据化简、揭示变量之间的关系和进行统计分类解释的目的。其基本理论叙述如下^[3~5,7,9~10]。

设实测了 n 个 p 维样本

$$\bar{x}_i = \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{pi} \end{bmatrix} \quad i = 1, 2, \dots, n$$

要求将这 n 个样本分类。用主成分分析法,首先计算样本平均值与样本协阵:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_{1i} \\ \frac{1}{n} \sum_{i=1}^n x_{2i} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{pi} \end{bmatrix} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}$$

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})' =$$
$$\begin{bmatrix} \frac{1}{n-1} \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{1i} - \bar{x}_1)' & \frac{1}{n-1} \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)' & \dots & \frac{1}{n-1} \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{pi} - \bar{x}_p)' \\ \frac{1}{n-1} \sum_{i=1}^n (x_{2i} - \bar{x}_2)(x_{1i} - \bar{x}_1)' & \frac{1}{n-1} \sum_{i=1}^n (x_{2i} - \bar{x}_2)(x_{2i} - \bar{x}_2)' & \dots & \frac{1}{n-1} \sum_{i=1}^n (x_{2i} - \bar{x}_2)(x_{pi} - \bar{x}_p)' \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n-1} \sum_{i=1}^n (x_{pi} - \bar{x}_p)(x_{1i} - \bar{x}_1)' & \frac{1}{n-1} \sum_{i=1}^n (x_{pi} - \bar{x}_p)(x_{2i} - \bar{x}_2)' & \dots & \frac{1}{n-1} \sum_{i=1}^n (x_{pi} - \bar{x}_p)(x_{pi} - \bar{x}_p)' \end{bmatrix}$$

再求方程的

$$Q = | \hat{\Sigma}_{pxp} - \lambda I_{pxp} |$$

的最大根 λ_1 与次大根 λ_2 ,即矩阵 $\hat{\Sigma}$ 的最大特征根与次大特征根 ,设为 λ_1 与 λ_2 。再解方程组

$$\begin{cases} (\hat{\Sigma}_{pxp} - \lambda_1 I_{pxp})y = 0 \\ (\hat{\Sigma}_{pxp} - \lambda_2 I_{pxp})y = 0 \end{cases}$$

将解得的 y 标准化 ,即 :

$$(\alpha)^{(1)} = \frac{y}{\sqrt{y'y}} = \begin{bmatrix} (\alpha)_1^{(1)} \\ (\alpha)_2^{(1)} \\ \vdots \\ (\alpha)_p^{(1)} \end{bmatrix}$$
$$(\alpha)^{(2)} = \frac{y}{\sqrt{y'y}} = \begin{bmatrix} (\alpha)_1^{(2)} \\ (\alpha)_2^{(2)} \\ \vdots \\ (\alpha)_p^{(2)} \end{bmatrix}$$

以上 $(\alpha)^{(1)}$ 与 $(\alpha)^{(2)}$ 即 $\hat{\Sigma}$ 的最大特征根与次大特征根所对应的特征向量。

再计算 n 个 p 维样品的第一、第二主分量 ,得

$$\begin{cases} Z_{1i} = \alpha^{(1)'} x_i = \sum_{k=1}^p \alpha_k^{(1)} x_{ki} \\ Z_{2i} = \alpha^{(2)'} x_i = \sum_{k=1}^p \alpha_k^{(2)} x_{ki} \end{cases} \quad i = 1, 2, \dots, n$$

将 n 个二维点 $(z_{1i}, z_{2i}) , i = 1, 2, \dots, n$,点在平面直角坐标系中。到此数学计算已完成 ,下一步结合专业知识进行目测 ,将直角坐标系中距离近者归为一类 ,共聚多少类视具体情况而定。如此分类的合理之处在于 :原来是 n 个 p 维空间的点 $x_i ; i = 1, 2, \dots, n$,经坐标轴的刚性旋转之后在新坐标系中坐标变为 z_i ,刚性旋转并不改变点与点之间的绝对距离。实质上只是改个记号而已 ,结构并未大改。但是此刚性旋转使得 z_i 中后几个指标差异不大 ,当然这些指标对分类也就不起作用了 ,故可舍去 ,从而压缩维数。多维

空间的点是难以作图的。在许多实际课题中 ,往往取前两个主分量 ,就已经提取了主要信息。当然有时仅取二维是不够的 ,取多少个主分量才合适 ,这要看 $\hat{\Sigma}$ 的各特征根的大小 ,如果 λ_3 也较大 ,可用简单的装置来标明三维空间中的位置 ,维数再高就难了。

2 地质灾害危险性评价多指标体系

我们所要做的工作就是将所要研究的区域 ,根据能反映物体性质的指标综合起来 ,通过定性分析与定量分析 ,把所要研究的区域分成几类 ,然后给出合适的工程地质解释。在处理地质问题时 ,首先要选好能反映实际地质问题属性的各种定性定量指标。指标是反映对象之间的一种关系的指标量 ,根据其把研究对象给以分类。一个度量指标是否合理 ,就是去度量对象有没有不一致或不符合实际的情况。指标之间具有相关性 ,一般来说 ,我们考察物体性质的指标有两大类 :基础指标和派生指标。派生指标往往是由研究的内容来定的 ,基础指标往往是事物固有的。为了评价的需要 ,需对多个指标加以综合。因为反映一个对象的情况 ,往往要从各个不同的角度予以比较 ,这样 ,对这个对象给出一个总的评价就需要把许多不同类型的指标综合。

定量分析与定性分析表面上是两个不同的内容 ,然而仔细了解一下 ,就知道这是无法分开的。没有主要方面的定性研究、定量分析就没有对象 ,对象之间的联系也无从考虑 ,更谈不上去描述、分析。没有定量分析的依据 ,定性分析的结论就显得贫乏 ,并且不易更加深入。很明显 ,在定量分析之前 ,必须把所关心的“性质”、“关系”等等的“质”给以量化 ,而量化绝不是一个简单的打分 ,量化必须符合所反映的“性质”、“关系”等特点。在处理多个样本数据时 ,首先遇到的问题是观测数据很多的总体 X 。总体 X 是一个 p 维随机向量 ,这样组成一个 $n \times p$ 维的矩

阵。一般 p 个指标即 p 个随机变量 x_1, x_2, \dots, x_p 之间的相关关系,使数据分析复杂化。另一方面, p 个变量(或指标)的大部分变差能够由它们的 K 个所谓“综合指标”来概括,那么包含在这 K 个“综合指标”中的信息与 p 个变量几乎一样多,可以用这 K 个“综合指标”代替原 p 个变量。

地质灾害的发生具有很大的不确定性,其潜在危险性受多种条件控制。因而必须全面考虑危险性本身及构成危险性各因素的相似性与差异性,进而选择一些相互联系的指标作为综合分析的依据。从总体上来说,地质条件、地形地貌条件、气候条件、水文条件、植被条件、地下水条件、断裂分布及活动性、有无地震、岩土体性质、人为活动等都是控制各种地质灾害活动的基本条件,见图 1。但是这些条件在不同类型地质灾害中的主次地位和所起作

用不尽相同,所以其评价指标也各异。历史地质灾害可以从其规模、频次及密度来反映^[8],见表 1。潜在地质灾害的形成与发展由表 2 的条件来控制^[8]。

3 地质灾害危险性区划综合评价步骤

地质灾害是指在地质动力作用下,地质自然环境恶化,造成人类生命财产损毁或人类赖以生存与发展的资源、环境发生严重破坏的现象或过程。危险性区划综合评价的基本途径是:将评价区域划分成若干单元,把控制地质灾害活动的各方面指标量化,根据这些指标对地质灾害的控制方式和作用程度,建立地质灾害危险性综合评价的数学模型。其危险性综合评价结构示意图如图 2。

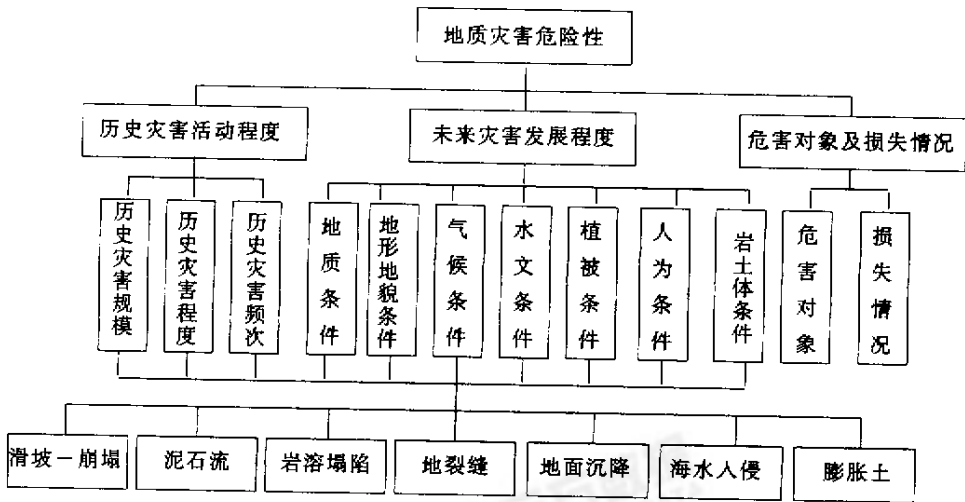


图 1 地质灾害评估系统结构及指标体系图

表 1 历史地质灾害种类及判别指标

项目	规 模		频 次		密 度	
灾害种类及判别指标	崩塌、滑坡(灾害体积)/m ³	泥石流(堆积物体积)/m ³	岩溶塌陷(影响面积)/m ²	崩塌、滑坡(次)	泥石流(次)	岩溶塌陷(次)

表 2 几种地质灾害的形成条件及潜在指标

灾害类型	地质	地形地貌条件	气候植被	人为活动
崩塌、滑坡	地质构造;新构造运动;岩石性质与岩土体结构	地貌类型;区域地形条件;相对高差;斜坡坡度;坡体形态	年降雨量及分布;暴雨洪水;森林覆盖率	矿山工程密度、人为爆破;植被破坏程度
泥石流	地质构造;新构造运动;松散堆积物数量	山地高原区;具备泥石流形成区、流通区和堆积区的局部地形地貌;相对高差;山坡坡度;主沟平均比降	年降雨量;暴雨比例;森林覆盖率	植被破坏;坡沟破坏
岩溶、塌陷	浅层可溶岩比例;岩溶发育程度;覆盖层厚度;断层构造;新构造运动;地下水活动	山间盆地、河谷盆地山麓倾斜平原、大型盆地等,呈裸露型、覆盖型和埋藏型	降雨量;植被覆盖率	开采地下水;矿井排水与突水

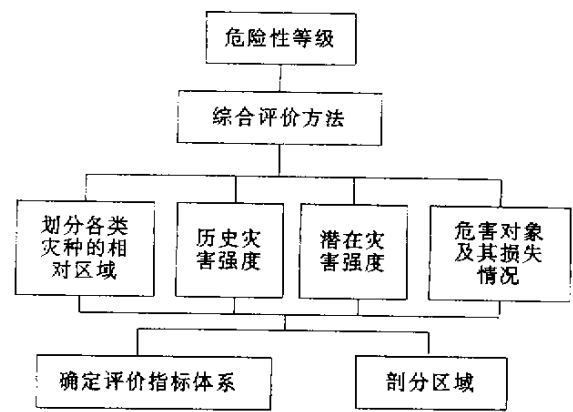


图 2 地质灾害危险性综合评价结构示意图

4 地质灾害危险性评价分级

根据地质灾害危险性的现状,划分研究地区不同稳定状态控制指标的特点,以及未来的趋势进行综合评估,进行稳定状态归类,同时根据其性质、变化、危害对象和损失情况,确定其分级特征(表 3)。

表 3 地质灾害危险性分级表

危险性分级	稳定状态	危害对象	损失情况
危险性大	差	城镇及主体建筑物	大
危险性中等	中等	有居民及主体建筑物	中
危险性小	好	无居民及主体建筑物	小

建立不同危险性大小的标准单元的指标特点,与剖分的各单元经主成分运算后进行比较归类,根据各单元与标准单元的相似程度进行归类,即可划分出研究区不同区域的危险性等级。

5 实例分析

依据以上分析的方法及步骤,对某高速公路所经场地的地质灾害进行了综合评估,评价指标包括岩土承载力、黄土湿陷量、液化指数、地形地貌条件、断裂构造、地下水条件等,见表 4。由于该段公路地质灾害的危害对象及损失情况基本相似,故此两项指标未参与计算。根据该公路段的工程地质特点,经综合分析,把研究区域分成 264 个单元,加上 3 个标准单元,共 267 个单元,每个单元有 9 个指标,组成 267×9 的矩阵,具体评价结果见图 3 与图 4。

根据计算得出综合评价结果,将地质灾害按稳定状态、危害对象及损失情况分为三级,并将线路划分为三种区段,即无地质灾害地段、地质灾害危险性小的地段、地质灾害危险性中等的地段。

6 结论

主成分分析法在地质灾害综合评估中的应用,能保证数据信息损失最小的前提下,经线性变换和

表 4 地质灾害分级评价指标表

级别	特 征	评价指标								
		地貌	潜水位埋深 /m	黄土湿陷性 /cm	液化 指数	岩土承载力 /kPa	断裂覆盖层 厚度/m	断裂影响带 宽度/m	断裂活动速率 /mm/a	断裂是否发震
地 质 灾 害 危 险 性小	地形平坦,地貌简单,地下水位埋深>8.0 m,黄土不具湿陷性,无砂土液化,无断裂构造,岩土承载力高,无发震断裂。	二 级 冲 洪 积阶地,冲 积平原区。	>8.0	<7	0	>200	>200	0	0	无
地 质 灾 害 危 险 性中等	地形略有起伏,地势相对较低,地下水埋深3.0~8.0 m,黄土具Ⅰ级(轻微)湿陷,砂土具轻微液化,岩土承载力较高,断裂活动性较弱,垂直埋深大,断裂无地震或有零星地震。	一 级 冲 洪 积阶地,缓 倾斜山前 平原。	3.0~8.0	7~30	0~5	120~200	80~200	0~25	<0.1	零星或无
地 质 灾 害 危 险 性大	地形有起伏,位于山前洪积扇及河道区,地下水埋深<3.0 m,黄土具Ⅱ级(中等)湿陷,砂土具中等液化,断裂具活动性,垂直埋深小,岩土承载力低,有发震断裂。	山 前 洪 积 扇, 河 道 区。	<3.0	30~60	5~15	<120	<80	>25	0.1~1	少量

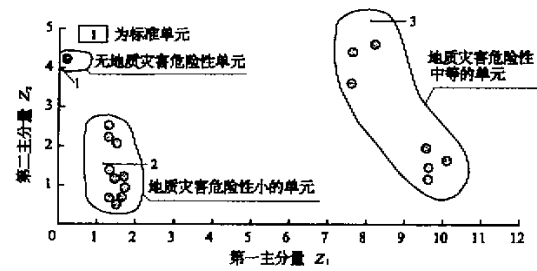


图 3 多指标综合评价分类图

舍弃一小部分信息,以少数新的综合变量取代原来采用的多维变量,较好地进行了地质灾害的分区,在地质灾害综合评估中是可行的。但也存在变量多重相关性的危害作用,有时会扭曲客观结论,故在指标的选取上要合理。该方法的具体计算结果的解释更加重要。

[参 考 文 献]

[1] 胡永宏,等.综合评价方法[M].北京:科学出版社,2000.
[2] 张尧庭.指标量化、序化的理论和方法[M].北京:科学出版社,1999.

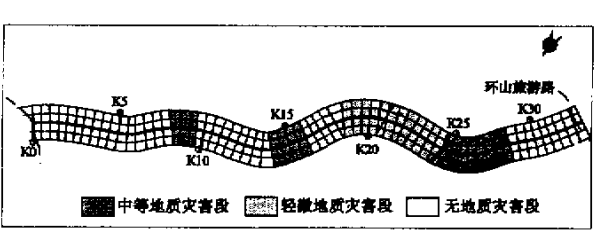


图 4 地质灾害分段评价结果图

[3] 任若恩,等.多元统计数据分析——理论、方法、实例[M].北京:国防工业出版社,1997.
[4] 吴翊,等.应用数理统计[M].长沙:国防科技大学出版社,1995.
[5] 张尧庭,等.多元统计分析引论[M].北京:科学出版社,1997.
[6] 陆雍森.环境评价[M].上海:同济大学出版社,1999.
[7] 卢崇飞,等.环境数理统计学应用及程序[M].北京:高等教育出版社,1988.
[8] 张梁,等.地质灾害灾情评估理论与实践[M].北京:地质出版社,1998.
[9] 李明远.多指标决策与评价的新方法——主成分投影法[J].数理统计与管理,2000,20(5):45~48.
[10] 高惠璇.处理多元线性回归中自变量共线性的几种方法[J].数理统计与管理,2000,20(5):49~55.

THE APPLICATION OF PRINCIPAL COMPONENTS ANALYSIS
IN SYNTHETICAL ESTIMATION OF GEOLOGICAL HAZARD

FAN Wen^{1,2}, LIU Xue-mei¹, GAO De-bin¹, SU Sheng-rui¹, CHANG Zhong-hua¹

(1. School of Geological Engineering and Surveying Engineering, Chang'an University, Xi'an 710054, China 2. School of Civil Engineering and Mechanics, Xi'an Jiaotong University, xi'an 710049, China)

Abstract : The principal components analysis method is a multivariate statistics analysis method in solve multi-index factor problems. This article introduces the principles and process of this method, and how to choose factors, and also multi-index factors estimation system of geological hazard is suggested. This method can be practicable in geological hazard estimation through practice this method.

Key words : principal components analysis method ; index factors system ; synthetical estimation ; geological hazard