

基于投影寻踪模型的震灾损失程度综合评价¹

李彦苍, 周书敬

河北工程大学 土木工程学院, 河北 邯郸 056038

【摘要】 为了克服现有的地震灾害损失程度综合评价方法中存在的主观性强、不易处理高维数据的缺陷, 本文提出了基于投影寻踪模型的评价新方法。该模型直接利用样本固有的数据, 将多维评价指标值投影为一维投影数据, 并据投影值的大小对损失进行评价, 克服了专家打分的主观性缺陷。同时, 利用改进蚁群算法实现了投影寻踪技术。应用实例表明, 该模型便于决策, 具有很强的客观性、适用性和可操作性。该研究对于为地震灾害发生后的求援及灾后重建决策提供了新的技术工具。

【关键词】 地震灾害; 灾害评估; 投影寻踪模型; 蚁群算法; 信息熵

Comprehensive assessment of earthquake destruction based on projection pursuit model

LI Yan-cang, ZHOU Shu-jing

College of Civil Engineering, Hebei University of Engineering, Handan 056038, China

Abstract: To overcome the shortcomings of the methods used in comprehensive assessment of earthquake destruction, a new method, named projection pursuit (PP) model was presented. The evaluation indexes values of each sample were synthesized as projection value with one dimension which indicates the destruction degree. The improved ant colony algorithm was employed to predigest the realized process of projection pursuit technique. Engineering practice shows that the model can realize the simple, objective and feasible assessment. This work has significance for the decision making of the succoring and rebuilding after the earthquake.

Keywords: earthquake disaster; destruction assessment; projection pursuit model; ant colony algorithm; information entropy

1 引言

地震损失综合评价是根据评价指标体系对地震破坏程度做出综合评定的过程, 其对于震灾地区救援、灾后重建决策具有重要意义, 该方面的研究具有重要的理论及现实意义。目前, 研究热点和瓶颈是如何合理地将多维评价指标转化为单维评价指标, 并克服人为赋权带来的主观性影响。近年来, 国内外学者提出的模糊综合优选模型、灰色综合优选模型、神经网络优选模型、层次分析法优选模型等对方案综合评价起到了很大的作用。但这些方法均需要专家打分, 同时, 由于方案优选涉及众多评价指标、各指标的量纲不尽相同、权重不易确定, 优选过程中存在着主观性强、不易处理高维数据的缺陷。

作为一种分析和处理非正态高维数据的降维分析方法, 投影寻踪模型最早是由 Friedman 于 1974 年提出的^[1]。其基本思想是利用计算机技术, 把高维数据 (尤其是高维非正态数据) 投影到低维 (1-3 维) 子空间上, 并通过优化投影指标函数, 寻找出能反应高维数据结构或特征的投影向量, 在低维空间上对数据进行分析, 以达到研究和分析高维数据的目的。其以成功应用于遥感图像解译、光(色)谱数据分析、环境毒理学分析等诸多领域^[2-3]。

2 基于投影寻踪方法的震灾后环境评价模型

震灾损失评价涉及技术、经济、环境、社会等多个因素, 是一个典型的多目标、多层次、多属性评价问题。以往的研究多通过专家对一系列评价指标打分来评价优劣, 存在较大的主观性。为此, 有必要采用 PP 这种有效的降维技术对灾后损失进行评价。模型的建立步骤如下:

(1) 评价指标规范化处理

设评价指标值的样本集为:

$\{x^*(i, j), i=1 \sim n, j=1 \sim p\}$, 其中, $x^*(i, j)$ 为第 i 个样本点的第 j 个指标值。 n, p 分别为样本的个数和指标的数目。首先对各指标值进行归一化处理。

对效益型指标, 采用:

$$x(i, j) = \frac{x^*(i, j) - x_{\min}(i, j)}{x_{\max}(i, j) - x_{\min}(i, j)} \quad (1)$$

对成本型指标, 采用:

$$x(i, j) = \frac{x_{\min}(i, j) - x^*(i, j)}{x_{\max}(i, j) - x_{\min}(i, j)} \quad (2)$$

式中, $x(i, j) \ i=1, 2, \dots, n; j=1, 2, \dots, p$ 为归一化处理后的评价指标集; $x_{\max}(i, j)$ 和 $x_{\min}(i, j)$ 分别为第 j 项评价指标值的最大值和最小值。

(2) 构造投影指标函数 $Q(a)$

¹ 河北省教育厅自然科学研究指导计划项目 (Z2003404)

Foundation Item: Natural Science Research Directed Foundation of Hebei Education Administration (No.Z2003404)

设 $a = a\{a(1), a(2), \dots, a(p)\}$ 为 p 维单位向量, PP 法就是将归一化处理后的 p 维评价指标集 $x(i, j)$ 投影到 $\bar{a}$ 上, 得到对应的一维线性空间投影值 $z(i)$:

$$z(i) = \sum_{j=1}^p a(j)x(i, j), \quad i = 1, 2, \dots, n \quad (3)$$

式中, $a(j), j = 1, 2, \dots, p$ 为投影方向单位长度向量。

这样, 找到能反应高维数据特征结构的最优投影方向 $\bar{a}$ 这一关键问题便转化为求有约束的投影指标函数 $Q(a)$ 的极值问题。约束条件是 $z(i)$ 的散布特征应满足局部投影点分类内尽可能满集, 类间尽可能展开。因此, $Q(a)$ 可表示为:

$$Q(a) = S_z D_z \quad (4)$$

式中, S_z 为类间散开度, 为投影值 $z(i)$ 的标准差; D_z 为类内密集度, 即投影值 $z(i)$ 的局部密度。两者的表达式分别为:

$$S_z = \left\{ \frac{\sum_{i=1}^n [z(i) - \bar{z}]^2}{n-1} \right\}^{\frac{1}{2}} \quad (5)$$

$$D_z = \sum_{i=1}^n \sum_{j=1}^p (R - r_{ij}) \cdot I(R - r_{ij}) \quad (6)$$

式中, $\bar{z}$ 为序列 $\{z(i), i = 1 \sim n\}$ 的平均值; R 为局部密度的窗口半径, 取值为 $r_{\max} + \frac{p}{2} \leq R \leq 2p$ [4]。 r_{ij} 表示样本之间的距离, $r_{ij} = |z(i) - z(j)|$; $I(R - r_{ij})$ 为单位越阶函数。

(3) 优化投影指标函数

各指标的样本集合给定时, 投影指标函数 $Q(a)$ 只随投影方向的变化而变化。可通过求解投影函数最

$$a = (0.1785, 0.1535, 0.2267, 0.1870, 0.2040, 0.3228, 0.1082, 0.0366, 0.1974, 0.2987, 0.1902)$$

。据投影值可确定城市 C 受损失最大, 城市 A、B 依次次之, 与现场实际情况相符。

4 结论

地震灾害评价在灾害防控及救援和灾后重建中具有重要意义。现有的评价方法存在主观性强、不易处理高维数据的缺陷, 本文引入基于投影寻踪模型的评价方法, 并采用改进蚁群算法实现了投影寻踪的操作。实例证明, 因为直接采用了各样本的原始数据, 该模型不会造成信息的丢失, 同时, 避免了专家赋权的人为干扰, 克服了传统方法的不足。该研究为安全工程及其他工程的灾害评价提供了一较为有效的方法。

大化问题来估计最佳投影方向, 即:

$$\text{Max: } Q(a) = S_z D_z$$

$$\text{s.t. } \sum_{j=1}^n A^2(j) = 1 \quad (7)$$

这是一个以 $\{a(j) | j = 1 \sim p\}$ 为优化变量的复杂非线性优化问题, 本文采用文献[5]提出的基于信息熵的改进蚁群算法求解, 具体的求解步骤同[5]。

(4) 评价

通过采用蚁群算法优化(7)式后, 可求得最佳投影方向 $a^*(i)$ 及相应的各样本点的投影值 $z(i)$, 将 $z(i)$ 按其值的大小排序, 便可实现对方案的评价。

3 工程实例

为验证方法有效性, 采用本文提出的模型, 对我国某地地震灾后三个城市的损失进行了综合评价, 各样本点评价指标及指标值见表 1。

表 1. 震灾后环境评价指标值

评价指标	城市 A	城市 B	城市 C
死亡人数 (人)	12660	12280	13170
重伤人数 (人)	4286	4800	5728
需安置人数 (人)	6990	6320	5230
最少需要救援人员 (人)	2900	2600	2800
房屋倒塌 (间)	62500	6500	5750
抢修道路所需天数 (天)	3.50	2.25	1.75
疾病防控所需物资 (kg)	4320	4050	4990
据铁路线距离 (m)	4280	3270	2370
重建所需投资 (万元)	190000	160000	120000
环境及生态平衡影响系数	0.218	0.272	0.509

将该方案集按式(1)和(2)进行归一化处理后, 依次代入(3)~(6)式, 采用改进蚁群算法对所得的(7)式进行优化, 得到最大投影指标函数值为 0.226, 最佳投影方向为

References (参考文献)

- [1] Friedman J. H., Turkey J W A. Projection pursuit algorithm for exploratory data analysis [J]. IEEE Trans on computer, 1974, 23(9): 881-890.
- [2] Vasil LovchinovStefan Tsakovski. Multivariate statistical approaches as applied to environmental physics studies [J]. Central European Journal of Physics, 2006, 4(2): 277-298.
- [3] W. Bobbie-Jo M, Jarman K. H, D Harvey S., et al. An improved optimization algorithm and a bayes factor termination criterion for sequential projection pursuit [J]. Chemo metrics and Intelligent Laboratory Systems, 2005, 77(1/2): 149-160.
- [4] Michael D S. Statistical Modeling of High-Dimensional Nonlinear Systems: A Projection Pursuit Solution [D]. Atlanta: Georgia Institute of Technology, 2005.
- [5] Y. Li, W. Li. Adaptive Ant Colony Optimization Algorithm Based on Information Entropy: Foundation and application [J]. Fundamenta Informaticae, 2007, 77(3): 229-242.