

# 主成分聚类分析在土地利用生态安全评价中的应用

鲍 艳, 胡振琪\*, 柏 玉, 郭瑞珊

(中国矿业大学(北京)土地复垦与生态重建研究所, 北京 100083)

**摘 要:** 土地资源是人类赖以生存的资本, 土地利用生态安全问题也日益被重视。该文采用主成分聚类分析的方法对土地利用生态安全进行评价并分类。先建立土地利用生态安全评价的指标体系, 然后用主成分分析剔除存在相关性、信息重叠的指标, 再利用主成分分析得到的具有代表性的主成分指标代替原来的评价指标, 对土地利用生态安全进行聚类分析, 所得结果客观, 为土地利用生态安全评价提供参考依据。

**关键词:** 土地利用; 生态安全; 主成分分析; 聚类分析

**中图分类号:** O212.4; S181

**文献标识码:** A

**文章编号:** 1002-6819(2006)08-0087-04

鲍 艳, 胡振琪, 柏 玉, 等. 主成分聚类分析在土地利用生态安全评价中的应用[J]. 农业工程学报, 2006, 22(8): 87- 90.

Bao Yan, Hu Zhenqi, Bai Yu, et al. Application of principal component analysis and cluster analysis to evaluating ecological safety of land use[J]. Transactions of the CSAE, 2006, 22(8): 87- 90. (in Chinese with English abstract)

## 0 引 言

人类的一切社会经济活动都离不开土地, 都要以占用一定数量的土地为前提条件。中国土地紧缺的压力越来越大, 土地问题已成为国民经济发展的一个严重制约因素。生态安全是近年来新提出的概念, 土地利用生态安全是土地资源可持续利用研究的前沿课题<sup>[1]</sup>。

国内对土地生态安全评价的研究并不多。王朝科对生态安全的概念及本质、特点、评价对象、指标体系建立的原则、评价的标准、构建指标体系的总体思路等几方面对生态安全进行了较全面的研究<sup>[2]</sup>; 冯孝杰认为土地利用是生态环境变化的重要动力, 生态环境变化是土地利用的积累结果<sup>[3]</sup>; 张建新等将选取的 24 个指标采用均方差进行赋权, 从各指标的统计特征值及全国的平均值出发, 确定湖南省的生态安全阈值  $S$ , 计算多指标加权综合平均值<sup>[4]</sup>; 陈浩等以怀来县为例对生态脆弱的荒漠化地区的安全状况进行研究, 采用关联度分析方法确定指标的权重, 并提出了衡量生态安全度的数量化指标、生态安全系数及其计算方法<sup>[5]</sup>; 王丽霞等从不安全的角度上分析, 用“不安全指数”来对黄土高原边缘地区的生态安全进行了评价与分析<sup>[6]</sup>; 曹新向对土地生态安全评价方法进行研究, 主要有综合指数法、景观指数法、景观生态安全格局法、层次分析法等<sup>[7]</sup>; 罗贞礼选取 24 个反映土地利用生态安全评价的指标, 并对指标作了系统聚类和分析<sup>[1]</sup>; 王耕等运用 GIS 的空间分析和空间数

据管理的功能讨论了西辽河生态安全空间差异的评价方法<sup>[8]</sup>; 李晓燕等在景观干扰度和脆弱度指标构建的基础上, 引入生态安全指数, 并利用区变量理论对研究区的生态安全状况进行时空动态分析<sup>[9]</sup>; 田克明将层次分析模糊评价法运用于农用地生态安全评价中<sup>[10]</sup>。

本文是在运用多元统计分析<sup>[11, 12]</sup>的相关理论与方法的基础上, 运用一种基于主成分聚类分析的客观分析方法对阜新市的土地利用生态安全进行综合评价研究, 旨在探索新的土地利用生态安全评价思路。

## 1 评价模型及评价步骤

### 1.1 主成分分析模型

主成分分析法<sup>[11-13]</sup>是将多个要素转化为少数综合指标的一种统计方法。简言之, 分析指标反映的信息在一定程度上有所重叠, 主成分分析是利用降维的思想, 将原来较多的指标用约化后较少的综合主成分指标来代替。综合指标保留了原始变量的绝大多数信息, 且彼此间互不相关, 能够使复杂问题简单化。基本原理如下。

1) 设有  $n$  个区域,  $p$  个指标, 初始样本矩阵  $X = (x_{ij})_{n \times p}$ ,  $i = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, p$ 。

2) 计算指标间相关系数矩阵  $R_{p \times p}$  及其特征值  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$  和正则化特征向量  $e_j$ ; 得到主成分  $Y_j = X e_j$ ;

3) 第  $j$  个主成分的方差贡献率为  $a_j = \lambda_j/p$ 。当累计方差贡献率  $a = \sum_{j=1}^q a_j$  达到一定数值(一般不小于 85%) 时, 取前  $q$  个主成分  $Y_1, Y_2, \dots, Y_q$ , 即认为这  $q$  个主成分就以较少的指标综合体现了原来  $p$  个评价指标的信息<sup>[11, 14]</sup>。

### 1.2 聚类分析模型

聚类分析(Cluster Analysis)是统计学所研究的“物以类聚”问题的一种方法, 它属于多变量统计分析的范畴。它是一种建立分类的方法, 能够将一批样本数据(或变量)按照它们在性质上的亲疏程度在没有先验知识的情况下自动进行分类。这里, 一个类就是一个具有

收稿日期: 2006-01-16 修订日期: 2006-02-24

基金项目: 教育部新世纪优秀人才计划(NCET-04-0484); 国土资源部重点科技项目

作者简介: 鲍 艳(1976-), 女, 山东烟台人, 博士生, 主要从事地理信息系统应用与土地利用研究。北京 中国矿业大学(北京)土地复垦与生态重建研究所, 100083。Email: baoyanok@126.com

\*通讯作者: 胡振琪(1963-), 男, 安徽五河人, 博士, 教授, 博士生导师, 主要从事土地复垦与生态重建研究。北京 中国矿业大学(北京)土地复垦与生态重建研究所, 100083。

Email: huzqbj@yahoo.com.cn

相似的个体的集合,不同类之间具有明显的非相似性。在分类过程中,不必事先给出一个分类标准,聚类分析能够从样本数据出发,客观地决定分类标准。

系统聚类法(Hierarchical Clustering Methods)也称层次聚类分析法,是目前国内外使用得最多的一种方法。这种方法的基本思想是:先将  $n$  个样品各自看成一类,然后规定样品之间的距离和类与类之间的距离。开始,因每个样品自成一类,类与类之间的距离与样品之间的距离是相等的,选择距离最小的一对并成一个新的类,计算新类和其他类的距离,再将距离最近的两类合并,这样每次减少一类,直至所有的样品都成一类为止。由此可见,系统聚类方法中,度量数据之间的亲疏程度是极为关键的。这里并没有给定分类的标准,也没有给出所有数据分成几类,而要求比较客观地从数据自身出发进行分类。类与类之间的亲疏程度有最短距离法、最长距离法、中间距离法、重心法、类平均法、离差平方和法等<sup>[12]</sup>。

### 1.3 评价步骤

- 1) 评价指标的选取;
- 2) 将选取的指标进行标准化处理;
- 3) 对标准化后的指标体系进行主成分分析,剔除存在相关性、信息重叠的指标,以减少数据的冗余;
- 4) 将进行主成分分析后的指标,进行聚类分析。

## 2 应用实例

研究区域—阜新市地处辽宁省西北部,位于东经  $121^{\circ}10' \sim 122^{\circ}56'$ ,北纬  $41^{\circ}41' \sim 42^{\circ}56'$  之间,下辖两县五区。东界与沈阳市毗邻,西与朝阳市接壤,北依内蒙古

自治区,南连锦州,是连接辽宁省中南部工业城市群、辽西走廊与内蒙、华北腹地的要冲。该地区地处内蒙古高原和辽河平原的中间过渡地带,属于辽宁西部的低山丘陵区。地势西北高,东南低。全地区总面积  $10323 \text{ km}^2$ 。

### 2.1 评价指标的选取

土地是区域发展的空间载体,任何社会、经济、生态的发展都离不开土地的支撑,没有土地资源保证的发展项目最终只能是空中楼阁化为泡影。而土地利用生态安全评价可以从一定程度上限制盲目滥用土地。土地利用生态安全评价指标的选取依据以下的原则。

1) 综合全面性原则。指标内容应该覆盖该区域社会经济的各个方面。包括人口特征,资源状况,经济发展水平,城镇化水平等。

2) 可比性原则。所选取的指标必须在该区域所辖的下一级单位具有横向可比性。

3) 易于获取性原则。指标的数据应该容易获取,且来源可靠,科学客观。

4) 简洁性原则。所选取的指标体系既要能全面反映研究对象,又要使指标个数尽可能地少。这要求选取一些代表信息量大且能反映事物本质特征的指标。

根据以上原则,选取了 11 个指标以反映该市的社会、经济、生态发展状况。其中,  $x_1$  为森林覆盖率(%),  $x_2$  为土地后备资源率(%),  $x_3$  为草地覆盖率(%),  $x_4$  为施用化肥量( $\text{kg}/\text{hm}^2$ ),  $x_5$  为农药用量( $\text{kg}/\text{hm}^2$ ),  $x_6$  为土壤污染率(%),  $x_7$  为水土流失率(%),  $x_8$  为人均耕地面积( $\text{hm}^2/\text{人}$ ),  $x_9$  为矿区土地复垦率(%),  $x_{10}$  为人均 GDP(元),  $x_{11}$  为城镇化水平(%)。具体指标数据见表 1。

表 1 土地利用生态安全评价指标数据

Table 1 Index data of land-use ecology safety assessment

区域	森林覆盖率 /%	土地后备资源率 /%	草地覆盖率 /%	施用化肥量 / $\text{kg} \cdot \text{hm}^{-2}$	农药用量 / $\text{kg} \cdot \text{hm}^{-2}$	土壤污染率 /%	水土流失率 /%	人均耕地 / $\text{hm}^2 \cdot \text{人}^{-1}$	矿区土地复垦率 /%	人均 GDP /元	城镇化水平 /%
海州区	2.93	2.521	0.01	608.55	4.20	0.18	0.00	0.007	24.44	1917.00	54.58
新邱区	12.37	7.788	3.38	334.20	6.00	0.17	0.00	0.036	16.75	2068.00	51.20
太平区	9.23	5.648	1.13	390.45	2.55	0.21	0.00	0.009	7.10	1496.00	46.96
细河区	5.10	2.874	0.71	444.60	8.85	0.18	0.01	0.027	0.00	1478.00	47.58
清河门区	5.46	9.763	2.12	704.70	9.30	0.15	0.01	0.060	1.99	4284.00	42.59
阜蒙县	21.64	8.868	8.02	485.85	3.75	0.01	0.03	0.298	0.90	2484.00	20.35
彰武县	25.29	5.526	5.73	590.55	3.45	0.01	0.01	0.298	0.80	2473.00	19.25

数据来源:《阜新市统计年鉴(2004)》、《阜新市土地利用总体规划(2003—2020)专题》、《阜新市矿产资源总体规划》。

## 2.2 指标数据的处理

### 2.2.1 数据标准化

由于不同的参数具有不同的量纲,它们的数量级差别很大。为了排除不同的量纲和不同的数量级对评价结果的影响,需要对各参数的原始数据进行标准化处理。本文采用标准差进行标准化处理。其标准化公式如下<sup>[15]</sup>

$$S = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

式中  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i (i = 1, 2, \dots, n)$ ;  $S$ ——样本的标准差;  $X_i$ ——总体  $X$  的样本;  $n$ ——样本的个数。

### 2.2.2 主成分分析

将标准化处理后的 11 个指标进行主成分分析,分析指标间的关系,剔除一些没有明显分异作用的指标或相互间存在明显的线性相关关系的指标,以确定最终的指标。整个过程利用 SAS(Statistics Analysis System) 统计软件<sup>[16]</sup>编程实现,运行结果如下。

相关矩阵的特征值

	特征值	方差	贡献率	累积值
1	6.11785062	3.79368148	0.5562	0.5562
2	2.32416915	1.29320129	0.2113	0.7675
3	1.03096786	0.12385846	0.0937	0.8612
4	0.90710940	0.54047252	0.0825	0.9436
5	0.36663688	0.11337080	0.0333	0.9770
6	0.25326608	0.25326608	0.0230	1.0000
7	0.00000000	0.00000000	0.0000	1.0000
8	0.00000000	0.00000000	0.0000	1.0000
9	0.00000000	0.00000000	0.0000	1.0000
10	0.00000000	0.00000000	0.0000	1.0000
11	0.00000000		0.0000	1.0000

从各成分对方差的贡献率可以看出,第 7 成分以后的上下特征值之差为 0。因此,对第 7 成分以后的成分忽略不计,这里只需对前 6 个主成分进行分析。

### 2.2.3 聚类分析

利用主成分分析得到的 6 个主成分指标数据组成聚类分析的样本矩阵,采用类平均法进行系统聚类分析,本文采用 SAS(Statistics Analysis System) 统计软件编程实现聚类分析的过程,并得到聚类分析的谱系图。运行结果如图 1 所示。

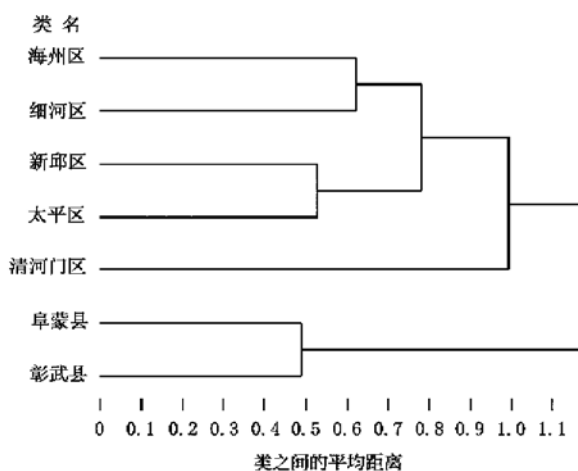


图 1 土地利用生态安全评价聚类图  
Fig. 1 Cluster graph of ecological safety assessment in land utilization

从图 1 中可以看出,取类间距离  $d = 1.0$  时,分为 2 类,有的类间距离太大,所以取  $d = 0.8$ ,分为 3 类比较合适。这样根据该地区的土地利用的具体情况将该地区的生态安全区分为 3 个区,其中阜蒙县与彰武县成一类,清河门区自成一类,海州区、新邱区、太平区、细河区成一类。这三类都各有特点:(1)阜蒙县与彰武县是市的两大大农业县,相对其它 5 个区来说,矿产资源较少,森林覆盖率较高;尽管受境外西北部沙漠的影响,但是有三北防护林的存在可以抵御西北风沙的侵袭,因此,土

地资源的生态环境较好。(2)清河门区是该地区中离是市中心较远的一个区,也是该市的煤炭生产基地,由于采煤的影响形成两个采煤沉陷区,对土地资源的生态环境也产生一定的破坏作用。(3)海州区是该地区的主城区所在地,同时也是煤炭资源的集聚地,有年产 800 万 t 煤炭的大型露天矿,采煤形成了 4 个采煤沉陷区,对矿区周边的土地生态环境影响较大。细河区位于市区的西部,由于阜新市城市发展“西进北挺”的思想,对细河区的土地资源开发利用程度较高,对周边土地资源的生态安全的影响是显而易见的。新邱区、太平区分别位于该市区的东部、北部。尽管新邱区、太平区的森林覆盖率较高,但是由于这 2 个区的煤炭资源比较丰富,采煤对周边的影响较大,在新邱区形成 3 个沉陷区、太平区形成了 4 个沉陷区,并且这 2 个区的人口密度较大。采煤沉陷区需要治理,城区面积有限,必然导致向外扩张,加大对土地资源的开发利用强度,因此土地利用的生态安全性较差。

### 3 结 论

利用主成分分析的结果作为聚类分析的样本矩阵,减少了数据的冗余,原理清晰,计算简单,所得的结论客观,比较符合阜新市目前发展的状况。该方法定量分析了阜新市的土地利用生态安全问题,并进行了类型分区:

- 1) 生态环境较好的区。阜蒙县与彰武县是两个农业大县,有三北防护林体系,农田防护林网较完善,生态绿化较好;
- 2) 生态环境待改善区。清河门区经济欠发达,煤炭资源是其经济发展的重点,而煤炭资源的开采不可避免地对土地资源生态环境造成破坏;
- 3) 生态环境脆弱区。海州区、太平区、细河区是阜新市的主城区,尤其细河区是阜新市经济发展的重点区域,经济转型时期的土地资源利用程度的加大必然会在一定程度上影响土地资源生态环境;海州区、太平区、新邱区是煤炭资源丰富区,尽管煤炭资源日渐枯竭,但是矿产资源开发对土地资源造成的影响、破坏还继续存在。

生态类型区的划分对阜新市日后土地资源的再利用提供了一定的参考和依据,该方法对其它地区的土地利用生态安全评价也将会有较好的借鉴作用。

#### [参 考 文 献]

[1] 罗贞礼. 土地利用生态安全评价指标体系的系统聚类分析[J]. 湖南地质, 2002, 21(4): 252- 254.  
 [2] 王朝科. 建立生态安全评价指标体系的几个理论问题[J]. 统计研究, 2003, (9): 17- 20.  
 [3] 冯孝杰, 高殿森, 顾宏波, 等. 区域的土地利用与生态安全及可持续发展的分析[J]. 后勤工程学院学报, 2004(3): 31 - 34.  
 [4] 张建新, 邢旭东, 刘小娥. 湖南土地资源可持续利用的生态安全评价[J]. 湖南地质, 2002, 21(2): 119- 121.  
 [5] 陈 浩, 周金星, 陆中臣, 等. 荒漠化地区生态安全评价

- 以首都圈怀柔县为例[J]. 水土保持学报, 2003, 17(1): 58- 62.
- [6] 王丽霞, 任志远. 黄土高原边缘地区生态安全评价与分析——以山西省大同市为例[J]. 干旱区研究, 2005, 22(2): 251- 255.
- [7] 曹新向, 郭志永, 雒海潮. 区域土地资源持续利用的生态安全研究[J]. 水土保持学报, 2004, 18(2): 192- 195.
- [8] 王耕, 吴伟. 基于GIS的西辽河流域生态安全空间分异特征[J]. 环境科学, 2005, 26(5): 28- 33.
- [9] 李晓燕, 张树文. 基于景观结构的吉林西部生态安全动态分析[J]. 干旱区研究, 2005, 22(1): 57- 62.
- [10] 田克明, 王国强. 我国农用地生态安全评价及其方法探讨[J]. 地域研究与开发, 2005, 24(4): 79- 82.
- [11] Richard A Johnson, Dean W Wichern. 实用多元统计分析(第四版)[M]. 北京: 清华大学出版社, 2001.
- [12] 张尧庭, 方开泰. 多元统计分析引论[M]. 北京: 科学出版社, 2003.
- [13] 于秀林, 任雪松. 多元统计分析[M]. 中国统计出版社, 2003.
- [14] 李玉民, 李旭宏, 毛海军, 等. 主成分聚类分析在省域物流规划中应用[J]. 东南大学学报(自然科学版), 2004, 34(4): 549- 552.
- [15] 马玲, 高运良. 数理统计[M]. 北京: 煤炭工业出版社, 2002.
- [16] 阮桂海, 等. SAS 统计分析实用大全[M]. 北京: 清华大学出版社, 2003.

## Application of principal component analysis and cluster analysis to evaluating ecological safety of land use

Bao Yan, Hu Zhenqi<sup>\*</sup>, Bai Yu, Guo Ruishan

(China University of Mining and Technology(Beijing), Research Institute of Land Reclamation and Ecological Restoration, Beijing 100083, China)

**Abstract:** Land resources is depending capital of human existence, therefore, more and more problems of ecological safety in land utilization are concerned. Principal component analysis and cluster analysis were used in appraising and classifying ecological safety in land utilization. The author established firstly the index system of ecological safety assessment in land utilization, then applied principal component analysis to eliminate the indexes having the relativities and overlap information. The representative indexes from principal component analysis process substitute for the primary indexes. And then cluster analysis method was used to analyze ecological safety in land utilization. The conclusion is objective, and it can provide reference to ecological safety.

**Key words:** land utilization; ecological safety; principal component analysis; cluster analysis