

文章编号: 1000-0585(2001)06-0761-07

典型相关分析在土地利用结构研究中的应用 ——以环渤海地区为例

张 明, 朱会议, 何书金

(中国科学院地理科学与资源研究所, 北京 100101)

摘要: 典型相关分析是一种揭示两组多元随机变量之间相关关系的统计模型方法。本文通过介绍这一多元统计方法的内涵、特点和思路, 将其引入有关地学问题的分析——判别土地利用类型分布与其影响因子之间的相关关系。以环渤海地区为例, 通过数据准备、操作过程和统计检验等几个方面, 全面介绍了典型相关分析在实际研究中的应用。

关键词: 土地利用; 典型相关分析; 环渤海地区

中图分类号: F301.24; O213 **文献标识码:** A

20 世纪 90 年代以来, 土地利用与土地覆被变化成为全球环境变化研究的重点领域。目前国际上有关土地利用与土地覆被变化的研究项目多侧重于土地覆被的分类、动态监测和环境影响评价, 对土地结构及其变化的动力机制的研究虽然也开展了一些工作, 但多难以深入。科学家们均已认识到人类活动是影响土地利用变化的主要驱动力, 但在建立分析模型时受到各种限制, 尚不能成功地将社会经济因子的驱动力贡献加以定量分析和模拟。

典型相关分析 (Canonical Correlation Analysis) 是近年来开始广泛普及的一种新型多元统计分析方法。它最早由 H. Hotelling 于 1936 年在《生物统计》期刊上提出^[1], 经过长期应用发展, 到 70 年代臻于成熟。由于典型相关分析涉及较大量的矩阵计算, 其应用在早期曾受到相当大的限制。现代计算机技术的飞速发展, 弥补了应用典型相关分析中的困难症结, 因此它在各个领域的应用开始逐渐普及并蓬勃发展起来。

典型相关分析揭示的是两组多元随机变量之间的关系。这两组多元随机变量中一组是自变量 (解释变量) 组, 另一组则是标准变量组。通过模型计算将变量组之间的相关关系剖析到可能的最为简化和清晰的方案。由于它是一种更一般性的方法, 具有较强的分析能力, 同时操作简单便利, 所以在解决实际问题中有着广泛应用, 如在生态学研究领域, 常用它来分析不同生物种群与各种生境因子间的相关关系等。和其他相关统计方法比较, 典型相关分析特别适合于所分析的标准变量组的各个变量之间本身具有较强的相关性。近年来随着全球变化研究的不断深入, 土地利用与土地覆被变化成为各项研究集中的焦点^[2], 不同的土地利用类型与不同的自然和社会经济因子之间的相互关系研究成为其中不可或缺的部分^[3,4], 由于不同的土地利用类型之间相互影响, 相互制约, 所以典型相关分析自然成为研究土地利用特征同自然—人为因素之间的关系的最佳统计工具。日本学者较早将典

收稿日期: 2001-09-08; 修订日期: 2001-10-26

基金项目: 国家自然科学基金项目 (49831020); 中科院知识创新工程项目 (KZCX2-310-01-06)

作者简介: 张明 (1970-), 男, 河南安阳人, 博士。主要从事土地利用与土地覆被变化研究。

型相关分析同其他经典数理统计方法结合一并引入到土地利用变化研究中,对区域土地利用结构变化的驱动因子进行了较为缜密的相关诊断,取得了满意的结果^[5]。对土地利用结构及其变化机制的研究不仅将大大丰富人类对土地覆被变化规律的认识,它也是土地利用变化前景预测和可持续土地利用规划的重要基石^[6,7]。本文将对这一方法的思路和操作进行介绍,并以环渤海地区为例,探讨它在土地利用结构及其变化分析中的具体应用。

1 基本思路

图 1 示意了典型相关分析的主要思路。设有两组观测变量,目标变量是土地利用类型,自然和社会经济变量为解释变量。我们来分析不同土地利用类型分布与自然和社会经济因子间的相关关系。具体做法是:在第一组变量中提出一个典型变量,在第二组变量中也提出一个典型变量,并使这一典型变量组合具有最大的相关;然后在每一组变量中提出第二个典型变量,使得在与第一个典型变量不相关的典型变量中,这两个典型变量组合之间的相关是最大的。将此过程继续进行,直至两组变量间的相关被提取完毕为止^[8]。可见,典型相关分析把原来较多变量转化为少数几个典型变量,通过这较少的典型变量之间的典型相关系数来综合的描述两组多元随机变量之间的相关关系。

本文主要介绍应用大型社会科学统计软件包 SPSS 中附带的命令程序来进行典型相关分析及专业分析。

2 操作程序

与一般的相关分析和回归分析一样,典型相关分析要求模型中所有的变量为间距测度等级,样本数目符合统计分析要求;除此之外,各组内的观测变量之间不能有高度的多重共线性。后者对于提高统计诊断模型精度是很重要的。SPSS 软件提供了多重共线性诊断程序,可以自动甄别输出具有多重共线性的变量来。

SPSS 中附带的典型相关程序是以 SYNTAX 命令直接编写的文件。其文件名为: CANCECORR.SPS。在 SPSS 中首先打开已经准备好的数据文件,将所有数据调入 SPSS 中的工作文件窗口。在调用 CANCECORR 程序之前需编写有关语句:打开 SYNTAX 窗口,输入调用上述命令程序及定义典型相关分析变量组的命令。

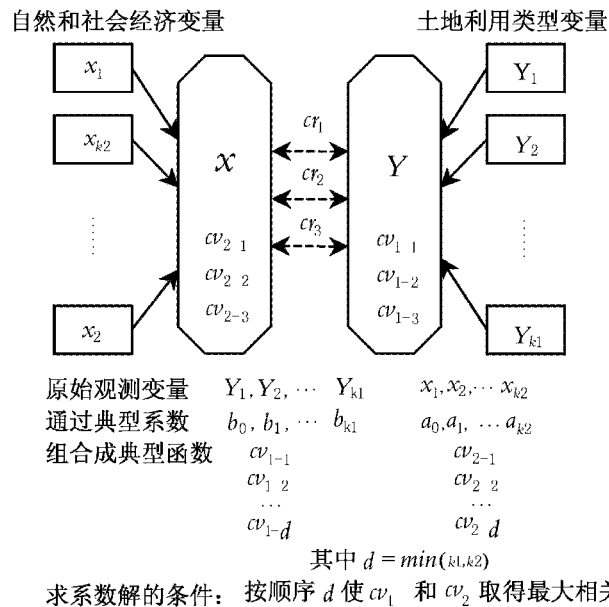


图 1 典型相关分析主要思路示意图

Fig. 1 The main thought of canonical correlation analysis

```

INCLUDE 'C:\SPSSWIN\CANCORR.SPS
CANCORR SET1 = Y1 Y2 Y3
/SET2 = x1 x2 x3 x4.

```

上述命令语句可以根据实际要求进行相应的修改，如添加对数据进行标准化处理的语句等，但一定要确保变量名与 SPSS 的工作文件中的变量名一致。

运行以上命令语句，即可得到所有典型相关分析的结果。

通过 SPSS 中的子程序宏指令 CANCORR，两组变量之间最大相关的系数可以被判定，运作产生的直接结果包括：典型相关系数；典型相关系数的平方；特征值及指标；检验典型相关系数（整体检验和维度递减检验）；典型系数；典型负载系数；交叉负载系数；变式对总方差的代表比例；冗余指数。其中，典型相关系数的平方、变式对总方差的代表比例及冗余指数可作为统计的检验指标，反映模型模拟的精度。此外，SPSS 的运行过程还在工作文件中计算出了每个随机变量在每个样本单元的典型得分（Canonical Score），从而能在空间上反映出典型相关模拟的程度。

3 具体应用

我们通过 SPSS 软件，运用典型相关分析对环渤海地区土地利用结构与自然和社会经济因子之间的相关关系进行揭示。

选取的目标变量组——土地利用类型分别为：耕地、林地、草地、水域、城建矿居用地和未利用地，为 1995 年遥感影像解译数据。自变量组——自然社会经济因子包括地形、地貌、气候以及同时期社会经济统计指标等 58 个变量。样本单元为行政县。

SPSS 运行结果所显示的第一对典型变量是

土地利用结构：

$$w_1 = -0.845y_1 + 0.931y_2 + 0.218y_3 - 0.182y_4 - 0.199y_5 - 0.289y_6$$

影响因子：

$$z_1 = 0.438x_1 - 0.741x_2 - 0.559x_3 - 0.818x_4 - 0.32x_5 + \dots - 0.301x_{14} + \dots + 0.153x_{58}$$

第二对典型变量是

土地利用结构：

$$w_2 = -0.248y_1 + 0.06y_2 - 0.387y_3 + 0.414y_4 + 0.3330y_5 + 0.768y_6$$

影响因子：

$$z_2 = 0.072x_1 - 0.049x_2 - 0.46x_3 + \dots - 0.387x_{32} + \dots + 0.362x_{41} + \dots - 0.028x_{58}$$

由于同一随机变量组内各典型变量之间的样本协方差为零，不同组不对应的典型变量间的样本协方差也为零，这使得分析变量组 Y 和变量组 X 之间的关系转化为只需分析从两组中提取出的相对应的典型变量之间的关系^[9]。上述函数中的变量系数我们称之为典型负载系数，它的统计含义就是这种转化关系的反映，它作为典型相关分析的主要结果，以桥梁的作用体现出标准变量组同自变量组之间的相关程度。

图 2 和图 3 用折线图较为直观地展示了运用典型相关分析对环渤海地区土地利用变化驱动力判别的结果（第一和第二个典型变量）。

第一个典型变量将耕地和林地从其他类用地中区分出来，其典型载荷为 -0.845 和

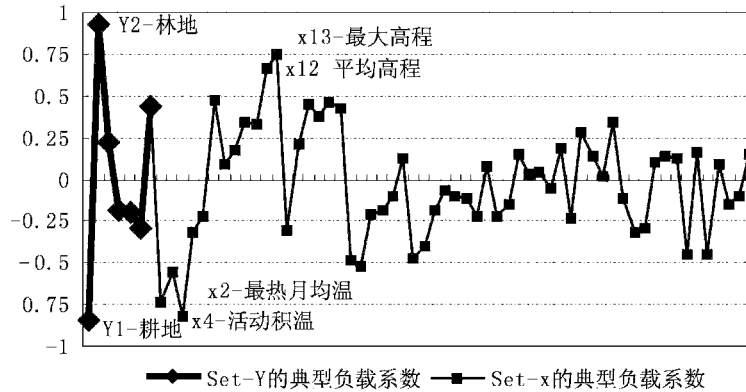


图 2 第一个典型变量的典型负载系数

Fig. 2 The canonical loadings of the first canonical variate

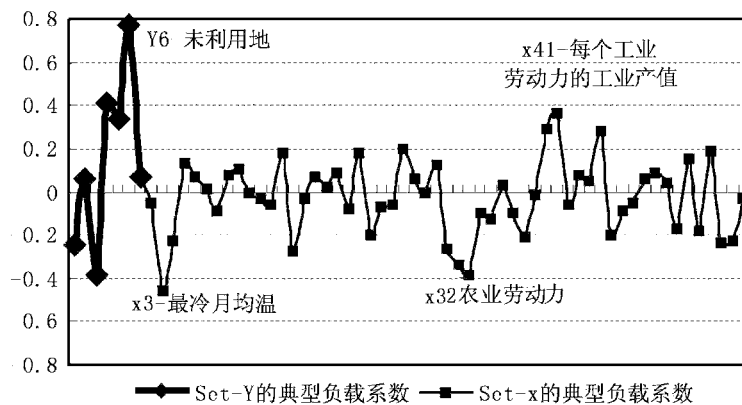


图 3 第二个典型变量的典型负载系数

Fig. 3 The canonical loadings of the second canonical variate

0.931, 自变量组中与之相对应的解释变量主要是活动积温、最大高程、最热月均温, 其典型载荷分别为 -0.818 、 0.749 和 -0.741 。在以平原为主的环渤海地区, 耕地和林地的分布主要受温度和地形条件的制约和影响。第二个典型变量将未利用地从其他用地中区分开来, 它的分布与最冷月均温、农业劳动力数目及每个工业劳动力的工业产值呈最大相关(相应的典型载荷分别为 0.768 和 -0.46 、 -0.387 和 0.362 。可见, 未利用的分布除受自然的温度条件限制外, 还受到农业和工业企业发展水平的限制。此外, 第四个典型变量将城居工矿用地从其它用地中区分出来, 解释变量中相关负载系数最大的是人口密度, 农用地和非农业用地人口分布的差异正好反映和解释了这种非农业用地和人口密度的最大相关^[10~12]。

用于检验典型相关分析结果的最一般方法是看其典型相关系数。为保证统计结果的可靠性, 我们采用典型变量之间的共享方差, 即典型相关系数的平方。图 4 给出了 6 个典型变量的典型相关系数的平方。可以看出, 这些值比较高, 它们表明相应的典型变量之间关

系密切。尤其是头两个典型相关变量对应的相关系数高达 0.91 和 0.78，表明判别出的解释变量能清晰充分地解释相应标准变量的分布。

检验典型相关分析结果的另一个指标是冗余指数。它是一组当中形成的典型变量对另一组观测变量总方差的解释比例，是一种组间交叉共享比例。这一比例可以反映自变量组各典型变量对于因变量组所有观测变量的解释能力^[13,14]。图 5 示意了冗余指数分析的结果。可以看出，第一、二和四个典型变量具有较高的解释百分比，如第一个典型变量的代表比例为 29.8%，其相应的解释比例为 27.1%。

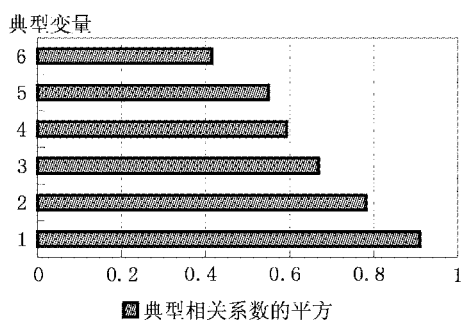


图 4 典型变量的典型相关系数的平方
Fig. 4 The squares of canonical correlation

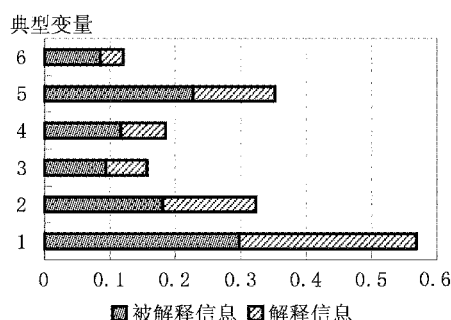


图 5 冗余指数分析
Fig. 5 The analysis of redundancy index

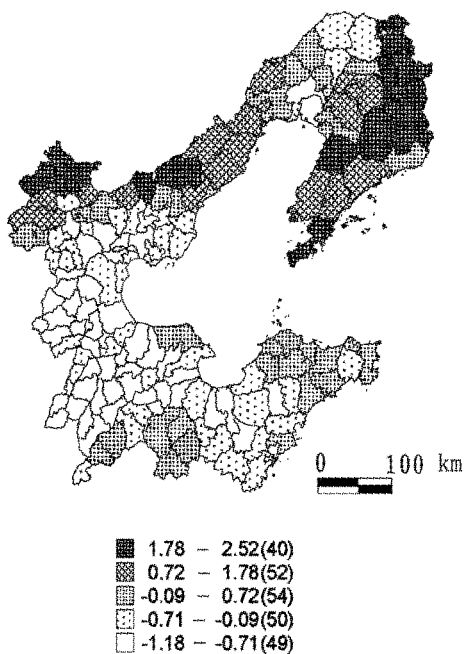


图 6 第一个典型变量被解释变量典型得分空间分布
Fig. 6 The spatial distribution of canonical scores of the explained variable for Set-1

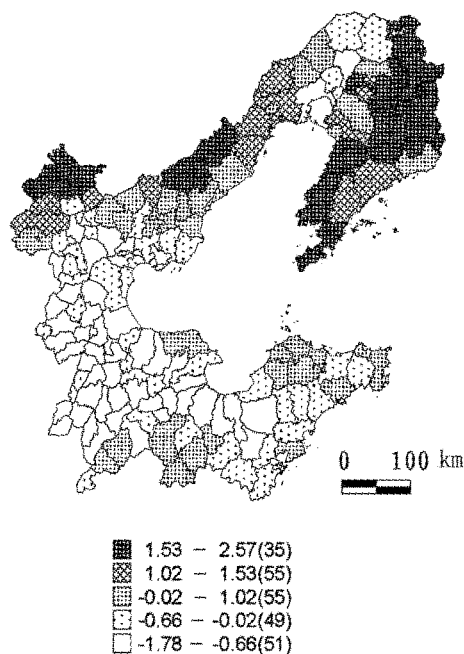


图 7 第一个典型变量解释变量典型得分空间分布
Fig. 7 The spatial distribution of canonical scores of the explanatory variable for Set-1

另外, 通过将本次数理统计的最小单元——行政县的分布图数字化, 将空间标识号同典型相关分析得出的每个单元的典型得分链接, 使解释变量和被解释变量之间的解释程度通过空间的单元得以体现。图 6 和图 7 就是根据典型得分所编制出的第一个典型变量相关指数的空间分布图, 可以看出, 解释变量和被解释变量之间的相关关系拟合得很好。

4 结论

典型相关分析用于描述两组变量之间所存在的相关关系。它广泛适用于地学中各类专业分析, 特别是对土地利用结构和变化与其影响因子之间相关关系的诊断。

与其它数理统计方法相比, 典型相关分析将每组变量作为一个整体看待, 通过典型负载系数可以定量判别不同因子对因变量的影响程度或对被解释变量的解释的贡献程度。这是它能广泛应用于地理专业统计分析的突出特点。

对于各维度上的典型相关关系进行检验, 可以精简不显著的维度, 以便将分析集中于那些能够肯定存在相关关系的维度。其中, 冗余分析是其中十分关键的一环, 它综合分析指标对两个变量组之间的整体相关性进行评价。

全面获取区域单元社会经济及自然因素数据是对统计分析精度的根本保证。建立全面综合的地理信息系统, 对认识土地利用结构, 准确定量判别其驱动因素具有重要意义。

参考文献:

- [1] Hotelling H. Relations between two sets of variates[J]. *Biometrika* 28, 1936; 321~377.
- [2] 李秀彬. 全球环境变化研究的核心领域[J]. *地理学报*, 1996, 51(6): 554~557.
- [3] Turner II B L, Skole D, Sanderson S, *et al.* Land-Use and Land-Cover Change, Science/Research Plan [A]. IG-BP Report No. 35 & HDP Report No. 7. Stockholm;IGBP, 1995. 52~60.
- [4] Turner II B L, Moss R H, Skole D. Relating Land Use and Global Land-Cover Change [A]. IGBP Report No. 24 & HDP Report No. 5, Stockholm;IGBP, 1993. 21~25.
- [5] Satoshi Hoshino. Statistical Analysis of Land-use Change and Driving Forces in the Kansai District, Japan[A]. IIASA Working Paper, wp-96-120. Laxenburg;IIASA, 1996. 20~25.
- [6] 李平, 等. 我国现阶段土地利用变化驱动力的宏观分析[J]. *地理研究*, 2001, 20(2): 129~131.
- [7] 冷允法, 等. 基于 GIS 的土地利用变更与时空动态分析[J]. *地理研究*, 1999, 18(增刊): 136~138.
- [8] Jobson J D. Applied Multivariate Data Analysis [A]. Volume II; Categorical and Multivariate Methods [C]. Springer-Verlag, 1992; 181~189.
- [9] William R. Dillon & Matthew Goldstein. Multivariate Analysis - Methods and Applications [A]. New York; Wiley & Sons, 1984, 337~359.
- [10] 张明. 榆林地区脆弱生态环境的景观格局与演化研究[J]. *地理研究*, 2000, 19(1): 31~33.
- [11] 陆大道. 中国环渤海地区持续发展战略研究[M]. 北京: 科学出版社, 1995. 175~186.
- [12] 刘卫东. 论环渤海地区资源结构与产业结构特点[J]. *地理研究*, 1995, 14(2): 83~85.
- [13] 王国梁, 何晓群. 多变量经济数据统计分析[M]. 西安: 陕西科学技术出版社, 1993. 17~29.
- [14] 郭志刚. 社会统计分析方法—SPSS 软件应用[M]. 北京: 中国人民大学出版社, 1999. 320~327.

An application of canonical correlation analysis to research on land-use patterns ——Taking Bohai Rim as an example

ZHANG Ming, ZHU Hui-yi, HE Shu-jin

(Institute of Geographic Sciences and Natural Resources Research, Beijing 100101, China)

Abstract: Land-use and land-cover change (LUCC) has become a key field for the research of global environmental change since the 1990s. More attention is paid to the identification and description of the driving forces of LUCC. For all kinds of limitations, it is difficult to incorporate different influencing factors into a whole model to carry out such an analysis.

Canonical Correlation Analysis (CCA) is a kind of traditional statistical model to show the relationship between two multiple stochastic variables. It is widely applied in different research fields to describe the relationship between the set of criterion measures and the set of explanatory factors. This paper introduces the connotations, characteristics and procedures of this multivariate analysis. It is combined with one of the analyses on geographic problems, to identify the relationship between the land-use patterns and its influencing factors. Taking the Bohai Rim in China as an example, through the preparation of the data, operation procedures and statistical checks, the application of CCA in a practical study is given comprehensively.

It is shown that CCA could be widely applicable to the geographic analysis, especially to the identification of the relationship between land-use structure and its influencing factors. Through the canonical loadings analysis, the influencing extension of different factors to the dependant variables or the contribution extension of explanatory variables to the explained variables could be described quantitatively. As CCA is carried out in combination with a series of data checks and precision tests, it makes the specific analysis more rational and scientific.

Key words: land-use change; canonical correlation analysis; Bohai Rim