

建立基于土地利用的交通发生量模型的研究*

张利, 戚浩平

(东南大学交通学院, 南京 210096)

摘要:针对出行生成预测中变量众多以及多重相关性问题,以常州调查数据为样本,分别采用逐步回归分析方法和偏最小二乘回归(PLSR)方法,建立基于土地利用形态的交通发生量模型。并对两者建模方法进行分析比较,结果表明PLSR能有效地解决变量多重相关性问题,模型意义清晰,效果显著。

关键词:偏最小二乘回归(PLSR);土地利用;交通发生量;多元回归

中图分类号:P962

文献标识码:A

土地利用是交通建设、人口和经济活动在空间上集聚的表现,同时也是社会经济活动在空间上的表现,是产生交通的根源^[1]。城市土地利用与交通相互联系、相互影响。不同的土地利用形态,决定了交通发生量和交通分布形态,在一定程度上也决定了交通的结构和形式。一般来说不同土地利用布局性质和强度对应着不同的交通发生量^[2]。因此,从土地利用的角度研究城市交通发生量模型是很有意义的。

由土地利用与交通生成之间的关系,城市交通发生量可按照土地利用——出行生成相关因素——出行交通量的程序进行^[3],建立交通发生量与土地利用相关指标等变量之间的定量关系,进而根据规划年的土地利用推算出规划年的交通发生量^[4]。事实上,由于交通发生量不仅与用地类型、地块面积有关,还与土地开发强度密切相关,所以本文研究中除了调查各类用地类型面积外,并引入建筑密度、建筑容积率等土地开发强度指标。文章结合常州新北区综合交通规划的实践,以城市土地利用特性为直接变量,分别基于多元回归分析方法和偏最小二乘回归方法建立交通生成预测模型,通过对两者建模方法的比较,指出偏最小二乘方法不仅可以有效分析各个参数对交通发生量的影响,而且可以准确预测交通发生量,对交通需求预测有着较大的意义。

1 多元回归分析方法

回归分析,就是研究变量之间的统计相关关系的一种统计方法。它从自变量和因变量的

* 收稿日期:2009-06-04

基金项目:中国地质调查局资助项目(121201010510218-5)。

第一作者简介:张利(1984~),女,江苏宿迁人,硕士研究生,主要研究方向为摄影测量与遥感。

一组或多组观测数据出发,寻找一个函数式,将变量之间的统计相关关系近似的表达出来。这个能够近似表达自变量与因变量之间关系的函数式,称为回归方程或者回归函数^[5]。在回归方程中,存在多个自变量的称为多元回归。回归分析是实际应用中一种很重要的统计分析方法。由于交通发生量往往和多个因素有关,本文涉及到的是多元回归分析。多元回归模型的基本原理为:所研究的对象 y 受多个因素 $x_1, x_2, \dots, x_k$ 的影响,假定各个影响因素与 y 的关系是线性的,则可建立多元线性回归模型:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \epsilon \quad (1)$$

其中, $x_1, x_2, \dots, x_k$ 为影响因素, ϵ 称为随机误差项。

多元回归模型的检验主要包括拟合优度检验、方程显著性检验、变量显著性检验^[6]。拟合优度检验通常采用 R^2 检验,检验回归方程对样本观测值的拟合程度;方程显著性检验即 F 检验,判断因变量与所有自变量之间是否为线性关系;变量显著性检验即 t 检验,用来挑选重要因素,剔除可有可无的因素。

2 偏最小二乘法回归建模

偏最小二乘方法是应实际需要而产生和发展起来的一个新型的、有广泛适用性的统计分析方法^[6]。1983年由S. Wold^[7]首次在化工领域得到了广泛的应用,从此偏最小二乘回归的应用迅速扩展到其它领域,如教育评测学,药物研究等^[8]。偏最小二乘法基于主成分回归思想,其建模思路^[9]是:首先,从自变量 $X(x_1, x_2, \dots, x_p)$ 中提取相互独立的成分 $t_h (h=1, 2, \dots)$,从因变量 $Y(y_1, \dots, y_q)$ 中提取相互独立的成分 $u_h (h=1, 2, \dots)$,然后建立这些成分与自变量的回归方程。

偏最小二乘法具有以下突出的优点:能够在自变量严重多重相关条件下进行回归建模;允许样本点少于变量个数的条件下建模;模型更易于辨识系统信息和噪声。为了推导方便,数据先标准化处理,其建模过程如下:

(1) 提取交通发生量的成分 u_1 和土地利用指标的第一个成分 t_1 。 u_1 是表示标准化的交通发生量。从 X_0 中抽取第一个成分 t_1 , t_1 为各土地利用形态变量的线性组合,满足 $t_1 = X_0 \omega_1$ 且 $\|\omega_1\|^2 = 1$, ω_1 为 X_0 的第一个轴。取 $\omega_1 = X_0^T Y_0 / \|X_0^T Y_0\|$,进而建立 Y_0 对 T_1 的回归及 X_0 对 T_1 的回归方程。

(2) 提取第 h 个土地利用的成分 t_h ,以土地利用变量的残差阵 X_1 取代 X_0 ,以交通发生量的残差阵 Y_1 取代 Y_0 ,用上述方法求第2个轴 ω_2 和第2个成分 t_2 ,有 $\omega_2 = X_1^T Y_1 / \|X_1^T Y_1\|$, $t_2 = X_1 \omega_2$,同样, X_1, Y_1 分别对 t_2 做回归。以此类推可以求出 h 个成分 t_h 。成分的个数可以通过交叉有效性采取截尾的方式选取。在得到成分 $t_1, \dots, t_m$ 后,有

$$Y_0 = \beta_1 t_1 + \dots + \beta_m t_m + Y_m \quad (2)$$

由于 t_h 均为 X_0 的线性组合,把 t_h 代入方程,即得到 Y_0 与 X_0 的线性表达式。建模精度以成分 $t_1, \dots, t_h$ 对交通发生量的累计解释能力表示,即 y 与 $t_1, \dots, t_h$ 的累计相关系数之和。

(3) 交叉有效性截取成分,记 y_i 为原始数据, $t_1, \dots, t_m$ 是在偏最小二乘回归过程中提取的成分, $\hat{y}_m$ 是使用全部样本点并取 $t_1 \sim t_h$ 个成分回归建模后,第 i 个样本点的拟合值。 $y_{h(\dots)}$ 是在建模时删去的样本点 i ,再用此模型计算的 y_i 的拟合值。记

$$PRESS_h = \sum_{i=1}^n (y_i - \hat{y}_{h(-i)})^2$$

(3)

PRESS 取最小时表明模型的拟合效果最好,此时提取的成分个数 *h* 为最佳成分数。

3 实例分析

常州高新区是常州市唯一的国家级高新技术产业开发区濒临长江,位于常州市北面,南接常州市老城区,区域总面积为 439.16 km²。根据交通小区现状与未来规划要求,将高新区分为 25 个交通小区,区域内没有特殊用地。依据交通规划的实践调查了 25 个交通小区的交通发生量,以及各区的建筑密度、建筑容积率、居住用地、公共设施用地、工业用地、仓储用地、对外交通用地、道路广场用地、市政公用设施用地、绿地、水域或其它用地等土地利用形态指标。

3.1 相关性分析

对各自变量和交通发生量进行相关性检验,其相关系数表见表 1。表 1 可见各土地利用形态变量之间存在多重共线性,即自变量之间存在着线性相关的现象,比如居住用地和容积率的相关性达到了 0.85,建筑密度和容积率达到 0.62,绿地和水域达到 0.65,公共设施和建筑密度达到 0.58。由上节所述采用 PLSR 可以解决变量的多重共线问题;对于变量的多重共线性,采用逐步回归法剔除不太重要的相关性变量是多元回归分析常用的方法。通过对实验数据的相关性分析分别采用 PLSR 和逐步回归方法建模。

表 1 变量相关系数

Table 1 Correlation coefficients of variables

相关系数	建筑密度	容积率	工业用地	市政用地	公共设施	居住用地	仓储用地	道路广场	对外交通	水域其它用地	绿地	出行发生
建筑密度	1.00											
容积率	0.62	1.00										
工业用地	0.25	-0.24	1.00									
市政用地	0.00	-0.07	-0.10	1.00								
公共设施	0.58	0.46	-0.05	0.21	1.00							
居住用地	0.34	0.85	-0.30	-0.12	0.19	1.00						
仓储用地	-0.15	-0.06	-0.05	-0.14	0.21	-0.07	1.00					
道路广场用地	-0.09	-0.12	-0.07	0.12	-0.03	0.00	-0.06	1.00				
对外交通	0.03	0.04	-0.19	-0.13	0.32	-0.12	-0.08	-0.06	1.00			
水域其它用地	-0.29	-0.26	0.16	-0.09	-0.54	-0.16	-0.09	0.05	-0.24	1.00		
绿 地	-0.19	-0.20	0.48	-0.14	-0.44	-0.19	0.06	-0.07	-0.18	0.65	1.00	
出行发生	0.35	0.85	-0.19	-0.14	0.19	0.91	-0.16	0.07	-0.08	-0.08	-0.08	1.00

3.2 模型建立

3.2.1 多元回归建模

针对变量间存在多重相关性,采用逐步回归法筛选变量是最常用的变量筛选方法。逐步回归法是采用边进边退的方法,对于模型外部的变量,只要可以提供显著的解释信息,就可以再次进入模型,对于已在模型内部的变量,只要其偏 F 检验不通过,则还可能从模型中删

除。采用SPSS 软件进行逐步回归建模,模型复相关系数为0.912。居住用地是交通发生量的重要因素,而其它对交通发生量有显著意义的变量已删除。

采用逐步回归建模的方法得出的交通发生量标准化模型为:交通发生量=0.912(居住用地)

3.2.2 PLSR 建模

根据交叉有效性检验来截取成分的个数,参数见表2。由表1 可见提取第1 个成分交叉有效性达到49.9%,提取第2 个成分时,累计交叉有效性为66.0%,提取第3 个成分时,累计交叉有效性减少为63.0%,说明第3 个成分的交叉有效性为负。根据公式(3),提取2 个主成分即可,其对交通发生量的复相关系数为87%。将主成分代入方程进行回归,得到土地类型和土地利用强度的系数,模型的标准化系数见表3。图1 为提取自变量集合中的第一个成分t 与因变量成分u 的散点图。从图中可知,提取一个成分时,自变量已与因变量存在很强的线性关系。

采用偏最小二乘法得到的标准化交通发生量模型为:

交通生成发生量=0.06(建筑密度)+0.415(容积率)-0.055(工业用地)-0.097(市政公用设施用地)-0.023(公共设施)+0.498(居住用地)-0.107(仓储用地)+0.083(道路与广场用地)-0.081(对外交通用地)+0.069(水域或其它用地)+0.079(绿地)。

表3 模型的标准化系数

Table 3 Standardization coefficients of model

指标	建筑密度	建筑容积率	工业用地	市政公用设施	公共设施	居住用地	仓储用地	道路广场用地	对外交通用地	水域和其它用地	绿地
系数	0.06	0.415	-0.055	-0.097	-0.023	0.498	-0.107	0.083	-0.081	0.069	0.079

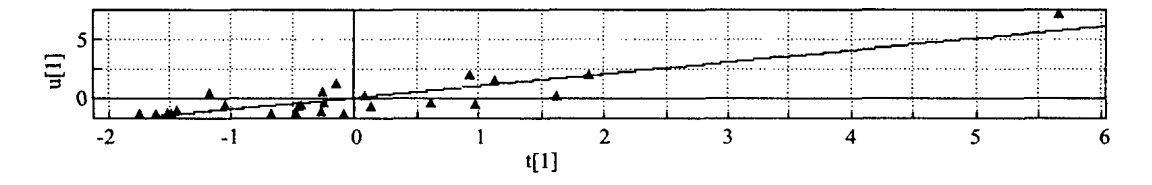


图1 t[1]与u[1]散点图

Fig.1 Dispel spot map of t[1] vs. u[1]

3.3 模型结果分析

由多元回归分析可知,采用逐步回归法建模,模型中把其它用地类型以及与交通发生量密切相关的建筑容积率,建筑密度等变量全剔除了,模型可靠性受到质疑。而采用偏最小二乘法建立的模型,不仅各用地类型以及建筑密度等土地利用强度指标保留,而且各变量系数的符号与实际情况大致相符的。比如建筑密度,建筑容积率和居住用地的系数不仅为正,而且居住用地的系数最大,表明这3 个变量与交通发生量是成正比例的,且居住用地与交通发

生量的关系最为密切,这在简单相关系数表里也得到了证实。将采用PLSR方法建模的样本点实测值

与模型拟合值进行对比,如图2所示。从图2可以看出,模型的整体精度很好。

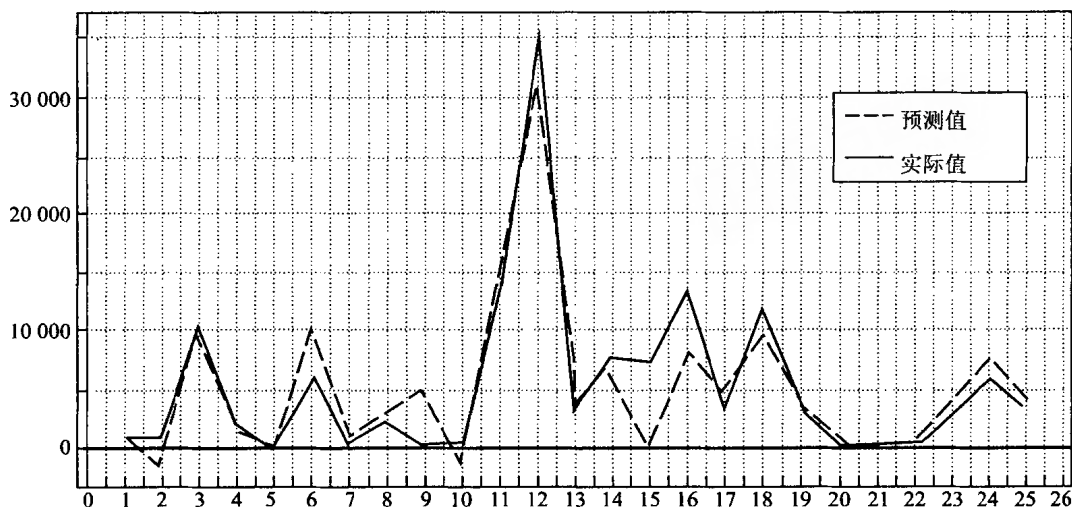


图2 样本点拟合图

Fig. 2 Fitting chart of swatch spots

4 结论

通过与普通多元回归分析建模的比较,表明偏最小二乘法在自变量强相关以及样本点很少的情况下建模的优势,且其模型精度是比较满意的,而且其变量系数的物理意义贴近实际,在简单相关系数表中也体现了这一点。模型中揭示了交通发生量与土地利用存在密切的关系,并对交通发生量与土地利用等变量进行了量化分析,为实现准确的交通需求预测以及后续计算奠定了基础。

由于实验点只有25个且变量之间存在多重相关性,采用偏最小二乘法建立的模型,物理意义明确,结论可靠,目前偏最小二乘法的应用已涉及很多领域。本文仅针对交通发生量模型进行了探讨,建立一套完整的交通生成模型还需进一步研究。

参考文献

- [1] 石飞,江薇,王伟,等.基于土地利用形态的交通生成预测理论方法研究[J].土木工程学报,2005,38(3):115-124.
- [2] 王丽,刘小明,任福田.城市土地开发交通生成预测方法的探讨[J].北京工业大学学报,2004,30(1):98-101.
- [3] 杨敏,陈学武,王伟,等.基于人口和土地利用的城市新区交通生成预测模型[J].东南大学学报(自然科学版),2005,35(5):815-819.
- [4] 杨明,曲大义,王伟.城市土地利用与交通需求相关关系模型研究[J].公路交通科技,2002,19(1):72-75.

- [5] 王松桂,陈敏,陈立萍. 线性回归与方差分析[M]. 北京:高等教育出版社,2004:1-10.
- [6] 韩悦臻,孔敏. 回归分析在交通生成预测中的应用[J]. 山西建筑,2007,33(20):297-298.
- [7] 王惠文,吴载斌,孟浩. 偏最小二乘回归的线性与非线性方法[M]. 北京:国防工业出版社,2006:97-112.
- [8] Wold,Herman. Partial Least Squares. In Samuel Kotz and Norman L[J]. Johnsoneds,Encyclopedia of Statistica Sciences,Vol. 6,New York:Wiley,1985:581-591.
- [9] 曾九孙,刘祥官,罗世华,等. 主成分回归和偏最小二乘法在高炉冶炼中的应用[J]. 浙江大学学报(理学版),2009,36(1):33-3.

Study on establishment of traffic generation model in combination with land-use

ZHANG Li,QI Hao-ping

(Transportation College of Southeast University,Nanjing,210096,China)

Abstract

There are serious multiple correlations of the urban land-use parameters which are the independent variables in the study of the relationships between land-use and traffic generation. The problem of multiple correlations of parameters is difficult to be solved by ordinary regression analysis methods. Taking the investigated parameters of traffic generation and land-use form of Changzhou city as a sample, this paper used partial least square regression (PLSR) and stepwise regression methods to establish the traffic generation model in combination with land-use form. Comparison between the two methods shows that the PLSR method is effective and reliable to solve the problem of multiple correlations of parameters and has clear physical meaning.

Key words:partial least square regression (PLSR); land-use; traffic generation; multiple regressions