

基于扩样技术和地理加权泊松回归模型的交通量估计

荆毅, 林航飞

(同济大学 道路与交通工程教育部重点实验室, 上海 201804)

摘要: 提出了扩样和地理加权泊松回归(GWPR)相结合的方法来估计有限观测值下的路网流量。首先,采用基于空间相似性的扩样方法对不平衡的观测流量进行纠正;然后,考虑道路几何特征和建成环境等因素的影响,采用地理加权泊松模型估计车道的小时交通量。结果表明,与传统线性回归模型和原始样本下的地理加权泊松模型相比,组合模型具有最佳的估计性能。此外,自变量与交通量关系的局部空间异质性也得到了很好的捕捉。

关键词: 交通流量估计;不均匀样本;地理加权泊松回归;扩样;悉尼协调自适应交通系统(SCATS)

中图分类号: U491.2

文献标志码: A

Estimating Traffic Volume Based on Sampling Expansion Technique and Geographically Weighted Poisson Regression

JING Yi, LIN Hangfei

(Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China)

Abstract: A method combining sampling expansion with geographically weighted Poisson regression (GWPR) was proposed to estimate the road network traffic volume with limited observation values. Firstly, a sampling expansion method based on the spatial similarity was employed to correct the imbalance missing data. Then, the GWPR was employed to estimate the hourly traffic volume of the lane considering the influence of the geometric characteristics of the road and the built environment. Results show that: compared with traditional linear models and GWPR with the original sample set, the proposed combination model has the best estimation performance. In addition, the local

spatial heterogeneity of the relationship between independent variables and traffic volume is also well captured.

Key words: traffic volume estimation; imbalanced sample; geographically weighted Poisson regression (GWPR); sampling expansion; Sydney coordinated adaptive traffic system(SCATS)

交通流量是交通工程中的基础数据。交通流量的用途包括但不限于交通设施的规划和建设、交通政策的设计和实施、道路碰撞风险的量化和交通污染的评估。正是由于道路交通量的重要性,地方政府投入了大量资金和精力,并采用不同的方法来采集交通流量数据。其中,自动流量检测器(例如线圈检测器和摄像机)是采集交通流量的主要设备。使用流量检测器面临的一个重要问题是存在大量缺失值^[1]。首先,出于成本考虑,流量检测器只布设于少数主要道路上,而绝大多数的低等级道路由于没有安装检测器而没有数据。第二,频繁的道路重建,可能会破坏埋在道路下的线圈检测器,导致数据缺失。第三,传输错误和设备故障也可能导致数据丢失。因此,利用有限的交通流量数据来估算整个路网的交通量就显得更加实际和重要。

在以往的研究中,不同研究人员用不同的方法来估计交通量。基于扩展因子的模型是一种传统的交通流量估计方法。在该模型中,根据交通量的变化模式对路段进行聚类,然后利用扩展因子将观测到的交通量扩展到同一个聚类中没有交通量的路段。该方法简单易行,在世界范围内被广泛采用^[2]。多元线性回归(MLR)是另一种常用的交通量估计

收稿日期: 2019-12-13

基金项目: 国家自然科学基金(71734004)

第一作者: 荆毅(1982—),男,工学博士,主要研究方向为交通运输规划与管理。E-mail: jingyi@tongji.edu.cn



论文
拓展
介绍

技术。在MLR中,最常用的自变量可分为道路设计特征、社会人口统计特征和土地利用特征^[3-4]。此外,还融入了一些基于空间位置的地理特征,例如,Morley等^[5]提出了一种由最短路径的通过次数确定的道路重要度特征来预测路网的年平均日交通量(AADT)。随着数据量的快速增长,近年来机器学习算法如神经网络^[6]、支持向量机回归^[7]和基于树的模型^[8]也被用来估计流量。虽然机器学习算法能提供比统计方法更高的精度和效率,但由于其黑箱特性,机器学习结果解释性并不高。此外,越来越多的研究表明,区域内的交通活动之间存在空间相关性,即流量不仅与研究路段的特征相关,而且与相邻路段上的流量相关。因此,能捕捉这种空间相关性的空间模型可以提高流量估计的精度,这些空间模型包括空间插值^[9-11]、空间回归模型^[12]和地理加权回归模型^[12-14]。

地理加权回归模型是一种可以捕捉研究区域内因变量和自变量变化关系的回归方法。与许多全局回归模型不同,地理加权回归模型是为每个研究单元构建一套模型参数^[15]。根据因变量分布的不同分为地理加权高斯模型、地理加权泊松模型(GWPR)和地理加权Probit模型等。其中,地理加权泊松模型更适合于对计数数据(如交通量)的估计^[12-13]。

使用空间模型的一个问题是道路网络中可用数据的不平衡分布。由于在估计路段交通量时,还考虑了附近的交通量,因此附近可用数据较少的道路估计性能较差。Eom等^[9]采用空间插值算法对北卡罗来纳州非高速公路道路的交通量进行预测,发现在流量检测器密度较高的区域估计性能更好。解决此问题的一种方法是对原始样本数据重新采样以平衡其分布。一些研究证明,在建立空间模型之前采用扩样方法可以提高模型的性能。例如,Tajmayer等^[16]在使用机器学习模型预测交通量之前,先使用基于扩展因子的方法扩大了样本量。Chen等^[17]采用合成少数过采样技术(SMOTE)对非主干路上的交通量进行扩样,结果表明,SMOTE有助于纠正样本的不平衡,并显著提高模型的估计性能。因此,在建立空间模型之前解决不平衡观测值问题以提高估计精度是有必要的。

本研究提出采用扩样与地理加权泊松回归相结合的方法来对路网缺失流量进行流量估计。流量数据来源于上海市杨浦区安装的悉尼协调自适应交通系统(SCATS),共计2 217条车道,其中只有560条车道有可用流量。首先,采用基于空间相似性的扩

样方法来纠正缺失流量的空间不平衡。然后,基于扩样后数据利用地理加权泊松模型估计路网中所有车道的交通流量。最后,为了验证组合模型的性能,采用10倍的交叉验证方法比较了3种不同的模型,即MLR模型、GWPR模型以及扩样与GWPR组合模型。本文研究结果为在观测值有限且不平衡的情况下提高流量估计的精度提供了一种可行方案。

1 方法

本研究的步骤如下:首先,采用Moran's I 指数来检验不同交叉口的交通量之间是否存在空间相关性。第二,采用扩样技术对空间分布不平衡的流量样本进行校正。第三,在扩样样本的基础上,采用GWPR方法提取交通流量与各自变量之间的关系。最后,采用10倍交叉验证法验证扩样和GWPR组合模型的有效性。

1.1 Moran's I 指数

Moran's I 指数广泛用于检验线性回归的残差之间是否存在空间自相关^[18]。本研究采用R语言中的“spdep”包来计算。Moran's I 指数的范围在-1到1之间,表示不同的空间相关类型。

(1) $0 < I \leq 1$: 正相关,在空间上表现为高值(低值)样本与高值(低值)样本聚集。

(2) $-1 \leq I < 0$: 负相关,在空间上表现为高值样本与低值样本聚集。

(3) $I = 0$: 不相关,在空间上表现为高值或低值样本随机分布。

1.2 基于空间相似性的扩样方法

本文提出了一种新的基于空间相似性的扩样方法来扩充流量样本。首先,定义空间相似性指标 d_{ij} ,如公式(1)所示;然后,对于缺失流量的车道 i ,计算车道 i 与其他有流量车道的空间相似性;最后,将空间相似性最高(最小的 d_{ij})对应车道的流量作为车道 i 的流量。

车道 i 和车道 j 的空间相似性定义为

$$S_{ij} = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2 + \sum_k (F_{ik} - F_{jk})^2} \quad (1)$$

式中: S_{ij} 为空间相似性; X_i 为车道 i 终点的 X 坐标; Y_i 为车道 i 终点的 Y 坐标; F_{ik} 为车道 i 的第 k 个特征; F_{jk} 为车道 j 的第 k 个特征。考虑到车道特征的不同尺度,在计算空间相似性之前,需要对所有特征进行min-max规范化。

1.3 地理加权泊松回归 (GWPR)

GWPR采用由距离定义的权重矩阵来处理空间自相关问题。在估计中,样本附近的观测值具有更高的权重。GWPR为每一个样本估计一组局部参数。本研究使用R语言中的“spgwr”包对GWPR模型的进行估计。GWPR模型定义如下:

$$\ln y_i = \beta_{0i} + \sum_{k=1}^p x_{ki} \beta_{ki} + \epsilon_i \quad (2)$$

式中: y_i 为样本*i*的因变量向量($i=1, 2, \dots, n$, n 为样本总数); x_{ki} 为样本*i*的第*k*个特征向量; β_{0i} 为截距项; β_{ki} 为相应系数; p 为特征(自变量)总数; ϵ_i 为样本*i*的误差项。

样本*i*的估计系数 $\hat{\beta}_i$ 可由下式得到:

$$\hat{\beta}_i = (X^T W_i X)^{-1} X^T W_i Y \quad (3)$$

式中: W_i 为样本*i*的空间权重矩阵,由公式(4)和(5)计算得到; Y 为因变量向量; X 为特征矩阵。

本研究采用高斯函数计算空间权重矩阵,它将权重作为距离的连续函数来计算,以确保有足够的局部观测值来估计模型^[19]。

$$W_i = \begin{bmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{in} \end{bmatrix} \quad (4)$$

$$w_{ij} = \exp \left[- \left(\frac{D_{ij}}{\theta} \right)^2 \right] \quad (5)$$

式中: D_{ij} 为样本*i*与样本*j*的欧式距离; n 为样本总数; θ 为带宽。

2 数据描述

2.1 研究区域和数据来源

本研究以位于上海中心城区东北部的杨浦区为研究区域。杨浦区土地面积60.73 km²,到2016年底拥有130.94万居民^[20]。当地政府从1986年开始引入悉尼协调自适应交通系统(SCATS),对重要交叉口的交通量进行监控和协调信号时间。SCATS依靠埋在进口车道的线圈检测器实时采集交通量,并根据采集的交通量自动调整交叉口的信号相位,以达到整个交叉口的最佳通行效率^[21]。

本研究的路网如图1所示。地面道路全长308.0 km,其中主干道占14%,次干道占23%,支路占63%。此外,该区还有20.5 km的快速路。车道流量来自于2016年10月11日的SCATS流量数据。

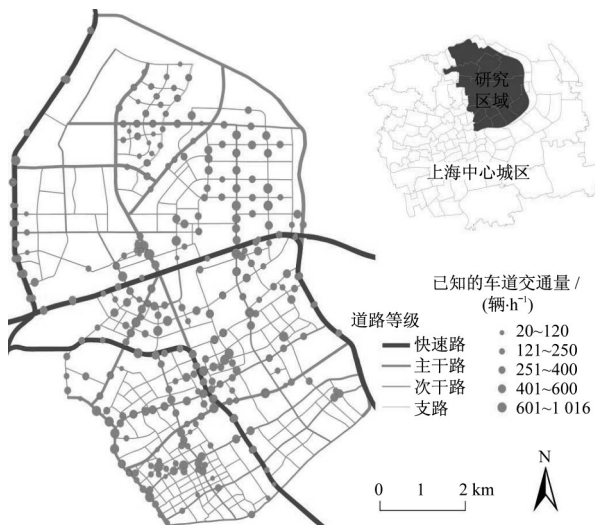


图1 研究区域和已知流量分布

Fig. 1 Study area and known traffic volume distribution

研究区共有2 217条车道,但只有560条具有完整的流量。如图1所示,具有已知流量的车道主要分布在等级较高的路段,这是合理的,因为在路网中高等级道路更有可能获得较好的维护。另外在路网中的一些区域完全没有已知流量,可知已知流量的空间分布极不平衡。

2.2 变量

本研究以早晨7:00—8:00车道小时交通量为因变量,以道路几何特征和建成环境特征为自变量。

在道路几何特征中,所有车道方向变量都是哑变量,其中直行方向为参考方向。其他道路几何特征包括路段长度、路段等级、下游道路等级、路段车道数量、公交线路数量、是否单行道、是否公交专用道、是否与快速路平行,以及上下游是否连接快速路。受数据条件的限制,建成环境特征包括车道附近的人口密度、平均房价、到最近三甲医院的距离、到最近购物中心的距离以及到最近大学的距离。变量选择使用MLR进行,统计上不显著以及方差膨胀因子(VIF)大于5.0的变量^[4-5]将被去掉。所有最终使用变量汇总见表1。

3 结果

3.1 扩样结果

使用第1.2节中提出的扩样方法,将原始数据集从560车道扩大到2 217车道(整个路网)。如表1所示,扩样前后两个样本集的同变量的均值和标准差相近。另外,一些变量,如车道数、公交线路数、

表1 自变量和因变量描述

Tab. 1 Description of dependent variables and independent variables

变量	变量名	描述	原始样本 (N=560)				扩样后样本 (N=2 217)			
			均值	标准差	最小值	最大值	均值	标准差	最小值	最大值
因变量	小时流量/(辆·h ⁻¹)	早晨7:00到8:00的车道小时流量	270.53	173.99	20.00	1 016	258.30	163.72	20.00	1 016
	右转	若车道前进方向为右转则为1,否则为0	0.08		0	1.00	0.11		0	1.00
	左转	若车道前进方向为左转则为1,否则为0	0.25		0	1.00	0.19		0	1.00
	直右	若车道前进方向为右转和直行则为1,否则为0	0.16		0	1.00	0.20		0	1.00
	左右	若车道前进方向为右转和左转则为1,否则为0	0.03		0	1.00	0.05		0	1.00
	车道数量/条	车道所在路段的车道总数	2.89	1.35	1.00	7.00	2.44	1.35	1.00	7.00
	公交线路数量/条	车道所在路段通过的公交线路总数	5.41	4.93	0	19.00	3.87	4.34	0	19.00
	公交专用道	若车道为公交专用道则为1,否则为0	0.03		0	1.00	0.03		0	1.00
	下游连接快速路	若车道下游路段有快速路进口则为1,否则为0	0.04		0	1.00	0.03		0	1.00
	上游连接快速路	若车道上游路段有快速路出口则为1,否则为0	0.03		0	1.00	0.02		0	1.00
	下游道路等级	车道下游路段的等级	1.99		1.00	3.00	2.15		0	3.00
	与快速路平行	若车道所在路段与快速路平行则为1,否则为0	0.18		0	1.00	0.10		0	1.00
	建成	人口密度/(万人·km ⁻²)	2.55	1.46	0.31	5.00	2.53	1.32	0.31	5.00
	环境	房价/(万元·m ⁻²)	5.17	1.57	1.17	13.19	4.91	1.96	0.94	25.45
		最近医院距离/km	1.59	0.78	0.16	4.56	1.70	0.82	0.16	5.08

注:上海市公交专用道在工作日上午7:00至10:00和下午16:00至19:00时段内限制非公交车辆通行。

人口密度和房价,在扩样后呈下降趋势。这是合理的,因为原始流量数据主要来自于较高等级道路的检测器,这些路段的变量,如车道数、公交线路数、人口密度等相对较大,当样本集扩大到整个路网时,即样本集中等级较低道路的比例增加时,这些自变量的均值随之减小^[17]。

3.2 空间相关性检验

Moran’s I 检验用于检验不同车道上小时交通量的空间相关性。原始样本的 Moran’s I 指数为 0.055 6, P 值为 0.000 6。扩样后的样本, Moran’s I 指数为 0.147 4, P 值为 0。Moran’s I 检验表明车道小时交通流量之间存在显著的正空间相关性。也就是说,在空间上高流量车道与高流量车道聚集,低流量车道与低流量车道聚集。因此,利用空间模型来处理空间相关性是非常必要的。

3.3 回归结果

本研究分别对 3 个模型进行了拟合和比较,它们分别是使用原始样本的 MLR(模型 1),使用原始样本的 GWPR(模型 2),和使用扩样后样本的 GWPR(模型 3),结果见表 2。

对于模型 1,所有方向相关变量与小时流量呈负相关(参考方向为直行)。其他变量,包括下游道路

等级、路段车道数、最近医院距离、是否为公交专用道以及是否与快速路平行,也都与流量呈负相关。相反,公交线路数、附近的房价、附近的人口密度以及上下游是否连接快速路都呈现出正相关关系。

对于模型 2 和模型 3,每条车道都有自己的回归系数。在表 2 中,系数由均值、最小值、最大值、正系数比例以及负系数比例来描述。如表 2 所示,3 个模型中同一变量的系数(或系数均值)的符号相同,表明该变量对流量的整体影响在所有模型中一致。然而,当聚焦于局部区域时会观察到差异。模型 3 在正系数比例和负系数比例中表现出最大的变化性,表明其捕捉自变量局部效应的能力最佳。与此相反,模型 2 在很大程度上与模型 1 一致,呈现较少的局部变化。如图 2 所示,左图表示原始的已知车道小时流量,右图表示由模型 3 估计的车道小时流量,对比可知,估计的车道流量与已知的原始流量高度一致。

3.4 模型验证

为了比较模型的估计性能,本文进行了 10 倍的交叉验证,并采用了 R 平方值(R²)、均方根误差(RMSE)和绝对百分比平均误差(MAPE)3 种指标来衡量对验证结果。如表 3 所示,模型 3 的 RMSE 和

表 2 模型 1、模型 2 和模型 3 的回归结果

Tab. 2 Regression results of model 1, model 2 and model 3

变量名	模型 1		模型 2					模型 3				
	系数	P 值	均值	最小值	最大值	正系数/%	负系数/%	均值	最小值	最大值	正系数/%	负系数/%
右转	-161.64***	0	-0.67	-0.95	-0.46	0	100	-0.64	-2.10	-0.05	0	100
左转	-131.09***	0	-0.54	-0.60	-0.37	0	100	-0.64	-1.36	-0.17	0	100
直右	-42.54**	0.02	-0.14	-0.19	0.02	1	99	-0.09	-0.73	0.51	39	61
左右	-138.62***	0	-0.45	-0.66	-0.06	0	100	-0.04	-1.00	2.15	44	56
车道数量	-29.56***	0	-0.11	-0.15	-0.05	0	100	-0.02	-0.23	0.39	37	63
公交线路数量	7.47***	0	0.02	0.01	0.04	100	0	0.03	-0.04	0.16	91	9
公交专用道	-183.19***	0	-0.84	-1.04	-0.63	0	100	-1.06	-1.78	0.01	0	100
人口密度	12.26**	0.02	0.06	0.01	0.15	100	0	0.04	-0.38	0.26	78	22
房价	8.82**	0.03	0.03	0.01	0.04	100	0	0.01	-0.09	0.44	51	49
下游连接快速路	115.97***	0.00	0.38	0.23	0.60	100	0	0.37	-0.15	1.29	85	15
上游连接快速路	81.25**	0.03	0.32	0	0.65	99	1	0.24	-1.29	1.11	78	22
与快速路平行	-72.11***	0.00	-0.33	-0.43	-0.19	0	100	-0.36	-1.09	0.50	10	90
最近医院距离	-20.57**	0.03	-0.05	-0.12	0.01	5	95	-0.02	-0.46	0.29	48	52
下游道路等级	-22.55**	0.01	-0.10	-0.19	-0.06	0	100	-0.05	-0.52	0.15	29	71
R^2	0.359				0.424					0.566		

注:***表示显著性水平为0;**表示显著性水平为0.001。

表 3 模型 1、模型 2 和模型 3 的交叉验证结果

Tab. 3 Cross validation results of model 1, model 2 and model 3

模型	模型 1	模型 2	模型 3
R^2	0.359	0.424	0.566
RMSE	145.186	145.049	128.651
MAPE/%	77.8	74.7	65.1

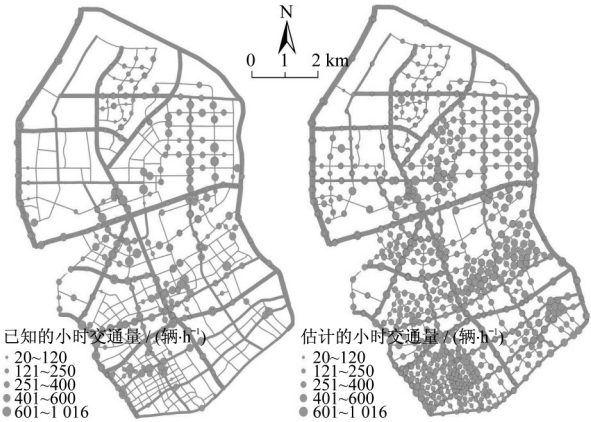


图 2 已知车道交通量和估计车道交通量

Fig.2 Known lane traffic volume and estimated lane traffic volume

MAPE 最小, 并且 R^2 最大。具体来说, 与模型 1 和模型 2 相比, 模型 3 的 RMSE 分别降低了 11.4% 和 11.3%, MAPE 分别降低了 16.3% 和 12.9%, R^2 分别提高了 57.7% 和 33.5%。

综上所述, 扩样技术和 GWPR 的联合使用不仅提高了模型的解释能力, 而且提高了预测精度。值

得注意的是, 使用原始样本的 GWPR 性能与 MLR 相近, 即对于空间分布不均衡或大量缺失数据的样本而言, GWPR 本身并不会提高模型的性能。这也再次验证拟合模型前对不平衡的缺失数据的处理是有意义的。

4 讨论

除了高精度外, 高解释性也是选择 GWPR 的原因。如表 2 中所示, 3 个模型的同一变量系数(或系数平均值)的符号一致。换言之, 在不同的模型中, 变量对交通量的全局影响的方向是一致的。然而, 由于模型 3 中各系数正负值所占比例的变化最大, 因此采用扩样样本的 GWPR 具有最大的空间异质性捕捉能力。在 GWPR 中, 所有自变量的估计系数和 P 值在空间上都是不同的, 这为解释自变量对流量的局部影响提供了更多的信息^[13,16,23]。考虑到篇幅限制, 本文只讨论几个信息量最大的自变量, 包括右转、车道数、是否公交专用道、人口密度, 共 4 个变量。如图 3 所示, 在每个子图中, 左图表示变量的系数值, 点的大小随系数的增大而增大, 正系数设为深灰色, 负系数设为浅灰色; 右图表示变量系数的显著性, 统计上显著的系数(即 P 值 < 0.1) 设置为浅灰色, 反之设置为深灰色。如图 3 所示, 系数和系数显著性都在研究区域内呈现出变化趋势。

对于与车道方向变量, 其中直行车道作为基本

参考。右转变量与研究区域的交通量始终呈负相关,并且在大多数车道上统计显著(P 值 <0.1)。左转变量与右转变量相似。表明右转和左转车道的交通量均低于直行车道,这是符合实际的,因为在大多数交叉口,直行车道流量都大于转弯车道流量。

在63%的车道中,路段车道数与交通量呈负相关,即路段车道数越多,每条车道的交通量越低。值得注意的是,这一结论似乎与一些现有研究不一致^[5,10-11,16]。原因在于现有研究的因变量是路段交通量,而在本研究中是车道交通量。路段上的更多车道对应更大的交通量,但是在某一路段上,由于更多的车道共享路段交通量,每条车道的分享交通量反而会减少。此外,仍有37%的车道与交通量呈正相

关,主要分布在研究区西南部。该区域靠近上海市中心,交通需求较大可能是导致与交通量正相关的主要原因。

另一个与交通量呈全局负相关的道路几何特征是公交专用道变量,表明公交专用道的设置显著降低了车道流量。人口密度与大多数车道(78%)的交通量呈正相关,这与之前的各种研究一致^[4,12]。然而,仍有一些地方,人口密度呈负相关。其中一些在统计上不显著,而另一些主要位于主干道处。一种解释是,主干道上的大多数交通都是过境交通,这与当地社会人口结构(如人口密度)没有密切关系^[19]。

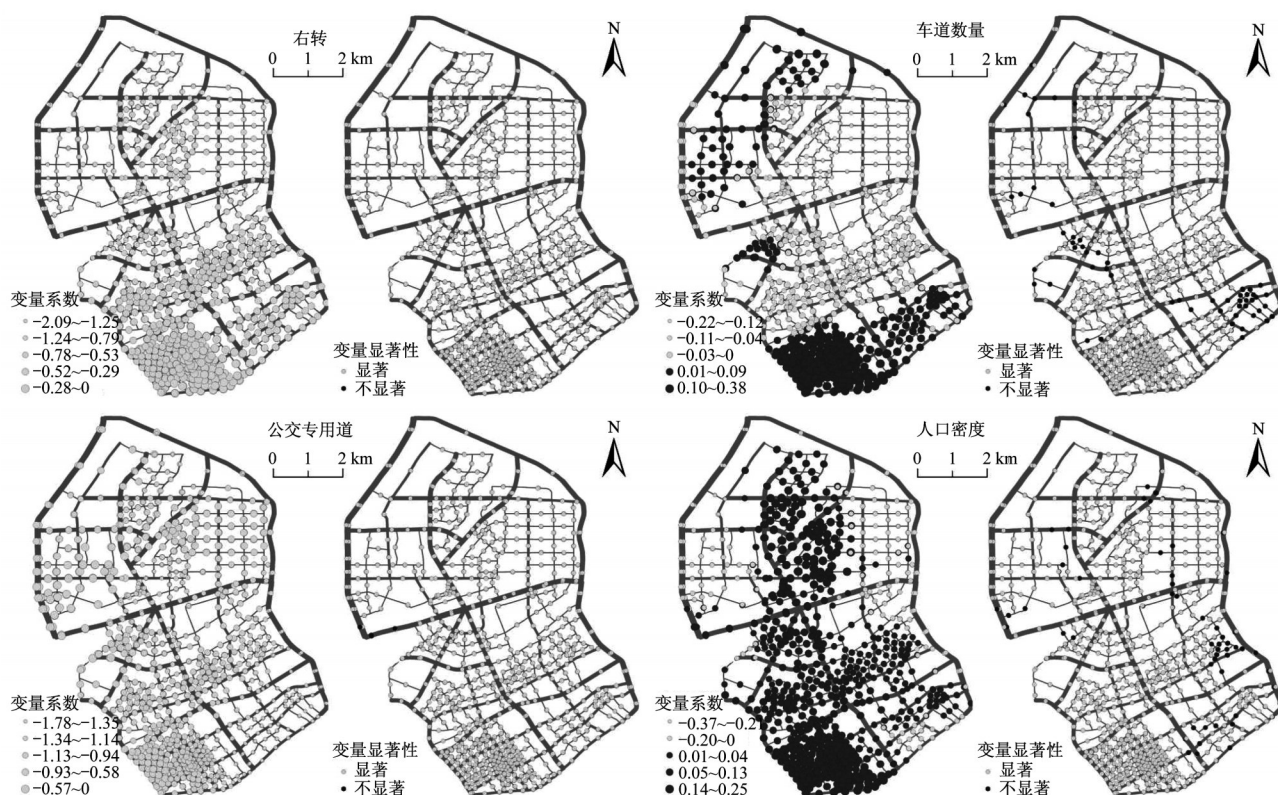


图3 部分变量系数值与显著性的空间分布

Fig. 3 Spatial distribution of the values and significance of some variables

5 结论

本研究为观测值有限且分布不均时的交通流量估计提供了一种新的方法。首先用扩样技术填充不平衡缺失数据,然后用地理加权回归模型(GWPR)估计车道交通量。结果表明,与传统的MLR或直接使用GWPR相比,采样扩展技术和GWPR结合的方

法具有更好的交通量估计性能。此外,GWPR可以很好地捕捉自变量与交通量关系的空间异质性,为解释自变量的影响提供更多的细节。

值得进一步研究的是,首先,由于数据条件的限制,本研究未能包含足够的土地利用和社会人口特征,例如岗位密度和土地利用等。这些因素都与交通量关系密切。若将它们纳入到模型中可以进一

步提高模型拟合优度和精度。此外,大量新的数据源如手机数据、GPS数据等也在不断涌现,可以将这些新数据源与传统SCATS数据集成,以获得更高的流量估计性能。

参考文献:

- [1] XIAO X, CHEN Y S, YUAN Y. Estimation of missing flow at junctions using control plan and floating car data[C]// 18th Euro Working Group on Transportation. Delft: Ewgt, 2015: 113-123.
- [2] Federal Highway Administration. Traffic monitoring guide[R]. Washington D C: Federal Highway Administration, 2012.
- [3] ANDERSON M D, SHARFI K, GHOLSTON S E. Direct demand forecasting model for small urban communities using multiple linear regression[J]. Transportation Research Record, 2006, 1981: 114.
- [4] ZHAO F, CHUNG S. Contributing factors of annual average daily traffic in a Florida county - exploration with geographic information system and regression models[J]. Transportation Research Record, 2001, 1769: 113.
- [5] MORLEY D W, GULLIVER J. Methods to improve traffic flow and noise exposure estimation on minor roads [J]. Environmental Pollution, 2016, 216: 746.
- [6] SEKULA P, MARKOVIC N, VANDER LAAN Z, *et al.* Estimating historical hourly traffic volumes via machine learning and vehicle probe data: a Maryland case study [J]. Transportation Research Part C-Emerging Technologies, 2018, 97: 147.
- [7] CASTRO-NETO M, JEONG Y, JEONG M K, *et al.* AADT prediction using support vector regression with data-dependent parameters [J]. Expert Systems with Applications, 2009, 36(2): 2979.
- [8] WU J Q, XU H. Annual average daily traffic prediction model for minor roads at intersections[J]. Journal of Transportation Engineering Part A-Systems, 2019, 145(10): 100.
- [9] EOM J K, PARK M S, HEO T Y, *et al.* Improving the prediction of annual average daily traffic for nonfreeway facilities by applying a spatial statistical method [J]. Artificial Intelligence and Advanced Computing Applications, 2006 (1968): 20.
- [10] SELBY B, KOCKELMAN K M. Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression [J]. Journal of Transport Geography, 2013, 29: 24.
- [11] WANG X K, KOCKELMAN K M. Forecasting network data spatial interpolation of traffic counts from texas data [J]. Transportation Research Record, 2009, 2105: 100.
- [12] YU H T, PENG Z R. Exploring the spatial variation of ridesourcing demand and its relationship to built environment and socioeconomic factors with the geographically weighted Poisson regression[J]. Journal of Transport Geography, 2019, 75: 147.
- [13] BRUNSDON C, FOTHERINGHAM A S, CHARLTON M E. Geographically weighted regression: a method for exploring spatial nonstationarity [J]. Geographical Analysis, 1996, 28 (4): 281.
- [14] YANG H T, LU X Z, CHERRY C, *et al.* Spatial variations in active mode trip volume at intersections: a local analysis utilizing geographically weighted regression [J]. Journal of Transport Geography, 2017, 64: 184.
- [15] ZHAO F, PARK N. Using geographically weighted regression models to estimate annual average daily traffic [J]. Transportation Research Record, 2004(1879): 99.
- [16] TAJMAJER T, SPLAWINSKA M, WASILEWSKI P, *et al.* Predicting annual average daily highway traffic from large data and very few measurements[C]// 2016 Ieee International Conference on Big Data. Washington D C: [s.n.], 2016: 1493-1501.
- [17] CHEN P, HU S H, SHEN Q, *et al.* Estimating traffic volume for local streets with imbalanced data [J]. Transportation Research Record, 2019, 2673: 598.
- [18] LESAGE J, PACE R K. Introduction to spatial econometrics [M]. Boca Raton: Chapman and Hall, 2009.
- [19] WANG C H, CHEN N. A geographically weighted regression approach to investigating the spatially varied built-environment effects on community opportunity [J]. Journal of Transport Geography, 2017, 62: 136.
- [20] 杨浦区统计局. 2017年杨浦统计年鉴[R]. 上海: 上海杨浦区统计局, 2017.
Yangpu District Bureau of Statistics. 2017 Yangpu district statistical yearbook[R]. Shanghai: Yangpu District Bureau of Statistics, 2017.
- [21] SCATS. Why choose SCATS? [EB/OL]. [2019-9-15]. <https://www.scats.com.au/why-choose-scats-performance.html>.