

基于轨迹大数据的城市交通状态识别研究

吴若楠, 冯慧芳

西北师范大学数学与统计学院, 甘肃 兰州
Email: 3181706916@qq.com, hffeng@nwnu.edu.cn

收稿日期: 2021年6月21日; 录用日期: 2021年7月26日; 发布日期: 2021年8月4日

摘要

对城市交通状态的有效识别与评估有助于提高城市交通精细化管理水平和应急响应能力。以轨迹大数据为基础, 提出基于多个交通要素的城市交通状态识别模型。首先建立基于路段平均速度、交通流密度、行驶时间指数、拥堵指数和车道占有率的交通状态识别的综合指标体系, 并以交通时空大数据为基础, 结合主成分分析方法构建综合指标体系中各个指标的客观权重。进一步根据交通状态的模糊性, 建立基于核模糊C均值聚类的交通状态识别模型。最后以兰州市出租车GPS数据为基础, 对兰州市的城市交通状态进行识别和划分, 分析了兰州市的交通状态的时空演化特征。实验结果表明, 兰州市交通状态分为5个等级, 城市交通整体上呈现拥堵状态, 且拥堵路段空间分布有较大差异, 集聚特征比较明显。

关键词

数据挖掘, 城市交通, 交通状态识别, 核模糊C均值聚类算法, 轨迹大数据

Research on Urban Traffic State Identification Based on Trajectory Big Data

Ruonan Wu, Huifang Feng

College of Mathematics and Statistics, Northwest Normal University, Lanzhou Gansu
Email: 3181706916@qq.com, hffeng@nwnu.edu.cn

Received: Jun. 21st, 2021; accepted: Jul. 26th, 2021; published: Aug. 4th, 2021

Abstract

The effective identification and evaluation of urban traffic state are helpful to improve the fine management level and emergency response capability of urban traffic. Based on trajectory big data, a traffic state recognition model based on multiple traffic elements is proposed. Firstly, a comprehensive index system for traffic state identification based on average speed of road sections, traffic

flow density, travel time index, congestion index and lane occupancy rate is established. Based on trajectory big data, combined with principal component analysis, the objective weight of each index in the comprehensive index system is constructed. Then, according to the fuzziness of traffic state, a traffic state recognition model based on kernel-based fuzzy C-means clustering is established. Finally, the urban traffic state of Lanzhou city is identified and divided based on the taxi GPS data of Lanzhou city, and the temporal and spatial evolution characteristics of the traffic state of Lanzhou city are analyzed. The experimental results show that the traffic state in Lanzhou is divided into five grades. The urban traffic is congested during all working hours. There are great differences in the spatial distribution of congested road sections, and the aggregative characteristics are obvious.

Keywords

Data Mining, Urban Traffic, Traffic State Identification, Kernel-Based Fuzzy C-Means Clustering Algorithm, Trajectory Big Data

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着城市化进程的加快和经济的飞速发展,城市人口不断增长,伴随着城市交通量也急剧增加,这给城市交通系统带来了巨大的压力。城市道路容量和交通流量的不平衡引起了严重的交通拥堵,严重影响了城市的发展和居民的日常生活[1]。交通拥堵会造成能源消耗、环境污染和降低城市运营效率,解决交通问题是政府和人民面临的重要挑战。实时、准确地识别与评估城市交通状态能够科学管理城市交通系统和引导居民出行。

近年来,交通拥堵问题引起了许多国内外学者的兴趣,他们对此已经做了大量的研究工作,这些研究工作分为理论研究和应用研究,理论研究主要包括交通状态评价指标的研究,应用研究主要是交通状态识别及交通流量预测。国内外学者针对识别交通状态和评价交通系统的交通评价指标做了多方面的探究。Anjaneyulu 等研究了各类速度指标与交通流量之间的关系并选择了车速变异系数作为拥堵指标识别交通拥堵情况[2]。Litman 等分析了影响交通拥堵成本的各种因素,总结了包括服务水平、行程时间、平均速度、平均通勤时间、拥堵持续时间、延误时间等交通拥堵评价指标,提出了各种减少拥堵的策略[3]。黄艳国等根据城市路网拓扑,通过仿真实验,验证了交通状态判别指数模型的有效性[4]。另外,世界各国根据自己国家的道路交通状态,制定了不同的交通状态评价体系。美国的许多大城市面临严重的交通拥堵问题,基于道路服务水平,美国道路状态评价指标根据平均行程速度、车辆流量与通行能力的比值以及负荷系数等三个指标分为6种[5]。我国制定了《城市交通运行状态评价规范》,采用基础性指标(平均交通流量、平均行程速度等)、特征性指标(行程时间比、延误时间比等)及综合性指标(拥堵里程、行程时间等)等作为交通运行状态的评价指标[6]。

准确识别城市的交通状态是缓解和解决交通拥堵的前提。大数据、云计算、物联网等计算机技术的迅速发展,为我们识别交通状态提供了新思路和新途径。An 等提出了一种基于网格级拥堵检测方法,该方法包括拥堵检测、反复拥堵检测及反复拥堵的演化过程等三个步骤,并应用哈尔滨市实地采集的GPS轨迹数据对该方法的有效性进行了验证[7]。何兆成等构建道路状态判断模型,提出城市区域交通状态模式识别与分类方法,以广州市出租车数据为例,研究分析了城市交通状态分布特性[8]。基于GPS轨迹数

据、道路网络和 POI 数据, Wang 等提出了一种三个阶段的预测框架, 以探讨路段之间的拥堵相关性[9]。研究结果表明道路拥堵具有明显的方向性和传递性。熊励等构建了基于神经网络的城市拥堵识别模型, 以上海部分区域的交通数据集为例验证了模型的有效性[10]。Guo 等根据路网拓扑和交通流量建立动态有向加权复杂网络模型, 提出了基于社团划分的城市交通网络动态拥堵预警方法, 并实证研究了济南市交通网络[11]。以城市多源数据为基础, Song 等结合 K-均值聚类 and 城市地理信息挖掘城市交通拥堵的时空模式, 确定影响交通拥堵的关键因素, 其研究结果能够识别潜在交通拥堵路段, 并有助于指导城市土地使用的规划[12]。以淮安市的出租车轨迹数据为基础, 黄子赫等提出了基于密度聚类方法的道路拥堵识别算法, 通过出租车运行情况识别城市道路拥堵状态, 并对拥堵程度进行区分[13]。

然而, 在现有交通状态识别问题研究中, 绝大部分研究都是从车流量、路段平均速度、延误时间、行程时间等单一要素出发, 没有综合考虑多个相关要素对城市交通拥堵的影响, 使得对城市交通状态的识别结果不能准确反映其交通状态。另外, 出租车 GPS 轨迹数据包含大量的连续型时空轨迹数据集, 它是城市大数据的重要组成部分, 为我们研究城市交通网络动态特征提供了有效数据。为了更准确地识别城市交通拥堵状态, 本文以车辆 GPS 大数据为基础, 结合主成分分析和核模糊 C 均值聚类方法, 构建基于多个交通要素的城市交通状态识别模型, 并对兰州市交通状态进行实证研究。

2. 综合交通状态评价指标体系建立

2.1. 交通状态评价指标的选择

在文献[2] [3] [4] [5] [6]中讨论了多种交通状态评价指标。如果仅采用一种指标进行交通状态评估, 结果可能存在一些偏差, 但若选择的指标参数太多, 则会增加计算复杂度并且难以平衡指标间的关系。选取评价指标遵循原则是用较少的指标将城市交通状态较为全面、精确地刻画和反映出来。基于这一原则, 本文选择路段平均速度、交通流密度、行驶时间指数、拥堵指数和车道占有率等五个指标综合评价城市交通状态。

2.1.1. 路段平均速度

路段平均速度是指给定时间段内, 某路段上所有行驶车辆的速度平均值。

$$v_j = \frac{\sum_{i=1}^N v_i^{(j)}}{N} \quad (1)$$

其中, $v_i^{(j)}$ 是指给定时间段内第 j 条道路路段上行驶的第 i 辆车; N 表示该时间段内第 j 条道路路段上所有车辆总数。

2.1.2. 交通流密度

交通流密度是指给定时间段内, 某路段单位长度上的车量数。

$$D_j = \frac{Vol^{(j)}}{L^{(j)}} \quad (2)$$

其中, $Vol^{(j)}$ 表示某时间段内第 j 条道路路段上的车流量数; $L^{(j)}$ 是指第 j 条道路路段的路段长度。

2.1.3. 行驶时间指数

行驶时间指数表示为给定时间段内路段上车辆实际行驶时间与自由流行驶时间的比值。

$$TTI_j = \frac{T_R^{(j)}}{T_F^{(j)}} \quad (3)$$

其中, $T_R^{(j)}$ 表示第 j 条道路路段车辆实际行驶所需时间; $T_F^{(j)}$ 表示第 j 个路段在自由流条件下所需行驶时间。

2.1.4. 拥堵指数

拥堵指数是车辆行驶延误时间与自由流行驶时间的比值:

$$CI_j = \frac{T_R^{(j)} - T_F^{(j)}}{T_F^{(j)}} \times 100\% \quad (4)$$

其中, $T_R^{(j)}$ 表示第 j 条道路路段车辆实际行驶所需时间; $T_F^{(j)}$ 表示第 j 个路段在自由流条件下所需行驶时间。

2.1.5. 车道占有率

车道占有率指给定时间段内所有车辆在某路段上的实际行驶时间之和与给定时间的百分比率:

$$OC_j = \frac{\sum_{i=1}^N t_i^{(j)}}{T_o} \quad (5)$$

其中, $t_i^{(j)}$ 表示在观测时间段内第 j 条道路路段上第 i 辆车的行驶时间; N 表示在观测时间段内通过第 j 条路段的车辆总数; T_o 表示观测时间段。

2.2. 指标权重的构建

基于上述五个交通拥堵评价指标, 我们建立综合交通状态评价指标体系, 但这些指标在评价模型中的重要程度不同, 通过权重反映了各指标在多属性评价模型中所起的作用。因此, 我们需要确定各类指标在评价模型中的权重。与主观权重法相比较, 基于主成分分析的客观赋权法[14]克服了主观赋权法在指标权重选择上的主观因素影响, 该方法确定的权重客观反映了指标间的真实关系, 既能体现出评价指标的相对重要性, 也能反映出指标间的差异性。本文采用该方法确定交通拥堵状态识别模型中各指标权重。基于主成分分析的客观赋权方法的具体步骤为:

1) 给定数据集 $X = \{x_1, x_2, \dots, x_n\}$, 其中 $x_i = \{x_{i1}, x_{i2}, \dots, x_{il}\}, i = 1, 2, \dots, n$, n 为样本总数, l 为样本的特征个数。建立数据集 X 的相关系数矩阵 $R = (r_{uv})_{l \times l}, r_{uv} = r_{vu}$ 。计算相关系数矩阵 R 的特征值 $\lambda_u, u = 1, 2, \dots, l$ 以及各特征值对应的特征向量 $e_u = (e_{u1}, e_{u2}, \dots, e_{ul})^T, u = 1, 2, \dots, l$ 。主成分载荷系数矩阵表示为 $(e'_1, e'_2, \dots, e'_l)$ 。

2) 计算各个主成分的贡献率 $\alpha_u = \frac{\lambda_u}{\sum_{c=1}^l \lambda_c}, u = 1, 2, \dots, l$ 和累积贡献率 $\alpha_{(u)} = \frac{\sum_{c=1}^u \lambda_c}{\sum_{c=1}^l \lambda_c}, u = 1, 2, \dots, l$ 。

3) 选取累积贡献率 $\geq 85\%$ 的前 p 个主成分。计算各指标在选取的 p 个主成分组合中的系数: $\phi_u = \frac{e'_u}{\sqrt{\lambda_u}}, u = 1, 2, \dots, p$, 其中 $\phi_u = (\phi_{u1}, \phi_{u2}, \dots, \phi_{ul})^T$ 。

4) 计算各指标权重: $(w'_1, w'_2, \dots, w'_l)^T = \frac{\phi_1 \times \alpha_1 + \phi_2 \times \alpha_2 + \dots + \phi_p \times \alpha_p}{\alpha_{(p)}}$, 然后对指标权重归一化得:

$$W = \{w_1, w_2, \dots, w_l\}, \text{ 其中 } w_u = \frac{w'_u}{\sum_{c=1}^l w'_c}。$$

2.3. 基于核模糊 C 均值聚类的交通状态识别模型

模糊聚类是以模糊理论为基础对数据进行分类的一种无监督机器学习算法。核模糊 C 均值聚类算法

[15] (kernel-based fuzzy C-means clustering algorithm, KFCM)应用最广泛模糊聚类算法之一, 它基于隶属度和聚类中心建立目标函数, 通过对目标函数的优化迭代, 计算聚类中心和隶属度, 直到目标函数收敛迭代停止, 样本数据分类完成。该算法通过核函数将原始空间中的点映射到高维特征空间中, 将原本线性不可分的数据变成线性可分或近似线性可分, 能够克服了 FCM 算法的不足之处。

给定数据集 $X = \{x_1, x_2, \dots, x_n\}$, 设 k 为分类数目, $m_j (j=1, 2, \dots, k)$ 为每个聚类的中心, $\mu_j(x_i)$ 是第 i 个样本对应第 j 类的隶属度函数, b 为指数权重因子, 将 x_i 通过非线性特征映射到高维线性空间, 则在高维特征空间下的目标函数和约束条件为:

$$J_{\text{KFCM}} = \sum_{j=1}^k \sum_{i=1}^n [\mu_j(x_i)]^b \|\psi(x_i) - \psi(m_j)\|^2 \tag{6}$$

$$\sum_{j=1}^k \mu_j(x_i) = 1, i = 1, 2, \dots, n \tag{7}$$

其中, $\psi(x_i)$ 和 $\psi(m_j)$ 分别表示 x_i 与 m_j 在高维空间中的像:

$$\begin{aligned} & \|\psi(x_i) - \psi(m_j)\|^2 \\ &= \psi(x_i) \cdot \psi(x_j) + \psi(m_i) \cdot \psi(m_j) - 2\psi(x_i) \cdot \psi(m_j) \\ &= K(x_i, x_j) + K(m_i, m_j) - 2K(x_i, m_j) \end{aligned} \tag{8}$$

高斯核函数由于其所对应的线性空间是任意维数, 映射的问题线性可解, 被应用的最为广泛, 其定义为:

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{\sigma^2}\right) \tag{9}$$

将(9)代入(8)得:

$$\|\psi(x_i) - \psi(m_j)\|^2 = 2(1 - K(x_i, m_j)) \tag{10}$$

从而得到 KFCM 更进一步的目标函数, 即:

$$J_{\text{KFCM}} = 2 \sum_{j=1}^k \sum_{i=1}^n [\mu_j(x_i)]^b [1 - K(x_i, m_j)] \tag{11}$$

根据核函数特征空间与内积的关系式可推导出 KFCM 算法的聚类中心和隶属度函数公式为:

$$m_j = \frac{\sum_{i=1}^n [\mu_j(x_i)]^b K(x_i, m_j) x_i}{\sum_{i=1}^n [\mu_j(x_i)]^b K(x_i, m_j)} \tag{12}$$

$$\mu_j(x_i) = \frac{(1 - K(x_i, m_j))^{-1/(b-1)}}{\sum_{s=1}^k (1 - K(x_i, m_s))^{-1/(b-1)}} \tag{13}$$

通过迭代不断求解(12)式和(13)式, 直至公式(11)收敛, 得到最优解。

模糊 C 均值聚类的缺点是聚类个数人为确定, 这导致确定聚类个数具有较强的主观性, 为了避免这个问题, 本文采用吸引力传播聚类算法[16] (Affinity Propagation Clustering, AP)确定最佳聚类个数。AP 算法通过计算数据点之间的两种信息, 即吸引信息和归属感信息, 并不断迭代更新吸引信息矩阵和归属

度信息矩阵, 直到选出 N 个聚类中心且结果保持不变, 并确定每个数据点与聚类中心的归属关系后算法结束。该算法能自动算出最佳聚类数而不需要在运算之前确定聚类的个数, 因此是计算最佳聚类数目的常用方法之一。本文采用 AP 算法确定最佳聚类个数。

Table 1. Traffic state recognition model based on kernel-based fuzzy C-means clustering

表 1. 基于核模糊 C 均值聚类的交通状态识别模型

输入: 规范化的数据集 \mathbf{X} , 指数权重因子 b , 收敛精度 ε , 高斯核函数参数 σ
输出: 交通网络各路段交通状态的等级
Step 1. 根据主成分客观赋权法确定各个评价指标的权重 $w_j, j=1,2,\dots,l$ 。
Step 2. 根据 AP 算法确定最佳聚类个数 k 。
Step 3. 用核模糊 C 均值聚类算法初始化聚类中心集 $\mathbf{m}_j^{(0)}$, 初始化迭代次数 t , 令 $t=0$;
Step 4. 根据聚类中心集 $\mathbf{m}_j^{(t)}$, 由式(13)计算出隶属度矩阵 $\mu_j(\mathbf{x}_i)^{(t)}$;
Step 5. 根据 $\mathbf{m}_j^{(t)}$, $\mu_j(\mathbf{x}_i)^{(t)}$, 由式(11)计算出 $J_{\text{KFCM}}^{(t)}$;
Step 6. 根据隶属度矩阵 $\mu_j(\mathbf{x}_i)^{(t)}$, 由式(12)计算出新的聚类中心集 $\mathbf{m}_j^{(t+1)}$;
Step 7. 根据聚类中心集 $\mathbf{m}_j^{(t+1)}$, 由式(13)计算出新的隶属度矩阵 $\mu_j(\mathbf{x}_i)^{(t+1)}$;
Step 8. 根据 $\mathbf{m}_j^{(t+1)}$ 、 $\mu_j(\mathbf{x}_i)^{(t+1)}$, 由式(11)计算出新的目标函数 $J_{\text{KFCM}}^{(t+1)}$ 值;
Step 9. 判断 $\ J_{\text{KFCM}}^{(t+1)} - J_{\text{KFCM}}^{(t)}\ < \varepsilon$ 是否成立, 若是, 终止运算, 转至 Step 10; 否则, $t=t+1$ 并转到 Step 6 继续执行;
Step 10. 得到聚类中心集为 $\mathbf{m}_j = (m_{j1}, m_{j2}, \dots, m_{jl})$, 其中 $j=1,2,\dots,k$;
Step 11. 计算综合评价指标 M_j , 根据 M_j 的值确定交通状态的等级;
Step 12. 将识别的交通状态与城市路段一一对应, 得到各路段状态的等级。

本文基于核模糊 C 均值聚类的交通拥堵识别算法的基本步骤为: 1) 根据数据集, 采用主成分分析客观赋权法确定各个评价指标的权重; 2) 对样本数据进行核模糊 C 均值聚类并得到聚类结果; 3) 定义并计算综合评级指标 $M_j = m_{j1} \times w_1 + m_{j2} \times w_2 + \dots + m_{jl} \times w_l$, 其中 $w_j, j=1,2,\dots,l$ 为指标权重, $\mathbf{m}_j = (m_{j1}, m_{j2}, \dots, m_{jl})$, $j=1,2,\dots,k$ 为聚类中心; 4) 根据 M_j 的值确定城市交通状态的等级。具体算法如表 1 所示。

3. 实验结果及分析

3.1. 研究区域

兰州市是甘肃省的省会城市, 地处中国西北, 是典型的带状组团式结构, 城区主要坐落于河谷地狭长的地带内。受南北两山限制, 交通拥堵在这个河谷型城市变得越来越严重。兰州是 2019 年中国十大拥堵城市之一[17]。本文以兰州市主要城区——城关区、七里河区、安宁区和西固区这四个区域为研究对象, 该区域中含有 227 条道路, 1022 个路段。

3.2. 数据集

本文数据集是 2017 年 3 月 6 日到 2017 年 3 月 12 日兰州市出租车的 GPS 轨迹数据。该数据集是连续 7 天的数据, 具有一定的代表性, 恰好体现了工作日和休息日的城市交通状态和居民出行规律。GPS 轨迹数据主要包含车辆 ID、经度、纬度、瞬时速度、记录时间、行驶方向和车辆载客状态等信息, 轨迹点采样间隔时间最大为 30 s。

在使用前对原始 GPS 数据需预处理, 首先, 清理原始 GPS 数据中的冗余值, 并通过地图匹配识别方法, 剔除区域边界外和不在路网上的离群点和异常值, 应用线性插值方法对缺失值进行补充; 然后, 通过 MNTG [18] (Minnesota Traffic Generator) 获得兰州市主城区路网拓扑信息; 最后, 采用几何特性的方法

对出租车轨迹数据和路网拓扑进行匹配, 获得有效的数据集。

3.3. 城市整体交通状态评估

由于各个交通状态评价指标数据量纲不同, 在建模之前, 我们先对其进行标准化处理来统一这些交通指标量纲。假设某交通评价指标 $S = \{s_i, i=1, 2, \dots, n\}$, $S_{\max} = \max_i S$, $S_{\min} = \min_i S$, 标准化公式为 $s'_i = (s_i - S_{\min}) / (S_{\max} - S_{\min})$, 经过数据标准化后, 每个交通评价指标都是具有相同尺度的无量纲量。数据分析之后, 如果需要评价指标的真实值, 可采用反标准化方法得到, 反标准化公式为 $s_i = s'_i (S_{\max} - S_{\min}) + S_{\min}$ 。本文中在进行核模糊 C 均值聚类算法的过程中, 我们设定参数取值分别为指数权重因子 $b = 2$, 收敛精度 $\varepsilon = 10^{-5}$, 高斯核函数参数 $\sigma = 2$ 。

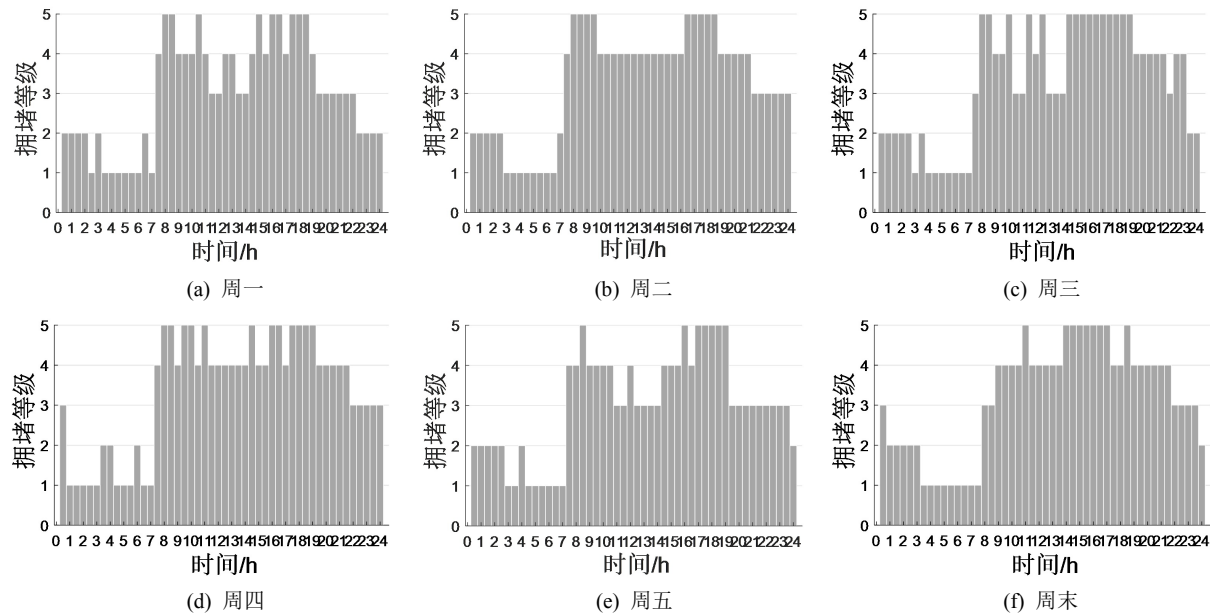


Figure 1. The overall traffic status of Lanzhou on weekdays and weekend

图 1. 兰州市工作日和周末的整体交通状态

我们以 30 分钟为时间间隔, 对周一到周日每天 24 小时的交通基础数据进行统计, 计算城市路网每个路段的各个交通状态评价指标均值, 建立城市交通状态识别模型, 本文在实验部分采用 AP 算法, 分别对兰州市周内连续五个工作日和周末的交通数据集进行最佳聚类个数确定, 我们发现最佳聚类个数基本分为五类, 只有周二的数据集的最佳聚类个数划分为四类, 再结合我国《城市交通运行状况评价规范》[6], 最终将兰州市交通状态划分为五个等级, 这五个等级分别对应畅通(1 级)、基本畅通(2 级)、轻度拥堵(3 级)、中度拥堵(4 级)和严重拥堵(5 级), 五个等级的综合评级指标值对应区间分别为 $(-0.16, -0.07]$, $(-0.07, -0.02]$, $(-0.02, 0.01]$, $(0.01, 0.09]$ 和 $(0.09, 0.2]$ 。图 1 为兰州市周内连续五个工作日和周末的交通状态。

根据图 1 可知, 在 7:30~22:00 时间段内, 城市拥堵状态转换频繁, 但路段基本处于拥堵状态, 而且中度拥堵和严重拥堵状态较多, 说明城市道路在一天内交通流量都比较大, 该结果与兰州是中国十大拥堵城市之一的结论一致。另外, 工作日和周末城市交通状态有区别, 工作日的 7:30~8:30, 17:00~19:00 时段拥堵级别为严重拥堵, 说明这两个时间段路段拥堵状况最为严重, 与工作日的早晚上下班、上学放学的高峰期契合, 另外, 工作日的交通拥堵状态相比周末拥堵状态更严重。和工作日相比较, 周末的交通高峰时段大约是 10:30~11:00 和 13:30~19:00。这是由于周末上班族和学生休息, 故没有明显早高峰,

居民出门较晚, 故第一个高峰时段是 10:30~11:00, 第二个高峰时段是午饭之后, 居民出门逛街娱乐, 即 13:30~18:30。实验结果与城市居民的出行特征相符。

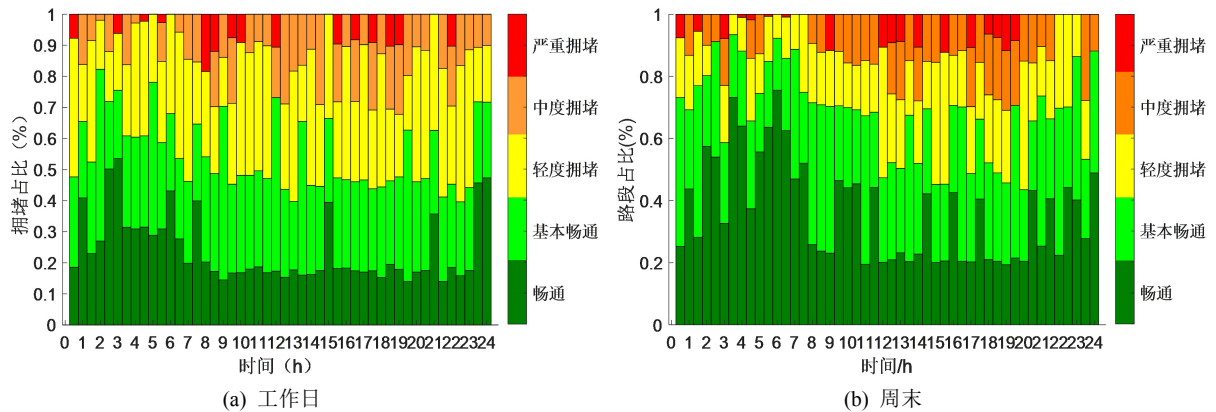


Figure 2. Proportion of road sections with different congestion levels within 24 hours during the weekdays and weekend
图 2. 工作日和周末 24 小时各拥堵等级路段占比

由图 1 可知, 城市交通状态在工作日呈现相似状态, 但工作日和周末的交通状态明显不同, 我们分别计算各个路段工作日和周末的平均交通统计量, 用识别算法识别各路段的交通状态, 分析不同时间段内呈现五种交通状态的路段的比例。实验结果如图 2 所示。

由图 2(a)可知, 工作日 7:30~9:00 和 12:00~20:00 的时段内, 严重拥堵和中度拥堵路段的比例较大, 接近 30%。比较图 2(a)和图 2(b), 我们发现和工作日相比, 周末畅道路段比例较大, 发生中度以上拥堵的时间段有差异, 与图 1 结果一致。工作日和周末的畅通和基本畅道路段的比例大约都是 50%, 这个和工作日的比较接近。不论是工作日还是周末, 从早晨 7:00 到晚上 23:00, 轻度、中度和严重拥堵比例之和都接近 50%, 这再次说明兰州市确实为全国十大堵城之一。

3.4. 路网交通状态动态演化及可视化

通过 48 个交通状态快照图, 可观察城市各个路段 24 小时的交通状态随时间的演化特征。由于篇幅所限, 我们分别选择工作日和周末的几个典型时段做分析。

图 3 工作日六个时段(5:30~6:00, 6:00~6:30, 7:30~8:00, 8:00~8:30, 18:00~18:30, 18:30~19:00)的交通状态的可视化, 其中 7:30~8:30 和 18:30~19:00 对应工作日的早高峰和晚高峰时段。由图 3(a) (b)与图 3(c) (f)比较可知, 5:30~6:00 和 6:00~6:30 道路交通状态良好, 部分路段呈现轻度拥堵。在早高峰和晚高峰时段, 中度拥堵和严重拥堵的道路路段激增, 许多区域出现交通拥堵现象。从区域来看, 四个城区中, 城关区拥堵路段最多, 其次为七里河区, 安宁区交通状态最好。城关区主要交通拥堵区域西起解放门立交桥, 东至瑞德大道, 在早高峰和晚高峰时段, 这些区域内的道路大部分都是中度拥堵和严重拥堵。该结果与兰州市城市功能区分布特征一致。兰州市的中心区是城关区。城关区是甘肃省的政治中心和文化中心, 是党、政、军机关驻地, 兰州市主要商圈都分布在城关区中心地带。另外, 该区域常驻人口密度大, 使得该区域交通流量比较大。在七里河区, 西起柳家营什字, 东至解放门立交桥的西津路是主要拥堵路段, 该路段经过了西站什字、兰州中心、小西湖等主要商圈。安宁区在高峰期仅有个别路段拥堵, 这些路段基本都位于中小学周边, 拥堵时段基本与学校上学、放学时间同步。另外, 安宁区是科教文化中心, 区域内 19 所高等学府和科研院所, 该区域没有大型商圈、大中型企业和政府职能部门, 常驻人口中学生占一半, 所以, 人口流动较小, 区域内交通状态比较好。

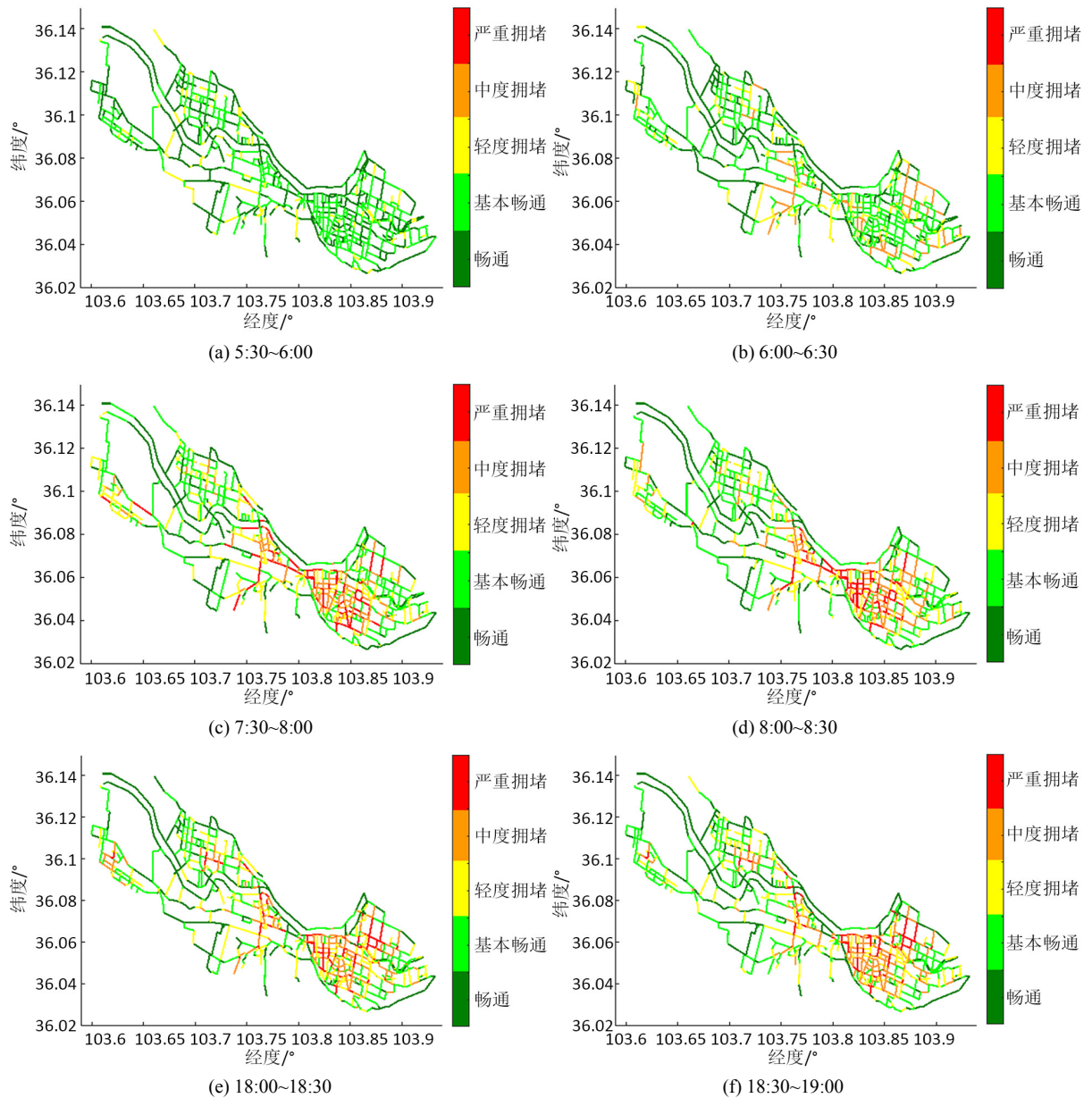


Figure 3. Visualization of traffic status at different periods of working days

图 3. 工作日不同时段交通状态可视化

图 4 周末四个时段(7:30~8:00, 8:00~8:30, 18:00~18:30, 18:30~19:00)的交通状态的可视化。显然, 7:30~8:00 和 8:00~8:30 两个时段并非周末的高峰期, 交通状态良好, 部分道路轻度拥堵, 仅有少量道路中度拥堵。18:00~18:30 和 18:30~19:00 是周末的交通高峰期, 中度拥堵和严重拥堵的路段较多, 但是, 和工作日相比较, 周末高峰期的交通状态略好。

4. 结论

本文建立了基于多指标的交通状态评价体系, 以城市交通大数据为基础, 采用主成分分析赋权方法确定各个指标的客观权重, 用于刻画每个指标对交通状态的影响力。根据交通状态的模糊特性, 采用核模糊 C 均值聚类算法对交通状态进行分类识别, 并以兰州市出租车 GPS 数据为基础, 实证研究了兰州市

的城市交通状态的分布特征和演化特征。研究表明,兰州市交通状态分为 5 个等级,在 7:30~22:00 的时段内,城市交通整体上呈现拥堵状态,拥堵程度与居民出行特征相符合。拥堵路段空间分布有较大差异,集聚特征比较明显,经常性拥堵路段出现在白银路、麦积山路、甘南路、南滨河路、中山路、平凉路,晏家坪路以及西津路等。

城市交通状态及演化规律研究,能够提供城市路网中拥堵路段的时空分布信息,可为居民出行提供指引,为城市交通管理和有效控制提供科学依据。但本文仍存在一些不足,出租车仅是城市众多交通方式的一种,难以全面反映城市整体交通状态。今后考虑增加地铁、公交以及共享单车等多源交通大数据,用于挖掘更具普适性、更深层次的城市交通的时空动态演化规律。

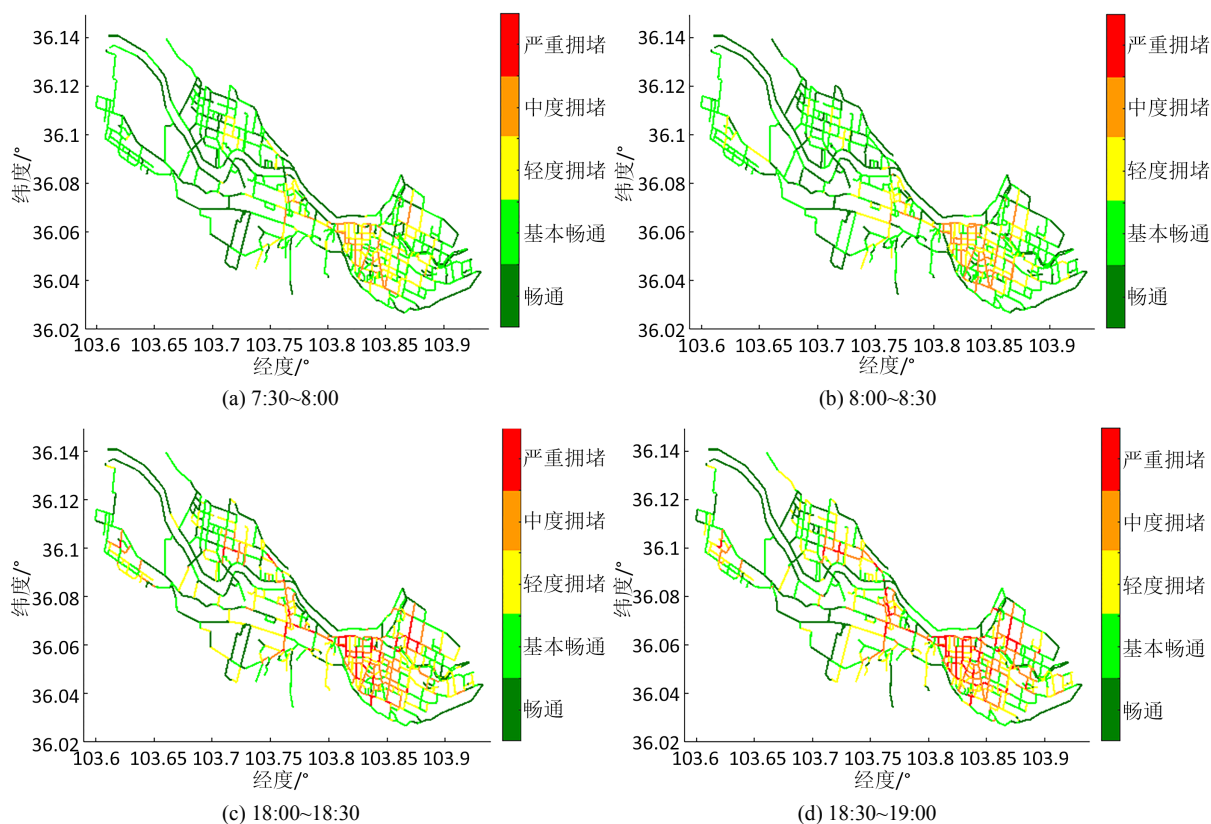


Figure 4. Visualization of traffic status at different periods of weekend

图 4. 周末不同时段交通状态可视化

基金项目

本论文得到国家自然科学基金(No. 71761031)资助。

参考文献

- [1] Rudolph, F. and Matrai, T. (2018) Congestion from a Multimodal Perspective. *Periodica Polytechnica Transportation Engineering*, **46**, 215-221.
- [2] Anjaneyulu, M.V.L.R. and Nagaraj, B.N. (2009) Modeling Congestion on Urban Roads Using Speed Profile Data. *Journal of the Indian Roads Congress*, **1**, 65-74.
- [3] Litman, T. (2014) Congestion Evaluation Best Practices. *International Transportation Economic Development Conference*, Dallas, USA, 9-11 April 2014, 1-20.
- [4] 黄艳国, 宋二猛, 钟建新. 城市区域路网交通状态分析与评价方法[J]. 重庆交通大学学报(自然科学版), 2017,

- 36(12): 91-96.
- [5] TRB and National Research Council (2000) Transportation Research Board. Highway Capacity Manual. Washington DC.
- [6] 北京交通发展研究中心. GB/T33171-2016 城市交通运行状况评价规范[S]. 北京: 中国标准出版社, 2016.
- [7] An, S., Yang, H.Q., Wang, J., Cui, N. and Cui, J.X. (2016) Mining Urban Recurrent Congestion Evolution Patterns from GPS-Equipped Vehicle Mobility Data. *Information Sciences*, **373**, 515-526. <https://doi.org/10.1016/j.ins.2016.06.033>
- [8] 何兆成, 周亚强, 余志. 基于数据可视化的区域交通状态特征评价方法[J]. 交通运输工程学报, 2016, 16(1): 133-140.
- [9] Wang, Y.Q., Cao, J.N., Li, W.G., Gu, T. and Shi, W.Z. (2017) Exploring Traffic Congestion Correlation From Multiple Data Sources. *Pervasive and Mobile Computing*, **41**, 470-483. <https://doi.org/10.1016/j.pmcj.2017.03.015>
- [10] 熊励, 杨淑芬, 张芸. 大数据背景下基于 5S 的城市交通拥堵评价模型研究[J]. 运筹与管理, 2018, 27(1): 117-124.
- [11] Guo, Y.J., Yang, L.C., Hao, S.X. and Gao, J. (2019) Dynamic Identification of Urban Traffic Congestion Warning Communities in Heterogeneous Networks. *Physica A: Statistical Mechanics and Its Applications*, **522**, 98-111. <https://doi.org/10.1016/j.physa.2019.01.139>
- [12] Song, J.C., Zhao, C.L., Zhong, S.P., Nielsen, T.A.S. and Prishchepov, A.V. (2019) Mapping Spatio-Temporal Patterns and Detecting the Factors of Traffic Congestion with Multi-Source Data Fusion and Mining Techniques. *Computers, Environment and Urban Systems*, **77**, Article ID: 101364. <https://doi.org/10.1016/j.compenvurbsys.2019.101364>
- [13] 黄子赫, 高尚兵, 蔡创新, 等. 多重处理的道路拥堵识别可视化融合分析[J]. 中国图象图形学报, 2020, 25(2): 409-418.
- [14] 颜惠琴, 牛万红, 韩惠丽. 基于主成分分析构建指标权重的客观赋权法[J]. 济南大学学报(自然科学版), 2017, 31(6): 519-523.
- [15] 万春圆, 叶明全, 姚传文, 等. 基于核模糊聚类优化算法的脑核磁共振图像分割研究[J]. 中国数字医学, 2020, 15(11): 10-15.
- [16] Frey, B.J. and Dueck, D. (2007) Clustering by Passing Messages between Data Points. *Science*, **315**, 972-976. <https://doi.org/10.1126/science.1136800>
- [17] 车轮快讯. 2019 最新数据: 全国最堵城市排名, 你的家乡上榜了吗? [EB/OL]. <http://news.bitauto.com/hao/wenzhang/30960711>, 2020-11-08.
- [18] Mokbel, M.F., Alarabi, L., Bao, J., Eldawy, A., Magdy, M., Sarwat, M., Waytas, E. and Yackel, S. (2013) MNTG: An Extensible Web-Based Traffic Generator. *International Symposium on Spatial and Temporal Databases*, Munich, 21-23 August 2013, 38-55. https://doi.org/10.1007/978-3-642-40235-7_3