

DOI:10.19651/j.cnki.emt.2519538

基于高斯加权动态时间规整的时间序列聚类

黄胜浩 黄文德

(广西大学计算机与电子信息学院 南宁 530004)

摘要: 针对动态时间规整及其传统加权变体在城市交通状态时序测量中存在的时段敏感度不足与权重分配机制不合理的问题,提出一种基于高斯函数加权的改进型动态时间规整方法,以提升轨迹相似性度量的判别精度。该方法通过构建时间依赖的多峰高斯权重函数,将城市交通流的拟高斯分布特性与高峰时段先验知识融入序列对齐过程,非线性地强化交通高峰时段在距离计算中的贡献度。实验以成都市出租车 GPS 轨迹数据为基础,构建以小时为时间粒度的空间网格出发点数量时间序列。结合 K-Medoids 聚类算法,引入轮廓系数和 Calinski-Harabasz 指数对 GS-WDTW、DTW 与 WDTW 的算法性能进行定量评估。结果表明,在最优聚类数 $K=4$ 时,GS-WDTW 的轮廓系数和 Calinski-Harabasz 指数分别相较于 DTW 提升了 39.9% 和 13.1%,相较于 WDTW 提升了 41.1% 和 13.0%,有效提升了出行模式时空特征的识别能力。空间分析验证了聚类结果与城市功能区分布的高度吻合性。GS-WDTW 方法能够更精准地捕捉时序数据中的关键非线性特征,为智慧交通系统的状态感知、资源优化和功能区识别提供参考。

关键词: GS-WDTW;出租车轨迹;时间序列聚类;K-Medoids

中图分类号: TP391.4;TN911.7 **文献标识码:** A **国家标准学科分类代码:** 520.60

Time series clustering based on Gaussian-weighted dynamic time warping

Huang Shenghao Huang Wende

(College of Computer and Electronic Information, Guangxi University, Nanning 530004, China)

Abstract: To address the limitations of dynamic time warping and its conventional weighted variants—namely, insufficient sensitivity to critical time periods and suboptimal weight allocation mechanisms—in the context of urban traffic state time-series measurement, this paper proposes an enhanced DTW method based on Gaussian function weighting, termed Gaussian-weighted DTW, to improve the discriminative accuracy of trajectory similarity measurement. The proposed approach constructs a time-dependent multi-peak Gaussian weighting function that integrates the quasi-Gaussian distribution characteristics of urban traffic flow and prior knowledge of peak-hour patterns into the sequence alignment process, thereby nonlinearly amplifying the contribution of peak traffic periods in distance computation. Experiments are conducted using taxi GPS trajectory data from Chengdu, from which hourly origin-count time series are constructed on a spatial grid basis. Coupled with the K-Medoids clustering algorithm, the performance of GS-WDTW is quantitatively evaluated against DTW and WDTW using the silhouette score and Calinski-Harabasz index. Results show that, at the optimal cluster number $K=4$, GS-WDTW achieves a 39.9% and 13.1% improvement over DTW, and a 41.1% and 13.0% improvement over WDTW, in terms of SS and CHI, respectively, demonstrating significantly enhanced capability in identifying spatiotemporal characteristics of travel patterns. Spatial analysis further confirms a high degree of consistency between the clustering results and the actual urban functional zone distribution. The GS-WDTW method thus enables more precise capture of critical nonlinear features in time-series data, offering a valuable reference for state perception, resource optimization, and functional zone identification in intelligent transportation systems.

Keywords: GS-WDTW; taxi trajectory; time-series clustering; K-Medoids

0 引言

交通状态的高精度时序测量是智慧城市感知系统的核

心环节。通过建立车辆、道路与控制系统之间的联系,多元化管理道路交通,能够低成本、高效率的缓解交通拥堵^[1]。随着物联网传感技术与边缘计算能力的提升,轨迹大数据

因其蕴含丰富的模式及规律^[2]成为动态交通监测的关键数据来源。此类数据通过车载 GPS 传感器实时采集,为交通流量建模、出行热点探测提供了高精度测量基础。

轨迹的相似性度量及聚类分析对时间序列数据的实时跟踪、预测与模式挖掘具有重要的意义^[3]。姚尧等^[4-6]将原始轨迹数据通过提取 OD(origin-destination)点,并映射为时间序列形式,利用时间序列相似性度量的方法进行识别以刻画特定区域内交通状态随时间的变化趋势。动态时间规整(dynamic time warping, DTW)作为一种经典的时间序列匹配算法能够有效处理序列长度不一致与局部时序错位的问题^[7],在轨迹聚类等任务中得到了广泛应用。然而,DTW 在计算序列对齐路径时,默认对所有时间点赋予均等权重。而城市交通活动具有显著的时段异质性特征,尤其是早晚高峰时段的交通模式往往对时序的识别具有决定性影响。这种均匀权重分配策略可能导致算法对关键交通特征的捕捉不足,进而影响测量准确性。近年来,针对 DTW 的改进研究主要集中在引入时间或距离相关的权重机制。早期研究者提出的加权动态时间规整(weighted dynamic time warping, WDTW)通过指数衰减函数对时间偏移施加惩罚,提升了分类性能,但其权重仅依赖于序列位置偏移量,未考虑实际应用场景中的语义重要性。后续研究尝试引入领域知识,如杨傲雷等^[8-11]提出基于固定窗口加权,但其权重函数仍为线性或分段常数形式,难以刻画交通流量的非线性波动特征。此外,现有方法普遍假设单一时段主导模式,无法有效建模城市交通中典型的“早晚双峰”结构。

值得注意的是,城市交通流量在高峰时段常呈现拟高斯分布特性,且全天分布多表现为双峰形态。受相关研究^[12-13]将高斯模型用于交通流特征提取的启发,本文提出一种基于高斯函数加权的动态时间规整方法(Gaussian-weighted dynamic time warping, GS-WDTW)。该方法通过构建时间依赖的多峰高斯权重核,融合交通高峰的先验知识与非线性权重分配机制,显著增强模型对关键时段的敏感性。实验基于规则网格划分城市空间,利用出租车上车点数据构建小时级出行时序曲线,并结合时序聚类与城市功能区的空间耦合分析,验证所提方法在出行模式识别与功能区划分中的有效性。

1 研究区域及数据

本研究使用的数据来源于 2014 年 8 月 3 日~8 月 30 日期间成都市约 14 000 辆出租车的 GPS 轨迹记录。数据集包含 5 个关键属性,如表 1 所示。为确保分析结果的代表性与稳定性,本文选取其中连续 5 个工作日的数据作为研究样本。

2 研究方法

首先,对原始出租车轨迹数据进行预处理。基于既有

表 1 出租车 GPS 轨迹数据格式

Table 1 GPS trajectory data format of taxis

字段	属性	示例
ID	出租车编号	10 001
Lon	经度位置坐标	30.624 72
Lat	纬度位置坐标	104.136 014
Status	车载状态	0
Time	时间戳	2014-8-19 19:06:06

研究^[14]对网格尺度的优化分析,将研究区域划分为 600 m×600 m 的规则网格,该网格尺度被证实能够在计算效率与空间分辨率之间取得良好平衡。然后,统计每个网格单元在不同时间段内的上车点数量,构建具有时空特性的网格化时间序列数据集。在此基础上,提出一种基于高斯函数加权的动态时间规整方法,以衡量不同网格间时间序列的相似性。该方法通过引入高斯权重函数,增强关键时段在序列匹配过程中的影响。进一步采用 K-Medoids 聚类算法^[15-16]对时间序列进行聚类分析,并引入轮廓系数(silhouette score)^[17]和 Calinski-Harabasz 指数(Calinski-Harabasz index, CH)^[18]对聚类结果进行量化评估。最终,利用成都市连续 5 天的出租车轨迹数据对所提方法进行实证验证,研究方法路线如图 1 所示。

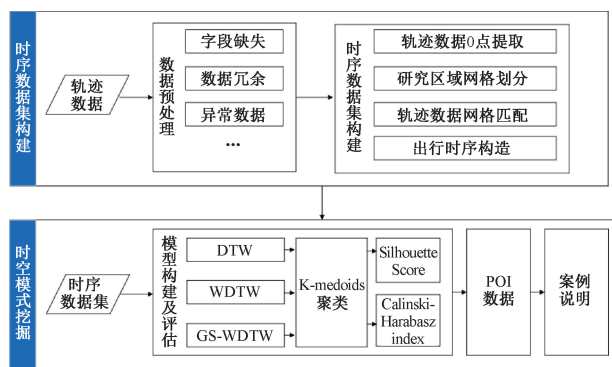


图 1 技术路线

Fig. 1 Technical roadmap

2.1 DTW

DTW 是一种经典的时间序列相似性度量方法,广泛用于交通流量分析。与传统的欧氏距离相比,DTW 能够有效处理时间序列在时间轴上的非线性形变。该方法通过动态规划计算两条序列之间的最优对齐路径,使序列在允许时间轴局部伸缩的条件下,累积距离达到最小。其计算过程主要包括构建距离矩阵、搜索最小累积距离路径和计算最终 DTW 距离 3 个步骤。路径搜索允许斜向移动以对齐相似模式,最终得到的距离越小则相似性越高。DTW 的优势在于能够处理时间偏移、对噪声和局部形变具有较强适应性。

设两个时间序列为:

$$S_1 = [s_1^{(1)}, s_2^{(1)}, \dots, s_n^{(1)}] \quad (1)$$

$$S_2 = [s_1^{(2)}, s_2^{(2)}, \dots, s_m^{(2)}] \quad (2)$$

式中: S_1 和 S_2 为两条待对齐的时间序列; $s_i^{(1)}$ 和 $s_j^{(2)}$ 分别为序列 S_1 和 S_2 在时间点 i 和 j 处的观测值; n 和 m 为序列长度。

定义局部代价函数为欧氏距离平方,即:

$$d_{DTW}(i, j) = (s_i^{(1)} - s_j^{(2)})^2 \quad (3)$$

式中: $d_{DTW}(i, j)$ 表示序列 S_1 中第 i 个点与 S_2 中第 j 个点之间的局部距离匹配代价。

构建累积代价矩阵 $D \in \mathbf{R}^{n \times m}$, 即:

$$D(i, j) = \begin{cases} \infty, & i = 0, j = 0 \\ d(i, j) + \min \begin{cases} D(i-1, j), \\ D(i, j-1), \\ D(i-1, j-1) \end{cases}, & \text{其他} \end{cases} \quad (4)$$

式中: $D(i, j)$ 表示从 $(1, 1)$ 到 (i, j) 的最小累积对齐代价; $\min\{D(i-1, j), D(i, j-1), D(i-1, j-1)\}$ 表示从上、左、左上 3 个方向转移的最小代价。

满足有界、连续且路径单调, 保证对齐稳定。边界条件为:

$$\begin{cases} D(0, 0) = d(0, 0) \\ D(i, 0) = \infty, i > 0 \\ D(0, j) = \infty, j > 0 \end{cases} \quad (5)$$

式中: $D(0, 0)$ 表示两条序列首元素之间的局部匹配代价; $D(i, 0)$ 和 $D(0, j)$ 分别为累积代价矩阵 D 在上边界和左边界的取值; ∞ 表示不可达路径的代价(禁止沿边界单向扩展)。

最终 DTW 距离定义为:

$$D_{DTW}(S_1, S_2) = \sqrt{D(n, m)} \quad (6)$$

2.2 WDTW

DTW 在序列对齐时忽略时序偏移代价差异, WDTW 通过距离自适应权重机制改进这一局限。其核心思想是对较大的时序偏移施加更强的惩罚, 具体实现如下:

定义序列对齐路径上任意两点 s_i 与 s_j 的基准偏移量为 $d_{base}(i, j) = |i - j|$, 通过逻辑函数转换权重, 即:

$$\omega_{WDTW} = \frac{1}{1 + e^{-g(d_{base}(i, j) - m)}} \quad (7)$$

式中: g 为权重衰减系数, m 为平均偏移量。

加权局部距离计算为:

$$d_{WDTW}(i, j) = \omega_{WDTW} \cdot (s_i^{(1)} - s_j^{(2)})^2 \quad (8)$$

累积距离矩阵更新规则与传统 DTW 相同(式(4)~(6))。

2.3 GS-WDTW

DTW 方法默认对所有时间点赋予均等权重, 而现有 WDTW 方法在交通时序分析中常采用线性或固定模式的权重分配, 二者均难以充分反映城市交通活动中关键时段(如早晚高峰)对整体出行模式的主导作用。为克服这一局限, 特别是针对交通高峰时段数据的高变异性和非线性波动问题, 本文引入高斯函数作为时间维度上的权重函数, 以

强化关键时段在序列对齐中的贡献。权重生成过程如下: 假设每个时间序列长度为 T , 则生成一个 $0 \sim (T-1)$ 的时间点序列 $t \in [0, T-1]$ 。

设定若干个高峰时段及其对应的方差, 构造多峰高斯函数生成时间点权重, 即:

$$\omega_t = 1 + \sum_{k=1}^K \exp\left(-\frac{(t - \mu_k)^2}{2\sigma_k^2}\right) \quad (9)$$

式中: t 表示时间点索引(以 h 为单位); K 表示高峰时段数量; μ_k 表示第 k 个高峰时间点; σ_k 表示第 k 个高峰影响范围的标准差; ω_t 表示时间点 t 的最终权重; 初始权重设为 1 以避免无高峰时段被忽视。

本文根据实证数据分析, 确定城市交通活动存在早晚两个显著高峰, 因此设置 $K = 2$ 。 μ_k 的值是基于平均出行量曲线中早晚两个最高峰的时间点。 σ_k 则决定了权重函数的非线性衰减速率, 控制了高峰影响的“宽度”或“持续时间”。通过将时间点 t 映射到高斯核的中心 μ_k 附近, GS-WDTW 赋予这些时间点较高的权重值, 从而非线性地、显著地放大了高峰时段数据的微小差异在最终距离度量中的贡献。这种非线性加权机制有效应对了高峰时段因出行量剧增导致的数据波动剧烈、形态非线性变化的挑战, 避免了 DTW 仅依赖欧氏距离被稀释的问题, 实现了对关键时段非线性特征的准确捕捉。权重计算通过高斯函数实现, 最终在 DTW 的累积距离计算中引入 $\omega_{GS-WDTW}$ 进行动态加权。

$$d_{GS-WDTW}(i, j) = \omega_{GS-WDTW} \cdot (s_i^{(1)} - s_j^{(2)})^2 \quad (10)$$

式中: i, j 分别为序列中当前对齐的时间点; 为避免因单侧权重过高而导致的过度拟合问题, $\omega_{GS-WDTW}$ 取时间点 i 和 j 的平均权重, 累积距离矩阵更新规则与 DTW 相同(式(4)~(6))。

2.4 K-Medoids 聚类

K-Medoids 是一种基于划分的经典聚类算法, 其核心思想是通过选择实际数据点作为簇中心, 将整个数据集划分为若干具有代表性的子集。与 K-Means 不同, K-Medoids 并不依赖于对所有数据点计算均值向量, 而是从数据集中选取最能代表某一簇特性的实际样本点作为中心点, 在每次迭代过程中最小化簇内对象到该 medoid 的总距离。这种基于真实数据点的聚类机制使 K-Medoids 在处理含有噪声、异常值或非线性结构的数据时展现出更强的鲁棒性与稳定性。尤其在数据分布复杂、对结果可解释性要求较高的应用场景中, 如城市功能区识别、交通模式分析等领域, K-Medoids 表现出优于传统 K-Means 的性能优势。此外, K-Medoids 算法支持多种类型的距离度量方式, 包括但不限于欧式距离、曼哈顿距离以及动态时间规整等, 这使其在时间序列聚类任务中具有更高的灵活性和适应性。尤其是在轨迹数据分析中, 结合 DTW 等非刚性匹配方法, K-Medoids 能够更准确地捕捉时间序列间的相似性特征。

2.5 评价指标

本文采用两种广泛应用于无监督学习领域的评估指标:轮廓系数和 Calinski-Harabasz 指数。选择这两种指标主要基于它们的互补性与鲁棒性。轮廓系数通过计算样本与其所在簇内其他样本的平均距离以及与最近邻簇的平均距离,综合衡量每个样本在当前聚类结构中的合理性。其基于距离的计算,用于评估聚类结果的内聚性和分离性,取值范围为 $[-1, 1]$,数值越高表示聚类效果越好。Calinski-Harabasz 指数则通过计算簇间离散度与簇内离散度的归一化比值来评估聚类质量,侧重于从方差角度衡量簇的紧凑性和分散度,数值越大表示聚类质量越高。这两种指标从几何距离和统计方差两个不同维度量化聚类结构的质量,避免了单一指标可能带来的片面性。此外,虽然也存在其他评估方法如戴维森-布尔丁指数(Davies-Boulding index, DBI)和邓恩指数(Dunn index, DI),但轮廓系数和 Calinski-Harabasz 指数因其计算效率和对数据分布的低依赖性,已被广泛证明在时间序列聚类评估中具有良好的可靠性。

3 实验与分析

本研究数据处理和算法实现均采用 Python 编程实现,使用了 NumPy, Pandas 等开源库,部分核心计算模块启用 Numba 技术进行加速。

3.1 高斯权重函数建模

城市交通出行通常在工作日期间表现出较强的规律性和稳定性。本文选取连续五个工作日的出行数据作为研究样本,绘制了上车点维度的平均载客量随时间变化的趋势曲线,如图 2 所示。

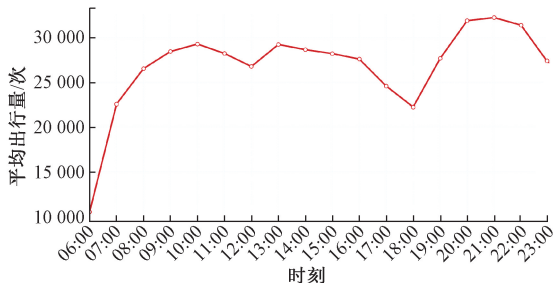


图 2 平均载客量变化趋势

Fig. 2 Temporal trend of average passenger volume

如图 2 所示,出发频次呈现出显著的“早晚双峰”特征。具体而言,载客量自 07:00 起逐步上升,并于 10:00 达到峰值,随后逐渐回落;至 18:00 后再次上升,21:00 左右达到峰值,之后逐步下降。数据显示,早晚高峰的上升与衰减趋势在时间分布上呈现一定对称性,符合拟高斯分布特征,为后续构建高斯权重函数提供了实证依据。参考全国主要城市的交通分析报告,早晚高峰的持续时间通常为 2~3 h。基于这一特征,本研究选取 10:00 和 21:00 作为早、晚高峰的中心时间点 μ_1 和 μ_2 。本文采用双高斯加权策略,即通

过两个高斯核函数的线性组合构建时间权重,以更准确地刻画双高峰时段对出行行为的影响。在实验中,分别设置 $\mu_1=4$ (对应 10:00)、 $\mu_2=15$ (对应 21:00),并考察 $\sigma=1, 2, 3$ h(对应高峰前后 1、2、3 h)的权重分配方案。通过 K-Medoids 聚类算法,结合轮廓系数和 Calinski-Harabasz 指数评估不同参数配置下的聚类效果,最终确定最优模型参数。

3.2 最佳聚类数目选择

城市交通出行模式主要由居民日常活动、通勤习惯和道路结构决定,通常表现为有限数量的核心行为特征,如早高峰通勤、晚高峰返程和非高峰休闲出行等。当聚类簇数过多时,新增簇往往只是已有模式的微小变体,而非独立的行为类别。同时,簇数增加还可能导致簇间边界模糊,部分样本被划分到语义不清的小簇中,影响聚类结果的稳定性与可解释性。

如图 3、4 所示,无论是 DTW、WDTW 还是 GS-WDTW 方法,其轮廓系数与 Calinski-Harabasz 指数均随簇数 K 的增加整体呈下降趋势,这符合聚类分析中关于“自然簇结构”的基本假设,即真实数据通常只包含有限数量的主导模式,并非可以无限细分。当簇数 K 在 2~7 范围内,GS-WDTW 在 $\sigma=1, 2, 3$ h 情形下,其轮廓系数与 Calinski-Harabasz 指数相较于 DTW 和 WDTW 均表现更优,这表明 GS-WDTW 能够更好地捕捉时间序列数据间的相似性和差异性,提升样本归属的合理性与簇间分离度。当 K 为 8、9 时尽管 DTW 在轮廓系数出现小幅波动,但此时聚类结果已缺乏实际意义,其实用价值也较为有限,需要注意的是,Calinski-Harabasz 指数对噪声数据敏感,且倾向于选择较小的簇数 K 。结合本研究的聚类目标及出行模式的多样性,选择聚类数 $K \geq 3$ 更为合理。当 $K=4$ 时,GS-WDTW 在不同评估指标下均表现出显著优于 DTW 和 WDTW 的聚类效果:轮廓系数分别提升了 39.9% 和 41.1%, Calinski-Harabasz 指数分别提升了 13.1% 和 13.0%。而当 $K > 4$ 时,性能提升幅度明显减弱,定量结果

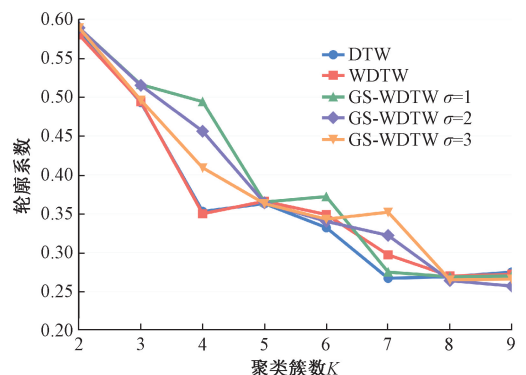


图 3 不同聚类数目下各算法轮廓系数的变化趋势

Fig. 3 Trend of silhouette coefficients for different algorithms under varying numbers of clusters

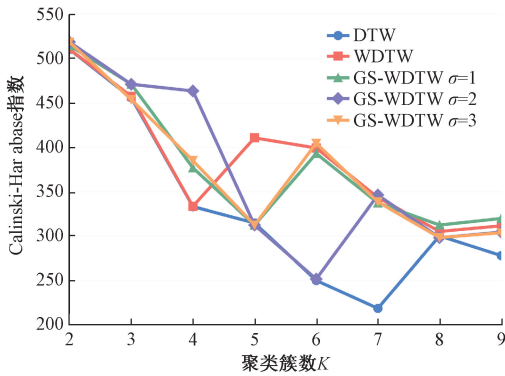


图4 不同聚类数目下各算法 Calinski-Harabasz 指数的变化趋势

Fig.4 Trend of Calinski-Harabasz index for different algorithms under varying numbers of clusters

如表 2、3 所示。因此,选择模型 GS-WDTW($\mu_1 = 4, \mu_2 = 15, \sigma = 1, K = 4$)能够表现出更好的模式识别能力,其各个时刻的权重分布情况如图 5 所示,在 10:00、21:00 点权重值达到最大值,与日出行平均载客量的峰值同步。

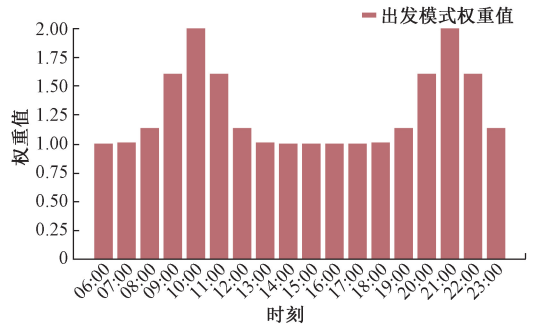


图5 各个时间点的权重分配值

Fig.5 Weight allocation values at each time point

表 2 不同聚类数目下各算法轮廓系数的定量结果

Table 2 Quantitative results of silhouette coefficients for various algorithms under different numbers of clusters

K	DTW	WDTW	GS-WDTW		
			$\sigma = 1$	$\sigma = 2$	$\sigma = 3$
2	0.584	0.580	0.589	0.589	0.589
3	0.493	0.494	0.516	0.515	0.496
4	0.353	0.350	0.494	0.456	0.409
5	0.363	0.366	0.365	0.364	0.363
6	0.332	0.349	0.372	0.340	0.343
7	0.267	0.297	0.275	0.322	0.352
8	0.269	0.270	0.269	0.264	0.265
9	0.275	0.272	0.270	0.257	0.266

表 3 不同聚类数下各算法的 Calinski-Harabasz 指数的定量结果

Table 3 Quantitative results of Calinski-Harabasz index for various algorithms under different numbers of clusters

K	DTW	WDTW	GS-WDTW		
			$\sigma = 1$	$\sigma = 2$	$\sigma = 3$
2	510.382	510.653	514.728	518.778	518.778
3	455.955	457.590	471.333	471.178	453.953
4	333.266	333.468	376.987	463.658	385.062
5	314.717	410.636	312.318	312.354	312.422
6	249.916	399.211	393.110	251.495	404.497
7	218.485	343.435	337.313	346.199	338.434
8	300.134	305.142	312.406	298.147	298.162
9	277.859	311.590	319.822	304.404	303.722

4 结果分析

基于上述的讨论,本文进一步探讨城市的出行规律,选择 GS-WDTW($\mu_1 = 4, \mu_2 = 15, \sigma = 1, K = 4$)模型对出行规律进行分析。

4.1 聚类结果分析

基于动态时间规整的相似性度量进行聚类,目的是更准确地识别具有相似“整体模式”的序列,同时将具有不同“整体模式”的序列分离。如图 6、7 所示,观察到 DTW 及 WDTW 生成的聚类中心曲线在形态上存在显著的相似性。例如,Cluster 1、Cluster 2 和 Cluster 4 的曲线呈现明显的相似性,这使得各自模式的本质差异难以辨别,反映出传统 DTW 及 WDTW 在处理此类复杂模式时的局限性。相比之下,图 8 中的 GS-WDTW 所生成的聚类中心则表现出较高的形态可区分性。每条曲线均具有独特的、易于理解的核心特征:Cluster 1 表现为平缓且稳定的低频次模式,Cluster 2 为典型的多峰模式,主要出现在早晚高峰;Cluster 3 呈现高频次且波动剧烈的多峰特征。GS-WDTW 不同簇聚类的效果如图 9 所示。此外,从图中对比观察出 GS-WDTW 的聚类结果覆盖了更宽泛的频次范围(如图 8 所示,频次范围从接近 0 到超过 250)。这一现象体现出该方法对数据集中不同尺度模式的全面识别能力。传统 DTW 由于其对所有时间点赋予同等权重,当数据集中存在极端高频或低频的模式时,这些异常模式可能会被“稀释”或与形态相似但频次差异巨大的正常模式混淆在一起,难以将这些边缘模式有效分离。而 GS-WDTW 通过放大关键极值点的影响,能够更有效地识别和聚类那些具有极高频次的独特行为模式。

4.2 功能区分布验证

结合上述分析思路,对到达模式的数据进行处理。以居民的出发模式为基础,结合工作日上下车点数据,利用 ArcMap 进行空间分析与聚类识别,最终划分出 8 类具有代表性的地块出行模式,如图 10 所示。结合表 4 进一步对各地块功能属性进行判别。其中,富集指数是量化特定地理区域内某类兴趣点(POD)的空间聚集程度相对于基准区

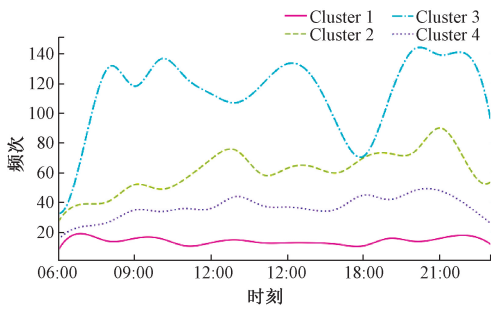


图 6 DTW 方法在 $K=4$ 时的聚类中心曲线

Fig. 6 Cluster centroid curves of the DTW method at $K=4$

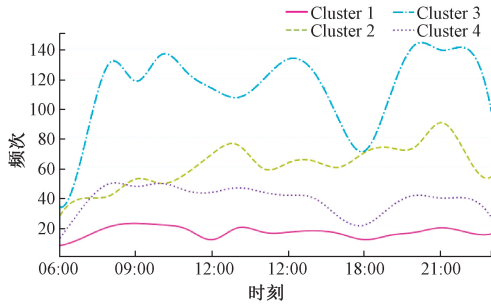


图 7 WDTW 方法在 $K=4$ 时的聚类中心曲线

Fig. 7 Cluster medoid curves of the WDTW method at $K=4$

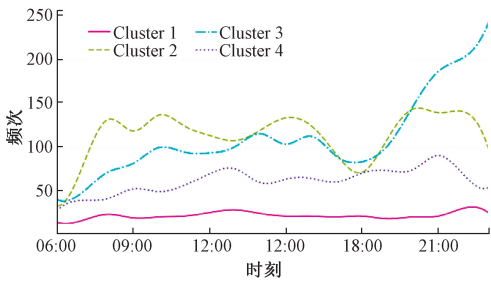


图 8 GS-WDTW 方法在 $K=4$ 时的聚类中心曲线

Fig. 8 Cluster medoid curves of the GS-WDTW method at $K=4$

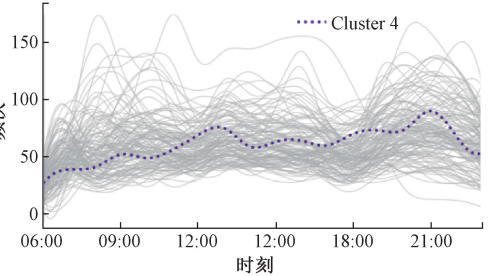
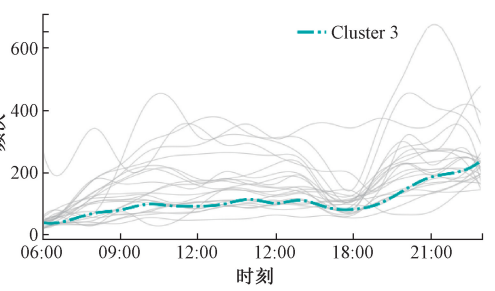
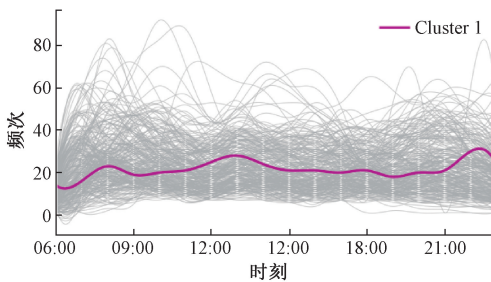


图 9 GS-WDTW 识别的不同聚类簇的结果

Fig. 9 Results for the distinct clusters derived from GS-WDTW method

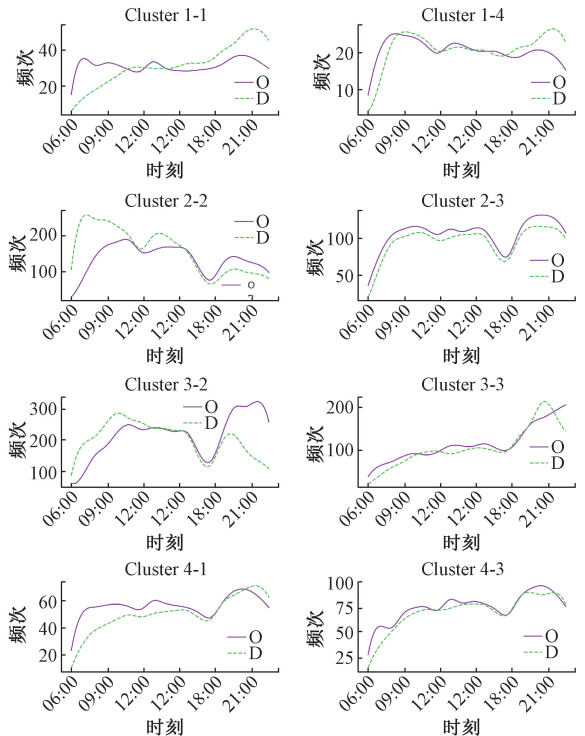
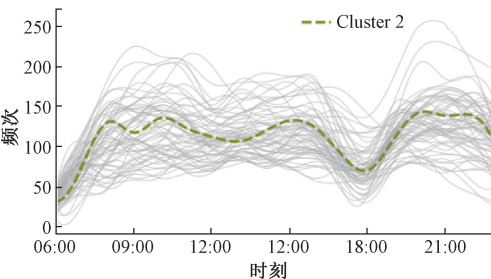


图 10 各地块出行模式曲线

Fig. 10 Travel pattern curves of individual land parcels



域的偏离程度。各地块功能属性识别结果如图 11 所示。本研究通过聚类分析识别出 6 类典型城市功能区：居住型区域(Cluster1-1)呈现全天均衡出行和晚高峰到达特征，商务住宅密集；外围居住区(Cluster1-4)表现为夜间到达高峰明显，且基础服务设施集中；中央商务区(Cluster2-2)具有典型早晚高峰和早间净流入特点，集聚企业、酒店和文化设

表4 不同地块的POI富集指数

Table 4 POI enrichment indices across different land parcels

POI类别	1-1	1-4	2-2	2-3	3-2	3-3	4-1	4-3
餐饮	1.11	0.94	0.89	0.94	0.89	1.18	1.01	1.14
公司	0.54	1.03	1.26	1.19	1.20	1.06	0.84	1.03
交通	1.24	1.25	0.95	0.88	0.81	0.81	1.02	0.79
科教文化	1.06	0.96	1.29	1.00	0.98	1.03	1.06	0.81
商务住宅	1.29	0.89	1.10	1.04	0.78	0.83	1.06	0.92
生活服务	1.14	1.05	0.75	0.93	0.96	0.86	1.06	1.00
政府	0.95	0.89	0.80	1.10	1.19	0.66	1.12	0.88
住宿	0.49	0.71	1.98	1.08	1.84	1.29	1.08	0.69

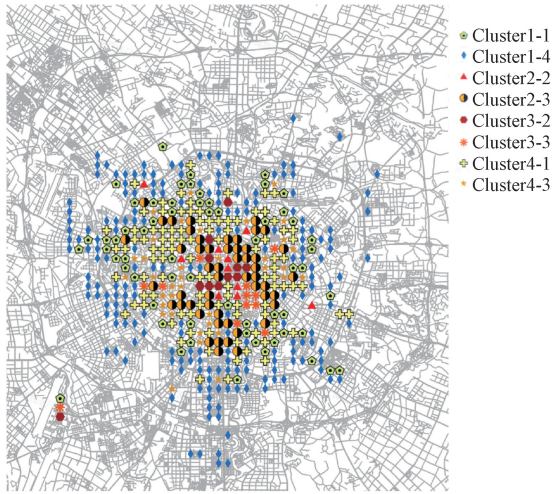


图11 各地块功能属性识别结果

Fig. 11 Functional attribute identification results for individual land parcels

施;职住过渡区(Cluster2-3、4-1与4-3)显示明显通勤潮汐现象,商住混合且配套完善;综合枢纽区(Cluster3-2)展现全天高强度出行和早进晚出模式,复合交通枢纽、政务和商业功能;休闲娱乐区(Cluster3-3)则以夜间活跃和餐饮住宿密集为特征。其中编号规则采用“出发模式聚类-到达模式聚类”的匹配方式,如Cluster1-4表示出发模式Cluster1与到达模式Cluster4的匹配区域。识别结果与成都市实际的城市功能区分布相吻合^[19],进一步验证了该方法的有效性。

5 结 论

本研究针对交通状态时间序列分析中,传统动态时间规整加权变体在关键时段敏感性不足、权重分配机制不合理等问题,提出一种基于高斯函数加权的改进动态时间规整方法。该方法引入具有时间依赖的高斯权重函数,将城市交通流量所呈现的拟高斯分布特征及高峰持续时间作为先验知识,融入相似性度量过程。通过构建双峰高斯核函数,显著增强了高峰时段在序列对齐路径计算中的权重贡

献,有效提升了对交通流关键特征的捕捉能力,对处理其他具有显著时段异质性的时序数据(如用电负荷、客流量、网络流量)具有借鉴意义。尽管本研究提出的GS-WDTW方法在交通时序分析中有较好的表现,但仍存在一些需要改进的方面:当前高斯权重参数仍需人工设定,泛化能力有限;未考虑交通模式的长期动态变化。针对这些不足,下一步的工作将从3个方面出发:一是结合机器学习实现权重参数的自适应优化,提升方法泛化能力。近期研究表明,通过构建观测质量因子模型与异常检测机制,实时评估并动态调节每个时间点的权重贡献,也成为提升时序相似性度量鲁棒性考量的重要方向^[20];二是融合手机信令、公共交通等多源异构数据,构建更全面的城市出行分析框架;三是引入时间序列预测方法,探索交通模式的动态演变规律。这些改进将显著提升该方法在智慧城市交通规划与管理中的实际应用价值。

参考文献

- [1] 赵谦,马文越,郑超,等. 基于单路口车流量统计的交通灯配时系统研究[J]. 电子测量技术, 2023, 46(8): 82-91.
ZHAO Q, MA W Y, ZHENG CH, et al. Research on traffic light timing system based on traffic flow statistics at single intersection [J]. Electronic Measurement Technology, 2023, 46(8): 82-91.
- [2] 骆钰波,陈碧宇. 面向路网轨迹的自适应数据模型与索引结构[J]. 地球信息科学学报, 2023, 25(1): 63-76.
LUO Y B, CHEN B Y. Adaptive data model and index structure for network-constrained trajectories[J]. Journal of Geo-information Science, 2023, 25(1): 63-76.
- [3] 程前,李建良. 基于改进动态时间规整的相似性度量及轨迹聚类[J]. 电子测量技术, 2021, 44(23): 1-5.
CHENG Q, LI J L. Improved dynamic time warping for similar metrics and trajectory clustering [J]. Electronic Measurement Technology, 2021, 44(23): 1-5.
- [4] 姚尧,张亚涛,关庆锋,等. 使用时序出租车轨迹识别多层次城市功能结构[J]. 武汉大学学报(信息科学版), 2019, 44(6): 875-884.
YAO Y, ZHANG Y T, GUAN Q F, et al. Multi-level urban functional structure recognition using time-series taxi trajectories[J]. Geomatics and Information Science of Wuhan University, 2019, 44(6): 875-884.
- [5] 马书红,张俊杰,陈西芳,等. 利用出租车时序数据识别城市功能区[J]. 吉林大学学报(工学版), 2025, 55(2): 603-613.
MA SH H, ZHANG J J, CHEN X F, et al. Identifying urban functional structures using time-series taxi data [J]. Journal of Jilin University

- (Engineering and Technology Edition), 2025, 55(2): 603-613.
- [6] 邸少宁, 朱杰, 郑加柱, 等. 出租车轨迹数据的南京人群出行模式挖掘[J]. 测绘科学, 2021, 46(1):203-212. DI SH N, ZHU J, ZHENG J ZH, et al. Movement pattern mining of Nanjing residents on taxi trajectory data[J]. Science of Surveying and Mapping, 2021, 46(1): 203-212.
- [7] 陈海燕, 刘晨晖, 孙博. 时间序列数据挖掘的相似性度量综述[J]. 控制与决策, 2017, 32(1):1-11. CHEN H Y, LIU CH H, SUN B. Survey on similarity measurement of time series data mining[J]. Control and Decision, 2017, 32(1): 1-11.
- [8] 杨傲雷, 周应宏, 杨帮华, 等. 基于 Transformer 的三维人体姿态估计及其动作达成度评估[J]. 仪器仪表学报, 2024, 45(4):136-144. YANG AO L, ZHOU Y H, YANG B H, et al. Transformer-based 3D human pose estimation and action achievement assessment evaluation[J]. Chinese Journal of Scientific Instrument, 2024, 45(4): 136-144.
- [9] YANG G Y, XIA SH Y. Weighted dynamic time warping for time series[J]. International Journal of Bifurcation and Chaos, 2023, 33(13): 23501493.
- [10] ZHAO X F, HU C Y, LIU ZH, et al. Weighted dynamic time warping for grid-based travel-demand-pattern clustering: Case study of Beijing bicycle-sharing system [J]. ISPRS International Journal of Geo-Information, 2019, 8(6): 281.
- [11] DAU H A, SILVA D F, PETITJEAN F. et al. Optimizing dynamic time warping's window width for time series data mining applications[J]. Data Mining Knowledge Discovery, 2018, 32: 1074-1120.
- [12] LI MAN, ZHU YE, ZHAO T G, et al. Weighted dynamic time warping for traffic flow clustering[J]. Neurocomputing, 2022, 472: 266-279.
- [13] BRUSTAD H K, MIDTBØ J E, TOMBA G S, et al. Extracting commuters from automated road traffic counters: A gaussian mixture approach [J]. Data Science for Transportation, 2025, 7: 9.
- [14] NING J, MA H ZH, SUN Y, et al. Research on the characteristic identification and multidimensional dynamic evolution of urban-rural fringe in Harbin, China[J]. Land, 2025, 14(2): 359.
- [15] HOLDER C, GUIJO-RUBIO D, BAGNALL A. Clustering time series with K-Medoids based algorithms[C]. Lecture Notes in Computer Science, 2023, 14343:39-55.
- [16] 程静, 刘家骏, 高勇. 基于时间序列聚类方法分析北京出租车出行量的时空特征[J]. 地球信息科学学报, 2016, 18(9): 1227-1239. CHENG J, LIU J J, GAO Y. Analyzing the spatio-temporal characteristics of Beijing's OD trip volume based on time series clustering method[J]. Journal of Geo-information Science, 2016, 18(9): 1227-1239.
- [17] 王恺亮, 朱新宇. 基于聚类分析的交通事故影响因素研究[J]. 科技创新与应用, 2024, 14(35):107-112. WANG K L, ZHU X Y. Research on influencing factors of traffic accidents based on cluster analysis[J]. Technology Innovation and Application, 2024, 14(35): 107-112.
- [18] NGO Q D, PHAN L H. Analysis of multivariate indoor building data: A comparative study of time-series clustering methods[J]. IEEE Access, 2025, 13: 139018-139032.
- [19] 李莹. 基于成都出租车数据的居民出行特征识别与聚类分析[D]. 武汉: 武汉大学, 2021. LI Y. Research on travel characteristics of urban residents and cluster analysis based on Chengdu taxi data[D]. Wuhan: Wuhan University, 2021.
- [20] 黄超, 黄予昕, 杨泽彬, 等. VIG-SLAM: 基于自适应多传感器融合的 SLAM 算法[J]. 电子测量与仪器学报, 2025, 39(5):67-74. HUANG CH, HUANG Y X, YANG Z B, et al. VIG-SLAM: A SLAM algorithm based on adaptive multi-sensor fusion[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(5): 67-74.

作者简介

黄胜浩, 硕士, 主要研究方向为信息分析与数据挖掘

E-mail:2571810799@qq.com

黄文德(通信作者), 博士, 教授, 主要研究方向为卫星导航定位技术。

E-mail:nn_hwd@126.com