Impact of Regional Driving Behavior Differences on Traffic Flow

Yuting Wang[1,2], Zhaocheng He[1,2,3,*] , Lijian Zhuang[4]

[1] School of Intelligent Engineering, Sun Yat-sen University
[2] Guangdong Provincial Key Laboratory of Intelligent Transportation System Sun Yat-Sen University
[3] The Pengcheng Laboratory
[4] Shenzhen Urban Transport Planning Center Co., Ltd.

* Corresponding author. E-mail: hezhch@mail.sysu.edu.cn

Supplementary Information

. Supplementary Tables

Supplementary Table 1. Mean values of performance metrics for each functional station type.

Cluster ID	Functional Type	No. of Stations	F_1	F_2	T_1	T_2	M_1	M_2	E_1	E_2	Connectors	Power (kW)
0	5413	71.49%	0.2	0.13	0.84	0.48	4.45	2.66	3.58	2.09	26.72	46.01
1	1236	16.32%	0.6	0.54	3.5	3.11	12.87	11.06	10.74	9.25	20.17	46.79
2	91	1.20%	3.16	1.08	2.26	0.93	63.94	9.36	120.74	51.98	19.34	132.43
3	832	10.99%	0.53	0.28	1.5	0.7	7.31	3.39	7.83	3.69	80.89	17.15

Supplementary Table 2. Description of charging station construction and operational data.

Field	Data Type	Description
station_id	Integer	A unique identifier for each physical charging station.
station_type	Enum	The category of the station (e.g., public, private, battery-swapping).
num_connectors	Integer	The total number of charging connectors (outlets) at the station.
num_piles	Integer	The total number of charging piles (physical units) at the station.
address	String	The detailed street address of the station.
site_context	String	The physical environment of the station (e.g., underground parking, roadside, shopping mall).
electricity_fee	Decimal	The price per kilowatt-hour (kWh) of electricity.
service_fee	Decimal	The service fee, charged either per session or per kWh.
longitude	Decimal	The longitude of the station's location (WGS84).
latitude	Decimal	The latitude of the station's location (WGS84).
total_power	Decimal	The total rated power of all charging equipment at the station, in kilowatts (kW).

Supplementary Table 3. Description of charging transaction data.

Field	Data Type	Description
station_id	Integer	A unique identifier for the charging station.
equipment_id	Integer	A unique identifier for a specific piece of charging equipment within a station.
connector_id	Integer	A unique identifier for a specific charging connector on a piece of equipment.
order_id	String	A unique identifier for each charging session.
start_time	Datetime	The start time of the charging session (to the second).
end_time	Datetime	The end time of the charging session (to the second).
energy_delivered	Decimal (kWh)	The total energy delivered during the session, in kilowatt-hours (kWh).

total_fee	Decimal	The total fee for the session, including both electricity and service charges.
electricity_fee	Decimal	The portion of the total fee corresponding to the electricity consumed.
service_fee	Decimal	The portion of the total fee corresponding to the service charge.

Supplementary Table 4. Description of network resilience metrics.

Category	Metric	Definition	Interpretation & Role in Analysis
Structural Integrity	Network size	The total number of nodes in the network.	Indicates the extent of node removal or failure.
	Number of connected components	The number of maximal subgraphs in which any two nodes are connected to each other by paths.	An increase signifies network fragmentation and reduced interconnectivity.
	Size of the largest connected component (LCC)	The number of nodes in the network's largest connected component.	A high value indicates the presence of a giant component maintaining overall function; a sharp decrease signals the collapse of the network backbone.
Core Robustness	Maximum k-core number	The largest integer k for which a k-core (a maximal subgraph where every node has a degree of at least k) exists.	A high value indicates a densely interconnected core. A rapid drop to zero during an attack signifies the failure of this core skeleton.
Accessibility & Efficiency	Average path length	The average of the shortest path lengths over all pairs of reachable nodes.	Reflects the average travel distance between nodes in a connected network. A sudden drop after fragmentation indicates that measurements are confined to small, local components.
	The average of the inverse of the shortest path lengths over all pairs of nodes (unreachable pairs contribute zero).	Measures the efficiency of information or resource flow across the entire network. A high value indicates high accessibility and efficiency. May exhibit complex behaviour post-fragmentation.	The average of the inverse of the shortest path lengths over all pairs of nodes (unreachable pairs contribute zero).

Supplementary Algorithm

Supplementary Algorithm 1

Algorithm 1 基于主成分降维的充电站聚类评级算法

Input: 特征矩阵 $X \in R^{n\times d}$；KMO 阈值 τ；显著性水平 α；累计解释率阈值 η；候选簇数集合 $\mathcal{K}$；候选算法 $\mathcal{A} = \{\text{K-means}, \text{Agglo}, \text{DBSCAN}\}$

Output: 每个站点的簇标签 c_i 及对应的聚类评价指标

```
                                        ▷ 特征标准化，参见式 (A1)
依据式 (3–5) 对 X 做 Z-score 标准化，得到矩阵 Z，并构造相关矩阵 R
                                        ▷ KMO 与 Bartlett 适用性检验，参见式 (A2)–(A3)
计算 KMO 统计量与 Bartlett 检验统计量 χ^2 及其 p 值
if KMO ≤ τ  or  p ≥ α then
    return "数据不适合进行主成分降维处理"
end if
                                        ▷ 主成分提取与主成分得分计算，参见式 (A4)–(A6)
对 R 做特征分解 R w_q = λ_q w_q，按 λ_q 由大到小排序
根据式 (A5) 计算累计解释率 η(Q)，选择满足 η(Q) ≥ η 且碎石图出现拐点的主成分个数 Q
构造投影矩阵 W = [w_1, ..., w_Q] 并计算主成分得分矩阵 Y = ZW
                                        ▷ 聚类参数确定与建模，参见式 (A7)
for all A ∈ 𝒜 do
    if A 为 K-means 或 Agglomerative then
        在 k ∈ 𝒦 上利用式 (3–11) 计算 WCSS(k) 并采用手肘法确定 k*
        在 Y 上用 A 和簇数 k* 进行聚类，得到标签向量 c^(A)
    else                                ▷ A 为 DBSCAN
        依据 k-distance 曲线选取 (ε, MinPts) 并在 Y 上运行 DBSCAN，得到 c^(A)
    end if
                                        ▷ 聚类质量评价，参见轮廓系数式 (A8)
    计算当前聚类结果的平均轮廓系数 S^(A)
end for
选择具有最大 S^(A) 的算法 A* 及其参数，令最终簇标签 c_i ⇐ c_i^(A*)
return {c_i}_{i=1}^n 及相应的聚类质量评价结果
```

基于主成分降维的充电站聚类评级算法详细步骤与公式：

（1）符号定义与数据标准化

设共有 n 个充电站、d 个评价特征，构建特征矩阵 $X = [x_{ij}]_{(n\times d)}$，其中 x_i 为站点 i 的特征向量。考虑不同指标量纲差异，采用 Z-score 标准化得到：

$$z_{ij} = (x_{ij} - \mu_j)/\sigma_j \quad \text{A1}$$

其中 μ_j、σ_j 分别为第 j 个特征的均值与标准差。标准化后的矩阵记为 Z，用于相关结构检验与降维计算。

（2）KMO 与 Bartlett 球形检验

由于特征指标之间可能存在相关性，降维前需检验数据结构是否适合提取潜在变量。

1）KMO 检验用于衡量变量间相关性是否适合做主成分/因子提取。令 r_{jk} 为相关系数，p_{jk} 为偏相关系数，则

$$\mathrm{KMO} = \left[\sum(\mathrm{j} \neq \mathrm{k})\, r_{jk}{}^2\right] / \left[\sum(\mathrm{j} \neq \mathrm{k}) r_{jk} + \sum(\mathrm{j} \neq \mathrm{k})\, p_{jk}{}^2\right] \quad \text{A2}$$

2）Bartlett 球形检验用于检验相关矩阵 R 是否为单位阵。设样本量为 n、变量数为 d，则统计量可写为

$$\chi^2 = -(\mathrm{n} - 1 - (2\mathrm{d} + 5)/6)\ln|\mathrm{R}| \quad \text{A3}$$

其自由度为 $d(d-1)/2$。当 $p < 0.05$ 时拒绝“相关矩阵为单位阵”的原假设，认为变量间存在足够相关性，适合进行降维处理。本研究中 KMO=0.63>0.5，巴特利特球形度检验 P<0.05，结果表明此数据适合进行降维聚类处理。

（3）降维：主成分提取与主成分得分计算

在通过适用性检验后，基于标准化矩阵 Z 构造相关矩阵 R（或协方差矩阵），并进行特征分解：

$$\mathrm{Rw_q} = \lambda_\mathrm{q} \mathrm{w_q},\ \ \mathrm{q} = 1, \ldots, \mathrm{d} \quad \text{A4}$$

其中 λ_q 为特征值，w_q 为对应特征向量。综合采用碎石图（Scree Plot）与方差解释准则确定主成分数量 Q。累计解释率定义为

$$\eta(\mathrm{Q}) = \left(\sum(\mathrm{q} = 1)^\mathrm{Q} \lambda_\mathrm{q}\right) / \left(\sum(\mathrm{q} = 1)^\mathrm{d} \lambda_\mathrm{q}\right) \quad \text{A5}$$

在满足累计解释率达到阈值且碎石图出现明显拐点的条件下确定 Q。随后计算站点在潜在空间的低维表示（主成分得分）：

$$\mathrm{y_i} = \mathrm{W^T z_i} \quad \text{A6}$$

其中 $W = [w_1, \ldots, w_Q]$ 为由前 Q 个主成分载荷向量组成的投影矩阵，y_i 为站点 i 的 Q 维潜在表示。将所有站点的 y_i 组成矩阵 Y，作为聚类输入数据。

（4）最佳聚类数确定

完成特征降维后，需要基于数据自身确定聚类数 k。对 K-means 等基于中心的聚类方法，采用手肘法（Elbow Method）分析簇内平方和（Within-Cluster Sum of Squares, WCSS）随 k 的变化关系：

$$\mathrm{WCSS(k)} = \sum(\mathrm{c} = 1)^\mathrm{k} \sum(\mathrm{y_i} \in C_c) \left|\left|\mathrm{y_i} - \mu_c\right|\right|^2 \quad \text{A7}$$

其中 C_c 表示第 c 个簇，μ_c 为其质心。通常 WCSS(k) 随 k 增大而下降，当曲线下降趋势出现明显“拐点”时，对应的 k 可视为较合理的聚类数。

（5）聚类模型构建

在确定聚类参数后，本研究采用三种典型无监督聚类方法进行对比分析：

1）K-means：适用于近似球状簇结构；

2）Agglomerative Clustering（层次聚类）：自底向上逐步合并簇，能够揭示层级结构；

3）DBSCAN（Density-Based Spatial Clustering of Applications with Noise）：基于密度可识别任意形状簇，并对异常点/噪声点具有较强鲁棒性。

（6）聚类效果评价

采用轮廓系数（Silhouette Coefficient）评价聚类质量。对样本 y_i，定义：a(i) 为 y_i 与同簇样本的平均距离；b(i) 为 y_i 与最近其他簇样本的平均距离，则轮廓系数为

$$\mathrm{s(i)} = b(i) - a(i)/\max[a(i), b(i)] \quad \text{A8}$$

其取值范围为 $[-1,1]$。s(i) 越接近 1 表明簇内更紧密、簇间更分离；若 $s(i) < 0$ 通常表示样本更接近其他簇，聚类效果较差。

本文聚类结果 K-Means 、DBSCAN、 Agglomerative Clustering 三种轮廓系数分别为 0.46，

-0.46，0.41。K-means 聚类算法表现最优。

Supplementary Algorithm 2

Algorithm 2 基于功率排序的渐进式节点失效策略

Input: 初始网络 $G^{(0)} = (V^{(0)}, E^{(0)})$；节点功率函数 $\text{sum\_power}(v)$；总步数 T

Output: 每一步的剩余网络 $\{G^{(t)}\}_{t=0}^{T}$ 及对应的失效节点集合 $\{V_{\text{remove}}^{(t)}\}_{t=1}^{T}$

$N \Leftarrow |V^{(0)}|$ ▷ 记录初始节点总数

▷ Step 1: 定义节点功率，参见 $P(v) = \text{sum\_power}(v)$

for all $v \in V^{(0)}$ **do**

  $P(v) \Leftarrow \text{sum\_power}(v)$

end for

▷ Step 2: 按功率从大到小排序生成攻击序列

将 $V^{(0)}$ 按 $P(v)$ 降序排序，得到序列 $\pi = (v_1, v_2, \ldots, v_N)$，使得 $P(v_i) \geq P(v_{i+1})$

▷ Step 3: 渐进式攻击与网络更新

$G^{(0)} \Leftarrow (V^{(0)}, E^{(0)})$

for $t = 1$ **to** T **do**

  $p_t \Leftarrow 0.1 \times t$ ▷ 本研究中攻击比例每步增加 10%

  $n_{\text{remove}}^{(t)} \Leftarrow \lfloor p_t \cdot N \rfloor$

  $V_{\text{remove}}^{(t)} \Leftarrow \{v_1, \ldots, v_{n_{\text{remove}}^{(t)}}\}$

  $V^{(t)} \Leftarrow V^{(0)} \setminus V_{\text{remove}}^{(t)}$

  $E^{(t)} \Leftarrow \{(u, v) \in E^{(0)} \mid u \in V^{(t)}, v \in V^{(t)}\}$

  $G^{(t)} \Leftarrow (V^{(t)}, E^{(t)})$

  在 $G^{(t)}$ 上计算韧性指标

  if $V^{(t)} = \emptyset$ **then**

    break ▷ 网络完全失效时提前终止

  end if

end for

return $\{G^{(t)}\}_{t=0}^{T}$, $\{V_{\text{remove}}^{(t)}\}_{t=1}^{T}$ 与指标

Supplementary Algorithm 3

Algorithm 3 基于功能分区的分阶段目标攻击策略

Input: 初始网络 $G^{(0)} = (V^{(0)}, E^{(0)})$；土地用途类型函数 $\mathrm{land\_type}(v) \in \{A, B, R\}$；攻击序列 $\mathcal{L} = (L_1, \dots, L_M)$

Output: 每一阶段的剩余网络 $\{G^{(s)}\}_{s=0}^{M}$ 及对应的失效节点集合 $\{V_{\mathrm{remove}}^{(s)}\}_{s=1}^{M}$

▷ Step 1: 功能类型定义

for all $v \in V^{(0)}$ **do**

  指定 $\mathrm{land\_type}(v) \in \{A, B, R\}$ ▷ 如 A 公共管理, B 商业服务业, R 居住

end for

▷ Step 2: 定义功能分区攻击顺序

给定攻击序列 $\mathcal{L} = (L_1, L_2, \dots, L_M)$，其中 $L_s \in \{A, B, R\}$ 表示第 s 阶段攻击的功能分区

▷ Step 3: 分阶段目标攻击与网络更新

$G^{(0)} \Leftarrow (V^{(0)}, E^{(0)})$

for $s = 1$ **to** M **do**

  $V_{\mathrm{remove}}^{(s)} \Leftarrow \{v \in V^{(s-1)} \mid \mathrm{land\_type}(v) = L_s\}$ ▷ 移除指定功能分区内所有节点

  $V^{(s)} \Leftarrow V^{(s-1)} \setminus V_{\mathrm{remove}}^{(s)}$

  $E^{(s)} \Leftarrow \{(u, v) \in E^{(s-1)} \mid u \in V^{(s)}, v \in V^{(s)}\}$

  $G^{(s)} \Leftarrow (V^{(s)}, E^{(s)})$

  在 $G^{(s)}$ 上计算韧性指标

  if $V^{(s)} = \emptyset$ **then**

    break ▷ 网络无剩余节点时提前终止

  end if

end for

return $\{G^{(s)}\}_{s=0}^{M}$, $\{V_{\mathrm{remove}}^{(s)}\}_{s=1}^{M}$ 与各阶段韧性指标