

基于路径拼接模型的路径行程时间估计方法*

李焯焯¹, 黄敏¹, 黄春婷¹, 张小兰²

1. 中山大学智能工程学院 / 广东省智能交通系统重点实验室, 广东 深圳 518107

2. 广东工贸职业技术学院, 广东 广州 510510

摘要: 依托卡口检测数据, 提出路径拼接模型以利用大量短路径行程时间对城市道路中的长路径行程时间进行估计。该模型提出子路径进行拼接时要满足的两个条件: 存在公共结点及交通状态相近; 在此基础上将目标路径按照整体方差最小的原则拆分为由多个子路径组成的拼接方案。基于提出的路径交通流基本图划分子路径的交通状态, 使用 Burr 分布拟合不同交通状态下的子路径行程时间得到行程时间分布函数; 将给定的目标行程时间分配至拼接方案中的子路径使得子路径行程时间之和等于目标行程时间, 累加所有时间分配方案对应的概率得到目标路径的行程时间概率, 计算多个目标行程时间对应的概率得到目标路径的行程时间概率分布。实验结果显示: 本文提出的路径行程时间估计方法得到的路径行程时间均值与真实轨迹均值的误差为 3.04%; 得到的行程时间分布与真实行程时间分布的 JS 散度为 0.05。路径行程时间估计值真实可靠, 可为后续研究提供数据基础。

关键词: 路径行程时间; 路径拼接; 交通状态; 行程时间分布

中图分类号: U491 **文献标志码:** A **文章编号:** 2097-0137(2023)06-0080-09

A method of path travel time estimation based on path splicing model

LI Yetao¹, HUANG Min¹, HUANG Chunting¹, ZHANG Xiaolan²

1. School of Intelligent Systems Engineering / Guangdong Provincial Key Laboratory

of Intelligent Transportation System, Sun Yat-sen University, Shenzhen 518107, China

2. Guangdong Polytechnic of Industry and Commerce, Guangzhou 510510, China

Abstract: In this paper, we propose a path splicing model to estimate the long path travel time on urban roads using the automatic vehicle identification (AVI) data. The model requires two conditions to be met when splicing the sub-paths: public nodes exist and the traffic status is similar. A target path is split into a splicing scheme composed of several subpaths whose variance is the least. Based on the proposed traffic state, the travel time distributions are obtained by fitting the travel time under different states using the Burr distribution. Next, the method assigns target travel time to the sub-paths, thus, the travel time probability of the target path is the accumulation of the probability of all time assignments. The study shows that the error between the average path travel time of the proposed method and the real path is 3.04%, and the JS divergence between the travel time distribution of the method and the real path is 0.05. The estimated path travel time is reliable, which can provide a data basis for subsequent research.

Key words: path travel time; path splicing; traffic status; distribution of travel time

* 收稿日期: 2023-02-13

录用日期: 2023-04-21

网络首发日期: 2023-07-26

基金项目: 国家重点研发计划(2020YFB1600400)

作者简介: 李焯焯(1997年生), 男; 研究方向: 智能交通系统; E-mail: liyd33@mail2.sysu.edu.cn

通信作者: 黄敏(1975年生), 女; 研究方向: 智能交通系统; E-mail: huangm7@mail.sysu.edu.cn

交通管理与控制技术的进步与发展, 交通信息收集设备的增加, 提高了道路交通信息采集与实时处理的能力。例如, 以AVI系统(电警卡口、RFID设备等)为代表的道路监控系统, 对路径行程时间的相关参数进行了直接测量(刘晓波等, 2019)。行程时间作为智能交通信息系统服务的重要组成部分, 对实现交通流诱导、提高道路服务质量具有重要的意义, 在交通政策制定、路径规划等方面具有基础性的作用。

基于车牌识别数据的路径行程时间估计相关的研究按照研究尺度的大小可以分为2个方面: 基于路段层面与基于路径层面。基于路段层面的研究思路一般为先对组成路径的单个路段进行行程时间估计, 再由路段行程时间得到路径行程时间(刘张雷和史忠科, 2010; 马闻博, 2020; 千梦晗, 2020)。将路段行程时间的简单叠加来估计路径行驶时间方便快捷, 但同时也将带来包括路段行程时间误差的累积以及经过信控交叉口时可能的延误等。基于路径层面的研究将路径作为一个整体, 通过把握路径整体特征避免了路段误差的累积, 实现对路径行程时间更加精准的估计。李先通等(2022)对路径的空间特征展开研究, 考虑路段空间依赖、时序依赖以及粗颗粒度中时间飘移问题, 利用卷积神经网络及长短期记忆网络对路径行程时间进行估计; Han et al.(2022)提出多语义模型, 考虑了路径长度、交叉口数量及不同驾驶员个体对于路径行程时间的影响; Li et al.(2019)构建了深度生成模型, 结合路径的动静态特征给出了路径行程时间分布。此类方法利用数据驱动的方法, 结合大数据技术、深度学习模型对行程时间进行估计, 十分依赖于给定路径上的出行轨迹的数量是否充足。已有研究(冯轩, 2019)给出了稀疏数据下路径行程时间估计的方法, 但相关方面文献较少, 值得进一步探究。

基于路段层面的路径行程时间估计方法会积累各路段上行程时间估计的误差, 造成路径行程时间总体偏差较大; 在路径层面整体预测路径行程时间, 误差较小, 这类方法需要大量长出行轨迹数据的作为支撑, 但在实际中由于种种原因, 待研究长路径上的出行轨迹数据可能不足。针对此现象, 本文利用大量卡口检测数据探究路径可拼接的条件, 综合考虑路段与路径层面上进行行程时间估计的优劣, 提出用于构建目标路径的路径拼接方案的选择方法, 利用此方法对城市道路

中长路径行程时间进行估计, 以概率分布函数的形式记录各路径的行程时间, 再次使用时可直接利用, 为后续研究提供数据支撑, 具体步骤见图1。

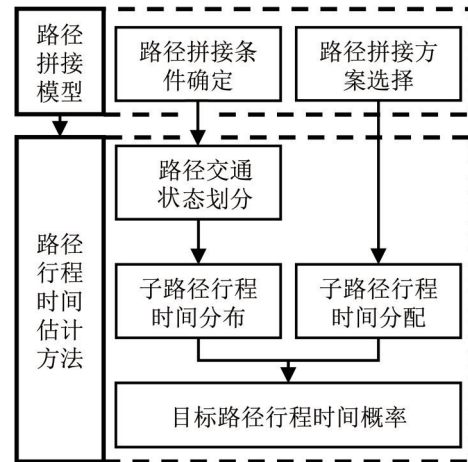


图1 基于路径拼接模型的行程时间估计方法
Fig. 1 Travel time estimation method based on path splicing model

1 路径拼接模型

本文提出一种路径拼接模型, 介绍了路径拼接时要满足的条件以及确定拼接方案的方法, 在交通状态相近的前提下将不同的短路径行程时间拼接得到待研究的长路径行程时间及其概率。首先根据研究需求选取待研究的目标路径, 再根据卡口检测数据计算各子路径所处路网状态, 提取目标路径所包含的子路径并计算行程时间, 分析子路径所处的交通状态, 选取若干满足路径拼接条件的子路径拼接为待研究的目标路径, 根据各子路径行程时间计算目标路径行程时间。

路径拼接旨在获取卡口检测数据中未记录到足够车辆出行轨迹的路径的行程时间, 利用卡口检测数据中若干有真实车辆出行轨迹覆盖的路径, 选取路径中与目标路径重合的部分, 在满足一定条件的前提下组合成目标路径。图2(a)为有真实出行轨迹覆盖的路径1~3; 图2(b)中所示的目标路径(蓝)为无真实出行轨迹覆盖的路径。为了得到目标路径的出行数据, 分别在路径1~3中选取结点O到A、结点A到C以及C到D的部分路径, 即可在空间上组合成目标路径。再根据不同路径在拼接结点(结点A和C)处交通状态是否相近确定该拼接方案是否成立。

1.1 路径拼接条件

将路径 L 表示为起点到终点经过的路段的序

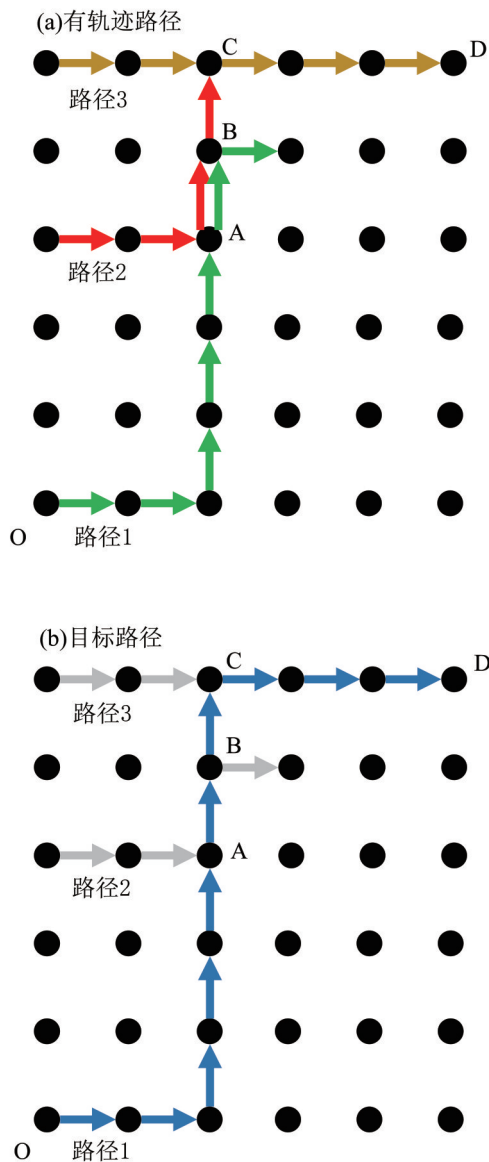


图2 路径拼接示意图

Fig. 2 Path splicing

列, 即路径 $L = \{l_1^2, l_2^3, \dots, l_n^{n+1}\}$. 假设路径目标路径 $L = \{l_1^2, l_2^3, \dots, l_n^{n+1}\}$ 可由 m 个不同的子路径 $L_1, L_2, L_3, \dots, L_m$ 组成, 即 $L = \{L_1, L_2, L_3, \dots, L_m\}$, 例如目标路径 $\{l_1^2, l_2^3, l_3^4\}$ 的一个拼接方案为 $L = \{\{l_1^2, l_2^3\}, \{l_3^4\}\}$, 即目标路径由子路径 $\{l_1^2, l_2^3\}$ 与 $\{l_3^4\}$ 组成. 为保证拼接得到的目标路径行程时间与现实相符, 相邻子路径的轨迹 L_i, L_{i+1} 应当在符合拼接所需的物理条件以及路径交通状态条件, 两方面的约束保证了拼接路径可以是真实存在的, 从而保证行程时间估计结果的准确性.

1) 物理约束. 物理条件保证了拼接路径在空间上是连续的, 用于拼接的两段相邻轨迹 L_i, L_{i+1} 必须有一部分是在空间上重合, 即两条轨迹至少

有一个共同的结点(拼接结点).

2) 交通状态约束. 交通状态反映了路网交通流运行情况, 可依据路径流量与密度判断交通状态通畅与否. 路径的行程时间受交通状态影响, 假设同一路径的路径路网状态相近时, 该路径上的行程时间服从相同的分布. 对于相邻的子路径 L_i, L_{i+1} , 要求下游子路径 L_{i+1} 在其出发的时刻所对应的实时路径路网状态与上游路径 L_i 到达拼接结点时路径 L_{i+1} 所对应的路径的初始路径路网状态相同.

设满足上述物理条件以及交通状态条件的拼接路径 $L = \{L_1, L_2, L_3, \dots, L_m\}$ 中各子路径的行程时间为 T_{L_i} , 则路径 L 的总行程时间为各子路径的行程时间之和.

1.2 拼接方案选择

对于一条由 n 个路段组成的路径而言, 最多可存在 $2^{n-1} - 1$ 种满足拼接条件的拼接方案, 以由 3 个路段组成的目标路径 $\{l_1^2, l_2^3, l_3^4\}$ 为例, 最多有 $\{\{l_1^2\}, \{l_2^3\}, \{l_3^4\}\}, \{\{l_1^2, l_2^3\}, \{l_3^4\}\}$ 和 $\{\{l_1^2\}, \{l_2^3, l_3^4\}\}$ 共 3 种可能的拼接方案. 将真实行程时间的均值作为行程时间的真实值, 为了使行程时间估计值与真实值之间偏差幅度最小, 选用各子路径行程时间方差之和最小的拼接方案. 拼接方案 $L = \{L_1, L_2, L_3, \dots, L_m\}$ 行程时间的整体方差为 Var , n_{L_i} 为经过子路径 L_i 的车辆数, $T_{L_i}^j$ 为第 j 辆车经过子路径 L_i 所用时间, \bar{T}_{L_i} 为经过子路径 L_i 的行程时间均值, 则 Var 的计算方法为:

$$\text{Var}(L, L_1, L_2, L_3, \dots, L_m) = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^{n_{L_i}} \frac{(T_{L_i}^j - \bar{T}_{L_i})^2}{n_{L_i}}.$$

选择行程时间误差最小的拼接方案可转为求解行程时间的整体方差最小的优化问题:

$$\begin{aligned} & \arg \min_{L_1, L_2, \dots, L_m} \text{Var}(L, L_1, L_2, \dots, L_m) \\ & \text{s.t. } L = \{L_1, L_2, L_3, \dots, L_m\}. \end{aligned}$$

使用遗传算法(靳文舟等, 2020)求解上述问题, 将拼接方案编码为一个 0-1 字符串, 涉及到 n 个路段的拼接问题染色体长度为 $n - 1$, 每一位用字符 0 或 1 表示对应相邻路段是否被打断. 如图 3 所示, 以目标路径为 $L' = \{l_1, l_2, l_3, l_4\}$ 为例, 目标路径由 4 个路段组成, 则其染色体长度为 3, 对应 4 个路段形成的 3 个路口, 染色体“101”表示第 1 个和第 3 个路口不打断, 第 2 个路口打断, 即拼接方案为 $L' = \{\{l_1, l_2\}, \{l_3, l_4\}\}$.

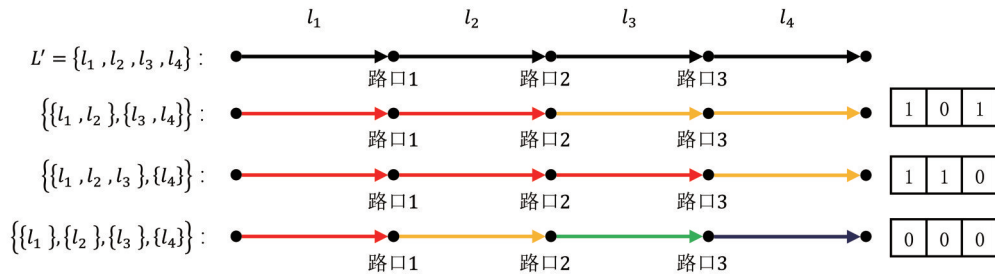


图 3 路径拼接方案编码示意图

Fig. 3 Example of chromosome composition

适应度代表了染色体存活概率的大小, 将适应度函数设置为 $F_i = -\text{Var}(L, L_1, L_2, \dots, L_m)$, 采用“轮盘赌”选择方法确定下一代个体, 则方案的整体方差越小染色体被选中的概率越大; 染色体交叉过程采用二点交叉法, 在两条父代染色体上分别随机抽取长度相等的两条染色体片段并交换; 变异过程采用点位变异, 随机选择多个基因依照设置的概率在子代染色体上进行变异, 由于采用了 0-1 编码, 变异时将基因编码取反即可。

2 路径行程时间估计

2.1 交通状态划分

交通流基本图模型由 Greenshields et al.(1935) 提出, 描述了宏观交通运行过程中流量、密度和速度之间关系。宏观基本图由 Daganzo(2007) 提出, 相较于传统交通流基本图, 宏观基本图更关注道路网络层面上交通流 3 参数之间的关系和性质。参考传统基本图与宏观基本图, 提出介于二者之间的描述路径 3 参数间关系的路径交通流基本图, 用于对路径所处的交通状态进行划分。参考 Edie(1963) 提出的交通流参数的观测方法, 定义路径 L 的流量:

$$Q_L = \frac{\sum_{i=1}^n [\text{lanenum}_{l_i^{i+1}} \cdot \text{len}_{l_i^{i+1}} \cdot Q(l_i^{i+1}, t_i^t)]}{\sum_{i=1}^n (\text{lanenum}_{l_i^{i+1}} \cdot \text{len}_{l_i^{i+1}})}, \quad (1)$$

路径 L 的密度为

$$K_L = \frac{\sum_{i=1}^n [\text{lanenum}_{l_i^{i+1}} \cdot \text{len}_{l_i^{i+1}} \cdot K(l_i^{i+1}, t_i^t)]}{\sum_{i=1}^n (\text{lanenum}_{l_i^{i+1}} \cdot \text{len}_{l_i^{i+1}})}, \quad (2)$$

其中 $\text{len}_{l_j^k}$ 为路段 l_j^k 的长度, $\text{lanenum}_{l_j^k}$ 为路段 l_j^k 进口道的车道数, t_i^t 为路径 L 上第 i 个路段起始节点对应的时刻, $Q(l_i^{i+1}, t_i^t)$ 与 $K(l_i^{i+1}, t_i^t)$ 分别为 t_i^t 时刻路段 l_i^{i+1} 上的流量与密度。图 4 为路径交通流基本图, 图中各点表示同一路径上路径流量与路径密度的对应关系。随着路径密度的增大, 路径流量

也随之上升, 当路径密度增大到临界密度 k_m^L 后, 路径流量随密度增大呈下降的趋势。

根据 Greenshields 模型中交通流速度与密度存在的正比关系, 可以推出流量与密度间存在二次函数关系, 利用二次函数拟合路径流量与路径密度之间的关系, 图 4 中绘制了拟合后的曲线。设路径 L 上流量与密度的关系为:

$$Q_L = aK_L^2 + bK_L + c,$$

则临界密度 $k_m^L = -(b/2a)$ 。根据路径 L 在时刻 t 的密度划分其所处的交通状态 $C_L(t)$, 参考 Lin(2019) 的密度划分依据, 可将 $C_L(t)$ 划分为畅通、基本畅通、拥堵、严重拥堵 4 类, 具体划分方法为

$$C_L(t) = \begin{cases} \text{畅通}, & K_L \in [0, 0.2k_m^L), \\ \text{基本畅通}, & K_L \in [0.2k_m^L, 0.6k_m^L), \\ \text{拥堵}, & K_L \in [0.6k_m^L, k_m^L), \\ \text{严重拥堵}, & K_L \in [k_m^L, \infty). \end{cases} \quad (3)$$

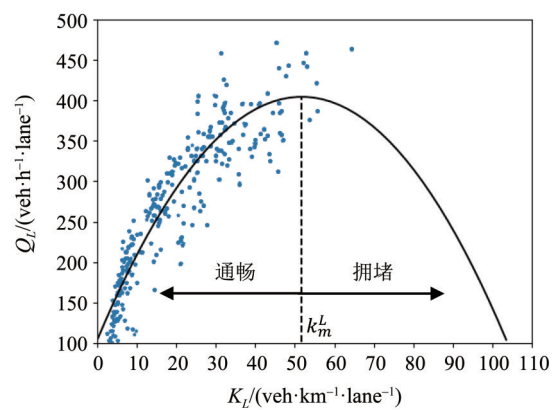


图 4 路径交通流基本图示意图

Fig. 4 Basic diagram of a path

2.2 考虑交通状态的路径行程时间分布

路径行程时间在不同交通状态下会呈现不同分布, 分别对路径在不同交通状态下的行程时间分布进行拟合, 得到路径 L 在交通状态 $C_L(t)$ 下的

路径行程时间概率密度函数为 $f(T_L, C_L)$, 则行程时间小于 T_x 的概率

$$P(0 \leq T_L < T_x) = \int_0^{T_x} f(T_L, C_L) dT_L.$$

由于出行者对连续变化的时间感知度有限, 为了便于后续计算, 可以将行程时间分布离散化处理。将行程时间为 $[T_x - \tau, T_x)$ 的概率作为行程时间为 T_x 的概率, 以 $P(T_x, C_L)$ 表示路径 L 上路径行程时间为 T_x 的概率, 即

$$\begin{aligned} P(T_x, C_L) &= P(T_x - \tau \leq T_L < T_x) \\ &= \int_{T_x - \tau}^{T_x} f(T_L, C_L) dT_L, \end{aligned}$$

其中 $T_x = h\tau$, $h = 1, 2, 3, \dots$, T_x 单位为s.

2.3 路径行程时间概率

在出发时间已知的情况下, 给定目标行程时间, 则概率计算步骤:

1) 将目标路径分解, 选取所有拼接方案中Var最小的拼接方案, 为该方案中各子路径分配行程时间使得各子路径的行程时间之和为给定的目标行程时间;

2) 根据子路径的起始时刻所对应的交通状态及相应的概率分布函数, 计算各子路径对应行程时间的概率, 将每段子路径行程时间概率相乘即得到该时间分配方案下路径行程时间的概率, 累加所有时间分配方案的概率, 即为目标路径的行程时间为给定值的概率(图5)。

假设待研究的拼接方案为

$$L = \{L_1, L_2, L_3, \dots, L_m\},$$

将目标路径行程时间视为所选拼接方案下各子路径的行程时间之和, 即 $T_L = \sum_{i=1}^m T_{L_i}$, 其中 T_{L_i} 为各子路径的行程时间。给定目标行程时间 T_x , 各子路径的行程时间 T_{L_i} 存在多种组合方式, 由于子路径的行程时间是离散化的, 因此可以列出所有可能的子路径行程时间组合。设每个子路径上存在最小行程时间 $T_{L_i}^{\min}$, 则当 $T_x \geq \sum_{i=1}^m T_{L_i}^{\min}$ 时, 给定行程时间是合理的, 存在一个概率值与之对应。设可用于分配的时间长度为

$$\Delta T_x = T_x - \sum_{i=1}^m T_{L_i}^{\min} = \alpha\tau,$$

在最小行程时间的基础上为每个子路径分配行程时间增量 ΔT_{L_i} , 则共有 m^α 种时间分配方案使得 $\Delta T_x = \sum_{i=1}^m \Delta T_{L_i}$, 故各子路径的行程时间 $T_{L_i} = T_{L_i}^{\min} + \Delta T_{L_i}$. 子路径上行程时间为 T_{L_i} 的概率为 $P(T_{L_i}, C_{L_i})$, 将该拼接方案下所有可能的子路径行

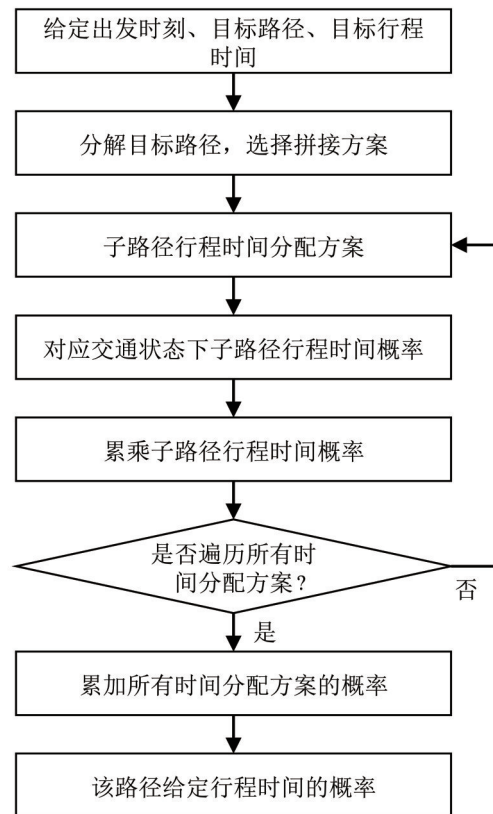


图5 路径行程时间概率计算过程

Fig. 5 Process of path travel time probability

程时间组合的概率和为目标路径行程时间为 T_x 的概率为

$$\sum_{T_x = \sum_{i=1}^m T_{L_i}} \prod_{i=1}^m P(T_{L_i}, C_{L_i}).$$

3 案例分析

为了验证基于路径拼接的行程时间估计方法可行性及准确性, 选取广州市工业大道、南田路、前进路、东晓路及新滘路所围的区域内, 于2021年8月29日卡口过车数据进行实例研究。研究区域内共有可观测路口21个, 可观测路段32条, 见图6。卡口过车数据记录车辆车牌、所处路段、进入和离开路段的时刻等信息, 结合路网拓扑结构提取车辆出行记录。剔除车辆出行轨迹中路径平均速度小于5 km/h或大于120 km/h的异常值后, 研究区域内单日提取车辆出行20万余次, 单次出行路径长度均值约为2 km, 单次出行时间均值为664 s。

选取图6中绿色路径为目标路径, 起点A为南田路与工业大道交叉口, 终点B为昌岗路与东晓路交叉口, 该路径长约4.5 km, 2021年8月29日目标路径上共有真实出行轨迹1232条, 平均

路径行程时间为 816.8 s。一天内该路径的行程时间分布如图 7 所示。

3.1 宏观基本图拟合

以目标路径 L 的拼接方案

$$L = \{L_1, L_2, L_3, L_4\}$$

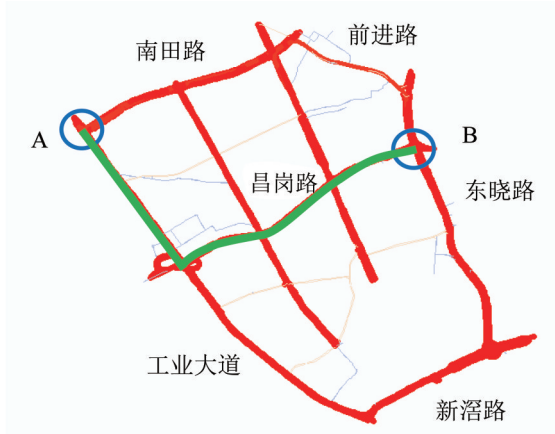


图 6 研究区域

Fig. 6 Study area

为例, 依据式(1)和(2)计算各子路径 L_i 的流量与密度, 并绘制各子路径的路径交通流基本图(图 8)。

绘制目标路径及其子路径的路径交通流基本图并使用二次函数进行拟合, 选取确定系数 R^2 以及均方根误差 RMSE 表征拟合优度。其中 $R^2 \in [0, 1]$, 越接近 1 则拟合程度越高; RMSE

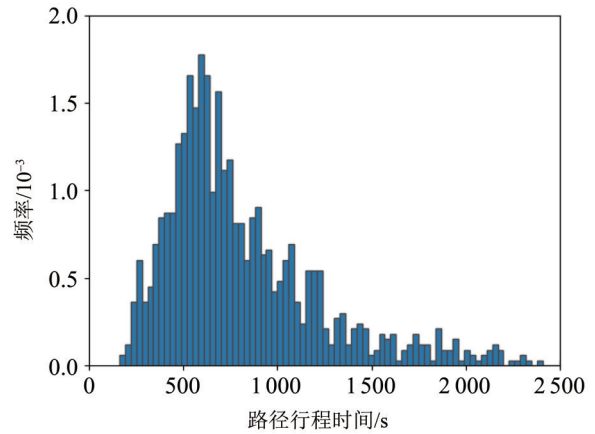


图 7 真实轨迹行程时间分布

Fig. 7 Travel time distribution of real trajectory

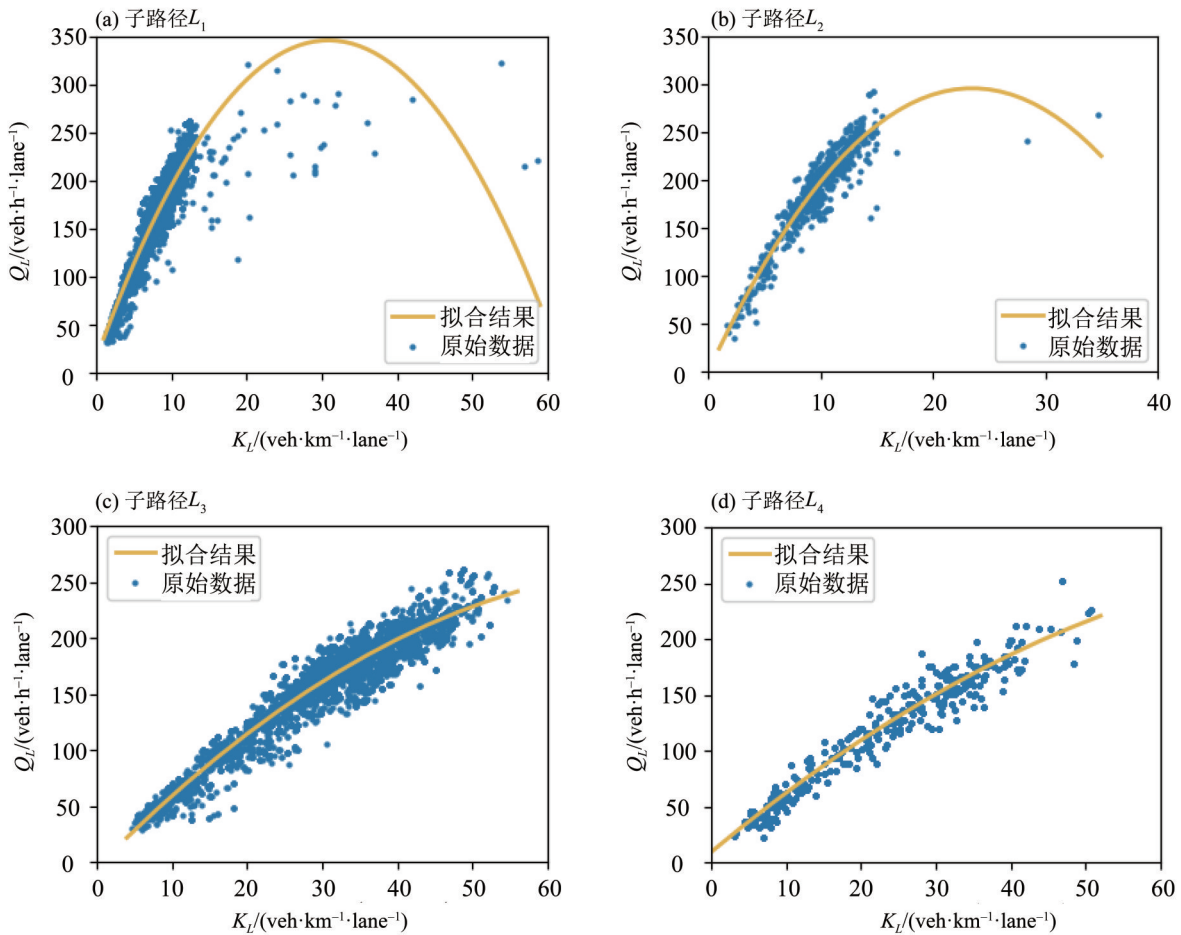


图 8 子路径交通流基本图

Fig. 8 Basic diagram of sub-paths

用于衡量拟合值与真实值之间的偏差;拟合情况见表1。由表1可知,利用二次函数拟合的结果中 $R^2>0.85$, $RMSE>10$,拟合度较好。

3.2 行程时间分布函数拟合

按照式(3)将子路径划分为“畅通”“基本畅

通”“拥堵”和“严重拥堵”等4种交通状态,以子路径 L_2 为例,不同交通状态下子路径行程时间分布情况如图9所示。

表1 路径交通流基本图拟合情况(部分)

Table 1 Fitting of basic diagram of path

| 子路径 | 拟合参数 | | | R^2 | RMSE |
|-------|--------|--------|--------|-------|--------|
| | a | b | c | | |
| L_1 | -0.348 | 21.498 | 14.375 | 0.875 | 12.854 |
| L_2 | -0.536 | 25.221 | -0.790 | 0.892 | 12.286 |
| L_3 | -0.694 | 27.325 | -5.157 | 0.913 | 11.039 |
| L_4 | -0.470 | 22.385 | 8.991 | 0.903 | 10.334 |

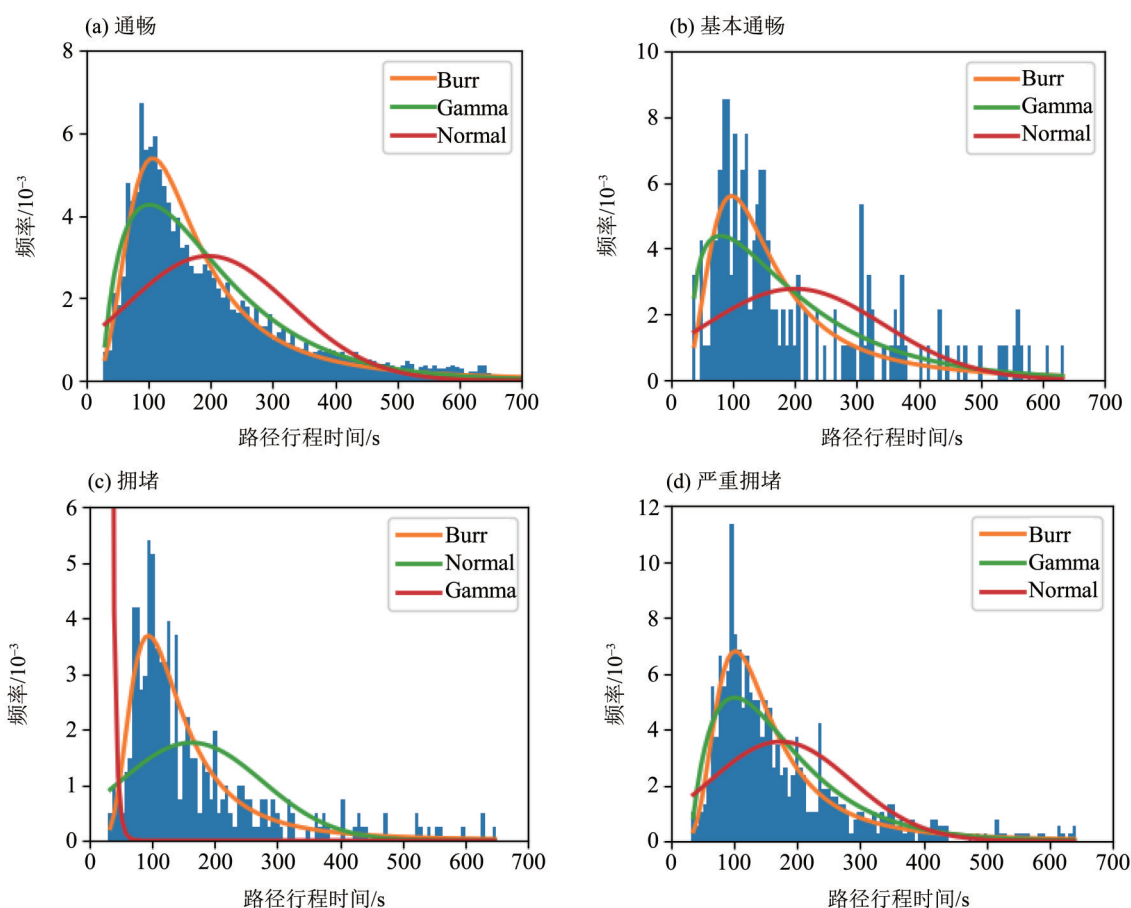


图9 不同交通状态下行程时间分布

Fig. 9 Travel time distributions under different traffic status

分别利用 Gamma、Normal 以及 Burr 分布对行程时间分布进行拟合,利用残差平方和评价分布拟合效果。残差平方和越小,则拟合效果越好。子路径在不同交通状态下的残差平方和见表2。由表2可知,Burr 分布在各种交通状态下拟合的残差

平方和均为最小,因此采用 Burr 分布对所有状态下的行程时间分布进行拟合。特别地,若交通状态为“严重拥堵”时的路径行程时间数据量不足时,则认为该路段不存在此状态,在进行路径拼接时不考虑该状态。

3.3 路径行程时间概率分布

将 2021 年 8 月 29 日 8:00 时刻的路径路网状态作为输入, 分别计算目标路径上行程时间为 0~2 000 s 的概率, 得到目标路径行程时间估计值分布。将其与该路径上 2021 年 8 月 29 日 8:00~9:00 时段内轨迹的行程时间真实值分布进行对比, 结果见图 10, 部分行程时间对应的概率密度与累计

概率如表 3 所示。在此时间段内, 真实出行轨迹的行程时间均值为 714 s, 本方法所得行程时间估计值均值为 693 s, 误差 3.04%。

使用 JS 散度(Chen & Liu, 2021)进一步判断上述 2 种分布的相似性。JS 散度是基于 KL 散度提出的一种概率统计方法(Kullback & Leibler, 1951), 在信息论中广泛应用于定量衡量两个概率

表 2 子路径在不同交通状态下路径行程时间分布拟合残差平方和

Table 2 RSS of travel time distribution of sub routes under different traffic status

10^{-6}

| 子路径 | 通畅 | | | 基本通畅 | | | 拥堵 | | | 严重拥堵 | | |
|-------|-------|--------|------|-------|--------|------|-------|--------|------|-------|--------|------|
| | Gamma | Normal | Burr | Gamma | Normal | Burr | Gamma | Normal | Burr | Gamma | Normal | Burr |
| L_1 | 5 | 35 | 3 | 28 | 52 | 25 | 395 | 110 | 59 | 28 | 84 | 18 |
| L_2 | 25 | 114 | 13 | 220 | 369 | 175 | 2 835 | 337 | 95 | 96 | 238 | 58 |
| L_3 | 35 | 96 | 15 | 196 | 325 | 138 | 686 | 420 | 86 | \ | \ | \ |
| L_4 | 135 | 26 | 5 | 61 | 118 | 50 | 774 | 1 067 | 656 | \ | \ | \ |

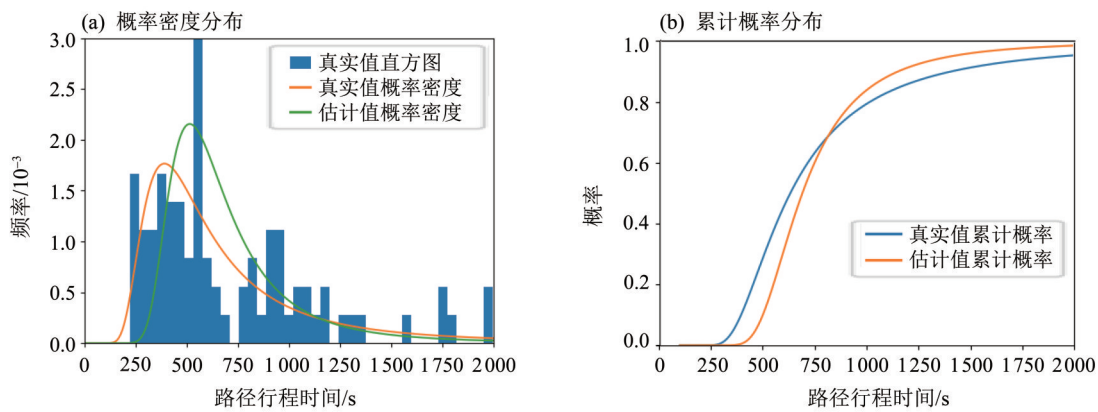


图 10 目标路径行程时间真实值与估计值对比

Fig. 10 Comparison of target path travel time distributions

表 3 行程时间真实值与估计值对应累计概率与概率密度

Table 3 Cumulative probability and probability density of different travel times

| 行程时间/s | 真实值累计概率 | 估计值累计概率 | 真实值概率密度/ 10^{-3} | 估计值概率密度/ 10^{-3} |
|--------|---------|---------|--------------------|--------------------|
| 600 | 0.536 | 0.438 | 1.220 | 1.790 |
| 630 | 0.571 | 0.490 | 1.120 | 1.660 |
| 660 | 0.603 | 0.538 | 1.030 | 1.520 |
| 690 | 0.633 | 0.581 | 0.941 | 1.380 |
| 720 | 0.660 | 0.620 | 0.860 | 1.250 |
| 750 | 0.685 | 0.656 | 0.785 | 1.120 |
| 780 | 0.707 | 0.688 | 0.716 | 1.010 |
| 810 | 0.728 | 0.716 | 0.654 | 0.901 |
| 840 | 0.746 | 0.742 | 0.597 | 0.806 |
| 870 | 0.763 | 0.764 | 0.546 | 0.722 |
| 900 | 0.779 | 0.785 | 0.500 | 0.646 |

分布间的差异性。离散型分布 U, V 的 KL 散度为

$$D_{\text{KL}}(U \| V) = \sum_{x \in X} U(x) \log \left(\frac{U(x)}{V(x)} \right).$$

KL 散度具有非对称性, JS 散度在 KL 散度基础上消除了非对称性, 计算公式为

$$\begin{aligned} \text{JSD}(U \| V) &= \text{JSD}(V \| U) \\ &= \frac{1}{2} D_{\text{KL}}(U \| V) + \frac{1}{2} D_{\text{KL}}(V \| U). \end{aligned} \quad (4)$$

JS 散度取值在 0~1 之间, 相同分布的 JS 散度为 0, JS 散度越大, 两个分布间的相似性越小。按照式(4)计算可得目标路段行程时间分布的估计值与实际值的 JS 散度为 0.05, 可以认为本文方法得到的行程时间分布与真实行程时间分布的趋势一致, 同一行程时间所对应的概率相近。因此, 从行程时间均值与分布两方面来看, 本方法对行程时间估计的

结果可靠, 可为后续相关研究提供数据基础。

4 总 结

针对行程时间相关研究中可能出现的路径行程时间数据量不足的问题, 本文提出了基于路径拼接模型的路径行程时间估计方法, 通过路径拼接模型确定路径拼接方案。使用 Burr 分布拟合不同路网状态下的路径行程时间分布, 结合路径拼接方案给出目标路径行程时间及其概率分布。该方法计算得到的行程时间与实际行程时间的均值误差较小, 所得行程时间分布与实际分布相似, 对路径行程时间的估计结果较为准确, 可为后续行程时间可靠性、个性化路径推荐等基于路径行程时间的研究提供数据支撑。

参考文献:

- 冯轩, 2019. 稀疏轨迹数据下的行程时间估计方法研究[D]. 长沙: 长沙理工大学.
- 靳文舟, 朱子轩, 巫威眺, 等, 2020. 联合配送外包策略的定位运输问题[J]. 中山大学学报(自然科学版), 59(4): 89-97.
- 李先通, 全威, 王华, 等, 2022. 基于时空特征深度学习模型的路径行程时间预测[J]. 吉林大学学报(工学版), 52(3): 557-563.
- 刘晓波, 蒋阳升, 唐优华, 等, 2019. 综合交通大数据应用技术的发展展望[J]. 大数据, 5(3): 55-68.
- 刘张雷, 史忠科, 2010. 一种基于路网变化的动态路径规划策略[J]. 交通运输系统工程与信息, 10(3): 147-152.
- 马闻博, 2020. 基于车牌识别数据的城市道路行程时间估计及短时预测研究[D]. 成都: 西南交通大学.
- 千梦晗, 2020. 城市干道行程时间动态估计及可靠性计算方法[D]. 重庆: 重庆交通大学.
- CHEN Z, LIU X C, 2021. Statistical distance-based travel-time reliability measurement for freeway bottleneck identification and ranking[J]. Transp Res Rec, 2675(11): 424-438.
- DAGANZO C F, 2007. Urban gridlock: Macroscopic modeling and mitigation approaches[J]. Transp Res B: Methodol, 41(1): 49-62.
- EDIE L C, 1963. Discussion of traffic stream measurements and definitions[M]. New York: Port of New York Authority.
- GREENSHIELDS B D, CHANNING W, MILLER H, 1935. A study of traffic capacity[C]//Proceedings of the Fourteenth Annual Meeting of the Highway Research Board. Washington D C.
- HAN L, DU B, LIN J, et al, 2022. Multi-semantic path representation learning for travel time estimation[J]. IEEE Trans Intell Transp Syst, 23(8): 13108-13117.
- KULLBACK S, LEIBLER R A, 1951. On information and sufficiency[J]. Ann Math Statist, 22(1): 79-86.
- LI X, CONG G, SUN A, et al, 2019. Learning travel time distributions with deep generative model[C]//The World Wide Web Conference. USA.
- LIN X, 2019. A road network traffic state identification method based on macroscopic fundamental diagram and spectral clustering and support vector machine[J]. Math Probl Eng, DOI:10.1155/2019/6571237.

(责任编辑 秦社彩)