引文格式：李萌,王伊丽,戴建光. 基于车牌识别数据的稀疏车辆出行轨迹还原研究[J]. 地理空间信息, 2026, 24(7): 33-37. doi:10. 3969/j.issn. 1672-4623.2026.07.007

基于车牌识别数据的稀疏车辆出行轨迹还原研究

李　萌[1]，王伊丽[2*]，戴建光[1]
（1. 常州市武进规划勘测设计院，江苏 常州 213161；2. 常州市武进规划管理服务中心，江苏 常州 213161）

摘　要：针对部分区域电警卡口设备分布稀疏、还原车辆出行轨迹难度大的问题，研究了一种基于车牌识别数据的稀疏车辆出行轨迹还原算法。首先对车牌识别数据进行清洗，并基于车牌号、检测时间排序的车牌识别数据提取车辆出行链；再通过出租车 GPS 数据标定分时段道路运行速度，根据道路运行速度上下限、出行最小停留时间阈值，将车辆出行链分成一个或多个 OD；然后以时间最小化为目标，运用 K 短路算法生成车辆候选出行轨迹集合，并基于 Max-Min 法对指标进行归一化；最后以候选出行轨迹指标组成的多维向量为输入，驾驶员选择的路径为输出，构建基于全连接神经网络的车辆出行轨迹决策模型，还原缺失轨迹。利用武进区某区域实际车牌识别数据对模型进行训练和测试，结果表明该算法的稀疏车辆出行轨迹还原综合准确率为 75.0%，比最短路径法、自编码器模型高 32.9 个百分点和 37.0 个百分点，验证了其准确性和良好的泛化能力。

关键词：车牌识别；稀疏轨迹；轨迹还原；全连接神经网络
中图分类号：P228　　**文献标志码：**B　　**文章编号：**1672-4623（2026）07-0033-05

Research on Sparse Vehicle Trajectory Reconstruction Based on License Plate Recognition Data

LI Meng[1], WANG Yili[2], DAI Jianguang[1]
(1. Changzhou Wujin Planning and Surveying Design Institute, Changzhou 213161, China;
2. Changzhou Wujin Planning and Management Service Center, Changzhou 213161, China)

Abstract: Aiming at the problems of sparse distribution of electrical police devices in some areas and the difficulty in reconstructing vehicle trajectories, we proposed a sparse vehicle trajectory reconstruction algorithm based on license plate recognition data. After cleaning the license plate recognition data, we extracted the vehicle travel chain from the license plate recognition data sorted by license plate and detection time. Then, we used taxi GPS data to calibrate road speeds in different time periods. Based on the upper and lower limits of road speeds and the minimum travel time threshold, the vehicle travel chain was divided into one or more ODs. Aiming to minimize travel time, we employed the K-shortest path algorithm to generate a set of candidate vehicle trajectories, and used the Max-Min method to normalize indicators. Finally, taking the multi-dimensional vector of candidate trajectory indicators as input and the path chosen by the driver as output, we constructed a vehicle trajectory decision-making model based on a fully connected neural network, and reconstructed the missing trajectories. We used actual license plate recognition data from a certain area of Wujin District, Changzhou City, Jiangsu Province to train and test this model. The results indicate that the proposed algorithm achieves a composite accuracy of 75.0% in reconstructing sparse vehicle trajectories, which is 32.9 percentage points higher than the shortest path method and 37.0 percentage points higher than the autoencoder model, verifying the accuracy and good generalization ability of this algorithm.

Key words: license plate recognition, sparse trajectory, trajectory reconstruction, fully connected neural network

电警卡口视频识别的车牌数据在城市交通中应用广泛。利用车牌识别数据可提取车辆出行链、溯源道路断面流量、标定道路运行速度，进而分析城市机动车出行需求、道路交通运行状况，支撑城市宏中微观交通模型参数标定与结果校核。受设备安装位置、天气情况、识别算法精度等因素影响，部分车辆轨迹数据不完整，需要设计相应算法还原车辆出行轨迹。国内外学者[1-4]对车辆出行轨迹还原进行了大量研究；部分学者[5-7]运用逼近理想方案的序数偏好方法（TOPSIS）、灰色关联法等经典多目标决策方法还原车辆出行轨

项目来源：深圳市科技重大专项资助项目（KJZD20230923115223047）。
第一作者简介：李萌（1983—），硕士，高级城市规划师，从事交通空间规划、交通时空数据分析、交通模型仿真等工作，E-mail：65944141@qq.com。
通信作者：王伊丽（1983—），硕士，高级城市规划师，从事交通空间规划、交通时空数据分析、交通模型仿真等工作，E-mail：108648611@qq.com。

迹；部分学者[8-11]以贝叶斯理论为基础研究了车辆轨迹还原算法；部分学者[12-13]引入自编码器等无监督学习算法解决车辆出行轨迹还原问题，并采用完整历史轨迹数据定义轨迹偏好。基于自编码器的车辆出行轨迹选择算法通过历史轨迹分布数据定义驾驶员出行路径偏好，适用于设备密集区域的车辆轨迹还原，候选路径较少，而在设备分布稀疏的区域，轨迹缺失率较高，难以获取完整的车辆历史轨迹数据，实际可选路径多，算法无法直接迁移使用。近年来部分学者[14-16]利用时空数据智能分析技术研究车辆出行轨迹，取得了一定的成果。针对上述算法的局限性，本文提出了一种基于车牌识别数据的稀疏车辆出行轨迹还原算法。根据浮动车GPS数据标定路段行程时间，通过车牌识别数据提取机动车出行链，并利用速度和时间阈值切分车辆OD；然后利用K短路算法生成候选车辆轨迹，依托设备分布密集区域的完整轨迹数据训练基于全连接神经网络的车辆出行路径决策模型，并将训练后的路径决策模型应用于设备稀疏区域，进行稀疏车辆出行轨迹的路径选择；最后以武进区某区域实际车牌识别数据为例验证算法的有效性。

1 数据预处理

本文采用车牌识别数据、道路网络数据和出租车GPS数据等还原车辆出行轨迹。由于受多种因素干扰，上述数据中可能存在错误信息，使用前需进行预处理。此外，不同来源的数据间存在一定相关关系，需要进行关联。

1）关联电警卡口设备与路网节点。根据电警卡口设备的地理位置，与对应的路网交叉口节点进行初步匹配，再结合电警卡口对应的道路名称对关联结果进行人工校核，验证各设备匹配的交叉口节点是否合理。若发现初步匹配的卡口名称与实际位置不符，应进行人工校正。由于城市路网和交通管理设施通常处于动态变化中，不同时期的路网结构与电警卡口设备之间易出现匹配偏差，因此应尽量确保路网数据、电警卡口设备信息和车牌识别数据的采集时间一致，以提高关联结果的准确性和可靠性。

2）剔除无车牌、问题车牌、仅检测到一次车牌等数据。设备距离拍摄车辆较远、夜晚多车场景、车牌遮挡等因素均会影响车牌识别结果[17]，产生部分无车牌、问题车牌等异常数据，需要进行剔除。此外，仅检测到一次的车牌数据因缺少车辆出行起终点无法进行路径还原，也需剔除。

3）剔除出租车、网约车车牌数据。由于出租车和网约车的行驶路径可能受拼车、打车平台规划路径等因素影响，与私家车存在较大差异，因此需剔除。

4）标定分时段道路平均运行速度。为精细化分析道路运行状况，基于出租车GPS数据标定分时段（以30 min为间隔）、分路段的平均运行速度。为处理缺少GPS数据的路段，将分时段、分道路等级的路段平均运行速度作为路段速度的缺省值。最后计算得到所有路段分时段平均运行时间。

2 出行OD切分

本文分别基于车牌编码、过车时间对车牌数据进行分组、排序，获取车辆途径电警设备集合；再根据电警卡口设备与路网节点的关联关系，构建每辆车的出行链条和路网节点集合，包括车牌号、经过的路网节点和经过时间等信息。车辆出行轨迹还原前需将出行链拆分为若干个独立的出行OD。基于相邻节点之间的行驶速度与时间阈值确定出行链的分离位置。对于一辆经过的电警卡口点位序列为 $S_c=(s_1,s_2,\cdots,s_{n-1},s_n)$ 的车辆，采用最短路径算法计算两个相邻电警卡口点 s_i 和 s_{i+1} 的最短路径距离 d，s_i 与 s_{i+1} 的时间差 t 可由检测时间相减得到，其平均速度为 d 与 t 的比值。假设速度运行阈值上下限为 v_l、v_h，出行最小停留时间阈值为 $\triangle t_{\min}$，若 $\bar{v}_{i,i+1}>v_h$，则 s_i 与 s_{i+1} 之间的出行速度异常，剔除 s_{i+1}；若 $\bar{v}_{i,i+1}<v_l$ 且 $t-t_{\min}>\triangle t_{\min}$，则将 s_i 标记为上一段OD的终点，s_{i+1} 标记为下一段OD的起点。

3 车辆出行轨迹还原

3.1 候选车辆轨迹生成

在电警卡口设备稀疏区域，出行OD各相邻电警卡口间的出行轨迹通常经过多个交叉口，车辆存在较多出行路径，驾驶员选择出行路径时通常受出行时间、出行距离、信号交叉口数量、转弯次数、驾驶员出行偏好等因素影响，不一定选择最短路径，但会选择出行时间较短的多条路径之一，因此出行时间较短的多条路径即为候选车辆出行轨迹集合。利用K短路算法[18]根据时间搜索 s_i 和 s_{i+1} 之间的前 K 条最短路

径，并对出行时间进行排序，构建候选车辆出行轨迹集合。

3.2　车辆轨迹指标

出行轨迹的补全需要合理模拟出行者的路径决策过程。出行者在出行决策过程中考虑的因素主要包括路径距离、路径时间一致性、路径选择偏好、信号交叉口数量、转弯次数、道路等级等[6]。由于电警卡口设备稀疏区域通常缺少完整路径数据，难以计算电警卡口间路径选择偏好指标，且交叉口左转和右转的平均期望时间不同，因此本文选择路径距离、路径时间差、信号交叉口数量、左转次数、右转次数、道路等级6个指标进行车辆路径决策。对比Max-Min法、基于指数的Max-Min法和Z-Score法[19]的指标归一化效果可知，Max-Min法的线性变换过程简单，有利于大规模轨迹数据的高效处理；且不要求数据符合特定分布，更适用于路径指标中存在偏分布或边界明显的情况，便于统一各指标量纲并保持原始数据之间的相对关系。因此，本文采用Max-Min法对指标进行归一化。

路径距离的归一化公式为：

$$x_1(p_i)=\begin{cases}\dfrac{L_i-\min(L)}{\max(L)-\min(L)}, & \max(L)\neq\min(L)\\ 0, & \max(L)=\min(L)\end{cases} \tag{1}$$

式中，L_i 为路径 p_i 的距离；L 为路径距离集合。

路径时间差的归一化公式为：

$$\Delta T_i=\left|T_i-T_{\text{true}}\right| \tag{2}$$

$$x_2(p_i)=\begin{cases}\dfrac{\Delta T_i-\min(\Delta L)}{\max(\Delta L)-\min(\Delta L)}, & \max(\Delta L)\neq\min(\Delta L)\\ 0, & \max(\Delta L)=\min(\Delta L)\end{cases} \tag{3}$$

式中，T_i 为路径 p_i 的预测时间；T_{true} 为实际时间，即相邻电警卡口设备的检测时间之差；ΔT_i 为路径 p_i 预测时间与实际时间之差的绝对值。

信号交叉口数量的归一化公式为：

$$x_3(p_i)=\begin{cases}\dfrac{N_i-\min(N)}{\max(N)-\min(N)}, & \max(N)\neq\min(N)\\ 0, & \max(N)=\min(N)\end{cases} \tag{4}$$

式中，N_i 为路径 p_i 经过的信号交叉口数量；N 为路径信号交叉口数量集合。

左转次数的归一化公式为：

$$x_4(p_i)=\begin{cases}\dfrac{\text{Left}_i-\min(\text{Left})}{\max(\text{Left})-\min(\text{Left})}, & \max(\text{Left})\neq\min(\text{Left})\\ 0, & \max(\text{Left})=\min(\text{Left})\end{cases} \tag{5}$$

式中，Left_i 为路径 p_i 左转次数；Left 为路径左转次数集合。

右转次数的归一化公式为：

$$x_5(p_i)=\begin{cases}\dfrac{\text{Right}_i-\min(\text{Right})}{\max(\text{Right})-\min(\text{Right})}, & \max(\text{Right})\neq\min(\text{Right})\\ 0, & \max(\text{Right})=\min(\text{Right})\end{cases} \tag{6}$$

式中，Right_i 为路径 p_i 右转次数；Right 为路径右转次数集合。

道路等级的归一化公式为：

$$H_i=\frac{\sum_j^n H_{ij}\times L_{ij}}{\sum_j^n L_{ij}} \tag{7}$$

$$x_6(p_i)=\begin{cases}\dfrac{H_i-\min(H)}{\max(H)-\min(H)}, & \max(H)\neq\min(H)\\ 0, & \max(H)=\min(H)\end{cases} \tag{8}$$

式中，H_i 为路径 p_i 的道路等级；n 为路径 p_i 中包含的路段数；H_{ij} 为路径 p_i 中第 j 段道路的等级；L_{ij} 为路径 p_i 中第 j 段道路的长度；H 为路径道路等级集合。

3.3　候选车辆轨迹生成

综合考虑候选车辆出行轨迹集合的各项指标后选择最优出行路径的过程本质上是一个多属性决策问题。经典的多属性决策方法包括层次分析法、主成分分析、灰色关联分析、TOPSIS方法等，这些方法将路径特征降维后再进行路径决策，带有一定的主观性。部分学者采用自编码器、变分自编码器等对路径特征进行降维，但降维过程易产生信息损失。由于路径各项特征对车辆出行决策的影响是非线性的，存在非线性可分的特点，因此本文采用神经网络对车辆出行选择进行监督学习。全连接神经网络作为深度学习的主要网络形式之一，可以拟合任何函数[20]。本文设计了一种基于全连接神经网络的车辆出行轨迹决策模型，通过标记数据对模型进行训练，可以更好地捕捉输入特征与输出之间的关系。为防止过拟合现象，在神经网络中加入Dropout层来随机丢失神经元节点[21]，以提高模型预测精度与泛化能力。

车辆出行轨迹决策模型由3个部分组成：第一部分通过两个隐藏层构建路径特征提取网络，提取每条路径的特征；第二部分通过两个隐藏层创建比较器模型，用于处理路径间的关系，采用ReLU函数作为激活函数；第三部分通过1个输出层计算备选路径选择概率，采用Softmax函数作为激活函数。车辆出行轨

迹决策模型将由候选轨迹指标组成的多维向量作为输入，驾驶员选择的最佳路径作为输出；优化器采用自适应学习率优化算法，根据参数的历史梯度自动调整学习率；损失函数采用适用于多类分类问题的稀疏类别交叉熵，计算实际标签和预测概率之间的交叉熵损失。

$$\mathrm{Loss}=-\frac{1}{M}\sum_{j=1}^{M}\log(P_j) \tag{9}$$

式中，j 为实际标签编号；M 为标签类别数量；P_j 为模型预测的相应标签类别的概率。

4 实验结果与分析

4.1 基础数据

以武进区某区域实际车牌识别数据为例，对算法进行测试。数据集主要包括：①道路网络数据，选取约24 km²区域内的道路网络数据（图1），由3 080条路段、2 390个节点组成，包括120个信号交叉口，路段包含长度、等级等属性；②电警卡口数据，选取研究区内395个电警卡口设备数据，区域1的263个电警卡口数据用于训练模型，区域2的132个电警卡口数据用于测试模型的泛化能力；③车牌识别数据，选取区域内某天7:00—9:00电警卡口设备识别车牌数据，包括加密的车辆唯一编号、识别车辆时间、车辆类型等信息；④出租车、网约车数据，区域内出租车GPS数据用于标定分时段路段平均运行速度和时间，加密的出租车、网约车唯一编号用于剔除车牌识别数据中的出租车、网约车数据。



图1 研究区范围

审图号：苏D(2019)1号

4.2 结果分析

为验证算法性能，本文将最短路径法和自编码器作为对比模型，最短路径法[22]将间断点之间的最短路径作为车辆还原轨迹；自编码器[23]通过神经网络对数据进行非线性降维，将每个候选轨迹的校验指标映射到两个维度。模型训练和测试采用的数据集分别为区域1和区域2内原始车牌识别数据的子集，包括不存在间断点的轨迹数据。对于每个完整OD轨迹，实验中随机删除首末样本点间 n 个连续节点，将OD轨迹分割成多条不完整轨迹，并计算不完整轨迹间K短路集合和真实轨迹对应的路径编号，分别基于全连接神经网络、自编码器构建车辆轨迹决策模型，进行轨迹还原。

实验在搭载 Intel Core i9-10900K CPU 和 128 GB RAM 的台式机上进行，模型的参数量为6 602个，在187 610个样本上训练30轮次，总耗时约为60 s。训练损失由0.398 8迅速降至0.141 6，准确率由90.84%提升至96.02%，模型收敛速度快，学习效率高。模型在1 800条测试样本上的总预测时间低于3 s，平均单条轨迹还原耗时约1.67 ms，吞吐量达到600条/s，具备优异的实时处理能力。

对比实验结果见图2，可以看出，随着中间节点缺失数的增加，最短路径法和自编码器模型的轨迹还原精度明显降低，主要原因在于，在电警卡口点位稀疏区域自编码器缺少车辆完整的历史出行轨迹，无法考虑车辆出行轨迹偏好特征，因此学习的参数未体现出行者主观偏好等对车辆路径的影响，轨迹还原的精确度与最短路径算法相差不大；而本文算法采用基于全连接神经网络的监督学习模型进行出行轨迹还原，学习了出行者的出行风格、固有习惯特征，具有更加优越的性能。此外，随着节点缺失数的增加，本文算法的轨迹还原准确率下降速率较慢，在节点缺失较多时仍具有较强的适应性（缺失中间节点数为8时，轨迹还原精确度接近60%）。本文算法在设备稀疏区域的车辆轨迹还原综合准确率为75.0%，分别比最短路径法和自编码器模型高32.9个百分点和37.0个百分点。

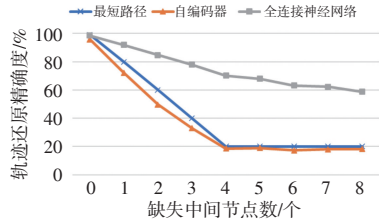


图2 对比实验结果

5 结 语

本文提出了一种基于车牌识别数据的稀疏车辆出行轨迹还原算法，以路径距离、路径时间差、信号交叉口数量、左转次数、右转次数、道路等级6个客观

特征为输入，通过基于全连接神经网络的监督学习模型学习车辆出行路径偏好，从而还原车牌识别数据稀疏区域的车辆出行轨迹。通过武进区某区域实际数据对算法进行校验，验证了算法的准确性和良好的泛化能力，为稀疏车牌出行轨迹还原提供了参考。然而，本文研究也存在一些局限性，如模型主要依赖静态路径属性，缺乏对动态交通状况的考虑，且在不同路网区域的泛化能力有待进一步验证。后续研究可考虑引入实时交通参数，并结合图神经网络等模型增强时空特征表达能力，以提升算法的适用性与精度。

参考文献

[1] Tang J J, Song Y J, Miller H, et al. Estimating the Most Likely Space-time Paths, Dwell Times and Path Uncertainties from Vehicle Trajectory Data: A Time Geographic Method[J]. Transportation Research Part C: Emerging Technologies, 2016, 66:176-194

[2] Quddus M, Washington S. Shortest Path and Vehicle Trajectory Aided Map-matching for Low Frequency GPS Data[J]. Transportation Research Part C: Emerging Technologies, 2015, 55:328-339

[3] Zhang Z, Liu H Q, Rai L, et al. Vehicle Trajectory Prediction Method Based on License Plate Information Obtained from Video-imaging Detectors in Urban Road Environment[J]. Sensors, 2020, 20(5):1258

[4] Sun Z B, Ban X G. Vehicle Trajectory Reconstruction for Signalized Intersections Using Mobile Traffic Sensors[J]. Transportation Research Part C: Emerging Technologies, 2013, 36:268-283

[5] 王龙飞. 基于车牌照的车辆出行轨迹分析方法与实践研究[D]. 西安:长安大学, 2011

[6] Yu H Y, Yang S, Wu Z H, et al. Vehicle Trajectory Reconstruction from Automatic License Plate Reader Data[J]. International Journal of Distributed Sensor Networks, 2018, 14(2):1-13

[7] 阮树斌, 王福建, 马东方, 等. 基于车牌识别数据的机动车出行轨迹提取算法[J]. 浙江大学学报(工学版), 2018, 52(5): 836-844

[8] Rao W M, Wu Y J, Xia J G, et al. Origin-destination Pattern Estimation Based on Trajectory Reconstruction Using Automatic License Plate Recognition Data[J]. Transportation Research Part C: Emerging Technologies, 2018, 95:29-46

[9] Yu F, Sun J, Chen P. Vehicle Trajectory Reconstruction Using Automatic Vehicle Identification and Traffic Count Data[J]. Journal of Advanced Transportation, 2015, 49:174-194

[10] 李广耀, 黄正锋, 楼乐依. 基于贝叶斯网络的稀疏出租车GPS轨迹路径还原方法[J]. 宁波大学学报(理工版), 2021, 34(2):17-24

[11] Li W H, Liu C K, Tang T, et al. An Innovative Supervised Learning Structure for Trajectory Reconstruction of Sparse LPR Data[J]. Transportation, 2024, 51:73-97

[12] Qi X Y, Ji Y J, Li W H, et al. Vehicle Trajectory Reconstruction on Urban Traffic Network Using Automatic License Plate Recognition Data[J]. IEEE Access, 2021, 9:49110-49120

[13] 王杰, 范晓武, 何逸昕, 等. 基于车牌自动识别数据的车辆OD轨迹还原[J]. 计算机系统应用, 2022, 31(7): 247-252

[14] 王冠, 肖昶, 代琦. 基于时空数据的武汉市绿色交通出行模型研究[J]. 地理空间信息, 2023, 21(1):20-25

[15] 李璇, 吴雷, 姜梦, 等. 基于轨迹大数据的司机出行知识图谱构建与实现[J]. 地理空间信息, 2023, 21(1):32-35

[16] 周琪姗, 杨璨, 李健, 等. 基于轨迹方向特征的道路交叉口提取算法[J]. 地理空间信息, 2023, 21(7):7-10

[17] 应申, 曾卓源, 张纪元. 面向数据脱敏的交通场景车牌检测方法[J]. 交通信息与安全, 2024, 42(6):84-94

[18] Hadjiconstantinou E, Christofide N. An Efficient Implementation of an Algorithm for Finding K Shortest Simple Paths[J]. Networks, 2015, 34(2):88-101

[19] 杨寒雨, 赵晓永, 王磊. 数据归一化方法综述[J]. 计算机工程与应用, 2022, 59(3):13-22

[20] 普运伟, 姜萤, 田春瑾, 等. 基于全连接神经网络的在线学习行为分类判别[J]. 现代电子技术, 2023, 46(17):89-94

[21] 徐海文, 史家财, 汪腾. 基于深度全连接神经网络的离港航班延误预测模型[J]. 计算机应用, 2022, 42(10): 3283-3291

[22] 徐鹏, 程钢, 黎旻懿. Dijkstra算法在GIS中的优化与应用[J]. 北京测绘, 2021, 35(6):737-740

[23] 袁非牛, 章琳, 史劲亭, 等. 自编码神经网络理论及应用综述[J]. 计算机学报, 2019, 42(1):203-230