

车辆轨迹地图匹配异常成因辨识方法

暨育雄¹, 陈丹璐^{1,2}, 郑玉靖¹, 沈煜¹

(1. 同济大学交通运输工程学院, 上海 201804; 2. 浙江数智交院科技股份有限公司, 浙江 杭州 310030)

摘要: 车辆轨迹地图匹配异常指轨迹被匹配到不合理的地图路段, 其主要成因可归纳为物理遮挡、地图路段缺失和复杂路网。针对这一现象, 提出一种数据驱动的匹配异常轨迹快速识别和成因归类方法。首先归纳不同成因导致的地图匹配异常数据特征, 并构建表征指标; 其次提出车辆轨迹分段方法, 区分正常和异常的地图匹配轨迹段; 进而建立基于随机森林的地图匹配异常轨迹段成因分类方法; 最后通过上海市地图路网和出租车营运轨迹数据验证方法的有效性。验证表明, 提出的方法准确率达 93.5%, 可有效辨识物理遮挡、社区路段缺失和复杂路网三种成因, 同时具有较好的区域可迁移性。

关键词: 交通工程; 地图匹配; 随机森林; 车辆轨迹; 成因辨识

中图分类号: U491

文献标志码: A

Methodology of Identifying Causes of Abnormal Map Matching

Ji Yuxiong¹, CHEN Danlu^{1,2}, ZHENG Yujing¹, SHEN Yu¹

(1. College of Transportation Engineering, Tongji University, Shanghai 201804, China; 2. Zhejiang Institute of Communications Co. Ltd., Hangzhou 310030, China)

Abstract: Abnormal map matching refers to that partial vehicle trajectory is matched to a road segment incorrectly, typically caused by physical occlusion, road missing, and complex road configuration. This study proposes a data-driven methodology of identifying the causes of abnormal map matching, which supports fast identification and classification of abnormal matching results. First, we summarize the statistical features of abnormal matching results with different causes and present criteria for representation. Second, we propose a segmentation algorithm to distinguish the abnormal results from the trajectory. Third, a classification algorithm is developed

based on the random forest model. Finally, the proposed methodology is validated based on the real-world data of urban taxi operations in Shanghai. The case study reveals that the algorithms are effective in identifying the abnormal matching results caused by physical occlusion, local road missing, and complex road configuration, with a precision of 93.5%. In addition, the experiments show the high spatial suitability of the proposed methodology.

Key words: transportation engineering; map matching; random forest; vehicle trajectory; cause identification

车辆轨迹数据由有序轨迹点组成, 描述车辆运动过程的位置变化。地图匹配将轨迹数据关联到电子地图路网, 识别车辆轨迹点所在的道路和位置, 并将轨迹点序列转换为地图坐标下的路段序列。地图匹配是道路交通运行态势判别^[1-2]、车辆智能调度^[3]、路径选择行为解析^[4]等研究和实践的重要基础。

常用的地图匹配算法可分为四类。基于道路几何信息的算法直接将轨迹点匹配到距离最近的路段上^[5]; 基于道路拓扑信息的算法进一步考虑路网连通性, 从而提升匹配路段序列的合理性^[6]; 针对轨迹数据低采样率和噪声等问题, 基于概率的算法假设定位误差分布已知, 通过轨迹点与周边多条候选路段的匹配概率确定最佳匹配路段^[7]; 综合匹配算法依托数据挖掘、机器学习等先进算法, 考虑上述信息和其他先验知识, 提高复杂路网环境下的匹配准确率, 如模糊逻辑^[8]、卡尔曼滤波^[9]和隐马尔科夫模型^[10]等。其中隐马尔可夫模型 (Hidden Markov Model, HMM) 是实践应用较广且准确度较高的高级地图匹配算法。该算法将轨迹点位置视为观测状态, 车辆实际经过的路段位置视为隐含状态, 根据观测概率和隐含状态转移概率, 采用 Viterbi

收稿日期: 2022-01-26

基金项目: 国家重点研发计划 (2021YFB1600400); 上海市级科技重大专项 (2021SHZDZX0100)

第一作者: 暨育雄 (1978—), 男, 教授, 博士生导师, 工学博士, 主要研究方向为交通全息感知与智能计算。

Email: yxji@tongji.edu.cn

通信作者: 郑玉靖 (1994—), 男, 博士后, 工学博士, 主要研究方向为交通系统数据挖掘。

Email: yujing_zheng@tongji.edu.cn



论文
拓展
介绍

算法^[11],求解概率最大的隐含状态序列,即地图匹配结果。

地图匹配异常指轨迹点被匹配到不合理路段,其主要成因可归纳为:物理遮挡,如高楼、桥隧等设施影响;地图路段缺失;复杂路网,如主辅路、高架路等。针对上述成因,诸多学者提出定制化地图匹配算法,降低匹配异常产生的概率。例如,Zhang等^[12]和Hsueh等^[13]分别通过识别转折点和增加备选路段的手段,提高物理遮挡下轨迹数据匹配准确率。针对路段缺失问题,Torre等^[14]和Haunert和Budig^[15]通过识别轨迹“越野”现象,辅助地图匹配并同时补全缺失路段。部分研究利用道路宽度^[16]、级别^[17]、限速^[18]、驾驶行为^[19]等信息,标定地图匹配算法参数,

从而降低轨迹在复杂路网中的匹配异常概率。然而,目前尚无方法可全面解决上述成因造成的地图匹配异常问题。

本文提出一种数据驱动地图匹配异常成因辨识方法,可实现大规模的匹配异常轨迹段快速识别和成因归类。

1 地图匹配异常数据特征

基于实测数据挖掘分析,总结不同成因导致的匹配异常数据特征。典型的异常匹配结果如图1所示。

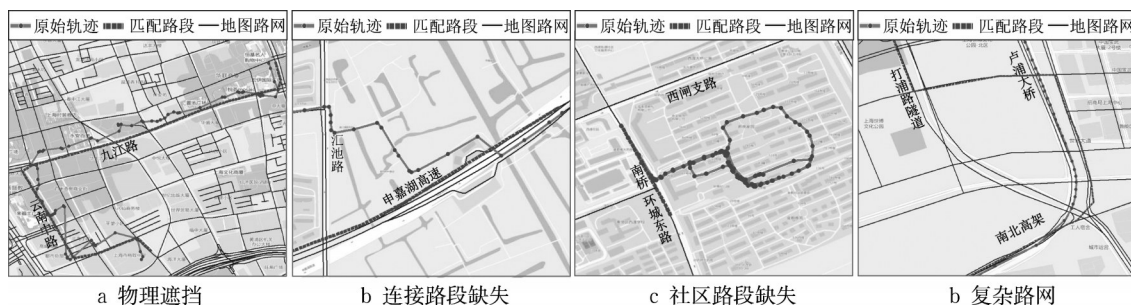


图1 地图匹配异常成因及表现特征示意

Fig.1 Causes of abnormal map matching and their characteristics

物理遮挡(如图1a)指高楼等物理设施导致定位信号漂移,轨迹点与车辆实际经过路段偏差较大,可能将轨迹点匹配到错误路段。即使轨迹点被匹配到正确路段,轨迹点与匹配路段的距离一般较大。此类匹配异常通常表现为,车辆轨迹方向和速度变化频繁,轨迹点与匹配路段和周边地图路段均存在较大距离。

路段缺失指地图更新滞后导致地图上部分路段缺失,引起地图匹配结果异常。根据缺失道路的类型不同,路段缺失可分为连接路段缺失(如图1b)和社区路段缺失(如图1c):

(1) 连接路段缺失:由于新建道路未及时更新,地图中两点间的路段缺失,导致实际经过该路段和前后路段的车辆轨迹点被匹配到其他路段上。此类异常通常表现为,轨迹点与匹配路段距离较大,且周边没有其他邻近地图路段。

(2) 社区路段缺失:由于地图缺少小区、公园等内部道路信息,导致经过这些道路的车辆轨迹点被匹配到外部较高等级道路上。此类异常通常表现为,车辆轨迹常呈环形或L型分布,轨迹点与匹配路段和周边地图路段的距离较大,车辆速度较低。

复杂路网(如图1d)指地图匹配算法无法适应所有复杂路网结构,如高架与地面道路并行、主辅路并行、

多进出口立交等区域,难以准确判定轨迹所在路段。此类异常通常表现为,车辆轨迹较为平顺,部分轨迹段与匹配道路距离较大,但贴合周边其他邻近地图路段。另外,与高楼密集区域、新建道路或社区道路车速相比,高架或立交周边车辆运行速度一般较高。

2 地图匹配异常识别与分类

本文提出一种数据驱动地图匹配异常识别与分类算法,流程框架如图2所示。算法以地图匹配结果为输入,通过轨迹点指标计算、匹配异常轨迹段识别和成因分类3个步骤,最终输出匹配异常成因分类结果。

2.1 轨迹点特征指标计算

地图匹配结果包括地图路网、原始轨迹点和匹配后的轨迹点-路段对应关系。地图路网由各路段信息构成,路段集合记为 S 。一条轨迹 $P=[p_1, \dots, p_i, \dots, p_I]$ 由 I 个有序轨迹点构成,每个轨迹点 p_i 包含4项信息(lon, lat, head, speed),分别为经度、纬度、航向角(行进方向与正北方向的夹角)和车速。轨迹 P 的匹配路段序列记为 $Q=[q_1, \dots, q_i, \dots, q_I](q \in S)$ 。

基于地图匹配结果,根据前文总结的不同成因的

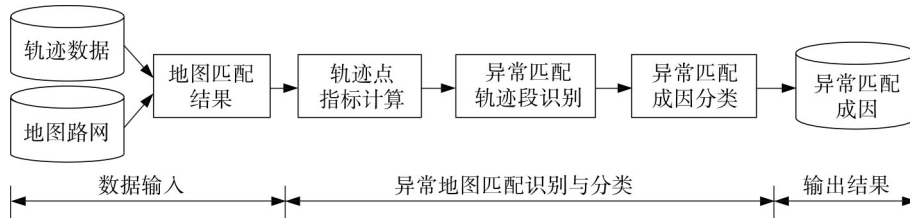


图2 算法流程

Fig.2 Flowchart of the Algorithm

匹配异常特征,针对每条轨迹 P 的每个轨迹点 p_i ,获取匹配距离、路网邻近距离、方向角变化率和速度指标,分别反映轨迹点与匹配路段的接近程度、轨迹点与周边路网的贴合程度、车辆轨迹平顺程度和车辆行驶快慢,作为匹配异常轨迹段识别和成因分类的依据。

指标1:匹配距离,为原始轨迹点与其匹配路段的距离,如式(1)。其中,函数 f^h 表示轨迹点到路段的距离。

$$E_i = f^h(p_i, q_i) \quad i = 1, \dots, I \quad (1)$$

指标2:路网邻近距离,为轨迹点到路网中最近路段的距离,如式(2)。

$$B_i = \min(f^h(p_i, s) | s \in S) \quad i = 1, \dots, I \quad (2)$$

指标3:方向角变化率,为车辆行进方向的偏转角度,如式(3)。其中,mod表示取余函数。

$$H_i = \begin{cases} 0 & i = 1 \\ |p_{i,\text{head}} - p_{i-1,\text{head}}| \bmod \pi & i = 2, \dots, I \end{cases} \quad (3)$$

指标4:速度,直接由轨迹点数据得到,如式(4)。

$$V_i = p_{i,\text{speed}} \quad i = 1, \dots, I \quad (4)$$

2.2 匹配异常轨迹段识别

研究主要关注轨迹点与匹配路段距离较大的异

常。为避免数据噪声对后继算法的影响,以轨迹段为基本单元进行匹配异常识别和成因辨识。提出一种可行的匹配正常和异常轨迹段分段流程,如图3所示:

步骤1:匹配点标记。基于式(1),计算所有轨迹点的匹配距离,结合设备定位误差,通过前序数据分析、预实验和经验判定等方法,定义阈值TD,将匹配距离超过阈值的轨迹点划分为匹配异常点,其他为匹配正常点。

步骤2:匹配正常轨迹段提取。定义参数TN,匹配正常轨迹段由连续出现超过TN个匹配正常点构成。

步骤3:匹配异常轨迹段提取。定义参数TU,在其余轨迹段(由剩余所有匹配异常点和连续个数不超过TN的匹配正常点构成)中,将包含轨迹点个数超过TU的轨迹段识别为匹配异常轨迹段。其余较短轨迹段中的匹配异常可认为由数据随机误差导致,在本文未予考虑。

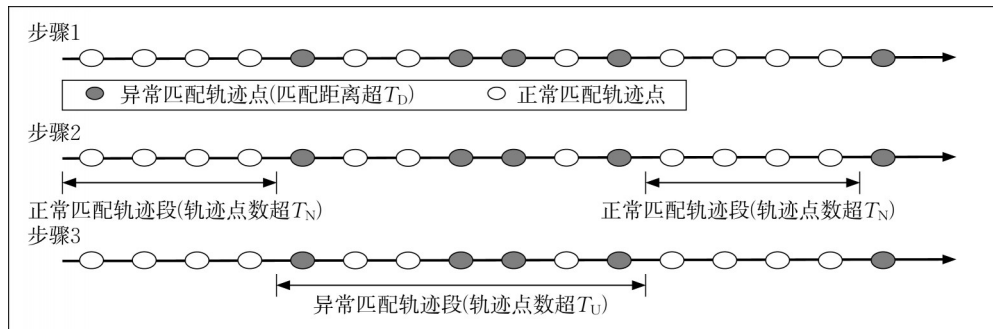


图3 轨迹分段方法示意图

Fig.3 Trajectory segmentation method

2.3 匹配异常轨迹段分类

基于2.1定义的特征指标,生成匹配异常轨迹段的特征向量,作为成因分类的依据。一条匹配异常轨迹段包含若干轨迹点,根据式(1)~式(4)计算其中所有轨迹点的匹配距离、路网邻近距离、方向角

变化率和速度4项指标,统计各自均值、中位数和方差,作为该匹配异常轨迹段的12个特征值。

本文采用随机森林算法进行匹配异常成因归类。随机森林是一种集成学习算法,其基本思想是将多个决策树集成,每棵决策树都是一个分类器。对一个输

入样本,每棵决策树均会产生一个分类投票,随机森林集成所有的分类投票结果,将该样本归为投票次数最多的类别。随机森林算法在分类任务中具备如下优势^[20]:训练速度较快,且具备抗噪能力,能够高效应用于大规模数据集;能够处理高维特征的输入样本,无需预先降维;具有较高的准确率。

基于随机森林的匹配异常轨迹段分类流程如下:

(1)训练集标定。记匹配异常轨迹段的训练集为 X ,每个样本 x_k 含12项特征值。通过人工标定得到成因类别 $y_k \in Y$,其中 Y 表示异常成因集合,包括物理遮挡、连接路段缺失、社区路段缺失、复杂路网4类。

(2)样本抽样。随机且有放回地从训练集 X 中抽取 n 个训练样本,形成决策树训练集。在12个特征指标中随机且无放回地选取 m 个特征指标,构建决策树特征集。

(3)决策树生成。从根节点开始不断分裂形成新的节点。如节点包含的样本属于同一成因,则节点不可分,称为叶节点。如节点可分,选取最优特征及其阈值,将该节点样本集 X_v 二分为 X_+ 和 X_- ,形两个子节点,特征 a 及其阈值的选取基于信息增益最大原则。当所有节点均不可分,决策树构建完成。信息增益 $G(X_v, a)$ 定义如式(5)所示,其中, X_v 表示当前节点样本集, a 表示特征及其取值, $H(\cdot)$ 表示样本集的信息熵。记样本集 X_v 中成因 y 的样本比例为 $r_{v,y}$,则 $H(\cdot)$ 如式(6)所示。

$$G(X_v, a) = H(X_v) - \left(\frac{|X_+|}{|X_v|} H(X_+) + \frac{|X_-|}{|X_v|} H(X_-) \right) \quad (5)$$

$$H(X_v) = - \sum_{y \in Y} r_{v,y} \log_2 r_{v,y} \quad (6)$$

(4)随机森林构建。重复(2)和(3),随机生成 N 棵形态不同的独立决策树,完成分类器训练。

(5)分类器应用。输入待分类匹配异常轨迹段的特征向量 X^* 后,每棵决策树根据自身节点判定条件对其分类并投票,记 N 棵决策树分类结果为 $\{y_1^*, y_2^*, \dots, y_N^*\}$ 。分类器最终输出投票次数最多的类别。

3 案例分析

3.1 标定数据集

本文采用的原始轨迹数据来源于上海市某出租车公司。数据集由16 022辆出租车2016年单日产

生的479 353条营运轨迹组成,每条轨迹表示出租车空车或重车的行驶过程。车载设备每10s实时上传车辆信息,包括车辆编号、时间、载客状态、经度、纬度、航向角、速度等。地图数据来自开放街道地图(OpenStreetMap)^[21],以上海市全域为研究范围。案例采用基于HMM的地图匹配算法^[10],得到原始轨迹对应的路段序列,即轨迹数据的地图匹配结果。

案例采用2.2的匹配异常轨迹段识别算法,通过预实验(采用不同阈值参数获取多组识别结果,与人工判定结果对比)选取合适参数,定义 $T_N = T_0 = 25$, $T_0 = 20m$,识别出12 373条匹配异常轨迹段。结合轨迹形态、路网结构、车速等信息,人工标定匹配异常的成因,共标定出物理遮挡类1 719条、连接路段缺失489条,社区路段缺失3 179条和复杂路网类6 986条。

不同成因的匹配异常轨迹段的特征值具有明显差异性。图4展示了各类匹配异常轨迹段12项特征值的中位数,图中特征值均已做最大值归一化处理。可以看出,物理遮挡类的匹配距离和路网邻近距离均较小,速度和方向变化频繁;连接路段缺失类的匹配距离较大,且路网邻近距离变化频繁;社区路段缺失类的原始轨迹整体速度较低,方向角变化较大,且远离周边路网,符合车辆在社区中行驶的特征;复杂路网类最显著的特征是轨迹与周边路段距离较近,但与匹配路段距离较远,且行驶速度比其他3类较快。

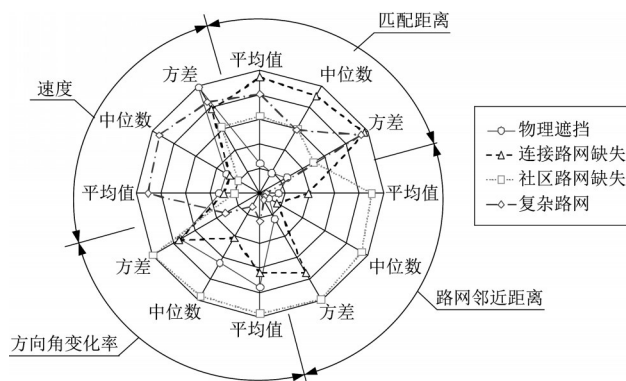


图4 不同类型匹配异常轨迹段指标统计结果

Fig.4 Median of indicators for different types of abnormal matching results

不同成因的匹配异常轨迹段的分布如图5所示。物理遮挡类主要分布于市区中心(图5a);路段缺失类分布较为均衡,其中社区路段缺失(图5c)显著多于连接路段缺失(图5b);复杂路网类主要分布于高架路、桥梁和隧道等基础设施附近(图5d)。

以上海虹桥枢纽区域为例,该区域路网复杂,高

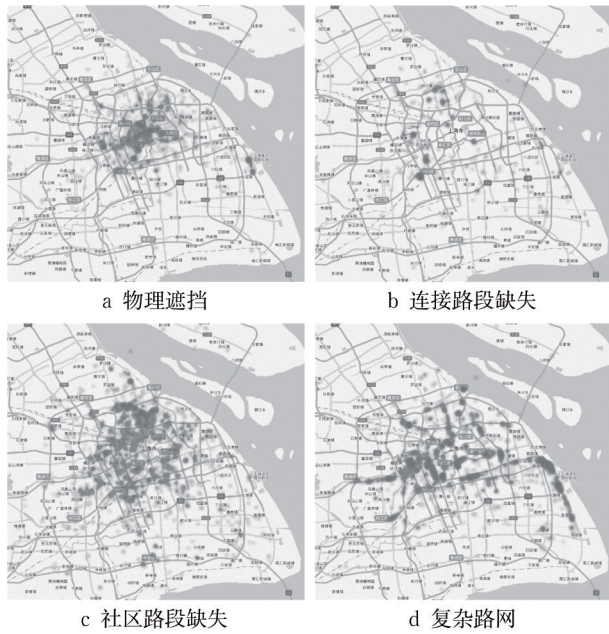


图5 不同类型匹配异常轨迹分布

Fig.5 Spatial distribution of different types of abnormal matching results

架和地面道路交叠。在该区域,共识别出889条匹配异常轨迹段,通过人工标定成因,复杂路网类占93.5%(831条),物理遮挡、连接路段缺失和社区路段缺失导致的异常分别占5.3%(47条),0.8%(7条)和0.4%(4条)。图6展示了该区域的一条轨迹,包含匹配正常和异常的轨迹段。轨迹起点应为虹桥火车站北侧出发层,却被异常匹配到邻近的地面辅道上;后续的匹配结果为正常匹配。

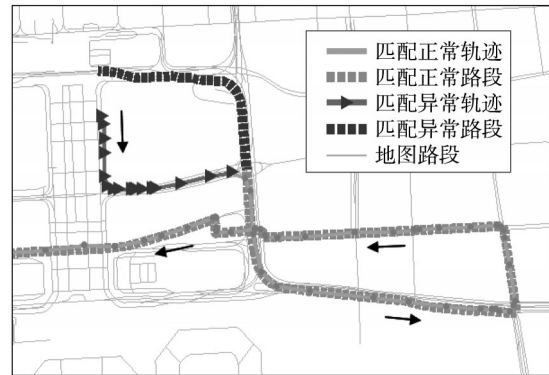


图6 上海虹桥枢纽区域匹配异常/正常轨迹段实例

Fig.6 Example of abnormal/normal matching results in the vicinity of Hongqiao Transportation Hub, Shanghai

3.2 算法验证

随机抽取全市匹配异常轨迹段的80%(9 898条)为训练集,其余轨迹段(2 475条)为测试集,验证2.3所述成因分类算法的有效性。本文设定随机森林分类器生成200棵决策树,每棵树随机抽取的样本个数与测试集规模相同,通过随机4个特征值进行训练,即, $N=200, n=9\,898, m=4$ 。表1展示了随机森林算法分类结果,并与CART决策树算法^[22](括号中数据)进行对比。结果表明,随机森林算法输出的各类精确率和召回率均高于决策树算法,验证了其有效性。随机森林算法可有效区分匹配异常轨迹段的成因,尤其是物理遮挡、社区路段缺失和复杂路网类,总体准确率达93.5%,单项精确率和召回率均超过87%。

表1 随机森林(括号中对比CART决策树)算法分类结果混淆矩阵

Tab. 1 Confusion matrix of classification results of the random forest (decision tree) algorithm

真实类别	算法判定类别				召回率/%
	物理遮挡	连接路段缺失	社区路段缺失	复杂路网	
物理遮挡	307 (275)	6 (5)	18 (22)	17 (46)	88.2 (79.0)
连接路段	13 (25)	46 (29)	17 (21)	30 (31)	43.4 (27.4)
社区路段	23 (19)	4 (20)	575 (551)	11 (23)	93.8 (89.9)
复杂路网	9 (27)	6 (13)	7 (16)	1386 (1352)	98.4 (96.0)
精确率/%	87.2 (79.5)	74.2 (43.3)	93.2 (90.3)	96.0 (93.1)	总体准确率: 93.5 (89.2)

然而,算法在判定连接路段缺失导致的匹配异常时表现不佳,精确率为74.2%,召回率为43.4%。一方面是由于该类异常数据量较小(只占总样本的4%左右),分类器难以得到充分训练从而识别这类异常特征;另一方面,这类异常在案例中的表现形式多样。如图7所示,轨迹段A、B、C均由于连接路段缺失导致匹配异常,其中A的异常成因被正确分类,而B和C分别被识别为复杂路网和物理遮挡类。轨迹段A所在的长段道路缺失,分类特征值符合图4规律,因此被正确归类;轨迹段B所在路段存在部分缺失,但却导致整段轨迹匹配到其他道路上,对匹配结果产生较大影响,导致路网邻近距离较小而匹配距离较大,同时车速较快,与图4中的复杂路网类异常特征相近;轨迹段C所在的缺失道路与匹配道路距离较小,同时轨迹方向角变化幅度较大,与图4中物理遮挡类异常特征相近。

然而,算法在判定连接路段缺失导致的匹配异常时表现不佳,精确率为74.2%,召回率为43.4%。一方面是由于该类异常数据量较小(只占总样本的4%左右),分类器难以得到充分训练从而识别这类异常特征;另一方面,这类异常在案例中的表现形式多样。如图7所示,轨迹段A、B、C均由于连接路段缺失导致匹配异常,其中A的异常成因被正确分类,而B和C分别被识别为复杂路网和物理遮挡类。轨迹段A所在的长段道路缺失,分类特征值符合图4规律,因此被正确归类;轨迹段B所在路段存在部分缺失,但却导致整段轨迹匹配到其他道路上,对匹配结果产生较大影响,导致路网邻近距离较小而匹配距离较大,同时车速较快,与图4中的复杂路网类异常特征相近;轨迹段C所在的缺失道路与匹配道路距离较小,同时轨迹方向角变化幅度较大,与图4中物理遮挡类异常特征相近。

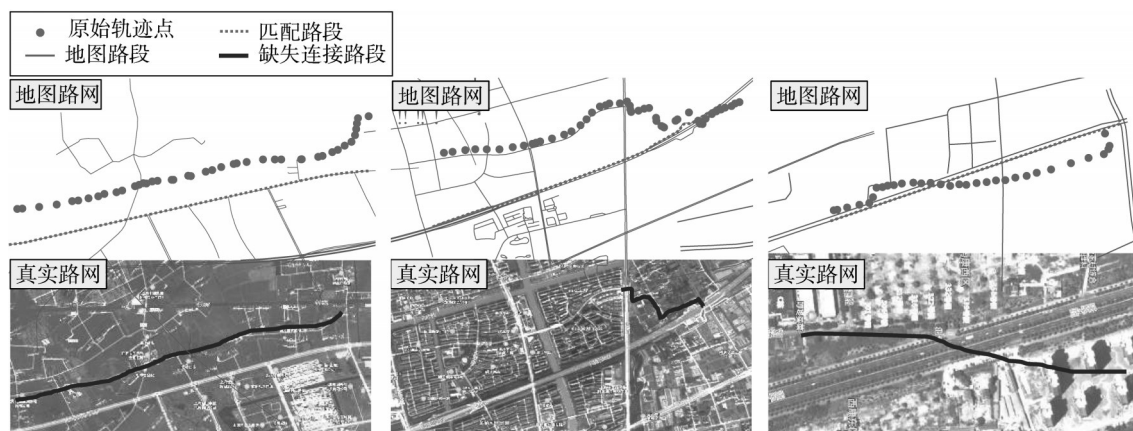


图7 连接路段缺失异常示例

Fig.7 Examples of abnormal matching results caused by missing connecting roads

图8展示了12项匹配异常轨迹段分类特征值的影响权重。可以看出,路网邻近距离的中位数和均值是分类任务中最为关键的指标,其影响权重分别达26.8%和19.9%;匹配距离中位数、方向角变化率方差和速度方差对分类的影响较低,权重均低于3%。

3.3 算法区域可迁移性

为验证算法的区域可迁移性,分别定义训练集和测试集的区域,方案如表2所示。方案A和B以上海中环快速路为边界,将城市划分为中心区和外围区,分别基于中心和外围区数据训练算法,应用于外围和中心区的成因分类;方案C和D以黄浦江为界,将城市划分为浦西和浦东,分别基于浦西和浦东匹配异常数据训练算法,应用于浦东和浦西的成因分类。

表2表明算法具有较好的区域可迁移性。在考虑区域数据差异性的情况下,各验证方案的分类准

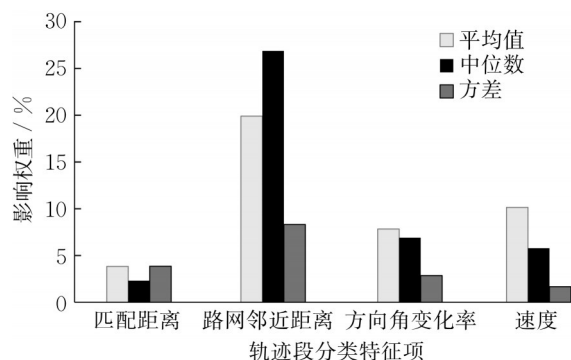


图8 随机森林分类器中每个特征项的权重

Fig.8 The weight for each feature element in the random forest classifier

确率均较高,超过90%。除连接路网导致的异常外,其余各类成因辨识单项召回率均高于88%,单项精确率均高于75%。

表2 算法区域可迁移性验证实验设计和结果

Tab. 2 Settings and results of the experiments for validating the spatial suitability of the algorithm

验证方案	训练集		测试集		总体准确率/%	单项召回率最小值 (除连接路网类)/%	单项精确率最小值 (除连接路网类)/%
	选取区域	样本数量	选取区域	样本数量			
A	中环内	3 190	中环外	9 183	92.2	90.7	90.5
B	中环外	9 183	中环内	3 190	92.3	88.4	75.1
C	浦西	6 995	浦东	5 378	95.6	89.8	85.5
D	浦东	5 378	浦西	6 995	90.8	89.2	86.9

4 结语

本文基于大量数据的挖掘分析,归纳了物理遮挡、路段缺失(连接路段或社区路段缺失)和复杂路网导致的地图匹配异常特征,据此提出了相应的表征指标,并建立了基于随机森林的匹配异常轨迹段识别和成因辨识方法。案例分析验证了方法的有效性。实例验证表明,方法可有效辨识物理遮挡、社区路段缺失和复杂路网成因,准确率达93.5%,单项精确率和召回率均超过

87%。此外,方法具有较好的区域可迁移性,基于某区域数据训练算法,应用于其他区域的匹配异常成因辨识时,准确率超85%,单项精确率和单项召回率分别超88%和75%。分析表明,在选取的特征值中,路网邻近距离均值和中位数对随机森林分类器影响较大,影响权重分别达到26.8%和19.9%。本文提出的方法实现了地图匹配异常结果的批量快速筛选和成因辨识,可作为地图匹配实践应用的前序步骤,支撑后续的地图修复、定制化算法研发和应用、路侧辅助定位设备布

设等研究和实践。

后续研究方向包括:选取适当特征值,提高连接路段缺失类异常的识别有效性;构建标准化评价体系,基于地图匹配异常结果和成因,分析不同地图匹配算法的有效性;探索不同成因导致的地图匹配异常的修正方法。

作者贡献声明:

暨育雄:概念提出、研究计划制定、论文撰写;

陈丹璐:试验设计与开展、结果分析与可视化;

郑玉靖:研究计划制定、试验设计与开展、论文撰写;

沈煜:结果分析与可视化。

参考文献:

- [1] FABRITHIS C D, RAGONA R, VALENTI G. Traffic estimation and prediction based on real time floating car data [C]// Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems. [S.l.]: IEEE, 2008: 197-203.
- [2] 廖律超,蒋新华,林铭榛,等.基于交通轨迹数据挖掘的道路限速信息识别方法[J].交通运输工程学报,2015,15(5):118.
LIAO Luchao, JIANG Xinhua, LIN Mingzhen, *et al.* Recognition method of road speed limit information based on data mining of traffic trajectory [J]. Journal of Traffic and Transportation Engineering, 2015, 15(5): 118.
- [3] 翁剑成,刘文韬,陈智宏,等.基于浮动车数据的出租车运营管理研究[J].北京工业大学学报,2010,36(6):779.
WENG Jiancheng, LIU Wentao, CHEN Zhihong, *et al.* Research on floating car data based taxi operation and management [J]. Journal of Bjing University of Technology, 2010, 36(6): 779.
- [4] 唐炉亮,常晓猛,李清泉,等.基于蚁群优化算法与出租车GPS数据的公众出行路径优化[J].中国公路学报,2011,24(2):89.
TANG Luliang, CHANG Xiaomeng, LI Qingquan, *et al.* Public travel route optimization based on ant colony optimization algorithm and taxi GPS data [J]. China Journal of Highway and Transport, 2011, 24(2): 89.
- [5] QUDDUS M A, OCHIENG W Y, NOLAND R B. Integrity of map-matching algorithms [J]. Transportation Research Part C: Emerging Technologies, 2006, 14(4): 283.
- [6] QUDDUS M A, OCHIENG W Y, ZHAO L, *et al.* A general map matching algorithm for transport telematics applications [J]. GPS Solutions, 2003, 7(3): 157.
- [7] BIERLAIRE M, CHEN J, NEWMAN J. A probabilistic map matching method for smartphone GPS data [J]. Transportation Research Part C: Emerging Technologies, 2013, 26: 78.
- [8] 宋洁,李国燕,李娜娜,等.基于模糊逻辑的GPS/DR地图匹配算法[J].计算机工程与科学,2008,30(10):30.
SONG Jie, LI Guoyan, LI Nana, *et al.* A fuzzy-logic-based map matching algorithm for the GPS/DR system [J]. Computer Engineering & Science, 2008, 30(10): 30.
- [9] CHU H J, TSAI G J, CHIANG K W, *et al.* GPS/MEMS INS data fusion and map matching in urban areas [J]. Sensors, 2013, 13(9): 11280.
- [10] NEWSON P, KRUMM J. Hidden Markov map matching through noise and sparseness [C]// Proceedings of the 17th ACM Sigspatial international conference on advances in geographic information systems. Seattle: ACM, 2009:336-343.
- [11] FORNEY G D. The viterbi algorithm [J]. Proceedings of the IEEE, 1973, 61(3): 268.
- [12] ZHANG D, DONG Y, GUO Z. A turning point-based offline map matching algorithm for urban road networks [J]. Information Sciences, 2021, 565(2): 32.
- [13] JIMENEZ F, MONZON S, NARANJO J E. Definition of an enhanced map-matching algorithm for urban environments with poor GNSS signal quality [J]. Sensors, 2016, 16(2): 193.
- [14] TORRE F, PITCHFORD D, BROWN P, *et al.* Matching GPS traces to (possibly) incomplete map data: bridging map building and map matching [C]// Proceedings of the 20th International Conference on Advances in Geographic Information Systems. Redondo Beach: ACM, 2012:546-549.
- [15] HAUNERT J-H, BUDIG B. An algorithm for map matching given incomplete road data [C]//Proceedings of the 20th International Conference on Advances in Geographic Information Systems. Redondo Beach: ACM, 2012:510-513.
- [16] GOH C Y, DAUWELS J, MITROVIC N, *et al.* Online map-matching based on hidden Markov model for real-time traffic sensing applications [C]//Proceedings of the 15th International IEEE Conference on Intelligent Transportation Systems. [S.l.]: IEEE, 2012: 776-781.
- [17] WU Z, XIE J, WANG Y, *et al.* Map matching based on multi-layer road index [J]. Transportation Research Part C: Emerging Technologies, 2020, 118: 102651.
- [18] HSUEH Y L, CHEN H C, HUANG W J. A hidden Markov model-based map-matching approach for low-sampling-rate GPS trajectories [C]//Proceedings of the 7th International Symposium on Cloud and Service Computing (SC2). [S.l.]: IEEE, 2017: 271-274.
- [19] 张涛,杨殿阁,杨扬,等.地图匹配中快速路出入口车辆行为模式识别[J].中国公路学报,2010,23(6):103.
ZHANG Tao, YANG Diange, YANG Yang, *et al.* Vehicle behavior pattern recognition at merging and diverging sections for expressways in map-matching [J]. China Journal of Highway and Transport, 2010, 23(6): 103.
- [20] BIAU G, SCORNET E. A random forest guided tour [J]. Test, 2016, 25(2): 197.
- [21] HAKLAY M, WEBER P. Openstreetmap: User-generated street maps [J]. IEEE Pervasive Computing, 2008, 7(4): 12.
- [22] RUTKOWSKI L, JAWORSKI M, PIETRUCZUK L, *et al.* The CART decision tree for mining data streams [J]. Information Sciences, 2014, 266: 1.