

数据驱动的大宗物流地图构建: 架构及进展*

毛嘉莉¹, 吴涛¹, 李思佳¹, 郭焯², 周傲英¹, 金澈清¹, 钱卫宁¹

¹(华东师范大学 数据科学与工程学院, 上海 200062)

²(京创智汇(上海)物流科技有限公司, 上海 200433)

通信作者: 毛嘉莉, E-mail: jlmao@dase.ecnu.edu.cn



摘要: 普通的城市道路地图未能覆盖(超)重卡货车的道路禁限行信息, 缺少标注适用于大宗货运的热门停放区域, 无法满足货运司机的大批量长距离公路运输需求. 为解决大宗货运交通事故频发、物流效率低下等问题, 进一步提升货运司机的出行体验感, 亟需结合运输货物类型、货车车型以及司机的线路选择偏好等因素, 研究适用于公路大宗货运的定制化物流地图构建方法. 随着移动互联网、车联网的普及, 大宗货运产生的时空数据迅猛增长, 与物流运营数据等一起构成物流大数据, 为构建物流地图提供了数据基础. 在梳理地图构建技术的基础上, 针对现有电子地图构建方法在大宗货运领域的局限性, 利用多源物流数据提出了一个数据驱动的物流地图构建框架, 主要研究内容包括: (1) 基于用户先验知识的多约束物流地图构建; (2) 动态时空数据驱动的物流地图增量更新. 物流地图将成为大宗货运发展新一代物流科技的 AI 基础设施. 研究成果为物流地图构建的技术创新提供了丰富的实践内容, 也为促进大宗物流降本增效提供了新的解决思路, 具有重要的理论意义和应用价值.

关键词: 时空数据; 物流地图; 大宗货运; 道路拓扑

中图法分类号: TP311

中文引用格式: 毛嘉莉, 吴涛, 李思佳, 郭焯, 周傲英, 金澈清, 钱卫宁. 数据驱动的大宗物流地图构建: 架构及进展. 软件学报, 2023, 34(1): 421-443. <http://www.jos.org.cn/1000-9825/6434.htm>

英文引用格式: Mao JL, Wu T, Li SJ, Guo Y, Zhou AY, Jin CQ, Qian WN. Data-driven Logistics Map Building for Bulk Commodity Transporting: Architecture and Research Progress. Ruan Jian Xue Bao/Journal of Software, 2023, 34(1): 421-443 (in Chinese). <http://www.jos.org.cn/1000-9825/6434.htm>

Data-driven Logistics Map Building for Bulk Commodity Transporting: Architecture and Research Progress

MAO Jia-Li¹, WU Tao¹, LI Si-Jia¹, GUO Ye², ZHOU Ao-Ying¹, JIN Che-Qing¹, QIAN Wei-Ning¹

¹(School of Data Science and Engineering, East China Normal University, Shanghai 200062, China)

²(Jing Chuang Zhi Hui (Shanghai) Logistics Technology, Shanghai 200433, China)

Abstract: Since ordinary city road map has not covered the road restrictions information for the lorry, and lacks of hot spots labeling, they cannot satisfy massive batches and long-distance road transportation requirements of bulk commodity transporting. In order to address the issues of frequent transportation accidents and low logistics efficiency, and further improve the truck drivers' travel experience, it is urgent to combine the type of goods transported with the type of truck as well as the driver's route selection preference to study the building method of customized logistics map for bulk commodity transporting. With the widespread applications of mobile Internet and Internet of vehicles, spatio-temporal data generated by bulk commodity transporting is growing rapidly. It constitutes logistics big data with other logistics operational data, which provides a solid data foundation for logistics map building. This study first comprehensively reviews the state-of-the-art work about the issue of map building using trajectory data. Then, to tackle the limitations of existing digital map building methods in the field of bulk commodity transporting, a data-driven logistics map building framework is put forward using multi-

* 基金项目: 国家自然科学基金(62072180, U1911203)

收稿时间: 2021-05-06; 修改时间: 2021-07-13, 2021-08-04; 采用时间: 2021-08-22; jos 在线出版时间: 2021-10-20

CNKI 网络首发时间: 2022-11-15

source logistics data. The following researches are focused on: (1) multi-constraint logistics map construction based on users' prior knowledge; (2) dynamic spatio-temporal data driven logistics map incremental updating. Logistics map will become AI infrastructure for new generation of logistics technology fit for bulk commodity transportation. The research results of this study provide rich practical contents for the technical innovation of logistics map building, and offer new solutions to promote the cost reduction and efficiency improvement of logistics, which have important theoretical significance and application values.

Key words: spatio-temporal data; logistics map; bulk commodity transporting; road topology

大宗货物是指可进入流通领域, 但非零售环节, 具有商品属性并用于工农业生产与消费使用的大批量买卖物质商品, 典型的有原油、煤炭、钢铁、矿石、水泥、粮食等. 近年来, 我国社会物流总费用占国内生产总值 (GDP) 的比率已逐年下降, 但大宗物流成本仍远高于欧美等发达国家. 以钢铁物流为例, 2019 年我国钢铁行业物流成本费用率约为 12%, 高于发达国家的 6%; 其中, 运输成本占比过半, 成为钢铁等大宗物流降本增效的关键. 作为国民经济的重要物质基础, 大宗物流一直备受国家和业界的关注. 2018 年 12 月, 国家发展改革委和交通运输部印发《国家物流枢纽布局和建设规划》, 将在未来 15 年内基本形成与现代化经济体系相适应的国家物流枢纽网络, 使物流运行效率和效益达到国际先进水平. 2019 年 9 月国务院印发的《交通强国建设纲要》指出, 要打造绿色高效的现代物流系统, 强化前沿关键科技研发, 推进大件运输等专业化物流的发展; 《纲要》还强调要将交通运输系统的主要发展内容部署到新一代基础设施建设工具、智能化载运工具、人性化的服务系统等建设要求中. 为促进大宗物流降本增效, 亟需重视人工智能 (AI) 基础设施建设, 确保与一流的物流枢纽和物流网络相配套.

区别于快递、快运等普通物流, 大宗货运具有大批量、远距离、重载等特点. 公路运输具有门到门的运输优势以及能实现铁路覆盖不到的区域运输, 成为大宗物流的主要运输方式. 公路大宗货运普遍使用重型卡车或超重型卡车, 由于货车车身具有高、重、宽、长的特点, 在城市道路网络中行驶存在一定的安全隐患. 我国大部分城市为了减少交通事故与环境污染, 在一些时段内对部分路段禁止货车通行, 对于部分桥梁等基础设施也设置了限高、限重等道路限制条件, 这给大宗货运带来严峻的挑战. 近年来, 货运车辆由于规划线路时未充分考虑道路的禁限行规定, 频繁发生交通事故, 给国家的交通安全与道路基础设施造成严重损失. 2019 年 10 月 10 日, 两辆运输钢卷的货车在行驶通过无锡 312 国道 K135 处高架桥时, 由于超重桥荷载 3 倍引致桥面发生侧翻事故, 直接经济损失 823.1 万元. 此外, 大宗公路货运还面临运输任务规划未充分考虑出发时的路况信息、导航地图不准确、货车司机不熟悉运输路线或故意绕路等一系列问题, 导致运输效率低下、时间与人力成本浪费等. 为解决上述问题, 亟需运用大数据、云计算、AI 等技术为传统大宗物流领域智慧赋能, 通过构建物流地图来支持公路运输任务实施中的规划、调度、监督等环节.

物流地图是保障公路大宗货运有效实施运输任务的基础支撑, 不仅应用于运输前运力的预测调度与车货匹配决策、运输过程中的任务追踪监控、车辆行驶位置 (或运输终点) 校准、最优线路推荐、运输异常检测, 以及运输后运费结算等物流场景, 还能对物流网络规划中的仓储、分拨节点、充电桩 (新能源货车适用) 等的选址发挥重大作用. 早期的商业地图构建涉及实地勘测与校准处理, 需要耗费大量的人力物力. 随着志愿地理信息 (volunteered geographic information, VGI) 技术的出现, 即使是商业利益较低区域的地图信息也能逐步完善, 相比商业地图地理覆盖范围更广. 因此, 该类地图已成为商业地图的补充数据集, 如 WikiMapia, OpenStreetMap 和谷歌地图制作者等. 然而, VGI 技术需要用户使用专门的软件工具创建、收集其自愿提供的地理数据, 对用户的专业能力要求较高.

随着以北斗卫星导航系统为代表的智能终端、车联网的广泛应用, 衍生了大规模流式增长的时空数据, 它们几乎覆盖了道路网络, 这些数据使得低成本的地图推断与更新成为可能. 在近 20 年里出现了一大批基于轨迹数据的电子地图构建算法, 它们利用车辆行驶产生的轨迹数据, 通过使用地图推断或地图更新算法提取道路拓扑结构 (包括道路的几何形状与连接属性). 与普通电子地图相比, 为满足大宗货物的公路运输需求, 物流地图需要标注货车的道路限行等信息. 2017 年, 高德地图结合交管部门提供的货车通行道路限制信息推出了货车地图导航产品; 2019 年, 百度智能云发布了百度物流地图正式版. 上述地图产品支持避开限行违章路段的货运线路规划与导航服务. 然而, 大宗货物种类繁多, 从供给端到需求端的长距离运输途中, 不仅承运车型不同 (比如, 原油采用罐车运输、

钢铁产品采用轴线车运输), 包括仓库、中转枢纽的布局以及运输线路亦有较大区别。为保障大宗货物远距离运输任务的顺利实施, 物流地图除关注货车车型相关的道路禁限行信息以外, 还需根据承运货物类型、货车车型(包含货高、载重等属性)等标注加油站、充电桩、仓库/分拨站点、维修点、餐馆、酒店等不同热门停车区域。与此同时, 还需提取适用于不同类型货物运输的可通行热门运输线路。因此, 为全面提升运输效率、降低运输成本, 亟需构建公路大宗货运的定制化物流地图, 确保运输任务的合理规划、调度与实施。

随着信息化应用程度的提高, 我国大宗物流已逐渐呈现出平台化的发展趋势, 涌现出一批互联网物流平台。它们以共享物流所有环节的信息为核心, 利用互联网模式提升车货匹配效率。大数据、云计算、人工智能等新一代信息技术的蓬勃发展, 为打造适于大宗货运的物流科技、构建以物流地图为中心的“数智化”基础设施提供了崭新的思路和方法。货物运输产生的车辆轨迹连同运输涉及的港口、道路等组成的基础信息, 以及物流平台产生的订单与运力调配信息, 形成了覆盖运输全程的物流大数据。以云计算为基础, 利用 AI 的推理能力与大数据分析能力构建物流地图能有效提升物流效率。经调研, 实际应用中公路大宗货运的运输线路规划普遍由货车司机的先验知识(即依赖于有经验的司机提供运输线路)获得。不但避开了货车的禁/限行路段, 还在不同程度上考虑了司机的运输停留热点偏好、行驶路段偏好(如避开高速)及线路选择偏好(更便捷、更快、更低油耗、轮胎磨损最少)等。由此, 通过挖掘多源物流数据可以识别适用于大宗货运的热门停车区域与各个时段内的经验路径, 在此基础上提取道路拓扑结构继而构建物流地图。所构建的物流地图可以应用于大宗货物运输的仓储(或分拨、充电桩)选址、车货匹配、运力调度、线路规划、运输终点校准、运输到达时间预测、运输监控等场景, 其研究具有广泛的应用前景。

本文第 1 节对基于轨迹的电子地图构建定义以及所采用的关键技术进行了梳理。第 2 节对不同类型的地图构建技术的基本原理进行了综述。第 3 节在梳理物流地图构建面临挑战的基础上, 提出了大宗物流地图的构建框架。第 4 节与第 5 节分别阐释了基于轨迹数据的物流地图构建与动态更新机制。第 6 节基于真实的钢铁物流数据集介绍了大宗物流地图构建的实践案例, 最后对全文进行总结。

1 基于轨迹的地图构建: 概念及技术

1.1 概念

电子地图已成为人们日常生活的重要组成部分, 在汽车导航和路线规划等智能交通应用中起着关键作用。随着传感器网络技术、移动通信技术与定位技术的高速发展, 衍生了海量的时空轨迹数据。该类数据表现为一系列包含移动对象产生的地理坐标、速度、方向、时间戳等属性的时序位置信息。由于轨迹数据可以实时记录不同级别道路范围内移动对象的活动情况, 利用轨迹数据可以洞察路网结构的时变演化规律。因此, 基于轨迹数据的地图构建方法业已成为目前采用的主流方法。

电子地图可视为一种高度地理化的图形, 包含交叉路口的实际位置与范围、路段的几何形状与长度、路段之间的连接属性等。地图的基本模型是一个几何图形, 以顶点表示交叉路口、以边表示街道(路段), 将图形嵌入在平面中。由此, 电子地图常用无向几何图表示, 其中, 每个顶点建模为嵌入平面中的一个点, 每条边则建模为嵌入平面中连接两个顶点的多边形曲线。

基于轨迹数据的地图构建问题定义如下: 给定一组输入的轨迹数据集, 其中每条轨迹由一系列位置信息的测量值组成。每个测量值包含一个位置点(以经/纬度坐标或投影后的 (x, y) 坐标表示)、时间戳以及该位置点的行驶相关信息, 如车辆的航向或即时速度等。采用选定的道路推断方法对轨迹数据进行分析并输出某种嵌入式图。根据地图构建过程是否依赖基准路网, 现有的构建方法可以划分为地图推断与地图更新两类。其中, 地图推断旨在无路网前提下, 将定位设备采集到的轨迹数据直接转换成路网。地图更新技术则是利用持续产生的轨迹数据与基准路网进行差分运算, 它们关注相对于基准路网的缺失路段(路口)发现、道路封闭/开通检测以及道路转向规则变化的辨识等。采用各类地图构建算法推断生成的路网可能包含不一样的道路特征, 具有不同的推断精度。

1.2 关键技术

由于移动对象的位置信息通常以不同采样率采集获得且伴有大量噪声, 给基于轨迹数据的地图构建工作带来

极大的挑战. 图 1 展示了一个利用车辆轨迹数据推断地图的样例, 其中, 图 1(a) 为北京 2013 年 10 月 20 日一个测试区域内采集到的车辆轨迹分布 (以绿色线条表示); 图 1(b) 显示在 OpenStreetMap 地图 (以灰色线条表示) 上叠加了使用地图推断算法生成的地图 (以浅蓝色线条表示). 由图 1 可以看出, 原始采集到的轨迹数据具有大量噪声且数据分布不均匀. 一个精准的地图推断算法应确保所生成的地图不受噪声影响、输出的道路拓扑结构能正确反映真实的道路形态. 因此, 面向大规模快速更新的轨迹数据, 如何通过数据预处理、道路拓扑结构获取等系列技术生成高质量的地图已成为基于轨迹的地图构建方法的一大挑战. 为保证推断地图的高精度, 在利用轨迹数据进行地图构建时需要采用轨迹去噪、轨迹简化、轨迹聚类、道路中心线拟合以及地图匹配等关键技术.

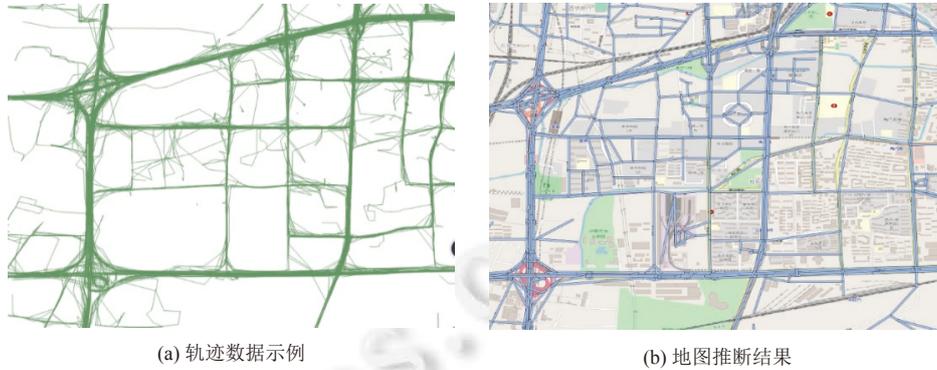


图 1 地图推断示例

1.2.1 轨迹数据的质量提升

定位设备容易受到大型建筑物遮挡、电磁干扰、车辆低速行驶等因素的影响, 使得采集到的位置数据与真值具有较大偏差. 例如, 车辆在等待红绿灯时, 容易产生大量偏离其真实移动方向的位置信息, 表现为与实际行驶主航向具有明显差异的位置点, 称为位置漂移. 同时, 部分人为因素 (例如, 出租车或货车司机为节约能源临时关闭车载定位设备) 或传感器故障等使得位置采集设备在少数时段无法准确获取移动对象的轨迹信息. 上述因素导致采集到的轨迹数据面临经/纬度出界、数据缺失、数值异常等问题, 严重影响了基于轨迹数据推断地图的准确性.

为了有效缓解低质轨迹数据对地图构建精度的影响, 可以先采用噪声过滤、预设阈值过滤等方法对轨迹数据进行清洗; 再对轨迹数据进行正确性校验, 包括数据缺失判断以及不一致数据检验等, 同时, 可结合车辆的历史移动趋势与近邻对象的实时移动行为对轨迹数据进行校准; 随后, 利用卡尔曼滤波器、粒子滤波器、均值/中值滤波器、基于航向/距离等特征的去噪技术消除异常的轨迹数据. 此外, 为了大幅度减少轨迹数据的实时处理规模, 可运用在线数据压缩、采样、基于时间间隔或轨迹形状的轨迹分段等技术, 在保留轨迹数据时空特征的前提下, 对去噪后的轨迹数据进行近似化地表示.

1.2.2 轨迹聚类

在基于轨迹数据构建地图的过程中, 聚类扮演了极其重要的角色. 作为一类无监督的移动模式发现技术, 聚类根据既定的相似性度量方法将采集到的轨迹数据集划分成若干相似的簇, 继而提取不同移动对象的代表路径或其共同移动趋势. 在地图构建时, 聚类常用于缺失路段的推断与转向路径的识别, 通过将轨迹按航向/速度等特征划分成簇来获取密集轨迹分布区域的道路拓扑结构, 常用方法包括 K-means 聚类、DBSCAN 聚类、Mean-shift 聚类等.

定义轨迹间的相似性度量方法 (亦称为距离函数)^[1]是轨迹聚类首要解决的核心问题, 常采用欧氏距离、DTW (动态时间归整)、LCSS (最长公共子序列)、ERP (实补偿编辑距离)、EDR (实序列编辑距离)、Hausdorff 距离、Fréchet 距离等. 其中, 欧氏距离适用于等长轨迹间的距离计算, 但由于其容易受噪声数据影响, 不适用于评测不等长、不同采集频率的以及时间尺度不一致的轨迹间的距离. DTW 方法允许位置数据元素根据距离度量的需要参与多次计算, 适用于任意长度轨迹间的距离度量, 但该方法也对噪声数据较为敏感. LCSS 方法通过计算轨迹间最长相

似子轨迹序列实现轨迹间的相似性评测.然而,该方法不考虑轨迹间相似子序列的不同间隙大小,这在一定程度上降低了相似性评测精度.ERP方法支持局部时间位移,但同样易受噪声数据的影响.EDR方法通过量化位置元素对之间的距离降低了噪声数据的影响.上述方法都是基于欧氏空间的相似性度量,计算复杂度较高.Hausdorff距离常用于测量两个数据集中最近数据点对之间最短距离的最大值.由于它在距离度量时结合了方向特性,常用于评测道路网络中具有不同航向的轨迹间的相似性,将其描述为路段之间的最长距离评测.对于实际路网中采集到的曲线轨迹,由于Fréchet距离在相似性度量时考虑了曲线的单调性和连续性,也常用作地图构建方法中的轨迹距离评测,但是该方法容易受轨迹测量误差影响导致地图推断精度不高.

1.2.3 道路中心线拟合

道路中心线拟合是利用轨迹数据进行地图构建的核心步骤,通过对由轨迹聚类等技术产生的来自同一路段的轨迹子集进行拟合,提取能代表其底层道路实际位置与形状的曲线.因此,中心线拟合的准确性将直接影响生成电子地图的精度.

现有的道路中心线拟合技术通过提取位于同一路段上的多个转向特征点序列实现中心线的拟合.根据其实现原理,可将现有技术分为4类:基于计算的拟合方法、基于采样的拟合方法、基于Voronoi图的拟合方法以及基于密度的拟合方法.其中,基于计算的拟合技术应用最为广泛,它先通过计算获取道路中心点(即转向特征点),再将其顺序连接形成道路中心线.代表性方法包括Sweep-Line^[2-5]、序列对齐法^[6]、K-means质心连接法^[7]、B样条拟合法^[8]、主曲线法^[9,10]等.基于采样的道路中心线拟合技术根据预设的采样规则,从轨迹集中筛选出能代表大多数轨迹形状的一条轨迹作为道路中心线.常用的采样规则包括距离最小和采样法^[11]、起终点间最短路径法^[12]等.为了提升基于采样的道路中心线拟合方法的有效性,可利用吸引力模型在采样前调整轨迹位置,使沿同一路段的轨迹集分布更为紧密^[13].基于Voronoi图的道路中心线拟合技术利用轨迹采样点重建曲线,在由轨迹建立的Voronoi图中提取Voronoi单元间公共边上的端点作为中心点,获取每个Voronoi单元内的中心线(曲)线段,再根据单元间的邻接关系连接各段生成道路中心线^[14].基于密度的道路中心线拟合技术则在提取轨迹点沿道路分布密度最高的区域作为道路中心线^[15].

基于轨迹数据的地图构建方法大多选择一种中心线拟合方法提取道路网络中的路口、路段等区域.然而,实际路网中的道路形态较为复杂.具体地,道路形状呈现不同的弯曲程度,表现为近直线型、转向平缓型、直角转弯型、急转弯型等;此外,道路形态复杂多样,如道路交叉口附近的自相交型道路、掉头路段等.目前未有适用于所有道路形态的中心线拟合方法.在构建地图的过程中,基于待推断道路的初步形态感知,结合拟合计算开销与道路精度等因素为不同形状道路确定恰当的中心线拟合策略,确保获取精准的道路拓扑结构.

1.2.4 地图匹配

鉴于位置信息采集的低精度误差容易导致采样获得的位置点与地图中的真实位置出现较大偏离,采用地图匹配(map-matching)技术能将轨迹映射到基准道路网络中实际对应的路段序列.因此,地图匹配常作为定位信息的纠错技术,通过结合车辆行驶产生的轨迹数据与电子地图的路网信息,将车辆的实际位置与电子地图中距离最近的道路进行比较,获取车辆所在路段的位置.具体地,根据车辆与路段间的距离以及航向夹角、路段间的上下文邻近关系等识别车辆在路段中的真实位置,从而实现定位误差的修正.此外,在缺失道路发现、路口缺失信息检测等地图构建工作中,地图匹配常作为利用真实轨迹校准基准道路网络的重要步骤.通过对轨迹簇进行中心线拟合获取的道路位置与几何形状,利用基准路网对其进行匹配以检测道路拓扑结构不一致的区域,为路网校准提供支持.

根据地图匹配应用的场景,大致可分为在线模式和离线模式.由于移动对象的位置信息连续采样,在线匹配方法以流式处理,即每次只对实时采集到的位置信息进行地图匹配.而离线模式以匹配开销较小的最优匹配路径为目标,在获得移动对象的整条轨迹后进行.常用的匹配算法遵循“最接近”思想,大致分为基于路网几何信息的方法、结合道路拓扑结构的方法与基于概率理论的方法等.其中,几何地图匹配方法关注路网的几何信息,不考虑道路之间的连接属性.拓扑地图匹配方法结合道路之间链接的连通性与连续性,能获得更为准确的匹配精度.隐马尔科夫模型(HMM)^[16]是应用最为广泛的基于概率理论的地图匹配模型,在匹配路网时考虑了映射路段序列的合理性,能较好地适应地图匹配过程.它将每个采集到的位置点信息作为观测值,将移动对象在道路上未知的实际位置作为

隐藏状态. 由于轨迹测量误差, 所有与观测值接近的道路都有一定概率是车辆的实际位置 (状态), 当车辆连续移动时, 两个连续时刻之间的转换由候选状态之间的移动可能性 (转换概率) 计算得到. 因此, 每个位置点的地图匹配可视为寻找连接一个候选对象的最优路径.

2 基于轨迹的地图构建研究进展

近 10 年来, 国内外学术界、工业界就地图推断与更新问题的理论建模及优化问题进行了大量探索实践. 但面向大宗货物运输的物流地图构建算法理论研究与应用仍处于探索阶段. 我们围绕电子地图的推断、更新方法以及相关核心技术讨论国内外的研究进展.

早期的电子地图构建主要依靠人工进行采集和测绘, 由商业地图公司调派配置定位设备车辆通过巡游路网对所有路段进行实地探测. 该方法虽然能够较为精准地获取道路信息, 但耗费了大量的人力、物力, 不能保证地图的及时更新. 图像处理技术的迅猛发展使得基于图像推断地图成为可能, 部分学者通过对卫星遥感影像与航空图像数据进行分析处理推断地图^[17-21]. 但是, 由于云层等大气现象的干扰, 许多道路部分或完全地被树冠、高层建筑、桥梁和交叉路口的阴影所遮挡, 基于图像处理技术推断的电子地图存在一定的失真现象; 同时, 遥感影像数字化还涉及大量的后期处理且精度有限, 不能保证电子地图的实时高精度更新.

随着移动定位技术的高速发展以及手机、车载导航等移动设备的广泛应用, 衍生了海量的时空轨迹, 因为它们蕴含了丰富的路网结构信息且采集成本较低, 基于轨迹数据的地图构建成为目前采用的主流方法. 国内外研究学者从不同角度对基于轨迹的地图构建算法进行了综合分析^[22-26]. Biagioni 等人于 2012 年首次将地图推断方法分为 3 类: K-means 聚类、核密度估计 (KDE) 与轨迹合并, 并对其进行了定量评估^[22]. 结果表明, 基于 KDE 的方法总体上具有更好的精度和更低的时间开销, 这也成为 Biagioni 等人进行后续 KDE 方法研究的动机之一. Ahmed 等人将 K-means 和 KDE 合并归结为基于聚类的方法, 增加了交叉链接类方法. 他们基于多个数据集, 选择 Hausdorff 距离、基于最短路径的距离等多种不同的评价标准, 对主流地图构建算法进行了综合评价与比较^[23,24].

轨迹数据的密度、精度及其底层网络的复杂性对地图推断算法构建的地图质量具有至关重要的作用. Hashemi 等人通过建立一个用于评估路网构建算法的实验平台, 基于不同密度和复杂性的街道和行人网络, 根据完整性、精度以及拓扑正确性等评价指标对各类算法进行了综合评估^[25]. Duran 等人基于徒步的轨迹数据, 选择了 4 个具有不同地理特征的地形区域, 从生成地图的局部细节方面对 5 种主流地图构建算法进行了对比分析^[26]. 作为最具代表性的调查, 这些有意义的综合评价能帮助新的研究人员快速了解基于轨迹数据的地图推断机理. 不同于文献^[22-26], 本文重点关注面向大宗货运的物流地图构建, 首先对已有地图构建算法进行了回顾, 对基于轨迹的地图构建技术的研究进展进行了分类概述, 同时指出了现有方法在解决物流地图构建问题上存在的局限性.

目前, 基于轨迹的地图构建技术根据不同的应用场景以及不同的实现原理, 大致可以划分为 4 类: 基于聚类的方法、基于密度的方法、基于轨迹增量插入的方法以及基于交叉路口链接的方法.

2.1 基于聚类的地图推断技术

基于聚类的地图推断方法^[3,7,8,27-32]是最基本的地图构建方法. 早期的方法将轨迹阐释为位置点的集合, 通过不同方式将位置点集合进行聚类, 获取交叉节点或街道 (路段) 继而生成整个街道地图. 输入的轨迹集可以是原始采集的位置点集合, 也可以是所有输入位置点的密集样本. 同时, 输入轨迹被假定是一条连续的曲线, 通过在输入位置点序列之间进行插值计算获得.

Edelkamp 等人率先提出了基于 K-means 聚类的地图构建方法^[27]. 该方法在距离和移动方向限制条件下, 将移动对象的轨迹点集划分成簇, 产生沿道路方向的轨迹簇中心, 通过使用拟合样条方法提取道路中心线. Schrödl 等人提出了包含道路分割、地图匹配和车道聚类的地图推断框架, 通过采用 K-means 算法对轨迹点进行聚类、使用最小二乘逼近法拟合生成道路中心线^[8]. 然而, 上述通过寻找聚类中心的方法将聚类中心连接形成道路中心线, 不能精准表示实际道路形状, 故地图推断精度较低.

Li 等人针对轨迹采样率低且伴有噪声等问题, 结合路网的线性特性提出了一种基于空间线性聚类 (SLC) 的地

图构建方法^[3]。由于考虑了位置信息测量误差的范围和连续采样点间的移动方向, 通过对空间线性分布的轨迹点聚类生成细长的空间线性簇, 能近似捕捉道路的几何形状; 随后, 他们设计了一种几何方法计算空间线性簇的代表轨迹, 实现地图推断。Qiu 等人针对轨迹的低采样率问题, 提出了一个三阶段的地图构建方法^[7]。先结合方向与密度特征对轨迹点集进行聚类, 将其划分为代表短距离路段的轨迹点簇; 随后, 基于自适应 K-means 算法提取轨迹点簇的道路中心线; 在此基础上, 依据格式塔原理设计道路中心线分组策略对代表短距离路段的道路中心线进行连接继而生成成长距离的路段。

Stanojevic 等人将路网推断问题描述为一个网络对齐优化问题^[28]。先对轨迹点进行 K-means 聚类推断顶点, 通过查找连接每个顶点(质心)对的轨迹推断顶点间的边, 并利用 Graph Spanner 算法去除冗余边, 构造保留连接信息的稀疏图作为道路地图。Chen 等人根据道路中心区域轨迹点密度较大的观察, 提出了基于均值滤波的方法获取位于道路中心线附近的采样点, 根据采样点的方向与距离特征提出了基于图的道路聚类方法连接采样点生成路段; 同时, 结合地图匹配方法获取轨迹数据映射的路段序列来捕捉路段间的上下文邻接关系^[29]。Worrall 等人利用位置与方向信息对轨迹点进行聚类, 获取车道级别的轨迹点簇及其方向, 再根据轨迹簇的方向一致性与距离等信息连接轨迹簇形成初始路网, 并利用加权最小二乘曲线拟合方法提取各簇的道路中心线^[30]。Jang 等人将路网区域按网格划分, 通过对分布在网格单元中的轨迹点进行聚类获取强关联的网格, 再使用样条曲线拟合道路以捕捉整个路网的拓扑结构^[31]。

与上述轨迹点聚类方法不同, Liu 等人针对稀疏采样的轨迹数据, 提出一个三阶段的轨迹线段聚类方法^[32]。该方法基于轨迹线段方向和轨迹点方向筛选与道路方向一致的轨迹线段集, 再采用基于层次聚类的方法对轨迹线段集进行迭代聚类, 同时结合 B 样条拟合技术提取轨迹线段簇的道路中心线。

与地图推断技术相比, 基于聚类的地图更新技术着重强调相对于基准地图的遗失(新)路段发现。先通过地图匹配获取相对于基准地图未匹配的轨迹数据, 再对其进行聚类以识别遗失路段。根据所处理的数据对象不同可划分为基于轨迹线段的方法^[10]与基于轨迹点^[33,34]的方法。两类方法在轨迹分布密集的区域具有较好的检测性能, 但是不适合处理轨迹分布稀疏区域的路段检测。文献[4]采用基于方向与基于距离的双重去噪预处理方法, 利用滑动窗口模型, 提出了结合轨迹点与轨迹线段处理的混合新路发现方法 HyMU, 有效地解决了轨迹稀疏分布区域的缺失道路检测问题。

2.2 基于密度的地图推断技术

基于密度的算法^[15,35-37]通过计算将输入点集转换为基于密度的离散化图像。其中, 基于核密度估计(kernel density estimation, KDE)的密度方法应用最为广泛, 该方法通过计算轨迹点的分布密度, 并根据轨迹点的密集程度将历史轨迹数据转化为栅格图像; 在此基础上, 利用 Voronoi 图或形态学方法捕捉连续的高密度栅格以获取道路中心线, 继而推断潜在的路网拓扑结构。当数据被高频次采样时, KDE 类算法表现较好。

Davies 等人提出了一个基于轨迹分段的 KDE 方法, 该方法先评测轨迹分段的分布密度, 采用轮廓跟踪方法生成道路的初始轮廓, 随后利用 Voronoi 图提取初始轮廓的道路中心线, 并通过预设阈值的方式修剪位于轮廓以外的分段^[35]。文献[36]利用非参数核密度估计方法提取车道级道路信息, 先在分段后的道路中心线上选择一个采样点作为垂直线, 再使用非参数核密度估计处理该垂直线与交点坐标得到核密度分布函数, 以函数的峰值作为车道所在位置, 而峰值的个数代表车道的数量。Biagioni 等人采用核密度函数, 通过设置不同的阈值获取连续的多尺度栅格图, 并使用骨架提取算法去噪获取道路中心线^[15]。Dey 等人将密度图视为地貌图形, 结合离散摩尔斯理论在密度图中以识别山脊线的方式提取道路中心线^[37]。

基于 KDE 的方法在高采样频率、数据分布均匀的情况下具有较高的地图推断精度。然而, 由于它们需要将轨迹转换成栅格图像, 容易丢失重要的局部位置信息以及轨迹的上下文关系, 在轨迹分布稀疏的区域, 所推断的道路拓扑结构与真实路网有较大偏差, 因此生成的地图质量不高。

2.3 基于轨迹增量插入的地图推断技术

基于轨迹增量插入的地图推断方法^[13,38,39]由空地图开始, 在每次插入一条或多条轨迹时, 根据轨迹点的空间

与方向特征不断地合并相似轨迹, 递增式地扩展地图. 具体地说, 该类方法借鉴了地图匹配思想, 使用第一条轨迹作为基础地图, 通过不断添加更多轨迹来逐步完善、细化地图. 在插入轨迹的过程中, 利用轨迹与基准地图匹配部分校准匹配道路的几何结构, 同时, 通过利用未匹配的轨迹推断新的顶点与边来完善地图.

为降低轨迹噪声的影响, Cao 等人在进行轨迹合并前, 先引入了一个应用物理引力的“澄清”步骤, 通过模拟粒子之间的引力与斥力调整轨迹位置^[13]. 在两种力中引力较强且作用距离较短, 可将属于同一方向道路的轨迹“吸引”在一起; 排斥力较弱且作用距离较长, 能避免轨迹受引力作用偏离其真实位置. 两种力的相互作用使来自同一条道路的相似轨迹更为紧密, 而不同方向道路上轨迹更加分散, 沿道路中心线形成紧密的带状. “澄清步骤”输出相对纯净的数据集, 增量插入算法在此基础上进行地图推断, 根据轨迹点与图节点的距离、方向差异, 依次对插入轨迹的轨迹点进行局部决策, 判断其是否与当前地图匹配, 将未匹配轨迹转换成地图中的新顶点与边, 匹配轨迹则进行地图合并.

Ahmed 等人提出了一个三阶段的增量插入方法: 先将轨迹与基准地图进行基于 Fréchet 距离的地图匹配, 再将未匹配轨迹插入到基准地图中以创建新的顶点与边; 此外, 基于匹配轨迹使用最小链接算法更新地图中已存在的边, 计算获取新的代表性道路^[38]. 不同于文献 [13] 仅根据距离和方向对插入轨迹点进行局部决策, Fréchet 距离度量更适合于连续曲线的形状, 使用 Fréchet 距离对轨迹曲线进行局部决策, 可以更好地利用输入轨迹的连续结构信息, 进而准确获取道路的真实形状.

基于轨迹增量插入的地图推断方法被广泛地用于地图更新, Zhang 等人提出了一种基于空间相似性的地图更新算法^[39]. 它使用一种类似轨迹合并的方法不断完善现有的道路地图. 根据轨迹与现有道路的距离、方向关系以及轨迹与道路之间的角度, 将轨迹与已有道路进行匹配筛选出道路的候选轨迹. 随后, 采用模糊 c 均值聚类的方法, 将同方向且距离邻近的轨迹划分成簇, 通过对其提取新的道路中心线以更新道路的几何形状, 并从轨迹中挖掘车道数和转弯限制等道路属性信息.

由于轨迹合并过程耗时长, 计算开销大, 同时, 轨迹数据伴有的大量噪声容易引致算法在增量插入过程中推断出错误的道路信息, 因此, 基于轨迹增量插入的方法不适用于大规模路网的地图推断.

2.4 基于交叉链接的地图推断技术

基于聚类、密度与基于轨迹增量插入的地图推断技术通过从轨迹中挖掘地图的点或边来生成地图骨架. 基于交叉链接的地图推断技术则采用两阶段方法, 先检测路网中的交叉路口节点, 再推断路口间的道路, 通过链接交叉路口形成完整的道路地图. 该类方法强调对交叉路口的正确检测以保证道路间的连通准确性, 从而提升地图推断的质量. 现有的交叉路口检测方法根据车辆大多在路口改变航向的假设, 重点关注道路交叉口区域内的转向轨迹, 通过对其聚类来确定交叉口位置. 检测技术大致分为基于路段交汇的路口检测算法与基于区域特征的路口检测算法.

交叉路口是道路的交汇处, 基于路段交汇的路口检测技术^[40,41]提取道路之间的交点作为道路交叉口. Xie 等人将地图划分为网格, 先采用动态规划方法搜索轨迹中的最长公共子序列, 再从中推导出连接点, 以最长公共子序列的两端为路口, 继而确定相交位置^[40], 该类方法计算开销大且对轨迹数据质量要求较高. Li 等人基于轨迹推断道路的骨架图, 在保留独立冲突方向的同时合并相似方向, 以交叉口周围路段的优势方向作为交叉路口^[41].

考虑车辆行驶通过交叉路口时常发生航向/速度变化, 基于区域特征的路口检测技术根据轨迹数据的航向、速度变化特性判断待测区域为交叉路口, 通过提取转向特征点并对其进行聚类获取路口中心位置. Fathi 等人提出了基于子轨迹特征的形状描述符, 通过为位置信息构造特征向量以及训练二分类模型识别待测位置为道路交叉口^[42]. Chen 等人基于图的聚类技术推断路段, 提出了尺度与方向恒定的 TRAJ-SIFT 特征, 并使用监督学习方法识别不同类型的交叉路口^[29]. Karagiorgou 等人根据预设的速度/方向阈值提取轨迹中的转向点, 基于转向点的空间邻近性与转向类型聚类生成不同的转向点簇, 以其质心作为交叉点^[5]. Wu 等人设计了一种基于转向点搜索和汇聚点聚类的交叉点识别机制^[43]. 由于道路网络中的交叉路口形状各异且大小不一, 为解决这个问题, Huang 等人提出先检测主路口与次路口再将两者组合的方式获取路口信息^[44]. 他们通过计算连续点对间的方向差异筛选转向点对、计算聚合点与约束聚合点; 随后, 采用 DBSCAN 方法对聚合点与约束聚合点进行聚类, 提取交叉路口的位置、方向与类型 (按主路口与次路口分类). Pu 等人提出了一个交叉路口检测框架, 通过提取基于方向统计分析的候选

单元、采用混合聚类策略进行交叉位置优化, 检测不同大小的道路交叉口的的位置与范围^[45]. Wang 等人采用基于 force attraction 的聚类方法对轨迹预处理, 再以轨迹交点为特征点进行聚类确定道路交叉口范围^[46].

上述方法大多侧重于识别道路交叉口的的位置和覆盖范围等信息, 较少关注道路交叉口的空间连通多样性以及复杂的可路由拓扑结构. 鉴于此, Zhao 等人将道路交叉口检测问题扩展为交叉口影响区内的道路拓扑结构校准问题, 提出了一个三阶段的路口拓扑校准框架 CITT^[11], 包括轨迹数据质量提升、核心区检测和影响区内道路拓扑结构校准等. CITT 方法不仅能确定道路交叉口的的位置和覆盖范围, 还可以自动获得高质量的道路交叉口影响区拓扑, 通过与基准地图匹配找出路口影响区域内不正确/缺失的转向路径, 具有较强的稳定性和鲁棒性.

在检测出交叉路口后, 部分研究^[5,42,44,46]使用链接算法将交叉路口链接起来, 推断路口间道路的形状与位置. Fathi 等人链接交叉点间的最短距离轨迹, 将其作为交叉点间的道路, 并将交叉点间的最短间接路径替代长度相近的直接路径, 继而解决链接过程中可能遗失中间节点的问题^[42]. Karagiorgou 等人通过发现对推断交叉点有贡献的轨迹, “捆绑”交叉点周围的轨迹以筛选交叉口链接序列, 使用 Sweeping Line 方法拟合路段中心线, 创建地图的边^[5]. Huang 等人使用滑动窗口模型沿着交叉路口的驶入或驶出方向搜索轨迹, 以搜索到的第一个路口作为邻接点, 不断链接邻接点生成路网^[44]. Wang 等人根据交叉口区域内外推断的道路中心线共享边界的同一入口/出口, 判断其链接关系, 据此合并交叉口范围内外两区域的子图生成路网^[46].

轨迹数据包含大量噪声, 具有不确定性与偏态分布性; 道路网络中的交叉路口拓扑复杂、类型多样, 从不确定性轨迹中准确识别交叉路口位置/范围以及提取路口范围内的道路拓扑, 是基于交叉链接的推断技术亟需解决的重要问题, 此外, 链接交叉路口推断路段时, 路段形状位置的准确性与节点间链接的正确性也是影响构建地图精度的重要因素.

2.5 已有地图推断技术构建大宗地图时面临的局限性

(1) 现有的地图推断方法利用道路网络内的车辆轨迹数据生成地图, 未考虑适用于不同车型车辆出行的热门停放区域标注, 未识别带有时空约束的禁/限行路段与车辆频繁行驶的路段, 因此, 不能直接用于公路大宗货运的物流地图构建. 为此, 需结合车辆历史轨迹数据、道路信息、物流运营信息等多源物流数据, 从基础路网中筛选不同时段适用于承运各类大宗货物车辆的停留热点与热门运输路径, 推断道路拓扑结构, 进而构建多约束的物流地图.

(2) 现有的地图更新方法利用相对于基准地图的不匹配轨迹发现缺失路段, 但物流地图的增量更新需要及时检测因交通限行规定调整的带约束的禁/限行路段与各时段内的可通行路段, 以及路段间转向关系变化引致的新邻接关系, 亟需结合持续增加的时空数据提取道路拓扑结构的变化特征, 定位变化影响的路网区域, 保证物流地图的持续更新.

此外, 大多数地图推断算法只能生成无向地图, 未关注转弯限制和速度限制等特征的提取, 需要通过轨迹地图匹配等后处理步骤完善地图. 因此, 面对公路大宗货运的物流地图定制需求, 已有地图构建方法仍存在一定的局限性, 迫切需要提出高效的方法在地图构建过程中获取适用于大宗货运的运输停留热点与多约束(带时间约束的禁行、限高、限重等)的热门路段, 并通过及时感知道路拓扑结构的动态变化保证物流地图的持续更新. 针对上述问题, 为大宗货运构建物流地图亟需采用数据驱动模式, 以大宗货运产生的时空数据为核心, 结合货运车辆、运单信息等多源物流数据, 挖掘运输停留热点与热点间的频繁路段序列, 构建适于各类大宗货物运输的专属物流地图; 与此同时, 利用持续增加的时空数据洞察道路拓扑结构的变化, 保持对物流地图的增量更新, 为运输线路规划等应用提供可靠保障.

3 挑战与框架

3.1 物流地图构建面临的挑战

数据的海量、不确定、稀疏、偏态分布、时变演化等特性, 以及地图推断的高效与高质量需求, 使得融合多源物流数据构建物流地图面临严峻挑战.

(1) 噪声: 由于定位技术的有限精度与 GPS 定位设备的计算误差、信号衰减及丢失导致采集到的位置数据伴

有噪声,表现为轨迹数据具有大量的位置偏差.具体地,当货车通过山区道路时,卫星的可见性较差,定位设备接收器的记录坐标可能偏离实际位置,甚至可能映射到道路以外区域;此外,货车运输通过连续弯道或等待红绿灯时常处于低速行驶状态,容易产生短距离范围内多个与主行驶方向不同的位置点,即位置漂移.这些噪声数据在道路推断过程中对于道路长度、形状与转向信息的正确提取有较大影响,因此,如何有效剔除噪声提升数据质量成为地图推断工作的一大挑战.

(2) 低采样率:对于连续移动的物体,其位置信息由定位设备定期采样获得,不同采样频率会带来位置数据的时序不确定性.因此,位置信息的采样率是影响轨迹数据表示精度的一个重要因素.在货车行驶过程中,部分司机为节省车辆能量消耗常选择关闭定位设备,定位设备信号丢失引致轨迹数据采集频率较低.对于低采样率轨迹,两个连续位置点连接生成的线段较长,可能与它们之间的实际路径不符.特别是当车辆行驶的路径不是直线时,采样得到的轨迹线段可能穿过农田、湖泊等,而不是沿着实际的道路.因此,如何基于低采样率轨迹推断高质量的物流地图是一大挑战.

(3) 轨迹偏态分布:在货车行驶的道路网络中,不同道路通行车辆的频率具有较大差异.部分货车司机由于个人运输偏好选择运输成本较低的省道,这些路段在很长时间内仅有少量车辆通过;而多个运输流向必经的主干道路在较短时间间隔内却有大量货车经过.相应地,轨迹数据常常表现为偏态分布.在地图推断过程中如何兼顾不同数据分布区域设计恰当的推断方法与停留点提取方法是一大挑战.

(4) 数据时变演化:货车行驶产生的轨迹数据实时产生且持续增加.只要处于活动状态,位置信息就会不断产生,因此,轨迹的数据量是无限的.为保证物流地图的增量更新,需要解决数据的高效分析问题.但是,在线地图更新算法在数据的计算过程中,无法保存采集到的全部轨迹数据:硬件设备没有足够大的空间存储无限增长的轨迹数据;也缺少恰当的方法有效管理海量数据.对于快速到达的轨迹数据,亟需设计能精确获取移动行为特征的实时分析方法.

(5) 停留热点样本分布不均衡:在货车运输过程中,卸货装载、连续弯道行驶以及交通拥堵产生了大量低速行驶点,增加了停留热点识别的难度.对于任一流向的货运线路,货车在货物装载/卸货地点停留较多,而在运输途中的加油站、维修点等区域停留较少,不同类别的运输停留热点样本分布极度不均衡,容易导致停留热点分类模型倾向于多数类样本、泛化能力较差.因此,如何解决类样本极度不均衡下运输热点的功能语义标注问题成为一大挑战.

(6) 经验路径计算开销高:在运输过程中,不同时段出发执行运输任务的货车司机会综合运输成本、货物应交付时间以及交通状况等因素选择不同的经验路径.单一时段内执行相同流向运输任务的货车司机也会根据个人偏好(如耗时最短、油耗最低、过路费最低等)等选择不同的运输路径.然而,在货运路网中基于轨迹数据搜索(差异性)频繁路径涉及较高的计算开销,为了避免多次扫描时空轨迹数据库,如何结合剪枝策略加快频繁路径搜索效率成为物流地图推断的一大挑战.

(7) 道路形态复杂:道路中心线能反映路网中各条路段的平面位置与曲直变化.大宗物流的长距离运输大多需要在崎岖山路区域行驶,受各地区的不同地形限制,连续弯道型、急转弯型道路等不规则形状道路居多.相比普通城市路网中占比较大的近直线型道路,复杂转向道路的中心线拟合难度更大.因此,针对物流路网内道路形态的复杂性与多样性,如何设计有效的中心线拟合策略从轨迹数据中准确识别道路形态成为物流地图推断的一大挑战.

3.2 物流地图构建框架

研究数据驱动的公路大宗货运的物流地图构建,需要针对不同类型的大宗货物运输.

- 基于历史数据挖掘,通过识别运输热点、提取可达热点对之间带有多约束(通行时段、限重、限高等)的热门运输线路,获取其覆盖的路段序列,构建适于各类大宗货物运输的多约束物流地图;

- 利用持续新增的时空数据进行在线分析,通过地图匹配、道路流量与路段间转向概率统计等方法发现道路拓扑结构变化区域,甄别新通行路段、禁行路段以及转向变化引致的路段新邻接关系等不同拓扑变化类型,采用聚类、中心线拟合等方法实现道路拓扑结构的增量更新(如图2所示).

(1) 多源数据融合

货车运输区别于城市公共交通出行方式,它没有固定的线路和站点,其行驶过程中的时空行为能够很好地反映不同类型货物运输的出行特点与城市交通的运行状况.结合轨迹数据与物流营运数据、基础路网数据等进行分析,可以有效获取道路网络中不同时段内的可通行路段.同时,还能通过提取货车司机的相似出行行为,识别运输途中的热门停留点以及停留热点间的频繁运输路径.因此,基于公路大宗货物运输中产生的多源物流数据进行关联分析,通过提取运输停留热点与热门运输路径构建适用于大宗货运的专属物流地图.

公路大宗货物运输过程中产生的物流数据包括:交通的基础地理信息(如道路网络、交叉路口布局等)、通过定位设备采集的货车轨迹、物流运营所产生的订单数据与货车派单数据、运输线路规划信息等.以订单实体为中心对不同来源的数据进行关联,并运用分布式架构

并行实现实体匹配.采用基于时空行为特征的实体关联方法提取不同车型货车在不同流向运输产生的物流数据.

(2) 轨迹数据的预处理

从动态的时空数据中发现运输热点与热门运输路径是一项极具挑战性的工作.以钢铁物流企业的货车运输轨迹数据为例,每个数据记录主要包括车牌号、数据接收时间、经度、纬度、瞬时车速、角度(以正北为起始值,顺时针方向计算的瞬时夹角)、定位时间、行驶总里程等属性.由于信号延迟等原因,货车的定位时间与信息中心接受车辆定位信息的时间存在差异,可采用货车定位时间作为标准时间.进一步分析数据可发现,轨迹具有大规模、采集间隔不稳定、冗余、位置漂移等特征.轨迹数据的低质量往往导致低效的挖掘效果,使得从数据中发现的模式不准确,因此进行有效的轨迹数据预处理工作是精确提取道路拓扑结构、构建物流地图的重要前提.

- 数据量大.货车单日平均定位数据记录有800多万条,由于定位信号弱以及货车停车导致部分记录重复现象,可能造成运输停留热点的误判.针对这个问题,可对轨迹数据进行压缩以提高数据分析效率,采用道格拉斯普克(Douglas-Peucker)算法保留有效特征点方式实现轨迹压缩;

- 数据采集不稳定.受天气、环境等因素影响,易造成定位信号丢失或数据无效等现象.针对这个问题,考虑货车行驶产生的连续轨迹点其速度与方向属性应保持一定范围的一致性,基于货车的出行行为通过插值计算方式填补轨迹空缺;

- 位置点漂移.货车位置采样点由于定位误差,不可避免地存在位置漂移现象,特别是在货车低速行驶状态下表现明显.具体地,同一辆货车在较短距离范围内产生的位置点信息具有多个方向.漂移现象为物流路网拓扑结构的推断带来较大干扰.鉴于货车位置点基本沿道路中心线分布,采用基于轨迹点主方向的去噪以及基于距离去噪等处理,避免位置漂移影响,进而提升轨迹数据质量.

(3) 物流地图的构建

从多源物流数据中挖掘用户的先验知识,通过获取运输停留热点与热点间的频繁路径发现其覆盖的路段序列,从而生成多约束的物流地图.

首先,通过挖掘货车的出行规律识别运输途中的热门停留点(区域),可获取构建大宗货物流网的基本要素.但是,由于通过连续弯道、交通拥堵、路口红绿灯等因素产生了低速行驶轨迹点,增加了运输停留热点的识别难度.结合多源物流数据,可利用车速、停留时长与停车位置距离道路的距离等停留行为特征、聚类停留点获取的停留区域特征,以及停留热点之间的关联特性等精准识别运输停留热点.

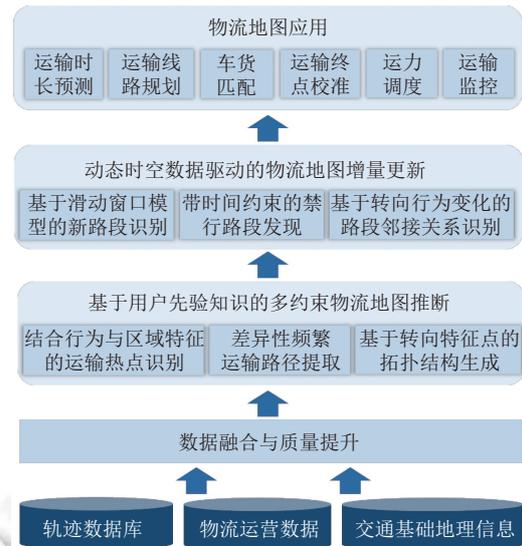


图2 物流地图构建框架

其次, 鉴于道路网络中部分路段存在带时间约束的禁止通行以及限高、限重等规定, 可利用历史轨迹提取运输停留热点间的直接可达关系, 获取不同时段内可达热点间不同车型货车的热门行驶路径. 分析历史物流数据能获取不同车型货车在不同时段内的路线选择经验, 可提取不同时段内停留热点对之间的频繁运输路径. 频繁运输路径实质是基于货车行驶路径的统计结果, 不但避开了对应车型在各时段内的交通禁/限行路段, 还包含了承运不同货物行驶经过的热门路段. 将不同车型货车运输的通行时段、货高、车重等属性作为频繁路径的交通限制约束. 选择恰当的轨迹相似性度量进行轨迹聚类, 再根据预设的轨迹数支持度阈值, 发现直接可达热点对之间差异的频繁轨迹簇.

最后, 对于停留热点对之间提取的不同频繁运输路径 (以轨迹簇表示), 通过拟合其映射到路网中的路段序列生成多约束的物流地图. 考虑停留热点对之间的长距离, 其对应的货车轨迹不仅量大, 还常表现为具有多转向的曲线形状, 可先提取转向特征点将各轨迹簇拆分成子簇, 再对拆分后的各轨迹子簇分别拟合道路中心线. 基于此, 停留热点对之间的多约束运输路径可由多段道路中心线的序列构成. 随后, 使用地图匹配方法将道路中心线序列映射为路网中的路段序列, 并采用邻接矩阵存储路段间的可达关系. 最终, 基于不同时段内运输路径覆盖的路段序列集生成多约束的物流地图.

(4) 物流地图的更新

城市道路网络随时间不断变化, 公路大宗货物运输的物流地图也随之动态演化. 一方面, 随着城市道路基础设施建设的飞速发展, 道路网络中不断增加新建的道路与临时封闭的道路; 另一方面, 出于城市建设的道路安全性考虑 (如防止疫情扩散封锁部分道路), 部分路段会增加带时间约束的禁行与限行规定. 上述原因导致可通行路段、禁行路段以及路段邻接关系的动态变化, 对物流地图的道路拓扑结构 (包括停留热点) 带来不同程度的影响, 需要利用持续新增的时空数据及时洞察道路拓扑变化, 实现物流地图的增量更新.

对于新增的带约束的可通行路段, 结合不同车型的货车运输货物 (包含其承载重量、车高、车长等属性) 产生的轨迹数据, 采用基于滑动窗口模型的连续观测方法, 对每个时间窗口 (例如, 以一天作为一个时间窗口) 中各时段内不同车型货车新产生的轨迹数据, 结合物流地图进行基准地图匹配提取未匹配区域的轨迹; 随后通过对未匹配轨迹进行聚类、中心线拟合等处理获得对应时段内新通行路段 (具有货高、载重约束) 的候选集. 再通过连续观测多个时间窗口, 将满足设定时间窗口数阈值的道路中心线映射为道路网络中的路段, 以此作为该时段内带约束的新通行路段添加至物流地图中实现更新.

对于各类车型货车在不同时段内的禁行路段, 研究基于流量与转向概率变化的禁行路段检测方法. 禁行路段通常表现为轨迹数相对于历史同一时段的明显减少、邻接路段相对该路段的转向概率降低等特征变化. 利用各时段内不同车型的历史轨迹数据统计其映射路段的平均流量, 将其与相同时段内新产生轨迹映射路段的流量进行对比, 以低于预设阈值的路段作为不同车型货车的候选禁行路段; 同时, 以候选禁行路段为观测对象, 利用新增轨迹数据统计邻接路段与该路段的转向概率, 将概率明显低于历史转向概率的视为对应车型货车在相应时段内的禁行路段, 将其从对应时段内的物流地图中删减以实现更新.

由于新增的通行路段或禁行路段会导致物流地图中路段间的邻接关系发生变化 (表现为转向规则变化), 可基于转向变化检测进行路段邻接关系更新. 基于新通行路段/禁行路段的位置, 采用凸包技术加快与基准物流地图的匹配效率, 确定新通行路段或禁行路段与邻接路段交接的路口区域. 基于新产生的轨迹数据, 利用轨迹聚类提取不同时段内路口区域代表不同转向的轨迹簇, 并通过转向中心线提取以及地图匹配技术获取各转向对应的路段序列, 继而更新物流地图中不同时段的路段邻接关系.

4 基于用户先验知识的物流地图构建

4.1 基于速度与停留时长的运输热点识别

物流地图可以表示为无向有权图 $G = (V, E)$, 其中 V 是顶点集合, E 是可通行的路段集合, $E = \{e_i | e_i = (N_{e_i}, L_i, t_i)\}$. e_i 表示一条路段, N_{e_i} 为路段 e_i 的直接可达 (邻接) 路段集, t_i 为路段 e_i 的可通行时段, 路段 e_i 的约束 $L_i = (w_i, h_i, \dots)$, 其中 w_i 为路段 e_i 的限重约束, h_i 为路段 e_i 的限高约束. 每条路段可关联多个权重, 表示经过该路段所需消耗 (如时间、

距离、成本等)。

大宗货物运输具有长距离的特征, 货车由于装卸货、在途吃饭、休息、加油、充电、维修等事件在运输途中会作较多停留, 涉及的停留点包括仓库、中转枢纽、港口、铁路站点、加油站、维修点、服务区、临时停留区等。这些有意义的停留点反映了货车出行的行为模式, 通过挖掘货车的出行规律识别它们, 可获取构建大宗货物流网络的基本要素。物流平台广泛积累的包括物流运营数据、货车运输的轨迹数据以及道路设施基础数据为识别运输停留热点提供了数据基础。但是, 在长距离的运输过程中, 并非所有的低速停留点都是值得关注的运输热点, 如果仅从空间上对低速停留点进行分析会将交通拥堵或等待红绿灯时产生的停留点误判为停留热点。可结合轨迹点的速度、停留时长以及停车位置距离道路中心的距离等因素进行分析和判断。

首先, 考虑位置数据中每条记录的瞬时速度与角度存在一定的误差, 利用相邻轨迹点的经纬度以及时间等信息计算得到每个轨迹点的速度与角度。随后, 可基于以下规则筛选所有可能的停留干扰点。

① 通过结合货车运输轨迹数据、物流订单数据、发货通知单数据与基础路网数据进行分析, 确定货物运输中的装载点与运输终点等位置。

② 为了避免误判货车因交通拥堵、等待红绿灯以及通过连续弯道处出现的停留点, 对于任一货车轨迹中的所有低速行驶点, 将连续低速点之间出现少数高速点(即速度高于预设速度阈值的位置点)的记录作为速度异常变化记录, 速度阈值根据所有低速行驶点速度的均值与方差的 3 倍之和设置。如果带有速度异常变化记录的相邻低速点之间行驶里程数超过该区域同时段平均距离阈值, 将其视为非停留点予以删除。

③ 根据交通路网中大部分高速、国道、省道的 4-4 车道, 按每条车道宽度估算, 结合定位误差, 可将低速轨迹点位于距离道路中心约 20 米处位置视同位于道路中, 筛除高速、国道、省道等道路中的低速行驶轨迹点。

鉴于货车在某一位置的停留时长通常可以反映出有关停车事件性质的语义信息, 将停留时长超过设定时间阈值的轨迹点视为有意义的停留点。再利用 DBSCAN 聚类方法将相同时段内地理上邻近的停留点进行聚类, 继而识别货车运输途中的停留热点(如图 3 所示)。在设置距离阈值与停留时间阈值时, 可通过统计同一区域提取的大多数停留点的距离分布与停留时间间隔分布的平均值获得。



图 3 运输停留热点示例

4.2 多约束的频繁运输路径提取

在一定的时空范围内货车运输货物时频繁经过的路径, 体现了不同车型货车个体或群体的移动模式, 将其映

射到路网中是由一条或若干条路段拼接而成的路段集合. 通过提取运输热点之间多约束的频繁路径, 能为获取特定时段内的热门行驶路段从而构建物流地图提供支持. 鉴于轨迹聚类的方法可以在保持原有轨迹数据时空特性不变的前提下分析货车的移动模式, 采用基于轨迹聚类的频繁路径提取方法.

为了尽可能覆盖运输热点之间的所有热门路径, 利用货车的历史运输轨迹数据提取差异性频繁路径. 为确保挖掘出货车在特定时间范围内的高流量、强连续的热门运输路径集合, 使用基于时空相似性的轨迹聚类方法, 即在考虑轨迹空间相似性的基础之上, 进一步考虑时间因素的影响. 由此, 轨迹间的相似性可定义为空间邻近性与时间接近性的加权和度量. 对于轨迹的空间邻近性度量, 采用 Fréchet 距离评测. 通过对停留热点间的轨迹数据采用 DBSCAN 方法进行聚类分析, 可以将具有时空相似性的轨迹汇聚成簇, 最后将任一时段内前 K 个轨迹数满足支持度阈值的簇作为运输停留热点间的 K 频繁路径. 同时, 将行驶通过不同频繁路径的各类车型的时段、货高与载重等属性作为对应路径的限制约束予以存储.

4.3 基于转向特征点提取的道路拓扑结构生成

基于所识别的运输热点以及不同时段内热点间的差异性频繁运输路径 (即轨迹簇集合), 可以生成各个时段内带约束的热门通行路段序列, 从而构建多约束的物流地图. 为提取频繁运输路径对应的路段序列, 采取先对各个轨迹簇进行道路中心线拟合, 再将其与基准路网进行地图匹配的方式. 考虑到任意两个运输热点间的轨迹簇包含大量的轨迹点且涉及路网空间范围较大, 为提升道路中心线的拟合精度, 先将各簇轨迹基于转向特征点进行分段, 在对各段轨迹子簇进行道路中心线拟合后顺序连接各段中心线形成运输热点间的道路中心线. 随后, 将道路中心线与地图进行匹配获取映射的路段序列, 并将频繁路径的约束作为映射路段的约束.

在提取转向特征点时, 对于任一轨迹簇中所有位置点集, 将方向变化明显不同于其前序位置点的视为特征点. 随后, 根据特征点所在位置将轨迹簇划分成子簇. 采用基于 MDL 技术的方法, 确保每个轨迹子簇划分后的 MDL 代价最小化, 其目标为发现一个尽可能长的方向一致的轨迹子簇. 在提取道路中心线时, 根据各子簇的形状选择恰当的中心线拟合方法. 具体地, 根据轨迹子簇的起点和终点的角度差值先确定其路段大致形状为近直线或曲线, 分别采用基于 Sweeping Line^[2]与基于 force attraction^[47]的拟合方法.

对于近直线, 采用 Sweeping Line 代表轨迹生成方法拟合道路中心线, 先计算所有轨迹线段的平均方向, 按照该方向从左至右扫描. 将每个轨迹线段的两个端点按照扫描的方向进行排序, 再依次枚举这些点以及预设的固定间隔的点, 每次画一条通过该点与扫描方向垂直的线, 称其为“扫描线”. 通过统计扫描线经过的轨迹线段数, 当该值超过预设阈值 $MinLns$ 则将扫描线与所有轨迹交点的平均值作为代表点, 最后将所有代表点连接形成所拟合的中心线.

对于曲线, 采用基于 force attraction 的聚类方法^[47], 该方法首先在一个类中随机采样一个轨迹点作为参考轨迹, 再使用同一类中其余轨迹点不断调整参考轨迹的位置实现道路中心线的拟合. 在调整过程中, 基于 force attraction 的方法假设任意一点上有两个力作用, 通过搜索二力平衡的位置, 获得参考轨迹点的新位置. 此外, 引入基于 Fréchet 距离的采样方法改善 force attraction 随机采样对中心线拟合精度的影响, 并对生成的道路中心线采用 B-Spline 方法平滑处理^[11].

对于各个轨迹子簇拟合的道路中心线连接而成的中心线, 采用基于隐马尔科夫的地图匹配方法, 以中心线上的每个点作为观测变量, 将基准路网视为隐藏变量, 通过路网与中心线中点的欧氏距离计算获得其观测概率, 由路段之间的距离计算状态转移概率. 利用 Viterbi 算法计算在各个时刻不同状态下的最大概率路径, 实现中心线与基准路网的匹配, 获得其映射的路段序列. 最后, 基于各个时段内带约束的热门路段序列 (以邻接矩阵方式保留路段间的上下文关系) 以及标注的运输热点构建多约束的物流地图.

5 动态时空数据驱动的道路增量更新

城市的道路网络随时变进化, 同时, 各城市的货车交通限行规定也在不断变化. 一方面, 快速的城市道路建设导致部分道路新开通, 而交通意外、交通管制与城市道路维护等事件引致部分路段不能通行达数小时、数天乃至数月; 另一方面, 由于经济建设需要或城市交通安全考虑, 部分道路调整货车禁止通行时段或限行规定. 及时检测

相对于货车通行的路网拓扑结构变化实现物流地图更新,对于保障运输任务的高效实施具有重要意义。大宗货物运输过程中持续产生时序位置信息,可以利用这些时空数据流有效地监测路网变化进而更新物流地图。货车轨迹相对于物流地图的道路拓扑结构变化检测包括带约束的新通行路段的检测、禁行路段的发现以及转向变化引致的路段间新邻接关系提取等。

5.1 多约束的新通行路段检测

基于货车持续产生的轨迹数据可发现物流地图上新增的可通行道路。先采用地图匹配方法获取相对于物流地图未匹配的轨迹数据,通过对其进行基于时间窗口连续观测的路网推断来发现新通行路段。采用基于隐马尔科夫的地图匹配方法将持续产生的轨迹数据与基准物流地图进行匹配,并将未匹配的轨迹分别进行基于轨迹线段与轨迹点的去噪处理。考虑噪声轨迹线段在方向与速度上通常具有以下特征:轨迹线段与其邻接的轨迹点在方向属性存在不一致(表现为转向行为)、轨迹线段的平均速度超过货车行驶的最大限速(一般由定位设备的故障引致)。而对于噪声轨迹点,考虑采用基于距离与方向的双重去噪方法,先根据预设的距离阈值筛除轨迹近邻数较低的轨迹点;再通过计算各轨迹点附近数据的方向分布,将与其近邻点方向分布不一致的轨迹点作为噪声予以删除。

在地图匹配与轨迹去噪后,对于未匹配的轨迹数据采用聚类方法推断新通行路段的候选集。鉴于路网中轨迹数据的倾斜分布特性,在轨迹稀疏分布区域,采用基于轨迹线段的方法容易推断错误的转向与路段,采用基于轨迹点的方法容易推断得到不连续甚至错误路段。因此,为提高路段推断精度,结合基于轨迹线段聚类的方法和基于轨迹点聚类的方法,分别获取基于轨迹点和基于轨迹线段的聚类。轨迹线段之间的相似性采用 Hausdorff 距离评测线段间的最远距离,其值越小,表示两条线段越相似。

轨迹点之间的相似性采用欧氏距离度量。为避免噪声数据导致路段误判,结合候选集提取中基于轨迹线段方法的高覆盖率以及基于轨迹点方法的高准确率优势,通过合并两种方法提取的候选集提升新通行路段的检测精度。具体地,如果两个方法提取的候选集出现在同一区域,且候选集数达到预设的阈值时,认为检测区域存在一条新通行路段,将候选集进行合并。随后,根据轨迹簇的形状分别采用 Sweeping Line 方法与基于 force attraction 的方法进行中心线拟合。

为进一步验证检测到的新通行路段是否允许车辆通行,采用基于时间窗口模型的连续观测方法。当检测路段在每天固定时间范围内允许货车通行,且通行时限达到预设时限时,认定其为对应时段内可允许通行的新路段;同时,根据新路段通行货车的车型属性(货高、载重等)确定该路段的限行约束,最后将具有多约束的新通行路段更新至物流地图中。更新时,可根据新通行路段的方向信息,分别从其道路端点按照平均方向往外延伸一定长度,使其与基准物流地图中近邻路段相交来实现地图更新。

5.2 带时间约束的禁行路段发现

由于交通意外、交通管制、重大活动、城市道路建设等导致城市部分路段不能通行达数小时、数天甚至数月;城市交通管理部门因道路安全或其他考虑对部分路段新增特定时段的货车禁止通行规定,引致货车在部分时段不能通过原来允许通行的路段。分析轨迹数据可以发现,被禁行路段相对于禁行前同路段以及禁行后邻近路段,货车轨迹数与转向行为明显减少,这反映了其通行流量的显著降低特性。通过监测物流地图各路段的轨迹流量变化以及与邻接路段的转向概率变化,可实现带时间约束的禁行路段的识别。

假设各时段内监测路段的历史流量呈数学分布(如高斯分布,泊松分布),可分时段(采用时间窗口模型)对监测路段的历史流量进行统计建模,获得各时段内监测路段的预计通行流量。根据持续产生的货车轨迹数据,将各时段内监测路段的轨迹数与历史同时间段内预计流量进行比较,当实际通过的轨迹流量明显低于预计值时,将其标记为该时段内的禁行路段。考虑到恶劣天气、举办重大活动以及节假日等因素可能导致监测路段发生较大的流量变化,采用深度学习技术结合外部因素预测各路段在不同时段内的通行流量。

基于历史轨迹数据提取包括监测路段的流量特征(即其不同时段内的历史流量)、监测路段的直接可达路段流量特征(即监测路段的近邻路段在不同时段内的历史流量)、监测路段所在区间特征(如通行流量下降区间影响路段的总长度、各时段内的通行流量下降率、观测时间范围内流量下降时段占比等)、天气与节假日等外部影响

因素. 鉴于所提取特征中包含时序特征与上下文关联特征, 采用长短期记忆网络 (LSTM) 和卷积神经网络 (CNN) 的混合模型对路段流量进行预测.

根据流量显著变化特性发现的禁行路段, 结合速度与转向概率变化统计进一步予以验证. 基于各时段内禁行路段对应区域的货车实时轨迹数据, 在进行地图匹配后利用轨迹数据计算获得禁行路段的平均通行速度, 以速度远低于同时段历史平均值以及上下文路段的平均速度为特征验证其为禁行路段; 同时, 根据实时轨迹统计禁行路段与同时段历史上下文路段的转向概率, 如转向概率值远低于历史路段间转向概率, 从而验证其为对应时段内的禁行路段.

5.3 基于转向变化识别的路段新邻接关系提取

在不同时段内, 随着通行路段以及禁行路段的新增, 其邻接路段亦会受到影响, 表现为与变化路段相连接的交叉路口范围内轨迹转向行为发生变化, 因此, 需要根据持续产生的轨迹数据识别车辆转向行为的变化, 确保物流地图的及时更新. 可先根据检测到的新通行路段以及禁行路段位置, 确定与其邻接路段相交的道路交叉口区域. 考虑到交叉路口附近物理结构复杂, 货车通行轨迹的转向行为具有多样性, 为区分不同转向模式, 采用基于 Fréchet 距离的 DBSCAN 聚类方法对轨迹进行聚类以提取不同转向行为模式.

具体来说, 对于每个监测路口区域, 先提取路口范围内的轨迹, 通过随机采样、轨迹压缩预处理数据减少后续处理的计算开销. 随后, 利用基于 Fréchet 距离的聚类方法进行轨迹聚类生成转向模式簇. 为实现相对于物流地图的轨迹差分进而发现道路交叉口的转向行为变化, 对各个转向轨迹簇进行拟合生成转向簇代表轨迹. 首先对各转向簇随机采样 M ($M \geq 3$) 个轨迹, 通过计算采样轨迹之间的 Fréchet 距离, 识别与簇内其他轨迹相似度值最小的轨迹, 再采用 B-Spline 方法对该条轨迹进行平滑处理. B-Spline 方法能通过多项式使用差值法对轨迹线段的曲线特征进行还原.

为检测物流地图缺失的或不正确的转向行为模式, 将提取的各路口不同转向簇代表轨迹与物流地图进行匹配, 通过发现未匹配的转向代表轨迹, 识别路口转向行为变化区域. 采用基于隐马尔科夫的地图匹配方法, 并结合凸包技术加快匹配过程. 将转向簇代表轨迹中的每个轨迹点视为观测变量, 将交叉路口范围内相关路网视为隐藏变量, 对于转向簇代表轨迹中的每个点设定最大窗口半径阈值, 确定与最大窗口相交的路网, 通过路网与转向代表轨迹中点的欧式距离计算获得观测概率, 并根据路段间的距离计算出状态转移概率; 利用 Viterbi 算法计算最大概率路径, 根据其判断转向代表轨迹是否与物流地图匹配. 最后, 根据所识别的路口范围内转向路径映射的路段序列, 更新路段间的邻接关系.

6 物流地图构建示例及应用验证

为了评估本文所提方法构建物流地图的质量, 我们使用来自第三方钢铁物流企业在山东省内的真实货车轨迹数据、货物的运单数据以及车辆信息等数据集评测所提地图构建方法的有效性. 对运输停留热点识别、差异性经验路径提取及道路中心线拟合等核心步骤的实验结果进行了可视化展示, 重点关注道路拓扑结构识别的准确性. 所有实验代码采用 Python 2.6 实现, 在运行 Spark 2.2.0 的 10 节点集群上进行实验, 每个节点配备了 8 核 Intel E5335 2.0 GHz 处理器和 16 GB 内存. 实验中的所有参数取值都是在真实数据集上进行多次实验得到. 关于数据集的描述如下.

车辆信息数据集, 包含钢铁物流平台 15 624 辆重型卡车的信息, 每条记录由车牌号、车量规格 (长×宽×高)、车辆最大载重等字段组成;

运单数据集, 包含 2020 年 11 月—2021 年 3 月产生的 610 059 个运单信息, 每条记录由车牌号、货物品名、货物重量、收货公司、货物装载时间、返单时间等字段组成;

货车轨迹数据集, 包含 15 624 辆货车于 2020 年 11 月—2021 年 3 月产生的轨迹采样点信息, 平均采样间隔为 30 s, 数据集大小为 28.21G. 每条记录由时间戳、车辆 ID、经度、纬度、瞬时速度、方向等字段组成. 我们从 OSM 地图选取了山东省内约 1 630 544 个顶点、652 603 条边组成的路网作为中心线拟合性能评测以及地图匹配的基准路网.

6.1 运输停留热点挖掘

停留热点挖掘涉及停留热点的识别与标注. 根据预设的距离阈值 (设为 50 m), 从轨迹数据中提取了低速轨迹点集. 为避免将交通拥堵和路口等待红绿灯等行为产生的停留点误判为停留热点, 保留停留时长大于 480 s 的低速轨迹点序列. 图 4 展示了运输停留热点的提取过程, 先从运输轨迹中提取低速轨迹点集 (如图 4(a) 所示的红色圆点), 采用 DBSCAN 方法对其进行聚类, 产生如图 4(b) 所示的停留点簇, 共提取了 9 个停留热点 (如图 4(c) 所示). 随后, 我们根据停留热点的功能类型为其进行标注. 为验证停留热点标注的有效性, 停留热点由多位货车司机人工标识, 包括 528 个运输终点, 312 个采购装载点, 128 个加油站, 56 个维修点, 62 个休息区. 为解决类样本极度不平衡对停留热点标注的影响, 采用基于支持向量机、随机森林为基分类器的集成学习框架. 结合停留时长分布、停留时段分布等停留行为统计特征以及邻近道路类型等停留区域特征对运输热点样本进行特征工程, 将其表征为 121 维的特征向量; 再以各类型运输热点样本集的 70% 作为分类标注模型的训练集, 30% 作为测试集. 运输停留热点部分标注结果如图 5 所示, 以红色圆点表示钢铁成品运输过程产生的停留点、绿色圆点表示钢铁原材料运输过程产生的停留点, 在轨迹对应区域的卫星图中用红色椭圆形框表示停留区域. 图 5 展示了钢铁运输的 4 类典型停留热点, 包括对图 4(c) 所提取停留热点标注的运输终点 (如图 5(a) 所示)、货车司机的临时停留区 (如图 5(b) 所示)、钢铁原材料采购装载点 (如图 5(c) 所示) 与车辆维修点 (如图 5(d) 所示).



图 4 停留热点识别示例

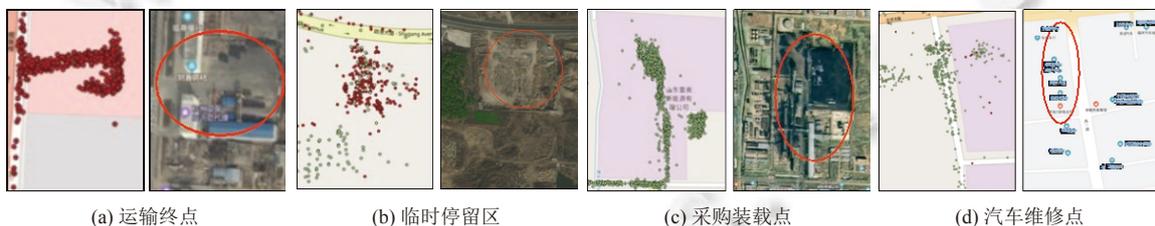


图 5 停留热点标注示例

在运输停留热点标注过程中, 恰当的停留时长阈值不仅能筛除货车因交通拥堵、等待红绿灯等短暂停留产生的无效停留点, 还能在一定程度上保证停留热点类型的有效标注. 我们进一步分析了停留时长阈值对典型运输停留热点的类型标注有效性 (以停留热点标注的准确率与召回率的调和平均数 $F1$ 值表示) 的影响, 如图 6(a) 所示. 可以看出, 随着停留时长阈值的增加, 有效减少了噪声停留点的影响, 当停留时长设为 8–10 min 时停留热点标注具有较高的 $F1$ 值, 能准确识别运输停留热点类型. 此后, 随着停留时长阈值的继续增加, 较大的停留时长阈值误将有意义的停留点删除导致停留热点标注有效性降低, 以加油站类停留热点的标注表现最为明显. 因此, 在实际应用中为准确标注停留热点类型, 停留时长阈值设置为 8–10 min 为宜.

6.2 差异性经验路径挖掘

通过轨迹聚类与 K 频繁路径挖掘, 我们提取了不同时段内的 K ($K=3$) 条差异性频繁路径 (路径频繁度由高到

低分别以红色、橙色与绿色线条表示),以日照至连云港为例对差异性经验路径挖掘结果进行了可视化展示,如图 7 所示.我们以货车司机最常选择的 3 个运输时段 (03:00–06:00、11:00–14:00 与 18:00–21:00) 为观测时段,对日照至连云港的钢铁运输轨迹数据进行挖掘,实验结果分别如图 7(a)–图 7(c) 所示.由图中可以看出,不同时段内货车行驶的运输线路存在较大差异.具体地,在夜间运输时,由于城市区域道路限制较少,货车司机常选择城市道路行驶,绕行道路较少,如图 7(a) 所示;而在 11:00–14:00 时段内运输货物车辆因城市道路设有货车通行限制,为保证货物的及时送达,货车司机倾向于选择高速公路行驶,如图 7(b) 所示;在 18:00–21:00 时段内运输货物的车辆,由于正值城市道路的通行晚高峰时段,出于运输时效与物流成本的考虑,货车司机常选择由高速公路行驶至赣榆港收费站再改经省道行驶的运输线路,如图 7(c) 所示.

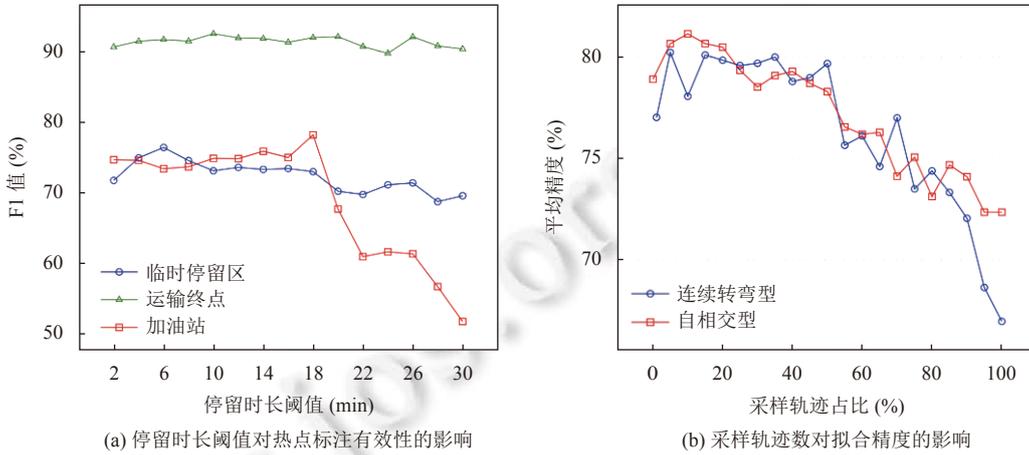


图 6 阈值分析

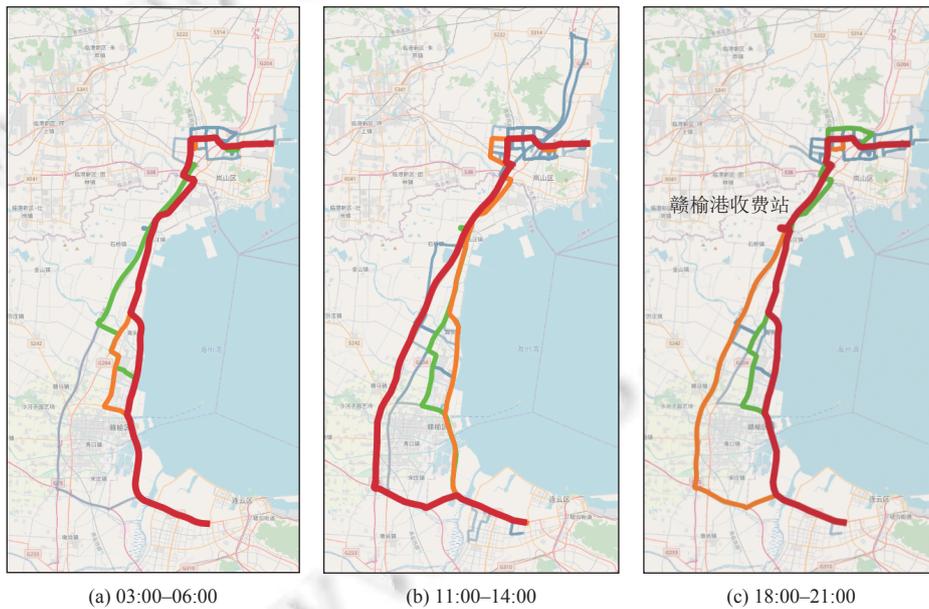


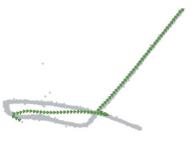
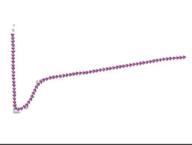
图 7 差异性经验路径挖掘示例

6.3 道路拓扑结构生成

对于各个时段内挖掘得到的差异性频繁路径(即频繁轨迹簇),我们采用 Sweeping Line 方法与基于 force attraction 的三阶段拟合方法对其进行中心线拟合,获取道路拓扑结构.为验证所采用的中心线拟合策略的有效性,

我们选择了物流路网中 4 种道路形态的轨迹簇, 将 Sweeping Line 方法与基于 force attraction 的三阶段拟合方法进行了对比, 可视化的实验结果如表 1 所示 (轨迹点数据以灰色点表示, 两种方法提取的中心线分别以绿色线条和紫色线条表示)。可以看出, 两种方法在 (近) 直线型道路上的拟合效果较好, 但在连续转弯型、U 型弯道型与自相交型 3 种道路形态上, 基于 force attraction 的拟合方法表现更好。

表 1 不同形状道路的中心线拟合

方法	(近)直线型	连续转弯型	U型弯道型	自相交型
轨迹簇				
Sweeping Line拟合方法				
基于force attraction的拟合方法				

随后, 我们进一步分析了基于 force attraction 的拟合方法中参数 k 取值对中心线拟合精度的影响。在该方法中, 随机采样的 k 条轨迹将作为每组转向簇的候选参考轨迹, 通过计算每条候选轨迹与簇内其他轨迹之间的 Fréchet 距离之和, 选取距离和最小的轨迹作为参考轨迹。实验中选择了两类复杂的转向轨迹簇 (连续转弯型与自相交型) 进行了 50 次采样, 由于各类转向簇的轨迹数差异较大 (分别是 96 条与 2031 条), 实验通过不断调整各转向簇内采样轨迹数所占比例验证了采样参数 k 对精度的影响, 最后获得 50 次实验的平均精度如图 6(b) 所示。可以看出, 随着各转向簇内采样轨迹数占比持续增加, 中心线拟合精度不断提升, 但当超过 25% 时, 两类转向道路拟合精度呈下降趋势。因此, 实际应用中, 采样轨迹数设为所在簇轨迹数的 15%~25% 之间为宜。

最后, 对于推断得到的每条道路, 根据历史数据分别统计货车可通行时段、货物最大载重以及货车车辆信息, 获得各条道路的通行时间限制、最大载重限制以及限高、限宽等信息, 生成如图 8 所示的多约束的物流地图。其中, 用于钢铁货运车辆的可通行路段以蓝色线条表示, 对于典型的运输停留热点, 运输终点 (以黄色菱形表示)、临时休息区 (以绿色圆形表示)、维修点 (以橘色圆形表示)、加油站 (以深绿色三角形表示) 与采购装载点 (以蓝色圆形表示) 等均在图 8 中进行了标注。

6.4 案例-基于物流地图的最优运输线路规划

为了评估所构建物流地图的实用性, 基于物流地图与商业地图软件分别对货车运输线路进行了路径规划。以山东省日照市华东国际物流城为货物装载地点、山东省淄博市中埠镇为运输终点, 采用商业地图软件与所构建物流地图进行线路规划, 结果如图 9 所示。基于商业地图软件的最短运输线路规划路径总长 266 km, 涉及高速路段总长 258 km, 由于全程为高速路段, 运输时长仅需 3 小时, 但需高速过路费用约 117 元 (如图 9(a) 所示); 选择商业地图软件进行了“避开高速路段”的线路规划, 推荐路线如图 9(b) 所示, 路径总长 275 km, 由于全程为非高速路段, 时长 5 小时, 途经五莲县部分城市道路区域, 无高速过路费; 基于我们的物流地图进行最优路径规划, 结果如图 9(c) 所示, 路径总长 280 km, 时长 5 小时, 途经 2 个加油站, 2 个临时休息区, 未产生高速过路费。相较于商业地图软件避开高速的线路规划结果, 基于物流地图规划的运输线路不仅避开了城市道路和高速, 物流成本较低, 同时, 货车司机在运输途中得到了充分的休息, 更符合大宗货运特点。

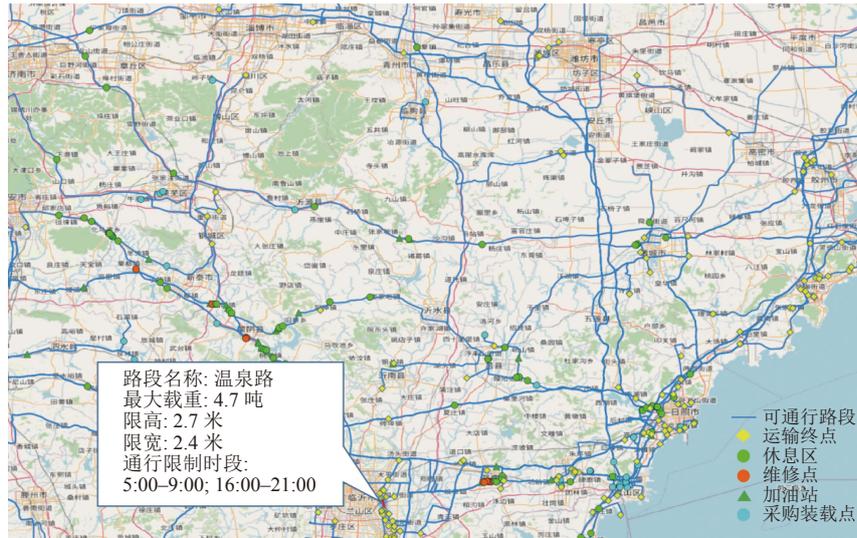


图 8 物流地图示例

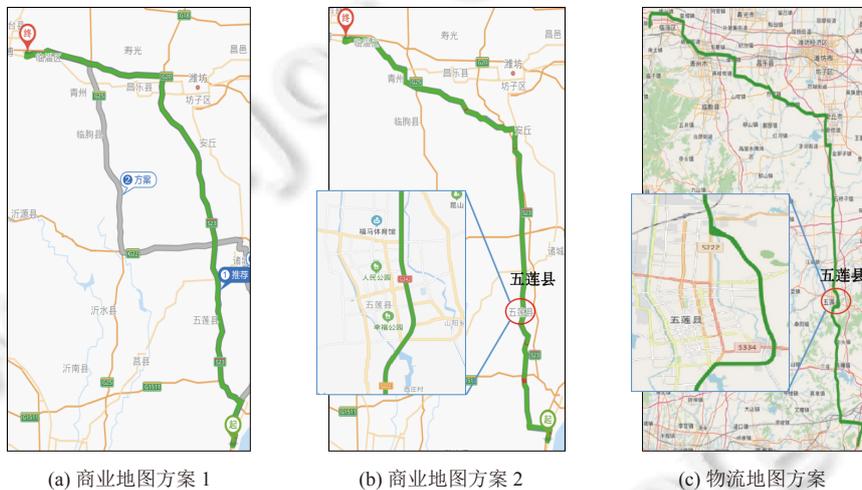


图 9 运输线路规划案例

7 结论

公路大宗货运的物流地图构建是典型的需求导向的问题研究, 具有重要的理论研究价值和明确的应用背景。物流大数据内涵丰富, 以动态变化的时空数据为核心, 包含静态交通数据与物流运营数据。为解决物流路网拓扑结构的精准提取与道路拓扑结构变化的高效识别问题, 本文采用机器学习方法从多源物流数据的挖掘着手, 探索公路大宗货运的运输线路规划机理, 研究基于用户先验知识的多约束物流地图构建机制。随后, 利用持续新增的时空数据, 研究道路拓扑结构的演化规律, 实现物流地图的增量更新。本文的研究成果为大宗货运的运输线路规划等物流场景提供切实可行的系统解决方案, 助力网络货运等物流新模式的创新发展。

References:

[1] Su H, Liu SC, Zheng BL, Zhou XF, Zheng K. A survey of trajectory distance measures and performance evaluation. The VLDB Journal, 2020, 29(1): 3-32. [doi: 10.1007/s00778-019-00574-9]

- [2] Lee JG, Han JW, Whang KY. Trajectory clustering: A partition-and-group framework. In: Proc. of the 2007 ACM SIGMOD Int'l Conf. on Management of Data. Beijing: Association for Computing Machinery, 2007. 593–604. [doi: [10.1145/1247480.1247546](https://doi.org/10.1145/1247480.1247546)]
- [3] Li HF, Kulik L, Ramamohanarao K. Automatic generation and validation of road maps from GPS trajectory data sets. In: Proc. of the 25th ACM Int'l on Conf. on Information and Knowledge Management. Indianapolis: Association for Computing Machinery, 2016. 1523–1532. [doi: [10.1145/2983323.2983797](https://doi.org/10.1145/2983323.2983797)]
- [4] Wang T, Mao JL, Jin CQ. HyMU: A hybrid map updating framework. In: Proc. of the 22nd Int'l Conf. on Database Systems for Advanced Applications. Suzhou: Springer, 2017. 19–33. [doi: [10.1007/978-3-319-55699-4_2](https://doi.org/10.1007/978-3-319-55699-4_2)]
- [5] Karagiorgou S, Pfoser D. On vehicle tracking data-based road network generation. In: Proc. of the 20th Int'l Conf. on Advances in Geographic Information Systems. Redondo Beach: Association for Computing Machinery, 2012. 89–98. [doi: [10.1145/2424321.2424334](https://doi.org/10.1145/2424321.2424334)]
- [6] Fränti P, Mariosescu-Istodor R. Averaging GPS segments competition 2019. Pattern Recognition, 2021, 112: 107730. [doi: [10.1016/j.patcog.2020.107730](https://doi.org/10.1016/j.patcog.2020.107730)]
- [7] Qiu J, Wang RS. Inferring road maps from sparsely sampled GPS traces. Journal of Location Based Services, 2016, 10(2): 111–124. [doi: [10.1080/17489725.2016.1183053](https://doi.org/10.1080/17489725.2016.1183053)]
- [8] Schröedl S, Wagstaff K, Rogers S, Langley P, Wilson C. Mining GPS traces for map refinement. Data Mining and Knowledge Discovery, 2004, 9(1): 59–87. [doi: [10.1023/B:DAMI.0000026904.74892.89](https://doi.org/10.1023/B:DAMI.0000026904.74892.89)]
- [9] Kégl B, Krzyzak A, Linder T, Zeger K. Learning and design of principal curves. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2000, 22(3): 281–297. [doi: [10.1109/34.841759](https://doi.org/10.1109/34.841759)]
- [10] Wang Y, Liu XM, Wei H, Forman G, Chen C, Zhu YM. CrowdAtlas: Self-updating maps for cloud and personal use. In: Proc. of the 11th Annual Int'l Conf. on Mobile Systems, Applications, and Services. Taipei: Association for Computing Machinery, 2013. 27–40. [doi: [10.1145/2462456.2464441](https://doi.org/10.1145/2462456.2464441)]
- [11] Zhao LS, Mao JL, Pu M, Liu GP, Jin CQ, Qian WN, Zhou AY, Wen X, Hu RB, Chai H. Automatic calibration of road intersection topology using trajectories. In: Proc. of the 36th Int'l Conf. on Data Engineering (ICDE). Dallas: IEEE, 2020. 1633–1644. [doi: [10.1109/ICDE48307.2020.00145](https://doi.org/10.1109/ICDE48307.2020.00145)]
- [12] Mariosescu-Istodor R, Fränti P. CellNet: Inferring road networks from GPS trajectories. ACM Trans. on Spatial Algorithms and Systems, 2018, 4(3): 8. [doi: [10.1145/3234692](https://doi.org/10.1145/3234692)]
- [13] Cao LL, Krumm J. From GPS traces to a routable road map. In: Proc. of the 17th ACM SIGSPATIAL Int'l Conf. on Advances in Geographic Information Systems. Seattle: Association for Computing Machinery, 2009. 3–12. [doi: [10.1145/1653771.1653776](https://doi.org/10.1145/1653771.1653776)]
- [14] Chen D, Guibas LJ, Hershberger J, Sun J. Road network reconstruction for organizing paths. In: Proc. of the 21st Annual ACM-SIAM Symp. on Discrete Algorithms. Austin: Society for Industrial and Applied Mathematics, 2010. 1309–1320. [doi: [10.1137/1.9781611973075.105](https://doi.org/10.1137/1.9781611973075.105)]
- [15] Biagioni J, Eriksson J. Map inference in the face of noise and disparity. In: Proc. of the 20th Int'l Conf. on Advances in Geographic Information Systems. Redondo Beach: Association for Computing Machinery, 2012. 79–88. [doi: [10.1145/2424321.2424333](https://doi.org/10.1145/2424321.2424333)]
- [16] Newson P, Krumm J. Hidden Markov map matching through noise and sparseness. In: Proc. of the 17th ACM SIGSPATIAL Int'l Conf. on Advances in Geographic Information Systems. Seattle: Association for Computing Machinery, 2009. 336–343. [doi: [10.1145/1653771.1653818](https://doi.org/10.1145/1653771.1653818)]
- [17] Hu JX, Razdan A, Femiani JC, Cui M, Wonka P. Road network extraction and intersection detection from aerial images by tracking road footprints. IEEE Trans. on Geoscience and Remote Sensing, 2007, 45(12): 4144–4157. [doi: [10.1109/TGRS.2007.906107](https://doi.org/10.1109/TGRS.2007.906107)]
- [18] Cheng GL, Wang Y, Xu SB, Wang HZ, Xiang SM, Pan CH. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. IEEE Trans. on Geoscience and Remote Sensing, 2017, 55(6): 3322–3337. [doi: [10.1109/TGRS.2017.2669341](https://doi.org/10.1109/TGRS.2017.2669341)]
- [19] Mokhtarzade M, Zoej MJV. Road detection from high-resolution satellite images using artificial neural networks. Int'l Journal of Applied Earth Observation and Geoinformation, 2007, 9(1): 32–40. [doi: [10.1016/j.jag.2006.05.001](https://doi.org/10.1016/j.jag.2006.05.001)]
- [20] Seo YW, Urmson C, Wettergreen D. Exploiting publicly available cartographic resources for aerial image analysis. In: Proc. of the 20th Int'l Conf. on Advances in Geographic Information Systems. Redondo Beach: Association for Computing Machinery, 2012. 109–118. [doi: [10.1145/2424321.2424336](https://doi.org/10.1145/2424321.2424336)]
- [21] Bastani F, He ST, Abbar S, Alizadeh M, Balakrishnan H, Chawla S, Madden S, DeWitt DJ. RoadTracer: Automatic extraction of road networks from aerial images. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 4720–4728. [doi: [10.1109/CVPR.2018.00496](https://doi.org/10.1109/CVPR.2018.00496)]
- [22] Biagioni J, Eriksson J. Inferring road maps from global positioning system traces: Survey and comparative evaluation. Transportation Research Record: Journal of the Transportation Research Board, 2012, 2291(1): 61–71. [doi: [10.3141/2291-08](https://doi.org/10.3141/2291-08)]

- [23] Ahmed M, Karagiorgou S, Pfoser D, Wenk C. A comparison and evaluation of map construction algorithms using vehicle tracking data. *GeoInformatica*, 2015, 19(3): 601–632. [doi: [10.1007/s10707-014-0222-6](https://doi.org/10.1007/s10707-014-0222-6)]
- [24] Ahmed M, Karagiorgou S, Pfoser D, Wenk C. *Map Construction Algorithms*. Cham: Springer, 2015. 1–120.
- [25] Hashemi M. A testbed for evaluating network construction algorithms from GPS traces. *Computers, Environment and Urban Systems*, 2017, 66: 96–109. [doi: [10.1016/j.compenurbsys.2017.08.003](https://doi.org/10.1016/j.compenurbsys.2017.08.003)]
- [26] Duran D, Sacristán V, Silveira RI. Map construction algorithms: A local evaluation through hiking data. *GeoInformatica*, 2020, 24(3): 633–681. [doi: [10.1007/s10707-019-00386-7](https://doi.org/10.1007/s10707-019-00386-7)]
- [27] Edelkamp S, Schrödl S. Route planning and map inference with global positioning traces. In: Klein R, Six HW, Wegner L, eds. *Computer Science in Perspective*. Berlin: Springer, 2003. 128–151. [doi: [10.1007/3-540-36477-3_10](https://doi.org/10.1007/3-540-36477-3_10)]
- [28] Stanojevic R, Abbar S, Thirumuruganathan S, Chawla S, Filali F, Aleimat A. Robust road map inference through network alignment of trajectories. In: *Proc. of the 2018 SIAM Int'l Conf. on Data Mining*. San Diego: Society for Industrial and Applied Mathematics, 2018. 135–143. [doi: [10.1137/1.9781611975321.15](https://doi.org/10.1137/1.9781611975321.15)]
- [29] Chen C, Lu CW, Huang QX, Yang Q, Gunopulos D, Guibas L. City-scale map creation and updating using GPS collections. In: *Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. San Francisco: Association for Computing Machinery, 2016. 1465–1474. [doi: [10.1145/2939672.2939833](https://doi.org/10.1145/2939672.2939833)]
- [30] Worrall S, Nebot E. Automated process for generating digitised maps through GPS data compression. In: *Proc. of the Australasian Conf. on Robotics and Automation*. Brisbane: ACRA, 2007.
- [31] Jang S, Kim T, Lee E. Map generation system with lightweight GPS trace data. In: *Proc. of the 12th Int'l Conf. on Advanced Communication Technology (ICACT)*. Gangwon: IEEE, 2010. 1489–1493.
- [32] Liu XM, Biagioni J, Eriksson J, Wang Y, Forman G, Zhu YM. Mining large-scale, sparse GPS traces for map inference: Comparison of approaches. In: *Proc. of the 18th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. Beijing: Association for Computing Machinery, 2012. 669–677. [doi: [10.1145/2339530.2339637](https://doi.org/10.1145/2339530.2339637)]
- [33] Wu H, Tu CC, Sun WW, Zheng BH, Su H, Wang W. GLUE: A parameter-tuning-free map updating system. In: *Proc. of the 24th ACM Int'l Conf. on Information and Knowledge Management*. Melbourne: Association for Computing Machinery, 2015. 683–692. [doi: [10.1145/2806416.2806425](https://doi.org/10.1145/2806416.2806425)]
- [34] Shan ZQ, Wu H, Sun WW, Zheng BH. COBWEB: A robust map update system using GPS trajectories. In: *Proc. of the 2015 ACM Int'l Joint Conf. on Pervasive and Ubiquitous Computing*. Osaka: Association for Computing Machinery, 2015. 927–937. [doi: [10.1145/2750858.2804286](https://doi.org/10.1145/2750858.2804286)]
- [35] Davies JJ, Beresford AR, Hopper A. Scalable, distributed, real-time map generation. *IEEE Pervasive Computing*, 2006, 5(4): 47–54. [doi: [10.1109/MPRV.2006.83](https://doi.org/10.1109/MPRV.2006.83)]
- [36] Uduwaragoda ERIACM, Perera AS, Dias SAD. Generating lane level road data from vehicle trajectories using kernel density estimation. In: *Proc. of the 16th IEEE Int'l Conf. on Intelligent Transportation Systems (ITSC 2013)*. The Hague: IEEE, 2013. 384–391. [doi: [10.1109/ITSC.2013.6728262](https://doi.org/10.1109/ITSC.2013.6728262)]
- [37] Dey TK, Wang JY, Wang YS. Improved road network reconstruction using discrete Morse theory. In: *Proc. of the 25th ACM SIGSPATIAL Int'l Conf. on Advances in Geographic Information Systems*. Redondo Beach: Association for Computing Machinery, 2017. 1–4. [doi: [10.1145/3139958.3140031](https://doi.org/10.1145/3139958.3140031)]
- [38] Ahmed M, Wenk C. Constructing street networks from GPS trajectories. In: Epstein L, Ferragina P, eds. *European Symposium on Algorithms*. Berlin: Springer, 2012. 60–71. [doi: [10.1007/978-3-642-33090-2_7](https://doi.org/10.1007/978-3-642-33090-2_7)]
- [39] Zhang LJ, Thiemann F, Sester M. Integration of GPS traces with road map. In: *Proc. of the 3rd Int'l Workshop on Computational Transportation Science*. San Jose: Association for Computing Machinery, 2010. 17–22. [doi: [10.1145/1899441.1899447](https://doi.org/10.1145/1899441.1899447)]
- [40] Xie XZ, Philips W. Road intersection detection through finding common sub-tracks between pairwise GNSS traces. *ISPRS Int'l Journal of Geo-Information*, 2017, 6(10): 311. [doi: [10.3390/ijgi6100311](https://doi.org/10.3390/ijgi6100311)]
- [41] Li L, Li DG, Xing XY, Yang F, Rong W, Zhu HH. Extraction of road intersections from gps traces based on the dominant orientations of roads. *ISPRS Int'l Journal of Geo-Information*, 2017, 6(12): 403. [doi: [10.3390/ijgi6120403](https://doi.org/10.3390/ijgi6120403)]
- [42] Fathi A, Krumm J. Detecting road intersections from GPS traces. In: Fabrikant SI, Reichenbacher T, Van Kreveld M, Schlieder C, eds. *Geographic Information Science*. Berlin: Springer, 2010. 56–69. [doi: [10.1007/978-3-642-15300-6_5](https://doi.org/10.1007/978-3-642-15300-6_5)]
- [43] Wu JW, Zhu YL, Ku T, Wang L. Detecting road intersections from coarse-gained GPS traces based on clustering. *Journal of Computers*, 2013, 8(11): 2959–2965. [doi: [10.4304/jcp.8.11.2959-2965](https://doi.org/10.4304/jcp.8.11.2959-2965)]
- [44] Huang YR, Xiao Z, Yu XY, Wang D, Havyarimana V, Bai J. Road network construction with complex intersections based on sparsely sampled private car trajectory data. *ACM Trans. on Knowledge Discovery from Data*, 2019, 13(3): 35. [doi: [10.1145/3326060](https://doi.org/10.1145/3326060)]

- [45] Pu M, Mao JL, Du YT, Shen YB, Jin CQ. Road intersection detection based on direction ratio statistics analysis. In: Proc. of the 20th IEEE Int'l Conf. on Mobile Data Management (MDM). Hong Kong: IEEE, 2019. 288–297. [doi: [10.1109/MDM.2019.00-46](https://doi.org/10.1109/MDM.2019.00-46)]
- [46] Wang J, Rui XP, Song XF, Tan XS, Wang CL, Raghavan V. A novel approach for generating routable road maps from vehicle GPS traces. Int'l Journal of Geographical Information Science, 2015, 29(1): 69–91. [doi: [10.1080/13658816.2014.944527](https://doi.org/10.1080/13658816.2014.944527)]
- [47] Wang J, Wang CL, Song XF, Raghavan V. Automatic intersection and traffic rule detection by mining motor-vehicle GPS trajectories. Computers, Environment and Urban Systems, 2017, 64: 19–29. [doi: [10.1016/j.compenvurbsys.2016.12.006](https://doi.org/10.1016/j.compenvurbsys.2016.12.006)]



毛嘉莉(1979—), 女, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为时空数据管理, 数据驱动的物流科技。



周傲英(1965—), 男, 博士, 教授, 博士生导师, CCF 会士, 主要研究领域为数据库, 数据管理, 数字化转型, 金融科技, 教育科技和物流科技等数据驱动的应用。



吴涛(1992—), 男, 博士生, 主要研究领域为时空数据管理, 数据驱动的物流科技。



金澈清(1977—), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为区块链, 数据质量, 不确定数据管理, 基于位置的服务。



李思佳(1997—), 女, 博士生, 主要研究领域为时空数据管理, 数据驱动的物流科技。



钱卫宁(1976—), 男, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为可扩展事务处理, 大数据管理系统基准评测, 海量数据分析处理及其应用, 数据驱动的计算教育学。



郭焯(1979—), 男, 博士生, CCF 专业会员, 主要研究领域为物流科技, 数据挖掘。