

智慧城市中的交通大数据分析 with 拥堵预测

樊祥至, 高 敏

(齐齐哈尔工程学院, 黑龙江 齐齐哈尔 161003)

摘 要 随着城市交通越来越拥挤, 对交通大数据进行实时采集、分析以及预测已经成为缓解交通堵塞的重要措施之一。通过对车辆 GPS、路侧传感器、移动设备轨迹等多源数据进行融合, 建立交通流时空特征模型, 运用机器学习与深度学习算法来进行短期、中期交通拥堵预警, 从而对交通流态进行实时掌控。根据预测结果对信号灯配时、出行路线选择以及交通疏导进行调整, 可以大大提升道路通行能力, 减少等待时间, 节约能源以及排放量, 为智慧城市的交通管理提供智能决策参考。

关键词 智慧城市; 交通大数据; 拥堵预测; 时空特征; 模型优化

中图分类号: U495; TP391

文献标志码: A

DOI: 10.3969/j.issn.2097-3365.2026.25.010

0 引言

随着工业化进程不断加快, 城市机动车数量持续增长, 交通堵塞已经成为制约城市发展的瓶颈性问题, 多源交通信息的实时获取以及智能化分析为我们探索交通流量变化规律带来机遇。应用机器学习、深度学习等技术对交通流量构建模型以及进行预测, 从而做到实时观察路况以及发出警报, 在合理使用预测结果的基础上可以调控红绿灯时序、路径计算以及调度方案, 进而提升道路利用率, 节约能源消耗, 降低环境污染, 对智慧城市的交通治理提供基于大数据的智能化决策建议。

1 智慧城市交通大数据概述

1.1 交通大数据的来源

城市智慧交通大数据来源是各类异构型的数据源, 主要有车辆 GPS 和 V2X 数据、路侧设备如摄像头、雷达、地磁线圈等、手机信令、公交地铁刷卡以及公众出行等信息。车辆的 GPS 及 V2X 设备可以得到非常精确的车辆的位置坐标、速度与行驶路径等信息, 可以作为城市道路中车流密度研究的基础数据。路侧设备则有摄像机、雷达以及地磁线圈等可以获取车辆的数量、速度、通行方式等信息, 用于城市道路中行车流量监测与突发事件发现^[1]。

此外, 移动端轨迹数据(手机位置记录、共享单车骑行轨迹)以及公众出行数据(公交刷卡、滴滴出行等打车应用)都可体现行人的行为特征以及出行需求的变化情况。对上述多种来源的数据加以整合就可以构建起全时空覆盖面广、角度多样性的交通大数据集, 对交通分析、拥堵预报、出行优化等问题都有参考作用。

1.2 数据特征与处理挑战

交通大数据具备海量性、时空性和多源异构性的特点, 数量庞大并持续增加, 符合“大数据量、高速产生、实时更新”的特点; 而交通数据除了包含空间位置与时间外, 还有包含速度、加速度、路段特征等多种属性, 解析和处理起来较为繁杂。因此, 为了适应实时交通预测以及动态调度的需求, 对数据处理的要求是需要做到高并发、低延迟、高可靠性的状态。

针对上述难点, 要对采集到的数据先经过预处理, 包括清洗、去噪、融合以及归一化等一系列的操作来去除传感器偏差所带来的误差以及过多的数据。另外, 使用高性能的流式大数据处理引擎(Spark Streaming、Apache Flink 等)来实现实时处理大规模的数据, 并进行分析与计算, 可以给之后的道路车流预测、拥堵判断提供精确的、实时的信息来源^[2]。

1.3 交通大数据分析技术

交通大数据研究借助新型的算法、计算平台深入理解复杂的交通体系。采用时空数据分析技术对道路交通网上的车辆时空分布特征进行研究, 可以找出交通瓶颈点、热点路段以及危险路段, 从而作为交通拥堵预报的一种特征输入项; 模式识别及聚类分析的方法可以从以往交通数据内获取典型的车流量变化状况, 用作模型构建与预测的参考数据。

实时流处理是一种对动态更新的交通数据持续不断地进行监测与分析的技术手段, 在事件触发或者流处理的过程中能够迅速地检测出诸如交通堵塞以及交通事故等突发性情况变化, 再配合时间和空间维度上

作者简介: 樊祥至(2003-), 男, 本科, 研究方向: 大数据分析。

的分析,从而使得交通大数据分析可以实现对当前的详细描绘,同时也给交通管理部门做出合理安排提供了有效的参考依据,使道路交通资源分配更合理并且实现智慧管理。

2 交通拥堵预测模型与算法

2.1 传统统计预测方法

传统的统计预测方法是最早的针对交通拥堵预测的方式,主要是借助经典的统计学方法的时间序列、回归方程以及马尔科夫链等方法。其中,时间序列法就是通过以往交通流的数据的历史进行时序上的相关性和线性的探索来建立预测模型。例如:自回归整合移动平均模型 (ARIMA) 及扩展的形式就可以在较短时间内很好地预测交通流量的变化情况;而基于马尔科夫模型的交通拥堵预测则是运用其状态转换的概率阵型表示交通状态从一个状态到另一个状态变化的过程,在较小的范围中进行预测^[3]。

虽然上述统计方法适用于线性的变化及有显著规律性的交通流量信息,在处理复杂的行车行为、多重模式的车流量以及非线性的特点时就显得力不从心了。为了提高模型的灵活性,本文采用了高阶多元马尔科夫模型并用增加调控变量、多种来源变量集成等手段增强模型对于动态变化交通的反应程度,从而大大提高了模型预测精准度以及稳定性,但是依然基于线性的前提条件。

2.2 机器学习预测模型

大数据体量以及计算性能的增强使得机器学习成为交通拥堵预报的一类主要手段。机器学习的模型可以自动从海量的历史交通信息中挖掘出输入特征到输出流量之间的复杂非线性联系,在多个输入参数的情况下提升了预报准确性。而支持向量机可以构造一个非常高维度的特征空间,并且通过最大间隔分类器来实现对非线性样本的区分,适合用于短期或者中期的流量预测工作。

随机森林是一种集成决策树的方法,在多颗决策树的投票中来减小过拟合问题,增强模型的泛化能力;而梯度提升树 (GBDT) 则是通过对误差的修正来进行预测优化的一种方式,可以提取到各个变量之间的复杂的相互关系。该类机器学习模型适用于高维、多源输入的数据,具有较好的预测效果,但是需要较多的特征工程以及参数调整工作,因此需要借助一定的专业知识来进行特征选择与设计。

2.3 深度学习模型

深度学习方法是目前交通流预测的研究前沿热点,可以实现对时空序列数据进行端到端的特征抽取。长

短时记忆网络 (LSTM) 采用门控策略来克服传统的递归神经网络存在的梯度爆炸或梯度消失现象,可以较好地学习交通量的时间变化过程中所蕴含的长程依赖性;而图卷积网络 (GCN) 则适用于对交通网络的空间结构进行建模,在表示交通网络中的道路连接以及其拓扑信息上具有较好的效果,可以通过将道路节点及其相互连接的关系映射成一种图结构特征的方式,从而方便后续的学习任务。

除此之外,利用注意力机制的 Transformer 模型可以实现并行运算并且可以对全序进行建模,因此能够用于交通流的预测任务当中,进而提高了对于复杂的时空关系的学习能力。深度学习算法适用于异构的大规模数据集,但是训练过程占用大量的处理器资源并且对于标记的数据量有较高的需求以防止模型过拟合等。

2.4 综合模型与混合集成算法

集成模型、混合集成算法在整合不同的预测技术的同时,也提高了对交通拥堵情况的预测效果。集成学习方法如袋装法 (Bagging)、提升法 (Boosting)、堆叠法 (Stacking) 等,可以将一系列独立的基模型的结果进行整合以得到一个更准确可靠的预测结果,降低某一个模型的误差,从而提高了集成的学习效果,比如把时间序列模型的短期趋势捕捉能力以及深度学习模型的非线性表示能力结合起来,既可以包含局部的变化,又可以包含全局的变化^[4]。

混合模型也能够结合使用统计方法以及机器学习或者神经网络模型来建立一个多层次的预测机制,利用不同的时间段,运用最适宜的方法来进行单独的预测之后加以整合。

3 交通拥堵预测系统架构设计

3.1 系统总体架构

交通拥堵预测系统通常由数据层、计算层、模型层和应用层构成,形成完整的智能预测体系。数据层负责多源交通数据的采集与存储,包括车辆 GPS、路侧传感器、手机定位以及出行记录等信息。通过分布式文件系统与消息队列技术 (如 HDFS、Kafka), 实现数据的高效传输与存储,并支持历史数据与实时数据的统一管理。

计算层承担数据预处理与计算任务,对采集数据进行清洗、格式转换及特征提取。采用批处理与流处理结合方式,在 Spark、Flink 等框架下完成数据过滤、缺失值补齐和时间对齐等操作,同时提取关键时空特征,为模型提供高质量输入数据。

模型层集中部署各类预测算法,包括统计模型、机器学习模型及深度学习模型。通过容器化与微服务

架构实现模型的灵活部署与扩展,并结合模型管理机制进行版本控制与自动更新,提升预测性能与稳定性。

应用层用来将预测结果进行可视化展示以及业务规则的应用,将模型输出转换成可以使用的交通管控措施,提供给交管部门仪表盘、报表生成功能模块等,可以跨设备使用。

3.2 实时数据处理与流分析

交通拥堵预报系统依靠实时数据流处理技术,能够实现快速交通监测预警。在实时数据流水线中,基础数据由各种各样的传感器以及车联网装置不断向系统中输送,运算环节运用流式计算框架对数据流实时进行运算。例如:以事件时间窗口来处理某段时间内车辆数量信息等,从而快速了解交通流量变动情况以及突发性交通堵塞警报等。

实时数据分析不仅涉及基本的统计汇总,还有复杂的事件处理(CEP)、模式匹配,可以找到数据流中的特定的事件串。基于实时计算系统,在短短数秒,甚至毫秒以内就可以做出拥堵趋势的预测,进而推送到模型层或者应用层,作出交通管控灯光调控与拥堵告警信息提醒,使整个系统更加灵敏、快速响应^[6]。

3.3 可视化与决策支持

交通拥堵预报系统最核心的目的就是给交通管理部门以及各级领导层提供一目了然的信息支撑依据。可视化层借助于开发交互式的图形界面以及动态地图将交通流、拥堵等级与未来发展的走势以热力图、时间线图表、流量图表等方式呈现出来,使调度人员能迅速了解掌握整个城市的路网运行状况及其发展趋势。

同时,可视化系统还具有对多维度查询、回溯比较、虚拟仿真的功能,方便交管部门根据分析结果实施方案论证、红绿灯调控、路径指引发布等工作内容,通过可视化加决策支持模块智能推荐的方式,在出现潜在拥堵预警情况后可以即时发送报警提示以及调度建议预案,提高城市智能化交通管控水平。

4 智慧城市交通拥堵预测面临的挑战、应用与未来发展

4.1 当前面临的挑战

智慧城市的道路交通拥堵预测实践中有诸多困难:其一,交通数据质量参差不齐以及存在隐私泄露的威胁,来源多样化的数据产生各种各样的误差问题,如传感器误差、空缺值、格式不一等,还存在有关车辆路线以及人员出行的信息需要在开放共享的同时考虑到个人隐私保护的问题;其二,模型的泛化性较差,不同的城市之间存在着巨大的区别,即道路格局的区别、交通流量的区别以及人们的出行习惯的区别等,

所以直接采用单个城市训练好的模型是不太现实的,应该进行调节优化并且调整特征以适应新的城市环境。

4.2 智慧城市交通拥堵预测的典型应用

交通拥堵预测技术的应用前景广阔,在智慧城市的建设中有着广泛的应用场景。智慧信号机控制根据预测结果适时调节绿灯时间及交叉口配时方案来解决局部路段拥堵问题;出行导航及路线规划依据最新的预测流量情况给司机以及公共交通系统提供最佳行驶方案,从而增大道路承载力,减少排队时间及空气污染等负面因素的影响。高峰预警及分流措施可以在突发性大流量之前及时发出警报并且指导车辆进行分流,以避免出现大规模堵塞的现象。

4.3 未来发展趋势

交通拥堵预测将向融合化与智能化方向发展。边缘计算与5G技术的结合提升了数据传输与处理效率,可在低时延条件下实现实时流量预测,并增强模型的可解释性,便于决策应用。多出行方式融合预测逐渐成为重点,在统一框架下综合公交、共享出行及步行等数据并进行分析。随着算法优化、算力提升及通信技术进步,交通预测模型精度与响应能力将持续提高,推动智能交通系统向更高水平发展。

5 结束语

智慧城市交通大数据挖掘与拥堵预报依托多源信息融合、算法建模及系统架构设计,实现道路交通流的实时监测与预测。通过融合统计学、机器学习与深度学习方法,并结合实时数据分析与智能决策平台,可有效支持交通疏导与路径优化,提升道路利用效率。当前仍存在数据质量、模型适应性及计算时效等方面的挑战。随着边缘计算、5G通信及多交通方式协同预测技术的发展,交通预报系统将持续向智能化、高效化发展,更好地服务城市治理与智慧城市建设。

参考文献:

- [1] 崔瑶.基于5G通信技术的城市智慧交通堵塞状况预测的方法[J].长江信息通信,2023,36(12):213-215,220.
- [2] 鹿皓元.基于深度学习的交通流补齐以及预测算法的研究[D].北京:北京邮电大学,2024.
- [3] 熊佳慧.基于深度学习的交通拥堵预测研究[D].沈阳:沈阳建筑大学,2025.
- [4] 杨瑞雪.“智慧城市”背景下衡城市数字交通的研究[D].保定:河北大学,2024.
- [5] 古明生.多模式大数据融合及大模型驱动智能体系统的智慧城市建设应用分析[J].长江信息通信,2025,38(08):231-233.