

智能网联环境下城市道路多源交通数据补全方法

文 / 王 鑫 济南城建集团有限公司

尹 健 济南城建集团有限公司

摘要：随着智能网联技术的快速发展，城市道路交通管理系统正逐步向智能化、高效化转型。在这一背景下，多源交通数据的采集与分析成了提升交通管理效率、优化交通流的关键。然而，由于设备故障、通信中断、数据记录误差等多种原因，交通数据往往存在缺失或异常，这严重制约了交通数据的有效利用和智能交通系统的发展。因此，研究智能网联环境下城市道路多源交通数据的补全方法显得尤为重要。

关键词：智能网联环境；城市道路；多源交通数据；补齐方法

【DOI】10.12254/j.issn.2096-6539.2025.13.050

引言

随着城市化进程的加速和智能网联技术的快速发展，城市道路交通系统日益复杂，交通数据呈现出海量、异构和多源的特点。然而，由于设备故障、通信中断、数据传输错误等原因，交通数据往往存在缺失和不完整的问题，这严重影响了交通状态评估、交通流预测和交通信号控制等应用的准确性和可靠性。因此，开展智能网联环境下城市道路多源交通数据补全方法研究具有重要的理论和现实意义。

一、智能网联技术概述

智能网联技术是一个综合了车辆、基础设施、通信

技术和人工智能的复杂系统。它通过信息通信技术（ICT）将车辆与基础设施、其他车辆、行人和云端服务连接在一起，实现信息的实时交换与处理，从而提高交通系统的智能化和自动化水平。智能网联技术的核心组成部分包括环境感知技术、智能决策技术、控制执行技术、V2X通信技术、云平台与大数据处理技术以及信息安全技术。这些技术的集成应用，使得智能网联汽车能够实时感知周围环境，自主决策并优化行驶路径，同时与其他交通参与者进行高效协同，提高交通系统的整体效率和安全性（见图1）。

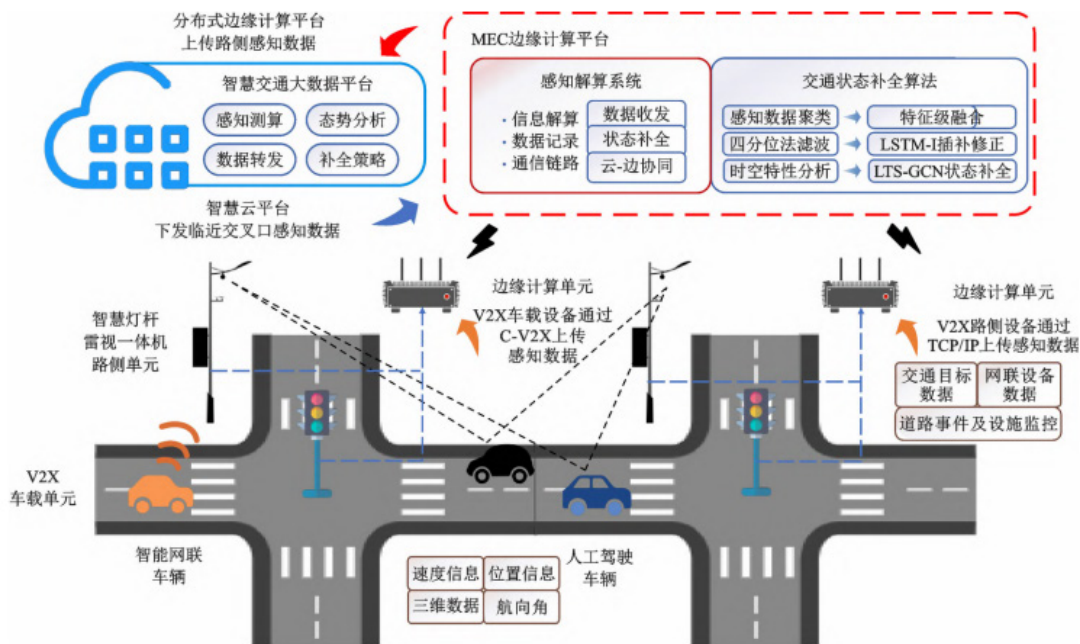


图1 智能网联交通感知系统示意

二、多源交通数据补全方法

(一) 数据预处理

在进行交通数据补全之前，数据预处理是一个至关重要的步骤，它直接影响到后续数据分析和建模的效果。数据预处理主要包括数据清洗、数据格式转换和数据归一化等步骤。首先，数据清洗是数据预处理的第一

步，其目的是去除数据中的噪声和不一致性，确保数据的质量。具体来说，数据清洗包括去除重复数据、处理缺失值和处理异常值。去除重复数据可以避免因数据采集系统的故障或人为操作失误导致的重复记录，从而减少数据处理负担和分析结果的偏差；处理缺失值可以通过删除含有缺失值的记录、使用均值或中位数填充，或

者通过插值法进行预测填充,具体方法取决于应用场景和数据特点;处理异常值则通过统计方法(如 3σ 原则)或机器学习算法(如孤立森林)来识别并处理那些明显偏离正常范围的数据点。其次,数据格式转换是为了将来自不同数据源(如传感器、GPS设备、交通监控系统等)的交通数据转换为统一的格式,便于后续处理和分析。这包括时间格式统一、数据字段对齐和数据编码转换。时间格式统一是将不同时间格式(如时间戳、日期时间字符串等)转换为统一的格式(如Unix时间戳或ISO8601格式);数据字段对齐是确保每个数据记录都包含相同的字段;数据编码转换则是将不同编码方式(如UTF-8、GBK等)转换为统一的编码(如UTF-8),以避免数据处理过程中出现乱码问题。最后,数据归一化是将数据缩放到一个特定的范围,以提高数据处理的效率和准确性。交通数据中的不同特征可能具有不同的量纲和取值范围,如车速的范围可能是0-120km/h,而交通流量的范围可能是0-1000辆/小时。如果不进行归一化处理,这些不同量纲的特征在后续的机器学习模型中可能会产生不同的影响,导致模型训练不稳定或预测结果不准确。常用的归一化方法包括最小-最大归一化、Z-score归一化和小数缩放归一化。最小-最大归一化将数据线性地映射到一个特定的范围(如[0,1]或[-1,1]),简单直观但对异常值敏感;Z-score归一化将数据转换为均值为0、标准差为1的标准正态分布,对异常值不敏感但要求数据近似服从正态分布;小数缩放归一化通过将数据除以某个固定的数(如最大值)来进行归一化,适用于数据分布较为均匀的情况。

(二) 基于图卷积神经网络的交通状态补全

图卷积神经网络(Graph Convolutional Neural Networks, GCNs)是一种专门用于处理图结构数据的神经网络模型,它通过将图结构数据转换为高维特征向量,并利用卷积操作提取节点间的局部特征,从而实现了对图结构数据的有效处理和分析。本文提出了一种基于图卷积神经网络的实时交通状态补全方法,该方法通过多个步骤实现对交通数据的有效补全。根据路网拓扑关系构建路网无向图模型,其中节点表示交叉口,边表示交叉口之间的连接关系。这一步骤是后续处理的基础,通过图模型能够直观地反映路网的结构特征。应用异常数据辨识算法对原始数据进行检测,识别出异常数据并进行插补处理,以构成有效数据集。异常数据可能由传感器故障、数据传输错误等原因引起,若不进行处理,将严重影响后续分析结果的准确性。因此,通过统计方法或机器学习算法识别并插补异常数据,确保数据的完整性和可靠性。在数据预处理完成后,根据实际路网时空关系,确定补全网络的隐藏层权重,以反映不同交叉口之间的相互影响关系。这一步骤是图卷积神经网络的核心,通过隐藏层权重能够捕捉节点之间的复杂依赖关系,为后续的特征提取奠定基础。随后,通过图卷积操作将原始数据映射至空间维度,提取交叉口特征的空间聚类信

息。图卷积操作能够有效地捕捉节点之间的局部特征,并通过多层卷积逐步提取更高层次的空间特征。与此同时,利用门控循环单元(Gated Recurrent Unit, GRU)在时间序列上游走记忆,提取数据的时间维度特征。GRU作为一种改进的循环神经网络,能够更好地处理时间序列数据中的长期依赖关系,从而捕捉交通状态在时间维度上的变化规律。将空间维度和时间维度的特征进行融合,通过全连接层进行状态数据补全计算,得到完整的交通状态信息。这一步骤是整个方法的核心,通过融合空间和时间特征,能够全面反映交通状态的时空变化规律,从而实现高精度的数据补全。该方法不仅能够有效处理交通数据中的缺失值和异常值,还能够捕捉路网中交叉口之间的复杂依赖关系,为交通管理和决策提供可靠的数据支持。通过基于图卷积神经网络的交通状态补全方法,可以显著提高交通数据的完整性和准确性,为智能交通系统的构建和优化提供强有力的技术支撑。

(三) 实验验证与结果分析

为了验证本文提出的基于图卷积神经网络的实时交通状态补全方法的有效性,本文选取了北京市高级别自动驾驶示范区的典型智能网联交叉口群进行实验验证。该区域具有复杂的路网结构和丰富的交通数据,能够充分体现本文方法的适用性和鲁棒性。实验数据包括多源交通数据,如交通流量、车速、占有率等,涵盖了不同时间段和交通状态下的数据分布。实验分为数据预处理、模型训练和结果分析三个阶段,旨在全面评估本文方法在智能网联环境下的性能。在数据预处理阶段,首先对原始数据进行清洗,去除重复数据、处理缺失值和异常值,确保数据的完整性和一致性。随后,将不同来源的数据转换为统一的格式,并进行归一化处理,以消除不同特征之间的量纲差异。这些预处理步骤为后续的模型训练和数据分析奠定了坚实的基础。在模型训练阶段,本文方法基于图卷积神经网络(GCN)和门控循环单元(GRU)构建了一个时空特征融合模型。首先,根据路网拓扑关系构建路网无向图模型,并确定隐藏层权重以反映交叉口之间的相互影响关系。接着,通过图卷积操作提取交叉口特征的空间聚类信息,并利用GRU提取时间序列特征。最后,将空间维度和时间维度的特征进行融合,通过全连接层进行状态数据补全计算。为了对比分析,本文还训练了长短期记忆神经网络(LSTM)等现有方法,以评估本文方法的性能优势。在结果分析阶段,本文采用均方根误差(RMSE)和平均绝对误差(MAE)等指标对补全结果进行定量评估。实验结果表明,本文方法在长时序数据下,补全结果与真实值之间的误差不高于10.64%,综合性能较LSTM等现有方法的RMSE降低了17.2%。这表明本文方法能够有效地捕捉交通状态的时空变化规律,实现高精度的数据补全。此外,本文方法在处理复杂路网和多源数据时表现出较强的鲁棒性和适应性,能够应对不同场景下的交通数据补全需求。进一步分析表明,本文方法的性能优势主要得益于图卷积神经

网络对空间特征的提取能力和 GRU 对时间特征的捕捉能力。通过融合空间和时间特征,本文方法能够全面反映交通状态的时空变化规律,从而提高补全结果的准确性和可靠性。此外,本文方法在处理异常数据和缺失数据时表现出较强的鲁棒性,能够有效应对实际应用中的各种挑战。

三、未来发展趋势

(一) 融合多种数据源

随着物联网、大数据和人工智能等技术的快速发展,交通数据补全的可用数据源日益丰富,融合多种数据源成为提高补全准确性和可靠性的重要途径。传统的交通数据主要依赖于固定传感器和交通监控设备,这些数据虽然具有一定的准确性,但覆盖范围有限,且容易受到设备故障或环境干扰的影响。而新兴的数据源,如车载传感器数据、手机信令数据和视频监控数据,能够提供更全面、实时的交通信息,为数据补全提供了新的可能性。车载传感器数据能够实时记录车辆的速度、位置、加速度等信息,这些数据具有高精度和高时效性,能够直接反映道路的交通状态。通过融合车载传感器数据,可以弥补固定传感器覆盖不足的问题,尤其是在城市边缘区域或复杂路网中。手机信令数据则通过移动通信网络获取用户的位置信息,能够间接反映道路的交通流量和速度分布。这种数据具有覆盖范围广、成本低的特点,适用于大范围的交通状态监测。视频监控数据通过图像识别技术提取交通流量、车速和车辆类型等信息,能够提供直观的交通状态信息,尤其是在交叉口和拥堵区域具有较高的应用价值。通过融合多种数据源,可以实现数据的互补和协同,从而提高数据补全的准确性和可靠性。例如,在固定传感器数据缺失的情况下,可以利用车载传感器数据或手机信令数据进行插补;在复杂路网中,可以通过视频监控数据验证其他数据源的准确性。此外,多源数据融合还能够提高模型的鲁棒性,降低单一数据源异常对补全结果的影响。

(二) 智能化和自动化

未来交通数据补全方法将更加注重智能化和自动化,通过引入先进的算法和模型,如深度学习、强化学习等,能够显著提升数据补全的效率和精度。深度学习模型,特别是图卷积神经网络(GCN)和长短期记忆网络(LSTM),能够从复杂的交通数据中自动提取时空特征,实现对缺失数据的精准预测和补全。强化学习则可以通过与环境的交互不断优化补全策略,适应动态变化的交通场景。这些智能化方法不仅能够减少人工干预,还能在处理大规模、高维度数据时表现出更高的效率和鲁棒性。与此同时,结合自动驾驶和智能交通系统(ITS)等先进技术,交通数据补全将更加紧密地融入城市交通管理的整体框架中。自动驾驶车辆通过车载传感器和通信设备实时采集和共享交通数据,为数据补全提供更丰富的信息源。智能交通系统则通过集成多源数据和智能化算法,实现对交通状态的实时监控和动态优化,从而提升道路通行

效率、减少拥堵和排放,构建更加安全、高效和环保的城市交通系统。

(三) 实时性和大规模数据处理

随着城市交通系统的日益复杂和庞大,实时性和大规模数据处理成为未来交通数据补全方法的重要挑战。城市交通数据具有高维度、高频率和大规模的特点,传统的数据处理方法难以满足实时性和高效性的需求。为了应对这一挑战,未来需要从算法和硬件架构两方面进行优化。在算法层面,可以通过分布式计算和并行化技术提升数据处理效率。例如,采用分布式图卷积神经网络(GCN)或分布式长短期记忆网络(LSTM)将计算任务分配到多个节点上,从而加速模型训练和推理过程。此外,还可以引入流式计算框架,如 Apache Flink 或 Apache Storm,实现对实时数据流的快速处理和分析,确保交通状态补全的实时性。在硬件层面,可以利用高性能计算设备,如 GPU(图形处理单元)和 TPU(张量处理单元),加速深度学习模型的训练和推理。同时,结合边缘计算技术,将部分计算任务下沉到靠近数据源的边缘设备上,减少数据传输延迟,提高实时响应能力。

结语

本文提出了一种基于图卷积神经网络的实时交通状态补全方法,用于智能网联环境下城市道路多源交通数据补全。该方法通过构建路网无向图模型、应用异常数据辨识与插补方法、确定补全网络隐藏层权重、进行图卷积操作和门控循环单元处理等步骤,实现了对缺失交通数据的有效补全。实验结果表明,该方法具有较高的准确性和可靠性,为智能网联环境下交通全息感知技术应用提供了理论基础和实现方案。未来,随着智能网联技术的不断发展和应用,城市道路多源交通数据补全方法将呈现出更加智能化、自动化和实时化的发展趋势,为城市交通管理和优化提供更加有力的支持。

参考文献

- [1] 卢家琪. 动态交通网络流量数据补全及连续推断研究[D]. 电子科技大学, 2024.
- [2] 杨玮. 基于改进神经网络算法的多源交通数据融合方法研究[D]. 东南大学, 2022.
- [3] 徐文权. 基于多源交通时空数据的数据补全技术研究[D]. 大连理工大学, 2021.
- [4] 杨盛宝. 基于多源数据集共有结构的交通数据张量补全算法研究[D]. 云南大学, 2021.
- [5] 胡雪, 彭敦陆. 张量表达下的多模态交通缺失数据补全算法[J]. 小型微型计算机系统, 2021, 42(01): 105-110.
- [6] 谢竞成, 丁楚吟, 吴越, 等. 基于多源交通数据的城市道路实时运行状态综合评价方法[C]// 中国智能交通协会. 第十三届中国智能交通年会大会论文集. 银江研究院; 浙江工业大学信息工程学院; 银江股份有限公司; 浙江大学智能交通研究所; , 2018: 12.