

多源大数据融合提升智慧交通环境监测效率与准确性研究

张默可

江西机电职业技术学院 江西南昌 330000

摘要: 在城市化进程不断加快的今天, 交通环境监测是城市管理的一个重要环节。交通环境不但关系着市民出行体验, 而且对城市空气质量、噪声水平和生态环境也有着直接影响。但是, 传统交通环境监测手段通常依靠固定监测站点等单一数据来源进行监测, 很难全面准确反映交通环境复杂多样情况。近年来大数据技术快速发展, 使得多源数据采集和融合得以实现。因此, 文章分析多源大数据概念, 探讨如何利用多元大数据提升智慧交通环境监测效率与准确性。

关键词: 多源大数据融合; 智慧交通; 环境监测

在信息技术不断进步的背景下, 多源大数据融合运用给智慧交通环境监测带来新机遇。通过融合交通流量数据、空气质量监测数据、气象数据和社交媒体用户反馈信息等不同来源的信息, 能够构建出更加全面和详细的交通环境画像。这种跨领域数据融合既可以揭示交通环境和各因素间的内在关系, 又可以为交通管理决策提供科学依据。所以, 探究智慧交通环境监测中多源大数据融合技术的运用, 对提升监测效率, 强化监测准确性, 推动交通环境可持续发展有着深远的意义。

1 多源大数据概述

1.1 定义

多源大数据是指从各种不同种类, 结构, 格式和来源渠道中收集和聚合而成的大量数据集。这些数据源是多元的, 涉及各种传感器设备, 信息系统, 社交网络平台和移动终端。以智慧交通为例, 传感器的数据来源于交通流量监测设备和气象监测传感器等; 信息系统的数据来源于交通管理系统, 车辆运营调度系统; 社交网络平台数据中包含用户对于交通状况的实时反馈信息; 在移动终端的数据中, 可以找到车辆的行驶轨迹以及出行 APP 所记录的行程详情等。多源大数据打破单一数据源的限制, 通过融合多维度数据为各个领域的深入分析和决策提供丰富而综合的数据基础, 从而满足复杂业务场景中信息在广度和深度上的要求^[1]。

1.2 特征

数据规模巨大: 多源大数据的体量异常巨大, 规模远远超出传统数据处理系统所承载的容量。在交通领域中, 每天生成的车辆行驶轨迹数据, 交通流量监测数据以及社交

媒体中关于交通状况的讨论等数据量累计都能达到 PB 乃至 EB 级。海量数据中包含丰富的信息, 但是同时对于存储, 计算资源的需求却非常高。

数据类型多样: 多样的数据来源, 决定多源大数据种类的丰富性。既有结构化数据库表数据例如交通管理部门车辆登记信息等; 还有半结构化 XML、JSON 格式的数据, 类似于智能交通设备所传递的配置和状态信息; 还存在着海量的非结构化的数据, 例如道路监控视频, 与交通有关的文本资料和音频。不同种类的数据处理方式和分析方法不一样, 加大数据融合和运用的困难。

数据速度快: 多源大数据生成迅速且实时性强。如交通传感器在秒级乃至毫秒级的频率下收集数据、车辆在运行时不断上传轨迹信息、社交网络中对交通事件的探索在一瞬间扩散等等。数据流速快需要系统具有实时处理的能力, 对信息进行及时捕获, 分析和反馈, 才能适应交通实时监测和应急响应的需要。

数据真实性复杂: 数据的真实性因来源的广泛性和数据质量的良莠不齐而有很大的不同。一些传感器会失效而造成数据异常, 社交网络中的数据含有虚假信息或者主观偏见, 并且不同系统的数据之间会发生冲突。保证数据真实性, 发现和排除错误数据是多源大数据价值挖掘的关键环节。

2 多源大数据融合在智慧交通环境监测中的流程

2.1 数据采集

智慧交通环境监测中数据的获取是基础。各种传感器发挥着各自的作用, 交通流量传感器布设在道路的关键节点上, 如十字路口和高速出入口, 采集车流量和车速的实时信

息。气象传感器主要关注天气情况，包括温度、湿度、风速和降雨等各种数据，为交通状况提供重要的气象信息。同时摄像头拍摄的路面画面将车辆行驶轨迹和交通违章行为一览无余。另外在车载设备和智能手机上设置定位系统可以得到汽车的实时位置和行驶状态。这些广布且种类繁多的数据源从各个角度对交通环境数据进行全方位的收集，并为之后的分析提供丰富的资料^[2]。

2.2 数据传输

需要迅速，稳定地将采集数据传输给处理中心。利用如光纤这样的有线传输手段，由于其高带宽和低延迟的特点，在城市的交通中心、固定的观测站等关键位置，能够实现大量数据的高速传输。通过 4G/5G 网络的无线传输，车载设备和移动监测终端可以实现数据的即时上传，从而充分发挥其灵活性。在某些偏僻或布线困难的地区，LoRa 等低功耗广域网技术能够以较低的能耗实现长距离的数据传输。为保证数据的安全性和完整性，在传输的过程中使用加密协议来避免数据的被盗用或者篡改。不同传输方式协同配合，搭建一条高效的数据源至处理端的数据传输通道以保证数据能够及时到达下一步。

2.3 数据预处理

刚刚收集到的资料有很多问题，对资料进行预处理是一个解决之道。首先对数据进行清洗以剔除重复，错误和缺失的数据。比如交通流量传感器会因为失效而出现异常值的情况，通过设置合理的阈值进行甄别，将此类无效数据排除。随后，开始对数据进行标准化处理，统一各种格式和量级的数据，以便于后续的数据分析工作^[3]。例如将不同传感器收集到的温度数据统一转换为摄氏度。在数据融合的过程中，将来自多个来源的数据从时间和空间等多个角度进行整合，例如将同一时刻不同路段的交通流量和气象数据相互关联。在这一系列预处理后，杂乱无章的数据规整可得，从而为深度分析和智慧交通决策打下良好的基础。

3 多源大数据融合提升智慧交通环境监测效率与准确性策略

3.1 实施异构数据协同整合，推进交通监测效率提升

在智慧交通大繁荣的今天，交通监测领域面临空前的数据挑战。在物联网，传感器技术和智能设备广泛应用的背景下，交通数据的来源异常广泛，涉及交通摄像头，各种传感器和车辆定位系统，由于数据格式的多样性，数据整合过

程面临着巨大的挑战。实现异构数据协同整合，已经成为提高交通监测工作效率的重点突破口^[4]。

基于此，构建统一的数据标准体系势在必行，通过以交通摄像头所拍摄图像数据为研究对象，通过建立详细数据格式规范，编码规则和数据定义、传感器输出物理量数据，车辆定位系统生成坐标数据等等，经过综合规范化处理后，具有统一一致格式。在此基础上，运用成熟的数据抽取、转换与加载（ETL）等技术，并借助有效的数据抽取工具对散布于城市每个角落的数据源进行数据抽取，在进行格式转换和数据清洗之后，把清洗出来的优质数据装载到统一数据仓库中，并对数据进行集中存储和管理。为进一步破除不同系统之间数据的语义隔阂，引入语义映射和本体匹配技术，通过构建语义模型，将不同数据源中的相同概念进行精准映射，实现数据的无缝对接，从而大幅提升数据整合效率，为后续深层次的分析与实时监测工作奠定坚实基础。

3.2 运用智能算法融合优化，助力交通数据准确研判

随着智能交通技术的持续进步，我们观察到交通数据量正在急剧上升，而传统的数据处理方法已经不能满足对交通信息进行精确分析的要求。采用智能算法进行融合优化，就成增强交通数据处理能力和分析准确性的不二选择。在深度学习算法中，卷积神经网络（CNN）和循环神经网络（RNN）各有其独到之处，将这两种网络完美融合可以大大增强对交通数据的处理能力。

特别是 CNN 对图像数据处理性能突出，交通摄像头采集到的实时路况画面中蕴含丰富而复杂的视觉信息，通过构造多层卷积层与池化层来实现 CNN，能自动地提取出车辆数量，车型和行驶方向等图像关键特征。而且 RNN 善于对时间序列数据进行处理，对交通流量随时间变化的信息，RNN 可以利用自身的循环结构有效地捕获时间序列中长期的依赖关系并对未来交通流量趋势进行精确预测。通过大范围地训练大量历史交通数据，使融合模型能够全面地学习交通数据所包含的复杂规律和内在规律。此外，使用集成学习技术，将多种不同的智能算法，例如支持向量机（SVM）和决策树等，进行整合，以最大化它们在分类任务中的应用、回归等优点，将多种算法输出的结果进行整合，有效提升数据研判精度，可对交通拥堵概率，事故风险等级以及其他重要交通状况进行较为准确的预判。

3.3 搭建多源数据实时交互，促进交通信息高效流通

为推动交通信息高效流转，构建多源数据实时交互平台成为智慧交通建设中至关重要的一环。现代交通系统需要将来自不同数据源的数据及时准确地传输到各需求方为实时决策提供支持。借助消息队列技术，例如在行业中被广泛采用的 Kafka，我们可以建立一个高效的数据传输路径，确保从不同的数据源获取的数据能够被实时地推送和接收。Kafka 以其高吞吐量和低延迟等特点能够快速处理大量数据传输任务。同时利用 Spark 等分布式计算框架并行处理实时传输数据。Spark 以内存计算为特点，可以大大提高数据处理速度，它通过把数据划分为若干分区并在簇内若干节点并行执行计算任务来明显减少数据处理时间。

磁瓦，搭建实时数据交互接口并采用 RESTful API，方便交通管理部门，出行者和运输企业各方面的工作、有效的数据获取方式使得可以实时地获得需要的交通信息例如路况，公交的实时位置以及货运车辆的运行情况。通过物联网技术，将交通基础设施（比如路，桥，信号灯）、车辆等设备连接起来，形成一个庞大且复杂的信息交互网络，保证各种交通信息能在各主体之间快速准确流转，从而为交通决策的制定提供及时可靠的数据支撑，有利于实现交通系统高效运行和科学管理。

3.4 构建动态权重分配模型，增强数据融合精准程度

在多源大数据进行融合时，为提高数据融合精准程度，建立动态权重分配模型成为有效方法。由于各种数据源在各种交通场景中扮演的角色及重要程度是有差别的，所以有必要结合实际情况合理地分配并动态地调整各数据源权重。

首先，利用历史数据和实时数据，采用层次分析法(AHP)来确定各个数据源的初始权重。AHP 对复杂决策问题进行层次结构模型的构造，并将其分解成若干等级，采用两两对比的方法来确定各个要素的相对重要程度，进而获得各个数据源初始权重。在真实的应用场景中，考虑到数据的及时性和准确性等核心要素，我们选择自适应权重调整方法，例如粒子群优化 (PSO) 策略，来对数据源的权重进行实时的动态调节。PSO 算法对鸟群觅食行为进行仿真，利用粒子之间

相互协作和信息共享来寻找解空间内的最优解。比如，在交通事故中，事故现场传感器及摄像头所采集到的数据对事故影响范围，交通拥堵情况等精确评价起着关键作用，这时采用 PSO 算法显著提高这些数据源权重，使得融合数据能更加准确地反映事故发生时周围的交通状况，从而达到动态准确融合交通数据，有效增强监测结果可靠性和可信度的目的。

4 结语

多源大数据融合显示出极大的潜力，它通过对多源异构数据的融合打破传统监测的局限性，给交通环境监测带来综合视角。它由基础流程构建走向提升效率和准确性策略，促使交通监测由静态，单一走向动态和多元。这样既革新交通管理模式又为城市的可持续发展带来动力，有助于城市走向交通与环境协调发展的新台阶，未来需持续深化研究，拓展其在智能交通体系全方位应用，挖掘更多潜在价值，以应对复杂多变的城市交通挑战。

参考文献：

- [1] 李美玲, 吴海伟, 李子辉, 等. 基于多源大数据的高速公路事故预测机器学习模型对比研究 [J]. 武汉理工大学学报, 2024, 46(12): 126-134.
- [2] 林港. 基于大数据的济南市交通流量预测模型研究 [J]. 智能城市, 2024, 10(12): 20-22.
- [3] 吴祥国, 张建嵩, 余梓冬, 等. 基于大数据的轨道线路初期运营客流预测实践 [C]// 中国城市规划学会城市规划专业委员会. 绿色数智 提质增效——2024 年中国城市规划年会论文集. 重庆市交通规划研究院; 重庆市交通规划研究院, 交通信息中心.; 2024: 1838-1849.
- [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. DOI: 10.27456/d.cnki.gyndu.2021.002626.

作者简介：

张默可 (1993 年 10 月—), 男, 汉族, 江西上饶市余干县, 助教, 硕士研究生, 研究方向: 智能交通、数据挖掘。