

基于多源数据融合的交通运输安全指数体系构建研究

杨 静

河南省运输事业发展中心（河南省邮政安全发展中心） 河南 郑州 450016

摘 要：交通运输安全评估正经历从单指标预警向多维度综合研判的范式转型，而数据环境的日益丰富为精准刻画安全态势提供了新机遇。当前交通运输领域已积累海量的交通流检测、气象环境、基础设施状态、驾驶行为及事故记录等异构数据，但各类数据源之间相互割裂，缺乏有效的融合机制与统一的量化标尺。本文提出构建交通运输安全指数体系，旨在通过多源数据的系统性融合，形成覆盖人、车、路、环境、管理五维度的综合安全测度框架。研究首先辨析交通运输安全评估的数据需求与现有数据资源的匹配关系，继而设计指数体系构建的理论框架，明确指标筛选、权重配置与合成方法的学理依据，并探讨数据融合过程中的时空对齐、质量清洗与动态更新等关键问题。在此基础上，阐述安全指数在态势感知、风险预警与治理决策中的功能定位。研究认为，安全指数体系的核心价值不在于生成单一数值，而在于建立一套可解释、可追溯、可比较的安全状态表征语言，为交通运输安全管理提供科学决策支点。

关键词：多源数据融合；交通运输安全；指数体系；安全评估；数据治理

引言

交通运输系统规模的持续扩张与运行环境的日趋复杂，使得安全管理的精准性要求与既有评估手段之间的差距不断拉大。传统的交通运输安全评估多依赖事故统计指标，以死亡人数、万车死亡率、事故起数等事后性数据作为评价依据，其局限性在于：滞后性强，无法反映安全状态的实时变化；归因能力弱，难以辨析事故背后的多因素耦合机制；覆盖度低，大量未酿成事故的安全隐患得不到量化关注。与此同时，物联网感知、移动定位、视频识别及气象监测等技术的广泛部署，使得交通系统运行状态的数字化描述成为可能，数据维度之丰富、更新频率之密集远超以往。然而，多源数据在格式、尺度、精度与时效上的异构性，使其难以直接整合为统一的决策参考信息。如何将分散的数据资源转化为系统性的安全认知工具，已成为交通运输治理能力现代化的重要课题。

1 交通运输安全评估的数据环境与融合需求

1.1 数据源的类型与特征

交通运输安全相关的数据源可归为六类。交通流检测数据（包括地磁线圈、雷达测速、卡口车牌识别及手机信令）实时反映车辆速度、密度与流量变化，是刻画运行秩序的基础信息。气象与环境数据（能见度、降雨强度、路面温度、风力等级）直接影响路面附着系数与驾驶视野，构成安全的外部胁迫因子。基础设施状态数据（路面平整度、摩擦系数、桥隧技术状况、护栏完整

性）表征道路本身的物理安全水平。驾驶行为数据（通过车载诊断接口或驾驶监测设备获取的急加速、急减速、疲劳时长等）映射驾驶员风险偏好与操作规范性。事故与事件记录（含事故位置、类型、伤亡等级及近邻事件）提供安全后果的直接测度，但属低频稀疏数据。此外，互联网开放数据（导航路况、社交媒体舆情、施工占道信息）可作为补充信息源，增强对突发扰动的感知能力。

1.2 数据异构性与融合障碍

上述数据源在三个维度上存在显著异质性。时间尺度上，交通流数据以分钟甚至秒级更新，气象数据以小时级更新，而基础设施检测数据可能以季度或年度为周期刷新，时间粒度的不匹配使同步融合面临技术困难。空间尺度上，卡口数据为断面采集，气象数据为区域面状插值，事故数据为离散点记录，三者在空间表达方式上的差异要求统一的空间化处理^[1]。语义尺度上，不同数据源对“安全状态”的描述维度不同——流量数据反映的是运行秩序，气象数据反映的是环境胁迫，行为数据反映的是人为因素，彼此之间缺乏共同的量化锚点。若不进行有效的融合处理，这些数据只能各自孤立地触发局部告警，而无法合成对整体安全态势的综合判断。

1.3 融合的增值逻辑

多源数据融合对安全评估的价值提升体现在三个层面。一是信息互补——事故数据的稀疏性可由交通流异常波动数据加以弥补，因大量安全隐患在事故前已表现

为运行特征畸变,融合可使评估从“事后统计”前移至“事中感知”。二是交叉验证——同一安全态势在不同数据源中应有对应信号,融合可实现相互校核,降低单一数据源误报或漏报的风险。三是因果增强——多维度数据叠加有助于辨析事故风险的贡献来源,例如在同一事故率上升背景下,若同期降雨与流量均无显著变化而急减速事件激增,则可初步判定风险源偏向人为因素或车辆因素。这种归因能力是单一数据源无法提供的。

2 安全指数体系构建的理论框架

2.1 安全指数的内涵界定

交通运输安全指数是对特定时空范围内交通系统安全状态的综合量化表达,其本质是一个合成的相对性指标。它不直接等同于“事故概率预测”,亦非简单的事件统计汇总,而是通过对多源观测数据的融合运算,生成反映安全状态相对优劣的序次性量值。指数必须具备三个基本属性:可比性(同一指数在不同时间、不同区域之间具有可比意义)、可解释性(指数变化能够回溯至具体指标变动,而非不可分解的黑箱输出)与敏感性(指数应能响应安全状态的实质性变化,而非过度平滑以至失去预警价值)。

2.2 构建原则

指数体系的构建应遵循五项原则。科学性原则要求指标的选取与权重配置具备理论依据或实证支撑,避免随意赋权。系统性原则要求维度设计完整覆盖影响交通运输安全的主要因素,防止关键变量遗漏造成评估盲区。可操作性原则强调指标应基于可获得、可更新的数据源,脱离数据支撑的指标设计无实际意义。独立性原则要求同一维度内部指标之间的相关性控制在合理范围内,避免多重共线性导致信息冗余和权重扭曲^[2]。动态性原则要求指标体系具备吸收新数据源、淘汰失效指标的更新机制,而非一经确立便固化不变。

2.3 多维框架结构

基于交通运输系统组成要素的经典分解,安全指数体系采用“人—车—路—环境—管理”多维框架。其中,“人”维度聚焦驾驶员生理心理状态、操作规范性及安全素养,反映人为风险因子;“车”维度涵盖车辆制动性能、轮胎状况、安全装置有效性及载重合规性,测度运载工具的本质安全水平;“路”维度评价道路几何线形、路面性能、交安设施完备性及防护能力,表征基础设施的安全供给条件;“环境”维度整合气象条件、光照状况及自然灾害预警,刻画外部时空场景的胁迫程度;“管理”维度纳入执法力度、应急预案完备

度、安全教育覆盖率及事故处置效率,反映治理能力的调节作用。五个维度之间并非简单并列,而是存在“环境胁迫—路车承载—行为响应—管理调节”的链式耦合关系,这一内在逻辑为后续的权重配置与指数合成提供了结构先验。

3 多源数据融合的关键方法与指标合成

3.1 数据预处理与时空对齐

融合的第一步是建立统一的数据参考框架。时间层面,需确定基本分析单元(如小时、日、周),各数据源按该单元进行时间聚合——高频数据取均值或极值,低频数据通过插值或最近邻赋值延展至每个分析单元。空间层面,须建立路段编码系统(如基于里程桩号或地理信息系统网格),将点状、线状与面状数据统一匹配至路段单元。对于缺失数据,不宜简单删除或均值填充,而应根据缺失模式选择线性插值、多重插补或基于相似路段迁移的方法。数据质量清洗环节需剔除明显的传感器漂移、传输丢包及逻辑矛盾记录,并建立数据可信度标签供后续融合时参考。

3.2 指标筛选与降维处理

从原始数据到最终指标集合须经筛选与降维。初选指标池可包含数十乃至上百个可计算变量,需通过相关性分析剔除两两相关系数高于0.85的冗余变量,保留解释力强且代表性高的指标。进一步可采用主成分分析或因子的分析,将同一维度内的多个指标压缩为少数综合因子,既降低维度又减少噪声干扰^[3]。但需注意,降维不应损害指标的可解释性——若某个主成分的载荷结构杂乱无章,则应放弃机械降维而改用理论导向的分组加权。最终保留的指标数量控制在20~30个为宜,过少则覆盖不足,过多则操作成本攀升且边际增益递减。

3.3 权重配置的主客观结合方法

权重设置是指数合成的核心环节,单一主观赋权(如层次分析法)易受专家偏好影响,单一客观赋权(如熵权法)又可能输出违背常识的权重分布。推荐采用主客观结合的组合赋权策略——先以层次分析法确立维度层级的理论权重(体现“理想状态下各维度重要性”的价值判断),再以熵权法或离差最大化法确定指标层级的客观权重(反映实际数据中指标的区分能力),最后通过乘法或线性组合生成综合权重。对于“管理”等难以量化但理论重要的维度,可适当提高主观权重占比,避免客观计算将其边缘化。

3.4 指数合成与标准化

在指标无量纲化处理(常用极值标准化或Z-score标

准化)和权重确定之后,采用加权线性合成或加权几何合成生成综合指数。线性合成直观且易于分解,便于追溯各维度贡献度;几何合成则能抑制极端值影响,适用于指标间具有较强互补性的情形。无论采用何种合成方式,最终指数均需经过标准化变换,设定基准值(如历史中位数或参考区域均值)使指数围绕100或0上下波动,便于跨期跨区域比较。应同时报告综合指数与各维度分项指数,避免单一总分掩盖维度间的此消彼长。

4 指数体系的运行机制与应用适配

4.1 动态更新与阈值预警

安全指数须纳入常态化更新机制,更新频率视数据供给能力与决策需求而定——省级宏观评估可按月度更新,市级管理可按周更新,重点路段或特殊时期(节假日、汛期)可按日更新。阈值的设定不应采用固定百分位截断,而应基于指数历史分布的控制图原理,以均值加减两倍标准差作为“关注阈值”,三倍标准差作为“警戒阈值”,当指数跌破阈值时自动触发分级预警流程。阈值的自适应性要求每年或每两年依据新增数据重新计算统计基线,避免因系统运行状况整体改善或恶化导致的阈值失效。

4.2 面向不同层级治理的适配输出

安全指数体系的服务对象涵盖决策层、管理层与作业层三个层级,输出形式应相应分化。对决策层,提供综合指数趋势图、区域排名与重大风险摘要,侧重宏观态势判断与资源分配依据^[4]。对管理层(如市级交通运输局、高速公路运营公司),提供分维度指数明细、主要短板指标及同期对比,用于制定整改措施与督查重点。对作业层(如路政巡查、执法中队),应提供路段级预警名单与重点关注指标原始值,直接指导现场排查与处置。同一指数体系下不同颗粒度的输出设计,是实现“数据—信息—决策”转化链的关键。

4.3 与既有管理制度的衔接

安全指数体系不宜游离于现有交通运输安全管理体制之外,而应嵌入已有的考核、通报与约谈制度。可将指数纳入年度安全生产责任制考核指标体系,权重不宜过高(建议5%~10%),以免诱发为追求指数而选择性上报数据等逆向行为。同时,指数作为研判辅助工具,

而非行政问责的直接依据,其功能侧重于“预警”与“诊断”而非“排名”与“惩罚”。在制度设计中明确指数异常不直接等同事故责任,但须触发专项分析报告机制,由专业团队核查原因并提交处置建议。这种非惩罚性的使用定位有助于维护数据上报的真实性,避免因顾虑问责而产生的信息扭曲。

5 结语

交通运输安全管理正处在从经验驱动向数据驱动转型的关键阶段,而多源数据融合与指数化表达是实现这一转型的方法论桥梁。本文构建的基于多源数据融合的交通指数体系,在理论上确立了五维评估框架与主客观组合赋权的方法论基础,在操作上明确了数据对齐、指标筛选、动态更新与分级输出的完整流程,在使用上区分了面向不同层级的输出形式及其与既有管理制度的接口设计。该体系的核心贡献在于将分散的异构数据转化为统一、可比、可解释的安全状态语言,使隐性风险得以显性化表达,使跨区域、跨时期的安全绩效比较获得量化依据。需要指出的是,指数体系的科学性与实用价值并非取决于算法的复杂性,而取决于数据基础的扎实程度、指标选取的理论合理性以及权重配置的透明性。未来研究应持续关注新型数据源(如车路协同报文、毫米波雷达点云、自动驾驶轨迹)的纳入规则,以及人工智能方法在模式识别环节的辅助作用,但须始终秉持“可解释优先”原则,确保指数体系的逻辑链条清晰、决策者与公众可理解,方能在交通治理实践中获得持续生命力。

参考文献

- [1] 王长君,刘东波,李平凡.道路交通安全评价方法研究综述[J].中国公路学报,2022,35(1):1-15.
- [2] 张兰芳,孙剑,王昊.基于多源数据融合的城市交通运行安全评估框架[J].交通运输工程学报,2023,23(3):112-124.
- [3] 陆化普,王建强,张毅.交通系统安全态势感知的理论与方法[J].清华大学学报(自然科学版),2021,61(10):1053-1063.
- [4] 裴玉龙,马国胜,赵晓华.道路安全评价指标体系的构建与实证检验[J].哈尔滨工业大学学报,2022,54(9):78-87.