ICS 03.220.20
CCS R 00

团 体 标 准

T/ITS 0242-2023

城市交通大数据质量评价体系 设备感知类

Urban traffic big data quality evaluation system--device sensing data

（征求意见稿）

本稿完成日期：2024 年 9 月 20 日

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上

2024-xx-xx 发布 2024-xx-xx 实施

中 国 智 能 交 通 产 业 联 盟 发 布

目　　次

前　言

本文件按照GB/T 1.1-2020给出的规则起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国智能交通产业联盟（C-ITS）提出并归口。

本文件起草单位：青岛海信网络科技股份有限公司、青岛市交通运输局、南京慧尔视智能科技有限公司、东南大学、交通运输部公路科学研究院、金陵科技学院、北京工业大学。

本文件主要起草人：***。

城市交通大数据质量评价体系 设备感知类

1 范围

本文件规定了设备感知类交通大数据进行质量评估的通用规则、规范流程与评价方法。

本文件适用于企业、公安交警等开展设备感知类数据质量评价、业务应用层级数据支撑度相关评价工作，保证数据集在业务应用中的可用性，促进业务引用的开展。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 29101 道路交通信息服务数据服务质量规范

GB/T 35775-2017 智慧城市时空基础设施 评价指标体系

GB/T 36344-2018 信息技术 数据质量评价指标

NB/T 11083-2023 风电信息管理数据质量评估及治理技术规范

DB 5227/T 112-2022智慧黔南 数据质量评价规范

T/CITSA 10-2021 城市交通时空大数据格式标准

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据质量 data quality

数据质量是指在特定的业务环境下，数据符合数据消费者的使用目的，能满足业务场景具体需求的程度。

3.2

数据质量评价体系 data quality evaluation system

数据质量评价体系是一套用于评估和监控数据质量的标准和方法,其目标是确保数据质量满足业务需求，提高数据的可用性和可信度。

3.3

设备感知类数据 device perception data

设备感知类数据指的是交通设备设施，如交通信号灯、交通监控摄像头、车载传感器等收集的关于交通状况的数据。主要包括过车数据、交通流量、车辆GPS等数据。这些数据对于交通管理、交通规划、智能交通系统等领域具有重要价值。

4 缩略语

下列缩略语适用于本文件。

GPS: 全球定位系统（Global Positioning System）

5 数据质量评价体系

5.1 数据范围

设备感知数据作为交通大数据的重要来源之一，其定义为城市交通运行过程中交通建设的设备设施收集到的交通信息数据。设备感知类数据包括基础信息与设备检测信息，按照来源包括电警卡口数据、交通流数据、视频数据、车载 GPS 数据四类：

a） 电警卡口数据：基础信息包括电警设备基础信息、设备安装点位信息；电警卡口检测数据为设备检测的过车记录或违法行为信息，包括号牌号码、号牌类型、过车时间等信息。主要应用交通运行态势、信号控制、指挥调度等场景；

b） 交通流数据：基础信息包括多目标雷达、微波、超声波等设备基础信息、设备点位信息；交通流检测数据为设备检测的交通流量信息，主要应用于交通运行态势、违法分析、缉查布控等场景；

c） 事件检测数据：基础信息包括设备基础信息、设备安装点位信息；事件检测包括了炸街车声呐检测器、不礼让行人、违法停车、大货车闯禁行、不礼让行人、抛洒物、逆行等事件检测信息；

d） 车载 GPS 数据：基础信息为车载机信息、车辆信息、车载机与车辆关联信息；GPS 数据是车辆出行实时的记录信息。

5.2 评价体系

数据质量评价体系是从数据综合应用的角度考虑，从单一数据质量评价与业务支撑度两个层次进行全面的考察和评价，为数据治理与数据决策提供支撑。

单一数据质量评价是对某一项数据开展质量评价，业务支撑度是对与业务相关联的多项数据项开展质量评价，对评价流程、评价维度与指标、评价方法、评价等级进行规范。

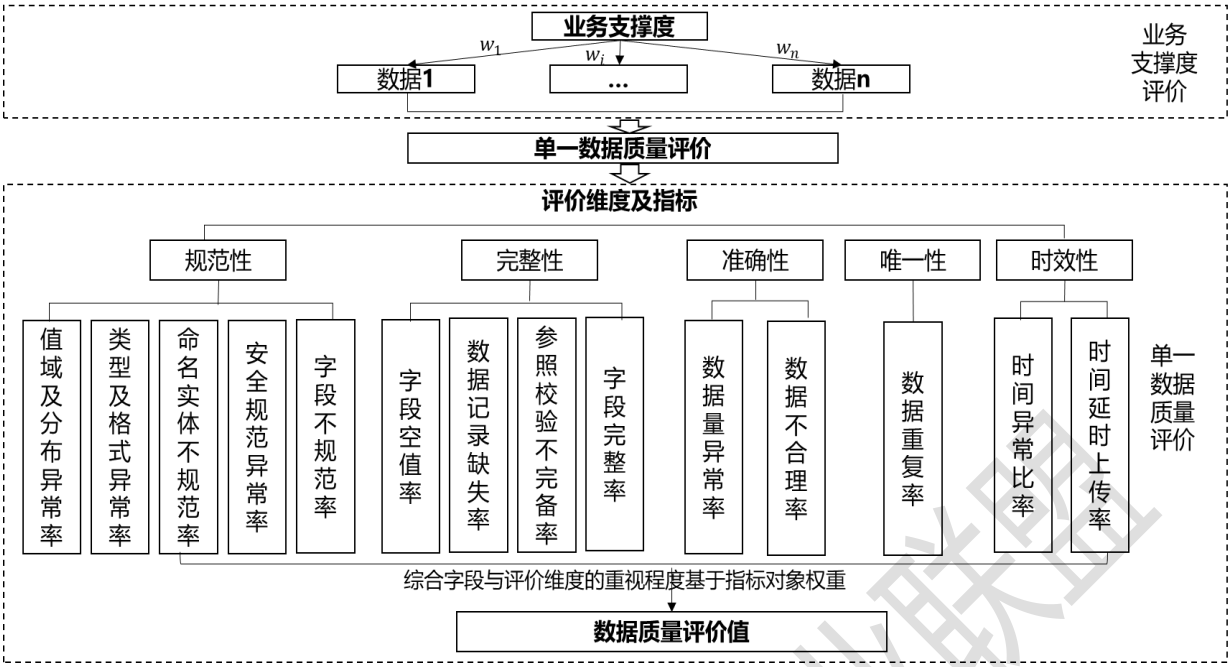

图 1 数据质量评价维度

6 质量评价模型

6.1 评价流程

数据质量评估一般步骤由数据评价与业务支撑度评价两部分组成，如图 2 所示。

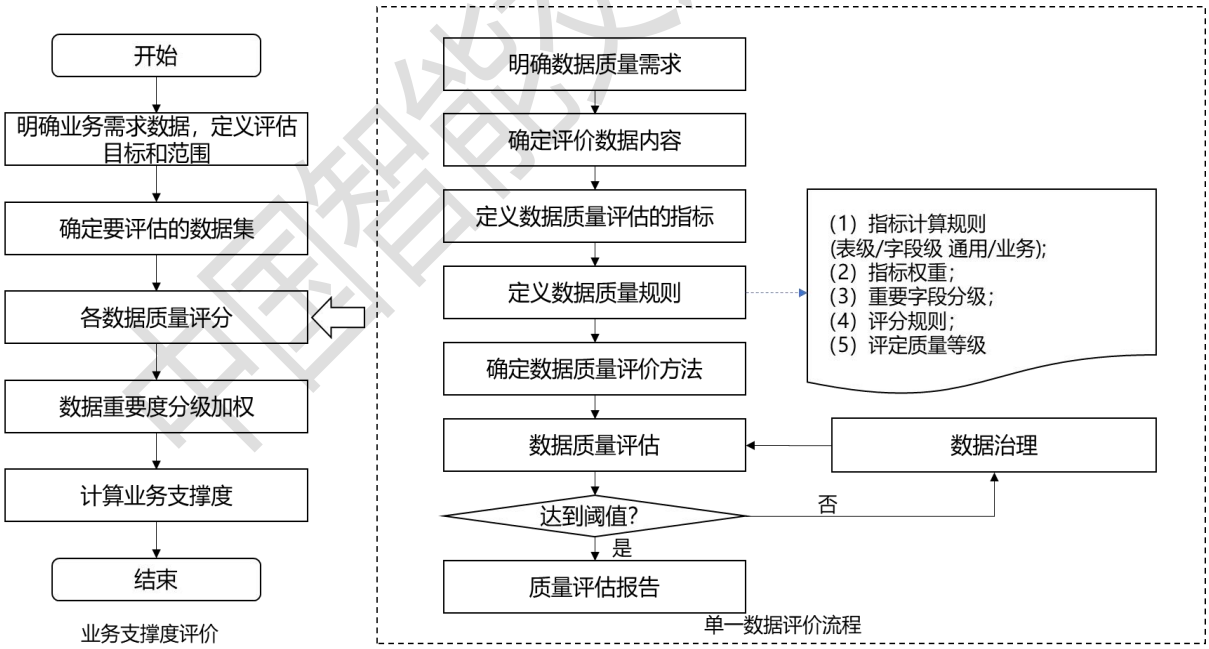

图 2 数据质量评价流程图

6.1.1 单一数据质量评价

单一数据评价流程包括：

a） 明确数据质量需求，定义评估目标和范围。根据实际的业务需要，明确数据质量需求目标，确定数

据的时空范围。

b） 确定评价数据内容，依据数据评价目标根据设立不同数据项的数据评价指标，得到数据质量评价规则项。

c） 定义数据质量规则，包括指标计算规则、各规则权重、字段重要程度分级、评分规则、评定质量分级。指标计算规则定义了各指标的计算方法，各规则权重与字段重要程度分级根据实际业务需要定义值，评分规则定义了综合评价得分的计算方法，评定质量分级定义了设定数据质量的三等级标准，明确各等级的划分依据。

d） 确定数据质量评估方法并开展数据质量评价。按照既定方法和规则开展数据质量评估，记录评估过程和结果。

e） 输出质量评价值。对比评估结果与预期的质量需求目标，进行数据信息的判断，可依据指标结果分析数据质量存在的问题及原因，输出质量评估报告。

6.1.2 业务数据支撑度评价

业务支撑度评价流程包括：

a） 明确业务的数据质量数据，定义评估目标和范围。根据实际的业务需求，明确开展某些业务需要的数据项以及需求目标，明确业务范围。

b） 确定要评估的数据集与范围，根据业务需求和范围确认涉及的数据源、数据源的时空范围。

c） 业务关联数据源进行综合评分判定，即该业务关联数据源评分按照单一数据质量评价流程进行。

d） 确认数据重要度，根据该业务对不同数据源的需求与依赖程度设置权重。

e） 输出业务支撑度，使用业务支撑度评价方法计算业务支撑度，对比评估结果与预期的质量需求目标，进行数据信息的判断，可依据指标结果分析数据质量存在的问题及原因，输出质量评估报告。

6.2 评价维度与指标

数据质量评价维度包括规范性、完整性、准确性、唯一性、时效性五个维度：

a） 数据规范性（A）：数据是否符合标准，数据规范性体现为数据格式、类型、值域和业务规则的有效性。

b） 数据完整性（B）：包括数据属性缺失和字段值缺失两部分。

c） 数据准确性（C）：与描述的客观实体是否一致，包括数据错误和数据异常两部分。数据异常主要指数据异常大、异常或数据值异常为零。

d） 数据唯一性（D）：主要用于衡量实体的重复性。

e） 数据时效性（E）:衡量数据时效是否符合用户需求，交通大数据中涉及的数据检测、存储、展示数据的时间属性，体现为数据更新及时性与数据校时准确性。

感知类交通大数据评价指标及计算方法如表 1 和表 2 所示。

表 1 数据质量评价指标

评价维度	一级指标	二级指标	类型
规范性（A）	值域及分布异常率	字典匹配异常率（A_1）	字段级
		取值范围异常率（A_2）	字段级
	类型及格式异常率	数据类型异常率（A_3）	表级
		字段格式及长度异常率（A_4）	字段级
		级联校验异常（A_5）	字段级
	命名实体不规范率	命名实体不规范率（A_6）	字段级
	安全规范异常率	安全规范异常率（A_7）	字段级
	字段不规范性率	字段不规范性率（A_8）	表级

表 1（续）

评价维度	一级指标	二级指标	类型
完整性（B）	字段空值率	字段空值率（B_1）	字段级
	数据记录缺失率	数据记录缺失率（B_2）	表级
	校验不完备率	参照校验/双向检验信息不完备（B_3）	表级
	字段完整率	字段不完整率（B_4）	表级
准确性（C）	数据量异常率	数据量同比骤升（C_1）	表级
		数据量同比陡降（C_2）	表级
	数据不合理率	数据不符合业务逻辑（C_3）	表级
		数据不符合常识（C_4）	表级
		数据校验不匹配率（C_5）	表级
		位置数据漂移率（C_6）	字段级
		其他自定义不合理校验（C_7）	表级
唯一性（D）	数据重复率	数据重复率（D_1）	表级
时效性（E）	时间异常比率	时间异常率（E_1）	表级
	数据延时上传率	数据延时上传率（E_2）	表级

表 2 数据质量评价指标计算方式

评价指标	计算方式	举例
字典匹配异常率（A_1）	定义字典的字段中非字典值行总数/字段总行数	如性别定义 1-男，2-女，0-未知，表中性别非字典值的比例
取值范围异常率（A_2）	字段取值范围超界行总数/字段总行数	如经度-180-180 之间，超出范围的比例
数据类型异常率（A_3）	物理表字段数据类型非标准类型/表总字段数	如过车时间字段的数据类型探查异常比例
字段格式及长度异常率（A_4）	字段格式和长度异常行总数/物理表字段总行数	如设备编号的格式和长度不符合规范比例
级联校验异常（A_5）	字段级联校验异常行总数/物理表字段总行数	如根据业务标准要求过车中点位编号前六位为区域编号，级联校验异常为区域编号非点位前六位的比率
命名实体不规范率（A_6）	命名实体字段中不规范行总数/该字段总行数	如号牌号码、手机号、身份证明号码、驾驶证号等命名实体规则不规范比率
安全规范异常率（A_7）	物理表不满足安全规范行总数/该字段总行数	安全规范是安全和隐私方面的规则，包括数据权限管理、数据脱敏处理等，如某字段为按照要求进行脱敏处理
字段不规范性率（A_8）	物理表不满足安全规范行总数/数据总行数	
字段空值率（B_1）	物理表字段空值行总数/物理表字段总行数	
数据记录缺失率（B_2）	表数据断流周期数/总周期数	如数据断流，一天划分 24 个时段，统计时段内无数据更新的时段比例；天气数据覆盖 365 天，缺失数据比例

表 2 （续）

评价指标	计算方式	举例
多表关联信息维护不全（B_3）	数据内容与参照内容相互不包含的行总数/数据总行数	验证校验数据包含在参照数据中的程度以及参照数据包含在验证数据中的程度，即检验校验数据和参照数据相互包含的程度。如过车中的点位编号关联点位信息，关联不到，该点位异常，计算点位异常比例
字段不完整率（B_4）	数据内容不符合字段完整性的行总数/数据总行数	
数据量同比骤升（C_1）	数据骤升周期数/总周期数	如一天划分 24 个时段，统计同时段不同周期同比数据量异常比例
数据量同比陡降（C_2）	数据陡降周期数/总周期数	如一天划分 24 个时段，统计同时段不同周期同比数据量异常比例
数据不符合业务逻辑（C_3）	数据校验不符合业务逻辑行总数/数据总行数	如过车数据一个设备关联多个方向，则该设备异常，计算设备异常比例
数据不符合常识（C_4）	数据校验不符合常识行总数/数据总行数	如过车数据时空不可达校验数据异常比例
数据校验不匹配率（C_5）	数据校验不匹配率行总数/数据总行数	如点位名称与点位经纬度信息不匹配数据异常比例
位置数据漂移率（C_6）	存在瞬时速度突变或距离漂移的行总数/数据总行数	如 GPS 漂移点比例
其他自定义不合理校验（C_7）	按照业务理解自定义不符合的行总数/数据总行数	如过车一天仅一条过车记录则根据业务判断数据异常，计算孤立过车记录数据异常比例
数据重复率（D_1）	数据重复行总数/数据总行数	如过车同一号牌同一设备相同或极短时间内两条及以上过车记录比例
时间异常率（E_1）	数据时间异常行总数/数据总行数	如设备校时不准导致的过车时间错误比例
数据延时上传率（E_2）	数据延时上传行总数/数据总行数	某过车延时上传数据比例

6.3 评价方法

6.3.1 重要字段分级定权重

根据数据表中关键字段，按照数据影响业务的重要性程度对数据字段进行分级定义，梳理待评价的数据项，数据项重要等级从高到低分别为一级、二级、三级。

表 3 数据项权重推荐值

字段等级	一级	二级	三级
权重	k_1	k_2	k_3
建议值	0.4-0.6	0.2-0.4	0.1-0.2
注：权重可根据业务与评价需要商定，权重递减。			

6.3.2 维度指标权重

根据实际业务与评价需求设定各维度与指标权重值。其中：

维度权重设置要求为：

$$w^a, w^b, w^c, w^d, w^e，\ w^a + w^b + w^c + w^d + w^e = 1。$$

指标权重设置要求为：

$$w_i^a, w_i^b, w_i^c, w_i^d, w_i^e；\ \sum_i w_i^a = 1；\ \cdots，\ \sum_i w_i^e = 1$$

6.3.3 质量评分

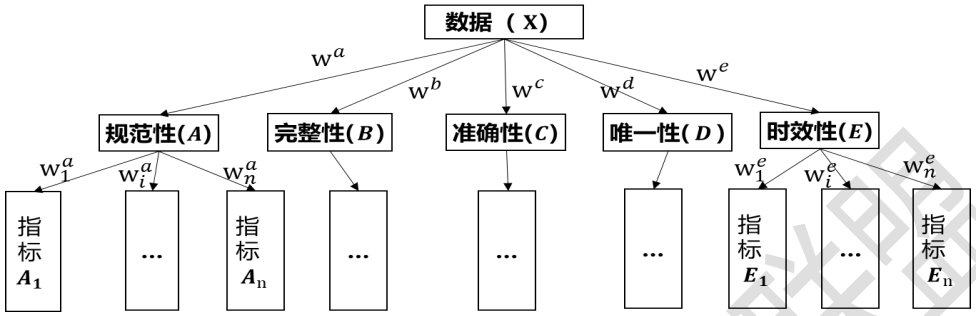

图 3 质量评分体系

综合质量得分值：

$$X = w^a \sum_i w_i^a A_i + w^b \sum_i w_i^b B_i + w^c \sum_i w_i^c C_i + w^d \sum_i w_i^d D_i + w^e \sum_i w_i^e E_i$$

指标得分值由指标计算公式得到，若单个指标涉及多个字段的数据评价内容，如指标字段空值率涉及到数据表中字段 1 的空值率、字段 2 的空值率，则根据字段等级权重计算指标字段空值率的综合得分。

6.4 评价等级

数据质量等级根据数据质量评分划分为质优、质中、质差，数据质量等级的划分规则按照表 4 的规定确定。

表 4 质量评价等级表

质量等级	质优	质中	质差
分值	[80, 100]	[60, 80)	[0, 60)

附录 A

（资料性附录）

A.1 过车数据

业务涉及过车数据需求字段：号牌号码、号牌种类、过车时间、设备编号、行政区划、数据来源。

表 A.1 过车数据评估指标

维度	指标	评价内容	类型
规范性	字典匹配异常率	号牌种类异常（枚举范围外）比率	字段级
		区域编号异常（枚举范围外）比率	字段级
		数据来源异常（枚举范围外）比率	字段级
	取值范围异常率	/	字段级
	数据类型异常率	数据类型非标准表类型比率	表级
	字段格式及长度异常率	过车时间格式不规范（字段标准）比率	字段级
		设备编号长度异常比率	字段级
	级联校验异常	设备编号前 6 位与区域编号不一致	字段级
	命名实体不规范率	号牌号码正则规范性检查	字段级
	安全规范异常率	/	字段级
	字段不规范性率	存在重要不规范行为的行/行总数	表级
完整性	字段空值率	号牌号码空值率	字段级
		号牌种类空值率	字段级
		过车时间空值率	字段级
		点位编号空值率	字段级
	数据记录缺失率	断流:过车数据设备小时断流比率	表级
	多表关联信息维护不全	不上数路口比率(信号路口关联过车无数据)	表级
		点位编号记录不完备率（点位编号关联点位信息取经纬度）	表级
	字段不完整率	/	表级
准确性	数据量骤升	过车数据骤升:同时段不同周期同比异常高异常时段比率	表级
	数据量陡降	过车数据骤升:同时段不同周期同比异常低异常时段比率	表级
	数据不符合业务逻辑	设备方向异常率(一个设备多个方向)	表级
	数据不符合常识	时空不可达过车比率	表级
	数据校验不匹配率	点位名称与经纬度不一致比率	表级
	位置数据漂移率	/	字段级
	其他自定义不合理校验	过车孤立点异常比率（一天仅一条过车比率）	表级
唯一性	数据重复率	过车记录重复比率	表级
时效性	时间异常比率	过车校时不准比率	表级
	数据延时上传率	过车延时上传比率	表级

A.2 设备数据

业务涉及设备数据需求字段：设备编号、设备名称、设备安装点位、经度、纬度、设备类型。

表 A.2 设备数据评估指标

维度	指标	评价内容	类型
规范性	字典匹配异常率	设备类型异常（枚举范围外）比率	字段级
	取值范围异常率	经度范围异常比率	字段级
		纬度范围异常比率	字段级
	数据类型异常率	数据类型非标准表类型比率	表级
		设备编号长度异常比率	字段级
	级联校验异常	设备编号前 6 位与区域编号不一致	字段级
	命名实体不规范率	/	字段级
	安全规范异常率	/	字段级
	字段不规范性率	存在重要不规范行为的行/行总数	表级
完整性	字段空值率	设备编号空值率	字段级
		经度空值率	字段级
		纬度空值率	字段级
	数据记录缺失率	/	表级
	多表关联信息维护不全	有过车记录无设备信息比率	表级
	字段不完整率	/	表级
准确性	数据量骤升	/	表级
	数据量陡降	/	表级
	数据不符合业务逻辑	/	表级
	数据不符合常识	/	表级
	数据校验不匹配率	点位名称与经纬度不一致比率	表级
	位置数据漂移率	/	字段级
	其他自定义不合理校验	/	表级
唯一性	数据重复率	设备记录重复比率	表级
时效性	时间异常比率	设备更新不及时比率	表级
	数据延时上传率	/	表级

A.3 车辆轨迹还原业务

车辆轨迹还原业务所需数据源：过车数据、设备数据。分别对过车数据和设备数据两类数据进行评价计算业务支撑度。

中国智能交通产业联盟
标准
城市交通大数据质量评价体系 设备感知类
T/ITS 0242-2023

北京市海淀区西土城路 8 号（100088）
中国智能交通产业联盟印刷
网址：http://www.c-its.org.cn

2024 年×月第一版　2024 年×月第一次印刷