

团体标准

T/ITS 0242-2024

城市交通设备感知类数据质量评价体系

Quality evaluation system of device perception data for urban traffic

2024-12-26 发布

2025-01-01 实施

中国智能交通产业联盟 发布

目 次

前 言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 数据范围 2

5 数据质量评价体系 2

6 评价流程 2

7 评价维度与指标 3

8 评价模型 5

附录 A 评价示例 8

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

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。

本文件主要起草单位：青岛海信网络科技股份有限公司、青岛市交通运输局、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。

本文件起草人：刘晓冰、于学秀、赵建龙、秦秀伟、高鹏、江筱薇、章庆、孟宪、孙永良、李建伟、时颖、刘爱华、王江涛、张珂、许晶玉、史世莲、谭磊、王成龙、刘祥。

城市交通设备感知类数据质量评价体系

1 范围

本文件规定了设备感知类城市交通大数据进行质量评价的体系框架、规范流程、评价维度与指标、评价方法以及评价等级的标准。

本文件适用于交通运输行业、交通管理部门开展城市交通设备感知类数据的质量评价以及业务应用层级数据支撑度评价。

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 29095-2012 道路交通管理数据字典

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

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

在指定条件下使用时，数据的特性满足明确的和隐含的要求的程度。

[来源：GB/T 36344-2018 2.2]

3.2

数据质量评价体系 data quality evaluation system

一套用于评价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

城市交通设备感知类数据 urban traffic device perception data

城市交通运行过程中设备设施收集到的交通行业专题数据，包括了交通检测器检测数据、交通事件检测数据、车载定位数据等。

4 数据范围

设备感知数据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行业专题数据。设备感知类数据包括基础信息与设备检测数据，按照来源包括了交通检测器检测数据、交通事件检测数据、车载定位数据，通过对城市交通设备感知类数据的挖掘和分析，可以再现城市交通系统运行特征。

- a) 交通检测器检测数据：包括电警、卡口、多目标雷达、微波等交通检测器基础信息以及设备安装点位信息；交通检测器检测数据为过车数据、交通流等设备检测数据。
- b) 事件检测数据：包括设备基础信息、设备安装点位信息；事件检测包括了声呐检测器、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。
- c) 车载定位数据：包括车载机信息、车辆信息、车载机与车辆关联信息；车载定位数据是车辆出行实时的记录信息。

5 数据质量评价体系

数据质量评价体系是从数据综合应用的角度考虑，从评价流程、评价维度和指标、评价模型三部分构建规范化数据质量评价体系。

评价流程部分从单一数据评价与业务支撑度两部分分别制定了规范的数据质量评价的规范化流程。

评价维度和指标部分从规范性、完整性、准确性、唯一性、时效性、一致性六个维度划分了 15 个一级指标、23 个二级指标。

评价模型部分对指标权重定义与设置方法、质量得分计算方式、评价等级划分标准进行了规范。

6 评价流程

数据质量评价一般步骤由单一数据评价与业务支撑度评价两部分组成，如图 1 所示。

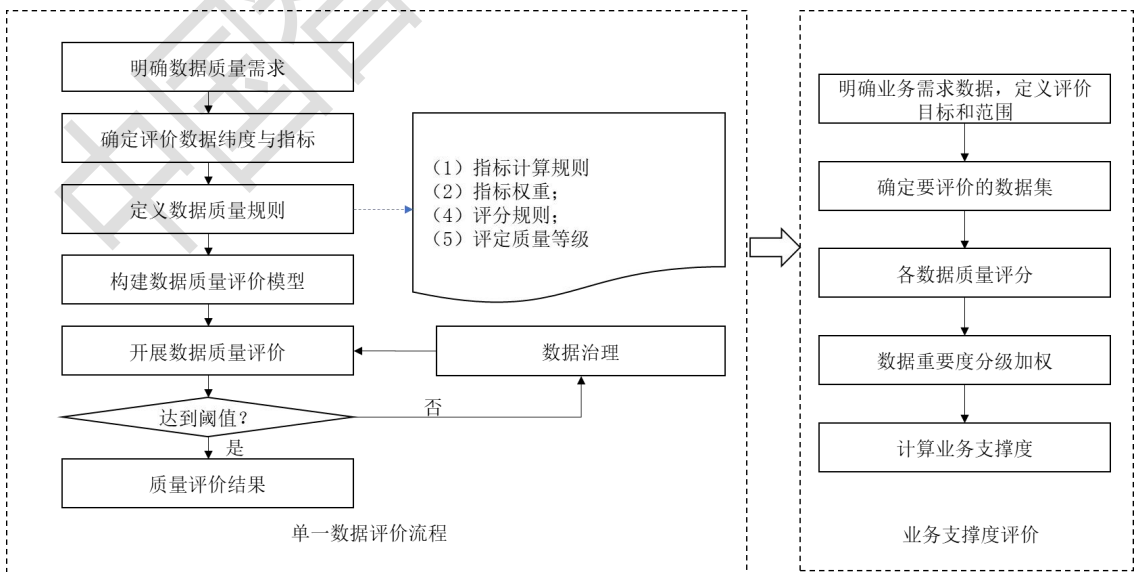


图 1 数据质量评价流程图

6.1 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，定义评价目标和范围。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) 确定评价数据纬度与指标，依据数据评价目标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) 定义质量评价规则，包括指标计算规则、规则权重、评分规则、评价质量分级。其中指标计算规则定义了各指标的计算方法；各指标权重根据业务需要定义具体数值，评分规则定义了数据质量综合评价得分的计算方法，评定质量分级定义了数据质量的三等级标准，明确了各等级的划分依据。
- d) 构建数据质量评价模型。
- e) 开展数据质量评价，按照既定方法和规则开展数据质量评价。
- f) 输出质量评价结果，对比评价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，输出质量评价报告。

6.2 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，定义评价目标和范围。根据业务需求，明确开展某些业务所需数据项以及需求目标，明确业务范围。
- b) 确定要评价的数据集，根据业务需求和范围确认涉及的数据集、数据集的时空范围。
- c) 各数据质量评分，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) 数据重要度分级加权，根据该业务对不同数据集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，对比评价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，输出质量评价报告。

7 评价维度与指标

数据质量评价包括规范性、完整性、准确性、唯一性、时效性、一致性六个维度：

- a) 数据规范性（*A*）：数据符合数据标准、业务规则与元数据的程度，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*）：按照数据规则要求，数据元素被赋予数值的程度。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、异常小或数据值异常为零的趋势异常；数据错误是数据描述脱离真实值。
- d) 数据唯一性（*D*）：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，城市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。
- f) 数据一致性（*F*）：数据与其他上下文使用数据无矛盾的程度。

数据评价指标及计算公式如表 1 和表 2 所示，附录 A 为过车数据、设备数据以及车辆轨迹还原业务的指标选取示例。

表 1 数据质量评价指标

	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
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
		自定义业务规则不合理校验 (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
一致性 (F)	相同数据一致率	相同数据一致率 (F ₁)	表级
	关联数据一致率	关联数据一致率 (F 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	如过车数据上传设备类型取值为字典枚举值，指标为非字典枚举值的比例。
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	如车载定位数据经纬度字段取值-180-180 之间，指标为超出范围的比例；
数据类型异常率 (A ₃)	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	如过车数据中的设备编号与设别信息表级联校验异常的比例。
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、设备安装点位编号等命名实体规则不规范比例。

表 2（续）

评价指标	计算方式	举例
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。
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	如流量数据重要字段车流量缺失的比率。
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。
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，计算点位缺失比例。
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	如过车表缺失号牌号码、号牌类型等重要字段信息。
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	如车载定位数据漂移点比例。
自定义业务规则不合理校验 (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	如过车延时上传数据比例。
相同数据一致率 (F_1)	同一数据满足一致性的数据行数/数据总行数	如同一数据在不同位置存储或被不同应用或用户使用数据一致性；数据发生变化时，存储在不同位置的同一数据被同步修改。
关联数据一致率 (F_2)	关联数据满足一致性的数据行数/数据总行数	如一致约束规则检查过车数据与设备基本信息关联数据存在一致性的比例。

8 评价模型

8.1 维度指标权重

同维度计算时权重相加必须等于 100%，同级指标计算时权重相加必须等于 100%或 0，评价模型中权重设置标识及定义如下：

维度权重设置要求见公式（1）

$$w^a + w^b + w^c + w^d + w^e + w^f = 1 \cdots \cdots \cdots (1)$$

式中：

$w^a, w^b, w^c, w^d, w^e, w^f$ 为各维度权重。

指标权重设置要求见公式（2）

$$\sum_i w_i^a = 1 \text{ 或 } 0, \cdots, \sum_i w_i^f = 1 \text{ 或 } 0 \cdots \cdots \cdots (2)$$

式中：

w_i^a, \dots, w_i^f 为各指标权重。

权重设置方法通过专家打分法确定，具体步骤如下：

- a) 选择专家，选取的专家应当熟悉当前业务与数据基本信息；
- b) 按照评价体系，设计权重设置意见征询表进行征询；
- c) 专家意见分类汇总形成行程最终结论。

8.2 质量得分计算方式

综合质量得分值由各评价维度与指标得分加权平均得出，见公式（3）

$$X = w^a \sum_i w_i^a A_i + w^b \sum_i w_i^b B_i + w^c \sum_i w_i^c C_i + w^d \sum_i w_i^d D_i + w^e \sum_i w_i^e E_i + w^f \sum_i w_i^f F_i \cdots \cdots (3)$$

式中：

$\sum_i w_i^a A_i$ 为规范性得分， w^a 为规范性权重；

$\sum_i w_i^a A_i$ 为完整性得分， w^a 为完整性权重；

$\sum_i w_i^a A_i$ 为准确性得分， w^a 为准确性权重；

$\sum_i w_i^a A_i$ 为唯一性得分， w^a 为唯一性权重；

$\sum_i w_i^a A_i$ 为时效性得分， w^a 为时效性权重；

$\sum_i w_i^a A_i$ 为一致性得分， w^a 为一致性权重。

8.3 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 3 的

规定确定。

表 3 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)

中国智能交通产业联盟

附录 A 评价示例

(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 过车数据评价指标

评价维度	评价指标	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
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	过车数据骤降：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
	自定义业务规则不合理校验	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
一致性	相同数据一致率	相同数据一致率	表级
	关联数据一致率	关联数据一致率	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 设备数据评价指标

评价维度	评价指标	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
	自定义业务规则不合理校验	/	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
一致性	相同数据一致率	相同数据一致率	表级
	关联数据一致率	关联数据一致率	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。车辆轨迹还原业务支撑度由业务关联的过车数据和设备数据进行综合评价。

中国智能交通产业联盟

标准

城市交通设备感知类数据质量评价体系

T/ITS 0242-2024

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>

2025 年 1 月第一版 2025 年 1 月第一次印刷