



基于 IOWA 算子的 RFID 数据有效性评价方法

摘要

智能交通系统应用的迅猛发展需要高质量的交通数据,而交通数据采集技术采集到的原始数据会不可避免地混入异常数据,降低数据的有效性,从而可能引起系统应用性能的降低.本文针对基于无线射频技术(RFID)的交通数据,设计了一套交通数据评价指标,并根据交通数据评价指标的特点,综合评价权重向量计算方法的适用条件及优缺点,建立了基于 IOWA (Induced Ordered Weighted Averaging) 算子的交通数据综合评价方法.基于实际数据的分析结果表明,该方法可以对 RFID 交通数据进行有效性评价.

关键词

RFID 数据;数据采集;数据评价指标;IOWA 算子;评价

中图分类号 O429

文献标志码 A

收稿日期 2019-06-30

资助项目 国家自然科学基金(61573106)

作者简介

龙思卿,男,硕士生,研究方向为交通信息工程与控制.1067487132@qq.com

郭建华(通信作者),男,博士,教授,博士生导师,研究方向为交通信息工程与控制.seugh@163.com

0 引言

智能运输系统(ITS)离不开交通数据,然而长期以来,交通数据的有效性一直受到质疑.由于各种 ITS 应用的迅猛发展,特别是向公众提供交通信息服务的出现,使得数据有效性问题变得更为迫切,对交通数据质量的需求变得更为复杂.一般情况下,使用交通数据采集技术采集到的原始数据会不可避免地混入来源各异的各种噪声,降低数据的有效性,从而对智能交通系统中的交通预测和交通管控等应用带来较大影响,因此需要对采集到的交通数据进行有效性评价.

数据质量有效性评价出现于 19 世纪末,其概念最初源于经济领域,以数据满足特定用户期望的程度来定义^[1].而后数据质量评价逐步扩展成多维概念,针对不同维度,Hermans 等^[2]首先利用数据准确性、数据完整性、数据一致性和数据时效性 4 个维度来对动态数据质量进行指标定义,并将该思路扩展到多个领域.随后 Turner 等^[3]对不同的交通数据使用者及其数据使用需求进行研究,在 6 个评价指标的基础上,提出了数据质量评价方法的指标详细度和交通数据质量综合评价方法的概念,并将数据质量定义为适用于所有需求的数据.对数据质量进行评价,首先需要了解该数据的所有预期用途^[4].通过分析国内外数据评价的研究现状发现,有效性评价包括微观评价、宏观评价、综合评价等方法,其中综合评价主要针对多属性决策问题,对不同属性进行评价指标计算,覆盖范围广、实用性高,使用广泛.

在交通数据处理方面,耿彦斌等^[5]针对 ITS 数据中的错误数据、缺失数据和不精确数据,采用交通流理论的方法建立了问题数据判别规则.袁高峰等^[6]针对南京市区内采集到的 RFID (Radio Frequency Identification, 无线射频技术)交通数据,提出基于统计的 RFID 数据匹配有效性检验方法,可有效检测出 RFID 匹配异常数据.杜威等^[7]、杨越思等^[8]对由 RFID 技术采集的交通数据进行数据质量控制,对 RFID 错误数据和冗余数据进行检测和统计.杨越思等^[9]分析了 RFID 数据特性,定义 RFID 交通数据时间覆盖率,分别计算 7 种不同时间汇集度下的时间覆盖率,用于数据有效性和交通状态的评价.

在自主式综合评价方法方面,其基本思想认为评价主体自身具有价值认识、利益诉求和竞争意识.在此假设条件下,评价主体会自发地将自身的优势属性权重最大化,同时将自身劣势属性权重最小化,以期在综合评价中得分较高、排序较前.自主式综合评价又分为主观

¹ 东南大学 智能运输系统研究中心,南京,210018

自主式和客观自主式综合评价两类.主观自主式指评价主体具有分析能力,并提供除客观信息外的主观有利信息;客观自主式指评价主体不具有分析能力,仅提供客观信息^[10].在交通参数的一致性检验方面,速度流量关系公式检验和交通数据曲线拟合方法^[11]是两种常用方法.本文将针对 ITS 的数据有效性需求,采用基于 IOWA (Induced Ordered Weighted Averaging)算子的方法,建立针对 RFID 数据的有效性综合评价方法.

1 RFID 数据采集方法

RFID 交通数据采集系统主要包括前端数据采集系统及后台数据处理系统两个部分.前端数据采集系统主要由电子标签、读卡器及 RFID 天线三部分组成.当装有电子标签的车辆进入 RFID 天线发射的磁场范围后,接收读卡器发出的射频信号,读卡器会从电子标签中读取存储的车辆信息并解码,将信息上传至信息处理中心进行处理.系统结构如图 1 所示.

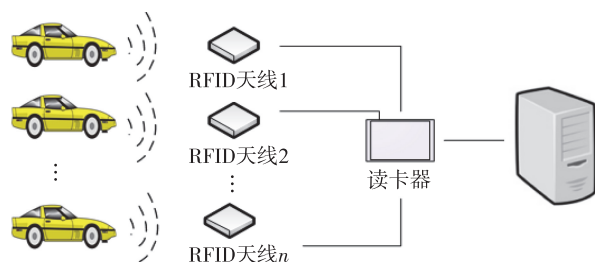


图 1 RFID 交通数据采集系统

Fig. 1 RFID traffic data acquisition system

RFID 基站采集的原始数据中主要包括车辆信息和通过信息.车辆信息主要是车辆的基本属性信息,如车牌号、车型、车身颜色等;通过信息是车辆的动态信息,主要包括车辆通过基站的时间、开始读卡时间、结束读卡时间等.可对 RFID 技术采集到的原始数据进行汇集,得到交通量及行程时间等.根据应用的目的,可以得到任意时间汇集度的交通量数据和行程时间数据.

2 RFID 数据评价指标

RFID 数据评价指标可以分为 5 个部分,分别是数据完整性、数据准确性、数据可用性、数据稳定性和数据一致性,叙述如下.

1) 数据完整性

RFID 数据的完整性,主要通过检测交通缺失数

据,统计不同缺失类型的数据量,进而计算得到不同缺失情况下的数据缺失率.RFID 交通数据缺失分为完全缺失和部分缺失两种类型.对于完全缺失的 RFID 数据,其完整性计算步骤为:首先统计时间汇集度为 15 min 的缺失率,分析其在 15 min 时间汇集度下的数据完整性,然后统计应有数据量和实际数据量,计算数据缺失率.对于部分缺失的 RFID 数据,其完整性计算步骤为:首先分析 RFID 交通数据,统计速度非空的数据量和数据总量,然后计算速度缺失率.

2) 数据准确性

RFID 数据的准确性,主要指 RFID 数据是否满足数据标准和交通参数的正常范围,即对数据集中存在的错误数据进行检测与数据统计.错误数据类型主要包括车牌错误及交通参数错误.车牌信息错误主要包括车牌汉字识别错误、字母错误及车牌号字符长度错误,然后根据车牌的正则表达式对错误车牌进行统计,计算得到车牌错误率.交通参数错误主要是针对交通流量和车辆速度参数,通过设定交通流量和速度阈值,检测阈值外的不准确数据,即计算交通参数平均值 v 和方差 σ ,当交通数据不在 $[v - 3\sigma, v + 3\sigma]$ 范围内时,则认为数据是错误的.

3) 数据可用性

RFID 交通数据可用性,是指通过检测与统计数据中存在的非法或异常的不可用数据,来评价交通数据集的可用性.一般采用箱型图对异常值检测.箱型图依据实际交通数据进行绘制,表现数据直观真实,同时箱型图以上四分位数 (Q_1)、下四分位数 (Q_3) 和四分位距 ($Q_3 - Q_1$) 为基础,对异常数据进行识别,具有一定的鲁棒性.交通数据异常值被定义为大于 $Q_1 + 1.5(Q_3 - Q_1)$ 或小于 $Q_3 - 1.5(Q_3 - Q_1)$ 的值,统计异常值数量后可计算异常率.

4) 数据稳定性

RFID 数据稳定性指交通流和速度在纵向时间上的相对稳定性,即通过对比不同日期相同时间段的交通参数的波动性来刻画其稳定性.数据稳定性评价指标是数据的方差值,计算步骤为:首先将交通数据按时间间隔进行排序和分组,其次将同一个时间间隔内的数据分为同一个小组,以 15 min 为汇集度,则每个基站共可得到 96 组数据,然后对每组数据求其交通参数的均方差值 (std),最后采用 96 组数据均方差的最大值、最小值、平均值来评价该基站数据的稳定性,即数据的 $\max(\text{std})$ 、 $\min(\text{std})$ 和

mean(std).

5) 数据一致性

根据交通参数之间相关性的大小,对交通状态和车辆速度的一致性进行判断,主要考虑道路流量和车辆速度是否在宏观上保持一致性.若该基站交通数据符合交通参数数据一致性规则,则认为该基站一致性为1,若不符合一致性规则,则为0.数据一致性的判断步骤为:首先汇集15 min的交通量数据,然后构建流量-速度($v-s$)散点图拟合曲线,再构建速度 v 自相关曲线,最后分析两张拟合曲线图的一致性,得出结论.

3 数据评价方法

3.1 数据评价方法简介

自主式综合评价主要是对评价对象参与评价或竞争优化等问题进行研究,包括两个方面:一是在评价意志约束的前提下,突出被评价对象的竞争优势;二是在数据包络(Data Envelop Analysis, DEA)模型中融入各决策单元之间的竞争.本文采用客观自主式综合评价方法对交通数据有效性进行评价.

1) 自主式综合评价的规则

在自主式综合评价中,被评价对象基于自身利益出发,尽量提升自身竞争优势,同时尽量打压竞争对手的竞争优势.基于这一前提,对自主式综合评价进行规则界定.规则一是竞争视野优化.评价主体按照一定的方法确定其竞争视野,在此基础上采取一定的策略,达到优化竞争视野的目的,其中,优化竞争视野的策略通常有赋权策略、评价信息主观打分策略等.规则二是权数非独裁性条件.当评价主体通过赋权策略实现竞争视野优化的目的时,若对其赋权行为不进行约束,则它将会对自己处于绝对优势的指标赋权为1,剩余指标赋权为0,从而出现独裁的现象.为避免这一问题,需要对评价主体的赋权行为进行非独裁性约束.

2) 自主式综合评价方法的基本过程

自主式综合评价方法的基本过程包括如下5个步骤:

步骤1:明确评价目的和对象,确定评价对象范围与评价对象的边界.

步骤2:收集评价信息,构建评价指标体系,评价对象指标值求取,指标类型一致化和无量纲化.

步骤3:确定竞争视野.依据指标值矩阵或评价值矩阵,确定评价主体与被评价对象构成竞争关系

的被评价对象集合.

步骤4:优化竞争视野.被评价对象通过一定的赋权策略提升自身的评价值,同时打压其竞争对手的评价值.

步骤5:再次集结优势评价信息.对各评价主体给出的评价信息(包括各被评价对象的评价值或优势权重向量等),再次集结后得到评价结果,并据此对各评价对象进行排序、选择或分类.

3.2 竞争视野及指标优势度

在客观自主式综合评价中,易平涛等^[12]、林德花等^[13]比较分析了各被评价对象指标取值的差异分布情况,分别采用建立指标绝对优势度和相对优势度来优化竞争视野的方法.他们认为存在 n 个被评价对象,集合表示为 $O = \{o_1, o_2, \dots, o_n\}$,存在 m 个评价指标,集合表示为 $X = \{x_1, x_2, \dots, x_m\}$.被评价对象 o_i 在指标 x_j 下的取值记为 $x_{ij} = x_j(o_i)$ ($i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$),则被评价对象 o_i 的指标值集为 $X_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$.

记 $N = \{1, 2, \dots, n\}$, $M = \{1, 2, \dots, m\}$ (设 $m, n \geq 3$).客观自主式综合评价之前,对指标进行一致化和无量纲化处理,预处理后的指标值集合仍标记为 $X_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$ ($i \in N$).

1) 竞争视野

设 o_i ($i \in N$)为评价主体,在进行客观自主式综合评价时,首先要确定其竞争视野,即与评价主体 o_i 构成竞争关系的其他被评价对象组成的集合.具体方法如下:若对于 $j \in M$,恒有 $x_{ij} \geq x_{kj}$ ($k \in N, k \neq i$),则第 i 个被评价对象 o_i 优于第 k 个被评价对象 o_k ,两者不存在竞争关系.若对于 $j \in M$,恒有 $x_{ij} \leq x_{kj}$ ($k \in N, k \neq i$),则第 i 个被评价对象 o_i 劣于第 k 个被评价对象 o_k ,两者不存在竞争关系.若对于 $j \in M$,有 $x_{ij} \geq x_{kj}$ 和 $x_{ij} \leq x_{kj}$ ($k \in N, k \neq i$)同时成立,则第 i 个被评价对象 o_i 和第 k 个被评价对象 o_k 存在竞争关系.

对于评价主体 o_i ($i \in N$),通过比较得到所有与其存在竞争关系的被评价对象,这样的被评价对象组成的集合称为 o_i 的竞争视野,记为 $C_i = \{o_1^{(i)}, o_2^{(i)}, \dots, o_{n_i}^{(i)}\}$,其中 n_i 表示与评价主体 o_i ($i \in N$)存在竞争关系的被评价对象的数量.

2) 指标优势度

对于评价主体 o_i ($i \in N$),设其竞争视野为 $C_i = \{o_1^{(i)}, o_2^{(i)}, \dots, o_{n_i}^{(i)}\}$,记 $N_i = (1, 2, \dots, n_i)$.通过分别比较评价主体与其竞争视野内被评价对象的指标取值的差异,对评价主体 o_i ($i \in N$)的指标优势度进行

确定,为优化竞争视野提供基础.

定义 1 对于评价主体 $o_i (i \in N)$, 其竞争视野为 $C_i = \{o_1^{(i)}, o_2^{(i)}, \dots, o_{n_i}^{(i)}\}$, 称 $d_{il}^{(j)}$ 为评价主体 o_i 在第 j 个指标 x_j 上相对于竞争视野内被评价对象 o_l 的竞争强度:

$$2d_{il}^{(j)} = x_{ij} - x_{lj}, \quad i \in N, \quad l \in N_i, \quad j \in M. \quad (1)$$

若 $d_{il}^{(j)} > 0$, 表明评价主体 o_i 在第 j 个指标 x_j 上的竞争强度优于 o_l ; 若 $d_{il}^{(j)} = 0$, 表明评价主体 o_i 在第 j 个指标 x_j 上的竞争强度与 o_l 相同; 若 $d_{il}^{(j)} < 0$, 表明评价主体 o_i 在第 j 个指标 x_j 上的竞争强度劣于 o_l .

定义 2 对于评价主体 $o_i (i \in N)$, 其竞争视野为 $C_i = \{o_1^{(i)}, o_2^{(i)}, \dots, o_{n_i}^{(i)}\}$, 称 $\lambda_i^{(j)}$ 为关于评价指标 x_j 、评价主体 o_i 的绝对优势度:

$$\lambda_i^{(j)} = \sum_{m=1}^{k_i} d_{im}^{(j)} / \sum_{l=1}^{n_i} |d_{il}^{(j)}|, \quad (2)$$

其中, k_i 表示评价主体 o_i 相对于竞争视野内其余所有评价对象, 竞争强度值为非负的数量.

定义 3 对于评价主体 $o_i (i \in N)$, 其竞争视野为 $C_i = \{o_1^{(i)}, o_2^{(i)}, \dots, o_{n_i}^{(i)}\}$, 称 $\lambda_i^{(j)}$ 为关于评价指标 x_j 、评价主体 o_i 上的相对优势度:

$$\lambda_i^{(j)} = e^{\sum_{l=1}^{n_i} d_{il}^{(j)}} / \sum_{l=1}^{n_i} e^{\sum_{l=1}^{n_i} d_{il}^{(j)}}, \quad (3)$$

其中, $\sum_{l=1}^{n_i} d_{il}^{(j)}$ 为第 i 个评价主体 ($i \in N$) 在第 j 个评价指标 ($j \in M$) 上的整体竞争强度, $\sum_{l=1}^{n_i} d_{il}^{(j)} \in (-\infty, +\infty)$.

3.3 IOWA 算子及位置加权向量

1998 年, Filev 等^[14] 提出有序加权平均 (OWA) 算子, 1999 年, Yager 等^[15] 进一步提出诱导有序加权 (IOWA) 算子. 本文以评价主体对于各评价指标的绝对优势度为第一诱导分量, 以评价主体对于各评价指标的相对优势度为第二诱导分量, 对评价对象下的各评价指标进行优势度重新排序, 求得指标位置权重, 并对评价信息集结.

定义 4 对评价主体 $o_i (i \in N)$, 称 Y_k 为在评价主体 o_i 下, 第 k 个被评价对象 o_k 的评价值:

$$Y_k(\lambda_i^{(1)}, \lambda_i^{(1)}, x_{k1}, \lambda_i^{(2)}, \lambda_i^{(2)}, x_{k2}, \dots, \lambda_i^{(m)}, \lambda_i^{(m)}, x_{km}) = \sum_{j=1}^m \omega_j a_{kj} \quad (k \in N), \quad (4)$$

其中, $\omega = (\omega_1, \omega_2, \dots, \omega_m)^T$ 是与 Y_k 相关联的位置

加权向量, $\omega_j \in (0, 1)$, $\sum_{j=1}^m \omega_j = 1$ 包含 3 个元素, 诱导分量 $\lambda_i^{(j)}$ 和 $\lambda_i^{(j)}$ 分别表示评价指标 x_j 的第一重要性和第二重要性; a_{kj} 是评价指标重新排序后的第 j 个重要评价指标在被评价对象 o_k 下的取值 ($j \in M$). 依据诱导分量 $\lambda_i^{(j)}$ 和 $\lambda_i^{(j)}$ ($j \in M$) 对评价主体 $o_i (i \in N)$ 的指标集 $\{x_2, \dots, x_m\}$ 进行重新排序, 规则如下: 当 $\lambda_i^{(j)} > \lambda_i^{(j+1)}$ 或 $\lambda_i^{(j)} = \lambda_i^{(j+1)}$, $\lambda_i^{(j)} > \lambda_i^{(j+1)}$ 时, 指标 x_j 的重要性大于指标 x_{j+1} ($j = 1, 2, \dots, m-1$).

定义 5 对评价主体 $o_i (i \in N)$, 设其位置加权向量为 $\omega = (\omega_1, \omega_2, \dots, \omega_m)^T$, 则 $\omega_j (j \in M)$ 的表达式为 $\omega_j = q \sum_{k=1}^j \eta_k / \sum_{k=1}^m \eta_k$, 式中, $0 < q < 1$, $\eta_k = 1 - (\alpha \lambda_i^{(k)} + \beta \lambda_i^{(k)})$ ($k \in M$), $\alpha \lambda_i^{(k)} + \beta \lambda_i^{(k)}$ 为第 k 个评价指标的竞争优势; α, β 为评价者对指标绝对优势度和指标相对优势度的偏好程度, 满足 $\alpha + \beta = 1$, $\alpha, \beta \in [0, 1]$. α, β 的值根据评价者偏好可事先确定, 无特殊情况令 $\alpha = \beta = 0.5$.

定义 6 对于位置加权向量 $\omega = (\omega_1, \omega_2, \dots, \omega_m)^T$, 可用参数 O_s 表示其 orness 度为

$$O_s = \frac{1}{m-1} \sum_{j=1}^m [(m-j)\omega_j]. \quad (5)$$

当 $0 < q < 1$ 时, $O_s > 0.5$; 当 $q = 1$ 时, $\omega = (1/m, 1/m, \dots, 1/m)^T$, $O_s = 0.5$; 当 $q > 1$ 时, $O_s < 0.5$. 基于值优化的自主式综合方法评价, 其评价主体将自身评价价值最大化, 同时将竞争对手的评价价值最小化, 因此评价主体在进行位置加权向量的确定时, 会自主偏向优势指标, 并赋予优势指标较大的权重, 故规定 $0 < q < 1$.

3.4 信息集结及方法步骤

设第 i 个评价主体 $o_i (i \in N)$ 下的各评价对象的评价值组成的评级向量为 $y^i = (y_1^i, y_2^i, \dots, y_n^i)^T$, 则所有评价向量组成的评价值矩阵为 $Y = (y^{(1)}, y^{(2)}, \dots, y^{(n)})$.

定理 1 对于 $\forall y \in \mathbf{R}^n$,

$$\max_{\|y\|_2=1} \sum_{i=1}^n y^T y^{(i)} = \sum_{i=1}^n [(y^*)^T y^{(i)}]^2 = \lambda_{\max}, \quad (6)$$

其中, λ_{\max} 为实对称矩阵 YY^T 的最大特征根, $Y = (y^{(1)}, y^{(2)}, \dots, y^{(n)})$, y^* 为 λ_{\max} 对应于 YY^T 的正特征向量, 且 $\|y\|_2 = 1$. 对各评级主体给出的评价值信息进行再集结, 得到最终的评价值向量 $Y^* = (y_1^*, y_2^*, \dots, y_n^*)^T$, 并以此对各评价对象进行比较判断.

综上, 数据评价方法的步骤为: 首先, 确定不同

评价主体的竞争视野,确定该评价主体下,各评价指标的绝对优势度和相对优势度;其次,按照IOWA算子的思想,对各评价主体下的评价指标进行重要度重新排序;然后,确定各评价主体下的各指标位置加权向量,并将其与重新排序后的指标进行信息集结,得到各评价主体的评价值向量;最后,对由所有评价主体的评价值向量组成的评价值矩阵 Y ,依据定理1对评价值矩阵 Y 进行集结,得到最终评价值 $Y^* = (y_1^*, y_2^*, \dots, y_n^*)^T$.

4 数据评价实例

4.1 数据采集

针对南京市安装的RFID检测基站,选择南京市龙蟠路、中山东路、明故宫路等交叉口位置的4个RFID基站采集的交通数据进行研究分析,基站具体信息如表1所示.数据采集区间为2014年3月1日—2014年3月31日.

4.2 评价指标标定

1) RFID数据完整性标定

表1 选择的RFID基站
Table 1 Selected RFID base stations

基站编号	基站位置
6027	龙蟠路中山东路以北南侧
6149	明故宫路中山东路以南西侧
6150	明故宫路中山东路以北东侧
6151	北安门街珠江路以南西侧

对采集到的RFID原始数据进行车牌缺失统计,并进行15 min时间段的汇集,统计汇集后的数据缺失量和数据缺失率.15 min汇集度下的RFID数据应有数据量 $31 \times 24 \times 4 = 2976$,数据缺失情况如表2所示,可见数据量缺失均小于1%(表2、图2).

2) RFID数据准确性标定

计算采集到的RFID数据准确性,得到RFID数据的错误量和错误率如表3和图3所示.

3) RFID数据可用性标定

RFID数据可用性标定主要是计算数据的异常值,即检测交通流量异常值和速度异常值.RFID交通数据的异常值检测,主要采用箱型图的方法检测

表2 RFID交通数据缺失概况

Table 2 Overview of missing RFID traffic data

基站	15 min 汇集度下数据			车牌号			车辆速度	
	缺失量	数据量	缺失率/%	缺失量	数据量	缺失率/%	数据量	缺失率/%
6027	22	2954	0.74	40597	558229	6.78	598826	0
6149	14	2962	0.47	10877	378642	2.79	389519	0
6150	16	2960	0.54	10439	324361	3.12	334800	0
6151	17	2959	0.57	7873	318775	2.41	326648	0

表3 RFID交通数据错误概况

Table 3 Overview of RFID traffic data errors

基站	车牌号			车辆速度(15 min)			交通流量(15 min)		
	数据量	错误量	错误率/%	数据量	错误量	错误率/%	数据量	错误量	错误率/%
6027	598826	113	0.0189	2954	16	0.54	2954	15	0.51
6149	389519	6	0.0015	2962	16	0.54	2962	21	0.71
6150	334800	6	0.0018	2960	21	0.71	2960	19	0.64
6151	326648	1	0.0003	2959	35	1.18	2959	22	0.74

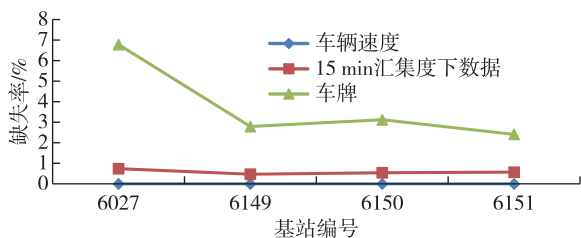


图2 RFID数据缺失率

Fig. 2 Missing rate of RFID traffic data

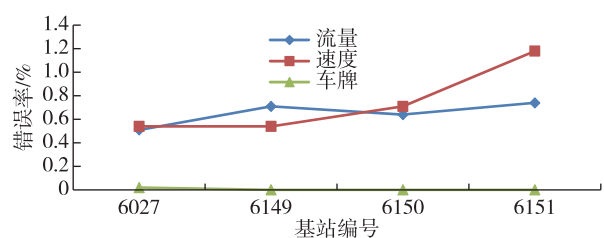


图3 RFID数据错误率

Fig. 3 RFID traffic data error rate

出异常,并对异常值进行统计.通过箱型图得到各基站 RFID 数据异常值如图 4 和表 4 所示.

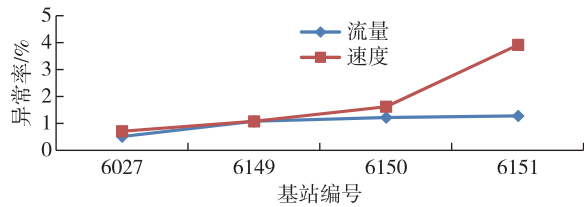


图 4 RFID 数据异常率

Fig. 4 Abnormal rate of RFID data

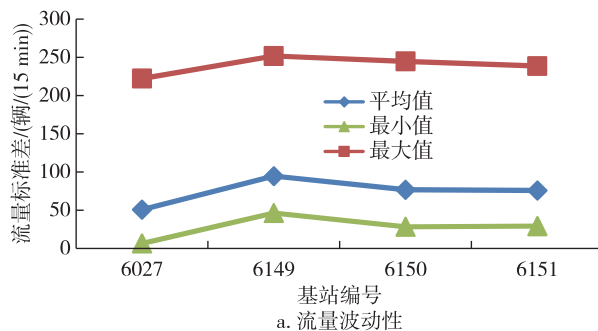
表 4 RFID 交通数据异常概况

Table 4 Overview of abnormal RFID traffic data

基站	车辆速度(15 min)			交通流量(15 min)		
	数据量	异常量	异常率/%	数据量	异常量	异常率/%
6027	2 954	21	0.71	2 954	15	0.51
6149	2 962	32	1.08	2 962	32	1.08
6150	2 960	48	1.62	2 960	36	1.22
6151	2 959	116	3.92	2 959	38	1.28

4) RFID 数据稳定性标定

本文采用不同日期同一时间间隔的数据均方差



a. 流量波动性

值判断数据的稳定性,即对每个基站的 96 个时间间隔,提取其均方差最大值、均方差最小值和均方差平均值,从而整体评价基站数据的稳定性.计算各基站 RFID 数据稳定性如图 5 和表 5 所示.

表 5 RFID 交通数据波动性概述

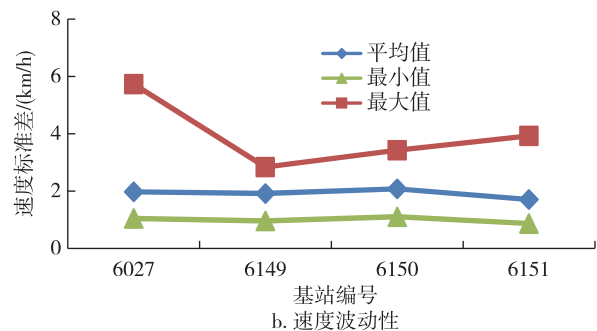
Table 5 Overview of volatility of RFID traffic data

基站	速度均方差/(km/h)			交通流量均方差/(辆/(15 min))		
	平均值	最大值	最小值	平均值	最大值	最小值
6027	1.98	5.73	1.05	51.02	222.18	6.82
6149	1.92	2.84	0.96	94.70	251.67	46.47
6150	2.08	3.43	1.11	77.02	244.71	28.50
6151	1.71	3.93	0.87	76.07	238.64	29.45

5) RFID 数据一致性标定

交通参数宏观一致性是对基站交通参数数据的宏观把握,本文根据交通流理论制定交通参数宏观一致性判断规则,判断 RFID 交通数据的宏观一致性,具体结果如表 6 和图 6 所示.

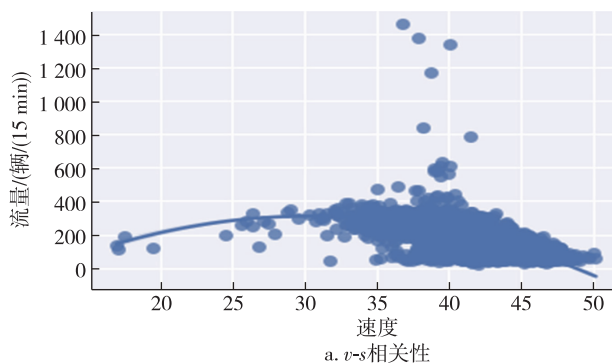
基站 6027 中,速度-流量($v-s$)相关性为 0.6,散点图中随速度增加,流量减少,速度受交通流量负面影响,但速度集中在 40 km/h,认为此时还未处于拥



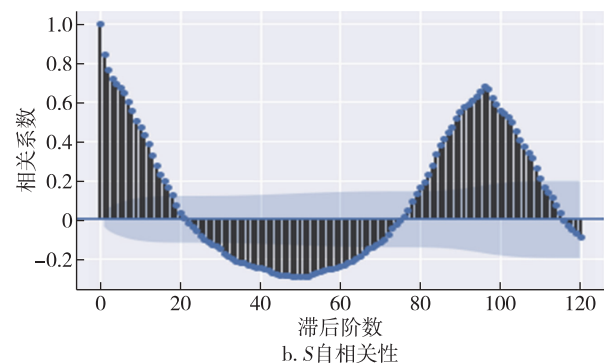
b. 速度波动性

图 5 RFID 数据波动性

Fig. 5 Volatility of RFID data



a. $v-s$ 相关性



b. S自相关性

图 6 基站 6027 交通参数宏观一致性

Fig. 6 Macro consistency of traffic parameters of base station 6027

表 6 RFID 交通参数宏观一致性

Table 6 Macro consistency of RFID traffic parameters

基站	v-s 相关性	趋势一致性
6027	-0.6	1
6149	-0.4	1
6150	-0.2	1
6151	-0.3	1

堵状态,速度自相关性较大;基站 6150 中,速度-流量相关性为 0.2,速度主要集中在 30 km/h,此时交通流处于拥堵状态,车辆走走停停,速度自相关较小。

4.3 基于 IOWA 算子的客观自主式道路数据评价

客观自主式综合评价最大的优点是可以根据各评价主体下评价指标的重新排序及其优势度,或通过各评价主体的位置加权向量来判断评价主体的主要优势和主要劣势,并结合交通数据使用需求,进而判断数据的利用价值。

1) 评价指标重新排序及其优势度

计算道路交通数据有效性指标的绝对优势度和相对优势度,并根据绝对优势度和相对优势度,对每个评价主体下的指标进行重新排序。以基站 6149 为例,数据有效性评价指标优势度示例如表 7 所示。

表 7 数据有效性评价指标重新排序及优势度

Table 7 Data validity evaluation index reordering and dominance degree

基站	评价值 i	λ_1	λ_2
6149	4	1.00	0.02
6149	13	1.00	0.02
6149	1	0.99	0.33
6149	11	0.95	0.54
6149	6	0.71	0.03
6149	0	0.66	0.01
6149	12	0.61	0.02
6149	3	0.60	0.02
6149	10	0.43	0.01

表 8 RFID 数据有效性最终评价结果

Table 8 Final evaluation results of RFID data validity

基站	评价值	最优指标	最劣指标	综合排序
6149	0.169	车牌错误	流量均方差最小值	1
6027	0.290	一致性判断	车牌错误	2
6150	0.510	流量错误	速度均方差平均值	3
6151	0.598	速度均方差平均值	速度均方差最大值	4

2) 位置加权向量和评价值向量

按照 3.3 计算位置加权向量的方法和内容,对各评价主体下的有效性评价指标计算 $\eta_k(\alpha = \beta = 0.5)$,并通过求解目标函数为位置加权向量最大的最优规划模型,得到每个评价主体的 q 值,进而计算不同评价主体下的位置加权向量和评价值向量。以基站 6149 为例, $q = 0.2225$,评价值为 0.031,位置加权向量 $\omega = (0.5, 0.24, 0.115, 0.069, 0.047, 0.018, 0.007, 0.002, 0.001, 0, 0, 0, 0, 0)^T$,评价值向量 $y = (0.615, 0.031, 0.053, 0.085, 0.096, 0.135, 0.047, 0.046, 0.379, 0.076, 0.043, 0.069, 0.163, 0.117)^T$ 。

3) 最终评价值确定

求解 RFID 交通数据有效性评价值向量,RFID 有效性评价值的最大特征值为 1.1633。根据最大特征值对应的最大特征向量,确定最终评价值。其中,位置加权向量最大值(0.5)对应的指标,同时也是重新排序后位于第一位的评价指标,为该评价主体的最优评价指标;位置加权向量最小值(0)对应的指标,同是也是重新排序后最后一位的评价指标,为该评价主体的最劣评价指标(表 8)。

本文对综合评价结果的排序是按照评价值升序排列的,即评价值越高,其排序号越大,也即有效性评价指标值越大,该数据的有效性越差,例如在 RFID 交通数据的实际使用过程中,数据的缺失值越大,则该数据的可用性就越低。

通过基于 IOWA 算子的客观自主式综合评价能够获得不同评价主体的优劣指标,在体现不同交通数据的有效性和优劣势上,更符合使用者对数据有效性的判断。从评价得分结果来看:该方法的评价基站最终得分在 0.004~0.598 之间,基站间评分差别相对较大,可以有效体现各基站数据有效性的差异性。从评价结果信息表现来看:该方法可以同时给出不同评价主体下的评价值向量、被评价主体的相对优势指标和相对劣势指标,而且评价结果可直接得出该基站的最优指标和最劣指标,用户可根据具体需求,基于指标优劣来选择合适的基站。

5 结束语

本文针对 RFID 交通数据,分析现有交通数据评价指标和评价方法,设计了交通数据有效性指标,并采用基于 IOWA 算子的客观自主式评价方法对 RFID 交通数据进行了综合评价.本文介绍了 RFID 数据检测方法,建立了基于完整性、准确性、可用性、稳定性和交通参数宏观一致性的交通数据有效性评价指标体系,分别检测和统计 RFID 不同问题类别数据,得到不同维度下的数据有效性指标评价.同时,采用基于 IOWA 算子的客观自主式综合评价,得到基于指标优势度的指标序列,通过求解最优规划模型,得到不同评价主体下的评价权重向量,获取评价矩阵的最大特征值和最大特征向量,进一步得到最终评价值.最后,综合评价了不同 RFID 基站的数据有效性,发现基于 IOWA 算子的客观自主式综合评价能够获得不同评价主体的优劣指标,在体现不同交通数据的有效性和优劣势上,满足使用者对数据有效性的判断需求.

参考文献

References

- [1] Cappiello C, Francalanci C, Pernici B. Data quality assessment from user's perspective [C] // Proceedings of the 2004 International Workshop on Information Quality in Information Systems. Paris, France, 2004: 68-73
- [2] Hermans L, Vander Auweraer H, Mathieu L, et al. Modal parameter extraction from in-operation data [C] // Proceedings of 15th International Modal Analysis Conference; Current Horizon for Structural Damage Detection. Orlando, FL: Society for Experimental Mechanics, 1997: 531-539
- [3] Turner S, Albert L, Gajewski B, et al. Archived intelligent transportation system data quality: preliminary analyses of san Antonio TransGuided data [J]. Transportation Research Record (Journal of the Transportation Research Board), 2000, 1719(1): 77-84
- [4] Turner S. Defining and measuring traffic data quality: white paper on recommended approaches [J]. Transportation Research Record: Journal of the Transportation Research Board, 2004, 1870(1): 62-69
- [5] 耿彦斌,于雷,赵慧. ITS 数据质量控制技术及应用研究 [J]. 中国安全科学学报, 2005(1): 82-87
GENG Yanbin, YU Lei, ZHAO Hui. ITS data quality control techniques and applications [J]. China Safety Science Journal, 2005(1): 82-87
- [6] 袁高峰,王扬,郭建华. RFID 匹配数据的有效性分析和统计检验方法 [J]. 无线互联科技, 2013(8): 113-115, 168
YUAN Gaofeng, WANG Yang, GUO Jianhua. RFID match data validity analysis and statistical test [J]. Wireless Internet Technology, 2013(8): 113-115, 168
- [7] 杜威,杨越思,宁丹,等. RFID 交通错误数据检测及分析 [J]. 交通信息与安全, 2016, 34(2): 81-87
DU Wei, YANG Yuesi, NING Dan, et al. An analysis and a novel detection method of erroneous traffic data from RFID devices [J]. Journal of Transport Information and Safety, 2016, 34(2): 81-87
- [8] 杨越思,杜威,宁丹,等. RFID 交通冗余数据检测及分析 [J]. 交通信息与安全, 2016, 34(3): 72-80
YANG Yuesi, DU Wei, NING Dan, et al. Detection and an analysis of redundant RFID traffic data [J]. Journal of Transport Information and Safety, 2016, 34(3): 72-80
- [9] 杨越思,杜威,宁丹,等. RFID 交通数据时间覆盖率分析 [J]. 交通信息与安全, 2017, 35(3): 43-50, 58
YANG Yuesi, DU Wei, NING Dan, et al. An analysis on time coverage rate of RFID traffic data [J]. Journal of Transport Information and Safety, 2017, 35(3): 43-50, 58
- [10] 李伟伟. 自主式综合评价理论与方法研究 [D]. 沈阳: 东北大学, 2011
LI Weiwei. Study on theory and methods of self-determine comprehensive evaluation [D]. Shenyang: Northeastern University, 2011
- [11] 王殿海. 交通流理论 [M]. 北京: 人民交通出版社, 2002
WANG Dianhai. Traffic flow theory [M]. Beijing: People's Transportation Press, 2002
- [12] 易平涛,李伟伟,郭亚军,等. 基于二维 IOWA 算子的客观自主式评价方法 [J]. 运筹与管理, 2011, 20(6): 182-187
YI Pingtao, LI Weiwei, GUO Yajun, et al. Research into objective self-determined evaluation based on two-dimension IOWA operator [J]. Operations Research and Management Science, 2011, 20(6): 182-187
- [13] 林德花,袁振洲. 基于 IOWA 算子的短时交通流预测方法研究 [J]. 科学技术与工程, 2013, 13(25): 7596-7600
LIN Dehua, YUAN Zhenzhou. Research on short-term traffic flow forecasting method based on IOWA operator [J]. Science Technology and Engineering, 2013, 13(25): 7596-7600
- [14] Filev D, Yager R R. On the issue of obtaining OWA operator weights [J]. Fuzzy Sets and Systems, 1998, 94(2): 157-169
- [15] Yager R R, Filev D P. Induced ordered weighted averaging operators [J]. IEEE Transactions on Systems, Man and Cybernetics, Part b (Cybernetics), 1999, 29(2): 141-150

An IOWA operator based RFID data validity evaluation method

LONG Siqing¹ ZHAO Panming¹ GUO Jianhua¹

¹ Intelligent Transportation System Research Center, Southeast University, Nanjing 210018

Abstract The rapid development of intelligent transportation system (ITS) applications demands high quality traffic data. However, presence of abnormal data in the raw data acquired through traffic data collection techniques is inevitable, thereby reducing the validity of traffic data and degrading the performance of ITS applications. In this paper, concerning the traffic data collected using radio frequency identification (RFID)-based traffic data collection technique, a series of evaluation measures are proposed. Based on the characteristics of these measures and comprehensively considering the applicability conditions, advantages and disadvantages of each weight vector, an IOWA operator-based comprehensive evaluation method is proposed. Based on an application of the method to real world RFID data, the proposed method is demonstrated to be effective in evaluating the RFID-based traffic data.

Key words radio frequency identification (RFID); data acquisition; data evaluation index; IOWA operator; evaluation