

# 基于多源速度数据的动态权值融合方法研究\*

王路曼<sup>1,2</sup>, 王理民<sup>1,2</sup>, 蔡铭<sup>1,2</sup>

1. 中山大学智能工程学院, 广东 深圳 518106
2. 广东省智能交通系统重点实验室, 广东 广州 510006

**摘要:** 导航平台路段速度和浮动车 GPS 速度是常用的道路运行车速数据源, 单独使用一种数据进行相关研究存在实时更新数据路网覆盖率低、数据质量不稳定等问题。基于数据可靠性的动态权值数据融合方法, 文章将导航平台路段速度和浮动车 GPS 速度进行了融合。先将两种数据进行空间和时间同步化处理, 并针对路段数据和 GPS 数据设计不同的数据可靠性评估算法; 再将实时更新的导航平台路段速度和浮动车 GPS 速度基于数据可靠性融合, 得到最终的路段速度。实验表明: 导航平台路段数据、GPS 路段数据的平均时间粒度更新路网覆盖率分别为 60.32% 和 33.43%, 基于数据可靠性的融合数据达到了 74.29%, 路段速度平均误差降低到了 3.82 km/h, 均优于导航平台路段速度、GPS 路段速度和固定权值的融合速度。

**关键词:** 数据可靠性; 动态权值; 数据融合; 路网覆盖率; 路段速度

**中图分类号:** U491 **文献标志码:** A **文章编号:** 2097-0137(2022)05-0031-10

## Research on dynamic weight fusion method based on multi-source speed data

WANG Luman<sup>1,2</sup>, WANG Limin<sup>1,2</sup>, CAI Ming<sup>1,2</sup>

1. School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen 518106, China
2. Key Laboratory of Intelligent Transportation System in Guangdong Province, Sun Yat-sen University, Guangzhou 510006, China

**Abstract:** Road section speed of navigation platform and GPS speed of floating car are common sources of road speed data. Using one of the data alone for related research will have problems such as low coverage of real-time update data, unstable data quality and so on. Therefore, a real-time dynamic weight data fusion method based on data reliability is proposed, which fuses the navigation platform road section speed and floating car GPS speed. Before the two kinds of data fusion, space and time are synchronized, and different data reliability evaluation algorithms are designed for the navigation platform road section data and floating car GPS data. Then, the real-time updated navigation platform road section speed and floating car GPS speed are fused based on data reliability to get the final road section speed. The experimental results show that: the average time granularity data updating network coverage rate of navigation platform road section data and GPS road section data are 60.32% and 33.43% respectively, the fusion data based on data reliability is 74.29%, and the average error of road section speed is reduced to 3.82 km/h, which are better than navigation platform road section speed, GPS road section speed and data fusion speed according to fixed weight.

**Key words:** data reliability; dynamic weight; data fusion; road network coverage; road section speed

\* 收稿日期: 2021-03-16

录用日期: 2021-04-22

网络首发日期: 2022-01-18

基金项目: 国家重点研发计划 (2020YFB1600400)

作者简介: 王路曼 (1996年生), 女; 研究方向: 交通大数据; E-mail: wanglm25@mail2.sysu.edu.cn

通信作者: 蔡铭 (1977年生), 男; 研究方向: 交通大数据和智能交通系统; E-mail: caiming@mail.sysu.edu.cn

随着智慧交通建设的展开,交通数据采集能力有所提升,数据种类与数量增长迅速,交通大数据的应用成为了交通信息化水平提升的关键。导航平台路段速度数据和浮动车 GPS 速度数据具有数据量大、数据更新稳定、容易获取等优点,被广泛应用于交通领域,是常用的道路运行车速数据源。导航平台路段速度数据是指高德或百度等互联网公司发布的实时路段速度数据,此类数据以不同长度的路段为单元进行反馈,包含路段位置、长度、道路等级、运行车速等信息;而浮动车 GPS 数据是指实时采集的出租车、公交车、网约车等车辆的位置及速度等数据,将其位置坐标与地图进行匹配,可以得到的路段速度数据。

大多数学者都是基于其中一种数据进行相关研究。刘瑶杰<sup>[1]</sup>、孟晓彤<sup>[2]</sup>、鲁武当<sup>[3]</sup>基于导航平台数据对城市交通拥堵时空特征及未来路况预测进行研究。邹蕊焯<sup>[4]</sup>、贾洪飞等<sup>[5]</sup>利用浮动车 GPS 数据进行城市路网状况分析。Li C<sup>[6]</sup>基于出租车 GPS 实时速度、位置等信息,分析了道路实时拥堵等交通运行状态指标,为出租车及其他车辆出行提供参考。他们在研究中都提到浮动车 GPS 数据存在因实时车辆覆盖强度不同所导致的部分路段数据缺失、不完整等问题。目前,将两种数据融合使用的研究相对较少。赵顺晶等<sup>[7]</sup>将导航平台路段速度和浮动车 GPS 速度基于固定权值进行融合,但未考虑两种数据的实时更新数据质量问题。在多源数据融合方面,Alwan<sup>[8]</sup>、丁玥等<sup>[9]</sup>提出的多源数据融合的基本框架包含三个模块,即模式匹配、实体对齐和实体融合。陈磊等<sup>[10]</sup>提出了一种基于思维进化的优化 BP 神经网络数据融合算法。崔青华等<sup>[11]</sup>通过构建最优加权融合模型,将浮动车 GPS 数据与微波车辆检测数据进行数据融合,并将融合数据应用于行程时间的短时预测。战昕<sup>[12]</sup>将视频传感器与磁敏传感器获得的车辆检测信息融合,用于提高车辆检测信息的准确度,采用的融合方法是基于数据可靠性的 Dempster-Shafer 证据融合理论。上述数据融合方法大多需要掌握终端检测设备和数据传输情况,在实际应用中存在一定局限性。同时,由于数据源格式等不同,并不能完全适用于导航平台路段速度和浮动车 GPS 速度的融合。

本文将导航平台路段数据和浮动车 GPS 数据进行对比分析,提出了数据可靠性概念。针对导航平台路段速度和浮动车 GPS 速度设计不同的数据可靠性评估算法,基于数据可靠性指数赋予两

类数据实时动态权值并进行融合。该研究方法不仅可以避免单一数据源出现异常造成较大误差的情况,同时可以提高实时更新数据的路网覆盖率,有助于进行区域交通研究。

## 1 数据源与问题描述

### 1.1 实验数据

本研究所用到的数据包括导航平台路段数据和浮动车 GPS 数据,导航平台路段数据采用的是高德地图的路段速度数据,浮动车 GPS 数据采用的是营运车辆车载 GPS 产生的速度数据,数据所在区域为佛山市,数据时间范围为 2019 年 10 月 15 日至 11 月 14 日。高德路段速度数据以不同长度的路段为单元进行反馈,共计 52 752 条路段产生 731 732 350 条数据记录,数据时间粒度为 2 min,路段数据的主要字段及释义如表 1 所示。

表 1 路段速度数据字段及释义

Table 1 Field and interpretation of road section speed data

字段	字段类型	字段释义
gid	int	路段编号
roadname	string	路段名称
roadclass	int	路段等级
length	float	路段长度/m
width	float	路段宽度/m
maxlanes	int	路段车道数
lane_wide	float	单车道宽度/m
direction	int	路段方向
ad_code	int	行政区代码
geom	geometry	地理范围
update_time	timestamp	更新时间
speed	int	速度/( $\text{km}\cdot\text{h}^{-1}$ )

浮动车 GPS 数据主要包括出租客运、公交客运、网约车、普通货运等 10 类营运车辆车载 GPS 所产生的数据,共计 92 827 辆浮动车产生约 50 亿条数据记录,数据更新时间不定,浮动车 GPS 数据的主要字段及释义如表 2 所示。

### 1.2 数据预处理

由于导航平台路段速度数据是以不同长度的路段为单元进行反馈,而浮动车 GPS 数据是以车辆 GPS 点为单元生成数据记录,且两种数据的更新频率不同,数据融合前需先将两种数据进行空间和时间同步化处理,主要是对浮动车 GPS 数据进行处理。处理过程包括 GPS 数据清洗、统一空

表2 GPS数据字段及释义

Table 2 Field and interpretation of GPS data

字段	字段类型	字段释义
registrationNo	string	车牌号码
carclass	string	车辆类型
gpstime	timestamp	更新时间
latitude	decimal	纬度
longitude	decimal	经度
altitude	float	海拔高度
speed	int	速度/(km·h <sup>-1</sup> )
direction	int	行车方向

间坐标系、路网匹配以及路段速度计算4个部分,处理成为时间粒度为2 min的GPS路段速度数据。

浮动车GPS数据清洗是为了剔除经纬度异常、速度异常、时间异常、长期停驶等异常数据。浮动车GPS数据以经纬度为坐标,采用WGS84空间坐标系,而高德路段数据采用GCJ02坐标系,因此需要将两者统一到同一坐标系下。本研究采用的方法是将GPS数据坐标和路网地图的坐标转化到同一投影坐标下进行计算,所采用的投影坐标是:Xian\_1980\_3\_Degree\_GK\_Zo-ne\_38, WKID为2362。使用投影坐标的优点是可以将经纬度转化成以m为单位,以便于浮动车GPS点与道路的距离和后续的路网匹配算法的计算。

路网匹配算法主要是利用方向匹配法和最短距离法进行路网匹配。根据浮动车GPS数据的行车方向与高德路段数据行车方向的一致性以及最短距离原则,将浮动车GPS数据匹配到距离最近且与行车方向一致的路段。随机选择一辆公交车在一定时段内产生的GPS数据进行验证,预处理后的GPS数据有4 162条,正确匹配的数据有3 965条,匹配准确率达到95.3%,则认为该匹配方法能够满足研究需求。路段速度计算是利用浮动车GPS数据的速度字段作为瞬时速度,计算2 min内在同一路段上的所有浮动车速度的平均值,作为这条路段在2 min内的速度。

### 1.3 数据问题分析

导航平台路段速度数据和浮动车GPS数据由于数据终端采集、传输等均不相同,数据的更新频率和更新范围也有差异。经统计分析发现路段速度数据实时更新较为稳定,更新数据空间范围也相对固定,但更新的部分路段数据质量较差,存在路段更新速度一直为0 km/h、连续多次更新速度为同一值以及个别速度值异常等问题。以2019

年11月14日为例,有数据更新的路段有32 614条,占佛山市总路段条数的61.83%,路段数据更新情况如图1(a)所示,平均每2 min有更新的路段为32 604条,但全天路段更新速度一直为0 km/h的路段有897条,出现连续30 min以上更新速度为同一值的路段有11 843条,这两类异常数据占佛山市总路段条数的24.15%。

将浮动车GPS数据清洗后匹配到路段,通过统计分析发现GPS路段数据的更新空间范围较广,但由于实时车辆路段覆盖强度不同,存在部分路段数据缺失、不完整等问题。以2019年11月14日为例,有数据更新的路段有51 708条,占佛山市总路段条数的98.02%,路段数据更新情况如图1(b)所示,但平均每2 min有GPS数据更新的路段为11 773条,仅占佛山市总路段条数的22.32%。

通过上述分析可知两种数据各有优缺点,导航平台路段速度数据更新稳定但范围不够广,浮动车GPS数据更新范围广但实时更新稳定性较差,而且由于两种数据源不同,两种数据实时更新的路段覆盖范围也不相同。结合两种数据的特征将其融合,不仅可以有效提高路段速度数据的准确率,同时可以提高实时更新数据的路网覆盖率,即提高了查准率和查全率。

## 2 基于数据可靠性的数据融合算法

### 2.1 算法概述及框架

本文所提出的基于数据可靠性的导航平台路段速度和浮动车GPS速度融合算法,如图2所示。框架包含3部分内容:1)高德路段速度数据可靠性评估,从历史数据中提取历史时刻数据集合,通过对比实时数据与历史数据分析得到实时数据的可靠性指数。2)GPS数据可靠性评估,首先将GPS数据清洗匹配到路段上形成GPS路段数据,再依据GPS历史时段数据更新情况进行可靠性评估。3)将两类数据基于数据可靠性指数进行加权平均得到路段融合后实时更新的数据。数据的可靠性评估是整个算法中最重要的部分,评估指标是可靠性指数 $r$ , $r \in (0, 1)$ , $r$ 越大表示数据越可靠。

### 2.2 高德路段数据可靠性评估算法

路段速度数据本质上是由交通出行者决定的,而出行者的出行特征总体上具有一定的时间空间分布规律,因此导致了路段速度数据也具有一定的时间周期性和时空相关性。本文以路段历史速度数据以及与之相邻的上下游路段历史速度数据

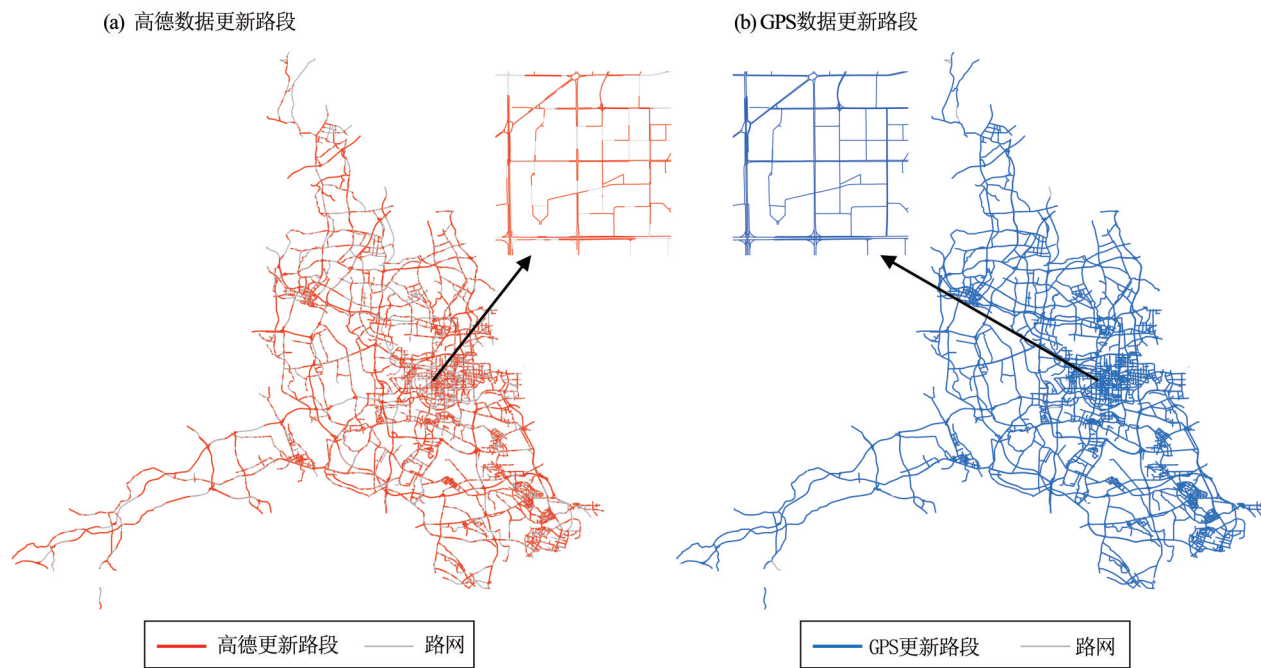


图1 两种数据更新空间范围

Fig. 1 The spatial range of two kinds of data updating

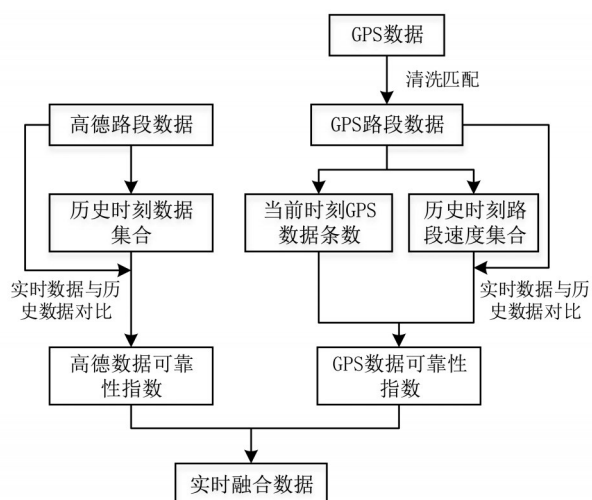


图2 数据融合算法框架

Fig. 2 Framework of data fusion algorithm

进行路段速度数据可靠性评估。城市路网中存在交叉口会将不同道路打断成多条封闭路段，但为排除交叉口对路段速度的影响，封闭路段不包括交叉口临近路段，封闭路段示意图如图3所示。路段的上下游路段是指封闭路段内与其接壤的上下游路段，即路段的上下游路段唯一或没有（交叉口临近路段无上下游路段），路段速度数据的可靠性评估算法如图4所示，与之相对应的字段释义如表3所示。

在高德路段数据中加入表征星期

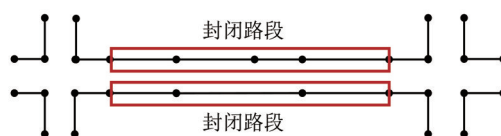


图3 封闭路段示意图

Fig. 3 Schematic diagram of closed road section

1~7。输入某路段当前更新时刻的高德路段速度  $v_t$  和该路段及其上下游路段历史数据中同星期几、同时、同分的速度值组成历史数据集  $H_t$ ，求出该数据集的阈值集合  $[H_{lmin}, H_{lmax}]$  和均值  $H_{lmean}$ 。当实时路段速度  $v_t \in [H_{lmin}, H_{lmax}]$ ，说明该速度数据较为可靠，则当前更新时刻路段数据的可靠性指数  $r_t$  为1；当实时路段速度  $v_t \notin [H_{lmin}, H_{lmax}]$ ，则该路段速度数据的可靠性指数与实时速度和历史数据集均值的差值成反比。若该路段的历史数据集均值等于0且当前实时速度等于0，则该路段速度数据的可靠性指数为0；若该路段的历史数据集均值等于0且当前实时速度不为0，则该速度数据的可靠性指数为1。

### 2.3 GPS数据可靠性评估算法

浮动车GPS数据的应用是通过将其匹配到具体的路段上来计算路段速度，因此浮动车GPS数据的可靠性评估包括两部分：1) 2 min时间间隔内成功匹配到某一路段上的GPS数据条数直接决定了GPS数据的可靠性。GPS数据条数越多数据可靠性就越高，反之则存在数据偶然性的概率越大；

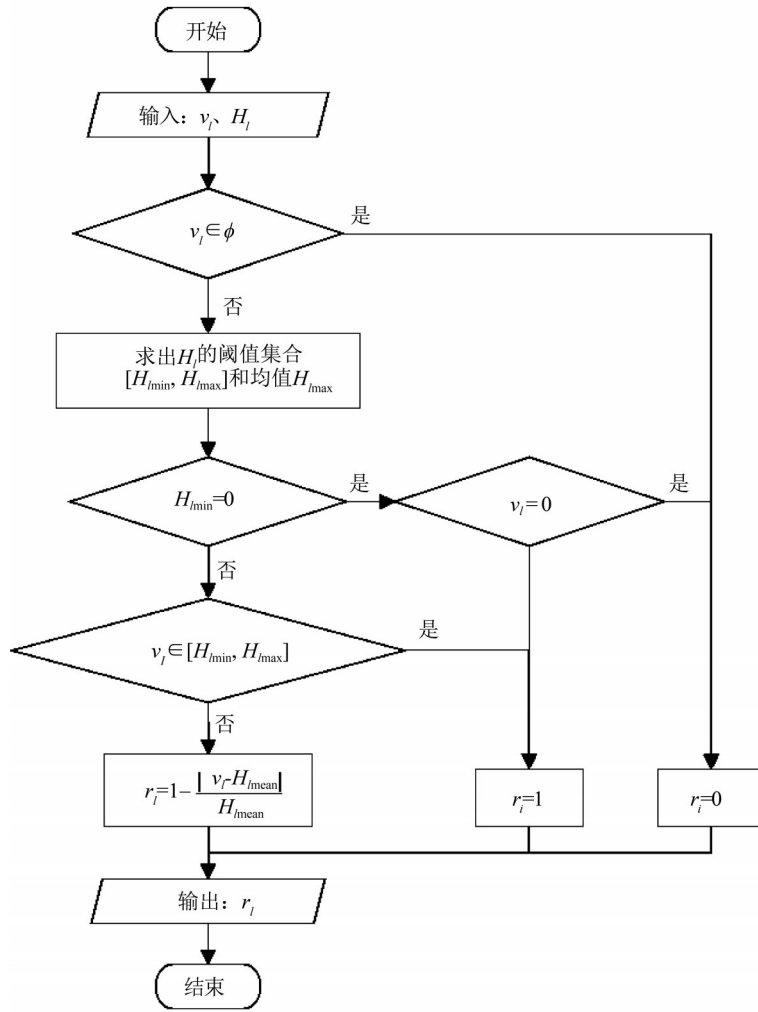


图 4 高德路段数据可靠性评估算法

Fig. 4 Reliability evaluation algorithm for Gaode road section data

表 3 数据可靠性评估算法字段释义表

Table 3 Field and interpretation of road section data reliability evaluation algorithm

字段	释义
$v_i$	某路段当前更新时刻的高德路段速度
$H_i$	某路段及其上下游路段历史数据中同星期几、同时、同分的速度值组成的历史数据集
$r_i$	某路段速度数据的可靠性指数

2) 基于该路段历史同更新时刻且匹配 GPS 数据条数较多的路段速度组成的历史数据集, 该路段当前更新时刻的路段速度与历史数据集的偏差决定了 GPS 数据的可靠性, 偏差越大, 可靠性越低。

利用数理统计方法确定基于路段平均速度估计的最小 GPS 数据样本量, 即对于某条路段, 若给定允许速度误差  $\delta$ 、样本标准差为  $s$ , 置信度水平为  $(1 - \alpha)$ , 则 GPS 数据样本数量  $n$  应满足

$$n \geq \left( t_{\frac{\alpha}{2}, n-1} \times s / \delta \right)^2, \quad (1)$$

式中  $t_{\frac{\alpha}{2}, n-1}$  为  $n - 1$  自由度的  $t$  分布值。

由上式可计算出, 当某条路段时间粒度 GPS 数据样本数量  $n \geq 5$  时, 置信度水平  $(1 - \alpha) \geq 0.9$ 。即认为若当前时刻成功匹配到某路段上的 GPS 点数  $\geq 5$  个, 则该路段的可靠性指数较高。路段浮动 GPS 数据可靠性评估算法如图 5 所示, 与之相对应的字段释义如表 4 所示。

在浮动 GPS 数据中加入表征星期几的列, 取值 1~7。整个可靠性评估算法包含两部分: 1) 由成功匹配到路段的 GPS 点数决定路段 GPS 样本量可靠性指数  $r_{g1}$ 。2) 将该路段历史数据中同星期几、同时、同分且匹配到该路段的 GPS 个数  $\geq 5$  的路段速度值组成历史数据集, 评估方法与高德路段数据相同, 得出 GPS 路段速度的可靠性指数  $r_{g2}$ , 最

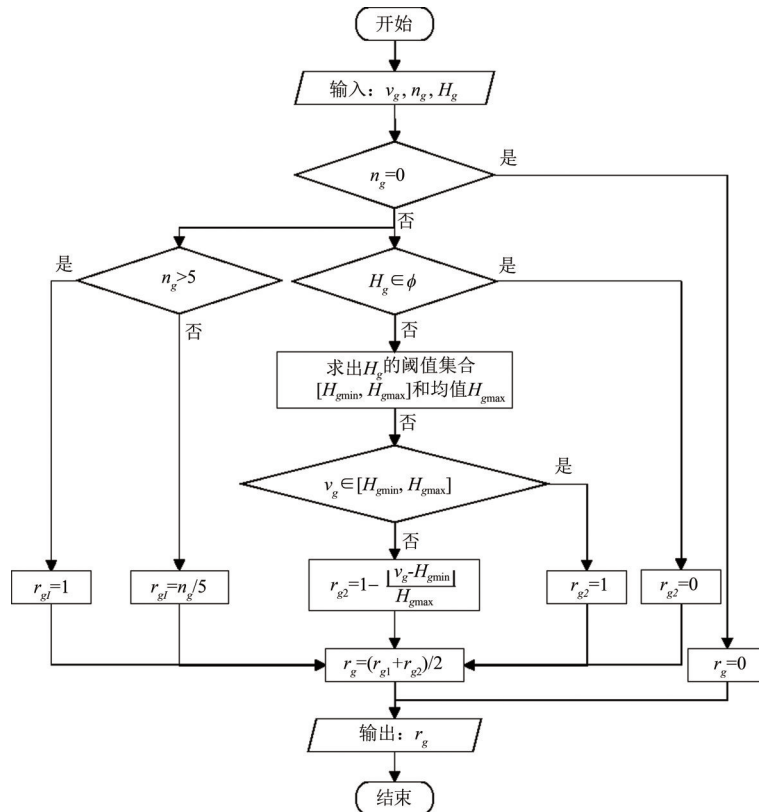


图 5 GPS 数据可靠性评估算法流程

Fig. 5 Procedure of data reliability evaluation algorithm for GPS data

表 4 GPS 数据可靠性评估算法字段释义表

Table 4 Field and interpretation of GPS data reliability evaluation algorithm

字段	释义
$v_g$	某路段当前更新时刻的 GPS 路段速度
$n_g$	匹配到该路段的 GPS 点数
$H_g$	某路段历史数据中同星期几、同时、同分且匹配到路段的 GPS 个数 $\geq 5$ 的路段速度值组成的历史数据集
$r_g$	某路段速度数据的可靠性指数

后将两部分的可靠性指数进行平均得到该路段当前更新时刻浮动车 GPS 数据可靠性指数  $r_g$ 。

### 2.4 数据融合

在进行高德路段速度数据和浮动车 GPS 数据的数据可靠性评估后, 可得出当前时刻的高德路段速度数据的可靠性指数  $r_l$  和 GPS 路段数据的数据可靠性指数  $r_g$ 。将实时更新的高德路段速度  $v_l$  和浮动车 GPS 平均速度  $v_g$ , 基于数据可靠性指数进行加权平均得到路段融合后的最终速度  $v$ , 路段当前更新时刻的最终速度可表示为

$$v = \frac{v_l r_l + v_g r_g}{r_l + r_g} \quad (2)$$

## 3 实验与分析

### 3.1 数据可靠性评估

实验中高德路段速度数据和浮动车 GPS 数据的历史数据集均为 2019 年 10 月 15 日至 11 月 13 日, 共 30 d。以 2019 年 11 月 14 日的数据为测试集, 实验数据样例如表 5 和表 6 所示。数据表均以路段为单元进行展示, 共计 52 752 条路段, 更新时间粒度为 2 min, 但由于表中只展示有更新的路段数据, 且某些路段某些时刻高德路段速度无更新或者无成功匹配到路段的 GPS 数据, 表中存在路段编号不连续现象。

### 3.2 结果与分析

将基于数据可靠性的实时动态权值融合数据与单独使用某一种数据或固定权值的融合数据的数据质量进行对比, 以验证本文所提出算法的有效性和实用性。本文以平均时间粒度数据更新的路网覆盖率 (RNC, road network coverage) 和路段速度平均绝对误差 (MAE) 为评价指标。定义

$$RNC = \frac{1}{NLT} \sum_{j=1}^T \sum_{i=1}^N l_i \times 100\%, \quad (3)$$

$$MAE = \frac{1}{T} \sum_{i=1}^T (|v - v_i|), \quad (4)$$

表5 2019年11月14日的高德路段数据可靠性评估

Table 5 Reliability evaluation of Gaode road section data at Nov. 14, 2019

编号	星期	更新时间	速度/(km·h <sup>-1</sup> )	历史数据阈值区间	历史数据集均值	$v_i$
3	4	00:02	50	[40, 90]	58.25	1.00
4	4	00:02	63	[51, 88]	60.92	1.00
5	4	00:02	67	[40, 90]	61.62	1.00
...	...	...	...	...	...	...
18 397	4	12:58	70	[54, 78]	67.83	1.00
18 398	4	12:58	58	[54, 78]	65.25	1.00
18 399	4	12:58	76	[61, 74]	68.50	0.89
...	...	...	...	...	...	...
52 747	4	24:00	69	[63, 75]	69.33	1.00
527 48	4	24:00	71	[63, 75]	69.75	0.85
52 749	4	24:00	52	[23, 40]	35.75	0.55

表6 2019年11月14日的GPS数据可靠性评估

Table 6 Reliability evaluation of GPS data at Nov. 14, 2019

编号	星期	更新时间	速度/(km·h <sup>-1</sup> )	GPS个数	$r_{gl}$	历史数据阈值区间	数据集均值	$r_{g2}$	$r_g$
2	4	00:02	48.75	8	1.00	[60.40, 73.50]	66.95	0.73	0.87
3	4	00:02	35.00	2	0.40				0.40
4	4	00:02	42.00	4	0.80				0.80
...	...	...	...	...	...	...	...	...	...
18 396	4	12:58	56.25	8	1.00	[33.36, 68.40]	47.97	1.00	1.00
18 397	4	12:58	61.00	2	0.40	[55.33, 80.75]	65.46	1.00	0.70
18 398	4	12:58	59.00	2	0.40	[21.25, 70]	45.12	1.00	0.70
...	...	...	...	...	...	...	...	...	...
52 716	4	24:00	63.00	2	0.40	—	—	—	0.40
52 743	4	24:00	67.20	10	1.00	[52.33, 70.50]	64.19	1.00	1.00
52 745	4	24:00	68.00	2	0.40	—	—	—	0.40

式中  $N$  表示每个时间粒度更新的路段数,  $l_i$  表示更新路段的长度,  $L$  表示路网总长度,  $v$  表示更新数据中的速度值,  $v_s$  表示实际测量的速度值,  $T$  表示时间粒度的个数。

**3.2.1 平均时间粒度路段更新的比例** 本文的融合数据、高德路段速度数据、浮动车GPS数据的平均时间粒度更新的路网覆盖率如图6所示。图6中, 由于白天路网中行驶的车辆较多, 产生的GPS数据路网覆盖率较高, 本文融合数据路网覆盖率也较高; 但GPS数据更新情况不稳定, 存在某个时刻路网覆盖率突降的情况。高德路段数据更新情况较为稳定, 更新数据路网覆盖率维持在60%左右。据统计GPS数据平均时间粒度数据更新路网覆盖率为33.43%, 高德路段数据为60.32%, 而本文的

融合数据达到74.29%, 显著高于其他两种数据。

**3.2.2 路段速度平均绝对误差** 采用了实地调查的方式, 对融合的速度数据的准确度进行校验。在佛山市主城区随机选取了6条两类数据更新状况均良好的路段, 调查人员站在调查路段中间位置, 所选路段及测量点如图7所示。

调查时间为2019年11月14日, 调查时段分为早高峰7:00~9:00, 晚高峰17:00~19:00, 平峰时段10:00~12:00和14:00~16:00, 每个时间段为2h, 选取其中1h进行不间断测量, 保证每条路段被测时段包含一个高峰时段和一个平峰时段。测量路段的编号、名称、道路等级、长度等基本属性, 以及测量时段如表7所示。

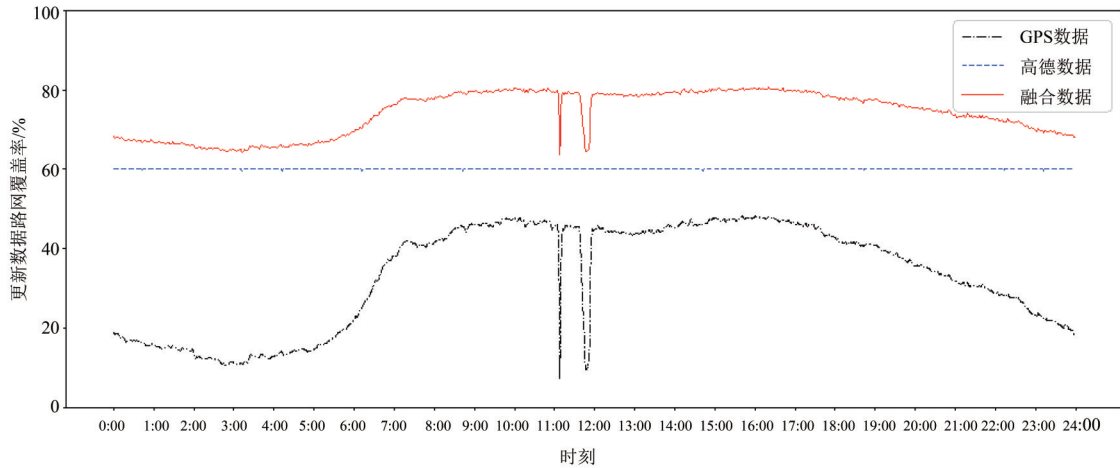


图 6 不同数据实时更新数据路网覆盖率

Fig. 6 Real time update of road network coverage with different data



图 7 测量路段示意图

Fig. 7 Schematic diagram of survey road

表 7 测量路段信息

Table 7 Survey road section information

编号	道路名称	长度/m	宽度/m	车道数	道路等级	行车方向	平峰测量时段	高峰测量时段
37732	佛山大道中	392.8	10.5	3	快速路	3	15:30~16:30	17:00~18:00
37999	佛山大道北辅路	272.1	9	3	主干路	3	15:30~16:30	17:00~18:00
38784	汾江南路	189.4	9	3	主干路	2	8:00~9:00	10:00~11:00
39830	佛山大道中	392.2	10.5	3	快速路	2	15:30~16:30	17:00~18:00
40315	汾江南路	164.8	9	3	主干路	2	8:00~9:00	10:00~11:00
42204	汾江中路	214	6	2	主干路	2	8:00~9:00	10:00~11:00

调查人员是由专业技术人员担任,使用雷达测速仪测速时随机选择被测车辆,避免总是选择车队的第一辆车、某一车道车辆比例过大、某一种车型比例过大或者高速行驶车辆比例过大等情况,同时调查人员在测速时隐蔽在驾驶员不易发

现的位置,尽量保证调查结果能够真实地反映车辆行驶状况。

调查人员调查时首先使用皮尺测量路段长度并确定路段中点位置,并选定固定测速点位置。正式测量开始时,调查人员使用雷达测速仪不间

断测速记录, 测量频率平均约为 5 s/次, 并以 2 min 为时间粒度进行分隔, 取 2 min 内所测得的所有速度样本的平均值为当前时间粒度的实测速度值。

将高德路段速度、GPS 路段速度和两种数据按 1 : 1 固定权值融合、按照本文提出的实时动态权值融合, 并与实测速度值进行对比, 结果如图 8 所示。可以看出, 各测量路段的不同数据速度曲线不连续。一是由于测量时段不连续, 在平峰时段与高峰时段断开; 二是由于某些时刻某种数据无更新。且, 当高德路段速度和 GPS 路段速度数据质量都较好时, 固定权值融合速度和动态权值融合速度曲线几乎重合, 并都与实测速度值拟合较

好, 如路段 37732 和 39830 所示。当高德路段速度和 GPS 路段速度数据质量不稳定, 尤其是 GPS 数据波动较大时, 动态权值融合速度曲线与实测速度值的拟合性明显优于其他三种数据源, 如路段 37999 和 40315 所示。计算得出各路段测量时段的速度平均绝对误差, 如表 8 所示。实验表明, 在随机选取的 6 条路段上, 基于动态权值融合的速度数据准确度较高, 所有测量时段的速度平均绝对误差为 3.82 km/h, 高峰测量时段的速度平均绝对误差为 3.69 km/h, 平峰测量时段的速度平均绝对误差为 3.95 km/h, 表现最优。

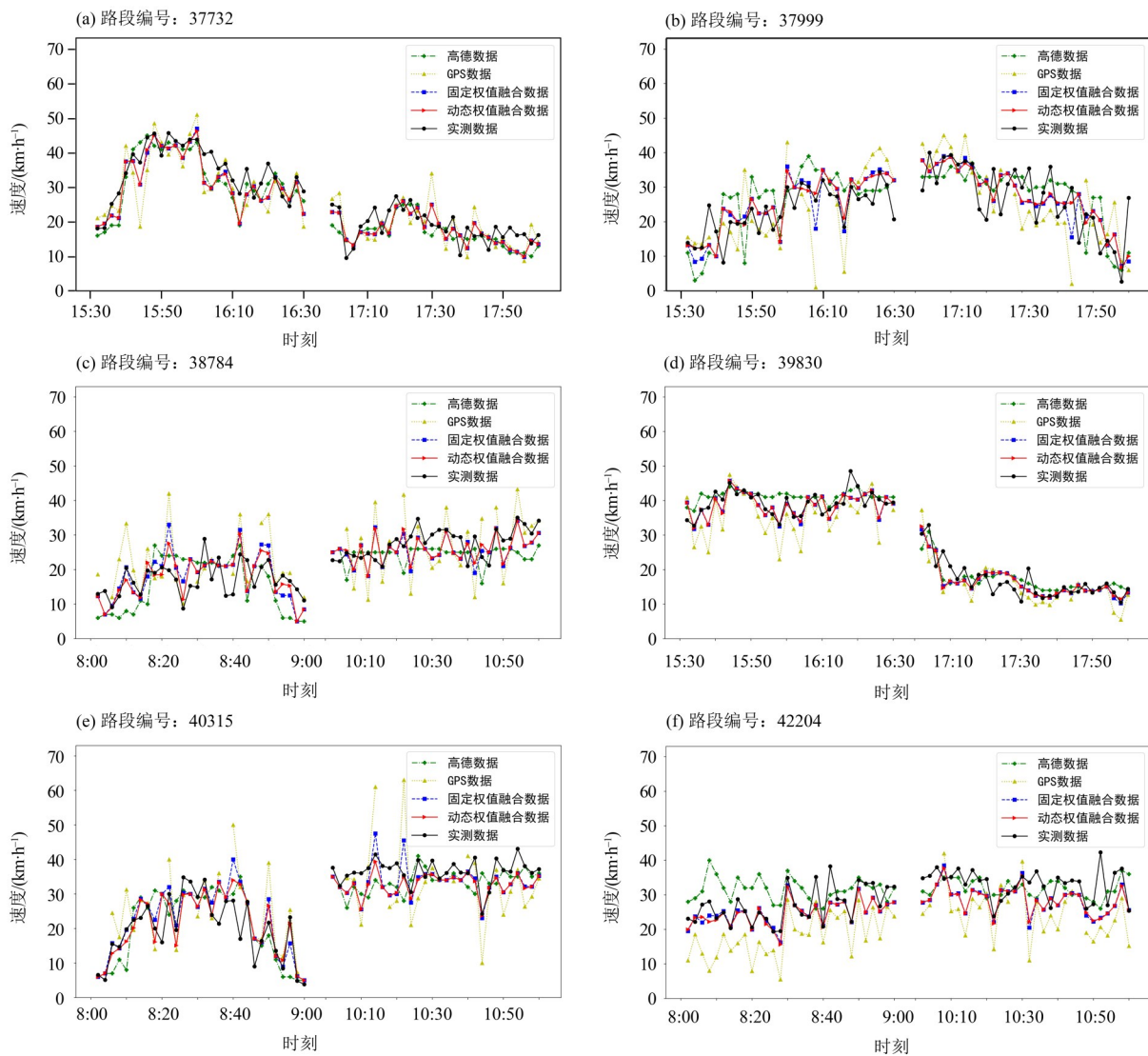


图 8 各测量路段的不同数据速度波动曲线

Fig. 8 Different data speed fluctuation curve of each measurement road section

表 8 不同数据源路段速度的平均绝对误差  
Table 8 Average absolute error of road section speed in different data

数据源	测量时段 MAE / (km·h <sup>-1</sup> )	高峰测量时段 MAE / (km·h <sup>-1</sup> )	平峰测量时段 MAE / (km·h <sup>-1</sup> )
高德数据	4.80	5.23	4.37
GPS 数据	6.67	6.25	7.09
固定权值融合数据	4.00	3.85	4.13
动态权值融合数据	3.82	3.69	3.95

## 4 结 论

本文提出了一种基于数据可靠性的动态权值融合方法,将导航平台路段速度数据和浮动车 GPS 速度数据进行了融合。以佛山市的高德路段速度数据和浮动车 GPS 数据为例,对两类数据进行空间和时间同步化处理,并针对高德路段速度数据和浮动车 GPS 数据设计不同的可靠性评估算法。将实时更新的高德路段速度和 GPS 路段速度基于数据可靠性指数进行加权平均得到了融合后最终的路段速度。验证实验表明:基于数据可靠性的

动态权值融合方法有效地提高了实时更新数据的路网覆盖率和数据准确度,不仅降低了单一数据源出现异常造成的误差,同时避免了固定权值融合不考虑数据源质量简单加权融合的弊端。

本文提出的方法虽然在导航平台路段速度数据和浮动车 GPS 速度数据融合中具有较好的应用效果,但由于不同类型数据源的特征不同,在后续的研究工作中,尝试将该方法应用到其他类型数据源融合场景中。此外,还可以挖掘更多的速度数据特征,进一步优化数据融合算法,提高融合数据精度。

### 参考文献:

- [1] 刘瑶杰. 基于实时路况的交通拥堵时空聚类分析——以北京市二环以内为例[D]. 北京:首都师范大学, 2014.
- [2] 孟晓彤. 基于实时路况的兰州市城关区交通拥堵时空特征及对策探究[D]. 兰州:兰州大学, 2017.
- [3] 鲁武当. 基于交通态势数据的城市道路平均行程速度时空分布特征及预测研究[D]. 西安:长安大学, 2019.
- [4] 邹蕊焯. 基于浮动车 GPS 数据的交通区域状况分析[D]. 西安:长安大学, 2018.
- [5] 贾洪飞, 郭明雪, 罗清玉, 等. GPS 数据下的城市路网关键路段识别[J]. 吉林大学学报(工学版), 2020, 50(4): 1338-1344.
- [6] LI C. The analysis of urban residential activities based on taxi GPS data[J]. Applied Mechanics & Materials, 2013, 253/254/255: 1816-1819.
- [7] 赵顺晶, 李浩, 汤夕根, 等. 一种基于 FCD 与互联网数据融合的交通拥堵指数评价方法: 中国, 11159672.7[P]. 2018-02-03.
- [8] ALWAN A A, NORDIN A, ALZEBER M, et al. A survey of schema matching research using database schemas and instances [J]. International Journal of Advanced Computer Science and Applications, 2017 (8): 102-111.
- [9] 丁玥, 王涓, 卢卫, 等. 面向多源关系数据的融合[J]. 中国科学: 信息科学, 2020, 50(5): 649-661.
- [10] 陈磊, 王江锋, 谷远利, 等. 基于思维进化优化的多源交通数据融合算法[J]. 吉林大学学报(工学版), 2019, 49(3): 705-713.
- [11] 崔青华, 聂庆慧, 基于多源交通数据可靠性的行程时间短时预测[C]// 第九届中国智能交通年会, 2020.
- [12] 战昕. 交通系统数据融合算法研究[D]. 西安:西安工业大学, 2018.

(责任编辑 王海蓉)

(终校 本刊编辑部)