

浮动车信息处理技术研究

诸彤宇 郭胜敏

(北京航空航天大学软件开发环境国家重点实验室, 北京 100191)

摘要 概括了浮动车技术的基本原理,重点介绍了浮动车信息处理系统的处理流程及地图匹配、路径推测和路况信息计算方法。针对GPS数据精度差、路网结构复杂等影响浮动车信息处理准确性提升的问题,给出了一系列改进型研究工作。最后,给出了评估浮动车信息处理准确性的方法,并指出了浮动车信息处理技术的后续工作。

关键词 浮动车 信息处理 地图匹配 路径推测 信息融合

中图法分类号:TP391.4 **文献标识码**:A **文章编号**:1006-8961(2009)07-1230-08

A Study on Floating Car Based Information Processing Technology

ZHU Tong-yu, GUO Sheng-min

(State Key Laboratory of Software Development Environment, Beijing University of Aeronautics and Astronautics, Beijing 100191)

Abstract The floating car technique is one of the key technologies in ITS (intelligent transportation system) to acquire the traffic information in recent years. In this paper, firstly, the fundamental principle of floating car technique is introduced. Then, the enhanced processing procedures of floating car information system, and some important approaches including map matching, path deriving and traffic information calculating are emphasized. To review the key issues affecting information processing accuracy, such as poor GPS data precision and complex road network, a serious of enhanced approaches are presented. Finally, some experiments are given to evaluate information processing accuracy and future work is pointed out.

Keywords floating car, information processing, map matching, path deriving, information fusion

1 引言

当前,城市交通日益拥堵、事故频发以及交通所带来的环境恶化和能源短缺已成为世界各国所面临的共同挑战,交通问题已成为各国政府亟待解决的重要问题。近年来中国道路交通基础设施建设水平虽有较大改观,但仍不能完全满足现代经济生活的需要。社会公众、政府和企业对全面、准确、实时的城市道路动态信息服务的迫切需求已达到了前所未有的状态。

解决城市交通问题的一个有效途径就是建设先进交通信息系统(advanced traffic information system, ATIS),希望在完善的信息网络基础上,通过装备在

道路上、车辆上、换乘站、停车场等场地上的传感器和传输设备,获得各类实时交通信息并进行综合处理,向社会提供全面、准确、实时的道路交通信息,为公众出行提供诱导服务。

浮动车(floating car data, FCD)技术^[1]作为一种新兴的交通信息采集方式,可利用车载GPS定位系统采集车辆的位置,同时将位置信息通过无线通讯系统传输到信息处理中心,在信息中心应用相关的计算模型和算法,得到城市的道路交通拥堵状况信息。与传统的道路检测技术和视频监控技术等固定点采集手段相比,浮动车具有低成本、易安装、易维护、测量范围广、实时性和动态性等优势。因此,基于浮动车的交通信息采集和处理方式在ATIS中逐渐趋于主导地位。

收稿日期:2009-04-21;改回日期:2009-04-28

第一作者简介:诸彤宇(1969~),男。副教授,硕士生导师。1999年于北京航空航天大学获计算机软件与理论专业硕士学位。主要研究方向为智能交通系统、人工智能、数据挖掘等。E-mail:zhutongyu@nlsde.buaa.edu.cn

本文简要介绍浮动车技术的基本原理,信息处理系统的功能模块和一系列关键技术,同时探讨目前尚存在的问题及可能的解决思路。

2 浮动车信息处理基本原理

2.1 浮动车技术框架

浮动车利用在车辆上安装的车载移动通信设备和定位装置,例如基于GPS定位系统或移动电话定位系统,将车辆动态信息(时间、速度、坐标和行驶方向等)实时传送到浮动车信息处理中心,通过定位数据与路网的匹配,以及其他数据的分析来获取道路交通路况信息。

浮动车技术的数据范围遍布整个地区,能全天候24小时进行数据采集;利用无线实时传输、中心式处理大大提高信息采集效率;通过测量的车辆瞬时状态数据,能准确反映交通流变化;利用浮动车技术还可以实现多参数测量,包括天气、道路状况、车辆安全等参数;利用现有的GPS和通信网络资源,采集设备维护和安装成本低。

2.2 浮动车信息处理架构

信息处理子系统是浮动车系统的核心,主要包含地图匹配、车辆行驶路径推测^[2]、路况信息计算3个子模块。

2.2.1 地图匹配

地图匹配根据数字地图的道路坐标数据调整GPS返回的车辆坐标数据,使其落到可能正确的行驶道路上。该部分主要完成车辆坐标数据点投影的工作,即通过将浮动车不同时刻的车辆定位数据向周围道路进行匹配投影,最终获得其可能行驶道路和相应投影点的信息。

2.2.2 路径推测

路径推测位于地图匹配之后,主要是综合每辆浮动车在某一时间段(通常是5分钟)内的所有车辆点数据,根据时间和道路的连续性,最终确定该车辆在本周期内的行车路线,然后进一步可以获得路线上每个定位点对应的唯一的投影点和道路信息。

2.2.3 路况计算

路况计算处理的对象是推测出的行车路线数据。计算每辆车相邻两个定位点之间途经道路的拥堵状况和旅行时间,对横跨相邻路链的路线在交点处进行分割,依次完成所有浮动车所行驶道路的路况信息计算。最后对每条路链上所有车辆产生的路

况信息做进一步的融合处理,生成以路链为单位的实时动态交通信息。

从上文的论述可以看出,浮动车信息处理需要两个必不可少的要素:路网结构表示的电子地图及浮动车GPS定位点。目前的技术条件下,以及基于成本的控制,浮动车GPS数据质量难以满足实际需求,突出表现在以下几个方面:

(1)可采集的参数有限,不能采集到判断车辆行驶特征的关键参数;

(2)GPS定位精度差,浮动车普遍装备的GPS接收机一般存在50 m以内的定位误差;

(3)采样间隔大,连续两个定位点跨越了较长距离,存在多条可能的车辆行驶路径。

同时,复杂的城市路网也制约了浮动车信息处理准确性的进一步提高。例如,北京市普遍存在的主辅路并行路网结构,主辅路之间的间隔一般在15 m以内,现有GPS定位精度很难将定位点准确匹配到主路或是辅路上,影响后续路况信息的计算。

3 浮动车信息处理关键技术

为了提高浮动车信息处理的准确性和计算效率,本文做了以下几个方面的尝试:(1)对基本道路空间数据进行分层处理和抽象,优化路网结构;(2)浮动车行驶模式判断模型可以甄别不良行驶模式下的干扰数据;(3)运用统计学原理和数据挖掘算法提升地图匹配的准确性;(4)启发式路径推测算法导入几何运算的约束条件,快速搜索车辆的行驶轨迹;(5)提出并行路网结构下的浮动车信息处理模型;(6)在生成路况信息的过程中考虑不同车辆所提供信息的可信度,等等。上述研究成果将在本章分节做具体介绍。

3.1 三层路网模型

处理浮动车数据需要城市交通地理信息系统(GIS-T)的支持。在一般的矢量编码方式路网数字地图中,道路以线或折线对象的方式存储,表示为一系列坐标点的有序集。然而,大城市路网密集,道路结构越来越复杂,高效的地图匹配算法和路径推测算法应该整合相应的路网数据结构。在这种情况下,通过对基本道路空间数据的分层处理和抽象,可以建立一种基于节点-路段-路链的三层路网数据结构^[2]。

$$\begin{cases} \mathbf{R}_n = (\mathbf{N}, \mathbf{S}) \\ \mathbf{N} = \{n | n = (x, y), x, y \in \text{Coordinate}\} \\ \mathbf{S} = \{s | s = \langle m, n \rangle, m, n \in \mathbf{N}\} \end{cases} \quad (1)$$

式中, $\mathbf{R}_n$ 代表路网结构, $\mathbf{N}$ 和 $\mathbf{S}$ 分别是节点集和路段集。每一个节点 n 包含一个经度和一个纬度; 路段 s 由两个节点的有序对 $\langle m, n \rangle$ 组成, m 和 n 分别是 s 的起始节点 $\text{begin}(s)$ 和终止节点 $\text{end}(s)$ 。路网结构如图 1 所示。

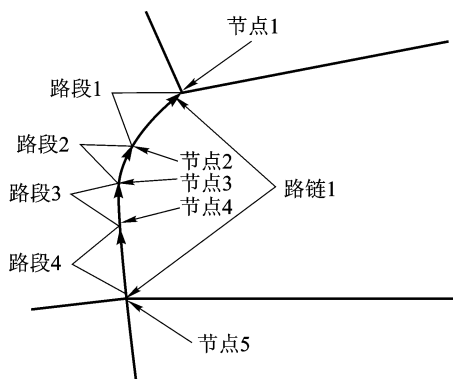


图 1 路网结构示意图

Fig. 1 The sample of road network

对于路网 $\mathbf{R}_n$ 中的任何一个节点 n , n 的出度 $o(n) = \{\langle n, z \rangle | \langle n, z \rangle \in \mathbf{S}\}$ 表示以 n 作为起始节点的路段的个数, 同理可以定义 n 的入度 $i(n)$ 。图 1 中, 节点 1 的入度和出度分别是 1 和 2。

对于路网 $\mathbf{R}_n$ 中的一个路段序列 $\langle s_1, s_2, \dots, s_n \rangle$, 如果满足如下条件

$$\begin{cases} i(\text{begin}(s_1)) \neq 1, o(\text{end}(s_n)) \neq 1 \\ i(\text{begin}(s_k)) = 1, o(\text{end}(s_k)) = 1, 1 < k < n \end{cases} \quad (2)$$

则 $\langle s_1, s_2, \dots, s_n \rangle$ 组成一条路链, 如图 1 所示。

3.2 行驶模式判断

在实际应用中, 通常选择出租车或公交车作为浮动车, 而这些车辆所具有的某些特殊行驶特征往往给路况计算的准确性造成消极影响。如出租车超过 40% 的时间处在停驶状态或在道旁及宾馆等地待客, 系统则会根据此时车辆的行驶速度而做出该区域交通拥堵的误判。鉴于采集的 GPS 定位数据中充斥着大量的上述干扰数据, 在进行进一步处理之前, 对浮动车的行驶状态进行预处理和分析, 甄别并剔除干扰数据, 可以提高路况信息计算的准确性并提升系统处理的整体效率。

通常认为浮动车存在 3 种行驶模式: (I) 正常行驶模式; (II) 停驶及待客模式; (III) 拥堵及缓慢

模式。根据浮动车记录包含的即时速度信息可以较准确地辨别正常模式, 但由于浮动车 GPS 定位数据具有一定的误差, 在低速情况下区分模式 II 和模式 III 具有一定的困难, 需要结合浮动车的经纬度坐标变化、事件触发等其他数据特征, 设计适当的分类器来判断。

文献[3]设计了一个浮动车行驶模式判断模型, 以期将模式 II 和模式 III 正确区分开来, 为干扰数据的剔除提供决策依据。分类器设计过程如下: 首先提取数据特征, 将原始数据映射到特征空间; 然后通过特征选择, 对特征空间进行降维, 选择对分类最有利的特征组合; 最后在低维空间中进行 SVM(支持向量机)分类器^[4]的设计。模型的体系架构如图 2 所示。

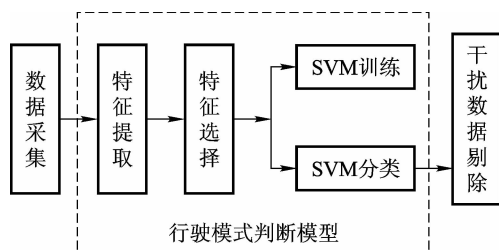


图 2 浮动车行驶模式判断模型架构图

Fig. 2 Driving mode classification model

3.3 地图匹配算法

传统的地图匹配算法只应用了 GPS 定位点和路链之间的位置参考信息, 而没有考虑路链拓扑特性与车辆行驶特性之间深层次的内在联系。经过深入细致的数据分析, 可以发现车辆在高速和载客的情况下倾向于在高等级道路上行驶。因此, 地图匹配可参考的因素就扩展至 4 个: (1) 定位点到路链的投影距离 distance ; (2) 车辆的行驶方向和路链方向的夹角 angle ; (3) 平均车速 speed ; (4) 车辆载客状态 state 。

地图匹配可以被看作一个二分类问题, 即根据一定的数据特征将 GPS 定位点和路链的关系分为匹配和不匹配两种。本文将分类样本定义为 (GPS 定位点, 路链) 关系对, 选取 SVM^[4] 作为分类器, 上述 4 个因素分别作为数据特征。同时, 为了提高分类的质量, 提取了 4 个特征对应的后验概率^[5] 作为二次数据特征。

对于 GPS 定位点 p 和路链 l 组成的样本, 可以提取 4 个后验概率组成的特征向量:

$$\vec{s} = \begin{pmatrix} p(\text{match}|\text{distance}), p(\text{match}|\text{angle}), \\ p(\text{match}|\text{speed}, \text{link} - \text{type}), \\ p(\text{match}|\text{state}, \text{link} - \text{type}) \end{pmatrix} \quad (3)$$

定位点 p 匹配到路链 l 的概率 $prob$ 定义为

$$prob = \omega^* \cdot \vec{s} \quad (4)$$

式中, ω^* 为归一化后的 SVM 最优分类面向量, 设分类阈值为 b^* , 即对于任何一个样本, 只要满足规则:

$$\omega^* \cdot \vec{s} + b^* \geq 0 \quad (5)$$

即认为 GPS 定位点成功地匹配到路链上。

3.4 路径推测算法

路径推测是浮动车数据处理过程中的关键技术之一, 即利用浮动车在不同路链上连续运动的轨迹点来搜寻车辆的真实行驶轨迹。经过路径推测过程, 浮动车数据就可以在时间和空间上同城市道路关联起来, 由车辆 GPS 定位点信息得到车辆在具体城市道路上的行驶状态, 从而反映出车辆所在道路的路况信息。

路径推测在本质上是路径搜索的过程, 即将当前定位点的候选匹配点作为起点, 搜索车辆可能经过的下一条或几条路链, 直到找到下一个定位点所在的路链, 从而确定两个定位点之间的车辆行驶轨迹。这可以归结为图的深度优先遍历问题, 文献[2]采用一种启发式的搜索方法, 对搜索过程进行剪枝, 提高了搜索效率。

将一个定位点的候选匹配点作为起点, 按时间顺序选择下一个定位点作为终点建立一条向量, 计算向量的方向与长度, 向量的方向为浮动车行驶路线的方向, 向量的长度为浮动车行驶路线的最小可能距离。设向量长度为 L_1 , 起点与其所在路链终点的距离为 L_2 , 如果 $L_1 < L_2$, 则说明车辆还未驶出当前路链, 否则说明车辆驶入当前路链的下一条或后续若干条路链。

当车辆驶入当前路链的后续路链时, 存在两种行车状态, 一种是保持直行(图 3 所示), 另一种是拐弯。在拐弯情况下, 向量和车辆所行驶过的路链形成类三角形, 如图 4 所示。按照一般道路建设的规划, 相邻路链的夹角大于 60° , 则根据三角形的性质, 向量为三角形最长边, 则车辆实际行驶的距离应当小于两倍的向量长度, 此性质可以作为路径搜索的一个极重要的约束条件。同时, 所搜索路链的方向与向量方向夹角应当小于 90° 。

依据上述约束条件, 路径推测算法可以在较短的时间内推测出车辆可能的行驶轨迹, 并计算出车

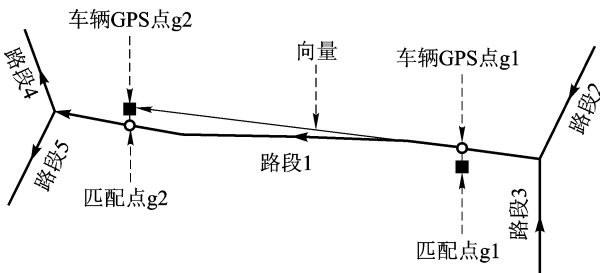


图 3 车辆直行示意图
Fig. 3 The sample of straightforward vehicles

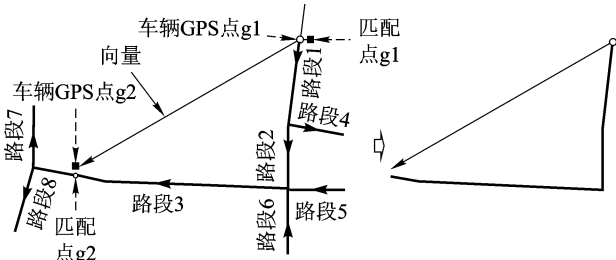


图 4 车辆跨路链转弯
Fig. 4 The sample of cross link and turning vehicles

辆的平均速度, 作为评估其途经路链的路况信息的基本标准。

3.5 并行路网结构模型

一些城市的环路和主干线均采用主辅路并行路网结构, 基于现有 GPS 定位精度很难将定位点准确匹配到主路或是辅路上, 进而产生多条路径推测结果, 导致计算出的路况信息不准确。

并行路网结构可以抽象如图 5 所示, 由于主辅路之间的距离远小于 GPS 的区分精度, 对于一个定位点而言, 其真实的坐标可能在主路上, 也可能在辅路上, 在这种情况下, 会产生大量候选路径集。在图 5 中, p_1 可以匹配到路段 1 和路段 6 上, p_2 可以匹配到路段 3 和路段 9 上, 则车辆的候选路径有:

$$path_1: \text{路段 1} \rightarrow \text{路段 2} \rightarrow \text{路段 3};$$

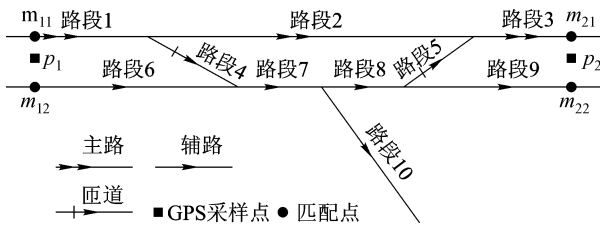


图 5 并行路网结构
Fig. 5 The structure of parallel road network

$path_2$: 路段 1 → 路段 4 → 路段 7 → 路段 8 → 路段 5 → 路段 3;

$path_3$: 路段 1 → 路段 4 → 路段 7 → 路段 8 → 路段 9;

$path_4$: 路段 6 → 路段 7 → 路段 8 → 路段 5 → 路段 3;

$path_5$: 路段 6 → 路段 7 → 路段 8 → 路段 9。

并行路网下路径推测的关键点是如何从候选路径中挑选出车辆的真实行驶轨迹。传统的策略主要基于某种规则,选择一条路径作为车辆的行驶轨迹,但缺乏足够的信息来保证这种选择的准确性,一旦路径选择错误,将给许多路链的路况计算造成消极影响。极端情况下,如果车辆的真实行驶轨迹是 $path_1$,而被错误地选择为 $path_5$,则 $path_1$ 和 $path_5$ 中所有路链的路况计算都将受到消极影响。

从融合的角度来讲,推测车辆的真实行驶轨迹,其目的在于将车辆的平均速度标记到轨迹中包含的路链上去,并以此作为获取路况信息的基础。在并行路网结构下,路径选择可以等价为从并行路网所覆盖的路链集中挑选车辆途经的路链子集,并将车辆的平均速度标记到这些路链上。

针对并行路网的特殊情况,在无法唯一确认车辆真实行驶轨迹的前提下,可以有第二种路径选择策略:将主路路径($path_1$)和辅路路径($path_5$)同时作为车辆的行驶轨迹输出。鉴于主路路径和辅路路径涵盖了并行路网的所有路链,故可以保证车辆的平均速度被标记到了其途经的路链上,从而避免了路径选择错误所带来的严重后果。同时,该方法也带来了相当数量的路链上标记的速度信息不精确的问题,但就目前的系统规模而言,环路和主干线上的浮动车密度足以保证在一个处理周期内有多辆浮动车先后经过同一路链,利用这些车辆所提供的速度信息的一致性,可以相对简单地对信息的可信度做出判断,进而摒弃不可信和不精确信息对融合过程的消极影响。通过引入 D-S 证据推理框架^[6-7],可以有效地解决信息不可信和信息冲突情况下的融合问题。

3.6 路况信息生成

随着浮动车规模的不断扩大,同一周期内往往有多辆浮动车先后经过同一条路链。由于各车辆提供的路况信息的时效性、准确性等存在差异,基于统计平均的多车辆交通信息融合机制不能准确体现非线性道路交通系统的时变特征。因此,路况信息的

生成必须考虑信息的可信度,以及合理处理信息之间的相互冲突。

D-S 证据推理^[6-7]理论是目前信息融合领域的主流算法,在智能交通领域有着广泛的应用。其最大的特点是对不确定信息的描述采用“区间估计”的方式,在区分不知道与不确定方面以及精确反映证据收集方面显示出很大的灵活性。当发生信息冲突时,它可以通过“悬挂”在所有目标集上共有的概念(可信度)使得发生的冲突获得解决,并保证原来高可信度的结果比低可信度的结果加权要大。

辨识框架是 D-S 理论的核心概念。为了降低推理过程的空间复杂度,保证浮动车信息融合的实时性,必须严格限制辨识框架的规模。浮动车系统对路况信息的表示有多种形式,可以是车辆通过路链的旅行时间或平均速度,也可以是粗粒度的路况等级。其中路况等级的定义包括:拥堵(车速 0 ~ 20 km/h);缓慢(车速 20 ~ 40 km/h);畅通(车速 40 km/h 以上)。因此,可以将拥堵等级定义为辨识框架,以保证融合的时效性;在获得正确路况等级的基础上,将包含在该等级内的车辆行驶速度进行平均,从而获得统计意义上的平均路况信息。

根据 D-S 证据推理理论,定义辨识框架为

$$\Theta = \{\text{畅通}, \text{缓慢}, \text{拥堵}\} \quad (6)$$

则有命题集为

$$2^\Theta = \left\{ \begin{array}{l} \emptyset, \{\text{拥堵}\}, \{\text{缓慢}\}, \{\text{畅通}\}, \\ \{\text{拥堵}, \text{缓慢}\}, \{\text{缓慢}, \text{畅通}\}, \\ \{\text{拥堵}, \text{畅通}\}, \{\text{拥堵}, \text{缓慢}, \text{畅通}\} \end{array} \right\} \quad (7)$$

式中, $\emptyset$ 代表空集,包含一个元素的命题为基元命题, $\{\text{拥堵}, \text{缓慢}, \text{畅通}\}$ 在下文中用 $\{\text{不明}\}$ 表示, $\emptyset$ 和 $\{\text{畅通}, \text{拥堵}\}$ 在本文无意义。

D-S 理论中的证据被定义如下:

$$S = \{s | s = vid, lid, spd, reliability\} \quad (8)$$

式中, vid 唯一标识一辆车, lid 标识一条路链, spd 为车辆通过该路链的平均速度, $reliability$ 是信息的可信度指标,由信息的时效性、车辆 GPS 定位点与路链的匹配概率,路径推测结果的确定性(如果路径推测结果唯一,则证据的可信度相对较高)等因素综合决定。

对于每一个证据 s_j ,其基本信任分配(basic belief assignment, BBA)函数 m_j 需满足:

$$\begin{aligned} m_j: 2^\Theta &\rightarrow [0, 1] \\ \sum_{A \in 2^\Theta} m_j(A) &= 1; m_j(\Phi) = 0 \end{aligned} \quad (9)$$

本文将对证据可信度的关注体现在 BBA 的分配上,如果证据的可信度较低,则命题{不明}上分配较大的 BBA,反之亦然。定义证据 s_j 在命题{不明}上的 BBA 为

$$m_j\{\text{不明}\} = 1 - \text{reliability} \quad (10)$$

在满足式(9)的前提下,其他命题的 BBA 可以根据车辆的平均速度来分配。

在获得证据的基本信任分配后,根据 D-S 融合算法^[8]即可以对同一路链上的多个证据进行融合操作,最终获得对各基元命题的支持度,决定路链的路况等级,然后将处在该等级范围内的所有速度进行平均,以得到车辆通过该路链的平均速度和期望旅行时间。

3.7 小结

在浮动车信息处理基本架构之下,结合本文列出的浮动车信息处理关键技术,可以将浮动车 GPS 定位数据与电子地图在时间和空间上联系起来,从而实现从浮动车 GPS 定位数据到路况信息的转变。在城市内部署一定规模的浮动车,能够动态获取城市内道路的实时路况信息,为市民出行提供引导服务。

4 数据评估

浮动车信息处理过程是一个有机的整体,各功能模块通过中间数据紧密地联系起来。对数据进行科学合理的评估,将有助于发现系统中存在的问题,为科学研究提供决策支持。作为结果数据,路况信息准确性的评估固然重要,然而对地图匹配及路径推测结果等中间数据的评估也十分必要。

4.1 地图匹配结果评估

在 3.3 节提到,地图匹配问题实质是一个二分类问题。在数据挖掘^[9]理论中,对于分类问题,比较通用的评估指标是查全率和查准率。

在分类器的设计中,系统提供了 11 240 个训练样本和 3 520 个测试样本,最终在测试样本集上获得的查全率和查准率分别为 97% 和 76%。

4.2 路径推测结果评估

为了评估路径推测的准确性,需要进行实际车辆测试。即从浮动车中指定测试车辆,并记录车辆在测试时段的实际行驶轨迹,与系统输出的该车辆路径推测结果进行比对,以评估路径推测模块的性能。

路径推测结果采用查全率和查准率指标评估,定义如下:

$$\text{recall} = \frac{\text{count}(LS_{\text{sys}} \cap LS_{\text{std}})}{\text{count}(LS_{\text{std}})} \quad (11)$$

$$\text{precision} = \frac{\text{count}(LS_{\text{sys}} \cap LS_{\text{std}})}{\text{count}(LS_{\text{sys}})} \quad (12)$$

式中, LS_{sys} 为系统输出的路径推测结果中包含的路链集合, LS_{std} 为测试车辆真实行驶轨迹中包含的路链集合, count 操作计算路链集合中包含的路链数。路测结果表明,应用并行路网结构下新的地图匹配和路径选择策略以后,系统路径推测的查全率和查准率均有显著提升,目前查准率平均为 83.7%,查全率达到 95.3%。

4.3 路况信息质量评估

在评估路况信息准确性的过程中,为了保证待测数据与标准数据之间的独立性,测试车辆独立于系统提供数据的浮动车。测试车辆按照事先划定的路线进行跑车测试,每辆测试车配备一名测试人员,负责操作手持式 GPS 采集设备,记录高密度和高精度的 GPS 数据,从而获取测试车辆进出路链的准确时间,并以此作为系统评测所参考的标准路况信息数据。

路测结束以后,系统处理测试期间的历史数据,将得到的路况信息分别与标准路况信息进行准确性的对比。评估的基本单位为路链,考察系统输出的路况信息与标准路况信息一致的路链占测试路线包含的路链总数的比例。评价指标包括定性的拥堵等级,定量的旅行时间和平均车速。对于定量的路况信息,根据相对误差率来确定其与标准路况信息的一致性:

旅行时间相对误差率

旅行时间相对误差率 =

$$\frac{|\text{系统输出的旅行时间} - \text{标准旅行时间}|}{\text{标准旅行时间}} \times 100\% \quad (13)$$

平均速度相对误差率

平均速度相对误差率 =

$$\frac{|\text{系统输出的平均速度} - \text{标准平均速度}|}{\text{标准平均速度}} \times 100\% \quad (14)$$

本文设定,当相对误差率在 $[0, t)$ 以内时,可以认为系统输出的路况信息与标准路况信息一致。相应地,路况信息的准确性定义如下:

旅行时间准确性

旅行时间准确性 =

$$\frac{\text{旅行时间相对误差率小于 } t \text{ 的路链数}}{\text{测试路线包含的路链总数}} \times 100\% \quad (15)$$

平均速度准确性

平均速度准确性 =

$$\frac{\text{平均速度相对误差率小于 } t \text{ 的路链数}}{\text{测试路线包含的路链总数}} \times 100\% \quad (16)$$

拥堵等级准确性

拥堵等级准确性 =

$$\frac{\text{系统拥堵等级与实测拥堵等级相同的路链数}}{\text{测试路线包含的路链总数}} \times 100\% \quad (17)$$

以 2008 年 6 月份环路的测试结果为例,实验结果表明,主路的路况结果计算要优于辅路,故路况信息准确性评估主辅路分开进行。去除匝道与过短的路链,测试路线中包含 119 条主路路链和 74 条辅路路链。根据路况信息准确性的 3 个评估指标,统计结果如表 1 所示。

表 1 路况信息准确性评估

Tab. 1 The precision of traffic condition

路链类型	准确性指标		
	旅行时间(%)	平均速度(%)	拥堵等级(%)
主路($t=30\%$)	85.71	87.39	88.24
辅路($t=40\%$)	66.20	67.16	56.76

其中,表中的 t 在式(15)和式(16)中定义。这里需要说明的是,由于系统输出的旅行时间、平均车速是经多辆浮动车提供的路况信息统计平均得到的,而标准路况信息仅由测试车辆提供,故 t 的取值在很大程度上应当体现个体数据与统计数据的差别。由于辅路受交通灯和行人的影响较大,各车辆(包括测试车辆)提供的路况信息之间的差异度较主路要明显得多,故其路况准确性评估的指标较主路应适当放宽。据此,在评估主路融合结果时设定 $t=30\%$,评估辅路融合结果时设定 $t=40\%$ 。

当车辆行驶在主路上时,车辆的行驶速度比较稳定,车辆之间的速度差异也不会很大。同时,主路的交通流变化比较平滑,因此,在主路上可以相对容易地获得高质量的路况信息。

辅路的情况要复杂得多,车辆拐弯、红绿灯和行人都可能造成车辆速度的急剧变化,因此,各车辆提供的路况信息差异性相对较大,即使是测试车辆也不一定能够代表实时路况信息。例如,当测试车排队上主路时,辅路外侧车道依旧保持较高车速,此时标准路况信息也可能与实际情况略有出入。综上,辅路路况信息质量与主路还存在较大差距,需要寻

求一种更为科学的评估方法。同时,对车辆在辅路上的行驶特征进行分析,可以获取更为精确的辅路路况信息。

5 结 论

浮动车信息处理技术的核心功能都是以交通参数信息的应用为基础展开的,服务的质量关键取决于能否获取准确、有效和实时的与交通运行状态有关的信息。针对交通参数信息数量和质量存在的不足,以及影响交通信息服务质量进一步提高的几个问题,本文介绍了浮动车信息处理过程中应用的关键算法。在浮动车信息处理架构下应用上述算法,可以完成从浮动车 GPS 定位数据到路况信息的转换。

在环路辅路以及其他城市主干道路上,由于存在交通灯控制,导致路况计算出现大量随机性的、不反映交通流本身特性的干扰性波动。这种波动不仅导致实时路况信息计算不准确,对旅行时间的计算也造成了很大的影响。问题的实质在于正常的交通流中混入了人工控制的成分,导致系统的处理结果不能很好地反应实际交通流的状况。解决问题的关键在于,将交通灯引起的控制延误和交通流本身的非控制延误相分离,用交通流本身的状况来衡量实时路况,同时将两种延误相结合来计算旅行时间。

由于浮动车行驶的随机性和浮动车数量不足,使得在同一个处理周期内,部分道路因没有浮动车经过而无法计算路况信息。信息填补是在生成路况信息过后,根据历史数据以及基于当前交通流的分析,对那些浮动车没有覆盖到的路链的路况进行计算,从而获得全面的道路路况信息。

交通灯对浮动车信息处理的影响、路况信息填补和预测,以及建立一套路况信息准确性的科学的评估体系,是浮动车信息处理技术下一步的工作重点。

参考文献 (References)

1 Brakatsoulas S, Pfoser D, Salas R, *et al.* On map-matching vehicle tracking data[A]. In: Proceedings of the 31st VLDB Conference [C], Trondheim, Norway, 2005: 853-864.

2 Lv Wei-feng, Zhu Tong-yu, Wu Dong-dong, *et al.* A heuristic path-estimating algorithm for large-scale real-time traffic information calculating[J]. Science in China Series E-Technological Sciences, 2008, 51(13): 165-174.

3 Zhu Tong-yu, Guo Sheng-min, Lü Wei-feng. SVM based float car driving mode classification model[J]. Journal of Beijing University of Aeronautics and Astronautics, 2008, **34**(8): 976-980. [诸彤宇,郭胜敏,吕卫锋. 基于 SVM 的浮动车行驶模式判断模型[J]. 北京航空航天大学学报, 2008, **34**(8): 976-980.]

4 Vapnik V N. The Nature of Statistical Learning Theory[M]. New York, USA: Springer-Verlag, 1995: 138-141.

5 Bian Zhao-qi, Zhang Xue-gong. Pattern Recognition[M]. Beijing: Tsinghua University Press, 2000. [边肇祺,张学工. 模式识别[M]. 北京: 清华大学出版社, 2000.]

6 Dempster A P. Upper and lower probabilities induced by a multivalued mapping[J]. Annals of Mathematical Statistics, 1967, **38**(2): 325-339.

7 Shafer G. A Mathematical Theory of Evidence[M]. Princeton NJ, USA: Princeton University Press, 1976.

8 He Bing, Mao Shi-yi, Zhang You-wei, *et al.* Evidence combination and decision based on DS evidence theory and evidence classification [J]. Journal of Electronics and Information Technology, 2002, **24**(7): 894-899. [何兵,毛士艺,张有为等. 基于证据分类的 DS 证据合成及判决方法[J]. 电子与信息学报,2002, **24**(7): 894-899.]

9 Han J W, Kamber M. Data Mining Concepts and Techniques[M]. Los Altos CA, USA: Morgan Kaufmann, 2001: 132-136.