并行路网结构下的浮动车信息处理模型

诸彤宇 郭胜敏 吕卫锋 王技方

(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: 受 GPS定位精度的限制,浮动车在主辅路并行路网结构下进行地图匹配和路径推测存在困难,直接影响路况信息获取的准确性. 在现有浮动车信息处理框架基础上,运用统计学方法提升地图匹配的精度,并将主辅路径同时作为车辆的行驶轨迹输出,有效降低了路径推测错误的可能性. 最后,引入证据推理框架以解决路径选择策略调整所带来的信息不可信和信息冲突问题. 对比实验表明,改进系统在主路和辅路上的路况准确性分别有 14. 26%和 9. 46%的提升.

关 键 词: 浮动车; 并行路网结构; 证据推理理论; 信息融合

中图分类号: TP 391. 4

文献标识码: A **文 章 编 号**: 1001-5965(2009)10-1232-05

FCD based information processing model in parallel road network

Zhu Tongyu Guo Shengmin Lü Weifeng Wang Jifang

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

Abstract: Limited by the poor GPS locating precision, it is quite difficult to implement map matching and path deriving procedures in the parallel road network by FCD (floating car data) technology, which affects the traffic information accuracy directly. On the basis of existing FCD information processing framework, some approaches were applied to improve map matching performance. And then, the freeway and service road paths were both viewed as vehicle trajectories, so as to reduce the probability of inaccurate path deriving effectively. Finally, the evidential reasoning framework was introduced to address the information unlikelihood and information conflict issues after the path selecting strategy adjustment. Contrast experiment shows that the traffic information accuracy on freeway and service road was improved about 14. 26% and 9. 46% respectively after employing our modified model.

Key words: floating car data; parallel road network; evidential reasoning theory; information fusion

浮动车技术是近年来先进交通信息系统中获取实时路况信息的先进技术手段之一. 其基本原理是通过装备车载全球定位系统 GPS和无线通信系统的浮动车在其行驶过程中发送车辆的位置、方向和速度等信息,在处理中心应用地图匹配、路径推测[1]和信息融合[2]等算法,最终获得浮动车所经道路的交通拥堵信息.

虽然基于浮动车的动态交通信息处理系统在理念和技术上具有先进性,但是受 GPS定位精度的影响,对于主辅路和立交桥等复杂的路网结构,仍亟待大量的基础研究工作来满足系统在信息处理准确性方面的需求. 以北京城市道路结构为例,环路和主干线多采用主辅路并行路网结构 (PRN, Parallel Road Network),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. 在这种情况下,往往产生错误的路径推测结果,进而导致计算出的路况信息不准确. 鉴于环路和主干线在城市路网中的重要作用,准确获取并行路网结构下的路况信息,已成为动态交通信息处理中的一个重要研究课题.

收稿日期: 2008-10-14

基金项目: 国家 863基金资助项目 (2006AA12Z315)

作者简介: 诸彤宇 (1969 -),男,北京人,副教授, zhutongyu@nlsde. buaa. edu. cn.

本文在现有浮动车信息处理框架的基础之上,优化了地图匹配算法,并对路径选择策略进行了调整,将并行路网结构下的浮动车信息处理问题转化为存在证据冲突的信息融合问题,进而引入了一种 D-S证据推理融合算法.

1　并行路网结构

浮动车信息处理需要两个必不可少的要素:路网结构[1]表示的电子地图及浮动车 GPS定位点.并行路网结构下浮动车信息处理模型的描述也需要对路链 (文中路网的原子结构)和 GPS定位点进行抽象定义.

定义 1　路链

路链是路网拓扑[1]中存在的一条有向通路,具有若干拓扑属性.

$$L = \{ l \mid l = (l_{\mathrm{id}}, l_{\mathrm{direction}}, l_{\mathrm{length}}, l_{\mathrm{type}}) \} \quad (1)$$

其中, l_{id}唯一标识一条路链; $l_{\mathrm{direction}}$和 l_{length}分别指路链的方向和长度; l_{type}指主路或辅路.

定义 2　GPS定位点

$$G = \{ g \mid g = (v_{\mathrm{id}}, g_{\mathrm{time}}, \langle x, y \rangle, g_{\mathrm{direction}}, g_{\mathrm{speed}}, g_{\mathrm{state}}) \} \quad (2)$$

其中, v_{id}唯一标识一辆浮动车; g_{time}, $\langle x, y \rangle$, $g_{\mathrm{direction}}$, g_{speed}和 g_{state}分别是该浮动车原始数据采集时的时间、经纬度坐标、车辆行驶方向、即时速度和载客状态.

定义 3　并行路网结构

并行路网结构的定义涉及两组路链.设 g_1和 g_2是同一辆浮动车两个连续的 GPS定位点,$LS_1 = \langle l_{11}, l_{12}, \cdots, l_{1m} \rangle$和 $LS_2 = \langle l_{21}, l_{22}, \cdots, l_{2n} \rangle$分别是构成主路和辅路的两组路链序列.如果 g_1可以匹配到 l_{11}和 l_{21}上, g_2可以匹配到 l_{1m}和 l_{2n}上,且 LS_1和 LS_2平行分布,则认为 LS_1和 LS_2组成并行路网结构,可以抽象如图 1所示.

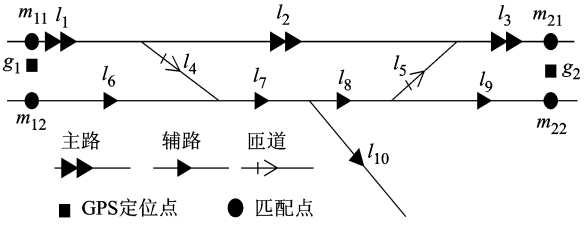


图 1　并行路网结构

在并行路网结构下,由于主辅路之间的距离远小于 GPS的圆概率误差,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,会产生数目可观的候选路径集.如图 1所示, g_1可以匹配到 l_1和 l_6上, g_2可以匹配到 l_3和 l_5上,则车辆的候选路径有:

path_1: $l_1 \rightarrow l_2 \rightarrow l_3$;

path_2: $l_1 \rightarrow l_4 \rightarrow l_7 \rightarrow l_8 \rightarrow l_5 \rightarrow l_3$;

path_3: $l_1 \rightarrow l_4 \rightarrow l_7 \rightarrow l_8 \rightarrow l_9$;

path_4: $l_6 \rightarrow l_7 \rightarrow l_8 \rightarrow l_5 \rightarrow l_3$;

path_5: $l_6 \rightarrow l_7 \rightarrow l_8 \rightarrow l_9$.

路径推测的关键问题是如何从为数众多的候选路径中挑选出车辆的真实行驶轨迹.传统的策略基于某种规则,选择一条路径作为车辆的行驶轨迹,即使没有足够的信息来保证这种选择的准确性.这样做是极具风险的,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影响.

从融合的角度看,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,忽略连接主辅路之间的匝道,在任何轨迹下车辆的实际行驶距离均相差无几,故其平均速度也大体一致.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,本文提出第 2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.诚然,此举的一个副作用是同时也有相当数量的路链上标记的速度信息是不精确的,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.因此,本文引入了 D-S证据推理框架[3],以解决信息不可信和信息冲突情况下的融合问题,并在实际应用中取得了良好的效果.

2　*PRN* 下的浮动车信息处理模型

为了实现并行路网结构下实时路况的准确获取,本文优化了地图匹配算法,提出了一种路径双选策略,并将 D-S证据推理框架引入了浮动车信息处理模型.

2.1 优化的地图匹配算法

地图匹配是浮动车信息处理的一个关键步骤,对于并行路网结构而言,地图匹配的重要性就更加不言而喻. 在图 1中,如果 g_1匹配到 l_1 上的 m_{11}, g_2匹配到 l_2 上的 m_{22},则车辆可能的行驶轨迹就仅有一条. 本文在此提出了一种基于支持向量机 (SVM, Support Vector Machines) [4]分类的地图匹配算法.

传统的地图匹配算法[1]只考虑了 GPS定位点与路链之间的位置参考信息,而没有利用路链特性与车辆行驶特性之间的内在联系,如车辆在高速或满载的情况下一般在主路上行驶. 因此,在原有基础上,本文将地图匹配考虑的因素拓展到4个: ①定位点到路链的投影距离 d; ②车辆行驶方向与路链方向夹角 θ; ③即时车速 g_{speed}; ④车辆载客状态 g_{state}.

定义 GPS定位点和路链关系对 $\langle g, l\rangle$为一个样本,则地图匹配问题可以看作一个分类问题,即根据已知的数据特征将 g和 l的关系分为匹配 m 和不匹配 $\bar{m}$ 两种. 为了提高分类器设计的质量,本文分别提取上述 4个因素对应于匹配关系的后验概率作为特征向量:

$$s = (p(m \mid d), p(m \mid g_{\text{speed}}, l_{\text{type}}), p(m \mid \theta), p(m \mid g_{\text{state}}, l_{\text{type}})) \quad (3)$$

则车辆 GPS定位点与路链相匹配的概率为

$$p_m = \omega^* \cdot s \quad (4)$$

其中,ω^*为归一化后的 SVM最优分类面向量. 设 b^*为归一化后的分类阈值,如果 $p_m \geq b^*$,则认为车辆成功匹配到路链上.

2.2 路径双选策略

提高地图匹配的精度可以大幅缩小候选路径集的规模,但在多数情况下依旧无法确认车辆的唯一行驶路径. 因此,必须从候选路径集中挑选主路和辅路路径分别输出.

路径是由一系列路链所组成的,虽然路链的 l_{type}属性可以标识其主辅路特性,但由于路链的长度不一,在车辆行驶距离一定的情况下,不同路径推测结果中包含的路链数目往往是不同的,很难基于路链属性给出主辅路径的确切定义. 特别的,当车辆途经立交桥等复杂路网结构时,由于一些短路链的 l_{type}属性定义不合理,致使路径选择的效果不佳.

因此,本文在选取路径的时候,不仅考虑了路链的 l_{type}属性,还考虑了路径的长度. 设路径候选集为 $\{\text{path}_1, \text{path}_2, \cdots, \text{path}_m\}$,以主路路径为例,路径选取规则如下:

1) 对路径 path_i ($i = 1, 2, \cdots, m$)计算其主路路链占所有路链的比例在所有路径候选中的排名 a_i;

2) 计算 path_i的长度在候选路径中的排名 b_i;

3) 计算总排名 $c_i = p \times a_i + (1 - p) \times b_i$;

4) 总排名 c_i ($i = 1, 2, \cdots, m$)值最小的路径为主路路径.

目前,权重 p的设置只是基于经验,亦可采用机器学习算法以获取精确的权重值. 辅路的选取规则与主路类似. 路径选择的结果是车辆可能途径的 1条或 2条完整的行驶轨迹,由匹配结果单元 (MRU, Matched Result Unit)表示.

定义 4 匹配结果单元

$$\Psi = \{\psi \mid \psi = (v_{\text{id}}, l_{\text{id}}, \psi_{\text{speed}}, \psi_{\text{time}}, \psi_{\text{m}}, \psi_{\text{r}})\} \quad (5)$$

其中, v_{id}唯一标识一辆浮动车; l_{id}标记一条路链; ψ_{time}和 ψ_{speed}分别是车辆经过该路链的时刻和平均速度; ψ_{m} 反映了结果单元的地图匹配概率; ψ_{r} 是信息可信度的一个度量,如果唯一确认车辆的行驶轨迹,或主辅路径在某些 ψ 上重合,则对应的 ψ具有高可信度. 在信息融合阶段,通过对同一路链上所有车辆的 ψ进行融合,可以获取该路链相对准确的实时路况信息.

2.3 信息融合模型

前文提到,一条路链上车辆提供的速度信息可能是不精确的,相互之间也可能是冲突的. 因此,融合过程必须判断信息的可信度,以及合理地处理信息之间的相互冲突.

D-S证据推理[3]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[5-6]. 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.

2.3.1 证据推理框架

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,实现浮动车信息融合的实时性,必须严格限制辨识框架的规模.

浮动车技术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.

定义 5　辨识框架

$$\Theta = \{\text{缓慢},\text{拥堵},\text{畅通}\} \tag{6}$$

定义 6　命题

根据 D-S理论,命题是辨识框架幂集的子集:

$$2^{\Theta} = \{\varnothing, \{\text{拥堵}\}, \{\text{缓慢}\}, \{\text{畅通}\}, \{\text{拥堵},\text{缓慢}\}, \{\text{缓慢},\text{畅通}\}, \{\text{拥堵},\text{畅通}\}, \{\text{拥堵},\text{缓慢},\text{畅通}\}\} \tag{7}$$

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

定义 7　证据

每一个匹配结果单元被定义为一个证据,即

$$S = \{s \mid s = \psi, \psi \in \Psi\} \tag{8}$$

证据是信息融合的基本操作单元.

定义 8　基本信任分配

对于每一个证据 s_j,其基本信任分配 (BBA, Basic Belief Assignment)函数 m_j需满足:

$$\left.\begin{aligned} m_j &: 2^{\Theta} \rightarrow [0, 1] \\ \sum_{A \subset 2^{\Theta}} m_j(A) &= 1 \\ m_j(\varnothing) &= 0 \end{aligned}\right\} \tag{9}$$

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

$$m_j\{\text{不明}\} = \sum_{i=1}^{I} \alpha_i (1 - f_i(s_j)) \tag{10}$$

其中,I为判断证据可信度时所考虑的因素个数;α_i ($i = 1, 2, \cdots, I$)是各因素对应的权重因子. 这里考虑的因素主要有 ψ 的时效性和可信度等,$f_i(s_j)$ ($i = 1, 2, \cdots, I$)是系统对各因素的度量值.

在满足公式 (9)的前提下,其它命题的 BBA可以根据车辆的平均速度来分配. 需要注意的是,如果车辆速度接近路况等级的分界线时,对于非基元命题 (如速度为 19 km/h对{拥堵,缓慢})的 BBA分配要适当提高.

在获得证据的基本信任分配后,融合算法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.

2.3.2　证据融合算法

虽然 D-S理论为不确定性信息的表达和合成提供了自然而有力的方法,但其组合规则却被证明存在漏洞. 因此,本文选用文献[7]提出的一种基于证据分类的融合方案.

分类策略以证据对基元命题的支持程度作为分类指标,设命题集为 $\{A_i,\ i = 1, 2, \cdots, M\}$,证据集为 $\{S_j,\ j = 1, 2, \cdots, N\}$,基元命题集为 $\{B_k,\ k = 1, 2, \cdots, K\}$. 证据 S_j对基元命题 B_k的支持度为

$$D_j(B_k) = \sum_{A_i \cap B_k \neq \Phi} \frac{m_j(A_i)}{J(A_i)} \tag{11}$$

其中 $J(A_i)$是命题 A_i中包含的基元命题的个数.

融合的过程为:①根据支持度将证据分为 L类 ($L \leq K$),以避免相互冲突的证据进行融合;②对每个分类内的证据按照经典的 D-S融合规则进行融合,可以得到一组合成证据,设为 m_f^l,$l = 1, 2, \cdots, L$;③对 m_f^l ($l = 1, 2, \cdots, L$)进行加权平均,获得最终的合成证据 m_F,计算 m_F对各基元命题的支持度,选取支持度最大的命题为路况等级判定的结果.

在文献[7]中,对融合过程第③步的各合成证据的加权因子主要考虑各分类中证据的个数,本文采用证据的 ψ_r和 ψ_m属性来共同决定证据在融合过程中的权重,即高可信度证据加权为 1,其它证据用 ψ_m加权.

对于并行路网结构,采用改进的地图匹配算法,如果确认车辆的行驶路径,则输出唯一的车辆行驶轨迹;否则按照路径双选策略分别输出主辅路径. 车辆的行驶轨迹信息用匹配结果单元表示并标记其可信度. 在证据推理框架下,应用基于证据分类的信息融合方法,首先获得路链的路况等级,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　实验验证

3.1　*SVM* 分类训练和评估

本文提供了 11 240个训练样本,3 520个测试样本. SVM训练的结果为:分类面法向量 $\omega^* = (0.432, 0.228, 0.232, 0.048)$,阈值 $b^* = 0.592$

在本文的应用中,一个 GPS定位点可能匹配到多条路链上,因此,本文选用两个阈值 κ_m和 κ_s将样本空间按照匹配概率 p_m划分为 3个等级,并认为如果存在高匹配等级的样本,则不考虑低等级样本的匹配关系. 通过设定 $\kappa_m = 0.650$, $\kappa_s = 0.410$,当查全率保持在 0.97的水平时,分类器的查准率从 0.56上升到 0.76.

3.2　模型评估

为了评估并行路网下浮动车信息处理质量,

实验安排若干辆浮动车沿着北京市三环主辅路行驶. 通过高密度和高精度的采样,可以获取车辆进出路链的准确时间,并以此作为并行路网结构下路况信息准确性评估的标准.

3.2.1 评估指标

在浮动车技术中,路链旅行时间(即车辆通过一条路链的时间期望)被广泛用于表示路况信息. 因此,本文将旅行时间相对误差率作为评估并行路网模型下浮动车信息准确性的指标. 设测试车辆通过路链的时间为 t_1,融合获得的旅行时间期望为 t_2,则旅行时间相对误差率为

$$e = | t_2 - t_1 | / t_1 \qquad (12)$$

3.2.2 准确性评估

除去匝道等过短的路链,测试路线中包括119条主路路链和 74条辅路路链. 从实验结果看,主路路链的融合效果要远远好于辅路路链. 因此,本文融合效果的评估主辅路链分开进行.

目前系统采用的是一种基于统计平均的融合方法,即通过平均所有车辆经过某条路链的旅行时间来获得当前周期内途经该路链的期望旅行时间. 本文将其作为对比方案,原有系统与改进系统旅行时间误差率对比效果如图 2所示.

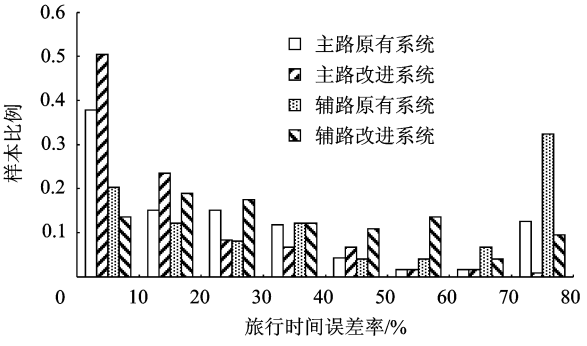


图 2 旅行时间准确性评估分布

3.2.3 误差分析

由图 2可见,统计旅行时间误差率在 30%以内的测试样本情况,改进后系统在主路和辅路上的路况信息准确性分别有 14.26%和 9.46%的提升. 其中,主路上的信息处理准确性优于辅路.

当车辆行驶在主路上时,车辆行驶速度比较稳定,而且车辆之间的速度差异也不是很大. 但是辅路上的情况要复杂的多,车辆拐弯、红绿灯和行人都是造成车辆在辅路上速度急剧变化的主要原因,即使测试车辆也不一定能够代表辅路上的真实路况,故在辅路上测试效果较差.

4 结束语

本文基于目前的浮动车信息处理框架,在地图匹配和路径选择等关键功能模块上对现有系统进行改进,将并行路网结构下的浮动车信息处理问题转化为存在信息冲突的融合问题,并据此引入了一种基于证据分类的 D-S融合算法. 实验结果表明,采用本文提出的模型后,系统在主路及其辅路上的路况准确性分别提高 14.26%和 9.46%. 对车辆在辅路上的行驶特性进行分析,以获取更为精确的辅路路况信息,是本文下一步的工作目标.

参考文献 (References)

[1] Wu D D, Zhu T Y, Lü W F, et al. A heuristic map-matching algorithm by using vector-based recognition[C]//Proc of the International Multi-Conference on Computing in the Global Information Technology (ICCGI'07). Piscataway, NJ: IEEE, 2007: 18 - 24

[2] 吕卫锋,诸彤宇,薛明,等. 一种交通信息融合处理方法和系统:中国. 200610168271[P]. 2006 - 06 - 06
Lü Weifeng, Zhu Tongyu, Xue Ming, et al. A traffic information fusion processing method and system: China. 200610168271[P]. 2006 - 06 - 06 (in Chinese)

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

[4] Vapnik V N. The nature of statistical learning theory[M]. New York: Springer-Verlag, 1995: 138 - 141

[5] 杨兆升. 基础交通信息融合技术及其应用[M]. 北京:中国铁道出版社, 2005: 134 - 141
Yang Zhaosheng. Traffic information fusion technology and application[M]. Beijing: China Railway Publishing House, 2005: 134 - 141 (in Chinese)

[6] 苏惠敏,周鹏,陈哲,等. 基于 D-S证据理论的 GPS/MM组合系统车辆定位算法[J]. 北京航空航天大学学报, 2001, 27(2): 157 - 160
Su Huimin, Zhou Peng, Chen Zhe, et al. Study of GPS/MM integrated navigation system for vehicle positioning based on D-S evidence reasoning[J]. Journal of Beijing University of Aeronautics and Astronautics, 2001, 27(2): 157 - 160 (in Chinese)

[7] 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
He Bing, Mao Shiyi, Zhang Youwei, 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 (in Chinese)

(编 辑:文丽芳)