



(12)发明专利

(10)授权公告号 CN 104331422 B

(45)授权公告日 2018.07.10

(21)申请号 201410542082.2

(22)申请日 2014.10.14

(65)同一申请的已公布的文献号
申请公布号 CN 104331422 A

(43)申请公布日 2015.02.04

(73)专利权人 广州市香港科大霍英东研究院
地址 511458 广东省广州市南沙区南沙资
讯科技园软件楼N301室

(72)发明人 丁焯 谭浩宇 高民 倪明选

(74)专利代理机构 广州三环专利商标代理有限
公司 44202

代理人 麦小婵 郝传鑫

(51)Int.Cl.
G06F 17/30(2006.01)

(56)对比文件

CN 101495835 A,2009.07.29,说明书第2-13页.

US 2002/0193938 A1,2002.12.19,

CN 1761980 A,2006.04.19,

US 2013/0304379 A1,2013.11.14,

US 2009/0138497 A1,2009.05.28,

刘经南等.位置大数据分析处理研究进展.《武汉大学学报信息科学版》.2014,第39卷(第4期),379-385.

Yu Zheng et al..Mining Interesting Locations and Travel Sequences from GPS Trajectories.《Proceedings of the 18th international conference on World wide web 》.2009,第2009年卷791-800.

审查员 郑岩

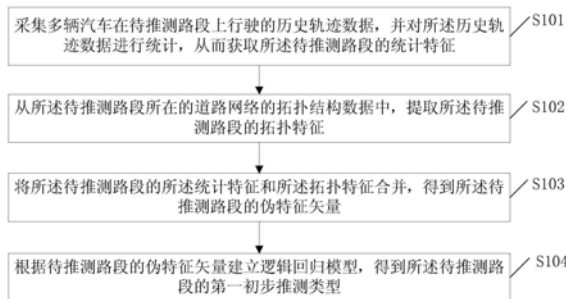
权利要求书2页 说明书10页 附图4页

(54)发明名称

一种路段类型推测方法

(57)摘要

本发明公开了一种路段类型推测方法,包括:采集多辆汽车在待推测路段上行驶的历史轨迹数据,并对所述历史轨迹数据进行统计,从而获取所述待推测路段的统计特征;从所述待推测路段所在的道路网络的拓扑结构数据中,提取所述待推测路段的拓扑特征;将所述待推测路段的所述统计特征和所述拓扑特征合并,得到所述待推测路段的伪特征矢量;根据待推测路段的伪特征矢量建立逻辑回归模型,得到所述待推测路段的第一初步推测类型。本发明推测精确性更高,使用更灵活。



1. 一种路段类型推测方法,其特征在于,包括:

采集多辆汽车在待推测路段上行驶的历史轨迹数据,并对所述历史轨迹数据进行统计,从而获取所述待推测路段的统计特征;

从所述待推测路段所在的道路网络的拓扑结构数据中,提取所述待推测路段的拓扑特征;

将所述待推测路段的所述统计特征和所述拓扑特征合并,得到所述待推测路段的伪特征矢量;

根据待推测路段的伪特征矢量建立逻辑回归模型,得到所述待推测路段的第一初步推测类型;

从所述道路网络的拓扑结构数据中,获取所述待推测路段与所述待推测路段的相连路段之间的连接角度;

根据获取的连接角度和所述待推测路段的相连路段的类型进行推测,得到所述待推测路段的第二初步推测类型;

根据所述第一初步推测类型和第二初步推测类型,采用集成算法计算得到所述待推测路段的最终推测类型。

2. 如权利要求1所述的路段类型推测方法,其特征在于,所述采集多辆汽车在待推测路段上行驶的历史轨迹数据,并对所述历史轨迹数据进行统计,从而获取所述待推测路段的统计特征,具体包括:

采集多辆汽车行驶的历史轨迹数据;

采用ST-Matching算法将所述多辆汽车行驶的历史轨迹数据与所述道路网络匹配,获得多辆汽车在待推测路段上行驶的历史轨迹数据;

对所述多辆汽车在待推测路段上行驶的历史轨迹数据进行统计,从而获取所述待推测路段的统计特征。

3. 如权利要求1所述的路段类型推测方法,其特征在于,在将所述待推测路段的所述统计特征和所述拓扑特征合并,得到所述待推测路段的伪特征矢量之后,还包括:

采用主成分分析法降低所述伪特征矢量的维度。

4. 如权利要求1所述的路段类型推测方法,其特征在于,在从所述道路网络的拓扑结构数据中,获取所述待推测路段与所述待推测路段的相连路段之间的连接角度之前,还包括:

根据道路网络的拓扑结构数据和道路网络中已知路段类型的路段,采用贝叶斯分类器进行学习,得到一个多项式分布;其中,所述多项式分布是当相连接的两个路段为不同的连接角度时,当其中一个路段为不同类型时,另一路段类型的概率分布。

5. 如权利要求1所述的路段类型推测方法,其特征在于,所述根据获取的连接角度和所述待推测路段的相连路段的类型进行推测,得到所述待推测路段的第二初步推测类型,具体包括:

根据获取的连接角度和所述待推测路段的相连路段的类型,利用多项式分布,采用贝叶斯算法计算得到所述待推测路段的第二初步推测类型。

6. 如权利要求1所述的路段类型推测方法,其特征在于,所述集成算法为堆叠泛化算法、支持向量机算法、随机森林算法中的任一种。

7. 如权利要求1所述的路段类型推测方法,其特征在于,所述多辆汽车具体为多辆出租

车;则所述统计特征包括出租车行驶速度、空闲出租车密度、载客出租车密度和载客事件次数。

8.如权利要求1所述的路段类型推测方法,其特征在于,所述拓扑特征包括路段长度、累积摆动值、相连路段和相邻路段。

一种路段类型推测方法

技术领域

[0001] 本发明涉及众包地图数据处理领域,尤其涉及一种路段类型推测方法。

背景技术

[0002] 近几年,众包地图服务已经成为公共商业地图服务提供商(例如谷歌地图)的一个有力的竞争者。不同于由小数量的专业人员通过遥感影像和测量数据生成的商业地图服务,众包地图是由成千上万的注册用户维护的,这些用户通过精密的地图编辑器不断地创建和更新地图,例如增加、删除和修改道路。因此众包地图服务比现有的商业地图服务更符合当今地图服务变化趋势。例如,据报道,世界最大的众包地图项目开放街道地图(Open Street Map,OSM),相比其他所有的数据集,可以提供更丰富和更及时的地图更新数据。

[0003] 与其他众包应用相似,众包地图服务依赖于众多用户的贡献,但是这些贡献是易于出错的,而且有严重的一致性问题的。其主要的原因是地图的主题(例如,道路和地区)通常是很复杂的,这样会导致制作既特征丰富又容易使用的地图编辑器很难。为了解决这个问题,现有技术提出了一个名字为CrowdAtlas的地图更新系统,用于通过大量的历史车辆轨迹探测地图的改变。CrowdAtlas可以从轨迹中自动的生成新的或已经改变的道路的形状,生成的道路形状作为地图编辑器的推荐,用户可以直接使用生成的道路或者根据经验稍稍调整,从而减少用户绘制道路的耗时。

[0004] 为了使众包地图的地图数据能够更适用于通用应用(例如导航系统和旅游规划服务),就不仅需要提供道路网络的拓扑结构和道路形状,还需要提供路段类型(例如,高速干道、主/次干道、住宅区道路等)。其中,通常一条道路包括若干路段,这些路段有可能是不同的类型。另一方面,为减少手动编辑地图的耗时,还需要提供合适的推荐以供用户选择。

[0005] 现有技术主要是根据大量的车辆历史轨迹生成道路形状,但是还没有解决根据车辆历史轨迹推测路段类型的问题。但是,道路的元数据对于很多基于地图的应用(例如导航系统和旅游规划服务)是非常重要的。典型的道路的元数据包括宽度、速度限制、方向限制和进入限制。这些元数据可以通过路段的类型有效地反应出来。例如:高速干道比次级干道的速度限制更高;高速干道或主干道通常是双向道路,但是住宅区道路很可能是单向道。因此,为了提高众包地图服务的质量,用户不仅需要提供道路形状,还需要提供路段类型。那么,为更进一步减少用户更新众包地图的耗时,就需要使道路类型的分类过程自动化。另外,虽然道路类型有可能直接从道路网络拓扑结构推测出来,例如,有相同方向的路段也许有相同的类型。但是,这种推测通常不是很精确。

发明内容

[0006] 本发明所要解决的技术问题在于,提供一种路段类型推测方法,其精确性更高。

[0007] 为了解决上述技术问题,本发明提供了一种路段类型推测方法,包括:

[0008] 采集多辆汽车在待推测路段上行驶的历史轨迹数据,并对所述历史轨迹数据进行统计,从而获取所述待推测路段的统计特征;

[0009] 从所述待推测路段所在的道路网络的拓扑结构数据中,提取所述待推测路段的拓扑特征;

[0010] 将所述待推测路段的所述统计特征和所述拓扑特征合并,得到所述待推测路段的伪特征矢量;

[0011] 根据待推测路段的伪特征矢量建立逻辑回归模型,得到所述待推测路段的第一初步推测类型。

[0012] 进一步的,在根据所述待推测路段的伪特征建立逻辑回归模型,得到所述待推测路段的第一初步推测类型之后,还包括:

[0013] 从所述道路网络的拓扑结构数据中,获取所述待推测路段与所述待推测路段的相连路段之间的连接角度;

[0014] 根据获取的连接角度和所述待推测路段的相连路段的类型进行推测,得到所述待推测路段的第二初步推测类型。

[0015] 进一步的,在得到所述待推测路段的第二初步推测类型之后,还包括:

[0016] 根据所述第一初步推测类型和第二初步推测类型,采用集成算法计算得到所述待推测路段的最终推测类型。

[0017] 其中,所述采集多辆汽车在待推测路段上行驶的历史轨迹数据,并对所述历史轨迹数据进行统计,从而获取所述待推测路段的统计特征,具体包括:

[0018] 采集多辆汽车行驶的历史轨迹数据;

[0019] 采用ST-Matching算法将所述多辆汽车行驶的历史轨迹数据与所述道路网络匹配,获得多辆汽车在待推测路段上行驶的历史轨迹数据;

[0020] 对所述多辆汽车在待推测路段上行驶的历史轨迹数据进行统计,从而获取所述待推测路段的统计特征。

[0021] 进一步的,在将所述待推测路段的所述统计特征和所述拓扑特征合并,得到所述待推测路段的伪特征矢量之后,还包括:

[0022] 采用主成分分析法降低所述伪特征矢量的维度。

[0023] 进一步的,在从所述道路网络的拓扑结构数据中,获取所述待推测路段与所述待推测路段的相连路段之间的连接角度之前,还包括:

[0024] 根据道路网络的拓扑结构数据和道路网络中已知路段类型的路段,采用贝叶斯分类器进行学习,得到一个多项式分布;其中,所述多项式分布是当相连接的两个路段为不同的连接角度时,当其中一个路段为不同类型时,另一路段类型的概率分布。

[0025] 其中,所述根据所述连接角度和所述待推测路段的相连路段的类型进行推测,得到所述待推测路段的第二初步推测类型,具体包括:

[0026] 根据获取的连接角度和所述待推测路段的相连路段的类型,利用所述多项式分布,采用贝叶斯算法计算得到所述待推测路段的第二初步推测类型。

[0027] 本发明采集多辆汽车在待推测路段上行驶的历史轨迹数据,并对所述历史轨迹数据进行统计,从而获取所述待推测路段的统计特征;从所述待推测路段所在的道路网络的拓扑结构数据中,提取所述待推测路段的拓扑特征;将所述待推测路段的所述统计特征和所述拓扑特征合并,得到所述待推测路段的伪特征矢量;根据待推测路段的伪特征矢量建立逻辑回归模型,得到所述待推测路段的第一初步推测类型。本发明既考虑了车辆行驶的

历史轨迹数据,又考虑了道路网络拓扑结构数据,因此推测精性高,推测结果更准确。另外同时还将相连路段类型之间的限制关系作为推测依据,整合两个初步推测结果,获得最终推测结果,精确度更高。其中采用的模型或算法可以使用其他算法灵活替换,灵活性更高。

附图说明

[0028] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0029] 图1是本发明提供的路段类型推测方法的一个实施例的流程示意图;

[0030] 图2是历史轨迹的采样点数据类型的示意图;

[0031] 图3是两个相邻路段的示意图;

[0032] 图4是路段类型的示意图;

[0033] 图5是本发明提供的路段类型推测方法的另一实施例的流程示意图

[0034] 图6是本发明提供的路段类型推测方法的又一实施例的流程示意图。

具体实施方式

[0035] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0036] 图1是本发明提供的路段类型推测方法的一个实施例的流程示意图,如图1所示,包括步骤:

[0037] S101、采集多辆汽车在待推测路段上行驶的历史轨迹数据,并对所述历史轨迹数据进行统计,从而获取所述待推测路段的统计特征。

[0038] 具体的,步骤S101包括步骤:采集多辆汽车行驶的历史轨迹数据;采用ST-Matching算法将所述多辆汽车行驶的历史轨迹数据与道路网络匹配,获得多辆汽车在待推测路段上行驶的历史轨迹数据;对所述多辆汽车在待推测路段上行驶的历史轨迹数据进行统计,从而获取所述待推测路段的统计特征。

[0039] 其中,路段是指两个十字路口或交叉点之间的道路,高速公路或大道在两个十字路口或交叉点之间可能会有两个不同的路段,因为这两个路段的方向可能不同。

[0040] 其中,所述多辆汽车具体为多辆出租车,可以理解的,所述汽车也可以为其他类型汽车,例如公交车、私人车等。因其他类型车辆的数据不容易提取,特别是私人车辆涉及个人隐私问题,因此优选采用出租车的历史轨迹数据。其中,历史轨迹数据是一系列的采样点,采样频率可以为20s,每个采样点的数据具体可以如图2所示,包括:出租车ID、时间戳、经纬度、速度、角度、状态。时间戳是采样时的时间戳,经纬度是采样点的GPS定位位置,速度是出租车采样时的速度,角度是出租车采样时的驾驶方向,状态是指出租车采样时处于载客状态还是空闲状态。

[0041] 其中,采样点的位置信息仅仅是经纬度信息,从经纬度信息无法得知该采样点属

于哪一路段,因此,需要采用地图匹配方法将采样点数据与道路网络匹配,获取属于每一路段的采样点数据。本文采用ST-Matching算法进行匹配。ST-Matching算法是在2009年的地理信息系统国际期刊(International Journal of Geographical Information Science)上第352~361页刊登的文献名为“Map-matching for low-sampling-rate gps trajectories”的文献中公开的地图匹配算法。ST-Matching既考虑了道路网络的空间几何/拓扑结构,又考虑了历史轨迹的时间特征。ST-Matching适合于处理低抽样率的轨迹,例如本文的出租车轨迹。

[0042] 在历史轨迹和道路网络进行匹配后,获取属于待推测路段的采样点数据,对其进行统计,获得待推测路段的统计特征。其中,所述统计特征包括出租车行驶速度、空闲出租车密度、载客出租车密度和载客事件次数。待推测路段的统计特征是通过对于属于待推测路段的采样点数据统计和计算得到的,即对一系列采样点的出租车ID、时间戳、经纬度、速度、角度、状态数据进行统计和计算得到的。

[0043] S102、从所述待推测路段所在的道路网络的拓扑结构数据中,提取所述待推测路段的拓扑特征。

[0044] 其中,道路网络的拓扑结构数据是已知的。采用 $\{\tau_i\}_{i=1}^n$ 表示道路网络,其中, τ_i 是路段*i*,道路网络中共有*n*条路段。对于网络的拓扑结构可以采用连接矩阵 $M^{n \times n}$ 表示,连接矩阵中的元素 m_{ij} 表示路段*i*和路段*j*的连接角度,如果路段*i*和路段*j*不连接,元素 m_{ij} 为0。

[0045] 其中,待推测路段的拓扑特征包括路段长度、累积摆动值、相连路段和相邻路段。其中,路段长度和累积摆动值可以有效地反映路段类型。例如,大道通常是进入限制的,并且距离较长,中间很少有交叉点。因此,有很长长度的路段很可能是大道或者高速干道。类似的,根据经验,当一个路段是直的时候,其更可能是大道,而当一个路段是弯曲的时候,其不太可能是大道。因此采用累积摆动值反应路段类型。对于相连路段,是指两个路段拓扑相连接。如果一个路段有很多相连路段,那么此路段几乎不可能是大道,因为大道或者高速干道通常是在出口或入口有一两个相连路段。对于相邻路段,是指两个路段之间的距离小于预设阈值(具体是一个小的距离,例如10米)。路段距离的计算是通过计算路段线段的每个顶点之间的平均距离。如图3所示, τ_1 是路段1, τ_2 是路段2,路段1和路段2的顶点之间的距离为 d_1 、 d_2 、 d_3 ,取 d_1 、 d_2 、 d_3 的平均值作为路段1和路段2之间的距离,当距离小于预设阈值时,判定为两个路段相邻。两个相邻的路段可能会有相同的类型,特别是当他们方向相反时。

[0046] S103、将所述待推测路段的所述统计特征和所述拓扑特征合并,得到所述待推测路段的伪特征矢量。

[0047] 具体的,待推测路段的统计特征和拓扑特征共同构成待推测路段的伪特征。因为,采集数据较多,因此,伪特征矢量的维数有可能很大,因此,在步骤S103之后还可以包括步骤:采用主成分分析法降低所述伪特征矢量的维度,获取伪特征矢量的主成分。

[0048] S104、根据待推测路段的伪特征矢量建立逻辑回归模型,得到所述待推测路段的第一初步推测类型。

[0049] 其中,路段类型是按照国家标准规定的路段类型,如图4所示,包括7种路段类型。第一初步推测类型是逻辑回归模型输出的推测结果。

[0050] 本发明实施例公开了一种路段类型推测方法,包括:采集多辆汽车在待推测路段上行驶的历史轨迹数据,并对所述历史轨迹数据进行统计,从而获取所述待推测路段的统

计特征;从所述待推测路段所在的道路网络的拓扑结构数据中,提取所述待推测路段的拓扑特征;将所述待推测路段的所述统计特征和所述拓扑特征合并,得到所述待推测路段的伪特征矢量;根据待推测路段的伪特征矢量建立逻辑回归模型,得到所述待推测路段的第一初步推测类型。本发明实施例既考虑了车辆行驶的历史轨迹数据,又考虑了道路网络拓扑结构数据,因此推测精度高,推测结果更准确。

[0051] 图5是本发明提供的路段类型推测方法的另一实施例的流程示意图,如图5所示,包括步骤:

[0052] S201、采集多辆汽车在待推测路段上行驶的历史轨迹数据,并对所述历史轨迹数据进行统计,从而获取所述待推测路段的统计特征。

[0053] 具体的,步骤S201包括步骤:采集多辆汽车行驶的历史轨迹数据;采用ST-Matching算法将所述多辆汽车行驶的历史轨迹数据与道路网络匹配,获得多辆汽车在待推测路段上行驶的历史轨迹数据;对所述多辆汽车在待推测路段上行驶的历史轨迹数据进行统计,从而获取所述待推测路段的统计特征。

[0054] 其中,路段是指两个十字路口或交叉点之间的道路,高速公路或大道在两个十字路口或交叉点之间可能会有两个不同的路段,因为这两个路段的方向可能不同。

[0055] 其中,所述多辆汽车具体为多辆出租车,可以理解的,所述汽车也可以为其他类型汽车,例如公交车、私人车等。因其他类型车辆的数据不容易提取,特别是私人车辆涉及个人隐私问题,因此优选采用出租车的历史轨迹数据。其中,历史轨迹数据是一系列的采样点,采样频率可以为20s,每个采样点的数据具体可以如图2所示,包括:出租车ID、时间戳、经纬度、速度、角度、状态。时间戳是采样时的时间戳,经纬度是采样点的GPS定位位置,速度是出租车采样时的速度,角度是出租车采样时的驾驶方向,状态是指出租车采样时处于载客状态还是空闲状态。

[0056] 其中,采样点的位置信息仅仅是经纬度信息,从经纬度信息无法得知该采样点属于哪一路段,因此,需要采用地图匹配方法将采样点数据与道路网络匹配,获取属于每一路段的采样点数据。本文采用ST-Matching算法进行匹配。ST-Matching算法是在2009年的地理信息系统国际期刊(International Journal of Geographical Information Science)上第352~361页刊登的文献名为“Map-matching for low-sampling-rate gps trajectories”的文献中公开的地图匹配算法。ST-Matching既考虑了道路网络的空间几何/拓扑结构,又考虑了历史轨迹的时间特征。ST-Matching适合于处理低抽样率的轨迹,例如本文的出租车轨迹。

[0057] 在历史轨迹和道路网络进行匹配后,获取属于待推测路段的采样点数据,对其进行统计,获得待推测路段的统计特征。其中,所述统计特征包括出租车行驶速度、空闲出租车密度、载客出租车密度和载客事件次数。待推测路段的统计特征是通过对于属于待推测路段的采样点数据统计和计算得到的,即对一系列采样点的出租车ID、时间戳、经纬度、速度、角度、状态数据进行统计和计算得到的。

[0058] S202、从所述待推测路段所在的道路网络的拓扑结构数据中,提取所述待推测路段的拓扑特征。

[0059] 其中,道路网络的拓扑结构数据是已知的。采用 $\{\tau_i\}_{i=1}^n$ 表示道路网络,其中, τ_i 是路段*i*,道路网络中共有*n*条路段。对于网络的拓扑结构可以采用连接矩阵 $M^{n \times n}$ 表示,连接矩阵

中的元素 m_{ij} 表示路段 i 和路段 j 的连接角度,如果路段 i 和路段 j 不连接,元素 m_{ij} 为0。

[0060] 其中,待推测路段的拓扑特征包括路段长度、累积摆动值、相连路段和相邻路段。其中,路段长度和累积摆动值可以有效地反映路段类型。例如,大道通常是进入限制的,并且距离较长,中间很少有交叉点。因此,有很长长度的路段很可能是大道或者高速干道。类似的,根据经验,当一个路段是直的时候,其更可能是大道,而当一个路段是弯曲的时候,其不太可能是大道。因此采用累积摆动值反应路段类型。对于相连路段,是指两个路段拓扑相连接。如果一个路段有很多相连路段,那么此路段几乎不可能是大道,因为大道或者高速干道通常是在出口或入口有一两个相连路段。对于相邻路段,是指两个路段之间的距离小于预设阈值(具体是一个小的距离,例如10米)。路段距离的计算是通过计算路段线段的每个顶点之间的平均距离。如图3所示, τ_1 是路段1, τ_2 是路段2,路段1和路段2的顶点之间的距离为 d_1 、 d_2 、 d_3 ,取 d_1 、 d_2 、 d_3 的平均值作为路段1和路段2之间的距离,当距离小于预设阈值时,判定为两个路段相邻。两个相邻的路段可能会有相同的类型,特别是当他们方向相反时。

[0061] S203、将所述待推测路段的所述统计特征和所述拓扑特征合并,得到所述待推测路段的伪特征矢量。

[0062] 具体的,待推测路段的统计特征和拓扑特征共同构成待推测路段的伪特征。因为,采集数据较多,因此,伪特征矢量的维数有可能很大,因此,在步骤S103之后还可以包括步骤:采用主成分分析法降低所述伪特征矢量的维度,获取伪特征矢量的主成分。

[0063] S204、根据待推测路段的伪特征矢量建立逻辑回归模型,得到所述待推测路段的第一初步推测类型。

[0064] 其中,路段类型是按照国家标准规定的路段类型,如图4所示,包括7种路段类型。第一初步推测类型是逻辑回归模型输出的推测结果,是第一个初步推测结果。

[0065] 但是,因为有可能伪特征数据较少,导致第一初步推测类型不是很准确,本发明实施例还将路段的连接角度作为推测依据。

[0066] S205、从所述道路网络的拓扑结构数据中,获取所述待推测路段与所述待推测路段的相连路段之间的连接角度。

[0067] 通常,相连路段的连接角度很大程度上决定了两个路段类型的关系。例如,在一个普通城市的道路网络中,如果两个路段的连接角度为 180° ,他们通常是同一名字的同一条道路。但是当连接角度变小时,例如连接角度为 90° ,那么这两条路通常是不同名字的不同道路。因此,已知一个路段的类型和与相连路段的连接角度,可以在一定程度上推测相连路段的类型。

[0068] S206、根据获取的连接角度和所述待推测路段的相连路段的类型进行推测,得到所述待推测路段的第二初步推测类型。

[0069] 其中,第二初步推测类型是通过连接角度和相连路段的类型推测得到的第二个初步推测结果。

[0070] S207、根据所述第一初步推测类型和第二初步推测类型,采用集成算法计算得到所述待推测路段的最终推测类型。

[0071] 其中,集成算法用一些相对较弱的学习模型独立地就同样的样本进行训练,然后把结果整合起来进行整体预测。本实施例中的集成算法是堆叠泛化算法、支持向量机算法、随机森林算法中的任一种,可以灵活变化。最终推测类型是综合考虑第一初步推测类型和

第二初步推测类型后,获得的最后推测结果。

[0072] 本发明实施例公开了一种路段类型推测方法,包括:采集多辆汽车在待推测路段上行驶的历史轨迹数据,并对所述历史轨迹数据进行统计,从而获取所述待推测路段的统计特征;从所述待推测路段所在的道路网络的拓扑结构数据中,提取所述待推测路段的拓扑特征;将所述待推测路段的所述统计特征和所述拓扑特征合并,得到所述待推测路段的伪特征矢量;根据待推测路段的伪特征矢量建立逻辑回归模型,得到所述待推测路段的第一初步推测类型;从所述道路网络的拓扑结构数据中,获取所述待推测路段与所述待推测路段的相连路段之间的连接角度;根据获取的连接角度和所述待推测路段的相连路段的类型进行推测,得到所述待推测路段的第二初步推测类型。本发明实施例既考虑了车辆行驶的历史轨迹数据,又考虑了道路网络拓扑结构数据,因此推测精性高,推测结果更准确。同时还将相连路段类型之间的限制关系作为推测依据,整合两个初步推测结果,获得最终推测结果,精确度更高。

[0073] 图6是本发明提供的路段类型推测方法的又一实施例的流程示意图,如图6所示,包括步骤:

[0074] S301、采集多辆汽车在待推测路段上行驶的历史轨迹数据,并对所述历史轨迹数据进行统计,从而获取所述待推测路段的统计特征。

[0075] 具体的,步骤S301包括步骤:采集多辆汽车行驶的历史轨迹数据;采用ST-Matching算法将所述多辆汽车行驶的历史轨迹数据与道路网络匹配,获得多辆汽车在待推测路段上行驶的历史轨迹数据;对所述多辆汽车在待推测路段上行驶的历史轨迹数据进行统计,从而获取所述待推测路段的统计特征。

[0076] 其中,路段是指两个十字路口或交叉点之间的道路,高速公路或大道在两个十字路口或交叉点之间可能会有两个不同的路段,因为这两个路段的方向可能不同。

[0077] 其中,所述多辆汽车具体为多辆出租车,可以理解的,所述汽车也可以为其他类型汽车,例如公交车、私人车等。因其他类型车辆的数据不容易提取,特别是私人车辆涉及个人隐私问题,因此优选采用出租车的历史轨迹数据。其中,历史轨迹数据是一系列的采样点,采样频率可以为20s,每个采样点的数据具体可以如图2所示,包括:出租车ID、时间戳、经纬度、速度、角度、状态。时间戳是采样时的时间戳,经纬度是采样点的GPS定位位置,速度是出租车采样时的速度,角度是出租车采样时的驾驶方向,状态是指出租车采样时处于载客状态还是空闲状态。

[0078] 其中,采样点的位置信息仅仅是经纬度信息,从经纬度信息无法得知该采样点属于哪一路段,因此,需要采用地图匹配方法将采样点数据与道路网络匹配,获取属于每一路段的采样点数据。本文采用ST-Matching算法进行匹配。ST-Matching算法是在2009年的地理信息系统国际期刊(International Journal of Geographical Information Science)上第352~361页刊登的文献名为“Map-matching for low-sampling-rate gps trajectories”的文献中公开的地图匹配算法。ST-Matching既考虑了道路网络的空间几何/拓扑结构,又考虑了历史轨迹的时间特征。ST-Matching适合于处理低抽样率的轨迹,例如本文的出租车轨迹。

[0079] 在历史轨迹和道路网络进行匹配后,获取属于待推测路段的采样点数据,对其进行统计,获得待推测路段的统计特征。其中,所述统计特征包括出租车行驶速度、空闲出租

车密度、载客出租车密度和载客事件次数。待推测路段的统计特征是通过属于待推测路段的采样点数据统计和计算得到的,即对一系列采样点的出租车ID、时间戳、经纬度、速度、角度、状态数据进行统计和计算得到的。

[0080] S302、从所述待推测路段所在的道路网络的拓扑结构数据中,提取所述待推测路段的拓扑特征。

[0081] 其中,道路网络的拓扑结构数据是已知的。采用 $\{\tau_i\}_{i=1}^n$ 表示道路网络,其中, τ_i 是路段 i ,道路网络中共有 n 条路段。对于网络的拓扑结构可以采用连接矩阵 $M^{n \times n}$ 表示,连接矩阵中的元素 m_{ij} 表示路段 i 和路段 j 的连接角度,如果路段 i 和路段 j 不连接,元素 m_{ij} 为0。

[0082] 其中,待推测路段的拓扑特征包括路段长度、累积摆动值、相连路段和相邻路段。其中,路段长度和累积摆动值可以有效地反映路段类型。例如,大道通常是进入限制的,并且距离较长,中间很少有交叉点。因此,有很长长度的路段很可能是大道或者高速干道。类似的,根据经验,当一个路段是直的时候,其更可能是大道,而当一个路段是弯曲的时候,其不太可能是大道。因此采用累积摆动值反应路段类型。对于相连路段,是指两个路段拓扑相连接。如果一个路段有很多相连路段,那么此路段几乎不可能是大道,因为大道或者高速干道通常是在出口或入口有一两个相连路段。对于相邻路段,是指两个路段之间的距离小于预设阈值(具体是一个小的距离,例如10米)。路段距离的计算是通过计算路段线段的每个顶点之间的平均距离。如图3所示, τ_1 是路段1, τ_2 是路段2,路段1和路段2的顶点之间的距离为 d_1 、 d_2 、 d_3 ,取 d_1 、 d_2 、 d_3 的平均值作为路段1和路段2之间的距离,当距离小于预设阈值时,判定为两个路段相邻。两个相邻的路段可能会有相同的类型,特别是当他们方向相反时。

[0083] S303、将所述待推测路段的所述统计特征和所述拓扑特征合并,得到所述待推测路段的伪特征矢量。

[0084] 具体的,待推测路段的统计特征和拓扑特征共同构成待推测路段的伪特征。因为,采集数据较多,因此,伪特征矢量的维数有可能很大,因此,在步骤S103之后还可以包括步骤:采用主成分分析法降低所述伪特征矢量的维度,获取伪特征矢量的主成分。

[0085] S304、根据待推测路段的伪特征矢量建立逻辑回归模型,得到所述待推测路段的第一初步推测类型。

[0086] 其中,路段类型是按照国家标准规定的路段类型,如图4所示,包括7种路段类型。

[0087] 但是,因为有可能伪特征数据较少,导致第一初步推测类型不是很准确,本发明实施例还将路段的连接角度作为推测依据。

[0088] S305、从所述道路网络的拓扑结构数据中,获取所述待推测路段与所述待推测路段的相连路段之间的连接角度。

[0089] 通常,相连路段的连接角度很大程度上决定了两个路段类型的关系。例如,在一个普通城市的道路网络中,如果两个路段的连接角度为 180° ,他们通常是同一名字的同一条道路。但是当连接角度变小时,例如 90° ,那么这两条路通常是不同名字的不同道路。因此,已知一个路段的类型和与相连路段的连接角度,可以在一定程度上推测相连路段的类型。

[0090] S306、根据道路网络的拓扑结构数据和道路网络中已知路段类型的路段,采用贝叶斯分类器进行学习,得到一个多项式分布。

[0091] 其中,所述多项式分布是当相连接的两个路段为不同的连接角度时,当其中一个路段为不同类型时,另一路段类型的概率分布。具体的,多项式分布是当相连接的两个路段

为不同的连接角度时,当其中一个路段为不同类型时,另一路段的类型分别为如图4所示的类型1~7的概率。

[0092] S307、根据获取的连接角度和所述待推测路段的相连路段的类型进行推测,得到所述待推测路段的第二初步推测类型。

[0093] 具体的,根据获取的连接角度和所述待推测路段的相连路段的类型,利用所述多项式分布,采用贝叶斯算法计算得到所述待推测路段的第二初步推测类型。

[0094] S308、根据所述第一初步推测类型和第二初步推测类型,采用集成算法计算得到所述待推测路段的最终推测类型。

[0095] 其中,集成算法用一些相对较弱的学习模型独立地就同样的样本进行训练,然后把结果整合起来进行整体预测。本实施例中的集成算法是堆叠泛化算法、支持向量机算法、随机森林算法中的任一种,可以灵活变化。

[0096] 本发明实施例公开了一种路段类型推测方法,包括:采集多辆汽车在待推测路段上行驶的历史轨迹数据,并对所述历史轨迹数据进行统计,从而获取所述待推测路段的统计特征;从所述待推测路段所在的道路网络的拓扑结构数据中,提取所述待推测路段的拓扑特征;将所述待推测路段的所述统计特征和所述拓扑特征合并,得到所述待推测路段的伪特征矢量;根据待推测路段的伪特征矢量建立逻辑回归模型,得到所述待推测路段的第一初步推测类型;从所述道路网络的拓扑结构数据中,获取所述待推测路段与所述待推测路段的相连路段之间的连接角度;根据道路网络的拓扑结构数据和道路网络中已知路段类型的路段,采用贝叶斯分类器进行学习,得到一个多项式分布;根据获取的连接角度和所述待推测路段的相连路段的类型进行推测,得到所述待推测路段的第二初步推测类型。本发明实施例既考虑了车辆行驶的历史轨迹数据,又考虑了道路网络拓扑结构数据,因此推测精性高,推测结果更准确。同时还将相连路段类型之间的限制关系作为推测依据,整合两个初步推测结果,获得最终推测结果,精确度更高。

[0097] 需要说明的是,在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者装置所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、方法、物品或者装置中还存在另外的相同要素。

[0098] 上述本发明实施例序号仅仅为了描述,不代表实施例的优劣。

[0099] 专业人员还可以进一步意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、计算机软件或者二者的结合来实现,为了清楚地说明硬件和软件的可互换性,在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本发明的范围。

[0100] 结合本文中所公开的实施例描述的方法或算法的步骤可以直接用硬件、处理器执行的软件模块,或者二者的结合来实施。软件模块可以置于随机存储器(RAM)、内存、只读存储器(ROM)、电可编程ROM、电可擦除可编程ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质中。

[0101] 对所公开的实施例的上述说明,使本领域专业技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的,本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下,在其它实施例中实现。因此,本发明将不会被限制于本文所示的这些实施例,而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

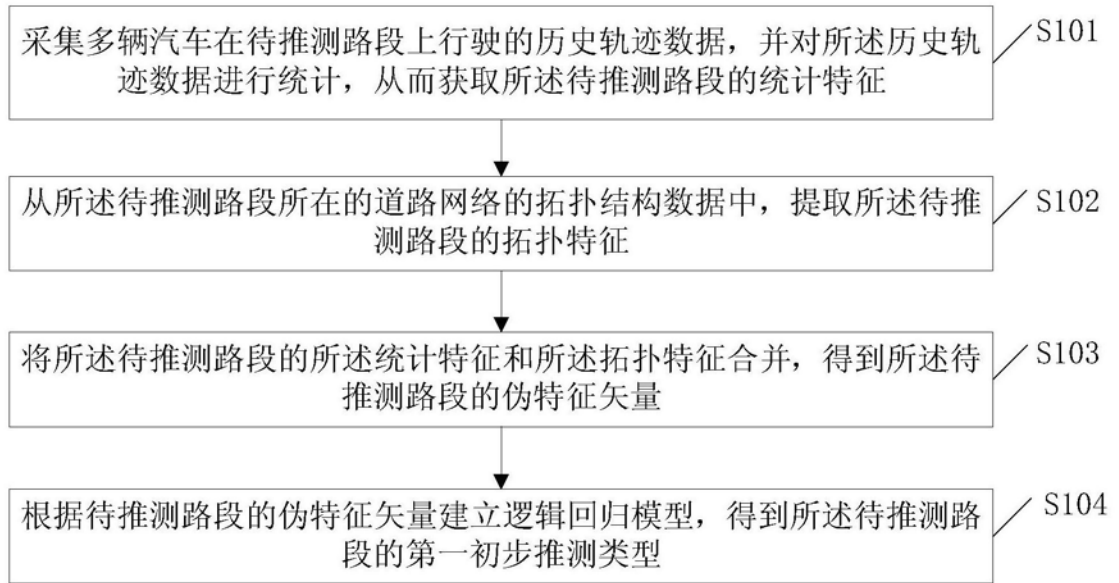


图1

数据类型	描述
出租车 ID	出租车注册号码
时间戳	采样时的时间戳
经纬度	采样点的 GPS 位置
速度	出租车采样时的速度
角度	出租车采样时的驾驶方向
状态	出租车处于载客状态还是空闲状态

图2

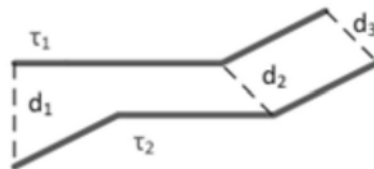


图3

类型 ID	路段类型
1	国道
2	省道
3	普通高速公路
4	大道
5	主干道
6	次干道
7	普通公路

图4

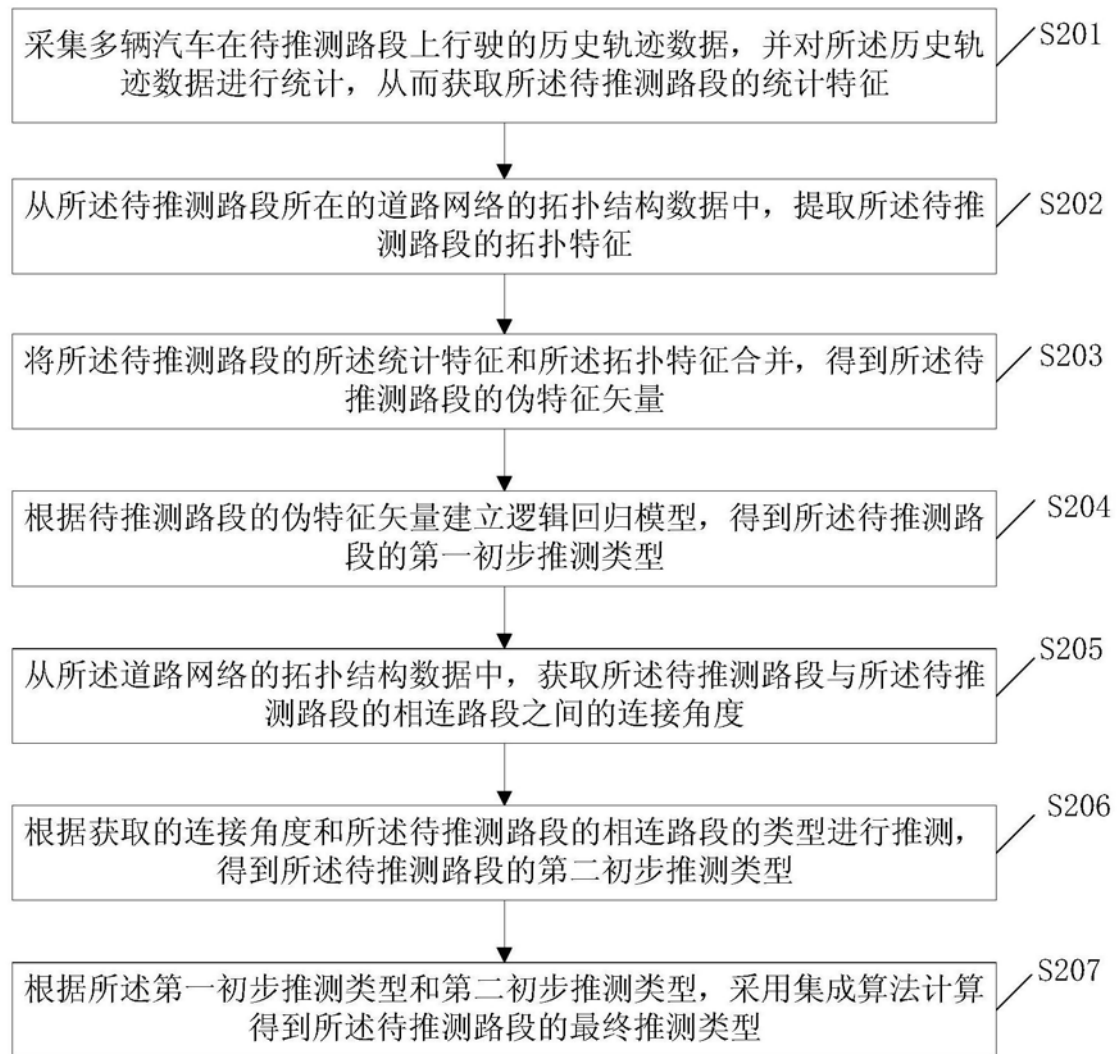


图5

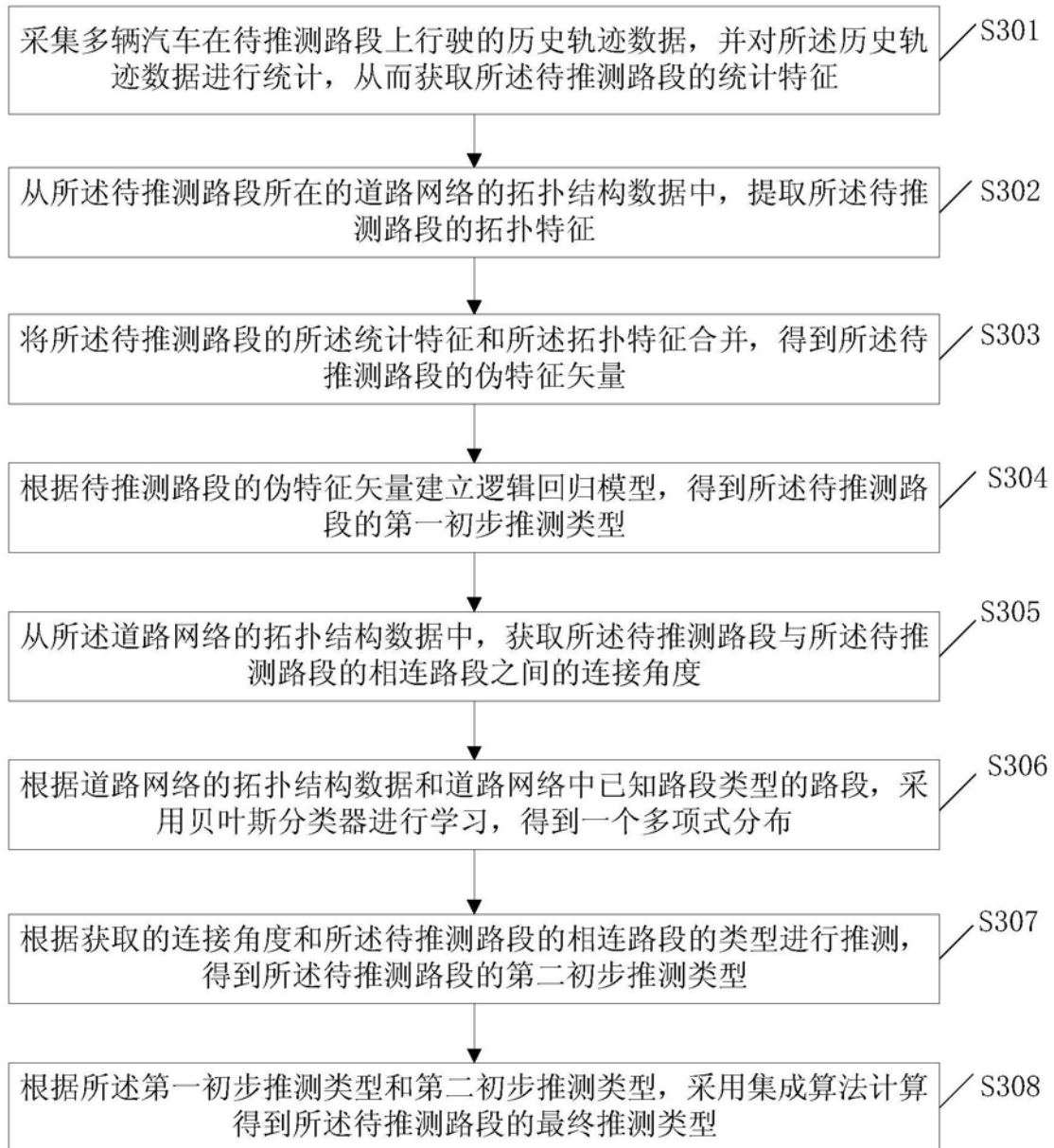


图6