



(12)发明专利

(10)授权公告号 CN 108062857 B

(45)授权公告日 2019.08.23

(21)申请号 201711405689.6

G06Q 10/04(2012.01)

(22)申请日 2017.12.22

G06Q 50/30(2012.01)

(65)同一申请的已公布的文献号

审查员 凌辰

申请公布号 CN 108062857 A

(43)申请公布日 2018.05.22

(73)专利权人 西南交通大学

地址 610031 四川省成都市二环路北一段
111号

(72)发明人 蒋阳升 罗孝羚 姚志洪 赵斌

吴奇 马媛 韩鹏 黄丹芮

(74)专利代理机构 成都点睛专利代理事务所

(普通合伙) 51232

代理人 孙一峰

(51)Int.Cl.

G08G 1/00(2006.01)

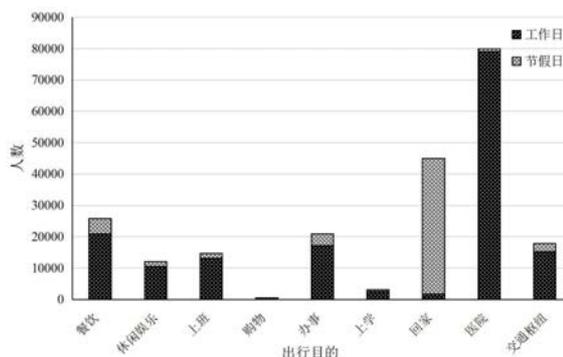
权利要求书3页 说明书10页 附图3页

(54)发明名称

用于出租车乘客出行目的的预测方法

(57)摘要

本发明属于交通规划技术领域,具体的说涉及一种用于出租车乘客出行目的的预测方法。本发明与传统方法不同的是,本发明从出行特征和乘客下车点的最终目的地所属POI点类型两个层面对乘客出行目的进行判断,该方法能够有效解决仅仅依靠出行特征推断出行目的的不足,即对出行特征较为类似的不同出行目的识别的准确率较差,而不需将不同类别的出行目的聚为一类。



1. 用于出租车乘客出行目的的预测方法,其特征在于,包括以下步骤:

a、构建乘客出行特征识别模型,具体方法为:

根据出租车运营数据获取乘客的出行特征;所述出租车运营数据至少包括:出租车辆的ID编号、每次触发GPS设备时车辆所处位置的经纬度、每次触发GPS设备时是否有乘客在车上、每次触发GPS设备时的时刻;所述乘客的出行特征至少包括:出行是否为节假日、下车点时刻、乘客出租车出行距离、下车点经纬度;其中,

所述出行是否为节假日、下车点时刻、下车点经纬度可根据乘客的下车点信息直接获取:

$$t_{s_i^{alight}}^w \Rightarrow A_i^w$$

$$t_{s_i^{alight}}^t \Rightarrow A_i^t$$

$$A_i^{\log} = t_{s_i^{alight}}^{\log}$$

$$A_i^{\text{lat}} = t_{s_i^{alight}}^{\text{lat}}$$

其中 A_i^w 为乘客i出行是否为节假日的特征标识符, $t_{s_i^{alight}}^w$ 为乘客i出行的为周几,当 $t_{s_i^{alight}}^w$ 为周一至周五,即为工作日时 A_i^w 为0,当 $t_{s_i^{alight}}^w$ 为周六或周日时 A_i^w 为1, A_i^t 为乘客i出行时段特征标识符, $t_{s_i^{alight}}^t$ 为乘客i到达下车点的时刻,其中时刻表以一个小时为区间单位进行划分, A_i^{\log} 为乘客i的下车点经度特征标识符, $t_{s_i^{alight}}^{\log}$ 为乘客i下车点的经度取值, A_i^{lat} 为乘客i的下车点纬度特征标识符, $t_{s_i^{alight}}^{\text{lat}}$ 为乘客i下车点的纬度取值;

所述获取乘客出租车出行距离的具体方法为:

根据车载GPS的触发时间,将出行轨迹划分为n段,则乘客的出行距离为:

$$l_{od}^i = \sum_{j=1}^n r_j^i$$

其中, l_{od}^i 为乘客i从起点o到终点d的行驶距离, j为将乘客i从起点o到终点d轨迹划分的小段序列标识符, n为轨迹分小段数量, r_j^i 为乘客i的第j段小段轨迹的距离长度;

其中 r_j^i 小段轨迹的距离长度为:

$$r_j^i = 2000R \times \arcsin \left[\sqrt{\sin^2 \left(\frac{r_j^{i-olat} - r_j^{i-dlat}}{2} \right) + \cos(r_j^{i-olat}) \times \cos(r_j^{i-dlat}) \times \sin^2 \left(\frac{r_j^{i-olog} - r_j^{i-dlog}}{2} \right)} \right]$$

其中R为地球半径, r_j^{i-olat} 为 r_j^i 小段轨迹起点的纬度, r_j^{i-dlat} 为 r_j^i 小段轨迹终点的纬度, r_j^{i-olog} 为 r_j^i 小段轨迹起点的经度, r_j^{i-dlog} 为 r_j^i 小段轨迹终点的经度;

则获得乘客的出行距离特征标识 A_i^{dis} 为:

$$A_i^{dis} = I_{od}^i ;$$

b、构建乘客出行目的识别模型,具体方法为:

b1、采用分类方法根据步骤a获得的乘客出行特征获取乘客出行目的:

$$f(A_i^w, A_i^t, A_i^{dis}) = purpose_i^a$$

其中 $purpose_i^a$ 表示乘客i通过其出行特征推断出的出行目的, f 表示推断出行特征所用的方法,至少包括神经网络、贝叶斯、决策树、支持向量机、最近邻方法;

b2、设置信息点对乘客出行目的进行判断,所述信息点为至少包括名称、类别、经度、纬度的预设乘客出行目的地,根据乘客下车点及与该下车点邻近的信息点,判断方式为:

$$purpose_i^{poi} = type_{poi} \quad dis(loca_i, loca_{type_{poi}}) \leq r$$

其中, $purpose_i^{poi}$ 为通过下车点邻近的信息点推断的乘客i的出行目的, $type_{poi}$ 为在下车点的距离为 r 的范围内存在的信息点的集合, $dis(loca_i, loca_{type_{poi}})$ 为乘客i下车点与最终目的地的距离, r 为设定的下车点与最终目的地距离阈值;

c、乘客出行目的预测:

根据获得的 $purpose_i^a$ 和 $purpose_i^{poi}$, 进行融合,有:

$$purpose_i^{poi} \cap purpose_i^a = \emptyset \quad purpose_i^{poi} = \emptyset$$

则最终识别结果 $purpose_i^{fin} = purpose_i^a$;

或:

$$purpose_i^{poi} \cap purpose_i^a = \emptyset \quad purpose_i^{poi} \neq \emptyset$$

则最终识别结果 $purpose_i^{fin} \in purpose_i^{poi}$;

或:

$$purpose_i^{poi} \cap purpose_i^a = purpose_i^a \quad purpose_i^{poi} = purpose_i^a$$

则最终识别结果 $purpose_i^{fin} = purpose_i^a$;

或:

$$purpose_i^{poi} \cap purpose_i^a = purpose_i^a \quad purpose_i^{poi} \neq purpose_i^a$$

则最终识别结果 $purpose_i^{fin} = purpose_i^a$ 。

2. 根据权利要求1所述的用于出租车乘客出行目的的预测方法,其特征在于,所述步骤b2中,还包括判断信息点是否位于下车点半径阈值 r 内:

对每位乘客下车点与所有信息点的经纬度进行预判,将信息点分为两个集合:

$$Set_{poi}^{i-1} = \{v_{POI}^i \mid v_{POI_{lat}}^i - v_{ALIGHT_{lat}}^i < \Phi_{lat}^r, v_{POI_{log}}^i - v_{ALIGHT_{log}}^i < \Phi_{lon}^r\}$$

$$Set_{poi}^{i-2} = Set_{poi} - Set_{poi}^{i-1}$$

其中, Set_{poi} 为所有POI点的集合, Set_{poi}^{i-1} 为满足乘客i下车点的经度差和纬度差均在预设范围内的信息点集合, Set_{poi}^{i-2} 为除了 Set_{poi}^{i-1} 中信息点元素的剩余信息点集合, v_{POI}^i 为乘客i满

足后续计算条件的信息点元素, $V_{POI_{lat}}$ 为信息点的经度, $v_{ALIGHT_{lat}}^i$ 为乘客 i 的下车点经度, $V_{POI_{log}}$ 为信息点的纬度, $v_{ALIGHT_{log}}^i$ 为乘客 i 下车点的纬度, Φ_{lat}^r 为预设的最大经度差, Φ_{lon}^r 为预设的最大纬度差;

针对每位乘客的下车点将所有信息点进行分类, 只需计算下车点与该下车点的集合 Set_{poi}^{i-1} 中的信息点的距离, 然后进行目标信息点判断。

用于出租车乘客出行目的的预测方法

技术领域

[0001] 本发明属于交通规划技术领域,具体的说涉及一种用于出租车乘客出行目的的预测方法。

背景技术

[0002] 出租车乘客的出行目的调查是交通调查中的重要内容,其对城市功能结构规划,城市道路网络规划,交通组织设计及优化具有重要的意义。传统出租车乘客出行目的调查方式主要有旅行日记调查、电话调查、邮件调查、小组调查等人工调查方式。其中这些调查方式又可以分为两类:一类为直接调查出行目的,另一类为调查出行特征,通过特征间接获得出行目的。传统的调查方式要求前期准备、试点调查、实地调查、调查结果整理和录入,需要耗费大量的人力和物力。

[0003] 近年来随着信息技术的发展,以及信息设备在交通领域的广泛应用,使得一些原始的交通数据得以自动传输到数据库系统,并保存下来。在这样的背景下,利用已有的原始数据,获取有价值的交通信息成为了研究热点。如公交数据方面有:利用公交车IC卡数据分析票价结构,利用公交GPS数据分析公交车运营服务可靠性,融合IC卡数据和GPS数据获取乘客出行需求;出租车数据有:利用出租车运营数据分析路段行程时间和速度,利用出租车GPS数据分析运营效率;地铁数据有:利用地铁数据对地铁运营现状进行分析。这些研究都说明通过有效的技术手段对一些已有的数据进行挖掘,能够快速有效的获取有价值的交通信息,相比于传统的人工调查方式,这种方式能够节约大量的人力物力。

[0004] 现有的研究具体到利用数据分析及挖掘获取乘客的出行目的研究主要原理是通过数据获取乘客出行特征,然后结合分类学习的方法对出行目的进行推断,包括Kusakabe T,Asakura Y.Behavioural data mining of transit smart card data:A data fusion approach[J].Transportation Research Part C:Emerging Technologies,2014,46:179-191、Tierney K,Decker S,Proussaloglou K,et al.Travel survey manual[R].1996、Tierney K,Decker S,Proussaloglou K,etal.Travel survey manual[R].1996。但这些方法通过实验表明这种方法主要问题是对于出行特征相似的不同出行目的识别准确率很低,因此,该研究针对这一问题,将具有相似特征的不同出行目的类别聚为一类,从而提高识别准确率。虽然,将具有相似性的出行特征的不同出行目的聚为一类能够有效提高出行目的的识别率,但是这种方法无法区分一类中不同出行目的具体比例,且是以损失了信息的完备性为代价提高最终的识别准确度,本质上还是没有提高相似特征不同出行目的的识别率,对于最终应用于规划设计具有一定的局限性。

发明内容

[0005] 本发明所要解决的,就是针对上述问题,在上述传统方法的基础上,提出一种基于出租车运营数据和信息点(Point of Interesting,POI)数据推断出租车乘客出行目的的方法。其中POI点可以简单理解为出行者潜在的最终的目的地。由于POI点中包含类别信息

即乘客潜在出行目的地的行业类别,因此,将POI点数据融入出租车运营数据能够有效提高出行目的识别精度。在所提出的方法中,首先通过所构建的特征识别模型获取出租车运营数据中乘客的出行特征,然后依据小样本调查数据,训练不同的识别模型,并选取识别精度最高的模型,最后利用所得到的出租车乘客出行的特征数据和训练好的识别模型实现对乘客出行目的识别。

[0006] 为实现上述目的,本发明采用如下技术方案:

[0007] 用于出租车乘客出行目的的预测方法,其特征在于,包括以下步骤:

[0008] a、构建乘客出行特征识别模型,具体方法为:

[0009] 根据出租车运营数据获取乘客的出行特征;所述出租车运营数据至少包括:出租车辆的ID编号、每次触发GPS设备时车辆所处位置的经纬度、每次触发GPS设备时是否有乘客在车上、每次触发GPS设备时的时刻;所述乘客的出行特征至少包括:出行是否为节假日、下车点时刻、乘客出租车出行距离、下车点经纬度;其中,

[0010] 所述出行是否为节假日、下车点时刻、下车点经纬度可根据乘客的下车点信息直接获取:

$$[0011] \quad t_{S_i}^{W, \text{alight}} \Rightarrow A_i^W$$

$$[0012] \quad t_{S_i}^t \Rightarrow A_i^t$$

$$[0013] \quad A_i^{\log} = t_{S_i}^{\log, \text{alight}}$$

$$[0014] \quad A_i^{\text{lat}} = t_{S_i}^{\text{lat}, \text{alight}}$$

[0015] 其中 A_i^W 为乘客i出行是否为节假日的特征标识符, $t_{S_i}^{W, \text{alight}}$ 为乘客i出行的为周几,当 $t_{S_i}^{W, \text{alight}}$ 为周一至周五,即为工作日时 A_i^W 为0,当 $t_{S_i}^{W, \text{alight}}$ 为周六或周日时 A_i^W 为1, A_i^t 为乘客i出行时段特征标识符, $t_{S_i}^t$ 为乘客i到达下车点的时刻,其中时刻表以一个小时为区间单位进行划分, A_i^{\log} 为乘客i的下车点经度特征标识符, $t_{S_i}^{\log, \text{alight}}$ 为乘客i下车点的经度取值, A_i^{lat} 为乘客i的下车点纬度特征标识符, $t_{S_i}^{\text{lat}, \text{alight}}$ 为乘客i下车点的纬度取值;

[0016] 所述获取乘客出租车出行距离的具体方法为:

[0017] 根据车载GPS的触发时间,将出行轨迹划分为n段,则乘客的出行距离为:

$$[0018] \quad l_{od}^i = \sum_{j=1}^n r_j^i$$

[0019] 其中, l_{od}^i 为乘客i从起点o到终点d行驶距离的计算方式, j为将乘客i从起点o到终点d轨迹划分的小段序列标识符, n为轨迹分小段数量, r_j^i 为乘客i的第j段小段轨迹的距离长度;

[0020] 其中 r_j^i 小段轨迹的距离长度为:

$$[0021] \quad r_j^i = 2000R \times \arcsin \left[\sqrt{\sin^2 \left(\frac{r_j^{i-olat} - r_j^{i-dlat}}{2} \right) + \cos(r_j^{i-olat}) \times \cos(r_j^{i-dlat}) \times \sin^2 \left(\frac{r_j^{i-olog} - r_j^{i-dlog}}{2} \right)} \right]$$

[0022] 其中R为地球半径, r_j^{i-olat} 为 r_j^i 小段轨迹起点的纬度, r_j^{i-dlat} 为 r_j^i 小段轨迹终点的纬度, r_j^{i-olog} 为 r_j^i 小段轨迹起点的经度, r_j^{i-dlog} 为 r_j^i 小段轨迹终点的经度;

[0023] 则获得乘客的出行距离特征标识 A_i^{dis} 为:

$$[0024] \quad A_i^{dis} = l_{od}^i;$$

[0025] b、构建乘客出行目的识别模型,具体方法为:

[0026] b1、采用分类方法根据步骤a获得的乘客出行特征获取乘客出行目的:

$$[0027] \quad f(A_i^w, A_i^l, A_i^{dis}) = purpose_i^a$$

[0028] 其中 $purpose_i^a$ 表示乘客i通过其出行特征推断出的出行目的, f表示推断出行特征所用的方法,至少包括神经网络、贝叶斯、决策树、支持向量机、最近邻方法;

[0029] b2、设置信息点对乘客出行目的进行判断,所述信息点为至少包括名称、类别、经度、纬度的预设乘客出行目的地点,根据乘客下车点及与该下车点邻近的信息点,判断方式为:

$$[0030] \quad purpose_i^{poi} = type_{poi} \quad dis(loca_i, loca_{type_{poi}}) \leq r$$

[0031] 其中, $purpose_i^{poi}$ 为通过下车点邻近的信息点推断的乘客i的出行目的, $type_{poi}$ 为在下车点的距离为r的范围内存在的信息点的集合, $dis(loca_i, loca_{type_{poi}})$ 为乘客i下车点与最终目的地的距离, r为设定的下车点与最终目的地距离阈值;

[0032] c、乘客出行目的预测:

[0033] 根据获得的 $purpose_i^a$ 和 $purpose_i^{poi}$, 进行融合,有:

$$[0034] \quad purpose_i^{poi} \cap purpose_i^a = \emptyset \quad purpose_i^{poi} = \emptyset$$

[0035] 则最终识别结果 $purpose_i^{fin} = purpose_i^a$;

[0036] 或:

$$[0037] \quad purpose_i^{poi} \cap purpose_i^a = \emptyset \quad purpose_i^{poi} \neq \emptyset$$

[0038] 则最终识别结果 $purpose_i^{fin} \in purpose_i^{poi}$;

[0039] 或:

$$[0040] \quad purpose_i^{poi} \cap purpose_i^a = purpose_i^a \quad purpose_i^{poi} = purpose_i^a$$

[0041] 则最终识别结果 $purpose_i^{fin} = purpose_i^a$;

[0042] 或:

$$[0043] \quad purpose_i^{poi} \cap purpose_i^a = purpose_i^a \quad purpose_i^{poi} \neq purpose_i^a$$

[0044] 则最终识别结果 $purpose_i^{fin} = purpose_i^a$ 。

[0045] 进一步的,所述步骤b2的具体方法判断方法为:

[0046] 对每位乘客下车点与所有信息点的经纬度进行预判断,将信息点分为两个集合:

$$[0047] \quad Set_{poi}^{i-1} = \{v_{POI}^i \mid v_{POI_{lat}} - v_{ALIGHT_{lat}}^i < \Phi_{lat}^r, v_{POI_{log}} - v_{ALIGHT_{log}}^i < \Phi_{lon}^r\}$$

$$[0048] \quad Set_{poi}^{i-2} = Set_{poi} - Set_{poi}^{i-1}$$

[0049] 其中, Set_{poi} 为所有POI点的集合, Set_{poi}^{i-1} 为满足乘客i下车点的经度差和纬度差均在预设范围内的信息点集合, Set_{poi}^{i-2} 为除了 Set_{poi}^{i-1} 中信息点元素的剩余信息点集合, v_{POI}^i 为乘客i满足后续计算条件的信息点元素, $v_{POI_{lat}}$ 为信息点的经度, $v_{ALIGHT_{lat}}^i$ 为乘客i的下车点经度, $v_{POI_{log}}$ 为信息点的纬度, $v_{ALIGHT_{log}}^i$ 为乘客i下车点的纬度, Φ_{lat}^r 为预设的最大经度差, Φ_{lon}^r 为预设的最大纬度差;

[0050] 针对每位乘客的下车点将所有信息点进行分类,只需计算下车点与该下车点的集合 Set_{poi}^{i-1} 中的信息点的距离,然后进行目标信息点判断。

[0051] 本发明的有益效果为,与传统方法不同的是,本发明从出行特征和乘客下车点的最终目的地所属POI点类型两个层面对乘客出行目的进行判断,该方法能够有效解决仅仅依靠出行特征推断出行目的不足,即对出行特征较为类似的不同出行目的识别的准确率较差,而不需将不同类别的出行目的聚为一类。

附图说明

[0052] 图1是半径阈值说明示意图;

[0053] 图2是出租车非直线轨迹截断示意图;

[0054] 图3是POI点是否位于下车点半径阈值r范围内判断方法示意图;

[0055] 图4是不同识别方法的识别准确度示意图;

[0056] 图5是传统方法与本发明所提出的方法识别精准度对比图;

[0057] 图6是对实际出租车运营数据的乘客出行目的识别结果示意图。

具体实施方式

[0058] 下面结合附图,详细描述本发明的技术方案:

[0059] 本发明是以出租车运营数据和POI点为依据,其中,出租车运营数据实际上是由车载终端生成,虽然不同城市的出租车运营数据记录的字段有所差别,但包含的主要内容基本一致。以本文中所用到的XX市的出租车数据为例,其包含的主要字段为:出租车辆的ID编号(该编号为车辆在出租车数据系统中的唯一编号,与其车牌号一一对应);每次触发GPS设备时车辆所处位置的经纬度;每次触发GPS设备时是否有乘客在车上(即出租车是否处于服务乘客的状态还是空驶状态);每次触发GPS设备时的时刻。如表1所示为出租车在按照车辆ID和GPS设备触发时间排序之后的部分运营信息表。

[0060] 表1 XX市出租车数据库系统记录数据

ID	LATITUDE	LONGITUDE	USER_ON	TIME
(ID 编号)	(纬度)	(经度)	(运营状态)	(时刻表)
2	30.599108	103.98768	0	10:12:39
[0061] 2	30.600106	103.98793	0	10:12:49
2	30.600311	103.988	0	10:12:59
2	30.60099	103.98817	0	10:13:09
2	30.60166	103.988296	0	10:13:19

[0062] 根据所展示的部分信息表可知,出租车运营数据主要是由一些离散记录了车辆位置和车辆服务状态的时刻序列数据组成,从数据来看,不同出租车服务年代及安装GPS设备时间不同,导致每辆车辆GPS触发的时间间隔有所不同(如表1中的GPS触发时间间隔为10秒钟),但总体数据都保持在30秒钟以内,会向数据库系统实时传送一次数据。

[0063] 本发明所述的POI点,可以利用爬虫技术从高德地图上获取的,具体步骤是首先在高德地图申请密钥,然后根据获得的密钥结合爬虫技术就可以从高德地图获取特定区域所有的POI点信息,每个POI点的信息包含其所属类别,具体类别与该POI点的信息功能相对应。由于乘客从下车点到最终目的地的距离处于一定范围内(具体范围为依据调查数据确定),我们将该乘客从下车点到最终目的地所能够容忍的距离定义为半径阈值 r ,即乘客下车点的最终目的地与其下车点的最大距离不超过半径阈值,因此,从POI点层面,可以对乘客的出行目的做出相应的判断。如图1所示,我们认为乘客下车之后,可能前往POI2点作为其最终目的地,由于POI1点与下车点的距离大于半径阈值 r ,乘客将不会选择POI1点作为最终出行目的地。

[0064] 特征识别模型构建:

[0065] 通过对现有的研究分析可知,要根据出租车运营数据获取乘客的出行目的,首先需要从出租车运营数据获取乘客出行特征,根据前面调查数据对出行特征的分析,确定从出租车运营数据中需要获取的出行特征分别为:出行是否为节假日、下车点时刻、乘客出租车出行距离、下车点经纬度。其中出行是否为节假日、下车点时刻、下车点经纬度直接根据乘客的下车点信息可以直接获取,出行距离则需要获取上下车点信息,同时结合轨迹信息判断。因此,在提取出行特征时,首先需要识别乘客的上下车点。

[0066] (1) 上车点识别

[0067] 在XX市出租车运营数据中,包含一个字段USER_ON记录在特定时刻状态下出租车是处于服务状态还是处于空驶状态,因此,乘客上车点为出租车由空驶状态变为服务状态的点,即USER_ON字段的值从0变为1,可以表示为

$$S_{\text{board}} = \{m \in M \mid p_m - p_{m-1} = 1\} \quad (1)$$

[0069] 其中 S_{board} 为上车数据记录序列标号集合, m 为数据库中所有数据记录序列标号, M

运营数据库中所有数据记录序列标号集合, p_m 为序列标号为 m 的USER_ON字段值, 当有乘客时值为1, 没有乘客时值为0。

[0070] (2) 下车点识别

[0071] 类似于上车点的识别方式, 乘客的下车点的识别为出租车由服务状态变为空驶状态, 即USER_ON字段的值从1变为0, 可以表示为

$$[0072] \quad S_{\text{alight}} = \{m \in M \mid p_{m-1} - p_m = 1\}, m \geq 2 \quad (2)$$

[0073] 其中 S_{alight} 为下车数据记录序列标号集合。

[0074] (3) 具体特征识别

[0075] 获取了上下车位置及时间之后, 是否为节假日、下车点的时刻、下车点的经纬度可以直接根据式(2)中下车点识别后的结果直接从数据库中读取

$$[0076] \quad t_{S_i^{\text{alight}}}^w \Rightarrow A_i^w \quad (3)$$

$$[0077] \quad t_{S_i^{\text{alight}}}^t \Rightarrow A_i^t \quad (4)$$

$$[0078] \quad A_i^{\log} = t_{S_i^{\text{alight}}}^{\log} \quad (5)$$

$$[0079] \quad A_i^{\text{lat}} = t_{S_i^{\text{alight}}}^{\text{lat}} \quad (6)$$

[0080] 其中 A_i^w 为乘客 i 出行是否为节假日的特征标识符, $t_{S_i^{\text{alight}}}^w$ 为乘客 i 出行的为周几, 当 $t_{S_i^{\text{alight}}}^w$ 为周一至周五, 即为工作日时 A_i^w 为0, 当 $t_{S_i^{\text{alight}}}^w$ 为周六或周日时 A_i^w 为1, A_i^t 为乘客 i 出行时段特征标识符, $t_{S_i^{\text{alight}}}^t$ 为乘客 i 到达下车点的时刻, 其中时刻表以一个小时为区间单位进行划分, 即6:00~24:00分别被划分为18个时间区段, 编号分别为1~18, A_i^{\log} 为乘客 i 的下车点经度特征标识符, $t_{S_i^{\text{alight}}}^{\log}$ 为乘客 i 下车点的经度取值, A_i^{lat} 为乘客 i 的下车点纬度特征标识符, $t_{S_i^{\text{alight}}}^{\text{lat}}$ 为乘客 i 下车点的纬度取值。

[0081] 以上特征都是可以直接根据下车点信息直接获取, 但是乘客的出行距离不能够直接获取, 这是因为出租车本身轨迹不是标准的直线, 直接通过上下车点只能获取乘客出行的直线距离, 显然不符合实际。为了获取实际的出行距离, 我们将该轨迹尽可能多的划分为如图2所示的 n 个小段, 则乘客的实际出行距离为

$$[0082] \quad l_{od}^i = \sum_{j=1}^n r_j^i \quad (7)$$

[0083] 其中 l_{od}^i 为乘客 i 从 o 点到 d 点行驶距离的计算方式, j 为将乘客 i 从 o 点到 d 轨迹划分的小段序列标识符, n 为轨迹分小段数量, r_j^i 为乘客 i 的第 j 段小段轨迹的距离长度。

[0084] 其中 r_j^i 小段轨迹的距离长度计算, 可以根据该小段的起始点经纬度进行计算,

$$[0085] \quad r_j^i = 2000R \times \arcsin \left[\sqrt{\sin^2 \left(\frac{r_j^{i-olat} - r_j^{i-dlat}}{2} \right) + \cos(r_j^{i-olat})} \right. \\ \left. \times \cos(r_j^{i-dlat}) \times \sin^2 \left(\frac{r_j^{i-olog} - r_j^{i-dlog}}{2} \right) \right] \quad (8)$$

[0086] 其中R为地球半径, r_j^{i-olat} 为 r_j^i 小段轨迹起点的纬度, r_j^{i-dlat} 为 r_j^i 小段轨迹终点的纬度, r_j^{i-olog} 为 r_j^i 小段轨迹起点的经度, r_j^{i-dlog} 为 r_j^i 小段轨迹终点的经度。

[0087] 因此具体到某个乘客的出行特征可以表示为

$$[0088] \quad A_i^{dis} = l_{od}^i \quad (9)$$

[0089] 其中 A_i^{dis} 为乘客i的出行距离特征标识符。

[0090] 需要说明的是乘客轨迹划分小段时要依据GPS触发时间间隔时间, 一般来说, 小段划分时间节点就是依据时间间隔或者是时间间隔的整数倍数。其中将以时间间隔的小段划分是在当前数据条件下最精细的划分, 其计算的距离与真实距离最为接近, 但该方式可能会造成计算机资源的损耗, 而以整数倍数划分可以节约部分计算机资源, 但是其与真实距离的接近程度相对较低。在本文中为了保证结果的精准性, 采用的以时间间隔作为小段划分条件。

[0091] 出行目的识别模型构建:

[0092] 本文提出的基于出租车数据和POI点数据的出行目的识别方法是从出行特征和所属POI点类别两个层面对出行目的进行识别, 然后依据两个层面的识别结果进行综合判断。

[0093] 出行特征判别:

[0094] 依据出行特征判断出行目的, 本质上属于分类问题, 目前该问题已经有多种方法可以解决, 如神经网络、贝叶斯、决策树、支持向量机、最近邻等方法。这些方法都是常用的依据特征推断判别类的方法, 由于不同的方法的识别效果不同, 因此在实际使用过程中需要通过测试然后选择分类方法。虽然这些方法的原理差别很大, 如决策树使用的是依据信息熵的增益, 贝叶斯是依据条件概率, 支持向量机是依据超平面等, 但是这些方法总体都可以表示为依据特征推到对象所属类别, 具体到该问题为

$$[0095] \quad f(A_i^w, A_i^t, A_i^{dis}) = purpose_i^a \quad (10)$$

[0096] 其中 $purpose_i^a$ 表示乘客i通过其出行特征推断出的出行目的, f 表示推断出行特征所用的方法。

[0097] 下车POI点类别归属:

[0098] 由于出租车具有灵活便利, 具备门到门的服务能力, 因此在调查中, 乘客乘坐出租车出行的“下车点离最终目的地的距离”这一内容进行了调查以确定乘客从下车点到达可能的最终目的地的阈值半径r。因此, 依据这个条件, 通过POI点层面也可以对其出行目的进行判断, 可以表示为

$$[0099] \quad purpose_i^{poi} = type_{poi} \quad dis(loca_i, loca_{type_{poi}}) \leq r$$

[0100] 其中 $purpose_i^{poi}$ 为通过下车点可能的POI点归属推断的乘客i的出行目的, $type_{poi}$ 为在特定下车点的距离为r的范围内存在的POI点种类集合, $dis(loca_i, loca_{type_{poi}})$ 为乘客i下车点与最终目的地的距离, r为下车点与最终目的地距离阈值。

[0101] 最终出行目的识别规则:

[0102] 由于基于POI点识别的最终结果可能出现一种出行目的、多种出行目的、无出行目的多种情况, 依据特征识别的出行目的必然是为一种出行目的, 而实际乘客出行目的为一

种出行目的。因此,在最终出行目的识别,需要将依据特征识别的出行目的和依据POI点类别识别的出行目的实现有效融合,才能够得到最终的出行目的。将两种不同的方式作为一个集合,则可能出现如下情况

$$[0103] \quad \text{purpose}_i^{\text{poi}} \cap \text{purpose}_i^a = \emptyset \quad \text{purpose}_i^{\text{poi}} = \emptyset \quad (11)$$

$$[0104] \quad \text{purpose}_i^{\text{poi}} \cap \text{purpose}_i^a = \emptyset \quad \text{purpose}_i^{\text{poi}} \neq \emptyset \quad (12)$$

$$[0105] \quad \text{purpose}_i^{\text{poi}} \cap \text{purpose}_i^a = \text{purpose}_i^a \quad \text{purpose}_i^{\text{poi}} = \text{purpose}_i^a \quad (13)$$

$$[0106] \quad \text{purpose}_i^{\text{poi}} \cap \text{purpose}_i^a = \text{purpose}_i^a \quad \text{purpose}_i^{\text{poi}} \neq \text{purpose}_i^a \quad (14)$$

[0107] 情况(11)由于基于POI点识别方式的结果为空集,即没有任何元素,最终出行目的只能为依据出行特征识别的出行目的;情况(13)由于两种方式识别的结果完全一致,最终出行目的即为任意一种方式识别的出行目的;情况(14)则是由于基于POI点识别方式的结果出行多种出行目的,且多种出行目的有一种恰好和使用特征识别的出行目的一致,最终出行目的为两者识别结果的交集,则以上三种情况的最终识别结果都可以表示为

$$[0108] \quad \text{purpose}_i^{\text{fin}} = \text{purpose}_i^a \quad (15)$$

[0109] 其中 $\text{purpose}_i^{\text{fin}}$ 为最终识别结果。

[0110] 情况(12)则是基于POI点识别方式的结果不为空且与基于特征识别的结果出现不一致的情况,因此在最终确定识别目的时存在两种方式,一种为以出行特征为主导的即最终识别目的按照式(14)(文献[19,21]的出行目的识别方法),另外一种则是以POI点为主导的,可以表示为式(16)

$$[0111] \quad \text{purpose}_i^{\text{fin}} \in \text{purpose}_i^{\text{poi}} \quad (16)$$

[0112] 式(16)表示出现两种方式识别不同情况时,采用POI点识别结果为准,但基于POI点识别的结果在非空集的情况下可能存在两种情况,即出现全是同一种出行目的或出现不全为一种目的。当全是一种出行目的时,最终出行目的即为该目的,但当出现不同种出行目的是,我们提出了两种方法,其中第一种为统计各类出行目的的数量,并将数量出现最多的类型作为最早出行目的,第二种为以离下车点距离最短的出行目的地类型作为最终出行目的。

$$[0113] \quad \text{purpose}_i^{\text{fin}} = \max(\text{countnum}(\text{purpose}_i^{\text{poi}})) \quad (17)$$

$$[0114] \quad \text{purpose}_i^{\text{fin}} = \min(\text{distance}(\text{purpose}_i^{\text{poi}})) \quad (18)$$

[0115] 其中式(17)表示以目的地类型出现次数最多的作为最终出行目的,式(18)表示以离下车点距离最短的出行目的地类型作为最终出行目的。

[0116] 依据本文所提出的方法对出租车乘客出行目的进行估计时需要判断每个POI点是否处于乘客下车点半径阈值r范围内,一种可行的方式就是对每位乘客的下车点计算其与所有POI点的距离值。但该方法在处理乘客数量较多,城市规模较大的出租车乘客的出行目的估计时将耗费大量的计算机资源。为了说明该问题,我们以图3为例进行说明,如图3所示为一位乘客下车点和需要判断的是否在半径阈值r范围内的5个POI点,依据最简单的方式即为需要计算下车点分别于POI1、POI2、POI3、POI4、POI5五个点的距离,然后再分别判断所计算的与半径r的大小。这种做法需要计算下车点与所有POI点的距离,而实际的情况

是一个城市包含大量的POI点和成千上万的乘客出行,导致计算所有下车点与所有POI点之间的距离的时间急剧增加。

[0117] 为了使所提出的方法能够应用于实际问题,我们对POI点是否位于下车点半径阈值 r 范围内的算法进行改进。为了节约计算机的计算时间,首先分析了经纬度差与距离之间的关系,在地球任意点,同一经度条件下,距离差可以直接转化为纬度差,同理,在同一纬度条件下,距离差也可以直接转化为经度差。其中在同一经度条件下,距离为 r 的两点的纬度差可以表示为如下式(19),同一纬度条件下,距离为 r 的两点的经度差可以表示为如下式(20)

$$[0118] \quad \Phi_{lat} = \frac{180r}{\pi R} \quad (19)$$

$$[0119] \quad \Phi_{lon} = \frac{180r}{\pi R \cos \varphi} \quad (20)$$

[0120] 从式(19)可以看出同一经度条件下,距离为 r 的两点的纬度差可以直接算出来,我们定义该纬度差为 Φ_{lat}^r ,根据地理知识可知,当距离为 r 的两点不在同一经度,则其纬度差小于 Φ_{lat}^r 。从式(20)可以看出,同一纬度条件下,距离为 r 的两点的纬度差与所在纬度值也有关系,在 $0 \sim 90$ 度范围内,纬度越大,则经度差越大,其中XX市纬度在北纬 $[30.083333, 31.433333]$ 区间内,因此定义在XX市最大纬度(31.433333度),距离为 r 的经度差为 Φ_{lon}^r ,依据地理知识也有,当距离为 r 的两点不在同一纬度,其经度差小于 Φ_{lon}^r 。

[0121] 通过上述分析可知如果乘客的下车点和POI点距离小于或者等于半径阈值 r 的话,则该POI点与该下车点的纬度差一定小于或等于 Φ_{lat}^r ,且经度差小于或等于 Φ_{lon}^r 。通过这个结论,我们可以直接对每位乘客下车点与所有POI点的经纬度进行预判,将POI点分为两个集合

$$[0122] \quad Set_{poi}^{i-1} = \{v_{POI}^i \mid v_{POI_{lat}}^i - v_{ALIGHT_{lat}}^i < \Phi_{lat}^r, v_{POI_{log}}^i - v_{ALIGHT_{log}}^i < \Phi_{lon}^r\} \quad (21)$$

$$[0123] \quad Set_{poi}^{i-2} = Set_{poi}^{i-1} - Set_{poi}^{i-1} \quad (22)$$

[0124] 其中: Set_{poi} 为所有研究范围区域内所有POI点的集合, Set_{poi}^{i-1} 为满足乘客 i 下车点的经度差和纬度差均在预设范围内的POI点集合, Set_{poi}^{i-2} 为研究范围内除了 Set_{poi}^{i-1} 中POI点元素的剩余POI点集合, v_{POI}^i 为乘客 i 满足后续计算条件的POI点元素, $v_{POI_{lat}}^i$ 为POI点的经度, $v_{ALIGHT_{lat}}^i$ 为乘客 i 的下车点经度, $v_{POI_{log}}^i$ 为POI点的纬度, $v_{ALIGHT_{log}}^i$ 为乘客 i 下车点的纬度, Φ_{lat}^r 为预设的最大经度差, Φ_{lon}^r 为预设的最大纬度差。

[0125] 针对每位乘客的下车点将所有POI进行分类,只需计算下车点与该下车点的集合 Set_{poi}^{i-1} 中的POI点的距离,然后进行判断,该方法可节约大量的计算时间。

[0126] 为了与现有的研究进行对比分析,首先采用了传统方法对乘客出行目的研究的方法,对出租车乘客的出行目的进行了识别,选取一个城市,获取出租车运营数据作为调查数据。即根据调查数据调查的特征:乘客的出行是否为节假日、到达目的地时间以及出行距离,对乘客出行目的进行识别,并与最终实际目的进行对比分析,验证识别模型的识别精

度。其中验证方法采用10折验证方法,即将样本随机分为10个样本数量大小一致的子样本,依次选择其中9个子样本作为训练集,1个子样本作为测试集,10次识别的准确度的平均值作为最终测试结果。采用Matlab编程,结合不同识别模型的最终识别效果如图4所示,通过识别结果可知,不同的识别模型其识别精度差别较大从47%到65.7%。

[0127] 因此,仅仅依靠出行特征进行识别,容易出行较大的误识别率。针对这个问题,我们提出的基于出租车运营数据和POI点数据的出行目的综合识别方法。

[0128] 由于在上述6种不同的识别方法中决策树的识别准确率最高,因此在出行特征层面的出行目的识别方法仍采用决策树,然后结合乘客下车POI点半径阈值 r 内的可能目的地的POI点类型进行综合决策。为了进行对比,数据集以及识别准确度的验证方法与上述一致,依据本发明方法的最终识别规则得到的对比识别结果如图5所示。

[0129] 其中决策树+POI (I) 为式 (17) 的出行目的的决策方法,决策树+POI (II) 为式 (18) 的出行目的的决策方法。通过上述不同方法的结果对比分析可知,文中所提出的在现有的出行特征识别出行目的基础上融合下车点所属POI点信息的出行目的决策方法具有明显的优势,相比现有的出行特征估计出行目的的方法,所提出的决策树+POI (I) 能够在提高6.31%的识别准确率,决策树+POI (II) 能够提高15.76%的识别准确率。

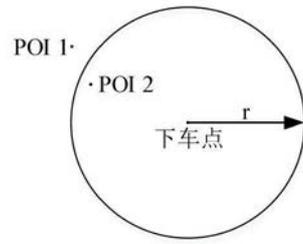


图1

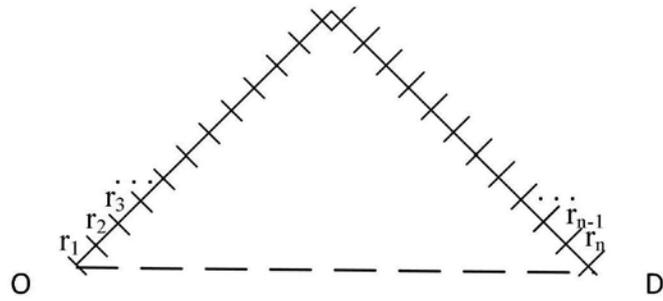


图2

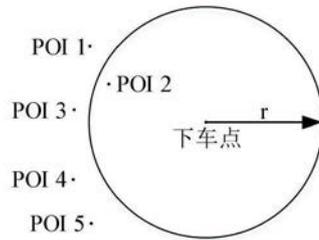


图3

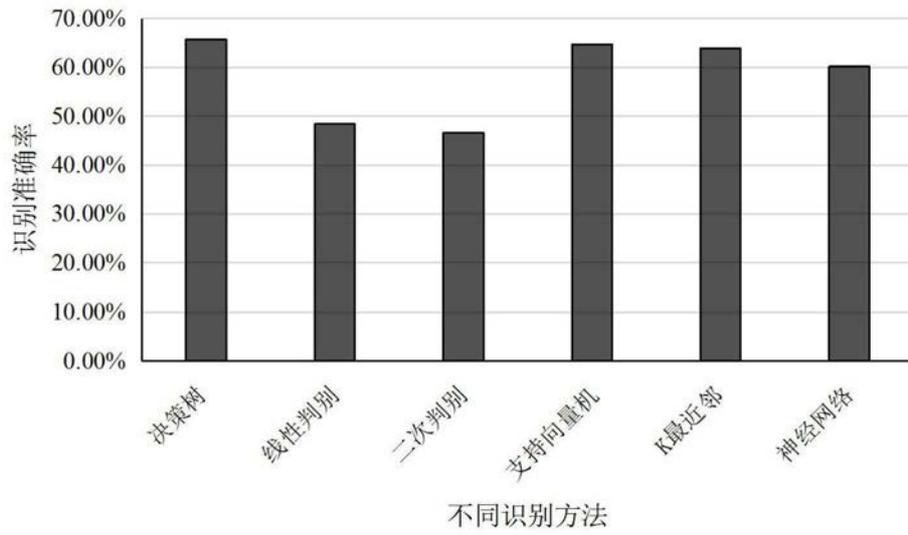


图4

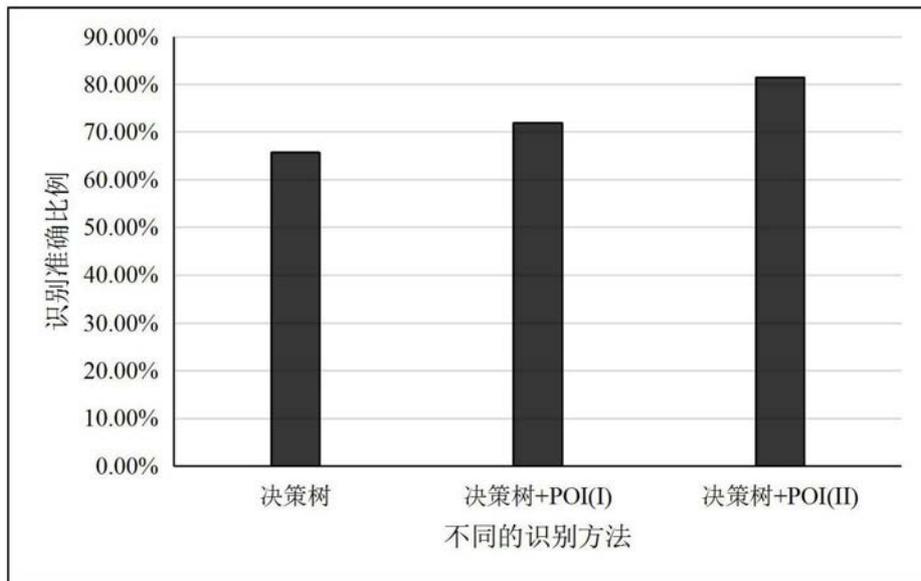


图5

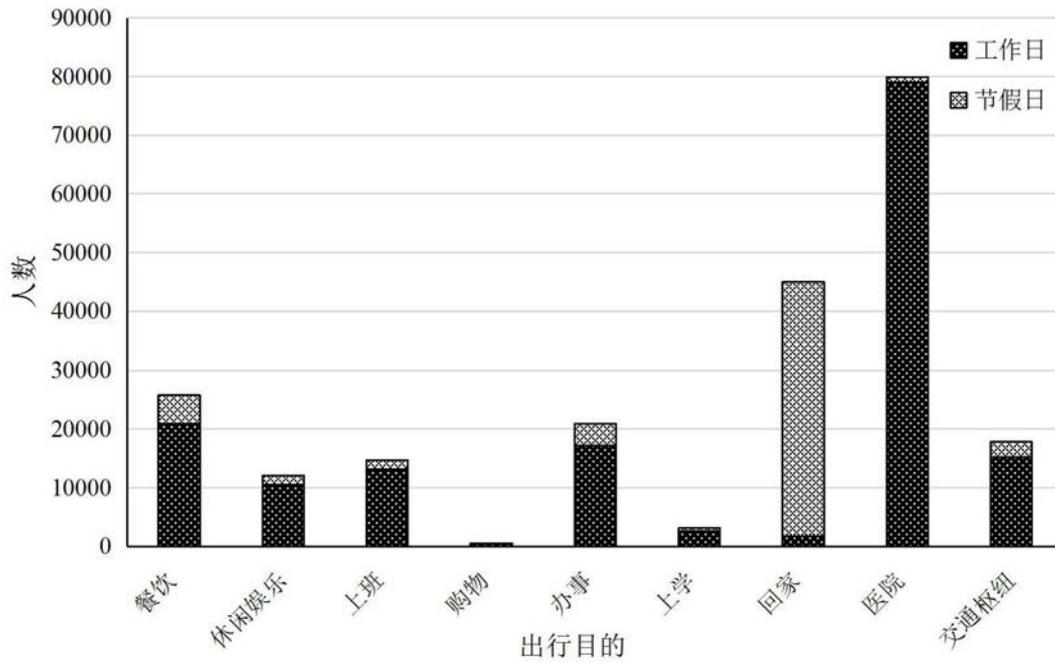


图6