



(12) 发明专利

(10) 授权公告号 CN 115017250 B

(45) 授权公告日 2024. 07. 09

(21) 申请号 202210756912.6

(22) 申请日 2022.06.30

(65) 同一申请的已公布的文献号

申请公布号 CN 115017250 A

(43) 申请公布日 2022.09.06

(73) 专利权人 东南大学

地址 210096 江苏省南京市玄武区四牌楼2号

(72) 发明人 王帅 梅洛瑜 张心睿 刘尧畅

(74) 专利代理机构 南京众联专利代理有限公司

32206

专利代理师 叶倩

(51) Int. Cl.

G06F 16/29 (2019.01)

G06Q 10/0835 (2023.01)

G06Q 10/083 (2024.01)

G06F 18/22 (2023.01)

G06N 3/047 (2023.01)

G06N 3/0442 (2023.01)

G06N 3/0455 (2023.01)

(56) 对比文件

CN 113065074 A, 2021.07.02

US 2018300400 A1, 2018.10.18

审查员 任俊鹏

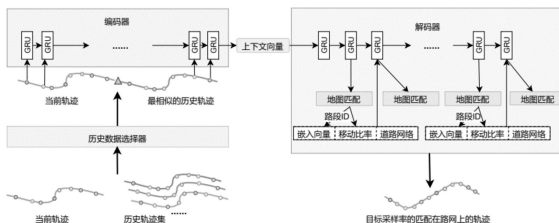
权利要求书2页 说明书5页 附图2页

(54) 发明名称

基于深度学习的稀疏轨迹数据恢复系统及方法

(57) 摘要

本发明公开了一种基于深度学习的稀疏轨迹数据恢复系统及方法,包括:历史数据选择器模块用于接收用户的历史轨迹集合和当前稀疏轨迹,对历史轨迹数据进行周期性提取和特征归一化处理后,找到与当前稀疏轨迹最相似的历史轨迹;编码器模块通过轨迹数据的序列特征获得稀疏轨迹与历史轨迹的上下文向量,同时通过独热编码对环境上下文特征进行表示,用一个全连接网络学习特征的嵌入得到特征嵌入向量,将其与上下文向量融合作为解码器输入;解码器模块引入注意力机制建模稀疏轨迹与最相似历史轨迹的全局相关性,将稀疏轨迹恢复到目标采样率,与路网匹配任务共享参数,同步进行道路网络的地图匹配,实现稀疏轨迹的数据恢复,完善轨迹路径,提高工作效率。



1. 基于深度学习的稀疏轨迹数据恢复系统, 其特征在于: 包括历史数据选择器模块、编码器模块和解码器模块,

所述历史数据选择器模块: 接收当前轨迹所属用户的历史轨迹集合和当前稀疏轨迹, 对历史轨迹数据进行周期性提取, 对其特征进行归一化处理, 并计算历史轨迹集合中与当前稀疏轨迹最相似的历史轨迹, 将其作为编码器模块的输入;

所述编码器模块: 基于轨迹数据的序列特征, 利用编码器获得稀疏轨迹与最相似历史轨迹的上下文向量;

所述编码器模块中还包括可抽取环境上下文特征的属性模块, 通过独热编码对环境上下文特征进行表示, 将独热编码的结果串联后, 通过一个全连接网络学习所有特征的嵌入, 得到特征嵌入向量;

将上下文向量与特征嵌入向量融合, 作为解码器的输入进行恢复和匹配任务;

所述解码器模块: 引入了注意力机制建模稀疏轨迹与最相似历史轨迹的全局相关性, 使用解码器中的当前隐藏状态与编码器输出的相似度来生成上下文向量, 将稀疏轨迹恢复到目标采样率, 并与路网匹配任务共享参数, 同步进行道路网络的地图匹配;

解码器模块中, 将解码器中的当前隐藏状态作为查询向量, 将编码器的输出作为键值向量, 在每个时间步 $j \in \{1, 2, \dots, n\}$ 中, 利用解码器的当前隐藏状态和编码器的输出生成上下文向量 a_j , 具体为:

$$a_j = \sum_{i=1}^n \alpha_{j,i} s_i$$

$$\alpha_{j,i} = \frac{\exp(u_{j,i})}{\sum_{i'=1}^n \exp(u_{j,i'})}$$

$$u_{j,i} = v^T * \tanh(W_h h_j + W_s s_i)$$

其中, v 、 W_h 和 W_s 是可学习参数, h_j 是解码器中的隐藏状态向量, s_i 是编码器输出的隐藏状态向量, $\alpha_{j,i}$ 是第 j 个编码器第 i 个时间步输出向量计算时的权重, $u_{j,i}$ 为更新前的权重, v^T 表示矩阵 v 的转置;

所述隐藏状态向量 h_j 为:

$$h_j = \text{GRU}(h_{j-1}, \text{id}_{j-1}, \text{ratio}_{j-1}, a_j)$$

其中, id_{j-1} 是路段ID, ratio_{j-1} 是位置点的移动比率, a_j 是解码器的当前隐藏状态和编码器的输出生成的上下文向量;

对隐藏状态 h_j 使用全连接层的Softmax函数来预测过程中的路段ID, 输出所有路段ID出现概率的预测矩阵, 矩阵中概率最大的一类即为对路段ID的预测结果; 使用全连接层的Sigmoid函数来预测移动比率; 根据路段ID和移动比率, 得到匹配在道路网络上的目标采样率的轨迹。

2. 如权利要求1所述的基于深度学习的稀疏轨迹数据恢复系统, 其特征在于: 所述历史数据选择器模块中, 通过计算欧几里得距离 d 来确定历史轨迹集合中与当前轨迹最相似的历史轨迹, 所述欧几里得距离 d 的计算公式如下:

$$d = \sqrt{\sum_{i=1}^N \frac{(x_i - y_i)^2}{S_i^2}}$$

其中, x_i, y_i 表示向量 x, y 在第 i 维的值, S_i 是向量在第 i 维的标准差;

欧几里得距离 d 最小的, 即为历史轨迹集合中与当前轨迹最相似的历史轨迹。

3. 如权利要求2所述的基于深度学习的稀疏轨迹数据恢复系统, 其特征在于: 所述编码器模块中, 使用门控递归单元GRU作为编码器来获得稀疏轨迹与最相似历史轨迹的上下文向量, 所述门控递归单元GRU使用更新门 z 和复位门 r 来控制时间同步中的信息, 在每个时间步 $i \in \{1, 2, \dots, n\}$ 中, 隐藏状态向量 s_i 如下:

$$z_i = \sigma(W_z * [s_{i-1}, u_i] + b_z)$$

$$r_i = \sigma(W_r * [s_{i-1}, u_i] + b_r)$$

$$\tilde{s}_i = \tanh(W_z * [r_i * s_{i-1}, u_i] + b_z)$$

$$s_i = (1 - z_i) * s_{i-1} + z_i * \tilde{s}_i$$

其中, W_x 是门 (x) 的权重, b_x 代表门 (x) 的偏置项, z_i 和 r_i 代表更新门和复位门, u_i 表示转成离散单元后的坐标点, \tilde{s}_i 代表更新前的状态向量 s_i , σ 代表激活函数sigmoid;

最后一个隐藏状态向量 s_n 为上下文向量, 它将被作为解码器的输入, 成为解码器的初始隐藏状态。

4. 如权利要求2所述的基于深度学习的稀疏轨迹数据恢复系统, 其特征在于: 所述属性模块中的环境上下文特征至少包括天气状况、是否是假期和以小时为度量的时间点, 所述环境上下文特征是固定的。

5. 使用如权利要求2所述系统的基于深度学习的稀疏轨迹数据恢复方法, 其特征在于, 包括如下步骤:

S1, 最相似的历史轨迹确定: 收集当前轨迹所属用户的历史轨迹集合, 对历史轨迹数据进行周期性提取, 对其特征进行归一化处理, 并计算历史轨迹集合中与当前稀疏轨迹最相似的历史轨迹, 确定最相似的历史轨迹;

S2, 上下文向量的获得: 基于轨迹数据的序列特征, 利用编码器获得稀疏轨迹与最相似历史轨迹的上下文向量;

S3, 环境特征嵌入向量的获得: 通过独热编码对环境上下文特征进行表示, 将独热编码的结果串联后, 通过一个全连接网络学习所有特征的嵌入, 得到特征嵌入向量;

S4, 向量融合: 将步骤S2获得的上下文向量与步骤S3获得的特征嵌入向量融合, 作为解码器的输入进行恢复和匹配任务;

S5, 解码器执行: 引入了注意力机制建模稀疏轨迹与最相似历史轨迹的全局相关性, 使用解码器中的当前隐藏状态与编码器输出的相似度来生成上下文向量, 将稀疏轨迹恢复到目标采样率, 并与路网匹配任务共享参数, 同步进行道路网络的地图匹配。

基于深度学习的稀疏轨迹数据恢复系统及方法

技术领域

[0001] 本发明属于计算机信息计算技术领域,尤其涉及轨迹移动数据的获取与恢复,主要涉及了一种基于深度学习的稀疏轨迹数据恢复系统及方法。

背景技术

[0002] 快递物流是电子商务中的重要组成部分,“末端物流配送”是全链路物流运转流程中快递员将包裹送达消费者手中的阶段,是完成物流配送服务重要环节。随着移动计算技术的进步与发展,物流运输全流程中许多环节尤其是末端配送环节都部署了移动定位设备,用于时空轨迹数据的采集,这些时空轨迹数据包含的信息能够表示人类、车辆的移动性。在末端物流配送中,对于快递员轨迹数据的分析利用有着非常重要的作用,如交付时间预测、快递员配送调度等,可以提高快递配送的效率,优化快递配送服务。

[0003] 由于全球定位系统(Global Positioning System,GPS)通信时的损耗丢失、能量限制等原因,获取到的轨迹数据往往不能直接满足下游应用的需求:采集到的GPS数据会存在缺失点、缺失段;同时,为了节约能量,采集设备选取的采样率一般较低,实际采集到的数据中,大部分GPS数据的采样间隔在2分钟以上,这种情况下采集的轨迹数据相对稀疏,并不能很好地支持基于轨迹数据的应用。同时,由于GPS漂移等问题,采集到的位置点定位并不完全准确,不能精确匹配在道路网络上。这种稀疏的、偏离道路网络的轨迹数据,在下游各类数据驱动的应用中,不能支持其达到满意的效果。尤其由于快递员的移动位置涉及城市主干道、社区道路、乡村小路等多种复杂的道路类型,其应用的有效性更加依赖于轨迹数据是否详细准确。因此,对于快递员配送的时空轨迹数据进行挖掘之前,大部分应用均需要对稀疏的轨迹数据做恢复,并将其匹配到道路网络上,才能准确、高效地完成相应的任务。

发明内容

[0004] 本发明正是针对现有技术中的问题,提供一种基于深度学习的稀疏轨迹数据恢复系统及方法,包括历史数据选择器模块、编码器模块和解码器模块,历史数据选择器模块用于接收当前轨迹所属用户的历史轨迹集合和当前稀疏轨迹,对历史轨迹数据进行周期性提取和特征归一化处理后,找到与当前稀疏轨迹最相似的历史轨迹,作为编码器模块的输入;编码器模块通过轨迹数据的序列特征获得稀疏轨迹与历史轨迹的上下文向量,同时通过独热编码对环境上下文特征进行表示,用一个全连接网络学习所有特征的嵌入得到特征嵌入向量,将其与上下文向量融合作为解码器的输入;解码器模块引入了注意力机制建模稀疏轨迹与最相似历史轨迹的全局相关性,将稀疏轨迹恢复到目标采样率,并与路网匹配任务共享参数,同步进行道路网络的地图匹配。本系统可实现稀疏轨迹的数据恢复,完善轨迹路径,有利于各种基于轨迹数据挖掘工作的顺利开展,提升路径准确度,提高工作效率。

[0005] 为了实现上述目的,本发明采取的技术方案是:基于深度学习的稀疏轨迹数据恢复系统,包括历史数据选择器模块、编码器模块和解码器模块,

[0006] 所述历史数据选择器模块:接收当前轨迹所属用户的历史轨迹集合和当前稀疏轨迹,对历史轨迹数据进行周期性提取,对其特征进行归一化处理,并计算历史轨迹集合中与当前稀疏轨迹最相似的历史轨迹,将其作为编码器模块的输入;

[0007] 所述编码器模块:基于轨迹数据的序列特征,利用编码器获得稀疏轨迹与最相似历史轨迹的上下文向量;

[0008] 所述编码器模块中还包括可抽取环境上下文特征的属性模块,通过独热编码对环境上下文特征进行表示,将独热编码的结果串联后,通过一个全连接网络学习所有特征的嵌入,得到特征嵌入向量;

[0009] 将上下文向量与特征嵌入向量融合,作为解码器的输入进行恢复和匹配任务;

[0010] 所述解码器模块:引入了注意力机制建模稀疏轨迹与最相似历史轨迹的全局相关性,使用解码器中的当前隐藏状态与编码器输出的相似度来生成上下文向量,将稀疏轨迹恢复到目标采样率,并与路网匹配任务共享参数,同步进行道路网络的地图匹配。

[0011] 为了实现上述目的,本发明还采取的技术方案是:基于深度学习的稀疏轨迹数据恢复方法,包括如下步骤:

[0012] S1,最相似的历史轨迹确定:收集当前轨迹所属用户的历史轨迹集合,对历史轨迹数据进行周期性提取,对其特征进行归一化处理,并计算历史轨迹集合中与当前稀疏轨迹最相似的历史轨迹,确定最相似的历史轨迹;

[0013] S2,上下文向量的获得:基于轨迹数据的序列特征,利用编码器获得稀疏轨迹与最相似历史轨迹的上下文向量;

[0014] S3,环境特征嵌入向量的获得:通过独热编码对环境上下文特征进行表示,将独热编码的结果串联后,通过一个全连接网络学习所有特征的嵌入,得到特征嵌入向量;

[0015] S4,向量融合:将步骤S2获得的上下文向量与步骤S3获得的特征嵌入向量融合,作为解码器的输入进行恢复和匹配任务;

[0016] S5,解码器执行:引入了注意力机制建模稀疏轨迹与最相似历史轨迹的全局相关性,使用解码器中的当前隐藏状态与编码器输出的相似度来生成上下文向量,将稀疏轨迹恢复到目标采样率,并与路网匹配任务共享参数,同步进行道路网络的地图匹配。

[0017] 与现有技术相比,本发明在轨迹恢复的方法中,加入历史数据选择器,利用周期性特征进行针对轨迹的恢复与路网匹配任务;设计了一个周期性感知系统,通过挖掘历史轨迹数据的周期性模式来提升复杂路网中地图匹配的轨迹恢复任务的效果;本方法和系统可实现稀疏轨迹的数据恢复,完善轨迹路径,提升路径准确度,提高工作效率,尤其适用于快递行业,对稀疏的快递员轨迹数据进行高采样率的恢复,提高快递员轨迹的采样率和匹配精度,并将之根据快递员配送的特点匹配到地图的道路网络上,得到的恢复并匹配后的数据是各类基于轨迹的下游应用的有效数据支撑,有利于各种基于快递员配送轨迹数据挖掘工作的顺利开展,从而加强末端物流配送的效率,提高整个物流领域的服务质量。

附图说明

[0018] 图1是本发明基于深度学习的稀疏轨迹数据恢复系统的结构示意图;

[0019] 图2是本发明实施例2中实验采集的部分位置点数据图;

[0020] 图3是本发明实施例2中地图匹配点表示方法的示意图;

[0021] 图4是本发明实施例2中实验结果可视化图。

具体实施方式

[0022] 下面结合附图和具体实施方式,进一步阐明本发明,应理解下述具体实施方式仅用于说明本发明而并不用于限制本发明的范围。

[0023] 实施例1

[0024] 基于深度学习的稀疏轨迹数据恢复系统,如图1所示,包括历史数据选择器模块、编码器模块和解码器模块,采用序列到序列模型来解决地图约束的轨迹恢复问题,采用序列到序列模型的编码器-解码器结构作为本发明模型的主要结构。

[0025] 所述历史数据选择器模块:接收当前轨迹所属用户的历史轨迹集合和当前稀疏轨迹,对历史轨迹数据进行周期性提取,对其特征进行归一化处理,并计算历史轨迹集合中与当前稀疏轨迹最相似的历史轨迹,将其作为编码器模块的输入;

[0026] 历史数据选择器模块接收当前轨迹所属的用户的历史轨迹集合和当前轨迹,首先对输入的轨迹数据特征进行编码建模并计算,其欧几里得距离 d , d 的计算公式如下:

$$[0027] \quad d = \sqrt{\sum_{i=1}^N \frac{(x_i - y_i)^2}{S_i^2}}$$

[0028] 其中 x_i, y_i 表示向量 x, y 在第 i 维的值, S_i 是其在第 i 维的标准差。历史轨迹集合中与当前轨迹归一化欧几里得距离最小的即为最相似的历史轨迹,将其与当前轨迹连接,共同作为编码器的输入。

[0029] 所述编码器模块:编码器将低采样率的轨迹与最相似的历史轨迹进行编码,作为上下文向量,输入到解码器中进行恢复和匹配任务。

[0030] 将历史数据选择器的输出作为编码器的输入进行恢复和匹配任务。将GPS坐标点 p_i 转换成离散单位 $u_i = \langle x_i, y_i, tid_i \rangle$,其中 x_i, y_i 代表网格单元, tid_i 是目标轨迹中该点的索引。通过轨迹数据的有序列特征,使用门控递归单元作为编码器来获得低采样率轨迹与历史轨迹的上下向量。GRU使用更新门 z 和复位门 r 来控制时间同步中的信息。更新门用于决定由多少以前的时间步的信息需要传递到下一时间步。重置门则决定有多少信息需要被遗忘。在每个时间步 $i \in \{1, 2, \dots, n\}$ 中,我们可以得到隐藏状态向量 s_i 如下:

$$[0031] \quad z_i = \sigma(W_z * [s_{i-1}, u_i] + b_z)$$

$$[0032] \quad r_i = \sigma(W_r * [s_{i-1}, u_i] + b_r)$$

$$[0033] \quad \tilde{s}_i = \tanh(W_z * [r_i * s_{i-1}, u_i] + b_z)$$

$$[0034] \quad s_i = (1 - z_i) * s_{i-1} + z_i * \tilde{s}_i$$

[0035] 其中, W_x 是门 (x) 的权重, b_x 代表门 (x) 的偏置项。将其简化为:

$$[0036] \quad s_i = \text{GRU}(s_{i-1}, u_{i-1})$$

[0037] 最后一个隐藏状态向量 s_n 将被视为上下文向量,也作为解码器的输入,成为解码器的初始隐藏状态。

[0038] 同时,由于在现实世界中快递员并非以匀速移动,恢复的准确性还受到交通状况的影响。其移动速度还受到天气、日期、时间等环境的影响,本发明还对编码器的输出结合

了属性模块,将环境上下文特征与编码器的最后一个隐藏状态向量,也就是上下文向量进行结合,共同作为解码器的输入。

[0039] 为了抽取环境上下文特征 f_e ,结合天气状况、假期和时间的属性来抽取环境上下文特征,天气状况包括晴天、多云、阴天等,假期指轨迹所在当天是否是假期,时间为一天中以小时度量的时间点。因为某条轨迹的环境上下文是固定的,其不会随着轨迹时间步的展开而发生变化,因此本发明将环境背景特征作为一个较为独立的模块来实现。由于上述因素属于分类值,不能直接输入神经网络且每个因素的维度较小,因此本发明使用独热编码来对这些特征进行表示,在将独热编码的结果串联起来之后,通过一个全连接网络(Fully Convolutional Network,FCN)来学习所有特征的嵌入,然后将得到的特征嵌入向量与编码器的最后一个隐藏状态也就是上下文向量融合,以作为解码器的输入,也就是初始隐藏向量:

$$[0040] \quad h_0 = \text{FCN}(s_n, \text{Emb}(f_e))$$

[0041] 由于环境上下文特征不会随着轨迹的移动而发生明显变化。因此,直接将嵌入向量与编码器的上下文向量结合,作为解码器的初始输入。

[0042] 所述解码器模块,解码器用于将低采样率轨迹恢复到目标采样率,并与路网匹配任务共享参数,同步进行道路网络的地图匹配。解码器接收来自编码器的结合环境上下文特征向量的上下文向量 s_i 作为输入,使用门控递归单元。

[0043] 解码器中引入了注意力机制使输入的低采样率轨迹与历史轨迹的全局相关性。在计算注意力的过程中,使用查询向量(query)与键值向量(key)的相似度来生成上下文向量 a ,其中将解码器中的当前隐藏状态作为查询向量,将编码器的输出作为键值向量。在发明工作中,每个时间步 $j \in \{1, 2, \dots, n\}$ 中,利用解码器的当前隐藏状态和编码器的输出生成上下文向量 a_j ,其中 a_j 表示为:

$$[0044] \quad a_j = \sum_{i=1}^n \alpha_{j,i} s_i$$

$$[0045] \quad \alpha_{j,i} = \frac{\exp(u_{j,i})}{\sum_{i'}^n \exp(u_{j,i'})}$$

$$[0046] \quad u_{j,i} = v^T * \tanh(W_h h_j + W_s s_i)$$

[0047] 其中 v 、 W_h 和 W_s 是可学习参数, h_j 是解码器中的隐藏状态, s_i 是编码器的输出。在每个时间步 $j \in \{1, 2, \dots, n\}$ 中,可以得到隐藏状态向量 h_j 如下:

$$[0048] \quad h_j = \text{GRU}(h_{j-1}, \text{id}_{j-1}, \text{ratio}_{j-1}, a_j)$$

[0049] 然后,对隐藏状态 h_j 使用全连接层的Softmax函数来预测过程中的路段ID用于道路、网络匹配任务,输出所有路段ID的出现概率的预测矩阵,矩阵中概率最大的一类即为对路段ID的预测结果,使用全连接层的Sigmoid函数来预测移动比率。每个时间步中编码器输出隐藏状态 h_j 后,直接在地图匹配任务中预测它的路段ID和移动比率ratio,最终得到匹配在道路网络上的目标采样率的轨迹。

[0050] 实施例2

[0051] 如图2所示,为北京市某快递员2021年7月某周内的GPS位置点,将其作为原始数据输入;经异常点去除、时间间隔的轨迹分割得到轨迹集合后,将其输入到历史轨迹选择器

中,选择器将遍历该快递员历史轨迹集合,与最新待恢复轨迹进行相似度计算,选择相似度最高的一条历史轨迹,与当前轨迹拼接,输入编码器;编码器中通过门控循环单元GRU逐个时间步处理得到最终的隐藏状态向量,也就是上下文向量;同时输入7月的日期分布与天气情况,对应待恢复轨迹所在日期,对相应环境特征进行独热编码,通过全连接网络学习得到特征嵌入向量,将其与上下文向量拼接;将拼接后的向量输入解码器,解码器中同时输入北京市对应地区的路网信息,在编码器中通过多任务学习共享参数,进行高采样率轨迹序列的恢复和道路匹配任务,输出匹配在道路网络上的高采样率的轨迹,其中为了保证轨迹约束在道路网络上,选取位置点表示方式如图3所示,每个点使用路段ID和位置比率两个值来表示,路段ID表示位置点所在路段,位置比率表示位置点在路段上的位置。如图4所示,即为北京市某快递员某段稀疏轨迹的道路网络约束的恢复结果,将稀疏的以2分钟为采样间隔的偏离路网的轨迹,恢复至以30秒为采样间隔的匹配在道路网络上的轨迹。

[0052] 需要说明的是,以上内容仅仅说明了本发明的技术思想,不能以此限定本发明的保护范围,对于本技术领域的普通技术人员来说,在不脱离本发明原理的前提下,还可以做出若干改进和润饰,这些改进和润饰均落入本发明权利要求书的保护范围之内。

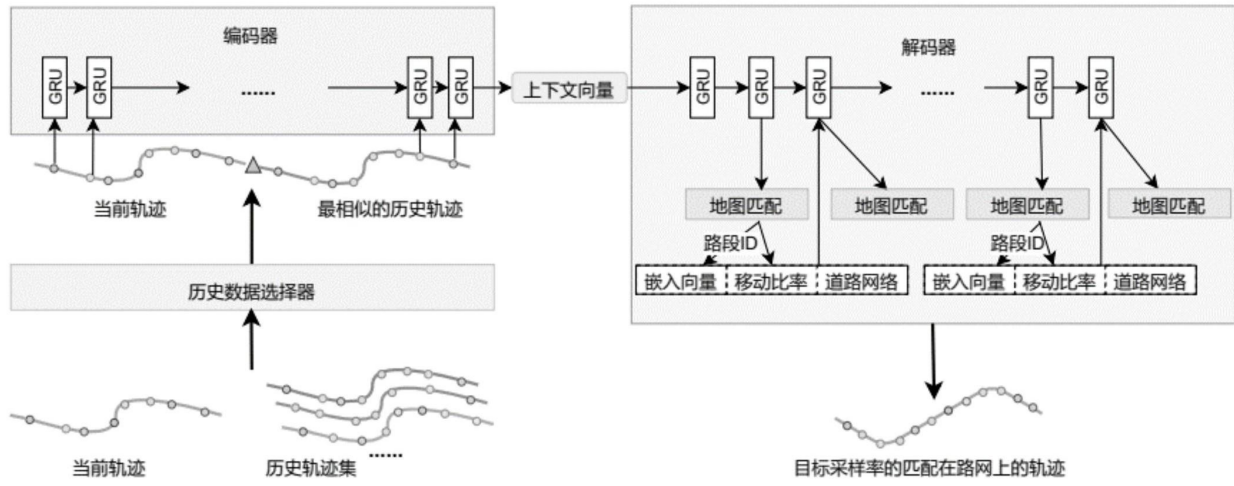


图1

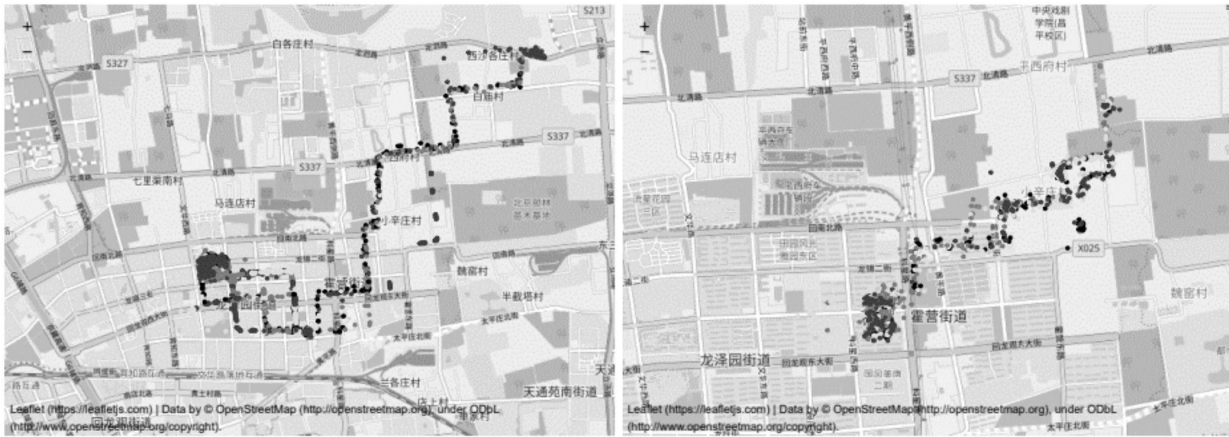


图2

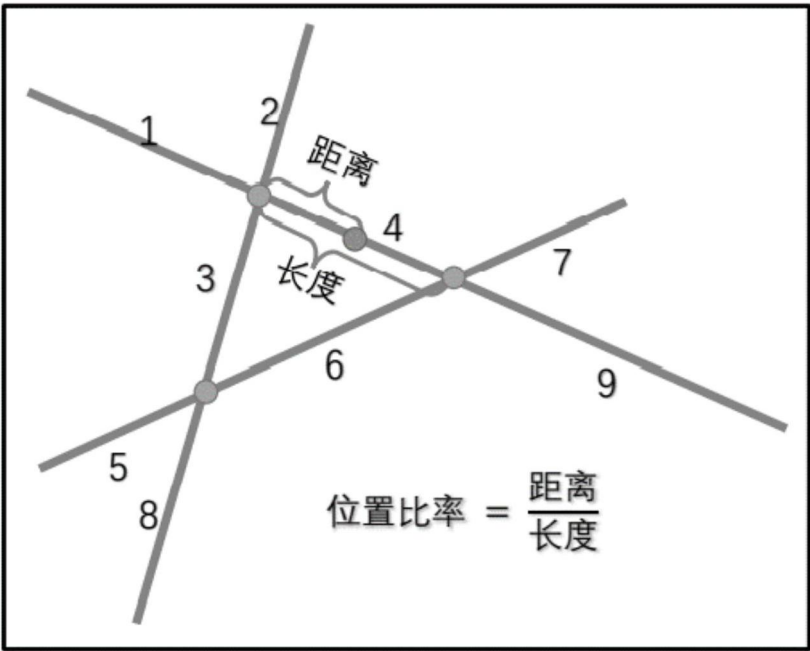


图3

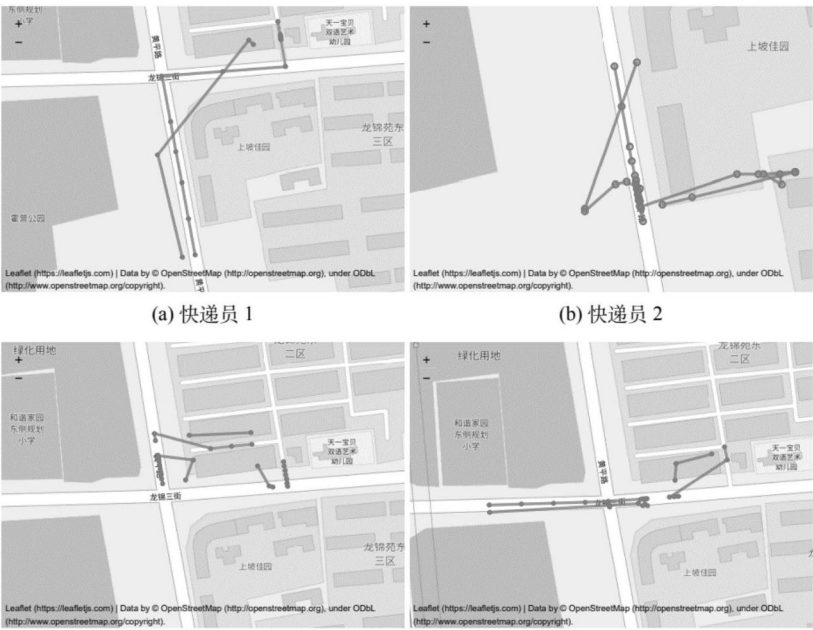


图4