

路网上基于时空锚点的移动对象群体和个体运动监测方法

韩京宇^{1,2,3} 许梦婕^{1,2} 朱曼^{1,2}

1 南京邮电大学计算机学院 南京 210023

2 江苏省大数据安全与智能处理重点实验室 南京 210023

3 东南大学计算机网络和信息集成教育部重点实验室 南京 211189

摘要 为了实时监控路网上移动对象(车辆)的运动,各移动对象不断向中心服务器汇报其位置,中心服务器存储数据以响应用户的各种查询。此类方法不仅通信开销巨大,增加服务器负载,而且不能同时满足群体态势感知和个体移动对象位置追踪的需求。因此,提出一种基于时空锚点的双粒度移动感知(Double-granularity Movement Detection Based on Spatial-temporal Anchors, DMDSA)框架,将移动对象嵌入时空网格,其经过时空锚点时向服务器汇报其运动模式,实现对群体运动的感知和个体移动的追踪。离线阶段,服务器从历史轨迹中挖掘运动模式;移动对象运动时,服务器结合挖掘的运动模式,在线计算聚合模式表征群体运动,并采用最大似然估计确定目标的运动模式,实现群体态势感知;进一步,采用锚点独立策略和锚点序列策略识别最可能的运动序列,实时追踪个体对象的运动。在模拟数据集和实际数据集上的实验表明,所提方法在大幅度减小位置汇报代价的前提下,不仅能够准确地监控区域的群体运动态势,并且能够有效地追踪和预测个体移动对象的位置,有助于智慧城市的建设。

关键词: 路网;移动对象;位置汇报;群体运动;位置追踪

中图法分类号 TP311.132

Detecting Group-and-individual Movements of Moving Objects Based on Spatial-Temporal Anchors of Road-network

HAN Jing-yu^{1,2,3}, XU Meng-jie^{1,2} and ZHU Man^{1,2}

1 School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

2 Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

3 Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing 211189, China

Abstract To detect the movement of moving objects (vehicles) in real-time fashion, every moving object continuously reports its latest positions to the sever and the server keeps the data to answer queries posed by users, which can incur great communication cost and pressing overload on the server. In particular, this method cannot effectively detect the group movements and track the individual objects simultaneously. Therefore, this paper proposes a framework of double-granularity movement detection based on spatial-temporal anchors (DMDSA), which can effectively detect group movement and track individual positions by embedding each moving object into a spatial-temporal grid and reporting its movement pattern to servers whenever the moving object passes the anchor of each grid cell. During the offline stage, movement patterns of each grid cell are mined from historical trajectories by the server. During the running of moving objects, the server detects the group movements using the maximum-likelihood estimation by aggregating the mined movement patterns. Furthermore, independent-anchor policy and sequenced-anchor policy are used to identify the most likely running path, thus tracking each moving object in real-time. The experimental results on synthetic and real data sets demonstrate that DMDSA framework can not only detect the group movement effectively but also track individual object precisely with the great reduction of communication cost.

Keywords Road-network, Moving object, Location reporting, Group movement, Position tracking

到稿日期:2019-11-11 返修日期:2020-04-26 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:东南大学计算机网络和信息集成教育部重点实验室(K93-9-2015-07C);国家自然科学基金项目(61003040,61602260,61876091);江苏省社科基金重点项目(18GLA004)

This work was supported by the Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education (K93-9-2015-07C), National Natural Science Foundation of China (61003040,61602260,61876091) and Social Science Foundation of Jiangsu Province (18GLA004).

通信作者:韩京宇(jyhan@njupt.edu.cn)

1 引言

随着计算机、通讯技术的飞速发展和智能移动终端的普及,路网上移动车辆的群体态势感知和实时追踪成为了交通导航、智能拼车、旅游推荐等位置相关服务(Location Based Service, LBS)的基础^[1-2]。路网上的移动对象根据一定策略不断向中心服务器汇报自己的运动状况,中心服务器存储并索引数据,响应群体态势查询或对移动对象的跟踪查询,如“南京中山南路新街口附近车辆平均速度是多少?通行是否顺畅?”“车辆苏 AE676 的当前位置?”。位置相关服务的核心是移动端的位置更新策略和服务器端的位置和态势计算。

位置更新策略分成单向位置更新和双向位置更新。在单向位置更新策略中,移动对象根据一定策略不断向服务器汇报其运动状况,服务器根据最近汇报的运动矢量来推断移动对象的最新位置;当移动对象的实际位置和推断位置的偏差超过一定阈值时向中心服务器汇报^[3-4]。双向位置更新策略中^[5-6],服务器根据一定策略向移动对象下发查询或者估算的位置,若移动对象发现自己落入查询范围或者与估算的位置相差超过一定阈值,则进行汇报。这类方法存在如下不足:1)频繁的位置汇报会占用大量带宽,导致高昂的通信成本,制约位置相关服务的普及;2)每次移动对象进行位置汇报时,服务器都要插入或者修改数据,加剧了服务器负载;3)单向位置更新和双向位置更新,均是以移动对象个体为基础的汇报,导致不同区域汇报密度相差巨大,不能有效地对区域交通状况进行全面监测。

为实现区域群体态势监测,文献[7-8]采用聚集计算实时探测路网交通流的变化情况,文献[9]提出通过构建索引,迭代式地获得空间聚集计算结果,文献[10]通过离线挖掘提炼运动模式,发现群体态势。这些方法或基于全网数据收集,代价高昂,或不适用于在线场景。近年来,传感探测得到长足发展^[11-12],但这类技术侧重群体信息的收集,不能对个体移动对象的位置进行追踪。

本文提出一个基于时空锚点的双粒度移动感知(DMD-SA)框架,由移动端的基于时空锚点的位置汇报、服务器端的群体态势监控和个体运动追踪组成。该框架先对历史数据进行离线挖掘,提取路网上不同时空区域的移动对象运动模式(速度分布和最大加速度分布)。移动对象运动时,为实现均衡取样汇报,将整个路网时空划分成若干时空单元格,将每个时空单元格中心作为汇报锚点,即位置汇报的期望时空点。在服务器端,为响应用户的群体态势查询,采用聚合模式表征区域交通运动态势,采用最大似然估计计算区域的目标运动模式。最后,分别依据锚点独立策略和锚点序列策略在线推断移动对象的运动,从而响应个体追踪查询。

本文框架是一个可伸缩、低能耗的解决方案,在大幅降低通信代价、减轻服务器负载的前提下,不仅可以实时探测路网上不同时空区域的群体运动态势,而且能有效追踪各个移动对象的运动。

2 相关研究

2.1 移动对象的位置更新和追踪

移动对象的位置汇报分成单向通信汇报和双向通信汇

报。单向通信汇报以 Wolfson 的 MOST (Moving Objects Spatio-Temporal)模型为代表^[3],移动对象的运动建模为一个三元组,包括更新时间、更新位置和运动函数。在查询时,服务器通过位置函数推断移动对象的当前位置,当实际位置与推断位置的偏差超过某个阈值时,移动端会向服务器发送位置更新。文献[4,13]在以元路段为主的位置汇报方式的基础上,进一步提出了基于道路的汇报,并结合加速度画像(Acceleration Profile)来进行位置汇报。文献[14]提出利用移动端和服务器的通讯尾时,合并移动端的多个应用发送位置更新。文献[15]提出通过局部链接(如 Wifi),收集周围移动对象的位置信息,汇总后通过通信链路向中心服务汇报,从而减小频繁汇报的通信代价。

双向通信汇报主要分成两大类。一类是服务器和移动端协同计算前方各个位置点的预期到达时间,如果实际运动的偏差超过阈值,则移动对象进行汇报。文献[6]采用一个时空安全区域(Spatial-Temporal Safe Region, STSR)来表征移动对象的位置和速度,在移动对象运动时采取主动更新和被动更新策略进行位置汇报。前者指移动对象发现自己超出预测区域,进行主动汇报;后者指服务器在回答范围查询时,若发现某些移动对象超出时空安全区域,则会向移动对象询问确认。文献[5]提出当移动对象上路时,首先把目的地发送到服务器,服务器计算各个地点的预期到达时间,并将其下发给移动对象。移动对象到达某个地点时,如果超过了预期时间,则向服务器汇报。另一类是服务器广播查询,移动对象根据查询是否与自己相关,来决定是否向服务器汇报。文献[16]提出了一种基于围区(precinct)和相遇点(encounter point)划分的双向通信汇报方式,每个移动对象只在自己的围区和查询相交时才进行汇报。文献[17]提出两种位置更新方式,即按需访问(on-demand access)和广播频道(broadcast channel),服务器周期性地广播查询,移动端若发现自己的位置落入查询范围则进行上报。文献[18]提出一种基于划分的延迟更新方法(Partition-based Lazy Update, PLU),服务器维护一个全局的位置信息表(Location Information Table, LIT),记录所有的查询;移动对象下载与自己相关的位置消息表,当服务器端接收到新的查询时,会向相关移动端广播。

2.2 群体交通状况探测和移动模式挖掘

群体交通状况探测方法可以分成两大类,即在线探测方法和离线探测方法。对于在线探测方法,文献[7]提出基于元路段的在线交通流统计方法,通过联机分析获得各个路段、各个时间窗口的交通流,从而获得实时的群体交通状况图景;文献[8]提出集成 GPS 数据和蓝牙监测数据以在线探测道路的交通状况;文献[9]提出通过构建索引,迭代式地获得空间聚集计算结果;文献[12]提出通过数据融合的方法在线检测高速公路的车速状况。对于离线探测方法,文献[10]通过离线挖掘来提炼运动模式,从而发现群体态势。以上方法或基于收集的全部轨迹数据,代价高昂,或不适用于在线场景。文献[11]提出了一种分布式的传感信息收集方法,但汇报分布不均匀,不能根据需要进行动态调整。

近年来,如何通过离线数据挖掘发现群体运动模式,以辅助决策和预测受到了前所未有的关注。文献[19]通过聚类发现城市热点、路况和人群流向。文献[20]提出在轨迹数据上

通过密度聚类,来发现簇的生成、扩展、收缩和合并,进而发现运动模式的变化。文献[21]提出在 Apache Spark 上采用并行发掘算法提取群体运动(co-movement)模式。文献[22]采用网格上剪枝的方法来逐步识别热点活动区域。文献[23]提出通过轨迹插点、时间弹性匹配和序列对齐发现同行护航模式。文献[24]提出采用数据挖掘的方法进行最优路径选择。

3 位置管理框架和数据预处理

本文提出的框架包含离线模式挖掘和在线运动两个阶段。在离线模式挖掘阶段,服务器提取路网时空网格的运动模式参数,并创建时空网格索引,移动端加载路网数据和运动模式参数。在在线运动阶段,移动对象在路网上运行时,根据配置参数动态计算所经过的时空单元格的锚点,并按照一定的策略,向中心服务器汇报自己的轨迹点;服务器根据收到的汇报轨迹点和预先挖掘的运动模式参数,响应查询。

3.1 相关定义和位置管理框架

定义 1(路网和元路径) 路网是一个无向图 $NT(V, E)$, V 是顶点集合, E 是连接任意两个顶点的元路径集合,每个元路径 $e \in E$ 由多线段构成,即 $e = \langle p_0, \dots, p_k, \dots, p_z \rangle$,其中 p_k 代表线段顶点的经纬度坐标。

定义 2(时空网格和时空单元格) 给定时间间隔 a 和空间距离 b ,元路径 $e \in E$ 的时空网格记为 $grid(e) = \langle \langle t_0, \dots, t_i, \dots, t_m \rangle, \langle p_0, \dots, p_j, \dots, p_n \rangle \rangle$,其中 t_i 是时刻点,满足 $|t_i - t_{i-1}| = a$; p_j 是经纬度坐标,满足 $ntdist(p_j, p_{j-1}) = b$; $ntdist$ 代表沿路网距离。元路径 e 的第 (i, j) 个时空单元格记为 $e.cl(i, j) = [t_{i-1}, t_i], [p_{j-1}, p_j]$ 。

定义 3(锚点和锚点邻域) 锚点 (ct, cs) 是每个时空单元格的中心,是移动对象汇报的期望时空点,锚点邻域为 $([ct - q/2, ct + q/2], [cs - r/2, cs + r/2])$, $[ct - q/2, ct + q/2]$ 是以 ct 为中心、长度为 q 的时间间隔, $[cs - r/2, cs + r/2]$ 是以 cs 为中心、长度为 r 的路网空间。

定义 4(轨迹) 给定移动对象 O ,其轨迹 $tr(O) = \langle mv_0, \dots, mv_i, \dots, mv_n \rangle$, $mv_i = (p_i, v_i, av_i, t_i)$ 是四元组轨迹点,其中 p_i, v_i, av_i, t_i 分别代表经纬度、瞬时速度、瞬时加速度和时刻。

移动对象根据速度和最大加速度的不同,呈现不同的运动模式,如高速一剧变模式、高速一稳态模式等,运动模式的定义如定义 5 所示。

定义 5(运动模式) 运动模式是一个时空单元格内若干移动对象呈现的整体运动特征,表征为一个二维高斯分布 $N(u_1, u_2, \delta_1, \delta_2)$,记为:

$$mp(v, w) = \frac{1}{2\pi \delta_1 \delta_2} \exp\left[-\frac{1}{2} \left(\frac{(v - \mu_1)^2}{\delta_1^2} + \frac{(w - \mu_2)^2}{\delta_2^2} \right)\right] \quad (1)$$

其中, v 代表速度, w 代表最大加速度, u_1 代表速度期望, u_2 代表最大加速度期望, δ_1 代表速度标准差, δ_2 代表最大加速度标准差。

服务器响应两类查询,即区域态势查询和个体追踪查询。

定义 6(态势查询) 态势查询 $Q_{rn}(e, target, x, y)$ 要求返回元路径 e 上以 $target$ 为中心、 x 为时间半径、 y 为

距离半径的运动模式。

定义 7(追踪查询) 追踪查询 $Q_{rk}(oid, t_l)$ 返回移动对象 oid 自 t_l 以来的时间位置序列 $\langle (t_l, p_l), (t_{l+1}, p_{l+1}), \dots, (t_j, p_j), \dots, (t_{l+n}, p_{l+n}) \rangle$,其中 $\langle t_j, p_j \rangle$ 表示在 t_j 时刻移动对象在 p_j 点处。

3.2 运动模式挖掘和时空网格索引的构建

由于环境和交通规则的限制,不同时空单元格会呈现一个或若干个有代表性的运动模式。本文提出采用离线的聚类算法,发现每个时空单元格代表性的运动模式。在聚类前,首先对轨迹数据进行预处理,归并属于同一个单元格的轨迹段,然后提取轨迹在各个时空单元格的速度和最大加速度。

经过预处理,每个单元格对应一个二维点 (v_i, w_i) 的集合, v_i 和 w_i 分别代表经过该单元格的对象在该单元格内的平均速度和最大加速度。为有效识别各运动模式中有代表性的平均速度和最大加速度分布,本文提出采用概率模型聚类来实现模式挖掘。设有 K 个类簇 $\{C_1, \dots, C_j, \dots, C_K\}$,每个类簇 C_j 用一个二维高斯分布 $N(u_1^j, u_2^j, \delta_1^j, \delta_2^j)$ 来代表,采用期望最大化(Expectation Maximization, EM)算法实现聚类,如算法 1 所示。

算法 1 运动模式挖掘

输入:训练数据集 $TD = \{o_i = (v_i, w_i) \mid 1 \leq i \leq n\}$,类簇数目 K ,平均速度期望变化阈值 θ_1 ,最大加速度期望变化阈值 θ_2 ,平均速度标准差变化阈值 ω_1 ,最大加速度标准差变化阈值 ω_2

输出: K 个类簇的高斯分布参数 $(u_1^j, u_2^j, \delta_1^j, \delta_2^j) (1 \leq j \leq K)$

对于每个类簇 C_j ,随机化参数 $(u_1^j, u_2^j, \delta_1^j, \delta_2^j)$;

$u_1^j, u_2^j, \delta_1^j, \delta_2^j \leftarrow u_1^j, u_2^j, \delta_1^j, \delta_2^j (1 \leq j \leq K)$;

$deta_{u_1} \leftarrow 1, deta_{u_2} \leftarrow 1, deta_{cov_1} \leftarrow 1, deta_{cov_2} \leftarrow 1$;

while $(deta_{u_1} > \theta_1 \parallel deta_{u_2} > \theta_2 \parallel deta_{cov_1} > \omega_1 \parallel deta_{cov_2} > \omega_2)$

{

/* 计算各对象属于不同类簇的概率 */

for (each o_i in TD)

{

计算 o_i 属于 C_j 的概率 $P(C_j | o_i, C) = \frac{p(o_i | C_j)}{\sum_{l=1}^K P(o_i | C_l)}$;

}

/* 调整参数 */

for (each cluster C_j)

{

计算 $(u_1^j, u_2^j) = \frac{1}{k} \frac{\sum_{i=1}^n (v_i, w_i) P(C_j | o_i, C)}{\sum_{i=1}^n P(C_j | o_i, C)}$;

$\delta_1^j = \sqrt{\frac{\sum_{i=1}^n P(C_j | o_i, C) (o_i \cdot v_i - u_1^j)^2}{\sum_{i=1}^n P(C_j | o_i, C)}}$;

$\delta_2^j = \sqrt{\frac{\sum_{i=1}^n P(C_j | o_i, C) (o_i \cdot w_i - u_2^j)^2}{\sum_{i=1}^n P(C_j | o_i, C)}}$;

}

$deta_{u_1} \leftarrow \max_{1 \leq j \leq K} |u_1^j - u_1^j|$;

$deta_{u_2} \leftarrow \max_{1 \leq j \leq K} |u_2^j - u_2^j|$;

$deta_{cov_1} \leftarrow \max_{1 \leq j \leq K} |\delta_1^j - \delta_1^j|$;

```

detacovt ← max1≤j≤K |δ2j - δ1j| ;
u1j, u2j, δ1j, δ2j ← u1i, u2i, δ1i, δ2i (1≤j≤K);
}
return (u1j, u2j, δ1j, δ2j) (1≤j≤K);

```

算法1的时间复杂度取决于变化阈值和训练集中的样本数目,用 $\Delta\theta$ 表示阈值变化,则时间复杂度为 $O(\Delta\theta |TD|)$ 。

服务器离线挖掘运动模式后,构建时空运动索引(Spatial-temporal Movement Index, SMI)。时空运动索引是一个双层索引:上层是R树索引,下层是元路径网格索引。上层中每个叶结点对应元路径集合;下层中每个元路径对应若干时空单元格,每个时空单元格 $e.cl(i, j)$ 对应一个二元组 (S_{mp}, S_{rp}) ,其中 $S_{mp} = \{mp_i, P(mp_i) | 1 \leq i \leq K\}$ 存储了各个运动模式及其概率, S_{rp} 是时空单元格的汇报轨迹点集合。

4 基于锚点的移动对象位置汇报

移动对象在时空网格中运动,每个时空单元格的中心即为时空锚点,记为 (ct, cs) ,其中 ct 是期望汇报时间, cs 是期望汇报位置。为了保证整个路网上汇报的均衡分布,每个进入锚点邻域的移动对象依照 $N(ct, cs, \delta t, \delta s)$ 高斯分布概率进行汇报,其密度函数记为 $f(t, s) = \frac{1}{2\pi\delta t\delta s} \exp[-\frac{1}{2}(\frac{(t-ct)^2}{\delta t^2} + \frac{(s-cs)^2}{\delta s^2})]$,以保证汇报的时空点依照概率收敛于锚点。标准差参数 $(\delta t, \delta s)$ 直接决定锚点附近汇报的集中程度,其设置方法如下:设 K_f 代表单元格的饱和流量, V_f 代表车辆在该单元格的自由行驶速度,依据交通流密度-速度关系^[23],根据车辆的当前速度 V 估算单元格内的移动对象的数目为:

$$K = K_f(1 - \frac{V}{V_f}) \quad (2)$$

不失一般性,设单元格内移动对象均匀分布,如果期望一个锚点邻域采样 γ 个汇报,则 γ 个采样应在 $(\frac{a}{\sqrt{K/\gamma}}, \frac{b}{\sqrt{K/\gamma}})$ 的邻域内完成,每个进入邻域 $(\frac{a}{\sqrt{K/\gamma}}, \frac{b}{\sqrt{K/\gamma}})$ 的移动对象依照二维高斯分布概率在锚点附近进行汇报。根据3-sigma准则,调整标准差参数 δt 和 δs ,使得 $(\frac{a}{\sqrt{K/\gamma}}, \frac{b}{\sqrt{K/\gamma}})$ 邻域覆盖3-sigma范围,从而保证汇报的时空点趋向锚点,以近似于1的概率在时空邻域内汇报。整个汇报过程如算法2所示。

算法2 运动汇报

输入:元路径 e ,单元格间隔参数 a 和 b ,单元格的饱和流量 K_f ,自由行驶速度 V_f ,单元格的期望采样数目 γ ,基准时间点 t_{std} ,高斯分布模拟取样间隔 λ

输出:汇报轨迹

/* 根据单元格参数 a 和 b 计算元路径 e 对应的时空网格 */

grid(e) ← computeGrid(e, a, b, t_{std});

/* 以下是汇报过程 */

while(移动对象在元路径 e 上运动)

```

{
  t ← 当前时间;
  s ← 当前位置;
  if ((t, s) 落入一个新的时空单元格 e.cl(i, j))

```

```

{
  K ← 根据式(2)计算单元格移动对象数目;
  (δt, δs) ← 根据 K, γ, a, b 计算标准差参数;
  (ct, cs) ← 计算 e.cl(i, j) 的时空锚点;
  while(对象在邻域半径范围 ∧ (t mod λ) = 0)
  { x ← 当前时间;
    y ← 相对锚点的路网距离;
    z = rand(0, 1) * rand(0, 1);
    prob = 1 / (2πδtδs) * exp[-1/2 * ((t-ct)2 / δt2 + (s-cs)2 / δs2)];
    if (z < prob) {
      对象向服务器汇报轨迹点 (pi, vi, avi, ti);
      break; }
    }
  }
}

```

5 服务器端的查询响应

5.1 在线的群体态势探测

为了响应群体态势查询 $Q_{rn}(e, target, r, t)$,服务器分两步处理数据:1)根据查询中心 $target$ 和半径 (r, t) ,确定相交的时空单元格集合;2)根据时空单元格内的汇报轨迹点,计算目标运动模式集合,进一步基于卡尔曼滤波返回聚合模式来表征群体运动,具体步骤如下:

在目标时空单元格 $e.cl(i, j)$ 中,服务器从汇报轨迹点 $\{mv_0, \dots, mv_j, \dots, mv_m\}$ 中提取速度和加速度分量,构建观测输入 $X = \{o_0, \dots, o_j, \dots, o_{m-1}\}$ 。然后,依据贝叶斯定理,采用极大似然估计确定可能的目标运动模式。设时空单元格 $e.cl(i, j)$ 有 K 个可能的运动模式 $MP = \{mp_0, \dots, mp_{K-1}\}$,给定汇报观测 $X = \{o_0, \dots, o_j, \dots, o_{m-1}\}$,在模式独立的假设下,计算不同模式组合出现 X 的概率,记为:

$$P(o_0, \dots, o_j, \dots, o_{m-1} | mp_0, \dots, mp_{K-1}) = \prod_{j=0}^{m-1} \frac{1}{2\pi \delta_1^j \delta_2^j} \exp[-\frac{1}{2}(\frac{(o_j \cdot v - \mu_1^j)^2}{(\delta_1^j)^2} + \frac{(o_j \cdot w - \mu_2^j)^2}{(\delta_2^j)^2})]$$

若 X 属于目标模式组合 $MET = \langle mp(o_0), \dots, mp(o_{m-1}) \rangle$, $mp \in MP$,根据贝叶斯定理,当且仅当:

$$P(X | MET)P(MET) \geq P(X | MP_l)P(MP_l)$$

其中, $P(MP_l) = \prod_{i=0}^{m-1} P(mp(o_i))$ 。时空单元格内的运动模式识别如算法3所示。

算法3 计算目标运动模式

输入:单元格 $e.cl(i, j)$ 内各个运动模式参数 $\{mp_j(v, w) = \frac{1}{2\pi \delta_1^j \delta_2^j}$

$$\exp[-\frac{1}{2}(\frac{(v - \mu_1^j)^2}{(\delta_1^j)^2} + \frac{(w - \mu_2^j)^2}{(\delta_2^j)^2})] | 1 \leq j \leq K\}$$
和对应的先验概率 $\{P(mp_j) | 1 \leq j \leq K\}$, m 个观测 $X = \{o_0, \dots, o_j, \dots, o_{m-1}\}$

输出:目标运动模式

$P_{max} \leftarrow 0$; /* 初始化目标模式概率 */

$MP_{max} \leftarrow \Phi$; /* 初始化目标模式集合 */

for(长度为 m 的不同的模式组合)

```

{
  计算 P(mp0, ..., mpm-1)P(o0, ..., oj, ..., om-1 | mp0, ..., mpm-1); if

```

```

(P(mp0, ..., mpm-1)P(o0, ..., oj, ..., om-1 | mp0, ..., mpm-1) > Pmax)
{
  Pmax ← P(mp0, ..., mpm-1)P(o0, ..., oj, ..., om-1 | mp0, ..., mpm-1);
  MPmax ← {mp0, ..., mpm-1};
}
}

```

消除 MP_{max} 中的重复模式;

return MP_{max};

不妨将计算出的若干个代表模式记为 $\{mp_0, \dots, mp_j, \dots, mp_{n-1}\} (n \leq K)$ 。由于各个模式分布彼此交错, 基于卡尔曼滤波思想, 进一步取多个代表性模式的乘积作为目标模式。由于每个代表性模式是高斯分布, 因此其乘积也是高斯分布, 聚合模式参数 $(u_1^g, u_2^g, \delta_1^g, \delta_2^g)$ 的计算如下:

$$u_1^g = \frac{1}{\sum_{j=1}^n (\delta_1^0)^2, \dots, (\delta_1^{j-1})^2, (\delta_1^{j+1})^2, \dots, (\delta_1^{n-1})^2} * \sum_{j=1}^n [(\delta_1^0)^2, \dots, (\delta_1^{j-1})^2, (\delta_1^{j+1})^2, \dots, (\delta_1^{n-1})^2] \cdot u_1^j \quad (3)$$

$$u_2^g = \frac{1}{\sum_{j=1}^n (\delta_2^0)^2, \dots, (\delta_2^{j-1})^2, (\delta_2^{j+1})^2, \dots, (\delta_2^{n-1})^2} * \sum_{j=1}^n [(\delta_2^0)^2, \dots, (\delta_2^{j-1})^2, (\delta_2^{j+1})^2, \dots, (\delta_2^{n-1})^2] \cdot u_2^j \quad (4)$$

$$\delta_1^g = \frac{\sqrt{\prod_{j=1}^n (\delta_1^{j-1})^2}}{\sqrt{\sum_{j=1}^n (\delta_1^0)^2, \dots, (\delta_1^{j-1})^2, (\delta_1^{j+1})^2, \dots, (\delta_1^{n-1})^2}} \quad (5)$$

$$\delta_2^g = \frac{\sqrt{\prod_{j=1}^n (\delta_2^{j-1})^2}}{\sqrt{\sum_{j=1}^n (\delta_2^0)^2, \dots, (\delta_2^{j-1})^2, (\delta_2^{j+1})^2, \dots, (\delta_2^{n-1})^2}} \quad (6)$$

5.2 追踪个体移动对象的位置

不失一般性, 设移动对象沿同一道路运行, 如果发生道路变更, 移动对象会触发汇报。给定移动对象最近汇报的轨迹点, 服务器采取锚点独立策略或锚点序列策略推断移动对象当前或未来的位置: 前者根据经过的单元格的汇报轨迹点推断移动对象在该单元格的运动, 不考虑单元格间运动的依赖关系; 后者根据经过的锚点序列来推断移动对象最可能的运动模式序列, 进而推断移动对象各个时刻的位置。

5.2.1 锚点独立推测策略

服务器根据移动对象经过的各个锚点的运动模式, 依次估算移动对象经过各个时空单元格的运动, 从而推断移动对象的完整运动过程。给定一个移动对象 ob_i , 以及最后一次汇报的轨迹点 $(p_i, v_i, a v_i, t_i)$, 采用算法 4 来追踪其运动。

算法 4 锚点独立推测

输入: 移动对象 ob 最后一次汇报的轨迹点 $start(ob) = (p_i, v_i, a v_i, t_i)$, 时空运动索引 SMI

输出: 移动对象从 t_i 到现在的运动

```

trret ← start(ob);
clnext ← 在 SMI 中查找 start(ob) 对应的时空单元格;
while (存在 clnext)
{
  (MP, RP) ← 在 SMI 中检索 clnext 对应的运动模式集合和汇报点集合;
  Cset ← detectMovementPattern(MP, RP);
  Ci ← 计算 Cset 的聚合模式;
  mvnext ← 根据 Ci 的期望速度计算到达 clnext 边界的轨迹点;

```

```

trret ← trret ∪ mvnext;
clnext ← 根据 mvnext 在 SMI 中计算下一时空单元格;
}
trret ← 提取 trret 中各轨迹点的位置和时刻分量;
return trret;
```

算法 4 连续推测下一个单元格的轨迹点, 其时间代价主要取决于访问磁盘索引 SMI 的次数。设 SMI 索引中 R 树的高度是 H^R , 在最坏的情况下, 对应一个单元格要访问磁盘 $H^R + 1$ 次。由于移动对象的最大速度为常数, 因此经过的单元格数目正比于时长 ΔT 。综上所述, 算法 4 的时间复杂度为 $O(\Delta T * H^R)$ 。

5.2.2 锚点序列推测策略

鉴于移动对象前一时空单元格的运动会制约或影响下一时空单元格的运动, 本文提出采用隐马尔可夫过程建模移动对象在各个时空单元格的运动。移动对象自最近一次汇报起, 在元路径上会经过一系列锚点, 根据各锚点的汇报轨迹点提取观测序列, 根据观测序列求解最可能的运动模式序列。

隐马尔可夫模型是一个五元组 (N, O, π, A, B) , 其中 N 是运动模式集合, 每个运动模式用一个二维高斯分布来表征; O 是观测状态集合, 每个元素是一个速度和加速度二元组 (v, av) ; π 是初始状态概率矩阵, 表征元路径 e 上各个运动模式的初始概率分布; A 是转移概率矩阵, 元素 $a_{ij} \in A$ 代表从运动模式 mp_i 向运动模式 mp_j 转变的概率; B 是观测分布, 元素 $b_{ij} \in B$ 表示在运动模式 mp_i 时观测到轨迹点 o_j 的概率。

为识别元路径上的运动模式, 在所有元路径上对各个时空单元格的运动模式进行聚类预处理, 然后根据式 (3) 一式 (6) 将所有运动模式合并成一个运动模式。针对不同元路径, 采用 Baum-Welch 算法离线学习 HMM 参数^[26], 并将其存储在服务器上。移动对象运动时, 服务器根据移动对象最后汇报的轨迹点及离线学习的运动模式参数, 采用 Viterbi 算法发现最可能的运动模式序列。由于每个时空单元格对应若干个汇报点, 根据对汇报点的不同利用方式, 服务器采用两种方法进行运动追踪。

(1) 基于虚拟观测点的运动追踪

该策略综合各个汇报观测点的权重计算一个虚拟观测点, 以虚拟观测点为对应时空单元格的观测。设一个锚点附近有 m 个汇报轨迹点 $\{mv_0, \dots, mv_j, \dots, mv_{m-1}\}$, 每个轨迹点 mv_i 的权重反比于它和锚点的距离。设轨迹点 $mv_i = (p_i, v_i, av_i, t_i)$ 与锚点 (ct, st) 的时空相似性定义为:

$$ts_i = 1 / \sqrt{dis_t^2 + dis_s^2} \quad (7)$$

其中, dis_t 和 dis_s 分别代表轨迹点和锚点的归一化的时间距离和路网距离。给定 m 个轨迹点, 设其与锚点的时空相似性分别为 $(ts_0, \dots, ts_i, \dots, ts_{m-1})$, 则 mv_i 的权重为 $w_i = \frac{ts_i}{\sum_{j=0}^{m-1} ts_j}$ 。

进一步, 根据权重计算虚拟观测点 $o = \langle v, av \rangle$, 其中 $v = \sum_{i=0}^{m-1} w_i \cdot v_i$, $av = \sum_{i=0}^{m-1} w_i \cdot av_i$ 。移动对象的运动追踪如算法 5 所示。

算法 5 基于虚拟观测点的追踪

输入: 移动对象 ob 最后一次汇报的轨迹点 $start(ob) = (p_i, v_i, a v_i, t_i)$, 隐马尔可夫模型 $HMM = (N, O, \pi, A, B)$, 时空运动索引

SMI,道路上的最大速度阈值 V_{\max}

输出:从 t_1 开始的后续运动

```
ret ← Φ;
next ← start(ob);
v ← max(Vmax, v1); /* 取最大可能速度 */
⟨cl1, ..., cli, ..., cln⟩, ⟨o1, ..., oi, ..., on⟩ ← 根据 v 在 SMI 中计算最长的
时空单元格序列和对应的虚拟观测点序列;
⟨mp1, ..., mpi, ..., mpn⟩ ← Viterbi(⟨o1, ..., oi, ..., on⟩, HMM);
foreach(mpi)
{
    next ← 根据 mpi 计算向前的单元格边界轨迹点;
    ret ∪ next;
}
ret ← 提取 ret 中各轨迹点的位置和时刻分量;
return ret;
```

(2) 基于观测路径的运动追踪

定义 8 (观测路径) 给定候选的时空单元格序列

$\langle cl_1, \dots, cl_i, \dots, cl_n \rangle$, 各时空单元格对应的汇报观测集合为 $\langle O_1, \dots, O_j, \dots, O_n \rangle$, 观测路径是一个观测点序列 $\langle o_1, \dots, o_j, \dots, o_n \rangle$ ($o_j \in O_j$), 其中 $o_j = (v_j, av_j)$ 是一个速度和加速度二元组。

本文提出选取概率最大的观测路径对应的运动模式序列作为目标运动模式序列, 即:

$$\langle mp_1, \dots, mp_n \rangle = \underset{o_j \in O_j}{\operatorname{argmax}} P(o_1, \dots, o_j, \dots, o_n) \quad (8)$$

进一步, 根据该运动模式序列来推断移动对象的运动, 整个过程如算法 6 所示。

算法 6 基于观测路径的追踪

输入: 移动对象 ob 最后一次汇报的轨迹点 $\text{start}(ob) = (p_1, v_1, av_1, t_1)$, 隐马尔可夫模型 $\text{HMM} = (N, O, \pi, A, B)$, 时空运动索引 SMI

输出: 移动对象 o_j 从 t_1 的后续运动状态

```
T ← Φ;
next ← start(ob);
v ← max(Vmax, v1); /* 取最大速度 */
⟨cl1, ..., cli, ..., cln⟩ ← 根据 v 在 SMI 中计算最长的时空单元格序列;
⟨O1, ..., Oi, ..., On⟩ ← 在 SMI 中检索单元格序列对应的观测集合;
Paths ← 计算 ⟨O1, ..., Oi, ..., On⟩ 中所有的观测路径;
for (each 观测路径 ⟨o1, ..., oi, ..., on⟩ ∈ Paths)
{
    ⟨mp1, ..., mpn⟩, P(⟨mp1, ..., mpn⟩) ← 对 ⟨o1, ..., oi, ..., on⟩ 调用
    Viterbi 算法计算最可能的模式序列和概率;
    T ← T ∪ ⟨mp1, ..., mpn⟩, P(⟨mp1, ..., mpn⟩);
}
⟨mp1t, ..., mpit, ..., mpnt⟩ ← 从 T 中选取概率最大的运动模式序列;
foreach(mpi)
{
    next ← 根据 mpit 计算对象的下一个边界轨迹点;
    ret ← ret ∪ next;
}
ret ← 提取 ret 中各轨迹点的位置和时刻分量;
return ret;
```

算法 5 和算法 6 中, 为了推测移动对象在各单元格的轨

迹点, 需要访问 SMI 索引获取各个单元格的信息, 然后在内存中进行计算, 因此磁盘访问是主要的时间代价。设 SMI 索引中 R 树的高度是 H^R , 在最坏的情况下, 对应一个单元格要访问磁盘 $H^R + 1$ 次。设单元格序列的最大长度为 N , 则算法 5 和算法 6 的时间复杂度为 $O(N * H^R)$ 。

6 实验分析

实验采用 Windows 10 平台, CPU 为 3.4 GHz, 内存为 8 GB。实验在模拟数据和实际数据集上进行, 从移动端汇报代价、服务器态势查询误差、服务器追踪查询误差和相关工作比较 4 个方面进行分析。

从 OpenStreetMap 网站下载北京地图^[27], 相关统计数据如表 1 所列。模拟移动对象参照 Thomas Brinkhoff 的移动对象生成器^[28]实现。

表 1 北京路网的统计数据

network parameter	Beijing
经度范围	116.0843~116.7119
纬度范围	39.6850~40.1757
道路数目	157955
节点数目	8
路段的平均长度	470

实际数据集采用北京出租车的轨迹数据^[29-30], 该数据集包含北京 2008 年 2 月某个星期的出租车轨迹, 平均取样间隔为 177 s, 平均取样距离为 623 m, 采用地图匹配算法将轨迹匹配到路网^[31]。由于有些数据点间的时间和距离较长, 无法细粒度地反映车辆的运动, 因此采取如下插值方法: 1) 根据 A^* 算法计算两个点在路网上的最短距离和经过路段; 2) 根据时间和距离计算平均速度; 3) 根据所有经过该路段的轨迹点, 计算速度的标准差参数; 4) 以平均速度和标准差为参数, 用高斯分布模拟其速度分布, 每隔 4 s 取样一次, 按照取样速度运行。这样既保持了运行的概貌, 又逼近了真实运动。在挖掘运动模式时, 采用肘方法来确定聚簇的个数^[32]; 设簇的数目为 K , 随着 K 的增大, 各个簇的簇内平均方差会越来越小, 寻找第一个拐点, 即簇的数目增大但方差没有明显降低的位置对应的 K 作为簇的数目。

移动端汇报采用两个指标来衡量: 1) 移动对象单位时空的平均汇报次数 (Average Number of Reports, ANR), 用于反映每个移动对象在单位时空区域内的通信代价; 2) 单元时空汇报次数标准差 (Spatial-temporal Variance, SVAR), 用于反映所有单元格 (单元格的时间长度为 a , 路网距离为 b) 汇报分布的倾斜程度。

给定 W 个时空单元格, 每个时空单元格经过 m_i 个移动对象, 汇报 r_i 次, 则:

$$\text{ANR} = \frac{1}{W} \sum_{i=1}^W \frac{r_i}{m_i * a * b} \quad (9)$$

可知每个时空单元格的平均汇报次数为 $\frac{r_i}{m_i}$, 则汇报次数的标准差为:

$$\text{SVAR} = \sqrt{\frac{\sum_{i=1}^W (a * b * \text{ANR} - \frac{r_i}{m_i})^2}{W}} \quad (10)$$

给定时空范围,实际车辆的平均速度为 V_1 ,服务器探测的速度为 V_2 ,则态势查询误差(Error of Group Movement, EGM)定义为:

$$EGM = \frac{|V_1 - V_2|}{|V_1|} \quad (11)$$

给定一个移动对象,其追踪误差通过比较各个时刻点的实际位置和计算位置的差异来衡量。设均匀选取若干个时间点 $\langle t_0, \dots, t_i, \dots, t_{m-1} \rangle$,其在路网上的实际位置是 $\langle p_0^a, \dots, p_i^a, \dots, p_{m-1}^a \rangle$,计算位置为 $\langle p_0^c, \dots, p_i^c, \dots, p_{m-1}^c \rangle$,则追踪查询误差(Error of Tracking, ET)定义为:

$$ET = \frac{\sum_{i=0}^{m-1} ndist(p_i^a - p_i^c)}{t_{m-1} - t_0} \quad (12)$$

6.1 影响性能因素分析

本节分析网格粒度大小(a 和 b 参数)、单元格汇报次数 γ 、移动对象密度 ω 因素对汇报代价和效果的影响。除特别说明,网格粒度大小参数 $a = 40$ s, $b = 400$ m, 单元格汇报次数 $\gamma = 5$, 移动对象密度 ω 为 20 个每单元格。

6.1.1 移动对象汇报效果的影响因素

(1) 网格粒度大小

固定其他参数,时空单元格大小参数 a 和 b 从小到大同步扩大:时间间隔 a 依次从 20 s 扩大到 180 s,每次增加 20 s;相应地,路网间隔依次从 200 m 扩大到 1 800 m,每次增加 200 m。图 1、图 2 分别显示了 ANR 和 SVAR 指标的变化情况。可以看出,随着时空单元格变大,由于每个单元格的汇报次数固定,因此 ANR 指标逐步降低,SVAR 指标基本平稳。

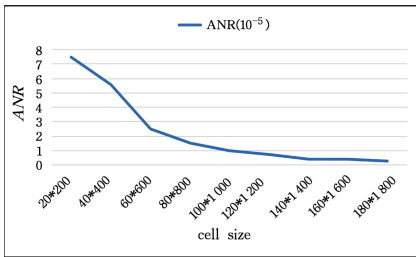


图 1 网格粒度对 AVR 指标的影响

Fig. 1 Effect of how grid granularity affects AVR

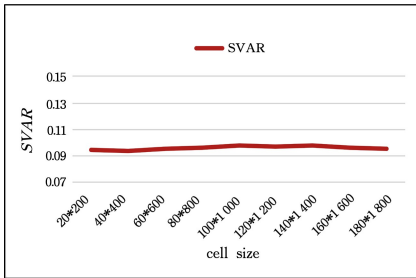


图 2 网格粒度对 SVAR 指标的影响

Fig. 2 Effect of how grid granularity affects SVAR

(2) 单元格汇报次数

固定其他参数,每个单元格汇报次数 γ 从 2 依次增大到 16,每次增加 2。图 3、图 4 分别显示了 ANR 和 SVAR 指标的变化情况。可以看出,随着汇报次数的增多,ANR 指标线性增加,SVAR 指标缓慢增加。

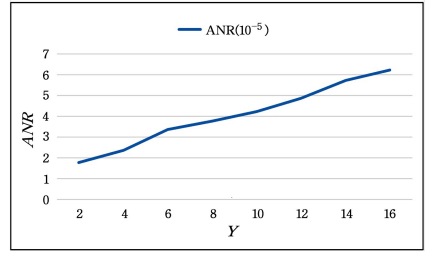


图 3 单元格汇报次数对 ANR 指标的影响

Fig. 3 Effect of how number of reports in each grid cell affects ANR

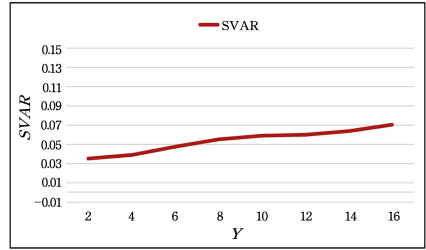


图 4 单元格汇报次数对 SVAR 指标的影响

Fig. 4 Effect of how number of reports in each grid cell affects SVAR

(3) 移动对象密度

固定其他参数,每个单元格的移动对象数目 ω 从 10 依次增大到 90,每次增加 10,图 5、图 6 分别给出了 ANR 和 SVAR 指标的变化情况。可以看出,ANR 指标线性降低,这是由于随着密度增大,分摊到每个移动对象的汇报代价降低,而 SVAR 指标先略微降低后逐渐平稳。

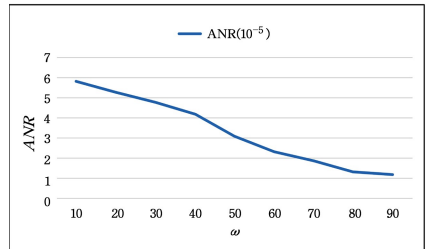


图 5 移动对象密度对 ANR 指标的影响

Fig. 5 Effect of how moving object density affects ANR

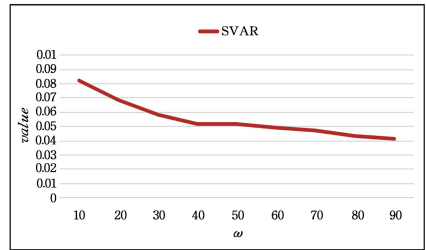


图 6 移动对象密度对 SVAR 指标的影响

Fig. 6 Effect of how moving object density affects SVAR

6.1.2 服务器端群体态势查询性能的影响因素

(1) 网格粒度大小

固定其他参数,网格粒度同步扩大:时间间隔 a 依次从 20 s 扩大到 180 s,每次增加 20 s;相应地,路网间隔距离依次从 200 m 扩大到 1 800 m,每次增加 200 m。图 7 显示了 EGM

指标的变化情况。可以看出,随着网格粒度变大,EGM 指标最初基本保持稳定,但当网格粒度增大到一定程度时,EGM 指标开始升高,这主要是因为此时固定时空内的汇报次数变少,不足以反映运动状态,导致探测误差变大。

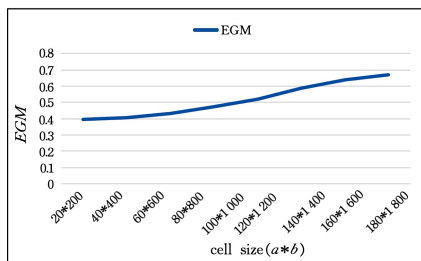


图7 网格粒度对 EGM 指标的影响

Fig.7 Effect of how grid granularity affects EGM

(2) 单元格汇报次数

固定其他参数,将每个单元格汇报次数从 2 依次增大到 16,每次增加 2。从图 8 可以看出,随着单元格汇报次数的增多,EGM 指标会稳步降低,这是因为服务器对群体会有更准确的把握。同时可以看出,当单元格内汇报次数增加到一定程度时,误差下降幅度会越来越小,这是因为汇报提供的增量信息会越来越少。

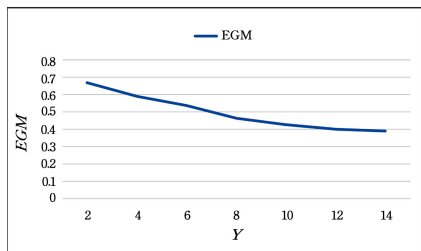


图8 单元格汇报次数对 EGM 指标的影响

Fig.8 Effect of how number of reports in each grid cell affects EGM

(3) 移动对象密度

网格粒度是 20 * 200,固定其他参数,将单元格的移动对象密度从 10 依次增大到 90,每次增加 10。从图 9 可以看出随着移动对象密度的增大,EGM 指标缓慢增加。

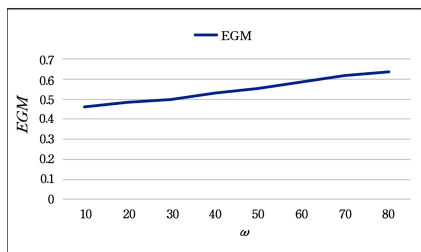


图9 移动对象密度对 EGM 指标的影响

Fig.9 Effect of how moving object density affects EGM

6.1.3 服务器端追踪性能的影响因素

移动对象追踪有 3 种策略,即锚点独立策略(Anchor Independent Policy, AIP)、基于虚拟观测点(Virtual Observation Point, VOP)的锚点序列策略和基于观测路径(Observation Path, OP)的锚点序列策略。下面从网格粒度、单元格汇报次数、移动对象密度 3 个方面进行分析。

(1) 网格粒度的影响

固定其他参数,时空单元格大小的参数 a 和 b 从小到大同步扩大:时间间隔 a 依次从 20 s 扩大到 180 s,每次增加 20 s;相应地,路网间隔依次从 200 m 扩大到 1800 m,每次增加 200 m。图 10 显示了 3 种策略的 ET 指标的变化情况。最初,随着网格粒度变大,抽样汇报点可以反映整体运动状况,性能基本平稳,ET 指标略微下降;但当粒度增大到一定程度时,汇报点稀疏,性能开始下降,从而 ET 指标增大。同时锚点序列策略优于锚点独立策略,但锚点序列策略在网格粒度超过一定阈值后,性能开始下降,这是因为各个锚点间运动的依赖性逐步变小。

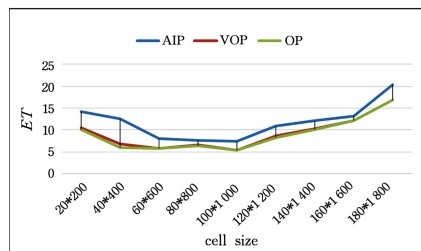


图10 网格粒度对 ET 指标的影响

Fig.10 Effect of how grid granularity affects ET measure

(2) 单元格汇报次数

固定其他参数,将每个单元格汇报次数 γ 依次从 2 增大到 16,每次增加 2。图 11 显示了单元格汇报次数对 ET 指标的影响。可以看出,随着单元格汇报次数的增多,最初误差快速减小,但当次数超过一定阈值时,ET 指标下降的幅度越来越小。

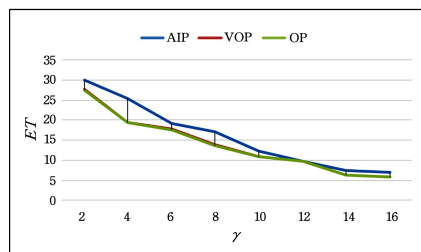


图11 单元格汇报次数对 ET 指标的影响

Fig.11 Effect of how number of reports in grid cell affects ET

(3) 移动对象密度

图 12 显示了当网格密度为 20 * 200 时,固定其他参数,每个单元格的移动对象数目从 10 调整到 90,每次增加 10 时 ET 指标的变化情况。可以看出,随着移动对象密度的增大,ET 指标逐步增加。

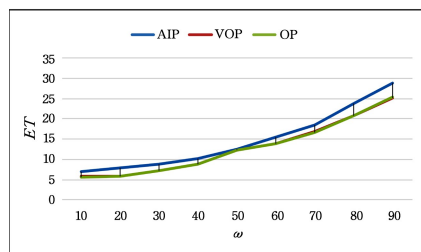


图12 移动对象密度对 ET 指标的影响

Fig.12 Effect of how moving object density affects ET measure

6.2 相关工作比较

对本文的 DMDSA 方法和 3 种策略进行比较, 3 种策略分别为基于路段的追踪策略 (Segment-Based Tracking, SBT)^[4]、基于安全区域追踪策略 (Safe-Region Tracking, SRT)^[6] 和移动机会众包策略 (Opportunistic Mobile Crowdsensing, OMC)^[11]。

(1) 基于路段追踪策略。在每个路段上, 结合挖掘的加速度画像, 服务器按照移动对象最近汇报的运动矢量推断其位置。如果移动端发现自己的实际位置和上次汇报的位置的偏差超过阈值, 则触发一次汇报。

(2) 安全区域追踪策略。移动对象采用时空安全区域 (Spatial-Temporal Safe Region, STSR) 表征可能的位置和速度, 避免频繁的位置更新, 即根据空间安全区域 (spatial safe region) 和由最大和最小速度构成的速度安全区域 (velocity safe region), 可以计算出未来一系列时刻的矩形预测区域 (Predicted Regions)。移动对象运动时, 如果其位置超出预测区域或者根据实时位置和速度预测出其在最近一个时间窗口将超出预测区域, 则向服务器汇报位置更新。目前的位置更新策略分成主动更新和被动更新两种。为了实现公平比较, 实验中只考虑在路网中运行的情况, 因此空间区域修正为沿道路运行的多线段。

(3) 移动机会众包策略。整个区域划分成 N 个不同区域, 在每个时间窗口中, 众包参与者上传一次汇报数据。服务器将时间划分成若干相互重叠的时间窗口, 每个时间窗口 $T_i = \{\Delta t_{i-w+1}, \dots, \Delta t_i\}$ 的长度为 w , 即不同时间窗口有 $w-1$ 个重叠时间单位。每个时间窗口 T_i 的满意指数计算如下: $SI_i = m_i / M_0$, 其中 m_i 表示在 T_i 时间窗口内收到的汇报次数, M_0 是预期收到的汇报次数。在每个时间窗口, 服务器会广播 SI_i 指标, 各个汇报者将根据其进行相应的调整。为实现公平比较, 区域设定为路网区域, M_0 参数设定为 DMDSA 策略的 γ 值。

下面从汇报倾斜程度 SVAR、态势查询误差 EGM 和追踪查询误差 ET 3 个方面进行比较。

(1) 汇报倾斜程度

图 13 显示了在相同的通信代价下, 北京路网的 SVAR 指标的变化情况。可见 DMDSA 方法的汇报倾斜程度最小, SBT 次之, OMC 方法具有最大的汇报倾斜程度。

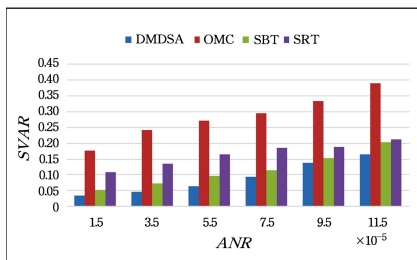


图 13 汇报倾斜程度比较

Fig. 13 Comparison of uneven distribution of reporting

(2) 态势查询误差比较

将通信代价从 1.5 增加到 11.5, 图 14 显示了在北京路网上的态势查询误差的比较。可以看出, SBT 和 SRT 方法的态势查询误差远远高于本文方法, 这是由于前两种方法以个体追踪为目标, 汇报分布不平衡, 导致群体态势估计精度性能

不佳。同时可以看出, DMDSA 方法的效果稳定地优于 OMC 方法, 这是由于 DMDSA 方法能够保证汇报点在整个路网内均匀分布, 避免汇报不能反映全体移动对象运动的情况。

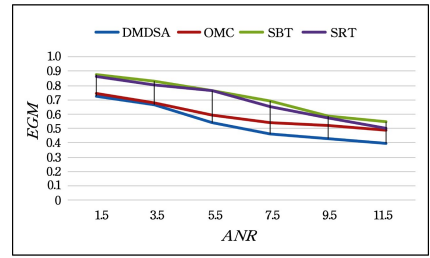


图 14 态势查询误差比较

Fig. 14 Comparison of situational query error

(3) 追踪查询误差比较

将通信代价从 1.5 增加到 11.5, 图 15 显示了在北京路网上的几种方法的追踪查询误差比较。可以看出, 本文方法虽然是以群体为基础的个体追踪, 但由于基于均匀分布的锚点来推测各个移动对象的位置, 避免了取样点倾斜, 能更全面、准确地反映移动对象经过路线的速度状况, 误差远小于 SBT 和 SRT 方法, 效果也稳定地优于 OMC 方法。

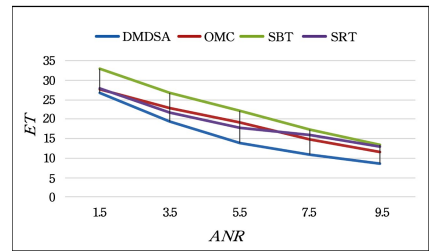


图 15 追踪查询误差比较

Fig. 15 Comparison of tracking query error

结束语 在智慧交通中, 需要对路网上的群体运动态势和个体移动对象的实时位置进行查询, 本文提出了一种同时满足这两类查询的汇报和服务器计算框架。该框架可以显著减小移动对象向服务器进行位置汇报的代价, 不仅能够实时探测区域交通状况, 而且能够准确追踪个体移动对象的位置, 满足不同粒度的交通监控和决策需求。理论分析和实验表明了本文方法的有效性。在未来的工作中, 我们将进一步探究如何结合 ad-hoc 网络来提高整体框架的性能。

参考文献

- [1] GUO C J, YANG B, HU J L, et al. Learning to Route with Sparse Trajectory Sets[C]//Proc of the 34th International Conference on Data Engineering. New York; IEEE, 2018; 1073-1084.
- [2] DING Z M. Data Model, Query Language, and Real-Time Traffic Flow Analysis in Dynamic Transportation Network Based Moving Objects Databases [J]. Journal of Software, 2009, 20 (7): 1866-1884.
- [3] WOLFSON O, XU B, CHAMBERLAIN S, et al. Moving object databases: issues and solutions[C]//Proc of the 10th International Conference on Scientific and Statistical Database Management, 1998: 111-122.
- [4] CHRISTIAN S, JENSEN S P. TRAX—Real-World Tracking of

- Moving Objects[C]// Proc of 33rd International Conference on Very Large Databases. 2007;1362-1365.
- [5] JO J, KIM J, MIN K W, et al. An Arrival Time Anticipation Approach for Real-Time Tracking of Moving Object in Mobile Networks [C]// Proceedings of the Third international conference on Ubiquitous Intelligence and Computing (UIC'06). Berlin: Springer, 2006;166-175.
- [6] CHEN S, OOI B C, ZHANG Z J. An AdaETive Updating Protocol for Reducing Moving Object Database Workload[C]// Proc of 36th International Conference on Very Large Databases. 2010;735-746.
- [7] DING Z M, HAN J Y, LI M, et al. Network Constrained Moving Objects Database Based Traffic Flow Statistical Analysis Model [J]. Journal of Computer Research and Development, 2008, 45(4):646-655.
- [8] BORRESEN J L, JENSEN C S, TORP K. FoGBAT: Combining Bluetooth and GPS Data for Better Traffic Analytics[C]// Proc of 17th IEEE International Conference on Mobile Data Management. 2016;325-328.
- [9] WANG L, CHRISTENSEN R, LI F F, et al. Spatial Online Sampling and Aggregation[C]// Proc of 41st International Conference on Very Large Databases. 2015;84-95.
- [10] LI Y, GUNOPULOS D. Urban Travel Time Prediction using a Small Number of GPS Floating Cars[C]// Proc of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. USA: ACM, 2017.
- [11] MONTORI F, BEDOGNI L, BONONI L. Distributed Data Collection Control in Opportunistic Mobile Crowdsensing[C]// Proc of the 3rd Workshop on Experiences with the Design and Implementation of Smart Objects. 2017;19-24.
- [12] CUI Y L, JING B H, ZHANG F S. Highway traffic condition detection with data fusion[J]. Chinese Journal of Computers, 2017, 40(8):1798-1812.
- [13] CIVILIS A, JENSEN C S, PAKALNIS S. Techniques for efficient road-network-based tracking of moving objects, Knowledge and Data Engineering[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(5):698-712.
- [14] BIAGIONI J, MUSA A B M, ERIKSSON J. Thrifty Tracking: Online GPS Tracking with Low Data Uplink Usage[C]// Proc of SIGSPATIAL'13. USA: ACM, 2013;486-489.
- [15] HAN J Y, CHEN K J, DING Z M, et al. An efficient location reporting and indexing framework for urban road moving objects [J]. Distributed Parallel Databases, 2014, 32(2):271-311.
- [16] PESTI P, LIU L, BAMBA B, et al. ROADTRACK: Scaling Location Updates for Mobile Clients on Road Networks with Query Awareness[C]// Proc of 36th International Conference on Very Large Databases. 2010;1493-1504.
- [17] LIU F Y, HUA K A, XIE F. On Reducing Communication Cost for Distributed Moving Query Monitoring Systems[C]// Proc of the 9th International Conference on Mobile Data Management. New York: IEEE, 2008;156-164.
- [18] HSUEH Y L, ZIMMERMANN R, WANG H J, et al. Partition-Based Lazy Updates for Continuous Queries over Moving Objects[C]// Proc of the 15th International Symposium on Advances in Geographic Information Systems. USA: ACM, 2007;1-8.
- [19] KUMAR D, WU H Y, LU Y, et al. Understanding Urban Mobility via Taxi Trip Clustering[C]// Proc of the 17th IEEE International Conference on Mobile Data Management. USA: IEEE, 2016;318-324.
- [20] TICIANA L, COELHO DA SILVA, JOSÉ A F, et al. Discovering Frequent Mobility Patterns on Moving Object Data[C]// Proc of the Third ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems. USA: ACM, 2014;60-67.
- [21] FAN Q, ZHANG D X, WU H Y, et al. A general and parallel platform for mining co-movement patterns over large-scale trajectories[C]// Proc of 42th International Conference of Very Large Database. 2016;313-324.
- [22] EFTELIOGLU E, SHEKHAR S, KANG, et al. Ring-Shaped Hotspot Detection [J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(12):3367-3381.
- [23] LI K, MCKENNEY M. Convoy Detection Using Sequence Alignment[C]// Proc of the 12th ACM SIGSPATIAL International Workshop on Computational Transportation Science. 2019.
- [24] GONZALEZ H, HAN J W, LI X L, et al. AdaETive fastest path computation on a road network: a traffic mining approach[C]// Proc of 33rd International Conference on Very Large Databases. 2007;321-333.
- [25] AN S, WANG J, XU Y G, et al. Intelligent Transport Management Technology and Application [M]. Beijing: Science Press, 2005;80-81.
- [26] BAUM L E, PETRIE T, SOULES G, et al. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains [J]. Ann. Math. Statist, 1970, 41(1):164-171.
- [27] OpenStreetMap [EB/OL]. <http://www.openstreetmap.org/>.
- [28] BRINKHOFF T. A framework for generating network-based moving objects[J]. GeoInformatica, 2002, 6(2):153-180.
- [29] YUAN J, ZHENG Y, XIE X, et al. Driving with knowledge from the physical world[C]// Proc of the 17th ACM SIGKDD international conference on Knowledge Discovery and Data mining. NY, USA: ACM, 2011;316-324.
- [30] YUAN J, ZHENG Y, ZHANG C Y, et al. T-drive: driving directions based on taxi trajectories [C]// Proc. of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. NY, USA: ACM, 2010;99-108.
- [31] BRAKATSOULAS S, PFOSER D, SALAS R, et al. On Map-Matching Vehicle Tracking Data[C]// Proc of the 31st International Conference on Very Large Databases. 2005;853-864.
- [32] HAN J W, KAMBER M, PEI J. Data mining: concepts and techniques (third edition) [M]. USA: Morgan Kaufmann Publishers, 2012.



HAN Jing-yu, born in 1976, Ph.D, professor, is a member of China Computer Federation. His main research interests include spatial-temporal database, big data and machine learning.