

多模式移动对象不确定性轨迹预测模型

乔少杰¹ 韩楠² 丁治明³ 金澈清⁴ 孙未未⁵ 舒红平⁶

摘要 以移动设备、车辆、飞机、飓风等移动对象不确定性轨迹预测问题为背景, 将大规模移动对象数据作为研究对象, 以频繁轨迹模式挖掘、高斯混合回归技术为主要研究手段, 提出多模式移动对象轨迹预测模型, 关键技术包括: 1) 针对单一运动模式, 提出一种基于频繁轨迹模式树 FTP-tree 的轨迹预测方法, 利用基于密度的热点区域挖掘算法将轨迹点划分成不同的簇, 构建轨迹频繁模式树, 挖掘频繁轨迹模式预测移动对象连续运动位置. 不同数据集上实验结果表明基于 FTP-tree 的轨迹预测算法在保证时间效率的前提下预测准确性明显优于已有预测算法. 2) 针对复杂多模式运动行为, 利用高斯混合回归方法建模, 计算不同运动模式的概率分布, 将轨迹数据划分为不同分量, 利用高斯过程回归预测移动对象最可能运动轨迹. 实验证明, 相比于基于隐马尔科夫模型和卡尔曼滤波的预测方法, 所提方法具有较高的预测准确性和较低的时间代价.

关键词 移动对象数据库, 多模式, 轨迹预测, 频繁轨迹模式

引用格式 乔少杰, 韩楠, 丁治明, 金澈清, 孙未未, 舒红平. 多模式移动对象不确定性轨迹预测模型. 自动化学报, 2018, 44(4): 608–618

DOI 10.16383/j.aas.2017.c160575

A Multiple-motion-pattern Trajectory Prediction Model for Uncertain Moving Objects

QIAO Shao-Jie¹ HAN Nan² DING Zhi-Ming³ JIN Che-Qing⁴ SUN Wei-Wei⁵ SHU Hong-Ping⁶

Abstract This study aims to solve the problem of predicting uncertain trajectories of moving objects, including mobile devices, vehicles, airplanes, and hurricanes. In order to design a general schema of trajectory prediction on large-scale moving objects data, techniques of frequent trajectory patterns mining and Gaussian mixture regression model are employed, and a multiple-motion-pattern trajectory prediction model is proposed. The proposed key techniques include: 1) as for simple motion patterns, a new trajectory prediction algorithm based on frequent trajectory pattern tree (FTP-tree) is proposed, which employs a density based region-of-interest discovery approach to partition a large number of trajectory points into distinct clusters. Then, it generates a frequent trajectory pattern tree to forecast continuous locations of moving objects. Experimental results show that the FTP-tree based trajectory prediction algorithm performs better than existing prediction approaches with the guarantee of time efficiency. 2) Gaussian mixture regression approach is used to model complex multiple motion patterns, which calculates the probability distribution of different types of motion patterns, as well as partitions trajectory data into distinct components, in order to predict the most possible trajectories of moving objects via Gaussian process regression. Experimental results show a high accuracy and low time consumption on trajectory prediction, as compared to the hidden Markov model approach and the Kalman filter one.

Key words Moving objects databases, multiple-motion-pattern, trajectory prediction, frequent trajectory patterns

Citation Qiao Shao-Jie, Han Nan, Ding Zhi-Ming, Jin Che-Qing, Sun Wei-Wei, Shu Hong-Ping. A multiple-motion-pattern trajectory prediction model for uncertain moving objects. *Acta Automatica Sinica*, 2018, 44(4): 608–618

收稿日期 2016-08-04 录用日期 2016-12-10

Manuscript received August 4, 2016; accepted December 10, 2016

国家高技术研究发展计划 (863 计划) (2014BAI06B01), 国家自然科学基金 (61772091, 61100045, 91546111, 61501063, 61501064, 61772138), 教育部人文社会科学研究规划基金 (15YJAZH058), 教育部人文社会科学研究青年基金 (14YJCZH046), 四川省教育厅资助科研项目 (14ZB0458), 成都市软科学项目 (2015-RK00-00059-ZF), 四川高校科研创新团队建设计划资助 (18TD0027), 成都信息工程大学中青年学术带头人科研基金 (J201701), 成都信息工程大学引进人才科研启动项目 (KYTZ201715, KYTZ201750) 资助

Supported by National High Technology Research and Development Program of China (863 Program) (2014BAI06B01), National Natural Science Foundation of China (61772091, 61100045, 91546111, 61501063, 61501064, 61772138), Planning Foundation for Humanities and Social Sciences of Ministry of Education of China (15YJAZH058), Youth Foundation for Humanities and Social Sciences of Ministry of Education (14YJCZH046), Foundation of Educational Commission of Sichuan Province (14ZB0458), Soft Science Foundation of Chengdu (2015-RK00-00059-ZF), Innovative Research Team Construction Plan in Universi-

ties of Sichuan Province (18TD0027), Scientific Research Foundation for Young Academic Leaders of Chengdu University of Information Technology (J201701), and Scientific Research Foundation for Advanced Talents of Chengdu University of Information Technology (KYTZ201715, KYTZ201750)

本文责任编辑 黎铭

Recommended by Associate Editor LI Ming

1. 成都信息工程大学网络空间安全学院 成都 610225 2. 成都信息工程大学管理学院 成都 610103 3. 北京工业大学计算机学院 北京 100124 4. 华东师范大学数据科学与工程学院 上海 200062 5. 复旦大学计算机科学技术学院 上海 201203 6. 成都信息工程大学软件工程学院 成都 610225

1. School of Cybersecurity, Chengdu University of Information Technology, Chengdu 610225 2. School of Management, Chengdu University of Information Technology, Chengdu 610103 3. College of Computer Science, Beijing University of Technology, Beijing 100124 4. School of Data Science and Engineering, East China Normal University, Shanghai 200062 5. School of Computer Science, Fudan University, Shanghai 201203 6. School of Software Engineering, Chengdu University of Information Technology, Chengdu 610225

随着无线通信和定位技术的高速发展, 人们对持续移动物体所处的空间位置的跟踪能力不断加强, 推动相关应用领域的研究不断深入, 如智能交通控制、智能导航和军事指挥等. 此类应用系统往往基于移动对象数据库构建, 其中存储了大量关于移动对象时空位置及运动状态的信息. 这些数据通常以轨迹的形式存储, 隐藏着关于移动对象大量的行为特征及运动规律. 然而, 移动对象所处的运动环境往往是动态变化的, 不能单纯依赖静态交通网络环境预测其运动行为, 需要综合考虑移动速度和方向的动态变化对对象运动行为的影响. 如何综合考虑主客观因素, 高效准确地查询和预测移动对象不确定性位置成为当前移动数据管理领域的研究热点^[1]. 尤其值得注意的是, 在包含较为复杂的轨迹模式场景下, 典型运动模式不止一个, 一条轨迹可能隶属于多个轨迹模式, 称之为多模式. 不同类型移动对象运动行为各有差异, 相同类型的移动对象由于主客观因素影响, 运动模式也不尽相同, 因此挖掘多模式移动对象运动行为特征更具实际意义.

多模式移动对象位置预测是一项非常困难和富有挑战意义的课题. 1) 由于移动对象运动模式复杂多变, 且通过定位系统获得的数据流信息量大, 具有不确定性, 需要更加稳定和具有可伸缩性的挖掘方法进行处理; 2) 位置预测需要尽可能快地对可能运动轨迹进行评价, 延时或滞后将产生无意义的预测结果; 3) 基于模拟技术的预测方法依赖于大量输入参数, 这些参数的设置极大地影响模型的准确性, 对于不确定性的轨迹数据进行预测, 需要考虑诸多客观因素和领域专家知识; 4) 如果算法设计不合理, 随着移动对象数目的增加, 模型的计算代价可能呈指数级增长. 下面以移动对象轨迹预测为主要背景, 以智能交通系统为例说明问题意义和概貌.

智能交通旨在建立一种在较大范围内, 全方位发挥作用的实时、准确和高效的综合运输和管理系统. 系统将采集到的各种道路交通及服务信息处理后, 传输到运输系统的各种用户 (例如驾驶员、居民、警察), 出行者可实时选择交通方式和交通路线, 交通管理部门可进行自动和合理的交通疏导、控制及事故处理. 决策者可知每天哪一时段, 具体在哪一区域是机动车出行的高峰期, 应该采用何种手段可以使路网上的交通流量处于最佳状态, 改善交通拥挤和阻塞状况, 最大限度地提高交通的通行能力, 提高整个公路运输系统的机动性、安全性和运输效率. 可以形式化描述为在一组约束下, 确定一组阈值 ε 和 σ , 使得

$$\left\{ \text{Num}(T_i) \mid \sum T_i < \varepsilon, i \in \{\text{vehicles}\} \right\} \rightarrow \{P_j < \sigma \mid j \in [0.0 \text{ am}, 24.0 \text{ pm}]\}, \text{Supp}, \text{Conf}(\ast)$$

上式 (*) 意义为求取当各种车辆总和 $\sum T_i < \varepsilon$,

且保证不同时段 j 内交通事故发生率 $P_j < \sigma$ 时, 各种车辆的数量 $\text{Num}(T_i)$, 并且在测试数据集上支持度超过 Supp , 置信度超过 Conf . 其中, ε 表示交通运输能力, σ 表示事故发生率门限值.

本文结构如下: 第 1 节对国内外在移动对象不确定性轨迹预测领域的成果进行综述; 第 2 节针对单一运动模式轨迹, 挖掘轨迹热点区域并利用频繁模式树准确高效预测移动对象连续位置信息; 第 3 节介绍一个面向复杂运动场景的基于高斯混合回归的多模式轨迹预测模型; 第 4 节验证本文提出的两种针对不同运动模式的轨迹预测算法的性能; 第 5 章总结全文工作并对未来工作进行展望.

1 国内外研究现状

针对移动对象的连续位置预测主要包括过去轨迹的重现及当前和未来轨迹的预测, 主要研究集中于轨迹频繁模式挖掘, 通过挖掘频繁模式找出典型运动路径. Hu 等^[2] 将连续轨迹视为离散状态点之间转变的过渡, 对轨迹离散状态分析的不足之处在于对大量时空数据进行离散化处理代价极高, 而且需要计算离散数据点之间的相关性. Parker 等^[3] 设计了移动对象运动目标概率函数可以满足的若干 axiom, 这些 axiom 用于精确度量预测轨迹与真实轨迹之间的差异. Song 等^[4] 在 *Science* 上发表了预测人类移动性的研究成果, 通过测量个体轨迹的信息熵, 定量地给出了人类动态运动轨迹具有 93% 的可预测性, 并证明了人类有规律的运动路线与距离无关. Centola^[5] 在 *Science* 上的工作是在原始 GPS 数据中使用层次型马尔科夫模型抽取重要地点, 进而检测用户的行为模式. 针对已有轨迹预测算法仅能预测短期内运动路径的不足, Jeung 等^[6] 提出了一种基于路网的移动预测模型, 用于准确计算移动对象在交叉路口运动变换模式和在不同路段上的速度信息, 但该方法局限于路网中应用. Zhou 等^[7] 利用动态选取的参考轨迹构建预测模型, 模型建立之前目标轨迹是已知的, 具备先验知识, 可以基于少量的历史轨迹构建精准的预测模型.

随着越来越多不同种类移动对象轨迹数据被收集, 近期多模式轨迹预测的工作得到学者的广泛关注, Pan 等^[8] 提出了基于多变元正态分布的线性预测器, 应用这一方法预测会产生延迟, 不能应用于交通流的实时监控环境中. Qiao 等^[9] 借助隐马尔科夫模型 (Hidden Markov model, HMM), 设计实现了一种可以自适应调整轨迹预测参数的动态预测算法, 根据不同类型轨迹数据预测最佳路线, 但是这一模型没有考虑大数据环境下算法的运行时间性能问题. 此外, 乔少杰等^[10] 提出了基于隐马尔科夫模型的自适应轨迹预测模型, 该算法通过对大数据环境下移动对象海量轨迹利用基于密度的聚类方法进行位置密度分区和高效分段处理, 根据输入轨迹自动选取

参数组合, 避免隐马尔科夫模型中隐状态不连续、状态停留等问题. Ding 等^[11] 近期提出了一种路网匹配的基于轨迹数据库的交通流分析方法, 用于预测移动对象位置信息. Xu 等^[12] 提出了用于避免大图上过度二次计算的有效策略, 更好地支持动态路网中路径动态规划. 为了支持轨迹大数据中个性化多模式运动路径预测, Dai 等^[13] 利用高斯混合模型, 描述行驶偏好矢量中随机变量的概率分布, 构建带权重的轨迹图, 权重反映对象候选路径的可能性, 并结合最短路径算法推荐个性化运动轨迹. Tripathy 等^[14] 根据不同对象的历史轨迹特征, 构建线性函数, 预测移动对象将来可能运动位置, 所提方法可以保证较小的预测偏差. Xu 等^[15] 提出一种通用位置表达模型, 定义不同对象可能的位置, 实现复杂场景不同运动模式下轨迹预测. 黄玉龙等^[16] 提出了一种改进的高斯近似滤波方法, 用于再入飞行器的目标跟踪. 陈成等^[17] 提出了一种基于四阶贝塞尔曲线的无人车可行轨迹规划算法, 用于在实际道路中行驶的车辆. 考虑到高纬度和不确定性对轨迹预测准确性的影响, Monfort 等^[18] 提出了一种逆最优控制方法, 用于计算人类活动轨迹出现的概率.

通过对上述移动对象轨迹预测工作进行分析, 可以得到如下结论: 1) 如果针对海量轨迹数据挖掘移动对象频繁轨迹模式, 已有算法需要多次扫描数据库, 代价很高, 需要设计新型频繁模式挖掘算法, 提高挖掘的效率和准确性. 2) 现有轨迹预测算法对具有单一简单模式的移动对象的位置跟踪预测效果较好, 对于复杂场景下多模式轨迹预测的研究内容相对较少.

2 基于频繁轨迹模式树的轨迹预测算法

移动对象的运动模式通常比较单一, 例如在城市交通道路上匀速运动. 针对单一运动模式, 本节提出一种基于频繁模式挖掘的轨迹预测算法^[19].

2.1 轨迹热点区域挖掘算法

在预测前改进 DBSCAN 算法对海量 GPS 轨迹点进行聚类挖掘轨迹热点区域, 可以将轨迹点与由信号不稳定等原因产生的噪点区分. 轨迹热点区域挖掘算法 RoIDiscovery 采用密度聚类思想, 通过遍历每个轨迹点的邻域生成簇. 如果轨迹点 p 的邻域内包含多于 θ 个轨迹点, 则创建一个新的簇, 将 p 作为该簇的核心对象. 然后, 递归地遍历核心对象直接密度可达的对象, 这个过程中包含簇的合并操作. 当没有新的点可以被添加到任何簇时, 过程结束. 基于密度的轨迹热点区域挖掘算法如算法 1 所示^[19].

算法 1. 基于密度的轨迹聚类算法

输入. 轨迹数据集 Tr , 聚簇半径 ε , 最少轨迹点数 θ .

输出. 聚类后簇集合 $R = \{R_1, R_2, \dots, R_n\}$.

1. $n = 0$;

```

2. for each  $p \in Tr$  do
3.   mark  $p$  as visited;
4.    $N = getNeighbours(p, \varepsilon)$ ;
5.   if  $size(N) < \theta$  then
6.     mark  $p$  as noise;
7.   else
8.     create a new cluster  $R_k$ ;
9.      $ExpandCluster(p, N, R_k, \varepsilon, \theta)$ .
```

算法 1 的基本思想为: 遍历 Tr 中所有轨迹点, 初始化簇个数 (第 1 行); 将轨迹点 p 标记为已访问 (第 3 行); 第 4~9 行利用 $getNeighbours(\cdot)$ 函数计算轨迹点 p 与其他轨迹点的距离, 将距离小于 ε 的轨迹点存入集合 N 中, 如果 $size(N) < \theta$, 则标记 p 为噪声点, 否则以 p 为核心建立新簇, 并调用 $ExpandCluster(\cdot)$ 函数递归访问 N 中轨迹点. $ExpandCluster(\cdot)$ 基本思想是对 N 邻域内的所有点进行半径检查, 如果大于 θ , 扩展 N 的数目, 将新节点加入扩展后的簇. 当没有新节点可以被添加到任何簇, 该过程结束.

2.2 频繁轨迹模式挖掘算法

本节介绍基于频繁轨迹模式树的轨迹预测算法 FTP-mining^[19], 首先给出频繁轨迹模式树定义.

定义 1 (频繁轨迹模式树 — FTP-tree)^[19]. FTP-tree 由一个标记为空的根节点及一系列由原子路段组成的前缀子树构成, 且包含一个头节点表 Header table. FTP-tree 的数据结构与 FP-tree 类似, 前缀子树上的每一个节点具有三个属性: id , $count$ 和 $pointer$, 其中 id 表示原子路段的标识, $count$ 表示从根节点到达某一节点的路径被访问次数, $pointer$ 表示指向 FTP-tree 中具有相同 id 的下一个节点指针. Header table 中每一行记录包含两个属性: 节点 id 和头指针 (指向 FTP-tree 中具有某 id 的第一个节点). 基于频繁轨迹模式树的轨迹预测算法 FTP-mining 如算法 2 所示^[19].

算法 2. FTP-mining (F, α)

输入. 具有根节点 t 的 FTP-tree F , 与 t 对应的轨迹热点路段 L , 空集合 α .

输出. 频繁轨迹模式集合 $R = \{f_1, f_2, \dots, f_n\}$.

```

1. 构建一个空的频繁轨迹模式集合  $R$ ;
2. if 有一条从  $t$  出发的路径  $Q$  then
3.   构建一条模式  $q = Q \cup \alpha$ ;
4.    $support\_count = \min\{count_i\}$ ; //  $count_i$  表示  $Q$ 
   中第  $i$  个节点被访问的次数;
5.    $Q.add(q)$ ;
6. else
7.   for each 候选项  $l_i \in L$  do
8.     构建模式  $p = l_i \cup \alpha$ ;
9.      $support\_count = \min\{count_i\}$ ;
10.    构建与  $p$  对应的 conditional pattern base 和
    conditional FTP-tree  $F'$ ;
11.    if  $F' \neq null$  then
```

12. FTP-mining(F', p);
13. **if** R 不变化 **then**
14. $R.add(p)$;
15. 输出 R 中频繁轨迹模式.

算法 2 的基本思想为: 首先生成长度为 1 的频繁序列集; 然后根据长度为 1 的频繁序列集生成长度为 2 的频繁序列集; 依此类推, 产生所有频繁序列模式, 与 FP-tree^[20] 类似. FTP-mining 与 FTP-tree 的不同之处在于: 1) FTP-tree 中的项集表示被移动对象访问的路段, 从根节点到叶节点的一条完整的路径表示一条频繁轨迹模式; 2) 每一条前缀树带有时间戳信息, 时间戳表示访问一条路段的时间间隔. 基于 FTP-tree 的轨迹预测算法包含 FTP-tree 构建和 FTP-mining 两个步骤, FTP-tree 的构建过程与 FP-tree 的构建相同, 不再赘述. FTP-mining 算法挖掘频繁轨迹模式的过程不同于 FP-tree, 其主要目标是找到大于给定门限值的最长的频繁路径, 算法 2 第 4~12 行计算生成的最长频繁序列模式没有考虑路段模式组合的情况, 第 10 行中 conditional pattern base 是具有相同前缀模式的频繁路段项子集合.

在上述两项技术基础上, 给出基于 FTP-tree 的移动对象轨迹预测算法 PathPrediction, 如算法 3 所示^[19].

算法 3. 基于 FTP-tree 的移动对象轨迹预测算法 PathPrediction

- 输入. 移动对象数据库 \mathcal{D} , 路网中路段集合 \mathcal{S} .
 输出. 预测轨迹集合 Tr .
1. $\mathcal{R} = \text{RoIDiscovery}(\mathcal{D}, \mathcal{S})$;
 2. **for each** RoI r in \mathcal{R} **do**
 3. $Tr = \text{FTPTreeGeneration}(r)$;
 4. FTP-mining(Tr, α);
 5. 输出最可能运动轨迹集合 Tr .

算法 3 首先利用 RoIDiscovery 算法从移动对象数据库 \mathcal{D} 找出热点区域集合 \mathcal{R} (第 1 行). 然后, 对 \mathcal{R} 中的每个轨迹热点区域构建 FTP-tree (第 2 行和第 3 行), 并利用 FTP-mining 计算出频繁轨迹模式 (第 4 行). 最后, 输出所有频繁轨迹模式 (第 5 行).

3 基于 GPR 的复杂多模式轨迹预测模型

本节主要解决复杂场景中包含多种运动模式且很难用单一一种运动模式刻画的复杂运动模式预测问题, 首先介绍利用高斯过程回归模型实现单一运动模式预测, 进而引出基于高斯混合回归的复杂运动模式预测模型.

3.1 高斯过程回归

一种典型运动模式路径可以由一个高斯过程表示, 利用高斯过程回归 (Gaussian process regres-

sion, GPR)^[21] 进行预测. 下面介绍针对简单运动模式的高斯过程回归模型的理论基础.

对于输入训练数据集 $D_{\text{train}} = \{(x_i, y_i)\}_{i=1}^N = (X, Y)$, 其中 $x_i \in \mathbf{R}^d$ 为 d 维输入矢量, $y_i \in \mathbf{R}$ 为相应输出. 对于给定输入数据集 X , 可构成一个随机变量集合 $\{f(x_1), f(x_2), \dots, f(x_N)\}$, 且具有联合高斯分布, 则该高斯过程 (Gaussian process, GP) 的全部统计数字特征可由均值函数 $m(x)$ 和协方差函数 $k(x, x')$ 确定, 即

$$f(x) \sim GP(m(x), k(x, x')) \quad (1)$$

其中, $m(x) = E[f(x)]$, $k(x, x') = E[(f(x) - m(x)) \times (f(x') - m(x'))]$, 本文所用核函数为标准指数协方差函数, 表示为

$$\text{Cov}(f(x), f(x')) = \theta_0 \exp \left[-\frac{(x - x')^2}{2\theta_1^2} \right] + \sigma^2 \delta_{ij} \quad (2)$$

其中, θ_0 为指数权重参数, θ_1 表示长度规模参数, δ_{ij} 是狄拉克函数, 当 $i = j$ 时, 函数 $\delta_{ij} = 1$, 否则为 0.

应用高斯过程回归的目的是在训练集输入数据 X 和输出数据 Y 之间建立映射关系 $f(\cdot): \mathbf{R}^d \rightarrow \mathbf{R}$, 继而预测出与新测试输入数据 x^* 对应最可能的观测输出值 $f(x^*)$, 求取回归估计函数过程如下.

求解过程: 由于输入输出的观测和测量数据经常伴随噪声数据干扰, 因此将噪声数据考虑进观测目标值 y , 建立高斯过程回归一般模型: $y = f(x) + \varepsilon$. 其中, ε 为独立的高斯白噪声, 符合高斯分布, 均值为 0, 方差为 σ^2 , 可记为 $\varepsilon \sim N(0, \sigma^2)$. 由于噪声 ε 独立于 $f(x)$, 所以当 $f(x)$ 服从高斯分布时, 有限观测值联合分布的集合 y 同样可形成一个高斯过程, 即

$$y \sim GP(m(x), k(x, x')) \quad (3)$$

其中, 协方差函数以矩阵方式表达, 即

$$C(X, X) = K(X, X) + \sigma^2 I \quad (4)$$

式中, I 表示 $N \times N$ 的单位矩阵, $C(X, X)$ 表示 $N \times N$ 协方差矩阵, $K(X, X)$ 表示 $N \times N$ 的核函数矩阵, 元素 $K_{ij} = k(x_i, x_j)$, $K(X, X)$ 表示如下:

$$\begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{bmatrix}$$

$K(X^*, X) = [k(x^*, x_1), k(x^*, x_2), \dots, k(x^*, x_n)]$ 表示 $K(X, X)$ 中某一行, 且 $K(X^*, X^*) = k(x^*, x^*)$.

由于整个高斯过程模型利用多元高斯分布表达数据, 根据贝叶斯原理可知, GP 在给定训练集数

据 D_{train} 集合上建立先验函数, 然后在测试数据集 $D_{\text{test}} = \{(x_i, y_i)\}_{i=n+1}^{n+j} = (X^*, Y^*)$ 下转变为后验分布, 则训练数据输入数据 X 的输出向量 y 和测试数据 X^* 的输出向量 y^* 之间的联合高斯密度分布为

$$\begin{bmatrix} y \\ y^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu \\ \mu^* \end{bmatrix}, \begin{bmatrix} K(X, X) + \sigma^2 I & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix}\right)$$

为了求取输出向量 y^* , 需要借助联合高斯密度回归定理.

定理 1 (联合高斯密度回归). 若 $[X_1, X_2]^T \sim \mathcal{N}_{p+q}(\mu, \Sigma)$, 其中 $X_1 \in \mathbf{R}^P$, $X_2 \in \mathbf{R}^Q$, $\mu = [\mu_1, \mu_2]^T$, $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, 于是

$$X_2|X_1 \sim \mathcal{N}_q(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1), \Sigma_{22.1}) \quad (5)$$

其中, $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$.

由此可知, 当 $(y, y^*)^T$ 服从联合高斯分布时, 条件概率 $p(y^*|y)$ 是高斯分布, 所以得到 y^* 关于 y 回归方程及相应方差如下:

$$f(y) = \mathbb{E}[y^*|y] = \mu_{y^*} + \Sigma_{y^*y}\Sigma_{yy}^{-1}(y - \mu_y) \quad (6)$$

$$\sigma^2 = \text{Var}[y^*|y] = \Sigma_{y^*} - \Sigma_{y^*y}\Sigma_{yy}^{-1}\Sigma_{yy^*} \quad (7)$$

根据定理 1, 可以得到关于预测输出数据的条件概率 $P(y^*|y)$ 和 GP 回归方程, 即

$$y^*|y \sim \mathcal{N}(\mu^* + K^*K^{-1}(y - \mu), K^{**} - K^*K^{-1}K^{*T}) \quad (8)$$

其中, $K^* = K(X^*, X)$, $K^{**} = K(X^*, X^*)$, 于是 y^* 关于 y 的最佳回归估计方程, 即 y^* 的预测值为

$$\begin{aligned} \hat{y}^* &= f(y) = \mathbb{E}[y^*|X, y, X^*] = \\ &\mu^* + K(X^*, X)(K(X, X) + \sigma^2 I)^{-1}(y - \mu) \end{aligned} \quad (9)$$

而关于 y^* 的预测不确定性区间用方差表示为

$$\begin{aligned} v(y) &= \text{Var}[y^*|X, y, X^*] = K(X^*, X^*) - \\ &K(X^*, X)(K(X, X) + \sigma^2 I)^{-1}K(X, X^*) \end{aligned} \quad (10)$$

3.2 基于高斯混合回归的轨迹预测模型

在轨迹包含较为复杂运动模式场景下, 需要采用多个高斯过程, 利用高斯混合回归模型 (Gaussian mixture regression, GMR)^[22] 进行轨迹预测分析.

3.2.1 训练模型

在具有多种轨迹模式的场景中, 一条轨迹可能隶属于多个轨迹模式, 准确地刻画一条轨迹需要采用混合模型. 例如, 对于轨迹 $S = \langle (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \rangle$, 本文将转化为 X 和 Y 方向上的 n 维矢量 $X = (x_1, x_2, \dots, x_n)^T$ 和 $Y = (y_1, y_2, \dots, y_n)^T$, n 表示轨迹的观测点个数, X 和 Y 方向上矢量在所有 M 个轨迹模式中出现总概率是由不同运动模式的高斯概率混合而成, 表示为

$$p(x_n|\lambda) = \sum_{i=1}^M w_i GP(x_n|\mu_{x,i}, \Sigma_{x,i}) \quad (11)$$

$$p(y_n|\lambda) = \sum_{i=1}^M w_i GP(y_n|\mu_{y,i}, \Sigma_{y,i}) \quad (12)$$

其中, M 表示混合轨迹模式的状态数量, w_i 表示第 i 个轨迹模式的权重, 且 $\sum_{i=1}^M w_i = 1$; $\mu_{x,i}$ 和 $\Sigma_{x,i}$ 分别表示第 i 个轨迹模式状在 X 方向上的均值和协方差, 参数 $\lambda = \{w_i, \mu_i, \Sigma_i\}$, $i = 1, 2, \dots, M$. $GP(X_n|\mu_{x,i}, \Sigma_{x,i})$ 表示轨迹矢量 X_n 相对第 i 个运动模式状态的高斯概率函数, 定义如下:

$$\begin{aligned} GP(x_i|\mu_{x,i}, \Sigma_{x,i}) &= \frac{1}{\sqrt{(2\pi)^d |\Sigma_{x,i}|}} \times \\ &\exp\left\{-\frac{(x_i - \mu_{x,i})^T \Sigma_{x,i}^{-1} (x_i - \mu_{x,i})}{2}\right\} \end{aligned} \quad (13)$$

其中, μ 和 Σ 表示轨迹高斯过程的典型运动模式在 X 方向上的均值和协方差矩阵, $\mu = [\mathbb{E}(x_1), \mathbb{E}(x_2), \dots, \mathbb{E}(x_d)]^T$, $\Sigma = (C_{ij})_{d \times d}$, $C_{ij} = \text{Cov}(x_i, x_j)$.

于是, 对于轨迹训练集 $S = \{X, Y\}$, 整个轨迹训练集高斯混合模型似然函数为

$$p(X|\lambda) = \prod_{n=1}^N p(x_n|\lambda) \quad (14)$$

$$p(Y|\lambda) = \prod_{n=1}^N p(y_n|\lambda) \quad (15)$$

模型中如何计算复杂运动模式中的参数 λ 是一个关键步骤, 本文通过使轨迹训练模型 (即式 (14) 和式 (15)) 概率达到最大, 从已知的训练轨迹矢量集 $S = \{X, Y\}$ 中学习训练出最佳参数 λ , 利用概率最大的方式, 使用最大似然法估计 (Expectation-maximization, EM), 选出构成运动模式的概率密度最大的那一组参数, 为轨迹回归分析预测时所用.

EM 算法在迭代中改善模型参数估计, 在每次迭代中不断地增加模型估计参数 λ 与观测训练轨迹 X 方向矢量 x_i (由若干轨迹点构成) 的匹配概率, 这里讨论 X 轴方向, Y 轴方向情况类似, 即每次迭代使式 (14) 满足 $p(X|\lambda_{k+1}) > p(X|\lambda_k)$, 其中 k 表示

迭代的次数. 通过不断地学习, 获得最佳匹配训练轨迹矢量集 $X = \{x_1, x_2, \dots, x_n\}$. 迭代训练的目的即为找到一组 λ , 使得 $p(X|\lambda)$ 最大, 即

$$\hat{\lambda} = \operatorname{argmax}_{\lambda} p(X|\lambda) \quad (16)$$

由于篇幅限制, 求解使 $p(X|\lambda)$ 达到最大的参数 λ 值的过程参见文献 [23].

3.2.2 预测模型

在上一节的基础上, 本节给出基于 GMR 的轨迹预测模型. 已知训练数据集为 $S = (x, y)$, 其中输入数据为 x , 输出数据为 y . 测试数据集 $S^* = (x^*, y^*)$, 输入测试数据为 x^* , 预测输出 y^* , 那么 $[y, y^*]^T$ 的联合概率密度函数遵从如下的 GMR 模型.

$$p_{yy^*}(y, y^*) = \sum_{i=1}^M w_i GP(y, y^* | \mu_i, \Sigma_i) \quad (17)$$

其中, $\mu_i = [\mu_{iy}, \mu_{iy^*}]^T$, $\Sigma_i = \begin{bmatrix} \Sigma_{iyy} & \Sigma_{iyy^*} \\ \Sigma_{iy^*y} & \Sigma_{iy^*y^*} \end{bmatrix}$, 并且满足 $\sum_{i=1}^M w_i = 1$. 每个高斯成分 $GP(\cdot | \mu_i, \Sigma_i)$ 可由式 (9) 计算出, 联合概率密度函数表示为

$$p_{yy^*} = \sum_{i=1}^M \omega_i GP(y^* | y; f_i(y), \sigma_i^2) \quad (18)$$

其中, $f_i(y) = E[y^* | y] = \mu_{iy^*} + \Sigma_{iy^*y} \Sigma_{iyy}^{-1} (y - \mu_{iy})$, $\sigma_i^2 = \operatorname{Var}[y^* | y] = \Sigma_{iy^*y^*} - \Sigma_{iy^*y} \Sigma_{iyy}^{-1} \Sigma_{iyy^*}$. 除此之外, 边缘密度 p_y 和条件密度 $p_{y^*|y}$ 可由式 (17) 和式 (18) 求得, y 的边缘密度函数为

$$p_y(y) = \int p_{yy^*}(y, y^*) dy = \sum_{i=1}^M \omega_i GP(y; \mu_{iy}, \Sigma_{iyy}) \quad (19)$$

条件密度函数为

$$p_{y^*|y}(y^* | y) = \sum_{i=1}^M \Phi_i(y) GP(y^* | y; f_i(y), \sigma_i^2) \quad (20)$$

其中混合权重

$$\Phi_i(y) = \frac{\omega_i \times GP(y; \mu_{iy}, \Sigma_{iyy})}{\sum_{i=1}^M \omega_i \times GP(y; \mu_{iy}, \Sigma_{iyy})} \quad (21)$$

因此, y^* 关于 y 的回归函数 (y^* 预测值) 为

$$\hat{y}^* = f(y) = E[y^* | X, y, X^*] = \sum_{i=1}^M \Phi_i(y) f_i(y) \quad (22)$$

对应的方差为

$$v(y) = \operatorname{Var}[y^* | X, y, X^*] = \sum_{i=1}^M \Phi_i(y) (f_i^2(y) + \sigma_i^2) - \left(\sum_{i=1}^M \Phi_i(y) f_i(y) \right)^2 \quad (23)$$

式 (22) 称作高斯混合回归模型, 基于 GMR 的轨迹预测算法思路即利用式 (20) 对预测数据概率密度函数建模, 应用 EM 算法估计相应参数, 依据正态分布数据的条件分布 (参见定理 1) 得到 m 个高斯分量回归函数, 利用式 (22) 将总体回归函数加权混合实现对整体数据的回归预测分析.

4 实验及算法性能分析

4.1 基于 FTP-tree 的轨迹预测算法性能分析

为了验证基于 FTP-tree 的轨迹预测算法预测准确性和时间性能, 对 Naive 算法 (没有采用轨迹热点区域挖掘 RoIDiscovery 算法, 仅应用 FTP-mining 算法), PutMode 算法^[24] (基于时间连续贝叶斯网络的轨迹预测算法) 和 PathPrediction 算法 (集成轨迹热点区域挖掘和 FTP-mining 算法) 进行对比实验. 实验数据集使用 New York, Kansas 和 Chengdu 真实矢量地图. New York, Kansas 和 Chengdu 地图分别由 $435\,186 \times 563\,287$, $348\,693 \times 437\,998$ 和 $23\,514 \times 25\,672$ 个像素构成, 合成的轨迹数据通过修改 Brinkhoff 轨迹生成器^[25] 生成. 为了进一步证明算法的真实有效性, 实验使用微软亚洲研究院 GeoLife 数据集^[26], 由 182 个用户 5 年间 1 129 天 GPS 轨迹数据组成, 包含 17 621 条轨迹.

4.1.1 轨迹预测准确性对比

实验利用预测命中率评价算法的有效性^[19].

定义 2 (预测命中). 已知一条轨迹序列 $S = \{s_1, s_2, \dots, s_k\}$, 根据 S 得到的预测轨迹序列为 $T_p = \{p_1, p_2, \dots, p_k\}$, $\operatorname{dist}(p, q)$ 表示时空点 p 和 q 之间的欧氏距离, θ 是一个给定的距离门限值, $\operatorname{dist}(s_i, p_i) \leq \theta$ 表明一次轨迹预测命中, 预测命中定义为

$$H(s_i, p_i) = \begin{cases} 1, & \text{若 } \operatorname{dist}(s_i, p_i) \leq \theta \\ 0, & \text{若 } \operatorname{dist}(s_i, p_i) > \theta \end{cases} \quad (24)$$

定义 3 (预测命中率). 已知一条轨迹序列 S 和预测得到的轨迹序列 T_p , 预测命中率定义为

$$\operatorname{Accuracy} = \frac{\sum_{i=1}^n H(s_i, p_i)}{|T_p|}, \quad s_i \in S \quad (25)$$

其中, $|T_p|$ 表示组成 T_p 轨迹点的数量.

图 1 和图 2 给出了不同算法随轨迹数量增加的轨迹预测准确率. 其中, 横轴表示轨迹数量, 纵轴表示预测准确率. 实验结果表明, 融合轨迹热点区域挖掘和基于 FTP-tree 的频繁轨迹模式挖掘方法 PathPrediction 的预测精度在 Chengdu 数据集上平均高出 PutMode 算法 8.0 个百分点, 高出 Naive 预测算法 29.2 个百分点, 在 GeoLife 真实数据集上平均高出 PutMode 算法 5.5 个百分点, 高出 Naive 预测算法 28.9 个百分点. 原因在于 PathPrediction 算法对经过聚类后的轨迹热点区域进行频繁轨迹模式的挖掘, 轨迹热点区域是由经过噪声处理后的轨迹点组成的, 因此预测的准确性更高. 此外, 构建的频繁轨迹模式树 FTP-tree 综合考虑了真实路网中的轨迹运动规则, 因此预测的结果与移动对象真实运动情况更加吻合, 而 PutMode 算法融合了轨迹聚类和时间连续贝叶斯网络, 相比于没有进行轨迹热点挖掘的 Naive 算法准确性要高.

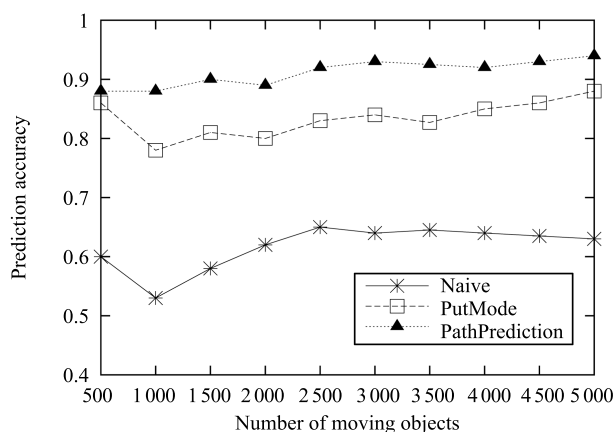


图 1 Chengdu 数据集下不同算法预测准确率比较

Fig.1 Prediction accuracy comparison of different algorithms under Chengdu datasets

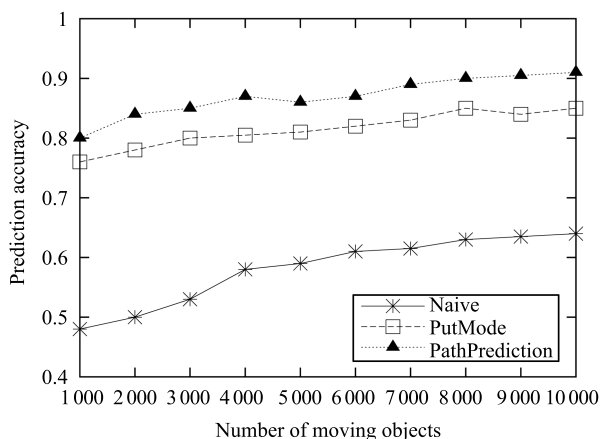


图 2 GeoLife 数据集下不同算法预测准确率比较

Fig.2 Prediction accuracy comparison of different algorithms under GeoLife datasets

4.1.2 轨迹预测时间性能对比

为了验证本文提出的基于 FTP-tree 的轨迹预测算法的时间性能, 本节实验在不同的电子地图 (New York, Kansas 和 Chengdu) 上生成轨迹数据集, 观察 PathPrediction 算法的预测时间, 结果如图 3 所示. 可以发现, 不同电子地图上, 随着生成轨迹数量的增加, PathPrediction 算法的运行时间均近似呈线性增长, 证明了算法的可伸缩性.

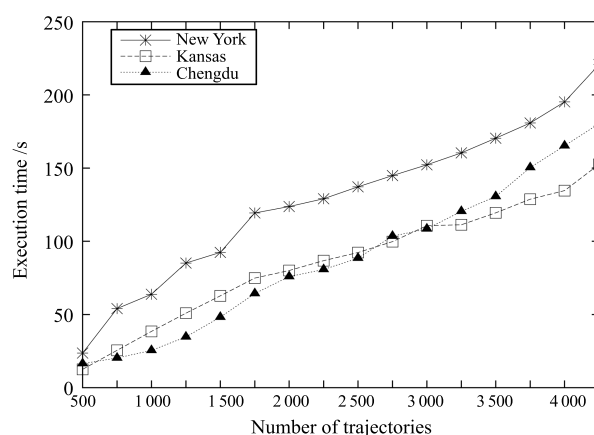


图 3 PathPrediction 算法在不同数据集下预测时间比较

Fig.3 Prediction time of PathPrediction algorithm under different datasets

为了进一步验证 PathPrediction 的时间性能, 随着移动对象数量增加, 观察算法 1~3 的运行时间, 实验数据集为 Chengdu 地图生成的轨迹数据集和 GeoLife 真实轨迹数据集, 结果如图 4 和图 5 所示.

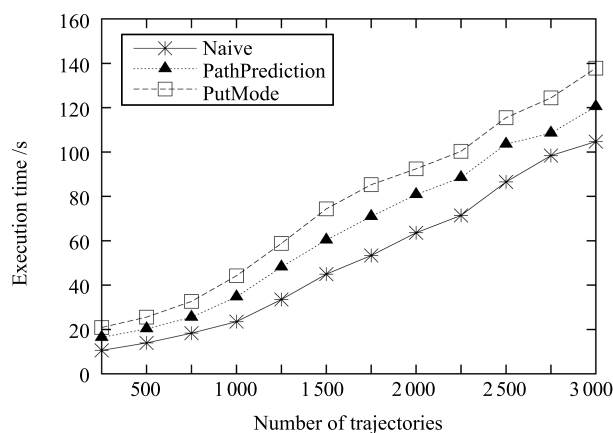


图 4 Chengdu 数据集下不同算法预测时间比较

Fig.4 Prediction time comparison of different algorithms under Chengdu datasets

通过图 4 和图 5 可以发现, PathPrediction 算法在所有情况下预测时间均小于 PutMode 算法, 略高于 Naive 算法. 原因在于 Naive 算法没有应

用轨迹热点区域挖掘算法, 因此预测时间要小于 PathPrediction 算法. 而 PutMode 算法需要训练时间连续贝叶斯网络并构建转换强度矩阵, 该过程非常耗时, 时间远远高于本文提出的频繁轨迹模式挖掘算法 FTP-mining. 此外, 可以发现 Naive 和 PathPrediction 算法运行时间接近, 说明本文提出的轨迹热点区域挖掘算法时间随着轨迹数量的增加不会发生巨大变化.

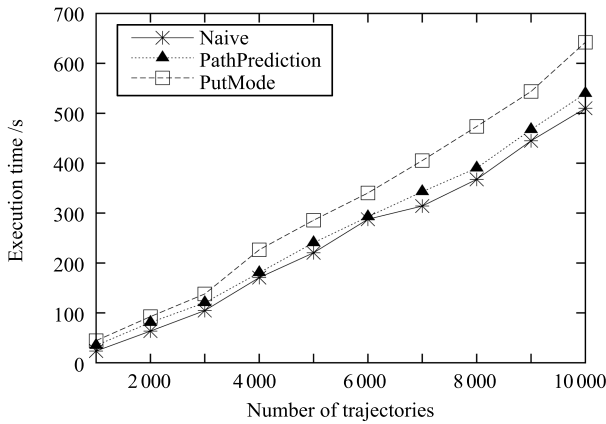


图5 GeoLife 数据集下不同算法预测时间比较

Fig.5 Prediction time comparison of different algorithms under GeoLife datasets

4.2 基于高斯混合回归的轨迹预测算法性能分析

为了验证本文提出的面向多模式移动对象轨迹预测算法的性能, 设计实现了基于高斯混合回归的轨迹预测算法 GM RTP, 基于卡尔曼滤波的预测算法 Kalman, 基于隐马尔科夫模型 (Hidden Markov model, HMM) 的轨迹预测算法 HMTP^[9]. 其中, Kalman filter^[27] 是一种应用普遍的回归预测方法, 与本文提出的预测算法的可比性较高. HMTP 算法基于 HMM 模型提取隐状态和观察状态, 预测效果较好, 通过与其比较, 可以充分证明本文算法的性能优势.

本节比较上述三种算法的准确性和时间性能. 实验数据来源于 MIT 停车场行驶车辆数据¹, 收集了 40 453 条真实轨迹.

4.2.1 高斯过程分量的选取

GM RTP 算法中高斯过程分量个数 M 的选取很重要, 如果选择的过程分量过少, 算法不能充分地对数据建模, 进而对测试轨迹进行准确的预测; 选择的过程分量太多, 则会增加轨迹预测时间代价. 本节实验观察在不同高斯过程分量个数下, 不同轨迹输入长度 (即已知轨迹点个数) 对预测误差的影响.

定义 4 (预测误差). 轨迹预测是将测试轨迹数据集输入预测模型得到预测输出轨迹, 其中输入测

试轨迹为部分轨迹点的集合, 采用均方根误差 (Root mean squared error, RMSE) 计算预测轨迹点与实际轨迹点的误差.

$$RMSE = \frac{\sum_{i=1}^k \sqrt{(x'_i - x_i)^2 + (y'_i - y_i)^2}}{k} \quad (26)$$

其中, (x_i, y_i) 表示真实位置, (x'_i, y'_i) 表示预测位置, k 为预测轨迹点的数量.

实验中每条观测轨迹由 60 个轨迹点构成, 已知输入轨迹点长度分别取 10, 30, 50, 对应预测的未来轨迹点数目分别是 50, 30, 10, 实验结果如图 6 所示.

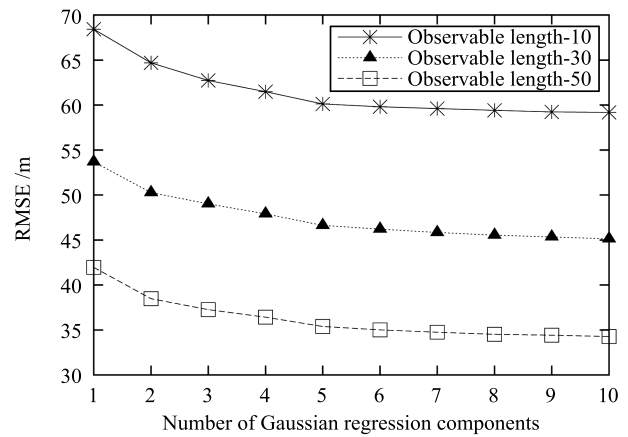


图6 不同高斯过程分量下轨迹预测误差比较

Fig.6 Prediction error comparison with distinct Gaussian regression components

从图 6 可以看出: 1) 在不同高斯过程分量下, 当观测的历史轨迹输入点个数越多时 (例如 observable length = 50), 算法预测误差越低. 因为预测模型有了更多的历史轨迹数据点信息, 包含更多的运动模式, 预测误差降低. 2) GP 混合分量个数从 1 增加到 5 时, 其预测误差变化较大. 当高斯过程分量达到 5 时, 预测误差收敛. 值得注意的是, 不同轨迹数据集其运动模式个数不同, 因此需要通过实验探寻最佳的高斯过程分量个数. 本节后面实验内容中 GM RTP 中高斯过程分量个数均设置为 5.

4.2.2 预测准确性分析

本节观察三种预测算法在训练轨迹数量为 39 000 条、不同测试集数据规模为 100~1 000 条轨迹下的预测准确性. 用预测命中率评价算法的准确性, 用预测误差 (RMSE) 评价算法的预测精度, 实验结果取每种测试集下所有轨迹预测结果的平均值, 如图 7 和图 8 所示.

从图 7 可以看出, 相比于其他两种算法, 随着测试轨迹数量的增加, GM RTP 预测误差最小, 一直保持在 40 m 上下波动, 相对较长的真实轨迹, 预测精

¹<http://www.ee.cuhk.edu.hk/~xgwan/MITtrajsingle.html>

度较高且比较稳定. 分析实验结果发现, HMTTP 预测的平均误差比 GM RTP 平均高出 5.5 m, GM RTP 相比 Kalman 算法, 预测误差平均减少 18.6 m. 原因在于 HMTTP 和 Kalman 轨迹预测算法对于处理较为简单运动模式的轨迹预测精度较好, 而实验中采用的数据集中轨迹模式较为复杂且种类繁多, 因此预测误差最大. 而基于 GMR 的模型通用性较好, 可以针对不同的多模式轨迹利用高斯过程回归预测.

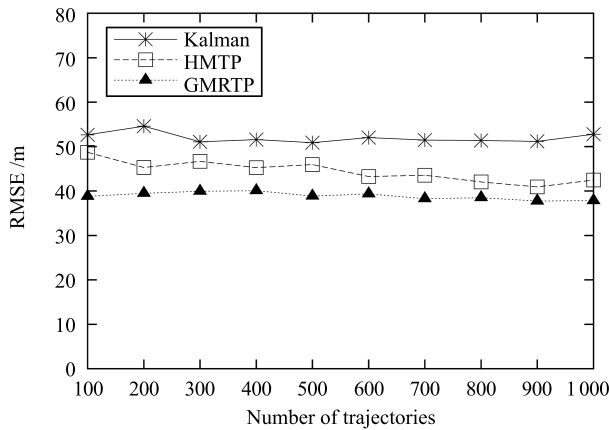


图 7 不同算法预测误差比较

Fig. 7 Prediction bias comparison of different algorithms

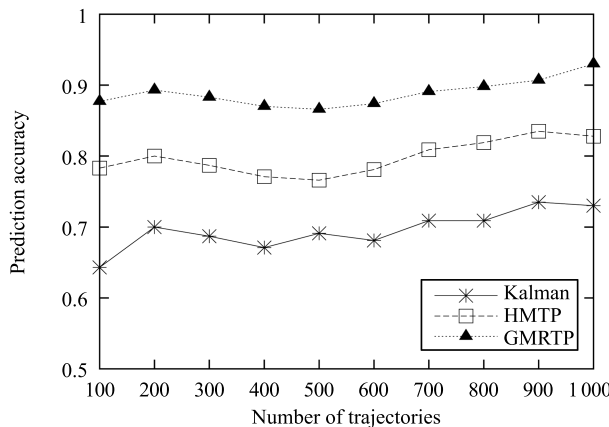


图 8 不同算法预测准确率比较

Fig. 8 Prediction accuracy comparison of different algorithms

从图 8 可以看出, 相比 HMTTP 和 Kalman 算法, GM RTP 算法的预测准确率平均提高 9.9% 和 21.2%, 且 GM RTP 比较稳定, 准确率维持在 90% 上下波动.

4.2.3 预测时间性能分析

本节实验观察上述三种算法随着移动对象数量增加的预测时间代价, 结果如图 9 所示.

从图 9 可以看出, GM RTP 的预测时间最少, 相比 HMTTP 和 Kalman 算法, 平均减少 49.3% 和 11.3%. 原因在于 Kalman 算法对每条轨迹中下一

个位置预测都要将前一个位置信息代入回归分析, 当预测的轨迹数量增加时, 预测时间近似呈线性增长. 而 GM RTP 预测时利用高斯过程刻画的轨迹仅需一次性预测即可, 因此当预测轨迹数量增加时, 只要没有增加较多的轨迹运动模式, 预测时间并不会产生较大的增加和波动. 而 HMTTP 算法预测最为耗时, 因为其在训练模型时需对轨迹进行分段处理, 并利用 HMM 模型提取隐状态和观察状态, 代价较高.

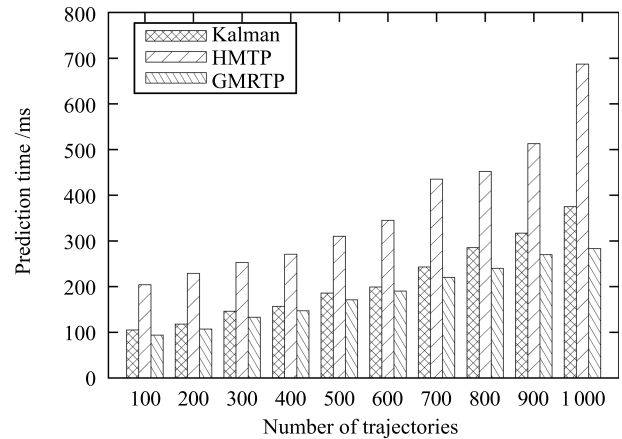


图 9 不同算法预测时间性能比较

Fig. 9 Prediction time comparison among different algorithms

4.2.4 单一运动模式轨迹预测准确性对比

为了进一步证明基于高斯过程回归的轨迹预测算法 GM RTP 的通用性, 本节比较在单一运动模式下, GM RTP 和 PathPrediction 的预测准确性, 实验采用 GeoLife 数据集, 结果如图 10 所示. 其中, 横轴表示轨迹的数量, 纵轴表示预测准确性.

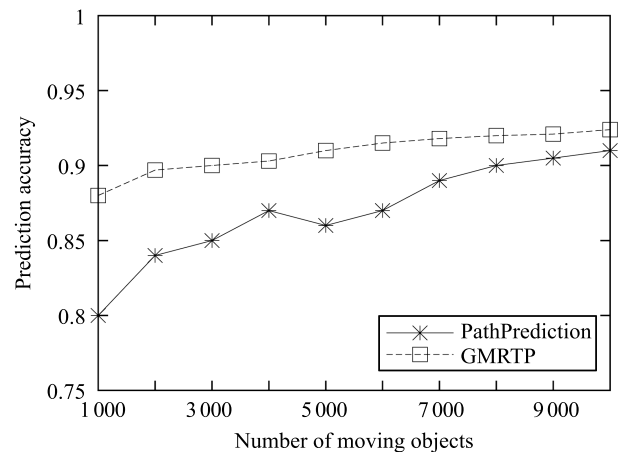


图 10 单一运动模式轨迹预测准确率比较

Fig. 10 Prediction accuracy comparison beyond single-motion-pattern trajectories

从图 10 可以看出, GM RTP 算法对具有单一运动模式的轨迹进行预测时仍然具有很高的准确性,

当移动对象数量大于 2000 时, 预测准确率在 90% 以上; 且随着移动对象数量增加, GM RTP 均优于本文提出的基于频繁模式树的单一模式轨迹预测算法 PathPrediction. 原因在于 GM RTP 对轨迹数据利用概率密度函数建模, 依据符合正态分布数据的条件分布得到回归函数完成回归预测, 能够准确地反应移动对象真实运动行为.

5 总结与展望

多模式移动对象不确定性轨迹预测是移动对象数据库领域新提出的充满挑战性的研究课题, 针对当前部分轨迹预测算法精度不高, 复杂多模式轨迹预测准确性不高、预测实时性不能保证、预测轨迹长度较短的不足, 本文提出一种基于轨迹频繁模式树的轨迹预测方法, 充分考虑真实路网中的轨迹运动规则, 预测结果准确性较高. 为了对复杂多模式移动对象位置进行预测, 提出了基于高斯过程回归的多模式轨迹预测模型, 算法运行过程中无需设置大量参数, 可以通过数据自身利用概率统计分布获得各种运动模式下的位置信息.

未来工作包括: 1) 对频繁模式树 FTP-tree 挖掘频繁项集进一步优化, 提高算法的挖掘效率; 2) 与国内交通部门合作将所提模型应用于卡口摄像头采集到的交通数据, 辅助跟踪和定位机动车辆和行人(主要针对未装备 GPS 定位系统的交通工具); 3) 为了保证预测的准确性, 充分考虑影响对象运动行为的主客观因素.

References

- Meng X F, Ding Z M, Xu J J. *Moving Objects Management: Models, Techniques and Applications*. Berlin: Springer-Verlag, 2014. 117–131
- Hu W M, Xiao X J, Fu Z Y, Xie D, Tan T N, Maybank S. A system for learning statistical motion patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(9): 1450–1464
- Parker A, Subrahmanian V S, Grant J. Fast and accurate prediction of the destination of moving objects. In: *Proceedings of the 3rd International Conference on Scalable Uncertainty Management*. Berlin, Germany: Springer, 2009. 180–192
- Song C M, Qu Z H, Blumm N, Barabási A L. Limits of predictability in human mobility. *Science*, 2010, **327**(5968): 1018–1021
- Centola D. The spread of behavior in an online social network experiment. *Science*, 2010, **329**(5996): 1194–1197
- Jeung H, Yiu M L, Zhou X F, Jensen C S. Path prediction and predictive range querying in road network databases. *The VLDB Journal*, 2010, **19**(4): 585–602
- Zhou J B, Tung A K H, Wu W, Ng W S. A “semi-lazy” approach to probabilistic path prediction in dynamic environments. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, USA: ACM, 2013. 748–756
- Pan T L, Sumalee A, Zhong R X, Indra-payoong N. Short-term traffic state prediction based on temporal-spatial correlation. *IEEE Transactions on Intelligent Transportation Systems*, 2013, **14**(3): 1242–1254
- Qiao S J, Shen D Y, Wang X T, Han N, Zhu W. A self-adaptive parameter selection trajectory prediction approach via hidden Markov models. *IEEE Transactions on Intelligent Transportation Systems*, 2015, **16**(1): 284–296
- Qiao Shao-Jie, Li Tian-Rui, Han Nan, Gao Yun-Jun, Yuan Chang-An, Wang Xiao-Teng, Tang Chang-Jie. Self-adaptive trajectory prediction model for moving objects in big data environment. *Journal of Software*, 2015, **26**(11): 2869–2883 (乔少杰, 李天瑞, 韩楠, 高云君, 元昌安, 王晓腾, 唐常杰. 大数据环境下移动对象自适应轨迹预测模型. 软件学报, 2015, **26**(11): 2869–2883)
- Ding Z M, Yang B, Güting R H, Li Y G. Network-matched trajectory-based moving-object database: models and applications. *IEEE Transactions on Intelligent Transportation Systems*, 2015, **16**(4): 1918–1928
- Xu J J, Gao Y J, Liu C F, Zhao L, Ding Z M. Efficient route search on hierarchical dynamic road networks. *Distributed and Parallel Databases*, 2015, **33**(2): 227–252
- Dai J, Yang B, Guo C J, Ding Z M. Personalized route recommendation using big trajectory data. In: *Proceedings of the 31st IEEE International Conference on Data Engineering*. Seoul, South Korea: IEEE Computer Society, 2015. 543–554
- Tripathy S, Howard C J. Multiple trajectory tracking. *Scholarpedia*, 2012, **7**(4): Article No. 11287
- Xu J Q, Güting R H. A generic data model for moving objects. *Geoinformatica*, 2013, **17**(1): 125–172
- Huang Yu-Long, Zhang Yong-Gang, Li Ning, Zhao Lin. An improved Gaussian approximate filtering method. *Acta Automatica Sinica*, 2016, **42**(3): 385–401 (黄玉龙, 张勇刚, 李宁, 赵琳. 一种改进的高斯近似滤波方法. 自动化学报, 2016, **42**(3): 385–401)
- Chen Cheng, He Yu-Qing, Bu Chun-Guang, Han Jian-Da. Feasible trajectory generation for autonomous vehicles based on quartic Bézier curve. *Acta Automatica Sinica*, 2015, **41**(3): 486–496 (陈成, 何玉庆, 卜春光, 韩建达. 基于四阶贝塞尔曲线的无人车可行轨迹规划. 自动化学报, 2015, **41**(3): 486–496)
- Monfort M, Liu A Q, Ziebart B D. Intent prediction and trajectory forecasting via predictive inverse linear-quadratic regulation. In: *Proceedings of the 29th AAAI Conference on Artificial Intelligence*. Austin, Texas, USA: AAAI, 2015. 3672–3678
- Qiao S J, Han N, Zhu W, Gutierrez L A. TraPlan: an effective three-in-one trajectory-prediction model in transportation networks. *IEEE Transactions on Intelligent Transportation Systems*, 2015, **16**(3): 1188–1198
- Han J W, Pei J, Yin Y W, Mao R Y. Mining frequent patterns without candidate generation: a frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, 2004, **8**(1): 53–87
- Rasmussen C E, Williams C K I. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. Cambridge, USA: The MIT Press, 2005. 12–22

- 22 Yu Jian-Bo, Lu Xiao-Lei, Zong Wei-Zhou. Wafer defect detection and recognition based on local and nonlocal linear discriminant analysis and dynamic ensemble of Gaussian mixture models. *Acta Automatica Sinica*, 2016, **42**(1): 47–59

(余建波, 卢笑蕾, 宗卫周. 基于局部与非局部线性判别分析和高斯混合模型动态集成的晶圆表面缺陷探测与识别. *自动化学报*, 2016, **42**(1): 47–59)

- 23 Dellaert F. The Expectation Maximization Algorithm, Technical Report GIT-GVU-02-20, College of Computing, Georgia Institute of Technology, USA, 2002.

- 24 Qiao S J, Tang C J, Jin H D, Long T, Dai S C, Ku Y C, Chau M. PutMode: prediction of uncertain trajectories in moving objects databases. *Applied Intelligence*, 2010, **33**(3): 370–386

- 25 Brinkhoff T. A framework for generating network-based moving objects. *GeoInformatica*, 2002, **6**(2): 153–180

- 26 Zheng Y, Xie X, Ma W Y. Geolife: a collaborative social networking service among user, location and trajectory. *IEEE Data(base) Engineering Bulletin*, 2010, **33**(2): 32–40

- 27 Deng Hu-Bin, Zhang Lei, Wu Ying, Zhou Jie, Liu Feng. Research on track estimation based on Kalman filtering algorithm. *Transducer and Microsystem Technologies*, 2012, **31**(5): 4–7

(邓胡滨, 张磊, 吴颖, 周洁, 刘枫. 基于卡尔曼滤波算法的轨迹估计研究. *传感器与微系统*, 2012, **31**(5): 4–7)

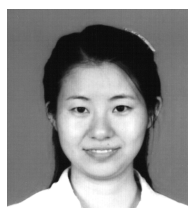


乔少杰 成都信息工程大学网络安全学院教授. 2009 年获得四川大学计算机学院工学博士学位. 主要研究方向为轨迹预测, 移动对象数据库, 大数据.

E-mail: sjqiao@cuit.edu.cn

(**QIAO Shao-Jie** Professor at the School of Cybersecurity, Chengdu University of Information Technology. He

received his Ph.D. degree from the College of Computer Science, Sichuan University in 2009. His research interest covers trajectory prediction, moving objects databases, and big data.)



韩楠 成都信息工程大学管理学院讲师. 2012 年获得成都中医药大学博士学位. 主要研究方向为数据挖掘. 本文通信作者. E-mail: hannan@cuit.edu.cn

(**HAN Nan** Lecturer at the School of Management, Chengdu University of Information Technology. She received her Ph.D. degree from Chengdu Uni-

versity of Traditional Chinese Medicine University in 2012. Her main research interest is data mining. Corresponding author of this paper.)



丁治明 北京工业大学计算机学院教授. 2002 年获得中国科学院计算技术研究所工学博士学位. 主要研究方向为轨迹大数据, 移动对象数据库.

E-mail: zmding@bjut.edu.cn

(**DING Zhi-Ming** Professor at the College of Computer Science, Beijing University of Technology. He received

his Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences in 2002. His research interest covers trajectory big data and moving objects databases.)



金澈清 华东师范大学数学科学与工程学院教授. 2005 年获得复旦大学计算机科学技术学院工学博士学位. 主要研究方向为基于位置的服务, 不确定数据管理, 数据质量.

E-mail: cqjin@sei.ecnu.edu.cn

(**JIN Che-Qing** Professor at the School of Data Science and Engineering, East China Normal University. He received his Ph.D. degree from the School of Computer Science, Fudan University in 2005. His research interest covers location-based services, uncertain data management, and data quality.)



孙未未 复旦大学计算机科学技术学院教授. 2002 年获得复旦大学计算机科学技术学院工学博士学位. 主要研究方向为空间数据处理, 基于位置的服务.

E-mail: wwsun@fudan.edu.cn

(**SUN Wei-Wei** Professor at the School of Computer Science, Fudan University. He received his Ph.D. de-

gree from the School of Computer Science, Fudan University in 2002. His research interest covers spatial data processing and location-based services.)



舒红平 成都信息工程大学软件工程学院教授. 2010 年获得四川大学计算机学院工学博士学位. 主要研究方向为数据挖掘. E-mail: shp@cuit.edu.cn

(**SHU Hong-Ping** Professor at the School of Software Engineering, Chengdu University of Information Technol-

ogy. He received his Ph.D. degree from the College of Computer Science, Sichuan University in 2010. His main research interest is data mining.)