

DOI:10.16356/j.1005-2615.2019.05.013

一种基于元路径拥塞模式挖掘的移动对象位置预测方法

王宁^{1,2} 韩京宇^{1,2,3} 王尚凌^{1,2} 万杨兰^{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(南京邮电大学), 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: 路网上移动对象位置预测是许多位置相关服务的基础。目前移动对象位置预测方法没有充分考虑到轨迹数据中所蕴含的道路拥塞信息,而路网上的道路拥塞状态对移动对象的位置更新会产生巨大影响。提出基于元路径拥塞模式挖掘的方法(Meta-congestion-pattern mining, MCPM)。在离线挖掘阶段,从历史轨迹的频繁路径(元路径)的紧集中挖掘当地的拥塞模式,并对运动模式进行建模,其中采用基于 k 均值的聚类算法解决数据稀疏性问题。在线预测阶段根据挖掘的拥塞模式和运动模式依概率进行预测。最后通过理论分析和实验验证得出了算法的有效性,与相同条件下的精度预测(WhereNext, WN)方法相比,平均预测准确性提高了近 20%,预测时间平均缩短了近 50%。

关键词: 移动对象;位置预测;GPS 数据;位置相关服务

中图分类号: TP391

文献标志码: A

文章编号: 1005-2615(2019)05-0675-06

Method to Position Prediction of Mobile Object Based on Meta-congestion-Pattern Mining

WANG Ning^{1,2}, HAN Jingyu^{1,2,3}, WANG Shangling^{1,2}, Wan Yanglan^{1,2}

(1. College of Computer Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China; 2. Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, Nanjing, 210023, China; 3. Key Laboratory of Computer Network and Information Integration Ministry of Education, Southeast University, Nanjing, 211189, China)

Abstract: The location prediction of mobile objects on the road network is the basis of many location-related services. At present, the method for predicting the position of a moving object does not fully consider the road congestion information which mining from the trajectory data, and the congestion state has a great influence on the position update of the moving object. A meta-congestion-pattern mining (MCPM) method is proposed. In the offline mining stage, local congestion patterns are mining from the tightly-concentrated frequent paths (meta-paths) of the historical trajectories, and the movement patterns are modeled. A mean-based clustering algorithm is used to solve the data sparsity problem. The online forecasting stage predicts the probability based on the mining congestion model and the motion pattern. Finally, the validity of the algorithm is obtained by theoretical analysis and experimental verification. Compared with the WN method under the same conditions, the average prediction accuracy is improved by nearly 20%, and the prediction time is shortened by an average of nearly 50%.

Key words: mobile object; location prediction; GPS data; location based services

基金项目: 国家自然科学基金(61003040, 61373139, 61602260)资助项目; 东南大学计算机网络和信息集成教育部重点实验室(项目号 K93-9-2015-07C)资助项目; 江苏省自然科学基金面上基金(BK20171447)资助项目; 江苏省高校自然科学研究面上基金(17KJB520024)资助项目。

收稿日期: 2019-06-10; **修订日期:** 2019-07-28

通信作者: 韩京宇, 男, 博士, 教授, E-mail: jyhan@njupt.edu.cn。

引用格式: 王宁, 韩京宇, 王尚凌, 等. 一种基于元路径拥塞模式挖掘的移动对象位置预测方法[J]. 南京航空航天大学学报, 2019, 51(5): 675-680. WANG Ning, HAN Jingyu, WANG Shangling, et al. Method to Position Prediction of Mobile Object Based on Meta-congestion-Pattern Mining[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2019, 51(5): 675-680.

许多位置相关服务依赖于准确预测路网上移动对象未来位置。比如滴滴打车中,乘客在地铁站发布打车任务,打车平台需要判断在最近时间内将要到达地铁站的车辆,并将需求发布给车辆。目前基于路网的移动对象位置预测面临的主要挑战在于^[1-4]:(1)不同地区的交通环境差异巨大;(2)交通拥堵状况瞬息万变,(3)车辆的高速移动导致位置变化十分迅速。目前,有许多路网移动对象位置预测方法,可以分为个性化预测方法^[5-8]和一般化预测方法^[9-12],但是许多方法并没有充分考虑到不同时间,由于拥塞状态的变化导致经过相同路径花费的行程时间差别。本文提出基于元路径拥塞模式挖掘的方法(Meta-congestion-pattern mining, MCPM)来预测路网上移动对象的位置,从历史轨迹中提取频繁路径(元路径),在其聚集上挖掘时空相关的拥塞模式,并进一步通过历史轨迹建模运动模式,在线预测移动对象的位置。但是由于成本、技术和隐私原因,轨迹数据稀疏问题尤为严重,具体表现为特定时空可用轨迹数据很稀少。为此本文采用基于 k 均值的聚类算法,丰富可用的轨迹数据。

1 相关工作

现存的移动对象位置预测方法大致分为两类:基于个体的预测方法和基于群体的预测方法。

基于个体的预测方法将每个移动对象的运动轨迹单独考虑,仅使用单独个体的历史轨迹,只能对特定移动对象位移预测。Jeung等^[5]首先定义运动函数,从Apriori算法的修改版本中提取的运动模式并建立线性或非线性模型。Ching等^[6]用前缀树表示历史运动模式,用降序树匹配当前部分轨迹,计算得分,返回最大分作为预测结果。其他方法如马尔可夫模型^[7]和隐马尔可夫模型(Hidden Markov model, HMMs)^[8]已经被广泛应用于个体位置预测问题。

基于群体的预测方法是基于所有移动对象的运动行为。Asahara等^[9]发现基于混合马尔可夫链模型(Mixed Markov-chain model, MMM)位置预测的新方法,行人被分为几个小组,每个小组中运动行为相似。Gambs等^[10]将移动马尔可夫链(Mobility Markov chain, MMC)扩展,并考虑移动对象之前访问的地方,对应于更高阶的马尔可夫模型。Mathew等^[11]提出了一种基于HMMs位置预测的混合方法。首先根据时间周期特征对历史位置进行聚类,然后对每个聚类进行HMM训练。Andy

等^[12]提出了一种解决数据稀疏问题的新方法,首先将轨迹分解为子轨迹,然后根据统计概率使用马尔可夫模型连接子轨迹,但忽略了零频率问题。

2 系统概览和基本概念

2.1 系统概览

如图1所示,MCPM系统从上到下分为3个阶段:预处理阶段、离线挖掘和在线预测阶段。预处理阶段,使用地图匹配算法,将GPS点匹配到路网上,并将点存储到以路径id为名建立的文件中。离线挖掘阶段,利用元路径间空间关系提取拥塞特征;建模移动对象的运动模式,采用基于 k 均值的聚类算法解决轨迹数据稀疏问题。在线预测阶段,输入移动对象当前信息,并预测移动对象位置。

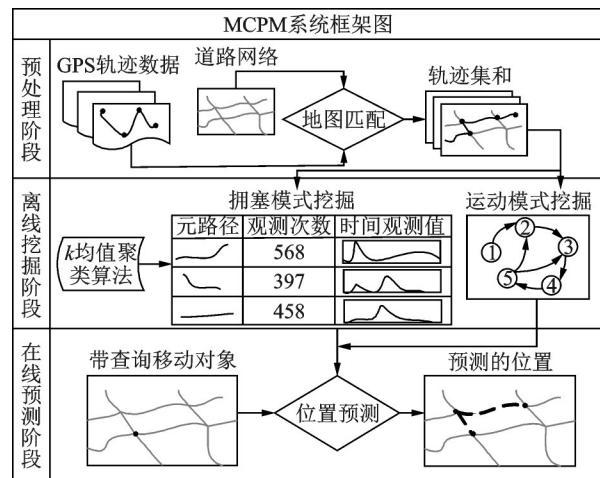


图1 MCPM系统框架图

Fig. 1 Framework of MCPM

2.2 基本概念

定义1 道路网络:道路网络 $G(V, E)$ 是一个有向图, V 为路口结点集, E 为路段边集,对于路段 $e, e \in E$ 。

定义2 GPS轨迹: $T = \{(p_i, t_i)\}_{i=1}^n$,将GPS轨迹表示为 n 个空间-时间点的序列,点 $p_i = (x_i, y_i)$ 是二维空间中的GPS点,表示经度和纬度, t_i 为样本的时间戳, n 表示 T 中GPS样本点的个数。

定义3 地图匹配轨迹:由一系列边的组合 $T_G = \{e_1, \dots, e_{|T_G|}\}$ 表示,给定路网 G 和移动轨迹 T ,通过地图匹配,将移动轨迹 T 在 G 中表示出来。

定义4 位置预测问题: T 为路网 G 中的历史轨迹集合。根据 T 中的轨迹,在已知移动对象当前位置 $p(p_x, p_y)$,速度 v_p 和时间 t_p ,计算经过时间 t 移动对象可能到达的位置 $p_t(p, v_p, t_p)$ 。

定义5 元路径:元路径 r 为路网 G 上两个路

口之间的轨迹,它可能被一条或者多条轨迹经过。

定义 6 路径字典(Path dictionary, PD):路径字典是元路径的集合。

3 拥塞模式挖掘

将 GPS 数据使用地图匹配算法匹配到路网上,将匹配后的轨迹分割成元路径存入路径字典中,挖掘每条元路径各个时间间隔的拥塞状态,使用基于 k 均值的聚类算法,填补特征矩阵中的空缺值。

3.1 元路径的拥塞建模

从历史数据中提取各条元路径的经过时间 $d(r)$,并计算其累积分布函数。根据累积分布函数计算元路径上的拥塞状态 $\hat{d}(r)$, $\hat{d}(r)$ 为 $[0, 1]$ 间的均匀分布,由此可以比较不同元路径的拥塞程度。图 2 将路径的拥塞状态用颜色表示,从 0 到 1 随着颜色的加深,拥塞程度越来越严重。

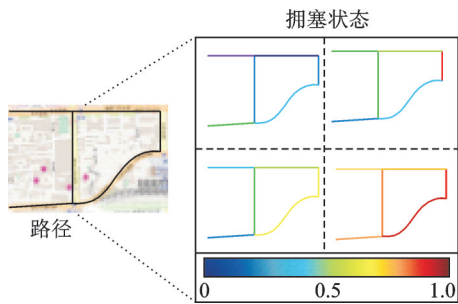


图 2 路径的拥塞状态
Fig. 2 Congestion status of path

现实交通模式中邻接的元路径间的交通状态不是独立的,因此利用元路径间的空间关系来捕获当地交通模式特征。

元路径 r 的邻域集合 $NB(r) = \{o_1, \dots, o_s\}$,表示与元路径 r 邻接的 s 条元路径(含 r),计算其动态拥塞状态——将整个时间范围离散成固定的时间间隔 f_i (本文中设置为 30 min),计算在各个时间间隔内,元路径上观测到的经过时间 $d_{f_i}(r)$;有些时间间隔 f_i 可能包含多个观测值,用观测值的期望表示,并计算拥塞状态 $\hat{d}_{f_i}(r)$ 。

给定元路径 r 和邻域 $\{o_1, \dots, o_s\}$,定义时间间隔 f_i 的特征向量 $M(r)_i, M(r)_i$ 为在时间间隔 f_i ,包含 $NB(r)$ 的拥塞状态的 s 维向量,即 $M(r)_i = [\hat{d}_{f_i}(o_1), \dots, \hat{d}_{f_i}(o_s)]$ 。将计算的所有 $M(r)_i$ 叠加到特征矩阵 $M(r)$ 中。设 N 为特征矩阵中特征向量的数量,即离散时间间隔的总量,矩阵 $M(r)$ 表示

$NB(r)$ 的动态拥塞状态,有

$$M(r) = \begin{bmatrix} M(r)_1 \\ \vdots \\ M(r)_N \end{bmatrix} \quad (1)$$

3.2 拥塞特征聚类

由于数据稀疏性,大多数时间间隔上没有观测值或只有很少的观测值。针对这个问题,使用 k-POD 方法^[13],一个迭代的基于 k 均值的聚类算法,适用于数据缺失率较高和缺失原因不清楚问题。

将 $M(r)$ 的行聚类为 k 个组,计算 k 个聚类中心 c_1, \dots, c_k ,找到距离 $M(r)$ 中的行最近的聚类中心,由此初始化 $M(r)$ 中的缺失值。不同元路径的特征矩阵尺寸变化比较大,采用 Gaussian-Means 方法^[14]找出最优的 k 值。

在拥塞特征中引入时间关系来优化初始聚类结果。给定特征矩阵 $M(r)$,关联矩阵 W 表示 $M(r)$ 中两个时间间隔的时间相似性约束,优化 k 个聚类中心,尽量减少与实际观测值的差距,同时确定 $M(r)$ 中的行与 k 个聚类中心中的软分配。

将相邻时间间隔之间的相似性设为第 1 种类型的时间相似性约束,即在连续的时间间隔内,元路径交通状况从完全畅通状态转换到完全拥塞状态概率较小。因此在第 i 个时间间隔和第 j 个时间间隔之间定义指数衰减函数,有

$$C_{sm}(i, j) = \exp\left(-\frac{(t_i - t_j)^2}{\sigma_{sm}^2}\right) \quad (2)$$

式中: t_i 和 t_j 为时间间隔 f_i 和 f_j 的开始时间; σ_{sm} 默认为 2。

第 2 种类型的时间相似性约束是基于城市交通流量的周期相似性。重点关注本时间间隔内到达高峰段的用时(Start time of day, SOD),称为 h_i ,即时间间隔 f_i 的 SOD 标志符,不考虑其是否处在工作日或周末。定义 f_i 和 f_j 之间 SOD 权重,有

$$C_{SOD}(i, j) = \exp\left(-\frac{\min\{|h_i - h_j|, h_{\max} - |h_i - h_j|\}^2}{\sigma_{SOD}^2}\right) \quad (3)$$

式中 $\sigma_{SOD} = \sqrt{2}$ 。边权重 $w(i, j)$ 由 C_{SOD} 和 C_{sm} 的线性组合计算得出,其中系数 θ 为 0.5,有

$$w_{i,j} = \theta C_{sm}(i, j) + (1 - \theta) C_{SOD}(i, j) \quad (4)$$

Q 为 $N \times k$ 阶聚类分配矩阵。 Q 的每行 q_i 是一个二元向量,如果时间间隔 f_i 被分配给聚类 j ,则 $q_i(j) = 1$,否则为 0。 k 个聚类中心为 $k \times s$ 阶矩阵 C 的行向量。使用上述结果初始化 Q 和 C ,然后通

过解决下列最小化问题找出最优值,有

$$\underset{Q,c}{\text{minimize}} \left\| QC - M(r) \right\|_F^2 + \gamma \text{Tr}(Q^T L Q) \quad (5)$$

式中 L 为拉普拉斯矩阵, $L = D - W$, $D = \text{diag}\left(\sum_{j=1}^n w_{i,j}\right)$ 。常系数 γ 控制聚簇过程中时间一致性问题的权重,式(5)使用交替方向优化方法来求解。

4 基于拥塞模式的位置预测

将移动对象的运动模式建模与挖掘的拥塞模式结合,说明如何进行位置预测,并给出相应的算法。

4.1 建模运动模式

建模运动模式,考虑移动对象到达终点前可能经过的元路径,总结任意相邻起点路口 s 到终点路口 s' 的路径 ξ 。假设路径 ξ 包含 n 个元路径 (r_1, \dots, r_n) , $p_{s,s'}(r_i)$ 表示从路口 s 到路口 s' , 经过元路径 r_i 的概率,且 $\sum_{i=1}^n p_{s,s'}(r_i) = 1$ 。给定元路径 r 的邻接元路径 r' , 由 r 转到 r' 的概率为 $p(r'|r) = \frac{p(rr')}{p(r)} = \frac{p(r)p(r')}{p(r)} = p(r')$ 。

根据历史数据计算所有相邻路口间经过各条元路径的概率,得到基于概率分布的运动模式。 $M_{s,s'} = (p_{s,s'}(r_1), \dots, p_{s,s'}(r_n))$ 为一个行向量,表示路口 s 到终点路口 s' , 经过各条元路径的概率,路网中共有 m 对相邻的路口,得到 m 个行向量。估算移动对象轨迹的方法总结在算法1中。

算法1: BriefCal

输入: 当前位置 p , 时间 t_p , 速度 v_p , 查询时间范围 t
 输出: 移动对象将要经过的轨迹集合 ξ

- (1) $road_id \leftarrow p$ // 得到路径 id
- (2) $g = v_p \times t$ // 估算移动对象经过的距离
- (3) while($g > 0$)
- (4) $road_length, s, s' \leftarrow road_id$ // 得到路长, 路口
- (5) $g = g - road_length$ // 更新距离
- (6) $\xi \leftarrow r_{road_id}$
- (7) $road_id \leftarrow M_{s,s'}$ // 依概率计算下一条路径 id
- (8) return ξ

4.2 基于拥塞模式的位置预测

将第3节挖掘出的拥塞模式加入位置预测,方法总结在算法2中。

算法2 PositionPredict

输入: 位置 p , 速度 v_p , 时间 t_p , 查询时间范围 t , CongesMol 为拥塞模式

输出: 移动对象位置 p_i

- (1) $Q = \emptyset$ // 保存移动对象经过的元路径
- (2) $\xi = (r_1, \dots, r_n) \leftarrow \text{Briefcal}(p, v_p, t)$ // 估算经过的元路径
- (3) for each r_i in ξ
- (4) $t_{r_i} \leftarrow \text{CongesMol}(r_i, t_p)$ // 计算时间
- (5) $t_p = t_p + t_{r_i}$ // 更新时间
- (6) $t = t - t_{r_i}$
- (7) if ($t \geq 0$) // 是否超出查询时间范围
- (8) $Q.add(r_i)$
- (9) else
- (10) $l = \frac{t + t_{r_i}}{t_{r_i}} |r_i|$ // 计算距离
- (11) $Q.add(l)$ // 将距离加入集合
- (12) break
- (13) $p_i \leftarrow Q$ // 计算位置

首先将当前拥塞模式转化为经过每条元路径的时间(第4行),判断移动对象经过此段时间是否继续前进,即时间差值是否大于0(第7行)。若时间差值大于等于0,将元路径 r_i 加入到结果集合 Q 中,继续循环;若小于0,在此条元路径上经过的距离为: $l = \frac{t + t_{r_i}}{t_{r_i}} |r_i|$ (第10行), $|r_i|$ 为元路径 r_i 的长度,将其加入到 Q 中。最后根据 Q 计算移动对象位置。

5 实验结果与分析

在本节中,通过实验来评估 MCPM 方法的效果。实验环境为 Intel(R) Core(TM) i5-5200U 2.20 GHz 的 CPU, 4 GB 内存, 使用 Windows 10 操作系统和 Eclipse 集成开发环境。算法用 Java 语言实现。

5.1 实验数据集

采用 T-drive 工程^[15-16]的真实数据集,包含2008年2月2日至2月8日期间北京地区10357辆出租车的GPS轨迹。轨迹由一系列带时间戳的点组成,包括出租车id、时间、经度和纬度。特别关注北京市5个行政区(东城区,西城区,朝阳区,丰台区和海淀区),比较其位置预测情况,表1列出了各个区域的特征。

在每个测试区域中,从原始位置信息中随机抽取85%作为训练集,剩余的15%作为测试用例。将文献[17]中的WN方法与本文提出的MCPM方法进行对比,WN方法从历史轨迹中提取频繁轨迹模式,并建立T-pattern决策树来进行位置预测。

表 1 区域参数表
Tab. 1 Specs of test regions

区名	轨迹数量	元路径数量
东城区	3 511	11 363
西城区	2 651	12 401
朝阳区	8 039	39 589
丰台区	7 574	30 681
海淀区	9 649	40 186

5.2 评估方法

实验采用控制变量法,对位置预测的准确性进行评估。计算预测得到的位置与真实位置点之间的距离偏差,误差率计算公式为

$$\delta = \frac{\sum_{i=1}^n \sqrt{(p_i - P_i)^2}}{n} \quad (6)$$

式中: p_i 为预测的移动对象位置, P_i 为移动对象真实位置, n 为预测的移动对象数量。

5.3 实验结果与分析

训练集为 5 000,比较表 2 中不同测试区域的位置预测结果,如表 2 所示。MCPM 方法的误差率比 WN 方法平均降低 23%。

西城区,东城区人流密,面积小,位置预测难度大。由于较高的路线多样性和较高的轨迹数量,朝阳区,丰台区和海淀区的准确率高于另外两个区。

表 2 各区域位置预测结果

Tab. 2 Regions location prediction result

区名	WN 误差率	MCPM 误差率	降低/%
东城区	0.011 3	0.009 3	0.17
西城区	0.012 0	0.009 1	0.19
朝阳区	0.008 9	0.006 2	0.30
丰台区	0.009 2	0.006 9	0.25
海淀区	0.007 1	0.005 2	0.26
平均	0.009 5	0.007 2	0.23

以下位置预测均基于朝阳区。在测试集量为 400,预测时间为 5 min,随着历史数据量的增大,MCPM 和 WN 的误差率变化如图 3 所示。

在图 3 中,误差下降越来越平缓,因为此时已包含大部分运动模式,能够处理各种复杂运动模式

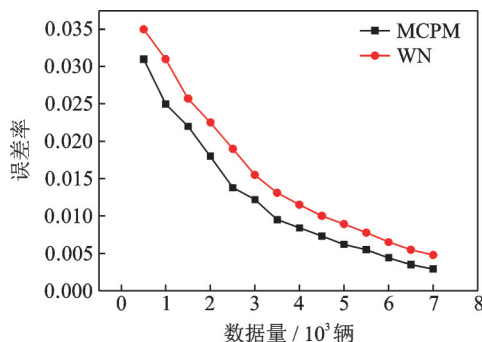


图 3 数据量与误差率关系结果图

Fig. 3 Relationship between data size and error rate

位置预测问题。WN 方法没有充分考虑同一条路径在 1 天中不同时刻,经过时间有所不同,因此误差率大。

进一步将 1 天中的时间划分为固定的时间间隔,详细分析每个时间间隔内各个预测方法的效果,如图 4 所示,选取了 1 天中的部分时间段,显示了在不同时间各个方法的误差率。其中的 MCPM* 方法从历史数据中挖掘拥塞模式,没有使用 k 均值聚类算法对稀疏数据进行处理,在数据丰富的早高峰和晚高峰时间段,误差率与 MCPM 方法相差不大,但在数据相对稀疏的时间段,比如:10:00~10:30,由于没有补齐特征矩阵中的空缺值,误差率明显高于 MCPM 方法,WN 方法由于没有考虑到轨迹数据的时效性,在不同的时间段误差率变化较大。

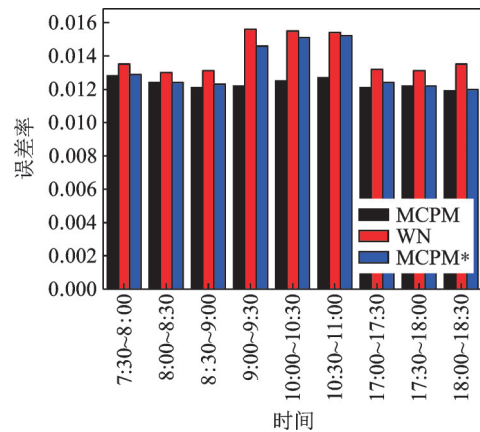


图 4 不同时间段与误差率关系结果图

Fig. 4 Relationship between different period and error rate

对位置预测时间参数进行修改,得到如图 5 所示结果。随着时间的增加,移动对象运动的随机性变大,而且出租车受到载客的影响,位置预测难度增大。

为了进一步验证本文方法的性能,观察算法的预测时间代价。图 6 显示了随着测试车辆数量的增大,预测时间的增长情况。WN 构建了基于全北京市的地区决策树,存储在磁盘上,预测时需要频繁 I/O 操作和对树的频繁遍历,非常耗时,预测时

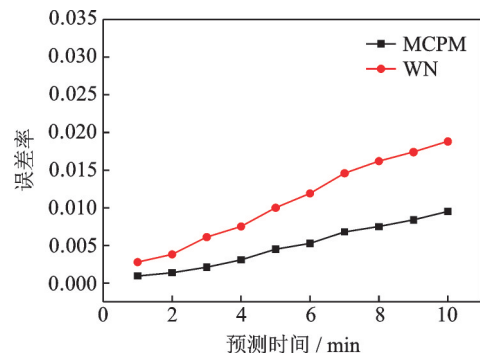


图 5 预测时间与误差率关系结果图

Fig. 5 Relationship between prediction time and error rate

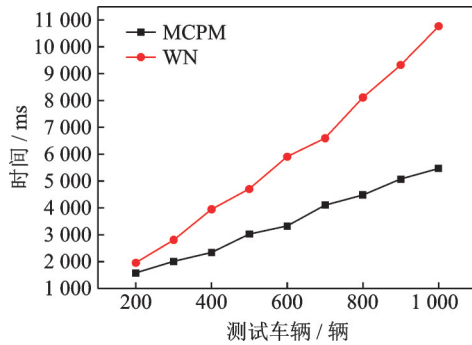


图6 测试车辆与时间关系实验结果图

Fig. 6 Relationship between test data size and time

间会呈线性增长;而MCPM方法将元路径的特征矩阵保存在内存中,时间代价低。

6 结 论

本文提出MCPM方法预测路网上移动对象的位置,利用邻域中元路径之间的空间关系,提取局部道路的拥塞特征,并针对稀疏轨迹数据;提出采用 k 均值聚类算法,实现准确、高效的位置预测。本方法的优点在于分解的元路径从更细的粒度捕捉拥塞变化规律,对数据进行聚类处理以填补特征矩阵中的空缺值,为预测提供精准支撑。

参考文献:

- [1] 乔少杰, 金琨, 韩楠, 等. 一种基于高斯混合模型的轨迹预测算法[J]. 软件学报, 2015, 26(5):1048-1063.
QIAO Shaojie, JIN Kun, HAN Nan, et al. Trajectory prediction algorithm based on Gaussian mixture model [J]. Journal of Software, 2015, 26(5):1048-1063.
- [2] GIANNOTTI F, NANNI M, PEDRESCHI D. Efficient mining of temporally annotated sequences [C]//Proceedings of the 2006 SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics. Bethesda, MD, USA: International Conference on Data Mining, 2006: 348-359.
- [3] 于彦伟, 贾召飞, 曹磊, 等. 面向位置大数据的快速密度聚类算法[J]. 软件学报, 2018, 29(8): 2470-2484.
YU Yanwei, JIA Zhaofei, CAO Lei, et al. Fast density-based clustering algorithm for location big data [J]. Journal of Software, 2018, 29(8): 2470-2484.
- [4] GIANNOTTI F, PEDRESCHI D, THEODORIDIS Y. Geographic privacy-aware knowledge discovery and delivery [C]//Proceedings of the 12th International Conference on Extending Database Technology. New York, USA: ACM, 2009: 1157-1158.
- [5] JEUNG H, LIU Q, SHEN H T, et al. A hybrid prediction model for moving objects [C]//International Conference on Data Engineering. [S. l.]: IEEE, 2008:70-79.
- [6] CHING YING J J, LEE W C, WENG T. Semantic

trajectory mining for location prediction [C]//Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. [S.l.]: ACM, 2011.

- [7] ASHBROOK D, STARNER T. Using GPS to learn significant locations and predict movement across multiple users [J]. Personal and Ubiquitous Computing, 2003, 7(5): 275-286.
- [8] ALVAREZ-GARCIA J A, ORTEGA J A, GONZALEZ-ABRIL L. Trip destination prediction based on past GPS log using a hidden Markov model [J]. Expert Systems with Applications, 2010, 37(12): 8166-8171.
- [9] ASAHARA A, MARUYAMA K, SATO A, et al. Pedestrian - movement prediction based on mixed Markov - chain model [C]//Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. [S.l.]: ACM, 2011: 25-33.
- [10] GAMBS S, KILLIJIAN M, DEL P C M N. Next place prediction using mobility Markov chains [C]//Proceedings of the First Workshop on Measurement. Privacy, and Mobility. [S.l.]: ACM, 2012: 254-265.
- [11] MATHEW W, RAPOSO R, MARTINS B. Predicting future locations with hidden Markov models [C]//Proceedings of the 2012 ACM Conference on Ubiquitous Computing. [S.l.]: ACM, 2012: 911-918.
- [12] ANDY Yuanxue, ZHANG Rui, ZHENG Yu, et al. Destination prediction by sub-trajectory synthesis and privacy protection against such prediction [C]//Data Engineering (ICDE), 2013 IEEE 29th International Conference on. [S.l.]: IEEE, 2013.
- [13] CHI J, CHI E, BARANIUK R. k-POD: A method for k-means clustering of missing data [J]. The American Statistician, 2016, 70(1): 91-99.
- [14] LI M, AHMED A, SMOLA A J. Inferring movement trajectories from GPS snippets [C]//Eighth ACM International Conference on Web Search and Data Mining. [S.l.]: ACM, 2015: 325-334.
- [15] YUAN J, ZHENG Y, XIE X, et al. Driving with knowledge from the physical world [C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [S.l.]: ACM, 2011: 316-324.
- [16] YUAN J, ZHENG Y, ZHANG C, et al. T-drive: Driving directions based on taxi trajectories [C]//SIGSPATIAL International Conference on Advances in Geographic Information Systems. [S. l.]: ACM, 2010: 99-108.
- [17] MONREALE A, PINELLI F, TRASARTI R, et al. Wherenext: A location predictor on trajectory pattern mining [C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM, 2009: 637-646.