

不确定性车辆路口的轨迹预测

毛莺池 陈 杨

(河海大学计算机与信息学院 南京 211100)

摘 要 在城市道路中,实时、准确、可靠地对移动车辆进行轨迹预测具有极高的应用价值,不仅可以提供准确的基于位置的服务,而且可以帮助过往车辆预知前方的交通状况。目前,移动车辆的轨迹预测方法主要基于历史轨迹的欧氏空间进行,并未考虑在受限路网中采用不确定性历史数据的车辆轨迹预测。针对这一问题,提出一种补全路径的基于马尔科夫链的轨迹预测方法,其优势在于:重新定义了补全路径算法,弥补了不确定性历史数据的不完整性,利用马尔科夫链低时间复杂度、高预测准确度的优势实现预测,避免了因频繁模式挖掘带来的查询时间过长而影响预测效率以及存在多余噪声影响轨迹预测准确率的问题。通过真实数据和实验分析表明:在参数设置相同的情况下,该方法比挖掘频繁轨迹模式算法的预测准确率平均提高了 18.8%,预测时间平均缩减了 80.4%。因此,该方法对于车辆路口的轨迹预测具有较高的预测准确率,并且能预测一系列的车辆未来轨迹。

关键词 受限路网,车辆轨迹预测,不确定性历史数据,补全路径,马尔科夫链

中图分类号 TP311 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.03.037

Uncertain Vehicle Intersection Trajectory Prediction

MAO Ying-chi CHEN Yang

(College of Computer and Information, Hohai University, Nanjing 211100, China)

Abstract In the city road, real-time, accurate and reliable trajectory prediction of mobile vehicles can bring very high application value, which can not only provide accurate location-based services, but also can help the vehicle to predict the traffic situation. At present, the trajectory prediction method of moving vehicles is mainly based on the precise historical trajectory in Euclidean space, and does not consider the vehicle trajectory prediction with uncertain historical data in restricted road network. A trajectory prediction method based on Markov chain was proposed to solve this problem. Its advantages include redefining the path algorithm of completion, making up for the incompleteness of uncertain historical data, and achieving prediction by using the characteristics of low time complexity and high prediction accuracy with Markov chain. This method avoids the problem of low prediction accuracy caused by too much query time due to the frequent pattern mining and the excess noise. The results show that under the same parameter setting, the prediction accuracy of the method is 18.8% higher than that of the mining frequent trajectory model, and the prediction time is reduced by 80.4% on average. Therefore, the method has high prediction accuracy for the trajectory prediction of the vehicle intersection, and achieves the prediction of a series of vehicle future trajectories.

Keywords Restricted road network, Vehicle trajectory prediction, Uncertainty historical data, Completion path, Markov chain

1 引言

随着无线通信技术和全球定位技术的日益发展以及移动设备的普及,越来越多基于位置的信息服务技术备受关注,尤其是轨迹预测技术。轨迹预测技术利用用户的当前位置信息、历史移动轨迹和其他辅助信息动态估计出移动用户的未来位置。在智能交通控制系统、军事数字化战场、辅助驾驶系统中,实时、精准、可靠地预测移动用户的轨迹具有极高的价

值^[1]。在位置数据中,轨迹数据拥有巨大的价值潜力^[2]。不同于一般的签到数据,轨迹数据具有很强的时效性,能够具体、完整地呈现出行程路线、驻留时间等。利用这些信息可以挖掘出人群的行动规律、驻流热点等,进而通过分析和选择,快速且有效地推送用户需求的信息^[3]。轨迹预测在基于位置的信息服务中,广泛的应用前景。譬如,在城市道路中通过轨迹预测技术可以提前监测和预判交通状况,从而为过往车辆推荐最佳路线;根据车辆的行驶道路为其推荐附近的商业信

到稿日期:2017-07-18 返修日期:2017-08-18 本文受国家重点研发计划项目(2016YFC0400910),重大科技专项(2017ZX07104001),中央高校基本科研业务费专项资金(2015B22214,2017B16814,2017B20914)资助。

毛莺池(1976—),女,博士,副教授,硕士生导师,CCF会员,主要研究方向为分布式计算与并行处理、分布式数据管理等,E-mail:yingchimao@hhu.edu.cn(通信作者);陈杨(1993—),男,硕士生,主要研究方向为分布式计算、轨迹预测,E-mail:1542551915@qq.com。

息;反映交通流量;影响道路规划;同时可被用于根据城市功能区的分布情况来完善城市的设施建设^[4]等。这些均是轨迹预测对城市规划的价值所在^[5]。因此,如何准确预测移动车辆的位置信息成为亟需解决的问题。目前已有一部分研究成果,如预测移动车辆的历史轨迹以及定位和位置信息等。但由于理论和技术尚不成熟,这些技术还不能很好地适应不断变化的移动技术的需求,不能高准确度地预测车辆轨迹。

本文第2节阐述相关工作;第3节给出基本概念和模型的基本思路;第4节将系统阐述不确定性车辆历史轨迹的数据分析;第5节阐述不确定性车辆路口轨迹预测的工作原理和算法实现;第6节通过仿真实验来证明所提算法的优势;最后总结全文并对未来工作进行展望。

2 相关工作

移动对象轨迹预测技术主要分为基于欧氏空间的轨迹预测和基于路网受限的轨迹预测。基于欧氏空间的预测方法可以分为基于模式的预测和基于向量的预测。基于模式的预测^[6]是通过频繁模式挖掘找出轨迹模式^[7],再根据轨迹模式进行轨迹预测。基于向量的预测方法分为线性模型和非线性模型。线性模型预测^[8]主要是假设移动对象从前一时刻到当前时刻的运动为匀速直线运动,并根据移动对象前一时刻的位置和速度预测当前时刻移动对象的位置。而非线性模型预测^[9-10]则是通过更改曲线模型的变量,使之变为直线模型,再用直线模型进行预测,即通过数学公式模拟移动对象的运动轨迹。

目前大多数移动对象都是在路网上运动的,并且运动的方向和行驶速度都会受到路网条件和环境的限制,因此基于欧氏空间的预测方法存在一定的局限性。国内外针对路网受限的轨迹预测也展开了相关的研究。文献^[11]通过测量个体轨迹的信息熵定量地给出了人类(车辆)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。Sung等人^[12]提出了基于轨迹预测的车辆自动巡航(ACC)方法,同时考虑了轨迹的偏航率和曲率对预测结果的影响,但预测精度不高,且实现过程繁杂。乔等人^[1]针对轨迹空间划分对预测准确率的影响问题,提出了基于隐马尔科夫模型(HMM)的轨迹预测算法 HMTTP,该算法将空间按照网格形式进行划分,划分后的轨迹网格对应 HMM 中的观察状态,训练轨迹划分出的段落对应 HMM 中的隐状态,由此建立轨迹预测模型,但模型的训练代价过高。文献^[13]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。文献^[14]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的运动变换模式和在不同路段上的速度信息。文献^[15]提出了一种基于轨迹时间连续贝叶斯网络的不确定性轨迹预测算法,其根据街区号、移动速度、移动方向3个变量构建了轨迹时间连续贝叶斯网络,并利用该网络预测移动对象的动态运动行为来计算可能的运动轨迹;但为了保证轨迹预测的准确率,该算法牺牲了时间性能。文献^[16]提出了一种环境自适应车辆轨迹预测方法,其通过车载移动终端实时获得轨迹数据,再将轨迹数据输入到轨迹回归预测模型中进行轨迹在线预测;但由于执行时

间过长,算法只能适用于对时间要求不高的环境,实际应用意义有所降低。文献^[17]提出了一种基于路网的不确定性轨迹频繁模式挖掘算法,其利用索引快速查询轨迹模式进行预测;但计算量较大,查询效率有待提升,并且预测准确率不高。

针对上述研究的不足,本文提出一种不确定性车辆路口轨迹预测算法。该算法利用补全路径的方法完善历史数据,再进行马尔科夫模型轨迹预测,其优势在于:避免了轨迹频繁模式挖掘带来的过多噪声,从而提升了预测的准确率,并且实现了车辆路口的一系列轨迹预测。

3 基本概念和模型的基本思路

已知移动车辆的数据库 D ,其中存在大量不同移动车辆在不同时间点的位置信息,由这些位置信息组成的有序集合被称为车辆轨迹。

定义1(路径概率) 如果有两条不连续的路径 P_1 和 P_2 ,并且两条路径是移动车辆连续经过的路径, S_1, \dots, S_n 是路径 P_1 和 P_2 之间连通可行的所有 n 条路径,每个路口都可以包含多个连通可行的路口,如图1所示,则从 P_1 到 P_2 选择路径 S_i 的路径概率公式^[21]如式(1)所示。

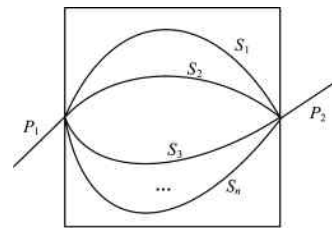


图1 连通可行的 n 条路径

Fig. 1 Connected and feasible n paths

$$P_{S_i|P_1} = \begin{cases} \frac{\sum_{j=1}^n L(S_j) - L(S_i)}{(n-1) \sum_{j=1}^n L(S_j)}, & n > 1 \\ 1, & n = 1 \end{cases} \quad (1)$$

其中, $L(S_j)$ 表示路径 S_j 的长度,当 $n=1$ 时,只有一条连通可行的路径,即移动车辆在经过已知路径 P_1 后选择可行路径 S_1 的概率为1。路径 S_i 越短,代表其路径概率越大,且 n 条连通可行的路径概率之和为1。

定义2(不确定性数据) 不确定性数据是由某一条可行路径与其路径概率组合得到的,由 $x:p$ 表示。其中 x 表示轨迹路径, p 表示该路径的路径概率,并且 $p < 1$ 。如果两条路径之间的路径只有一条,则表示为 $x:1$ 。

定义3(不确定轨迹数据集) 不确定性轨迹数据集是由不确定性数据组成的集合,表示为 $L = (x_1:p_1, x_2:p_2, \dots, x_n:p_n)$ 。

定义4(轨迹点) 存在于不确定性轨迹数据集中的每个路口点被称为轨迹点。

定义5(不确定性数据轨迹序列) 不确定性数据轨迹序列是由所有可行路径的不确定性数据集组成的有序序列,表示为 $S = (L_1, L_2, \dots, L_n)$,如图2所示。此轨迹序列包括3条路径信息 (a, b, c) 和两次路径补全,图中矩形内的路径就是

根据真实路网补全的可行路径。路径 a 可以用 $a:1$ 表示;路径 a 和 b 之间有 3 条可行路径,用数据集合表示为 $(x_1:0.3, x_2:0.6, x_3:0.1)$;路径 b 和 c 之间有 3 条可行路径,用数据集合表示为 $(x_4:0.2, x_5, x_7:0.5, x_6, x_7:0.3)$ 。那么该移动车辆的不确定性数据轨迹序列表示为 $S = \{a:1, (x_1:0.3, x_2:0.6, x_3:0.1), b:1, (x_4:0.2, x_5, x_7:0.5, x_6, x_7:0.3), c:1\}$ 。

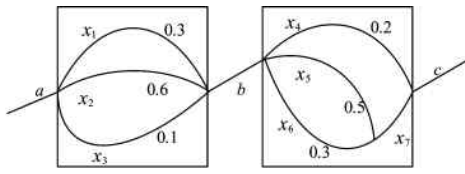


图 2 不确定性数据轨迹序列

Fig. 2 Track sequence of uncertain data

定义 6 (马尔科夫链) 设离散随机过程表示为 $\{X(t), t \in T\}, T = 0, 1, 2, \dots$; 状态空间 $S = \{0, 1, 2, \dots\}$ 。如果对于任意正整数 m, n, k 及任意非负整数 $j_m > j_{m-1} > \dots > j_2 > j_1 (n > j_m)$ 和 $i_{n+p}, i_n, i_{j_m}, \dots, i_{j_2}, i_{j_1}$, 存在 $P\{X(n+k) = i_{n+k} | X(n) = i_n, X(j_m) = i_{j_m}, \dots, X(j_2) = i_{j_2}, X(j_1) = i_{j_1}\} = P\{X(n+k) = i_{n+k} | X(n) = i_n\}$, 则称 $\{X(t), t \in T\}$ 为马尔科夫链^[19]。

定义 7 (k 步转移矩阵) 对于条件概率 $P\{X(n+k) = j | X(n) = i\}$, 系统在 n 时刻处于状态 i 的条件下, 经过时间间隔 k 之后, 在 $n+k$ 时刻转移到状态 j 的条件概率记为 $p_{ij}(n, n+k)$, 或简记为 $p_{ij}^{(k)}(n)$, 该概率被称为马尔科夫链的 k 步转移概率^[18-20]。当 $k=1$ 时, 通常记为 $p_{ij}^{(1)}(n) = p_{ij}$, 并称 p_{ij} 为马尔科夫链的转移概率。矩阵 $p^{(k)}$ 由转移概率 $p_{ij}^{(k)}(n)$ 组成, 表示为:

$$p^{(k)} = \begin{bmatrix} p_{11}^{(k)}(n) & p_{12}^{(k)}(n) & \dots & p_{1n}^{(k)}(n) \\ p_{21}^{(k)}(n) & p_{22}^{(k)}(n) & \dots & p_{2n}^{(k)}(n) \\ \dots & \dots & \dots & \dots \\ p_{n1}^{(k)}(n) & p_{n2}^{(k)}(n) & \dots & p_{nn}^{(k)}(n) \end{bmatrix}$$

称 $p^{(k)}$ 为马尔科夫链的 k 步转移矩阵。当 $k=1$ 时, 马尔科夫链的一步转移矩阵 p 为:

$$p = \begin{bmatrix} p_{11}(n) & p_{12}(n) & \dots & p_{1n}(n) \\ p_{21}(n) & p_{22}(n) & \dots & p_{2n}(n) \\ \dots & \dots & \dots & \dots \\ p_{n1}(n) & p_{n2}(n) & \dots & p_{nn}(n) \end{bmatrix}$$

本文所提模型的基本思路可以分为以下 3 个步骤:

- 1) 利用补全路径的方法处理不确定性历史数据, 然后将完整的数据储存在历史数据库中;
- 2) 根据公式, 基于历史数据计算路口转移概率, 并将所有转移概率进行组合以得到马尔科夫状态转移矩阵;
- 3) 利用马尔科夫链模型计算状态转移矩阵, 得到车辆未来的轨迹。

4 不确定性车辆历史轨迹的数据分析

4.1 问题分析

目前, 移动对象的服务结构由基于位置的服务 (Location Based Service, LBS) 和历史轨迹管理服务共同组成。其中, LBS 对移动对象发来的位置更新信息进行基本处理后, 会将其发送到历史轨迹管理服务中, 然后由历史轨迹管理服务对

移动对象的轨迹信息进行有效管理。由于轨迹信息在传输过程中可能受实际环境等因素的影响, 例如时间间隔过长导致位置更新延迟和包丢失等问题, 移动对象的位置更新路径可能会被部分遗漏, 即移动对象的历史轨迹并不是连续的, 因此在轨迹移动数据库中存在大量的不确定性轨迹。为了保证移动对象轨迹信息的完整性和连续性, LBS 需要补全其路径信息^[17]。

4.2 不确定性历史数据的处理

为了补全路径信息, 假设移动对象在路网中连续两次的位置信息 P 和 Q 是非连续的, 那么就不能完全依靠某一种单一标准 (如 P 和 Q 之间的最短路径) 对位置 P 和位置 Q 进行路径补全, 而需要充分考虑两个位置信息之间多条路径的选择可能性。根据路网有限道路信息的特点, 可以充分利用真实的路径长度作为多条路径的最优选择标准, 即由定义 1 可知, 将真实的路径长度作为路径概率的基础信息, 是补全路径方法的前提和基本要求。

对不确定性历史轨迹数据进行信息补全时, 大多学者针对不确定性轨迹数据集, 采用挖掘频繁轨迹模式再由轨迹模式进行轨迹预测的方法。该方法的实质是从初始轨迹数据集中挖掘频繁序列集合, 再利用轨迹模式索引树的方法对频繁序列集合进行比对, 最终找到支持度最高的轨迹模式作为预测结果。这种方法虽然有效地弥补了不确定性历史数据的不完整性, 但由于挖掘频繁轨迹模式的计算量较大, 查询效率有待提升, 并且预测准确率仅约 44.1%。因此, 本文根据马尔科夫链的特点重新定义了补全路径的方法: 首先利用真实的路径长度, 对具有不确定性的两个路口位置, 由定义 2 定义 5 计算得到概率信息, 再根据真实路口的车辆数据, 由概率信息计算每个路口的修补数据信息, 进而实现不确定性轨迹数据的补全目标。本文算法既弥补了不确定性历史数据的不完整性, 又通过低阶马尔科夫链低时间复杂度、高预测准确率的特点有效地避免了挖掘频繁轨迹模式的不足, 减少了查询时间, 避免了噪声带来的影响, 具有较高的轨迹预测准确率。

5 轨迹预测

5.1 工作原理

车辆在交叉路口的轨迹预测是由当前状态轨迹的历史信息和历史轨迹的统计数据共同决定的。距离当前状态较远的历史状态可以忽略不计, 距离当前状态越近的历史状态对轨迹预测的影响度越大, 由定义 4 和定义 5 可知, 马尔科夫状态转移矩阵的 k 值越大, 算法的复杂度越高, 因此对 k 步转移矩阵的选择至关重要。实验证明^[16] 低阶马尔科夫链也具有较好的预测效果, 因此可根据车辆历史轨迹数据构建一阶马尔科夫状态转移矩阵。

设某道路共有 n 个路口, 马尔科夫状态转移矩阵为一个 $n \times n$ 的矩阵, 一步状态转移矩阵 P 中的第 i 行第 j 列元素 p_{ij} 表示其经过路口 $i (i \geq 1)$ 后再通过路口 $j (j \leq n)$ 的概率, 其计算公式如式 (2) 所示:

$$p_{ij} = \frac{N_{ij}}{\sum_{j=1}^n N_{ij}}, 1 \leq i, j \leq n \quad (2)$$

其中, N_{ij} 表示车辆历史轨迹数据中车辆经过路口 i 后再经过

路口 j 的次数,即 i 和 j 必须是相邻路口,由此得到一步状态转移矩阵 P ,进而得到任意的 k 步状态转移矩阵, $P^{(k)} = P^k$ 。

在第 2 节中,根据定义 2 和定义 3 统计车辆历史轨迹数据时需要不确定性历史数据进行补全处理,得到非连续路口之间多条路径的路径概率;再由车辆在路网中经过不同路口的行驶次数,根据路径概率计算得到当前路口转到下一步所有路口的行驶次数,以完善轨迹数据。图 3 给出了一组路径缺失的路口,即 a, b 两个路口存在路径缺失。已知 x_1, x_2, x_3 是两路口之间的可行路径,其路径概率分别为 0.3, 0.5, 0.2, 并且路径 x_1 经过路口 c, x_2 经过路口 d, x_3 经过路口 e 。根据历史数据得知,车辆经过路口 a 再经过路口 b 的行驶次数为 420,则选择 x_1 路径的有 $420 * 0.3 = 126$ 次,选择 x_2 路径的有 $420 * 0.5 = 210$ 次,选择 x_3 路径的有 $420 * 0.2 = 84$ 次,这样就可以得到车辆经过路口 a 再经过路口 c 的行驶次数为 210,经过路口 c 再经过路口 b 的次数为 210,以此类推来完善历史轨迹数据库。

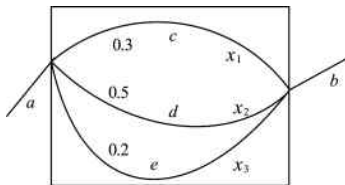


图 3 一组路径缺失的路口

Fig. 3 A set of path-missing intersections

有了完整的车辆历史轨迹,数据就可以基于马尔科夫链对车辆在路口的轨迹进行预测。根据公式计算得到的一步状态转移矩阵 P 表示为:

$$P = \begin{bmatrix} P_{11}(n) & P_{12}(n) & \cdots & P_{1n}(n) \\ P_{21}(n) & P_{22}(n) & \cdots & P_{2n}(n) \\ \vdots & \vdots & \vdots & \vdots \\ P_{n1}(n) & P_{n2}(n) & \cdots & P_{nn}(n) \end{bmatrix}$$

其中,行号代表路口总数为 n 的历史数据构成的当前已知的轨迹路口,列号代表未来通过的下一路口。根据当前位置扫描矩阵得到行号,选取对应行号所在的一行数据进行比较。概率最大的列号作为未来的下一路口,然后将此路口作为当前位置再次代入状态转移矩阵进行迭代求解,最终得到未来一系列的车辆路口轨迹。

5.2 算法实现

本文提出的不确定性路网车辆轨迹预测算法包括补全路径算法和车辆轨迹预测算法。

算法 1 补全路径算法

输入:不确定性历史轨迹数据集 $D_{\text{lack}} = \{L_1, L_2, \dots, L_n\}$

输出:补全路径历史轨迹数据集 $D_{\text{full}} = \{L_1, L_2, \dots, L_n\}$

1. 轨迹点 p ;
2. for each $p \in D_{\text{lack}}$
3. if ($p = \text{null}$)
4. mark p as D_{miss} ;
5. else
6. mark p as visited;
7. $M = \text{getPath}(D_{\text{miss}})$;
8. $D_{\text{full}} = \text{getFull}(M)$;
9. end

算法 1 的基本思路如下:

遍历 D_{lack} 中的所有轨迹点,若轨迹点 p 为空,则将其标记为信息遗漏集合 D_{miss} (第 4 行);若不为空,则将其标记为已访问(第 6 行);第 7 行利用 $\text{getPath}()$ 函数计算轨迹点集合 D_{miss} 中每个轨迹点与相邻轨迹点的路径概率,并将其存入集合 M 中;第 8 行利用 $\text{getFull}()$ 函数将补全的轨迹点及其路径概率添加到原数据中,得到补全路径数据集 D_{full} 。

算法 2 车辆轨迹预测算法

输入:补全路径历史轨迹数据集 $D_{\text{full}} = \{L_1, L_2, \dots, L_n\}$

输出:预测的下一个路口点 T

1. $D_{\text{full}} = \{L_1, L_2, \dots, L_n\}$;
2. for $i = 1$ to n
3. for $j = 1$ to n
4. $p_{ij} = \text{Transfer}(D_{\text{full}})$;
5. $P[i][j] = p_{ij}$;
6. end
7. end
8. $q = \text{Scan}(\text{now})$;
9. $T = q * P$.

算法 2 的基本思路是:第 1—4 行在得到不确定性历史轨迹数据集 D_{full} 的情况下,利用函数 $\text{Transfer}()$ 对补全路径数据集中的每个轨迹点计算路口转移概率 p_{ij} ;第 5 行由路口转移概率组合得到一阶马尔科夫状态转移矩阵 P ;第 8 行利用函数 $\text{Scan}()$ 扫描当前路口位置 q ,即得到当前路口在状态转移矩阵中的位置;第 9 行将当前路口 q 与一阶马尔科夫状态转移矩阵相乘,预测得到当前路口的下一个路口 T 。

不确定性车辆路口轨迹预测算法的优势在于,其在保持高预测准确度的同时,还拥有较低的时间复杂度。本文针对因环境等因素造成的历史数据缺失问题,用补全路径的方法完善数据,提高了数据的真实性和预测准确性。低阶马尔科夫链又具有低时间复杂度、高预测准确度的优势,在路径补全算法中,时间复杂度 $O(m)$ 由信息遗漏个数决定,一阶马尔科夫链的时间复杂度 $O(n^2)$ 是行数与列数之积,因此不确定性车辆路口轨迹预测算法的时间复杂度为 $O(n^2 * m)$,其中 n 表示路网中的路口数, m 表示补全路径的条数。

6 实验结果与分析

本节验证本文所提轨迹预测算法的性能。因数据种类不同(基于模式为路径次数,基于向量为速度),选择将其与两种基于模式的预测方法进行对比,设计并实现了不确定性车辆路口轨迹预测算法、挖掘频繁轨迹模式算法^[17]、PrefixTP 轨迹预测算法^[22]。PrefixTP 轨迹预测算法是采用挖掘频繁轨迹模式的方法,通过轨迹特征提取算法对数据集提取轨迹特征点,然后将轨迹数据转化为由特征点表示的特征序列,最后基于特征序列的轨迹预测算法进行预测。实验数据来源于微软 T-Drive 项目,其中包含 2008 年北京 10000 多辆出租车在一周内的轨迹数据,收集了 10340 条真实轨迹数据。从中随机选择 7000 条轨迹数据用于实验,剩下的轨迹数据用于验证轨迹预测的准确度。整个路网的交叉路口有 105 个,路段有 156 条,选取 4900 条长度为 138 的轨迹作为车辆的历史轨迹数据。本节从以下 3 个方面进行对比分析:1)3 种预测算法

的预测准确率和时间性能;2)3种预测算法的车辆连续轨迹的预测准确率和时间性能;3)影响预测准确率的因

表 1 实验数据及其设置

Table 1 Experimental data and setting

参数内容	取值
路网路口总数	105
路网路段总数	156
车辆历史轨迹数据	10340
车辆轨迹长度	138
车辆历史轨迹数据	4900

1)从 105 个路口中随机选择 7 个,使用不确定性车辆路口轨迹预测、挖掘频繁轨迹模式算法、PrefixTP 轨迹预测算法分别进行轨迹预测,实验结果如图 4 所示。可以得出,不确定性车辆路口轨迹预测算法的预测准确率保持在 54.08%左右,比挖掘频繁轨迹模式算法的预测准确率平均提高了 18.8%,比 PrefixTP 轨迹预测算法的预测准确率平均提高了 5.2%。这主要是由于挖掘频繁轨迹模式算法和 PrefixTP 轨迹预测算法在挖掘轨迹模式时带入了过多的数据噪声,检索轨迹模式不能很好地排除噪声影响,导致预测准确率偏低。因此,不确定性车辆路口轨迹预测算法有效地提高了车辆路口轨迹预测的准确率。

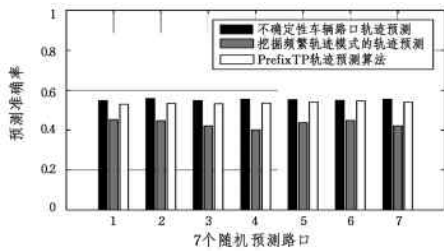


图 4 3种算法的轨迹预测准确率

Fig. 4 Accuracy of trajectory prediction of three algorithms

为了进一步验证算法的性能,检测算法的预测时间,在不同轨迹数量下对 3 种预测方法进行分析,实验结果如图 5 所示。由图 5 可知,不确定性车辆路口轨迹预测算法的预测时间很短,比挖掘频繁轨迹模式算法平均缩减了 80.4%,比 PrefixTP 轨迹预测算法平均缩减了 84.7%。这主要是由于不确定性车辆路口轨迹预测算法利用状态转移矩阵只需要进行一次预测即可,而挖掘频繁轨迹模式算法和 PrefixTP 轨迹预测算法需要索引查询轨迹模式再进行预测,计算量大,因此预测时间长。

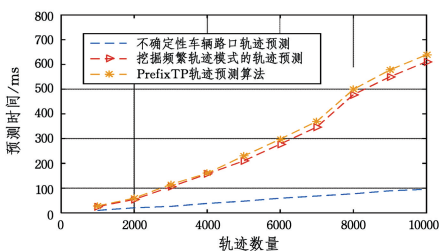


图 5 3种算法的预测时间

Fig. 5 Prediction time of three algorithms

2)从历史轨迹数据中随机选择 5000 条轨迹数据作为预

测样本,使用不确定性车辆路口轨迹预测算法和两种挖掘频繁轨迹模式算法分别预测车辆路口的连续轨迹,实现对车辆未来 1~7 个路口的轨迹预测,实验结果如图 6 所示。由图 6 可知,3 种预测算法的预测准确率随着预测路口数的增加均平缓降低,从而得出,随着预测路口数的增多,路口的每次迭代求解都会降低预测准确率;但不确定性车辆路口轨迹预测算法在每个层次都有较高的预测准确率。

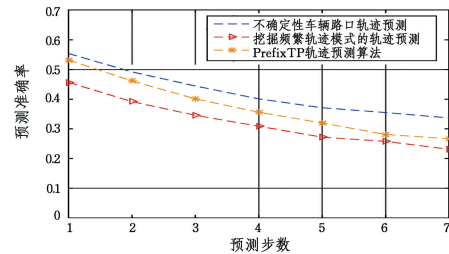


图 6 车辆连续轨迹的预测

Fig. 6 Continuous trajectory prediction of vehicles

3 种算法的时间性能比较如图 7 所示。可以看出,不确定性车辆路口轨迹预测算法的预测时间明显少于挖掘频繁轨迹模式算法和 PrefixTP 轨迹预测算法的预测时间。原因在于:不确定性车辆路口轨迹预测算法在处理多个路口的预测时只须返回前一个路口的位置信息,再由状态转移矩阵求得下一路口,没有多余噪声的影响,时间短、准确率高;而挖掘频繁轨迹模式算法和 PrefixTP 轨迹预测算法都需要重新返回求得特征序列后再进行预测,时间消耗量大,再次求取特征序列也会增加噪声,从而影响预测准确率。因此,不确定性车辆轨迹预测算法在对车辆进行多个路口的预测时不仅具有较高的预测准确率,而且具有较短的预测时间。

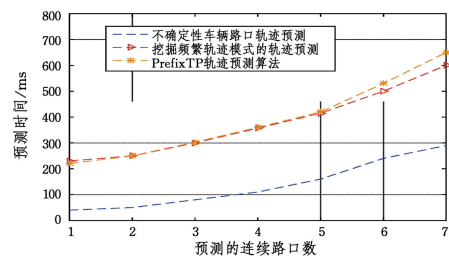


图 7 车辆的连续预测时间

Fig. 7 Continuous forecasting time of vehicles

3)对于当前已知的历史数据,包括路口数量、路段数量、轨迹长度、轨迹重复次数等,其数量越庞大,则车辆的轨迹重复率越高,轨迹的规律性也就越强。从历史数据中选择重复次数小于 50 的轨迹和重复次数大于 200 的轨迹各 50 条作为历史轨迹数据来分别进行轨迹预测,实验结果如图 8 所示。由图可知,重复次数小于 50 的轨迹预测率保持在 42.42%左右,重复次数大于 200 的轨迹预测率保持在 54.76%左右。从历史数据中选择轨迹长度为 20 的轨迹和轨迹长度为 100 的轨迹各 10 条来分别进行轨迹预测,实验结果如图 9 所示。由图可知,轨迹长度为 20 的轨迹预测率保持在 35.72%左右,轨迹长度为 100 的轨迹预测率保持在 50.86%左右。因此,若已知的历史数据量越大,则车辆轨迹的规律性越强,车辆路口的轨迹预测率也就越高。

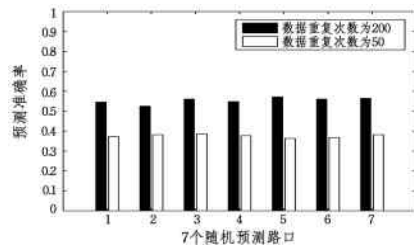


图8 数据重复次数对准确率的影响

Fig. 8 Impact of data repetition on accuracy

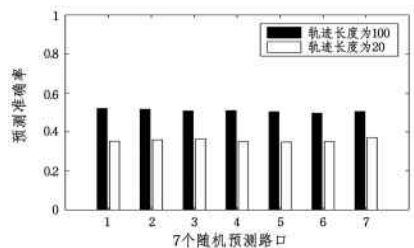


图9 轨迹长度对准确率的影响

Fig. 9 Impact of track length on accuracy

结束语 针对移动对象历史轨迹的不确定性,本文提出了一种不确定性车辆路口轨迹预测算法,不仅解释了不确定性历史轨迹产生的原因并给出了处理方法,还避免了因频繁模式挖掘带来的噪声而影响轨迹预测准确率的问题;在此基础上,使用马尔科夫算法对车辆在路口的未来轨迹进行连续轨迹预测。仿真实验证明,本方法在车辆单一路口预测及路口连续预测方面均具有较高的预测准确率。未来的研究方向包括补全算法的改进、多阶马尔可夫链的应用、移动人群的轨迹预测等。

参考文献

[1] QIAO S J, JIN K, HAN N, et al. Trajectory Prediction Algorithm Based on Gaussian Mixture Model[J]. Journal of Software, 2015, 26(5): 1048-1063. (in Chinese)
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.

[2] ZHENG Y. Trajectory Data Mining: An Overview[J]. ACM Transactions on Intelligent Systems & Technology, 2015, 6(3): 1-41.

[3] BAO J, ZHENG Y, WILKIE D, et al. Recommendations in location-based social networks: a survey[J]. Geoinformatica, 2015, 19(3): 525-565.

[4] YUAN N J, ZHENG Y, XIE X, et al. Discovering urban functional zones using latent activity trajectories[J]. IEEE Transactions on Knowledge & Data Engineering, 2015, 27(3): 712-725.

[5] LV Y, DUAN Y, KANG W, et al. Traffic Flow Prediction With Big Data: A Deep Learning Approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(2): 865-873.

[6] LÓOPEZ-DÍAZ M. Influence Diagrams with Super Value Nodes Involving Imprecise Information[J]. European Journal of Operational Research, 2007, 179(1): 203-219.

[7] DIEHL M, HAIMES Y Y. Influence Diagrams with Multiple Objectives and Trade off Analysis[J]. IEEE Transactions on

Systems, Man, and Cybernetics-Part A: Systems and Humans (S1083-4427), 2004, 34(3): 293-304.

[8] YAO H L, WANG H, ZHANG Y S, et al. Effective graph model selection algorithm based on structural decomposition[J]. Computer Science, 2007, 34(1): 133-139. (in Chinese)
姚宏亮, 王浩, 张佑生, 等. 一种基于结构分解的影响图模型选择算法[J]. 计算机科学, 2007, 34(1): 133-139.

[9] TATMAN J A, SHACHTER R D. Dynamic Programming and Influence Diagrams[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1999, 20(2): 365-379.

[10] SHACHER R D, NDILIKILIKESHA P P. Using Potential Influence Diagrams for Probabilistic Inference and Decision Making[C]// Proceedings of the 9th Conference on Uncertainty and Artificial Intelligence. 1993: 383-390.

[11] SONG C, QU Z, BLUMM N, et al. Limits of predictability in human mobility[J]. Science, 2010, 327(5968): 1018-1021.

[12] YI S G, KANG C M, LEE S H, et al. Vehicle trajectory prediction for adaptive cruise control[C]// Intelligent Vehicles Symposium. IEEE, 2015.

[13] CENTOLA D. The spread of behavior in an online social network experiment[J]. Science, 2010, 329(5996): 1194-1197.

[14] JEUNG H, YIU M L, ZHOU X, et al. Path prediction and predictive range querying in road network databases[J]. The VLDB Journal, 2010, 19(4): 585-602.

[15] QIAO S J, LI T R, ZHU Y, et al. Uncertain Trajectory Prediction of Moving Objects Based on CTBN[J]. Journal of University of Electronic Science and Technology of China, 2012, 41(9): 759-763. (in Chinese)
乔少杰, 李天瑞, 朱焱, 等. 基于 CTBN 的移动对象不确定轨迹预测算法[J]. 电子科技大学学报, 2012, 41(9): 759-763.

[16] XIA Z Q, HU Z Z, LUO J P. EVATP: An Adaptive Vehicle Trajectory Prediction Method[J]. Small Microcomputer System, 2016, 37(10): 2375-2379. (in Chinese)
夏卓群, 胡珍珍, 罗君鹏. EVATP: 一种环境自适应车辆轨迹预测方法[J]. 小型微型计算机系统, 2016, 37(10): 2375-2379.

[17] GUO L M, DING Z M, HU Z L, et al. Study on Uncertain Trajectory Prediction Based on Road Network [J]. Journal of Computer Research and Development, 2010, 47(1): 104-112. (in Chinese)
郭黎敏, 丁治明, 胡泽林, 等. 基于路网的不确定性轨迹预测[J]. 计算机研究与发展, 2010, 47(1): 104-112.

[18] 盛骤. 概率论与数理统计[M]. 北京: 高等教育出版社, 2003.

[19] 刘嘉焜. 应用概率统计[M]. 北京: 科学出版社, 2004.

[20] 龚光鲁, 钱敏平. 应用随机过程教程[M]. 北京: 清华大学出版社, 2004.

[21] SONG L B, KOTZ D, JAIN R, et al. Evaluating next-cell predictors with extensive Wi-Fi mobility data[J]. IEEE Transactions on Mobile Computing, 2006, 5(12): 1633-1649.

[22] WANG X T. Study on Uncertainty Trajectory Prediction Model of Moving Object [D]. Chengdu: Southwest Jiaotong University, 2016. (in Chinese)
王晓腾. 移动对象不确定性轨迹预测模型研究[D]. 成都: 西南交通大学, 2016.