

推论 1. 轨迹 α 称为频繁轨迹,当且仅当轨迹满足引理 1 和引理 2 给出的递归性质.

结合上述定义和性质,轨迹预测问题定义如下.

定义 10(轨迹预测). 轨迹预测问题定义为求以待预测轨迹序列后序轨迹作为前序轨迹的频繁轨迹序列.

由于引理 2 给出了频繁轨迹挖掘递归过程的原理,因此对于轨迹序列数据,其频繁轨迹挖掘过程可以根据以下步骤进行.

- 第 1 步:首先考虑其每一项在训练数据集中出现的次数,对数据集进行一步扫描,得到所有出现次数满足最小支持度的项,记为 F .
- 第 2 步:划分搜索空间.在第 1 步中得到了长度为 1 的频繁项集合 F .使用 F ,可以将轨迹序列集 S 划分成 $|F|$ 个相互独立的子集.其中,每个子集中的所有元素都对应一个频繁 1 项,且该子集中的所有元素都以该频繁 1 项为前序轨迹.
- 第 3 步:挖掘子集中序列模式.子集中序列模式可以通过构造相应的投影集来进行递归挖掘.若有投影集 $S|\alpha$,那么 α 的增长项必然出现在投影集 $S|\alpha$ 中,每次增长长度为 1.遍历投影集 $S|\alpha$,计算每一序列中第一项的支持度,取满足最小支持度的项作为 α 的增长项,同时将该项对应的序列进行再次划分.
- 重复第 3 步.

例 1(频繁轨迹挖掘示例):假设在轨迹数据库中存在有如表 1 所示的轨迹序列集合,设最小支持度为 2.

在轨迹序列集中存在 $\{a,b,c,d,e,f\}$ 项,各项在数据集中的支持度见表 2,其中, $\{f\}$ 项的支持度为 1,不满足最小支持度的要求,因此被舍弃;余下的 $\{a,b,c,d,e\}$ 项满足最小支持度要求,因此可以作为频繁轨迹的起始项进行下一步挖掘,对每一项构建投影集.

Table 1 Example of trajectory sequences

表 1 轨迹序列举例

轨迹 ID	轨迹序列
1	$\langle a b d e \rangle$
2	$\langle a b d c \rangle$
3	$\langle a c k e \rangle$
4	$\langle b c a d e f \rangle$
5	$\langle d a b c e \rangle$
6	$\langle e b a d e \rangle$

Table 2 Support counts of trajectory items

表 2 轨迹项支持度计数值

项	支持度
a	6
b	5
c	4
d	5
e	5
f	1

以 $\langle a \rangle$ 的投影集及其挖掘过程中频繁轨迹序列的增长过程为例,见表 3.频繁轨迹挖掘过程递归进行,第 1 次挖掘首先以 $\langle a \rangle$ 为频繁轨迹作为输入,计算得到其投影集为 $\{\langle b d e \rangle, \langle b d c \rangle, \langle b c e \rangle, \langle c k e \rangle, \langle d e f \rangle, \langle d e \rangle\}$,然后计算投影集中各序列第 1 项的支持度,取满足最小支持度的项作为频繁序列增长项,同时将该项对应的投影集中的项带入下一次迭代,作为其投影集.具体来说,在第 1 次迭代中, $\langle b \rangle$ 和 $\langle d \rangle$ 满足最小支持度要求,因此分别作为频繁轨迹增长项,进行第 2 次迭代.在表 3 中,第 2 次迭代被分为以 $\langle a b \rangle$ 和 $\langle a d \rangle$ 为频繁轨迹分别进行挖掘的过程,其中, $\langle a d \rangle$ 的投影集由第 1 次迭代中以 $\langle b \rangle$ 为前序轨迹的轨迹项组成.以此类推,最终当挖掘过程进行到第 3 次迭代时,投影集中已无满足最小支持度的轨迹项,因此挖掘过程结束.

Table 3 Example of mining frequent trajectory sequences

表 3 频繁轨迹序列挖掘举例

次数	1		2		3	
	频繁轨迹	投影集	频繁轨迹	投影集	频繁轨迹	投影集
过程	$\langle a \rangle$	$\langle b d e \rangle, \langle b d c \rangle$ $\langle b c e \rangle, \langle c k e \rangle$ $\langle d e f \rangle, \langle d e \rangle$ 支持度 $b:3, d:2, c:1$	$\langle a b \rangle$	$\langle d e \rangle, \langle d c \rangle, \langle c e \rangle$ 支持度 $d:2, c:1$	$\langle a b d \rangle$	$\langle e \rangle, \langle c \rangle$ 支持度 $e:1, c:1$
			$\langle a d \rangle$	$\langle e f \rangle, \langle e \rangle$ 支持度 $e:2$	$\langle a d e \rangle$	$\langle f \rangle$ 支持度 $f:1$

例 1 中展示的频繁轨迹挖掘过程最终产生了 5 条频繁轨迹,分别是 $\langle a \rangle, \langle a b \rangle, \langle a d \rangle, \langle a b d \rangle, \langle a d e \rangle$.

完成频繁轨迹挖掘后,轨迹预测过程即可简化为对频繁轨迹进行匹配查找.以例 1 中表 1 数据为例,例 2 给出了轨迹预测的具体过程.

例 2(轨迹预测示例):设待预测轨迹序列为 $\langle e c a b \rangle$,首先在频繁轨迹集合中查找以 $\langle e c a b \rangle$ 为前序轨迹的频繁轨迹,在集合中未找到,因此缩短待预测序列,取 $\langle e c a b \rangle$ 的后序轨迹序列 $\langle c a b \rangle$ 为待预测轨迹序列重复上一步骤,直至取 $\langle a b \rangle$ 作为待预测序列,通过匹配查询,找到频繁轨迹序列 $\langle a b d \rangle$.频繁轨迹 $\langle a b d \rangle$ 以待预测轨迹序列 $\langle e c a b \rangle$ 后序轨迹 $\langle a b \rangle$ 为前序轨迹,满足定义给出的描述,因此取 $\langle d \rangle$ 作为轨迹预测结果(如图 1 所示).

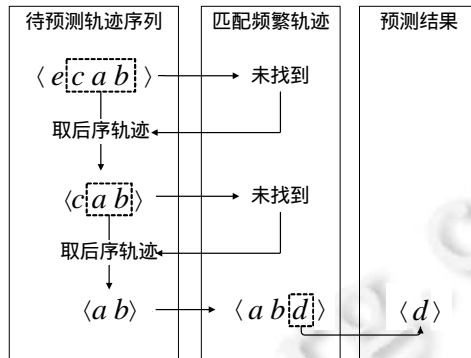


Fig.1 Example of trajectory prediction

图 1 轨迹预测示例

在进行轨迹预测时,可以采用两种轨迹匹配方法:(1) 精确匹配,从频繁轨迹集合中找到包含待预测轨迹后序轨迹的频繁轨迹;(2) 模糊匹配,从挖掘的频繁轨迹集合中找到所有将待预测轨迹后序轨迹作为子序列的所有频繁轨迹.本文所提出的预测算法采用精确匹配,针对待预测轨迹给出一条最可能的轨迹路线.

3 轨迹预测框架

轨迹预测功能的实现需要一系列相应辅助模块共同协作,本文所提出的 PPTP 轨迹预测算法也不例外.为了更好地进行轨迹预测,本文提出一种新的轨迹预测框架,其内部工作原理如图 2 所示.

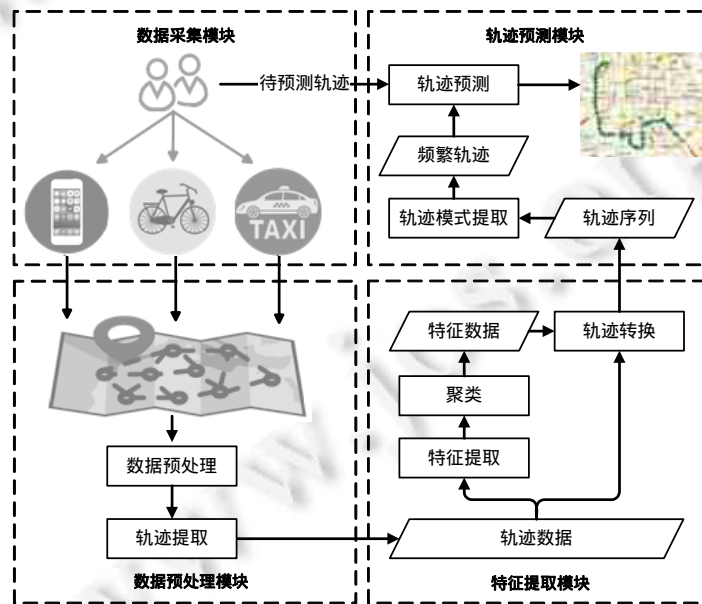


Fig.2 Architecture of trajectory prediction system

图 2 轨迹预测系统框架图

数据采集模块首先完成对用户数据的采集工作,然后将原始数据交由数据预处理模块;数据预处理模块将原始轨迹数据中的噪声去除,同时应用轨迹提取算法从中提取出轨迹数据,并按照时间、距离等条件进行轨迹分割,最终形成格式统一的轨迹数据;在特征提取模块中,首先对轨迹数据按角度变化特性进行特征点提取、聚类等操作,然后使用提取的特征数据对轨迹进行转换,从而得到轨迹序列数据;在最终的预测模块,算法利用轨迹序列数据进行轨迹模式提取,得到频繁轨迹作为预测所需的候选数据,然后根据用户的实时轨迹数据,使用PPTP轨迹预测算法完成最终的预测.

4 轨迹预测算法及性能分析

4.1 算法描述

前面通过举例说明了基于前缀投影技术的大规模轨迹预测算法的思想及工作原理,下面给出算法的形式化描述,见算法1.表4给出算法中使用参数的说明.

算法1. 基于前缀投影技术的增量式轨迹预测算法——PPTP.

输入:待预测序列 t , 预测步数 n .

输出:预测结果.

```

1.  $F \leftarrow \text{getPrefix}(t[0]);$ 
2. IF  $F = \emptyset$ 
3.   THEN RETURN  $\text{PPTP}(t.\text{suffix}(1), n);$ 
4. END IF
5.  $F' \leftarrow \emptyset;$ 
6. FOR EACH  $f$  IN  $F$ 
7.   IF  $f.\text{len} < t.\text{len} + n$ 
8.     THEN CONTINUE;
9.   END IF
10.   $i \leftarrow 0;$ 
11.  WHILE  $i < t.\text{len}$ 
12.    IF  $f[i] \neq t[i]$ 
13.      THEN BREAK;
14.    END IF
15.     $i \leftarrow i + 1;$ 
16.  END WHILE
17.  IF  $i = t.\text{len}$ 
18.    THEN  $F'.\text{add}(f);$ 
19.  END IF
20. END FOR
21.  $F'.\text{sort}();$ 
22. IF  $F' = \emptyset$ 
23.   THEN RETURN  $\text{PPTP}(t.\text{suffix}(1), n);$ 
24. END IF
25. RETURN  $F'[0].\text{get}(t.\text{len} + n - 1);$ 

```

Table 4 Parameter introduction of the proposed algorithm

表4 算法参数说明

参数	说明
t	待预测轨迹序列
n	预测步数
F	频繁轨迹集合
F'	满足条件的频繁轨迹集合

算法第 1 行~第 4 行获取待预测序列 t 的第 1 项,调用 `getPrefix` 函数获取以该项为前缀的频繁模式,若获取失败,则以 t 的第 2 项开始的子串作为输入,递归调用 PPTP 方法;第 6 行~第 20 行遍历之前获取到的 F 集合,其中,第 7 行~第 9 行判断 f 的长度是否大于待预测序列 t 的长度与预测步数之和,如果不满足,则说明 f 无法提供足够长度的预测步数,因此跳过 f ;第 11 行~第 16 行逐项匹配 f 和 t ,若失败,则跳出循环;第 17 行~第 19 行判断 f 的长度,若满足预测条件,则将 f 加入结果集;第 22 行~第 24 行判断 F' 是否为空,若为空,则以 t 的子串作为输入,递归调用 PPTP 方法;第 25 行返回 P' 中第 1 项的相应步作为预测结果.

4.2 算法性能分析

通过分析算法 1,可以得出其时间复杂度为 $O(m^n)$,其中, m 表示频繁轨迹数量; n 表示预测步数,即算法迭代次数.通过第 5.3 节图 5 给出的不同预测步数下算法轨迹预测准确性对比实验得到的结论可知:预测步数通常介于 2 步~4 步之间可以保证较高的预测准确性,而频繁轨迹的数量远远小于原始轨迹数据,因此算法 1 的整体复杂性不高.算法的空间复杂度为 $O(k)$,其中, k 表示所有轨迹点的数量.

算法 1 的正确性和预测结果完整性可以通过引理 1 和引理 2 得到证明,分析算法时间性能得到如下结论.

- (1) 算法不会生成多余的候选序列模式.PPTP 算法仅通过较短的频繁序列模式以增量式方式生成轨迹模式,不会产生投影数据库中不存在的候选轨迹序列模式.
- (2) 投影数据库的规模不断缩小.通过表 3 可以发现,经过 3 次迭代操作,投影数据库中项的规模不断缩小,因为其中的候选项是从频繁轨迹序列扩展生成的后序轨迹序列.
- (3) 算法的主要时间开销是投影数据库的构建.最坏的情况是,PPTP 算法对每一个轨迹序列模式构建一个投影数据库,如果候选频繁轨迹序列模式规模不是很大,则时间代价不会很高.

5 实验及算法性能分析

5.1 实验环境及数据集描述

本实验所使用的轨迹数据集来源于微软亚洲研究院郑宇研究员所领导的 T-Driver 项目^[21],数据采集自北京市真实路网中的出租车 GPS 设备,包含 10 357 辆出租车超过 1 周的行驶轨迹数据.该轨迹数据集中的轨迹点总量超过 15 000 000,总行驶长度超过 9 000 000km,具体描述见表 5.

Table 5 Description of experimental datasets

表 5 实验数据集描述

参数	值
轨迹时间跨度	2008/02/02~2012/02/08
车辆数	10 357
轨迹数量	>25000
轨迹点数量	>15000000
总长度	>9000000km

本文中所提到的算法均采用 Java 程序设计语言实现,使用 Eclipse Juno 作为开发环境,实验硬件平台为: Intel(R) Core(TM)2 Duo P8700 2.53GHz CPU,3GB 内存,操作系统平台为 Windows 7.实验通过对比实验评价所提出方法的性能优劣,对比算法包括:采用 1 阶马尔可夫链轨迹预测算法^[4](当前位置由前一个位置点决定)、2 阶马尔可夫链轨迹预测算法^[22](代表高阶马尔可夫链,当前位置由前面两个位置点确定,本文以 2 阶马尔可夫链为例,更高阶马尔可夫链情况类似)和本文提出的基于前缀投影技术的 PPTP 轨迹预测算法.实验中针对不同数据集随机选取 90% 作为训练数据,其余 10% 的轨迹数据用作测试数据.

5.2 性能评价指标

本文提出的轨迹预测算法工作过程包括:首先,在数据集上使用轨迹特征提取算法,提取轨迹特征点;然后,通过轨迹转换技术将 GPS 轨迹数据转化为由轨迹特征点表示的特征序列;最后,使用基于特征序列的轨迹预测

算法进行预测.为了方便量化表示算法的性能优势,本文采用如下性能评价指标^[2].

定义 11(预测命中). 已知轨迹序列 $T=\{T_1, T_2, \dots, T_k\}$, 预测轨迹序列 $P=\{P_1, P_2, \dots, P_n\}$, $k < n$, $dist(m, n)$ 表示时空轨迹点 m 和 n 间的欧氏距离, δ 表示距离阈值, 则 $dist(T_i, P_i) < \delta$ 时, 表示预测命中, 定义为

$$H(T_i, P_i) = \begin{cases} 1, & dist(T_i, P_i) < \delta \\ 0, & dist(T_i, P_i) > \delta \end{cases} \quad (3)$$

定义 12(预测准确率). 已知轨迹序列 T , 预测轨迹序列 P , 则预测准确率定义为

$$Accuracy = \frac{\sum_{i=1}^n H(T_i, P_i)}{|P|} \quad (4)$$

其中, $|P|$ 表示预测轨迹序列的长度.

5.3 轨迹预测准确性对比

本节实验首先观察轨迹预测算法在不同规模数据集上的预测准确性, 图 3(a)、图 3(b) 分别展示了在小规模和大规模轨迹数据集上进行的对比实验结果. 其中, 横轴表示训练轨迹的数量, 纵轴表示预测准确率 Accuracy.

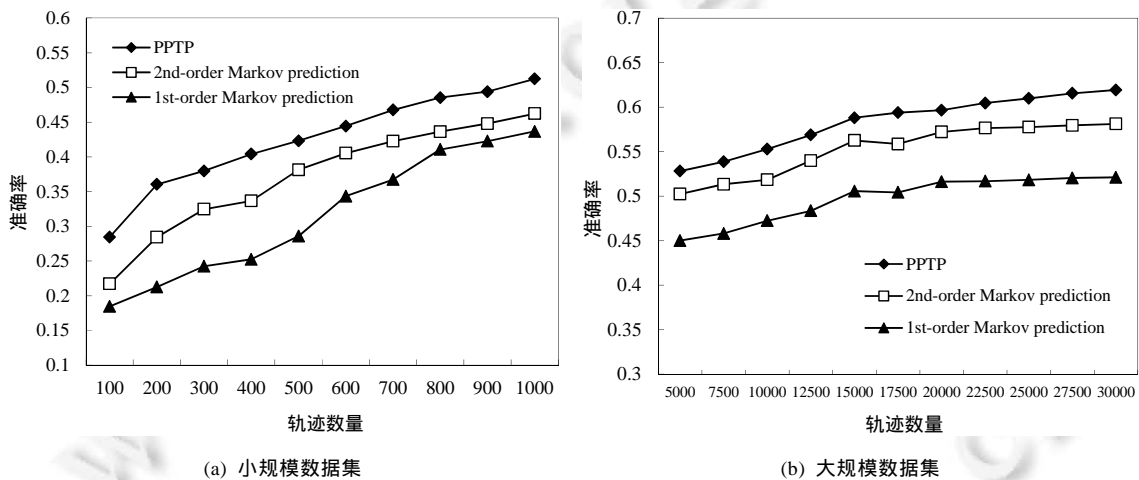


Fig.3 Prediction accuracy comparison with different number of trajectories

图 3 不同数量训练轨迹下预测准确性对比

实验所用训练数据规模依次递增, 通过观察发现:

- (1) 本文提出的 PPTP 轨迹预测算法在预测准确率上明显优于其他两种算法. 在小规模数据集图 3(a) 上实验, 相对于 2 阶马尔可夫链和 1 阶马尔可夫链轨迹预测算法平均提高了 15.8% 和 39.8%; 在大规模数据集图 3(b) 上实验, 相对于 2 阶马尔可夫链和 1 阶马尔可夫链轨迹预测算法平均提高了 5.5% 和 17.4%. 主要原因在于: 1 阶马尔可夫链预测算法和 2 阶马尔可夫链预测算法仅考虑了轨迹序列中较短项之间的相互影响, 对训练数据的利用不够全面, 因此当训练数据量不断增加时, 其预测准确率上升空间有限. 反观本文所提出的 PPTP 轨迹预测算法, 充分利用了训练数据集, 将不同长度的训练数据都作为依据, 对训练数据的使用更为全面, 因此具有更好的预测准确性.
- (2) 随着训练数据集规模的逐渐增大, 3 种轨迹预测算法的预测准确率均不断提高, PPTP 算法比较稳定, 尤其是在大规模训练数据集下, 准确性保持在一个较高的水平, 且增长趋势比较平缓; 而其他两种基于马尔可夫链的预测算法均会出现波动, 进一步说明本文提出的基于前缀投影技术的增量式轨迹预测算法具有较好的稳定性.

为了更加全面地展示 PPTP 轨迹预测算法对不同数据集的适应性, 使用 8 种不同规模的数据集进行如下实

图4展示了3种轨迹预测算法在不同数据集上的预测准确率,通过对比可以发现,PPTP 轨迹预测算法的预测准确率均优于其他两种算法,相对于 2 阶马尔可夫链和 1 阶马尔可夫链轨迹预测算法,平均提高了 7.1%和 19.9%。进一步证明了本文提出的算法不依赖于训练数据集,具有很好的普适性,且 PPTP 算法的预测准确性均高于其他两种算法,原因与上面的描述相同。

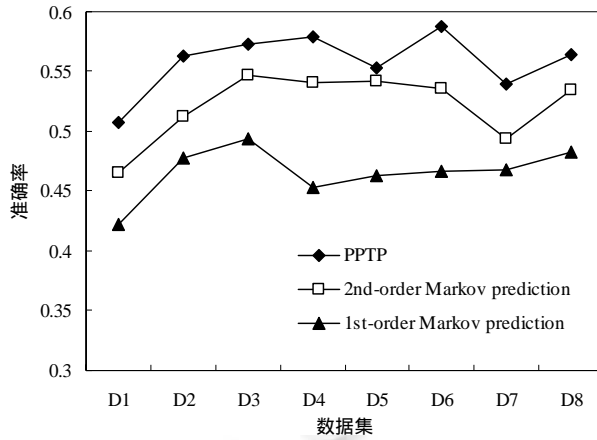


Fig.4 Prediction accuracy comparison with different datasets
图 4 不同数据集下算法预测准确率对比

在轨迹预测问题中,未来 n 步位置预测也是重点考虑的问题。 n 步预测指的是以当前位置为起始,预测 n 步后移动对象的位置,可以验证算法的长轨迹预测能力。针对这一问题,使用上述 3 种算法进行对比验证,实验结果如图 5 所示。

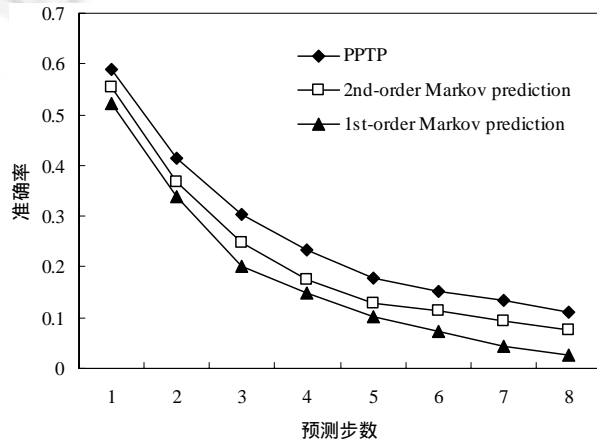


Fig.5 Relationship between prediction steps and accuracy
图 5 预测步数与准确率的关系

通过观察可以发现,随着预测步数的增加,3 种算法的预测准确率均有所下降,但 PPTP 算法在各个阶段均优于其他两种算法。原因在于,在进行 n 步预测对比实验时,基于马尔可夫链的对比算法采用 n 步转移概率,即条件概率 $p_{ij}^{(n)} = P\{X_{m+n}=j|X_m=i\} (i,j \in I, m \geq 0, n \geq 1)$ 来计算其 n 步转移矩阵 $P^{(n)} = (p_{ij}^{(n)})$ 。相对于 PPTP 算法使用可变长模式序列进行预测的方式,对比算法仅考虑了固定长度的训练轨迹预测的作用,对于数据的适应性较差,因此在实验中,PPTP 轨迹预测算法的预测准确率相对较高。

在 PPTP 轨迹预测算法中,支持度表示训练数据中某一序列模式出现的次数,该参数的选择对预测准确率

有一定影响,实验针对支持度的选择问题,在不同数据集上进行对比实验,实验结果如图 6 所示.通过观察实验结果可以发现,预测准确率随支持度的不同而产生波动,在所有实验数据集上都表现出了先上升后下降的趋势.当支持度位于 2~4 区间时,算法在所有实验数据集上都表现出了较高的预测准确率,因此可以认为在实验所用数据集上进行实验,支持度在 2~4 区间内选择可以获得最佳的预测准确率.

由于支持度这一概念本身与训练数据集的规模具有较大关联,因此实验本身并不能给出十分精确的参数选取范围.结合实验结果和支持度的定义可以总结出结论:当训练数据集的规模有显著提升时,支持度的选取范围也应该相应提高.

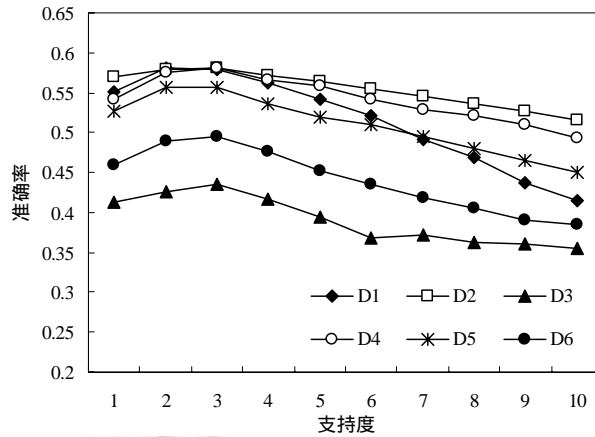


Fig.6 Effect of support counts on prediction accuracy

图 6 支持度对准确率的影响

5.4 轨迹预测时间性能对比

为了验证 PPTP 算法在时间上相对于其他算法的优劣,进行了时间性能对比实验.图 7 给出了 3 种算法在轨迹数量不断递增情况下的时间消耗情况.

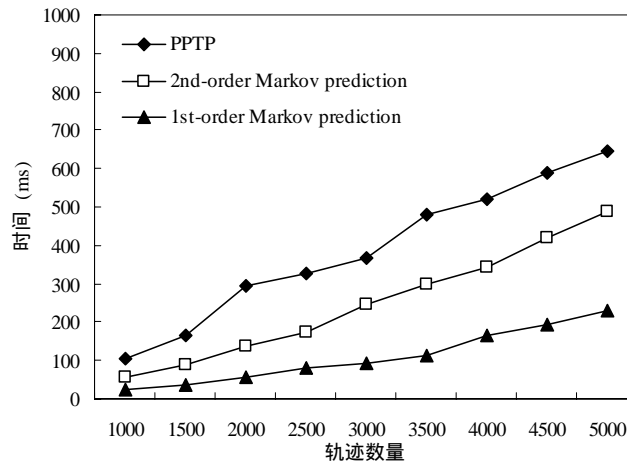


Fig.7 Prediction time comparison of algorithms with different number of trajectory

图 7 不同轨迹数量下算法预测时间比较

通过观察可以发现,PPTP 轨迹预测算法相对于其他两种算法,在模型训练过程中要花费相对多一些的时间,因为需要构建投影数据库.PPTP 算法相对于 2 阶和 1 阶马尔可夫链轨迹预测算法平均高出 0.138s 和 0.277s,

总体时间开销差异维持在 ms 级别,即使在实时性要求较高的轨迹预测系统中,这一差异也是可以接受的.

3 种算法在不同数据集上的时间代价如图 8 所示,PPTP 轨迹预测算法的时间代价略高于其他两种对比算法,但是总体上维持在 ms 级的差异.主要原因在于,构建前缀投影数据库的时间开销略大于马尔可夫链预测算法中计算条件概率转移矩阵的时间代价.

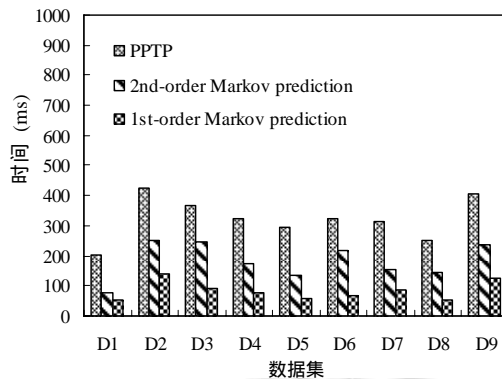


Fig.8 Prediction time comparison of algorithms on different datasets

图 8 不同数据集上算法预测时间比较

5.5 轨迹预测系统展示

在本文提出的 PPTP 算法的基础上设计实现了轨迹预测可视化系统,该系统包括 3 个主要功能模块:轨迹数据加载、数据可视化模块和轨迹预测模块.该系统是普适的轨迹预测系统,可以按照数据格式加载不同的地图信息和轨迹数据集.这里以轨迹预测模块为例,介绍系统主要功能.

轨迹预测模块以前面所提到的增量式前缀投影技术为理论基础,提供轨迹实时预测功能.模块分为训练部分和预测部分,训练部分首先对轨迹数据进行特征提取、聚类等操作,训练过程所得结果存储于移动数据库中.预测算法使用训练结果作为依据,结合当前输入轨迹进行预测.轨迹预测模块的功能界面如图 9 所示.

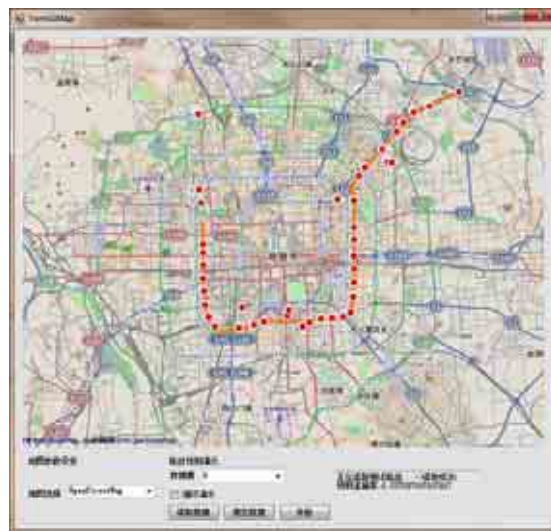


Fig.9 Graphical user interface of trajectory prediction system and visualization results

图 9 轨迹预测系统图形化用户界面及可视化结果

可视化界面中,主要包括地图选择、数据加载、清除、轨迹预测等功能,在加载数据并开始执行预测后,预测详细信息由右侧窗口打印输出至界面,直观显示给客户。

如图 9 所示,用户可以通过左下方“地图选择”功能加载需要的地图,通过点击“读取数据”按钮添加不同的轨迹数据集,点击“开始”按钮完成训练和预测的过程,轨迹预测准确率结果展示在右下方的文本框区域内,并在地图上可视化输出一条供用户参考的最佳路线。

6 结束语

针对海量移动对象轨迹数据,结合频繁序列模式挖掘算法,本文提出了一种面向大规模位置数据的前缀投影轨迹预测模型,提出前序、后序轨迹和投影数据库的概念,并给出前缀投影轨迹序列模式挖掘的相关性质和定理.利用真实轨迹数据进行多角度实验,对算法性能进行全面检验.本文提出的轨迹预测算法在实验中表现出较高的预测准确率,相对于 1 阶和高阶马尔可夫链轨迹预测算法,其平均预测准确率可以得到提升.以本文提出的预测模型为理论依据,开发了一个普适的轨迹预测系统,提供轨迹可视化、轨迹预测等功能。

未来的研究工作包括:(1) 结合实时交通情况,如交通拥堵、天气变化等因素,改进轨迹预测算法,提供更为准确的轨迹预测算法;(2) 引入社交数据,分析个体用户的出行模式,对用户进行聚类分析,提取群体的出行特征,以此为辅助改进轨迹预测算法;(3) 优化轨迹预测算法的时间效率,可对部分数据采用预计算的方式合理安排计算策略,从而降低实时计算量,提高预测效率。

References:

- [1] Guo C, Liu JN, Fang Y, Luo M, Cui JS. Value extraction and collaborative mining methods for location big data. Ruan Jian Xue Bao/Journal of Software, 2014,25(4):713–730 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4570.htm> [doi: 10.13328/j.cnki.jos.004570]
- [2] Qiao SJ, Han N, Zhu W, Gutierrez LA. TraPlan: An effective three-in-one trajectory prediction model in transportation networks. IEEE Trans. on Intelligent Transportation Systems, 2015,16(3):1188–1198. [doi: 10.1109/TITS.2014.2353302]
- [3] Yuan J, Zheng Y, Xie X, Sun G. T-Drive: Enhancing driving directions with taxi drivers' intelligence. IEEE Trans. on Knowledge and Data Engineering, 2013,25(1):220–232. [doi: 10.1109/TKDE.2011.200]
- [4] Qiao SJ, Shen DY, Wang XT, Han N, Zhu W. A self-adaptive parameter selection trajectory prediction approach via hidden markov models. IEEE Trans. on Intelligent Transportation Systems, 2015,16(1):284–296. [doi: 10.1109/TITS.2014.2331758]
- [5] Zheng Y. Trajectory data mining: an overview. ACM Trans. on Intelligent Systems and Technology, 2015,6(3):Article 29. [doi: 10.1145/2743025]
- [6] Bao J, Zheng Y, Wilkie D, Mokbel M. Recommendations in location-based social networks: A survey. Geoinformatica, 2015,19(3): 525–565. [doi: 10.1007/s10707-014-0220-8]
- [7] Zheng K, Zheng Y, Yuan J, Shang S, Zhou XF. Online discovery of gathering patterns over trajectories. IEEE Trans. on Knowledge and Data Engineering, 2014,26(8):1974–1988. [doi: 10.1109/TKDE.2013.160]
- [8] Prentow T, Thom A, Blunck H, Vahrenhold J. Making sense of trajectory data in indoor spaces. In: Proc. of the 16th IEEE Int'l Conf. on Mobile Data Management. Washington: IEEE, 2015. 116–121. [doi: 10.1109/MDM.2015.44]
- [9] Dai J, Ding Z, Xu J. Context-Based moving object trajectory uncertainty reduction and ranking in road network. Journal of Computer Science and Technology, 2016,31(1):167–184. [doi: 10.1007/s11390-016-1619-5]
- [10] Tao Y, Faloutsos C, Papadias D, Liu B. Prediction and indexing of moving objects with unknown motion patterns. In: Proc. of the ACM SIGMOD Int'l Conf. on Management of Data. New York: ACM Press, 2004. 611–622. [doi: 10.1145/1007568.1007637]
- [11] Morzy M. Mining frequent trajectories of moving objects for location prediction. In: Proc. of 5th Int'l Conf. on Machine Learning and Data Mining in Pattern Recognition. Berlin: Springer-Verlag, 2007. 667–680. [doi: 10.1007/978-3-540-73499-4_50]
- [12] Song C, Qu Z, Blumm N, Barabási AL. Limits of predictability in human mobility. Science, 2010,327(5968):1018–1021. [doi: 10.1126/science.1177170]
- [13] Pentland A. Society's nervous system: Building effective government, energy, and public health systems. IEEE Computer, 2012, 45(1):31–38. [doi: 10.1109/MC.2011.299]
- [14] Ding Z, Yang B, Güting RH, Li Y. Network-Matched trajectory-based moving-object database: Models and applications. IEEE Trans. on Intelligent Transportation Systems, 2015,16(4):1918–1928. [doi: 10.1109/TITS.2014.2383494]

- [15] Qiao SJ, Jin K, Han N, Tang CJ, Gesangduoji, Gutierrez LA. Trajectory prediction algorithm based on Gaussian mixture model. *Ruan Jian Xue Bao/Journal of Software*, 2015,26(5):1048–1063 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4796.htm> [doi: 10.13328/j.cnki.jos.004796]
- [16] Dai J, Yang B, Guo C, Ding Z. Personalized route recommendation using big trajectory data. In: *Proc. of the 31st IEEE Int'l Conf. on Data Engineering*. Washington: IEEE, 2015. 543–554. [doi: 10.1109/ICDE.2015.7113313]
- [17] Yuan G, Zhao J, Xia S, Zhang Y, Li W. Multi-Granularity periodic activity discovery for moving objects. *Int'l Journal of Geographical Information Science*, 2017,31(3):435–462. [doi: 10.1080/13658816.2016.1205194]
- [18] Zhang L, Liu L, Xia Z, Li W, Fan Q. Sparse trajectory prediction based on multiple entropy measures. *Entropy*, 2016,18(9):No.327. [doi: 10.3390/e18090327]
- [19] Chaulwar A, Botsch M, Utschick W. A hybrid machine learning approach for planning safe trajectories in complex traffic-scenarios. In: *Proc. of 15th IEEE Int'l Conf. on Machine Learning and Applications*. Washington: IEEE, 2016. 540–546. [doi: 10.1109/ICMLA.2016.0095]
- [20] Pei J, Han J, Mortazavi-Asl B, Wang J, Pinto H, Chen Q, Dayal U, Hsu M. Mining sequential patterns by pattern-growth: The prefixspan approach. *IEEE Trans. on Knowledge and Data Engineering*, 2004,16(11):1424–1440. [doi: 10.1109/TKDE.2004.77]
- [21] Yuan J, Zheng Y, Zhang C, Xie W, Xie X, Sun G, Huang Y. T-Drive: Driving directions based on taxi trajectories. In: *Proc. of the 18th SIGSPATIAL Int'l Conf. on Advances in Geographic Information Systems*. New York: ACM Press, 2010. 99–108. [doi: 10.1145/1869790.1869807]
- [22] Gambs S, Killijian M, Cortez DP, Miguel N. Next place prediction using mobility Markov chains. In: *Proc. of the 1st Workshop Measurement, Privacy, and Mobility*. New York: ACM Press, 2012. [doi: 10.1145/2181196.2181199]

附中文参考文献:

- [1] 郭迟,刘经南,方媛,罗梦,崔竞松.位置大数据的价值提取与协同挖掘方法. *软件学报*,2014,25(4):713–730. <http://www.jos.org.cn/1000-9825/25/713.htm> [doi: 10.13328/j.cnki.jos.004570]
- [15] 乔少杰,金琨,韩楠,唐常杰,格桑多吉,Gutierrez LA. 一种基于高斯混合模型的轨迹预测算法. *软件学报*,2015,26(5):1048–1063. <http://www.jos.org.cn/1000-9825/4796.htm> [doi: 10.13328/j.cnki.jos.004796]



乔少杰(1981 -),男,山东招远人,博士,教授,CCF 高级会员,主要研究领域为轨迹数据挖掘,机器学习.



李斌勇(1982 -),男,博士,讲师,CCF 专业会员,主要研究领域为大数据,云服务.



韩楠(1984 -),女,博士,讲师,主要研究领域为移动对象数据库.



王晓腾(1991 -),男,硕士,主要研究领域为移动对象数据库,轨迹预测.



李天瑞(1969 -),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为智能信息处理,数据挖掘.



Luis Alberto GUTIERREZ(1980 -),男,博士,Researcher,主要研究领域为数据挖掘.



李荣华(1985 -),男,博士,助理教授,CCF 专业会员,主要研究领域为图数据挖掘,机器学习.