

基于运动趋势的移动对象位置预测

李雯¹, 夏士雄¹, 刘峰², 张磊¹, 袁冠¹

(1. 中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116; 2. 中国煤炭工业协会, 北京 100713)

摘要: 提出一种基于运动趋势的移动对象位置预测算法, 不仅借鉴马尔可夫模型思想对移动对象的历史活动轨迹进行建模, 而且将对象的运动趋势作为位置预测的重要因素。以全体历史停留区域作为未来位置的候选, 根据位置的特征, 将结果分为预测位置和推荐位置。真实数据实验表明, 算法在保证较高时间效率的同时, 预测精度较同类算法提高近 10%。

关键词: 位置预测; 移动对象; 马尔可夫模型; 运动趋势; 区域可达性

中图分类号: TP311

文献标识码: A

文章编号: 1000-436X(2014)02-0046-08

Location prediction algorithm based on movement tendency

LI Wen¹, XIA Shi-xiong¹, LIU Feng², ZHANG Lei¹, YUAN Guan¹

(1. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China;

2. China National Coal Association, Beijing 100713, China)

Abstract: A location prediction algorithm based on movement tendency (LP-MT) was proposed, which not only builds moving object's historical activity model borrowing from Markov thinking, but also added movement tendency as an important reference of location prediction. Besides, each of the historical stay regions was considered as a candidate of future location. According to the characteristics of the future location, algorithm results were divided into prediction locations and recommendation locations. Experiment results on real dataset show LP-MT method can ensure good efficiency and improve the accuracy by 10% compared with the existing algorithms.

Key words: location prediction; moving object; Markov model; movement tendency; region reach-ability

1 引言

移动对象空间定位技术的快速发展极大地推动了基于位置服务的广泛应用, 为了使服务具有前瞻性, 不仅要移动对象当前活动的位置进行分析, 更要能够对其位置进行预测。移动对象位置预测是一项具有挑战性的工作, 主要原因如下: 由于 GPS 等定位设备具有采样盲点, 移动对象在封闭空间内的位置信息无法被捕获, 造成移动对象轨迹数据的部分缺失, 这些缺失的部分往往可能是移动对象活动较为重要的区域, 因此, 轨迹数据的缺失和不完整, 大大增加了移动对象活动的预测难度; 移

动对象的运动具有不确定性, 不同周期内移动对象的活动模式不完全相同; 移动对象轨迹的时空信息具有离散性和琐碎性, 无法将原始轨迹直接用于移动对象未来位置的预测。

现有的位置预测方法大多在移动对象历史行为模式^[1,2]或关联规则^[3]的基础上进行位置预测。文献[1]提出基于轨迹模式的位置预测方法 WhereNext, 考虑移动对象的群体模式, 却忽略了对对象间的相似性以及相似对象活动模式间的局部匹配性, 预测准确性不高。文献[3]使用匹配函数从提取的关联规则中发现最佳规则, 进而对位置进行预测。文献[4]中提出 HDTP 算法考虑移动对象位

收稿日期: 2012-05-18; 修回日期: 2012-12-25

基金项目: 教育部高等学校博士学科点专项科研基金资助项目 (20110095110010); 江苏省普通高校研究生科研创新计划基金资助项目 (CXZZ11_0933); 中央高校基本科研业务费专项资金基金资助项目 (2012DXS02)

Foundation Items: The Doctoral Foundation of Chinese Education Ministry (20110095110010); The Postgraduate Innovation Project of Jiangsu Province (CXZZ11_0933); The Fundamental Research Funds for the Central Universities (2012DXS02)

置的时空信息，使用轨迹常见运动模式对位置进行预测。文献[5]提出 PPM-C 算法将移动对象近期活动区域序列与历史活动模式进行完全匹配，发现最长匹配序列，并将匹配序列的后续区域作为未来位置的候选。此外，多种预测模型被应用至位置预测中，例如，贝叶斯网络^[6]、马尔可夫^[7,8]或隐马尔可夫模型^[9]、神经网络方法^[10]以及状态预测^[11]等方法。文献[12]提出基于对象运动模式预测未来路径和目的地的方法。文献[13]将轨迹定义为移动对象运动时穿过的网格的边的有序集合，在此基础上提出 Traj-PrefixSpan 算法发现频繁轨迹以及移动对象运动模式，进而进行预测。另外，文献[5,8]中介绍了使用手机基站数据集进行预测的方法，应用熵计算从当前基站到达下一个基站的最大概率上限。

以上文献大多从宏观角度分析移动对象兴趣区域间的转换关系，考虑移动对象兴趣区域间运动的研究不多，这样存在 2 个问题：1) 对象活动模式强调移动对象若干活动间的顺序性，由于运动的不确定性，距离较近的区域间的访问顺序可能不固定；2) 在移动对象历史活动记录中，一个兴趣区域之后的候选访问区域可能包含多个，仅依靠兴趣区域间的历史访问习惯很难准确地筛选最佳兴趣区域。

本文提出一种基于运动趋势的移动对象位置预测算法(LP-MT, location prediction algorithm based on movement tendency)，使用基于马尔可夫的移动对象活动模型记录相邻访问区域间的转换关系，在保证区域访问有序性的同时，有效地摆脱对传统模式匹配的依赖，降低运动不确定

性带来的影响。另外，由于对象从同一兴趣区域出发，至不同区域间的运动路径不可能完全相同，有各自的运动特征，因此移动对象的运动趋势对位置预测具有指导意义。以往算法仅将历史记录中最近停留区域后的已访问区域作为预测的候选，本文将移动对象的所有历史停留区域作为未来位置的候选，并根据位置的特征，将算法结果分为预测位置和推荐位置。真实数据的实验表明，算法可以对移动对象未来位置进行有效、高效的预测，较同类算法相比，预测精度提高 10% 左右。

2 基于运动趋势的移动对象位置预测模型

基于运动趋势的移动对象位置预测算法 (LP-MT) 包括移动对象历史活动模型的训练以及未来位置的预测 2 个阶段。

在训练阶段 (如图 1 所示)，对移动对象历史活动模式进行学习。首先，采用文献[14]中介绍的基于时间距离约束的网格聚类算法，从原始轨迹中提取历史停留区域集合，集合中的停留位置将作为未来位置的候选，该集合称为潜在停留区域集合 (图 1 中潜在停留区域集合为 {A,B,C,D})。其次，移动对象历史轨迹被表示为潜在停留区域的序列 (如图 1 中 A->B->C->A->B->D->A)，序列反映了移动对象的历史活动模式，访问顺序以及访问次数反映了移动对象运动的规律以及对各区域的喜好程度。最后，以潜在停留区域序列为输入，建立基于马尔可夫的移动对象活动模型。

未来位置预测过程 (如图 2 所示) 中，首先将

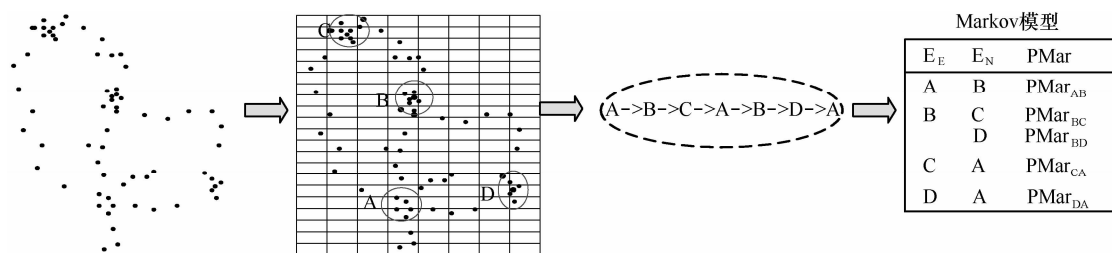


图 1 LP-MT 算法训练阶段示意

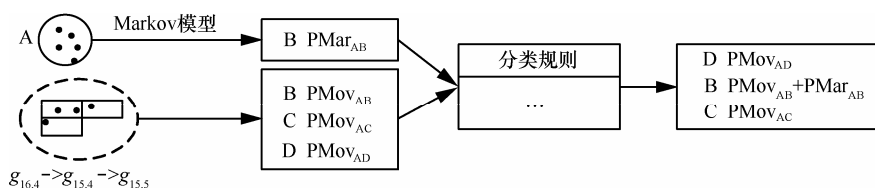


图 2 LP-MT 算法预测阶段示意

移动对象最近停留区域结合基于马尔可夫的移动对象活动模型进行初步预测。其次，将最近停留区域至当前位置间的轨迹段对应至停留区域提取时划分的网格中，将轨迹点序列转换为网格序列（如图 2 中 $g_{16,4} \rightarrow g_{15,4} \rightarrow g_{15,5}$ ），使用网格序列中提取的移动对象运动特征描述移动对象的近期运动趋势，根据运动趋势对移动对象未来位置进行进一步的预测。最后，根据未来位置的分类规则，给出综合预测结果。

2.1 基于移动对象活动模型的预测

目前，移动对象未来位置预测算法大多通过挖掘移动对象轨迹的频繁模式来发现其历史停留区域间的关联关系。在现实世界中，部分停留区域可能较为集中，移动对象对这些区域的访问顺序有很大的随机性。假设 3 个停留区域 A、B、C 间的距离较小，D、E 为距离 A、B、C 较远的 2 个区域，移动对象历史活动模式中包含以下序列：A->B->C->D，A->B->C->E，B->A-C->E，B->C->A->D。若移动对象的近期停留区域序列为 B->A->C，完全匹配历史活动模式时，预测结果为 E。然而，A、B、C 间的距离较近，区域访问具有不确定性，因此 D 作为未来停留区域的概率也较大。另外，对象近期停留区域序列的长度很难选取。本文借鉴马尔可夫模型的思想，提出基于马尔可夫的移动对象活动模型，考虑相邻停留区域间的转换，对移动对象历史活动建模。

马尔可夫模型^[15]由一系列状态以及状态转移矩阵组成，第 n 次转换获得的状态只和第 $n-1$ 次的状态有关。移动对象的历史停留区域序列 $SRL = \{SR_1, SR_2, \dots, SR_N\}$ 可以被表示为状态变量序列 $XL = \{X_1, X_2, \dots, X_N\}$ 。若移动对象潜在停留区域的数量为 m ，则状态空间集合 $EL = \langle E_1, E_2, \dots, E_m \rangle$ ，每个停留区域对应一个状态，移动对象在某一时刻只能处于一种状态。根据移动对象的当前状态以及状态间转移关系，可以粗略估计移动对象的后续状态。

若移动对象最近停留区域为 SR_r ，历史记录中区域 SR_r 后已访问的停留区域集合为 $SRL_N = \{SR_{i_1}, SR_{i_2}, \dots, SR_{i_s}\}$ ， SR_r 和 SR_{i_s} 是移动对象连续访问区域， SR_{i_s} 称为 SR_r 的后续停留区域，本文将移动对象最近停留区域的后续停留区域集合 SRL_N 称为预测停留区域集合。最近停留区域 SR_r 在状态空间集合中的对应状态为 E_r ， E_r 的后续状态集合为

$EL_N = \{E_{i_1}, E_{i_2}, \dots, E_{i_s}\}$ 。状态 E_r 至后续状态 E_{i_s} 的一步转移概率 $PMar_{r,i_s} = P(E_r \rightarrow E_{i_s}) = P(X_{n+1} = E_{i_s} | X_n = E_r)$ 体现了停留区域 SR_{i_s} 作为停留区域 SR_r 的未来位置的可能性。预测停留区域集合 SRL_N 中的区域作为最近停留区域 SR_r 的未来位置的概率集合 $PMarL = \{PMar_{r,i_1}, PMar_{r,i_2}, \dots, PMar_{r,i_s}\}$ 。

2.2 基于移动对象运动趋势的预测

移动对象活动模型从停留区域的角度对移动对象活动习惯进行描述，而对象近期运动趋势也能够指导移动对象未来位置的预测。例如，对象从停留区域 A 出发可能到达 B 和 C，但 B、C 分别在区域 A 的左侧和右侧。若对象从 A 出发向左走，距离 B 越来越近，则尽管历史记录中到达 C 的概率较大，根据移动对象的近期运动趋势可以发现，B 更可能为移动对象的未来停留位置。因此本文在考虑停留区域间关系的同时，也考虑移动对象的近期运动趋势，使用移动对象运动方向以及与各潜在停留区域的距离对移动对象运动趋势进行描述。

移动对象轨迹中距离当前时间越近的信息越能反映移动对象的近期运动趋势，对预测结果的影响越大。本文将移动对象最近停留区域至当前位置间的轨迹段转换为网格序列，首先计算网格序列至各潜在停留区域的距离概率和方向概率，然后根据距离权重和方向权重得到综合概率，用来描述运动趋势对各潜在停留区域作为未来位置的影响。

1) 至潜在停留区域的距离概率

从宏观角度，移动对象至目标位置的距离应越来越短。本文认为移动对象趋向于向距离当前位置较近的停留区域运动，若一段时间内，移动对象距离区域 X 越来越近，则移动对象下一个停留区域为 X 的概率较大。为了将距离标准转换为概率，假设当前网格到各潜在停留区域的最远、最近距离分别为 D_{max} 、 D_{min} ，到达 X 的距离为 D_X ，则从当前网格到达 X 的概率为 $P_d(X) = 1 - (D_X - D_{min}) / (D_{max} - D_{min})$ 。网格序列 $GL = \{G_1, G_2, \dots, G_s\}$ 至潜在停留区域 X 的距离概率集合为 $PL_d = \{P_{d_1}(X), P_{d_2}(X), \dots, P_{d_s}(X)\}$ ，则网格序列至 X 的综合距离概率为

$$P_D(X) = \sum_{i=1}^s P_{d_i}(X) / s \quad (1)$$

2) 至潜在停留区域的方向概率

移动对象向目标位置移动过程中，整体行进方向会趋向于目标位置的方向。当前网格与前一网格

以及停留区域代表网格间都存在方向向量。若随对象运动、方向向量间夹角越来越小,则停留区域为目标位置的概率较大。假设当前网格的第一个样本点为 p_{i+1} ,前一网格中最后一个样本点为 p_i ,则方向向量 $\overline{p_i, p_{i+1}} = (p_{i+1}.x - p_i.x, p_{i+1}.y - p_i.y)$,区域 X 代表网格的中心点为 $r' = (x, y)$,则当前网格至 X 的方向向量为 $\overline{p_{i+1}, r'} = (r'.x - p_{i+1}.x, r'.y - p_{i+1}.y)$,当前网格至区域 X 的方向概率为 $P_a(X) = \cos(\overline{p_i, p_{i+1}, p_{i+1}, r'})$ 。网格序列 $GL = \langle G_1, G_2, \dots, G_s \rangle$ 至区域 X 的方向概率集合为 $PL_a = \langle P_{a_1}(X), P_{a_2}(X), \dots, P_{a_s}(X) \rangle$,则网格序列至 X 的综合方向概率为

$$P_A(X) = \sum_{i=1}^s P_{a_i}(X) / s \quad (2)$$

若综合距离概率和综合方向概率的权重分别为 W_D 和 W_A ,则网格序列至 X 的综合概率

$$PMov(X) = W_D P_D(X) + W_A P_A(X) \quad (3)$$

假设移动对象潜在停留区域集合为 $SRL = \{SR_1, SR_2, \dots, SR_m\}$,反映移动对象近期运动趋势的网格序列 $GL = \langle G_1, G_2, \dots, G_s \rangle$,则根据网格序列计算至潜在停留区域集合 SRL 中的各个区域的综合概率集合 $PMovL = \{PMov_1, PMov_2, \dots, PMov_m\}$ 。

2.3 移动对象未来位置的分类

移动对象未来位置可以分为三类:一是预测停留区域集合中的位置,即历史数据中记录的最近停留区域的后继停留区域;二是潜在停留区域集合中预测停留区域外的其余停留区域,这类区域单纯依靠轨迹模式是无法得到的;最后一类是移动对象未出现过的区域,这类区域无法通过单个对象的历史数据得到,需要分析和当前对象相似的多个对象才能给出合适的结果。本文只考虑前两类位置,即预测位置和推荐位置。

移动对象可能在其历史停留区域集合中的任意位置停留,因此潜在停留区域集合中的区域将作为未来位置的候选。潜在停留区域 X 作为移动对象未来位置的可能性大小被定义为区域 X 的可达性,表示为 PR_X 。若一段时间内,移动对象到达停留区域 X 的可达性都较高,则移动对象运动到区域 X 的可能性最大。

针对预测位置和推荐位置,LP-MT 算法得到预测区域可达性列表和推荐区域可达性列表。预测区域可达性列表对应上述第一类,包含预测区域列表以及各个区域的可达性,其中预测区域列表按照区

域可达性递减的顺序排列。通过基于马尔可夫的移动对象活动模型和运动趋势的综合计算可得到预测区域的可达性。推荐区域可达性列表为上述第二类位置,包含推荐区域列表以及各个区域的可达性,其中推荐区域列表按照区域可达性递减的顺序排列。推荐区域的可达性可通过对移动对象近期运动趋势的分析得到。

若移动对象最近停留区域 SR_r 与基于马尔可夫的移动对象活动模型匹配得到的移动对象预测停留区域的概率集合为 $PMarL = \{PMar_{i_1}, PMar_{i_2}, \dots, PMar_{i_s}\}$,根据移动对象近期运动趋势得到的所有潜在停留区域的综合概率集合 $PMovL = \{PMov_1, PMov_2, \dots, PMov_m\}$ 。预测区域列表由预测停留区域集合组成,预测区域可达性列表的概率集合为 $PRL_{Pre} = \{PR_{i_1}, PR_{i_2}, \dots, PR_{i_s}\}$,其中 $PR_{i_i} = PMar_{i_i} + PMov_{i_i}, PR_{i_i} > PR_{i_{i+1}}, i \in \{1, s\}$ 。推荐区域可达性列表集合为 $SRL_{Com} = \{SR_j | SR_j \in SRL \wedge SR_j \notin SRL_N\}$,列表中的区域按照可达性递减的顺序排列,区域 SR_j 的可达性为 $PR_j = PMov_j, PR_j > PR_{j+1}$ 。

3 基于运动趋势的移动对象位置预测算法

3.1 算法描述

LP-MT 包含 2 个阶段:1) 训练阶段,移动对象历史数据作为训练数据集,提取历史停留区域及历史活动序列,并以此建立基于马尔可夫的移动对象活动模型;2) 预测阶段,根据移动对象最近停留区域以及运动趋势进行分步预测。

LP-MT 算法 输入: T_{Train} (训练轨迹), T_{Test} (测试轨迹), D_{th} (最小距离阈值), T_{th} (最小时间阈值), $G_{Hor}G_{Ver}$ (网格数量), W_D (距离权重), W_A (方向权重)

输出: SRL_{Pre} (预测可达性列表), SRL_{Com} (推荐可达性列表)

/*训练阶段*/

1) $SR = \text{FindStayRegions}(T_{Train}, D_{th}, T_{th}, G_{Hor} \times G_{Ver});$ /*发现停留区域集合*/

2) $SRL = \text{ExtractActivitiesList}(SR);$ /*提取历史活动序列*/

3) $MarkovM = \text{BuildMarkovModel}(SRL);$ /*建立基于马尔可夫的移动对象活动模型*/ /*预测阶段*/

4) $SR_r = \text{FindRecentSR}(T_{Test}, D_{th}, T_{th}, G_{Hor} \times G_{Ver});$ /*发现移动对象最近停留区域*/

5) $G_L = \text{AbstractGridsList}(T_{Test}, SR_r, G_{Hor} \times G_{Ver});$

/*提取反映运动趋势的网格序列*/

6) $PMarL=CalPredReachList(SR_r, MarkovM)$;

/*计算至预测停留区域的概率集合*/

7) $PL=CalSynP(G_L, SRL, W_d, W_a)$; /*计算至潜在停留区域的概率集合*/

8) $PMovL=CalPontitalReachList(PL, SRL)$; /*计算至潜在停留区域可达性*/

9) $SRL_{Pre}=SynPreReachList(PMarL, PMovL)$;

/*预测可达性列表*/

$SRL_{Com}=SynCommReachList(PMovL)$; /*推荐可达性列表*/

3.2 算法性能分析

LP-MT 算法时间主要消耗在移动对象历史活动的发现以及基于对象运动趋势的预测上。历史活动模式发现的时间复杂度为 $O(n)$, n 为移动对象训练轨迹样本点数量。利用运动趋势进行预测的时间复杂度为 $O(ms)$, 其中 m 为移动对象历史停留区域的数量, s 为最近停留区域至当前位置的网格数量。为确保算法具有较高的时间效率, 当真实停留区域确定后, 可以对活动模型进行调整和更新。当预测次数小于对象活动模型最大更新上限时, 历史活动模式重建的时间被更新对象活动模型的时间代替, 对象活动模型更新的时间复杂度为 $O(m)$ 。

4 实验及分析

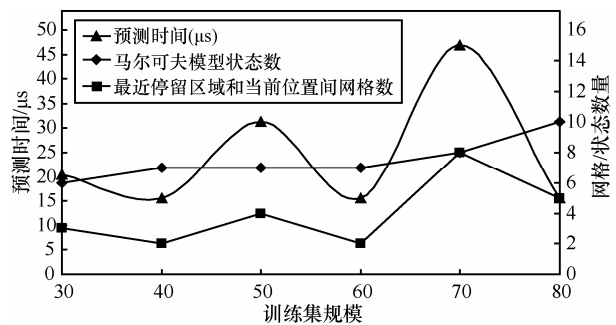
为了验证本文提出的基于运动趋势的移动对象位置预测算法, 采用 Visual C# 2005 开发了移动对象未来位置预测系统 LP-MT, 轨迹数据存储在 SQLServer 2000 中。实验进行的软硬件环境包括: Windows XP, Lenovo ThinkCentre (Duro 2, 2.8G 的 CPU, 3G 内存)。实验数据集使用 2007 年希腊雅典市区卡车(Trucks)的运动轨迹 [16]。

4.1 参数影响

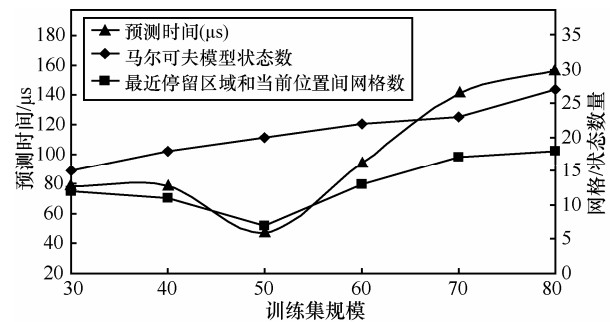
本实验测试网格数量对于基于马尔可夫的移动对象活动模型的健壮性的影响, 选取 Trucks 数据集中 ID=900 的轨迹进行实验, 设定最小距离阈值 $D_{th}=400$ m, 最小时间阈值 $T_{th}=5$ min, 由于对象运动时对方向和距离的无意识性, 设定 $W_D=W_A=0.5$ 。图 3 显示不同网格数量和训练集规模时, 马尔可夫模型状态数、网格序列数以及运行时间的变化情况。

通过图 3 可以看出, 基于马尔可夫的对象活动模型的状态数随着网格数量增加而增加, 相同网格

大小, 不同数据样本规模时, 对象活动模型的状态数增势平缓; 另外, 预测时间随网格数量增加略微增加, 但增速缓慢。以轨迹 10% 的数据量为梯度, 选择数据量的 30%~80% 作为预测模型的训练样本, 进行 6 次未来位置的预测。当网格数量为 10×10 时, 共计产生 3 次推荐和 3 次预测, 预测准确率为 100%。3 次推荐中, 真实停留区域分别位于推荐区域可达性列表的 1、2、2 位, 此时状态数较少, 因此实验结果较理想。网格数量为 30×25 时共产生 4 次预测和 2 次推荐。4 次预测中实际停留区域分别位于预测区域可达性列表的 1、1、2、4 位, 推荐中实际位置分别位于推荐区域可达性列表的 3、4 位。网格数量增加并没有使预测准确率增加, 主要原因是: 虽然网格数量增加使得对象活动模型更加健壮, 但是精细划分的网格使得原本联系紧密的状态被分割, 对象很小的动作被分离出来, 增加预测误差。而且反映对象运动趋势的网格序列至潜在停留区域的距离以及偏转趋势受网格数量的影响, 很小的距离或方向偏差可能对预测结果有很大影响。因此, 网格大小的选择应该适中。



(a) $G_{Hor} \times G_{Ver} = 10 \times 10$



(b) $G_{Hor} \times G_{Ver} = 30 \times 25$

图 3 不同网格数量和训练集规模对 LP-MT 算法的影响

4.2 预测精度

为了检验预测精度, 本文采用增量验证方式, 选取 Trucks 数据集中 ID=870 的轨迹进行分组实验。为

了衡量基于运动趋势的移动对象位置预测算法的准确性，给出评价预测精度的预测准确率和推荐准确率、预测误差率和推荐误差率 2 组概念。假定预测次数为 N ，真实停留区域位于预测可达性列表首位的次数为 PN ，则预测准确率 $PA_{Pre}=PN/N$ 。同理，假设真实停留区域位于推荐可达性列表首位的次数为 RN ，则推荐准确率 $PA_{Com}=RN/N$ 。假设移动对象当前位置至真实停留区域的距离为 D_{c-r} ，预测位置至真实停留区域的距离为 D_{p-r} ，则预测误差率 $PF_{Pre}=D_{p-r}/D_{c-r}$ 。同理，若推荐位置至真实停留区域的距离为 D_{rec-r} ，则推荐误差率 $PF_{Com}=D_{rec-r}/D_{c-r}$ 。当前位置至真实停留区域的距离越大，预测产生误差的概率越大，因此根据 D_{c-r} 与 D_{p-r} 的比值来衡量算法的误差率比较合理。预测误差率或推荐误差率越小，则预测位置或推荐位置越接近真实位置，算法精度越高。

1) 不同训练样本比例

该实验主要用于测试在历史数据充分学习的情况下，算法的预测精度。选取整条轨迹 30%~85% 的数据量作为训练样本，训练样本的结束点作为移动对象的当前位置开始预测。

从表 1 可以看出，训练样本比例与马尔可夫模型状态数成正比。实验表明马尔可夫模型中 10%~20% 的状态有 3 个及 3 个以上的转换状态，20%~30% 的状态有 2 个转换状态，其余包含一个转换状态，极少数情况下，不包含转换状态，此时，移动对象首次到达该停留区域且没有移动至下一区域。使用不同数据集建立的马尔可夫模型的状态转换情况有所不同。

表 1 Truck 数据不同训练样本比例对马尔可夫模型状态数量的影响

训练样本比例/%	马尔可夫模型状态数量
30~35	10
35~40	11
40~45	12
45~50	13
50~57	14
57~65	15
65~70	16
70~85	17

算法结果由预测可达性列表以及推荐可达性列表共同决定。预测中无法判断真实停留区域位于哪类结果中，因此预测误差率和推荐误差率的最小值决定了算法的综合误差率。图 4 描述了不同训练样

本比例下误差率的变化情况，从图中的曲线可以看出，算法综合误差率基本在 0.1 以下，预测精度较高。

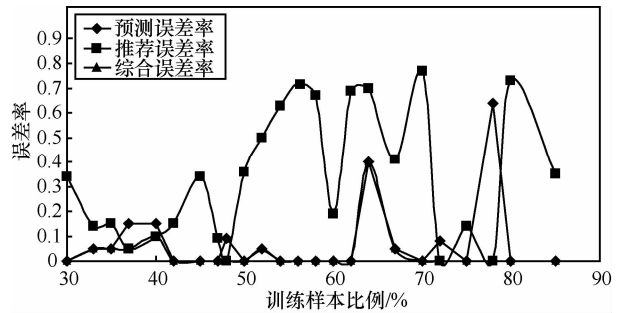


图 4 不同训练样本比例下的误差率

以 3% 的数据集为一个梯度，25 次实验中共产生 5 次推荐和 20 次预测。推荐准确率达到 40%。20 次预测结果中，仅根据基于马尔可夫的移动对象活动模型预测的准确率达到 50%。增加考虑移动对象运动趋势可提高 24% 的预测成功率，其余情况下，结合运动趋势也可以使准确率在简单的一步预测的基础上有显著提高。实验表明算法在不同训练样本比例下都具有较高准确性。

2) 训练集规模

该实验主要测试当历史数据量较少时，预测结果的准确性。选取整条轨迹 30%~90% 间不同梯度的数据量作为训练样本，以 90% 处的样本点作为移动对象当前位置。

从图 5 中的曲线可以看出，训练集规模较小时，推荐结果的准确率较高，这是由于此时移动对象历史数据较少，仅依靠历史停留区域间的转换关系无法准确预测，但是移动对象近期运动趋势能够反映对象活动意图，从而提高推荐准确性。图 5 显示推荐误差率趋近于 0，表明本文选取距离和方向作为反映运动趋势的标准较为合理。随着训练集规模的增加，预测结果更加准确，综合误差率基本小于 0.3。因此，历史数据量较少时，运动趋势能够很好地对未来位置进行推荐，随着历史数据集规模的增大，预测精度则越来越高。

3) 学习充分性

该实验用于检验基于马尔可夫的移动对象活动模型能否充分学习对象的历史活动模式。选取整条轨迹 50%~100% 之间的数据进行训练，从轨迹的 45% 处开始预测。以 5% 的数据集为一个梯度，20 次实验的预测误差率全部为 0，实验表明对象活动模型能够准确、全面地反映移动对象的历史活动模式。

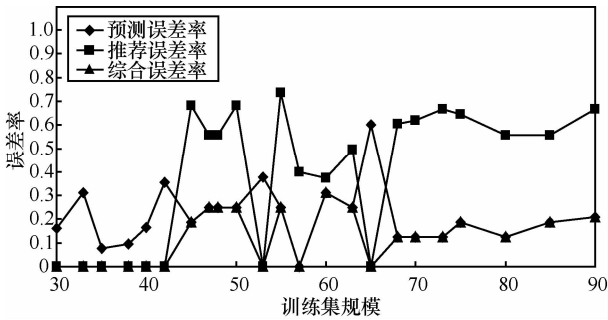


图 5 训练集规模与误差率间关系

4) 算法稳定性

该实验主要用于测试算法的稳定性。选取整条轨迹为训练集，从轨迹的 30%~90%之间随机选取若干位置作为移动对象的当前位置开始预测。若随机选取一系列位置，误差率集合 $PFL = \{PF_1, PF_2, \dots, PF_n\}$ ，则 $\forall 1 \leq i \leq n$ ，第 i 次预测的平均误差率为 $Ave_PF_i = \sum_{j=1}^i PF_j / i$ 。平均误差率反映了算法在多次预测中的持续预测能力，体现了算法的稳定性。图 6 显示算法的平均综合误差率维持在 0.1 以下，预测精度较高且较为稳定。另外，对于不同的数据集都有较好的效果，算法的适应性和可移植性较强。

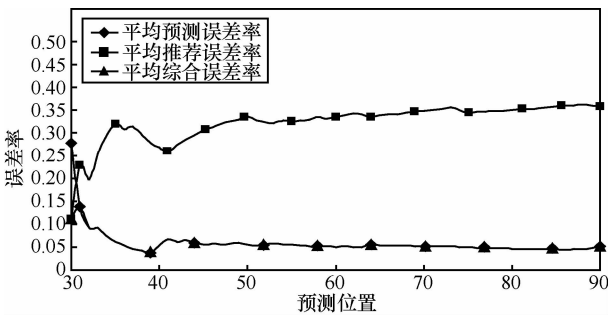


图 6 测试位置与平均误差率的关系

4.3 移动对象活动模型的更新有效性

该实验主要测试算法更新的有效性以及最大的更新次数。以 50% 的数据量为训练数据集对移动对象位置开始连续预测，根据真实停留区域对对象活动模型进行 55 次更新。在实际情况中，移动对象的真实停留区域会与历史停留区域有一定的偏差，但如果重叠面积较大，本文认为是同一个停留区域。从图 7 中的曲线可以看出，前 20 次更新中，综合误差率基本在 0.2 以下波动，为可接受的误差范围。然而，在 20 次更新之后，算法的综合误差率开始增加，甚至达到 0.5。多次更新过程中，移

动对象在同一区域出现的位置可能发生变化。因此，更新策略可保证算法在一定次数内的更新是有效的。本实验环境下，20 次更新后需要对对象活动进行重新建模。

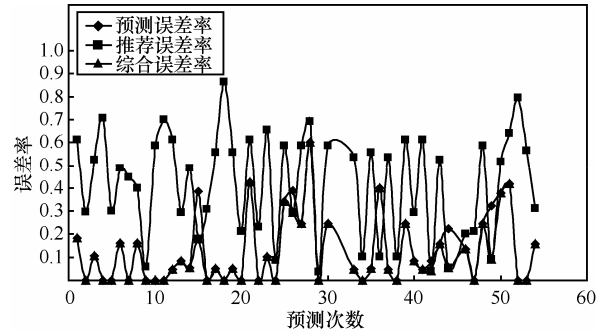


图 7 马尔可夫模型的更新对误差率的影响

4.4 同类算法性能比较

PPM-C 算法与 LP-MT 算法都是对移动对象位置进行预测的方法，在同类算法中，PPM-C 算法的性能较好，具有一定代表性。为了比较 LP-MT 算法与 PPM-C 算法的性能，本文选取整条轨迹数据量的 30%~90% 作为训练样本，以 90% 处的样本点作为预测起始位置。图 8 和图 9 显示与 PPM-C 算法相比，LP-MT 具有较好的预测效果和运行速度。

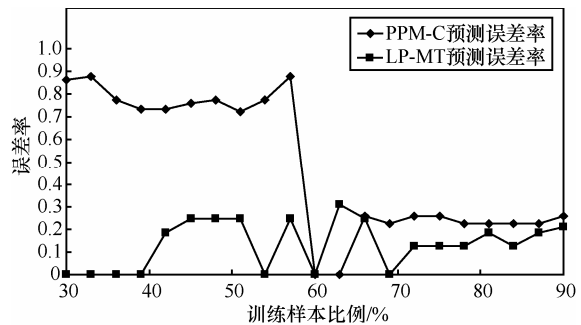


图 8 PPM-C 与 LP-MT 算法误差率比较

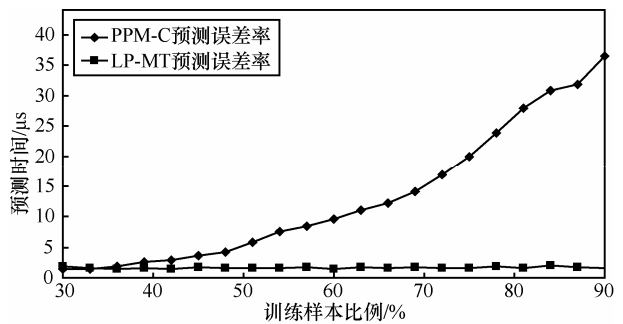


图 9 PPM-C 与 LP-MT 算法运行时间比较

图 8 比较了 LP-MT 与 PPM-C 使用相同训练集，从同一位置开始预测的综合误差率。当训练样

本规模占数据集总量的 30%~60% 时, PPM-C 算法的综合误差率基本在 0.7 以上, 同等情况下, LP-MT 算法的综合误差率小于 0.3。训练样本规模大于数据集总量的 70% 时, 2 种算法的综合误差率均小于 0.3, LP-MT 算法的综合误差率略低。主要由于: 1) PPM-C 主要依靠学习历史数据来对移动对象未来位置进行预测, 当训练样本较小时, PPM-C 对历史活动学习不够充分, 无法得到较为准确的预测结果, 而 LP-MT 算法在学习不充分时, 使用移动对象近期运动趋势对结果进行补充, 确保算法具有较好的预测效果; 2) PPM-C 仅考虑历史数据中最近停留区域的后继停留区域集合, 忽略了移动对象运动的不确定性, LP-MT 算法在考虑预测停留区域的同时, 计算至所有潜在停留区域的可达性, 提高预测精度。

图 9 比较了 LP-MT 与 PPM-C 算法的运行时间, PPM-C 算法的运行时间随训练样本规模的增大迅速增加, 而 LP-MT 算法的运行时间变化较为平缓。若移动对象潜在停留区域的数量为 n , 停留区域序列的长度为 $N(N \gg n)$, PPM-C 算法中 order 表的规模与 N 成正比, 随着训练样本规模的增加, order 表的规模迅速增加, 而 LP-MT 只需根据 n 个历史停留区域集合进行建模。另外, 每次预测时, PPM-C 算法依次缩短近期停留区域序列与 order 表中各层进行完全匹配, 使得预测时间较长。因此, 训练样本规模较大时, PPM-C 算法的运行时间远远大于 LP-MT 算法的运行时间。

5 结束语

针对目前移动对象位置预测大多只考虑移动对象历史活动模式, 导致预测精度不高且预测结果较为局限的问题, 提出一种基于运动趋势的移动对象位置预测算法, 不仅使用移动对象最近停留区域和移动对象活动模型中存储的移动对象历史活动模式进行匹配给出初步预测, 还综合考虑移动对象最近运动趋势, 估计各潜在停留区域作为未来位置的概率。根据移动对象未来位置的特征, 将潜在停留区域分为预测停留区域以及非预测停留区域, 分别对应预测位置和推荐位置。应用真实数据的实验表明, 与其他同类算法相比, LP-MT 算法预测精度提高近 10%。在保证较高时间效率的同时, 具有较好的准确性。下一步将在

本文研究的基础上, 考虑移动对象的群体性以及多移动对象的局部相似性, 使用多个对象的轨迹数据对未来位置进行预测。

参考文献:

- [1] MONREALE A, PINELLI F, TRASARTI R. Where next: a location predictor on trajectory pattern mining[A]. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining[C]. New York, 2009. 637-646.
- [2] YOON H, SHAHABI C. Robust time-referenced segmentation of moving object trajectories[A]. ICDM '08, Eighth IEEE International Conference on Data Mining[C]. 2008. 1121-1126.
- [3] MORZY M. Mining frequent trajectories of moving objects for location prediction[A]. Machine Learning and Data Mining in Pattern Recognition[C]. 2007. 667-680.
- [4] LI H Y, TANG C J, QIAO S J. Hotspot district trajectory prediction[A]. Web-Age Information Management[C]. 2010. 74-84.
- [5] BURBEY I E. Predicting Future Locations and Arrival Times of Individuals[D]. Virginia: Virginia Polytechnic Institute and State University, 2011.
- [6] PETZOLD J, PIETZOWSKI A, BAGCI F, *et al.* Prediction of indoor movements using Bayesian networks[A]. Lecture Notes in Computer Science[C]. 2005. 173-184.
- [7] 彭曲, 丁治明, 郭黎敏. 基于马尔可夫链的轨迹预测[J]. 计算机科学, 2010, 37(8): 189-193.
- [8] PENG Q, DING Z M, GUO L M. Prediction of trajectory based on Markov chains[J]. Computer Science, 2010, 37(8): 189-193.
- [9] SONG C M, QU Z H, BLUMM N. Limits of predictability in human mobility[J]. Science, 2010, 327(5968): 1018-1021.
- [10] JEUNG H, SHEN H T, ZHOU X F. Mining trajectory patterns using hidden Markov models[A]. Data Warehousing and Knowledge Discovery[C]. 2009. 470-480.
- [11] KOSKELA T, LEHTOKANGAS M, SAARINEN J. Time series prediction with multilayer perceptron, FIR and Elman neural networks[A]. Proceedings of the World Congress on Neural Networks[C]. 1996. 491-496.
- [12] BAGCI P J, TRUMLER F, *et al.* Global state context prediction techniques applied to a smart office building[A]. The Communication Networks and Distributed Systems Modeling and Simulation Conference[C]. San Diego, 2004.
- [13] CHEN L, LV M Q, CHEN G C. A system for destination and future route prediction based on trajectory mining [J]. Pervasive and Mobile Computing, 2010, 6(6): 657-676.
- [14] ASHBROOK D, STARNER T. Using GPS to learn significant locations and predict movement across multiple users[J]. Personal and Ubiquitous Computing, 2003, (7): 275-286.

(下转第 62 页)

机学报, 2012, 35(4): 645-653.

TIAN Y L, MA J F, PENG C G, *et al.* Universally composable mechanism for group communication[J]. Chinese Journal of Computers, 2012, 35(4): 645-653.

[18] 张好, 胡杰. UC 安全计算的一种信任模型[J]. 四川大学学报(工程科学版), 2012, 44(3): 106-111.

ZHANG Y, HU J. A trust model of UC secure computation[J]. Journal of Sichuan University(Engineering Science Edition), 2012, 44(3): 106-111.

[19] BONEH D, FRANKLIN M. Identity based encryption from the weil pairing[J]. SIAM J of computing, Extended Abstract in Crypto, 2003, 32(3): 586-615.

作者简介:



田有亮 (1982-), 男, 贵州盘县人, 博士, 贵州大学副教授、硕士生导师, 主要研究方向为算法博弈论、密码学、信息安全。



彭长根 [通信作者] (1963-), 男, 侗族, 贵州锦屏人, 博士, 贵州大学教授、博士生导师, 主要研究方向为密码学、信息安全。E-mail: sci.cgpeng@gzu.edu.cn.



马建峰 (1963-), 男, 陕西西安人, 博士, 西安电子科技大学教授, 博士生导师, 主要研究方向为网络与信息安全、密码学等。

林辉 (1977-), 男, 福建福州人, 西安电子科技大学博士生, 主要研究方向为下一代网络、网络 QoS 技术。

杨凯 (1983-), 男, 山东莱芜人, 博士, 武警工程大学讲师, 主要研究方向为无线 mesh 网络、OPNET 仿真。

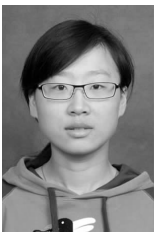
(上接第 53 页)

[14] ZHENG V W C, ZHENG Y, XIE X. Collaborative location and activity recommendations With GPS history data[A]. Proceeding of International Conference on World Wide Web[C]. 2010. 1029-1038.

[15] http://en.wikipedia.org/wiki/Markov_model[EB/OL]. 2011.

[16] <http://www.chorochronos.org/Default.aspx?tabid=75> [EB/OL]. 2012.

作者简介:



李雯 (1988-), 女, 河北邢台人, 中国矿业大学博士生, 主要研究方向为轨迹数据挖掘。

夏士雄 (1961-), 男, 辽宁抚顺人, 中国矿业大学教授、博士生导师, 主要研究方向为数字化矿山、智能信息处理等。

刘峰 (1967-), 男, 甘肃平凉人, 中国矿业大学兼职教授、主要研究方向为轨迹数据挖掘。

张磊 (1977-), 男, 江苏沛县人, 博士后, 中国矿业大学副教授、硕士生导师, 主要研究方向为轨迹数据挖掘。

袁冠 (1982-), 男, 江苏睢宁人, 博士, 中国矿业大学讲师, 主要研究方向为轨迹数据挖掘。