

基于隐马尔可夫模型的移动终端定位算法^①

路锦博, 蔡丽萍, 孙宗星

(中国石油大学(华东) 计算机与通信工程学院, 青岛 266580)

摘要: 移动互联网技术的快速发展, 对移动终端定位方法提出了新的要求. 针对传统方法在功耗、精度、通用性方面往往不能兼顾的问题, 在充分挖掘个人行为习惯的基础上, 提出了一种基于隐马尔可夫模型的定位算法. 该算法根据地图信息对基站覆盖区域进行路段划分, 通过记录个人的行走轨迹, 利用其统计规律来辅助建立 GSM 信号的隐马尔可夫模型, 通过测量基站的信号强度序列, 完成对当前位置给出较为精确的估计. 实验结果表明该算法在保证低功耗的同时有效提高了定位精度.

关键词: 基于位置的服务; 手机定位; 隐马尔可夫模型; 行为习惯

引用格式: 路锦博, 蔡丽萍, 孙宗星. 基于隐马尔可夫模型的移动终端定位算法. 计算机系统应用, 2017, 26(8): 195-200. <http://www.c-s-a.org.cn/1003-3254/5907.html>

Positioning Algorithm of Mobile Terminal Based on Hidden Markov Models

LU Jin-Bo, CAI Li-Ping, SUN Zong-Xing

(College of Computer and Communication Engineering, China University of Petroleum, Qingdao 266580, China)

Abstract: With the rapid development of mobile internet technology, new requirements have been set for the mobile terminal positioning method. In view of the problem that we cannot take into account of energy consumption, accuracy and versatility, we propose a Hidden Markov Model-based and personal habits-based network location algorithm. The algorithm divides the roads into sections in the area covered by the base stations according to the map, and records the personal walking trajectory, which helps build the hidden markov model of GSM signals with the statistical law. Thus, with only a station's signal sequence, we can get an accurate estimate of the user's current position. Experiments show that the algorithm can guarantee a good positioning accuracy with lower energy consumption.

Key words: location based service; mobile location; Hidden Markov Model; personal habits

基于位置的服务(Location Based Service, LBS)得到飞速发展, LBS 的基础是移动终端定位技术^[1-3], 目前最常用的定位方式是全球定位系统(Global Positioning System, GPS), 利用太空中的人造卫星对移动终端进行定位^[4,5], 定位精度较高, 但是能耗较大, 而且有覆盖不到的盲区, 也不适用于没有配备 GPS 装置的移动终端. 而以手机为代表的移动终端对于低能耗要求越来越高, 同时现存的大量的功能手机要求在不具备 GPS 功能的情况下实现定位. 虽然借助于电子罗盘、陀螺仪等

传感器的辅助可以降低 GPS 的能耗^[6], 但这会进一步限制其通用性. 在这种情况下, 依托移动通信网络的网络定位方式受到了广泛关注, 仅利用网络基站等基础设施, 不需要配备额外的装置, 即可对移动终端进行定位, 能耗低、盲区少, 但定位精度比较低. 本文在兼顾低能耗的基础上, 提出了一种基于隐马尔可夫模型(Hidden Markov Model, HMM)结合个人行为习惯的高精度定位算法.

^① 基金项目: 青岛市黄岛科技计划项目(2014-1-45)

收稿时间: 2016-12-08; 采用时间: 2017-01-04

1 相关工作

移动通信网络的基础结构是由一系列的基站构成, 这些基站把整个通信区域划分成蜂窝小区, 小则几十米, 大则几千米, 移动设备必须通过某一个蜂窝基站接入移动网络, 这就为利用基站进行定位提供了基础. 最简单的是利用基站的 ID 号定位, 即根据手机所处的小区基站的位置来确定自身位置, 显然, 定位的精度取决于蜂窝小区的半径, 从几十米到几千米不等. TOA^[7]是基于电波传播时间的定位方法, 如果得到 3 个位置已知的基站信号, 则可以计算位置. TOA 算法对系统的时间同步要求很高, 任何很小的时间误差都会被放大很多倍, 同时易受多径效应的影响. TDOA^[8]是 TOA 的改进, 通过对传输时间求差, 抵消一部分的时间误差以及多径效应带来的误差, 可以提高定位的精度. AOA^[9]是一种基于信号入射角度的定位方法, 需要两个基站配合, 同时也要求移动终端能够得到信号入射角度.

近几年基于接收信号强度 (RSSI) 的定位方法得到了广泛研究, 文献^[10]考虑到无线信号强度易受多种因素影响, 提出结合指纹的方法进行定位, 预先测量记录不同位置接收到的基站信号强度, 多个基站信号构成一个向量, 当移动终端测量当前信号强度之后, 在数据库中寻找与自己最相近的一组信号, 其位置即为当前位置. 文献^[11]认为无线信号在某个位置不是确定的, 而是符合某种概率分布, 提出基于 RSSI 的概率指纹定位算法.

文献^[12]提出了一种 Cell-ID 辅助的定位系统, 该算法基于用户行走路线的连续性, 会产生一段连续的 Cell-ID 序列的事实, 从而在用户经常行走的路线上, 利用 Cell-ID 序列匹配技术可以唯一的确定位置. 文献^[13]利用与之关联的信号塔的信号强度来获得精确的 GSM 定位, 将区域划分成小方格, 然后在得到一系列 GSM 信号的情况下估计用户的位置. 该算法不需增加其他辅助手段, 在保持了低功耗的同时, 定位精度得到较大提升. 文献^[14]提出了一种根据服务基站与相邻基站的稳定程度划分网格, 然后利用 HMM 滤波的方法. 但是, 该类算法忽略了不同个人在行走路线上的不同, 在多交叉路口易出现较大误差.

在室内定位领域, 采用客户端/服务端(C/S)模式将 HMM 广泛的应用于位置追溯^[15]或者和滤波算法结合实现目标定位^[16]. 但是, 这种使用方式算法的复杂度比

较高, 主要应用于客户端/服务端(C/S)模式的定位系统, 对于移动客户端而言, 算法过于复杂.

本文利用室外的特点, 根据道路和场地(例如广场、体育场等公共场所)进行区域划分与路径分段, 以此降低离线指纹库的体量. 此外, 结合一段时间内人的行为习惯, 优化了转移概率矩阵, 提高了定位精度. 相比已有工作, 主要贡献如下: (1)充分利用每个人的行为习惯来确定模型参数, 从而使模型更加精确. (2)只对可达的道路、广场和建筑进行区块划分, 根据每个人行走路线的不同, 分别剔除不可达的区域, 有效地减少了相似信号的干扰.

2 基于隐马尔可夫模型的移动终端定位算法设计

2.1 基本思想

在 GSM 网络中, 移动终端可以非常容易地收集周边基站的信号, 但无线信号的传播易受到多种因素影响, 导致测得的基站信号强度并不能与地理位置一一对应, 一个基站信号强度可能出现在多个位置, 一个位置在不同的时间也会测出不同的信号强度. 但是终端的移动总是从一个位置到紧邻的另一个位置, 不可能跳跃, 因此如果已知连续的多个位置(P_0-P_n)的信号强度(X_0-X_n), 此时测得 X_{n+1} , 则完全有可能预测位置 P_{n+1} , 这类问题的求解可以借助于隐马尔可夫模型. 同时移动终端一般被人持有, 而人的行为模式在时间和空间上都表现出一定规律性^[17,18], 例如在交叉路口不同的人左转和右转的概率不同; 再例如在居住的小区里面有些地方经常去, 而某些地方却始终没有去过, 这些规律都可以为我们的定位算法提供帮助. 本文的基本思想就是利用隐马尔可夫模型, 结合个人不同的行为模式设计低功耗、高精度的定位算法.

2.2 隐马尔可夫模型建立

隐马尔可夫模型是一个双重随机过程, 包含一组不能直接观测的状态和一组可观测的状态序列, 当获知到一个可观测状态序列时, 利用 Viterbi 算法寻找某种意义上最优的隐含状态序列, 不仅仅考察前后两个观察状态的关系, 而是全面考察整个观测序列, 得出一个最大似然的结果. 隐马尔可夫模型一般用五元组 $\{S, O, \pi, A, B\}$ 进行描述, 五元组参数的含义如下:

S: 系统的隐含离散状态集合, $S = \{S_0, S_1, \dots, S_N\}$. 势为 N.

O: 系统的可观察状态序列集合, $O = \{O_0, O_1, \dots, O_M\}$, 势为 M.

π : 系统的初试概率矩阵, 表示初始时刻, 系统处于 S 中各状态的概率. $\pi = [\pi_i]$, 且 $\sum_{i=1}^N \pi_i = 1$.

A: 转移概率矩阵, $A = \{a_{ij}, 1 \leq i, j \leq N\}$, 其中每个元素 a_{ij} , 表示 t 时刻状态为 S_i , 而 t+1 时刻状态为 S_j 的概率.

B: $B = \{b_{ij}, 1 \leq i \leq N, 1 \leq j \leq M\}$, 其中每个元素 $b_{ij} = P(O_j | S_i)$, 且有 $\sum_{j=1}^M b_{ij} = 1, 1 \leq i \leq N$.

很显然本文中提出的定位算法非常适合用隐马尔可夫模型求解, 模型元素含义如下:

S, 将某一区域的道路划分成的各个小段或者区域块.

O, 指手机检测到的 GSM 信号强度.

π , 指初始时刻, 位于某个路段的概率.

A, 指从一个路段到另一个路段的概率.

B, 指在各路段上, 出现各种信号强度的概率.

定位问题可以描述为, 在得出路段集 S, 初始时刻位于各路段的概率矩阵 π , 各路段之间的转移概率矩阵 A, 以及各路段上出现各种信号强度的概率矩阵 B 的前提下, 再给出一个观测到的 GSM 信号序列, 利用 Viterbi 算法得出用户最有可能走过的路段序列, 用户的当前位置, 即为最优路段序列的最后一个路段所在位置.

2.2.1 区域划分与路径分段

首先要将一个基站区域划分为足够小的小区块, 平均划分的方法是不可取的, 因为一些无法或者不易到达的地方是可以忽略的. 实际上整个区域只有两种地形: 道路和场地(例如广场、体育场等公共场所). 对于道路, 将其划分为小的路段, 一个基本原则是: 交叉路口不能包含在路段中央, 即交叉路口处作为一个段的终点和另一个段的起点, 因为在这些地方行走方向会变化. 同时路段长度是决定测量精确度的一个重要因素, 一方面路段越长, 可以更好地统计各路段上的 GSM 强度信息, 使得不同路段之间差异增大, 有利于提高算法精度; 另一方面较长路段内的估计误差也将越大, 降低了定位精度. 因此需要在路段长度上进行均衡, 得出一个比较合适的长度.

对于较大面积的场地, 需要进行小区域网格划分, 同样网格的边长也是影响定位精确度的一个重要因素.

对于 HMM 算法而言, 路段和网格是没有区别的, 为了叙述方便, 统称为路段.

2.2.2 建立矩阵 B

为了建立 B, 需要在各路段上选取多个观测点, 分别测量 GSM 信号强度, 然后统计该路段上的信号强度直方图, 由此可得出该路段上出现各种信号强度的概率值. 同理对其他路段进行类似处理, 就可以得出整个 B. 另外由于 GSM 信号受天气影响较大, 考虑到不同天气状况下 GSM 信号的差异性, 可以选择在不同天气条件下, 分别进行采样, 单独记录, 具体使用时, 根据实际情况, 系统自动选择对应的 GSM 样本下得出的 B, 这样有利于提高精确度.

2.2.3 根据个人行为习惯建立转移矩阵 A

概率转移矩阵 A 是 HMM 模型的重点, 也是决定定位算法精度的核心. 在基于 HMM 模型的定位算法中, 一般假设各相邻路段之间等概率转移, 从而确定概率转移矩阵 A. 但是, 此类方法忽略了不同个人在行走路线上的不同.

为了充分利用个人行为以提高定位精度, 提前统计用户在一段时间内的行走路线.

根据用户在一段时间内的行走路线, 重点关注了三种行为习惯: 在不同路口的选择、路段的选择及工作日与休息日的路线选择差异. 对于路口绘制每个路口的转移状况图, 得出转移概率, 并根据用户工作日与休息日的路线差异及路段差异, 建立不同的转移矩阵 A. 在定位阶段, 根据当前的日期以及用户所在路段, 自动选择采用哪种转移矩阵.

2.2.4 初始概率矩阵的建立

如果已知初始时刻位于各路段的概率, 则可直接使用; 如果不知道, 可以利用转移概率矩阵的稳定状态概率矩阵来代替.

2.3 算法设计

一般一个路段最多和 4-5 个路段相连, 因此, 如果采用传统的矩阵进行存储的话, 矩阵的规模会很大, 严重浪费存储空间, 带来较大的计算复杂性. 同时注意到矩阵大部分元素为 0, 因此采用稀疏矩阵来对转移概率矩阵 A 进行存储, 一方面节约存储空间, 另一方面可以很大程度地降低计算复杂性.

用户在行走过程中, 可能存在某一地点停留一段时间的情况, 在这一地点采集到的信号差别不大. 考虑一种极端情况, 用户在某一地点停留的时间足够长, 将

会导致用来定位的整个 GSM 信号序列的信号值大小都一样, 因此就失去了隐马尔可夫模型预测作用. 此时需要在定位前对信号序列进行处理. 如果采集到的信号序列中的信号有 70% 的在一个阈值范围内可以看作是一致的, 则可认为用户在这几个采集点没有移动, 只保留其中一个, 再向前取几个信号, 补足信号传长度; 对连续定位的应用, 如果信号序列中有超过 70% 的信号强度是一致的, 则认为用户位置基本没动, 可取上一次的定位作为当前用户的位置.

考虑到 GSM 信号容易受到天气情况的影响, 以及用户的出行规律有一定的周期性, 采用自适应的方法来应对, 当利用该算法进行定位时, 系统会自动根据天气选择一种合适的 B, 并根据日期选择合适的 A.

算法的具体流程如下所示:

Step1. 地图获取与路段划分, 通过此步骤可以得到隐藏状态集 S;

Step2. 对各路段, 不同天气情况下进行 GSM 信号信息采集, 获取各种天气状况下的矩阵 B;

Step3. 统计用户在一周或更长时间内的行走轨迹, 分别统计工作日和休息日的情况并得出各路段之间的转移概率矩阵 A;

Step4. 利用公式 $\pi_s A = \pi_s$, 计算出初试概率矩阵 π ;

Step5. 采集 GSM 信号一段时间, 得到信号强度序列, 并对信号强度序列进行处理, 移除信号强度序列中的重复值, 并产生有效的信号序列, 从而排除用户是否在此信号序列采集过程中基本没有移动的情况;

Step6. 系统自动获取天气和日期, 选择合适的 A 和 B;

Step7. 利用 Viterbi 算法计算出用户最有可能走过的路段序列, 用户的当前位置估计为此路段序列的最后一个路段.

3 实验验证与结果分析

3.1 实验方案与评价指标

以大学校园作为实验场地, 选用 Android 系统实现本文定位算法, 选取十名匿名同学参与实验. 定义系统的两个评价指标: 平均能量消耗(Average Energy Consumption, AEC)和平均定位错误(Average Location Error, ALE).

(1) 平均能量消耗: 能量消耗是指连续运行定位应用一段时间之后, 消耗的手机电量与时间的比值, 即

单位时间的平均能量消耗.

(2) 平均定位错误: 将 n 个时刻的测量值序列与真实值序列做差, 然后求平均.

3.2 实验模型建立

首先将道路进行路段划分, 需要对路段的长度进行控制, 这一点将在模型检验部分详细介绍. 各路段构成了隐藏状态集 S, 然后在各路段上选取多个点采集 GSM 信号信息, 绘制直方图, 如图 1 所示. 利用各路段的 GSM 信号强度统计信息构造出矩阵 B.

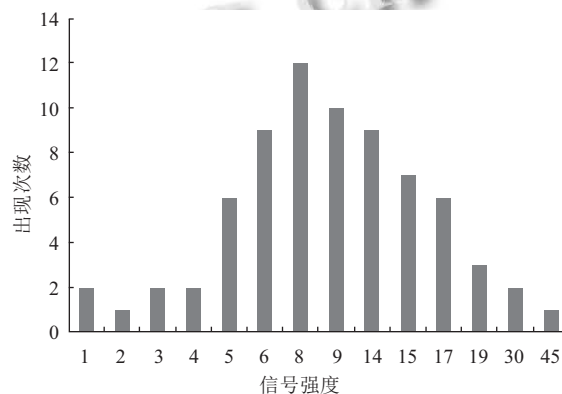


图 1 某一路段 GSM 信号强度直方图

对十名匿名同学, 分别记录他们一周内行走路线的情况. 图 2 展示了一名同学的行走轨迹, 红色线条的粗细代表走过路线的频率, 由此图可验证个人的行走路线的规律性, 然后结合之前构造的 S 中的各路段, 统计出各路段之间的转移情况, 由此得出转移概率矩阵 A.



图 2 某同学在一周内移动轨迹示意图

利用公式 $\pi_s A = \pi_s$ 计算出转移矩阵 A 的稳定概率矩阵, 作为初始概率矩阵 π . 至此, 该算法的隐马尔可夫模型已经基本建立起来, 系统自动记录用户刚才走过的一段轨迹的 GSM 信号强度, 来估测用户的当前位置.

线条粗细,表示走过的次数,越粗,表示走过次数越多.

3.3 实验结果分析

从五个方面对系统性能进行评估.

(1) 路段长度的影响

图3展示了不同路段长度对定位精度的影响,可以看出随着路段的长度增加,定位误差值,先减小,后增大.这是因为路段越长,就可以更好地统计各路段上的GSM强度信息,能够得到效果更好的矩阵B,有利于提高定位精度;同时路段越长,在该路段内的估计误差也将越大,降低了定位精度.实验表明在路段长度为70m左右时,可以达到最好的定位精度,因此在以后的试验中,我们使用的路段长度为70m.

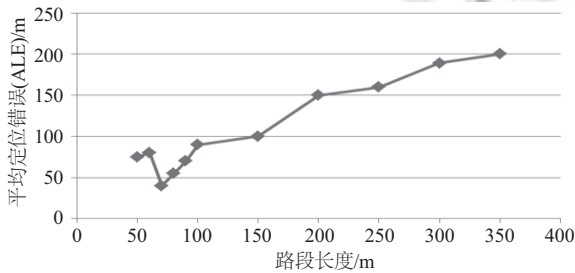


图3 路段长度对定位精确度影响图

(2) 序列长度的影响

图4展示了选取不同的隐马尔可夫链的长度对于定位精确度的影响.可以看出,当序列长度为11时,定位精确度较高.

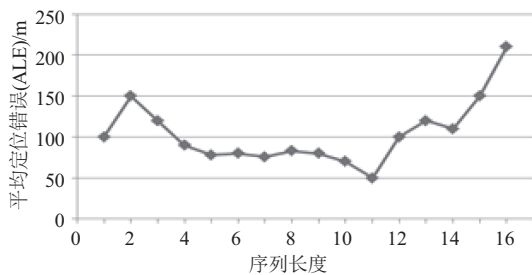


图4 隐马尔可夫模型序列长度对定位精确度影响图

(3) 是否结合个人行为习惯对系统的影响

建立不结合个人行为习惯的隐马尔可夫模型,也即各相邻路段之间等概率转移.然后随机选择10个点,分别利用结合与不结合个人行为习惯的隐马尔可夫模型进行定位,同时记录GPS定位准确值,对比二者与准确值之间的误差,如图5所示.可以看出相对于不学

习而言,结合个人行为习惯的算法,在性能上有很大提高.

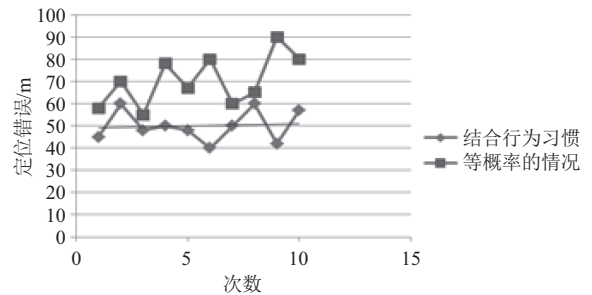


图5 结合行为习惯与不结合行为习惯定位错误对比

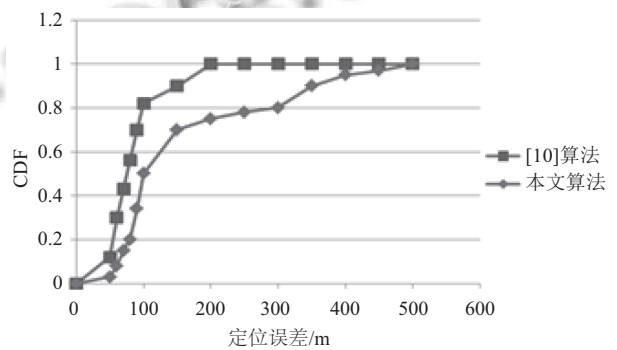


图6 各种定位算法的最佳定位误差累积示意图

(4) 定位精度比较

将本文算法与文献[10]中算法进行对比.图6展示测量误差的累积分布函数($F(x) = P(x < X)$).实验结果表明,本文算法要优于文献[10]中算法.本文提出的定位算法可以达到50m左右的精度,具有很好的定位效果.

(5) 能量消耗情况

使用本系统连续定位5小时,每隔5分钟定位一次.在另一个手机上利用GPS进行同样频率的定位.分别每隔1h记录一次手机剩余电量,手机剩余电量如图7所示,可以看出5个小时的定位本文算法只消耗了约200mAh左右的电量,而GPS定位消耗了约600mAh的电量.本文算法达到了很好的节能目的.

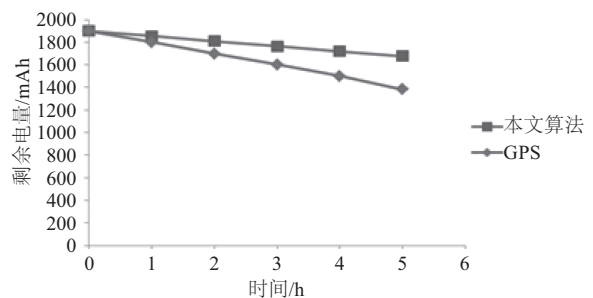


图7 定位过程中,手机电量消耗情况

定位过程的平均能量消耗图,如图8所示,可以看出,本文算法每小时的能耗大约为40 mAh,相比于GPS优势明显。

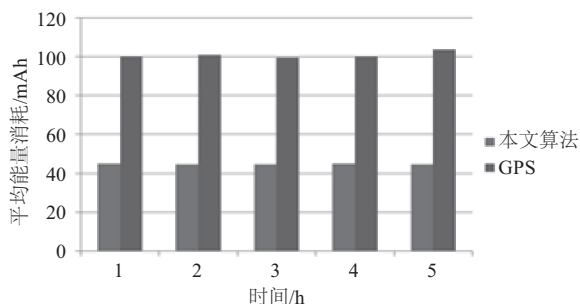


图8 平均能量消耗图

4 结语

本文提出了一种利用隐马尔可夫模型和个人行为习惯的GSM信号定位方法,该算法的特色是充分利用个人的行为习惯,来确定转移概率矩阵,从而使模型更加适应于使用此系统的个人,并且对天气和时段有一定的自适应性,进而达到了提高定位精度和个性化定位的目的。实验的结果表明,本方法可以在保证较低能耗的前提下有效地提高定位精度。

参考文献

- 1 罗军舟, 吴文甲, 杨明. 移动互联网: 终端、网络与服务. 计算机学报, 2011, 34(11): 2029–2051.
- 2 周傲英, 杨彬, 金澈清, 等. 基于位置的服务: 架构与进展. 计算机学报, 2011, 34(7): 1155–1171.
- 3 Xue MH, Liu Y, Ross KW, *et al.* I know where you are: Thwarting privacy protection in location-based social discovery services. Proc. of 2015 IEEE Conference on Computer Communications Workshops. Hong Kong, China. 2015. 179–184.
- 4 Lin K, Kansal A, Lymberopoulos D, *et al.* Energy-accuracy trade-off for continuous mobile devices location. Proc. of the 8th International Conference on Mobile Systems, Applications, and Services. New York, USA. 2010. 285–298.
- 5 Paek J, Kim J, Govindan R. Energy-efficient rate-adaptive GPS-based positioning for smartphones. Proc. of the 8th International Conference on Mobile Systems, Applications, and Services. New York, USA. 2010. 219–324.
- 6 Youssef M, Yosef MA, El-Derini M. GAC: Energy-efficient hybrid GPS-accelerometer-compass GSM localization. Proc. of 2010 IEEE Global Telecommunications Conference.

- Miami, FL, USA. 2010. 1–5.
- 7 Wann CD, Chin HC. Hybrid TOA/RSSI wireless location with unconstrained nonlinear optimization for indoor UWB channels. Proc. of 2007 IEEE Wireless Communications and Networking Conference. Kowloon, China. 2007. 3940–3945.
- 8 Vankayalapati N, Kay S, Ding Q. TDOA based direct positioning maximum likelihood estimator and the cramer-rao bound. IEEE Trans. Aerospace and Electronic Systems, 2014, 50(3): 1616–1635. [doi: 10.1109/TAES.2013.110499]
- 9 Elnahrawy E, Francisco JA, Martin RP. Bayesian localization in wireless networks using angle of arrival. Proc. of the 3rd International Conference on Embedded Networked Sensor Systems. New York, USA. 2005. 272–273.
- 10 Paek J, Kim KH, Singh JP, *et al.* Energy-efficient positioning for smartphones using cell-id sequence matching. Proc. 9th International Conference on Mobile Systems, Applications, and Services. New York, USA. 2011. 293–306.
- 11 Chen MY, Sohn T, Chmelev D, *et al.* Practical metropolitan-scale positioning for gsm phones. International Conference on Ubiquitous Computing. Berlin Heidelberg, Germany. 2006. 225–242.
- 12 Ibrahim M, Youssef M. CellSense: An accurate energy-efficient GSM positioning system. IEEE Trans. Vehicular Technology, 2012, 61(1): 286–296. [doi: 10.1109/TVT.2011.2173771]
- 13 Ibrahim M, Youssef M. A hidden markov model for localization using low-end gsm cell phones. Proc. of 2011 IEEE International Conference on Communications. Kyoto, Japan. 2011. 1–5.
- 14 竹博, 王建辉, 胡捍英, 等. 基于隐马尔可夫模型的Cell-ID定位跟踪方法. 太赫兹科学与电子信息学报, 2013, 11(4): 561–566.
- 15 Nicoli M, Morelli C, Rampa V, *et al.* HMM-based tracking of moving terminals in dense multipath indoor environments. Proc. of the 13th European Signal Processing Conference. Antalya, Turkey. 2005. 1–4.
- 16 Viol N, Link JAB, Wirtz H, *et al.* Hidden markov model-based 3D path-matching using raytracing-generated Wi-Fi models. Proc. of 2012 International Conference on Indoor Positioning and Indoor Navigation. Sydney, NSW, Australia. 2012. 1–10.
- 17 González MC, Hidalgo CA, Barabási AL. Understanding individual human mobility patterns. Nature, 2008, 453(7196): 779–782. [doi: 10.1038/nature06958]
- 18 Simini F, González MC, Maritan A, *et al.* A universal model for mobility and migration patterns. Nature, 2012, 484(7392): 96–100. [doi: 10.1038/nature10856]