

Essay

1. Fingerprinting[14]

1.1 模型

- $S = \{s_1, s_2, \dots, s_N\}$
状态: 路段编号
N: 路段总数
- $V = \{v_1, v_2, \dots, v_M\}$
观测向量: 一条观测向量是一组在 z 个采样点的信号强度序列
M: 所有可能的观测向量数
- $A(N \times N)$: 状态转移概率矩阵
两条假设: 车辆遵循相同道路; 匀速运动. $\Rightarrow a_{ij} = a_k$, where $k = j - i$ and $a_k = 0$ for $k < 0$.
每条道路 \rightarrow 一个矩阵A
- $B(N \times M)$: 每个路段对于每个信号强度序列的概率分布(收敛到高斯分布).
以下这段话不是很懂:

The HMM techniques developed for GSM RSSI-based localization, on the other hand, use either a deterministic function of the Euclidean distance and the number of matching base stations between the observed vector and each RSSI vector in the training database as the cost of the observed RSSI vector [11] or the histogram of the RSSI readings of the associated base station in each road segment [12].

- $\pi(N \times 1)$: 初始化时,各个路段编号的可能性(总和为1)

1.2 训练

- a_k 公式(公式1)
- Kolmogorov–Smirnov test
- b_{jk} 公式(公式2, 归一化处理)

1.3 计算

- RSSI序列(观测向量O) \rightarrow 路段序列(状态Q)
- 维特比算法

1.4 比较

- Ctrack的中值定位误差对于小和大的段长度来说非常大，而S-HMM由于包括驱动器和RSSI统计信息而具有稳定的定位误差;
- S-HMM的中值误差小于所有段长度的Ctrack的中值误差;
- 对于所有段长度，S-HMM的定位误差总是在Ctrack和KNN算法的左侧;
- 随着段长度的增加，Ctrack的成本表现最差，而曼哈顿的成本更稳定，但是仍然比S-HMM差;
- 随着段长度的增加，我们可能需要在HMM中包含状态持续时间密度(As the segment length increases, we may need to include the state duration density in the HMM);
- S-HMM计算需要消耗较多的电量

2. Ctrack[11]

2.1 算法

- Grid Sequencing
- Smoothing & Interpolation
- Segment matching
- Sensor Hint

2.2 HMM建模

- 第一层: Grid Sequencing (去噪)
 - S: Grid cells编号(N: Grid cell数量);
 - V: 一条观测向量是一组GSM fingerprints序列(M: 可能的观测向量数)
一条记录的计算方式(windowing): 聚集每5s扫描到的fingerprints为一个window, 每一个包含到的Grid Cell成为candidate
 - A: Grid cell A -> Grid cell B 概率: transition scores (T (G 1 , G 2))
 - B: 每个状态对于每一个observable的分布概率(Manhattan distance反比): emission score (E P (F 1 , F 2))
 - π : 起点Grid cell分布概率

Viterbi: fingerprint序列 -> windowed序列 ----Viterbi----> Grid Cells的最大似然序列(即使得emission和transition的乘积最大的Grid Cells序列)

- 第二层: Segment Matching
 - S: 三元组{S, H M , H T} (S:路段编号, HM: 当前movement hint, HT: 当前turn hint)
 - V: 一条观测向量是一个点:(lat, lon, H M , H T)
 - A: transition scores
如果S1, S2不相邻为0, 否则为1, 再乘movement penalty & turn penalty (0.1&0.1, 影响小)

- B: emission score

如果HM和HM'或HT和HT'不相等则为0, 否则为高斯距离;

- π : 初始化三元组分布概率

Grid -> 路段编号

2.5 评价

- 节能

CTrack has a significantly better energy-accuracy trade-off than sub-sampling GPS data to save energy, reducing energy cost for the same level of accuracy.

- 效果不如前者

Target

Input: MR数据

Output: 经纬度

前向-后向算法: 模型 -> 观测序列O 概率

Viterbi算法: O+模型 -> 状态序列Q

Baum-Welch算法: 给定O, 确定HMM模型参数

Model Based on Courseware

1. 单层HMM(借鉴fingerprinting):

- S: GridID编号(N: GridID数量)
- V: 一条观测向量是一组在z个主基站的RSSI序列(M: 可能的观测向量数)
一条记录的计算方式: 一条轨迹 -> 按主基站顺序排列RSSI, 空为0
- A: GridID A -> GridID B 概率
- B: 每个GridID对于每一个RSSI序列的分布概率
- π : 第一个主基站的GridID分布概率

MR数据 -> 基站 -> GridID(将中心坐标作为结果)

问题:

数据不是在同一条路上采集, 论文中的方法不可用; 如此计算一条观测向量实际上是将"时间"视为单次采样, 忽略了路径和速度的问题; 另一种计算方法是对数据进行预处理, 判断轨迹所属的类别, 按类计算, 一条记录是一条轨迹以时间先后顺序排列而成的RSSI序列.

2. 双层HMM(借鉴CTrack):

○ 第一层

- S: 栅格编号(N: 栅格数量)
- V: 一条观测向量是一组在z个主基站的RSSI序列(M: 可能的观测向量数)
一条记录的计算方式: 一条轨迹 -> 按主基站顺序排列RSSI, 空为0
- A: 栅格 A -> 栅格 B 概率: transition score
- B: 每个栅格对于每一个RSSI序列的分布概率: emission score
- π : 第一个主基站的栅格分布概率

MR数据 -> 栅格编号

○ 第二层

- S: 二元组{经度, 纬度}(N: 经纬度对总数(总数据量))
- V: 栅格编号(M: 栅格总数)
- A: transition score (考虑定义, 如同栅格相邻或到本身为**1**, 否则为**0**)
- B: emission score (考虑定义, 如同点到栅格中心点的欧氏距离的反比)
- π : 每一个栅格对应的取样点分布概率

栅格编号 -> 经纬度