

高级地图匹配算法:研究现状和趋势

于娟¹, 杨琼², 鲁剑锋¹, 韩建民¹, 彭浩¹

(1. 浙江师范大学数学与计算机科学学院, 浙江金华 321004; 2. 杭州电子科技大学计算机学院, 浙江杭州 310018)

摘 要: 地图匹配是许多位置服务与轨迹挖掘应用的基础. 随着定位技术和位置服务应用的发展, 地图匹配研究不断演进, 从早期基于高采样率GPS(Global Position System)的实时匹配, 到近期基于低采样率GPS轨迹的离线匹配、再到当前非GPS定位数据或高精度地图匹配. 迄今已有许多地图匹配算法相继提出, 但鲜有研究对这些算法进行全面总结. 为此, 对近十年提出的地图匹配算法进行调研, 归纳出地图匹配算法的统一框架及常用时空特征. 从模型或实现技术角度分类发现: 现有算法大都采用HMM(Hidden Markov Model)模型, 其次是最大权重模型; 深度学习技术近期开始用于地图匹配, 将是未来高精度地图匹配研究的趋势.

关键词: 地图匹配; 路网数据; 轨迹数据; HMM; CRF(Conditional Random Fields); 路径推断

中图分类号: TP301.6

文献标识码: A

文章编号: 0372-2112(2021)09-1818-12

电子学报URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20200379

Advanced Map Matching Algorithms: A Survey and Trends

YU Juan¹, YANG Qiong², LU Jian-feng¹, HAN Jian-min¹, PENG Hao¹

(1. College of Mathematics and Computer Science, Zhejiang Normal University, Jinhua, Zhejiang 321004, China;

2. Computer and Software, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China)

Abstract: Map matching is a necessary procedure for many trajectory data mining and various location-based applications. Map matching algorithms are continuously evolving with the development of positioning techniques and application requirements. Research on map matching has undergone several stages, from real-time GPS data map matching, to low-sampling rate GPS trajectories offline map matching, to recently non-GPS positioning data or high resolution map matching. Various advanced map matching algorithms have been proposed. However, there is a short of a complete review of recent map matching algorithms. To bridge this gap, this paper conducts a comprehensive survey on map matching algorithms proposed in the last decade. A general framework of map matching algorithms is extracted, and spatial or spatial-temporal features commonly used in these algorithms are summarized. From the technical perspective, the HMM is the most commonly-used model in existing algorithms, before the maximum weights model. The deep learning technique has been recently applied into map matching, and is becoming a future trend for high resolution map matching.

Key words: map matching; road network; trajectory data; HMM; CRF; route inference

1 引言

地图匹配是结合移动对象的观测位置与路网拓扑或空间布局, 推断其真实位置、或重构真实轨迹的过程. 观测位置是指采用GPS、WiFi、移动蜂窝网等定位技术测量的位置, 与真实位置存在误差. 地图匹配研究源于导航, 是许多位置服务应用的基础. 室内导航或追踪应用中, 地图匹配因无需增加额外硬件成本而成为提升定位精度的常用技术^[1], 典型的有: HTrack^[2]、FreeTrack^[3]、PDMatching^[4]等.

轨迹采集任务中, 地图匹配通过降低采样率要求, 以节省设备能耗、数据传输和存储开销, 同时确保数据质量. 例如: EnAcq^[5]利用地图匹配实现轨迹数据采集的自适应采样; Dong等^[6]结合地图匹配与轨迹压缩, 以降低GNSS(Global Navigation Satellite System)数据采样率要求.

地图匹配也是轨迹数据挖掘的必要预处理过程. 由于收集的位置或轨迹数据是对移动对象活动轨迹的不精确、不完备观测, 存在定位误差和数据缺失, 例如:

收稿日期: 2020-04-20; 修回日期: 2021-05-20; 责任编辑: 覃怀银

基金项目: 国家自然科学基金(No.61702148, No.61672648, No.62072411, No.62072412); 浙江省自然科学基金(No.LR21F020001); 浙江省教育厅一般项目(No.Y201941364)

车辆 GPS 轨迹, 移动手机、智能卡轨迹, 社交媒体签到轨迹等^[7]. 在不对这些轨迹数据进行地图匹配的情况下, 很难直接对其进行分析和挖掘. 地图匹配用低质量的观测轨迹修复或重构真实轨迹, 丰富轨迹数据的空间和语义信息, 进而实现各种轨迹挖掘任务^[8-10].

地图匹配的准确性与观测位置的精度(定位误差大小)、采样率, 及路网拓扑(或室内空间布局)的复杂性密切相关. 早期的地图匹配研究主要关注高采样率 GPS 定位数据的实时或在线匹配, 分为: 基于几何^[11]和基于拓扑的方法^[12]. 基于几何的方法仅考虑路段形状, 未考虑路段连通性, 而基于拓扑的方法则综合考虑了路段形状和连通性. 这两类算法思想简单、易于实现, 但易受定位误差和采样率影响, 仅适用于采样率高且定位误差小的场景, 如: GPS 导航.

近期地图匹配研究关注: (1) 低采样率 GPS 轨迹的离线匹配; (2) 定位误差大的数据的匹配, 例如: 蜂窝网、WiFi 等定位数据; (3) 低采样率且定位误差大的轨迹数据的离线匹配. 采样率降低或定位误差增大都会给地图匹配带来新的挑战. 为适应新挑战, 各种基于 HMM、CRF、智能优化等模型的高级地图匹配算法相继提出. 文献[13]对 2007 年前提出的地图匹配算法进行了综述. 文献[14]对 2012 年前的地图匹配算法进行了总结. 文献[15, 16]对在线地图匹配算法进行分类总结. 文献[17]从应用角度对地图匹配算法进行分类总结, 并探讨了不同应用对地图匹配算法的要求. 文献[18]对基于 HMM 的地图匹配算法进行了总结.

尽管已有一些综述文献, 但它们尚未对近期提出的地图匹配算法进行全面综述. 为此, 本文对近十年的地图匹配算法进行全面调研. 首先, 总结出地图匹配算法的统一框架, 并对地图匹配研究中所采用的特征进行分类. 其次, 从实现技术角度将地图匹配算法分为: 基于 HMM、基于 CRF、基于最大权重、基于局部路径推断、基于深度学习, 以及其他类型. 再次, 从适用场景和数据需求角度对比算法. 最后, 分析现有匹配算法存在的问题, 探讨可能的发展方向.

2 地图匹配的统一框架

为便于描述, 本节先给出相关的基本概念.

2.1 基本概念

定义 1 (采样点) 采样点或位置观测点, 可描述为至少包含时间和经纬度的元组, 即 $p=[t, \text{lat}, \text{lon}]$ 或 $p=[t, \text{lat}, \text{lon}, v, h]$, 其中 t 为采样时间, lat 为纬度, lon 为经度, v 为瞬时速度, h 为实时方向.

以下将用 $p.t, p.\text{lat}, p.\text{lon}, p.v, p.h$ 分别表示采样点 p 的采样时间、纬度、经度、瞬时速度和实时方向.

定义 2 (轨迹) 轨迹定义为按时间顺序收集的同一

移动对象的采样点序列, 记为 $T=p^1 \rightarrow p^2 \rightarrow \dots \rightarrow p^i \rightarrow \dots$, 其中 $p^i.t < p^{i+1}.t, 1 \leq i < |T|, |T|$ 为轨迹中所包含采样点的数量.

定义 3 (子轨迹) 子轨迹定义为轨迹的连续片段, 例如: T 的子轨迹记为: $T_{ij}=p^i \rightarrow \dots \rightarrow p^j$, 其中 $1 \leq i < j \leq |T|$.

定义 4 (路网) 路网定义为一个有向图 $G=\{V, E\}$, 其中, 顶点 $v \in V$ 表示一个路口或路段的终点, $e \in E$ 表示有向边, 对应路网中一个单向路段. 每个顶点包含经纬度和唯一标识, 记为 $v=\{\text{lat}, \text{lon}, \text{id}\}$. 每条边记为 $e=\{\text{id}, \text{end}, \text{start}, \text{geo}, \text{vel}, \text{len}\}$, 其中, id 为路段唯一标识, start 为路段起始点, end 为路段终点, geo 为路段几何形状, vel 为路段限速, len 为路段长度.

定义 5 (候选路段集) 给定采样点 p^i 和路网 $G=\{V, E\}$, p^i 在 G 上的候选路段集定义为: $E_C^i=\{e \in E \mid d(p^i, e) < \delta, e \in E\}$, 其中 $d(p^i, e)$ 为 p^i 到 e 的最短距离, δ 为指定距离阈值.

定义 6 (候选点集) 候选点定义为候选路段上离对应采样点最近的点. 对每个候选路段 $e_j \in E_C^i$, 记其候选点为 c_j^i , 上标 i 表示采样点, 下标 j 表示候选路段, $1 \leq j \leq |E_C^i|$. 采样点 p^i 的所有候选点构成的集合称候选点集, 记为 C^i .

定义 7 (路径) 路径定义为路段序列, 例如: $R_{c_k^i, c_j^{i+1}}=\{e_1 \rightarrow \dots \rightarrow e_m\}$, 表示候选点 c_k^i 和 c_j^{i+1} 在路网上的一条路径, 其中, $e_1=e_k^i$ 为候选点 c_k^i 所在的路段, $e_m=e_j^{i+1}$ 为候选点 c_j^{i+1} 所在的路段, $m=|R_{c_k^i, c_j^{i+1}}|$ 为路径中包含的路段数.

定义 8 (路径长度) 路径长度或路径距离定义为路段长度之和, 例如: 路径 $R_{c_k^i, c_j^{i+1}}$ 的长度 $L(R_{c_k^i, c_j^{i+1}})=d_e(c_k^i, e_1.\text{end})+\sum_{t=2}^{m-1} e_t.\text{len}+d_e(e_m.\text{start}, c_j^{i+1})$, 其中, $d_e(c_k^i, e_1.\text{end})$ 为路段 e_1 上从 c_k^i 到其终点的长度, $d_e(e_m.\text{start}, c_j^{i+1})$ 为路段起点到 c_j^{i+1} 的长度.

定义 9 (路网最短距离) 给定两个相邻采样点 p^i 和 p^{i+1} , 它们之间的路网最短距离定义为其对应候选点 c_k^i 和 c_j^{i+1} 在路网上的最短路径距离, 记为 $L(R_{c_k^i, c_j^{i+1}}^*)$, 其中 $R_{c_k^i, c_j^{i+1}}^*$ 为 c_k^i 和 c_j^{i+1} 在路网上的距离最短路径.

2.2 统一框架

地图匹配可概括为图 1 所示的统一框架, 包含四个模块: (1) 候选路段/点搜索, 据观测位置搜索移动对象在路网中的可能路段、及其在该路段上的可能位置, 搜索方法有: k -近邻搜索、区间搜索、对齐高^[19], 其中, 前两种最常用; (2) 基于候选路段/点的特征提取和时空关系建模, 构建一个以候选路段/点为顶点的加权候选图, 常用特征见 2.3 节. 时空关系建模方法有: HMM, CRF 等; (3) 最优路径搜索, 从加权候选图中寻找最大化(或最小化)目标函数的最优路径(即最优候选点或候选路段序列), 常用

搜索方法有 Viterbi 算法、双向 Dijkstra 算法^[20], 其中 Viterbi 算法最常用; (4) 轨迹预处理, 在地图匹配前对轨迹

进行简化^[21]、降噪^[22]、力导引绘图^[23]等预处理, 以提升匹配准确率或效率. 轨迹数据预处理是可选模块.

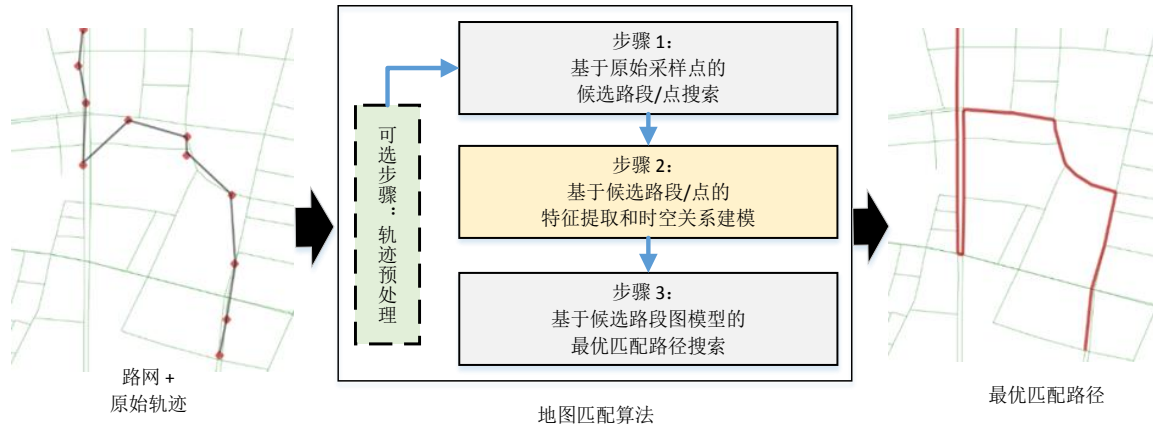


图1 地图匹配的统一框架

2.3 地图匹配常用特征

2.3.1 空间特征

地图匹配算法中常用的空间特征主要包括:

- (1) d_k^i : 位置采样点 p^i 到候选点 c_k^i 的直线距离;
- (2) $A(p^i, e_k^i)$: p^i 的运动方向与候选路段 e_k^i 的方向夹角;
- (3) $A(e_t, e_{t+1})$: 路段 e_t 与 e_{t+1} 的方向夹角;
- (4) $A(H(p^i, p^{i+1}), H(c_k^i, c_j^{i+1}))$: 方向夹角, $H(p^i, p^{i+1})$ 表示从 p^i 到 p^{i+1} 的方向, $H(c_k^i, c_j^{i+1})$ 表示从 c_k^i 到 c_j^{i+1} 的方向;
- (5) $d(p^i, p^{i+1})$: 相邻采样点间的直线距离;
- (6) $d(c_k^i, c_j^{i+1})$: 相邻候选点间的直线距离;
- (7) $L(R_{c_k^i, c_j^{i+1}})$: 局部路径长度;

$$(8) \psi(A(e_t, e_{t+1})) = \begin{cases} 0, & \text{if } A(e_t, e_{t+1}) < \frac{\pi}{4}, \\ 1, & \text{if } \frac{\pi}{4} \leq A(e_t, e_{t+1}) \leq \frac{3\pi}{4}, \\ 2, & \text{if } \frac{3\pi}{4} < A(e_t, e_{t+1}) < \pi, \\ 10, & \text{if } A(e_t, e_{t+1}) = \pi. \end{cases}$$

相邻路段的转弯开销^[24];

- (9) $TC(R_{c_k^i, c_j^{i+1}}) = \sum_{t=1}^{m-1} \psi(A(e_t, e_{t+1}))$: 局部路径转弯开销^[24];

- (10) $S_{c_k^i, c_j^{i+1}}^{d,1} = d(p^i, p^{i+1}) / L(R_{c_k^i, c_j^{i+1}})$: 直线距离与路径距离相似度^[25];

$$(11) S_{c_k^i, c_j^{i+1}}^{d,2} = \frac{\min(d(p^i, p^{i+1}), L(R_{c_k^i, c_j^{i+1}}))}{\max(d(p^i, p^{i+1}), L(R_{c_k^i, c_j^{i+1}}))}: \text{直线距}$$

离与路径距离的相似度^[26];

$$(12) \Delta_{c_k^i, c_j^{i+1}}^{d,1} = |L(R_{c_k^i, c_j^{i+1}}) - d(p^i, p^{i+1})|: \text{距离偏差}^{[27]};$$

- (13) $\Delta_{c_k^i, c_j^{i+1}}^{d,2} = |L(R_{c_k^i, c_j^{i+1}}) - v\Delta t_i| / L(R_{c_k^i, c_j^{i+1}})$: 距离偏差^[28,29], 其中 v 为观测平均速度;

- (14) $\Delta_{c_k^i, c_j^{i+1}}^{d,3} = (L(R_{c_k^i, c_j^{i+1}}) - d(c_k^i, c_j^{i+1})) / \Delta t_i$: 路径迂曲度 (circuitousness)^[30];

- (15) $\text{Freq}(R_{c_k^i, c_j^{i+1}})$: 路径使用频率 (根据历史数据统计)^[31].

采样点与候选点间的距离特征 d_k^i 用以刻画采样点的定位误差, 是最常用的一种空间特征, 且大都假设该距离特征服从正态分布. 特征 $S_{c_k^i, c_j^{i+1}}^{d,2}$ 是对 $S_{c_k^i, c_j^{i+1}}^{d,1}$ 的改进. $S_{c_k^i, c_j^{i+1}}^{d,1}$ 的取值落在区间 $[0, \infty)$ 上, 当相邻采样点的候选点重合时, $L(R_{c_k^i, c_j^{i+1}}) = 0$, 有 $S_{c_k^i, c_j^{i+1}}^{d,1} = \infty$; 当相邻采样点对应的候选点落在同一路段上时, 可能有 $d(p^i, p^{i+1}) > L(R_{c_k^i, c_j^{i+1}})$, 则有 $S_{c_k^i, c_j^{i+1}}^{d,1} > 1$. 而 $S_{c_k^i, c_j^{i+1}}^{d,2}$ 可以确保其值落在区间 $[0, 1]$ 上, 具有归一化效果.

2.3.2 时空特征

地图匹配算法中常用的时空特征包括:

- (1) $\Delta t_i = p^{i+1}.t - p^i.t$: 实际时间开销 (s);
- (2) $\Delta \tilde{t}_i = \sum_{t=1}^m (e_t \cdot \text{len} / v_t^f)$: 基于路段自由车速度的时间开销 (s)^[30], v_t^f 为路段 e_t 的自由车流速度 (区别于路段限速);
- (3) $\Delta \hat{t}_i = \sum_{t=1}^m (e_t \cdot \text{len} / v_t^h)$: 基于历史平均速度的时间开销 (s)^[32], v_t^h 为路段 e_t 的历史速度;

(4) $\Delta_{e_k, e_j^{i+1}}^{t,1} = \max\{(\Delta \tilde{t}_i - \Delta t_i), 0\} / \Delta t_i$: 时间开销偏离度^[30];

(5) $S_{e_k, e_j^{i+1}}^{t,1} = \min\{\Delta t_i, \Delta \tilde{t}_i\} / \max\{\Delta t_i, \Delta \tilde{t}_i\}$: 时间开销相似度^[32];

(6) $\tilde{v}_{e_k, e_j^{i+1}} = (\mathbf{p}^i \cdot \mathbf{vel} + \mathbf{p}^{i+1} \cdot \mathbf{vel}) / 2$: 基于瞬时速度的平均速度估计;

(7) $\bar{v}_{e_k, e_j^{i+1}} = L(R_{e_k, e_j^{i+1}}) / \Delta t_i$: 基于距离的平均速度估计;

(8) $\Delta_{e_k, e_j^{i+1}}^v = |\bar{v}_{e_k, e_j^{i+1}} - \tilde{v}_{e_k, e_j^{i+1}}|$: 平均速度差^[33];

(9) $\mathbf{v}_{e_k, e_j^{i+1}}^1 = [\hat{v}_1, \hat{v}_2, \dots, \hat{v}_m]$: 路段限速, $\hat{v}_\ell = \mathbf{e}_\ell \cdot \mathbf{vel}$, $1 \leq \ell \leq m$;

(10) $\mathbf{v}_{e_k, e_j^{i+1}}^2 = [v_1^h, v_2^h, \dots, v_m^h]$: 路段历史平均速度^[34];

(11) $\mathbf{v}_{e_k, e_j^{i+1}}^3 = [\bar{v}_1, \bar{v}_2, \dots, \bar{v}_m]$: 路段加权速度(历史平均速度与路段速度的加权和)^[35];

(12) $S_{e_k, e_j^{i+1}}^{v,1} = S_c\left(\mathbf{v}_{e_k, e_j^{i+1}}^1, \left[\bar{v}_{e_k, e_j^{i+1}}\right]_m\right)$: 速度相似性(路段限速 \hat{v}_ℓ 与 $\bar{v}_{e_k, e_j^{i+1}}$)^[25,29], 其中 S_c 为余弦相似度函数;

(13) $S_{e_k, e_j^{i+1}}^{v,2} = S_c\left(\mathbf{v}_{e_k, e_j^{i+1}}^2, \left[\bar{v}_{e_k, e_j^{i+1}}\right]_m\right)$: 速度相似性(路段历史平均速度 v_ℓ^h 与 $\bar{v}_{e_k, e_j^{i+1}}$)^[34];

(14) $S_{e_k, e_j^{i+1}}^{v,3} = S_c\left(\mathbf{v}_{e_k, e_j^{i+1}}^3, \left[\bar{v}_{e_k, e_j^{i+1}}\right]_m\right)$: 速度相似性(路段加权速度 \bar{v}_ℓ 与 $\bar{v}_{e_k, e_j^{i+1}}$)^[35];

(15) $S_{e_k, e_j^{i+1}}^{v,4} = \hat{v} / \left(\left| \hat{v} - \bar{v}_{e_k, e_j^{i+1}} \right| + \hat{v} \right)$: 速度相似性(加权路段限速 $\hat{v} = \sum_{\ell=1}^m \omega_\ell \cdot \hat{v}_\ell$ 与 $\bar{v}_{e_k, e_j^{i+1}}$)^[36];

(16) $S_{e_k, e_j^{i+1}}^{v,5} = \hat{v}_1 / \left(\left| \hat{v}_m - \bar{v}_1 \right| + \hat{v}_1 \right)$: 速度相似性(路段 e_k^i 限速 \hat{v}_1 与路段 e_j^{i+1} 限速 \hat{v}_m)^[36];

(17) $P_{e_k, e_j^{i+1}}^{v,1} = \mathbf{e}_k^i \cdot \mathbf{v} / (\max\{0, \mathbf{p}^i \cdot \mathbf{v} - \mathbf{e}_k^i \cdot \mathbf{v}\} + \mathbf{e}_k^i \cdot \mathbf{v})$: 超速惩罚因子(观测速度与候选路段限速)^[28];

(18) $P_{e_k, e_j^{i+1}}^{v,2} = \frac{\sum_{\ell=1}^m (\mathbf{e}_\ell \cdot \mathbf{len}) \cdot (\mathbf{e}_\ell \cdot \mathbf{v})}{\bar{v}_{e_k, e_j^{i+1}} \cdot L(R_{e_k, e_j^{i+1}})}$: 动量特征^[28].

上述时空特征可分为:以时间间隔为参照的特征和速度相关特征.局部路径的时间开销偏离度 $\Delta_{e_k, e_j^{i+1}}^{t,1}$ 的值域是 $[0, +\infty)$, 越接近于 0, 对应局部路径为真实路径的可能性越大.时间开销相似度 $S_{e_k, e_j^{i+1}}^{t,1}$ 的值域是 $[0, 1]$, 越接近 1, 对应局部路径为真实路径的可能性越大.速度相关特征 $S_{e_k, e_j^{i+1}}^{v,1}, S_{e_k, e_j^{i+1}}^{v,2}, S_{e_k, e_j^{i+1}}^{v,3}$ 值域都是 $[-1, 1]$, 越接近 1, 对应局部路径为真实路径的可能性越大.但由于车辆在路网上

的行驶速度会受路况、红绿灯等诸多因素影响,导致实际速度与路段限速可能相差较大,进而导致 $S_{e_k, e_j^{i+1}}^{v,1}$ 很难真实反映时空关联性.于是就有改进的基于路段历史平均速度和路段加权速度的相似度特征 $S_{e_k, e_j^{i+1}}^{v,2}$ 和 $S_{e_k, e_j^{i+1}}^{v,3}$, 但在无历史数据的情况下都无法使用. $S_{e_k, e_j^{i+1}}^{v,4}$ 与 $S_{e_k, e_j^{i+1}}^{v,5}$ 的值域是 $[0, 1]$, 越接近 1, 局部路径为真实路径的可能性越大.超速惩罚因子 $P_{e_k, e_j^{i+1}}^{v,1}$ ^[32] 的值域为 $(0, 1]$, 值越小表明对候选路段为真实路段的可能性越小.动量特征 $P_{e_k, e_j^{i+1}}^{v,2}$ 度量的是局部路径运动速度的平滑性.

3 典型地图匹配算法分类

图 2 从实现技术或模型角度对近十年提出的算法进行分类,箭头标记算法间的继承关系.从图 2 可见, HMM 模型是主流, 其次是基于最大权重的模型. HMM-News-on^[27]、ST-Matching^[25]、IVMM^[37] 和 HRIS^[38] 被引用对比最多, 是具有开创性的工作.另外, 从 2019 年开始, 有研究采用深度学习技术来解决地图匹配问题.

3.1 基于 HMM 的地图匹配算法

基于 HMM 的地图匹配算法将位置采样点视为观测, 将真实位置或真实路段视为隐状态, 地图匹配即为 HMM 的解码问题, 即求解最优隐状态序列, 可形式化定义如下.

给定轨迹 T 和路网 $G = \{V, E\}$, 基于 HMM 的地图匹配即为求解最大联合概率的状态序列:

$$R^* = \arg \max_{c_{k_1}^1 \rightarrow \dots \rightarrow c_{k_n}^n} \left\{ \Pr(c_{k_1}^1, \dots, c_{k_n}^n, \mathbf{p}^1, \dots, \mathbf{p}^n) \right\} \quad (1)$$

其中, $\{c_{k_1}^1, c_{k_2}^2, \dots, c_{k_n}^n\}$ 为状态序列, 即为匹配结果. 根据 HMM 的独立性假设, 式(1)的联合概率密度可简化为:

$$\Pr(c_{k_1}^1, c_{k_2}^2, \dots, c_{k_n}^n, \mathbf{p}^1, \mathbf{p}^2, \dots, \mathbf{p}^n) = \Pr(\mathbf{p}^1 | c_{k_1}^1) \cdot \prod_{i=2}^n \Pr(\mathbf{p}^i | c_{k_i}^i) \cdot \Pr(c_{k_i-1}^{i-1} | c_{k_i}^i) \quad (2)$$

其中, $\Pr(\mathbf{p}^i | c_{k_i}^i)$ 为观测概率, $\Pr(c_{k_i}^i | c_{k_{i-1}}^{i-1})$ 为状态转移概率.

表 1 从观测概率和转移概率的计算角度对基于 HMM 的算法进行分类. 从表 1 可见, 观测概率和转移概率的计算都依赖于具体的空间或时空特征(详见 2.3), 且大都假定特征满足某种模型. 例如: 假设采样点与候选点间的距离服从高斯分布, 相邻采样点的地表距离与对应的候选点在路网上的局部路径长度的差异服从指数分布等. 另外, 还有部分算法通过设计有效的解码算法实现实时地图匹配. 基于 HMM 的算法之间的主要区别在于观测概率和转移概率的计算方式.

3.2 基于 CRF 的地图匹配算法

基于 CRF 的地图匹配算法由模型训练和匹配两阶段构

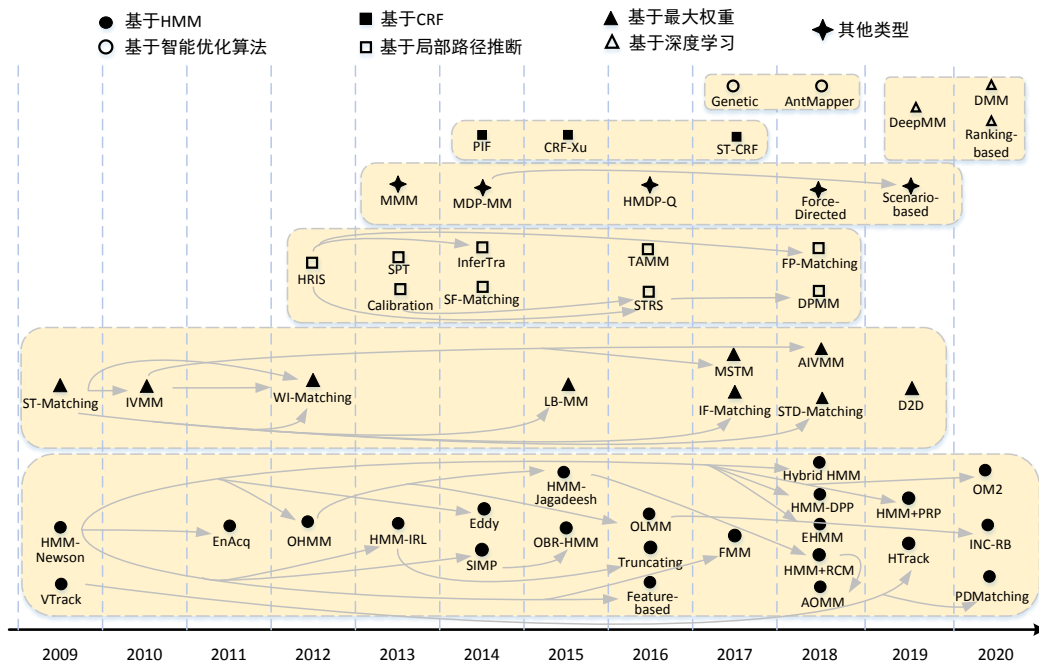


图2 地图匹配算法发展历程 (2009~2020)

表1 典型的基于HMM的地图匹配算法

算法	观测概率 $\Pr(p^i c_j^i)$	转移概率 $\Pr(c_j^{i+1} c_k^i)$
HMM-Newson ^[27] EnAcq ^[5] , OM2 ^[39] Eddy ^[40]	$f_N(d_k^i \mu, \sigma^2)$, 其中 f_N 为正态分布函数, μ, σ 为参数, 下同	$f_E\left(A_{c_k^i, c_j^{i+1}}^{d,1} \beta\right)$, 其中 $f_E(\cdot)$ 为指数分布函数, β 为参数, 下同
OHMM ^[28]	$\frac{1}{w} \int_{-w/2}^{w/2} f_N(x d_k^i, \sigma^2) dx$, 其中 w 为候选点 c_k^i 所在路段宽度	$f_E\left(\omega_1 A_{c_k^i, c_j^{i+1}}^{d,2} + \omega_2 P_{c_k^i, c_j^{i+1}}^{n,2} \beta\right)$, 其中 ω_1, ω_2 为权重参数
HMM-IRL ^[24]	$f_N(d_k^i \mu, \sigma^2)$	$f_E\left(A_{c_k^i, c_j^{i+1}}^{d,1} + \omega \cdot \text{TC}\left(R_{c_k^i, c_j^{i+1}}\right) \beta\right)$, 其中 ω 为权重系数
HMM-Jagadeesh ^[41]		$f_E\left(A_{c_k^i, c_j^{i+1}}^{d,3} \beta_1\right) \times f_E\left(A_{c_k^i, c_j^{i+1}}^{d,1} \beta_2\right)$
Feature-based ^[42] INC-RB ^[43]		$f_E\left(X \frac{ T_{i,i+1} }{ T d(p^i, p^{i+1})} \beta\right)$, 其中 $X = \sum_{t=1}^{m-1} Y \cdot \left(e_t, \text{len} + \omega \cdot \psi\left(A(e_t, e_{t+1})\right)\right)$, ω 为权重系数, $Y = \min\{d(e_t, T_{i,i+1}), \delta_d\}$, δ_d : 给定距离阈值
HMM-DPP ^[33]		$\begin{cases} f_E\left(A_{c_k^i, c_j^{i+1}}^{d,1} \beta_1\right) \times f_E\left(A_{c_k^i, c_j^{i+1}}^{d,2} \beta_2\right), & \text{if } \Delta t_i \leq 80s \\ \omega_1 \left(f_E\left(A_{c_k^i, c_j^{i+1}}^{d,1} \beta_1\right) \times f_E\left(A_{c_k^i, c_j^{i+1}}^{d,2} \beta_2\right)\right) + \omega_2 M, & \text{otherwise} \end{cases}$ 其中 M 为路径偏差模型 ^[37] , ω_1 和 ω_2 为权重系数, $\omega_1 + \omega_2 = 1$
EHMM ^[44]	$\begin{cases} 0, & \text{if } A(p^i, e_k^i) \geq \delta_A \\ Z, & \text{otherwise.} \end{cases}$ 其中 δ_A 为给定夹角阈值, $Z = f_E(d_k^i \beta) \times \cos(A(p^i, e_k^i)) $	$\begin{cases} f_E(x \beta_1) \times f_E\left(A(p^i, H(c_k^i, c_j^{i+1})) \beta_2\right) \times \\ f_E\left(A(H(p^i, p^{i+1}), H(c_k^i, c_j^{i+1})) \beta_3\right), & \text{if } \Delta t_i < 60s \\ f_E\left(1/S_{c_k^i, c_j^{i+1}}^{d,1} - 1 \beta_4\right), & \text{otherwise} \end{cases}$ 其中当 $e_k^i \cdot \text{end} = e_j^{i+1} \cdot \text{start}$ 时, $x=1$, 当 $e_k^i = e_j^{i+1}$ 时, $x=0$, 否则 $x=\infty$
AOMM ^[29]	$f_N(d_k^i \mu, \sigma^2)$	$\left(1 - A_{c_k^i, c_j^{i+1}}^{d,2}\right) \times S_{c_k^i, c_j^{i+1}}^{n,1}$
Hybrid HMM ^[34]		$L\left(R_{c_k^i, c_j^{i+1}}\right) \times \left(2 - S_{c_k^i, c_j^{i+1}}^{n,2}\right)$

成:(1) 模型训练:用极大对数似然估计学习模型参数,即

$$w^* = \arg \max_w \sum_i \log \Pr(R^i | T^i; w) \quad (3)$$

其中 w 为模型参数, T^i 表示第 i 条轨迹, R^i 表示 T^i 对应的匹配路径。(2) 匹配:用 Viterbi 算法求解最优状态序列作为匹配结果,即

$$R^* = \arg \max_R \Pr(R | T, w) \quad (4)$$

$$\Pr(R | T, w) =$$

$$\exp \left\{ \sum_{i=1}^{|T|} \omega_i \cdot \varphi(p^i, c_k^i) + \sum_{i=1}^{|T|-1} \sum_{k=1}^K \omega_k \cdot \phi(c_k^i, c_k^{i+1}) \right\} \quad (5)$$

其中 $\varphi(p^i, c_k^i)$ 为状态特征函数, $\phi(c_k^i \rightarrow c_k^{i+1})$ 为转移特征函数, K 为转移特征函数的个数。

基于 CRF 的算法主要有:CRF-Liao^[45], CRF-Xu^[32], ST-CRF^[46], 以及 PIF (Path Inference Filter)^[47]. 其中 PIF 是一个基于 CRF 的通用地图匹配框架, 可根据数据和具体应用需求合理选择特征函数. 但 PIF 存在过拟合、效率低等问题. 为此, 文献[48]结合 ℓ_1 正则化以避免 PIF 的过拟合问题. 现有基于 CRF 的算法都采用线性链 CRF, 主要区别在于转移特征函数不同.

表 2 典型的基于最大权重的地图匹配算法

算法	权重 $W(c_k^i, c_k^{i+1})$ 计算
ST-Matching ^[25] IVMM ^[37] , LB-MM ^[50]	$f_N(d_j^{i+1} \mu, \sigma^2) \times S_{c_k^i, c_k^{i+1}}^{d,1} \times S_{c_k^i, c_k^{i+1}}^{v,1}$
WI-Matching ^[51]	$\omega_d f_d + \omega_h f_h + \omega_c f_c + \omega_i f_i$, 其中 $f_d = (\delta_e - 0.5(d_k^i + d_j^{i+1})) / \delta_e$, $\delta_e = 100$; $f_h = (\delta_A - 0.5(A(p^i, e_k^i) + A(p^{i+1}, e_j^{i+1}))) / \delta_A$, $\delta_A = 60$ 度; $f_c = (I_{s,k}^{i-1,i} + I_{k,t}^{i,i+1}) / 2$, 如果 c_s^{i-1} 到 c_t^i 在路网上存在可达路径则有 $I_{s,k}^{i-1,i} = 1$, 否则 $I_{s,k}^{i-1,i} = 0$; $f_i = (\delta_T - 0.5(TC(R_{c_k^{i-1}, c_t^i} + TC(R_{c_k^i, c_j^{i+1}}))) / \delta_T$, δ_T 为给定阈值; $\omega_d, \omega_h, \omega_c, \omega_i$ 为实验确定的权重系数.
IF-Matching ^[35]	$f_N(d_j^{i+1} \mu, \sigma^2) \times S_{c_k^i, c_k^{i+1}}^{d,1} \times (1 - \arccos(A(p^{i+1}, e_j^{i+1})) / 2\pi) \times S_{c_k^i, c_k^{i+1}}^{v,3}$
FMM ^[26]	$f_N(d_j^{i+1} \mu, \sigma^2) \times S_{c_k^i, c_k^{i+1}}^{d,2}$
MSTM ^[52]	$f_N(d_k^i \mu, \sigma^2) \times Q \times f_N(d_j^{i+1} \mu, \sigma^2)$, 当 $ R_{c_k^i, c_j^{i+1}} \leq 2$ 时, $Q=1$, 否则 $Q = f_E \left(\left(2 - R_{c_k^i, c_j^{i+1}} \right) / \beta \right)$
STD-Matching ^[53]	$f_N(d_j^{i+1} \mu_1, \sigma_1^2) \times S_{c_k^i, c_k^{i+1}}^{d,1} \times S_{c_k^i, c_k^{i+1}}^{v,1} \times f_N(A(p_j^{i+1}, e_j^{i+1}) \mu_2, \sigma_2^2)$
AIVMM ^[36]	$f_N(d_j^{i+1} \mu, \sigma_1^2) \times S_{c_k^i, c_k^{i+1}}^{d,1} \times S_{c_k^i, c_k^{i+1}}^{v,4} \times S_{c_k^i, c_k^{i+1}}^{v,5}$
FP-based ^[31]	$\frac{(1+\omega)f_N(d_k^i \mu, \sigma^2) f_N(d_j^{i+1} \mu, \sigma^2)}{\omega f_N(d_k^i \mu, \sigma^2) + f_N(d_j^{i+1} \mu, \sigma^2)} \times \text{Freq}(R_{c_k^i, c_j^{i+1}})$, 其中 ω 为权重系数

3.3 基于最大权重的地图匹配算法

给定轨迹 T 和路网 $G = \{V, E\}$, 基于最大权重的地图匹配算法可以描述为:

$$R^* = \arg \max_{R^c \in G^c} \{W(R^c)\} \quad (6)$$

其中 $G^c = \{V^c, E^c\}$ 为候选图, 是以所有采样点的候选点为顶点集, 相邻采样点的候选点间的联系为边集的有向无环图, 即 $V^c = \{C^1 \cup C^2 \cup \dots \cup C^n\}$, $E^c = \{(c_k^i, c_j^{i+1}) | c_k^i \in C^i, c_j^{i+1} \in C^{i+1}, 1 \leq i < n\}$. R^c 为候选图上以候选点集 C^1 中的任一候选点为起点, 候选点集 C^n 中的任一候选点为终点的一条路径. $W(R^c)$ 表示路径 R^c 的权重, 即路径上所有边的权重之和:

$$W(R^c) = \sum_{i=1}^{n-1} W(c_k^i, c_j^{i+1}) \quad (7)$$

其中 $W(c_k^i, c_j^{i+1})$ 为候选图中边的权重, $c_k^i \in C^i, c_j^{i+1} \in C^{i+1}$.

表 2 从局部权重计算方法角度, 对基于最大权重的地

图匹配算法进行分类总结. 此类算法之间的区别主要在于: 局部路径和候选点特征的定义和选择. 类似地, MDP-MM (Multi-criteria Dynamic Programming Map Matching) 算法^[49] 将地图匹配问题建模为最小化距离权重优化问题.

3.4 基于智能优化的地图匹配算法

为简化问题, 基于 HMM、CRF 和最大权重的算法采用的特征都源于单个或相邻采样点以及其对应的候选路段/点, 而未考虑轨迹的整体特征. 为考虑全局信息, 研究设计了基于智能优化算法的匹配算法: 基于遗传算法的 Genetic 算法^[54] 和基于蚁群算法的 AntMapper^[55]. 两者的区别是: (1) 适应性函数定义不同; (2) 优化算法不同.

3.5 基于局部路径推断的地图匹配算法

前几类算法都直接假定相邻候选点间的局部路径为距离最短或时间最快路径. 但该假设并不适用于所有场景, 尤其是数据采样率低且路网密集的场景. 为此, 研究者们提出基于局部路径推断的匹配算法, 关注推断采样间隔时间长的相邻采样点间的局部路径.

基于局部路径推断的典型算法有:HRIS^[38]、Infer-Tra^[56]、SF-Matching^[57]、STRS^[58]、TAMM^[59]、FP-Matching^[31]、HMM+RCM^[30]。另外,单个采样点匹配是极端情况,相关研究很少,仅Wu等^[60]提出基于历史GPS数据的生成式模型。

3.6 基于深度学习的地图匹配算法

深度学习近来也开始用于地图匹配,例如:Deep-MM^[61]、基于排名学习的算法^[62]和DMM^[63]。但深度学习在地图匹配中的应用尚处于起步阶段。这类算法的劣势在于:需要大量的标记数据训练模型参数,而带标记的训练数据不易获取。

3.7 其他类型地图匹配算法

考虑到车辆在路网上的运动并不满足Markov条件独立假设,尤其是已知目的地时。为此,文献[64]提出基于 k -最短路径的地图匹配算法。区别于其他算法,它仅根据轨迹的起始和终点搜索候选路径。类似还有SPT(Shortest Path in Time)算法^[65],SPT选择最优路径时综合考虑路径的空间和时间全局特征,而前者仅考虑空间全局特征。

为利用大量GPS轨迹中的规律,提升匹配效率和

准确性,文献[66]提出多轨迹地图匹配算法MMM(Multi-track Map Matching)。基于场景的地图匹配^[67]在MDP-MM^[49]基础上根据场景选用不同的匹配策略。HMDP-Q^[68]是基于强化学习的地图匹配算法。

为克服缺乏标记数据的问题,有研究提出交互式地图匹配方法,借助可视化手段,让人参与到地图匹配过程中,利用人的知识优化地图匹配算法的参数配置、纠正错误匹配,例如:VIMM(Visual Interactive Map Matching)^[69]等。

4 典型地图匹配算法对比分析

4.1 应用需求角度的对比分析

表3从应用场景、改进方法和目标的角度对典型算法进行对比。其中基于GPS定位的离线匹配的研究居多,大都关注采样率对匹配准确率的影响。根据改进目标的不同,可分为:(1)匹配准确率提升。(2)匹配效率提升。提升地图匹配效率的方法可以概括为:(a)空间索引,(b)轨迹简化,(c)并行计算,(d)高级路径规划算法。(3)离线算法改造为在线算法,通过设计在线的最优路径搜索算法实现,要解决的主要问题是匹配准确率与时延的权衡。

表3 典型地图匹配算法应用场景角度的对比

应用场景	算法	采样间隔	改进方法和目标
实时GPS导航	EnAcq ^[5]	5~30s	利用自适应采样,降低采样率
	OHMM ^[28]	1~300s	利用变长滑动窗,降低时延
	Eddy ^[40]	1~120s	利用自适应滑动窗,降低时延
	OLMM ^[70]	/	利用变长滑动窗,降低时延
	HMM-RCM ^[30]	1~5min	利用路径选择模型,提高准确率、降低时延
	AOMM ^[29]	30~210s	优化转移概率、利用自适应滑动窗,降低时延
	HMM+PRP ^[71]	1~20s	利用概率路径预测,降低时延
离线WiFi轨迹挖掘	INC-RB ^[43]	60~180s	利用回滚机制,降低时延
	HTrack ^[2]	/	优化转移概率,提高准确率
	PDMatching ^[4]	1~6m	
离线GPS轨迹挖掘	SIMP ^[21]	1~60s	简化轨迹,提高效率
	OBR-HMM ^[72]	1~64s	逐段匹配,提高效率
	FMM ^[26]	/	利用预计算,提高效率
	Truncating ^[73]	30~300s	优化最短路径计算,提高效率
	Feature-based ^[42]	1~5min	优化转移概率,提高效率
	EHMM ^[44]	5~150s	优化观测、转移概率,提高准确率
	OM2 ^[39]	/	利用轨迹分割、后处理,提高准确率
	HMM-IRL ^[24]	30~360s	优化转移概率,提高准确率
	HMM-DPP ^[33]	20~180s	
	PIF ^[47]	1~120s	利用模型改进,提高准确率
	CRF-Xu ^[32]	180~420s	优化转移概率,提高准确率
	ST-CRF ^[46]	5~120s	
	CRF-Yang ^[48]	/	利用 ℓ_1 正则化,提高准确率
	ST-Matching ^[25]	30~300s	优化权重度量,提高准确率
	LB-MM ^[50]	2~6min	

续表			
应用场景	算法	采样间隔	改进方法和目标
离线 GPS 轨迹挖掘	IF-Matching ^[35]	/	
	MSTM ^[52]	5~30min	
	STD-Matching ^[53]	1~120s	
	AIVMM ^[36]	/	
	FP-based ^[31]	1~5min	
	IVMM ^[37]	2. 5~10. 5min	考虑采样点间影响,提高准确率
	WI-Matching ^[51]	/	优化权重度量、插值,提高准确率
	HRIS ^[38]	3~15min	利用参考轨迹,提高准确率
	InferTra ^[56]	5~15min	网络移动性建模,提高准确率
	SF-Matching ^[57]	2~8min	利用路径选择模型,提高准确率
	STRS ^[58]	15~90s	利用概率局部路径推断,提高准确率
	TAMM ^[59]	20~260s	利用重力模型,提高准确率
	FP-Matching ^[31]	30~300s	利用频繁路径,提高准确率
	DPMM ^[74]	30~300s	引入个性化路径选择偏好,提高准确率
	Genetic ^[54]	/	考虑全局特征,提高准确率
	AntMapper ^[55]	10~300s	

4.2 数据需求角度的对比分析

表 4 从数据需求角度对典型算法进行分类.位置采样点的时间、经纬度和路网基本拓扑数据(路口位置和路段形状)是地图匹配所需基本数据.匹配时可用信息越多越有利.因此,很多算法都尽可能挖掘可用信息以

提升准确性,例如:瞬时运动速度、运动方向、路段宽度、限速、历史平均速度、自由车流速度等.但这些数据不易获取.所以,仅依赖于基本数据的算法通用性更好.而从匹配准确性角度,则应选择能充分利用可用数据的算法.

表 4 典型地图匹配算法的数据需求对比

算法	轨迹数据包含:				路网数据包含:				
	时间+ 经纬度	瞬时 速度	瞬时 方向	历史 轨迹	路网 拓扑	路宽	限速	历史 平均 速度	自由 流速
CRF-Liao ^[45] , HMM-Newson ^[27] , HMM-IRL ^[24] , Eddy ^[40] , OBR-HMM ^[72] , MSTM ^[52] , OM2 ^[39] , Feature-based ^[42] , Genetic ^[54] , FMM ^[26] , INC-RB ^[43]	√				√				
WI-Matching ^[51] , AntMapper ^[55] , EHMM ^[44] , STD-Matching ^[53]	√		√		√				
OHMM ^[28]	√				√	√			
ST-Matching ^[25] , IVMM ^[37] , EnAcq ^[5] , LB-MM ^[50] , ST-CRF ^[46] , AIVMM ^[36]	√				√		√		
CRF-Xu ^[32] , HMM+RCM ^[30] , Hybrid HMM ^[34]	√				√			√	
HMM-Jagadeesh ^[41]	√				√		√		√
AOMM ^[29]	√	√			√			√	
STD-Matching ^[53]	√		√		√			√	
HRIS ^[38] , SF-Matching ^[57] , STRS ^[58] , InferTra ^[56] , FP-based ^[31] , DPMM ^[74]	√			√	√				
Hybrid HMM ^[34]	√			√	√	√			
HMM-DPP ^[33]	√	√		√	√				
IF-Matching ^[35] , TAMM ^[59]	√	√	√	√	√				

4.3 实现技术的优缺点

基于HMM的算法相对早期基于几何、拓扑的算法而言,对噪声、采样率具有更好的鲁棒性.但当定位精度或采样率降低时,如:采样间隔超过1分钟,其准确率会急剧下降,路网密集情况下更甚.基于CRF算法的特

征关系建模较HMM更具灵活性^[48],可克服HMM的标注偏好问题,但模型参数更多,需大量带标记的数据训练,而带标记的数据很难大量获取.总之,这两类算法的准确率对定位精度有一定的鲁棒性,但对采样率比较敏感^[25,40].

基于最大权重的算法主要用于离线地图匹配. 相对于基于 HMM 的算法, 其缺点是: 未对特征归一化, 可能导致匹配结果错误. 与前两类算法类似, 基于最大权重的算法准确率也会随采样率降低而急剧下降.

基于局部路径推断的算法主要用于低采样率轨迹的离线地图匹配, 采样间隔允许 2 分钟及以上, 摒弃了最短/快路径假设. 但它们需要大量历史轨迹以提取路径选择偏好和出行规律等辅助匹配. 基于深度学习的算法在效率和准确率上都有提升, 但也需要大量带标记训练数据学习模型参数.

5 总结展望

地图匹配研究已持续近三十年, 已有许多匹配算法相继提出. 但它们仍不完善, 存在问题有: (1) 特征选择停留在人工阶段; (2) 未充分利用相似轨迹辅助匹配; (3) 未考虑路网的不完整性; (4) 缺乏统一的评估方法和基准测试数据.

后续研究中, 可考虑从以下几方面展开: (1) 应用深度学习技术; (2) 自动驾驶车辆的高精度地图匹配, 例如: 车道级地图匹配^[75]为自动驾驶提供导航; (3) 三维立体路网上的地图匹配; (4) 泛在应用场景下的地图匹配.

参考文献

- [1] GU F Q, HU X K, et al. Indoor localization improved by spatial context-a survey[J]. ACM Comput. Surv., 2019, 52 (3): 1 – 35.
- [2] WU Y, CHEN P, et al. HTrack: An efficient heading-aided map matching for indoor localization and tracking[J]. IEEE Sensors Journal, 2019, 19(8): 3100 – 3110.
- [3] ZHOU R, TANG M, et al. FreeTrack: device-free human tracking with deep neural networks and particle filtering[J]. IEEE Systems Journal, 2020, 14(2): 2990 – 3000.
- [4] CHEN P, ZHENG X P, et al. Path distance-based map matching for Wi-Fi fingerprinting positioning[J]. Future Generation Computer Systems, 2020, 107: 82 – 94.
- [5] FANG S K, ZIMMERMANN R. EnAcq: energy-efficient GPS trajectory data acquisition based on improved map matching[A]. ACM SIGSPATIAL GIS'11[C]. New York: ACM, 2011. 221 – 230.
- [6] DONG J X, HICKS C, LI D J. A heuristics based global navigation satellite system data reduction algorithm integrated with map-matching[J]. Annals of Operations Research, 2020, 290(1): 731 – 746.
- [7] GONG X R, HUANG Z, et al. High-performance spatio-temporal trajectory matching across heterogeneous data sources[J]. Future Generation Computer Systems, 2020, 105: 148 – 161.
- [8] CHAO P F, HUA W, et al. A survey and quantitative study on map inference algorithms from GPS trajectories[J/OL]. IEEE Transactions on Knowledge and Data Engineering (Early Access), doi:10.1109/TKDE.2020.2977034, 2020.
- [9] ZONG F, WU T, JIA H. Taxi drivers' cruising patterns—insights from taxi GPS traces[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(2): 571 – 582.
- [10] LI L, JIANG R, et al. Trajectory data-based traffic flow studies: A revisit[J]. Transportation Research Part C: Emerging Technologies, 2020, 114: 225 – 240.
- [11] CHEN D, DRIEMEL A, et al. Approximate map matching with respect to the Fréchet distance[A]. ALLENEX 2011 [C]. San Francisco: SIAM, 2011. 75 – 83.
- [12] VELAGA N R, QUDDUS M A, BRISTOW A L. Developing an enhanced weight-based topological map-matching algorithm for intelligent transport systems[J]. Transportation Research Part C: Emerging Technologies, 2009, 17(6): 672 – 683.
- [13] QUDDUS M A, OCHIENG W Y, NOLAND R B. Current map-matching algorithms for transport applications: State-of-the art and future research directions[J]. Transportation Research Part C: Emerging Technologies, 2007, 15 (5): 312 – 328.
- [14] WEI H, WANG Y, et al. Map matching: Comparison of approaches using sparse and noisy data[A]. ACM SIGSPATIAL GIS'13[C]. New York: ACM, 2013. 444 – 447.
- [15] HASHEMI M, KARIMI H A. A critical review of real-time map-matching algorithms: Current issues and future directions[J]. Computers, Environment and Urban Systems, 2014, 48: 153 – 165.
- [16] SINGH J, SINGH S, et al. Evaluating the performance of map matching algorithms for navigation systems: an empirical study[J]. Spatial Information Research, 2018, 27: 63 – 74.
- [17] KUBICKA M, CELA A, et al. Comparative study and application-oriented classification of vehicular map-matching methods[J]. IEEE Intelligent Transportation Systems Magazine, 2018, 10(2): 150 – 166.
- [18] 高文超, 李国良, 塔娜. 轨迹路网匹配算法综述[J]. 软件学报, 2018, 29(2): 225 – 250.
- [19] GAO W C, LI G L, TA N. Survey of map matching algorithms[J]. Journal of Software, 2018, 29(2): 225 – 250. (in Chinese)
- [19] KUMAR P, GANTI R K, RAINA G. A new technique to find candidate links for map matching for transportation

- applications[A]. The 8th International Conference on Communication Systems and Networks [C]. New Jersey, USA: IEEE, 2016. 1 – 6.
- [20] KOLLER H, WIDHALM P, et al. Fast hidden Markov model map-matching for sparse and noisy trajectories[A]. ITSC 2015[C]. New Jersey, USA: IEEE, 2015. 2557 – 2561.
- [21] LI H F, KULIK L, RAMAMOCHANARAO K. Spatio-temporal trajectory simplification for inferring travel paths [A]. ACM SIGSPATIALGIS'14[C]. New York: ACM, 2014. 63 – 72.
- [22] LIU X L, YU L, et al. ST-PF: Spatio-temporal particle filter for floating-car data pre-processing[A]. (IF&IGIS'17 [C]. Switzerland: Springer, 2018. 197 – 211.
- [23] RAPPOS E, ROBERT S, CUDRÉ -MAUROUX P. A force-directed approach for offline GPS trajectory map matching[A]. ACM SIGSPATIAL GIS'18[C]. New York: ACM, 2018. 319 – 328.
- [24] OSOGAMI T, RAYMOND R. Map matching with inverse reinforcement learning[A]. IJCAI'13[C]. Palo Alto, California: AAAI, 2013. 2547 – 2553.
- [25] LOU Y, ZHANG C Y, et al. Map-matching for low-sampling-rate GPS trajectories[A]. ACM SIGSPATIAL GIS'09 [C]. New York: ACM, 2009. 352 – 361.
- [26] YANG C, GIDÓFALVI G. Fast map matching, an algorithm integrating hidden Markov model with precomputation[J]. International Journal of Geographical Information Science, 2018, 32(3): 547 – 570.
- [27] NEWSON P, KRUMM J. Hidden Markov map matching through noise and sparseness[A]. ACM SIGSPATIAL GIS'09[C]. New York: ACM, 2009. 336 – 343.
- [28] GOH C Y, DAUWELS J, et al. Online map-matching based on Hidden Markov model for real-time traffic sensing applications[A]. ITSC'12[C]. New Jersey, USA: IEEE, 2012. 776 – 781.
- [29] AN Q, FENG Z, et al. A green self-adaptive approach for online map matching[J]. IEEE Access, 2018, 6: 51456 – 51469.
- [30] JAGADEESH G R, SRIKANTHAN T. Online map-matching of noisy and sparse location data with hidden Markov and route choice models[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(9): 2423 – 2434.
- [31] HUANG Y, RAO W, et al. Frequent pattern-based map-matching on low sampling rate trajectories[A]. IEEE MDM'18[C]. New Jersey, USA: IEEE, 2018. 266 – 273.
- [32] XU M, DU Y M, et al. Map matching based on conditional random fields and route preference mining for uncertain trajectories[J]. Mathematical Problems in Engineering, 2015, 717095: 1 – 13.
- [33] SONG C, YAN X, et al. Hidden Markov model and driver path preference for floating car trajectory map matching [J]. IET Intelligent Transport Systems, 2018, 12(10): 1433 – 1441.
- [34] OZDEMIR E, TOPCU A E, OZDEMIR M K. A hybrid HMM model for travel path inference with sparse GPS samples[J]. Transportation, 2018, 45(1): 233 – 246.
- [35] HU G, Shao J, et al. IF-Matching: towards accurate map-matching with information fusion[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(1): 114 – 127.
- [36] ZHANG Y, HE Y. An advanced interactive-voting based map matching algorithm for low-sampling-rate GPS data [A]. IEEE ICNSC 2018 [C]. New Jersey, USA: IEEE, 2018. 1 – 7.
- [37] YUAN J, ZHENG Y, et al. An interactive-voting based map matching algorithm[A]. IEEE MDM'10[C]. New Jersey, USA: IEEE, 2010. 43 – 52.
- [38] ZHENG K, ZHENG Y, et al. Reducing uncertainty of low-sampling-rate trajectories[A]. IEEE ICDE'12[C]. New Jersey, USA: IEEE, 2012. 1144 – 1155.
- [39] XIE Y, ZHOU K, et al. High-accuracy off-line map-matching of trajectory network division based on weight adaptation HMM[J]. IEEE Access, 2020, 8: 7256 – 7266.
- [40] WANG G F, ZIMMERMANN R. Eddy: An error-bounded delay-bounded real-time map matching algorithm using HMM and online viterbi decoder[A]. ACM SIGSPATIAL GIS'14[C]. New York: ACM, 2014. 33 – 42.
- [41] JAGADEESH G R, SRIKANTHAN T. Probabilistic map matching of sparse and noisy smartphone location data [A]. IEEE ITSC'15 [C]. New Jersey, USA: IEEE, 2015. 812 – 817.
- [42] YIN Y F, SHAH R R, ZIMMERMANN R. A general feature-based map matching framework with trajectory simplification[A]. ACM SIGSPATIAL IWGS'16[C]. Burlingame, California, USA: ACM, 2016. 7:1 – 7:10.
- [43] LUO L B, HOU X T, et al. Incremental route inference from low-sampling GPS data: An opportunistic approach to online map matching[J]. Information Sciences, 2020, 512: 1407 – 1423.
- [44] CHE M L, WANG Y L, et al. An enhanced hidden Markov map matching model for floating car data[J]. Sensors,

- 2018, 18(6): 1758.
- [45] LIAO L, FOX D, KAUTZ H. Extracting places and activities from GPS traces using hierarchical conditional random fields[J]. *The International Journal of Robotics Research*, 2007, 26(1): 119 – 134.
- [46] LIU X L, LIU K, et al. A ST-CRF map-matching method for low-frequency floating car data[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2017, 18(5): 1241 – 1254.
- [47] HUNTER T, ABBEEL P, BAYEN A. The path inference filter: model-based low-latency map matching of probe vehicle data[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2014, 15(2): 507 – 529.
- [48] YANG J, MENG L Q. Feature selection in conditional random fields for map matching of GPS trajectories[A]. *Progress in Location-Based Services 2014*[C]. Switzerland: Springer, 2015. 121 – 135.
- [49] CHEN B Y, YUAN H, et al. Map-matching algorithm for large-scale low-frequency floating car data[J]. *International Journal of Geographical Information Science*, 2014, 28(1): 22 – 38.
- [50] CHANDIO A A, TZIRITAS N, et al. An approach for map-matching strategy of GPS-Trajectories based on the locality of road networks[A]. *Internet of Vehicles - Safe and Intelligent Mobility*[C]. Switzerland: Springer, 2015. 234 – 246.
- [51] LIU S Y, LIU C, et al. Calibrating large scale vehicle trajectory data[A]. *IEEE MDM 2012*[C]. New Jersey, USA: IEEE, 2012. 222 – 231.
- [52] TUO Y, YUN X, ZHANG Y. MSTM: A novel map matching approach for low-sampling-rate trajectories[A]. *IEEE PIMRC 2017* [C]. Montreal, QC, Canada: IEEE, 2017. 1 – 7.
- [53] HSUEH Y L, CHEN H C. Map matching for low-sampling-rate GPS trajectories by exploring real-time moving directions[J]. *Information Sciences*, 2018, 433-434: 55 – 69.
- [54] NIKOLIĆ M, JOVIĆ J. Implementation of generic algorithm in map-matching model[J]. *Expert Systems with Applications*, 2017, 72: 283 – 292.
- [55] GONG Y J, CHEN E, et al. AntMapper: an ant colony-based map matching approach for trajectory-based applications[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2018, 19(2): 390 – 401.
- [56] BANERJEE P, RANU S, RAGHAVAN S. Inferring uncertain trajectories from partial observations[A]. *IEEE ICDM 2014*[C]. New Jersey, USA: IEEE, 2014. 30 – 39.
- [57] CHEN C, ZHANG X D, et al. Map-matching based on driver behavior model and massive trajectories[A]. *IEEE ITSC2014*[C]. New Jersey, USA: IEEE, 2014. 2817 – 2822.
- [58] WU H, MAO J Y, et al. Probabilistic robust route recovery with spatio-temporal dynamics[A]. *ACM SIGKDD'16* [C]. New York: ACM, 2016. 1915 – 1924.
- [59] CINTIA P, NANNI M. An effective Time-Aware map matching process for low sampling GPS data[DB/OL]. arXiv:1603.07376, 2016.
- [60] WU H, SUN W W, Zheng B H. Is only one GPS position sufficient to locate you to the road network accurately[A]. *ACM UbiComp'16*[C]. New York: ACM, 2016. 740 – 751.
- [61] ZHAO K, FENG J, et al. DeepMM: deep learning based map matching with data augmentation[A]. *ACM SIGSPATIAL GIS'19*[C]. Chicago: ACM, 2019. 452 – 455.
- [62] 卢家品, 罗月童, 等. 基于排名学习和多源信息的地图匹配方法[J]. *浙江大学学报(理学版)*, 2020, 47(1): 27 – 35, 66.
- LU J P, LUO Y T, et al. An information fusion map matching method based on ranking learning[J]. *Journal of Zhejiang University(Science Edition)*, 2020, 47(1): 27 – 35, 66. (in Chinese)
- [63] SHEN Z H, DU W, et al. DMM: fast map matching for cellular data[A]. *MobiCom'20*[C]. New York: ACM, 2020. 1 – 14.
- [64] SRIVATSA M, GANTI R, et al. Map matching: facts and myths[A]. *ACM SIGSPATIALGIS'13*[C]. New York: ACM, 2013. 484 – 487.
- [65] RAHMANI M, KOUTSOPOULOS H N. Path inference from sparse floating car data for urban networks[J]. *Transportation Research Part C: Emerging Technologies*, 2013, 30: 41 – 54.
- [66] LI Y, HUANG Q X, et al. Large-scale joint map matching of GPS traces[A]. *ACM SIGSPATIAL GIS'13*[C]. New York: ACM, 2013. 214 – 223.
- [67] KONG X F, YANG J W. A scenario-based map-matching algorithm for complex urban road network[J]. *Journal of Intelligent Transportation Systems*, 2019, 23(6): 617 – 631.
- [68] 孙文彬, 熊婷. 历史数据和强化学习相结合的低频轨迹数据匹配算法[J]. *测绘学报*, 2016, 45(11): 1328 – 1334.
- SUM W B, XIONG T. A low-sampling-rate trajectory matching algorithm in combination of history trajectory and reinforcement learning[J]. *Acta Geodaetica et Carto-*

- graphica Sinica, 2016, 45(11): 1328 – 1334.(in Chinese)
- [69] KRUGER R, SIMEONOV G, et al. Visual interactive map matching[J]. IEEE Transactions on Visualization and Computer Graphics, 2018, 24(6): 1881 – 1892.
- [70] LIANG B W, WANG T J, et al. Online learning for accurate real-time map matching[A]. PAKDD 2016[C]. Switzerland: Springer, 2016. 67 – 78.
- [71] TAGUCHI S, KOIDE S, YOSHIMURA T. Online map matching with route prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(1): 338 – 347.
- [72] LI H F, KULIK L, RAMAMOCHANARAO K. Robust inferences of travel paths from GPS trajectories[J]. International Journal of Geographical Information Science, 2015, 29(12): 2194 – 2222.
- [73] IMAMICHI T, OSOGAMI T, RAYMOND, R. Truncating shortest path search for efficient map matching[A]. IJCAI 2016[C]. Palo Alto, California: AAAI, 2016. 589 – 595.
- [74] 高需, 武延军, 等. 基于偏好的个性化路网匹配算法[J]. 软件学报, 2018, 29(11): 3500 – 3516.
- GAO X, WU Y J, et al. Personalized map-matching algorithm based on driving preference[J]. Journal of Software, 2018, 29(11): 3500 – 3516.(in Chinese)
- [75] MAAREF M, KASSAS Z M. Ground vehicle navigation in GNSS-Challenged environments using signals of opportunity and a closed-loop map-matching approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(7): 2723 – 2738.

作者简介



于娟 女, 1983年9月出生于湖南省张家界市. 现为浙江师范大学数学与计算机科学学院讲师、硕士生导师. 主要研究方向为时空数据挖掘、数据隐私保护.
E-mail: yujuan@zjnu.edu.cn



韩建民 男, 1969年7月出生于辽宁大连. 现为浙江师范大学数学与计算机科学学院教授, 博士生导师. 主要研究方向为数据发布隐私保护.
E-mail: hanjm@zjnu.cn



杨琼 女, 1982年5月出生于湖南省吉首市. 杭州电子科技大学计算机学院博士研究生. 主要研究方向为时空数据挖掘、隐私保护.
E-mail: yangqiong525@hdu.edu.cn



彭浩 男, 1982年12月出生于江苏省泰兴市. 现为浙江师范大学双龙学者特聘教授、硕士生导师, 从事网络安全、人工智能安全方面的研究工作.
E-mail: hpeng@zjnu.edu.cn



鲁剑锋(通信作者) 男, 1982年10月出生于湖北嘉鱼县. 现为浙江师范大学数学与计算机科学学院教授, 博士生导师. 主要研究方向为群智感知和博弈论.
E-mail: lujianfeng@zjnu.cn