



基于隐马尔可夫模型的压缩轨迹地图匹配算法

北京航空航天大学 计算机学院 北京 100191

摘 要: 本文提出了一种新型的用于压缩轨迹的地图匹配系统。系统在匹配过程中充分考虑了压缩轨迹的特点（压缩后误差有界，原始轨迹点集中于压缩后轨迹线段一侧等）。我们使用两个大规模真实轨迹数据集与两种当前最先进的基于隐马尔可夫模型的地图匹配算法进行了比较。实验结果显示，大多数情况下，本文提出的算法能够给出最高的匹配精度。平均来看，本文提出的算法准确率比 Newson 算法和基于特征的地图匹配算法分别高 14% 和 13%。在效率方面，本文提出的算法比 Newson 算法和基于特征的地图匹配算法分别快 14% 和 13%。实验结果说明了本文提出的系统用于压缩轨迹的地图匹配的有效性和效率。

关 键 词: 轨迹压缩; 地图匹配; 隐马尔可夫模型

中图分类号: TP399

文献标识码: A

文章编号: 2018060621

Map Matching for Compressed Trajectories Based on HMM

Department of Computer Science Beihang University Beijing 100191

Abstract: In this paper, we proposed a novel map matching framework for compressed trajectories. Our framework fully considers the attributes of compressed trajectories such as error bound and space split. We compared the map matching accuracy and efficiency of the proposed method with two state-of-the-art algorithms, namely the Newson's and general feature-based algorithm, using two large-scale real datasets. The experimental results show that, in most cases, the proposed method gives the best accuracy. On average, the proposed method is 13% and 13% better than algorithms Newson's and general feature-based algorithm, respectively. In terms of efficiency, our framework runs 2.1 and 2.5 times faster than algorithms Newson's and general feature-based algorithm, respectively. All these results demonstrate the effectiveness and efficiency of the proposed method in the map matching of compressed trajectories.

Key words: trajectory compression; map matching; HMM

引言¹

随着 GPS 传感器和移动设备的广泛使用，产生了大量的轨迹数据。这些轨迹数据是基于位置的服务、城市交通流量分析、轨迹模式挖掘等应用的基础，具有重要的意义。然而，采集、传输和管理大量的轨迹数据又会对终端、网络和服务系统带来巨大的压力。首先，原始轨迹数据需要从设备上传到服务器中，大量轨迹数据的上传会造成网络传输的压力。其次，大量、重复的原始轨迹数据，需要越来越大的存储空间来进行存储。最后，越来越大的存储空间又

会增加查询的计算，降低查询速度，影响系统体验。因此，学者提出了轨迹压缩算法 [1][2][3][4][5]，用于去掉原始轨迹中的冗余轨迹点，仅保留关键轨迹点，从而简化了原始轨迹，大幅减少需要传输和管理的数据量，缓解系统压力。

与此同时，地图匹配是轨迹管理与应用的一个基础算法。许多基于位置的服务，如导航和追踪系统 [6][7][8]，区域功能分析 [9]，异常事件检测 [10][11]，实时交通流量预测 [12] 等需要使用地图匹配作为重要的预处理步骤，将轨迹点序列匹配到路

网上。然而,传统的地图匹配方法将原始轨迹点序列匹配到路网上,而对于压缩轨迹的地图匹配方法还缺少研究。

压缩轨迹有其自身不同于原始轨迹的特点,因此压缩轨迹的地图匹配与原始轨迹的地图匹配有明显的区别。首先,对轨迹进行压缩后,轨迹变得更稀疏,在压缩后的轨迹上进行地图匹配固然可以减少需处理的数据量,进而减少执行时间,但是同时会对地图匹配带来困难。若按常规的地图匹配方法进行匹配,将导致匹配正确性的显著下降。其次,目前的主流轨迹压缩是误差有界的方法,能够保证任意原始轨迹点到压缩后的轨迹之间的距离在一个误差界限之内,意味着压缩轨迹具有特定的几何特征,这些几何信息可以用于辅助地图匹配过程,从而提高匹配准确率。

针对压缩轨迹与原始轨迹的不同,我们提出一种用于压缩轨迹的地图匹配系统。本文的贡献主要有以下三个方面:

1. 我们首先提出一种新型的用于压缩轨迹的地图匹配系统。由于系统需要处理的轨迹点数更少,本文提出的系统的运行效率明显好于传统的基于隐马尔科夫模型的地图匹配算法,而匹配准确率相对在原始轨迹上进行压缩也没有损失太多。
2. 我们开发出一种用于压缩轨迹的地图匹配算法,算法的准确率和执行效率都好于当前最先进的算法。
3. 我们使用两个大规模真实轨迹数据集与两种当前最先进的算法进行了实验对比,实验结果验证了本文提出的系统的有效性和效率。

1 相关工作

在过去的几十年中,地图匹配问题吸引了大量研究者的关注。接下来,我们介绍一下地图匹配问题的 *state-of-the-art* 解决方法,尤其是对于稀疏轨迹的地图匹配问题。

1.1 一般地图匹配方法

Quddus 等人把地图匹配方法分为 4 个主要大类:基于几何的方法,基于拓扑的方

法,基于统计的方法和其他高级方法[13]。

基于几何的匹配主要考虑 GPS 点和道路网络的几何信息。使用搜索算法,在点对点匹配中,计算出 GPS 点最近的节点或道路段,在点到曲线的匹配中,找出最近的由分段线段组成的曲线,在曲线到曲线的匹配中,使用已知道路与轨迹比较。由于忽视了路网拓扑结构的限制,这类方法对于测量误差十分敏感。

基于拓扑的匹配同时考虑轨迹点和路网的几何信息以及道路网络的拓扑属性(道路转向、弯曲、连接等)来进行轨迹匹配。然而,这类方法对于测量误差较大的轨迹及稀疏轨迹匹配的效果仍然较差。

为进一步提高匹配的准确率,概率匹配算法考虑轨迹位置的置信椭圆或矩形置信区域。高级匹配算法综合使用多方面信息来进行匹配,主要有卡尔曼滤波、模糊逻辑模型、隐式马尔科夫链模型等。

1.2 稀疏轨迹的地图匹配

近几年,一些研究开始关注对于稀疏轨迹的地图匹配问题。这类算法基本都是使用序列标注模型的方法,其中隐马尔科夫模型和条件随机场是使用最为广泛的两种模型。

Lamb 和 Thiebaut 通过与卡尔曼滤波器结合,首先将隐马尔科夫模型应用于地图匹配问题中[14]。在文献[15]中,Newson 和 Krumm 提出了一种基于隐马尔科夫模型的算法,算法中在转移概率的计算中考虑了路网距离并对不同精度和采样率的轨迹取得了较好的效果。Wang 等人提出了一种名为 Eddy 的基于隐马尔科夫模型的在线地图匹配算法,并对算法的误差和延迟进行了分析。随后,Osogami 和 Raymond 提出了一种使用行为图的算法[16],该方法能够在最优路径搜索过程中加入道路转弯的考虑,并使用了逆强化学习从标注数据中学习模型权重。为了进一步提高算法在稀疏轨迹上匹配的准确率,Yin 等人提出了 *general feature-based framework*[17],该方法采用了一种新的基于轨迹段的概率建模策略,在最短路径搜索时,只对轨迹化简算法

检测到的关键点进行。

除了基于隐马尔科夫模型的方法之外,一些工作利用条件随机场来解决地图匹配问题。Hunter 等人提出了一种名为路径推断过滤(PIF)的方法[18],模型中考虑了包括路段长度,信号灯的数量,最短行驶时间等在内的 10 种影响因素。然而,在一般情况下,大部分影响因素是很难得到的,这限制了模型的推广。Liu 等人提出了一种名为 ST-CRF 的地图匹配算法[19],用于稀疏轨迹的地图匹配。模型中只使用了基本的 GPS 轨迹信息,并且在特征函数中考虑了 5 种影响因素。然而,如其他基于条件随机场的方法一样,这种方法需要大量历史轨迹数据来对特征函数中的权重进行学习。

2 相关概念描述

本章给出地图匹配问题以及相关概念。

2.1 轨迹相关概念

轨迹点(P): 轨迹点 $P = (x, y, t)$ 定义为一个三元组 (x, y, t) , 表示移动对象在时刻 t 位于经度 x , 纬度为 y 的位置。

原始轨迹(T): 原始轨迹 $T = \{P_1, P_2, \dots, P_n\}$ 是 GPS 轨迹点按时间戳 t 递增排列的一条序列 $(P_i.t < P_j.t, \forall 0 \leq i < j < n)$ 。

压缩轨迹线段(L): 压缩轨迹线段定义为 $L = \overline{p_s p_e}$, 表示连接两个端点 p_s 和 p_e 的直线段。同时, 线段包含属性 $L.L_{pos}$ 和 $L.L_{neg}$ 分别记录了原始轨迹在压缩轨迹线段两侧的长度。压缩轨迹线段 L 实际上表示了原始轨迹 T 中 p_s 到 p_e 之间的所有点。

压缩轨迹线段表示(\bar{T}): 压缩轨迹的线段表示 $\bar{T} = \{L_1, L_2, \dots, L_{\bar{n}}\}$ 是一个压缩轨迹线段的序列, 其中 $L_0.p_s = p_0, L_{\bar{n}}.p_e = p_n$,

$$L_i.p_e = L_{i+1}.p_s, \forall i \in [0, \bar{n} - 1]。$$

给定一条轨迹 T 及其压缩轨迹 \bar{T} , 如果对于原始轨迹 T 中的所有点 P , 在压缩后轨迹 \bar{T} 中都存在点 P_j 及 P_{j+1} , 使得 $d(P, \overline{P_j P_{j+1}}) \leq \epsilon$, 则称压缩轨迹 \bar{T} 对于界限 ϵ 是误差有界的。本文所有的压缩轨迹都是指误差有界的压缩轨迹。

2.2 地图匹配相关概念

路段(r): 一个路段 $r = (v_s, v_e)$ 表示地图中的连接两个端点的一条边。

路径(R): 径 $R = \{r_1, r_2, \dots, r_m\}$ 是连续的路径序列, 其中各条路段满足 $r_i.v_e = r_{i+1}.v_s$ 。

路网(G(V,E)): 路网 $(G(V,E))$ 是一个有向图, 其中 V 是所有端点的集合, E 是所有路段的集合。

地图匹配: 给定一条 GPS 轨迹 (T) 及相应的路网 $(G(V,E))$, 地图匹配的目标是在路网中找到一条路径 R , 该路径为用户真实经过的路径的可能性最大。

压缩轨迹的地图匹配问题: 给定压缩轨迹 (\bar{T}) 及相应的路网 $(G(V,E))$, 压缩轨迹的地图匹配算法在路网中找到一条路径 R , 该路径为用户真实经过的路径的可能性最大。

3 系统概览

越来越多的基于统计的地图匹配算法被提出来解决地图匹配问题, 其中, 隐马尔科夫模型是使用最为广泛的模型之一。该模型通过对路段发射概率和转移概率进行建模, 最终输出联合概率最大的一条序列作为最终的匹配结果。

尽管基于隐马尔科夫模型的算法在处理稀疏轨迹时有良好的表现, 但这种方法的计算效率不高, 不适合处理大量 GPS 轨迹点。Imamichi 在文献[20]中指出, 基于隐

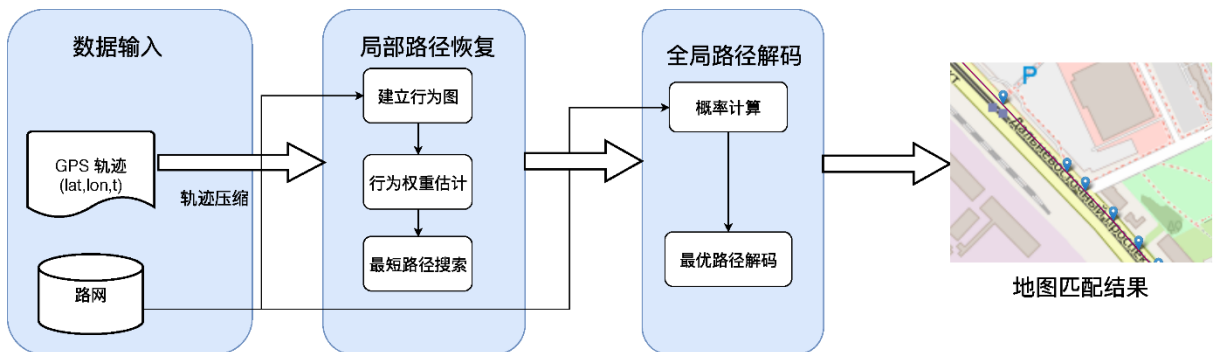


图 1 系统结构图

马尔科夫模型的地图匹配算法的计算效率瓶颈在于转移概率的计算。对每一个轨迹点的每一个候选路段，算法在路径解码过程中都需要计算到其前一个候选路段的局部最优路径。这需要大量的最短路径计算，造成算法效率低下。

由于压缩后的轨迹包含的轨迹点比原始轨迹显著减少，在压缩轨迹上进行地图匹配可以减少匹配过程所需的计算量，提高系统效率。

针对压缩轨迹存在的误差有界，轨迹点集中的特点，本文提出了一种针对压缩轨迹的地图匹配方法和系统。该系统结构如图 1 所示，系统以压缩后的轨迹作为输入，分别经过局部路径恢复模块和全局路径解码模块进行处理，最终输出路网匹配的结果。

局部路径恢复模块用来在两个相邻轨迹点之间产生局部最优路径。该模块在一张特殊的行为图上进行操作，行为图不仅使用了路网的信息，还利用了来自原始轨迹的信息。局部最优路径通过在行为图上进行最短路径搜索产生。

全局路径解码模块是隐马尔科夫模型的核心部分。基于局部路径恢复模块产生的局部最优路径，该模块计算各条候选路径的联合概率，并通过动态规划选择一条联合概率最大的路径作为全局最优路径作为整个系统的输出和地图匹配的最终结果。

4 局部路径恢复

由于进行轨迹压缩后，相邻两个轨迹点之间的距离增大，其间存在多条可能的路径。局部路径恢复模块用来在这些可能的路径中搜索一条作为最可能的路径。

4.1 候选路段搜索

给定一个待匹配的轨迹点 P ，局部路径恢复模块首先需要为该轨迹点找到它可能存在的路段集合。在本文实现中，我们把以点 P 为圆心，搜索半径 $r_s = 200$ 米范围内的路段作为点 P 的候选路段集合。我们采用基于 R-tree 的空间索引结构，加快候选路段的查询速度。

4.2 路网子图提取

由于压缩轨迹是误差有界的，如图 2 所示，这一距离阈值界定了原始轨迹可能存在的范围，我们可以根据这个范围从原始路网 $G(V, E)$ 中提取一个路网子图 $(G_s(V_s, E_s))$ ，后续的匹配步骤在这个子图上进行计算，可以缩小搜索范围，提高执行效率。具体地，子图范围为：以压缩轨迹线段为对称轴，宽度为 $w = 2 * (\epsilon + r_s)$ ，长度为 $l = \bar{T}.Len + 2 * r_s$ 的矩形范围。

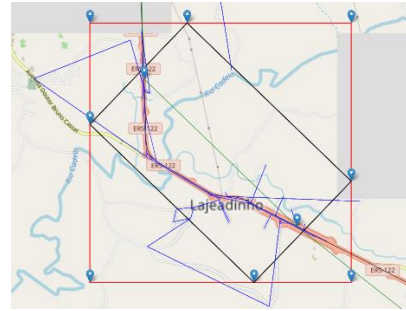


图 2 根据距离阈值提取子图

4.3 行为图

在传统的地图匹配方法中，局部路径恢复时通过直接在路网的有向图 $G(V, E)$ 中进行最短路径搜索进行的。然而，当轨迹变得稀疏时，最短路径搜索得到的结果可能不是一条合理的路径。此时，应该在搜索过程中加入其他路径合理性因素的考虑。

行为图是一张从路网中提取出来的有向图，Osogami 和 Raymond 首次在其文章 [16] 中使用。如图 3 所示，行为图中一个结点表示路网中的一条路段，一条边表示描述了在路网中从一条路段转移到另一条相邻路段的行为。图中边的权重可以用来表示做出这一转移行为的可能性。权重的计算中可以考虑路径长度，转弯角度等多种因素，从而产生尽可能合理的恢复路径。

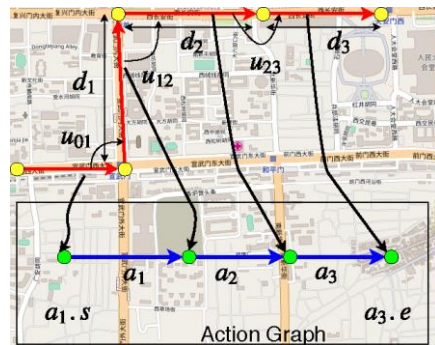


图 3 行为图

4.4 权重估计

在文献[16]中,行为的权重估计考虑了路段长度和路段之间的转弯角度两个因素。然而,针对压缩轨迹的匹配,我们还需要考虑压缩轨迹的几何特征,以提高匹配的准确率。

我们统计了真实轨迹数据中原始轨迹点在压缩轨迹线段两侧的分布,统计结果如图 5 所示。结果显示在不同的距离阈值下经过压缩后,大部分原始轨迹点集中于压缩轨迹线段的一侧,这一辅助信息可以用于提高地图匹配的准确率。



图 4 轨迹点集中于压缩轨迹一侧

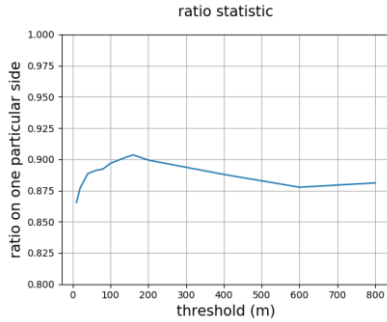


图 5 不同阈值下轨迹点集中的比例

针对压缩轨迹的这一特点,我们使用三项权值的加权和来作为行为权重的估计:

$$\omega = \omega_{len} + \alpha\omega_{turn} + \beta\omega_{ratio} \quad (1)$$

其中, ω_{len} 为该行为终点路段的长度, ω_{turn} 为该行为中从起点路段到终点路段的转弯角度带来的权值估计, ω_{ratio} 为根据路段在压缩轨迹线段两侧的长度分布与原始轨迹在压缩轨迹线段两侧的长度分布之差得到的权值。

在模型中,我们使用 Osogami 和 Raymond 提出的分段函数来估计 ω_{turn} :

$$\omega_{turn} = \begin{cases} 0, & angle < \pi/4 \\ 1, & \pi/4 \leq angle < 3\pi/4 \\ 2, & 3\pi/4 \leq angle \leq \pi \end{cases} \quad (2)$$

这一权值函数可以解释为三种不同转弯类型权值的加权和:“直行” ($|\theta_{s,e}| < \pi/4$), “正常转弯” ($|\pi/4 \leq \theta_{s,e} \leq 3\pi/4|$) 和 “急转弯” ($|3\pi/4 \leq \theta_{s,e} \leq 3\pi|$)。

权值估计中的最后一项是 ω_{ratio} , 同样使用一个分段函数来进行估计:

$$\omega_{ratio} = \begin{cases} 1, & (r_L < 0.25 \wedge r_R < 0.25) \\ 1, & (r_L > 0.75 \wedge r_R > 0.75) \\ 10, & (r_L > 0.75 \wedge r_R < 0.25) \\ 10, & (r_L < 0.25 \wedge r_R > 0.75) \\ 5, & \text{其他} \end{cases} \quad (3)$$

其中, r_L 为原始轨迹中在压缩轨迹线段正侧的长度所占比例, r_R 为该路径在压缩轨迹线段正侧的长度所占比例:

$$r_L = T.L_{pos} / (L.L_{pos} + L.L_{neg}) \quad (4)$$

$$r_R = R.L_{pos} / (R.L_{pos} + R.L_{neg}) \quad (5)$$

最终的行为权值按照公式(1)计算得到。其中各项的权重作为可调节的参数,在不同环境下可以有不同的设置。局部最优路径可以使用迪杰斯特拉算法在行为图中搜索权值最小的路径得到。

5 全局路径解码

系统使用隐马尔科夫模型来根据路径恢复模块得到的局部最优路径计算出一条整条轨迹的全局最优路径。在地图匹配问题中,将路段作为隐马尔科夫模型中的隐状态,把 GPS 点作为状态的观测。下面具体介绍模型的细节。如图 6 所示,在时刻 T_1, T_2, T_3 分别观测到 GPS 点 P_1, P_2, P_3 。对 GPS 点 P_1 有两个候选路段 r_1^1 和 r_1^2 , 作为模型在 T_1 时刻的隐状态。由路段到 GPS 观测点的可能性由观测概率描述,由前一个路段到当前路段的转移,由转移概率描述。模型根据观测概率和转移概率计算出对于当前观测序列的一条最优路径作为匹配结果。

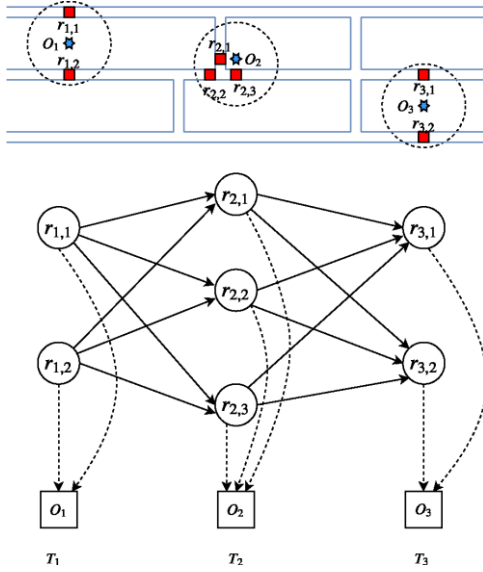


图 6 隐马尔科夫模型用于地图匹配建模

5.1 发射概率

发射概率描述了已知物体位于路段 r_i^j 的条件下, 观测到 GPS 点 P_i 的概率。对位于观测点周围的候选路段而言, 观测概率的大小取决于该路段到观测点之间的距离。距离观测点越近的候选路段应具有更大的观测发生概率。对于正确匹配的路段来说, 观测点到路段的距离是由于 GPS 测量误差造成的。Newson 和 Krumm 使用正态分布来对测量误差进行建模。具体地, 发射概率定义为:

$$P(r_i^j | P_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(d(P_i, r_i^j))^2}{2\sigma^2}} \quad (6)$$

其中, $d(P_i, r_i^j)$ 表示在地球表面上, GPS 观测点 P_i 到候选路段 r_i^j 的大圆距离, 参数 σ 表示 GPS 观测误差的标准差, 其值可以使用参数估计的方法得到。

5.2 转移概率

转移概率描述了用户从前一个候选路段移动到当前候选路段的可能性大小。转移概率的正确建模是整个模型的关键之处。在 Newson 和 Krumm 的方法中, 转移概率使用 GPS 观测点之间的地表大圆距离与它们之间的路网距离 (两点之间的路径由路径恢复模块计算得到) 之差来进行建模。在其文章[15]中, 他们从真实轨迹中统计了这一距离差, 并发现服从指数分布。因此, 他们使用这一指数分布定义转移概率。

然而, 在轨迹变得十分稀疏时, 两个 GPS 点之间可能存在多条长度相近的路径, 使用这种计算转移概率的方式会选择出一条不合理的路径。为了正确计算轨迹稀疏时的转移概率, 我们充分考虑了压缩轨迹的特点, 利用辅助信息来计算转移概率。具体地, 转移概率定义为:

$$P(r_i^k | r_{i-1}^j) = \lambda_L e^{-\lambda_L d_L} \lambda_R e^{-\lambda_R d_R} \quad (7)$$

其中, $d_L = |d(P_i, P_{i-1})_{raw_traj} - d(P_i, P_{i-1})_{route}|$ 为在原始轨迹中两个 GPS 观测点之间的距离与在路网中两点之间的行驶距离之差的绝对值。行驶距离为路径恢复模块所计算出的局部最优路径的长度。 $d_R = |r(P_i, P_{i-1})_{raw_traj} - r(P_i, P_{i-1})_{route}|$ 表示在原始轨迹中位于压缩后轨迹线段正侧的轨迹长度比例与局部恢复路径位于压缩后轨迹线段正侧的路径长度比例之差。在距离差的基础上利用这一项, 可以尽可能选择出与原始轨迹相似的路径。根据真实轨迹数据, 这两项都服从指数分布, 其中 λ_L, λ_R 为各自指数分布的系数, 可以通过参数估计得到。

5.3 最优路径解码

有了发射概率和转移概率的定义之后, 我们使用维特比算法来计算全局最优路径。维特比算法是一种动态规划算法, 它可以从所有可能的状态序列中快速识别出一条使整个序列的联合概率最大的序列。

整个系统的处理过程如图 7 所示:

Algorithm 1 The MMOC Algorithm

```

1: Input:  $\vec{T}, \epsilon, G(V, E)$ 
2: Output: map matching result  $P$ 
3:  $\vec{T} = \text{CISED-S}(\vec{T}, \epsilon)$ 
4:  $E_1 \leftarrow \text{search}(G(V, E), p_1)$ 
5: Initialize  $f[r_1^k] = p(r_1^k | p_1), k = 1, 2, \dots, 10$ 
6: for  $i = 1 \rightarrow n$  do
7:    $P_i \leftarrow \vec{T}[i].e$ 
8:    $P_{i-1} \leftarrow \vec{T}[i].s$ 
9:    $E_i \leftarrow \text{search}(G(V, E), P_i)$   $\triangleright$  find candidate set of  $P_i$ 
10:   $sbGraph \leftarrow \text{extract}(G(V, E), \vec{T}[i])$ 
11:  for  $r_i^j \in E_i$  do
12:    compute  $p(r_i^j | P_i)$  by equation 1
13:     $f[r_i^j] = -\infty$ 
14:    for  $r_{i-1}^k \in E_{i-1}$  do
15:      compute  $p(r_{i-1}^k, r_i^j)$  by equation 2
16:       $Conj \leftarrow f[r_{i-1}^k] * p(r_{i-1}^k, r_i^j) * p(r_i^j | P_i)$ 
17:      if  $Conj \geq f[r_i^k]$  then
18:         $f[r_i^k] = Conj$ 
19:         $Pre[r_i^k] = r_{i-1}^k$ 
20:  $R = \text{argmax}_{r_1^{k_1} \rightarrow r_2^{k_2} \rightarrow \dots \rightarrow r_n^{k_n}} f[r_n^{k_n}]$ 
21: return  $R$   $\triangleright$  The matched result is  $R$ 
    
```

图 7 系统算法描述

6 实验验证

本节中,我们展示本文提出的地图匹配系统与两种当前最先进的基于隐马尔科夫模型的地图匹配系统的实验比较结果。我们使用两个大规模真实轨迹数据集进行了三组实验来验证本文提出的系统的有效性和效率。

6.1 数据集与实验设置

本文使用了专车和公开数据集两个大规模真实轨迹数据集,如表 1 所示。其中,专车数据集是由一家中国的汽车租赁公司在 2015 年 8 月到 11 月之间采集的轨迹数据,采样率在 3-5 秒之间,平均每条轨迹有 114.1K 个轨迹点。我们手工标注了其中的 100 条轨迹用于实验验证。公开数据集是在世界各地采集的 100 条轨迹数据[21],Kubicka 等人对其进行了标注,并为每条轨迹附加了标注所有的地图。轨迹数据的采样率为每秒一个轨迹点。

数据集	轨迹 点数	采样率	轨迹平 均点数	总点 数
专车数据	30	3-5	4	110
公开数据	34	1	5	111

表 1 轨迹数据集

本文实现了两种当前最先进的基于隐马尔科夫模型的地图匹配系统作为对比:

1. Newson 算法: Newson 和 Krumm 于 2009 年提出的地图匹配算法。
2. 基于特征的地图匹配算法: 该算法在路径搜索时使用行为图,并使用一种基于轨迹段的策略来计算路段之间的转移概率。

本文所有算法采用 Java 实现,实验运行于一台。。。

6.2 验证指标与参数估计

本文从效率和准确率两个方面来验证所提出的地图匹配系统的有效性。

执行效率衡量了系统对轨迹进行匹配的速度。我们对每条轨迹进行加载和匹配,并且只记录地图匹配过程的运行时间。对于系统的匹配准确性,我们采用 Newson

和 Krumm 提出的路径错误匹配比例(Route Mismatch Fraction)以及 F-Score 作为评价指标。如图 8 所示,其中 L 表示真实路径的总长度, R 表示算法匹配结果的路径总长度。 L_- 为在真实路径中但不在算法匹配结果路径中的路径总长度, L_+ 表示在算法匹配结果路径中但不在真实路径中路径总长度,路径错误匹配比例定义为 $(L_- + L_+)/L$ 。

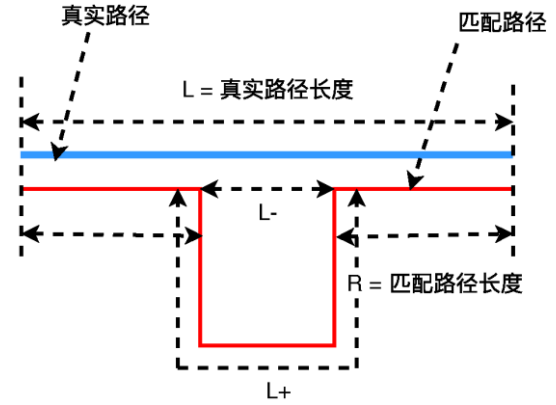


图 8 匹配准确率评价指标

我们还使用 F-score 作为匹配准确率的度量, F-score 是准确率和召回率的调和均值,通常用来评价分类算法的分类性能。在地图匹配问题中,准确率定义为算法匹配结果中,真实路径长度所占的比例,召回率定义为算法正确匹配出的路径长度在真实路径的总长度中所占的比例。各种评价指标的形式化定义如下:

$$RMF = (L_- + L_+)/L \quad (8)$$

$$\text{precision} = (R - L_+)/R \quad (9)$$

$$\text{recall} = (R - L_-)/L \quad (10)$$

$$F = 2 \times \frac{(\text{precision} \times \text{recall})}{\text{precision} + \text{recall}} \quad (11)$$

参数估计

在整个实验中,我们把 GPS 测量误差的标准差 σ 设置为 10 米。在 Newson 算法中用来计算转移概率的系数 β 设置为 2 米。在基于特征的算法中,平衡因子 ω 设置为 200。在本文提出的系统中,有两组参数需要确定,一组用来确定路径恢复模块,另一组用来计算转移概率。

我们首先通过参数估计来确定转移概率模型中的参数。如下图所示,通过对真实轨迹数据的统计发现,转移概率计算公式

(7)中,距离差 d_L 和两侧长度比例差 d_R 都

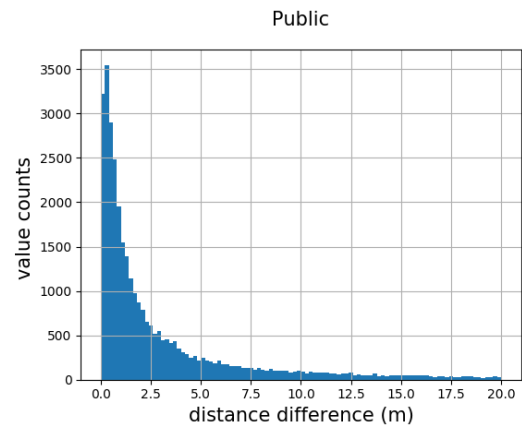
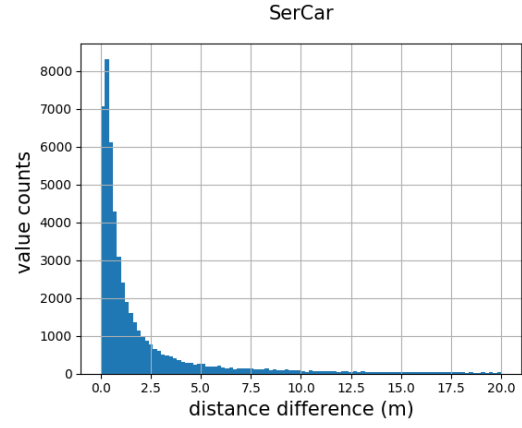
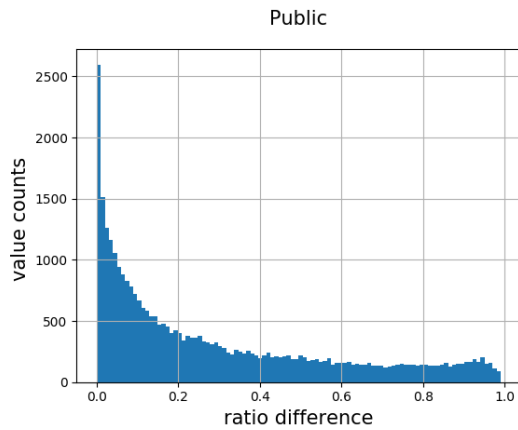
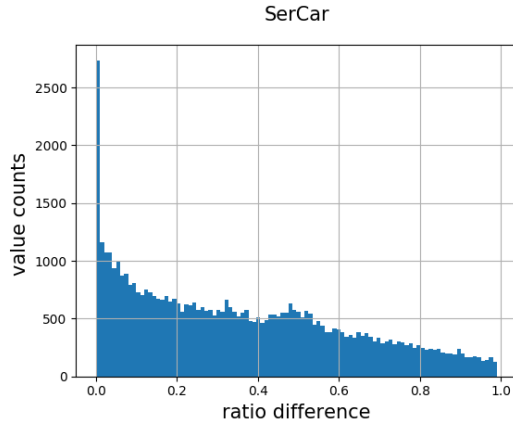
近似服从指数分布。我们使用统计值均值的倒数来作为参数的估计值。具体地，在专车数据集中， λ_L 和 λ_R 的估计值分别为 0.49 和 2.84，在公开数据集中， λ_L 和 λ_R 的估计值分别为 0.34 和 3.46。

确定了概率模型的参数之后，我们通过网格搜索来确定路径恢复模块的参数。最终确定的参数如表格 2 所示。

ϵ	10	40	200	400	800	1000
α	100	100	100	200	400	400
β	0	0.5	1.5	7.0	3.75	4.75

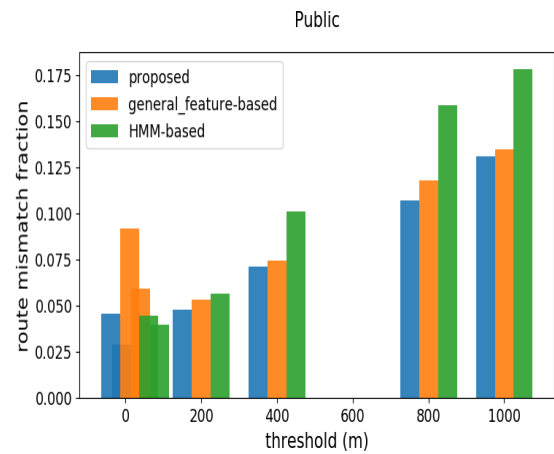
表 2 参数搜索结果

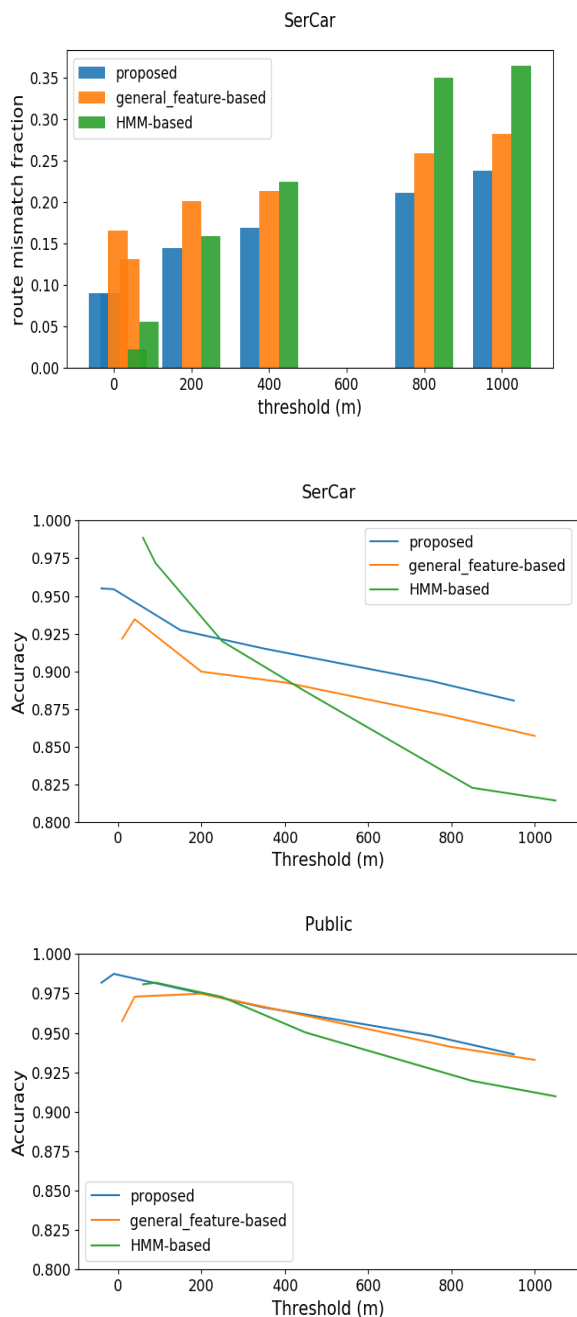
从结果中可以看出，当距离阈值较小时，相应的权重估计值也较小，当阈值增大时，权重估计值也相应增大。这是因为在阈值较小时，轨迹相对密集，此时使用最短路径作为路径恢复较为合理，随着阈值增大，轨迹变得稀疏，路径恢复时需要考虑其他因素的影响。



6.3 实验结果

我们首先进行实验与两种当前最先进的系统对比，验证本文提出的系统的匹配准确率。





从实验结果中，可以看出：

本文提出的地图匹配系统在两个数据集上都给出了最高的匹配准确率。本文提出的系统的平均路径错误匹配比例在公开数据集上只有 Newson 算法的 0.74，只有基于特征的算法的 0.81；在专车数据集中，分别只有 0.80 和 0.75。

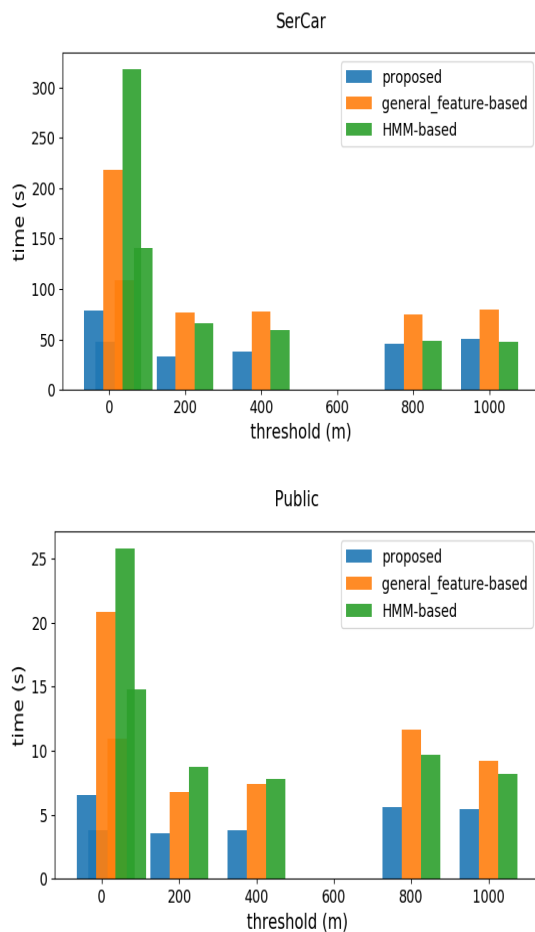
在距离阈值设置较小时，Newson 算法的匹配准确率是最高的，基于特征的算法准确率最差，本文提出的算法的匹配准确率与 Newson 算法接近。比如，当距离阈值为 10 米时，Newson 算法，本文算法，和基于特征的算法在专车数据上的路径错误匹

配率分别为 0.044，0.046 和 0.092；在公开数据集上，分别为 0.023，0.090 和 0.165。

在距离阈值设置较大时，Newson 算法的匹配准确率是最差的，基于特征的算法准确率最差，本文提出的算法的匹配准确率最好。比如，当距离阈值为 1000 米时，Newson 算法，本文算法，和基于特征的算法在专车数据上的路径错误匹配率分别为 0.364，0.237 和 0.282；在公开数据集上，分别为 0.178，0.131 和 0.135。

实验结果说明了本文提出的算法用于压缩轨迹地图匹配的有效性。

接下来对比系统的运行效率。



从图中可以看出，本文提出的系统在两个数据集中，在所有距离阈值下的运行速度都是最快的。具体地，本文提出的系统分别比 Newson 算法和基于特征的算法快 13%和 14%。

7 结论与展望

在本文中，我们提出了一种用于压缩轨迹的地图匹配系统。该系统在匹配过程中充

分考虑了压缩轨迹的特点。我们使用大规模真实轨迹数据集与两种当前最先进的地图匹配算法进行了实验对比,以验证本文提出的系统的有效性和执行效率。实验结果显示,大多数情况下,本文提出的系统都能给出最好的匹配准确率。在执行效率方面,本文提出的系统比 Newson 算法和基于特征的算法分别快 2.1 倍和 2.5 倍。所有实验结果说明了本文提出的地图匹配系统由于压缩轨迹地图匹配的有效性。

参考文献 References

- [1] DOUGLAS DH, PEUCKER TK. Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or Its Caricature. *Cartogr Int J Geogr Inf Geovisualization* 1973;10:112–22.
- [2] Meratnia N, Rolf A. Spatiotemporal compression techniques for moving point objects. *Int. Conf. Extending Database Technol., Springer*; 2004, 页 765–82.
- [3] Muckell J, Olsen PW, Hwang J-H, Lawson CT, Ravi SS. Compression of trajectory data: a comprehensive evaluation and new approach. *Geoinformatica* 2014;18:435–60. doi:10.1007/s10707-013-0184-0.
- [4] Liu J, Zhao K, Sommer P, Shang S, Kusy B, Jurdak R. Bounded quadrant system: Error-bounded trajectory compression on the go. *2015 IEEE 31st Int. Conf. Data Eng., IEEE*; 2015, 页 987–98.
- [5] Lin X, Ma S, Zhang H, Wo T, Huai J. One-pass Error Bounded Trajectory Simplification. *Proc VLDB Endow* 2017;10:841–52. doi:10.14778/3067421.3067432.
- [6] Quddus MA, Noland RB, Ochieng WY. A High Accuracy Fuzzy Logic Based Map Matching Algorithm for Road Transport. *J Intell Transp Syst* 2006;10:103–15. doi:10.1080/15472450600793560.
- [7] Yin Y, Zhang L, Zimmermann R. Exploiting Spatial Relationship Between Scenes for Hierarchical Video Geotagging. *Proc. 5th ACM Int. Conf. Multimed. Retr., New York, NY, USA: ACM*; 2015, 页 363–70. doi:10.1145/2671188.2749354.
- [8] Gustafsson F, Gunnarsson F, Bergman N, Forssell U, Jansson J, Karlsson R 等. Particle filters for positioning, navigation, and tracking. *IEEE Trans Signal Process* 2002;50:425–37. doi:10.1109/78.978396.
- [9] Yuan J, Zheng Y, Xie X. Discovering Regions of Different Functions in a City Using Human Mobility and POIs. *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., New York, NY, USA: ACM*; 2012, 页 186–94. doi:10.1145/2339530.2339561.
- [10] Zhang J. Smarter Outlier Detection and Deeper Understanding of Large-scale Taxi Trip Records: A Case Study of NYC. *Proc. ACM SIGKDD Int. Work. Urban Comput., New York, NY, USA: ACM*; 2012, 页 157–62. doi:10.1145/2346496.2346521.
- [11] Chawla S, Zheng Y, Hu J. Inferring the Root Cause in Road Traffic Anomalies. *Proc. 2012 IEEE 12th Int. Conf. Data Min., Washington, DC, USA: IEEE Computer Society*; 2012, 页 141–50. doi:10.1109/ICDM.2012.104.
- [12] Jenelius E, Koutsopoulos HN. Travel time estimation for urban road networks using low frequency probe vehicle data. *Transp Res Part B Methodol* 2013;53:64–81. doi:https://doi.org/10.1016/j.trb.2013.03.008.
- [13] Quddus MA, Ochieng WY, Noland RB. Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transp Res part c Emerg Technol* 2007;15:312–28.
- [14] Lamb P, Thiébaux S. Avoiding explicit

- map-matching in vehicle location. Proc. 6TH WORLD Congr. Intell. Transp. Syst. (ITS), HELD TORONTO, CANADA, Novemb. 8-12, 1999, 1999.
- [15] Newson P, Krumm J. Hidden Markov map matching through noise and sparseness. ACM Sigspatial Int. Conf. Adv. Geogr. Inf. Syst., 2009, 页 336–43.
- [16] Osogami T, Raymond R. Map Matching with Inverse Reinforcement Learning. Proc. Twenty-Third Int. Jt. Conf. Artif. Intell., AAAI Press; 2013, 页 2547–53.
- [17] Yin Y, Shah RR, Zimmermann R. A General Feature-based Map Matching Framework with Trajectory Simplification. Proc. 7th ACM SIGSPATIAL Int. Work. GeoStreaming, New York, NY, USA: ACM; 2016, 页 7:1--7:10. doi:10.1145/3003421.3003426.
- [18] Hunter T, Abbeel P, Bayen AM. The Path Inference Filter: Model-Based Low-Latency Map Matching of Probe Vehicle Data. 收入: Frazzoli E, Lozano-Perez T, Roy N, Rus D, 编辑. Algorithmic Found. Robot. X Proc. Tenth Work. Algorithmic Found. Robot., Berlin, Heidelberg: Springer Berlin Heidelberg; 2013, 页 591–607. doi:10.1007/978-3-642-36279-8_36.
- [19] Liu X, Liu K, Li M, Lu F. A ST-CRF Map-Matching Method for Low-Frequency Floating Car Data. IEEE Trans Intell Transp Syst 2017;18:1241–54. doi:10.1109/TITS.2016.2604484.
- [20] Imamichi T, Osogami T, Raymond R. Truncating shortest path search for efficient map-matching. IJCAI Int Jt Conf Artif Intell 2016;2016–Janua:589–95.
- [21] Kubička M, Cela A, Moulin P, Mounier H, Niculescu SI. Dataset for testing and training of map-matching algorithms. 2015 IEEE Intell. Veh. Symp., 2015, 页 1088–93. doi:10.1109/IVS.2015.7225829.