

出租车 GPS 数据的地图匹配算法研究

程歆

(深圳市华南信息技术有限公司, 广东 深圳 518000)

摘要:针对出租车 GPS 数据精确的地图匹配算法是交通大数据基本应用的第一步, 本文基于隐性马尔科夫链算法对出租车 GPS 数据进行地图匹配, 算法首先根据时间序列寻找每个 GPS 点所在的可能路段, 并对潜在可能匹配到的路段分配匹配概率, 最后根据维特比算法确定一段时间轨迹序列下最可能匹配的路段。通过验证精确度可达到 90% 以上。

关键词:出租车数据; 地图匹配; 隐性马尔科夫链; 发射概率

DOI:10.19474/j.cnki.10-1156/f.001402

交通信息采集是智能交通系统的基础, 对交通状况信息的采集技术包括感应线圈、摄像头、地磁等, 而 GPS 数据是分析城市交通运行状况的重要来源, 也逐渐成为国内外研究的热点。GPS 数据具备实时性强、数据量覆盖相对宽广等优点, 是可以大规模实时采集交通数据的可靠手段。目前, GPS 数据已广泛应用于多个与交通领域, 包括交通监管、路径推荐、行程时间估计等与预测、城市规划等。

一、基于 HMM 的地图匹配算法

(一) 隐性马尔科夫链

隐马尔可夫模型是一种统计模型, 用来描述一个含有隐含未知参数的马尔可夫过程, 模型从可观察的参数中确定该过程的隐含参数, 然后利用这些参数来作进一步的分析。通常的隐性马尔科夫链有三个状态: 隐含状态, 可观察的输出, 状态转换概率, 输出概率。

对于地图匹配来说, 其隐藏的状态即为 GPS 数据点实际所在的路段, 而观察到的状态可以简单的理解为 GPS 经纬度信息, 基于隐性马尔科夫链的地图匹配过程假设隐藏的状态是一个简单的一阶马尔科夫过程, 并且他们两两之间都可以相互转换。

(二) HMM 地图匹配算法

HMM 模型处理的是 GPS 点组成的路径, 其中包含了许多潜在可能的路段, 相邻 GPS 点潜在可能的路段之间的转换概率也是不同的, 通常相邻路段间的转换概率要高一些, 同时这些路段间转化又是不确定的。HMM 模型通常被用在语音识别, 由于其考虑了前文与当前词组的搭配, 模型的精确度是比较令人满意的。在隐性马尔科夫链的地图匹配算法中, HMM 中所说的状态即为路网上的路段, 对于路段的评测手段即为 GPS 数据与路段的距离。算法对每个 GPS 点在路网上找出相应的待选路段, 并根据其与前一个 GPS 点的备选路段为当前点估算和分配可能的概率; 再根据维特比算法反向推算出概率最高的路径, 从而确定出每个 GPS 点所匹配到的路段; 最后根据点到线段的距离计算出 GPS 点匹配到路段上的位置 (通常为 GPS 点到路段的垂足)。

1. 发射概率。发射概率给出了在给定状态下对似然概率的估计, 这个估计是基于某一特定状态下似然值, 而且是只基于这一特定状态的。对于地图匹配来说, 给定了 GPS 的记录点坐标 Z_t , 则对于候选的路段 i 来说, 其发射概率就为 $p(z_t | i)$, 由于分配的是似然概率, 通常可以认为离记录点坐标 Z_t 越近的路段, 其发射概率就越大。

另外, 为了简化 HMM-MM 的计算过程, 模型中与 GPS 点的空间距离大于 200m 的路段不会被考虑作为当前点的备选路

段; 另外, 如果 GPS 点周边点备选路段过多, 也将严重影响算法的执行效率, 考虑到这个因素, 在计算时将 GPS 点的备选路径限制在距离当前点距离最近的 10 条以内, 从而保证算法的计算效率。

2. 转换概率。对于当前的 GPS 点 Z_t 来说, 与其后续时间记录点一样, 都有一定数量的备选路段, 转移概率给出的就是相邻两个 GPS 点备选路段之间的概率。基于以上的考虑, 在计算转移概率的时候, 综合考虑了 GPS 点与备选路段之间的最短距离和匹配后点到下一个备选匹配点之间路网距离。由于出租车 GPS 点时间上点稀疏性, 相邻点间距离可能相差较大, 在寻找潜在匹配点之间点的路网距离是, 采用迪杰斯特拉算法计算路网两点的最近距离。

(三) 应用

在找出 GPS 点的备选路径以及各个备选点之间的转移概率之后, 就可以采用维特比算法对最可能对备选点的组合进行筛选。对于算法匹配的准确性, 通过对深圳市以及武汉市出租车 GPS 数据进行测试, 其中两个城市出租车的行驶路线基本覆盖了城市的各种复杂道路情况, 包括立交桥、高架桥、主辅路出入口、环岛等; 跟车测试中, 不但记录浮动车数据, 还实地考察跟车测试轨迹, 避免电子地图不准确对验证结果产生影响。验证结果表明, 此匹配模型的方案的准确率达到 90% 以上, 在路网相对稀疏的路段, 准确度可以达到 98% 以上; 同时所有算法在设计之初均考虑了并行计算以及云计算的可能性, 可直接在集群上运行程序, 提高了算法的效率, 极大的节省了匹配时间。

二、小结

本文给出了一个基于隐性马尔科夫模型的地图匹配算法, 该算法实现比较简单, 且在处理低频采样的 GPS 轨迹时仍有较好的性能。在真实数据集上对算法进行了测试, 实验结果显示所提算法与其他复杂的算法拥有相近的性能; 此外, 实验结果显示在稀疏数据的情况下, 算法依然拥有较好的性能, 同时在路网结构稀疏的情况下其匹配精度可达到 98% 以上。

参考文献:

- [1] 冯通, 杜文才. “一种基于理想隐性马尔科夫模型的地图匹配算法.” 海南大学学报自然科学版, 2014.
- [2] 唐小我, 曾勇, 曹长修. “市场预测中马尔科夫链转移概率的估计.” 电子科技大学学报, 1994.
- [3] 唐思静. “唐思静. 车辆定位导航系统中地图匹配和路径规划算法研究.” 硕士论文, 西安电子科技大学, 2009.

作者简介:

程歆, 深圳市华南信息技术有限公司。