

doi: 10.3969/j.issn.1005-3026.2017.04.004

基于隐马尔可夫模型和遗传算法的地图匹配算法

吴刚, 邱煜晶, 王国仁

(东北大学 信息科学与工程学院, 辽宁 沈阳 110819)

摘 要: 综合采用隐马尔可夫模型(HMM)和遗传算法,提出了一种新的地图匹配算法.首先初始化HMM 概率矩阵,然后使用前向后向算法进行参数学习,用 Viterbi 算法预测一组路段序列,最后将路段序列作为种群,通过遗传算法得到最优的路段序列.采用北京市 2012 年出租车 GPS 定位数据分别对传统的基于隐马尔可夫模型的算法和新算法进行测试,实验结果表明,传统的基于隐马尔可夫模型的算法的匹配精确度低于 90%,新算法的匹配精确度高达 90% 以上.

关 键 词: 地图匹配;隐马尔可夫模型;遗传算法;匹配精确度;路网数据

中图分类号: TP 399 **文献标志码:** A **文章编号:** 1005-3026(2017)04-0472-04

Map Matching Algorithm Based On Hidden Markov Model and Genetic Algorithm

WU Gang, QIU Yu-jing, WANG Guo-ren

(School of Information Science & Engineering, Northeastern University, Shenyang 110819, China. Corresponding author: QIU Yu-jing, E-mail: 724734215@qq.com)

Abstract: A new map matching algorithm was proposed on the basis of the hidden Markov model and the genetic algorithm. Firstly, the HMM probability matrix was initialized. Then, the parameters were learned by using the forward-backward algorithm, and a set of road sections was predicted by using the Viterbi algorithm. Finally, taking section sequence as population, the optimal section sequence was obtained by using the genetic algorithm. By using the taxi GPS data from Beijing in 2012 to test the traditional algorithm based on hidden Markov model and the proposed algorithm, the results showed that the traditional algorithm based on hidden Markov model has a matching accuracy below 90% and the proposed algorithm has a matching accuracy above 90%.

Key words: map matching; hidden Markov model; genetic algorithm; matching accuracy; road network data

随着车载导航系统的应用越来越广泛,人们对导航系统的定位精确度提出了更高的要求.由于存在多种不可避免的影响因素,利用全球定位系统(global positioning system, GPS)技术进行车辆定位的方法存在较大误差.因此,需要将 GPS 定位技术和电子地图信息融合来提高车辆的定位精确度.然而,目前已有的基于车辆历史轨迹信息的地图匹配算法存在着匹配精度低^[1]、匹配效率差^[2]等缺点,因此研究低复杂度、低成本、高精度、高效率的地图匹配算法具有重要的现实意义.

地图匹配技术^[3]是一种定位纠错技术,其基本思想是结合车辆的定位轨迹和电子地图中的路网信息,将 GPS 定位的车辆信息与电子地图数据按照一定的逻辑进行比较和匹配,找到车辆所在路段,根据一定的方法计算出车辆在路段上的准确位置,将车辆定位的点投影到路段上,从而矫正定位误差,提高车辆定位的精度.

本文设计了一种将隐马尔可夫模型(hidden Markov model, HMM)和遗传算法(genetic algorithm)相结合的新型地图匹配算法,即 HMM-

genetic (HG) 地图匹配算法. 该算法包括 4 个主要过程: 初始化 HMM 概率矩阵 (状态转移矩阵、观察概率矩阵、初始状态矩阵); 参数学习; 寻找一组匹配度较高的路段序列; 最后把这组路段序列作为种群, 通过遗传算法寻找最优的路段序列. 实验结果表明, 新算法在基于车辆历史轨迹的地图匹配问题上的匹配精确度高达 90% 以上.

1 HG 地图匹配算法

本文首先参考了 Newson 等^[4]提出的地图匹配算法, 使用传统的 HMM 模型编写地图匹配算法, 得出的实验结果显示匹配精确度不高, 预测出来的最优路径序列存在较大的偏差, 而遗传算法可以对已有的结果进行优化, 所以本文使用 HMM 算法得到一组概率较大的路径序列, 然后使用遗传算法对这组路径序列进行优化, 得到最优路径序列.

1.1 初始化 HMM 概率矩阵

观察概率矩阵的初始化^[5]: 观察概率矩阵是观察值在某个状态下的近似值. 对于地图匹配算法来说, 给定一个定位点 z_t , $t = 1, 2, 3 \dots$, 该定位点 z_t 对于每个路段 r_i , $i = 1, 2, 3 \dots$, 会有一个似然概率 $p(z_t | r_i)$. 这个概率表示假设车辆在该路段 r_i 上能被观察到的概率. 对于一个给定的定位点 z_t 和路段 r_i , 将路段上距离给定点最近的点记为 $x_{t,i}$, 定位点和候选匹配点在地球表面上的距离表示为 $\|z_t - x_{t,i}\|_{\text{greatcircle}}$. 可以将 GPS 噪音建模成一个零均值的高斯过程, 那么, 观察概率矩阵中元素初值的计算式为

$$p(z_t | r_i) = \frac{1}{\sqrt{2\pi}g} e^{-0.5(\|z_t - x_{t,i}\|_{\text{greatcircle}})^2 / g^2}. \quad (1)$$

式中, g 是 GPS 测量数据的标准差. 使用式 (2) 来评估该常量:

$$g = 1.4826 \text{median}_i(\|z_t - x_{t,i^*}\|_{\text{greatcircle}}). \quad (2)$$

初始状态概率矩阵的初始化^[6]: 初始状态概率矩阵中的元素表示为 π_i , $i = 1, 2, \dots, N_r$, 该矩阵在算法中表示在行驶开始时车辆在所有的路段中选择某条路段作为初始路段的概率. 本算法中的 $\pi_i = p(z_1 | r_i)$.

状态转移概率矩阵的初始化^[7]: 每个定位点有一系列可能的匹配路段. 状态转移概率矩阵给出了车辆在候选匹配路段之间转移的概率. 对于一个定位点 z_t 和候选路段 r_i , 在该路段上离定位点最近的点是 $x_{t,i}$. 对于下一个定位点 z_{t+1} 和候选路段 r_j , 对应的点是 $x_{t+1,j}$. 行驶距离记为 $\|x_{t,i} - x_{t+1,j}\|_{\text{route}}$, 两个测量点间地球表面上的距离记为

$\|z_t - z_{t+1}\|_{\text{greatcircle}}$, 见图 1. 正确匹配的地球表面距离和行驶距离的绝对值的差值满足

$$p(d_t) = \frac{1}{\beta} e^{-d_t/\beta}. \quad (3)$$

$$\text{式中: } d_t = \left| \|z_t - z_{t+1}\|_{\text{greatcircle}} - \|x_{t,i^*} - x_{t+1,j^*}\|_{\text{route}} \right|, \quad (4)$$

i^* 和 j^* 表示地面真实路段; β 描述行驶距离与地球表面距离间的差值. β 的评估方法为^[8]

$$\beta = \frac{1}{\ln 2} \text{median}_t \left(\left| \|z_t - z_{t+1}\|_{\text{greatcircle}} - \|x_{t,i^*} - x_{t+1,j^*}\|_{\text{route}} \right| \right). \quad (5)$$

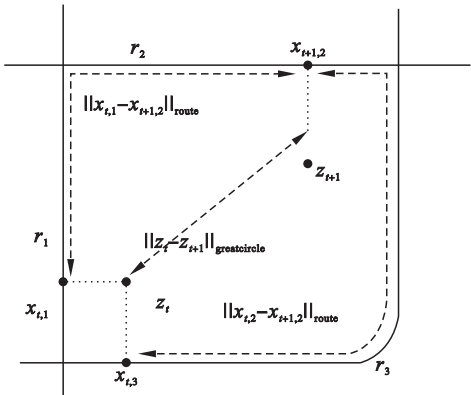


图 1 行驶距离和地球表面距离示意图
Fig. 1 Schematic diagram of route distance and great circle distance on surface of the earth

1.2 参数学习

在参数学习方面本算法采用的是标准的前向后向算法^[9], 即通过递归来更新 HMM 中的参数系统使其收敛, 最后得到的参数系统是和当前样本最匹配的.

1.3 预测路段序列

用 Viterbi 算法^[10]对输入的测试数据预测出一组路段序列. Viterbi 算法的原理是: 如果最优路径在时刻 t 通过结点 v_t , 那么这一路径从节点 v_t 到终点 v_T 的部分路径对于从 v_t 到 v_T 的所有可能的部分路径来说, 必须是最优的.

考虑到以上情况, 本文在此不直接得到最优路径, 而是利用 Viterbi 算法找出一组候选路径序列, 然后使用遗传算法继续寻找最优路径.

1.4 利用遗传算法优化路径

由于隐马尔可夫模型的分值计算方法的差异会影响序列预测结果, 所以采用遗传算法^[11]来解决这一问题. 本文采用遗传算法对预测出的路径组进行处理^[12], 从而得到与真实路径更匹配的路径.

本算法中的基因编码是路段的标号, 用适度函数评估基因. 车辆行驶往往会寻找一条最短的

本不在道路上的点,都能匹配到道路上.

2.3 实验结果对比

本实验利用多组 GPS 定位点数据对只利用基于隐马尔可夫模型的地图匹配算法和结合后的新算法分别进行了测试,根据定位点匹配的路段与真实路段匹配比例分别计算了它们的匹配精确度,并进行了对比.表 1 是利用真实数据匹配出的部分实验结果,表 2 是改进前的算法和改进后的算法匹配精度对比,其中,020564.xls,083406.xls,211254.xls,431050.xls 是包含 GPS 定位数据的四个 Excel 表,该数据主要包括采集时间、经度、纬度三个特征.

从上述实验结果可以发现,先使用基于隐马尔可夫模型的地图匹配算法对车辆历史轨迹进行匹配处理,再使用遗传算法进行改进,匹配精确度有所提高.

表 1 改进前和改进后算法对同一组 GPS 定位点匹配后的经纬度对比
Table 1 Comparison of GPS data of same set before and after improved algorithm

定位点	算法改进前	算法改进后
1	116.711 792,39.923 696 0	116.711 792,39.923 696 0
2	116.701 542,39.927 699 0	116.701 542,39.927 699 0
3	116.708 542,39.926 270 0	116.707 314,39.926 359 0
4	116.705 292,39.926 743 0	116.702 224,39.926 877 0
5	116.698 334,39.926 944 0	116.702 224,39.926 877 0
6	116.683 319,39.928 029 0	116.702 858,39.926 844 0
7	116.682 411,39.929 894 0	116.676 201,39.927 595 0
8	116.708 660,39.926 364 0	116.712 910,39.925 222 0

表 2 改进前和改进后的算法对四组 GPS 定位点的匹配精确度的对比
Table 2 Comparison of GPS data of four groups before and after improving algorithm

GPS 定位点数据	改进前精确度	改进后精确度
020564.xls	84.3%	91.2%
083406.xls	86.2%	92.3%
211254.xls	86.6%	90.0%
431050.xls	84.4%	88.1%

3 结 语

本文使用遗传算法对基于隐马尔科夫模型的地图匹配算法进行了改进.将隐马尔科夫模型预

测出来的一组路径序列,利用遗传算法进一步优化,提高了地图匹配的匹配精确度.实验结果显示,本文提出的新算法在提高匹配精确度方面具有较好的效果.

参考文献:

[1] Bernstein D,Kornhauser A. An introduction to map matching for personal navigation assistants [J]. *Geometric Distributions*,1998,122(7):1082-1083.

[2] White C E, Bernstein D, Kornhauser A L. Some map matching algorithms for personal navigation assistants[J]. *Transportation Research Part C: Emerging Technologies*, 2000,8(1):91-108.

[3] Srivatsa M,Ganti R,Wang J,et al. Map matching: facts and myths [C]//ACM Sigspatial International Conference on Advances in Geographic Information Systems. Orlando, 2013:484-487.

[4] Newson P,Krumm J. Hidden Markov map matching through noise and sparseness [C]//International Conference on Advances in Geographic Information Systems. New York: ACM,2009:336-343.

[5] Krumm J. A Markov model for driver turn prediction[C]//SAE 2008 World Congress. Detroit,2008:1-25.

[6] Szwed P,Pekala K. An incremental map matching algorithm based on hidden Markov model [C]//International Conference on Artificial Intelligence and Soft Computing. Zakopane,Poland,2014:579-590.

[7] Chen B Y,Yuan H,Li Q,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.

[8] Gather U, Schultze V. Robust estimation of scale of an exponential distribution[J]. *Statistica Neerlandica*, 1999,53(3):327-341.

[9] Welch H,Lloyd R. Hidden Markov models and the Baum-Welch algorithm [J]. *IEEE Information Theory Society Newsletter*,2003,53(2):194-211.

[10] Forney G D. The Viterbi algorithm[J]. *Proceedings of the IEEE*,2015,61(5):268-278.

[11] Reeves C R, Rowe J E. Genetic algorithms principles and perspectives [J]. *Operations Research/Computer Science Interfaces*,2014,20(2):354-355.

[12] Brakatsoulas S,Pfoser D, Salas R, et al. On map-matching vehicle tracking data[C]//International Conference on Very Large Data Bases. Trondheim,2005:853-864.

[13] Kubicka M,Cela A,Moulin P,et al. Dataset for testing and training of map-matching algorithms [C]//Intelligent Vehicles Symposium (IV). Barcelona,2015:1088-1093.