

Research on Abnormal Trajectory Behavior of Vehicles

Song Yu¹, Rui Wang¹, Lei Liu², Zhengying Cai¹

¹School of Computer and Information, China Three Gorges University, Yichang Hubei

²School of Science, China Three Gorges University, Yichang Hubei

Email: 1569229957@qq.com

Received: Jul. 3rd, 2019; accepted: Jul. 16th, 2019; published: Jul. 23rd, 2019

Abstract

In this paper, an outlier detection model is established. After removing data redundancy, the outlier data is cleaned up, that is partial selective removal of trajectory data drift. Then a map matching algorithm based on Hidden Markov Model and real-time electronic map are used to match a segment of the trajectory offset to the nearest road network, which completes the matching of the transportation road map of the transport vehicle and corrects the abnormal section of the vehicle trajectory better.

Keywords

Data Preprocessing, Hidden Markov Model, Map Matching Algorithms

车辆轨迹异常行为研究

余松¹, 王蕊¹, 刘磊², 蔡政英¹

¹三峡大学计算机与信息学院, 湖北 宜昌

²三峡大学理学院, 湖北 宜昌

Email: 1569229957@qq.com

收稿日期: 2019年7月3日; 录用日期: 2019年7月16日; 发布日期: 2019年7月23日

摘要

本文建立异常点检测模型, 去除数据冗余后对异常数据进行数据清洗, 即对轨迹数据漂移的部分选择性剔除。然后运用基于隐马尔科夫模型的地图匹配算法, 结合实时电子地图, 将轨迹偏移的一段匹配到与之最近的路网上, 完成对运输车运输路线图的匹配, 较好的纠正了车辆轨迹中的异常路段。

文章引用: 余松, 王蕊, 刘磊, 蔡政英. 车辆轨迹异常行为研究[J]. 数据挖掘, 2019, 9(3): 109-115.

DOI: 10.12677/hjdm.2019.93014

关键词

数据预处理, 隐马尔科夫模型, 地图匹配算法

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 研究背景

当今社会物联网科技发展速度迅猛, 伴随着全球汽车行业的飞速发展, 车联网技术渐渐进入我们的视野, 越来越多的专家、企业对车联网更加关注。同时道路交通行业作为我国现代运输方式的重要组成部分, 其运输方式的安全性能、运输效率以及运输能源的使用都将直接关系到国家的发展[1]。车联网系统往往通过在车辆仪表台安装车载终端设备, 实现对车辆所有工作情况和静、动态信息的采集、存储并发送。车联网系统一般具有实时实景功能, 利用移动网络实现人车交互。车联网获得的数据可以用于对车辆行驶轨迹的分析, 通过其了解道路交通状况, 合理规划外出行程, 缓解交通压力等。但是由于地理环境影响以及仪器故障等原因使得车联网采集的数据往往会出现异常, 所以我们需要对车辆行驶轨迹数据进行分析与挖掘, 对异常数据进行智能校正[2]。

2. 问题分析

针对各种可能的因素会造成车联网采集的数据缺失或者数据异常, 所以首先对轨迹数据做预处理。车联网在工作过程中, 由于传感器所采集的数据量是巨大的, 常常会造成大量数据冗余, 通过预处理去掉轨迹中经纬度坐标值一直无变化的点, 然后对轨迹数据中异常偏移的点进行筛选剔除[3]。

预处理去除错误以及冗余数据之后, 由于车辆在行驶过程中经纬度坐标变化比较连续, 基于运输车辆路线是走最短路径, 所以根据电子实时地图按照最大概率进行路网匹配, 通过编写程序代码, 可得到补全之后车辆完整的行驶路线图。

3. 建立模型

3.1. 轨迹数据预处理

运输车辆的运动轨迹在地理空间通常可以由一系列按照时间序列分布的坐标点来表示, 它们的轨迹可以由 $P_1 \rightarrow P_2 \rightarrow P_3 \rightarrow \dots \rightarrow P_n$ 表示, 其中每个点由对应的经纬度坐标和一个时间戳组成, 由 $P = (x, y, t)$ 表示。由于各种可能的因素会造成被记录的点数据缺失或者数据异常, 首先应对轨迹数据做预处理[4]。

3.2. 去除数据冗余

传感器采集的数据量是巨大的, 但是往往也会造成大量数据冗余, 例如在分析车辆轨迹时, 可以去除掉经纬度坐标值一直无变化的点, 在这一段时间内说明运输车辆是未运动的。以此筛选后的数据在地图上进行标注时会极大的提高效率。

3.3. 样条插值

计算运输车辆需要经纬度坐标变化比较连续, 减少毛刺现象的产生, 对存在缺失的异常值采用样条

插值的方式。数学上将具有一定光滑性的分段多项式称为样条函数，具体的说给定区间 $[a, b]$ 的一个分划：

$$\Delta: a = x_0 < x_1 < \dots < x_n = b$$

如果函数 $S(x)$ 满足在每个小区间 $[x_i, x_{i+1}] (i = 0, 1, \dots, n-1)$ 上 $S(x)$ 是 m 次多项式且有 $S(x)$ 在 $[a, b]$ 上具有 $m-1$ 阶连续导数，则称 $S(x)$ 为关于划分 Δ 的 m 次样条函数，其图形为 m 次样条曲线。

3.4. 噪点过滤

在空间路径不是特别精确的情况下，噪点过滤算法的作用是删除或者修正轨迹中位置存在有错误的点，如图 1 所示，存在偏差过大的点，需要在开始数据挖掘任务前先将其过滤，主要从三种过滤算法入手：均值过滤，卡尔曼过滤和基于启发式的异常检测模型。

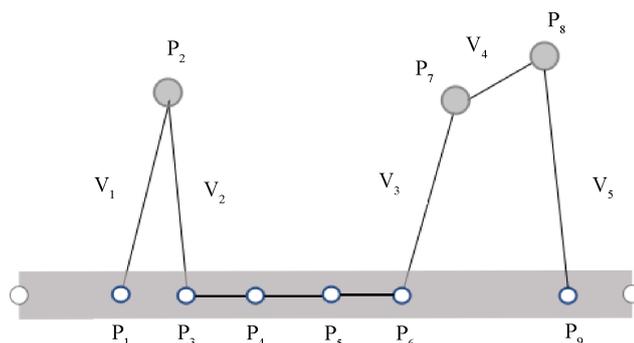


Figure 1. Schematic diagram of path noise
图 1. 路径噪声示意图

1) 均值过滤。如果某一个异常点对于它前后两个观测点位置偏移量极大时，可以采用均值过滤的方法，使车辆在单位时间内的位移距离更加合理。第 K 个点的位置估值可以表示为：

$$P_k = \frac{P_{k-1} + P_{k+1}}{2}$$

采用均值过滤的方法可以去除明显不合理的噪点。不过当出现连续的噪点时，会导致估值与实际值的差距增大，会使均值滤波效果变差。

2) 卡尔曼滤波过滤。卡尔曼过滤根据测量值和运动趋势得到轨迹估值，卡尔曼过滤基于轨迹符合线性模型的假设，根据运动物体转移概率矩阵的关系，对异常值的采集点预测并行校正[5]。

卡尔曼滤波通过系统输入输出的观测数据对系统状态进行最优的估计，能够实现实时预测的卡尔曼滤波动态轨迹预测的状态方程为：

$$X(k+1) = A(k)X(k) + T(k)W(k)$$

其中 $X(k)$ 表示为系统的状态向量， $A(k)$ 表示为状态转移矩阵，可以用来描述前一时刻到当前时刻下的运动状态转移方式，其中 $T(k)$ 为干扰转移矩阵， $W(k)$ 表示运动模型的系统状态噪声。

观测方程为：

$$Z(k) = H(k)X(k) + V(k)$$

其中 $Z(k)$ 为观测向量，描述了在某个时刻的观测值， $H(k)$ 在本文中为一维的观测矩阵， $V(k)$ 为运动估计过程中产生的观测噪声。

卡尔曼滤波算法的核心是运用递归算法来达到最优状态估计的估计模型。主要是利用前一时刻的估

计值和现时刻的观测值来更新当前状态的变量的估计，基于前 k 个观测值得到在 k 时刻下的最优状态估计，计算最小方差计算的公式表示为：

$$M = E \left[(x_i - x'_i)(x_i - x'_i)^T \right]$$

对时间和状态分别进行更新。时间的更新过程是根据前一时刻最优状态估计预测出当前时刻的状态，同时也要更新预测状态的协方差，时间更新方程为：

$$\begin{aligned} X(k+1, k) &= A(k)X(k, k) \\ Z(k+1, k) &= H(k)X(k+1, k) \end{aligned}$$

当预测出轨迹点后，根据观测值和预测值通过观测更新方程来推测出最优估计点。

3) 基于启发式的异常检测模型。

第一条：该类模型基于 GPS 坐标点与其后续点之间的时间间隔和距离间隔来计算出运输车在每个点的行驶速度。

如果一个路段相邻两个采集的经纬度坐标值的距离与时间间隔的比值，表示为：

$$V_k = \frac{S}{t}$$

当 V_k 明显大于某个阈值即高于 200 km/h 时，认为为异常点。

第二条：如果相邻两个 GPS 数据小于 10 米的点被认为是异常点。

第三条：GPS 数据的速度低于 5 千米每小时和相邻两条 GPS 数据的时间大于 1 分钟为异常点。

第四条：如果一条轨迹最多有 4 个 GPS 数据则被认为是异常数据并删除。

3.5. 基于隐马尔科夫的地图匹配算法

地图匹配就是通过车辆的行驶轨迹和电子地图数据库中的道路网进行比较，并结合车辆原来的运行趋势，在地图上找出相近的路线，然后将实际定位数据映射到直观的显示地图上，如图 2 所示。

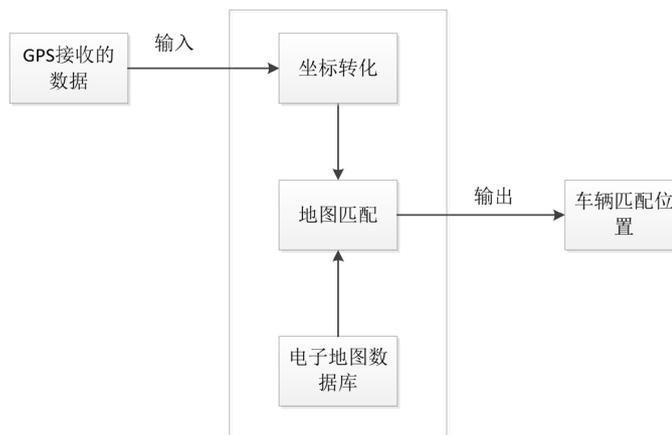


Figure 2. Map matching principle
图 2. 地图匹配原理

Step 1: 确定参数:

HMM 的 5 个基本因素，由状态变量数，观测变量数，状态转移矩阵，观测概率矩阵，初始状态概率向量组成[6]。

- 1) 确定观察变量。从车联网采集的数据中得到的位置信息(包括经度, 纬度);
- 2) 隐常状态。此时在路网上的运输车的实际所在位置;
- 3) 观测概率。观测经纬度离旁边道路近, 那么真实点在这个路段的概率越大[7], 假设服从高斯分布设有:

$$p(z_t | r_i) = \frac{1}{\sqrt{2\pi\sigma_z}} e^{-0.5 \left(\frac{\|z_t - x_{t,i}\|}{\sigma_z} \right)^2}$$

$$\sigma_z = 1.4826 \text{median}_t (\|z_t - x_{t,i}\|)$$

- 4) 状态转移概率。在两个时间点候选路段的转换概率:

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

$$d_t = \left\| \|z_t - z_{t+1}\| - \|x_{t,i} - x_{t+1,i}\| \right\|, \quad \beta = \frac{1}{\ln(2)} \text{median}_t (d_t)$$

- 5) 状态转移矩阵 A :

$$a_{ij} = P(q_{t+1} = r_j | q_t = r_i) = e^{-r_{ij}}$$

r_j 是时刻 $t+1$ 时所在路段, r_i 是时刻 t 所在路段, r_{ij} 有三种状态, 当 r_{ij} 为 1 时, 相邻, r_{ij} 为无穷大时, 不相邻, r_{ij} 等于 0 时, 在同一路段。

Step 2: Viterbi 算法求解

原理基于动态规划利用概率最大找最优路径, 其递推公式为:

$$\delta_t(i) = \max P(i_t = i, i_{t-1}, \dots, i_1, o_t, \dots, o_1 | \lambda)$$

$$\delta_{t+1}(i) = \max_{i_1, \dots, i_t} P(i_t = i, i_{t-1}, \dots, i_1, o_t, \dots, o_1 | \lambda)$$

其中 $i = 1, 2, \dots, N$ 。那么定义在时刻 t 状态为 i 的所有单个路径中概率最大的路径的第 $t-1$ 个结点为:

$$\psi_t(i) = \arg \max [\delta_{t-1}(j) a_{ij}]$$

4. 实验求解

图 3 为车联网系统采集到某辆运输车行驶轨迹的经纬度坐标所做出的行驶路线图, 从图 3 中可以明显的看出行驶路线中存在大量的异常值, 由于一辆车的速度等方面是有限的, 所以短时间内不会在经纬度路线图上有很大的变化(尖端毛刺), 所做出的路线图应该是平滑的才符合实际情况。

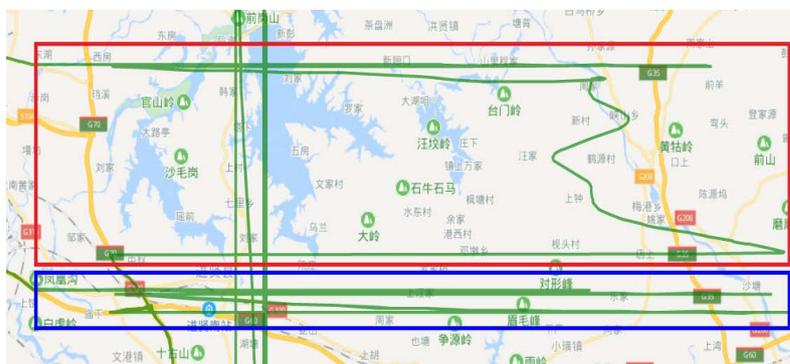


Figure 3. Actual data roadmap
图 3. 实际数据路线图

由于各种可能的因素会造成被记录的点数据缺失或者数据异常，所以首先对轨迹数据做预处理后。根据上述的异常点检测模型的检测方法，对轨迹数据进行噪点过滤，剔除异常点后路线图如图 4 所示。

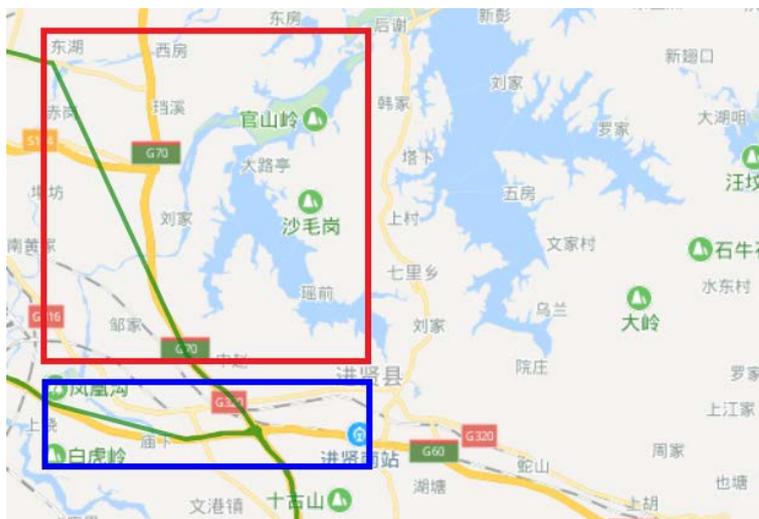


Figure 4. Removing the anomaly data graph
图 4. 去除异常数据图

预处理去除异常以及冗余数据之后，由于车辆在行驶过程中经纬度坐标变化比较连续，并且考虑到运输车行驶方向的趋势，再结合实时电子地图，采用基于隐马尔科夫模型的地图匹配算法可得到补全之后的路线图如图 5 所示。

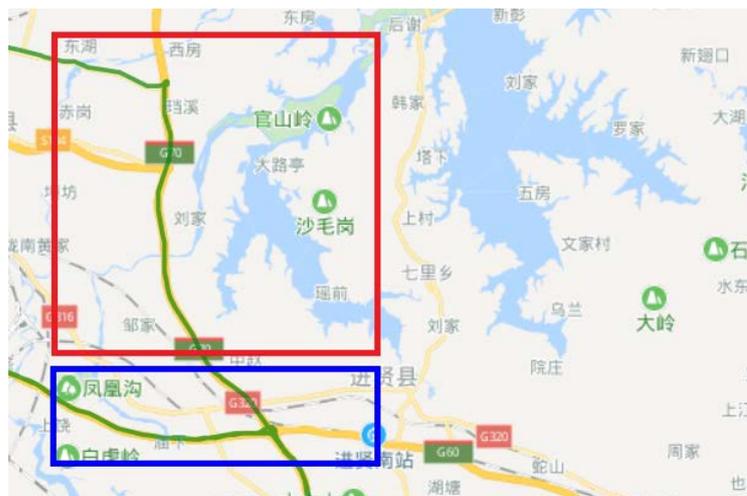


Figure 5. Map matching algorithm after processing
图 5. 地图匹配算法处理后

从图形中可以看出，地图匹配算法的补全效果的非常好，减少了许多毛刺，能较好的将缺失路段补全，非常贴合实际情况路线。

5. 总结

本文通过实验验证了异常点检测模型的检测效果比较好，能迅速找到轨迹数据存在漂移的地方，对

异常点合理的进行剔除，并且基于隐马尔科夫模型的地图匹配算法根据车辆行驶方向的趋势并且同时结合电子地图信息，能大概率的将偏离数据匹配到路网上，得到的轨迹路线图效果较好。但是本处理方法也存在一些问题，利用地图匹配算法需要联网更新实时电子地图，效率偏低，对轨迹的匹配仍然存在一定的误差。

参考文献

- [1] 张永鹏, 甘庆华. 物联网下的智慧交通大数据挖掘系统探析[J]. 数字技术与应用, 2018, 36(4): 191-192.
- [2] 范恒飞. 基于车联网的车辆运动轨迹追踪技术研究[学位论文]. 北京: 北京邮电大学, 2017.
- [3] 孙川, 吴超仲, 褚端峰, 杜志刚, 田飞. 基于车联网数据挖掘的营运车辆驾驶速度行为聚类研究[J]. 交通运输系统工程与信息, 2015, 15(6): 82-87.
- [4] 燕荣杰, 王国庆, 戴汝泉, 钟勤俭, 张竹林. 车联网数据预处理[J]. 物联网技术, 2017, 7(1): 81-82+85.
- [5] 张迎亚. 基于隐马尔科夫模型的车辆轨迹预测算法的研究[学位论文]. 南京: 南京邮电大学, 2017.
- [6] 乔少杰, 韩楠, 朱新文, 舒红平, 郑皎凌, 元昌安. 基于卡尔曼滤波的动态轨迹预测算法[J]. 电子学报, 2018, 46(2): 418-423.
- [7] 王潇. 基于车联网数据挖掘的货运车辆驾驶行为聚类研究[J]. 无线互联科技, 2017(11): 113-114+130.

知网检索的两种方式:

1. 打开知网首页: <http://cnki.net/>, 点击页面中“外文资源总库 CNKI SCHOLAR”, 跳转至: <http://scholar.cnki.net/new>, 搜索框内直接输入文章标题, 即可查询;
或点击“高级检索”, 下拉列表框选择: [ISSN], 输入期刊 ISSN: 2163-145X, 即可查询。
2. 通过知网首页 <http://cnki.net/>顶部“旧版入口”进入知网旧版: <http://www.cnki.net/old/>, 左侧选择“国际文献总库”进入, 搜索框直接输入文章标题, 即可查询。

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: hjdm@hanspub.org