

基于 HMM 模型的老年人出行异常检测

刘翔, 邹婷婷, 李仁旺, 王亚沛

(浙江理工大学机械与自动控制学院, 杭州 310018)

摘要: 针对老年人出行异常的检测问题, 提出一种基于 HMM 模型的出行行为检测方法。将出行轨迹的坐标点作为模型的训练数据, 改进了 HMM 模型中隐藏状态的确定方法, 建立老年人日常行为模型, 然后提出一种判别分析方法用于找出检测轨迹对应的状态序列, 计算该轨迹序列的输出概率并与阈值相比较, 从而判断是否出现异常行为。通过实验证明了该方法对老年人行为异常检测有较高的准确性。

关键词: 老年人; 行为检测; 聚类分析; HMM; DBSCAN

中图分类号: TP391

文献标志码: A

0 引言

我国已经进入老年型国家, 而人口老龄化带来诸多社会问题。由于老人的身体机能衰退, 很容易在日常出行中出现走失情况。随着科技的发展, 佩戴式智能设备已广泛地应用到人们的日常生活中。利用 GPS 设备采集监测对象的运动轨迹, 可以抽象出人的出行活动状态信息^[1]。

隐马尔可夫模型(hidden Markov model, HMM)已被广泛应用于语音识别领域并取得重大成功。HMM 模型改进算法在行为分析领域也有了较好的应用。余涛等^[2]将 Bayes 规则与 HMM 相结合的步态识别, 对检测出的目标特征序列采用直线拟合提取对称轴, 并等效转化为具有方向性的点线距序列, 来简化运算, 降低失真度。王相等^[3]采用基于车辆运行轨迹点序列方向角的车辆轨迹特征值表示方法和基于方向角区间划分的 HMM 观察值序列生成方法, 以方向角的区间变化来区分不同轨迹模式的特征。朱旭东等^[4]将马尔可夫模型与 4 层贝叶斯模型结合得到主题隐马尔可夫模型, 用于把运动词汇聚类成简单动作, 再聚类成全局行为, 从而消除了实时行为识别过程中的行为类型歧义。

在利用 HMM 模型建立老年人的行为习惯模型的

过程中, 需要对可观测的运动轨迹进行预处理, 来确定隐藏的状态序列。孔令富等^[5]利用 FCM (fuzzy C-means) 聚类算法确定 HMM 模型中的隐藏状态序列, 并改进 HMM 模型的重估公式, 再加入反馈滑动窗口提高轨迹检测的准确率。丁勇等^[6]利用 NONEPI++ 算法扫描给定的事件序列, 从中发现所有的频繁情节并记录每个事件发生的时间戳, 如果时间戳集合的长度大于给定的阈值, 则该情节是频繁的, 即为隐藏状态。

在日常生活环境下, 人的运动轨迹具有一定的规律性, 会形成一些活动较频繁的区域, 但是受地形、城市建筑布局等因素影响, 这些区域常常是不规则的图形, 很难直接地描述其特征。本文提出了一种基于 HMM 模型的出行行为检测方法, 改进了 HMM 模型中隐藏状态的确定方法, 并通过判别分析法找出检测轨迹对应的状态序列, 从而提高轨迹检测的准确性。

1 DBSCAN 算法

DBSCAN (density-based spatial clustering of applications with noise) 算法是一种基于密度的聚类算法, 可从有噪声的空间数据集合中发现任意形状的聚类并自动确定簇的数量。算法的核心思想是: 对于构成簇的每个对象, 在其半径 Eps 内所包含的对象个数, 必须不小于一个给定值 $MinPts$ 。

收稿日期: 2014-07-03

作者简介: 刘翔(1987-), 男, 浙江杭州人, 硕士研究生, 主要从事企业信息化方面的研究。

通信作者: 王亚沛, E-mail: lx_zjstu@163.com

DBSCAN算法的主要定义如下。

a) 定义1 Eps 邻域

给定一个数据对象 p , 其 Eps 邻域 $N_{Eps}(p)$ 的定义为

$$N_{Eps}(p) = \{p' \in D \mid dist(p, p') \leq Eps\} \quad (1)$$

其中: D 为空间数据集, $dist(p, q)$ 表示 D 中的对象 p 和 p' 之间的距离。

b) 定义2 直接密度可达

给定半径 Eps 和 $MinPts$, 如果对象 p 和 p' 满足: $p \in N_{Eps}(p')$; $|N_{Eps}(p')| \geq MinPts$, 则称对象 p' 从对象 p 直接密度可达。

c) 定义3 密度可达

给定对象集合 D , 当存在一个对象链 p_1, p_2, \dots, p_n , 其中 $p_1 = p', p_n = p$, 对于 $p_i \in D$, p_{i+1} 是从 p_i 关于 Eps 和 $MinPts$ 直接密度可达的, 则称对象 p 从对象 p' 关于 Eps 和 $MinPts$ 密度可达。密度可达是非对称的。

d) 定义4 密度相连

如果空间数据集 D 中存在一个对象 r , 使得对象 p 和 p' 是从 r 在 Eps 和 $MinPts$ 条件下密度可达的, 则称对象 p 和 p' 是在该条件下密度相连。密度相连是对称的。

e) 定义5 簇和噪声

从任意一个核心点对象开始, 将该对象密度可达的所有对象划分成一个簇。不属于任何簇的对象为噪声。

f) 定义6 核心点

对于数据对象 $p \in D$, 给定 $MinPts$ 如果 p 的 Eps 邻域内的对象数满足 $|N_{Eps}(p)| \geq MinPts$, 则称 p 为在 Eps 和条件下的核心点^[7]。

在日常生活环境中, 为了标识运动轨迹, 有的论文在环境内选取一些标志点, 并把这些标志点作为轨迹序列对比的参考点, 对于不同的方法, 其标志点选取的差异性较大。选取标志点作为轨迹特征状态的方法, 用于圆形区域会比较适合, 但监测目标的运动轨迹可能是任意形状, 就很难确定标志点及其数量。因此本文采用 DBSCAN 聚类算法, 只需给定 $MinPts$, 就能对轨迹数据进行分类。对轨迹序列进行分类, 假定可以分为 C 个类, 则对于轨迹序列中的所有轨迹点, 其隶属度 $u_{ij} \in \{0, 1\}$, 表示第 j 个轨迹点对第 i 类聚类的隶属度。将划分后的类所形成的区域作为运动轨迹的标志区域, 即 HMM 模型的隐藏状态。

2 HMM 模型

HMM 模型是一种具有双重随机过程的统计学

模型。其状态无法直接观测到, 但是能够通过观测向量序列间接观察到, 而观测向量都是通过概率密度分布表现为各种隐藏状态, 并由一个具有相应概率密度分布的隐藏状态序列产生。其核心问题是从可观察的序列中确定该过程的隐含参数, 然后利用这些参数来做进一步的行为分析。

HMM 模型可用参数 $\lambda = (A, B, \pi)$ 来表示, A 为状态转移概率分布, B 为观测值概率分布, π 为初始状态分布。HMM 模型的学习过程可以通过 Baum-Welch 算法实现。BW 算法的目标是找出有隐性变量的概率模型的最大可能性解, 其核心思想是对 HMM 模型参数进行 $\lambda = (A, B, \pi)$ 递归估计, 使重估的 λ 对于给定的观测序列的似然概率不断增大直至收敛, 从而使模型的参数逐渐地逼近真实值^[8-10]。

模型训练算法描述如下:

a) 假设时间长度为 T , t 为该时间段内的某一时刻, $1 \leq t \leq T$, 观测值集合 O 为所有的轨迹序列的坐标集合 $\{p_1, p_2, \dots, p_T\}$ 。利用 DBSCAN 算法对集合 O 聚类得到 C 个类, 则观测值所对应的类的集合为 $\{q_1, q_2, \dots, q_T\}$, 即状态集合 S 。

b) 计算前向变量 $\alpha_t(i)$ 及后向变量 $\beta_t(i)$,

$$\alpha_t(i) = P(p_1, p_2, \dots, p_t, q_t = S_i \mid \lambda) \quad (2)$$

$$\beta_t(i) = P(p_{t+1}, p_{t+2}, \dots, p_T, q_t = S_i \mid \lambda) \quad (3)$$

其中 S_i 为隐藏状态序列中的某个状态。

c) 计算在 t 时刻位于状态 S_i , 而在 $t+1$ 时刻处于状态 S_j 的概率 $\xi_t(i, j)$ 以及在 t 时刻处于 S_i 状态的概率 $\gamma_t(i)$,

$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_j \mid O, \lambda) = \frac{\alpha_t(i) a_{ij} b_j(p_{t+1}) \beta_{t+1}(j)}{P(p \mid \lambda)} \quad (4)$$

$$\gamma_t(i) = P(q_t = S_i \mid O, \lambda) = \frac{\alpha_t(i) \beta_{t+1}(j)}{P(p \mid \lambda)} \quad (5)$$

其中: a_{ij} 是隐藏状态 S_i 转移到 S_j 的概率, $b_j(p_{t+1})$ 是隐藏状态 S_j 输出观察结果 p_{t+1} 的概率。

d) 对参数进行重估, 得到隐马尔科夫模型 $\lambda = (A, B, \pi)$ 。

$$\bar{\pi}_i = \gamma_1(i) \quad (6)$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (7)$$

$$\bar{b}_j(p_t) = \frac{\sum_{\substack{t=1 \\ p_t=O_k}}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \quad (8)$$

其中 O_k 为观测值序列中的某个值。

通过多次的迭代, 循环直至满足收敛条件为止,

最终得到隐马尔科夫模型 $\lambda = (A, B, \pi)$, $A = \{\bar{a}_{ij}\}$, $B = \{\bar{b}_j(p_i)\}$, $\pi = \{\pi_i\}$ 。最终得到的参数是该隐马尔科夫模型的极大似然估计。

由于用 BW 算法训练 HMM 模型非常耗时,则需控制序列的长度来减少计算开销。在实际的应用中,工作日和节假日的出行规律会有显著不同,则可分为工作日和节假日,在此基础上再分时间段,分别建立对应的 HMM 模型。建立不同时间段的 HMM 模型所用的方法相同,因此本文只将相同时期中的相同时间段的数据作为实验数据。

3 轨迹识别与检测

利用 HMM 模型建立正常的运动轨迹后,实时检测运动轨迹,并对异常行为报警。先对实时监测的一段轨迹 $O' = \{p'_1, p'_2, \dots, p'_T\}$ 的数据点进行判别分析。

$$y = \operatorname{argmin}_{1 \leq x \leq T} [\operatorname{dist}(p'_t, p_x)] \quad (9)$$

其中: p'_t 为实时监测的数据点, p_x 为训练数据中已分类的数据点(除去噪声点)。

由式(9),找出与 p'_t 相邻的点 p_y 。用 DBSCAN 算法对数据集聚类,同一个类中两点之间的距离不大于 Eps 。因此,如果 $\operatorname{dist}(p'_t, p_y) \leq Eps$,则 p'_t 与 p_y 属于同一个类;反之,则 p'_t 属于异常点。用该判别方法得到实时监测轨迹所对应的类的集合 $\{q'_1, q'_2, \dots, q'_T\}$ 。

$$P(O' | \lambda) = \sum_{i=1}^C \alpha_T(i) \quad (10)$$

然后根据式(10)计算该轨迹序列的输出概率。如果 $\log(P(O' | \lambda)) < \epsilon$,则将该序列视作异常并报警。而作为判断异常的阈值 ϵ 需要根据实验数据进行选择。

4 实验

用 GPS 设备记录监测对象的运动轨迹,为了计算方便,要将轨迹进行坐标转换,以初始点作为坐标原点,将所有数据点转换到该坐标系。采用 Matlab 进行实验仿真来验证本文的方法。

首先,将采集的运动轨迹分别用 K-medoids, FCM 和 DBSCAN 进行聚类分析,得到 C 个簇。K-medoids, FCM 算法需要通过人工的方法观察轨迹点的分布区域,来最终确定 C 的值。而 DBSCAN 只需给定 $MinPts$,就能对轨迹数据进行分类。根据已有的研究成果, $MinPts$ 一般取 4 就能得出较好的聚类效果。这 3 种算法的聚类效果如图 1 所示。

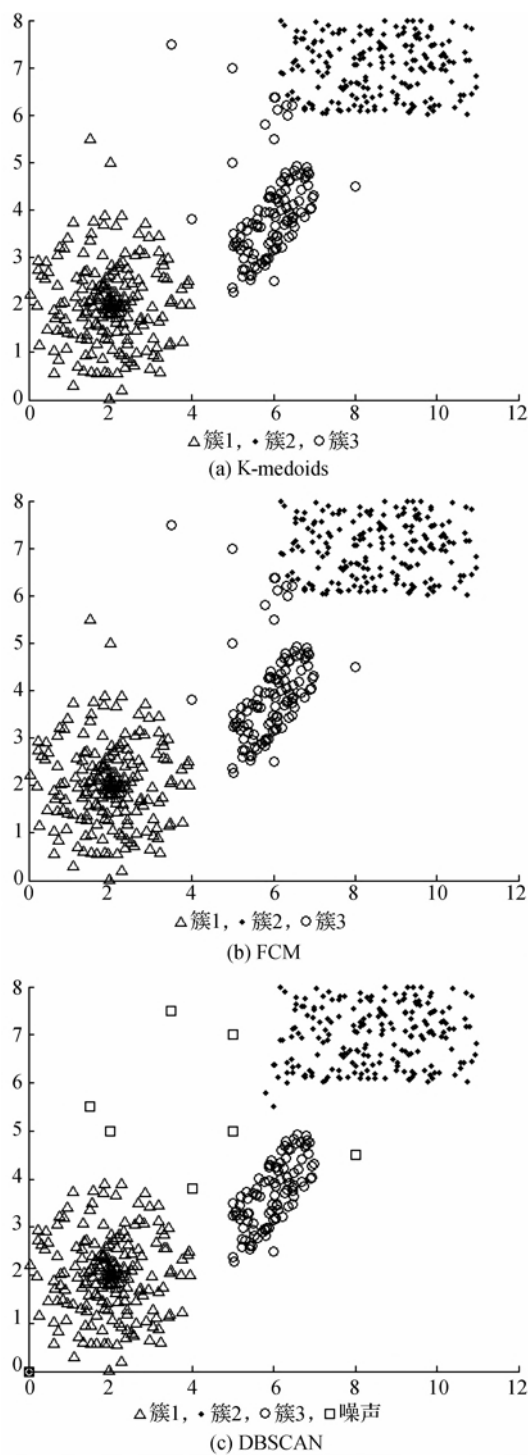


图1 聚类效果图

从图 1 和表 1 可以得出,虽然 K-medoids 算法和 FCM 算法得到聚类结果相同,但他们的聚类标志点不同,这会对接下去的异常检测产生影响。这两种算法都要预先设定 C 值,都将属于簇 2 的点错划到簇 3 中,并且都不能识别噪声点。而 DBSCAN 算法不会得到聚类标志点,而是得到聚类标志区域。同时该算法能识别噪声点,且不需要预先设定 C 值,可以减少人为的干预。

表1 聚类结果统计表

聚类算法	C 值	聚类标志
K-medoids	3	(2.01, 2.02), (5.89, 3.96), (8.66, 6.99)
FCM	3	(1.99, 2.03), (5.95, 3.93), (8.54, 6.99)
DBSCAN	3	聚类标志区域

然后在异常测试阶段,把待测试的轨迹样本数据做一些处理,人工加入一些异常轨迹点用于异常轨迹检测。

表2中的耗时包括了聚类分析,训练HMM模型和异常检测的运算时间。从表2可以得出,本文算法与其他两种算法相比,耗时有一定的增加,这是因为本文算法在检测轨迹的识别过程中,对检测数据的判别分析比另两种复杂。

测评结果表明,本算法训练模型和检测异常所消耗的时间大于另外两种,但是对异常检测的准确率要高于另两种。

表2 异常检测结果统计表

异常检测算法	耗时/s	准确率/%
K-medoids+HMM	65.4	78.6
FCM+HMM	67.3	76.3
本文算法	84.2	86.6

5 结束语

针对老年人出行异常的检测问题,本文提出了一种基于HMM模型的出行行为检测方法。将出行轨迹的坐标点作为模型的训练数据,改进了HMM模型中隐藏状态的确定方法,建立老年人日常行为模型,然后提出一种判别分析方法用于找出检测轨迹对应的状态序列,计算该轨迹序列的输出概率,并与阈值相比较,从而判断是否出现异常行为。通过实验证明了该方法对老年人行为异常检测有较高的准确性。

参考文献:

- [1] 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.
- [2] 余涛, 邹建华. 基于Bayes规则与HMM相结合的步态识别方法研究[J]. 计算机学报, 2012, 35(2): 386-396.
- [3] 王相海, 丛志环, 方玲玲. 基于HMM的车辆行驶状态实时判别方法研究[J]. 自动化学报, 2013, 39(12): 2131-2142.
- [4] 朱旭东, 刘志镜. 基于主题隐马尔科夫模型的人体异常行为识别[J]. 计算机科学, 2012, 39(3): 251-255.
- [5] 孔令富, 李海涛. 基于位置和运动轨迹的老年人异常行为检测[J]. 计算机工程与设计, 2012, 33(2): 735-739.
- [6] 丁勇, 朱辉生, 曹红根. 基于混合EHMM模型的数据流预测[J]. 计算机科学, 2014, 41(6): 391-393.
- [7] Campello R J G B, Moulavi D, Sander J. Density-based clustering based on hierarchical density estimates[C]//Advances in knowledge Discovery and Datamining. Berlin: Springer Berlin Heidelberg, 2013: 160-172.
- [8] Rabiner L. A tutorial on hidden Markov models and selected applications in speech recognition [J]. Proceedings of the IEEE, 1989, 77(2): 257-286.
- [9] Shehzad K. Activity classification and anomaly detection using m-medoids based modeling of motion patterns[J]. Pattern Recognition, 2010, 43(10): 3636-3647.
- [10] 张德丰. MATLAB 概率与数理统计分析[M]. 北京: 机械工业出版社, 2010: 319-329.
- [11] Chen M, Madden M G, Liu Y. Refined learning of hidden Markov models with a modified Baum-Welch algorithm and informative components[C]//Progress in Informatics and Computing (PIC), 2010 IEEE International Conference. IEEE, 2010: 165-169.

Detection of Abnormal Travel Behavior for the Old Based on Hidden Markov Model

LIU Xiang, ZOU Ting-ting, LI Ren-wang, WANG Ya-pei

(School of Mechanical Engineering & Automation, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: This paper puts forward travel behavior detection method based on HMM model in allusion to abnormal travel behavior for the old. This paper improves hidden state determination method and sets up daily behavior mode of the old with coordinate points of travel track. Then, a discriminant analysis is proposed to find out stage series corresponding to detection track, calculate output probability of this track series and judge whether there is abnormal behavior through comparing with the threshold value. The experiment proves this method has high accuracy in detecting abnormal behavior of the old.

Key words: elderly; behavior detection; cluster analysis; HMM; DBSCAN

(责任编辑: 陈和榜)