

基于智能移动终端 的行为识别方法研究

JIYU ZHINENGYIDONG ZHONGDUAN
DE XINGWEISHIBIE FANGFA YANJIU

赵中堂 著



电子科技大学出版社

基于智能移动终端 的行为识别方法研究

赵中堂 著



电子科技大学出版社

图书在版编目(CIP)数据

基于智能移动终端的行为识别方法研究 / 赵中堂
著. —成都: 电子科技大学出版社, 2015. 4

ISBN 978-7-5647-2877-9

I. ①基… II. ①赵… III. ①移动终端—行为
分析—方法研究 IV. ①B848.4

中国版本图书馆 CIP 数据核字(2015)第 049026 号

基于智能移动终端的行为识别方法研究

赵中堂 著

出 版: 电子科技大学出版社(成都市一环路东一段 159 号
电子信息产业大厦 邮编:610051)

策划编辑: 谭炜麟

责任编辑: 谭炜麟

主 页: www.uestcp.com.cn

电子邮箱: uestcp@uestcp.com.cn

发 行: 新华书店经销

印 刷: 郑州宏达印务有限公司

成品尺寸: 145mm×210mm 印张 6.5 字数 164 千字

版 次: 2015 年 4 月第一版

印 次: 2015 年 4 月第一次印刷

书 号: ISBN 978-7-5647-2877-9

定 价: 25.00 元

■ 版权所有 侵权必究 ■

◆ 本社发行部电话:028—83202463;本社邮购电话:028—83201495。

◆ 本书如有缺页、破损、装订错误,请寄回印刷厂调换。

内容简介

本书以移动智能终端为数据采集工具,针对终端携带者的行为识别展开研究,发现现有的行为识别方法存在一些技术难点,比如,离线阶段基于某个加速度传感器建立的模型不能在在线阶段很好地区分来自其他不同型号参数加速度传感器的数据、离线阶段基于某些特定位置的行为数据建立的模型不能在在线阶段很好地区分新位置的行为数据、离线阶段基于一人或者多人的行为数据建立的模型不能在在线阶段很好的区分新用户的行为数据、离线阶段基于某些特定类别行为数据构建的模型在在线阶段不能区分新类别行为数据。对这些问题,本书进行了深入剖析,揭露了导致这些问题的本质原因——目标域与源域的数据分布不同。为弥补两个数据域之间的分布差异,本书从迁移学习的角度出发,分别提出了相应的解决方案,获得了一定的效果。由于该领域的研究还处于初步阶段,有进一步挖掘的价值,本书的出版希望能起到抛砖引玉的作用,引起广大研究人员对该问题的兴趣。

前 言

行为识别是普适计算领域中重要的研究内容,具有广阔的应用前景,如“空巢”老人生活远程监护、肥胖病人热量消耗估计、心脑血管等慢性病人的疾病监测、老年人的跌倒检测与报警等。现有的行为识别方法一般分为离线训练和在线识别两个阶段,在离线阶段基于标定的行为数据训练行为识别模型,在在线阶段应用离线阶段训练的模型对未标定的行为数据进行预测。然而,现有的行为识别方法存在一些技术难点,比如离线阶段基于某个加速度传感器建立的模型不能在在线阶段很好的区分来自其他不同型号参数加速度传感器的数据,我们称之为非特定传感器行为识别问题;离线阶段基于某些特定位置的行为数据建立的模型不能在在线阶段很好的区分新位置的行为数据,我们称之为非特定位置行为识别问题;离线阶段基于一人或者多人的行为数据建立的模型不能在在线阶段很好的区分新用户的行为数据,我们称之为非特定人行为识别问题;离线阶段基于某些特定类别行为数据构建的模型在在线阶段不能区分新类别行为数据,我们称之为非特定类别行为识别问题。本书以迁移学习思路为指导,针对不同问题分别提出各自的解决方法,具体包括以下几个方面:

(1)提出了基于静止状态检测和时频域变换的异质加速度

传感器数据预处理方法。针对由于加速度传感器参数不同而导致的源域(离线阶段)和目标域(在线阶段)数据分布不同的问题,该方法通过检测静止状态,计算每种加速度传感器静止时的合成加速度,应用各自的静止合成加速度去规范各自的加速度数据,达到统一量纲的目的,极大地减少相同行为在时域特征的不同。然后,采用在每秒时间窗口内进行重采样的方法,使得各种传感器的采样频率相同,极大地减小相同行为在频域特征的差别。实验结果表明该方法能够将不同参数加速度传感器的输出值映射到一致的数据空间,从而能够实现模型在内置有异质加速度传感器的便携式移动终端间的迁移,达到构建非特定传感器行为识别模型的目的。

(2)提出了基于特征迁移的非特定位置行为识别方法。针对不同位置加速度传感器感知到的数据分布不同的问题,该方法首先利用主成分分析技术,寻找不同位置数据的主成分,生成各个位置数据的共有特征,并基于极速学习机算法离线训练一个行为识别模型。然后,在线识别来自不同位置的数据,给每个样本一个预测标号和对应的置信度,并选择高置信度的样本在线增量更新已有模型,当高置信度样本中含有新位置的样本时,更新后的模型就逐渐拥有了对新位置样本的区分能力。实验结果表明本书提出的方法能够在保持模型对已有位置行为数据识别精度的同时提高模型对新位置行为数据的识别能力,达到构建非特定位置行为识别模型的目的。

(3)提出了基于参数迁移的非特定人行为识别方法。针对使用模型的用户和训练模型的人员行为数据分布不同的问题,我们首先假设每个人的行为识别模型结构(本书指决策树模型)是相同的,不同之处在于模型的参数的值(本书指决策树非叶子结点的值),由此我们提出在源域和目标域之间迁移决策树及其

参数,并针对新用户的数据,对参数的值进行调整,使模型成为一个个性化的行为识别模型。实验结果表明该方法能够在在线阶段自动地利用用户的数据对模型进行更新,达到针对目标用户构建个性化模型的目的。

(4)提出了基于结构迁移的非特定类别行为识别方法。针对目标域中新类行为数据的出现而导致的整体数据分布发生变化的问题,我们基于源域数据训练具有疑似异常行为识别能力的极速学习机,用其对目标域数据进行分类,若某些样本的预测置信度处于异常区间之内,则认为其为疑似异常行为。若当前用户积累了一定数量疑似异常行为数据,系统将其上传到“云端”,通过无监督聚类算法可将它们划分成不同的类别,并邀请相关专家对每个类别提供标号,在此基础上采用类增量极速学习的方法更新模型,使得模型越来越能描述目标域的数据。实验结果表明由于该方法能够保持从源域数据学习得到的结构知识,因此能够在保持模型对已有类别行为识别精度的基础上,逐渐加入对新类别行为的识别能力,能够达到构建非特定类别行为识别模型的目的。

(5)面向实际应用的需求,实现了基于智能移动终端的社区老人健康感知公共服务原型系统。在日常生活中可以识别老人的行为,检测老人的运动量,为促进其合理运动提供帮助;在老人跌倒时,可以及时地发出警报,极大程度地缩短从老人跌倒到得到救助之间的时间,实现对老人健康的监护。

本书由八章组成,各章节组织如下:

第一章,介绍了本书的背景,提出了本书所研究的问题及问题的研究现状。

第二章,介绍了基于加速度传感器的行为数据的采集和特征提取,得到的样本集是后续各章节实验的基础。

第三章,提出了基于静止状态检测和时频域变换的异质加速度传感器数据预处理方法,能够将不同参数加速度传感器的数据映射到统一的数据空间。

第四章,针对基于某些特定位置样本的行为识别模型不能很好地识别新位置样本的问题,提出了基于特征迁移的位置无关识别方法。

第五章,针对基于某些用户样本的行为识别模型不能很好地识别新用户样本的问题,提出了基于参数迁移的非特定人为识别方法。

第六章,为解决目标域中新类数据的出现而导致的数据分布不同的问题,我们提出了基于结构迁移的非特定类别行为识别方法。

第七章,基于本书的部分研究成果,我们构建了一个基于智能移动终端的社区老人健康感知公共服务原型系统。

第八章,对本书做出的工作进行回顾和总结,并且指出了一些未来可能的研究方向。

在本书的编写和出版过程中,得到了郑州航空工业管理学院以及计算机科学与技术系有关领导和专家的大力支持,在此谨向他们以及本书引用的参考文献的作者表示衷心的感谢。

虽然本书是作者近年来研究工作的总结,有一定的理论和实验基础,但由于作者水平有限,书中错误在所难免,敬请广大读者不吝赐教。

编 者

2014 年 12 月

目 录

第一章 引言

1.1 什么是行为识别	1
1.2 为什么需要行为识别	2
1.3 研究行为识别的工具和方法有哪些	5
1.4 我们的研究内容、研究工具、研究方法是什么	10
1.4.1 研究内容——个人日常行为识别及其流程	10
1.4.2 研究工具——行为识别中的传感器设备	14
1.4.3 研究方法——机器学习方面的研究现状	16
1.5 我们遇到的挑战有哪些	48
1.6 本书的贡献	50

第二章 行为数据采集和特征提取

2.1 数据采集要求	54
2.1.1 现有的公共数据集	54
2.1.2 数据采集设备的多样性	58
2.1.3 数据采集要求	62
2.2 数据采集过程	62
2.3 数据分割和标定	63

2.4 特征提取	64
2.4.1 时域特征	65
2.4.2 频域特征	67
2.5 样本集统计信息	69
2.6 小结	73
第三章 非特定传感器行为识别方法	
3.1 问题描述	74
3.2 相关研究工作	77
3.3 面向异质加速度传感器数据的预处理方法	79
3.3.1 合成加速度的规范化方法	80
3.3.2 基于时间窗的采样频率一致化方法	81
3.4 实验及结果分析	83
3.4.1 数据准备	83
3.4.2 量纲对数据分布的影响	84
3.4.3 采样频率对数据分布的影响	87
3.4.4 方法的效果	91
3.5 小结	95
第四章 非特定位置行为识别方法	
4.1 问题描述	96
4.2 相关研究工作	98
4.3 基于特征迁移的非特定位置行为识别模型	101
4.3.1 合成加速度数据	102
4.3.2 基于 PCA 的特征降维处理	103
4.3.3 极速学习机分类器	104

4.3.4 样本的自标定	108
4.3.5 模型的在线增量更新	109
4.4 实验及结果分析	111
4.4.1 分类器性能比较	113
4.4.2 未进行 PCA 变换时的交叉位置行为识别实验	114
4.4.3 进行 PCA 变换后的交叉位置行为识别实验	115
4.4.4 自适应调整后模型对新旧位置数据的识别能力实验	117
4.5 小结	120

第五章 非特定人行为识别方法

5.1 问题描述	121
5.2 相关研究工作	122
5.3 一种基于迁移学习的嵌入决策树算法	124
5.3.1 算法框架	125
5.3.2 二叉决策树的构建及迁移	126
5.3.3 新用户高置信度样本的筛选过程	127
5.3.4 二叉决策树模型的更新过程	130
5.4 实验及结果分析	132
5.4.1 数据准备	132
5.4.2 决策树模型对个人数据的区分能力	133
5.4.3 不同用户两两之间模型的交叉验证实验	134
5.4.4 算法性能	135
5.4.5 决策树模型和训练样本数量的关系	135
5.4.6 个性化模型和新用户样本的关系	136
5.5 小结	137

第六章 非特定类别行为识别方法

6.1 问题描述	139
6.2 相关研究工作	140
6.3 基于结构迁移的非特定类别行为识别方法	142
6.3.1 算法框架	142
6.3.2 具有疑似异常行为识别能力的极速学习机	143
6.3.3 新类别样本的发现和标定	145
6.3.4 极速学习机模型结构的迁移和更新	146
6.4 实验及结果分析	150
6.4.1 数据准备	151
6.4.2 模型选择	152
6.4.3 识别异常行为的实验	152
6.4.4 分类器性能 VS 新样本数量	153
6.4.5 行为识别模型稳定性实验	154
6.4.6 批量学习与增量学习的性能比较实验	155
6.5 小结	156

第七章 基于智能移动终端的社区老人健康感知公共服务系统

7.1 问题描述	157
7.2 系统架构	159
7.3 系统的基本功能	160
7.3.1 行为识别	161
7.3.2 热量消耗估计	161
7.3.3 跌倒报警	162

7.4 系统功能的设计与实现	163
7.4.1 行为识别功能的实现	163
7.4.2 行为统计分析功能的实现	167
7.4.3 热量消耗估计功能的实现	168
7.4.4 跌倒检测功能的实现	171
7.5 小结	172
第八章 结束语	
8.1 本书工作总结	173
8.2 未来工作展望	175
参考文献	178

第一章 引言

本章通过综合消化各方面的文献,首先探讨什么是行为识别,明确研究对象;然后通过具体的例子,例证为什么需要行为识别;接着从技术的角度出发,列出研究行为识别的工具和方法有哪些;最后重点强调我们的研究内容、选择的研究工具和方法、在研究中遇到的各种问题,这是本文要探讨的核心问题。

1.1 什么是行为识别

行为是指人们受思想支配而表现出来的外表活动,人的行为可分为外显和内在行为。外显行为是可以被他人直接观察到的行为,它是由一系列简单动作构成的,是在日常生活中所表现出来的一切动作的统称。而内在行为则是不能被他人直接观察到的行为,如意识、思维活动等,即通常所说的心理活动。一般情况下,可以通过观察人的外显行为,进一步推测其内在行为。

行为识别的目标是通过传感器设备获取与人行为相关的信息,通过技术手段从行距信息反推出行为的名称,并能通过连续的行为序列预测用户进一步的行为。

当前的行为识别还处于初级阶段,有待进一步发展。

1.2 为什么需要行为识别

个人日常行为识别具有重要的研究意义。比如,在“空巢”老人生活远程监护、肥胖病人热量消耗估计、心脑血管等慢性病人的疾病监测、老年人的跌倒检测与报警等领域个人日常行为识别有着广阔的应用前景,具体而言:

(1)个人日常行为识别在老年人远程监护领域有广阔的应用前景。第六次全国人口普查结果显示,全国 60 岁及其以上人口已占 13.26%。根据国际公认标准,60 岁以上人口比例超过 10%,或者 65 岁以上人口超过 7%,就算进入老龄化社会[岳俊峰,2004]。按此标准,我国已经步入了老龄化社会。随着老龄化社会的到来,养老问题日益成为一个社会问题。在各种养老方式中,“居家养老”是人们的首选。但是随着社会人口结构的调整,“空巢”老人的比例日益增大,这些老年人日常起居无人看护,一旦发生意外,后果不堪设想。因此,自动识别老年人日常行为,在发现异常行为时能及时报警,是实现“居家养老”的重要保障,这对于像我国这样的人口大国意义尤其重大。

(2)个人日常行为识别在肥胖病人热量消耗估计领域有重要的市场需求。据统计,2007 年美国人大约有 30%罹患肥胖症。肥胖症导致的高血压、冠心病、糖尿病、痛风、胆石症、脂肪肝等疾病,每年大约消耗了 1 000 亿美元。而当年中国人中超重者的比例大约为 28.9%。肥胖已成为困扰世界的难题。大量研究证明,运动是一个能够保持健康的好方法,只要长期进行有规律地进行运动,就能使人们保持健康。减肥也是这样,适量

运动有助于燃脂、减重、强身健体。然而,目前没有便携的工具能准确地检测用户每天的日常行为,进而计算用户运动量和热量消耗,也就无法科学地衡量用户摄入的能量和消耗的能量之间的关系,从而无法给予用户科学的指导。因此自动识别个人日常行为,寻找行为产生的运动量与热量消耗之间的运动关系,是当前的研究热点,有着重要的市场需求。

(3)个人日常行为识别在心脑血管等慢性病人的疾病监测方面能够发挥越来越重要的作用。据统计,全球每年因冠心病、脑梗塞等心脑血管疾病死亡的人数约为 1 750 万,占全球总死亡的 29%,而这些死亡患者有 80%处于发展中国家。据卫生部统计,我国中老年人高血脂患者有 2 亿人、高血粘患者有 1.5 亿人、高血压患者有 1.5 亿人、冠心病患者有 1 亿人、脑中风患者有 2 000 万人、高血糖和糖尿病患者有 9 000 万人,且患病人数和发病率呈现逐年增加的趋势。心脑血管相关疾病现已成为我国需要重点关注和亟待解决的公共卫生领域的重大问题,做好该病的预防性工作控制其发生发展、进而提高人们生活质量的关键。心脑血管疾病具有发病率高、致残率高、死亡率高、复发率高、并发症多等“四高一多”的特点,但目前病人一般是在症状已经较为明显甚至病情已然严重时才入院进行诊治。而流行病学数据显示,有效改善入院前病情监护水平比住院强化治疗更有意义,同时,治疗越早康复效果越好。因此,必须缩短初始症状出现到确定治疗方法的时间周期,逐步从院内监护扩展到院外长期实时监护,及早诊断,及时救治。基于此,实时准确地监测心脑血管等慢性病人的个人日常行为,发现异常后能够及

时地报警,能够有效地保护他们的健康。

(4)个人日常行为识别能够检测老人的跌倒行为并及时报警[Shi,2012]。老年人由于身体衰弱,平衡能力较差,跌倒已成为威胁他们健康的重要因素。除了平时加强锻炼,防患于未然外,在跌倒事实已经形成后,及时地通知其监护人或者医疗救助机构,也能极大地减轻身体受到的危害。据统计,中国每年有不少于2000万老年人不幸跌倒,大约有不少于2500万人次的跌倒,由此造成的住院治疗费用高达50亿元以上,社会代价大约为160到800亿元。跌倒已经是导致老年人死亡的首位因素。意外伤害中大约有一半就是老年人跌倒,而跌倒导致的死亡率已经高居第4位,排在其前面的且处于第一位的是心脑血管疾病,处于第二位的是肿瘤疾病,处于第三位的是呼吸道疾病。老人意外伤害多在家中发生,占57.75%,在公共场所发生的占21.13%,在街道马路发生的占18.31%,在其他场所发生的占2.82%。故实时监测老人的个人日常行为,准确地检测出跌倒行为并及时报警,能够极大地缩短从跌倒发生到获得救助之间的时间,为挽救老人生命争取更多的时间。

基于便携设备的非干扰行为识别正成为普适可行的方案。随着微电子机械系统(Micro Electrical Mechanical Systems, MEMS)的发展,微型加速度传感器得以研发并被嵌入小型移动设备中,这大大促进了基于便携式设备的行为识别系统的研究和部署。已有少数设备能够识别个人日常行为,比如文献[Consolvo,2006]中提到的计步器(Step Counter)以用户走过的步数量化用户的行为,非常适合测量散步和跑步的运动量。一