

清华大学优秀博士学位论文丛书



电子鼻信号分析 关键算法研究

闫轲 著 Yan Ke

Research on Key Signal Analysis Algorithms
for Electronic Noses

清华大学出版社
TSINGHUA UNIVERSITY PRESS

清华大学优秀博士学位论文丛书

电子鼻信号分析 关键算法研究

闫轲 著 Yan Ke

Research on Key Signal Analysis Algorithms
for Electronic Noses

清华大学出版社
北 京

内 容 简 介

电子鼻是一种利用气敏传感器阵列识别气味的设备,呼气分析是电子鼻的重要应用之一。通过检测呼出气体中与特定疾病相关的标志物,电子鼻可以辅助完成疾病筛查和监控,该方法具有无创、易操作等优点。本书侧重研究电子鼻系统的信号分析算法,在搭建电子鼻系统的基础上,针对电子鼻数据的特点研究信号分析过程中的关键算法,旨在提高电子鼻的准确性和实用性。本书内容包括三个方面:基于电子鼻的呼气分析系统设计,针对相关特征的迭代特征删除算法,针对设备差异和时变漂移的补偿算法。本书可以作为电子鼻、生物医学工程和机器学习领域研究者的参考书籍。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

电子鼻信号分析关键算法研究/闫轲著. —北京:清华大学出版社,2019
(清华大学优秀博士学位论文丛书)
ISBN 978-7-302-52369-7

I. ①电… II. ①闫… III. ①智能传感器—信号分析—研究 IV. ①TP212.6

中国版本图书馆 CIP 数据核字(2019)第 039035 号

责任编辑:石磊 王倩

封面设计:傅瑞学

责任校对:王淑云

责任印制:李红英

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印装者:三河市铭诚印务有限公司

经 销:全国新华书店

开 本:155mm×235mm 印 张:8.75 插 页:2 字 数:151千字

版 次:2019年5月第1版 印 次:2019年5月第1次印刷

定 价:69.00元

产品编号:077353-01



图 2.2 电子鼻内部结构

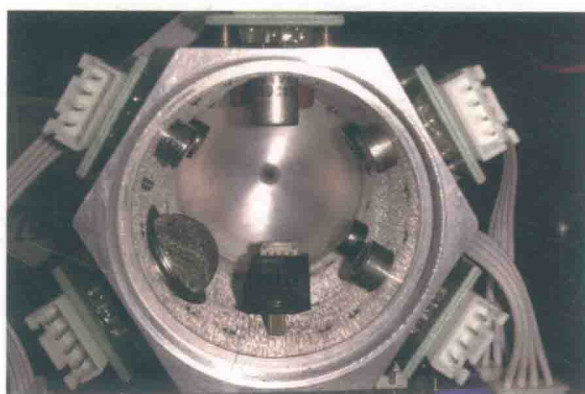


图 2.3 气室内部结构

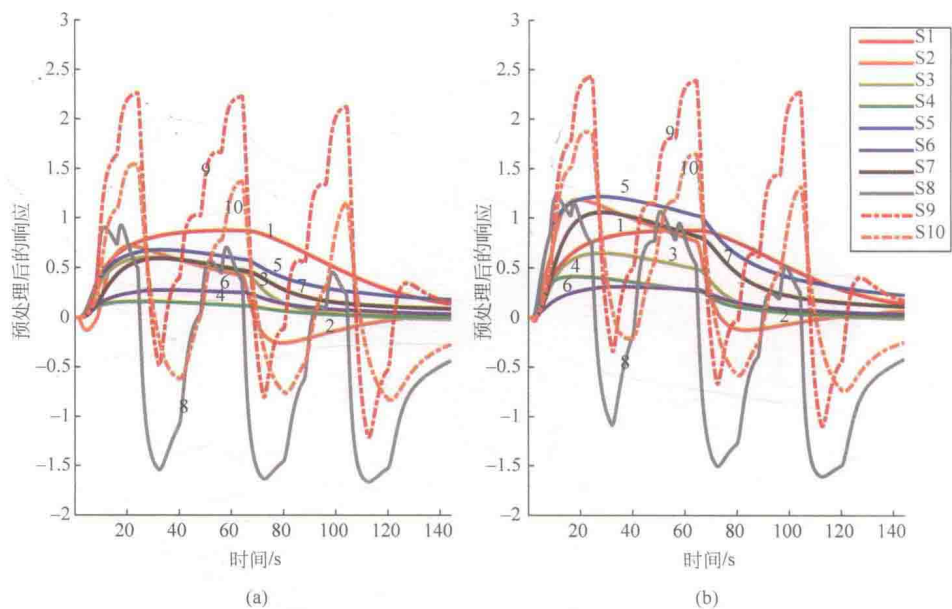


图 2.8 两类样本的平均响应

(a) 健康人; (b) 糖尿病患者

每条曲线代表一个气敏传感器,具体型号见表 2.2

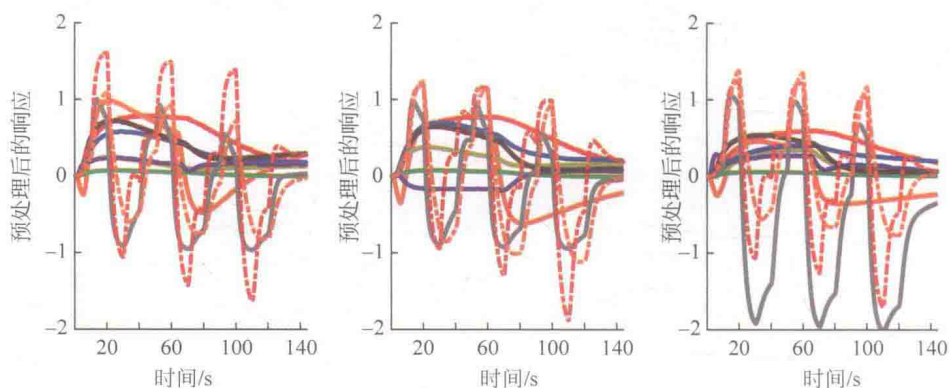


图 4.1 设备差异示例: 三台同样型号的电子鼻对同一个呼气样本的响应曲线

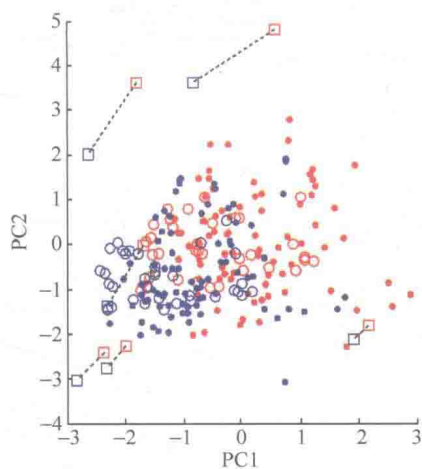


图 4.2 两台电子鼻上采集的被测样本和标定样本分布

蓝色和红色分别代表主设备和从设备,点和圈为两类被测样本,用虚线连接的方形为——对应的标定样本。坐标轴为被测样本的前二维主成分

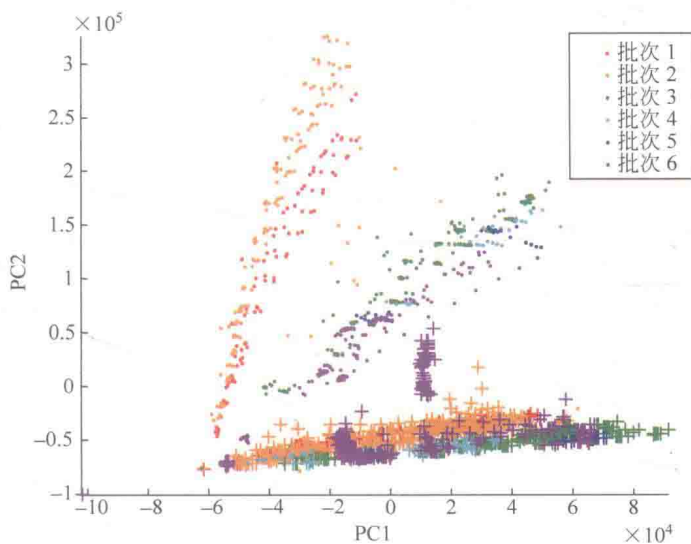


图 4.4 阵列漂移数据集里批次 1~6 的样本分布

点和加号分别代表氨气和乙醛样本;不同颜色代表不同批次

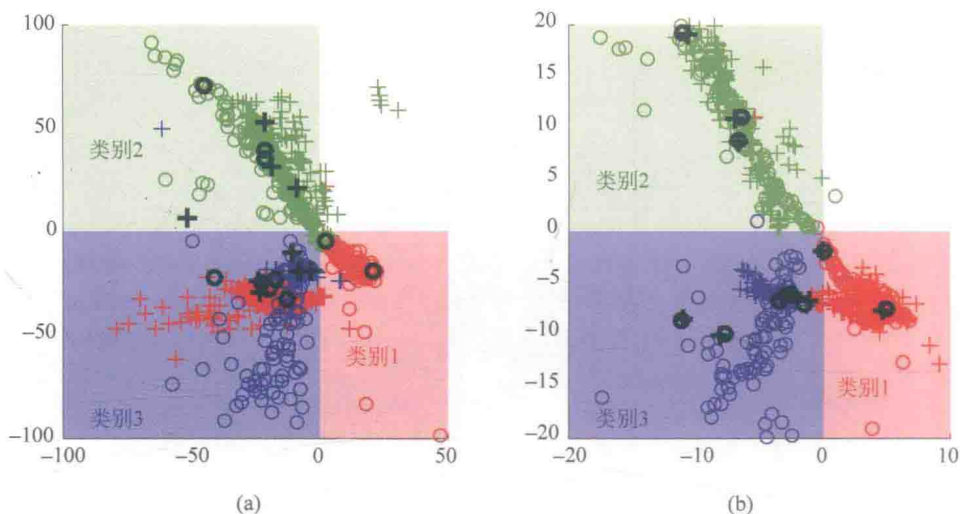


图 4.10 TMTL 效果图示

(a) 批次 1 和批次 2 的样本均由批次 1 的 LR 模型投影；(b) 用 TMTL 同时学习两个批次的模型，每个批次的样本由各自的模型投影

黑色粗线样本代表标定样本，其他不同颜色的样本代表不同类别。圆圈代表批次 1 的样本（源领域），加号代表批次 2 的样本（目标领域）。不同着色区域代表每个类别样本的正确分布区域

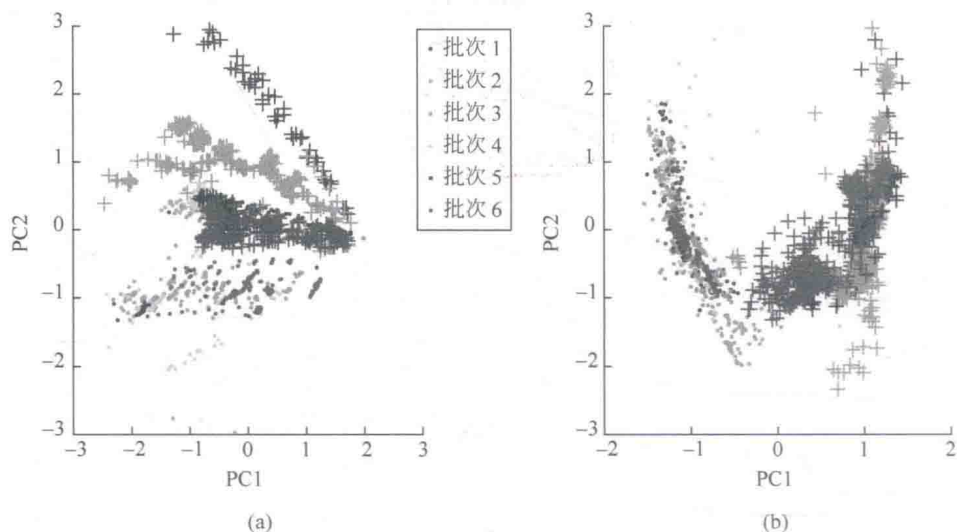


图 5.3 阵列漂移数据集中批次 1~6 的样本分布

(a) 原始特征；(b) 含补偿层的 DCAE 学习得到的特征

点和加号代表两个类别的样本；不同颜色代表不同批次

一流博士生教育 体现一流大学人才培养的高度(代丛书序)^①

人才培养是大学的根本任务。只有培养出一流人才的高校,才能够成为世界一流大学。本科教育是培养一流人才最重要的基础,是一流大学的底色,体现了学校的传统和特色。博士生教育是学历教育的最高层次,体现出一所大学人才培养的高度,代表着一个国家的人才培养水平。清华大学正在全面推进综合改革,深化教育教学改革,探索建立完善的博士生选拔培养机制,不断提升博士生培养质量。

学术精神的培养是博士生教育的根本

学术精神是大学精神的重要组成部分,是学者与学术群体在学术活动中坚守的价值准则。大学对学术精神的追求,反映了一所大学对学术的重视、对真理的热爱和对功利性目标的摒弃。博士生教育要培养有志于追求学术的人,其根本在于学术精神的培养。

无论古今中外,博士这一称号都是和学问、学术紧密联系在一起,和知识探索密切相关。我国的博士一词起源于2000多年前的战国时期,是一种学官名。博士任职者负责保管文献档案、编撰著述,须知识渊博并负有传授学问的职责。东汉学者应劭在《汉官仪》中写道:“博者,通博古今;士者,辩于然否。”后来,人们逐渐把精通某种职业的专门人才称为博士。博士作为一种学位,最早产生于12世纪,最初它是加入教师行会的一种资格证书。19世纪初,德国柏林大学成立,其哲学院取代了以往神学院在大学中的地位,在大学发展的历史上首次产生了由哲学院授予的哲学博士学位,并赋予了哲学博士深层次的教育内涵,即推崇学术自由、创造新知识。哲学博士的设立标志着现代博士生教育的开端,博士则被定义为独立从事学术研究、具备创造新知识能力的人,是学术精神的传承者和光大者。

^① 本文首发于《光明日报》,2017年12月5日。

博士生学习期间是培养学术精神最重要的阶段。博士生需要接受严谨的学术训练,开展深入的学术研究,并通过发表学术论文、参与学术活动及博士论文答辩等环节,证明自身的学术能力。更重要的是,博士生要培养学术志趣,把对学术的热爱融入生命之中,把捍卫真理作为毕生的追求。博士生更要学会如何面对干扰和诱惑,远离功利,保持安静、从容的心态。学术精神特别是其中所蕴含的科学理性精神、学术奉献精神不仅对博士生未来的学术事业至关重要,对博士生一生的发展都大有裨益。

独创性和批判性思维是博士生最重要的素质

博士生需要具备很多素质,包括逻辑推理、言语表达、沟通协作等,但是最重要的素质是独创性和批判性思维。

学术重视传承,但更看重突破和创新。博士生作为学术事业的后备力量,要立志于追求独创性。独创意味着独立和创造,没有独立精神,往往很难产生创造性的成果。1929年6月3日,在清华大学国学院导师王国维逝世二周年之际,国学院师生为纪念这位杰出的学者,募款修造“海宁王静安先生纪念碑”,同为国学院导师的陈寅恪先生撰写了碑铭,其中写道:“先生之著述,或有时而不章;先生之学说,或有时而可商;惟此独立之精神,自由之思想,历千万祀,与天壤而同久,共三光而永光。”这是对于一位学者的极高评价。中国著名的史学家、文学家司马迁所讲的“究天人之际、通古今之变,成一家之言”也是强调要在古今贯通中形成自己独立的见解,并努力达到新的高度。博士生应该以“独立之精神、自由之思想”来要求自己,不断创造新的学术成果。

诺贝尔物理学奖获得者杨振宁先生曾在20世纪80年代初对到访纽约州立大学石溪分校的90多名中国学生、学者提出:“独创性是科学工作者最重要的素质。”杨先生主张做研究的人一定要有独创的精神、独到的见解和独立研究的能力。在科技如此发达的今天,学术上的独创性变得越来越难,也愈加珍贵和重要。博士生要树立敢为天下先的志向,在独创性上下功夫,勇于挑战最前沿的科学问题。

批判性思维是一种遵循逻辑规则、不断质疑和反省的思维方式,具有批判性思维的人勇于挑战自己、敢于挑战权威。批判性思维的缺乏往往被认为是中国学生特有的弱项,也是我们在博士生培养方面存在的一个普遍问题。2001年,美国卡内基基金会开展了一项“卡内基博士生教育创新计划”,针对博士生教育进行调研,并发布了研究报告。该报告指出:在美国和

欧洲,培养学生保持批判而质疑的眼光看待自己、同行和导师的观点同样非常不容易,批判性思维的培养必须要成为博士生培养项目的组成部分。

对于博士生而言,批判性思维的养成要从如何面对权威开始。为了鼓励学生质疑学术权威、挑战现有学术范式,培养学生的挑战精神和创新能力,清华大学在2013年发起“巅峰对话”,由学生自主邀请各学科领域具有国际影响力的学术大师与清华学生同台对话。该活动迄今已经举办了21期,先后邀请17位诺贝尔奖、3位图灵奖、1位菲尔兹奖获得者参与对话。诺贝尔化学奖得主巴里·夏普莱斯(Barry Sharpless)在2013年11月来清华参加“巅峰对话”时,对于清华学生的质疑精神印象深刻。他在接受媒体采访时谈道:“清华的学生无所畏惧,请原谅我的措辞,但他们真的很有胆量。”这是我听到的对清华学生的最高评价,博士生就应该具备这样的勇气和能力。培养批判性思维更难的一层是要有勇气不断否定自己,有一种不断超越自己的精神。爱因斯坦说:“在真理的认识方面,任何以权威自居的人,必将在上帝的嬉笑中垮台。”这句名言应该成为每一位从事学术研究的博士生的箴言。

提高博士生培养质量有赖于构建全方位的博士生教育体系

一流的博士生教育要有一流的教育理念,需要构建全方位的教育体系,把教育理念落实到博士生培养的各个环节中。

在博士生选拔方面,不能简单按考分录取,而是要侧重评价学术志趣和创新潜力。知识结构固然重要,但学术志趣和创新潜力更关键,考分不能完全反映学生的学术潜质。清华大学在经过多年试点探索的基础上,于2016年开始全面实行博士生招生“申请-审核”制,从原来的按照考试分数招收博士生转变为按科研创新能力、专业学术潜质招收,并给予院系、学科、导师更大的自主权。《清华大学“申请-审核”制实施办法》明晰了导师和院系在考核、遴选和推荐上的权利和职责,同时确定了规范的流程及监管要求。

在博士生指导教师资格确认方面,不能论资排辈,要更看重教师的学术活力及研究工作的前沿性。博士生教育质量的提升关键在于教师,要让更多、更优秀的教师参与到博士生教育中来。清华大学从2009年开始探索将博士生导师评定权下放到各学位评定分委员会,允许评聘一部分优秀副教授担任博士生导师。近年来学校在推进教师人事制度改革过程中,明确教研系列助理教授可以独立指导博士生,让富有创造活力的青年教师指导优秀的青年学生,师生相互促进、共同成长。

在促进博士生交流方面,要努力突破学科领域的界限,注重搭建跨学科的平台。跨学科交流是激发博士生学术创造力的重要途径,博士生要努力提升在交叉学科领域开展科研工作的能力。清华大学于2014年创办了“微沙龙”平台,同学们可以通过微信平台随时发布学术话题、寻觅学术伙伴。3年来,博士生参与和发起“微沙龙”12000多场,参与博士生达38000多人次。“微沙龙”促进了不同学科学生之间的思想碰撞,激发了同学们的学术志趣。清华于2002年创办了博士生论坛,论坛由同学自己组织,师生共同参与。博士生论坛持续举办了500期,开展了18000多场学术报告,切实起到了师生互动、教学相长、学科交融、促进交流的作用。学校积极资助博士生到世界一流大学开展交流与合作研究,超过60%的博士生有海外访学经历。清华于2011年设立了发展中国家博士生项目,鼓励学生到发展中国家亲身体验和调研,在全球化背景下研究发展中国家的各类问题。

在博士学位评定方面,权力要进一步下放,学术判断应该由各领域的学者来负责。院系二级学术单位应该在评定博士论文水平上拥有更多的权力,也应担负更多的责任。清华大学从2015年开始把学位论文的评审职责授权给各学位评定分委员会,学位论文质量和学位评审过程主要由各学位分委员会进行把关,校学位委员会负责学位管理整体工作,负责制度建设和争议事项处理。

全面提高人才培养能力是建设世界一流大学的核心。博士生培养质量的提升是大学办学质量提升的重要标志。我们要高度重视、充分发挥博士生教育的战略性、引领性作用,面向世界、勇于进取,树立自信、保持特色,不断推动一流大学的人才培养迈向新的高度。



清华大学校长

2017年12月5日

丛书序二

以学术型人才培养为主的博士生教育,肩负着培养具有国际竞争力的高层次学术创新人才的重任,是国家发展战略的重要组成部分,是清华大学人才培养的重中之重。

作为首批设立研究生院的高校,清华大学自20世纪80年代初开始,立足国家和社会需要,结合校内实际情况,不断推动博士生教育改革。为了提供适宜博士生成长的学术环境,我校一方面不断地营造浓厚的学术氛围,一方面大力推动培养模式创新探索。我校已多年运行一系列博士生培养专项基金和特色项目,激励博士生潜心学术、锐意创新,提升博士生的国际视野,倡导跨学科研究与交流,不断提升博士生培养质量。

博士生是最具创造力的学术研究新生力量,思维活跃,求真求实。他们在导师的指导下进入本领域研究前沿,吸取本领域最新的研究成果,拓宽人类的认知边界,不断取得创新性成果。这套优秀博士学位论文丛书,不仅是我校博士生研究工作前沿成果的体现,也是我校博士生学术精神传承和光大的体现。

这套丛书的每一篇论文均来自学校新近每年评选的校级优秀博士学位论文。为了鼓励创新,激励优秀的博士生脱颖而出,同时激励导师悉心指导,我校评选校级优秀博士学位论文已有20多年。评选出的优秀博士学位论文代表了我校各学科最优秀的博士学位论文的水平。为了传播优秀的博士学位论文成果,更好地推动学术交流与学科建设,促进博士生未来发展和成长,清华大学研究生院与清华大学出版社合作出版这些优秀的博士学位论文。

感谢清华大学出版社,悉心地为每位作者提供专业、细致的写作和出版指导,使这些博士论文以专著方式呈现在读者面前,促进了这些最新的优秀研究成果的快速广泛传播。相信本套丛书的出版可以为国内外各相关领域或交叉领域的在读研究生和科研人员提供有益的参考,为相关学科领域的发展和优秀科研成果的转化起到积极的推动作用。

感谢丛书作者的导师们。这些优秀的博士学位论文,从选题、研究到成文,离不开导师的精心指导。我校优秀的师生导学传统,成就了一项项优秀的研究成果,成就了一大批青年学者,也成就了清华的学术研究。感谢导师们为每篇论文精心撰写序言,帮助读者更好地理解论文。

感谢丛书的作者们。他们优秀的学术成果,连同鲜活的思想、创新的精神、严谨的学风,都为致力于学术研究的后来者树立了榜样。他们本着精益求精的精神,对论文进行了细致的修改完善,使之在具备科学性、前沿性的同时,更具系统性和可读性。

这套丛书涵盖清华众多学科,从论文的选题能够感受到作者们积极参与国家重大战略、社会发展问题、新兴产业创新等的研究热情,能够感受到作者们的国际视野和人文情怀。相信这些年轻作者们勇于承担学术创新重任的社会责任感能够感染和带动越来越多的博士生们,将论文书写在祖国的大地上。

祝愿丛书的作者们、读者们和所有从事学术研究的同行们在未来的道路上坚持梦想,百折不挠!在服务国家、奉献社会和造福人类的事业中不断创新,做新时代的引领者。

相信每一位读者在阅读这一本本学术著作的时候,在吸取学术创新成果、享受学术之美的同时,能够将其中所蕴含的科学理性精神和学术奉献精神传播和发扬出去。



清华大学研究生院院长

2018年1月5日

导师序言

传统中医主要采用四诊合参诊察患者的健康生理状况,通过“望、闻、问、切”四个角度了解患者信息。其中,闻就是指听患者的说话、咳嗽、喘息,并嗅其口中或身上是否有异味。欧洲中世纪也有通过气味来诊断疾病的记录。这些传统方法其实是有科学依据的:细胞新陈代谢的产物通过血液运输到肺部,其中挥发性成分会通过呼吸排出。而其中某些成分的浓度与特定疾病相关,因此通过呼气分析人的身体状况是可行的。现代医学研究已经证明,包括糖尿病、慢性肾病、肺癌在内的多种疾病都会造成呼气中的成分异常。相比血检、医学影像等检查手段,呼气检查的主要优点在于无创、样本易采集。在人们越来越关心自己身体状况的今天,这样一种辅助诊断手段是非常有前景的。我们希望患者能够像伸出胳膊测血压那样简单地对一台仪器吹一口气,片刻之后就能得到一些疾病的风险指标,从而决定是否到医院进行进一步详细的检查。

要实现这一点,我们首先要解决的问题是:如何方便快速地测量呼气中的标志物?当代传感器技术的发展为这个问题提供了很好的解答。气相色谱、质谱分析等传统手段,虽然精确,但较为繁琐,成本也较高。而近年来兴起的气敏传感器则给出了另一种思路:通过敏感材料与气体分子的反应,将气体浓度信息转化为电信号。这种方法简单、快速、成本也较低。人们把利用气敏传感器测量气体浓度的仪器形象地称为电子鼻。事实上,电子鼻已经被广泛地应用于检测空气污染物、工厂中的易燃易爆气体等,当然最普及的应用还是酒驾的检测。不过检测酒驾和检测疾病的难度不在一个数量级上:疾病标志物的浓度要远低于饮酒后呼气中酒精的浓度,前者的浓度只有百万分之一左右。因此我们一方面需要更精确的传感器,另一方面需要用更好的算法对传感器信号进行分析。

信号分析算法是本书的主题,为了研制一套准确、实用的呼气分析系统,我们从仪器的硬件制作开始,“软硬兼施”,进行了全面的研究。现有商业传感器已经能够达到百万分之一级别的气体检测,因此我们经过反复试

验,选择了最适合呼气检测的传感器阵列,设计了一台便携、易操作的呼气分析仪。没有数据就无法进行信号分析,因此我们与广东省中医院等单位合作,采集了近万例患者的呼气样本,其中辛苦自不待言,在此我们也感谢所有付出了辛劳的患者、医护朋友和实验室师生。

信号分析算法是呼气分析系统的重要一环,我们针对气敏传感器信号和呼气分析的特点,对多种模式识别、机器学习算法进行了改进和创新。首先,针对呼气成分存在的个体差异问题,我们对预测模型的建立方式进行了改进,提出了一种特征增强策略。其次,为了建立更准确的预测模型,我们从气敏传感器信号上提取了上千维特征,改进了现有特征选择算法并得到了最优的特征组合。最后,我们重点关注了其他呼气分析研究者较少涉及的一个主题,也就是电子鼻的鲁棒性问题。气敏传感器存在设备间差异和时变漂移问题,我们针对不同应用场景设计了四种漂移补偿算法,应用多任务学习、深度学习、子空间学习等前沿机器学习方法进行漂移补偿,大大提升了电子鼻的实用性。

综上所述,本书是一个多学科交叉的产物,结合了医学、传感器、电子和计算机等领域的研究成果;同时兼顾了理论和实践,既设计了机器学习算法,也制作了呼气分析系统。在这个富有挑战性的课题上,本书的作者做出了重要的贡献,将呼气分析系统的准确性和实用性向前推进了一步。与人体相关的信号都是复杂的,未来我们需要挖掘已有数据中的信息,进一步深入分析呼气标志物与不同疾病的联系,从而得到更全面的算法模型。我们的最终目标是造福社会和患者,因此如何让呼气分析系统从实验室中走出、尽快走向实用,是未来研究者们需要努力的方向。

张大鹏

清华大学电子工程系

2018年6月

摘 要

电子鼻是一种利用气敏传感器阵列识别简单或复杂气味的设备,应用场景丰富,因此受到越来越多的重视。本书在搭建电子鼻系统的基础上,针对电子鼻数据的特点,研究信号分析过程中的关键算法,从而提高电子鼻的准确性和实用性。本书主要从三个方面进行研究:基于电子鼻的呼气分析系统设计,针对相关特征的迭代特征删除算法,以及设备差异和时变漂移的补偿算法。

呼气分析是电子鼻的应用之一。通过检测呼出气体中与特定疾病相关的标志物,电子鼻可以辅助疾病筛查和监控。该方法具有无创、易操作等优点。我们针对糖尿病的呼气标志物筛选了气敏传感器阵列,设计并优化了呼气分析系统的硬件和软件,之后采集了包含 500 余个健康人和糖尿病患者呼气样本的数据集,通过实验证明了该系统在糖尿病筛查和血糖预测问题上的有效性。

从电子鼻的信号曲线上可以提取多种特征,对候选特征进行排序和选择有利于提高预测准确率。本书发现,当部分候选特征高度相关时,经典的支持向量机迭代特征删除算法存在相关偏差问题,因此提出了一种改进策略。通过分析特征的排序结果,我们得到了不同气敏传感器和特征提取算法的有效性排名。

漂移是电子鼻实用中的难点问题,也是本书的研究重点。具体来说,同一型号的传感器敏感特性并不完全一致;同时,由于老化等原因,传感器的响应会随着时间变化。这些因素造成了电子鼻的设备差异和时变漂移。本书利用迁移学习的思想解决这两个问题,提出了四种漂移补偿算法。结合标准化误差的模型改进算法考察不同特征对漂移的鲁棒性,然后在已有预测模型中对不稳定的特征施加更高的惩罚权重。基于标定样本的多任务学习算法对每一个数据领域(即设备或时间段)学习一个模型,并针对两种漂移的不同先验分别采用并行和串行两种迁移策略;为了更精确地补偿时变漂移,我们还提出了一种动态模型策略。漂移补偿自编码器算法采用了深

度学习框架,在学习更有鉴别能力的特征的同时,显式地建模和补偿特征中的设备差异和复杂时变漂移。上述算法都使用了标定样本,而最大独立领域适配算法能够在标定样本无法获取时利用漂移后的无标签样本进行补偿。该算法通过最大化子空间中样本与背景特征的独立性来学习领域不变子空间。这些算法拓展了多种迁移学习算法框架的适用范围。实验证明,它们能有效地提高对漂移后样本的预测精度。

关键词: 电子鼻; 特征选择; 漂移补偿; 迁移学习; 呼气分析