云端融合的神经系统疾病多通道辅助诊断研究

李洋¹, 黄进¹, 田丰^{1,2*}, 韩冬奇¹, 范俊君¹, 陈辉¹, 彭晓兰¹, 戴国忠¹,
王宏安^{1,2}

1. 中国科学院软件研究所人机交互北京市重点实验室, 北京 100190

2. 中国科学院软件研究所计算机科学国家重点实验室, 北京 100190

* 通信作者. E-mail: tianfeng@iscas.ac.cn

收稿日期: 2017-04-18; 接受日期: 2017-07-27; 网络出版日期: 2017-09-05

国家重点研发计划 (批准号: 2016YFB1001405)、国家自然科学基金 (批准号: 61232013, 61422212) 和中国科学院前沿科学重点研究计划 (批准号: QYZDY-SSW-JSC041) 资助项目

摘要 对于神经系统疾病, 其主要的生理特性可以反映在患者日常的人机交互行为中. 对交互过程中产生的相关生理信息进行获取和分析, 有助于神经系统疾病辅助诊断与早期预警. 传统的神经系统疾病自动检测系统常常只分析单一交互通道中包含的信息, 会造成其他通道中包含的重要信息的缺失. 为此, 本文提出了一个云端融合的神经系统疾病多通道辅助诊断系统. 首先, 提出一种基于多通道信息的疾病自动诊断方法, 获取并分析了用户在多个交互通道中释放的生理信息, 并将结果进行融合, 提升自动诊断的精确性与鲁棒性; 其次, 提出了一个云端融合的框架, 将从不同地区、不同时刻提取的用户生理信息在云端存储, 降低了数据采集的地域限制与时间限制; 再次, 基于云计算强大的运算能力, 系统亦可以实时地对用户在自然交互过程中产生的多通道生理信息进行分析, 做出实时、精确的自动诊断; 最后, 通过语音与笔的交互混合诊断系统实例, 验证基于云端融合的神经系统疾病多通道辅助诊断方法的有效性.

关键词 云端融合, 多通道, 人机交互, 神经系统疾病, 辅助诊断

1 引言

1.1 神经系统疾病

神经系统疾病为发生在中枢神经系统、周围神经系统、植物神经系统的由细菌感染或遗传等多种病因引起的疾病, 其症状通常表现为运动障碍及记忆认知障碍等多方面问题, 例如运动能力下降、语言能力下降、记忆力减弱以及记忆缺失等. 对于不同的神经系统疾病, 其运动功能障碍表现会略有不同, 主要包括震颤、运动迟缓、步态异常、共济失调等. 即使表现均为震颤, 不同类型的神经系统疾病

引用格式: 李洋, 黄进, 田丰, 等. 云端融合的神经系统疾病多通道辅助诊断研究. 中国科学: 信息科学, 2017, 47: 1164–1182, doi: 10.1360/N112017-00075

Li Y, Huang J, Tian F, et al. Multimodal aided neurological disease diagnosis with synergy of cloud and client (in Chinese). Sci Sin Inform, 2017, 47: 1164–1182, doi: 10.1360/N112017-00075

对震颤的触发条件也并不相同,如在病人不动的时候产生静止性震颤,病人摆出一个特定姿势时触发的姿势性震颤和病人运动的时候触发的运动性震颤等.虽然表现方式复杂,但这些运动功能障碍都具有一定的特点.

常见的神经系统疾病如 Parkinson 症^[1],是一种常见的、缓慢发生的由多巴胺能神经元明显减少,引起基底神经节功能失调的老年性神经系统疾病,临床表现为静止性震颤、肌强直和平衡障碍等运动障碍,以及声带损伤等; Alzheimer 病^[2]是一种以认知功能障碍和记忆丧失为主要特征的神经系统疾病,临床表现为记忆力减退、语言和定向力障碍等;脑小血管病^[3]是指脑内小血管病变导致的疾病,临床表现主要为认知、情感和运动功能障碍及其他临床综合征.

由于环境的变化以及人口的老齡化,罹患神经系统疾病的患者数量越来越多^[4~6],这些疾病在很大程度上影响着患者及其家庭的生活质量.由于此类疾病发病机制复杂,检查步骤繁琐,医疗费用昂贵,因此尚未获得普及的、有效的治疗方案.随着计算机科学技术的不断发展以及人类对于生理信息与疾病之间关系认知的逐渐深入,使得通过计算机获取并分析生理信息,检测神经系统功能疾病变得可行,且具有很大的研究价值与发展空间.与人工诊断相比,通过分析生理信息的计算机辅助诊断系统可以降低主观因素对判决结果的影响,有利于疾病的诊断以及早期预警,对于提升神经系统疾病患者的存活率有着重要的意义^[7,8].

1.2 疾病诊断中的人机交互技术

人机交互是“以人为中心”的自然交互准则,用户可使用语音、笔、步态等多种自然的方式与计算机进行通信.在传统的人机交互过程中,计算机接收用户在语音通道、视觉通道、笔通道等某个单通道释放出的信号,并加以分析,进而了解用户当前的状态与需求,进行相应的响应.在日常交流时,人们会同时接受并理解语音、视觉、手势姿态等来自多个通道的输入信息,也会同时通过多个通道进行信息的反馈表达.因此,基于单通道的人机交互技术往往无法兼顾其他通道中包含的有效信息,而多通道人机交互技术的出现弥补了上述不足.作为计算机视觉、心理学、人工智能等领域的交叉学科,多通道人机交互技术同时对用户在多个交互通道中产生的有效信息进行获取、分析和融合,并做出更精确决策和更自然的响应,使用户在交互过程中感觉更加真实,产生更加自然的交互行为^[9,10].随着计算机技术的不断发展以及云计算、云存储能力的不断提升,多通道人机交互技术受到的计算及存储方面的限制越来越小,也受到了越来越多的关注.

近年来,基于人机交互技术的疾病辅助检测与评价方法受到了广泛的关注,并发展成为人机交互领域较为重要的一个分支.言语病理学家已经证实某些病人的某些病症与发音之间存在一定的联系,并积极尝试在一个可控的临床条件下搭建辅助诊断治疗的声音分析系统^[11,12].常见的声音分析系统普遍用持续元音发音(sustained phonations)^[13]或连续语句发音(running speech)^[14]进行声音质量评价.分析连续性语句发音的病理情况受到准确发音、构词连接等因素影响,具有一定复杂性,不便于在实际系统中使用.大量学者和临床工作表明,通过对持续性元音发音(例如“ah...”)进行评估也能满足疾病诊断的需求^[15~18],例如 Majidnezhad 等^[17]通过提取 Mel 频谱倒谱系数(MFCC)和小波包分解(wavelet packet decomposition)作为音频特征检测用户是否患有疾病得到较好的判别结果;牛津大学的 Tsanas 等^[18]针对病理特征提取幅值扰动、信噪比等多种音频特征进行模型训练也取得了较好的诊断结果.

笔交互作为一项成熟的交互技术,具有交互效率高、认知负荷低等特点.运笔状态和图形绘制结果可有效反应神经系统疾病患者在运动以及认知功能上异常的信息.因此通常从笔交互任务中提取运笔特征和绘图特征进行神经系统疾病诊断.如 Ünlü 等^[19]从数字笔的笔迹、压力和倾角上提取多

项特征, 并将其组合进行模型训练进行 Parkinson 疾病诊断取得了很好的效果. Smits 等^[20] 利用数字笔提取的信息研究 Parkinson 疾病, 找到了用于定量评估运动徐缓、字体过小和手部震颤症状的特征. 绘图理解作为检测神经系统疾病的重要手段^[21], Davis 等^[22] 利用算法对画钟测试 (clock drawing test, CDT) 进行理解, 得到钟面、数字等组成部分和几何参数用于神经系统疾病诊断, 连线测试 (trail making test, TMT) 进一步将完成时间、错误次数、连线的笔直程度这些特征与时序相关的运笔特征结合^[23] 评估了其对于不同神经系统疾病的诊断效果.

步态作为人类日常运动中表现的行为特征, 可有效反映人类的移动能力、体能和健康状况^[24]. 步态分析通过对人类行走过程中身体各部分, 特别是下肢的运动形态与受力情况进行量化分析可有效的反映神经系统疾病的运动障碍. 现阶段, 步态分析检测参数主要由空间参数和时间参数组成, 其中空间参数包括步长、步宽及步频等^[25]; 而时间参数以步态周期性为代表^[26,27]. 如 Lan 等^[28] 提出在智能手机中安装步行者航位推算系统, 监测并记录步长、步进频率等步态特征, 对特征集进行分类训练判定用户是否患有 Parkinson 疾病, 取得了准确度较高的判别结果; Chen 等^[29] 提出用一种由粗到细的特征选择方法, 选择有效步态特征集, 可以较为准确地判别患者是否患有脑小血管疾病.

生理计算 (physiological computing)^[30] 作为人体生理与计算机的直接接口, 将解析人类的生理信号转换为计算机的实时输入从而改善用户的交互体验, 近几年在神经系统疾病检测和预警中取得了很好的效果. 针对神经系统疾病的特点, 通常从时域和频域两个角度对用户产生的生理信号数据进行特征提取. 其中时域特征包括生理信号的最大最小值、平均值、标准差等, 主要表征用户对肢体的控制程度; 频域特征包括低频区域的能量值、频率谱波峰能量和频率特征, 主要表征用户上肢的震颤情况和动作的能量集中频段^[31]. 如 Patel 等^[32] 证实了使用加速度计的数据来估计 Parkinson 病患者的症状和运动并发症的严重程度的可行性. 而 Robichaud 等^[33] 的研究表明, 肌肉电信号的分析方法可以作为一种快速诊断 Parkinson 疾病的有效方法. Jordan^[34] 在其工作中使用脑电信号的频谱和分布来对脑中风进行评价与监控.

综上, 神经系统疾病的症状通常体现在运动能力下降、语言能力下降、记忆力减弱以及记忆缺失等多个方面. 因此通过单通道人机交互技术进行分析与诊断, 会忽略其他交互通道产生的生理信号中包含的有效信息, 导致判别结果精度不足. 随着社会人口平均寿命的提升, 多通道人机交互技术的发展, 将患者在听觉、触觉、视觉等多个通道的交互信息应用到病理识别中, 可获取更精确的诊断结果. 伴随而来的, 是越来越多的需要处理和计算的生理信息与单机服务无法承受的计算负载, 使用大规模服务器集群做后台处理成为智能交互平台的主要选择, 因此如何存储海量交互信息并进行快速分析与诊断变为急需解决的问题. 云计算技术^[35] 与云存储技术^[36] 的发展, 存储海量病理数据以及快速分析变得可能. 首先, 将分散在不同地区、不同规格的医疗机构中获取的相关病理信息统一存储在云端, 真正实现数据的共享. 同时借助于云计算技术强大的计算能力, 对来自多个交互通道的大量交互信息进行实时处理, 做出精准的决策. 基于上述概念, 本文提出一种云端融合的神经系统疾病多通道辅助诊断系统, 系统可以使用户在任何时间、任何地点, 使用任何设备将终端采集的多通道的人机交互生理数据传入云端, 同时实时反馈疾病的判定结果, 具有可扩展、易维护、安全性高、实时性、精度高等特点, 有效解决传统医疗诊断过程的独立、分散、效率低下以及诊断结果不精确的问题^[37,38].

本文首先分析了神经系统疾病的产生原因与表现, 并分别介绍了人机交互领域语音、笔、步态等多个通道神经系统疾病检测及分析方法. 根据现有计算机辅助诊断治疗方法存在的问题及云存储与云计算存在的优点, 提出云端融合多通道疾病辅助诊断的概念. 其次, 对云端融合疾病诊断的基本框架与各个模块, 多通道融合的疾病诊断方法进行介绍. 再次, 通过笔和语音通道的系统实例, 验证了云端融合疾病诊断方法能够满足神经系统疾病辅助诊断的需求. 最后, 对基于云端融合多通道人机交互疾

图 1 云端融合的神经系统疾病多通道辅助诊断框架

Figure 1 Cloud-Client synergy based multi-modal aided neurological disease diagnosis framework

病诊断的重要性和可行性进行了总结与描述, 并对未来可进一步发展的方向进行了展望.

2 云端融合的疾病诊断框架

2.1 模块划分

云端融合的疾病诊断框架如图 1 所示, 可以划分为 3 部分, 即终端部分、云端部分和融合部分.

终端部分. 包含数据获取与诊断反馈两部分. 数据获取部分对用户交互行为的词法及语法解析, 通过预处理生成用于疾病诊断的交互行为数据并进行本地存储; 诊断反馈部分提供通用交互界面, 供用户对诊断过程和结果进行管理.

云端部分. 由多通道特征融合、诊断模型训练以及综合诊断决策 3 部分组成. 多通道特征融合负责将上传的各通道交互行为数据分别进行特征提取, 并依据情况进行统一地特征选择; 诊断模型是一组能够适配单个通道或多个组合通道特征的机器学习模型, 它们由云端收集的病例数据训练得来; 综合诊断决策部分实现结合模型诊断、评估量表、医生经验的综合评定, 并向医生提供安排诊断交互任务的服务. 为了支撑上述服务, 云端部署了数据存储单元, 负责病例样本、诊断模型、医学知识和病人历史信息的存储.

融合部分. 终端提供了诊断数据和在线诊断界面, 云端提供了自动诊断和数据聚合服务, 融合部分则通过一系列应用协议实现两者之间的桥接和互补. 交互行为提交协议规定了由病人主动发起的诊断任务数据的提交规则; 病例收集协议规定了云端向终端获取新样本的方法, 以确保为模型取得相匹配且标记过的训练数据; 病人的诊断任务可由自己主动发起, 也可由医生或系统指派, 指派协议则规定了定期或临时的诊断任务定制规则. 在线诊断协议对医生介入和管理整个诊断过程提供了接口.

2.2 框架分析

基于本框架构建的辅助诊断系统所面向的用户主要包含病人和医生两个角色, 病人执行交互任务, 所得数据由终端负责采集并上传至云端进行分析, 最终由终端将诊断结果反馈给病人。此外, 医生通过终端参与疾病的整个诊断过程, 包括制定诊断交互任务、查阅单项检查报告、综合多项检查做出诊断、发布诊断结果等。每个病人或者医生都有可能拥有多个终端, 这些终端可能是个人电脑、移动设备或者服务器, 病人可以在家中、医院或者随时随地地进行疾病监测和诊断, 医生也能按需地参与整个诊断过程。因此基于本框架构建的辅助诊断系统, 能够充分利用不同时间和空间分布上的诊断信息, 对海量用户数据进行统一存储与资源共享, 并对数据进行快速分析, 提高诊断的时效性。

基于本框架构建的辅助诊断系统是可以与多种不同类型的交互设备相连的, 这些设备可以提供多通道的交互数据采集, 例如麦克风提供语音交互数据, 数位笔提供笔交互数据等。为了利用这些来自于不同设备、不同通道的交互信息, 终端通过原语解析、任务分析和预处理步骤, 生成了与设备无关的、用于疾病诊断的行为数据, 并且对这些数据的生成设备、通道类型进行标注, 云端根据数据的设备、通道类型匹配对应的特征提取操作, 最终使这些来自于不同设备、不同通道的交互信息都转化成病理特征。得到病理特征后, 云端会根据诊断模型的特征选择列表 (通过训练得到), 对多通道特征进行统一的特征筛选, 从而实现多通道交互数据的融合和利用。

本框架下的云端诊断模型是一组能够适配单个通道或多个组合通道特征的机器学习模型, 当用户在终端完成一个或多个诊断交互任务并发送一个诊断请求时, 发送的交互行为数据可能是单个通道或者不同通道组合而成的, 云端为完成这个请求需要匹配不同的诊断模型, 因此这里云端提供一个模型组实现不同条件下的自动诊断。此外, 这个模型组的训练数据来源于多个终端和多种采集通道, 通过样本采集协议, 使得云端不断吸收终端数据, 将其拆分组合, 实现各模型样本库的扩充, 最后通过定期重新训练, 保持诊断模型组的持续迭代和优化。

在诊断决策方面, 本框架通过医生管理界面、混合决策机制和在线诊断协议, 实现结合模型诊断、评估量表和医生专业知识的综合诊断决策。此外, 通过本框架, 病人在不同地点、由不同医生甚至是从不同医院得出的单项检查或诊断报告都保存在了病人的历史信息当中, 医生能够通过查阅这些信息更加全面地了解病人的身体状况。

3 多通道融合的疾病诊断方法

3.1 模型概述

基于神经系统疾病的特性, 本文提出了一个多通道融合的疾病诊断方法, 如图 2 所示。方法主要由原语分析器、任务分析器、数据处理器、融合管理器、病理特征选择模块、病理模型、混合决策与多通道交互信息分流优化模块构成。

首先, 用户利用不同通道的输入设备采集的原始信息传输到计算机终端, 通过原语分析器得到对应通道的交互原语, 并利用任务分析器生成相应的场景任务。

其次, 对任务数据进行处理得到适于训练病理模型的多通道交互特征参数, 本部分主要涉及的模块包括数据处理器的特征提取、融合管理器的多通道特征信息融合和病理特征选择模块。实际数据采集时的交互过程会受到交互环境、交互上下文等多种因素影响, 因此在特征提取前, 需要对数据进行预处理, 去除输入设备、交互通道噪声或其他因素对交互产生的影响, 加强有用的交互信息。

图 2 多通道人机交互神经系统疾病诊断流程

Figure 2 Neurological disease diagnosis based on multi-modal interaction

最后, 将特征选择后的多通道融合特征信息进行模型训练得到病理模型. 将病理模型得到的计算机辅助判别结果与用户数据库、医学知识库、医疗人员等相关意见进行混合决策, 通过多通道交互信息分流优化将决策结果通过不同通道对用户进行反馈. 本节将对上述模块及其功能进行详细介绍.

3.2 交互原语和任务

交互原语 (interaction primitive, IP) 是用户通过交互设备到计算机的一个词法输入, 表示一个独立的、最小的、不可分割的操作. 交互任务则是通过原语组成, 根据不同的场景需求分析得到的任务描述集合. 文献 [39] 将交互原语表示为一个四元组 $IP\langle Type, Data, Time, Device \rangle$, 这里 $Type$ 表示交互原语, $Data$ 表示交互数据, $Time$ 表示发生时间, $Device$ 表示交互通道, 通过不同的原语组合, 即可得到相应的场景交互任务.

原语分析器通过对每个通道的交互信息进行词法分析, 从而得到意义相同而形式不同的交互原语; 同时, 原语分析器与交互环境进行实时反馈, 为下一次的词法分析增加上下文信息. 任务分析器根据不同的评价任务需求, 将原语分析器输出的交互原语按照一定的语法规则进行融合与识别. 这里的融合为多通道信息语义的分析与融合, 旨在根据不同通道交互原语的特点和交互环境与上下文对用户操作的影响, 输出带有明确、完整语义的交互任务.

针对神经系统疾病的病理特征与评价方法, 文献 [40] 定义了 9 类交互原语与 6 类交互任务. 以听觉通道为例, 交互信息通过麦克风采集传入系统后, 通过原语分析器得到与设备无关、与通道相关的声音原语幅值大小, 任务分析器根据评价任务的不同, 得到听觉通道中不同交互内容的音频. 目前, 业界普遍通过对患者持续元音发音分析判定是否患有语音疾病, 如文献 [18] 中, 就是通过检测用户元音发音 /a/ 判定其是否患有 Parkinson 疾病.

3.3 多通道交互特征

数据处理器针对神经系统疾病的病理特性提取能表述其时间与空间特性的线性或非线性特征参

数, 此时的特征参数是从单一交互通道提取的, 经过变换能够反映此通道的特性并体现用户在此通道的病理反应, 常用空间向量表示。例如, 神经系统疾病在声音通道常表现为声波非周期性变化, 发音夹杂着大量噪声以及发音细节的不清楚等, 因此可通过提取音频的幅值、信噪比、频域特征等描述相应的状态。

融合管理器是将听觉、视觉、触觉、体感等只能描述用户某一种交互行为的单一通道特征信息进行融合, 得到多元交互特征信息, 以反映在复杂交互环境中用户行为的动态变化。去除空间上各个输入通道测量数据与时间上交互轮次起止点模糊和多通道语义信息的不同步导致的“非精确性”, 得到鲁棒的混合特征, 是后续认知计算和反馈呈现的前提。针对神经系统疾病的病理特点, 将提取的笔、语音、步态、生理等通道特征信息进行多通道信息融合, 提高感知的鲁棒性。如用户受时间或地域限制只采集单通道数据信息, 则无需特征融合。

病理特征选择是去除融合特征集合的冗余信息, 得到能够反映用户病理特征的最优子集。随着多通道交互的发展, 获取的多通道融合信息特征维度极有可能大于训练的样本量, 过高维度的特征会造成分类器计算性能的降低, 同时也会因为包含较多的冗余信息或者噪声导致识别精确度的不足。因此, 可通过有监督或无监督的特征选择方法将维数较高的特征变为维数较低的特征, 降低类内方差, 增大类间方差, 得到与用户行为相关度较大的特征信息^[41,42]。特征选择的方法通常可分为过滤器 (filter) 和包装器 (wrapper) 的方法, 前者通常寻找单维表现较好的特征进行特征组合, 根据某种统计方法计算单维特征与响应结果相关度, 根据分值进行排序选取分值较高的特征维度。此种方法注重对单维特征属性的评价, 虽然效率较高, 但是精度有限, 忽略特征之间的相关性, 当数据存在较大的噪声时, 通常得到的并非最优组合。后者以高分类精度为目标, 通过搜索若干种特征组合方案, 对组合结果进行评估, 侧重对特征子集进行评价。特征选择方法与具体的问题具有十分密切的联系, 因此需要选择与目标相关的特征选择方法, 得到与后续的分类结果更加相关的特征^[43], 本文选择近年来分类领域常用的 wrapper 特征选择方法。

3.4 病理模型判别机制与交互反馈

病理模型是将多通道交互特征集合作为输入样本, 训练得到引起损失或是错误识别率最小时收敛的分类器。分类器的选取对训练过程及最终结果的判定具有很大的影响, 有的分类器对于某一通道的相关特征较为敏感, 而有的分类器则对另外某些通道的相关特征比较敏感。此外, 分类器的泛化能力和拟合能力是此消彼长的关系, 一个过拟合的分类器可以在训练集上得到较好的拟合训练结果, 但是在测试集上的测试结果却可能较差; 而一个泛化能力强的分类器可以在测试集得到较好的效果, 其拟合能力也会相对较弱。因此分类器需要在泛化能力与拟合能力之间达到平衡。同时训练数据量与分类器的性能存在较强的关系, 一个好的分类器或分类系统应该能够平衡训练数据的大小, 调整拟合能力与泛化能力之间的偏重。集成学习 (ensemble learning) 作为一种强大的机器学习方法, 对于较庞大及复杂的学习任务, 通过对弱分类器的结合, 可避免弱分类器学习能力的缺陷与精度的不足, 得到更加真实的分类模型, 具有比单分类器更精确的分类潜力^[44], 实验表明, 集成学习的方法对冠心病^[45]、白内障^[46]、Parkinson^[47]等疾病的诊断具有很好的效果。

为了更好地适应不同的用户群体与用户在不同交互通道中的疾病表现, 本文选择集成学习的方法训练神经系统病理模型, 构建包含单独通道和多通道融合的病理模型库。这种基于集成学习的多层分类模型可以考虑到不同分类器对不同通道中包含的交互信息的敏感程度与识别精度的不同, 将多个分类器的决策结果进行有机的融合, 从而得到更精确的, 能够反映用户多通道交互生理信息状态的自动诊断结果。

图 3 基于笔和语音的多通道人机交互神经系统疾病诊断实例流程

Figure 3 Case study of neurological disease diagnosis based on pen and audio interaction

综合考虑计算机自动判别的结果, 用户数据库中存储的用户个人信息, 医学数据库中病情相关的评价参数的规范与标准, 以及医学人员的专业诊断结果, 建立多属性人机交互混合决策模型, 得到用户当前状态和日常生活习惯等多种因素对判别结果的影响. 系统在计算得到最终的评价结果后, 凭借多通道交互信息优化管理器对多通道融合人机交互的判别结果对用户进行反馈.

4 实例研究: 语音和笔交互混合诊断系统

4.1 系统框架

根据云端融合诊断框架和多通道融合的疾病诊断方法, 本文通过对语音通道和笔通道交互结果的判定作为混合诊断系统的实例, 说明如何利用多通道人机交互进行疾病诊断. 诊断系统的数据处理流程如图 3 所示.

首先, 通过麦克风与触摸屏采集用户的声音信号与笔信号, 将采集的输入信号经过原语分析器统一处理, 得到音频幅值以及笔的手势交互原语. 这里, 任务分析器通过对交互原语进行词法分析, 生成用户执行“音频 /ah:/”与“平面绘图”任务的结果.

其次, 数据处理器对生成的交互任务信息进行如降噪、剔除无效信息等预处理. 将生成的交互行为数据传输到云端并针对神经系统疾病提取与病理相关的特征参数. 在声音交互任务中, 从声音夹杂噪声、发音器官运动状态等多方面提取声音通道特征; 在笔交互任务中, 从运笔状态和图形绘制结果提取笔通道特征.

再次, 将声音通道与笔通道提取的特征融合, 得到多通道融合特征. 为了避免信息冗余, 选择基于相关性的特征选择方法 (correlation-based feature selection, CFS). 对语音特征、笔特征、语音与笔的融合特征分别进行特征选择, 得到与分类结果更相关, 维数更低的多通道交互特征.

最后, 利用交叉验证的方法对语音通道、笔通道、语音与笔多通道交互特征进行模型训练与测试, 得到语音通道诊断模型、笔通道诊断模型、语音和笔通道融合诊断模型. 为了避免单一分类器对某些

交互信息的特征较为敏感, 过度拟合影响判定结果, 提出基于集成学习的病理模型训练方法, 使用包含时空交互信息的语音与笔融合特征对“基”分类器进行模型训练, 并将对应的预测结果作为特征输入到第二层模型, 得到语音和笔通道融合的集成学习诊断模型. 综合考虑用户数据与医学知识得到诊断结果, 将诊断结果从云端返回到终端, 并通过信息的分流优化以麦克风与显示屏结合的方法对用户进行实时诊断反馈.

4.2 实验数据采集

本文主要通过笔通道和语音通道进行多通道神经系统疾病诊断, 用户数据采集地点为北京市顺义区各乡镇以及北京协和医院, 共 176 例用户数据. 其中 88 人次在北京协和医院经过系列严格筛查, 临床诊断患有神经系统疾病, 包括 14 例 Parkinson、62 例脑小血管病变、12 例运动障碍 (平均年龄 62.8 岁, 男性 64 人), 其中运动障碍为神经系统异常引起的运动障碍, 但无法判定具体的疾病类型; 88 人次健康用户 (平均年龄 62.7 岁, 男性 64 人), 为保证训练数据的均衡, 健康用户组从招募的 729 个普通测试人群中选出与每一个病人样本性别、年龄相匹配的样本. 样本挑选步骤如下: 对于每一个疾病用户样本, (1) 从健康测试人群中挑选性别匹配的用户数据集合; (2) 从步骤 (1) 的数据集合中选择与病人年龄 N 相同的用户数据; (3) 如果步骤 (2) 没有与年龄 N 完全匹配的用户数据, 逐步扩展年龄搜索范围到 $N \pm i$ ($i = 1, 2, 3, \dots$) 直到获取至少包含一个用户数据; (4) 步骤 (3) 得到的子集里, 找年龄最为接近的正常人作为对照样本, 如果年龄相同则随机选择一个. 这些测试作为北京市协和医院门诊检查的医疗测试程序之一, 所有的数据标签都经过医院认证.

数据采集的实验系统由一个麦克风、一个显示器、一个笔输入平板显示器和一台主机组成, 主机内嵌自主研发的客户端软件程序, 完成采集后由实验人员将数据发往云端服务器进行处理和分析, 录制数据时需要先对用户进行简单的培训, 具体如下:

语音数据. 用户使用 Audio Technica AT9944 型号高质量麦克风录制元音 /ah:/ 音频, 录制时将麦克风放置在桌面上, 麦克风与用户嘴唇距离大约 5 cm, 声音信号的采样率为 44.1 kHz, 16 bit^[15~18]. 当用户听到滴的一声后开始发音, 发音时尽量长时间的保持一个平稳的发音状态, 液晶显示器实时显示录制音频情况, 当音量低于一定阈值或发音时长大于 30 s 则自动停止录制, 系统将采集数据自动暂存到客户端.

笔数据. 用户使用 Wacom Cintiq DTK-1300 型号笔输入平板显示器和一支长 14.4 cm, 直径 1 cm, 重 19 g, 采样率为 100 Hz 的数字笔进行一个“连线测试”的交互任务, 并记录笔交互信息. 连线测试是一种应用广泛的神经心理学测验, 其要求受试者将无序分散的数字 1~25 按从小到大的顺序依次连接, 如图 4 所示, 用户在进行正式测试前, 将在测试人员的指导下首先完成一个与正式测试类似的简化测试 (连接数字 1~8)^[48,49]. 测试过程中的所有笔迹都将自动暂存到客户端.

4.3 语音和笔式病理特征

本文针对神经系统疾病在语音交互通道的病理表征, 提出由周期振动类特征、噪声类特征、发音器官运动情况类特征以及高阶统计量类特征构成 433 维音频特征集合, 包括反映声襞振动产生声波周期性变化的特征描述, 例如幅值扰动和频率扰动 (jitter 和 shimmer)^[50], Recurrence Period Density Entropy (RPDE) 和 Pitch Period Entropy (PPE)^[51] 等为代表的基频相关特征 231 维; 发声者声带闭合不完全导致声音夹杂着大量噪声的相关特征描述, 例如 Harmonic to Noise Ratio (HNR)^[52], Detrended Fluctuation Analysis (DFA)^[53] 和 Glottal to Noise Excitation (GNE)^[54] 等噪声相关特征 24 维; 发

图 4 (网络版彩图) (a) 语音通道数据采集; (b) 基于笔的连线测试
Figure 4 (Color online) (a) Voice data collection; (b) pen based trail making test

表 1 语音通道提取特征及其物理含义

Table 1 Extraction characteristics of voice data and their meanings

特征分类	特征名称	特征维度	特征与病理特性的关系
声襞振动周期	Jitter	22	声带振动声波相邻周期时间的变化
	Shimmer	22	声带振动声波相邻周期幅值的变化
	GQ	3	用熵的概念评估声带振动周期内声门张开时间和闭合时间
	RPDE	1	用熵的概念评估序列时间内声带振动频率的变化
	PPE	1	声带持续振动内声波基频变化的程度
	wavelet	182	声带振动产生基频的时域变化
声音夹杂噪声	HNR	2	声带振动或器官病变等因素得到的语音谐波信号和噪声的比例
	NHR	2	声带振动或器官病变等因素得到的语音噪声和谐波信号的比例
	GNE	6	整段语音信号夹杂噪声的程度
	DFA	1	一段时间内声波信号的包络变化趋势, 噪声的随机自相似性
	EMD_ER	6	分解各频段量化能量噪声的比重
	VFER	7	用能量、熵等衡量语音信号夹杂噪声的程度,
发音器官运动	MFCC	84	声波的整体衡量, 频域表示, 短时能量谱, 波形的细微扰动
高阶统计量	Skewness	47	声波病理状态与对称中心的偏离程度
	Kurtosis	47	声波病理状态与峰值的偏移量

声者无法有效操控发音器官对发音产生的影响的特征描述, 如常用于语音识别和生物医学领域相关的 MFCC 特征^[55] 82 维. 达到增强信息的高阶统计量特征描述, 本文为针对不同病理信息的 jitter, shimmer, HNR, NHR, GNE, MFCC 特征计算的三阶 skewness、四阶 kurtosis 统计量得到特征 94 维, 用于多通道交互的神经系统疾病诊断. 相关特征详细信息如表 1 所示.

与之类似, 针对神经系统疾病对运动功能的影响, 从笔交互数据上提取运笔特征与绘图特征 1004 维, 运笔特征是指从笔迹采样序列的运动参数中提取的, 反映用户运笔时局部或全局运动特点的一类特征. 这类特征一般以笔画作为基本分析单元, 将笔通道的各项运动参数看作时序信号进行处

表 2 笔通道提取特征及其物理含义
Table 2 Extraction characteristics of pen gesture and their meanings

特征分类	特征名称	特征维度	特征与病理特性的关系
位移	Position	72	笔尖的绝对位置
	Speed	112	笔尖的瞬时速度
	Acceleration	112	笔尖的瞬时加速度, 反映手部运动的力量变化
	UMO	40	笔尖位置与当前笔迹意图中心的偏离程度, 反映无意识运动程度
	Curvature	76	当前笔迹在该笔尖位置上的弯曲程度, 反映笔迹的笔直程度
压力	Pressure	76	绘图时用户施与笔尖的压力
	Pressure speed	36	笔尖压力的变化速度
	Pressure accuracy	36	笔尖压力的变化加速度, 反映手部运动的纵向力量变化
笔身姿态	Posture	148	笔身的高度角、方位角和自转角
	Posture speed	148	笔身角度的变化速度
	Posture accuracy	148	笔身角度的变化加速度, 反映手部运动在三维空间中的力量变化

理, 其运动参数一般包括位置、压力、高度角、方位角、自转角以及通过对上述参数计算得到的速度、加速度、曲率等. 然后对运动参数经过多种处理方法得到运笔特征, 包括利用统计方法求得均值、方差等统计特征, 利用时域分析方法求得信号能量、穿越次数等时域特征, 利用频域分析方法求得主要频率、频率能量等频域特征, 利用近似熵计算方法求得混乱度特征^[19]. 绘图特征是指通过比较实际绘制图形与期望图形 (或参考图形) 之间的差异, 得到反映用户认知和几何规划能力的一类特征. 这类特征一般以整体图形作为分析对象, 且其特征多数是针对特定绘图任务的, 采用反映连线的笔直程度^[23]的特征. 本文将基础原语数据通过统计方法、时域与频域分析方法、近似熵计算和经验模式分解处理获得位移相关特征 412 维、压力相关特征 148 维、笔身姿态相关特征 444 维, 对神经系统疾病做出较为全面的定量评价与诊断. 相关特征详细信息如表 2 所示.

4.4 语音和笔通道的融合机制与特征选择

为排除样本的空值、异常特征值和特征取值范围对后续处理的影响, 本文对数据特征进行了清洗和归一化, 具体步骤如下:

- (1) 剔除缺失 30% 以上特征数据的样本;
- (2) 在非空数据里面, 计算每一维特征的均值和标准差;
- (3) 使用每一维的特征均值替换样本数据对应维度的空值特征;
- (4) 使用每一维特征均值替换各维在 3 倍标准差以外的特征数据;
- (5) 将针对病理特性提取的特征向量利用最大最小值将数据范围归一化到 [0, 1].

语音通道和笔通道的交互行为分别反映神经系统疾病对发音和运动功能的影响, 某些患有神经系统疾病的用户在发音或是运动功能的某种异常行为表现并不明显. 将针对神经系统疾病对交互行为的影响提取的两个通道特征进行融合, 得到比单独语音通道或是单独笔通道更加全面、有效反应神经系统疾病的特征向量 1437 维.

由于样本数量较少, 而样本特征维度会增加运算的复杂度并引起分类模型的过拟合, 因此为了增加运算速度并且不丢失最主要的特征信息, 对双通道交互数据进行特征选择. 如 3.3 小节所示, 本文选择分类准确率较高的 wrapper 的 CFS 搜索最优子集, CFS 是根据属性子集中每一个特征的预测能

图 5 (网络版彩图) Stacking 集成学习分类器图示

Figure 5 (Color online) A diagram of Stacking ensemble learning classifier

力以及特征之间的相关性进行评估, 选择与预测结果高度相关而特征子集间相关性较低的特征集合作为最终的特征选择结果. 利用与病理诊断结果最相关而特征间冗余性最低的特征向量进行病理模型训练, 可降低冗余特征或噪声对结果的影响 [56]. 基于语音通道与笔通道的数据采集系统获取用户不同交互行为的特征数据, 利用 CFS 对语音与笔融合的特征信息搜索得到与病理模型最相关的特征集合, 共 55 维特征, 其中包括 18 维语音特征和 37 维笔特征.

4.5 多通道融合诊断模型

集成学习通过选择多样性的“基”分类器, 利用某种方法组合分类模型达到比单独某一种分类器更好的性能, 得到一个对输入数据更加独立与稳定的诊断模型.

支持向量机 (support vector machine, SVM) 主要通过损失函数最小化划分超平面, 在统计学习中具有结构风险最小化特点. 即使不进行特征选择, 维度较高时也不容易发生拟合的情况, 不容易陷入局部极小值的问题, 泛化能力强. 神经网络是模拟人脑执行分析结果的过程, 其中 RBF 神经网络可以根据具体的问题确定网络拓扑结构, 是一种广泛使用并成熟的神经网络技术, 在训练样本不充足的情况下也会得到较好的测试结果. Bayes 网络是通过概率和判别损失选择最优的标记类别, 是一种概率网络. 以上 3 种分类器经常被用于声音疾病、老年痴呆、轻度认知障碍等疾病检测 [57,58], 因此根据语音与笔利用不同设备获取不同交互通道信息的情况, 系统使用上述 3 种常用的分类器作为集成学习模块的“基”分类器.

集成学习的主要组成方式有 AdaBoost, Bagging 和 Stacking 等. 其中, AdaBoost 以及 Bagging 均是基于单一算法, 并根据投票表决确定最终结果; 而 Stacking 则是基于多种算法, 在组合分类器上根据再次学习的方法而非简单的投票方式, 进而学习到更多的信息, 通常比任何单一模型的效果都要好 [44]. 本文利用 Stacking 的方法对分类器进行组合. 所谓 Stacking, 指用交叉验证的方法, 将数据集进行分割得到特征子集, 利用特征子集的组合训练多个模型得到第一层分类器, 用不同的子集对训练模型进行测试, 得到的测试结果作为输入训练第二层新的分类模型. 如图 5 所示, 本文首先用交叉验证的方法将降维后的多通道融合数据集划分为 3 等份, 对于第一层分类器中的 SVM、RBF 神经网络、Bayes 网络 3 个“基”分类器, 每个使用 2 份数据进行训练, 剩余一份数据进行测试, 并将 3 个“基”分类器

表 3 五个不同模型的分类诊断效果
Table 3 Classification effects of five different models

Model	Accuracy (%)	Precision	Recall	<i>F</i> value
DMEM (Ex)	89.22	0.905	0.879	0.885
DMEM	88.35	0.901	0.863	0.875
DMM	86.30	0.873	0.851	0.854
PM	82.54	0.855	0.788	0.810
SM	77.80	0.805	0.740	0.759

得到的预测结果作为输入训练第二层的线性分类模型, 得到分类结果作为最终的自动判别结果.

4.6 结果分析

为了说明本文提出的云端融合的多通道神经系统疾病诊断方法有效, 我们分别设计了 3 组实验, 其中实验一将本文实现的包含额外信息的双通道融合诊断模型与其他模型进行了对比, 说明这一方法能够达到更好的分类效果. 实验二针对不同年龄段与性别训练多通道融合诊断模型, 与实验一包含额外信息的双通道融合诊断模型进行对比. 实验三对云端从客户端不断收集病例样本实现模型的迭代优化过程进行了模拟, 体现云端融合疾病诊断框架中病例样本不断扩充带来的优势.

实验一. 对 5 个不同模型的分类诊断效果进行了对比, 它们包括仅使用语音特征的模型 (sound model, SM)、仅使用笔特征的模型 (pen model, PM)、语音和笔双通道融合的模型 (dual modal model, DMM)、双通道融合的集成学习模型 (dual modal ensemble model, DMEM)、包括额外信息的双通道融合的集成学习模型 (dual modal ensemble model with extra information, DMEM (Ex)). 其中前 3 种仅使用线性内核的 SVM 分类器进行疾病诊断, 第 4 种则使用基于集成学习的方法训练的病理模型进行诊断, 第 5 种与第 4 种模型训练方法相同, 但使用语音、笔并增加年龄、性别、MMSE 量表、MOCA 量表的分值得到病理模型特征. 表 3 给出了同一样本集中不同模型的平均准确率、病人组的查准率、查全率和 *F* 值, 所有指标均为 1000 次十折交叉验证结果求平均得到.

病人组的 *F* 值是疾病查全率和查准率的调和平均数, 是一个综合考虑二者的评估指标, 本文使用病人组的 *F* 值来评估不同诊断模型对疾病的诊断效果. 从结果中可以看出, 单独使用语音通道和笔通道模型得到的 *F* 值分别为 0.759 和 0.810, 低于多通道融合模型的准确率. 其原因在于两个通道对于不同病症的敏感程度和表现形式不同. 由于实验选择了 3 种不同的病症作为样本集合, 而这些病症有的重点体现在病人语音特征方面的异常, 有的则重点体现在病人笔特征方面的异常, 因此对于某个单通道来说, 其分析的特征是不全面的, 容易忽略其他通道上的重要病理信息, 导致查全率偏低; 另外, 由于样本噪声的存在, 使得单通道模型更容易受到自身通道噪声的影响, 鲁棒性较差, 也会导致准确率和查准率的降低.

语音和笔双通道融合的模型则很好地解决了这一问题. 该模型通过将语音通道和笔通道的有效信息进行融合和筛选, 保证了所有重要的病理信息都被模型考虑在内, 进而提升了诊断识别的查全率; 与此同时, 由于同时考虑两个通道的特征, 模型受到单个通道噪声的影响也会降低, 进而提升识别的精度, 实验结果表明, 基于语音和笔双通道融合的模型的诊断系统的 *F* 值可以达到 0.854.

由于病理样本存在分布不均、噪声较多等问题, 因此使用单个机器学习模型进行样本拟合和进行诊断决策可能会导致过拟合或欠拟合的情况. 而基于集成学习的病理模型则同时考虑了多个机器学习模型的诊断结果, 并将各个模型的输出作为中间结果, 统一输入给最终的分类器做最终的决策, 从一定

表 4 总体模型在不同年龄段和性别样本集上的表现
Table 4 Classification effects of DMEM (Ex) on different ages and genders

Sample set	Accuracy (%)	Precision	Recall	<i>F</i> value
Age 0 ~ 64	86.95	0.898	0.827	0.846
Age 65 ~ 74	90.40	0.846	0.852	0.834
Age above 75	94.62	0.808	0.841	0.819
Male	88.69	0.914	0.856	0.874
Female	90.73	0.833	0.875	0.840

程度上避免了过拟合情况的发生, 并且减少了样本噪声对于最终结果的影响. 实验结果也证明了本文使用的多通道融合与基于集成学习的病理模型在神经系统疾病诊断上的优势, 其实验结果的 *F* 值可以达到 0.875. 在此基础上, 考虑年龄等因素对神经系统疾病产生的影响, 在语音与笔特征基础上加入额外信息, 利用集成学习方法进行模型训练, 最终得到较好的筛查效果, *F* 值可以达到 0.885, 实现对神经系统疾病的早期预警.

实验二. 由于神经系统疾病相对不同年龄段和性别的发病几率存在差异, 不同人群间病理表现也各不相同, 因此, 在实验二中对实验一表现最好的 DMEM (Ex) 模型的通用性和学习能力进行了测试. 首先, 本文统计了 DMEM (Ex) 模型在不同年龄段以及性别的样本集上进行十折交叉验证的表现, 如表 4 所示; 然后, 本文将样本总体根据年龄段和性别划分为 5 个不同的样本集, 将 DMEM (Ex) 在这 5 个样本集上分别重新训练并进行十折交叉验证, 得到了模型对于根据年龄段及性别划分的数据集的学习效果, 如表 5 和 6 所示.

研究表明^[59,60], 在 65 岁以前, 神经系统疾病发病几率相对较低同时疾病症状较为稳定; 随着年龄的增大与人体的衰老, 65 ~ 74 岁疾病的发病几率增加并伴随某些症状加剧; 75 岁以上更易患有疾病并伴有某些症状指标更加严重. 因此将年龄段划分为 0 ~ 64, 65 ~ 74, 75 及以上 3 个阶段, 以此划分条件得到的样本集合数量分别为 94 个、52 个、30 个, 每个年龄阶段均由均衡的疾病与健康用户的样本集合组成; 表 4 ~ 6 分别给出了上述模型分类器对应样本集合 1000 次十折交叉验证结果求平均得到的平均准确率、病人组的查准率、查全率和 *F* 值.

从表 4 的类模型 *F* 值可以看出, DMEM (Ex) 对不同年龄段、性别患有神经系统疾病的用户均有较好的诊断效果, 具有一定的鲁棒性. 随着年龄段的生长, 模型的诊断效果有所降低. 女性神经系统疾病的诊断效果相比男性有所降低, 结合本文训练样本的分布, 65 岁以上样本数量的降低以及男女比例的差异一定程度影响 DMEM (Ex) 对 65 岁以上及女性用户诊断结果的准确程度.

从表 5 和 6 的结果看, 本文方法在不同年龄段和性别数据集上均获得了较好的学习效果. 从表 5 对于不同年龄段的分类模型 *F* 值可以看出, 0 ~ 64 年龄区间模型样本的结果最好, *F* 值为 0.873; 65 ~ 74 岁年龄区间结果相对较差, *F* 值为 0.813, 75 岁以上年龄段的模型样本的效果最差, *F* 值为 0.807. 年龄的增长对机体的衰老程度、疾病症状的表现与严重程度产生一定的影响, 因此较大程度影响了判别的精度.

由表 6 的实验结果可知, 不同性别样本模型的分类结果较为均衡, *F* 值都为 0.85, 虽然性别对某些疾病的发病几率具有一定影响, 但疾病对应的症状表现并不存在太大区别, 因此性别对神经系统疾病在语音与笔通道的交互特性影响相对较小.

实验三. 模拟了一个简单的病例样本扩充并重新迭代更新模型的过程, 从总实验样本集中随机挑选 36 个数据组成测试集合, 从剩下的样本 (140) 中随机挑选 40 个、60 个、80 个、100 个、120 个、140 个

表 5 不同年龄段模型的分类诊断效果

Table 5 Classification effects of DMEM (Ex) on different ages

Age range	Accuracy (%)	Precision	Recall	<i>F</i> value
0 ~ 64	89.14	0.916	0.858	0.873
65 ~ 74	86.99	0.845	0.824	0.813
Age above 75	94.47	0.806	0.820	0.807

表 6 不同性别模型的分类诊断效果

Table 6 Classification effects of DMEM (Ex) on different genders

Model	Accuracy (%)	Precision	Recall	<i>F</i> value
Male	86.27	0.884	0.838	0.849
Female	90.15	0.854	0.873	0.8489

图 6 (网络版彩图) 5 个模型对不同数据集的分类效果图

Figure 6 (Color online) Classification effects of five models on different datasets

逐次增加的样本模拟病例库增量过程训练病理模型, 这个过程重复 1000 次十折交叉验证得到平均诊断结果. 图 6 给出了实验一所述的 5 个病理模型随着训练样本容量扩充, 病理模型诊断概率变化的趋势图.

通过上述结果可以看出, 虽然使用语音通道, 笔通道的单通道特征, 可以对神经系统疾病进行诊断, 但是与单一通道病理检测模型相比, DMM 的检测效果相对较好; DMEM 进一步地提高识别的精度; DMEM (Ex) 的识别精度最好. 随着训练数据量的增加, 模型的准确率都在不断增加, 直至达到理想的效果. 因此, 基于云端融合的系统可以凭借其海量数据存储能力与数据统一管理的能力不断提升病理样本的数据量, 进而得到越来越准确, 越来越鲁棒的疾病诊断模型.

5 结论

本文提出了一个云端融合的神经系统疾病多通道辅助诊断框架. 利用不同的采集设备同时捕捉并分析用户在日常生活中的交互操作, 并将交互行为中包含的信息传输至云端. 随后, 在云端提取各个交互通道中包含的有效信息, 通过分析提取到的交互特征与病症之间的关系, 给出用户是否患有

神经系统疾病的自动判别结果,并将该结果与用户数据库内的用户信息,医学知识库的参数数据以及医疗人员对用户的疾病史给出的建议相结合,得到最终的诊断结果.最后,通过信息的分流优化将诊断结果实时反馈给用户.本方法可有效地利用多通道融合的交互信息,弥补单通道重要交互信息缺失的不足,进而给出更加精确和鲁棒的诊断结果,提升神经系统疾病检测的正确率;在数据的储存与运算方面,将采集到的数据保存在云端,保证了数据的一致性,也降低了数据采集的地域限制与时间限制;同时,基于云计算强大的运算能力,本文方法可以对用户在不同交互通道产生的信息进行实时分析,并结合先验信息,对疾病进行辅助诊断与早期预警.

为了证明上述方法的可行性,本文设计了一个语音和笔交互混合神经性疾病诊断系统.系统首先对语音通道和笔通道中的交互信息进行获取和分析,得到交互原语并生成场景任务.随后,系统对上述数据进行处理,并提取与病理信息相关的特征.之后,系统将获取到的多通道特征进行融合,并进行特征选择,通过预先训练得到的集成学习病理模型得到自动诊断结果.最后,系统综合考虑计算机诊断结果以及医学知识库的参数数据等先验信息,得到基于混合判决的诊断结果,并实时反馈给用户.实验表明,系统可以同时用户对用户在多个交互通道中产生的交互信息进行实时分析,并做出优于任一单独交互通道的疾病诊断结果.

除了语音通道和笔通道,步态通道和生理计算通道也是人机交互过程中比较重要的通道,也包含着大量的与神经系统疾病相关的重要信息.在后续工作中,可以尝试将步态、生理计算等更多通道信息加入到融合的通道中,从更多角度分析神经系统功能情况,进而更加准确地诊断用户是否患有神经系统疾病.同时,随着数据量的增大,可用的训练样本越来越多,可以考虑使用深度学习模型代替现有的传统机器学习模型,并利用其精度高、鲁棒性强等特点,得到更稳定、更精准的诊断结果.

参考文献

- Wirdefeldt K, Adami H O, Cole P, et al. Epidemiology and etiology of Parkinson's disease: a review of the evidence. *Eur J Epidemiol*, 2011, 26: 1–58
- Morgado P M. Automated diagnosis of Alzheimer's disease using PET images a study of alternative procedures for feature extraction and selection electrical and computer engineering. Dissertation for M.S. Degree. Lisbon: Technical University of Lisbon, 2012
- Smith E E, O'Donnell M, Dagenais G, et al. Early cerebral small vessel disease and brain volume, cognition, and gait. *Ann Neurol*, 2015, 77: 251–261
- Rijk M C D, Launer L J, Berger K, et al. Prevalence of Parkinson's disease in Europe: a collaborative study of population-based cohorts. *Neurology*, 2000, 54: 21–23
- Prince M, Ali G C, Guerchet M, et al. Recent global trends in the prevalence and incidence of dementia, and survival with dementia. *Alzheimer's Res Ther*, 2016, 8: 23
- Kim J S. Small vessel disease: not a small problem. *J Stroke*, 2015, 17: 1
- Skodda S, Gronheit W, Mancinelli N, et al. Progression of voice and speech impairment in the course of Parkinson's disease: a longitudinal study. *Parkinson's Dis*, 2013, 2013: 389195
- Khedher L, Ramírez J, Górriz J M, et al. Early diagnosis of Alzheimer disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing*, 2015, 151: 139–150
- Jaimes A, Sebe N. Multimodal human-computer interaction: a survey. *Comput Vis Image Und*, 2007, 108: 116–134
- Zhang W, Smith M L, Smith L, et al. Eye center localization and gaze gesture recognition for humancomputer interaction. *J Opt Soc Am A-optics Image Sci Vis*, 2016, 33: 314–325
- Whitling S, Rydell R, Ahlander V L, et al. Design of a clinical vocal loading test with longtime measurement of voice. *J Voice*, 2015, 29: 13–27
- Roy N, Barkmeier-Kraemer J, Eadie T, et al. Evidence-based clinical voice assessment: a systematic review. *Am J Speech-Lang Pathol*, 2013, 22: 212–226

- 13 Maccallum J K, Zhang Y, Jiang J, et al. Vowel selection and its effects on perturbation and nonlinear dynamic measures. *Folia Phoniatr Et Logo*, 2011, 63: 88–97
- 14 Ali Z, Elamvazuthi I, Alsulaiman M, et al. Automatic voice pathology detection with running speech by using estimation of auditory spectrum and cepstral coefficients based on the allpole model. *J Voice*, 2016, 30: 7–19
- 15 Moers C, Mobius B, Rosanowski F, et al. Vowel- and text-based cepstral analysis of chronic hoarseness. *J Voice*, 2012, 26: 416–424
- 16 Alnasheri A, Muhammad G, Alsulaiman M, et al. An investigation of multidimensional voice program parameters in three different databases for voice pathology detection and classification. *J Voice*, 2016, 31: 9–18
- 17 Majidnezhad V, Kheidorov I. The SVM-based feature reduction in vocal fold pathology diagnosis. *Int J Future Gener Commun Netw*, 2013, 6: 45–56
- 18 Tsanas A, Gómez-Vilda P. Novel robust decision support tool assisting early diagnosis of pathological voices using acoustic analysis of sustained vowels. In: *Proceedings of Multidisciplinary Conference of Users of Voice, Speech and Singing, Las Palmas*, 2013. 3–12
- 19 Ünlü A, Brause R, Krakow K. Handwriting analysis for diagnosis and prognosis of parkinson's disease. In: *Proceedings of the 7th International Conference on Biological and Medical Data Analysis, Thessaloniki*, 2006. 441–450
- 20 Smits E J, Tolonen A, Cluitmans L, et al. Standardized handwriting to assess bradykinesia, micrographia and tremor in Parkinson's disease. *PLOS One*, 2014, 9: e97614
- 21 Kim H, Cho Y S, Do E Y. Context-bounded refinement filter algorithm: improving recognizer accuracy of handwriting in clock drawing test. In: *Proceedings of National Conference on Artificial Intelligence, Atlanta*, 2010. 53–60
- 22 Davis R, Libon D J, Au R, et al. THink: inferring cognitive status from subtle behaviors. *AI Mag*, 2015, 36: 49–60
- 23 Hagler S, Jimison H B, Pavel M. Assessing executive function using a computer game: computational modeling of cognitive processes. *IEEE J Biomed Health Inf*, 2014, 18: 1442–1452
- 24 Drake J M, Griffen B D. Early warning signals of extinction in deteriorating environments. *Nature*, 2010, 467: 456–459
- 25 Chen D Y, Ding H, Zhou X Y, et al. A system of measuring and analyzing gait of human walking. *Chin J Biomed Eng*, 1997, 16: 133–141 [陈大跃, 丁洪, 周新宇, 等. 人体行走的步态测试与分析系统. *中国生物医学工程学报*, 1997, 16: 133–141]
- 26 Sinha A, Chakravarty K, Bhowmick B, et al. Person identification using skeleton information from kinect. In: *Proceedings of International Conference on Advances in Computer-Human Interaction, Nice*, 2013. 101–108
- 27 Paul M, Haque S M, Chakraborty S, et al. Human detection in surveillance videos and its applications a review. *EURASIP J Adv Signal Process*, 2013, 2013: 176
- 28 Lan K C, Shih W Y. Early diagnosis of Parkinson's disease using a smartphone. *Procedia Comput Sci*, 2014, 34: 305–312
- 29 Chen Y, Huang M, Hu C, et al. A coarse-to-fine feature selection method for accurate detection of cerebral small vessel disease. In: *Proceedings of International Symposium on Neural Networks, Saint Petersburg*, 2016. 2609–2616
- 30 Dirican A C, Gokturk M. Psychophysiological measures of human cognitive states applied in human computer interaction. *Procedia Comput Sci*, 2011, 3: 1361–1367
- 31 Altun K, Barshan B. Human activity recognition using inertial/magnetic sensor units. In: *Proceedings of International Conference on Pattern Recognition, Istanbul*, 2010. 38–51
- 32 Patel S, Lorincz K, Hughes R, et al. Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensor. *IEEE Trans Inf Technol Biomed*, 2009, 13: 864–873
- 33 Robichaud J A, Pfann K D, Leurgans S, et al. Variability of EMG patterns: a potential neurophysiological marker of Parkinson's disease? *Clin Neurophysiol*, 2009, 120: 390–397
- 34 Jordan K G. Emergency EEG and continuous EEG monitoring in acute ischemic stroke. *Clin Neurophysiol*, 2004, 21: 341–352
- 35 Armbrust M, Fox A, Griffith R, et al. A view of cloud computing. *Commun ACM*, 2010, 53: 50–58
- 36 Kamara S, Lauter K. Cryptographic cloud storage. In: *Proceedings of International Conference on Financial Cryptography and Data Security, Tenerife*, 2010. 136–149
- 37 Subashini S, Kavitha V. Review: a survey on security issues in service delivery models of cloud computing. *J Netw Comput Appl*, 2011, 34: 1–11
- 38 Wang H, Yi X, Bertino E, et al. Protecting outsourced data in cloud computing through access management. *Concurr*

- Comput Pract Exper, 2016, 28: 600–615
- 39 Tian F, Dai G Z, Chen Y D, et al. Specification and structure design of 3D interaction tasks. *J Softw*, 2002, 13: 2099–2105 [田丰, 戴国忠, 陈由迪, 等. 三维交互任务的描述和结构设计. *软件学报*, 2002, 13: 2099–2105]
 - 40 Huang J, Chen Y N, Liu J, et al. Multimodal human-computer interaction model for nerve function assessment in mobile environment. *J Softw*, 2016, 27: 156–171 [黄进, 陈毅能, 刘杰, 等. 移动环境下神经功能评价多通道交互模型. *软件学报*, 2016, 27: 156–171]
 - 41 Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intel*, 2005, 27: 1226–1238
 - 42 Chandrashekar G, Sahin F. A survey on feature selection methods. *Comput Electr Eng*, 2014, 40: 16–28
 - 43 Kumari B, Swarnkar T. Filter versus wrapper feature subset selection in large dimensionality micro array: a review. *Int J Comput Sci Inf Technol*, 2011, 2: 1048–1053
 - 44 Polikar R. Ensemble based systems in decision making. *IEEE Circ Syst Mag*, 2006, 6: 21–45
 - 45 Miao K H, Miao J H, Miao G J. Diagnosing coronary heart disease using ensemble machine learning. *Int J Adv Comput Sci Appl*, 2016, 7: 30–39
 - 46 Yang J, Li J, Shen R, et al. Exploiting ensemble learning for automatic cataract detection and grading. *Comput Meth Prog Biomed*, 2016, 124: 45–57
 - 47 Inzamam-Ul-Hossain M, MacKinnon L, Islam M R. Parkinson disease detection using ensemble method in PASW benchmark. In: *Proceedings of IEEE International Advance Computing Conference, Bangalore*, 2015. 666–670
 - 48 Lezak M D. *Neuropsychological Assessment*. New York: Oxford University Press, 1995. 650–680
 - 49 Arnett J A, Labovitz S S. Effect of physical layout in performance of the trail making test. *Psychol Assess*, 1995, 7: 220–221
 - 50 Sofranko J L, Prosek R A. The effect of levels and types of experience on judgment of synthesized voice quality. *J Voice*, 2014, 28: 24–35
 - 51 Tsanas A, Little M A, Mcsharry P E, et al. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Trans Biomed Eng*, 2010, 57: 884–893
 - 52 Yumoto E, Gould W J, Baer T. Harmonicstonoise ratio as an index of the degree of hoarseness. *J Acoust Soc Am*, 1982, 71: 1544–1550
 - 53 Srinivasan B, Spinner T, Rengaswamy R, et al. Control loop performance assessment using detrended fluctuation analysis (DFA). *Automatica*, 2012, 48: 1359–1363
 - 54 Godinollorente J I, Osmaruz V, Saenzlechon N, et al. The effectiveness of the glottal to noise excitation ratio for the screening of voice disorders. *J Voice*, 2010, 24: 47–56
 - 55 Ariaslondono J D, Godinollorente J I, Saenzlechon N, et al. An improved method for voice pathology detection by means of a HMM-based feature space transformation. *Pattern Recogn*, 2010, 43: 3100–3112
 - 56 Hall M A. Correlation-based feature subset selection for machine learning. *Dissertation for Ph.D. Degree*. Hamilton: University of Waikato, 1998
 - 57 Sellam V, Jagadeesan J. Classification of normal and pathological voice using SVM and RBFNN. *J Signal Inf Process*, 2014, 5: 1–7
 - 58 Seixas F L, Zadrozny B, Laks J, et al. A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment. *Comput Biol Medicine*, 2014, 51: 140–158
 - 59 Kalyani R R, Corriere M D, Ferrucci L, et al. Age-related and disease-related muscle loss: the effect of diabetes, obesity, and other diseases. *Lancet Diabetes Endocrinol*, 2014, 2: 819–829
 - 60 Ransmayr G, Kunig G, Neubauer M, et al. Effect of age and disease duration on parkinsonian motor scores under levodopa therapy. *J Neural Transm*, 1995, 9: 177–188

Multimodal aided neurological disease diagnosis with synergy of cloud and client

Yang LI¹, Jin HUANG¹, Feng TIAN^{1,2*}, Dongqi HAN¹, Junjun FAN¹, Hui CHEN¹, Xiaolan PENG¹, Guozhong DAI¹ & Hongan WANG^{1,2}

1. *Beijing Key Laboratory of Human-Computer Interaction, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China;*

2. *State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China*

* Corresponding author. E-mail: tianfeng@iscas.ac.cn

Abstract The dominating physiological characteristics of neurological diseases are reflected in patients' daily behaviors. The aided diagnosis and early warning of neurological diseases will benefit from obtaining and analyzing related physiological information generated during the interaction process. Traditional systems for detecting neurological conditions only analyze a single interaction modal, which may lose important features contained in other modalities. Based on the above, we propose a multi-modal aided neurological disease diagnosis system with synergy of cloud and client. First, we propose an automatic disease diagnosis method based on multi-modal information; users' physiological information collected from multiple modals is then analyzed and the results are integrated to improve accuracy and robustness. Second, a framework of cloud-client synergy is proposed that stores the user's physiological information from different regions and different times in the cloud, thus reducing the geographical restrictions and time constraints of data collection. Third, based on the powerful computing capacity, the system can make real-time and precise automatic diagnosis by analyzing multi-modal physiological information produced by users during natural interaction. Finally, the effectiveness of this multi-modal method for diagnosing aided neurological diseases based on cloud-client synergy is verified by using a hybrid diagnostic system of voice and pen.

Keywords synergy of cloud and client, multi-modal, human computer interaction, neurological disease, aided diagnosis

Yang LI is an engineer in Intelligence Engineering Lab, Institute of Software, Chinese Academy of Sciences, China. Her research interests include human-computer interaction.

Jin HUANG is currently a doctoral student in Intelligence Engineering Lab, Institute of Software, Chinese Academy of Sciences, China. His research interests include human-computer interaction and computer graphics.

Feng TIAN is a professor and the deputy chief engineer in Institute of Software, Chinese Academy of Sciences. He is interested in interaction techniques and tools of pen-based user interface, multimodal user interface, 3D user interface, and ubiquitous computing.

Dongqi HAN is currently a graduate student at Intelligence Engineering Lab, Institute of Software, Chinese Academy of Sciences, China. His research focus is human computer interaction.