第 42 卷 第 6A 期
2015 年 6 月
计 算 机 科 学
Computer Science
Vol. 42 No. 6A
June 2015

基于嗅觉网络传输的重症疾病诊断机制与算法研究

李联宁
(西安交通大学城市学院 西安 710018)

摘 要 目前在医学界普遍存在的一个问题是:许多重症疾病的患者到医院就医确诊时已进入疾病晚期,失去了在早期发现本可医疗治愈的机会,但同时也缺乏必要的手段对海量的潜在患者人群进行有效筛查。研究的目的是探索通过嗅觉网络传输并远距离诊断的模式,实现超大规模采集早期患者的呼出气味数据,达到重症疾病早期预警的效果。主要讨论使用基于智能手机平台的微机电系统(MEMS)构造的光谱吸收型光化学气体传感器,在使用者使用移动电话的同时获得呼出气体的特征光谱,将其发送至云计算网络中查询云存储气味数据库,使用 BP 算法进行光谱疾病特征码比对。对疾病特征码符合率超过一定比率的潜在患者进行短信报警、督促就医核查,以便早期发现重症疾病,提高医疗效果,降低病患死亡率。

关键词 嗅觉,气体传感器,云计算,BP 算法,智能手机,重症疾病,诊断
中图法分类号 TP393 文献标识码 A

Study of Severe Disease Diagnosis Mechanism and Algorithm of Olfactory Based on Network Transmission

LI Lian-ning
(Xi'an Jiaotong University City College, Xi'an 710018, China)

Abstract This paper mainly discussed the use of MEMS spectrum absorption photochemical gas sensor based on intelligent mobile phone platform. The characteristic spectrum of exhaled gas was obtained when the users use mobile phones, at the same time, it was sent to the cloud computing cloud storage odor database query in network, and BP algorithm was used for spectral features of disease code comparison. To these patients whose potential rate is more than a certain percentage the potential rate is more than a certain percentage, SMS alarm on disease characteristic code conforms to supervise the medical check. For early detection of severe disease, it can improve the medical effect, reduce the mortality of patients.

Keywords Olfactory, Gas sensor, Cloud computing, BP algorithm, Intelligent mobile phone, Severe disease, Diagnosis

1 引言

重症疾病一般包括恶性肿瘤、高血压、糖尿病、冠心病、肝硬化、再生障碍性贫血、重症肌无力、帕金森综合症、系统性红斑狼疮、精神病、血友病、类风湿性关节炎、肺心病、慢性肾功能衰竭等。许多重症疾病患者到医院就医确诊时已进入疾病晚期,失去了在早期发现本可医疗治愈的机会。

例如,糖尿病是一种严重的代谢性疾病,患上糖尿病的人每年都在不断的增长,由于糖尿病严重危害着患者的健康,很多人都担心患上这种疾病。糖尿病筛查可以帮助我们及早地发现糖尿病,避免病情恶化。但是糖尿病筛查需要做糖化血红蛋白检查(HbA1C)、尿酮体检查、胰岛素或 C 肽检查等医院临床检查,会消耗大量的人力、时间与资金。如何对存在的海量潜在患者人群进行有效筛查一直没有有效的途径和方法。

又如对于胃癌患者来说,目前,医生在诊断胃癌时所采取的方法是通过探针和微型摄像机对患者的食道和胃壁进行检查,仅有 20% 患者在确诊时还能够接受手术治疗,大部分患者在发现时已是晚期。如果在癌症早期就对患者进行确诊,将大大提高胃癌患者的存活率。

2 国内外嗅觉识别疾病及诊断研究现状

2.1 重症疾病病人呼出气体检测机制

通过研究证实,当疾病影响了病人的细胞代谢时,大量的外泄挥发性化学物质可能会进入病人的呼吸气体、尿液和汗液中。患不同疾病的人所呼出的气体中会出现某些特定成分,如肝硬化患者的呼气中会出现脂肪酸,肾衰竭患者的呼气有三甲氨,肝癌患者的呼气中存在烷类和苯的衍生物等,通过分析病人呼出的气体中所含乙烷分子含量,可断定病人是否患上肺癌等疾病。各种癌症都有它的气味,通过闻气味通常可以诊断癌症,但人的鼻子要想嗅出癌症的气味是很困难的。

现在,世界各地的研究实验室都在寻找使用气体传感器来探测疾病的方法。我们期望研制出一种可完成气体信号远程传输的设备,它可以通过网络将气味信号传输到远方,从病人的气味描绘出病人的“呼吸图”,医生可以从中看出病人是否受到疾病的侵害。这样,医生就可以对糖尿病及肺部肿瘤等病症做非侵入性的远程医学诊断,也可以准确地测定环境地域影响的高发疾病比率。

本文受西安交通大学城市学院第二轮科研项目基金(2013KZ07)资助。
李联宁(1949—),男,教授,主要研究方向为计算机网络、物联网、多媒体技术,E-mail:li-ln@263.net。

2.2 人工嗅觉研究发展现状

2.2.1 人工嗅觉系统的基本组成

人工嗅觉系统模拟人的嗅觉器官，主要包括3大部分：气敏传感器、信号处理及模式识别。气味分子被人工嗅觉系统中的气敏传感器阵列吸附，产生信号，信号经处理电路加工处理，并完成信号转换与传输，最后经计算机模式识别做出判断。

气体传感器是用来检测气体的成份和含量的传感器。一般认为，气体传感器是一种将某种气体体积分数转化成对应电信号的转换器。探测头通过气体传感器对气体样品进行调理和检测。

目前气敏传感器主要有金属氧化物半导体、石英晶振、声表面波、导电有机聚合物膜与红外线光电等类型，其中以SnO_2为代表的金属氧化物半导体气敏传感器应用最为广泛[1]。光纤气体传感器根据传感原理分两大类：一类是传光型光纤气体传感器，光纤在传感系统中只起到传输光波的作用，探头则为外加的换能器；另一类是传感型光纤气体传感器，光纤不仅具有光波传导作用，还有气体探头的作用。

信号处理主要完成传感器信号的放大和滤波，进行特征提取，得到多维有用的响应信号，并由A/D转换成数字信号输入计算机。由气敏传感器阵列产生的电信号经处理后，可用绝对电压、电阻或电导来表示，也可用相对值如归一化的电压、电阻或电导来表示。

在人工嗅觉领域中，模式识别的研究和讨论始终较为活跃。模式识别是对传感器阵列的输出信号进行适当的处理，以获得混合气体组分信息和浓度信息。

在人工嗅觉系统中，常用的模式识别方法有统计模式识别方法和智能识别方法。统计模式识别方法主要有主成分分析法(Principal Components Analysis，PCA)、偏最小二乘法(Partial Least Squares，PLS)和聚类分析法(Cluster Analysis，CA)等，智能识别方法主要是人工神经网络(Artificial Neural Network，ANN)方法和模糊推理法(Fuzzy Illation，FI)[2]。

人工嗅觉系统的响应机理及其模型比较复杂，非线性严重，数学模型难以建立。统计模式识别方法普遍是基于线性的分析方法，只模仿了人的逻辑思维，它就数据处理后所得到的结果与人的感官感受之间无法对应起来，应用具有较大的局限性。在很多情况下，人们只需要得到与人的感官感受相一致的结果，对气味的化学组成与浓度高低并非十分关心，而人工神经网络法则显示了其优越性。

人工神经网络法是接近人类大脑思维方法的一种算法，人工神经网络由大量简单的处理单元即神经元，广泛地互为连接而形成复杂的网络系统。人工神经网络的一个显著特征就是优秀的学习能力。它通过学习，自动掌握和挖掘隐藏在事物内部的、不能用明确数学表达式表示的关系，并能够处理非线性数据，具有良好的容错性能和较高的预测精度。

模糊推理法模仿人的判断，不给出气味浓度的精确值，而是根据其表示的模糊逻辑，变换成“很高、高、中等、低、很低”等与人的亲和性较高的语言变量。对于系统误差所引起的传感器输出包含一定误差的情况，模糊推理法十分有用。人工神经网络、模糊理论、遗传算法以及混沌等融合在一起的信息处理技术的兴起为人工嗅觉系统的发展注入了新的活力。

2.2.2 人工嗅觉系统在临床医学中的应用

人体疾病与人的体味如口腔气味、汗液、尿液的直接关系，从远古就已为人们所认识。中医运用望、闻、问、切的手段来诊断病人，早就非常重视人体体味对疾病的诊断作用。从近代医学的观点来看，人体各部位的不同疾病都会引起血液、细胞新陈代谢的异常。如在肺部血液的氧化交换时，疾病患者和健康人的肺部呼出的气味和随尿液排出的味道是不同的。

研究表明，不同疾病患者所呼出的气体中会出现某些特定的成份，如胃溃疡患者呼出的气体中氢气成分比常人高，糖尿病患者呼出的气体会含有酮类气体成分，肺癌患者呼出的气体中乙烷含量比正常人高，肝癌患者呼出的气体中存在烷类和苯的衍生物，肾衰竭患者呼出的气体中会含有三甲氨等。因此通过检测分析患者呼出气体的成份与含量便可进行相应的定性定量的疾病诊断。

医学方面，利用人呼出的气体进行疾病的诊断，近年来已成为国际上的一个研究热点。为寻求一种非侵入式、无伤害、快速准确的检测手段，人工嗅觉系统为研究新型的疾病诊断仪器提供了可能。

目前，已有研究人员开展了人工嗅觉系统的相关研究工作，初步研制了基于MEMS技术的半导体气敏传感器阵列，并进行了一定深度的理论研究与识别气味的实验，获取了相应的图谱[3]。

3 研究探索方向与关键科学问题

研究的探索方向是用研制的气敏传感器阵列进行人体呼出气体初步检测，能在较短时间内获得被测对象呼出的相应图谱，区别出健康人与疾病患者的不同特征图谱。进一步的研究工作将把人工嗅觉系统用于临床诊断，尤其是恶性肿瘤的非侵入式诊断。

关键科学问题主要有如下几点：

①嗅觉感受的具体重症疾病生物学机制与系统框架设计。

②气体传感器MEMS微型化设计及制造方案；

③气味分析算法和疾病特征码获取；

④分布式云存储数据库建立及查询研究；

⑤适用于手机平台的嗅觉传导的BP神经网络算法。

3.1 嗅觉感受的具体重症疾病生物学机制

人体嗅觉系统是由嗅觉感受器接收气味信号的，人体嗅觉感受器是位于上鼻道及鼻中隔后上部的嗅上皮。从仿生学角度考虑，人体嗅觉感受器构成与机理可概括为3个部分：

① 鼻腔上皮组织(初级)，是接受气体并产生信号的第一个地方；

② 嗅觉球(二级)，气体的种类通过“镜像”在这里形成；

③ 大脑皮层，信息之间的联系在这里形成并存储。通过这个过程，各种不同的气味得以被人所识别。

人的嗅觉形成过程是嗅觉传感器工作原理的模拟基础，可以简单地从结构上将传感器阵列、信号预处理、模式识别分别与嗅觉膜、嗅小球、神经中枢相类比，更重要的是在功能上电子鼻系统也具有生物嗅觉系统的特点：对多种气体或气味敏感；通过必要的处理，能够识别所感受到的气体。

当待测气体呈现在一种敏感的传感器面前时，传感器将化学输入转换成电信号，由多个传感器对某种气体的响应便构成了传感器阵列对该气体的响应谱。显然，气体中的各种化学成分均会与敏感材料发生作用，所以这种响应谱为该气体的广谱响应谱。为实现对气体的定性或定量分析，必须将传感器的信号进行适当的预处理(消除噪声、特征提取、信号

放大等)后采用合适的模式识别分析方法对其进行处理。

理论上,每种气体都有其特征响应谱,根据其特征响应谱可区分不同的气体。同时还可利用气敏传感器构成阵列对多种气体的交叉敏感性进行测量,通过适当的分析方法,实现混合气体分析[4]。

3.2 系统框架设计

系统的总体框架设计主要分成3部分:

(1)典型疾病气味数据库

包括典型病人呼出气体的采样、质谱仪分析、疾病特征码的提取、气味数据库的构造、疾病特征码存储等工作。

(2)手机用户呼气采集与分析

包括呼气采集、气体传感器传感、样本光谱分析、远程传输、云计算分析、疾病气味数据库比对、验证确诊分析计算、确诊病例报警、相关数据输入确诊病人数据库等工作。

(3)确诊病人数据库

包括确诊病人数据库的构造、地区病人分析与统计、分析报表的生成等工作。

具体系统构造与概要设计见图1系统数据流图。

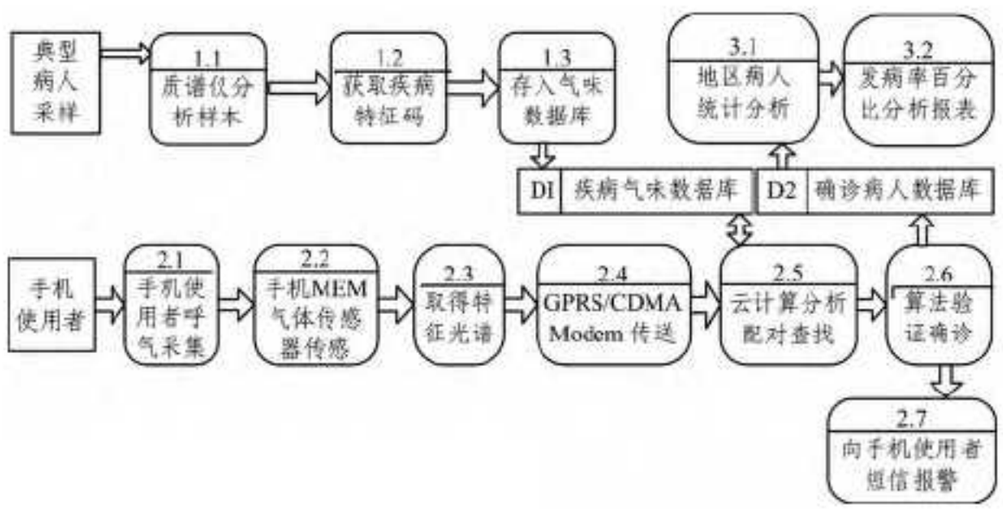

图1 系统数据流图

3.3 光谱吸收型气体传感器

从本质上讲,气体传感器是一种将某种气体体积分数转化成对应电信号的转换器。根据气体传感器使用的气敏材料以及气敏材料与气体相互作的效应的不同,可以大致将气体传感器分为以下几类:半导体气体传感器、电化学气体传感器、固体电解质气体传感器、接触燃烧式气体传感器、光化学型气体传感器等。

光学式气体传感器包括红外吸收型、光谱吸收型、荧光型、光纤化学材料型等。其主要以光谱吸收型为主。光谱吸收型气体传感器是最重要、最简单的一类光纤气体传感器。它的工作原理是不同的气体物质由于其分子结构不同、浓度不同和能量分布的差异而有各自不同的吸收光谱。这就决定了光谱吸收型气体传感器的选择性、鉴别性和气体浓度的唯一确定性。若能测出这种光谱,便可对气体进行定性、定量分析。其主要优点是灵敏度高、可靠性好。

正是由于光谱吸收型光学式气体传感器可对气体进行定性、定量分析,因此在本研究项目中其成为对重症疾病病人呼出气体进行检测传感器的唯一选择。

1. 吸收型光纤气体传感器的基本原理

光谱吸收型光纤气体传感器利用气体在石英光纤透射窗口($0.8\mu m\sim1.7\mu m$)内的吸收峰,测量由于气体吸收产生的光强衰减,得到气体的浓度。常见的气体(如CO、CH_4、NO_2、CO_2)在石英光纤透射窗口都有泛频吸收线,在这一波段发光器件和接收器件都是比较理想的光电转换器件。因此用这种方法可以对大多数的气体浓度进行较高精度的测量。

由于每一种气体都有固有的吸收光谱,因此当光源的发射谱与气体的吸收谱相吻合时,就会发生共振吸收,其吸收强度与该气体的体积分数有关,通过测量光谱的吸收强度就可测量气体的体积分数。

当一束光强为I_0的输入平行光通过待测气体时,如果光源光谱覆盖一个或多个气体的吸收谱,则光通过气体时发生衰减。

根据Beer-Lambert定律,出射光强I与入射光强I_0和气体的体积分数之间的关系为

$$I=I_0\exp[-\alpha CL] \tag{1}$$

式中,α为气体吸收系数;L为吸收路径的长度,单位为m;C为气体的体积分数。对式(1)进行变换

$$C=\ln(I_0/I)/(\alpha L) \tag{2}$$

从式(2)可知,如果L、α已知,那么,通过检测I、I_0就可以得到待测气体的体积分数C。

吸收型光纤气体传感器的一大优点是具有简单可靠的气室结构,而且只需要调换光源,对准另外的吸收谱线,就可以用同样的系统来检测不同的气体。

光谱吸收型光纤气体传感器是应用较为广泛的一类气体传感器。它采用的是普通光纤或多模光纤,这种传感器由光源、气室、双波束或双波长的光路以及信号处理4个环节组成。光源通常采用半导体激光器,包括发光二极管、激光二极管和分布反馈式半导体激光器(Distributed Feedback Laser Diode,DFBLD)。

2. 新型传感器工艺

在微电子和微机械迅速发展的基础上,基于MEMS的新型微结构气敏传感器主要有硅基微结构气敏传感器和硅微结构气敏传感器。硅基微结构气敏传感器是衬底为硅、敏感层为非硅材料的微结构气敏传感器,主要有金属氧化物半导体、固体电解质型、电容型、谐振器型。硅微结构主要有金属氧化物_半导体_场效应管(MOSFET)型和钯金属_绝缘体_半导体(MIS)二极管型。

MEMS技术将传感器与IC电路集成一起,而且精度高、体积小、质量轻功耗低、选择性高、稳定型高,同种器件之间的互换型高,可以批量生产。所以它是传感器工艺的发展方向,而且基本上所有的传感器都可以用MEMS技术生产。

MEMS技术和纳米技术的发展将为气敏传感器的发展提供更广阔的前景。同时实现传感器陈列,即电子鼻集成成为可能,并将有很大的发展空间,给传感器带来新的发展篇章。

3.4 气味分析算法

当使用差分吸收检测算法时,当光波通过气体时,考虑到光路的干扰因素,式(1)的比尔-朗伯定律表示为

$$I(\lambda)=I_0(\lambda)\exp[-\alpha(\lambda)CL+\beta(\lambda)] \tag{3}$$

式中,$I(\lambda)$为透射光强;$I_0(\lambda)$为入射光强;$\alpha(\lambda)$为给定波长的单位浓度、单位长度气体的吸收系数;L为待测气体与光相互作用的长度;C为待测气体的浓度;$\beta(\lambda)$为光路干扰系数。

当用光纤传感系统检测气体时,式(3)还应包含一个比例系数K,则改写为

$$I(\lambda)=I_0(\lambda)K(\lambda)\exp[-\alpha(\lambda)CL+\beta(\lambda)] \tag{4}$$

仅从式(4)确定待测气体的浓度C是困难的,因为$\beta(\lambda)$是一个随机变量。如果用两个波长(λ_1、λ_2)相近(但在吸收系数上有很大差别)的单色光同时或分别通过待测气体(但时间间隔很小),有

$$I(\lambda_1)=I_0(\lambda_1)K(\lambda_1)\exp[-\alpha(\lambda_1)CL+\beta(\lambda_1)] \quad (5)$$

$$I(\lambda_2)=I_0(\lambda_2)K(\lambda_2)\exp[-\alpha(\lambda_2)CL+\beta(\lambda_2)] \quad (6)$$

由式(5)和式(6)得，待测气体的浓度可以表示为

$$C=\frac{1}{[\alpha(\lambda_1)-\alpha(\lambda_2)]L}\{\ln\frac{K(\lambda_1)I_0(\lambda_1)}{K(\lambda_2)I_0(\lambda_2)}-\ln\frac{I(\lambda_1)}{I(\lambda_2)}-[\beta(\lambda_2)-\beta(\lambda_1)]\} \quad (7)$$

由于 λ_1、λ_2 相差很小，光几乎同时接近和通过待测气体，可以认为

$$\beta(\lambda_1)\approx\beta(\lambda_2)$$

因此式(7)可以简化为

$$C=\frac{1}{[\alpha(\lambda_1)-\alpha(\lambda_2)]L}\ln\frac{K(\lambda_1)I_0(\lambda_1)I(\lambda_2)}{K(\lambda_2)I_0(\lambda_2)I(\lambda_1)} \quad (8)$$

适当调节光学系统，使

$$K(\lambda_1)I_0(\lambda_1)=K(\lambda_2)I_0(\lambda_2) \quad (9)$$

则式(8)又可以简化为

$$C=\frac{1}{[\alpha(\lambda_1)-\alpha(\lambda_2)]L}\ln\frac{I(\lambda_2)}{I(\lambda_1)} \quad (10)$$

实际应用中，波长为 λ_1 的光对应于检测其他的吸收谱线，波长为 λ_2 的光不被检测气体吸收(即为参考波长)，在测试过程中为一空值，因此有 $I(\lambda_1)/I(\lambda_2)<1$。将对数 $\ln[I(\lambda_1)/I(\lambda_2)]$ 在 λ_1 和 λ_2 附近进行泰勒展开，可得到

$$\ln\frac{I(\lambda_2)}{I(\lambda_1)}=-\ln[1+(\frac{I(\lambda_1)}{I(\lambda_2)}-1)]\approx\frac{I(\lambda_2)-I(\lambda_1)}{I(\lambda_2)} \quad (11)$$

于是气体浓度为

$$C=\frac{1}{\alpha(\lambda_1)-\alpha(\lambda_2)L}\frac{I(\lambda_2)-I(\lambda_1)}{I(\lambda_2)} \quad (12)$$

在波长 λ_1、λ_2 下，若气体的吸收系数 α_1、α_2 可以测量，则气体浓度就可以从 $I(\lambda_2)-I(\lambda_1)$ 和 $I(\lambda_2)$ 的测量中求出，称为差分吸收技术。从式(12)可以看出，差分技术不仅从理论上完全消除了光路的干扰因素，而且还消除了光源输出光功率不稳定的影响

3.5 疾病特征码获取

呼气时会产生各种各样的挥发性有机化合物，这些化合物是新陈代谢造成的。癌症患者的新陈代谢会发生变化，呼出气体中的挥发性有机化合物也会随之变化。

研究人员发现，癌细胞散发出的分子与健康细胞不同，使用质谱仪或电子鼻可以察觉到这种差别。每个人所呼出的气体成分都是不同的，就像人的指纹一样，因此可称为“气体指纹”。医生可以通过对“气体指纹”进行仪器检测，来对诸如癌症等疾病做出诊断。在传统的中医疗法中，医生往往通过望闻问切就能完成对患者疾病的诊断，其中包括呼气测试法。

瑞士苏黎世联邦理工学院公布的一项研究成果表明，每个人在呼吸时呼出的化合物和人类的指纹一样独一无二，类似于“气体指纹”，医生甚至可以根据这些“气体指纹”来诊断疾病(如癌症)。在该项研究中，研究人员在为期 9 天的时间里，分别对 11 名志愿者进行了 4 次呼气测试，他们利用质谱仪对志愿者呼气中的化合物成分进行了分析。结果显示，每个人所呼出的气体中都含有水蒸气和二氧化碳，但其他成分却不尽相同，同时在 4 次呼吸检测中，每个人呼气的成分构成几乎都是独一无二并且基本保持不变的。

1. 质谱分析

质谱分析是一种测量离子荷质比(电荷_质量比)的分析方法，其基本原理是使试样中各组分在离子源中发生电离，生成不同荷质比的带正电荷的离子，经加速电场的作用，形成离子束，进入质量分析器。在质量分析器中，再利用电场和磁场使其发生相反的速度色散，将它们分别聚焦从而得到质谱图，进而确定其质量。

2. 气体质谱仪

质谱仪的基本工作原理是以电子轰击或其他的方式使被测物质离子化，形成各种质荷比(m/e)的离子，然后利用电磁学原理使离子按不同的质荷比分离并测量各种离子的强度，从而确定被测物质的分子量和结构。气体质谱仪多用于生产研究中以监测气体和进行多种痕量气体分析。

3. 疾病特征码获取

首先大范围采集已确诊的重症疾病患者的呼出气体，将患者呼出的气体先通过气相色谱分离后，每一成分被电离成为离子，进入质谱仪中，进行分析，得到该疾病患者的质谱图；在收集了大量该种疾病的质谱图样本后，进而分析确定该种疾病的典型质谱图。为方便起见，我们将这种疾病的典型质谱图称为疾病的“特征码”，并将这种疾病特征码存入网络数据库。其目的是将远程传输来的潜在未知疾病患者的呼出气体数据与数据库中疾病特征码进行比对，如与某种疾病特征码高度相似，则可以初步确诊为该种疾病。

3.6 分布式云存储数据库建立及查询研究

由于各个国家及地区的重症疾病发病率有很大的区别，全方位采集疾病样本的难度很大，因此建立统一的全球疾病特征码数据库并不现实。比较现实的办法是在各种重症疾病的高发区集中采集某种疾病的患者的呼出气体样本，使用气体质谱分析的方法分析出该种疾病的特征码，并存入地区性分布疾病特征码数据库，然后在使用时供全球范围的检索查询。目前云计算技术的发展给全球分布式云存储疾病特征码数据库提供了很好的技术平台。

1. 分布式云存储疾病特征码数据库的建立

近年来，云存储成为信息存储领域的一个研究热点，云存储可以实现存储完全虚拟化，大大简化应用环节，节省客户建设成本，同时提供更强的存储和共享功能。云存储中所有设备对使用者完全透明，任何地方任何被授权用户都可以通过一根接入线与云存储连接，进行空间与数据访问。该技术具有较高的应用价值。

以云计算技术为背景，基于互联网远程网络服务平台，应用方案可划分为数据存储、数据组织与数据检索 3 部分。

通过云计算核心技术之一的高性能分布式存储技术，结合分布式疾病特征码数据库检索要求，基于分布式通讯服务平台，实现海量数据实时写入、分布式存储、数据检索准确高效的应用要求。

分布式云存储疾病特征码数据库可以考虑建立在 Hadoop[3] 分布式文件系统(HDFS)的基础上，设计成适合运行在通用硬件上的分布式文件系统，即 HDFS(Hadoop Distributed File System)。Hadoop 由于是 Apache 基金会支持的一个开源的分布式计算平台项目，是 GFS 的开源实现，因此可以在有限的经费预算下实现。HDFS 是一个高度容错性的系统，适合部署在廉价的机器上，能提供高吞吐量的数据访问，非常适合大规模数据集上的应用。

分布式疾病特征码数据库中主要存储地区性分布疾病特征码数据库，在使用时供全球范围的检索查询。

2. 基于云技术的分布式疾病特征码数据库检索机制研究

(1)谱图库检索

质谱系列的测定是计算机在质谱应用中的一次飞跃，这

种质谱系列可提供给我们大量的质谱信息，并使质量数作为时间的函数成为可能。如在低分辨质谱中，为了说明实验中所测得的大量信息（如质谱系列），谱图库检索（Lib：ary Soach）是一种应用最广泛的方法。

所谓谱图库检索就是应用所贮存的大全质谱信息来辨认或评定联机运转中质谱仪所提供的大量质谱图。也就是说，将未知谱图与参考谱图相比较，以寻找相同或相似的质谱，从而确定化合物的结构或化合物的组成。其中主要涉及到质谱分析中谱图库检索的谱图编码、谱图库、检索系统等。

在谱图库检索中，对原始谱图进行简化和编码是十分必要的；否则，存贮量大，检索速度慢，且效果不一定会好。

(2)计算机辅助谱图解析和检索方法

计算机辅助谱图解析方法可以粗略地分为两大类：直接谱图库手段，即谱图库检索(library search)；间接谱图库手段，包括波谱模拟、模式识别和人工智能。

① 特征峰检索

谱图中某一质量峰的出现，就意味着在化合物中具有某种特定的子结构。根据子谱图与子结构具有相关性的一般原理，可以建立特征峰检索算法。在这种检索中，要尽可能选取特征性比较明显的谱峰。如首先键入某一峰值（设为 S1），之后计算机作出应答。其响应（即具有所键入谱峰的化合物）设为集合 A，则

$A=\{S1 \mid ID\#(S1)=$包含询问谱峰 S1 的谱图$\}$

若集合 A 过大，即匹配的化合物过多，则可继续进行询问。如键入谱峰 S2，设此次结果为集合 B，则

$B=\{S2 \mid ID\#(S2)=$包含询问谱峰 S2 的谱图$\}$

然后通过计算机进行交运算，设交集为 C，则

$C=A\cap B$

重复如上操作，当匹配化合物的数量减少到一定量（即交集足够小）时，则可将谱图信息输出。这种连续的、以交互问答方式进行检索的方法，对于具有特征谱峰的化合物是颇为有效的。

②相似检索

在任何一种谱库检索中，相似检索都是最为重要的。道理很简单，因为不管库中谱图量再大，而最新合成的化合物也难以包括进去。也就是说最普通、最大量的检索是相似检索。特别是当未知物不在谱图库中时，相似检索更为重要。

项目研究初期，可以使用德国马克斯普郎克学会煤炭研究所(Max—Planck—Institut für Kohlenforschung)的质谱谱图库系统 Masslib，目前该质谱谱图库拥有 130000 余张谱图，是世界上最大的质谱谱图库系统之一，SISCOM 为该系统中的相似检索子系统。检索中主要依据如下 6 个参数：

Nc：在未知谱与参考谱中共同存在的谱峰数；

NR：仅在参考谱图中存在的谱峰数；

Ns：仅在未知试样谱图中存在的谱峰数；

IR：NR 个谱峰总强度值与参考谱图中总谱峰强度值的比值；

Is：Ns 个谱峰总强度值与未知谱图中总谱峰强度值的比值；

Pc：未知谱与参考谱图的相似系数。

SISCOM 为一多参数系统，适应性很强，可用于多种类型有机及有机混合物的检索。

③加密检索

随着存储系统和存储设备越来越网络化，在海量的加密信息存储中，加密检索是实现信息共享的主要手段，是加密存储中必须解决的问题之一。加密检索技术有线性搜索算法、基于关键词的公钥加密搜索、安全索引、引入相关排序的加密搜索算法。

其中基于关键词的公钥加密搜索算法[4]比较适合分布式疾病特征码数据库的检索，该算法由 Boneh 等人提出的，其目的是可以在用户端存储、计算资源不足的情况下，通过访问远端数据库获取数据信息。此算法首先生成公钥、私钥，然后对存储的明文关键词用公钥进行加密，生成可搜索的密文信息。

此算法可以解决两方面的问题：1)存储、计算资源分布的不对称性，即用户的计算存储能力不能实时满足其需求；2)用户在移动情况下对存储、检索数据的需求，如 Email 服务等。

3.7 适用于手机平台的嗅觉传导的 BP 神经网络算法

人工神经网络是由大量的简单基本元件——神经元相互联接而成的自适应非线性动态系统。每个神经元的结构和功能比较简单，但大量神经元组合产生的系统行为却非常复杂。人工神经网络反映了人脑功能的若干基本特性，但并非生物系统的逼真描述，只是某种模仿、简化和抽象。与数字计算机相比，人工神经网络在构成原理和功能特点等方面更加接近人脑，它不是按给定的程序一步一步地执行运算，而是能够自身适应环境，总结规律，完成某种运算、识别或过程控制。

BP(Back Propagation)网络是 1986 年由 Rumelhart 和 McCelland 为首的科学家小组提出的一种按误差逆传播算法训练的多层前馈网络，是目前应用最广泛的神经网络模型之一。BP 网络能学习和存贮大量的输入_输出模式映射关系，而无需事前揭示描述这种映射关系的数学方程。其学习规则使用最速下降法，通过反向传播来不断调整网络的权值和阈值，使网络的误差平方和最小。BP 神经网络模型拓扑结构包括输入层(input)、隐层(hide layer)和输出层(output layer)[5]。

BP 算法的基本思想是：学习过程由信号正向传播和误差的反向回传两个部分组成；正向传播时，输入样本从输入层传入，经各隐层依次逐层处理，传向输出层，若输出层输出与期望不符，则将误差作为调整信号逐层反向回传，对神经元之间的连接权矩阵做出处理，使误差减小。经反复学习，最终使误差减小到可接受的范围。具体步骤如下：

①从训练集中取出某一样本，把信息输入网络中；

②通过各节点间的连接情况正向逐层处理后，得到神经网络的实际输出；

③计算网络实际输出与期望输出的误差；

④将误差逐层反向回传至之前各层，并按一定原则将误差信号加载到连接权值上，使整个神经网络的连接权值向误差减小的方向转化；

⑤对训练集中每一个输入_输出样本对重复以上步骤，直到整个训练样本集的误差减小到符合要求为止。

算法步骤如下：

(1)初始化，随机给定各连接权$[w]$，$[v]$及阈值θ_i，r_t；

(2)由给定的输入输出模式对计算隐层、输出层各单元的输出

$$b_j=f(w_{ij}a_i-\theta_j)\quad c_t=f(v_{jt}b_j-r_t)$$

式中，b_j 为隐层第 j 个神经元实际输出；c_t 为输出层第 t 个神经元的实际输出；w_{ij} 为输入层至隐层的连接权；v_{jt} 为隐层至输出层的连接权。

$$d_{tk}=(y_{tk}-c_t)c_t(1-c_t)$$

$$e_{jk}=[d_{tv}j_t]\ j_b(1-b_j)$$

式中，d_{tk}为输出层的校正误差；e_{jk}为隐层的校正误差。

(3)计算新的连接权及阈值，计算公式如下：

$$v_{jt}(n+1)=v_{jt}(n)+d_{tk}b_j\ w_{ij}(n+1)=w_{ij}(n)+e_{ka}i_k$$

$$r_t(n+1)=r_t(n)+d_{tk}\ \theta_j(n+1)=\theta_j(n)+e_{jk}$$

(4)选取下一个输入模式对返回第(2)步反复训练直到网络设定输出误差达到要求结束训练。

在原则上由非常简单的单元连接在一起组成的“网络”可以对任何逻辑和算术函数进行计算。因为网络的单元有些像大大简化的神经元，它现在常被称作“神经网络”。

结束语　但由于课题基于手机(移动电话)的平台进行，必须考虑到手机的智能化程度，核心处理器的处理能力、存储空间、电源供应等诸多特殊环境。特别是手机设备而具备的微型化，虽然带来作为移动终端设备而具备的便携、用户诸多的好处，但必须进一步研究按照将主要的计算工作交由“云计算”的模式进行，以适应手机工作模式[7]。

参 考 文 献

[1] Gopel W, Schierbaum K. SnO_2 sensors: Current status and future prospects [J]. Sensors and Actuators B, 1995, 26: 1-12

[2] Mielle P, et al. An alternative way to improve the sensitivity of electronic olfactometers[J]. Sensors and Actuators B, 1999, 58: 526-535

[3] Noll M G. Running Hadoop On Ubuntu Linux (Single-Node Cluster), August 5, 2007. http://www.michael-noll.com/

[4] Piazza L, Benedetti S. Investigation on the rheological propertiesof agar gels and their role on aroma release in agar/limonene solidemulsions [J]. Food Research International, 2010, 43: 269-276

[5] 洪雪珍，王俊. 基于逐步判别分析和 BP 神经网络的电子鼻猪肉储藏时间预测[J]. 传感技术学报，2010，23(10)：376-380

[6] Hirano S H, Truong K N, Hayes G R. uSmell: a gas sensor system to classify odors in natural, uncontrolled environments, the 2012 ACM Conference on Ubiquitous Computing[C]// ACM 2012 Article Poster Bibliometrics Data Bibliometrics, 2012: 657-658

[7] 李联宁. 基于手机平台的嗅觉网络传输研究与探索[J]. 计算机科学，2013，4(6A)：223-227

[8] Ki Sato. Prof. Shoji Takeuchi, Chemical Vapor Detection Using a Reconstituted Insect Olfactory Receptor Complext [J]. Angewandte Chemie International Edition, 2014, 53(44): 11798-11802

(上接第 117 页)

度；而 BN 方法是在忽略了电路扇出分支信号同步的情况下获得电路可靠度的。

4.2.3　基本门的敏感性比较

在电路设计的早期阶段，为了更好地指导可靠性分析与容错设计，在全加器、schneider 和 C17 上，分析 PTM 模型与 BN 模型计算结果对各基本门电路的敏感性并与 Monte Carlo 方法的模拟结果进行比较。实验结果如表 4 所列，其中，第 2 列至第 7 列数值指通过 PTM 模型与 BN 方法得到的敏感性单元与通过 Monte Carlo 方法得到敏感性单元的重合度。

表 4　不同方法计算结果比较

电路	PTM 模型			BN 方法		
	最敏感单元	前 2 敏感单元	前 3 敏感单元	最敏感单元	前 2 敏感单元	前 3 敏感单元
C17	100%	100%	100%	0	25%	20%
全加器	100%	100%	100%	0	0	50%
Schneider	100%	100%	100%	0	67%	40%

从表 4 可知，通过 PTM 模型得到的敏感基本门与 Monte Carlo 方法模拟结果一致，而通过 BN 方法得到的敏感基本门则与 Monte Carlo 方法模拟结果存在差异。一方面，与电路拓扑结构相关；另一方面，由两种方法的计算差异所导致。因此，精确的可靠性评估方法有利于在电路设计的早期阶段以较小代价改善设计以提高产品的可靠性与容错性。

结束语　集成电路的可靠性评估对于电路的高可靠设计至关重要。本文选取了目前最为典型的两种可靠性评估方法进行对比研究。通过定性比较和定量的实验分析可以看出，PTM 方法与 BN 方法原理相似，且均仅适用于小规模电路。PTM 方法可精确评估电路的可靠性，具有良好的完备性，而且模型简单准确，所以值得进行深入的研究，重点在于解决其时空复杂度过大、不能直接用于大电路的问题。BN 方法在精度上有一定程度的损失，但有较小的空间复杂度；PTM 方法可借鉴 BN 方法原理以改善其空间复杂度过大问题。

参 考 文 献

[1] Hennessy John L, Patterson David A. Computer architecture: a quantitative approach [M]. Elsevier, 2012

[2] 肖杰. 结合版图结构信息的门级电路可靠性评估方法的研究[D]. 上海：同济大学，2013

[3] Karthikeyan L, Sanjukta B. An error model to study the behavior of transient errors in sequential circuits[C]// Proceedings of the 22nd International Conference on VLSI Design, New Delhi. IEEE Computer Society: Los Alamitos, 2009, 485-490

[4] Thara R, Karthikeyan L, Sanjukta B. Probabilistic error modeling for nano-domain logic circuits [J]. IEEE Transactions on Very Large Scale Integration Systems, 2009, 17(1): 55-65

[5] Smita K, Viamontes George F, Markov Igor L, et al. Accurate reliability evaluation and enhancement via probabilistic transfer matrices[C]// Proceedings of the the Conference on Design Automation and Test in Europe, Munich. IEEE Computer Society: Los Alamitos, 2005, 282-287

[6] Xiao Jie, Jiang Jian-hui, Zhu Xu-guang, et al. A method of gate-level circuit reliability estimation based on iterative PTM model [C]// Proceedings of the IEEE 17th Pacific Rim International Symposium on Dependable Computing, Pasadena, USA . IEEE Computer Society: Los Alamitos, 2011, 276-277

[7] 肖杰，江建慧. 考虑时间因素的不同基本门故障概率计算[J]. 电子学报，2013，41(4)：666-673

[8] Yu Chien-Chih. Probabilistic Analysis for Modeling and Simulating Digital Circuits [D]. Michigan: The University of Michigan, 2012

[9] L Levin V. Probability analysis of combination systems and their reliability [J]. Engineering Cyberneitcs, 1996, 11(6): 78-84

[10] Michael B, Agrawal Vishwani D. Essentials of electronic testing for digital, memory, and mixed-signal VLSI circuits [M]. Springer, 2000