

第 1 篇 绪 论



测试智能信息处理概述

1.1 测试智能信息处理的产生及发展

1.1.1 测试系统的组成与特点

当前科技界普遍认为,信息技术由 4 大部分组成,即信息获取、信息传输、信息处理与信息应用。这 4 部分组成了一个如图 1.1 所示的信息链,信息链的源头是信息获取,属于仪器测试与检测技术的研究范畴。

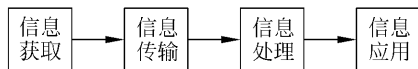


图 1.1 信息技术的 4 个组成部分及其信息链

仪器测试与检测技术的基本任务是研究信息获取技术及信息相关物理量的测量方法,并解决如何准确地获得和处理信息的问题,为被测信号(或数据)正确、可靠的传输提供必要的技术支持。同时针对信息获取、变送传输、数据处理和执行控制等部分的需要,研究在相关的信号产生、对象追踪、状态反馈、信息传送、动作控制、结果输出等技术环节中应用的控制技术与方法。

仪器测试与检测技术是一门工程应用技术,具有鲜明的时代性,其内涵随着科学技术的发展与时俱进。仪器已从单纯的机械结构、机电结构发展成为集传感器技术、计算机技术、电子技术、现代光学、精密机械等多种高新技术于一身的产品,其用途也从单纯的数据采集发展为集数据采集、信号传输、信号处理以及控制为一体的测控过程。进入 21 世纪以来,随着计算机网络技术、软件技术、微纳米技术的发展,仪器技术出现了智能化、虚拟化、远程化和微型化的发展趋势。

仪器科学与技术学科的结构内容与信息类学科(如自动化学科、电子信息学科)虽有重叠,但各自的研究目标和侧重点不同,“测量、控制、仪器”是描述仪器仪表类专业的关键词,也是区别于其他信息类专业的特征点。仪器科学与技术学科所进行的测量研究,主要从信息获取技术上掌握相关物理量的测量方法并解决如何准确获得信息的信号与数据处

理方法的问题,为被测信号(或数据)正确、可靠的传输提供必要的技术支持。仪器科学与技术学科所涉及的控制技术是针对信息获取、变送传输、数据处理和执行控制等部分的需要,研究在相关的信号产生、对象跟踪、状态反馈、信息传送、动作控制、结果输出等技术环节中应用的控制技术与方法;仪器测试技术则体现了该学科系统性、完整性、集成性的特征。

仪器测试技术的实质是信息获取、信息处理、信息利用的工具,而测试技术是研究以获取信息为目的的信息转换、处理、传输、存储、显示与应用等技术与应用科学。

著名科学家钱学森明确指出:“发展高新技术信息技术是关键,信息技术包括测量技术、计算机技术和通信技术。测量技术是关键和基础。”而测量技术则是仪器技术中的一项重要内容。

现代仪器仪表是对物质世界的信息进行测量与控制的基础手段和设备,因而美国商务部 1999 年的报告在关于新兴数字经济部分提出,信息产业包括计算机软硬件行业、通信设备制造及服务行业、仪器仪表测试行业。信息技术包括信息获取、信息处理、信息传输 3 部分内容。其中,信息的获取是靠仪器测试来实现的。仪器测试技术中的传感器、信号采集系统就是完成这一任务的具体器件。如果不能获取信息,或信息获取不准确,那么信息的存储、处理、传输都是毫无意义的。因而,信息获取是信息技术的基础,是信息处理、信息传输的前提。仪器测试是获取信息的工具,没有仪器测试,进入信息时代将是不可能的。因而,仪器技术是信息技术中信息获取—信息处理—信息传输的源头技术,也是信息技术中的关键技术。

一般说来,测试系统由传感器、中间变换装置和记录存储装置 3 部分组成,如图 1.2 所示。

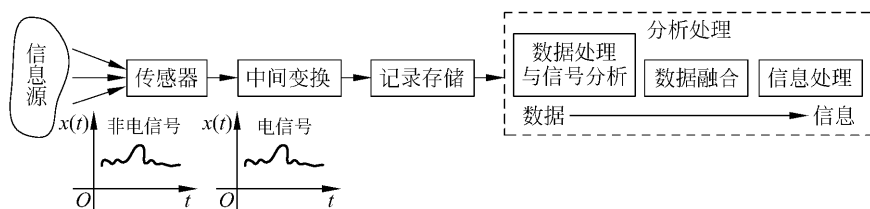


图 1.2 信息-信号的转换、传输与处理过程

测试系统中的记录存储部分主要以计算机为主体构成。若想对测量的数据进行处理,首先要进行信号分析,如通常采用快速傅里叶变换 FFT、频谱分析、小波分析等;测试系统要对来自多个传感器检测到的信号进行数据处理,需要进行数据融合完成对信号的深入分析,在此基础上采用智能计算方法进行信息处理,进而实现测试系统的最终测量目标。

测试是人类认识自然、掌握自然规律的实践途径之一,是科学研究中获得感性材料、接受自然信息的途径,是形成、发展和检验自然科学理论的实践基础。测试属于信息科学范畴,又被称为信息探测工程学。

信息(information),一般可理解为消息、情报或知识。例如,语言文字是社会信息,商品报导是经济信息,遗传密码是生物信息等。然而,从物理学观点出发来考察,信息是物质所固有的,是其客观存在或运动状态的特征。信息本身不是物质,不具有能量,但信息的传输却依靠物质能量。一般来说,传输信息的载体称为信号(signal),信息蕴涵于信号之中。

人类认识世界是以感官感知自然信息开始的。物质的颜色、形状、声响、温度变化,可以由人的视觉、听觉、触觉等器官感知,但人的感官感知事物的变化有局限性,人类感官的延伸——传感器,是近代信息探测工程学中的重要内容,传感技术的发展扩展了人类感知信息的智能。

信息探测涉及任何一项工程领域,无论是生物、海洋、气象、地质、雷达、通信以及机械、电子等工程,都离不开测试与信息处理。

按照信号变化的物理性质,可将信号分为非电信号与电信号。例如,随时间变化的力、位移、加速度等,为非电信号;而随时间变化的电压、电流、电荷等,则为电信号。电信号与非电信号可以比较方便地互相转换,因此,在工程中常常将各种非电物理量变换为电信号,以利于信息的传输、存储和处理。

工程中的信息处理,是指传感器第一次从敏感元件获得初始信息,直到用一定设备手段进行分析处理的过程,包括信息的获取、传输、转换、分析、变换、处理、检测、显示及应用等过程。通常又把研究信号的构成和特征值的过程称为信号分析;把信号再经过必要的加工变换,以期获得有用信息的过程称为信号处理。信号分析对信号本身的信息结构没有影响,而信号处理过程中,往往有可能对信号本身的信息结构有所改变。传感器是测试、控制系统中的信息敏感和检测部件,它直接感受被测信息并输出与其成一定比例关系的物理量(信号),以满足系统对信息传输、处理、记录、显示和控制的要求。

人们常常习惯于把传感器比作人的感官,把计算机比作人的大脑。因此,传感与计算机技术的发展促进了测试系统的智能化。从信息化角度出发,智能应体现在3个方面,即感知——信息的获取,思维——信息的处理,行为——信息的利用。

本章以下内容将简要介绍测试智能信息处理技术的发展过程和智能技术的相关技术,如神经计算、模糊计算、进化计算的概念、基本特点和应用,并对测试智能信息处理技术的应用等进行讨论。

1.1.2 智能计算的产生与发展

智能计算是测试智能信息处理的核心技术。20世纪90年代以来,在智能信息处理

研究的纵深发展过程中,人们特别关注精确处理与非精确处理的双重性,强调符号物理机制与连接机制的综合,倾向于冲破“物理学式”框架的“进化论”新路,一门称为智能计算(computational intelligence)的新学科分支被概括地提出并以更加明确的目标蓬勃发展。1994年IEEE为了促进多学科渗透和结合,把模糊系统(fuzzy systems)、神经网络(neural networks)和进化计算(evolutionary computation)3个年会合并举行,于1994年6月25日至7月3日在美国佛罗里达州(Florida)的奥兰多(Orlando)召开了全球第一届智能计算大会(WCCCI),出版了《智能计算、模仿生命》的论文集。大会决定,智能计算会议每3年召开一次。此次会议是智能计算的第一次综合性大会,共收集了来自世界各国学者的约1600篇论文,大会的主题是智能计算。

首次给出智能计算定义的是美国学者J. C. Bezdek。1992年,他在《国际近似推理杂志》(*International Journal of Approximate Reasoning*)上论道:智能计算依靠生产者提供的数字材料,而不是依赖于知识,而人工智能使用的是知识精华。Bezdek还说:人工神经网络应称为计算神经网络,即“人工”两字应改为“计算”。在人工智能AI和智能计算CI的关系上,Bezdek认为CI是AI的子集,即 $CI \in AI$ 。这次大会的主席J. M. Zurada却认为,CI和AI只有部分重合。J. C. Bezdek在题为“什么是智能计算”的报告中讲到:智能有3个层次,第一层是生物智能(biological intelligence, BI),它是由人脑的物理化学过程反映出来的,人脑是有机物,它是智能的物质基础。第二层是人工智能(artificial intelligence, AI),它是非生物的,是人造的,常用符号表示,AI的来源是人的知识精华和传感器数据。第三层是智能计算(computational intelligence, CI),它是由数学方法和计算机实现的,CI的来源是数值计算和传感器。以上三者第一个英文字符取出来称之为ABC。显然,从复杂性看有3个层次,即B(有机)、A(符号)、C(数值),而且BI包含了AI, AI又包含了CI。

按Bezdek的看法, AI是CI到BI的中间过渡,因为AI中除了计算算法外,还包含符号表示和数值信息处理。模糊集和模糊逻辑是CI到AI的平滑过渡,因为它包含了数值信息和语义信息。Bezdek还认为:计算神经网络CNN是一个最底层、最基本的环节,也是CI的一个重要基石,主要用于模式识别。CNN由以下4个点决定:功能、结构(连接拓扑和更新策略)、形式(集成和传递的结点函数式)、数据(用于训练/测试的数据)。按以上几点, CNN有多种形式,如前馈、自组织以及与Fuzzy结合的模糊神经网络等。

目前国际上提出,智能计算就是以人工神经网络为主导,与模糊逻辑系统、进化计算以及信号与信息处理学科的综合集成。新一代的智能计算信息处理技术应是神经网络、模糊系统、进化计算、混沌动力学、分形理论、小波变换、人工生命等交叉学科的综合集成。

尽管对智能计算的定义、内容以及与其他智能学科分支的关系尚没有统一的看法,但智能计算的下列两个重要特征却是人们比较共同的认识:

(1) 与传统人工智能不同,智能计算主要依赖于生产者提供的数字材料,而不是依赖

于知识；它主要借助数学计算方法（特别是与数值相联系的计算方法）的使用。这就是说，一方面，CI 的内容本身具有明显的数值计算信息处理特征；另一方面，CI 强调用计算的方法来研究和处理智能问题。需强调的是，CI 中计算的概念在内涵上已经加以拓广和加深。一般的，在解空间进行搜索的过程都被称为计算。

(2) 智能计算概念的提出，显然远不止于具有科学研究分类学的意义，其积极意义在于促进基于计算的或基于计算和基于符号物理相结合的各种智能理论、模型、方法的综合集成，以便在智能计算这个主题下发展思想更先进、功能更强大、能够解决更复杂问题的大系统的智能科学成果。

由此看来，当前智能计算发展的重要方向之一就是不断引进深入的数学理论和方法，以计算和集成作为学术指导思想，进行更高层次的综合集成研究。这种综合集成研究不仅不局限在模型及算法层次的综合集成的框框，而且还进入了感知层次及认知层次的综合集成。

一般来说，智能信息处理可以划分为两大类：一类为基于传统计算机的智能信息处理，另一类为基于神经计算的智能信息处理。基于传统计算机的智能信息处理系统包括智能仪器系统、自动跟踪监测仪器系统、自动控制制导系统、自动故障诊断系统等。在人工智能系统中，它们具有模仿或代替与人的思维有关的功能，通过逻辑符号处理系统的推理规则来实现自动诊断、问题求解以及专家系统的智能。这种智能实际上体现了人类的逻辑思维方式，主要应用串行工作程序按照一些推理规则一步一步进行计算和操作，目前应用领域很广。基于计算机（包括高速信号处理器开发系统）和人工智能的智能信息处理系统仍在继续向高新技术发展，但其发展速度已不太适应社会信息数量增长速度的需求，因而促使人们注意到新型智能信息处理系统的研究。

人工神经网络是模仿延伸人脑认知功能的新型智能信息处理系统。由于大脑是人的智能、思维、意识等一切高级活动的物质基础，构造具有脑智能的人工智能信息处理系统，可以解决传统方法所不能或难以解决的问题。以连接机制为基础的神经网络具有大量的并行性、巨量的互连性、存储的分布性、高度的非线性、高度的容错性、结构的可变性、计算的非精确性等特点，它是由大量的简单处理单元（人工神经元）广泛互连而成的一个具有自学习、自适应和自组织性的非线性动力系统，也是一个具有全新计算结构模型的智能信息处理系统。它可以模仿人脑处理不完整的、不准确的信息，甚至具有处理非常模糊的信息的能力。这种系统能联想记忆和从部分信息中获得全部信息。由于其非线性，当不同模式在模式特征空间的分界面极为复杂时，仍能进行分类和识别。由于其自适应、自学习功能，系统能从环境及输入中获取信息来自动修改网络结构及其连接强度，以适应各种需要而用于知识推广及知识分类。由于分布式存储和自组织性，使系统连接线即使被破坏了 50%，它仍能处在优化工作状态，这在军事电子系统设备中有着特别重要的意义。因此，基于神经计算的智能信息处理是模拟人类形象思维、联想记忆等高级精神活动的人工

智能信息处理系统。

近年来,以概率统计为基础的支持向量机理论和主分量分析方法已经迅速得到发展和应用,成为神经计算中一个崭新的研究热点和应用方法。

作为智能信息处理中重要组成部分的模糊逻辑及其模糊推理在得到迅速发展和应用的同时,粗糙集理论方法的出现为模糊计算提供了新的扩展空间和处理知识的方法。

进化计算作为智能信息处理中的另一个重要的研究领域,近年来的发展可谓更加迅速,除了在传统遗传算法上又有新的发展之外,在群智能的理论和方法方面也有所突破。特别是在粒群智能和蚁群智能的理论和方法方面正逐渐建立起新的理论和方法体系,展现了良好的发展前景和空间。

1.2 智能信息处理的主要技术

智能信息处理是一种多学科的交叉技术,迄今尚未建立起完整的理论体系。

1.2.1 神经计算技术

脑神经系统是以离子电流机构为基础,由神经细胞组成的非线性的(nonlinear)、适应的(adaptive)、并行的(parallel)和模拟的(analog)网络(network),简称 NAPAN。在脑神经系统中,信息的收集、处理和传送都在细胞上进行。各个细胞基本上只有兴奋与抑制两种状态。神经细胞的响应速度是毫秒级,比半导体器件要慢得多。神经细胞主要依靠网络的超并行性来实现高度的实时信息处理和信息表现的多样性。神经细胞上的突触机构具有很好的可塑性,这种可塑性使神经网络具有记忆和学习功能。突触结合的连接形成了自组织特性,并随学习而变化,使神经网络具有强大的自适应功能。

由于脑神经系统的复杂性,至今还没有可用于分析和设计 NAPAN 的理论。尽管人们早已经知道在人的大脑中存在着 NAPAN,但由于研究 NAPAN 的难度很大,而且电子计算机的功能已经十分强大,因而人们一直未能对它进行深入的研究。只有在开始注重到数字计算机的局限性的今天,人们才感到必须研究 NAPAN,希望能通过它实现崭新的超并行模拟计算机。

的确,由于神经科学的进展,从分子水平到细胞水平的详细构造和功能开始得到了了解,对脑神经系统所实现的信息处理的基本性质的理解也逐步深入。然而,即使细胞的结构以及生理的和物理的机理都弄清楚了,但对涉及 140 亿个神经细胞所组成的脑神经系统的超并行性、层次和分布式构造所形成的系统本质特性人们还知之甚少。目前,需要从系统论的立场出发来研究复杂的 NAPAN,在网络层次上弄清其功能和信息处理原理,确定使其体系化的理论。

在这种趋势中,对神经网络模型和学习算法的研究独具魅力。这种研究把许多简单

的神经细胞模型并行地、分层地相互结合成网络模型,提供了实现新的信息处理的一种手段,为建立 NAPAN 理论提供了一种途径。例如对于颜色等感觉信息,在脑内的表现是多重而广泛的多样体,只有采用新的信息处理方法才能对它进行描述。也就是说,原先靠生理心理实验是难以分析高级中枢中的信息处理行为的,但采用神经网络理论与技术却使之成为可能。这里所说的神经网络指的是人工神经网络,它是对真实脑神经系统构造和功能予以极端简化的模型。对神经网络的研究,有助于人类对 NAPAN 的理解,有助于探明大脑的信息处理方式,建立脑的模型,进一步弄清脑的并行信息处理的基本原则,并从应用角度寻求其工程实现的方法。

神经网络的主要特征是大规模的并行处理、分布式的信息存储、良好的自适应性和自组织性以及很强的学习功能、联想功能和容错功能。与冯·诺依曼计算机相比,神经网络的信息处理模式更加接近人脑,主要表现在以下几个方面:

(1) 能够处理连续的模拟信号(例如连续变换的图像信号)。

(2) 能够处理不精确的、不完全的模糊信息。

(3) 冯·诺依曼计算机给出的是精确解,神经网络给出的是次最优的逼近解。

(4) 神经网络并行分布工作,各组成部分同时参与运算;单个神经元的动作速度不快,但网络总体的处理速度极快。

(5) 神经网络具有鲁棒性,即信息分布于整个网络各个权重变换之中,某些单元的障碍不会影响网络的整体信息处理功能。

(6) 神经网络具有较好的容错性,即在只有部分输入条件,甚至包含了错误输入条件的情况下,网络也能给出正确的解。

(7) 神经网络在处理自然语言理解、图像识别、智能机器人控制等疑难问题方面具有独到的优势。

神经网络的上述特点是否反映了人脑神经网络的所有特征和功能呢?事实上,神经网络以连接主义为基础,是人工智能研究领域的一个分支。它从微观出发,认为符号是不存在的,认知的基本元素就是神经细胞。认知过程是大量神经细胞的连接引起神经细胞不同兴奋状态和系统表现出的总体行为。传统的符号主义与其不同。符号主义认为,认知的基本元素是符号,认知过程是对符号表示的运算。人类的语言、文字、思维均可用符号来描述,而且思维过程只不过是这些符号的存储、变换和输入、输出而已。以这种方法实现的系统具有串行、线性、准确、易于表达的特点,体现了逻辑思维的基本特性。20 世纪 70 年代的专家系统和 80 年代日本的第五代计算机研制计划就体现了典型的符号主义思想。

现代研究表明,基于符号主义的传统人工智能和基于连接主义的神经网络分别描述了人脑左、右半球的功能,反映了人类智能的两重性:一方面是精确处理,另一方面是非精确处理,分别对应认知过程的理性和感性两个方面。两者的关系是互补的,不能相互替

代。理想的智能系统及其表现的智能行为应是两者相互结合的结果。

主元分析(principal component analysis, PCA)就是这样的一种降维技术,是神经计算中近些年来发展的一种方法。通过把数据投影到能够准确表征过程状态的低维空间,降维技术可以简化和改进过程监控过程。它以某种方式产生低维表示,这种方式保留了过程变量间的关系结构,按获取数据的变化度来说是最优的。

作为检测工业过程的一种降维工具,PCA 的应用在学术界和工业界都进行了研究。DuPont 和其他公司已经开展了 PCA 对工厂数据的应用,其大部分成果也已在会议论文集和期刊杂志中发表。有些研究人员根据计算机过程仿真得到的数据进行了类似的研究。对于某些应用,数据的大部分变化都可以收集到二维或三维中,并且过程的变化可以用单张图形象地显示出来。单图显示(one-plot visualization)和从多维数据中提取的结构,都有助于操作员和工程师解释过程数据的显著变化趋势。

PCA 可以产生数据的低维表示,与用全维观测空间相比,可以更好地推广到独立于训练集的数据,从而提高检测和诊断故障的效率。无论是产生被测系统故障的变量,还是严重受故障影响的变量,用 PCA 提取的结构都可以有效地对它们进行识别。PCA 能够把观测空间分离成一个获取过程的系统趋势的子空间与一个基本上包含随机噪声的子空间。因为某些被测量故障主要影响两个子空间中的一个,这已被人们广泛接受,所以可以用一种方法针对一个子空间,用另一种方法针对另一个子空间,这可以从总体上提高准确性。

支持向量机(support vector machine, SVM)是建立在统计学习理论基础之上的一种新型的神经网络。统计学习理论是针对小样本情况研究统计学习规律的理论,是传统统计学的重要发展和补充,为研究有限样本情况下机器学习的理论和方法提供了理论框架,其核心思想是通过控制学习机器的容量实现对推广能力的控制。在这一理论中发展出的支持向量机方法是一种新的通用学习机器,较以往方法表现出很多理论和实践上的优势。

统计学在解决机器学习问题中起着基础性的作用。传统的统计学所研究的主要是渐近理论,即当样本趋向于无穷多时的统计性质。在现实的问题中,所面对的样本数目通常是有限的,有时还十分有限。虽然人们实际上一直知道这一点,但传统上仍以样本数目无穷多为假设来推导各种算法,希望这样得到的算法在样本较少时也能有较好的表现。然而,相反的情况是很容易出现的。近年来经常听到的所谓神经网络过学习问题就是一个典型的代表:当样本数有限时,本来很不错的一个学习机器却可能表现出很差的推广能力。

在人类即将迈进一个新世纪的时候,人们开始逐渐频繁地接触到一个词,就是统计学习理论。它是早在 20 世纪 70 年代就已经建立了其基本体系的一门理论,系统地研究了机器学习的问题,尤其是有限样本情况下的统计学习问题。在 90 年代,这一理论框架下产生出了支持向量机(SVM)这一新的通用机器学习方法。由于统计学习理论为人们系