

生物体的深层次信息提取的研究

欧阳楷

(首都医科大学生物医学工程系,北京 100054)

1 深层次信息提取的重要意义

在各种过程中蕴含着大量的信息要通过对信号进行各种各样的处理方能获得,例如心电信号蕴含着心血管活动的大量信息,地震信号中蕴含着地下各种结构的大量信息。这就是信息科学中常说的信息是蕴含在信号之中,信号是信息的载体的含义。现在人们已经能够把表面的、易于获得的信息从信号中提出,现今是要研究如何提取深层次的难于提取的信息。例如从ECG提出心脏内部的工作状况、疾病的位置和性质的信息,从地质勘探信号中准确地提出地层中含油的地层的深度、形状及性质的信息,等等。这些信息的提取是比较困难的,需要特殊的技术和方法,我们称之为深层次信息的提取。深层次信息的提取存在巨大的社会和经济意义,例如,从雷达回波中提取飞机类型的信息就可以对敌我识别作准确的判断,准确地判断地下的含油层,准确地诊断某些疾病以提高医疗水平,等等。所以,各国对于这方面的技术的发展都是不遗余力,竞争是激烈的,有人甚至提出:谁掌握了深层次信息提取的技术,谁就占领了21世纪的制高点。

2 “提取深层次信息”的含义

“深层次信息”的第一个含义是在复杂背景下提取特定的信息。例如强干扰下的动态信息提取;干扰与信息的关系不是加法性的而是乘法性的,甚至是卷积性情况下的动态信息提取;信息不完整、有缺损甚至有矛盾时的信息提取,这些都称为深层次的信息提取。它的第二个含义是无损的提取内部的、动态的信息。例如从体表上收集到的各种信号中提取生物体内各器官的信息、内部各组织的信息、以至细胞层次和分子层次的信息,而且这些信息应该是动态的(即是生命过程的信息)。由于这种信息的定量提取是非常困难的,因此信息的收集就要空间域、时域、频域多维并举。特别是,不同的信息载体有不同的特性,应取长补短,提取的方法也要多种方法并举,其中特别强调同胚模型(Homemorphic model)的方法。第三个含义是强调系统。不但注意到生物系统是由各功能子系统组成的,而且要注意到各子系统是有多个层次的。因此深层次信息的提取既要顾及各系统、各层次内的信息,又要顾及各系统各层次之间的相互联系。在这样的思想的指导下提取的信息就不只是各种信息的堆积,而且可以发现新的知识、新的规律,否则也就不会有真正深层次信息的提取。

综上所述,深层次信息的提取是一个内涵丰富、有待开拓的领域,随着时间的推移,深层次信息的提取也可以有不同的含义,在当前它确实是一个激励我们不断进取的领域。

3 深层次信息提取的方法

现代信息科学中各种深层次信息提取方法大体上分为三类:(1)基于人类认识事物的方法。人类在信息提取方面常常是异乎寻常的,有的基于已有经验,有的基于联想,还有许多是直觉的,甚至是灵感的。它们常常是没有很多根据,说不出多少道理,但确实有效。把这些方法连

同基于符号逻辑的人工智能嵌入到传统的信息提取技术中,是会改进现有技术的效能的。(2)基于数学方法:把现代数学应用于深层次信息提取是很有效的,例如综合利用空间域、时域、频域多维信息的方法,包括正问题与逆问题;又如目前流行的 Wavelet 方法,微分几何中的同伦论方法,特别是非线性动力过程中的混沌,非线性科学中的高阶方法,优化计算中的非最小二乘法,多维泛函指标法等。由于这些方法用到的数学既广又深,有些内容是现代数学正在研究或尚未解决的问题。因此对于研究人员的素质要求极高,这也是此方向的困难之一。(3)基于生物系统信息加工机制的方法。这是目前深层次信息提取最令人注目也是效果最突出的方法。例如应用极广的人工神经网络方法,优化计算中的遗传算法等。其他如利用视觉中的广义 Garbor 函数来做指纹识别,利用操作条件反射原则来做高精度自适应控制(如倒立摆的控制),特别是将生物知识和已知生物信息加工机制归纳而形成的同胚模型,反过来用于生物中深层次的信息提取等等,方法很新而且也都是很有效的。究其成功的理由也还是因为这些信息加工的机理是千万年进化所筛选出来的,自有其优美之所在,用得恰当,可以认为成功是必然的。

4 对生命科学的推动

在生物系统中,深层次提取获得成功将对生命科学以很大的推动。这里仅举几例:

(1)60 年代,AD 报告发表了美国军方对瞬时心率影响因素的报告(温度、紧张、交感神经活动),近年来,从心率变异性来了解神经体液对心率调控的兴趣增长。显然,瞬时心率含有植物神经活动的大量消息(交感神经、副交感神经、迷走神经)。如能对此深层次信息加以提取,不但开发了了解神经的另一窗口(加上希氏束电位所提取的神经体液调控机理),还将对心血管疾病的诊断和治疗大有好处。

(2)从头皮上拾取的脑电信号,包括自发脑电信号(EEG)和诱发脑电信号(EP),显然隐含着大量脑工作的信息。然而,数十年来虽然人们对此进行了很大的努力,也取得了部分的成功,但过高的期望始终困惑着这种深层次的信息提取,希望在方法学上取得突破,如加上生理解剖知识、计算机处理技术的改进、以及宏观脑模型的建立等等,就有可能提取更深层次中的脑的工作机理,它的应用及学术意义是不可限量的。

(3)人们都说“眼睛是心灵的窗户”。从现代的科学研究中已经证实,有关眼的信息处理关系是极端复杂的,仅从眼动这一信号来看,它受众多的信号影响,从眼动信号中可以观察到脑中许多部分活动的痕迹(小脑、脑干、以至大脑)。利用计算神经科学中的同胚模型方法,目前已可以对这些活动进行研究,提取到许多宝贵的信息,为早期神经系统准确的诊断打下了基础,近来还有人用眼动信号提取人类对文字进行处理的信息,发现了汉字识别的某些优点,中文拉丁化的提法就要慎重地考虑。

(4)超声波在临床上的意义是众所周知的,然而其深层次的信息仍有待提取,例如提取组织和肌肉的弹性分布等可以开阔一个全新的超声波定诊领域,这些目前都尚未涉及。

5 向识别方法进行拓展

“基于嗅觉的模式识别新方法”是一个典型的从生物信息加工机制发展出的一个大有前途的方向。1947 年,Piffs McCulloch 从蛙的顶盖解剖与生理研究中发现蛙的顶盖中枢可分为四区,分别敏感于四个特征:(1)反差变化(2)凸边快速运动(3)边缘(4)运动图像边缘变暗。而这四个特征联合起来正好就是飞着的昆虫,它也正好就是蛙的食物,当蛙看到这四个特征就跳起

来将其吞下,所以蛙的识别就是特征识别。这个结论还可以从两个方面获得证实:一是蛙边有不动的苍蝇,蛙也不会有任何动作;二是可以利用跳着的物件在水边钓蛙。

从以上试验获得一个概念即任何事物只要找到相应的特征(特征提取),就能用一定的方法识别(模式识别)。这种方法已形成很大的领域,取得了很大的成就。但是,什么是特征?这是没有先验知识的。因此癌细胞的特征提取数百个,其识别率依然不能如意,那么是否存在着的识别方法呢?蛙是两栖动物,进化程度不高,较为高级的动物是否存在更高级的识别原理呢?Freeman 在兔的嗅觉的研究中,搞清了它们的解剖结构,即嗅神经的神经元的相互连接关系,并加以适当简化构成 128×14 二阶微分方程组,用计算机解此系统获得以下概念:(1)设有输入的系统处于混沌状态,而且于初始条件无关。(不是状态轨迹而是相空间的混沌域与初始条件无关。)(2)有某种输入时系统处于一种较为复杂的极限环,输入除去,又恢复到混沌,另一种输入则系统处于另一种复杂的极限环(不是简单的数学上的极限环)。

由此看出,高级动物的识别过程不全是特征提取与特征分类的过程,而可能是另一种机制。由此启发我们可以发展出一种新的模式识别方法,也许它的识别机制在精确大量识别方面有其特有的优点。当然,这方面的工作刚刚开始,大量的工作有待进行,成败尚难预料,此处只是说明基于生物系统信息加工机制的新的工程技术的可能性。有了新的方法就能提取到别的方法所不能提取的信息。其它的生物系统机制如小脑的运动控制,生物中的突触可塑性机制的研究都有可能提出新的信息加工机制,也可提取到新的深层次的信息。

李国杰在谈到智能计算机的基础研究时,提出新的计算模型将有四个发展源泉:第一是基于心理学。因为心理学是研究高级思维活动规律,从这里有可能提出全新的计算模型概念,潘云鹤基于联想的形象思维机制的研究就是一个值得注意的方向;第二是基于社会学。分布人工智能系统(DAI)曾大量借用社会学的概念;第三是基于生命系统的演化原理。最近这方面的研究很活跃,如基于生物遗传过程的遗传算法等。生物进化理论、生态系统所揭示的一些特定基理都很值得注意;第四是基于生理学的,也就是前面所提到的那些例子。上述四个观点,都是与生物有关的,都可概括为基于生物的,也就是说,从生物中的深层次信息提取所获得的成果中有可能发展出新型计算机的计算模型。

这只是基于生物的一个方面,如能结合前述的基于数学,基于人类知识的各种新方法就能构成新一代的信息科学。

本文是在吕维雪教授的启发下写成的,并在与杨福生教授、白净教授多次探讨中获益很多,仅致谢意。

A STUDY OF THE DEEPLY IMPLICIT INFORMATION IN ORGANISM

Ouyang Kai

(Department of Biomedical Engineering, Capital University of Medical Sciences, Beijing 100054, China)