

信号、数据和信息的历史及展望

——纪念《数据采集与处理》创刊 30 周年

黄德¹ 张弓¹ 谭晓阳² 刘彦东³ 黄志球²

(1. 南京航空航天大学电子信息工程学院, 南京, 210016; 2. 南京航空航天大学计算机科学与技术学院, 南京, 210016; 3. 《数据采集与处理》编辑部, 南京, 210016)

摘要: 信号、数据和信息相关理论及其应用是《数据采集与处理》杂志 30 年来主要的刊登内容。从早期的 A/D 转换到 DSP, FPGA 处理芯片, 从分形技术、小波理论到语音信号识别和图像信号处理, 从神经网络到深度学习, 从无线通信技术到光通信技术, 《数据采集与处理》杂志就像一面镜子, 折射出信号、数据和信息处理技术的飞速发展。本文对信号、数据和信息处理的历史、发展状况进行了梳理, 并对其未来发展方向进行了展望。本文的结论将是《数据采集与处理》未来刊登内容的重要遵循依据。

关键词: 信号; 数据; 信息; 科技期刊

中图分类号: TP391 **文献标志码:** A

History and Outlook of Signals, Data and Information

——The 30th Anniversary of Journal of Data Acquisition and Processing

Ben De¹, Zhang Gong¹, Tan Xiaoyang², Liu Yandong³, Huang Zhiqiu²

(1. College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China; 2. College of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China; 3. Editorial Department of Journal of Data Acquisition and Processing, Nanjing, 210016, China)

Abstract: The DAPJ (Journal of Data Acquisition and Processing) has focused on publishing papers on the theory and applications of signal, data and information for 30 years, ranging from A/D converter in the early days to modern DSP and FPGA processing chip, from fractal technology and theory of wavelet to speech recognition and image processing, from neural network to deep learning, and from wireless communication to optical communication. The DAPJ, like a mirror, always reflects the latest development in the field of signal, data and information processing. This paper makes a brief summary about the history and current research of signal, data and information processing, and looks forward to the future developments of DAPJ, which will serve as the guidelines for the contents to be published on this journal.

Key words: signal; data; information; scientific journal

引 言

自从香农确定信息理论的基础以来, 人们把信息定义为事件发生前后所掌握的熵的变化, 这是一个

与事件的概率大小有关的量,于是信息可以被计量。人们可以具体地研究一个信号究竟包含多少信息、能够携带多少信息等。然后又发现了信号与信息之间的一个最基本的关系,那就是携带一定量的信息,需要一定的信号功率、时间和频带宽度。当利用信号携带信息时,存在一些最基本的规律,从而可能从根本上理解、解释、优化使用信号的过程^[1]。数据是对事实、概念或指令的一种表达形式,可由人工或自动化装置进行处理。数据经过解释并赋予一定的意义之后,便成为信息。因此,信号和数据都是获取信息的来源。但是,由于某些历史原因,信号和数据被分别进行研究。一般而言,信号及信号处理以更深入地了解信号所代表的物理对象的性质为目标,因而和电子科学与技术、自动控制等学科密切相关;而数据及数据处理以建立适当的模型来刻画数据内在的规律为主要目的,因而和数理统计、人工智能、机器学习等学科有密切联系。

信息作为构成现实世界的基本要素之一,是指音讯、消息、通讯系统传输和处理的对象、泛指人类社会传播的一切内容。信息资源区别于物资、能源资源有一个显著的特点,它可以被重复使用,可被共享,在使用过程中不仅可以不减少,并且还可以产生新的增量^[3]。信息技术是研究信息产生、采集、存储、变换、传递以及处理。信息处理技术的发展及其应用的广度和深度,极大地影响着人类社会发展的进程。

本文对信号、数据和信息处理的历史、发展状况进行了梳理,并对其未来发展方向进行了展望。本文的结论将是《数据采集与处理》未来办刊方向的重要参考。

1 数据采集系统的历史与现状

将模拟信号转换为数字信号,并进行存储和计算机处理显示的过程称为数据采集,而相应的系统则为数据采集系统。数据采集技术是信息科学的一个重要分支,它研究信息数据的采集、存储、处理及控制等内容。它是研究瞬间物理过程的有力工具,也是后期进行数据处理的重要基础。

数据采集系统始于20世纪50年代,1956年美国首先研究了用在军事上的测试系统。大约在60年代后期,国外有成套的数据采集设备产品进入市场。20世纪70年代中后期,随着微型机的发展,诞生了采集器、仪表和计算机溶于一体的数据采集系统。20世纪90年代至今,数据采集技术已经在军事、航空电子设备及宇航技术,工业等领域被广泛应用。由于集成电路制造技术的不断提高,出现了高性能、高可靠性的单片数据采集系统(Data acquisition system, DAS)。目前DAS产品精度已经达到了16位,采样率达到近10 GS/s。基于标准总线 and 高速DSP的超高速数据采集系统对满足不断发展的雷达、通信、电子对抗、航天测量、图像以及多媒体等领域广泛的应用需求具有重要意义。

随着科技的发展和数据采集系统的广泛应用,对数据采集系统的许多技术指标,如采样率、分辨率、存储深度、数字信号处理速度以及抗干扰能力等方面提出了越来越高的要求。数据采集处理设备正向着小型化、智能化和一体化方向发展,特别是随着移动技术、云计算、光纤传感以及超高速全光采样等新技术的引入,数据采集领域将面临着进一步的革新。

2 信号及信号处理

信号是运载信息的工具,是信息的载体。信号是表示信息的物理量,如电信号可以通过幅度、频率、相位的变化来表示不同的信息。这种电信号有模拟信号和数字信号两类。从广义上讲,信号包含光信号、声信号和电信号等。按照实际用途区分,信号包括电视信号、广播信号、雷达信号、通信信号等;按照所具有的时间特性区分,则有确定性信号和随机性信号等。

当前的信号在传统物理表现形式的基础上,出现了丰富多彩的信号形式,例如音频信号、视频信号、语言信号、图像信号、多媒体信号、传感器信号、通信信号、地理信号、声纳信号、雷达信号、化学信号、分子信号、基因信号、生物信号、医学信号以及音乐信号等。从信号带宽来看,信号由窄带信号扩展到宽带信号;从维度来看,信号由一维信号发展到高维信号,医疗、多光谱遥感图像等高维信号,具有大数据量。从分辨率来看,信号从低分辨率信号发展为高分辨率信号,视频产品的分辨率和帧频都在不断升高;

从采集方式来看,信号采集从单一传感器到信号采集网络化^[2]。

文献[3]指出,典型信息过程模型的含义是:认识主体通过获取信息和传递信息来把对象的本体论信息转变为第一类认识论信息,通过处理信息来深入认识系统的运动状态和方式,形成知识,达到认知;然后在此基础上“再生”第二类认识论意义的信息(策略信息),后者就是指明如何把系统由初始状态转变到目的状态的控制策略;控制的作用则是执行策略信息,产生控制行为,引导系统达到预定的目的状态,完成主体对对象施行的变革。这与人类通过自己的信息器官(感觉器官、神经系统、思维器官、效应器官)认识世界和改造世界这一活动的信息模型完全一致,它把信息科学的研究和人类自身的信息过程天然地联系起来,使人类自身信息过程和信息科学研究两者交相辉映、相得益彰。这也预示着信息技术的发展趋势正朝着从传统的对拥有固定描述的信号进行处理向对多维度信息恰当描述的探索。

2.1 雷达信号处理的发展

雷达信号处理作为信号处理领域的典型研究方向,它的发展变化可以在一定程度上反映出信号处理研究的发展轨迹。20世纪60年代,雷达信号处理开始应用数字技术和采用新型模拟器件(如表面声波器件),使其性能得到较大的提高,其中最具有代表性的成就是20世纪70年代初研制成功的动目标检测处理器。然而,早期的雷达信号处理器仍然以硬件构成为主,这样的结构必然导致设备量大,灵活性差等问题的出现。另一方面,由于数字器件速度不高,早期雷达信号处理在应用上受到很大的制约。70年代后期至80年代中期,超大规模集成电路,特别是超高速集成电路的迅速发展,又为雷达数字信号处理的进一步发展开拓了新的途径。

进入21世纪以来,雷达新体制的研究步入了蓬勃发展期,基于相控阵雷达思想的多输入-多输出雷达以及利用感知环境实现智能信息处理的认知雷达,都给雷达信号处理领域带来了新的机遇与挑战。同时,信号的检测方法也趋于信息化与智能化。在探索信号检测新方法的同时,学者们开始致力于雷达体制与信息融合的新研究,进入将信号处理向信息处理发展的新阶段。在1996年与1999年,何友院士分别发表了“多传感器数据融合算法综述”与“两级混合多传感器信息融合中的状态估计”2篇论文^[4-5],阐述了多传感器数据融合的特点、分布检测、空间、属性融合、态势评估和威胁估计算法,提出了信息融合的新方法,对日后的数据融合技术研究和选题方向提供了重要的参考与指导。何友院士的研究轨迹恰好可以在一定程度上反映了雷达信号处理领域研究的正在由低维信号处理向高维信息处理方向演进,向更加智能化的方向发展。

2.2 通信技术的发展

在通信研究领域,尤其在无线通信中也存在着类似的变化,正由传统无线通信向着智能无线电技术方向发展。软件无线电、认知无线电等智能无线电技术逐渐成为通信领域关注的热点,并给无线通信带来新的发展空间。

软件无线电可以使整个系统(包括用户终端和网络)采用动态的软件编程对设备特性进行重配置,通过软件定义来完成不同的功能。认知无线电是一种具有频谱感知能力的智能化软件无线电,由瑞典 Joseph Mitola 博士于1999年8月提出,它可以自动感知周围的电磁环境,与通信网络进行智能交流,寻找“频谱空穴”,将通信双方的信号参数实时地调整到与规则相适应、与环境相匹配最佳状态,达到通信系统的高可靠性以及频谱利用的高效性。认知无线电的架构设计原则是将软件无线电、传感器、感知和自主机器学习融合在一起,利用在射频端和用户域中的观察(传感、感知)、导向、规划、决策、行动和学习能力来提供更好的信息质量,并且利用软件无线电、传感器、感知和自主机器学习集成在一起创造意识、自适应和认知无线电,在射频和用户域完成从简单的感知或自适应变换为确定的认知无线电^[6]。

进入新世纪,3G/4G技术、WiFi/WiMAX技术与光通信迎来了飞速发展,信息的载体进一步多样化。3G/4G技术的发展趋于单模、双模和多模终端共存,对终端的要求不断提高,同时具有明显的智能化、多媒体化的趋势。如果说3G/4G技术赋予智能手机以羽翼,那么WiFi/WiMAX技术则是给予个人电脑以慧眼。WiMAX正逐步由一个宽带移动无线接入标准演变为一个准移动通信标准。光通信研究

的重点也从大容量、超高速转变为智能化、自动化和边缘化。

从上述雷达信号处理、通信技术等典型的信号处理与传输发展趋势可以看出,信号作为对物理世界描述的初级形式正在向多维、多谱、多分辨率、多源融合方向发展,伴随着信息量的爆炸式增长,并与各种智能技术相结合,呈现出显著的向高级形式方向发展的趋势。

3 数据及数据处理

数据是有意义的实体,它涉及到事物的存在形式,是人们通过观察、实验或计算得到的结果。数据可分为模拟数据和数字数据两大类,其表现形式是数字、文字、图像、声音和视频等。数据处理是对数据的采集、存储、检索、加工、变换和传输,其基本目的是从大量的、杂乱无章和难以理解的数据中抽取并推导出有价值、有意义的数据。

随着数据采集器能力的提高,相较于以前,人们已经能获得更加完整、多样、大量的数据。这类数据具有如下特点:(1)数据规模大(从 TB 扩大到 ZB, PB)。由于数据加工处理技术的提高,网络带宽的成倍增加,智能移动设备的持有量急剧增长以及社交网络技术的迅速发展,数据的产生量和存储量呈现出成倍增长的趋势;(2)数据移动性强。随着移动智能设备的增长,人们对数据的实时应用需求更加普遍,比如通过智能移动终端关注天气、物流、新闻、交通等信息;(3)数据的多样性高。随着网络技术的发展,各种各样的设备通过网络连成一个整体,人们除了可以获取到简单的文本外,还可以得到传感器数据、音频、视频、日志文件、点击流以及其他任何可用的信息;(4)数据价值的稀疏性大。人们需要从海量不确定和多样的数据流中获取对自己有帮助的信息。

现代的数据因为具有规模大、移动性强、多样性高和价值更稀疏等特点,常被称为大数据。大数据一方面给传统数据处理技术带来更大挑战,另一方面也有效促进了现代数据处理技术的发展。目前,在产业界已经有大批成功应用了人工智能技术进行大数据处理的智能系统^[7-9],涉及的领域有语音识别、工业机器人、物流、银行业软件、医疗诊断和搜索引擎等。以下以自然语言处理、图像处理和语音识别为例来加以说明。

自然语言处理是计算机科学领域与人工智能领域中一个重要的研究方向,它以实现人与计算机之间用自然语言进行有效通信为目标。最近,基于神经网络的深度学习方法已经在诸如机器翻译、段落检测和情感分析等自然语言处理应用方面展现出非常好的性能。比如文献[10]在一个大型文集库上利用递归神经网络来预测句子的情感,将其分类精度由 80% 提高到了 85.4%,在精细划分的短语情感标签预测方面,则将 bag-of-features 的方法提高了 9.7%。大量训练数据带来的一个挑战是处理效率问题。文献[11]成功地利用图形处理器来帮助提高训练效率,使其能够在包含五千万个词汇的单词库上训练神经网络模型,完成机器翻译的任务。

在图像处理方面,大数据也有效促进了机器的图像理解,并使得研究者们能够设计更为复杂的模型来研究图像中各目标的性质及其相互关系,并得出对图像内容含义的理解以及对原客观场景的解释。例如,著名的 ImageNet 数据集包含了 120 万张高分辨率的图像,类别数多达 1 000 类。而研究者们利用大数据和复杂的机器学习模型,已经能够在这个数据集上达到超过人类的分类精度^[12]。

语音识别技术就是让机器通过识别和理解过程把语音信号转变为相应的文本或命令的技术。近年来,特别是 2009 年以来,借助机器学习领域深度学习研究的发展以及大数据语料的积累,语音识别技术得到突飞猛进的发展。在此方面,微软公司的研究人员率先取得了突破性进展,他们使用深层神经网络模型后,语音识别错误率降低了 30%,是近 20 年来语音识别技术方面最快的进步。目前人们已经能够使用长达 10 万小时的大规模语音数据训练集和规模庞大的基于图形处理器的深度学习基础设施来训练语义识别模型。

总之,大数据不断挑战和促进着处理数据的能力,带来了机器学习等相关科技技术的蓬勃发展。人们现在已经能从大数据中自动筛选、分类和关联出有一定价值的信息。今后,人们将更加依赖能够获得

隐藏在现有庞大的数据和不断增长的数据中的信息,并对其进行深度挖掘。

4 结束语

互联网上有海量的文字、图像、语音和视频等数据需要处理,人们可以借助神经网络这一强有力的引擎和如此丰富大量的数据燃料通过深度学习的方式帮助计算机理解所有的数据。从数据缺乏转变为数据丰富,大数据背景下的信号处理,需要研究4个方面的关键问题:(1)感知压缩,以降低信号的数据量;(2)多传感器信息融合,以处理复杂多源信息;(3)智能传感网技术,以提取有价值信号;(4)高速信号处理,以提高信号处理速度。可以想象,人们很有可能借助它来改变身边所有其他技术的交流方式,也可以带来其他领域如金融、医疗保健和教育这些方面的变化,这将给互联网和人们的生活带来变革性的影响。

近30年来,随着集成电路、计算机和通信技术的飞速发展,信息技术已经成为最活跃、与人们生活最密切相关的科技领域,并且正日益改变着人类的生活方式。《数据采集与处理》杂志创刊30年来,经历了信息技术从最初关注的传感器、A/D变换到信号处理、通信技术,再到数据分析、数据挖掘、数据融合、人工智能等重要发展阶段,见证了信息技术的快速发展和信息领域学术研究的繁荣振兴。《数据采集与处理》将把握信息处理技术发展动态,搭建更好的学术交流平台,努力为信息领域科学研究发挥更加重要的作用。

参考文献:

- [1] 胡来招. 信号与信息[M]. 北京:电子工业出版社,2010.
- [2] 谢维信,陈曾平,裴继红,等. 大数据背景下的信号处理[J]. 中国科学信息科学,2013,43(12): 1525-1546.
Xie Weixin, Chen Zengping, Pei Jihong, et al. Signal processing in the context of big data[J]. Scientia SINICA Informatio- nis, 2013,43(12): 1525-1546.
- [3] 钟义信. 信息科学与技术导论[M]. 北京:北京邮电大学出版社,2007.
- [4] 何友,陆大绘,彭应宁. 多传感器数据融合算法综述[J]. 火力与指挥控制,1996,21(1):12-21.
He You, Lu Dajin, Peng Yingning. A survey of multisensor data fusion algorithms[J]. Fire Control & Command Control, 1996,21(1):12-21.
- [5] 何友,陆大绘,彭应宁. 两级混合多传感器信息融合中的状态估计[J]. 电子科学学刊,1999,21(5):698-701.
He You, Lu Dajin, Peng Yingning. State estimation for two level hybrid multisensor data fusion systems[J]. Journal of Elec- tronics, 1999,21(5):698-701.
- [6] 阎毅,贺鹏飞. 软件无线电与认知无线电概论[M]. 北京:电子工业出版社,2013.
- [7] Stuart Russell, Peter Norvig. Artificial intelligence: A modern approach[M]. Upper Saddle River, New Hersey: Prentice Hall,2003.
- [8] Committee on Innovations in Computing and Communications; Lessons from History, National Research Council. Funding a revolution: Government support for computing research[M]. Washington D C: National Academy Press,1999.
- [9] Don Corey. Are you talking to me? [M]. Douglas County, Colorado, United States: Outskirts Press,2007.
- [10] Socher R, Perelygin A, Wu J Y, et al. Recursive deep models for semantic compositionality over a sentiment treebank[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). Seattle, Washington, USA: ACL,2013, 1631-1642.
- [11] Schwenk H. Continuous space translation models for phrase-based statistical machine translation[C]// Proceedings of Com- putational Linguistics. Mumbai, India :ACM, 2012: 1071-1080.
- [12] Olga Russakovsky, Jia Deng, Hao Su, et al. ImageNet large scale visual recognition challenge[EB/OL]. [http://www. im- age-net. org](http://www.image-net.org),arXiv:1409. 0575, 2014-12-01.

作者简介:贲 德(1938-),男,院士,研究方向:新体制雷达,E-mail:bende@sohu.com;张 弓(1964-),男,教授,博士生导师,研究 方向:雷达信号处理、目标探测与识别、图象分析与处理;谭晓阳(1972-),男,教授,博士生导师,研究方向:机器学习、模 式识别和人脸识别;刘彦东(1970-),男,编辑,研究方向:计算机应用和编辑学;黄志球(1965-),男,教授,博士生导师,研 究方向:软件工程和数据工程。

