

# 深度学习中的数字信号处理问题

靳嘉诚<sup>1)</sup>

<sup>1)</sup>(自动化 2104 班, 西安交通大学, 西安, 中国, 715100)

**摘要** 随着时代的进步, 信号处理越显重要, 从图像、语音信号的识别到无人驾驶, 技术不断发展, 由其发展带来的产品迭代与进化也愈发便利人们的生活, 本文将结合不同场景下的实例, 讨论深度学习中的数字信号处理问题。深度学习在生活中日益应用, 遍历着人们的生活, 在 OpenAI 公司发布 ChatGPT 等产品后, 深度学习、人工智能等词语越来越被人们提及。而对于深度学习涉及到的一些技术, 如自动驾驶、图形识别、语音识别等, 更是和信号处理无法分开。

**关键词** 深度学习 数字信号处理 无监督学习 监督学习

## 1 数字信号处理

### 1.1 概念简述

信号是各种自然或人工系统所携带或产生的各类信息表现的一种基本形式。我们人类被各种各样的信号所包围, 它们来自于不同的环境或物理系统。如语音或语言, 这是人类沟通与交流的基本形式。还有与人们日常生活密切相关的电话、电视信号, 股票市场每日的收盘价, 等等, 这些信号都是自变量为时间的函数。<sup>[1]</sup>

### 1.2 模拟信号与数字信号

对于信号, 主要分为数字信号和模拟信号, 自然界中的信号绝大多数为模拟信号, 但模拟信号却较数字信号难直接处理, 因而在实际生活中, 我们往往将自然界中的模拟信号转换为数字信号进行处理, 又再次将处理好的数字信号转化为模拟信号便于我们使用, 这就是模数转换和数模转换, 即 AD 转换、DA 转换。

### 1.3 信号的时域分析与频域分析

除了信号的模拟与数字之间的转换, 信号的时域、频域分析也很重要。对于人们来说, 信号的时

域较频域时更加直观的, 但在处理信号时, 往往需要将其转换为频域信号, 方法主要有傅里叶变换、s 变换、z 变换等。

#### 1.3.1 傅里叶变换

实际上, 计算机存储的所有数据都是离散的, 它们需要运用时域和频域都是离散的离散傅里叶变换 (Discrete Fourier Transform, DFT) 进行处理。TD (Time-Domain) 连续信号经采样后, 通过快速傅里叶变换成为 FD (Frequency-Domain) 采样。通过数学表达式绘图, 不难看出输入 DFT 进行变换的时域信号和变换后输出的频域信号均为有限长序列, 即主值序列。在实际应用中常采用快速傅里叶变换计算 DFT: <sup>[2]</sup>

$$X(k) = DFT(x(n)) = \sum_{n=0}^{N-1} x(n)e^{-j\omega n}, 0 \leq k \leq N-1 \quad (1)$$

$$x(n) = IDFT(X(k)) = \sum_{k=0}^{N-1} X(k)e^{j\omega k}, 0 \leq n \leq N-1 \quad (2)$$

## 1.4 数字信号处理的一般步骤

### 1.4.1 模拟信号数字化

模数转换 (Analog-to-Digital Conversion, 简称 ADC) 是将模拟信号转换为数字信号的过程。在许多应用中, 需要将来自传感器、音频设备或其他模拟源的连续信号转换为数字形式, 以便进行数字信号处理、存储或传输。ADC 是执行这种转换的关键组件。

ADC 的基本原理是将连续的模拟信号离散化为一组数字值, 即样本。这个过程包括采样和量化两个主要步骤。采样是以固定的时间间隔获取模拟信号的快照, 而量化是将每个采样值映射为一个离散的数字值, 通常通过舍入或截断的方式实现。

ADC 有不同的类型, 包括逐次逼近型 (Successive Approximation ADC)、闪存型 (Flash ADC)、积分型 (Integrating ADC) 等。每种类型都有其独特的优势和适用场景, 选择取决于应用的要求和性能需求。

ADC 的性能参数通常包括分辨率、采样速率、信噪比等。分辨率表示 ADC 输出的数字位数, 采样速率表示 ADC 每秒能够处理的采样数, 信噪比则衡量了 ADC 在转换过程中引入的噪音和失真。

模数转换在各种领域中都得到广泛应用, 包括通信系统、医疗设备、音频处理、自动控制系统等。它是数字化世界中的关键环节, 使得模拟信号能够与数字系统进行交互, 并在数字领域中得到处理和分析。<sup>[3-6]</sup>

ADC 处理一般要经过**采样, 保持, 量化和编码**四个步骤。

### 1.4.2 信号处理

信号处理前, 一般对信号要进行预处理, 剥离信号的冗余与噪声并对其进行有效分析, 进而再根据实验、工程需求, 对数据进行加工、分析, 提取其特征参数等信息。

### 1.4.3 数字信号模拟化

数模转换 (Digital-to-Analog Conversion, 简称 DAC) 是将数字信号转换为模拟信号的过程。在数

字系统中, 信息通常以离散的形式存在, 而在某些应用中, 需要将这些数字信息转换为连续的模拟信号。数模转换器是执行这种转换的设备或电路。

数模转换的过程涉及将离散的数字数据转换为相应的模拟电压或电流。这种转换在许多领域都有广泛的应用, 比如音频处理、通信系统、图像处理等。

数模转换的主要类型包括两种: 脉冲宽度调制 (PWM) 和脉冲幅度调制 (PAM)。在 PWM 中, 数字信号被转换为脉冲宽度, 而在 PAM 中, 数字信号则被转换为脉冲幅度。这两种方法都能实现高质量的数模转换, 但选择取决于具体的应用需求。

DAC 的性能通常由几个指标来衡量, 包括分辨率、采样速率、非线性度等。分辨率表示 DAC 能够产生的不同输出水平的数量, 采样速率表示 DAC 每秒能够处理的数字样本数, 而非线性度则衡量输出信号与输入信号之间的误差。

数模转换在数字信号处理系统、嵌入式系统和各种通信设备中扮演着关键的角色, 它使得数字和模拟系统之间的互联变得更加灵活和可实现。<sup>[3-6]</sup>

## 2 深度学习

### 2.1 概念介绍

深度学习 (DL, Deep Learning) 是机器学习 (ML, Machine Learning) 领域中一个新的研究方向, 它被引入机器学习使其更接近于最初的目标——人工智能 (AI, Artificial Intelligence)。

深度学习是学习样本数据的内在规律和表示层次, 这些学习过程中获得的信息对诸如文字、图像和声音等数据的解释有很大的帮助。它的最终目标是让机器能够像人一样具有分析学习能力, 能够识别文字、图像和声音等数据。深度学习是一个复杂的机器学习算法, 在语音和图像识别方面取得的效果, 远远超过先前相关技术。<sup>[7]</sup>

深度学习在搜索技术、数据挖掘、机器学习、机器翻译、自然语言处理、多媒体学习、语音、推荐和个性化技术, 以及其他相关领域都取得了很多成

果。深度学习使机器模仿视听和思考等人类的活动，解决了很多复杂的模式识别难题，使得人工智能相关技术取得了很大进步。

## 2.2 深度学习分类

### 2.2.1 无监督学习

学习算法在学习过程完全不需要标签化的数据，称为无监督学习。<sup>[8]</sup>基于无监督学习的深度学习系统主要包括栈式自编码神经网络和深度信任网络，其基本单元分别是自编码器和受限玻尔兹曼机，采用栈式结构搭建深层神经网络。<sup>[9]</sup>因为世界上的只是大多数是由无监督学习获取到的，所以，无监督学习在深度学习中也发挥着重要的作用。无监督学习分为三层，分别是可见层、隐含层和输出层。

### 2.2.2 监督学习

监督学习是利用一组已知类别的样本调整分类器的参数，使其达到所要求性能的过程，也称为监督训练或有教师学习。

监督学习是从标记的训练数据来推断一个功能的机器学习任务，它使用一个训练集来教模型产生期望的输出。这个训练数据集包括输入和正确的输出，这使得模型可以随着时间的推移而学习。该算法通过损失函数测量其精度，调整直到误差被充分最小化。训练数据包括一套训练示例。

在监督学习中，每个实例都是由一个输入对象（通常为矢量）和一个期望的输出值（也称为监督信号）组成。监督学习算法是分析该训练数据，并产生一个推断的功能，其可以用于映射出新的实例。一个最佳的方案将允许该算法来正确地决定那些看不见的实例的类标签。这就要求学习算法是在一种“合理”的方式从一种从训练数据到看不见的情况下形成。<sup>[10]</sup>

## 2.3 深度学习中的信号处理

### 2.3.1 概述

深度学习在生活中日益应用，便利着人们的生活，在 OpenAI 公司发布 ChatGPT 等产品后，深度学习、人工智能等词语越来越被人们提及。而对

于深度学习涉及到的一些技术，如自动驾驶、图形识别、语音识别等，更是和信号处理无法分开。并且，运用数字信号处理 (DSP)，可以为深度学习增添深度，提升学习能力。<sup>[11]</sup>以下，将介绍部分深度学习中的信号处理问题。

### 2.3.2 语音识别

在语音识别中，先加入数字信号处理 (DSP) 部分，对音频信号进行预处理，根据深度学习技术，设计了基于语音特征识别并采用 DNN-HMM 的声学模型，通过底层网络过滤噪声，采用高层网络提取语音特征。此外，通过改进算法，利用线性单元代替非线性激活函数，从而获得更好的性能。在深度学习出现之前，语音建模方法通常使用 MFCC 特征，通过 GMM-HMM 模型进行训练和识别。深度学习被广泛应用后，DNN 和 DNN 都是在跨越多层网络结构后使用 softmax 分类器进行分类。<sup>[12-14]</sup>

数字信号处理在语音识别环节中起到了修整信号的作用，提升了识别的效率和正确度。语音信号结合深度学习，大大提升了识别的效率与准确度。语音识别系统有两个基本操作：信号建模和模式匹配。信号建模是将语音信号转换成一组参数的过程。模式匹配是从记忆中找到与输入语音信号中得到的参数集紧密匹配的参数集的任务。<sup>[15]</sup>

信号建模中，运用多种的 DSP 算法，提取到音频信号过零率、短时能量、MFCC 系数等有效信息，进而便于后期对信号内容进行分类与识别。

目前，有论文将语音信号处理部分分为两阶段，在第一阶段，使用传统的动态时间扭曲 (Dynamic Time Warping, DTW) 将测试阶段的话语与入学阶段的话语进行匹配，从而生成分数，然后将生成的 DTW 分数作为特征传递给一个 32 混合高斯混合模型 (GMM)，该模型是通过对注册数据进行内部匹配获得的 DTW 分数来学习的。<sup>[16]</sup>这就是 DSP 与深度学习的充分结合。

### 2.3.3 图像识别

图像识别，是指利用计算机对图像进行处理、分析和理解，以识别各种不同模式的目标和对象的技

术,是应用深度学习算法的一种实践应用。现阶段图像识别技术一般分为人脸识别与商品识别,人脸识别主要运用在安全检查、身份核验与移动支付中;商品识别主要运用在商品流通过程中,特别是无人货架、智能零售柜等无人零售领域。

图像的传统识别流程分为四个步骤: **图像采集** → **图像预处理** → **特征提取** → **图像识别**。

运用 DSP 算法的图像识别,使用了阈值分割算法。阈值分割算法是通过对目标区域的检测来实现图像的分割。因为图像分割的目的是提取目标区域。阈值分割算法是区域分割算法中具有代表性且非常重要的分割算法,其基本处理方法是:首先从图像的灰度值范围中选择一个灰度阈值,并将该阈值应用于图像中的所有像素。一般来说,该阈值分割后的两部分像素属于不同的区域,这使得使用该阈值分割图像区域成为可能。<sup>[17]</sup>选择合适的阈值是关键,一般的阈值分割可分为三个步骤。首先确定阈值,然后与像素值进行比较,最后对像素进行分类。结合三个步骤,确定好合适的阈值后,就可以正确的分割图像了。

还有的论文中采用了模型并行(Parallel model)的方法,运用多核心讯号处理器(multi-core DSP),进行图像的模式识别。<sup>[18]</sup>

### 3 总结

深度学习在数字信号处理领域的崛起为信号处理任务带来了全新的范式。一系列研究和论文在探索如何利用深度学习技术提高信号处理任务的性能。这个领域的发展呈现出多个关键趋势和重要挑战。

首先,深度学习在数字信号处理中的应用主要体现在特征提取和表示学习方面。相较于传统的手工设计特征,深度学习模型能够自动学习从原始信号中提取有用信息的表示。卷积神经网络(CNN)<sup>[19]</sup>等结构通过卷积操作在信号中捕获局部模式,为音频、图像等信号处理任务带来了显著的性能提升。

其次,循环神经网络(RNN)和长短时记忆网络(LSTM)在时序信号处理中发挥着重要作用。这些

网络结构能够有效地捕捉信号中的时序依赖性,为语音识别、时序预测等任务提供了强大的建模能力。这对于处理具有时间动态性的信号,如语音、生物信号等,具有重要意义。

时频分析一直是数字信号处理的核心任务,而深度学习方法为时频表示学习提供了新的思路。通过将深度学习与传统的时频分析方法结合,研究者们能够更准确地理解信号在时间和频率上的特征,从而提高对信号的理解和分析水平。

此外,深度学习还涉及到信号合成、生成对抗网络(GANs)等方面的研究。通过训练生成模型,可以合成具有真实感的信号数据,这对于模型的泛化能力和数据增强有着积极的影响。

然而,深度学习在数字信号处理中也面临一些挑战,包括对大量数据的依赖、模型的可解释性、实时性要求等方面。解决这些问题需要继续深入研究新的模型结构、训练方法以及硬件加速技术。

总体而言,深度学习在数字信号处理中的应用为信号处理任务提供了新的思路和解决方案。通过不断深化对深度学习在信号处理中的理解,我们有望进一步推动数字信号处理技术的发展,为实际应用提供更为有效和可靠的解决方案。

### 参考文献

- [1] 郑南宁. 数字信号处理简明教程[M]. 西安: 西安交通大学出版社, 2019.
- [2] 赵子曦. 基于 Matlab 的信号分析与数字滤波器设计[J/OL]. 电脑知识与技术, 2021, 17(114-115+119). DOI: 10.14004/j.cnki.ckt.2021.2823.
- [3] OPPENHEIM A V, SCHAFER R W, BUCK J R. Discrete-time signal processing[M]. Prentice Hall, 1999.
- [4] JOHNS D A, MARTIN K. Analog integrated circuit design[M]. John Wiley & Sons, 1997.
- [5] RAZAVI B. Principles of data conversion system design[M]. IEEE Press, 1995.
- [6] GRAY P R, MEYER R G. Analysis and design of

- analog integrated circuits[M]. John Wiley & Sons, 2009.
- [7] 张效荣. 基于卷积神经网络的深度学习算法与应用研究[D]. 西安电子科技大学, 2017.
- [8] 岳永鹏. 深度无监督学习算法研究[D]. 西南石油大学, 2015.
- [9] 殷瑞刚李晗 于洪. 深度学习中的无监督学习方法综述[J/OL]. 计算机系统应用, 2016, 25(1-7). DOI: 10.15888/j.cnki.csa.005283.
- [10] IBM. What is supervised learning?[EB/OL]. 2021. <https://www.ibm.com/topics/supervised-learning>.
- [11] GORDAY P, ERDÖL N, ZHUANG H. Lms to deep learning: How dsp analysis adds depth to learning[C/OL]//ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2019: 7765-7769. DOI: 10.1109/ICASSP.2019.8682669.
- [12] DI FATTA D, CAPUTO F, EVANGELISTA F, et al. Small world theory and the world wide web: linking small world properties and website centrality[J/OL]. International Journal of Markets and Business Systems, 2016, 2: 126. DOI: 10.1504/IJMABS.2016.080237.
- [13] WANG P. Research and design of smart home speech recognition system based on deep learning[C/OL]//2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL). 2020: 218-221. DOI: 10.1109/CVIDL51233.2020.00-98.
- [14] BOROOJENI K G I S S, AminiMH. Smart grids: security and privacy issues[M]. Springer International Publishing, 2017.
- [15] MUZAFFAR F, MOHSIN B, NAZ F, et al. Dsp implementation of voice recognition using dynamic time warping algorithm[C/OL]//2005 Student Conference on Engineering Sciences and Technology. 2005: 1-7. DOI: 10.1109/SCONEST.2005.4382877.
- [16] BHUKYA R K. Text-dependent speaker verification under practical scenarios using dtw-gmm based two-tier authentication[C/OL]//2022 IEEE 6th Conference on Information and Communication Technology (CICT). 2022: 1-5. DOI: 10.1109/CICT56698.2022.9997861.
- [17] HANG K. Real-time image acquisition and processing system design based on dsp[C/OL]//2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE): Vol. 4. 2010: 492-496. DOI: 10.1109/ICCAE.2010.5451614.
- [18] PEI X, ZHAO F, LU H. Multi-core dsp based parallel processing algorithm for imaging target detection[C/OL]//2020 International Symposium on Computer Engineering and Intelligent Communications (ISCEIC). 2020: 139-143. DOI: 10.1109/ISCEIC51027.2020.00036.
- [19] PERDIOS D, VONLANTHEN M, MARTINEZ F, et al. Cnn-based image reconstruction method for ultrafast ultrasound imaging[J/OL]. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 2022, 69(4): 1154-1168. DOI: 10.1109/TUFFC.2021.3131383.