

关于机器学习的五篇论文

文献综述

摘要

本综述旨在对关于机器学习和深度学习的一些方面进行总结，首先是对最基础的数据层面，虽然大模型在当今很流行，效果也很强，但它毕竟不是万能的，对于只有小样本数据集的一些领域，大模型却无处施展拳脚，所以需要加深研究关于小样本学习的方法，主题内容就对小样本学习方法进行一些总结。另外《聚类算法综述》这篇文献系统地分析聚类算法的原理并汇总了在机器学习和深度学习中关于聚类的算法，并给出了新的算法分类方法。图像识别是深度学习的一个非常重要的研究方向。主题内容中先介绍了关于卷积神经网络的一些综述性内容，后根据卷积神经网络的缺陷介绍了 Hinton 针对数据降维题出的神经网络降维方法。

关键词：机器学习；小样本；卷积；CNN，数据降维

1 引言

近年来，人工智能领域取得了突破性进展，其中深度学习作为机器学习的一个重要分支，展现出前所未有的特征学习与数据建模能力。在这一发展历程中，2006 年 Hinton 与 Salakhutdinov 在《Science》上发表的《Reducing the Dimensionality of Data with Neural Networks》具有里程碑意义。该文首次系统提出了深度神经网络的“逐层预训练”策略，有效解决了深层网络训练中的梯度消失与局部最优等长期难题，为深度学习的复兴奠定了理论基础【Hinton et al., 2006】。由此，研究者开始重新关注并发展多层非线性网络结构，推动了以卷积神经网络为代表的深度模型在多个领域的广泛应用。卷积神经网络作为一种特殊的深度学习架构，其设计灵感源于生物视觉皮层中“感受野”的层次化结构【Hubel & Wiesel, 1962】。早在 1980 年代，Fukushima 提出的神经认知机被视为 CNN 的雏形，而 LeCun 等人于 1998 年提出的 LeNet-5 模型则首次将卷积、池化和权值共享等机制成功应用于手写数字识别，奠定了现代 CNN 的基本结构【LeCun et al., 1998】。然而，由于当时数据规模与计算资源的限制，CNN 的发展一度陷入停滞。

直到 2012 年，Krizhevsky 等人提出的 AlexNet 在 ImageNet 大规模视觉识别挑战赛中取得突破性成果，将 Top-5 错误率降至 16.4%，远优于传统方法，重新点燃了研究者对 CNN 的热情【Krizhevsky et al., 2012】。AlexNet 的成功不仅验证了 CNN 在处理复杂视觉任务中的强大能力，也凸显了 ReLU 激活函数、Dropout 正则化与 GPU 并行训练等技术在深层网络优化中的关键作用。

随着 ImageNet 等大规模数据集的建立与 GPU 计算能力的飞速发展，CNN 的研究进入高速发展期。随后涌现出包括 ZF-Net、VGG、GoogLeNet、ResNet 等在内的多种深

层网络结构，它们在不断加深网络层数的同时，也通过 Inception 模块、残差连接、批量归一化等创新机制有效缓解了梯度退化与过拟合问题【Szegedy et al., 2015; He et al., 2016】。在物体检测方面，以 R-CNN、Fast R-CNN、Faster R-CNN 和 YOLO 为代表的系列模型，将区域提议机制与端到端学习相结合，显著提升了检测精度与效率【Girshick et al., 2014; Ren et al., 2015】。尽管 CNN 已在图像理解中展现出强大的表示与泛化能力，其发展仍面临诸多挑战，如深层网络的训练复杂度、对遮挡与形变的鲁棒性、模型可解释性以及大规模标注数据的依赖等。因此，系统梳理 CNN 的理论基础、结构演进与典型应用，并展望其未来发展方向，对推动计算机视觉与图像理解领域的进一步发展具有重要意义。

除了介绍卷积神经网络外，本文还介绍了关于小样本学习方法基本原理和面临的困境与未来发展，以及关于聚类算法的介绍和分类。

2 主题内容

2.1 小样本学习方法

现代的机器学习总是鼓吹大模型，使用大量数据训练一个非常复杂且极深的模型，这样得到的结果毋庸置疑是好的，但现实中仍有很多大模型无法解决或者说无法应用的场合，有些领域天生就没法产生大量的训练数据，典型的如医疗行业，实验样本就已经异常珍贵，更别说从中提取数据作为训练数据了，所以小样本学习显得尤为重要。

小样本学习旨在通过少量样本训练出有效的模型，解决现实世界中数据稀缺或标注成本高的问题。与人类快速学习能力形成对比，传统深度学习依赖大量数据，而小样本学习推动机器学习向“少样本、强泛化”方向发展。

论文《小样本学习》（本小节之后统称该论文）总结了三类方法：

1. 基于模型微调

在大规模源数据集上预训练一个基础模型，然后在目标小样本数据集上对模型参数进行微调。典型的方法如 ULMFiT, Nakamura 的微调方法。

优点：简单直接，是迁移学习的基础做法。

缺点：极易过拟合。当源域（如 ImageNet）与目标域（如医疗影像）分布差异大时，微调效果有限。

2. 基于数据增强

从根本上缓解数据量太少的缺陷。基于无标签数据、数据合成、特征增强等方法增加数据量或特征量。如使用 GAN 生成更多的数据。

3. 基于迁移学习

迁移学习是一种机器学习方法，其核心思想是将在一个任务（源领域）上训练好的模型的知识（尤其是其特征提取能力），迁移到一个新的、但相关的任务（目标领域）上，从而显著提升新任务的学习效率和性能。简单来说，它就是“举一反三”、“站在巨人的肩膀上”。

这些方法虽然可以一定程度上解决在小样本上训练模型遇到的困难，但仍面临很多难题。预训练悖论：很多小样本学习方法仍需在大规模数据集上预训练，这与“小样本”的初衷相悖。未来需要探索不依赖大规模预训练的方法。可解释性差：深度学习黑盒问

题在小样本学习中依然存在，难以理解模型究竟迁移了哪些知识。文本领域发展滞后：相比图像，小样本文本处理更具挑战（如语言复杂性、噪声），且缺乏像ImageNet那样统一的预训练和评估基准。

2.2 聚类算法

聚类是一种无监督学习方法，旨在将无标签数据集划分为若干个簇，使得类内样本相似性高，类间相似性低。

聚类分析包括四个主要步骤：

1. 特征选择或变换：

特征选择：选取最优特征子集，保留物理意义。

特征变换：通过映射提取隐含特征，消除冗余。

2. 聚类算法选择或设计

根据数据类型（数值型、非数值型）选择合适算法。相似性度量是关键。

3. 聚类结果评价：使用外部准则、内部准则、相对准则进行评估。

4. 聚类结果物理解析：结合领域知识解释聚类结果，形成可用的知识。

聚类最重要的就是对相似性的量，其主要方法有：

1. 对连续性变量可以使用欧氏距离、切比雪夫距离、曼哈顿距离、闵可夫斯基距离、马氏距离、对称点距离、皮尔逊相关系数、余弦相似度等。

2. 对离散变量，二值变量：使用匹配系数计算相似度，如海明距离。多值变量：通过单匹配策略或转化为二值变量处理。

3. 对混合变量：结合连续与离散变量的相似性度量方法，如加权求和或统一公式。

对聚类算法分类，论文的作者提出新的分类方式，以数据量大小为界，分为小数据聚类（ $<10^{12}$ 字节）和大数据聚类（ $>10^{12}$ 字节）。稳重还列举了几种典型聚类算法的性能对比。

2.3 卷积神经网络综述

CNN是学习深度学习神经网络的“入门级”核心算法，是每个人学习神经网络的“必经之路”，理解CNN有助于后续深度计算机视觉的学习。

CNN 是深度学习的重要分支，起源于对生物视觉皮层的研究（Hubel & Wiesel, 1962）。经历了从感知器、多层感知器（MLP）到反向传播网络（BP）的演变。2006年 Hinton 提出“逐层预训练”方法，推动了深度学习的复兴。

CNN的核心结构包括卷积层，池化层，全连接层和特征层。卷积层：使用卷积核提取局部特征，低层提取边缘等低级特征，高层提取更抽象的特征。详细介绍了输出特征图大小和参数数量的计算公式。强调了ReLU激活函数的优势。池化层：进行下采样，聚合相似特征，增强模型的平移不变性和鲁棒性。对比了最大池化、平均池化、随机池化等方法。全连接层：整合卷积/池化层学习到的局部特征，用于最终分类。常使用Dropout技术防止过拟合。特征面：讨论了特征面数量的重要性及设置策略，通常采用人工试错或金字塔架构。与传统模式识别算法的比较：CNN通过局部连接和权值共享，大大减少

了参数数量，降低了模型复杂度，更易于训练，且具有更强的表达能力和泛化能力。

CNN的一些改进算法。网内网络（NIN）：使用微型神经网络（MLP）替代传统的线性卷积滤波器，以增强模型的非线性表达能力；同时使用全局平均池化替代全连接层，大幅减少参数，防止过拟合。空间变换网络（STN）：在CNN中插入一个可学习的空间变换模块，能够自适应地对输入数据进行空间对齐和变换，使网络对平移、缩放、旋转等更具不变性。反卷积网络：作为CNN的“逆过程”，用于可视化各层学习到的特征，有助于理解和调试网络结构。Zeiler等人通过此方法发现CNN对平移和缩放具有不变性，但对旋转敏感。

2.4 用神经网络减少数据的维度

《Reducing the Dimensionality of Data with Neural Networks》这篇论文提出使用一种名为“栈式自编码器”的深层神经网络，通过“预训练+微调”的策略，来学习高维数据的高效、低维非线性表示，从而实现了比线性方法（如PCA）强大得多的数据降维和可视化效果。

在2006年之前，神经网络正处于第二个“寒冬”。人们普遍认为：1. 训练深层神经网络非常困难：误差在反向传播时，经过多层传递后会变得非常小（梯度消失）或非常大（梯度爆炸），导致底层的权重几乎无法得到有效更新。2. 主流的降维和特征提取方法是线性的：如主成分分析（PCA），它只能捕捉数据中的线性相关性。但对于图像、文本等复杂数据，其内在结构通常是非线性的，PCA无法很好地处理。

论文的核心方法包含两个部分：一个特殊的网络结构，和一个革命性的训练策略

网络结构：栈式自编码器。是一种无监督的神经网络，其目标是让输出尽可能等于输入。它由编码器和解码器两部分组成，编码器将高维数据映射到一个低维的编码，解码器将这个低维编码重建回原始的高维空间，训练目标是最小化重建误差，也就是让输入和输出的差别尽可能小。将多个自编码器堆叠起来形成栈式自编码器。

训练策略：贪婪逐层预训练。第一步：只训练网络的第一层（即第一个自编码器）。使用原始数据作为输入，学习第一层的编码器和解码器。第二步：固定第一层编码器的权重，将其输出（第一层的编码 h_1 ）作为第二层的输入。然后训练第二层自编码器。第三步：重复此过程，直到所有层都完成了初步训练。“贪婪”指的是每一层只关心优化自己这一层的重建误差，而不考虑整个网络的最终性能。这个过程为整个网络的权重提供了一个非常好的初始值，这个初始点位于一个相对平滑且易于优化的参数空间区域，有效避免了梯度消失和局部最优陷阱。在预训练完成后，将解码器部分移除，只在编码器末端连接一个分类器（如Softmax层），或者直接使用编码器的输出作为降维结果。然后，使用标准的反向传播算法，基于特定的任务目标（如分类误差或最终的重建误差）来微调整个网络的权重。这样的预训练提供了一个好的起点，而微调则根据最终任务对整个网络进行精细调整，使其性能达到最优。

2.5 图像理解中的卷积神经网络

深度学习，特别是CNN，因大数据（如ImageNet）和GPU算力支持，在图像理解中

取得突破。CNN通过端到端学习，避免了传统方法中复杂的手工特征提取，显著提升了图像理解的性能。

卷积网络一般结构包括卷积层，下采样层，全连接层，上一篇论文已经解释过，不再赘述。

CNN在图像理解中的应用：

2.5.1 图像分类与物体检测

AlexNet：开创性工作，使用ReLU、Dropout、数据增强。

Overfeat：统一分类、定位、检测任务。

R-CNN系列：R-CNN：使用候选区域 + CNN特征。SPP-net：引入空间金字塔池化，支持任意尺寸输入。Fast R-CNN：引入RoI池化，端到端训练。Faster R-CNN：引入区域提议网络（RPN），实现近实时检测

其他模型：VGG：使用小卷积核堆叠加深网络。GoogLeNet：Inception模块，参数少、性能强。ZF-Net：通过可视化优化网络结构。

2.5.2 人脸识别与验证

DeepID：提取高层特征，LFW上97.45%准确率。

DeepFace：3D 对齐 + 9 层 CNN，LFW 上 97.35%。

DeepID2：联合识别与验证信号，LFW上99.15%。

挑战：姿态、遮挡、光照变化等会影响模型的效率，而人脑可以通过全局信息对图片上的被干扰的局部信息进行想象推测，未来还需要提升模型对全局信息的理解。

2.5.3 语义分割与深度恢复

FCN：全卷积网络，支持任意尺寸输入，端到端分割。

CRF+RNN：结合条件随机场与循环神经网络，优化分割结果。

多任务学习：如Eigen等人同时预测深度、法向、语义标签。

弱监督方法：仅使用图像级标签进行训练。

2.5.4 人体关节检测

Tompson等人：从深度图像中检测人手关节。

Jain等人：加入空间约束，优化人体姿态估计。

挑战：自遮挡、精度提升，这方面的研究还未深入。

3 总结与展望

应对小样本学习面临的问题包括预训练悖论：很多小样本学习方法仍需在大规模数据集上预训练，这与“小样本”的初衷相悖；可解释性差；文本领域发展滞后，任务梯度冲突。未来可以将数据增强和迁移学习、元学习与图神经网络进行结合；设计更高效的元学习器、更强大的神经网络度量模块；引入新范式，结合主动学习、强化学习等框架，更智能的方式利用有限样本。

CNN通过权值共享、局部连接、池化等结构，适应图像特性，实现端到端学习。整体上卷积神经网络在图像理解上的挑战包括深层网络计算复杂度高，多目标遮挡、运动模糊等问题，对数据标注依赖强。未来还需要结合领域知识设计更优网络结构，进行理

论上的稀疏性。收敛性分析，大规模深度网络的高效训练与并行化。

《Reducing the Dimensionality of Data with Neural Networks》它向学术界和工业界证明了，深度神经网络是可以成功训练的，并且能解决复杂的现实问题。直接引领了随后的深度学习研究大爆发。Hinton提出了“预训练”范式，推动了表示学习的发展，强调了让机器自动学习数据表示的重要性，而不是依赖人工设计的特征，这成为了现代AI的核心思想之一。

4 参考文献

- [1] 赵凯琳,靳小龙,王元卓. 小样本学习研究综述[J]. 软件学报,2021,32(2): 349-369.
- [2] 章永来,周耀鉴. 聚类算法综述[J]. 计算机应用,2019,39(7): 1869-1882.
- [3] 周飞燕,金林鹏,董军. 卷积神经网络研究综述[J]. 计算机学报,2017,40(6): 1229-1251.
- [4] 李彦冬,郝宗波,雷航. 卷积神经网络研究综述[J]. 计算机应用,2016,36(9): 2508-2515, 2565.
- [5] 常亮,邓小明,周明全,等. 图像理解中的卷积神经网络[J]. 自动化学报,2016,42(9): 1300-1312.
- [6] G. E. Hinton, R. R. Salakhutdinov ,Reducing the Dimensionality of Data with Neural Networks.Science313,504-507(2006).DOI:10.1126/science.1127647
- [7] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报,2000,26(1): 32-42.
- [8] 余凯,贾磊,陈雨强,等. 深度学习的昨天、今天和明天[J]. 计算机研究与发展,2013,(9): 1799-1804.
- [9] 卢宏涛,张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述[J]. 数据采集与处理,2016,31(1): 1-17.
- [10] 杨丽,吴雨茜,王俊丽,等. 循环神经网络研究综述[J]. 计算机应用,2018,38(202): 1-6, 26.
- [11] 黄立威,江碧涛,吕守业,等. 基于深度学习的推荐系统研究综述[J]. 计算机学报,2018,41(7): 1619-1647.
- [12] 刘全,翟建伟,章宗长,等. 深度强化学习综述[J]. 计算机学报,2018,41(1): 1-27.
- [13] 郑远攀,李广阳,李晔. 深度学习在图像识别中的应用研究综述[J]. 计算机工程与应用,2019,55(12): 20-36.
- [14] 赵永强,饶元,董世鹏,等. 深度学习目标检测方法综述[J]. 中国图象图形学报,2020,25(4): 629-654.
- [15] 田萱,王亮,丁琪. 基于深度学习的图像语义分割方法综述*[J]. 软件学报,2019,30(2): 440-468.
- [16] 朱张莉,饶元,吴渊,等. 注意力机制在深度学习中的研究进展[J]. 中文信息学报,2019,33(6): 1-11.
- [17] 章永来,周耀鉴. 聚类算法综述[J]. 计算机应用,2019,39(7): 1869-1882.
- [18] 奚雪峰,周国栋. 面向自然语言处理的深度学习研究[J]. 自动化学报,2016,42(10): 1445-1465.
- [19] 杨剑锋,乔佩蕊,李永梅,等. 机器学习分类问题及算法研究综述[J]. 统计与决策,2019,35(6): 36-40.
- [20] 邵延华,张铎,楚红雨,等. 基于深度学习的YOLO目标检测综述[J]. 电子与信息学报,2022,44(10): 3697-3708.
- [21] 许德刚,王露,李凡. 深度学习的典型目标检测算法研究综述[J]. 计算机工程与应用,2021,57(8): 10-25.
- [22] 方路平,何杭江,周国民. 目标检测算法研究综述[J]. 计算机工程与应用,2018,54(13): 11-18, 33.

[23] 王晓刚. 图像识别中的深度学习[J]. 中国计算机学会通讯,2015,11(8): 15-21.