深度学习文献综述

院　 系 信息科学与技术学院

专 业 班 计算机技术

姓　 名　 余冠源

学　 号　 13250988

## 概述

### 深度学习的历史

深度学习是机器学习的一个分支。它的主要特点是通过多层次的学习而得到对于原始数据的不同抽象层度的表示，进而提高分类和预测等任务的准确性。2006年，机器学习领域的泰斗，加拿大多伦多大学教授Geoffrey Hinton和他的学生Ruslan Salakhutdinov在《科学》上发表了文章[1]，从此开启了在学术界和工业界对深度学习的浪潮。在他发表的这篇文章当中，主要有两个观点：第一，多隐层的人工神经网络具备着优异的学习特征的能力，它学习到的特征对样本数据有着更本质的刻画，使其更加有利于图像可视化或者文本处理等的分类任务；第二，深度神经网络在训练的时候存在一定的难度，这些可以通过“逐层初始化”的方法来有效的克服，在文章中，采用无监督学习来完成逐层初始化的工作。Hinton等在文献[2]中基于深信度网络（DBN，Deep Belief Nets）提出了非监督学习的贪心逐层训练算法，给解决深层结构中相关的优化难题带来了希望，之后，提出了多层自动编码器的深层结构。Hinton在文献[3]中提到的深度学习方法是机器学习研究的一个新的领域，它的动机是建立、模拟人脑的分析学习的神经网络，模范人脑的机制进行数据的解释，比如声音，图像和文本。深度学习是一种无监督的学习技术，在文章中提出了使用DBN快速学习的方法，主要分为两步：第一步，每次训练其中一层网络，第二步是调优，保持原始的表示x向上产生的高级表示r和向下产生的表示x’尽可能的保持一致。这一思想在文本提出的基于深度学习的方法处理情感分析问题时，依然被采纳。

此外，Lecun等人在文献[4]中采用卷积神经网络（CNNs， Convolutional Neural Networks），这是第一个真正具有多层结构的学习算法，它使用空间的相对关系来减少参数数目进而提高BP训练性能。在深度学习研究中，还有许多的变形结构，比如文献[5]中采用的去噪自动编码器（Denoisiong autoencoders），文献[6]使用的DCN方法，文献[7]提出的sum-product方法等。

### 深度学习的现状与应用

### Google Brain谷歌大脑

Google的科学家们，用16000台电脑模拟了一个模拟人脑神经网络出来，并向这个网络展示了1000万段随机从Youtube上选取的视频，看看它能学会什么。结果在完全没有外界干涉的条件下，它自己识别出了猫脸。而该项目使用的正是深度学习。深度学习无需人工干预，是一种“无监督的学习”。它基于一种学术假设：即人类对外界环境的了解过程最终可以归结为一种单一算法，而人脑的神经元可以通过这种算法，分化出识别不同物体的能力。这个识别过程甚至完全不需要外界干预。

通过向谷歌大脑的神经网络输入了一个单词“cat”，这个神经网络中并没有辞典，不了解这个单词的含义。但在观看了一千万段视频，它最终确定，cat就是那种毛茸茸的小动物。这个学习过程，与一个不懂英语的人，在没任何人教他的时候，通过独立观察学会“cat”的过程几乎一致。

### 百度深度学习研究院IDL与其成果

2013年百度年会上提出，2013年百度将建立初期专注于Deep Learning(深度学习)的研究院，并命名为Institute of Deep Learning(简称IDL)。百度是中国互联网企业中第一个把Deep Learning提到核心技术创新地位的企业。

目前，IDL研究成果包括：

* 商品图像检索，也就是用户上传一张图片，系统可以根据图片内容从海量商品库里面找到相同或者类似的商品。
* 三维计算视觉，利用图片、视频以及各类深度传感器信息，采用集合、统计以及优化等数学工具对现实世界进行三维建模
* 百度人脸搜索
* 百度自然场景OCR

### 微软“Adam计划”

微软的研究人员称，他们的深度学习系统—名叫Adam(亚当)取得了突破性的成果，比起之前的深度学习系统而言更为成熟。微软的研究人员举例说，比如在图片识别方面，这个系统不仅可以识别出指定的物品，还能够在该类目分类项下，进行更精确的识别。和先前的“Google大脑”作对比，如果说“Google大脑”能做到的是，在看完一周Youtube视频后，识别出猫，那么Adam可以识别出猫及猫的品种，并且使用的机器数量只有之前的三十分之一。

“Adam计划”寻找的并不是在模拟神经网络的规模上有所突破，而是构建方式上的改进。这些改进指的是：优化和微调了机器处理数据和机器间数据沟通的方式。当计算系统变得更为复杂时，每个部件之间都在同一时间传送信息—这件事会随着系统越来越庞大而越来越困难，而Adam系统允许了异步算法的存在。

## 传统神经网络

人工神经网络（NN, Neural Network），时通过对人脑的组织架构和活动原理的认知，从信息处理的角度，以数学和物理方法进行抽象而建立起来的对人脑生物神经网络的简化模型，能够完成一种映射学习，可以基于大规模历史数据进行训练、聚类。

神经网络时由许多被称为节点的简单非线性模拟处理要素密集互联而成的，时一种模仿生物神经元的系统模型。网络通过令每一个节点的输出与一些其他的节点输入连接形成，类似于真实神经元的突触连接。每个神经元表达一种特殊的输出函数，称为激励函数，每两个神经元之间的连接都包含一个连接强度，也就是作用于通过该连接的信号的加权值。经过训练之后的神经网络，具有信息特征抽取、知识概括和学习记忆的能力，而模型学习到的信息或知识则存储在每个单元节点之间的连接矩阵上。一个神经网络的表现是由节点定义，拓扑结构以及学习算法这三个特征组成的集体性质决定的。

神经元一般是一种多输入、单输出的非线性结构，作为神经网络的基本运算单元，可以有反馈输入和阈值参数，一个通用的单元模型在图1-1可以看到，

Ni

X1

W1i

X2

W2i

.

.

Xj

Wji

f(·)

Yi

Vi

图 2-1 神经元模型

其中Vi代表基函数，是一个单输出多输入的函数 V = V(X, W, N)，f(·)称为激励函数，也被称为传递函数或者传输函数。

神经网络从互联结构来看，包括前反馈网络，如径向基网络，反馈型局部互联网络，有反馈的前向反馈网络，全反馈型网络，如Hopfield神经网络和波尔兹曼机，以及自组织网络等几种典型结构。

所谓神经网络就是将许多个单一“神经元”联结在一起，这样，一个“神经元”的输出就可以是另一个“神经元”的输入。例如，下图就是一个简单的神经网络：

X1

X2

X3

+1

+1

a1 (2)

a2 (2)

a3 (2)

层L3

层L1

层L2

hW,b (x)

图 2‑2 神经网络模型

传统机器学习和信号处理技术仅仅包含单层非线性变换的浅层学习结构。浅层模型的一个共性是仅含单个将原始输入信号转换到特定问题空间特征的简单结构。典型的浅层学习结构包括传统隐马尔可夫模型（HMM）、条件随机场（CRFs）、最大熵模型（MaxEnt）、支持向量机（SVM）、核回归及仅含单隐层的多层感知器（MLP）等。例如，SVM用包含一层（使用核技巧）或者零个特征转换层的浅层模式分离模型。浅层结构的局限性在于有限的样本和计算单元情况下对复杂函数的表示能力有限，针对复杂分类问题其泛化能力受到一定制约。神经科学研究表明，人的视觉系统的信息处理是分级的。

人类感知系统这种明确的层次结构极大地降低了视觉系统处理的数据量，并保留了物体有用的结构信息。有理由相信，对于要提取具有潜在复杂结构规则的自然图像、视频、语音和音乐等结构丰富数据，深度学习能够获取其本质特征。受大脑结构分层次启发，神经网络研究人员一直致力于多层神经网络的研究。

## 深度学习神经网络

深度学习是为了尽可能的接近人类的工作方式，通过编写学习算法，利用大脑架构的模拟，神经网络学习的实现经历了浅层架构到深度架构的过渡。通过生物神经元的启示，机器学习经历仅仅有两层左右的学习架构到多层次结构的发展，从而改进现有机器学习方法。

假如一个系统X, 它有n层（X1, X2, …, Xn），它的输入是I，输出是O，形象的表示为I > X1 > X2 > … > Xn > O，如果输出O等于输入I，即输入I经过这个系统之后保持不变，没有任何信息损失，这意味着输入I经过每一层Xi，都可以当作原有信息的另一种表示。我们需要自动的学习特征，加入输入I（如一段语音、图像或者文本），假设我们设计一个系统X，该系统有n层，我们通过调整系统的参数，使得它的输出仍然为输入I，那么我们就可以自动获取输入I的一系列层次特征，即X1, …, Xn。

### 深度学习的方法

历史上，深层学习的概念起源于神经网络的研究。带有多隐层的前馈神经网络或者多层感知器通常被成为深层神经网络（DNNs），DNNs就是深层构架的一个很好的例子。BP算法作为传统训练多层网络的典型算法，实际上对于仅含几层网络，该训练方法就已很不理想。

在学习中，一个主要的困难源于深度网络的非凸目标函数的局部极小点普遍存在。反向传播是基于局部梯度下降，通常随机选取初始点。使用批处理BP算法通常会陷入局部极小点，而且随着网络深度的增加，这种现象更加严重。此原因在一定程度上阻碍了深度学习的发展，并将大多数机器学习和信号处理研究从神经网络转移到相对较容易训练的浅层学习结构。

经验上，有3种技术可以处理深层模型的优化问题：

* 大量的隐藏的单元
* 更好的学习算法
* 以及更好的参数初始化技术。

使用带有大量神经元的DNN可以大大提高建模能力。由于使用带有大量神经元的DNN得到较差局部最优值的可能性要小于使用少量神经元的网络，即使参数学习陷入局部最优，DNN仍然可以很好的执行。但是，在训练过程中使用深而广的神经网络，对计算能力的需求要求很大。 更好的算法也有助于了DNNs的训练。例如，现在随机BP算法已经代替了批处理BP算法用来训练DNNs。部分原因是由于，当训练是单学习器和大训练集上进行时，随机梯度下降（SGD）算法是最有效的算法(Bottou and LeCun,2004)。但更重要的是SGD算法可以经常跳出局部最优。 其它算法，如 Hessian free (Martens 2010)或Krylov子空间方法（Vinyals and Povey 2011）有类似的能力。

很明显，对于高度非凸的DNN学习的优化问题，更好的参数初始化技术将导致更好的模型.然而，如何高效的初始化DNN的参数却不是很显然的事情。最近，学者们给出很好的结果(Hinton et al. 2006; Hinton and Salakhutdinov, 2006; Bengio, 2009;Vincent et al., 2010; Deng et al., 2010; Dahl et al., 2010, 2012; Seide et al. 2011). 最著名的DNN参数初始化技术就是无监督预训练（pre-training）技术的提出(Hinton et al. 2006; Hinton and Salakhutdinov, 2006). 在上述文章中，引入了一个被称作深层信念网(DBN)的深层贝叶斯概率生成模型。为了学习DBN中的参数，提出非监督贪心逐层训练算法，算法把DBN中的每两层作为一个限制玻耳兹曼机（RBM）。这使得优化DBN参数的计算复杂度随着网络的深度成线性增长。DBN参数可以直接用作MLP或DNN参数，在训练集较小的时候，可以得到比随机初始化的有监督BP训练要好的MLP或DNN。带有无监督DBN预训练，随后通过反向微调（fine-tuning）的DNNs

有时候也被称作DBNs(e.g., Dahl et al., 2011; Mohamed et al., 2010, 2012). 最近,研究人员已经更小心区分

DNNs 和DBNs(Dahl et al., 2012; Hinton et al., 2012) ,当DBN用于初始化一个DNN的参数,由此产生的网络叫做DBN-DNN(Hinton et al., 2012). DBN预训练过程不是唯一有效的DNNs初始化方法。另一种效果同样好的无监督方法是通过把每两层作为一个去噪自动编码器来逐层预训练DNNs(Bengio, 2009; Vincent et al., 2010) .另一种方法是用收缩自动编码器，它对于输入变化的敏感度较低(Rifai et al., 2011).而且，Ranzato et al. (2007) 提出了稀疏编码对称机（SESM），它与RBMs非常类似，都作为一个DBN的构造模块。原则上，SESM也可以用来有效的初始化DNN训练。除了半监督预训练外，监督预训练（有时也叫作区别预训练）也被证明是有效的(Seide et al., 2011; Yu et al., 2011) 。在有标签样本数据充足的时候表现要优于无监督预训练技术。区别预训练的主要思想是从一个隐层MLP开始，用BP算法训练。然后，每次我们想要增加一个新的隐藏层，我们通过随机初始化一个新的隐藏和输出层来代替原来输出层，再用BP算法训练这个新的MLP（或DNN）。与无监督预训练技术不同，区分与监督需要标签。

参考文献

1. Hinton G E, Osindero S, The Y W. A Fast Learning Algorithm for Deep Belief Nets [J]. Neural Comp., 2006, 18(7):1527-1544.
2. 百度深度学习实验室，<http://idl.baidu.com/IDL-research-list.html>