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，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，比如文献中采用的去噪自动编码器（Denoisiong autoencoders），文献使用的DCN方法，文献提出的sum-product方法等。

## 深度学习的现状与应用

深度学习算法已经被广泛应用于各大IT巨头公司，例如苹果，谷歌，微软等。深度学习算法可以提高搜索的精确度，提高图像识别的能力，从而构造出如自动驾驶，人机对话等一类的颠覆性的产品。

### 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2-1可以看到，

*Ni*

.

.

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2-2就是一个简单的神经网络：

*+1*

*+1*

层*L3*

层*L1*

层*L2*

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（），它的输入是*I*，输出是*O*，形象的表示为，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。

## 深度学习的方法

深度学习大体可以分为两种，无监督学习和有监督学习。例如卷积神经网络，就是一种有监督学习下的深度结构学习模型，而深度信念网络，就是一种无监督学习下的深度结构学习模型。

### 卷积神经网络

卷积神经网络，包含卷积层和次抽样层，这两种特殊的结构层，是一种非全连接的神经网络结构。卷积层定义了相应的的接受域，对于每个神经元，只接受从其接受域传输的信号，如图3-1所示。每个特征平面由神经元构成，由多个特征平面共同组成卷积层，完成特征抽取任务，而相同特征平面上的所有神经元具有一致的连接权值。相同特征平面神经元的连接矩阵，因为神经元的接受域大小相同而一致。

图 3‑1卷积神经网络结构

卷积和子采样过程为：卷积过程包括输入数据与训练过滤器进行卷积运算，再加上训练用的偏置向量，从而产生卷积层；子采样过程包括邻域相加，加权标量，再加上偏置向量bx+1，所得结果通过*S*形函数产生大概小了*2x*的特征映射。

### 深度信念网络

另一种代表性的深度学习结构是深度信念网络模型。它是多个限制玻尔兹曼机的累加。一个玻尔兹曼机是具有热力学的能量函数定义的概率分布玻尔兹曼分布，一种基于能量理论的概率模型。设状态随机变量，能量函数为，一个典型的玻尔兹曼机是一个无向循环图，其能量函数定义为

如果对玻尔兹曼机加以约束条件，层内无互联，可得到限制玻尔兹曼机，能量函数定义为：

一个典型的信念网络，可以看成是一系列RBM的堆叠组成，可以看成由许多个随机变量组成的有向无环图。用联合概率分布刻画一个有个隐藏层的深度信念网络的输入和隐含向量的关系

为条件概率， 。深度信念网络属于深度架构中产生的结构。

Hinton教授与其他几位学者于2006年，提出了一种“深度信念网络”（Deep Belief Nets， DBN），并给出了一个高效的学习算法。这个贪婪的逐层无监督训练算法成为了其后至今深度网络算法的主要框架。一个典型的深度信念网络就是一个高复杂度的有向无环图，可以看成是一系列限制玻尔兹曼机的对对组成。DBN在训练过程中，通过由低到高逐层训练这些限制玻尔兹曼机来实现。大量实验证明，深度信念网络可以解决传统反向传播算法训练多层神经网络的难题：需要大量的含标号训练样本集，较慢的收敛速度，以及因不合适的参数选择而陷入局部最优。

# 总结

深度学习在信号处理，图像处理，文本处理范畴已经被广泛应用。应用对象从传统的音频，语音，图像和视频，到文字，语言，文本等更高层次的人类语义信息。深度学习在计算机视觉，语音识别，自然语言处理等领域都得到了成功的应用。

例如，2011年以来，微软研究院和Google的语音识别研究人员先后采用DNN技术降低语音识别错误率20％~30％，是语音识别领域十多年来最大的突破性进展。2012年，DNN技术在图像识别领域取得惊人的效果，在ImageNet评测上将错误率从26％降低到15％。

Google也采用了深层神经网络进行声音建模，是最早突破深层神经网络工业化应用的企业之一。但Google产品中采用的深度神经网络只有4-5层，而百度采用的深度神经网络多达9层。这种结构差异的核心其实是百度更好地解决了深度神经网络在线计算的技术难题，因此百度线上产品可以采用更复杂的网络模型。这将对于未来拓展海量语料的DNN模型训练有更大的优势。

将来，深度学习算法已不仅仅是计算机领域的范畴，未来可能会和各个学科相交融，例如生物神经网络，语言学，物理学，概率统计学等学科。借鉴这些学科中的科学研究方法，并应用于深度学习研究当中，更多更有效的算法会被发现。计算机的思考越来越接近人的思维模式，人工智能的未来将会离我们越来越近。

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. Alex Krizhevsky, Hya Sutskever, Geoffery E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks[J], 2012
3. 百度深度学习实验室，<http://idl.baidu.com/IDL-research-list.html>
4. 陈硕，深度学习神经网络在语音识别中的应用研究[D]. 硕士学位论文，华南理工大学，2013
5. 朱少杰，基于深度学习的文本情感分类研究[D]. 硕士学位论文，哈尔滨工业大学，2014
6. 林妙真，基于深度学习的人脸识别研究[D]. 硕士学位论文，大连理工大学，2013
7. 梁军，柴玉梅，原慧斌，昝红英，刘铭，基于深度学习的微博情感分析[J]，中文信息学报，2014