深度学习

来源：360百科

深度学习的概念源于人工神经网络的研究。含多隐层的多层感知器就是一种深度学习结构。深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征，以发现数据的分布式特征表示。

深度学习的概念由Hinton等人于2006年提出。基于深信度网(DBN)提出非监督贪心逐层训练算法，为解决深层结构相关的优化难题带来希望，随后提出多层自动编码器深层结构。此外Lecun等人提出的卷积神经网络是第一个真正多层结构学习算法，它利用空间相对关系减少参数数目以提高训练性能。

深度学习是机器学习研究中的一个新的领域，其动机在于建立、模拟人脑进行分析学习的神经网络，它模仿人脑的机制来解释数据，例如图像，声音和文本。

同机器学习方法一样，深度机器学习方法也有监督学习与无监督学习之分.不同的学习框架下建立的学习模型很是不同.例如，卷积神经网络(Convolutional neural networks，简称CNNs)就是一种深度的监督学习下的机器学习模型，而深度置信网(Deep Belief Nets，简称DBNs)就是一种无监督学习下的机器学习模型。

简介

深度学习是机器学习研究中的一个新的领域，其动机在于建立、模拟人脑进行分析学习的神经网络，它模仿人脑的机制来解释数据，例如图像，声音和文本。深度学习是无监督学习的一种。 深度学习的概念源于人工神经网络的研究。含多隐层的多层感知器就是一种深度学习结构。深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征，以发现数据的分布式特征表示。 深度学习的概念由Hinton等人于2006年提出。基于深信度网(DBN)提出非监督贪心逐层训练算法，为解决深层结构相关的优化难题带来希望，随后提出多层自动编码器深层结构。此外Lecun等人提出的卷积神经网络是第一个真正多层结构学习算法，它利用空间相对关系减少参数数目以提高训练性能。

基础概念

[折叠](https://baike.so.com/doc/7000002-7222884.html)深度

从一个输入中产生一个输出所涉及的计算可以通过一个流向图(flow graph)来表示:流向图是一种能够表示计算的图，在这种图中每一个节点表示一个基本的计算并且一个计算的值(计算的结果被应用到这个节点的孩子节点的值)。考虑这样一个计算集合，它可以被允许在每一个节点和可能的图结构中，并定义了一个函数族。输入节点没有父亲，输出节点没有孩子。

这种流向图的一个特别属性是深度(depth):从一个输入到一个输出的最长路径的长度。

传统的前馈神经网络能够被看做拥有等于层数的深度(比如对于输出层为隐层数加1)。SVMs有深度2(一个对应于核输出或者特征空间，另一个对应于所产生输出的线性混合)。

[折叠](https://baike.so.com/doc/7000002-7222884.html)解决问题

需要使用深度学习解决的问题有以下的特征:

深度不足会出现问题。

人脑具有一个深度结构。

认知过程逐层进行，逐步抽象。

在许多情形中深度2就足够表示任何一个带有给定目标精度的函数。但是其代价是:图中所需要的节点数(比如计算和参数数量)可能变的非常大。理论结果证实那些事实上所需要的节点数随着输入的大小指数增长的函数族是存在的。

我们可以将深度架构看做一种因子分解。大部分随机选择的函数不能被有效地表示，无论是用深的或者浅的架构。但是许多能够有效地被深度架构表示的却不能被用浅的架构高效表示。一个紧的和深度的表示的存在意味着在潜在的可被表示的函数中存在某种结构。如果不存在任何结构，那将不可能很好地泛化。

例如，视觉皮质得到了很好的研究，并显示出一系列的区域，在每一个这种区域中包含一个输入的表示和从一个到另一个的信号流(这里忽略了在一些层次并行路径上的关联，因此更复杂)。这个特征层次的每一层表示在一个不同的抽象层上的输入，并在层次的更上层有着更多的抽象特征，他们根据低层特征定义。

需要注意的是大脑中的表示是在中间紧密分布并且纯局部:他们是稀疏的:1%的神经元是同时活动的。给定大量的神经元，仍然有一个非常高效地(指数级高效)表示。

人类层次化地组织思想和概念;

人类首先学习简单的概念，然后用他们去表示更抽象的;

工程师将任务分解成多个抽象层次去处理;

学习/发现这些概念(知识工程由于没有反省而失败?)是很美好的。对语言可表达的概念的反省也建议我们一个稀疏的表示:仅所有可能单词/概念中的一个小的部分是可被应用到一个特别的输入(一个视觉场景)。

核心思想

把学习结构看作一个网络，则深度学习的核心思路如下:

①无监督学习用于每一层网络的pre-train;

②每次用无监督学习只训练一层，将其训练结果作为其高一层的输入;

③用监督学习去调整所有层

优势

用更多的数据或是更好的算法来提高学习算法的结果。对某些应用而言，深度学习在大数据集上的表现比其他[机器学习](https://baike.so.com/doc/3290018-3465759.html)(ML)方法都要好(稍后将讨论例外情况)。这些又如何转化为现实生活中的情形呢?深度学习更适合无标记数据，因而它并不局限于以实体识别为主的自然语言处理(NLP)领域。[[1]](https://baike.so.com/doc/7000002-7222884.html%22%20%5Cl%20%22refff_7000002-7222884-1)

五个能力级别

[折叠](https://baike.so.com/doc/7000002-7222884.html)C

C级别作者的定义是Classification Only，可以理解为只能做分类。涉及到的系统包含全链接神经网络(FCN)、卷积网络(CNN)以及它们的各种组合。这些系统采用高维向量作为输入并实现单个结果，通常是输入向量的一个分类。你可以把这些系统当作是无状态函数，这意味着当前的行为仅仅是输入的一个函数。生成模型也是热门研究领域中的一员，也归于此类。简而言之，这些系统本身都是相当有能力的。[[2]](https://baike.so.com/doc/7000002-7222884.html%22%20%5Cl%20%22refff_7000002-7222884-2)

[折叠](https://baike.so.com/doc/7000002-7222884.html)CM

CM是Classification with Memory的缩写，理解成在分类的基础之上进行记忆。该级别包括与C级网络合并的存储器元件，例如LSTM，其中存储器单元被嵌入在LSTM节点内部。其它变体则是来自DeepMind的神经图灵机(NMT)和可微分神经计算机(DNC)，它们在计算时都处于维护状态。

[折叠](https://baike.so.com/doc/7000002-7222884.html)CK

CK即是Classification with Knowledge的缩写，即可以形成自己的知识系统。该级别与CM级别有点类似，但并不是原始存储器，C级网络能够访问的信息是符号知识库。实际上，我发现了三种符号集成，一种是转移学习方法、一种是自上而下方法、最后一种是自下而上的方法。第一种主要是作为正则化矩阵的符号系统;第二种方法处于层次结构的顶部，具有在底部由神经表示构成的符号元素;最后一种则与第二种方法相反，其中C级网络实际上被附加在了符号知识库上。

[折叠](https://baike.so.com/doc/7000002-7222884.html)CCIK

CCIK是Classification with Imperfect Knowledge缩写，可以对不完整的知识库进行分类。该级别系统建立在CK之上，但是能对不完整信息进行推理。该级别的典型例子是AlphaGo和扑克系统，这些系统可以通过对自身模拟来训练自己。

[折叠](https://baike.so.com/doc/7000002-7222884.html)CCIK

CCIK的全称是Collaborative Classification with Imperfect Knowledge，即可以对不完整的知识库进行分类并且可以互相协作。这个级别与“心理理论”非常类似，实际上，我们有多种代理神经网络组合来解决问题，这些系统旨在实现多个目标。实际上，我们在对抗网络上做过这个原始版本，通过执行泛化的竞争鉴别器和生成网络将这个概念进一步扩展到游戏理论驱动的网络，能够执行战略和战术解决多个目标。

每个级别都有其不同的功能产生，例如，C级系统只能够预测anti-causal关系，CM拥有非常好的翻译能力，CIK级别系统能够进行战略游戏。

例题

a). AutoEncoder

最简单的一种方法是利用人工神经网络的特点，人工神经网络(ANN)本身就是具有层次结构的系统，如果给定一个神经网络，我们假设其输出与输入是相同的，然后训练调整其参数，得到每一层中的权重，自然地，我们就得到了输入I的几种不同表示(每一层代表一种表示)，这些表示就是特征，在研究中可以发现，如果在原有的特征中加入这些自动学习得到的特征可以大大提高精确度，甚至在分类问题中比目前最好的分类算法效果还要好!这种方法称为AutoEncoder。当然，我们还可以继续加上一些约束条件得到新的Deep Learning方法，如如果在AutoEncoder的基础上加上L1的Regularity限制(L1主要是约束每一层中的节点中大部分都要为0，只有少数不为0，这就是Sparse名字的来源)，我们就可以得到Sparse AutoEncoder方法。

b). Sparse Coding

如果我们把输出必须和输入相等的限制放松，同时利用线性代数中基的概念，即O = w1\*B1 + W2\*B2+....+ Wn\*Bn， Bi是基，Wi是系数，我们可以得到这样一个优化问题:

Min |I - O|

通过求解这个最优化式子，我们可以求得系数Wi和基Bi，这些系数和基础就是输入的另外一种近似表达，因此，可以用它们来特征表达输入I，这个过程也是自动学习得到的。如果我们在上述式子上加上L1的Regularity限制，得到:

Min |I - O| + u\*(|W1| + |W2| + ... + |Wn|)

这种方法被称为Sparse Coding。

c) Restrict Boltzmann Machine (RBM)

假设有一个二部图，每一层的节点之间没有链接，一层是可视层，即输入数据层()，一层是隐藏层()，如果假设所有的节点都是二值变量节点(只能取0或者1值)，同时假设全概率分布p()满足Boltzmann 分布，我们称这个模型是Restrict Boltzmann Machine (RBM)。下面我们来看看为什么它是Deep Learning方法。首先，这个模型因为是二部图，所以在已知的情况下，所有的隐藏节点之间是条件独立的，即p(|) =p(1|).....p(n|)。同理，在已知隐藏层的情况下，所有的可视节点都是条件独立的，同时又由于所有的v和h满足Boltzmann 分布，因此，当输入的时候，通过p(|) 可以得到隐藏层，而得到隐藏层之后，通过p(v|h) 又能得到可视层，通过调整参数，我们就是要使得从隐藏层得到的可视层与原来的可视层如果一样，那么得到的隐藏层就是可视层另外一种表达，因此隐藏层可以作为可视层输入数据的特征，所以它就是一种Deep Learning方法。

如果，我们把隐藏层的层数增加，我们可以得到Deep Boltzmann Machine (DBM);如果我们在靠近可视层的部分使用贝叶斯信念网络(即有向图模型，当然这里依然限制层中节点之间没有链接)，而在最远离可视层的部分使用Restrict Boltzmann Machine，我们可以得到Deep Belief Net (DBN) 。

当然，还有其它的一些Deep Learning 方法。总之，Deep Learning能够自动地学习出数据的另外一种表示方法，这种表示可以作为特征加入原有问题的特征集合中，从而可以提高学习方法的效果，是业界的研究热点。

转折点

2006年前，尝试训练深度架构都失败了:训练一个深度有监督前馈神经网络趋向于产生坏的结果(同时在训练和测试误差中)，然后将其变浅为1(1或者2个隐层)。

2006年的3篇论文改变了这种状况，由Hinton的革命性的在深度信念网(Deep Belief Networks, DBNs)上的工作所引领:

Hinton, G. E., Osindero, S. and Teh, Y.,A fast learning algorithm for deep belief nets.Neural Computation 18:1527-1554, 2006

Yoshua Bengio, Pascal Lamblin, Dan Popovici and Hugo Larochelle,Greedy LayerWise Training of Deep Networks, in J. Platt et al. (Eds), Advances in Neural Information Processing Systems 19 (NIPS 2006), pp. 153-160, MIT Press, 2007

Marc’Aurelio Ranzato, Christopher Poultney, Sumit Chopra and Yann LeCunEfficient Learning of Sparse Representations with an Energy-Based Model, in J. Platt et al. (Eds), Advances in Neural Information Processing Systems (NIPS 2006), MIT Press, 2007

在这三篇论文中以下主要原理被发现:

表示的无监督学习被用于(预)训练每一层;

在一个时间里的一个层次的无监督训练，接着之前训练的层次。在每一层学习到的表示作为下一层的输入;

用有监督训练来调整所有层(加上一个或者更多的用于产生预测的附加层);

DBNs在每一层中利用用于表示的无监督学习RBMs。Bengio et al paper 探讨和对比了RBMs和auto-encoders(通过一个表示的瓶颈内在层预测输入的神经网络)。Ranzato et al paper在一个convolutional架构的上下文中使用稀疏auto-encoders(类似于稀疏编码)。Auto-encoders和convolutional架构将在以后的课程中讲解。

从2006年以来，大量的关于深度学习的论文被发表。

成功应用

1、计算机视觉

, Alex Krizhevsky, Ilya Sutskever, Geoffrey E Hinton, NIPS 2012., Clement Farabet, Camille Couprie, Laurent Najman and Yann LeCun, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013., Koray Kavukcuoglu, Pierre Sermanet, Y-Lan Boureau, Karol Gregor, Micha&euml;l Mathieu and Yann LeCun, Advances in Neural Information Processing Systems (NIPS 2010), 23, 2010.

2、语音识别

微软研究人员通过与hintion合作，首先将RBM和DBN引入到语音识别声学模型训练中，并且在大词汇量语音识别系统中获得巨大成功，使得语音识别的错误率相对减低30%。但是，DNN还没有有效的并行快速算法，很多研究机构都是在利用大规模数据语料通过GPU平台提高DNN声学模型的训练效率。

在国际上，IBM、google等公司都快速进行了DNN语音识别的研究，并且速度飞快。

国内方面，科大讯飞、百度、中科院自动化所等公司或研究单位，也在进行深度学习在语音识别上的研究。

3、自然语言处理等其他领域

很多机构在开展研究，但深度学习在自然语言处理方面还没有产生系统性的突破。

参考资料：

* 1．[为什么会有深度学习？如何看待？](http://www.chinarobots.cn/ShenDuXueXi/16.html%22%20%5Ct%20%22_blank) . 中国机器人 . 2016-02-29 . [2017-3-9]
* 2．[深度学习的五个能力级别](http://www.chinarobots.cn/ShenDuXueXi/1520.html%22%20%5Ct%20%22_blank) . 中国机器人 . 2016-11-30 . [2017-3-10]