DOI: 10.3969/j.issn.1009-6868.2017.04.006
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.TN.20170706.1411.006.html

深度神经网络学习的结构基础：自动编码器与限制玻尔兹曼机

Architectures of Deep Neural Networks: Auto-Encoders and Restricted Boltzmann Machines

康文斌/ KANG Wenbin[1]
彭菁/ PENG Jing[2]
唐乾元/ TANG Qianyuan[3]
（1. 湖北医药学院，湖北 十堰 442000；
2. 平安科技（深圳）有限公司，广东 深圳 518000；
3. 香港浸会大学，香港 九龙塘 999077）
(1. Hubei University of Medicine, Shiyan 442000, China;
2. Ping An Technology (Shenzhen) Co, Ltd., Shenzhen 518000, China;
3. Hong Kong Baptist University, Hong Kong 999077, China)

中图分类号：TN929.5 文献标志码：A 文章编号：1009-6868 (2017) 04-0032-004

摘要： 自动编码器（AE）和限制玻尔兹曼机（RBM）是在深度神经网络领域广泛使用的两种常见的基础性结构。它们都可以作为无监督学习的框架，通过最小化重构误差，提取系统的重要特征；更重要的是，通过多层的堆叠和逐层的预训练，层叠式自动编码器和深度信念网络都可以在后续监督学习的过程中，帮助整个神经网络更好更快地收敛到最小值点。

关键词： 深度学习；神经网络；AE；RBM

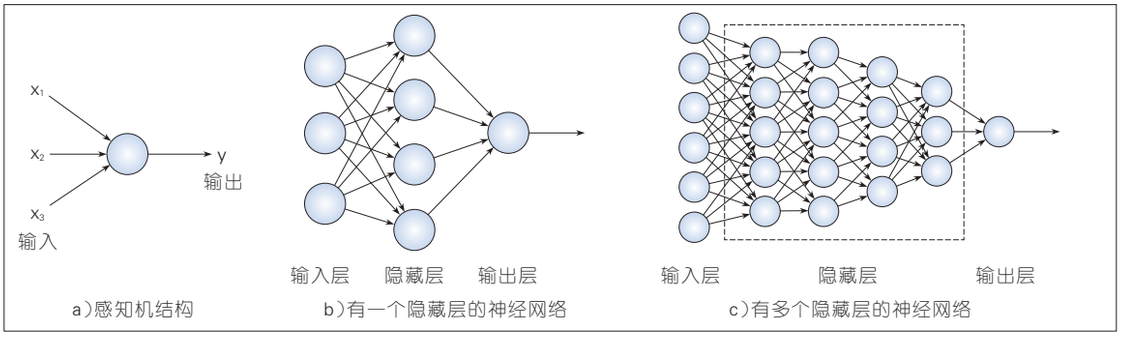
Abstract: Auto-encoders (AE) and Restricted Boltzmann Machines (RBM) are two kinds of basic building blocks which are widely used in the architectures of deep neural networks. By minimizing the reconstruction errors, both the AE and the RBM can extract the key characteristics of the input data and can work as the basic framework of the unsupervised learning. Moreover, with the layer-by-layer stacking and layer-wise pre-training, both the stacked AE and the deep belief networks can help neural networks converge faster and better in the following supervised fine-tuning process.

Key words: deep learning; neural network; AE; RBM

近年来，深度学习在图像和语音识别、自然语言处理、推荐系统等诸多领域中取得了许多重要的突破，深度学习的许多重大进展为解决许多长期以来难以解决的困难问题提供了崭新的思路[1-2]。深度学习以人工神经网络为结构基础，在一个神经网络中，如图 1a）所示，每个神经元都是一个感知机，输入端的数据在线性组合后，经过激活函数引入了非线性因素。在一个神经网络的输入层和输出层之间常常会有一个或者多个隐藏层，如图 1b）和 c）中所示。通过许多个包含不同连接权重的感知机的组合和叠加，一个神经网络因而具有了极强的表示能力。“深度学习”这一名词中的深度指的是神经网络中隐藏层的数量。多个隐藏层让深度神经网络能够表示数据中更为复杂的特征，例如：在用深度卷积神经网络（CNN）进行人脸识别时，较为底层的隐藏层首先提取的是图片中一些边缘和界面的特征，随着层级的提高，图片中一些纹理的特征可能会显现，而随着层级继续提高，一些具体的对象将会显现，例如：眼睛、鼻子、耳朵等，再到更高层时，整个人脸的特征也就被提取了出来。在一个深度神经网络上，较高层的特征是低层特征的组合，而随着神经网络从低层到高层，其提取的特征也越来越抽象、越来越涉及“整体”的性质[3]。

神经网络的训练在本质上是一个非线性优化问题，要求在已知的约束条件下，寻找一组参数组合，使该组合确定的目标函数达到最小。反向传播（BP）算法是人工神经网络训练中的常见方法，在训练的过程中，BP 算法要计算对网络中所有连接的权重计算损失函数的梯度，根据这一梯度值来更新连接的权值，进而最小化损失函数[1]。BP 算法最早在 20 世纪 70 年代被提出，这一算法在浅层的神经网络训练中取得了重要的成功，然而在面对深度神经网络时，这一算法会遇到“梯度消失问题”，即前面的隐藏层中的神经元的学习速度

收稿日期：2017-05-28
网络出版日期：2017-07-06



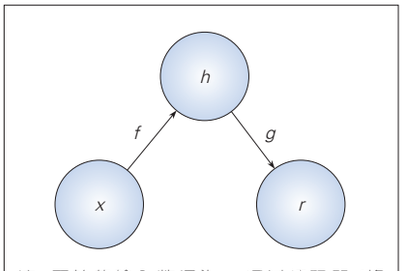
◀图1
神经网络结构示意

要慢于后面的隐藏层，这一问题如果无法解决，那么神经网络将在事实上无法有效利用多个隐藏层。这一困难直到2006年才被加拿大多伦多大学教授Geoffrey Hinton解决，这成为了深度学习领域的标志性事件，它使得神经网络和深度学习重新被学术界所重视[4]。在短短10余年的时间里，深度学习成为了学术界和工业界最为热门的研究主题，在许多不同的领域得到了广泛的应用。深度神经网络也发展出了诸多不同种类的变形。要想真正理解这些不同形式的深度神经网络的工作原理，我们首先必须对这些网络的结构基础进行深入的研究。在文章中，我们将以自动编码器（AE）和限制玻尔兹曼机（RBM）为例，介绍其工作原理和训练方法，在此基础上，我们将讨论这些基本结构在深度学习中的应用。

1 自动编码器

在许多复杂的深度学习问题中，我们都能见到AE的身影。一个AE包括两个基本的组成单元：编码器f和解码器g，两者本身可以是多层的神经网络，它能将输入端的信号在输出端复现出来。AE为了实现这种复现，就必须提取那些输入数据中最为核心的特征，从而实现有效的复现，对原始输入数据的复现称为一次“重构”。一个AE的结构如图2所示，编码器将输入数据x编码到隐藏层h，这一编码过程可以用映射表示为$h = f(x)$，随后解码器将隐藏层的表示解码为输出端的重构结果r，这一解码过程用映射表示即为$r = g(h)$。从概率的角度来看，我们也可以将编码器映射f推广为一个编码分布$p_f(h|x)$，解码器分布可以推广为$p_g(r|h)$。一个好的重构需要尽可能与输入数据相近。因此，学习过程可以简单描述为最小化损失函数$L(x, g(f(x)))$。

长期以来，AE被认为是无监督学习的一种可能的方案，其“编码—解码”过程不是对原始数据的简单重复，在一个编码—解码过程中，我们真正关心的是隐藏层h的特性，例如：当隐藏层h的数据有着比原始输入数据更低的维度，则说明编码器f将复杂的输入数据在隐藏层h用较少的特征重现了出来，这一过程不需要外加的其他标签，因而被称为无监督学习[5]。这与传统的主成分分析（PCA）方法有类似之处，换而言之，在线性的情况下，如果我们按照均方误差来定义惩罚函数L，此时的AE就是经典的PCA。AE提供了一种更为通用的框架，它可以有更好的非线性特性。AE可以用于进行数据降维、数据压缩、对文字或图像提取主题并用于信息检索等。

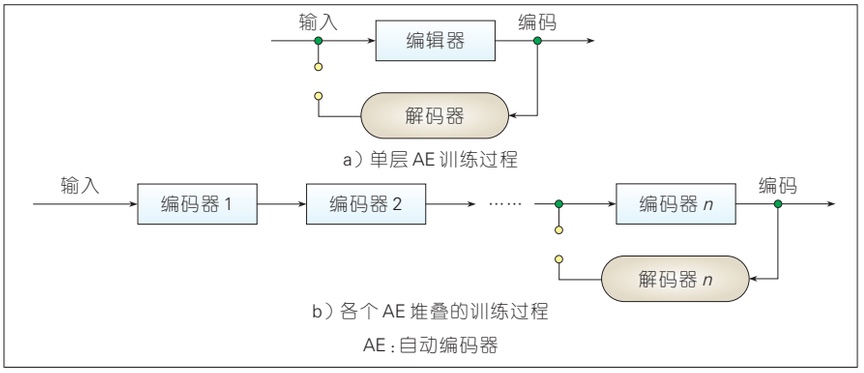


注：原始的输入数据为x，通过编码器f将输入数据映射到h，随后再通过解码器g将h解码为输出数据r。

▲图2 AE的结构

根据所解决的问题不同，AE可以有许多种不同形式的变形，例如：去噪自编码器（DAE）、变分自编码器（VAE）、收缩自编码器（CAE）和稀疏自编码器等[5]。下面我们主要以DAE为例介绍AE的训练过程[6]。当输入的信息中包含噪声时，输入数据x'中包含了真实信息x与噪声的叠加，此时，如果考虑最小化损失函数$L(x, g(f(x')))$，那么我们就得到了一个DAE，其输入为包含噪声的信号x'，输出为消除了噪声的信号x。考虑一个信息的损坏过程$C(x'|x)$，该条件分布代表了给定真实样本x产生包含噪声的输入数据x'的概率。通过最小化损失函数$L = -\log p_g(x|h=f(x'))$，这一最小化过程可以通过经典的BP算法来进行梯度下降，一个训练好的DAE可以从输入数据x'中很好地重构原始数据x。DAE学习了输入信号的更鲁棒的表示方式，其泛化能力比一般的AE更强[1,6]。

多个AE可以逐层堆叠，组合出堆叠式自动编码器（SAE）[7]。SAE的训练通常采用逐层训练的方法来进行：如图3 a）中所示，对于单层的AE，可以通过最小化输入端和经编码—解码得到的输出信号之间的重构误差来对其进行训练；而对于如图3 b）所示的多个AE的堆叠，在训练的过程中，可以采用类似于3 a）中的训练方法逐层对各层编码器进行训练，如果单层的AE已经被训练好，那么可以认为其编码已经能够较好地重构输入数据。在训练第n层编码



▲图3 AE的训练过程

器时，我们考虑对第 $n-1$ 层的编码结果进行再次编码，最小化输入端（即 $n-1$ 层的编码结果）与经编码—解码得到的输出信号之间的重构误差。在逐层训练的过程中，解码器本身对于训练编码器是重要的，而当整个编码器被逐层训练好之后，解码器本身也就不再需要了。

在具体的应用中，除了直接用AE无监督地提取特征以外，更常见的应用是在一个SAE后再接一个分类器，即用提取出来的特征来对系统的状态进行分类——这就变成了一个监督学习的问题。要解决这样的问题，通常分为2个阶段：在第1阶段进行无监督的预训练，通过逐层训练的方式得到一个容易进行下一步训练的神经网络；随后再进行有监督学习进行微调，最终得到一个能较好地处理分类问题的神经网络。这种逐层训练的方法是深度学习的基础[4]，如果直接训练一个深层的自动编码器，那么常常会由于遇到梯度扩散等问题而导致训练效果不佳，而逐层训练的方法可以有效地避免这些问题。

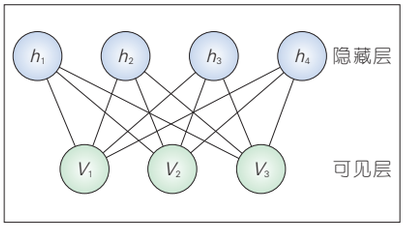
2 从限制玻尔兹曼机到深度信念网络

RBM是实现深度学习的另一种神经网络结构。RBM的基本功能与AE类似，RBM同样也可以对数据进行编码，多层RBM也可以用于对神经网络进行初始化的预训练。RBM的基本结构如图4所示，一个基本的RBM构成了一个二分图，其连接仅在隐藏层和可见层二者之间存在，在隐藏层或可见层的内部并不存在连接，这种特殊连接结构使得RBM的可见层（或隐含层）内各单元间保持了相互独立，这可以简化后续的条件概率计算及采样[1, 8]。

RBM是一种基于能量的神经网络模型，它具有深刻的统计物理背景。在统计物理学中，由于系统倾向于处在能量较低的状态，因此如果某一种状态的能量越低，系统就会有越大的概率处在这一状态。在平衡态的情况下，在服从玻尔兹曼分布时，系统出现某一状态 s 的概率 P 与这一状态的能量 E 的负指数成正比，即 $P \sim e^{-E}$。统计物理中能量与概率分布之间的关系启发我们将概率分布转变成基于能量的模型。此时，系统中复杂而抽象的各种状态出现的概率也就被用一个简单的能量函数给表示出来了。对于一个有 n 个隐含层节点和 m 个输入层节点的RBM，其能量函数 $E(v, h)$ 是这样定义的：

$$E(v,h)=-\sum_{i=1}^{n}\sum_{i=1}^{m}w_{ij}h_i v_j-\sum_{j=1}^{m}b_j v_j-\sum_{i=1}^{n}c_i h_i \quad (1)$$

式（1）中的各 w_{ij} 构成了RBM的连接矩阵，而 b_j 和 c_i 则分别表示了可见层和隐藏层的偏置，它们构成了模型的参数，而 v_j 和 h_i 则是对系统状态的刻画。这个函数与物理学中经典的Ising模型的能量函数是类似的[9]。有了能量函数，系统处在各状态的概率



▲图4 RBM的基本结构示意

分布及各种边缘分布则可得到。

RBM训练的目标即为让RBM网络表示的概率分布与输入样本的分布尽可能地接近，这一训练同样是无监督式的。在实践中，常常用对比散度（CD）的方法来对网络进行训练，CD算法较好地解决了RBM学习效率的问题[10]。在用CD算法开始进行训练时，所有可见神经元的初始状态被设置成某一个训练样本，将这些初始参数代入到激活函数中，可以算出所有隐藏层神经元的状态，进而用激活函数产生可见层的一个重构[10]。通过比照重构结果和初始状态，RBM的各参数可以得以更新，从这一点来看，用CD算法对RBM进行训练与AE的训练是非常相似的。近年来，CD算法有许多改进，例如：持续性对比散度（PCD）和快速持续性对比散度（FPCD）等；而在训练RBM时，也可以利用非CD式的算法，例如比率匹配等[1]。

在RBM的基础上，深度信念网络（DBN）由Hinton等人提出，DBN的工作开启了深度学习的序章[1, 2, 8]。在传统的BP神经网络中，一旦网络的深度增加，训练的过程将会变慢，同时不适当的参数选择将导致网络收敛于局部最优。此外，在训练时还常常需要为训练提供一个有标签的数据集，而DBN则很好地解决了这一问题。与我们提到的SAE类似，也可以通过叠加RBM来建立DBN。第1个RBM是整个网络的输入层，第 l 层RBM的隐藏层作为第 $l+1$ 个RBM的可见层。在训练DBN时，首先用CD算法来训练第1个RBM，然后将其隐藏层作为第2个RBM的可见层，用这

些数据来对第2个RBM用CD算法进行训练，以此类推。最终，当整个网络的所有层都被训练完之后，整个DBN也就被预训练好了。这样一个训练过程是无监督的学习过程，各层RBM都在尽可能反映其上一层数据的特征，这一过程与SAE的训练过程类似，是对网络各参数的初始化。在DBN的最后一层也可以再接一个分类器，进行有监督学习，用BP算法微调整个DBN，这种方法使得DBN克服了直接对深度神经网络进行训练时易出现的局部最优等问题，使得深层神经网络真正有了应用价值[8,11]。

3 自动编码器与限制玻尔兹曼机的区别

AE与RBM两种算法之间也有着重要的区别，这种区别的核心在于：AE以及SAE希望通过非线性变换找到输入数据的特征表示，它是某种确定论性的模型；而RBM以及DBN的训练则是围绕概率分布进行的，它通过输入数据的概率分布（能量函数）来提取高层表示，它是某种概率论性的模型。从结构的角度看，AE的编码器和解码器都可以是多层的神经网络，而通常我们所说的RBM只是一种两层的神经网络。在训练AE的过程中，当输出的结果可以完全重构输入数据时，损失函数 L 被最小化，而损失函数常常被定义为输出与输出之间的某种偏差（例如均方差等），它的偏导数便于直接计算，因此可以用传统的BP算法进行优化。RBM最显著的特点在于其用物理学中的能量概念重新描述了概率分布，它的学习算法基于最大似然，网络能量函数的偏导不能直接计算，而需要用统计的方法进行估计，因此需要用CD算法等来对RBM进行训练。

4 结束语

深度学习是机器学习领域中的重要算法，它通过构建多个隐含层的神经网络和海量的训练数据，来提取更有用的特征，并最终提升分类或预测的准确性。AE和RBM是深度学习的结构基础，它们本身都可以作为一种无监督学习的框架，通过最小化重构误差，提取系统的重要特征；更重要的是，通过多层的堆叠和逐层的预训练，SAE和DBN都可以为后续的监督学习提供一个良好的初值，从而让神经网络可以更好更快地达到收敛[11]。正是这种重要的性质使得深度学习在过去10余年的发展中取得了重要的成功。深度学习能解决的问题变得越来复杂，同时其精度不断提高。

深度学习是一个人工智能领域迅速发展的方向，随着计算能力的提高以及规模更大的数据集合的出现，深度学习的规模也在不断增长，深层神经网络的优势也在不断体现——许多曾经被认为较为抽象、难以分类的复杂特征在深度学习的框架下也变成了可以解决的问题，这使得深度学习算法在许多不同的领域都发挥了重要的应用，并且还有着广阔的应用前景[1,2,12]。更重要的是，深度学习还为解决大量的无监督学习问题提供了可能性。从无标签的数据中进行无监督学习一直以来都是研究人员所面临的一个主要挑战，在这方面，深度学习仍远远无法与人类智能相比，不过近年来，在无监督学习领域也已经出现了许多重要的突破。相信在不久的未来，我们将看到越来越多深度学习和人工智能领域的重大突破，也将看到相关算法在许多新领域的应用和机遇。

参考文献
[1] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep Learning[M]. USA:MIT Press, 2016.
[2] LECUN Y, BENGIO Y, HINTON G. Deep Learning[J]. Nature, 2015, 521(7553): 436–444. DOI:10.1038/nature 14539
[3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet Classification with Deep Convolutional Neural Networks[C]//Advances in Neural Information Processing Systems. USA: Neural Information Processing Systems Foundation, 2012: 1097–1105
[4] HINTON G E, SALAKHUTDINOV R R. Reducing the Dimensionality of Data with Neural Networks[J]. Science, 2006, 313 (5786): 504–507
[5] BALDI P. Autoencoders, Unsupervised Learning, and Deep architectures[J]. ICML Unsupervised and Transfer Learning, 2012, 27(37–50): 1
[6] VINCENT P, LAROCHELLE H, BENGIO Y, et al. Extracting and Composing Robust Features with Denoising Autoencoders [C]// Proceedings of the 25th International Conference on Machine Learning. USA: ACM, 2008: 1096–1103
[7] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion[J]. Journal of Machine Learning Research, 2010 (11): 3371–3408
[8] HINTON G E, OSINDERO S, TEHY W. A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2006, 18(7): 1527–1554
[9] MEHTA P, SCHWAB D J. An Exact Mapping Between the Variational Renormalization Group and Deep Learning[J]. DOI:1410.3831, 2014
[10] HINTON G E. Training Products of Experts by Minimizing Contrastive Divergence[J]. Neural Computation, 2002, 14(8): 1771–1800
[11] ERHAN D, BENGIO Y, COURVILLE A, et al. Why Does Unsupervised Pre–Training Help Deep Learning?[J]. Journal of Machine Learning Research, 2010(11): 625–660
[12] SCHMIDHUBER J. Deep Learning in Neural Networks: An Overview[J]. Neural networks, 2015(61): 85–117

作者简介

康文斌，湖北医药学院教师，湖北医药学院Bio–X研究中心主要负责人；研究方向为软凝聚态物理与生物物理、蛋白质物理学问题以及与机器学习算法在药物设计等领域中的应用；已发表2篇SCI论文，2篇EI论文以及3篇其他期刊论文。

彭菁，平安科技（深圳）有限公司数据分析师；研究方向为无监督学习在用户行为数据分析中的应用，尤其关注用户在各金融领域（保险、信贷、理财等）行为轨迹研究以及相关问题的商业价值探索。

唐乾元，香港浸会大学物理系博士后；研究方向为统计物理在生命科学和人工智能领域中的应用，包括深度学习算法的统计物理基础以及深度学习在生物大数据分析中的应用；已发表SCI论文2篇。