深度学习理论及其应用专题讲座(一)

第 2 讲 深度学习基本理论概述

陈栩杉[1]，张雄伟[2]，乔 林[1]，王 健[3]，李治中[4]
(1. 解放军理工大学指挥信息系统学院研究生 1 队，江苏 南京 210007；2. 解放军理工大学指挥信息系统学院；
3. 中国人民解放军驻 9373 厂军事代表室，安徽 蚌埠 233010；4. 解放军理工大学野战工程学院)

摘 要：深度学习作为机器学习领域的新课题，在学术界和工业界引起了广泛关注，掀起了大数据与人工智能发展的新浪潮。深度学习通过模拟人脑的分层结构，建立了从底层到高层逐级提取输入数据特征的模型，能够深刻揭示从底层信号到高层语义的映射关系。文章从深度学习在互联网、语音图像处理等领域取得的显著成就出发，介绍了深度学习的理论框架，详细阐述了深度学习最为关键的训练过程，概述了三种典型的深度学习模型，包括自动编码器模型、受限玻尔兹曼机模型和深信度网络模型，最后探讨了深度学习所面临的机遇和挑战，以及有待进一步研究解决的问题。

关键词：人工智能；机器学习；深度学习；自动编码器；受限玻尔兹曼机；深信度网络
中图分类号：TP18 **文献标识码**：A **DOI**：10.16464/j.cnki.cn32-1289.2015.04.020

Overview of Basic Theory in Deep Learning

CHEN Xu-shan[1], *ZHANG Xiong-wei*[2], *QIAO Lin*[1], *WANG Jian*[3], *LI Zhi-zhong*[4]
(1. Postgraduate Team 1 CCIS, PLAUST, Nanjing 210007, China;
2. College of Command Information Systems, PLAUST;
3. PLA Military Agency Office of 9373 Factory, Bengbu 233010, China;
4. College of Field Engineering, PLAUST)

Abstract: As a new research field of machine learning, deep learning has attracted wide attention in both academic and industrial community, and has become a huge development of big data and artificial intelligence. Deep learning builds up the hierarchical model which extracts the data features from the bottom level to the top level by simulating the hierarchical structure of the human brain. The model can reveal the mapping relationship from the underlying signal to the high-level semantics. According to the remarkable achievements of deep learning, such as Internet, speech and image processing, the theoretical framework of deep learning was introduced. Then the training process was elaborated which is crucial for deep learning, and three state-of-the-art deep learning networks reviewed including autoencoder model, restricted Boltzmann machine model and deep belief network model. Finally the opportunities and challenges of deep learning, and the prospecting research fields were discussed.

Key words: artificial intelligence; machine learning; deep learning; auto-encoder; restricted Boltzmannmachine; deep belief network

收稿日期：2015-10-09；**修回日期**：2015-10-19
基金项目：国家自然科学基金资助项目(61402519,61471394)、江苏省自然科学基金资助项目(BK20140071,BK20140074,BK2012510)
作者简介：陈栩杉(1987—)，男，博士生.

1　深度学习的来源与发展

深度学习 DL(Deep Learning)[1-3]是指通过多层神经网络拟合训练样本分布的一种机器学习方法，它缓解了传统神经网络算法在训练多层神经网络时出现的局部最优问题[4]，且其训练过程不依赖于样本标签信息。深度学习的这一特性决定了其适合于处理非线性的自然信号，如图像识别、语音识别、自然语言处理、大数据特征提取等方面，为许多面临瓶颈的信号处理问题提供了新的尝试方法。

2000 年，Hinton 等人提出一种适合训练马尔可夫随机场模型 MRF(Markov Random Fields)[5]的新算法，称为对比散度 CD(Contrastive Divergence)[6-7]算法，其中受限玻尔兹曼机 RBM(Restricted Boltzmann Machine)[7]就是一种典型的马尔可夫随机场模型，这为深度学习的诞生奠定了基础。

2006 年，Hinton 等人提出一种新的深度神经网络模型——深信度网络 DBN(Deep Belief Net)[8]，对比散度算法可以用来对该模型进行高效训练，深信度网络模型也逐渐成为深度学习的主流框架。在该框架中，一个 DBN 由若干个 RBM 堆叠起来，训练采用由低到高的逐层训练方法，由于单个 RBM 可以通过 CD 算法快速训练，因此整个框架的训练简化为多个 RBM 的训练问题，解决了深度网络训练的高复杂度问题。

自 2010 年以来，深度学习在实际应用中取得了引人瞩目的成功。谷歌公司的 Google Brain[9]项目利用 16 000 个处理器构建了一个超过 10 亿节点的大规模深度网络，研究成果已经应用于谷歌的图像搜索、无人驾驶和 Google Glass 等项目中。在语音识别方面，深度学习给出了更好的声学模型——深度神经网络 DNN (Deep Neural Network)[10]，大幅提高了语音识别准确率，微软研究院和谷歌的语音识别研究团队采用 DNN 技术将语音识别错误率降低了 20%～30%，这是语音识别领域十多年来最大的突破性进展。在图像处理方面，研究者将 DNN 技术应用于图像识别领域，在 ImageNet 上进行评测，将识别错误率从 26%降低到 15%。除此之外，DNN 技术还广泛应用在行人检测、图像分割、交通标志分类等方面，取得了惊人的效果。

2　深度学习的理论框架

解决许多人工智能问题的前提是从样本中设计并提取出有用的特征，再根据该特征采用合适的机器学习算法进行分类或预测处理。特征选取得好坏对最终结果的影响很大，因此，如何选取特征对于解决一个实际问题至关重要。例如对于说话人识别问题来说，基音周期(pitch)是最为关键的特征，基音周期是声带振动频率的周期，通过基音检测可以判断出说话人是男人、女人或孩子。

然而，在许多问题中事先并不知道需要提取哪种特征，同时一些外在因素的变化也会影响观测到的原始数据。例如对于一副夜晚的红色汽车的图像，汽车的某些像素近似于黑色，汽车轮廓的形状依赖于观察的角度等。这些高层抽象的特征很难直接从原始数据中提取出来，人工选取特征又需要大量经验和技巧，费时费力。那么是否能在无监督的条件下，不需要人为参与特征的选取过程，自主的从原始数据中学习出相应特征呢？深度学习给出了肯定的答案。

2.1　深度学习的基本思想

假设系统 S 有 n 层，S_1，…，S_n，它的输入是 I，输出是 O，如果输出 O 等于输入 I，即输入 I 经过这个系统之后没有变化，这意味着输入 I 经过每一层 S_i 都没有任何的信息损失，或者丢失的信息是冗余的。换句话说，在任何一层 S_i 的输出都是输入 I 的另一种表现形式。深度学习的基本思想就是堆叠多个层，把上一层的输出作为下一层的输入，通过这种方式实现对输入信息的分级表达。图 1 形象的说明了深度学习模型是如何从一副图像中提取简单的特征，用于目标分类识别的。从图 1 中可以看出，直接建立从像素到目标身份的映射十分困难，深度学习将其转化为一系列简单映射，每种映射作为模型的一层。

由此看出，这种分层的无监督的特征学习是深度学习的重要基础，它通过逐层的特征变换，将样本在原

空间的特征表示映射到一个新的特征空间中，用大量简单的特征构建复杂的表示，消除输入数据中与学习任务无关因素的改变对学习性能的影响，保留对学习任务有用的信息。

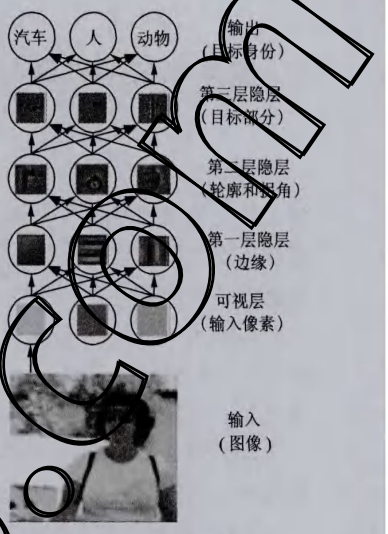


图 1 深度学习模型举例

2.2 深度学习与神经网络

多层神经网络是较早提出的一种深度结构。1986 年，反向传播算法 BP (Back Propagation)[11-12]的出现给机器学习带来了希望，掀起了基于统计模型的机器学习热潮。BP 算法可以让一个人工神经网络模型从大量训练样本中学习出统计规律，从而对未知事件进行预测。与以往的基于人工规则的系统相比，BP 算法在许多方面具有优越性。然而，由于 BP 算法属于有监督学习，需要训练样本的标签信息，而在处理实际问题时，常常无法获取大量训练样本的标签信息。另外，多层神经网络由若干层非线性单元构成，每一层非线性单元的代价函数都是非凸函数，在优化过程中容易陷入局部最优的情况，因此当多层叠加之后，BP 算法对模型的训练结果往往很差。

虽然以 BP 为基础的人工神经网络也被称为多层感知机 MLP(Multi-Layer Perceptron)[13]，由于多层网络训练的困难，实际使用的大多是只含有一层隐藏节点的浅层模型。为此，Hinton 等人经过二十多年的潜心研究，最终提出了一个切实可行的深度学习框架。如图 2 所示，深度学习可以理解为神经网络的延伸和发展，是机器学习研究中的一个新领域，其动机在于建立模拟人脑进行分析学习的神经网络，仿照人脑的机制来解释数据，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，具体训练流程如图 3 所示。虚线方框表示从原始数据中学习得到的信息。

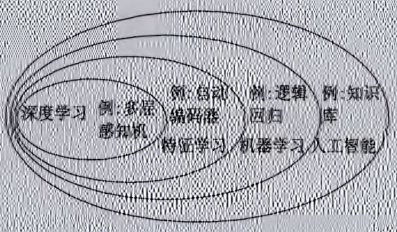


图 2 不同人工智能方法关系图

从图 2 和图 3 可以看出，深度学习与传统的神经网络之间有着千丝万缕的联系。两者的相同点在于深度学习框架采用了类似神经网络的分层结构，系统是一个多层网络，包括输入层、隐藏层（多层）、输出层，只有相邻层节点之间有连接，同一层以及跨层节点之间相互无连接，每一层可以看作是一个逻辑斯蒂回归 LR(Logistic Regression)[14]模型，这种分层结构是比较接近人类大脑的结构的。

而两者的区别在于，为了克服神经网络训练中的问题，深度学习采用了与神经网络完全不同的训练机制。传统神经网络中，采用的是 BP 算法来训练整个网络，整体是一个梯度下降方法，即随机设定初值，计算当前网络的输出，然后根据当前输出和标签之间的误差去改变前面各层的参数，直到收敛。这样做带来的缺陷是梯度越来越稀疏，会出现梯度扩散现象，且从顶层越往下误差校正信号越来越小。深度学习采用逐层训练的机制能够很好的解决上述问题，每次训练一层网络，同时使本层特征表示 r_k 向上生成的高级表示 r_{k+1} 与该高级表示 r_{k+1} 向下生成的 r'_k 尽量保持一致。

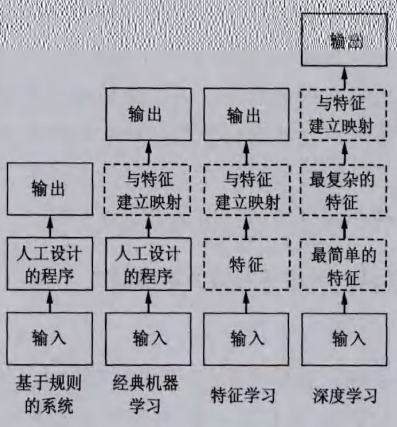


图 3 各类人工智能系统训练流程图

2.3 深度学习的训练过程

正如之前所说，非凸目标代价函数收敛到局部最优是深度网络训练困难的主要因素。如果对所有层同时训练，时间复杂度会太高；如果每次训练一层，误差就会逐层传递，最终出现输入和输出严重欠拟合的问题。

为此，深度学习采用“两步走”的方式学习各层的参数，即自下而上的非监督学习和自上而下的有监督学习。下面分别进行介绍。

2.3.1　自下而上的非监督学习

这一步是一个无监督的训练过程，从最底层开始，采用无标签的数据分层训练各层参数，将第 $n-1$ 层的输出作为第 n 层的输入，以此类推，逐层向顶层训练。这一步类似于传统神经网络的初值随机初始化过程，但这里的初值是通过学习输入数据的结构得到的，比随机初始化更接近全局最优。

当所有层训练完之后，除最顶层之外，将其它层与层之间的连接变为双向连接，这样一来，最顶层仍然保持一个单层神经网络，其它层则变成了图模型。对于连接的权重，向上的表示“认知权重”，向下的表示“生成权重”，采用 Wake-Sleep 算法调整所有的权重，使得认知和生成达成一致，尽可能保证生成的最顶层表示能够恢复最底层的结点。

(1) Wake 过程。醒的时候是一个认知过程，通过外界的特征和认知权重产生各层的节点值，同时采用梯度下降方法修改各层之间的生成权重。这一过程可以形象的描述为“如果现实跟我想象的不一样，那就改变我的生成权重，使我想象的东西就是现实这样的”。

(2) Sleep 过程。睡着或者做梦的时候是一个生成过程，通过醒的时候学习的顶层表示和生成权重，生成底层的状态，同时修改各层之间的认知权重。这一过程可以形象的描述为“如果我想象的事物与梦到的景象不一致，那就改变我的认知权重，使这种景象在我看来就是我所想象的事物”。

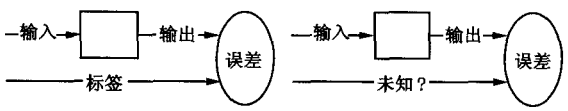
2.3.2　自上而下的有监督学习

这一步是一个有监督的训练过程，在第一步学习得到的各层参数的基础上，通过带标签的数据去训练，误差自顶向下传递，对各层参数进行微调。

3　常用的深度学习模型

3.1　自动编码器

在传统的神经网络中，输入的数据是有标签的，根据输出和标签之间的误差来调节网络各层的参数，如图 4(a)所示，直到收敛。但是如果事先并不知道输入数据的标签，如图 4(b)所示，那么如何去训练整个网络的参数呢？这就是设计自动编码器 AE(Auto-Encoder)[15] 的动机和初衷。



(a) 有标签训练　　(b) 无标签训练

图 4　神经网络训练示意图

自动编码器让输入数据经过一个编码器得到一个编码输出，再将该输出导入一个解码器得到最终的输出，由于输入数据是无标签数据，此时的误差来自于输出和原输入之间的比较。通过调整编码器和解码器的参数，使误差达到最小，就能得到输入信号的另一种表示 r，如图 5(a)所示。将多个编码器串联起来，把第 k 层输出的表示 r_k 看作是第 $k+1$ 层的输入，同理，最小化通过解码器重构的输出与输入之间的误差就能得到第 $k+1$ 层的参数，并且得到第 $k+1$ 层输出的 r_{k+1}，即原输入数据的第 $k+1$ 个表示，如图 5(b)所示，虚线框表示之前训练出的各层参数已经固定，不再变化。

假设经过多层的训练，自动编码器已经学习到一个良好的特征来表示原输入数据，那么可以在自动编码器的最顶层添加一个分类器，如 LR 回归、支持向量机 SVM(Support Vector Machine)[16] 等，利用梯度下降方法对整个网络进行有监督的微调。一旦完成了这个有监督训练，整个神经网络就可以用来分类了。如果对自动编码器加上一些约束就能得到新的深度学习方法，例如稀疏自动编码器、降噪自动编码器等。

3.2　受限玻尔兹曼机

受限玻尔兹曼机是一个二分图模型，由可视层 v 和隐藏层 h 组成，所有的节点都是随机二值变量节点(0 或 1)，全概率分布 $P(v, h)$ 满足 Boltzmann 分布，本层节点之间没有连接，如图 6 所示。无论在可视层还是隐藏层，节点都是独立的，因此在已知可视层 v 的情况下，$P(h|v)=\prod_{i=1}^{n} P(h_i|v)$，同理在已知隐藏层 h 的

前提下，根据 $P(v'|h)$ 可以得到可视层，v' 表示根据隐藏层估计出的可视层单元。通过调整参数，如果从隐藏层得到的 v' 与原来的可视层 v 相同，就可以认为隐藏层是可视层输入数据的特征表示，图7给出了RBM的训练过程。实际上，RBM是一种能量模型，其能量函数为：

$$E(\boldsymbol{v}, \boldsymbol{h}, \theta) = -\sum_{i,j} W_{ij} v_i h_j - \sum_i b_i v_i - \sum_j a_j h_j \quad (1)$$

其中 $\theta = \{W, a, b\}$ 是模型参数，E 为期望值。通过正则分布为RBM定义了可视节点和隐藏节点的联合概率分布：

$$P_\theta(\boldsymbol{v}, \boldsymbol{h}) = \frac{1}{Z(\theta)} e^{-E(\boldsymbol{v}, \boldsymbol{h}, \theta)} = \frac{1}{Z(\theta)} \prod_{i,j} e^{W_{ij} v_i h_j} \prod_i e^{b_i v_i} \prod_j e^{a_j h_j}$$

$$Z(\theta) = \sum_{\boldsymbol{v}, \boldsymbol{h}} e^{-E(\boldsymbol{v}, \boldsymbol{h}, \theta)} \quad (2)$$

给定可视层 $\boldsymbol{v}$，第 j 个隐藏层节点为1的概率为：

$$P(h_j = 1 \mid \boldsymbol{v}) = \frac{1}{1 + e^{-\sum_i W_{ij} v_i - a_j}} \quad (3)$$

同理，给定隐藏层 $\boldsymbol{h}$，第 i 个可视层节点为1的概率为：

$$P(v_i = 1 \mid \boldsymbol{h}) = \frac{1}{1 + e^{-\sum_j W_{ij} h_j - b_i}} \quad (4)$$

给定一个原始数据集 $d = \{v_1, v_2, \cdots, v_n\}$，满足独立同分布，构建下列对数似然函数：

$$L(\theta) = \frac{1}{N} \sum_{i=1}^{n} \log P_\theta(v_i) - \frac{\lambda}{N} \sum_{i=1}^{n} \sum_{j=1}^{n} |W_{ij}|^2 \quad (5)$$

上式对 $\boldsymbol{W}$ 求偏导得到：

$$\frac{\partial L(\theta)}{\partial W_{ij}} = E_{P_d}(v_i h_j) - E_{P_\theta}(v_i h_j) - \frac{2\lambda}{N} W_{ij} \quad (6)$$

令上式为0即可得到 W_{ij}：

$$W_{ij} = \frac{N}{2\lambda} [E_{P_d}(v_i h_j) - E_{P_\theta}(v_i h_j)] \quad (7)$$

如图8所示，如果把隐藏层的层数增加，我们可以得到深度玻尔兹曼机DBM（Deep Boltzmann Machine）[17]，如果在靠近可视层的部分使用贝叶斯信念网络，而在最远离可视层的部分仍然使用RBM，可以得到深信度网络。

3.3 深信度网络

深信度网络是一种贝叶斯概率生成模型，由多层随机隐藏变量组成，上面两层具有无向对称连接，下面的层得到来自上一层的有向连接。如图8所示，DBN的基本结构单元是RBM，每个RBM单元的可视层节点个数等于前一RBM单元的隐藏层节点个数。

在深信度网络框架中，最上面两层构成联想记忆（associative memory），其余各层之间的连接是通过自顶向下的生成权值来指导确定的。在训练过程中，先将可视层单元的值映射给隐藏层单元，然后可视层单元由

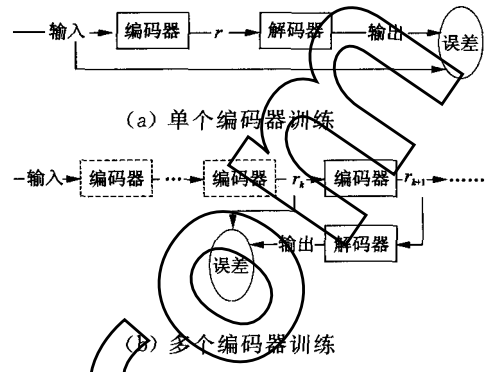


图5 自动编码器训练示意图

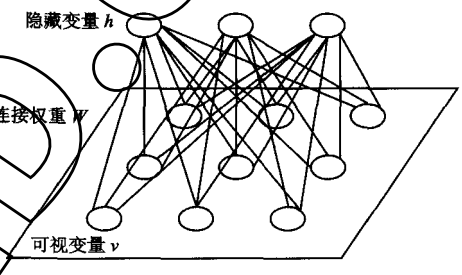


图6 RBM模型

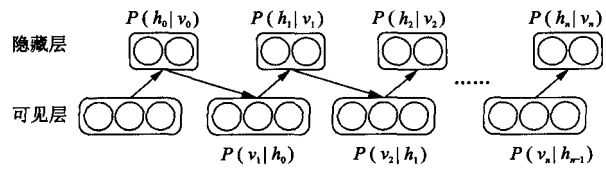


图7 RBM训练过程

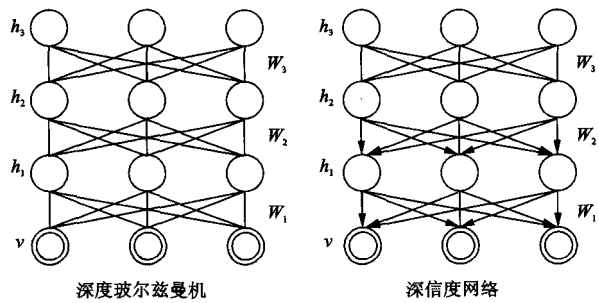
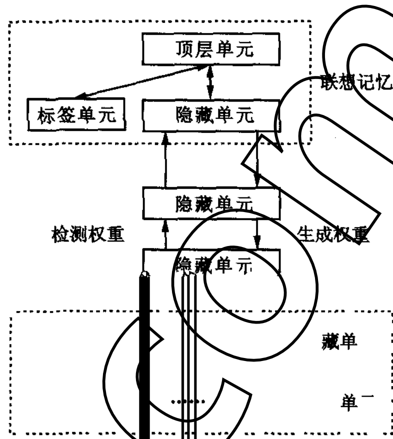


图8 从RBM衍生出的深度网络

隐藏层单元重建，这些新的可视层单元再次映射给隐藏层单元，获取到新的隐藏层单元，这种反复前进和后退的步骤称为吉布斯采样(gibbs sampling)[18]。由于可视层输入与隐藏层激活单元之间的概率分布差异是权值更新的主要依据，因此采用对比散度算法去预训练获得生成模型的权值，训练时间会显著减少，只需要很少几步就可以收敛。



对于最顶两层，下层的输出给顶层提供了相关的参考线索，顶层 据 索 输出 系 的 … 中有 … 顶层，DB … 训练之 … 以 … 的数据，用B 算法去 … 性能 …。经过这种 … 理的 … 性能会比单 … 用B … 法训练的要 … ，因为 B 的B 算法 … 要 … 权 … 参数 … 一个 … 的 … 索，训练 收 … 。

R M … 像是 … 的 … 的学习， 得DBN 有 … 的 … 属性 … ，深 … 度网络 CDBN Convol t onal Dee Be ie … et … 是其中的一种 DBN … 图像的 … 像 … 的 … 过卷 … RBM 模 … 到 … 型的变 … 性，而 … 以 … 变 … 得 … 。 … 层 … 机 … 卷 … 机 … M … m … l … nvol … o … Ma … ）等深度 … 构 … 模型 … 问题 … 来了 … 的未来 …

深度学 … 在多个 … 的 … 性 … 进展，比如 … 算机 … 识别等， … 别 … 在 … 模 … 集 … 的优 … 有 … 然而深度学习 … 是 … 的，然 … 在 … 多有 … 进一步 … 的问题， … 在深度学习 … 深度模型 … 深度网络 … 训练和优 … 要 … 。

（ … 要 … 少训练 … 据 … 学习 … 的深度网络模 … ？深度学习模 … 训练 … 的 … 因是 … ？重构 … ，是 … 其 … 合适的 … 示 … 度 … 网络的训练过 … ？ … 问题 … 广的 …

（ … 应用问题，如 … 计一个 … 适合的深度 … 型 … 门 … 如 … 一个 … 用的深度模型 … 的 … 来处理 … 图 … 语 … 。

（ … 于最 … 处理的 … 机 … 度 … 算 … 很 … 再多 … 计算机 … 并行 … 用 … 处理单元 … U … 学习过 … 是单 … 机 … 的 … U … 法用于 … 理 … 模数据集的学习，因 … 有 … 的并行 … 算法来训练 … 度模 … 。

：

[1] … he … vk A，Su … kever I … inton … E. I … agenet classi ication with deep con ol tiona … ne … ralne … orks[C] / Advances in … eura … n … or … ation Processi … g S stems. … ed Hook，USA：Curran A … ciates Press，2 … ：… 09 -1105.

[2] … ahl … ng L，et al. … onte … depen … en … p … e traine … deep neural net … orks for la … ge- … oca … lar speech recognit on …]. IEEE Trans … t ons on … ud … o Speech and Lang … age Proce … sing 2012， 0 … 1)：30-42

[3] … e … H Senior … ch … ster … . Statistic … l … ara … etric … peech synth … e … using deep ne … ral ne … orks[…]/ IE … E I … ternational … onference on … oustics Speec … and Si … nal Processing. Piscata … ay，USA：IEEE … ress … 7962 … 966.

[4] Hinton … E. How neural networks learn fro … e perience[… . Scientific American，1 … 2，2 … 7 … 3 … ：1 … 5 15 .

[5 … I … h … ka … a … . E … a … t optim … ation for … arkov random fie ds with con … e … priors[ ]. EEE Transactions o … Patte n Anal … sis … a … h … ne … ntelli … ence，003， … 0 … ：1333-133 … .

[6] Hinton G E. Training products of experts by minimizing contrastive divergence[J]. Neural Computation, 2000, 14(8): 1771-1800.

[7] Hinton G E. A practical guide to training restricted boltzmann machines[EB/OL]. (2010-08-02)[2015-10-19]. http://www.cs.toronto.edu//~hinton/absps/guideTR.pdf.

[8] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.

[9] Markoff J. How many computers to identify a cat? [N]. The New York Times, 2012-06-25(10).

[10] Sainath T N, Kingsbury B, Sindhwani V, et al. Low-rank matrix factorization for deep neural network training with high-dimensional output targets[C]// 2013 IEEE International Conference on Acoustics Speech and Signal Processing. Vancouver, Canada: IEEE Press, 2013: 6655-6659.

[11] Nebauer C. Evaluation of convolutional neural networks for visual recognition[J]. IEEE Transactions on Neural Networks, 1998, 9(4): 685-696.

[12] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533-536.

[13] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators[J]. Neural Networks, 1989, 2(5): 359-366.

[14] Pregibon D. Logistic Regression Diagnostics[J]. Annals of Statistics, 1981, 9(4): 705-724.

[15] Vincent P, Larochelle H, Lajoie I, et al. Stacked denoisingautoencoders: learning useful representations in a deep network with a local denoisingcriterion[J]. Journal of Machine Learning Research, 2010, 11(6): 3371-3408.

[16] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.

[17] Raiko T, Cho K H, Ilin A. Gaussian-Bernoulli deep Boltzmann machine[C]// International Joint Conference on Neural Networks. Dallas, USA: IEEE Press, 2013: 1-7.

[18] George E I, Mcculloch R E. Variable selection via Gibbs sampling[J]. Journal of the American Statistical Association, 1993, 88(3): 881-889.

[19] Honglak L, Roger G, Rajesh R, et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations[C]// International Conference on Machine Learning. New York: ACM Press, 2009: 609-616.

[20] Lockett A J, Miikkulainen R. Temporal convolution machines for sequence learning[EB/OL]. (2009-08-26)[2015-10-19]. http://www.cs.utexas.edu/users/nn/downloads/papers/lockett-tcm.pdf.

(上接第 40 页)

tions on Microwave Theory and Techniques, 2002, 50(3): 594-611.

[2] Deng H W, Zhao Y J, Fu Y, et al. Compact and high selectivity cross-coupled broadband microstrip bandpass filter[J]. Microwave and Optical Technology Letters, 2013, 55(10): 2501-2504.

[3] Deng P H, Lin Y S, Wang C H, et al. Compact microstrip bandpass filters with good selectivity and stopband rejection[J]. IEEE Transactions on Microwave Theory and Techniques, 2006, 54(2): 533-539.

[4] Djaiz A, Denidni T. A new compact microstrip two-layer bandpass filter using aperture-coupled SIR-hairpin resonators with transmission zeros[J]. IEEE Transactions on Microwave Theory and Techniques, 2006, 54(5): 1929-1936.

[5] Tu W H. Sharp-rejection broadband microstrip bandpass filter using penta-mode resonator[J]. Electronics letters, 2010, 46(11): 772-773.

[6] 邓 哲, 程崇虎, 吕文俊, 等. 微带发夹型谐振器滤波器的实验研究[J]. 微波学报, 2005, 21(1): 122-126.

[7] Hong J S, Lancaster M J. Couplings of microstrip square open-loop resonators for cross-coupled planar microwave filters [J]. IEEE Transactions on Microwave Theory and Techniques, 1996, 44(12): 2099-2109.

[8] 李明洋, 郭陈江. 微带抽头线发夹型滤波器设计[J]. 电子工程师, 2003, 29(9): 57-60.