江苏大学学报(自然科学版)
JOURNAL OF JIANGSU UNIVERSITY(Natural Science Edition)
2015 年 3 月 Mar. 2015
第 36 卷 第 2 期 Vol. 36 No. 2
doi: 10.3969/j.issn.1671 -7775.2015.02.013

深度学习的研究与发展

张建明, 詹智财, 成科扬, 詹永照
(江苏大学 计算机科学与通信工程学院, 江苏 镇江 212013)

摘要: 针对以往浅层学习对特征表达能力不足和特征维度过多导致的维数灾难等现象,深度学习通过所特有的层次结构和其能够从低等级特征中提取高等级特征很好地解决了这些问题,并给人工智能带来了新的希望. 首先介绍了深度学习的发展历程,并介绍了基于 restricted boltzmann machines (RBM)、auto encoder (AE)和 convolutional neural networks (CNN)的 deep belief networks (DBN)、deep boltzmann machine (DBM)和 stacked auto encoders (SAE)等深度模型. 其次,对近几年深度学习在语音识别、计算机视觉、自然语言处理以及信息检索等方面的应用的介绍,说明了深度学习结构在相比较于其他结构的优越性和在不同任务上更好的适应性. 最后通过对现有的深度学习在在线学习能力、大数据上和深度结构模型的改进上的思考和总结,展望了今后深度学习的发展方向.

关键词: 浅层学习; 深度学习; 层次结构; 人工智能; 机器学习
中图分类号: TP301 **文献标志码**: A **文章编号**: 1671 -7775(2015)02 -0191 -10
引文格式: 张建明, 詹智财, 成科扬, 等. 深度学习的研究与发展[J]. 江苏大学学报:自然科学版,2015,36(2):191 -200.

Review on development of deep learning

Zhang Jianming, Zhan Zhicai, Cheng Keyang, Zhan Yongzhao
(School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China)

Abstract: To solve the insufficiency of shallow learning expression ability and the excessive dimension disaster due to the feature dimension, deep learning was used due to the unique hierarchies and capability of extracting high level features from low-level features, and brought new hope for artificial intelligence. The development of deep learning during different periods was introduced. The basic models of RBM, AE and CNN were analyzed to present the deep hierarchical structures of DBN, DBM and SAE. The applications of deep learning in the fields of speech recognition, computer vision, natural language processing and information retrieval in recent years were introduced to illustrate the superiority and flexibility of deep learning compared with other shallow learning algorithms. Some future research directions were predicted based on the analysis, and some conclusions were made according to the improvement of deep learning on algorithm generalization, adaptation of big data and modifying on deep structure.

Key words: shallow learning; deep learning; hierarchical structure; artificial intelligence; machine learning

收稿日期: 2014 -05 -08
基金项目: 国家自然科学基金资助项目(61170126)
作者简介: 张建明(1964—),男,江苏丹阳人,教授(zhjm@ujs.edu.cn),主要从事图像处理与模式识别、虚拟现实的研究.
詹智财(1989—),男,江西上饶人,硕士研究生(342888106@qq.com),主要从事模式识别、深度学习的研究.

从2006年开始,深度学习作为机器学习领域中对模式(音频、图像、文本等)进行建模的一种方法已经成为机器学习研究的一个新领域.深度学习旨在使机器学习能够更加地接近其最初的目标——人工智能[1].

近年来,随着深度学习的出现,许多研究者致力于深度学习原理和应用的研究,主要体现在各大会议、高校研究组和企业应用上的热潮.会议包括:2013年声学,语音和信号处理国际会议(international conference on acoustics, speech, and signal processing, ICASSP)讨论关于语音识别和相关应用的深度神经网络学习的新类型;2010,2011和2012年神经信息处理系统(neural information processing systems, NIPS)讨论关于深度学习和无监督特征学习;2011,2013年机器学习国际会议(international conference on machine learning, ICML)讨论关于音频,语音和视觉信息处理的学习结构,表示和最优化[2].高校团队有:多伦多大学的Geoffrey Hinton研究组;斯坦福大学的Andrew Ng研究组;加拿大蒙特利尔大学的Yoshua Bengio研究组;纽约大学的Yann LeCun研究组等[3].企业团队有:百度公司的Andrew Ng与余凯团队;微软公司的邓力团队;Google公司的Geoffrey Hinton团队和阿里巴巴,科大讯飞以及中科院自动化所等公司或研究单位.

在深度学习中,深度指代在学到的函数中非线性操作组成的层次的数目.早在1969年Minsky和Papert在所著的《感知机》中就指出:单层感知机(浅层结构)不能实现“异或”(XOR)功能,即不能解决线性不可分问题.而多层感知机,即深度结构是可以求解线性不可分的问题的,深度结构将低等级特征组合或者变换得到更高等级形式的特征,并从中学习具有层次结构的特征,这种特有的结构允许系统在多层次的抽象中自动的学习并能够拟合复杂的函数.因为无监督自动学习数据中隐藏的高等级特征的能力会随着数据的规模的扩大和机器学习方法的应用范围增大而变得越来越重要,深度学习也会被越来越多的研究者重视.文中意在通过对深度学习的基本模型的介绍以及在几大领域上的应用,使读者能够对深度学习有大致的了解[4].

1 深度学习的发展历程

机器学习的发展历程可以大致分为2个阶段:浅层学习和深度学习.直到近些年,大多数机器学习的方法都是利用浅层结构来处理数据,这些结构模型最多只有1层或者2层非线性特征转换层.典型的浅层结构有:高斯混合模型(GMMs)[5]、支持向量机(SVM)[6]、逻辑回归等等.在这些浅层模型中,最为成功的就是SVM模型,SVM使用一个浅层线性模式分离模型,当不同类别的数据向量在低维空间中无法划分时,SVM会将它们通过核函数映射到高维空间中并寻找分类最优超平面.到目前为止,浅层结构已经被证实能够高效地解决一些在简单情况下或者给予多重限制条件下的问题,但是当处理更多复杂的真实世界的问题时,比如涉及到自然信号的人类语音、自然声音、自然语言和自然图像以及视觉场景时他们的模型效果和表达能力就会受到限制,无法满足要求[2].

早在1974年Paul Werbos提出了反向传播(back propagation, BP)算法[7],解决了由简单的神经网络模型推广到复杂的神经网络模型中线性不可分的问题,但反向传播算法在神经网络的层数增加的时候参数优化的效果无法传递到前层,容易使得模型最后陷入局部最优解,也比较容易过拟合.在很长一段时间里,研究者们不知道在有着多层全连接的神经网络上怎样高效学习特征的深度层次结构.

2006年,Hinton提出了深度置信网络(deep belief network, DBN)[8],这个网络可以看作是由多个受限玻尔兹曼机(restricted boltzmann machines, RBM)[9]叠加而成.从结构上来说,深度置信网络与传统的多层感知机区别不大,但是在有监督学习训练前需要先无监督学习训练,然后将学到的参数作为有监督学习的初始值.正是这种学习方法的变革使得现在的深度结构能够解决以往的BP不能解决的问题.

随后深度结构的其他算法模型被不断地提出,并在很多数据集上刷新了之前的一些最好的记录,例如2013年Wan Li等[10]提出的drop connect规范网络,其模型在数据集CIFAR-10上的错误率为9.32%,低于此前最好的结果9.55%,并在SVHN上获得了1.94%的错误率,低于此前最好的结果2.8%等等.

2 深度学习的基础模型及其改进

深度学习出现的时间还不算长,所以大部分模型都是以最基础的几种核心模型为基元,例如RBM,AE(atuo encoders)[11],卷积神经网络(convo-

lutional neural networks, CNN)[12] 等进行改进而得到的. 文中首先介绍这几种基础的模型,然后介绍这几种基础模型上的深度结构模型或者其改进模型.

2.1　受限玻尔兹曼机

RBM 有着一个丰富的原理架构,是由 1985 年 D. H. Ackley 等[13]提出的统计力学的随机神经网络实例玻尔兹曼机(boltzmann machines, BM)发展而来的. BM 具有强大的无监督学习能力,能够学习数据中复杂的规则. 但是,它无法确切计算 BM 所表示的分布. 为了解决这个问题,Smolensky 引入了受限玻尔兹曼机,他将 BM 原来的层间连接进行限定,使得同一层中不同的节点互相独立,只有层与层之间的节点才有连接,这样就可以较为容易地求得它的概率分布函数[14-15]. 本节介绍 RBM 的原理及基于 RBM 的 2 个深度结构:DBN 和深度玻尔兹曼机(deep boltzmann machine,DBM)[16].

2.1.1　受限玻尔兹曼机原理

RBM 是有着 2 层结构的马尔可夫随机场的特殊情况[17](见图 1),它包含了由 m 个可视的单元 $\boldsymbol{V}=(v_1,v_2,\cdots,v_m)$ 构成的可视层,一般是服从伯努利或者高斯分布;n 个隐藏的单元 $\boldsymbol{H}=(h_1,h_2,\cdots,h_n)$ 构成的隐藏层,一般是服从伯努利分布. 图 1 中上层表示 n 个隐藏单元构成的隐藏(输出)层,下层表示 m 个可视单元构成的可视(输入)层.

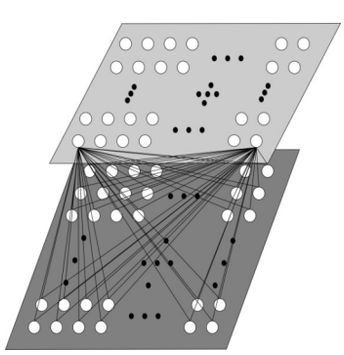


图 1　受限玻尔兹曼机

如图 1 所示,RBM 的可视单元层和隐藏单元层间有权值连接,但层内单元之间无连接.

统计力学中能量函数[8-9,11]可估算一个系统的能量,当系统按其内动力规则进行演变时,其能量函数总是朝减少的方向变化,或停留在某一固定值,最终趋于稳定. 所以可以借由能量函数来对 RBM 进行状态的估计. 一个 RBM 中,在当给定模型的参数 θ(即为权重 w,可视层偏置 b,隐藏层偏置 c)的情况下,它关于可是单元 v 和隐藏单元 h 的联合分布 $p(v,h;\theta)$ 可以由能量函数 $E(v,h;\theta)$ 给出,即为

$$p(v,h;\theta)=\frac{\exp(-E(v,h;\theta))}{Z},\tag{1}$$

式中 $Z=\sum_v\sum_h\exp(-E(v,h;\theta))$ 为一个归一化因子. 这个模型的可视单元 v 边缘概率是 $p(v;\theta)=\dfrac{\sum_h\exp(-E(v,h;\theta))}{Z}$.

对于一个伯努利 - 伯努利 RBM 模型来说,其能量函数为

$$E(v,h;\theta)=-\sum_{i=1}^{n}\sum_{j=1}^{m}w_{ij}h_iv_j-\sum_{j=1}^{m}b_jv_j-\sum_{i=1}^{n}c_ih_i,\tag{2}$$

式中:$i\in\{1,2,\cdots,n\}$;$j\in\{1,2,\cdots,m\}$;w_{ij} 为一个介于单元 v_j 和单元 h_i 之间的边的实数权重;b_j 和 c_i 为第 j 个可视变量和第 i 个隐藏变量各自的实数偏置项. 模型的条件概率为

$$p(h_i=1\mid v;\theta)=\sigma\Big(\sum_{j=1}^{m}w_{ij}v_j+b_j\Big),\tag{3}$$

$$p(v_j=1\mid h;\theta)=\sigma\Big(\sum_{i=1}^{n}w_{ij}h_i+c_i\Big),\tag{4}$$

式中 $\sigma(x)=\dfrac{1}{1+\exp(-x)}$,为 sigmoid 函数.

同样地,对于高斯 - 伯努利 RBM 来说,其能量函数为

$$E(v,h;\theta)=-\sum_{i=1}^{n}\sum_{j=1}^{m}w_{ij}h_iv_j-\frac{1}{2}\sum_{j=1}^{m}(v_j-b_j)^2-\sum_{i=1}^{n}c_ih_i.\tag{5}$$

与它相对应的条件概率为

$$p(h_i=1\mid v;\theta)=\sigma\Big(\sum_{j=1}^{m}w_{ij}v_j+b_i\Big),\tag{6}$$

$$p(v_j\mid h;\theta)=N\Big(\sum_{i=1}^{n}w_{ij}h_i+c_j,1\Big),\tag{7}$$

式中 v_j 为连续值,是服从均值为 $\sum_{i=1}^{n}w_{ij}h_i+c_j$,方差为 1 的高斯分布. 高斯 - 伯努利 RBM 能够将连续值的随机变量转换成二值的随机变量,然后能够通过伯努利 - 伯努利 RBM 进行处理. 在训练 RBM 时,采用 k 步对比散度(contrastive divergence, CD)[18]算法. 具体的 k-CD 算法如下.

输入:RBM$(v_1,v_2,\cdots,v_m,h_1,h_2,\cdots,h_n)$ 的训练集 S.

输出:Δw_{ij}, Δb_j, Δc_i 的近似梯度结果,$i=1,2,\cdots,n$,$j=1,2,\cdots,m$.

1 初始化：$\Delta w_{ij}=\Delta b_j=\Delta c_i=0$，for $i=1,2,\cdots,n$，$j=1,2,\cdots,m$

2 for all the $v\in S$ do

3 $v^{(0)}\leftarrow v$

4 for $t=0,1,\cdots,k-1$ do

5 for $i=1,2,\cdots,n$ do sample $h_i^{(t)}\approx p(h_i|v^{(t)})$

6 for $j=1,2,\cdots,m$ do sample $v_j^{(t+1)}\approx p(v_j|h^{(t)})$

7 for $i=1,2,\cdots,n,j=1,2,\cdots,m$ do

8 $\Delta w_{ij}\leftarrow\Delta w_{ij}+p(h_i=1|v^{(0)})\cdot v_j^{(0)}-p(h_i=1|v^{(k)})\cdot v_j^{(k)}$

9 for $j=1,2,\cdots,m$ do

10 $\Delta b_j\leftarrow\Delta b_j+v_j^{(0)}-v_j^{(k)}$

11 for $i=1,2,\cdots,n$ do

12 $\Delta c_i\leftarrow\Delta c_i+p(h_i=1|v^{(0)})-p(h_i=1|v^{(k)})$

2.1.2 基于受限玻尔兹曼机的深度结构

图1为一个RBM结构，其中下层为输入层，上层为输出层，当在上面再增加一个相同的RBM结构时就形成了部分的DBN结构，即在预训练阶段将一个RBM的输出作为另一个RBM的输入，然后采用BP微调来进行权值更好的训练. 本节将基于RBM介绍DBN和DBM，并简要的分析二者的不同.

2.1.2.1 深度置信网

深度置信网（Deep Belief Networks，DBN）即为若干个RBM模型的叠加，是有着多层隐藏解释因子的神经网络，由G. E. Hinton等[19]在2006年提出.

一个有着ℓ层的DBN模型，可对介于可视变量v_j和ℓ层隐藏层$h^{(k)},k=1,2,\cdots,\ell$间的联合分布进行建模，其中每一层隐藏层由二值单元$h_i^{(k)}$构成，整个DBN的联合概率$p(v,h^{(1)},h^{(2)},\cdots h^{(t)})$为

$$p(v,h^{(1)},h^{(2)},\cdots h^{(\ell)})=P(v|h^{(1)})P(h^{(1)}|h^{(2)})\cdots P(h^{(\ell-2)}|h^{(\ell-1)})P(h^{(\ell-1)},h^{(\ell)}),\tag{8}$$

式中$v=h^{(0)}$，$P(h^{(k)}|h^{(k+1)})$为k层与第$k+1$层间的阶乘条件分布：

$$P(h^{(k)}\mid h^{(k+1)})=\prod_i P(h_i^{(k)}\mid h^{(k+1)}).\tag{9}$$

简要来说就是通过预训练和反向微调来训练整个DBN：在预训练的时候是先单独训练每一个RBM，逐层叠加将下一层的RBM的输出作为上一层RBM的输入；在反向微调的时候可以通过BP训练根据误差函数进行反向调节.

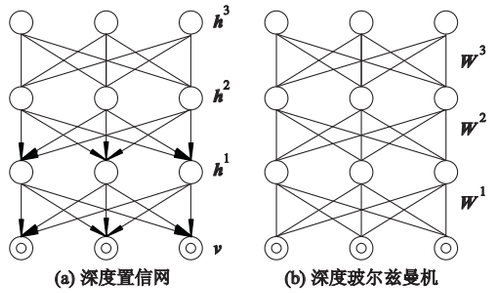
评估一个模型优劣的标准是模型的性能瓶颈，例如在一个分类任务上的测试，DBN可以在预训练后使用标签数据并使用BP算法去微调模型，提升预测性能. 这里需要说的是BP算法在这里只用在DBN中与传统的前向神经网络相关的局部权重微调上，使之加快训练速度和缩短收敛所需时间. 当Hinton在MNIST手写特征识别任务上使用DBN时，试验结果证实了DBN优于传统的前向网络的提升效果. 在文献[11]中Bengio首先通过分析Hinton提出的DBN模型的成功之处，并针对原有的DBN的输入只能是二值数据，通过将第1层的伯努力分布的输入改成高斯分布的输入，从而扩展成可以输入任意值来进行学习和训练. 自从深度置信网被提出后，研究者们针对DBN已经发展了很多的变种，比如卷积深度置信网（Convolutional Deep Belief Networks，CDBN）[20]，稀疏深度置信网（Sparse Deep Belief Networks，SDBN）[21]等等.

2.1.2.2 深度玻尔兹曼机

DBM是包含输入层为D个可视单元的$v\in\{0,1\}^D$集，和F_i个隐藏单元组成的$h^i\in\{0,1\}^{F_i}$集，$h^i\in\{0,1\}^{F_i}$集按序排列，如$h^1\in\{0,1\}^{F_1}$，$h^2\in\{0,1\}^{F_2}$，$\cdots$，$h^L\in\{0,1\}^{F_L}$. 在相邻层间只有隐藏单元之间才有连接，就像第1层中可视单元和与它相近的隐藏单元之间一样. 考虑到一个3层隐藏层的DBM（图2b），这个状态$\{v,h\}$的能量被定义成：

$$E(\boldsymbol{v},\boldsymbol{h};\boldsymbol{\theta})=-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{W}^1\boldsymbol{h}^1-\boldsymbol{h}^{1\mathrm{T}}\boldsymbol{W}^2\boldsymbol{h}^2-\boldsymbol{h}^{2\mathrm{T}}\boldsymbol{W}^3\boldsymbol{h}^3,\tag{10}$$

式中：$\boldsymbol{h}=\{\boldsymbol{h}^1,\boldsymbol{h}^2,\boldsymbol{h}^3\}$为隐藏单元集；$\boldsymbol{\theta}=\{\boldsymbol{W}^1,\boldsymbol{W}^2,\boldsymbol{W}^3\}$为这个模型的相对应的权重参数[16].



(a) 深度置信网 (b) 深度玻尔兹曼机

图2 深度置信网与深度玻尔兹曼机结构图

对于单层RBM来说，如果把RBM隐藏层的层数增加，就可以得到图2b所示的DBM结构；如果在靠近可视层的部分使用贝叶斯置信网络（即有向图模型），而在输出层的部分使用RBM，可以得到图2a所示的DBN结构. DBM有潜力去学习那些越来越复杂的内在的表征，这被认为是在处理对象识别和语音识别问题上一个新的方法，有可能提升深度学习领域在这方面的应用. 此外，DBM能从大量无标签

的自然信息数据(自然世界中存在的信息)中构建高等级表征,通过使用人为定义的有标签数据对模型进行微调,从而进一步达到期望的分类结果. 再之,除了都是自下而上的生成结构且都能够进行自顶向下的反馈外,DBM 允许更鲁棒性地处理模糊的输入数据且更好地进行传播,减少传播造成的误差[22].

2.2　自动编码器

Y. Bengio 等[11]在 2007 年通过理解 DBN 的训练策略的成功之处,即通过无监督预训练来更好地初始化所有层的权值从而减缓深度网络的优化困难的问题,并通过将 DBN 结构中的 RBM 建筑块替换成 AE 来验证这个想法. 本节先介绍 AE 的基本原理,然后再介绍基于 AE 的堆叠自动编码器(stacked auto encoders,SAE)[23].

2.2.1　自动编码器的原理

AE 通过将可视层的输入变换到隐藏的输出层,然后通过隐藏层进行重构使得自动编码器的目标输出与原始输入自身几乎相等,如图 3a 所示. AE 的目标函数为

$$J(\theta,\theta') = \frac{1}{n}\sum_{i=1}^{n} L(\boldsymbol{x}^{(i)},\boldsymbol{z}^{(i)}) + \frac{\lambda}{2}(\|\theta\|^2 + \|\theta'\|^2), \tag{11}$$

式中:第 1 项为最小化模型的重构误差;第 2 项为权重衰减项. 首先,假设一个自动编码器的输入为 d 维的向量 $\boldsymbol{x}\in[0,1]^d$,通过一个函数映射,映射到输出层为 d'维的表征向量 $\boldsymbol{y}\in[0,1]^{d'}$,映射函数为 $\boldsymbol{y}=f_\theta(\boldsymbol{x})=s(\boldsymbol{W}\boldsymbol{x}+\boldsymbol{b})$,模型的构造参数为 $\theta=\{\boldsymbol{W},\boldsymbol{b}\}$,且 $\boldsymbol{W}$ 是一个 $d'\times d$ 的权重矩阵,$\boldsymbol{b}$ 是偏置向量,s 是逐元素计算的逻辑 sigmoid 函数,$s(t)=\frac{1}{1+\exp(-t)}$,$t\in\{1,m\}$,$m$ 为所需传播的后一层的单元个数. 得到的输出表征 $\boldsymbol{y}$ 随后映射到“重构”向量 $\boldsymbol{z}\in[0,1]^d$,$\boldsymbol{z}=g_{\theta'}(\boldsymbol{y})=s(\boldsymbol{W}'\boldsymbol{x}+\boldsymbol{b}')$,模型的重构参数 $\theta'=\{\boldsymbol{W}',\boldsymbol{b}'\}$,$\boldsymbol{W}'$是一个 $d\times d'$的权重矩阵. 图 3b 是一个自动编码过程的简略表示.

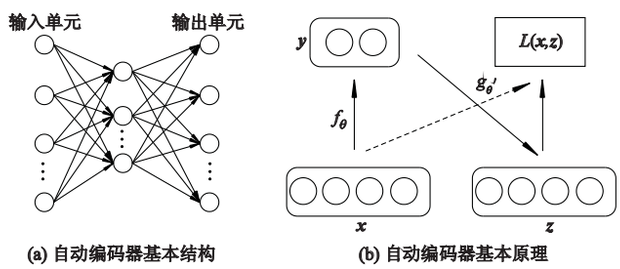


图 3　自动编码器的基本结构及其基本原理

最优化这个模型的参数$\{\theta,\theta'\}$即为最小化模型的平均重构误差:

$$\frac{1}{n}\sum_{i=1}^{n} L(\boldsymbol{x}^{(i)},\boldsymbol{z}^{(i)}) = \frac{1}{n}\sum_{i=1}^{n} L(\boldsymbol{x}^{(i)},g_{q'}(f_q(\boldsymbol{x}^{(i)}))), \tag{12}$$

式中:n 为样本数据的大小;$\boldsymbol{x}$ 为原始输入向量;$\boldsymbol{z}$ 为重构向量. 依据输入输出的不同,损失函数 L 可以是连续值的传统的方差损失函数 $L(\boldsymbol{x},\boldsymbol{z})=\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{z}\|^2$ 或者是二值的交叉熵损失函数 $L(\boldsymbol{x},\boldsymbol{z})=-\sum_{j=1}^{d}[x_j\log(z_j)+(1-x_j)\log(1-z_j)]$[24].

另外,为了防止过拟合,通过将权重衰减项作为正则化项加入到目标函数中,即为公式(11)的第 2 项. 权重衰减参数 λ 表明这个重构误差和权重衰减项的相关重要性.

2.2.2　基于自动编码器的深度结构

AE 结构简单,而且其数学表示通俗易懂,加之能够很好地进行堆叠形成深层结构,本节将介绍基于 AE 形成的 SAE 结构.

文献[4,11]中自动编码器的训练过程是和 RBM 一样使用贪心逐层预训练算法,但因为是通过重构误差来进行训练,相比较而言比训练 RBM 容易,所以常常用来代替 RBM 构建深度结构. 通过将 DBN 中的 RBM 替换成 AE,形成 SAE. SAE 的特点就是它与 RBM 一样也是一个生成模型,但是数据样本在作为 SAE 的输入的同时还能够作为 SAE 的输出目标,从而检测 SAE 中间层学到的特征是否符合要求,通过逐个 AE 的训练,最终完成对整个网络进行训练.

堆叠自动编码器(见图 4)是由多层自动编码器构成的深层神经网络,它被广泛地用于深度学习方法中的维数约简[25]和特征学习[26].

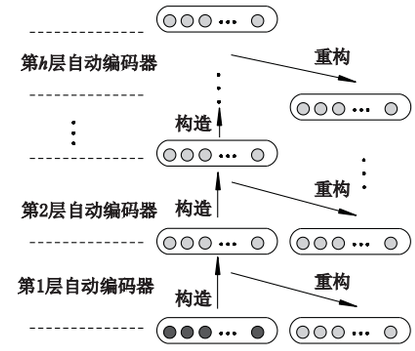


图 4　堆叠自动编码器

正如图 4 中展现的,假设有 h 个自动编码器,并以从底向上的顺序逐层进行训练. 具体的训练过程

如下[4]：① 训练第1个AE，最小化其原始输入（图4中黑色部分）的重构误差；② 将上一个AE的输出用作下一个AE的输入，按照步骤②中的方式进行训练；③ 重复②步的过程，直到完成下面层的训练；④ 将最后一层隐藏层的输出作为一个有监督层的输入，初始化其参数（保持剩余层的参数固定，最顶层的参数可以是随机或者有监督训练得到）；⑤ 按照有监督的标准，可以对所有层进行微调，或者仅对最高层进行微调.

最顶层AE的隐藏层就是这个SAE的输出，这个结果能够馈送到其他应用中去，例如在输出端使用一个SVM分类器. 这个无监督预训练能够自动地利用大规模的无标签数据在神经网络中获得比传统随机初始化更好的权重初始化.

若干个自动编码器的堆叠就成为了深层结构，如果在每个自动编码器的损失函数上加上一个稀疏惩罚值，那么就成为了稀疏堆叠自动编码器（stacked sparse auto encoders，SSAE）[27]：

$$J_{\text{sparse}}(\theta,\theta') = J(\theta,\theta') + \beta\frac{1}{n}\sum_{J}^{d'}\sum_{i=1}^{n} y_j^{(i)}, \tag{13}$$

式中：β为稀疏正则化常量；$\frac{1}{n}\sum_{J}^{d'}\sum_{i=1}^{n} y_j^{(i)}$为稀疏惩罚项. 在堆叠自动编码器的基础上，输入的时候将原始数据加上噪声项，然后在输出层能够获得原始无噪声的输出，那么就是堆叠消噪自动编码器（stacked denoising auto encoders，SDAE）[28]；如果在堆叠自动编码器的基础上加上卷积结构，那么就是堆叠卷积自动编码器（stacked convolutional auto encoders，SCAE）[29].

2.3 卷积神经网络

在1989年Yan Lecun等基于前人工作，提出了一个可以将BP成功用于训练深度网络的结构：CNN，它组合局部感受野、权重共享、和空间或时间上的子采样这3种结构去确保平移和变形上的不变性，一个典型的CNN网络如图5所示.

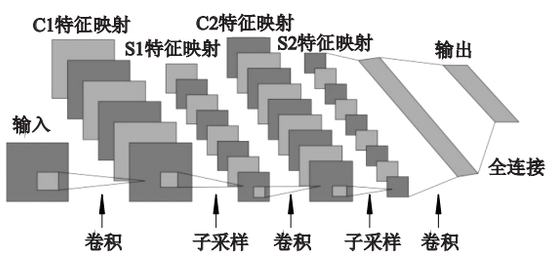


图5 卷积神经网络基本结构

局部感受野：图5中第1个隐藏层有着6个特征图，每个对应于输入层中的小方框就是一个局部感受野，也可以称之为滑动窗口.

卷积：卷积层l中第j个特征映射的激活值a_j^l为

$$a_j^l = f\Big(b_j^l + \sum_{i\in M_j^l} a_i^{l-1} * k_{ij}^l\Big). \tag{14}$$

权值共享：这里f是一个非线性函数，通常是tanh函数或是sigmoid函数，b_j^l是第l层的第j个单元的偏置值，$\boldsymbol{M}_j^l$是$l-1$层中特征映射i的索引向量，而在第l层中特征映射j是需要累加的，$*$是一个2维卷积操作且k_{ij}^l是作用在第$l-1$层中的特征映射i上的卷积核心，能够生成第l层中特征映射j的累加的输入部分. 一个卷积层通常由几个特征图构成，而这里的k_{ij}^l即为权重，在同一个特征图中是相同的，这样就减少了自由参数的数量.

子采样：如果平移这个卷积层的输入将会平移其输出，但是却不会改变它，而且一旦一个特征被检测到，其准确的位置就会不那么重要了，只要相对于其他特征的近似位置被保存即可. 因此，每个卷积层后面会有一个额外的层去执行局部的均值化，即子采样[30-31]去减少输出时关于平移和变形的灵敏度. 对于一个子采样层l中的特征映射j，有

$$a_j^l = down(a_j^{l-1}, N^l), \tag{15}$$

式中：$down$为基于因子N^l进行下采样的函数；N^l为第l层子采样层所需要的窗口边界大小，然后对大小为$N^l\times N^l$的窗口非重叠区域进行均值计算. 假设神经元的输出层为C维，那么就能对C类进行的鉴别，输出层是前层的连接特征映射的输出表征：

$$output = f(\boldsymbol{b}^o + \boldsymbol{W}^o\boldsymbol{f}_v), \tag{16}$$

式中：$\boldsymbol{b}^o$为偏置向量；$\boldsymbol{W}^o$为权重矩阵；$\boldsymbol{f}_v$为特征向量，模型的参数是$\{k_{ij}^l, \boldsymbol{b}_j^l, \boldsymbol{b}^o, \boldsymbol{W}^o\}$. 卷积层和子采样层通常是逐层交替，而特征图的数量是随着空间解析度的减少而增加.

在CNN的应用上一个很成功的实例是Y. LeCun等[32]于1995年提出的LeNet-5系统，在MNIST上得到了0.9%的错误率，并在20世纪90年代就已用于银行的手写支票识别.

近年来，关于CNN的模型逐渐成为研究的热点. 2012年A. Krizhevsky等[33]将CNN构造成深度卷积神经网络（deep convolutional neural network，DCNN），在ILSVRC-2012数据集上获得了top-5测试错误率为15.3%的好结果. 2014年Zheng Yi等提出的多通道深度卷积神经网络（multi-channels deep

convolutional neural networks, MC-DCNN)[34]在 BID-MC 数据集上获得最好的准确度(94.67%),优于之前这个数据集上的最好结果.

3　深度学习的应用

深度学习从 2006 年开始在语音识别、计算机视觉、自然语言处理和信息检索上面都取得了较好效果,在不同的数据集以及工业应用上都表现出远超以往浅层学习所能达到的最好的效果.

3.1　语音识别

在过去几十年中,语音识别领域的研究者们都把精力用在基于 HMM-GMM 的系统[35],而忽略了原始语音数据内部原有的结构特征. 深度神经网络 DNN 在 2010 年开始被引入处理语音识别问题,因为 DNN 对数据之间的相关性有较大的容忍度,使得当 GMM 被 DNN 替换时,效果明显有了飞跃.

2012 年,微软公司一个基于深度学习的语音视频检索系统(Microsoft audio video indexing service, MAVIS)成功问世,将单词错误率降低了 30%(从 27.4% 到 18.5%)[36]. 2014 年 IBM 的沃森研究中心的 T. N. Sainath[37]的工作结果显示 DNN 比以往过去的 GMM-HMM 模型有 8% ~15% 的提升,而 CNN 相比于一般 DNN 来说能对数据间强烈的相关性有更强的适应力,同时足够深的网络还有对数据的平移不变性的特性.

3.2　计算机视觉

深度学习在计算机视觉上的成功应用,主要体现在对象识别[38]和人脸识别领域[39]上. 过去很长一段时间,机器视觉中的对象识别一直依赖于人工设计的特征,例如尺度不变特征转换(scale invariant feature transform, SIFT)[40]和方向梯度直方图(histogram of oriented gradients, HOG)[41],然而像 SIFT 和 HOG 这样的特征只能抓取低等级的边界信息.

针对以往小规模样本所无法表现的真实环境中更复杂的信息,2010 年人们引入了更大的数据集,例如 ImageNet 数据集中有着 15 百万的标记高分辨率图像和超过 2 万 2 千个类别. A. Krizhevsky 等[33]在 2012 年通过训练一个大的深度神经网络来对 ImageNet LSVRC-2010 中包含着1 000个不同类别的 1.2百万个高分辨率图像进行分类. 在测试数据中,他们在 top-1 和 top-5 上的错误率是 37.5% 和 17.0%,刷新了这个数据集的最好记录.

2014 年 Sun Yi 等[42]提出了深度隐藏身份特征(deep hidden identity feature, DeepID)的方法去学习高等级特征表征来进行人脸识别. 通过将人脸部分区域作为每个卷积网络的输入,在底层中提取局部低等级特征,并在深度卷积网络的最后一层隐藏层的神经元激活值中形成 DeepID 特征,试验结果显示 Yi 等在 LFW 上获得了 97.45% 的准确度.

3.3　自然语言处理

自然语言处理(natural language processing, NLP)[43]意在将人类语言转换到能够容易地被计算机操作的表征的过程. 大多数的研究者将这些问题分离式考虑,例如词性标注、分块、命名实体识别、语义角色标注、语言模型和语义相关词等,而没有注重到整体性,使得自然语言处理领域中的进展不是很乐观. 具体来说现有的系统有 3 个缺陷[44]:① 它们都是浅层结构,而且分类器通常是线性的;② 对于一个效果好的线性分类器来说,它们必须事先用许多人工特征来预处理;③ 从几个分离的任务中进行串联特征以至于误差会在传播过程中增大.

2008 年 R. Collobert 等[44]通过将一个普通的深度神经网络结构用于 NLP,在“学习一个语言模式”和“对语义角色标签”任务上通过将重点关注到语义角色标签的问题上进行了没有人工设计特征参与的训练,其错误率为 14.3% 的结果刷新了最好记录.

3.4　信息检索

信息检索(information retrieval, IR)就是用户输入一个查询到一个包含着许多文档的计算机系统,并从中取得与用户要求所需最接近的文档[2]. 深度学习在 IR 上的应用主要是通过提取有用的语义特征来进行子序列文档排序,由 R. Salakhutdinov 等[25]在 2009 年提出,他们针对当时最广泛被使用在文档检索上的系统 TF - IDF[25]上的分析,认为 TF - IDF 系统有着以下的缺陷:在词计数空间中直接计算文档的相似性,这使得在大词汇量下会很慢;没有使用词汇间的语义相似性. 因为在 DNN 模型的最后一层中的隐藏变量不但在使用基于前向传播的训练后容易推导,而且在基于词计数特征上给出了对每个文档更好的表征,他们使用从深度自动编码器得到的紧凑的编码,使得文档能够映射到一个内存地址中,在这个内存地址中语义上相似的文档能够被归类到相近的地址方便快速的文档检索. 从词计数向量到紧凑编码的映射使得检索变得高效,只需要更便捷的计算,更少的时间.

2014 年 Shen Yelong 等[45]提出了卷积版的深度结构语义模型(convolutional deep-structured semantic

modeling, C-DSSM),C-DSSM 能将上下文中语义相似的单词通过一个卷积结构投影到上下文特征空间向量上,从之前 43.1% 的准确率提高到了 44.7%.

不同于以往浅层结构只能解决许多简单的或者许多约束条件下的问题,深度结构能够处理许多复杂的真实世界中的问题,例如人类语音、自然声音和语言、自然图像、可视场景等问题,它们可以直接从数据中提取数据所包含的特征而不受具体模型的约束,从而更具有泛化能力.

4 深度学习的研究展望

随着研究的深入,深度学习已经成为机器学习中一个不可或缺的领域,然而,关于深度学习的研究现在仍然才处于萌芽状态,很多问题仍然没有找到满意的答案[46]. 如对在线学习的能力的提升,以及在大数据方面的适应能力以及在深度层次结构上的改进.

在线学习方面:当前几乎所有的深度学习所应用到的深度结构训练的算法都是先在搭建好的结构上进行逐层训练,并在逐层训练之后加上一个全局微调得到更好的拟合数据的参数集. 这种训练算法在纯粹的在线环境下不是很适用,因为在线数据的数据集是在不断扩充的,一旦在在线环境下引入全局微调的方法,那么结果极有可能陷入局部最小. 如何将深度学习用于在线环境是值得思考的一个问题.

在对大数据的适应能力上:大数据中包含着很多有价值的信息,但是如何从大数据中找到能够表达这个数据的表征是研究者关心的问题. 2012 年的 Google 大脑团队在一个超大多节点的计算机网络上并行地训练深度网络结构,结果显示数据仍然呈现欠拟合的状态[47]. 对此,如何衡量训练复杂度与任务复杂度的关系,使得深度学习可以充分地用在大数据上,还有待于研究和实践.

在深度结构的改进上:深度结构的层次模型虽然比浅层模型在结构上具有突破,模拟了生物的视觉系统分层结构,但是未能完全匹配皮层的信息处理结构. 比如研究者们发现现有的主流的深度结构并未考虑到时间序列对学习的影响,而作为真正的生物皮层在处理信息上来说,对信息数据的学习不是独立静态的,而是随着时间有着上下文的联系的.

人类的信息处理机制表明深度结构可以从丰富的感知信息中提取复杂的结构和建立数据中内在的表征. 因为深度学习尚在初步阶段,很多问题还没有解决,所以还无法真正达到人工智能的标准,但是深度学习现有的成功和发展表明,深度学习是向人工智能迈进的一大步.

5 总 结

1)文中首先通过对现有的深度学习所使用的深度结构的分类,介绍了 RBM,AE,CNN 等深度学习所使用的几大基础模型具有的原理及特点,并相对应地分析了如何在这几个模型的基础上来得到 DBN、DBM 以及 SAE 等真正的深度层次结构模型.

2)通过在语音识别、计算机视觉、自然语言处理和信息检索几大领域上深度学习应用的介绍,说明了深度学习在机器学习领域有相比较于其他浅层结构学习具有更好的优越性和更少的错误率.

3)通过对深度学习在在线学习方面和大数据上的适应能力以及对深度结构的改进等方面对当前深度学习所面临的问题作了总结和思考. 当前深度学习还尚未成熟,仍有大量的工作需要研究,但是其展现的强大的学习能力和泛化能力表明,今后它将是机器学习领域中研究的重点和热点.

参考文献(References)

[1] 孙志军,薛 磊,许阳明,等. 深度学习研究综述[J]. 计算机应用研究,2012,29(8):2806-2810.
Sun Zhijun, Xue Lei, Xu Yangming, et al. Overview of deep learning[J]. *Application Research of Computers*, 2012,29(8): 2806-2810. (in Chinese)

[2] Deng Li, Yu Dong. Deep learning for signal and information processing[R]. Microsoft Research, 2013.

[3] 胡晓林,朱 军. 深度学习——机器学习领域的新热点[J]. 中国计算机学会通讯,2013,9(7): 64-69.
Hu Xiaolin, Zhu Jun. Deep learning—new hot spot in the field of maching earning [J]. *Communications of the CCF*,2013,9(7):64-69. (in Chinese)

[4] Bengio Yoshua. Learning deep architectures for AI [J]. *Foundations and Trends in Machine Learning*, 2009, 2(1): 1-27.

[5] Duarte-Carvajalino J M, Yu G S, Carin L, et al. Task-driven adaptive statistical compressive sensing of gaussian mixture models[J]. *IEEE Transactions on Signal Processing*, 2013, 61(3): 585-600.

[6] Abdel-Rahman E M, Mutanga O, Adam E, et al. Detecting sirex noctilio grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random

forest and support vector machines classifiers[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014, 88: 48 – 59.

[7] 刘国海, 肖夏宏, 江　辉, 等. 基于 BP-Adaboost 的近红外光谱检测固态发酵过程 pH 值[J]. 江苏大学学报: 自然科学版, 2013, 34(5): 574 – 578.
Liu Guohai, Xiao Xiahong, Jiang Hui, et al. Detection of PH variable in solid-state fermentation process by FT-NIR spectroscopy and BP-Adaboost [J]. *Journal of Jiangsu University: Natural Science Edition*, 2013, 34(5):574 – 578. (in Chinese)

[8] Sarikaya R, Hinton G E, Deoras A. Application of deep belief networks for natural language understanding[J]. *IEEE Transactions on Audio, Speech and Language Processing*, 2014, 22(4): 778 – 784.

[9] Fischer A, Igel C. Training restricted Boltzmann machines: an introduction[J]. *Pattern Recognition*, 2014, 47(1):25 – 39.

[10] Wan L, Zeiler M, Zhang S X, et al. Regularization of neural networks using dropconnect[C]// *Proceedings of the 30th International Conference on Machine Learning*. Atlanta: IMLS, 2013: 2095 – 2103.

[11] Bengio Y, Lamblin P, Popovici D, et al. Greedy layer-wise training of deep networks[C]// *Proceedings of 20th Annual Conference on Neural Information Processing Systems*. Vancouver: Neural information processing system foundation, 2007: 153 – 160.

[12] Palm R B. Prediction as a candidate for learning deep hierarchical models of data[D]. Technical University of Denmark, Denmark, 2012.

[13] Ackley D H, Hinton G E, Sejnowski T J. A learning algorithm for Boltzmann machines [J]. *Cognitive Science*, 1985, 9: 147 – 169.

[14] Yu Dong, Deng Li. Deep learning and its applications to signal and information processing [J]. *IEEE Signal Processing Magazine*, 2011, 28(1): 145 – 149, 154.

[15] Cho K Y. Improved learning algorithms for restricted Boltzmann machines [D]. Espoo: School of Science, Aalto University, 2011.

[16] Cho K H, Raiko T, Ilin A, et al. A two-stage pretraining algorithm for deep boltzmann machines[C]// *Proceedings of 23rd International Conference on Artificial Neural Networks*. Sofia: Springer Verlag, 2013: 106 – 113.

[17] Shu H, Nan B, Koeppe R, et al. Multiple testing for neuroimaging via hidden markov random field [DB/OL]. [2014 – 05 – 08]. http://arxiv.org/pdf/1404.1371.pdf.

[18] Hjelm R D, Calhoun V D, Salakhutdinov R, et al. Restricted Boltzmann machines for neuroimaging: an application in identifying intrinsic networks [J]. *NeuroImage*, 2014, 96: 245 – 260.

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

[20] Lee H, Grosse R, Ranganath R, et al. Unsupervised learning of hierarchical representations with convolutional deep belief networks [J]. *Communications of the ACM*, 2011, 54(10): 95 – 103.

[21] Halkias X C, Paris S, Glotin H. Sparse penalty in deep belief networks: using the mixed norm constraint[DB/OL]. [2014 – 05 – 08]. http://arxiv.org/pdf/1301.3533.pdf.

[22] Poon-Feng K, Huang D Y, Dong M H, et al. Acoustic emotion recognition based on fusion of multiple feature-dependent deep Boltzmann machines[C]// *Proceedings of the 9th International Symposium on Chinese Spoken Language Processing*. Singapore: IEEE, 2014: 584 – 588.

[23] Wang W, Ooi B C, Yang X Y, et al. Effective multi-modal retrieval based on stacked auto-encoders [J]. *Proceedings of the VLDB Endowment*, 2014, 7(8): 649 – 660.

[24] Arnold L, Rebecchi S, Chevallier S, et al. An introduction to deep learning [C]// *Proceedings of the 18th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. [S.l.]: i6doc.com publication, 2010:477 – 478.

[25] Salakhutdinov R, Hinton G. Semantic hashing [J]. *International Journal of Approximate Reasoning*, 2009, 50(7):969 – 978.

[26] Goroshin R, LeCun Y. Saturating auto-encoders [DB/OL]. [2014 – 05 – 08]. http://arxiv.org/pdf/1301.3577.pdf.

[27] Jiang Xiaojuan, Zhang Yinghua, Zhang Wensheng, et al. A novel sparse auto-encoder for deep unsupervised learning [C]// *Proceeding of 2013 Sixth International Conference on Advanced Computational Intelligence*. Hangzhou: IEEE Computer Society, 2013: 256 – 261.

[28] 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.

[29] Masci J, Meier U, Cireşan D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction

[C]// *Proceedings of* 21*st International Conference on Artificial Neural Networks*. Espoo: Springer Verlag, 2011: 52 - 59.

[30] Pinheiro P O, Collobert R. Recurrent convolutional neural networks for scene labeling[C]// *Proceedings of the* 31*st International Conference on Machine Learning*. Beijing: IMLS, 2014: 82 - 90.

[31] Zeiler M D, Fergus R. Stochastic pooling for regularization of deep convolutional neural networks[DB/OL]. [2014 - 05 - 08]. http://arxiv.org/pdf/1301.3557.pdf.

[32] LeCun Y, Jackel L D, Bottou L, et al. *Learning Algorithms for Classification: A Comparison on Handwritten Digit Recognition*[M]. Korea: World Scientific, 1995, 261 - 276.

[33] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]// *Proceeding of* 26*th Annual Conference on Neural Information Processing Systems*. Lake Tahoe: Neural information processing system foundation, 2012: 1097 - 1105.

[34] Zheng Yi, Liu Qi, Chen Enhong, et al. Time series classification using multi-channels deep convolutional neural networks[C]// *Proceedings of* 15*th International Conference on Web-Age Information Management*. Macau: Springer Verlag, 2014: 298 - 310.

[35] Mohamed A R, Dahl G E, Hinton G. Acoustic modeling using deep belief networks[J]. *IEEE Transactions on Audio, Speech and Language Processing*, 2012, 20(1): 14 - 22.

[36] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(8): 1798 - 1828.

[37] Sainath T N. Improvements to deep neural networks for large vocabulary continuous speech recognition tasks[R]. IBM T. J. Watson Research Center, 2014.

[38] Sohn K, Jung D Y, Lee H, et al. Efficient learning of sparse, distributed, convolutional feature representations for object recognition[C]// *Proceeding of* 2011 *IEEE International Conference on Computer Vision*. Barcelona: IEEE, 2011: 2643 - 2650.

[39] Cui Zhen, Chang Hong, Shan Shiguang, et al. Joint sparse representation for video-based face recognition[J]. *Neurocomputing*, 2014, 135: 306 - 312.

[40] 关海鸥, 杜松怀, 许少华, 等. 基于改进投影寻踪技术和模糊神经网络的未受精种蛋检测模型[J]. 江苏大学学报:自然科学版, 2013, 34(2): 171 - 177.
Guan Haiou, Du Songhuai, Xu Shaohua, et al. Detection model of un-fretilized egg based on improved projection pursuit and fuzzy neural network[J]. *Journal of Jiangsu University: Natural Science Edition*, 2013, 34(2): 171 - 177. (in Chinese)

[41] 王国林, 周树仁, 李军强. 基于模糊聚类和形态学的轮胎断面特征提取[J]. 江苏大学学报:自然科学版, 2012, 33(5): 513 - 517.
Wang Guolin, Zhou Shuren, Li Junqiang. Feature extraction of tire section based on fuzzy clustering and morphology[J]. *Journal of Jiangsu University: Natural Science Edition*, 2012, 33(5): 513 - 517. (in Chinese)

[42] Sun Yi, Wang Xiaogang, Tang Xiaoou. Deep learning face representation from predicting 10,000 classes[C]// *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE Computer Society, 2014: 1891 - 1898.

[43] Cambria E, White B. Jumping NLP curves: a review of natural language processing research[J]. *IEEE Computational Intelligence Magazine*, 2014, 9(2): 48 - 57.

[44] Collobert R, Weston J. A unified architecture for natural language processing: deep neural networks with multitask learning[C]// *Proceedings of* 25*th International Conference on Machine Learning*. Helsinki, Finland: Association for Computing Machinery, 2008: 160 - 167.

[45] Shen Yelong, He Xiaodong, Gao Jianfeng, et al. Learning semantic representations using convolutional neural networks for Web search[C]// *Proceedings of the companion publication of the* 23*rd international conference on World wide web companion*. Seoul: IW3C2, 2014: 373 - 374.

[46] Arel I, Rose D C, Karnowski T P. Deep machine learning a new frontier in artificial intelligence research[J]. *IEEE Computational Intelligence Magazine*, 2010, 5(4): 13 - 18.

[47] Bengio Y. Deep learning of representations: looking forward[C]// *Proceedings of* 1*st International Conference on Statistical Language and Speech Processing*. Tarragona, Spain: Springer Verlag, 2013: 1 - 37.

(责任编辑 梁家峰)