

深度自动编码器的研究与展望

曲建岭 杜辰飞 邸亚洲 高峰 郭超然

(海军航空工程学院青岛校区, 山东 青岛 266041)

摘要: 深度学习是机器学习的一个分支, 开创了神经网络发展的新纪元。作为深度学习结构的主要组成部分之一, 深度自动编码器主要用于完成转换学习任务, 同时也在无监督学习及非线性特征提取过程中也扮演着至关重要的角色。首先介绍深度自动编码器的发展由来、基本概念及原理, 然后介绍它的构建方法以及预训练和精雕的一般步骤, 并对不同类型深度自动编码器进行总结, 最后在深入分析深度自动编码器目前存在的问题的基础上, 对其未来发展趋势进行展望。

关键词: 深度学习; 深度自动编码器; 预训练; 精雕; 神经网络

中图分类号: TP391.9

文献标识码: A

doi: 10.3969/j.issn.1006-2475.2014.08.028

Research and Prospect of Deep Auto-encoders

QU Jian-ling, DU Chen-fei, DI Ya-zhou, GAO Feng, GUO Chao-ran

(Qingdao Branch, Naval Aeronautical Engineering Institute, Qingdao 266041, China)

Abstract: Deep learning, which is a branch of machine learning, inaugurates new era in the development of neural network. As a key component of deep structure, the deep auto-encoder is used to fulfill a task of transforming learning and plays important role in both unsupervised learning and non-linear characters extraction. We firstly introduced the origin of deep auto-encoder as well as its basic concept and principle, secondly, the construction procedure, pre-training and fine-tune procedure of depth auto-encoders were generally introduced, meanwhile, a comprehensive summarization of different kinds of DAE was made. At last, the direction of future work was proposed based on an in-depth study of current DAE researches.

Key words: deep learning; deep auto-encoder (DAE); pre-train; fine-tune; neural network

0 引言

BP 神经网络是有效的机器学习算法之一, 它具有较强的非线性拟合能力和容错能力, 能自动提取数据中的“合理规则”, 并将其植入网络权值, 自适应效果好。但是, 它受到人工特征提取的限制, 步骤繁琐, 运算量大, 代价函数复杂不易优化, 因而收敛速度慢, 并且该算法易陷入局部最小, 泛化能力差, 不能精确拟合部分高维复杂函数^[1]。为解决此类问题, Hinton 等人改进原有浅层结构, 提出深度学习神经网络的概念及其训练策略, 继而产生了深度自动编码器 (DAE)^[2]。DAE 的产生和应用免去了人工提取数据特征的巨大工作量, 提高了特征提取的效率, 降低了原始输入的维数, 得到数据的逆向映射特征, 展现了

从少数类标样本和大量无类标数据中学习输入数据本质特征的强大能力, 并将学习到的特征分层表示, 为构建深度结构奠定了基础, 成为神经网络研究的一个里程碑。

本文在阐述深度自动编码器的发展由来、基本概念和原理及其构建方法的基础上, 根据深度自动编码器的建模机理对其进行分类和总结, 指出深度自动编码器目前存在的问题, 明确它的发展方向。

1 DAE 的发展由来

1986 年 Rumelhart 提出自动编码器的概念, 并将其用于高维复杂数据处理, 促进了神经网络的发展^[3]。2006 年, Hinton 对原型自动编码器结构进行改进, 进而产生了 DAE, 先用无监督逐层贪心训练算

收稿日期: 2014-05-20

作者简介: 曲建岭 (1968-) 男, 山东莱阳人, 海军航空工程学院青岛校区教授, 博士生导师, 博士, 研究方向: 人工智能、信号处理、仪器仪表。

法完成对隐含层的预训练,然后用 BP 算法对整个神经网络进行系统性参数优化调整,显著降低了神经网络的性能指数,有效改善了 BP 算法易陷入局部最小的不良状况^[2]。2007 年,Benjio 提出稀疏自动编码器的概念,进一步深化了 DAE 的研究^[4]。2008 年,Vincent 提出降噪自动编码器,在输入数据中添加腐坏向量,防止出现过拟合现象,并取得了良好的效果^[5]。2009 年,Benjio 总结了已有的深度结构,阐述了利用堆叠自动编码器构建深度学习神经网络的一般方法^[6]。2010 年,Salah 对升维和降维的过程加以限制,提出收缩自动编码器^[7]。2011 年,Jonathan 提出卷积自动编码器,用于构建卷积神经网络^[8]。2012 年,Taylor 对 DAE 与无监督特征学习之间的联系进行了深入的探讨,详细介绍了如何利用自动编码器构建不同类型深度结构^[9]。Hinton, Bengio 和 Vincent 等人对比了原型自动编码器、稀疏自动编码器、降噪自动编码器、收缩自动编码器、卷积自动编码器和 RBM 等结构的性能,为以后的实践和科研提供了参考^[10-14]。2013 年 Telmo 研究了用不同代价函数训练的 DAE 的性能,为代价函数优化策略的发展指明了方向^[15]。

2 DAE 的基本概念及原理

DAE 是一种利用经过无监督逐层贪心预训练和系统性参数优化的多层非线性网络从无类标数据中提取高维复杂输入数据的分层特征,并得到原始数据的分布式特征表示的深度神经网络结构^[16]。DAE 由编码器、解码器和隐含层组成。编码器是输入 x 到隐含表示 h 的映射,表示为:

$$h = f(x) = S_f(Wx + b_h) \quad (1)$$

其中 S_f 是非线性激活函数,一般为逻辑函数,其表达式为:

$$\text{sigmoid}(z) = 1 / (1 + z^{-1}) \quad (2)$$

解码器函数 $g(h)$ 将隐含层数据映射回重构 y ,表示为:

$$y = g(h) = S_g(W'h + b_y) \quad (3)$$

其中 S_g 是解码器的激活函数,一般为线性函数或 sigmoid 函数。训练 DAE 的过程是在训练样本集 D 上寻找参数 $\theta = \{W, b_y, b_h\}$ 的最小化重构误差,重构误差的表达式为:

$$J_{AE} = \sum_{x \in D} L(x, g(f(x))) \quad (4)$$

其中 L 为重构误差函数,一般可用平方误差函数或交叉熵损失函数,二者分别表示为:

$$L(x, y) = \|x - y\|^2 \quad (5)$$

$$L(x, y) = - \sum_{i=1}^{d_x} x_i \log y_i + (1 - x_i) \log (1 - y_i) \quad (6)$$

其中,平方误差用于线性 S_g ,交叉熵损失函数用于 sigmoid^[17]。

3 DAE 的构建

DAE 的构建主要有 2 个步骤,第一步是改进原型自动编码器的结构(原型自动编码器如图 1 所示),即增加隐含层和神经元的数量,调整隐含层节点的分布并改变权值的分享方式等,构建 DAE 的基本框架。第二步是根据不同任务的需要选取合适的代价函数及其优化策略、隐含层品质因数和系统性参数优化时的性能指数等,确定 DAE 的训练方案。

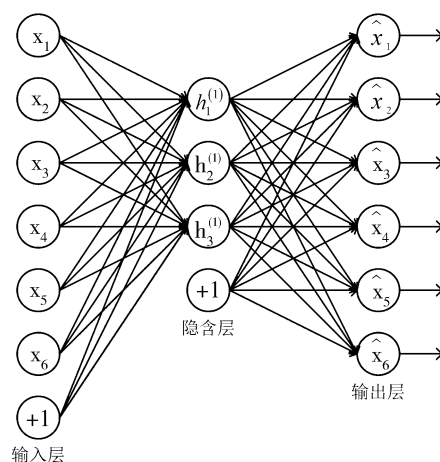


图 1 原型自动编码器

3.1 DAE 的预训练

预训练的目的是将所有权值链接和偏置限定在一定的参数空间内,防止随机初始化的发生进而降低每个隐含层的品质因数,便于对整个神经网络进行系统性参数优化,该算法的核心是用无监督的方法将 DAE 的输入层和隐含层全部初始化,然后再用逐层贪心训练算法将每个隐含层训练为自动关联器,实现输入数据的重构^[18],其基本步骤可总结如下:

- 1) 以无监督的方式训练神经网络的第一层,将其输出作为原始输入的最小化重构误差;
- 2) 每个隐含单元的输出生作为下一层神经网络的输入,用无类标数据样本对下一层进行训练,将误差控制在一定范围内;
- 3) 重复步骤 2),直到完成规定数量隐含层的训练为止;
- 4) 将最后一个隐含层的输出作为有监督层的输入,并且初始化有监督层的参数^[18],预训练过程如图 2 所示。

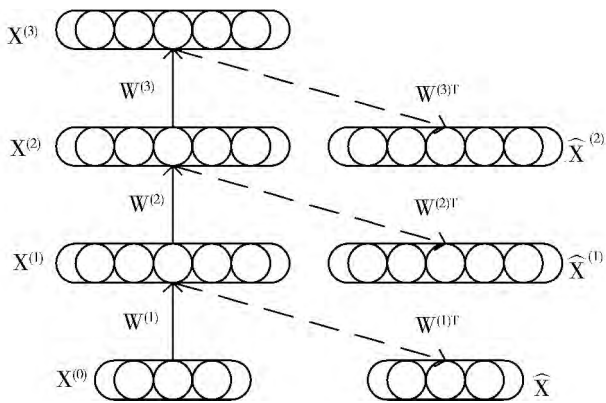


图2 预训练过程

通过持续的预训练,DAE 可以代替 RBM 构建深度结构,由它构建的深度结构性能指数更低,稳定性更强更可靠。更重要的是,能根据不同任务的需要来调整网络参数,可控性更强^[20]。

3.2 DAE 的精雕

精雕又叫微调,是构建 DAE 的必要步骤,通常采用 BP 算法来完成这一任务^[21](牛顿法、共轭梯度法、MOBP 和 SDBP 等,BP 算法的变形也可用于精雕)。精雕的核心思想是将自动编码器的输入层、输出层和所有隐含层视为一个整体,用有监督学习算法进一步调整经过预训练的神经网络,经过多次迭代后,所有权值及偏置均能被优化。由于最后一个隐含层只能输出原始数据的重构,因而不具有分类识别功能。为了让 DAE 具有分类识别的功能,需要在完成精雕的神经网络的输出层之后加入 Soft Max 分类器,将整个神经网络训练成能完成分层特征提取和数据分类任务的多重感知器。其基本步骤可总结如下:

- 1) 对权值、偏置和阈值赋值,对网络进行初始化;
- 2) 随机选取类标数据样本用 BP 算法对神经网络进行训练,计算各层的输出;
- 3) 求出各层的重构误差,并根据误差修正权值和偏置;
- 4) 根据性能指数判定误差是否满足要求,如果未能满足要求则重复步骤 2) 和 3),直到整个网络输出满足期望要求。

需要注意的是,由于 DAE 的训练难度较大,代价函数的优化需要很多技巧,并且网络输出易受到小扰动的影响,为了改善这一状况笔者将权值衰减策略用于 BP 算法,此时的重构误差函数为:

$$E = 0.5 \sum_i (T_i - O_i)^2 + 0.5 \mu \sum_{ij} [W_{ij} / (1 + W_{ij}^2)] \quad (7)$$

其中, T_i 是目标输出, O_i 是网络输出, μ 是权值衰减激活因子,当 $\mu \neq 0$ 时,权值衰减函数被激活,权值更

新规则为:

$$\Delta W_{ij} = -\eta \partial E / \partial W_{ij} = \eta H_j (T_i - O_i) O_i (1 - O_i) - \eta \mu [W_{ij} / (1 + W_{ij}^2)] \quad (8)$$

其中, H_j 是隐含层节点的输出, η 是学习速率。权值衰减策略强化了较大权值,弱化了较小权值,对不同阶的权值施以不同程度的惩罚,加快了收敛速度,降低了重构误差,显著增加了神经网络的鲁棒性^[22],完成全部训练任务的三隐含层 DAE 如图 3 所示。

完成预训练和精雕的 DAE 具有生物神经系统的基本特征,在一定程度上反映了人脑的若干功能,是对生物系统的成功模拟,实现了无监督学习和监督学习的有机结合,且不同隐含层能学到不同的特征,低阶隐含层为高阶隐含层提供特征模式,便于高阶隐含层进行模式组合。同时,DAE 的拓扑结构完整,具有强大的非线性拟合能力,能发现数据的分布式特征表示,提取数据的潜在分层特征,获取原始输入的“层次型分组”或“部分—整体分解”结构,因而可以更好地表示输入数据,具有大规模并行、分布式处理、自组织和自学习等优点^[23]。

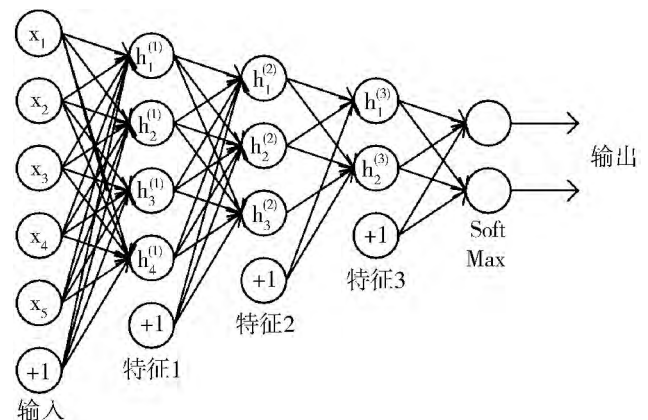


图3 三隐含层深度自动编码器

4 DAE 的分类与分析

DAE 的构建是通过改进原型自动编码器的结构和训练策略完成的,为了适应不同的条件,满足不同任务的需求,研究人员对原型自动编码器进行调整、改进、堆叠、预训练和精雕后产生了 4 类 DAE,分别是基于稀疏理论的 DAE、基于统计理论的 DAE、基于鲁棒理论的 DAE 和基于卷积理论的 DAE。

4.1 基于稀疏理论的 DAE

基于稀疏理论的 DAE 对原型自动编码器的隐含层添加了约束条件并增加了隐含层数量,因而当隐含层神经元的数量很大时,该类自动编码器依然能发现输入数据的结构性特征。设 x 是给定的神经网络输

入 ρ 是稀疏参数 j 为所有隐含层神经元的平均激活值。为了达到约束每个神经元的目的, 取 $\hat{\rho}_j = \rho$, 一般的 ρ 的取值在 0 附近。稀疏自动编码器在原型自动编码器的代价函数中增加了稀疏惩罚项, 其表达式为:

$$\sum_{j=1}^{s_2} \rho \log \frac{\rho}{\rho_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \rho_j} \quad (9)$$

此时代价函数为:

$$J_{\text{sparse}}(W, b) = J(W, b) + \beta \sum_{j=1}^{s_2} \text{KL}(\rho || \hat{\rho}_j) \quad (10)$$

其中 β 表示稀疏惩罚项的权重^[24]。

基于稀疏理论的 DAE 是目前应用最为广泛的 DAE。2013 年, 邓俊将基于稀疏理论的 DAE 用于特征转换和语义情感辨识取得了良好的效果^[25], 马云龙将其用于数据流的异常检测^[26], 张开旭将其用于中文词汇特征提取^[27]。

基于稀疏理论的 DAE 能提取高维数据变量的稀疏解释性因子, 保留原始输入的非零特征, 增加表示算法的鲁棒性, 增强数据的线性可分性, 使分类边界变得更加清晰, 并且能在一定程度上控制变量的规模, 改变给定输入数据的结构, 丰富了原有信息, 提高了信息表述的全面性和准确率。但由于原始数据分布稠密程度不同, 经过信息解锁后的稀疏变量难以控制。另外 KL 散度在惩罚激活值时需要预先给定稀疏目标的期望, 并且该类 DAE 要求每个隐含单元的稀疏程度基本一致, 致使远零激活值的惩罚效果较差, 欠拟合的问题时常出现。

4.2 基于统计理论的 DAE

基于统计理论的 DAE 又叫降噪自动编码器, 其核心思想是: 编码器将含有一定统计特性的噪声(腐坏向量)加入输入数据, 便于隐含层对输入数据进行编码, 保存输入数据中的信息, 解码器根据噪声统计特性从未受到干扰的数据中估计出受干扰数据的原始形式, 输出原始数据的分布参数, 从而消除背景噪声^[28], 该类 DAE 的代价函数为:

$$J_{\text{DAE}} = \sum_i E_q(\tilde{x} | x^i) \quad (11)$$

其中 $E_q(\tilde{x} | x^i)$ 是输入噪声的期望, 一般用高斯噪声作为腐坏向量, 其表达式为:

$$\tilde{x} = x + \varepsilon, \varepsilon \sim N(0, \sigma^2 I) \quad (12)$$

其中 σ 表示整个自动编码器的正规化程度。为了便于计算, 有时可用二项随机隐藏噪声代替高斯噪声^[29]。降噪自动编码器的训练规则为重构对数似然函数:

$$-\log P(x | c(\tilde{x})) \quad (13)$$

其中 x 是未受噪声干扰的输入数据, \tilde{x} 是腐坏向量, $c(\tilde{x})$ 是从 \tilde{x} 中获取的数据编码。将对数似然函数用于该类 DAE 的训练, 能最大限度地利用无类标数据, 用未受到噪声干扰的数据将原始数据估计出来。

基于统计理论的 DAE 是一种经过正规化的深度自动编码器, 其内部映射均具有鲁棒性, 腐坏向量的加入有效降低了 DAE 对微小随机扰动的敏感性, 缩小了重构误差, 打破了训练数据必须在输入数据概率密度曲线附近的局限性^[30]。因此, 基于统计理论的 DAE 常被用于构建生成性模型, 在区域自适应方面取得了良好的效果, 为深度理论的发展打下了坚实的基础^[31]。但由于背景噪声的统计特性不同, 其边缘分布对原始参数的估计过程有较大影响, 尤其是当原始输入是高维复杂函数时, 重构误差显著增加。另外, 该类 DAE 的计算效率极低, 对硬件及软件的要求较高。目前, CPU 的浮点运算和向量运算能力远不如 GPU, GPU 的硬件能够管理数千个并行线程而不需要开发人员进行任何编程。GPU 拥有高带宽的独立显存, 适合处理并行重复计算任务, 大幅度降低系统成本。Matlab 的循环运算效率较低, 程序封装性差, 而 PYTHON 具有非常简捷而清晰的语法, 适合完成各种高层任务, 其动态反射语言不受操作系统限制, 几乎可以在所有的操作系统中运行, 并能在运行中获取类标之下的方法和属性, 通过字符串描述反射出真实的实例化对象, 结合模式设计, 最大限度地满足变化中的需求。降噪自动编码器原理如图 4 所示。

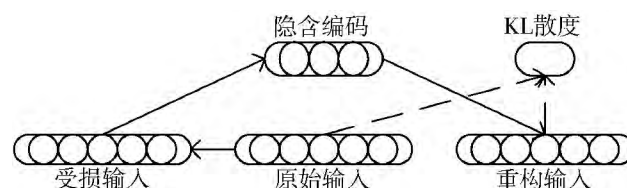


图4 降噪自动编码器原理

4.3 基于鲁棒理论的 DAE

为了进一步提高表示学习算法的鲁棒性, 研究人员在原型自动编码器的代价函数表达式中加入解析性收缩惩罚因子, 以减少特征表示的自由度, 使隐含层神经元达到饱和状态, 进而将输出数据限制在参数空间的一定范围内。该惩罚因子实际上是编码器 Jacobian 矩阵的 Frobenius 范数, 其作用是降低极小化变量对编码器的影响, 辅助编码器学习数据特征。令 $J(x) = \partial f_0(x) / \partial a$, 其中 x 是编码器 Jacobian 矩阵的估计, 收缩自动编码器的代价函数为:

$$J_{\text{CAE}} = \sum_i L(x^i, g_0(f_0(x^i))) + \lambda \|J(x^i)\|_F \quad (14)$$

其中 λ 是反映矩阵正规化程度的活跃参数。收缩自

动编码器的传递函数为 sigmoid 函数,因此解析性收敛惩罚因子可表示为:

$$\|J(x^1)\|^2 = \sum_j (f_0(x)_j(1-f_0(x)_j))^2 \|W_j\|^2 \quad (15)$$

由于解析惩罚因子的限制,基于鲁棒理论的 DAE 对输入数据中的小扰动敏感性较小,且重构特征不受惩罚因子影响,数据表示的准确率较高,便于神经网络计算代价函数。同时,活跃参数 λ 能够权衡代价函数的鲁棒性和重构误差,辅助 BP 算法优化神经网络参数^[32]。为了克服 DAE 的解析惩罚因子只对输入数据中的极小扰动具有鲁棒性的缺陷,研究人员利用统计理论对其代价函数提出进一步的修正,进而惩罚不同阶的偏差,使得代价函数变为:

$$J_{\text{CAE+H}} = \sum_i L(x^{(i)}, g_0(x^{(i)})) + \lambda \|J(x^{(i)})\|_F^2 + \gamma E_\varepsilon[\|J(x) - J(x + \varepsilon)\|_F^2] \quad (16)$$

其中 $\varepsilon: N(0, \sigma^2 I)$, γ 是关联性正规化活跃参数。

经过改进后,DAE 的鲁棒性进一步得到提高,能出色地完成无监督特征转换任务,深化了表示学习算法的研究^[33]。但是,由于基于鲁棒理论的 DAE 原理相对复杂,构建、训练和调整的难度较大,因而针对该类 DAE 的报道较少。

4.4 基于卷积理论的 DAE

上文中讨论的 DAE 均不能有效解决图像(2D)数据中的池化与白化问题,并且大量冗余参数被强迫参与计算使得运算效率较低,而基于卷积理论的 DAE 是一种专门用于处理图像数据的神经网络,该结构利用重要的局部特征重构原始数据,且输入数据的所有局部特征共享权值矩阵,因而该类 DAE 的隐含层能完整保存受局部空间限制的边缘特征。

对于单通道的输入 x 其 k 阶特征映射的隐含表示是:

$$h^k = \sigma(x * W^k + b^k) \quad (17)$$

其中 σ 是激活函数,一般采用 sigmoid 函数,“*”表示 2D 卷积运算, W^k 表示权值矩阵, b^k 表示偏置向量,卷积运算表达式为:

$$E(\theta) = \sum_{i=1}^n (x_i - y_i)^2 / 2n \quad (18)$$

该类 DAE 的重构函数为:

$$y = \sigma(\sum_{k \in H} h^k * \hat{W}^k + c) \quad (19)$$

其中 c 为每个数据通道的偏置, H 是隐含特征映射集, \hat{W} 是权值矩阵的批处理,权值更新的规则为随机梯度下降^[34]。值得注意的是,在该类 DAE 中,一个隐含映射对应一个偏置值,偏置向量 b 对整个映射均有效,并且每个映射负责捕捉数据的一个特征,便于

对神经网络进行预训练和精雕,有效缩短了特征提取的时间,简化了特征提取的过程,实现了数据特征的分层提取,卷积自动编码器原理如图 5 所示。

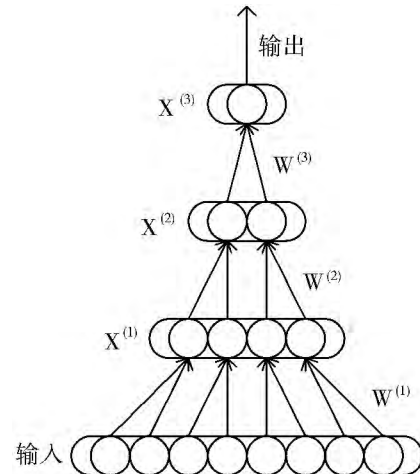


图5 卷积自动编码器原理

基于卷积理论的 DAE 用基本影像块的线性组合来拟合原有图像,显著提升了图像识别的速度和准确率。目前,该类 DAE 已成功完成了目标辨识、动态跟随和视觉模拟等任务,有效解决了原有 DAE 处理图像数据时出现的识别速度慢、准确率低、需要大量类标数据等问题^[8]。但是,由于理论和结构的一些不足,每个隐含层的输出都受到极大的限制,因而对预训练的依赖性很强,需要反复迭代,训练时间极其漫长(一般为几周甚至几千个小时)。另外,节点分布方式对图像识别的效果影响较大,但没有具体的理论指出节点分布具体如何影响训练结果,且该类 DAE 的构建和训练需要很多技巧,操作性较差。

5 DAE 存在的问题与未来展望

5.1 深度理论方面

1) 无监督逐层贪心预训练只是在一定程度上解决了局部最小问题,随着隐含层个数、神经元数量和数据复杂程度的增加,梯度稀释越发严重,现有方法依然不能遏制局部最小。基于梯度理论的随机初始化往往不能达到预期效果,使 DAE 不能拟合一些高维复杂函数,但没有文献指出其原因^[35]。

2) DAE 的研究有时需要结合计算机技术和统计理论,相应的神经生理科学发展缓慢,未能及时指导 DAE 的深入研究^[36-37],使得 DAE 的发展陷入困境。

3) DAE 的训练极其繁琐,先要用无监督逐层贪心预训练算法对每个隐含层进行逐层训练,达到一定程度后才能用 BP 算法、牛顿法、共轭梯度法、MOBP 或 SDBP 等算法对整个神经网络的参数进行系统性

优化调整,时间长,误差大,需要很多技巧,且学习到的知识表示的物理意义不明确。

另外,现有的 DAE 都是由不同类型的建模单元堆叠而成,这些建模单元均有理论缺陷,因而由它们构建的深度结构必然不完善。综上所述,完善深度学习理论,彻底搞清梯度理论与神经网络初始化的关系,找到能将逐层预训练与系统性参数优化结合的算法,并开发出新的建模单元和功能更加强大的深度结构,最终把各层学习到的知识表示成有物理意义的知识是 DAE 理论创新的必然要求。

5.2 建模策略方面

1) DAE 产生以后,神经网络的研究进入了崭新的阶段,但在以往的研究中没有文献说明设计神经网络时,层数与节点数的分布关系和权值分享方式对不同类型 DAE 造成何种影响^[38-39]。

2) 随着神经网络的隐含层数量和节点数的持续增加,目前的建模方法(堆叠、预训练、精雕)相对单一,不能满足用户日益提高的要求^[40]。

3) 现有的预训练算法和系统性参数优化策略对类标数据集有较强依赖性,远没有达到“真正意义上的无监督”^[41],并且没有文献指出二者以何种比例搭配能辅助训练,也没有文献讨论过不同的搭配方式对训练时间、收敛状况和不同类型代价函数等指标造成何种影响。随着今后研究的深入,搞清节点分布、隐含层数量和权值分享方式与不同类型 DAE 间的关系,提高无类标数据的利用率,用模块拼装的方法建立深度结构,将不同类型自动编码器单元引入同一个神经网络,实现 DAE 的多样化和模块化是 DAE 建模策略变革的必然趋势。

5.3 工程实践方面

1) DAE 处理的数据规模大,结构复杂,现有的软件和硬件难以满足用户需求。2) 不同类型 DAE 的原理不同,结构不一,现有的 DAE 不具有任务针对性,效率和准确率也迥然不同^[42]。3) DAE 的研究与应用跟传统神经网络相比有很大差异,因而相对独立。在以后的工程实践中,提高现有软件的数据处理能力和硬件的运算能力,开发针对深度结构的芯片和软件,快速实现大规模高维数据的分层特征提取,验证 DAE 的性能^[43],并根据不同类型 DAE 的特点构建神经网络完成相关任务,实现特定任务与特定种类 DAE 的对应,将 DAE 与传统神经网络结构有机结合,实现多种神经网络间的功能互补与强化是 DAE 应用实践的必经之路。

参考文献:

- [1] Arel L, Rose D C, Karnowski T P. Deep machine learning: a new frontier in artificial intelligence research [J]. Computational Intelligence Magazine, 2010, 5(4): 13-18.
- [2] Hinton G E, Osinder S, Teh Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527-1554.
- [3] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors [J]. Nature, 1986, 323: 533-536.
- [4] Bengio Y, Lamblin P, Popovici D, et al. Greedy layer-wise training of deep networks [C]. Proc. of the 20th Annual Conference on Neural Information Processing System. 2006: 153-160.
- [5] Vincent P, Larochelle H, Bengio Y. Extracting and composing robust features with denoising auto-encoders [C]. Proc. of the 25th International Conference on Machine Learning. 2008: 1096-1103.
- [6] Bengio Y. Learning deep architectures for AI [J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1-127.
- [7] Salah R, Vincent P, Muller X, et al. Contractive auto-encoders: Explicit invariance during feature extraction [C]. Proc. of the 28th International Conference on Machine Learning. 2011: 833-840.
- [8] Masci J, Meier U, Ciresan D. Stacked convolutional auto-encoders for hierarchical feature extraction [C]. Proc. of the 21th International Conference on Artificial Neural Networks. 2011, 6791: 52-59.
- [9] Guyon I, Dror G, Lemaire V, et al. Auto-encoders unsupervised learning and deep architectures [C]. Proc. of the 28th International Conference on Machine Learning, 2012: 37-50.
- [10] Vincent P, Larochelle H, Lajoie I, et al. Stacked denoising auto-encoders: Learning useful representations in a deep network with a local denoising criterion [J]. Journal of Machine Learning Research, 2010, 11: 3371-3408.
- [11] Mitchell B, Sheppard J. Deep structure learning: Beyond connectionist approaches [C]. Proc. of the 11th International Conference on Machine Learning and Applications. 2012: 162-167.
- [12] Erhan D, Bengio Y, Couville A, et al. Why does unsupervised pre-training help deep learning [J]. Journal of Machine Learning Research, 2010, 11: 625-660.
- [13] Li Deng, Selzer M L, Yu Dong, et al. Binary coding of speech spectrograms using a deep auto-encoder [C]. Proc. of the 11th International Speech Communication Association. 2010: 1692-1695.
- [14] Lee H, Kanadhamk C E. A sparse deep belief net model for

- visual area [C]. Proc. of the 21th Neural Information Process Systems International Conference. 2007: 873-880.
- [15] Amaral T, Silva L M, Alexandre L A, et al. Using different cost functions to train stacked auto-encoders [C]. Proc. of the 12th Mexican International Conference on Artificial Intelligence. 2013: 114-120.
- [16] Bengio Y, Delalleau O. On the expressive power of deep architectures [C]. Proc. of the 22nd International Conference on Algorithmic Learning Theory. 2011: 18-36.
- [17] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313: 504-507.
- [18] Gehring J, Miao Yajie, Metze F, et al. Extracting deep bottleneck features using stacked auto-encoders [C]. Proc. of the 26th IEEE International Conference on Acoustics, Speech and Signal Processing. 2013: 3377-3381.
- [19] Lange S, Riedmiller M. Deep auto-encoder neural networks in reinforcement learning [C]. Proc. of the International Joint Conference on Neural Networks. 2010: 18-23.
- [20] Suk H, Lee S W, Shen Dinggang. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis [J]. Brain Structure Function, 2013, 218(6): 1017-1036.
- [21] Iamsat S, Horata P. Handwritten Character Recognition Using Histograms of Oriented Gradient Features in Deep Learning of Artificial Neural Network [C]. Proc. of the 3rd International Conference on IT Convergence and Security. 2013: 1-5.
- [22] Gupta M, Lam S M. Weight decay back propagation for noisy data [J]. Neural Network, 1998, 11(6): 1127-1137.
- [23] 李海峰, 李纯果. 深度学习结构和算法比较分析 [J]. 河北大学学报(自然科学版), 2012, 32(5): 538-543.
- [24] Luo Xuxi, Li Wan. A novel efficient method for training sparse auto-encoders [C]. Proc. of the 6th International Congress on Image and Signal Processing. 2013: 1019-1023.
- [25] Deng Jun, Zhang Zixing, Erik M, et al. Sparse auto-encoder based feature transfer learning for speech emotion recognition [C]. Proc. of Humaine Association Conference on Affective Computing and Intelligent Interaction. 2013: 511-516.
- [26] Ma Yunlong, Zhang Peng, Gao Yanan. Parallel auto-encoder for efficient outlier detection [C]// Proceeding of IEEE International Conference on Big Data. 2013: 15-17.
- [27] 张开旭, 周昌乐. 基于自动编码器的中文词汇特征无监督学 [J]. 中文信息学报, 2013, 27(5): 1-7, 92.
- [28] Y Yang, S Guang, Shah M. Semi-supervised learning of feature hierarchies for object detection in a video [C]// Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 1650-1657.
- [29] Wong W K, S Mingming. Deep learning regularized fisher mapping [J]. Neural Networks, 2011, 22(10): 1668-1675.
- [30] C Minmin, Killan Q W, Xu Zhixiang. Marginalized denoising auto-encoders for domain adaptation [C]. Proc. of the 29th International Conference on Machine Learning. 2012: 538-542.
- [31] Q Hao, Y Zhe, Z Yajin. A new training principle for stacked denoising auto-encoders [C]. Proc. of the 7th International Conference on Image and Graphics. 2013: 384-389.
- [32] Schulz H, Cho K, Raiko T, et al. Two layers contractive encoding with shortcuts for semi-supervised learning [C]. Proc. of the 20th International Conference on Neural Information Processing. 2013: 450-457.
- [33] Gutmann M, Hyvarinen A. Noise-contrastive estimation: a new estimation principle for un-normalized statistical models [C]. Proc. of the 13th International Conference on Artificial Intelligence and Statistics. 2010: 297-304.
- [34] Firat O, Vural F T Y. Representation learning with convolutional sparse auto-encoders for remote sensing [C]. Proc. of the 21st Signal Processing and Communications Applications Conference. 2013: 24-26.
- [35] 孙志军, 薛磊, 许阳明. 深度学习研究综述 [J]. 计算机应用研究, 2012, 29(8): 2806-2810.
- [36] 庄永文. 关于系数编码理论及其应用 [J]. 现代电子技术, 2008(7): 157-160.
- [37] 刘海宁. 基于稀疏编码的设备状态识别及其重型轧辊磨床监测应用 [D]. 上海: 上海交通大学, 2011.
- [38] 孙志军, 薛磊, 许阳明. 基于深度学习的边际 Fisher 分析特征提取算法 [J]. 电子与信息报, 2013, 35(4): 805-811.
- [39] 殷力昂. 一种在深度结构中学习原型的分类方法 [D]. 上海: 上海交通大学, 2012.
- [40] Hinton G E, Krizhevsky A, Wang S. Transferring auto-encoders [C]. Proc. of Internet Corporation for Assigned Names and Numbers. 2011: 387-396.
- [41] 夏丁胤. 互联网图像高效标注和解译的关键技术研究 [D]. 杭州: 浙江大学, 2010.
- [42] Hinton G E, Li Deng, Dahl G E. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups [J]. Signal Processing, Magazine, 2012, 29(6): 82-97.
- [43] Guyon G, Dror V, Lemaire G. Auto-encoders, unsupervised learning, and deep architecture [J]. Journal of Machine Learning Research, 2012, 13: 37-49.