

浙江工业大学

本科毕业设计文献综述

(2015届)



论文题目: 自动编码器的研究与展望

作者姓名: 姜楠

指导教师: 龙胜春 副教授

学科(专业): 计算机科学与技术+自动化1101

所在学院: 计算机科学与技术学院

提交日期: 2015年2月

摘 要

深度学习是机器学习的一个重要新分支，开创了神经网络发展的新纪元。作为深度学习领域中的重要组成部分，自动编码器主要用于完成学习特征的任务，同时，在无监督学习以及非线性特征提取过程中也扮演了至关重要的角色。本文首先介绍稀疏自动编码器的历史发展流程，进而介绍其基本概念及原理，然后介绍它的构建方法以及训练方法的一般步骤，接着介绍了通用的代码框架，最后在深入分析自动编码器存在的问题的基础上，对其未来发展趋势进行展望。

关键词： 自动编码器，稀疏编码，无监督学习，神经网络

目 录

摘要	I
目录	
1 引言	1
2 自动编码器和稀疏编码的历史发展	2
3 数学模型及其原理	4
3.1 自动编码器的基本概念及原理	4
3.2 基于稀疏理论的自动编码器	5
3.3 基于统计理论的自动编码器	6
3.4 基于鲁棒理论的自动编码器	7
3.5 基于卷积理论的自动编码器	8
4 系统构建工具库	10
4.1 Python开发框架	10
4.2 Caffe开发框架	11
4.3 Matlab开发框架	11
5 总结与展望	12
5.1 深度理论方面	12
5.2 建模策略方面	12
5.3 工程实践方面	13
参考文献	14

1 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中，自适应效果好。但是，它受到人工特征提取的限制，步骤繁琐，运算量巨大，代价函数复杂且不易优化，因而收敛速度缓慢，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，继而产生了稀疏自动编码器（*Sparse Auto-encoder*）。自动编码器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，降级了原始输入的维度，得到数据的逆向映射特征，展现了从少数有标签样本和大量无标签样本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。

稀疏编码算法是一种无监督学习方法，它用来寻找一组“超完备”基向量来更高效地表示样本数据。其目的就是找到一组基向量 ϕ_i ，使得我们能将输入向量 \mathbf{x} 表示为这些基向量的线性组合： $X = \sum_{i=1}^k a_i \phi_i$ 。虽然形如主成分分析技术（PCA）能使我们方便地找到一组“完备”基向量，但是这里我们想要做的是找到一组“超完备”基向量来表示输入向量 $\mathbf{x} \in \mathbb{R}^n$ （也就是说， $k > n$ ）。超完备基的好处是它们能更有效地找出隐含在输入数据内部的结构与模式。然而，对于超完备基来说，系数 a_i 不再由输入向量 \mathbf{x} 唯一确定。因此，在稀疏编码算法中，我们另加了一个评判标准“稀疏性”来解决因超完备而导致的退化（degeneracy）问题。

本文在阐述稀疏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，更具深度自动编码区的建模机理对其进行分类和总结，指出稀疏自动编码器目前存在的问题，明确它的发展方向。

2 自动编码器和稀疏编码的历史发展

1. 1959年，David Hubel 和Torsten Wiesel^[2]通过对猫的视觉条纹皮层简单细胞感受野的研究得出一个结论：主视皮层V1区神经元的感受野能对视觉感知信息产生一种“稀疏表示”，基于这一知识，1961年，H.B.Barlow^[3]提出了“利用感知数据的冗余”进行编码的理论。
2. 1969年，D.J.Willshaw和O.P.Buneman^[4]等人提出了基于Hebbian学习的局部学习规则的稀疏表示模型，这种稀疏表示可以使模型之间有更少的冲突，从而使记忆能力最大化，Willshaw模型的提出表明了稀疏表示非常有利于学习神经网络的联想。
3. 1972年，Barlow^[5]推论出在稀疏性（Sparsity）和自然环境的统计特性之间必然存在某种联系。随后，有许多计算方法被提出来论证这个推论，这些方法都成功表明了稀疏表示可以体现出在大脑中出现的自然环境的统计特征。
4. 1986年，Rumelhart提出了自动编码器的概念，并将其用于高维复杂数据处理，促进了神经网络的发展^[6]。
5. 1987年，Field^[7]提出主视皮层V1区简单细胞的感受野非常适于学习视网膜成像的图像结构，因为它们可以产生图像的稀疏表示，基于这个结论，1988年，Michison明确提出了神经稀疏编码的概念^[8]，然后由牛津大学的E.T.Roll 等人正式引用。
6. 1989年，D.J.Field^[9]提出了稀疏分布编码方法。这种编码方法使响应于任意特殊信息的神经细胞数目被减少，信号的稀疏编码存在于细胞响应分布的四阶矩即峭度中。
7. 1996年，B.A.Olshausen和D.J.Field^[10]在Nature发表了一篇题为Emergence of simple cell receptive field properties by learning a sparse code for natural images 的重要论文，指出自然图像经过稀疏编码后得到的基函数类似于V1区简单细胞的感受野的反应特性。
8. 1997年，B.A.Olshausen和D.J.Field^[11]又提出了一种超完备的稀疏编码算法。

9. 1997年, Bell和Sejnowski^[12]把ICA用于自然图像分析, 并且得出一个重要的结论: ICA实际上就是一种特殊的稀疏编码。
10. 2006年, Hinton对原型自动编码器的结果进行改进, 进而产生了深度自动编码器, 先用无监督逐层贪心训练的算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能参数调整, 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情况^[13]。
11. 2007年, Benjio提出了稀疏自动编码器的概念, 进一步深化了自动编码器的研究^[14]。
12. 2008年, Vincent提出了降噪自动编码器, 在输入数据中添加噪声向量, 防止出现过拟合现象, 并取得了良好的效果^[15]。
13. 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^[16]。
14. 2010年, salah对升维和降维的过程加以限制, 提出了收缩自动编码器^[17]。
15. 2011年, Jonathan提出了卷积自动编码器, 用于构建卷积神经网络^[18]。
16. 2012年, Taylor对深度自动编码器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的深度结构^[19]。
17. 2012年, 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、卷积自动编码器和限制玻尔兹曼机等结构的性能, 为以后的实践和科研提供了参考^[20-24]。
18. 2013年, Telmo研究了用不同代价函数训练的自动编码器的性能, 为代价函数的优化策略的发展指明了方向^[25]。

3 数学模型及其原理

3.1 自动编码器的基本概念及原理

自动编码器是一种利用无监督预训练的非线性网络结构，从无类表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学习神经网络结构^[26]。

自动编码器由编码器、解码器组成。编码器是输入 x 到隐含层特征 h 的映射，计算公式为：

$$h = f(x) = S_f(Wx + b_n) \quad (3.1)$$

其中， S_f 是非线性激活函数，一般为 $sigmoid$ 函数，其表达式为：

$$S_f(x) = \frac{1}{1 + z^{-1}} \quad (3.2)$$

或者也可以是 $tanh$ 函数，其表达式为：

$$S_f(x) = \frac{1 - z^{-1}}{1 + z^{-1}} \quad (3.3)$$

解码器的函数 $g(h)$ 将隐含层数据映射回重构 y ，表示为：

$$y = g(h) = S_g(Wh + b_y) \quad (3.4)$$

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

$$J_{AE} = \sum_{x \in D} L(x, g(f(x))) \quad (3.5)$$

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：

$$L(x, y) = ||x - y||^2 \quad (3.6)$$

$$L(x, y) = - \sum_{i=1}^{d_x} [x_i \log y_i + (1 - x_i) \log(1 - y_i)] \quad (3.7)$$

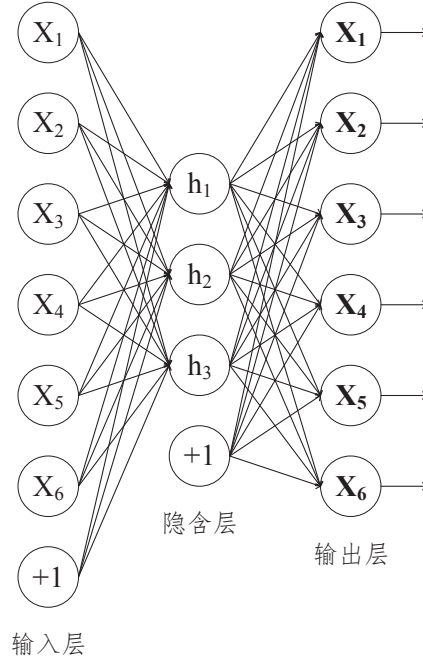


图 3.1 原型自动编码器，它由输入层、隐藏层和输出层构成。

其中，平方误差用于线性 S_g ，交叉熵损失函数用于 $sigmoid$ ^[27]。图3.1展示了原型自动编码器的结构。

为了适应不同的条件，满足不同的任务需求，研究人员对原型自动编码器进行调整产生了大致4类自动编码器，分别是：基于稀疏理论的自动编码器、基于统计理论的自动编码器、基于鲁棒理论的自动编码器、基于卷积理论的自动编码器。

3.2 基于稀疏理论的自动编码器

基于稀疏理论的自动编码器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 \mathbf{x} 是给定神经网络输入， 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 $\rho_j = \rho$ ，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：

$$KL(\rho||\rho_j) = \rho \log \frac{\rho}{\rho_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \rho_j} \quad (3.8)$$

此时的代价函数为：

$$J_{sparse}(W, b) = J(W, b) + \beta \sum_{j=1}^{s_2} KL(\rho||\rho_j) \quad (3.9)$$

其中， β 表示稀疏惩罚项的权重^[28]。

基于稀疏理论的自动编码器是目前应用最为广泛的自动编码器。2013年，邓俊将给予稀疏理论的自动编码器用于特征转换和语义情感识别取得了良好的效果^[29]。马云龙将其用于数据流的异常检测^[30]，张开旭将其用于中文词汇特征提取^[31]。

基于稀疏理论的自动编码器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，是分类边界变得更加清晰，并且能在一定程度上空盒子变量的规模，改变给定输入数据结构，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，并且该类自动编码器要求每个隐含层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。

3.3 基于统计理论的自动编码器

基于统计理论的自动编码器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，便于硬汉层对数据进行编码，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数据的分布参数，从而消除背景噪声^[32]，该类自动编码器的代价函数为：

$$J_{AE} = \sum_t E_q(\hat{x}|x^t) \quad (3.10)$$

其中， $E_q(\hat{x}|x^t)$ 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：

$$\hat{x} = x + \varepsilon, \varepsilon \sim N(0, \sigma^2 I) \quad (3.11)$$

其中， ε 表示整个自动编码器的正规化程度。为了便于计算，又是可用二项随机隐藏噪声代替高斯噪声^[33]。降噪自动编码器的训练规则为重构对数似然函数：

$$-\log P(x|c(\hat{x})) \quad (3.12)$$

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

基于统计理论的自动编码器是一种经过正规化的自动编码器，其内部映射均具有鲁棒性，腐坏向量的加入有效降低了自动编码器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^[34]。因此，基于统计理

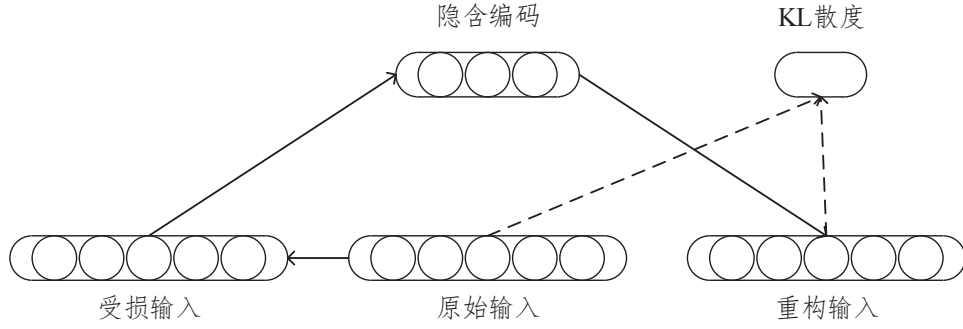


图 3.2 降噪自动编码器的设计原理

论的自动编码器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^[35]。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的影响，当原始输入是高维复杂函数时，重构误差显著增加。另外，该类自动编码器的计算效率极低，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拥有高速带宽的独立显存，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具有非常简洁而清晰的语法，适合完成各种高层任务，其动态反射语言不受操作系统的限制，可以在主流的Windows，Linux/Unix，Mac上面运行，并能面向对象编程，结合模式设计手段，能最大限度地满足系统设计变化需求。降噪自动编码器的设计原理如图3.2。

3.4 基于鲁棒理论的自动编码器

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 $J(x) = \frac{\partial f_\theta(x)}{a}$ ，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：

$$J_{CAE} = \sum_t L(x^t, g_\theta(f_\theta(x^t))) + \lambda \|J(x^t)\|_F^t \quad (3.13)$$

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，因此解析性收缩惩罚因子可表示为：

$$\|J(x^t)\|^2 = \sum_j (f_\theta(x)_j(1 - f_\theta(x)_j))^2 \|W_j\|^2 \quad (3.14)$$

由于解析惩罚因子的限制，基于鲁棒理论的自动编码器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，便意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^[36]。为了克服自动编码器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，使得代价函数变化为：

$$J_{CAE+H} = \sum_t L(x^{(t)}, g_\theta(x^{(t)})) + \lambda \|J(x^{(t)})\|_F^2 + \gamma E_\varepsilon[\|J(x) - J(x + \varepsilon)\|_F^2] \quad (3.15)$$

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，自动编码器的鲁棒性进一步得到提高，能出色地完成无监督特征转换热身，深化了特征表示学习算法的算法研究^[37]。但是由于基于鲁棒理论的自动编码器原理相对复杂，构建、训练和调整的难度较大，因而针对该类自动编码器的研究相对较少。

3.5 基于卷积理论的自动编码器

上文中讨论的自动编码器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制参与计算，使得计算效率较低，然而基于卷积理论的自动编码器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，因而该类自动编码器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：

$$h^k = \sigma(x * W^k + b^k) \quad (3.16)$$

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：

$$E(\theta) = \frac{\sum_{i=1}^n (x_i - y_i)^2}{2n} \quad (3.17)$$

该类自动编码器的重构函数为：

$$y = \sigma\left(\sum_{k \in H} h^k * \hat{W}^k + c\right) \quad (3.18)$$

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^[38]。值得注意的是，在该类自动编码器中，一个隐含映射对应一

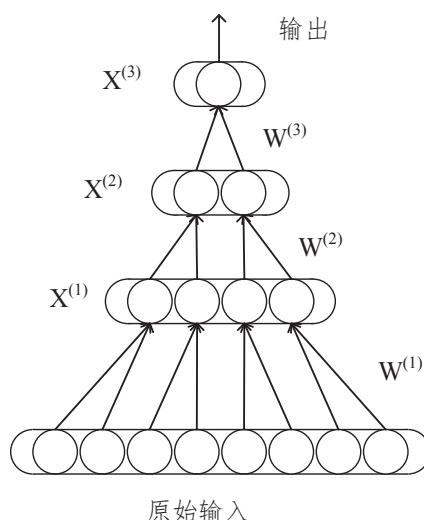


图 3.3 卷积自动编码器的设计原理

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调优，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3.3。

基于卷积理论的自动编码器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，该类自动编码器已成功完成目标辨识、动态跟随和视觉模拟等任务，有效解决了原有自动编码器处理图像数据是出现的识别速度慢，准确率低，需要大量类表数据等问题^[18]。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，训练时间及其漫长（一般为几周甚至几千小时）。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，且该类自动编码器的构建和训练需要很多技巧，操作性较差。

4 系统构建工具库

4.1 Python开发框架

Theano是Python中重要的第三方开发库，它与numpy能高效处理计算矩阵和向量。Theano可以抽象表达计算公式，然后被Theano的C编译器和GPU编译器编译成动态加载模块。Theano能高效稳定地优化图像抽象执行与计算。Theano支持图像的动态差分和GPU运算。Theano提供的功能足以满足先进的深度学习算法（包括了卷积算法），并能高效执行。

- **与numpy紧密结合：** 能直接接受numpy.ndarray类型数据输入，输出，并用于内部CPU计算。
- **自动设置使用GPU：** 能抽象的表示图像计算模型，所以用户无需考虑是否使用GPU参与计算，用户只需要在theano框架中配置GPU参数选项，让theano框架自动分配计算任务给CPU和GPU（GPU的矩阵运算比CPU快140倍）。
- **符号微分：** Theano能自动导出表达式的梯度计算式，它极大方便了梯度学习算法的代码表示。
- **速度和稳定性的优化：** 在生成字节码之前，Theano能重新排列并优化用户的计算式，它会将你的计算公式转换成范式，难以计算的函数会（例如： $\log(1+x)$ ）被转换成多项式模型，并对其一些特殊情况进行速度优化。
- **动态生成C代码：** Theano中提供的许多表达式都有对应的Python实现和C实现。这些C代码被特殊优化，能异常快速的计算，其缓存系统避免了冗余的再次编译。
- **丰富的特使和验证手段：** Theano中包含数以千计的单元测试样例，它们能快速检测计算过程中引入的回归误差。Theano同时也包含自我验证执行模式，以检测和诊断许多种的内部计算误差。

4.2 Caffe开发框架

Caffe¹是一个清晰而高效的深度学习框架，其作者是博士毕业于UC Berkeley的贾扬清²，他目前在Google工作。Caffe是纯粹的C++/CUDA架构，支持命令行、Python和MATLAB接口；可以在CPU和GPU直接无缝切换。Caffe的优势：

- **上手快**：模型与相应参数化都是通过配置文件设置，而非在实现代码中给出。Caffe给出了模型的定义、最优化设置以及预训练的权重，方便立即上手。
- **速度快**：能够出色承担实验研究和商业产品部署。Caffe与cuDNN结合使用，测试AlexNet模型，在K40上处理每张图片只需要1.17ms，说明这是迄今为止最快的运算框架。
- **模块化与开源化**：在它被开发的第一年中，来自全世界超过1,000开发者为期优化工作做出了贡献。由于这些优质的代码，我们才能实现最佳的模型设计和代码的高效执行。
- **社区开发维护**：Caffe已经成为学术研究项目和商业开发模型，甚至大规模的视觉、语音、多媒体的主要参考模型。Github上有其开发社区。

4.3 Matlab开发框架

Rasmusberg Palm学者开发了对应的深度学习的Matlab开发框架^[39]（DeepLearnToolbox³），并在github上进行开源。其主要功能包括：

- 深度信念网络模型；
- 堆叠自动编码器模型；
- 卷积神经网络模型；
- 卷积自动编码器模型；
- 简单神经网络模型。

¹<http://caffe.berkeleyvision.org/>

²<http://daggers.com/>

³<https://github.com/rasmusbergpalm/DeepLearnToolbox>

5 总结与展望

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、神经元数量和书籍复杂程度的增加，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，使自动编码器不能拟合一些高维复杂函数，但没有文献指出其原因^[40]。
2. 自动编码器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，未能及时指导自动编码器的深入研究^[41;42]，使得自动编码器的发展陷入困境。
3. 自动编码器的训练及其繁琐，先要用无监督逐层贪心预训练算法对隐含层逐层训练，达到一定程度后才能使用BP算法、牛顿算法、共轭梯度法、或SDBP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，现有的自动编码器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自动编码器理论创新的要求。

5.2 建模策略方面

1. 自动编码器产生以后，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，层数与节点数分布关系和权值分享对不同类型自动编码器造成何种影响^[43;44]。

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、调优）相对单一，不能满足用户日益提高的要求^[45]。
3. 现有的预训练算法和系统性参数优化策略对类标签数据集有较强的依赖性，远没有达到“真正意义上的无监督^[46]”，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自动编码器间的关系，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，实现自动编码器的多样化和模块化是自动编码器建模策略变革的必然趋势。

5.3 工程实践方面

1. 自动编码器处理的数据规模大，结构复杂，现有的软件和硬件难以满足用户需求。
2. 不同类型的自动编码器原理不同，结构不一，现有的自动编码器不具有任务针对性，效率和准确率也迥然不同^[47]。
3. 自动编码器的研究与应用跟传统神经网络相比有很大的差异，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，验证自动编码器的性能^[48]，并根据不同类型的自动编码器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的自动编码器的对应，将自动编码器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自动编码器应用实践的必经之路。

参考文献

- [1] Itamar Arel, Derek C Rose, Thomas P Karnowski. Deep machine learning-a new frontier in artificial intelligence research [research frontier][J]. Computational Intelligence Magazine, IEEE, 2010. 5(4):13–18.
- [2] David H Hubel, Torsten N Wiesel. Receptive fields of single neurones in the cat’s striate cortex[J]. The Journal of physiology, 1959. 148(3):574–591.
- [3] Horace B Barlow. The coding of sensory messages[J]. Current problems in animal behaviour, 1961:331–360.
- [4] David J Willshaw, O Peter Buneman, Hugh Christopher Longuet-Higgins. Non-holographic associative memory.[J]. Nature, 1969.
- [5] Horace B Barlow. Single units and sensation: a neuron doctrine for perceptual psychology?[J]. Perception, 2009. (38):795–8.
- [6] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams. Learning internal representations by error propagation[R]. :DTIC Document, 1985.
- [7] David J Field. Relations between the statistics of natural images and the response properties of cortical cells[J]. JOSA A, 1987. 4(12):2379–2394.
- [8] G Mitchinson. The organization of sequential memory: Sparse representations and the targeting problem[C]//Proceedings of the Bad Homburg meeting on Brain Theory. volume 1619. 1986.
- [9] David J Field. What the statistics of natural images tell us about visual coding[C]//OE/LASE’89, 15-20 Jan., Los Angeles. CA. International Society for Optics and Photonics, 1989:269–276.
- [10] Bruno A Olshausen, et al. Emergence of simple-cell receptive field properties by learning a sparse code for natural images[J]. Nature, 1996. 381(6583):607–609.
- [11] Bruno A Olshausen, David J Field. Sparse coding with an overcomplete basis set: A strategy employed by v1?[J]. Vision research, 1997. 37(23):3311–3325.
- [12] Anthony J Bell, Terrence J Sejnowski. The “independent components” of natural scenes are edge filters[J]. Vision research, 1997. 37(23):3327–3338.
- [13] Geoffrey Hinton, Simon Osindero, Yee-Whye Teh. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006. 18(7):1527–1554.

- [14] Yoshua Bengio, Pascal Lamblin, Dan Popovici, Hugo Larochelle, et al. Greedy layer-wise training of deep networks[J]. *Advances in neural information processing systems*, 2007. 19:153.
- [15] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders[C]//*Proceedings of the 25th international conference on Machine learning*. ACM, 2008:1096–1103.
- [16] Yoshua Bengio. Learning deep architectures for ai[J]. *Foundations and trends® in Machine Learning*, 2009. 2(1):1–127.
- [17] Salah Rifai, Pascal Vincent, Xavier Muller, Xavier Glorot, Yoshua Bengio. Contractive auto-encoders: Explicit invariance during feature extraction[C]//*Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. 2011:833–840.
- [18] Jonathan Masci, Ueli Meier, Dan Cireşan, Jürgen Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction. //*Artificial Neural Networks and Machine Learning–ICANN 2011*. Springer, 2011. 52–59.
- [19] Pierre Baldi. Autoencoders, unsupervised learning, and deep architectures.[J]. *ICML Unsupervised and Transfer Learning*, 2012. 27:37–50.
- [20] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion[J]. *The Journal of Machine Learning Research*, 2010. 11:3371–3408.
- [21] Ben Mitchell, John Sheppard. Deep structure learning: Beyond connectionist approaches[C]//*Machine Learning and Applications (ICMLA)*, 2012 11th International Conference on. volume 1. IEEE, 2012:162–167.
- [22] Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, Samy Bengio. Why does unsupervised pre-training help deep learning?[J]. *The Journal of Machine Learning Research*, 2010. 11:625–660.
- [23] Li Deng, Michael L Seltzer, Dong Yu, Alex Acero, Abdel-Rahman Mohamed, Geoffrey E Hinton. Binary coding of speech spectrograms using a deep auto-encoder.[C]//*Interspeech*. Citeseer, 2010:1692–1695.
- [24] Honglak Lee, Chaitanya Ekanadham, Andrew Y Ng. Sparse deep belief net model for visual area v2[C]//*Advances in neural information processing systems*. 2008:873–880.
- [25] Telmo Amaral, Luís M Silva, Luís A Alexandre, Chetak Kandaswamy, Jorge M Santos, Joaquim Marques de Sá. Using different cost functions to train stacked auto-encoders[C]//*Artificial Intelligence (MICAI)*, 2013 12th Mexican International Conference on. IEEE, 2013:114–120.
- [26] Yoshua Bengio, Olivier Delalleau. On the expressive power of deep architectures[C]//*Algorithmic Learning Theory*. Springer, 2011:18–36.
- [27] Geoffrey E Hinton, Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks[J].

- Science, 2006. 313(5786):504–507.
- [28] Yi Wan Yuxi Luo. A novel efficient method for training sparse auto-encoders[C]//Image and Signal Processing (CISP), 2013 6th International Congress on. 2013:1019 – 1023.
- [29] Marchi E. Schuller B. Jun Deng, Zixing Zhang. Sparse autoencoder-based feature transfer learning for speech emotion recognition[C]//Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on. 2013:511 – 516.
- [30] Yunlong Ma, Peng Zhang, Yanan Cao, Li Guo. Parallel auto-encoder for efficient outlier detection[C]//Big Data, 2013 IEEE International Conference on. IEEE, 2013:15–17.
- [31] 周昌乐张开旭. 基于自动编码器的中文词汇特征无监督学习[J]. 中文信息学报, 2013. 27:1–7.
- [32] Yang Yang, Guang Shu, Mubarak Shah. Semi-supervised learning of feature hierarchies for object detection in a video[C]//Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013:1650–1657.
- [33] Wai Keung Wong, Mingming Sun. Deep learning regularized fisher mappings[J]. Neural Networks, IEEE Transactions on, 2011. 22(10):1668–1675.
- [34] C Minmin, QW Killan, Xu Zhixiang. Marginalized denoising auto-encoders for domain adaptation[C]//Proc. of the 29th International Conference on Machine Learning. 2012:538–542.
- [35] Qianhaozhe You, Yu-Jin Zhang. A new training principle for stacked denoising autoencoders[C]//Image and Graphics (ICIG), 2013 Seventh International Conference on. IEEE, 2013:384–389.
- [36] Hannes Schulz, Kyunghyun Cho, Tapani Raiko, Sven Behnke. Two-layer contractive encodings with shortcuts for semi-supervised learning[C]//Neural Information Processing. Springer, 2013:450–457.
- [37] Michael Gutmann, Aapo Hyvärinen. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models[C]//International Conference on Artificial Intelligence and Statistics. 2010:297–304.
- [38] O Firat, FTY Vural. Representation learning with convolutional sparse autoencoders for remote sensing[C]//Signal Processing and Communications Applications Conference (SIU), 2013 21st. IEEE, 2013:1–4.
- [39] R. B. Palm. Prediction as a candidate for learning deep hierarchical models of data, 2012.
- [40] 孙志军, 薛磊, 许阳明, 王正. 深度学习研究综述[J]. 计算机应用研究, 2012. 29(8):2806–2810.
- [41] 庄永文. 关于稀疏编码理论及其应用[J]. 现代电子技术, 2008. 31(7):157–160.
- [42] 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][D]. :, 2012.
- [43] 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]. 电子与信息学报, 2013. 35(4):805–811.
- [44] 殷力昂. 一种在深度结构中学习原型的分类方法[D]. :上海交通大学, 2012.
- [45] Geoffrey E Hinton, Alex Krizhevsky, Sida D Wang. Transforming auto-encoders. //Artificial Neural Net-

-
- works and Machine Learning–ICANN 2011. Springer, 2011. 44–51.
- [46] 夏丁胤. 互联网图像高效标注和解译的关键技术研究[D][D]. :杭州: 浙江大学, 2010.
- [47] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups[J]. Signal Processing Magazine, IEEE, 2012. 29(6):82–97.
- [48] Pierre Baldi. Autoencoders, unsupervised learning, and deep architectures.[J]. ICML Unsupervised and Transfer Learning, 2012. 27:37–50.