

自编码器理论与方法综述^①

贾文娟, 张煜东

(南京师范大学 计算机科学与技术学院, 南京 210023)

通讯作者: 张煜东, E-mail: yudongzhang@ieee.org

摘要: 自编码器作为深度学习的一个重要分支, 吸引了该领域内大量杰出的研究者. 研究者们深入研究其本质并在此基础上提出了很多的优化方法, 如稀疏自编码器、降噪自编码器、收缩自编码器和卷积自编码器等. 在深入阅读了多篇基于自编码器方法的文献之后, 我们发现优化后的自编码器在图像分类、自然语言处理、目标识别等方面都取得了较好的实验结果. 因此, 本文将详细地分析优化后自编码器的基本结构和原理, 并对文献中的实验结果进行多方面的评价与分析.

关键词: 自编码器; 自然语言处理; 图像分类; 目标识别

引用格式: 贾文娟, 张煜东. 自编码器理论与方法综述. 计算机系统应用, 2018, 27(5): 1-9. <http://www.c-s-a.org.cn/1003-3254/6336.html>

Survey on Theories and Methods of Autoencoder

JIA Wen-Juan, ZHANG Yu-Dong

(School of Computer Science and Technology, Nanjing Normal University, Nanjing 210023, China)

Abstract: Autoencoder, as an important branch of deep learning, has appealed many outstanding researchers in this field. Researchers studied its essence, and proposed many optimized approaches, such as sparse autoencoder, denoising autoencoder, contractive autoencoder, and convolutional autoencoder. After reading a number of articles on autoencoder methods, we found that the optimized autoencoder had sound experimental results in terms of image classification, natural language processing, and object recognition. Therefore, this review analyzes the basic principle and structure of optimized autoencoder in details. In addition, the multi-perspectives evaluation and analysis on the experimental results in literatures are carried out as well.

Key words: autoencoder; natural language processing; image classification; object recognition

1 引言

自编码器是一种无监督学习算法, 其输出能够实现对输入数据的复现. 自编码器的概念是 Rumelhart 等^[1]最早提出的, 自编码器是一种数据压缩算法, 它使用编码器实现数据压缩, 使用解码器实现解压缩. 编码阶段将高维数据映射成低维数据, 减少数据量; 而解码阶段正好相反, 从而实现对输入数据的复现.

2006 年, Hinton 等^[2]将其作为一种降维方法发表在 Science 上. 文章通过使用“互补先验”的方法推导出

一个快速贪婪算法, 逐层学习深度定向网络, 然后对整个网络的参数进行优化, 改善了传统方法易陷入局部极小的情况, 由此产生了深度自编码器. Bengio 等^[3]在深度自编码器的研究过程中, 加入了稀疏性的限制, 使用尽可能少的神经元学习提取有用的数据特征, 进而产生了稀疏自编码器. Vincent 等^[4]为了提高算法的健壮性, 在输入数据中加入了噪声, 形成了去噪自编码器. 2010 年, Rifai 等^[5]在维度处理的过程中加以限制, 提出了收缩自编码器. 次年, Masci 等^[6]提出卷积自编码器

① 基金项目: 国家自然科学基金(61602250, 61503188); 江苏省自然科学基金(BK20150983, BK20150982); 江苏省高校自然科学研究面上项目(16KJB520025, 15KJB470010)

收稿时间: 2017-08-29; 修改时间: 2017-09-15; 采用时间: 2017-09-22; csa 在线出版时间: 2018-03-12

用于构建卷积神经网络.

自编码器在其优化发展过程中, 被应用到很多领域, 比如图像分类、人脸识别、自然语言处理以及其他领域, 并取得了较好的成绩. 本文也将从这些方面着重介绍自编码器的特点及性能, 并对这些方法进行分析 and 比较.

2 基础介绍

自编码器通过对输入数据进行压缩, 将其映射到低维空间中, 从而避免了“维度灾难”; 在解码阶段实现对输入数据的还原, 从而实现非监督的特征学习.

2.1 模型介绍

一个自编码器有编码器和解码器两部分组成 (如图 1 所示), 对于输入 x , 通过编码函数:

$$h = f(x) := s_f(Wx + p) \quad (1)$$

得到编码 h , 实现数据压缩. 其中 s_f 为编码器的激活函数, 通常取为 sigmoid 函数, 即:

$$s_f(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

W 表示输入层和隐层之间的权值矩阵. 在解码阶段使用解码函数:

$$r = g(h) := s_g(W^T h + q) \quad (3)$$

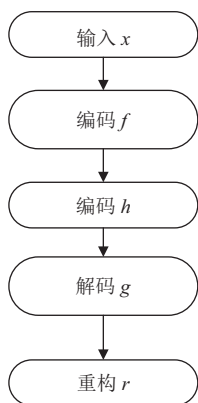


图 1 自编码器的基本结构

将编码 h 映射到原始高维空间, 从而达到对输入的复现. 其中 W^T 为隐层和输出层之间的权值矩阵 (为简单起见, 这里的权值矩阵取为 W 的转置), s_g 为解码器的激活函数, 通常取为 sigmoid 函数或恒等函数, 即:

$$s_g(z) = \frac{1}{1 + e^{-z}} \quad (4)$$

或:

$$s_g(z) = z \quad (5)$$

根据解码器激活函数的不同, $L(x, y)$ 通常有以下两种取法:

1) 当 s_g 为恒等函数时, 取平方误差函数, 即:

$$L(x, y) = \|x - y\|^2 \quad (6)$$

2) 当 s_g 为 sigmoid 函数 ($x_i \in [0, 1]$) 时, 取交叉熵 (cross-entropy) 函数, 即:

$$L(x, y) = - \sum_{i=1}^n [x_i \log(y_i) + (1 - x_i) \log(1 - y_i)] \quad (7)$$

在重构误差的基础上, 本文针对训练数据集 S 设置整体的损失函数 (loss function) 为:

$$\begin{aligned} J(W, b) &= \left[\frac{1}{m} \sum_{i=1}^n J(W, b; x^{(i)}, y^{(i)}) \right] \\ &+ \frac{\lambda}{2m} \sum_{l=1}^{m_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \\ &= \left[\frac{1}{m} \sum_{i=1}^n \left(\frac{1}{2} \|h_{W, b}(x^{(i)}) - y^{(i)}\|^2 \right) \right] \\ &+ \frac{\lambda}{2m} \sum_{l=1}^{m_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \end{aligned} \quad (8)$$

基本的自编码器可以看作一个三层的神经网络结构: 输入层、隐层、输出层, 其网络结构如图 2 所示. 图中 x_i 表示输入节点, x'_i 表示输出节点, “+1”表示偏置项, $h_{W, b}(x)$ 表示经过三层网络结构之后对输入数据的近似输出.

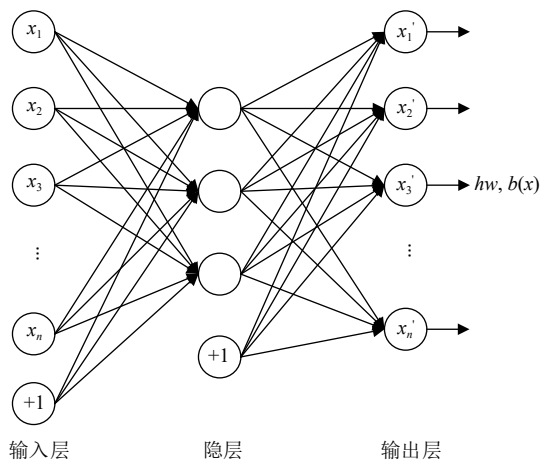


图 2 自编码器网络结构

2.2 自编码器算法流程

2.2.1 算法思路

自编码器算法的设计思路如下:

(1) 对于给定的无标签数据, 本文使用无监督学习方法学习特征. 对于无类标签的数据, 文中通过编码器将输入数据进行编码, 然后使用解码器得到一个输出信息, 如果输出近似等于输入数据, 则可以通过调整编码和解码阶段的参数, 使得重构误差最小.

(2) 将编码器产生的特征作为输入, 逐层训练下面层的网络. 由于第一个训练层得到的编码可以看作是输入数据的复现, 所以下面层的训练过程类似第一层.

有监督地微调整个神经网络. 在训练完所有层网络之后, 自编码器学习获得了可以较好地代表输入的特征, 并且这些特征可以最优地表示原始输入信号. 因此, 本文就可以在自编码器之后加入一个分类器(如 SVM 等), 然后通过有监督训练的方法实现对整个网络进行微调.

2.2.2 算法步骤

(1) 求各层激活值.

通过前向传播的方法, 对每一层的神经元进行前向传导计算, 利用前向传播公式, 得到各层的激活值.

(2) 计算 W 和 b 的残差, 利用梯度下降法更新 W 和 b , 使得输出更好的表示输入.

使用反向传播的方法, 求出最后输出层与个层神经元的残差, 利用梯度下降法不断更新 W 和 b , 使得输出更接近输入.

自编码器的基本流程如图 3 所示.

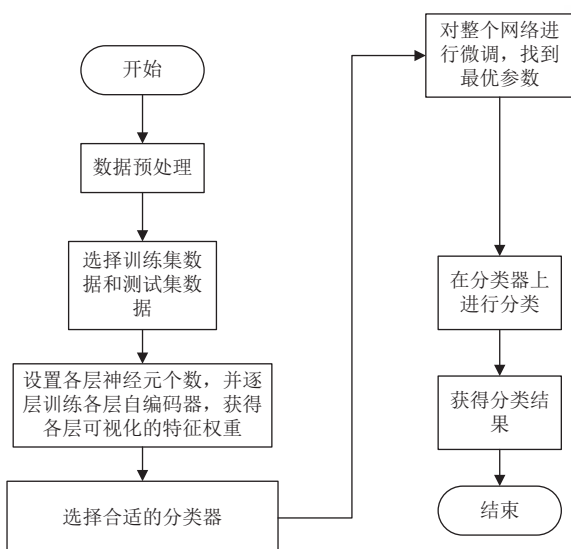


图 3 自编码器算法基本流程图

2.3 自编码器算法实现

这里本文使用 Matlab 代码来实现自编码器算法. 首先, 本文将训练数据加载到内存, 定义并训练自编码器神经网络, 设置各参数初始值, 在训练过程中, 本文可以得到特征权重的可视化结果. 对于分类问题而言, 文中使用 softmax 分类器对样本进行分类. 表 1 中定义了算法中使用的函数及其功能.

表 1 自编码器算法 Matlab 实现基本语句

函数语句	功能定义
trainAutoencoder()	无监督地训练自编码器
view()	可视化模型拓扑
plotWeights()	可视化学习得到的权重矩阵
TrainSoftmaxLayer()	应用于分类问题时, 使用softmax分类器进行分类

在表 1 中的第一条 Matlab 语句中, 本文需要设置自编码器神经网络中参数的初始值, 并定义隐层神经元的数量. 通过上述方法, 文中使用一幅磁共振病理脑图像进行训练, 设置隐层神经元个数为 100, 则可以得到病脑图像的可视化权重如图 4 所示.

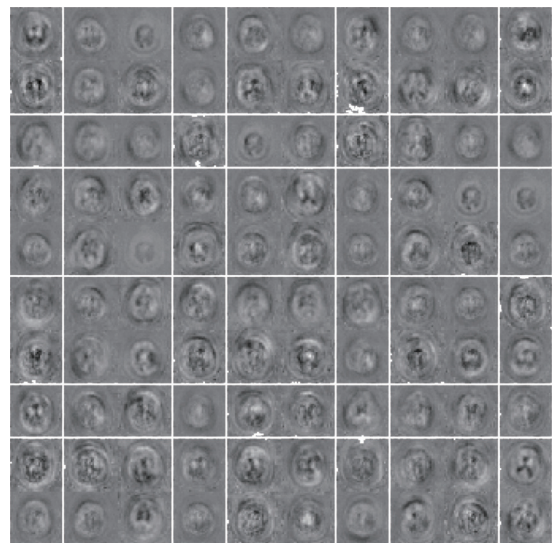


图 4 第一层自编码器训练得到的可视化权重

3 实际应用

在实际生活中, 自编码器已经应用到了很多方面, 如自然语言处理、图像分类、人脸识别等, 这些应用使得自编码器拥有了更广阔的发展前景.

3.1 自然语言处理

张开旭等^[7]利用自编码器能够实现无监督学习的特点, 从大量无标注语料中提取词汇的高维分布信息,

使用自编码器学习得到的低维特征来提升中文分词词性标注任务的效果. 文中还将自编码器方法与主成分分析和 k 均值聚类的方法进行了比较. 在宾州中文树库 5.0 数据集上的实验表明, 自编码器能够显著提高从大规模无标注语料中学习词汇特征的效果.

刘勘等^[8]使用微博数据作为分析对象, 利用基于深层噪声稀疏自编码器的短文本聚类算法进行无监督学习. 文章数据来源于大数据共享平台, 可分为 IT、财经、健康三大类, 每一类取 1500 条微博. 通过 Entropy 和 Precision 两种方法衡量聚类效果, 最终得到深层噪声稀疏自编码器 (DSAE)+ k -means 的方法效果最好, 综合信息熵为 0.207, 综合准确度为 87.8%.

胡庆辉等^[9]基于递归自编码器 (RAE) 方法, 使用真实的搜索引擎广告系统中的检索词-拍卖词对 (共 11 205 例), 并在其中随机选取词对进行人工标注. 实验结果表明, RAE 方法取得的实验准确率为 85.12%, 高于传统的 LDA 和 BM25 方法.

王贵新等^[10]采用稀疏自编码器和支持向量机 (SVM) 相结合的方法进行学习和分类仿真实验. 42 560 条垃圾短信和 70 000 条正常短信作为测试样本, 采用深度学习理论, 考虑词语在短信中的权重, 在文本分类中取得了较好的分类效果, 但是使用该方法的训练时间较长.

秦胜君等^[11]利用稀疏自编码器算法自动提取文本特征, 结合深度置信网络形成 SD 算法进行文本分类. 实验数据集来源于搜狗全网的新闻语料库, 可分为 20 个栏目的新闻数据, 从中随机抽取样本, 训练集数目分为 300, 800 和 1300. 实验结果表明, 文中使用的 SD 算法在 3 个训练集上的准确率分别为: 78.6%, 88.0%, 92.3%. 在同一数据集上, 与 SVM 方法进行比较得出, 当样本数量较少时, SD 算法容易出现过拟合现象, 准确率不如 SVM, 但是当数据集较大时, SD 算法的优势渐渐体现, 准确率和召回率都有所提升.

李翰清等^[12]采用 6000 余条家庭服务语料库 (包含 60 个意图) 进行实验, 利用深度去噪自编码器方法提取指令的高阶特征、高斯核支持向量机进行训练和预测, 在所建语料库上进行多折交叉验证, 结果表明指令意图理解的平均准确率达到 96% 以上.

魏扬威等^[13]首先根据英语作文写作技巧提取了大量语言学特征, 再使用自编码器、特征离散化方法对特征进行重构, 然后用分层多项模型来输出文章的最终得分. 实验表明, 在使用自编码器将特征压缩到

100 维时, 相关系数达到 0.787, 与直接使用支持向量机进行回归的实验相比, 准确率提升了 3.4%.

刘广秀等^[14]通过自编码器模型网络建造 Adaboost 框架下的弱分类器, 并在自编码器神经网络部分引入噪声, 更改不同参数以及层数构造弱分类器, 利用稀疏性来提高分类器的泛化能力. 实验使用 NLPiR 分词系统提取文本特征、TFIDF 作为词语的权值, 根据权值来选择特征词, 并统计词频作为文本特征训练集. 以调节自编码器参数等方式实现每个分类器的结构互异性, 提高各分类器的分类各异性, 从而实现网络的结构设计使得不同弱预测器具有不同的预测倾向性, 提高系统泛化能力, 提高分类器分类效果.

李阳辉等^[15]使用改进的降噪自编码器 (IDAE) 对微博情感进行分析, 模型的训练集和测试集通过 COAE2014 的微博评测数据集获取, 为检验模型的抗干扰能力, 作者采用 Python 方法从微博网页上直接爬取实时数据, 并将其作为样本集特意加入实验当中. 结果表明该方法聚类效果较好, 且准确度和稳定性与 DAE 相比有所提升.

刘红光等^[16]在对中文新闻文本进行分类时, 采用降噪自编码器实现对文本的压缩及分布式的表示, 并使用 SVM 算法进行分类. 实验结果证明, 随着样本数目的增大, 分类准确率、召回率等都在提高, 平均分类准确率达到 95% 以上, 但数据量依然较小, 并不能完全发挥该方法并行处理大容量数据的优势.

表 2 列出了自编码器在自然语言处理中的应用以及其优化后的方法, 方法的特点中也介绍了该方法应用于该领域的优势及不足.

3.2 图像分类

林少飞等^[17]将稀疏自编码器方法应用到数字识别中, 通过多个系数自编码器的堆叠构建成的深度学习数字图像的特征, 最后使用 softmax 分类器进行数字分类. 实验通过 matlab 平台实现, 使用 MNIST 数据库 (训练集 60 000 张手写数字图像, 测试集 10 000 张图像), 设置 softmax 层回归迭代 100 次, 第一层稀疏自编码器迭代 100 次, 第二层迭代 100 次, 微调部分迭代达到 100 次之后, 实验准确流畅高达 98.00%.

尹征等^[18]采用标准卷积自编码器方法, 选取 MINIS 和 CIFAR-10 数据集进行训练, 实验表明将卷积层相关理论运用到自编码器中能提高网络的识别性能, 并且实验识别率提高了 2%, 识别错误率降低了近 6%, 并且该方法对于加入了噪声的数据具有较强的鲁棒性.

表2 自然语言处理中自编码器的应用

文献	方法	适用范围	优点	缺点
张开旭等 ^[7]	DAE + SAE	中文分词词性标注	避免了由于特征过多造成的稀疏问题	没有解决词串、句法关系等的特征表示
刘勘等 ^[8]	DSAE + k-means	微博数据分析	能够在降低向量维度的同时,有效地提取数据的本质特征	数据只分为IT、财经、健康三类,并不能应用到其他类别
胡庆辉等 ^[9]	RAE	搜索引擎广告系统	融入了权重信息,降低了权重较大的词语在短语中的贡献	选取词对进行人工标注,耗时很多
王贵新等 ^[10]	SAE + SVM	生活短信	能有效处理垃圾短信,并且短信数量较大	训练时间较长
秦胜君等 ^[11]	SD	搜狗全网的新闻语料库	当数据集较大时,过拟合等问题可被解决,方法优势明显	样本较少时,容易产生过拟合
李翰清等 ^[12]	DAE + SVM	家庭服务语料库	采用多折交叉验证的方法,算法准确率高	数据集数据较少,泛化能力不强
魏扬威等 ^[13]	SAE + HMM	英语作文写作技巧	提取出的特征具有一定的主题无关性	没有考虑到人工评分的主观性
刘广秀等 ^[14]	Adaboost + DSAE	文本分类	泛化性能较强,分类效果较高	文中没有具体的实验证明
李阳辉等 ^[15]	IDAE	微博情感分析	提高了情感分析系统的稳定性和抗干扰能力	实验中,并没有为降噪自动编码器设计深度
刘红光等 ^[16]	DAE + SVM	中文新闻文本分类	实验准确率较高,实验充分体现了该方法与其他方法相比较的优势	数据量较小,泛化能力较弱

樊养余等^[19]使用一种基于卷积稀疏自编码器的方法,对少量有标记抽象图像进行自适应的学习,并使用平均梯度准则对所学到的权重进行排序,用于对基于不同领域的特征学习结果进行比较.实验使用 Abstract100、Abstract280 数据库中的数据进行分类,结果表明,该方法学习到的特征不仅能被用于认知层面的图像识别,还能够被用于情绪语义层面的图像辨识,这些结论也给深度学习在其他有限样本集合中的应用带来启发.

戴晓爱等^[20]引用栈式稀疏自编码器自学习原始数据的特征表达,然后将其连接到支持向量机分类器,完成微调过程.实验采用挪威 NEO 公司 Hyspex 成像光谱仪以地面成像方式在选定试验区获取的数据进行模拟性能测试,选取全部 7 类 108 个波段数据,实验精度达到 87.82%.但是该方法模型参数调整复杂,易陷入局部最小或过拟合状态,从而影响分类结果.

王杰等^[21]使用稀疏自编码器的方法来对烟叶成熟度进行分类,降低了人工成本.实验样本取自河南农业大学科教实验园区的烟叶试验田.该算法直接从像素层面提取所需特征,减少了人为选定特征的时间,并且避免了某些重要特征的丢失,分类精确度高达 98.63%.

张一飞等^[22]为了提高传统遥感图像分类准确率较低的问题,使用一种基于栈式去噪自编码器的分类方法,采用无监督逐层分类的方法获得更具鲁棒性的特

征表达,然后通过误差反向传播方法对整个网络参数进行微调.实验采用国产高分一号遥感数据,分类精度高达 95.7%,高于传统的支持向量机和 BP 神经网络方法,但是该方法时间消耗很大,因此可能限制其大规模、深层次模型的应用.

刘洋洋等^[23]提出将多尺度单层自编码器的方法应用到医学图像自动分类中,选择 ImageCLEFMed 2009 数据集中 20 类共 4780 幅图像,其中 3350 幅作为训练集,选择原测试数据集中的 1430 幅作为本实验测试样本,实验学习到的特征表示和总体准确率明显由于传统方法.

徐守晶等^[24]基于 Matlab 平台,使用 CIFAR-10 数据库中的 STL-10 数据集,利用栈式去噪自编码器 (SDA) 的方法,对包括飞机、鸟、小汽车、猫、狗、轮船、卡车、马、猴子、鹿等 10 类图像进行分类,结合 SIFT 和 ISDA 特征检索查准率为 91.3%.

Zeng 等^[25]开发了一种将单一图像 SR 耦合到深度自动编码器的数据传输模型,通过探索广泛的数据自动学习 LR/HR 图像补丁的内在表示及其关系,将 LR 表示精确地映射到其对应的 HR 表示的大数据传输函数.与 Set5 和 Set14 数据集上的其他先进的算法相比较,实验证明了文中提出方法具有卓越的有效性和效率.

表 3 列出了自编码器在图像分类中的应用以及其优化后的方法,方法的特点中介绍了该方法应用于该领域的优势及不足.

表3 图像分类中自编码器的优化与应用

文献	方法	适用范围	优点	缺点
林少飞等 ^[17]	SSAE + softmax	数字识别	算法准确率很高	未能深入研究网络层数对特征学习的影响
尹征等 ^[18]	CDAE	数字图像识别	算法的鲁棒性较强, 与传统的自编码器相比, 准确率较高	时间复杂度作为检验算法性能的指标之一, 文中并没有介绍
樊养余等 ^[19]	CDAE	认知层面和情绪语义层面的图像识别	将自编码器应用于情绪识别, 具有一定的创新性	人工方法对图像进行评价打分具有一定的主观性
戴晓爱等 ^[20]	SSAE + SVM	高光谱影像分类	分类性能高于文中其他算法	该方法模型参数调整复杂, 易陷入局部最小或过拟合状态
王杰等 ^[21]	SAE	烟叶成熟度分类	避免了重要特征的丢失, 降低了人工成本	没有考虑试验田中的烟叶与市场上烟叶的差别
张一飞等 ^[22]	SDAE	遥感图像分类	总体分类精度高达95.7%, 且高于文中其他方法的实验准确率	时间消耗大
刘洋洋等 ^[23]	多尺度单层 SAE + SRC	医学图像分类	方法简单, 适用性广, 精确度高	文中没有考虑该算法的时间复杂度
徐守晶等 ^[24]	SDAE + softmax	图像识别	一定程度上解决了训练过程缓慢的问题	但是算法鲁棒性较弱
Zeng等 ^[25]	CDA	具有超分辨率的单图像	在不同数据集上与其他先进的算法相比, 具有较好的有效性及效率	研究工作没有涉及该算法在其他领域的应用

3.3 目标识别

张宁等^[26]基于深度自编码器设计了机场安检中的人脸识别系统, 使用 Gabor 小波函数作为卷积核函数, 在对原始图像分块的基础上进行卷积运算, 采用多个 RBM 堆叠组成的自编码器, 通过比较差异算法训练 RBM, 实现自动提取人脸特征, 最后构建一个三层的 BP 神经网络, 将自动提取的人脸特征特作为输入, 将图像标签作为输出层, 通过反向传播算法训练网络进行人脸识别. 实验使用 YALE 数据库中的图片进行训练, 算法的平局正确率高达 99.08%.

赵飞翔等^[27]提出了一种基于栈式降噪稀疏自编码器的雷达目标识别方法, 通过设置不同隐层数量和迭代次数, 从雷达数据中高效地提取出各层次特征. 实验采用暗室仿真数据, 通过多次实验, 设置隐层数为 4 层, 隐层单元数分别为 150-100-50-10, 训练迭代次数依次为 10、30、50, 实验结果均高于 90%.

康妙等^[28]提出基于栈式自编码器 (SAE) 特征融合的合成孔径雷达 (SAR) 图像车辆目标识别算法, 然后利用 softmax 分类器微调网络, 提高了网络的融合性能. 实验基于 MSTAR 数据集的 10 类目标分类精度达 95.88%.

张彦等^[29]使用所有已知样本训练深度自编码器, 得到广义深度自编码器, 然后使用但样本目标的单个样本微调广义深度自编码器, 得到特定类别的深度自编码器, 识别时, 将识别图像输入每个特定类别的深度自编码器, 得到包含与测试图像相同类内变化的该类别的重构图像, 并用其训练 softmax 回归模型, 分类测

试图像. 实验在公共测试库 Extended Yale Face Database B 及 AR Database 上进行, 在相同环境下, 比其他算法获得了更高的识别率, 并且识别一幅图像的所需平均时间更少.

李江等^[30]针对传统方法识别人脸表情具有较差的鲁棒性的问题, 提出了一种基于降噪自编码器的方法, 同时为了避免由于样本不足造成的过拟合现象, 在全连接层采用了 Dropout 技术. 实验数据集采用 CK+、JAFPE 和 Yale, 每个数据集的平均准确率分别为 99.65%、98.92%、99.67%, 该方法也表明了该方法较强的鲁棒性和抗身份信息干扰的能力.

冉鹏等^[31]在处理人脸识别问题时, 采用了栈式降噪自编码器的方法. 该方法可分为无监督的预训练和有监督的训练两个阶段. 无监督预训练阶段能够自动学习人脸的一般特性, 有监督训练阶段进行人脸识别的测试. 实验使用 YALE、AR、FERE 三个人脸数据库进行实验, 实验表明, 在不同数据库中, 文中提出的方法得到的测试准确率均由于其他方法.

于万钧等^[32]为了提高气体识别的准确率, 提出了一种基于栈式降噪自编码器的方法, 并与千层机器学习算法的气体定性识别方法进行比较. 实验使用 UCI 机器学习气体数据集进行训练, 测试准确率达到 82.2%, 高于传统手工提取特征的方法, 并且过程简单, 改善了传统方法的复杂过程.

表 4 列出了自编码器在目标识别中的应用以及其优化后的方法, 方法的特点中也介绍了该方法应用于该领域的优势及不足.

表4 目标识别中自编码器的优化与应用

文献	方法	适用范围	优点	缺点
张宁等 ^[26]	Gabor + DAE	机场安检中的人脸识别	识别率高, 收敛速度快	没有考虑人脸被遮挡等情况
赵飞翔等 ^[27]	SDAE	雷达目标识别	实验准确率均高于90%	该方法不能确定其抗噪性
康妙等 ^[28]	SAE + softmax	SAR图像车辆目标识别	使用softmax分类器微调网络, 提高了网络的融合性能	SAR图像获取比较困难, 代价较高
张彦等 ^[29]	DAE + softmax	人脸识别	识别率高, 耗时少	随着样本复制倍数增加时, 使用该方法会产生过拟合现象
李江等 ^[30]	DAE + Dropout	人脸表情识别	该方法具有较强的鲁棒性和抗身份信息干扰的能力	该方法只针对大小尺度较小的图像, 使用范围有一定的局限性
冉鹏等 ^[31]	SDAE	人脸识别	避免了过拟合情况的发生	测试集数据较少, 并且文中没有分析时间复杂度
于万钧等 ^[32]	SDAE	气体识别	改善了传统方法的复杂过程, 识别准确率较高	实验数据获取困难, 无法验证出模型学习到的具体内容

3.4 其他领域

陈国定等^[33]将堆栈式稀疏自编码器应用到评估尾矿库的安全状况中, 实验数据采自于淳安某尾矿库, 该算法解决了多层网络结构权值易陷入局部最小的问题, 采用微调技术之后, 该方法的准确率高达 99.27%, 但该方法作为深度神经网络的一种, 仍处于学术界的广泛讨论中, 将其应用于尾矿库安全评价仍需进一步的研究和实验, 以获取更优的效果。

李荣雨等^[34]将平行堆栈式自编码器应用于过程工业预测建模中, 首先通过聚类算法对输入数据的属性进行聚类, 然后按结果将数据分类后输入并行的稀疏自编码器中进行特征的模块式提取, 联合学习得到的特征再进行逐层学习得到拟合结果. 在对加氢裂化的预测建模研究过程中, 本方法获得的平均预测准确率在 95% 以上, 在预测能力和稳定性方面高于栈式自编码器和多层感知器。

李远豪^[35]提出基于深度自编码器的人脸美丽吸引力预测模型, 利用深度自编码器网络提取人脸图像美丽特征, 然后用支持向量机分类器来预测人脸美丽. 在中国科学院 CAS-PEAL-R1 共享人脸库和互联网手机的人脸图像中, 得到该模型的预测性能达到了 77.3%. 但是所学到的特征鲁棒性较低, 没有达到理想的预测效果, 这有待于后续深入研究。

于乃功等^[36]基于深度自编码器与 Q 学习, 提出了移动机器人路径规划方法, 利用深度自编码器学习机器人所处环境特征, Q 学习根据环境信息选择机器人要执行的动作, 从而让机器人通过与环境的交互, 实现自主学习. 机器人仿真实验证明了该方法的有效性。

寇茜茜等^[37]将自编码器应用于金融领域, 提出了汇率时间序列预测的方法. 首先通过自编码器对时间序列进行无监督的特征识别与学习, 然后利用有监督微调的方法调整权值. 实验使用来源于真实金融市场的 Time Series Data Library 数据库, 文中使用的方法的预测效果优于传统的 BP 神经网络方法和 SVR 方法. 但是反向传导算法在网络层次过深时, 会出现梯度扩散的现象, 因此更适合网络层次不是特别深的预测。

王宪保等^[38]利用堆栈式降噪自编码器的方法进行胶囊缺陷检测, 首先对样本进行降噪训练, 获取网络初始权值, 之后通过 BP 算法进行微调, 得到样本与无缺陷模版之间的映射关系, 通过对比实现测试样本的缺陷检测. 训练样本为包括裂纹、空洞和胶囊帽缺失的数据, 实验取得了 100% 的缺陷胶囊检出率, 但是该方法只能处理 64×64 维的图像, 处理高精度的缺陷图像将是今后研究主要工作。

SUN 等^[39]为了实现感应电动机的故障诊断, 提出了一种将去噪自编码融入稀疏自编码器的深度神经网络的方法, 提出了算法的鲁棒性, 结合反向传播算法对网络进行微调, 提高了分类的准确度, 通过引入“Dropout”训练方法, 降低了由过拟合带来的预测误差. 实验选用了 50 Hz 供电频率下的电动机, 以 Y 轴方向的振动信号作为试验处理信号, 信号采样频率为 20 kHz. 实验得到的不同故障的平均识别正确率为 97.61%。

表 5 列出了自编码器在其他领域中的应用以及其优化后的方法, 方法的特点中也介绍了该方法应用于该领域的优势及不足。

表5 自编码器在其他领域的应用

文献	方法	应用领域	优点	缺点
陈国定等 ^[33]	SSAE + Fine-tune	尾矿库安全评估	该算法不易陷入局部最小的问题,并且实验准确率较高	仍处于初步尝试阶段,没有一定的理论证明
李荣雨等 ^[34]	PSAE	工业过程建模预测	在预测能力和稳定性方面高于栈式自编码器和多层感知器	在将训练数据按特征簇类型分类时,没有考虑聚类算法的准确率
李远豪 ^[35]	DAE + SVM	人脸美丽吸引力预测	该模型的预测性能高于文中的其他算法	学到的特征鲁棒性较低
于乃功等 ^[36]	DAE + Q学习	移动机器人路径规划	提高了系统的收敛速度,缩短了学习时间	文中强化学习处理的状态时离散的,而真实环境中机器人的状态时连续的
寇茜茜等 ^[37]	SAEP + Fine-tune	汇率时间序列预测	实验结果优于文中的其他方法	文中提出的模型只适合于低配置计算资源中网络层次不是特别深的预测中
王宪保等 ^[38]	SDAE + BP算法	胶囊缺陷检测	实验中算法的准确率高达100%	无法处理更高像素精度的图像
SUN等 ^[39]	DSAE + BP算法+ Dropout	感应电动机的故障诊断	(1) 鲁棒性更强; (2) 减少了过拟合现象的发生; (3) 具有更强的优化能力	在进行数据采集时,没有考虑不同供电频率以及信号采样频率的影响

4 总结与展望

目前,自编码器在图像分类、自然语言处理等多个方面都有了广泛的应用,并且取得了不错的成果,这也促进了自编码器的发展,使其更有效地为人类服务。

(1) 自编码器能够自动提取特征,这有效降低了传统手动提取特征的不足;

(2) 对于大数据训练问题,自编码器能有效地避免过拟合情况的发生;

(3) 在各文献中的实验中,与传统的神经网络相比,自编码器方法取得的准确率都较高;

(4) 通过总结各文献的应用领域,我们发现自编码器方法的应用范围较为广泛。

我们希望通过日后的研究工作,自编码器能够更有效、更广泛地应用于我们的日常生活需要当中。因此,针对自编码器存在的以下不足:

(1) 深度自编码器需要逐层训练每一隐层,然后才能使用某种算法对整个网络进行优化调整,学习时间较长;

(2) 自编码器无监督学习的特性使得其学习获得的特征表示的物理意义并不明确;

(3) 深度自编码器处理的数据规模大,结构较为复杂,现有的软硬件设备均难以满足用户需求。

我们日后的研究重点将放在如何解决上述问题当中,使自编码器更有效地发挥其优势,为人类社会服务。

参考文献

1 Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*, 1986, 323(6088): 533-536. [doi: 10.1038/323533a0]

2 Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006, 18(7): 1527-1554. [doi: 10.1162/neco.2006.18.7.1527]

3 Bengio Y, Lamblin P, Popovici D, et al. Greedy layer-wise training of deep networks. *Proceedings of the 19th International Conference on Neural Information Processing System*. Canada, 2006. 153-160.

4 Vincent P, Larochelle H, Bengio Y, et al. Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th International Conference on Machine Learning*. Helsinki, Finland, 2008. 1096-1103.

5 Rifai S, Vincent P, Muller X, et al. Contractive auto-encoders: Explicit invariance during feature extraction. *Proceedings of the 28th International Conference on Machine Learning*. Bellevue, WA, USA, 2011. 833-840.

6 Masci J, Meier U, Cireşan D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction. *Proceedings of the 21st International Conference on Artificial Neural Networks*. Espoo, Finland, 2011. 52-59.

7 张开旭, 周昌乐. 基于自动编码器的中文词汇特征无监督学习. *中文信息学报*, 2013, 27(5): 1-7, 92.

8 刘勘, 袁蕴英. 基于自动编码器的短文本特征提取及聚类研究. *北京大学学报(自然科学版)*, 2015, 51(2): 282-288.

9 胡庆辉, 魏士伟, 解忠乾, 等. 基于递归自编码器的广告短语相关性. *计算机应用*, 2016, 36(1): 154-157, 187. [doi: 10.11772/j.issn.1001-9081.2016.01.0154]

10 王贵新, 彭娟, 郑孝宗, 等. 基于稀疏自编码器和 SVM 的垃圾短信过滤. *现代电子技术*, 2016, 39(17): 145-148.

11 秦胜君, 卢志平. 稀疏自动编码器在文本分类中的应用研究. *科学技术与工程*, 2013, 13(31): 9422-9426. [doi: 10.3969/j.issn.1671-1815.2013.31.051]

12 李瀚清, 房宁, 赵群飞, 等. 利用深度去噪自编码器深度学习的指令意图理解方法. *上海交通大学学报*, 2016, 50(7):

- 1102–1107.
- 13 魏扬威, 黄萱菁. 结合语言学特征和自编码器的英语作文自动评分. 计算机系统应用, 2017, 26(1): 1–8. [doi: [10.15888/j.cnki.csa.005535](https://doi.org/10.15888/j.cnki.csa.005535)]
- 14 刘广秀, 宋单单. 基于 Adaboost 框架下自动编码器提升方法的文本分类. 电子世界, 2016, (11): 195, 197.
- 15 李阳辉, 谢明, 易阳. 基于降噪自动编码器及其改进模型的微博情感分析. 计算机应用研究, 2017, 34(2): 373–377.
- 16 刘红光, 马双刚, 刘桂锋. 基于降噪自动编码器的中文新闻文本分类方法研究. 现代图书情报技术, 2016, 32(6): 12–19. [doi: [10.11925/infotech.1003-3513.2016.06.02](https://doi.org/10.11925/infotech.1003-3513.2016.06.02)]
- 17 林少飞, 盛惠兴, 李庆武. 基于堆叠稀疏自动编码器的手写数字分类. 微处理机, 2015, (1): 47–51.
- 18 尹征, 唐春晖, 张轩雄. 基于改进型稀疏自动编码器的图像识别. 电子科技, 2016, 29(1): 124–127.
- 19 樊养余, 李祖贺, 王凤琴, 等. 基于跨领域卷积稀疏自动编码器的抽象图像情绪性分类. 电子与信息学报, 2017, 39(1): 167–175.
- 20 戴晓爱, 郭守恒, 任滔, 等. 基于堆栈式稀疏自编码器的高光谱影像分类. 电子科技大学学报, 2016, 45(3): 382–386.
- 21 王杰, 贾育衡, 赵昕. 基于稀疏自编码器的烟叶成熟度分类. 烟草科技, 2014, (9): 18–22.
- 22 张一飞, 陈忠, 张峰, 等. 基于栈式去噪自编码器的遥感图像分类. 计算机应用, 2016, 36(S2): 171–174, 188.
- 23 刘洋洋, 唐奇伶. 基于多尺度单层自编码器的医学图像分类. 科教导刊, 2017, (15): 20–21, 192.
- 24 徐守晶, 韩立新, 曾晓勤. 基于改进型 SDA 的自然图像分类与检索. 模式识别与人工智能, 2014, 27(8): 750–757.
- 25 Zeng K, Yu J, Wang RX, *et al.* Coupled deep autoencoder for single image super-resolution. IEEE Transactions on Cybernetics, 2017, 47(1): 27–37. [doi: [10.1109/TCYB.2015.2501373](https://doi.org/10.1109/TCYB.2015.2501373)]
- 26 张宁, 朱金福. 基于深度自动编码器的机场安检人脸识别系统设计. 计算机测量与控制, 2015, 23(2): 644–647.
- 27 赵飞翔, 刘永祥, 霍凯. 基于栈式降噪稀疏自动编码器的雷达目标识别方法. 雷达学报, 2017, 6(2): 149–156.
- 28 康妙, 计科峰, 冷祥光, 等. 基于栈式自编码器特征融合的 SAR 图像车辆目标识别. 雷达学报, 2017, 6(2): 167–176.
- 29 张彦, 彭华. 基于深度自编码器的单样本人脸识别. 模式识别与人工智能, 2017, 30(4): 343–352.
- 30 李江, 冉君军, 张克非. 一种基于降噪自编码器的人脸表情识别方法. 计算机应用研究, 2016, 33(12): 3843–3846. [doi: [10.3969/j.issn.1001-3695.2016.12.072](https://doi.org/10.3969/j.issn.1001-3695.2016.12.072)]
- 31 冉鹏, 王灵, 李昕. 一种基于栈式降噪自编码器的人脸识别方法. 工业控制计算机, 2016, 29(9): 100–101, 104.
- 32 于万钧, 安改换, 鹿文静, 等. 基于栈式降噪自动编码器的气体识别. 计算机工程与设计, 2017, 38(3): 814–818, 836.
- 33 陈国定, 姚景新, 沈佳红. 基于堆栈式自编码器的尾矿库安全评价. 浙江工业大学学报, 2015, 43(3): 326–331.
- 34 李荣雨, 徐宏宇. 平行堆栈式自编码器及其在过程建模中的应用. 电子测量与仪器学报, 2017, 31(2): 264–271.
- 35 李远豪. 基于深度自编码器的人脸美丽吸引力预测研究 [硕士学位论文]. 江门: 五邑大学, 2014
- 36 于乃功, 默非凡. 基于深度自动编码器与 Q 学习的移动机器人路径规划方法. 北京工业大学学报, 2016, 42(5): 668–673. [doi: [10.11936/bjtxb2015100028](https://doi.org/10.11936/bjtxb2015100028)]
- 37 寇茜茜, 何希平. 基于栈式自编码器模型的汇率时间序列预测. 计算机应用与软件, 2017, 34(3): 218–221, 247.
- 38 王宪保, 何文秀, 王辛刚, 等. 基于堆叠降噪自动编码器的胶囊缺陷检测方法. 计算机科学, 2016, 43(2): 64–67. [doi: [10.11896/j.issn.1002-137X.2016.02.014](https://doi.org/10.11896/j.issn.1002-137X.2016.02.014)]
- 39 Sun WJ, Shao SY, Yan RQ. Induction motor fault diagnosis based on deep neural network of sparse auto-encoder. Journal of Mechanical Engineering, 2016, 52(9): 65–71. [doi: [10.3901/JME.2016.09.065](https://doi.org/10.3901/JME.2016.09.065)]