

一种改进的 5 层深度学习结构与优化方法

毛勇华^{1,2}, 代兆胜¹, 桂小林¹

(1. 西安交通大学 电子与信息工程学院, 西安 710049; 2. 西安工程大学 理学院, 西安 710048)

摘 要: 已有研究针对 5 层神经网络结构中各隐层节点数的选取并不明确。为此, 提出一种改进的 5 层深信度网络(DBN)结构设计方案与节点数量优化选择方法。将第一隐层、第二隐层节点数预估为前一层节点数的 1/3 至 2/3 间的某值, 第三隐层、第四隐层节点数分别等于第一隐层和输入层的节点数, 然后采用样条插值方法优化选择第一隐层、第二隐层节点数。该结构特征只需预训练前 2 层权重, 简化了 DBN 的受限玻尔兹曼机预训练方法。MNIST 数据集上的实验结果验证了该网络结构的高效性与高准确率。

关键词: 深信度网络; 预训练; 节点选择; 样条插值; 受限玻尔兹曼机

中文引用格式: 毛勇华, 代兆胜, 桂小林. 一种改进的 5 层深度学习结构与优化方法[J]. 计算机工程, 2018, 44(6): 147-150.

英文引用格式: MAO Yonghua, DAI Zhaosheng, GUI Xiaolin. An improved 5-layers depth learning structure and optimization method[J]. Computer Engineering, 2018, 44(6): 147-150.

An Improved 5-layers Depth Learning Structure and Optimization Method

MAO Yonghua^{1,2}, DAI Zhaosheng¹, GUI Xiaolin¹

(1. School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China;
2. School of Science, Xi'an Polytechnic University, Xi'an 710048, China)

[Abstract] According to previous studies, the selection of the number of hidden nodes in the 5-levels neural network structure is not clear. To solve the problem, an improved 5-levels Deep Belief Networks (DBN) structure design and optimization method of nodes number is proposed. The number of the first hidden layer and the second hidden layer nodes is estimated to be 1/3 to 2/3 between the number of the first layer nodes. The number of the third hidden layer and the fourth hidden layer nodes equals the first of the number of hidden layer and input layer nodes, and then the first hidden layer and the second hidden layer nodes number is optimized by spline interpolation method. The structure features only 2-layers of weight before pre-training, which simplifies the Restricted Boltzmann Machine (RBM) pre-training method of DBN. Experimental results on the MNIST dataset verify the efficiency and high accuracy of the network structure.

[Key words] Deep Belief Networks (DBN); pre-training; node selection; spline interpolation; Restricted Boltzmann Machine (RBM)

DOI: 10.19678/j.issn.1000-3428.0046691

0 概述

深信度网络 (Deep Belief Networks, DBN) 模型^[1-2]和贪婪层学习算法^[3]的提出, 使深度神经网络再度成为当前的研究热点。用受限玻尔兹曼机 (Restricted Boltzmann Machine, RBM)^[4-5]栈式无监督逐层预训练, 再运用反向传播法对 DBN 进行微调。RBM 预训练已经将 DBN 的权值训练到较准确

的初始状态, 反向传播法做进一步的有监督修正或微调, 从而解决直接训练深度神经网络时因层数增加而无法快速收敛到全局最优解的问题。深度学习^[6-7]是如今机器学习领域的新热点, DBN 则是深度学习的主要框架结构之一。

文献[8-9]中介绍的深度学习技术与 GPU 的混合应用在图像处理、语音识别等领域取得了较多成果^[10-11]。随着 ReLU、Dropout 等技术^[12-13]的出现, 在

基金项目: 国家自然科学基金 (61472316); 陕西省重大基础 Research 计划项目 (2016ZDJC-05); 中央高校基本科研业务费专项资金 (XKJC2014008); 陕西省重点研发项目 (2017ZDXM-GY-011)。

作者简介: 毛勇华 (1979—), 男, 讲师、博士研究生, 主研方向为深度学习、机器学习、分支预测; 代兆胜, 硕士研究生; 桂小林 (通信作者), 教授、博士、博士生导师。

收稿日期: 2017-04-07 **修回日期:** 2017-06-27 **E-mail:** xlgui@mail.xjtu.edu.cn

海量数据样本情况下,对于 5 层~20 层的深度学习网络,预训练不再是必须的。然而,对于标签训练数据集相对较少的应用问题,这种预训练的 5 层 DBN 依然可取,5 层网络结构在特征提取和训练算法复杂度方面都取得了较好的效果^[6]。在如何设定各隐层节点数量的问题上,有文章提出了通用指导思想,比如各隐层节点数量相同^[14]、隐层节点数大于第一层数据维数^[15]、隐层节点总数量与训练样本数量相关联等。目前主要根据这些经验和先验数据表现来选取每层节点数量,但各隐层节点数的选取方法并不明确。

基于上述问题,本文根据图像压缩思想,归纳一种对称隐层结构的节点数设定方式,并针对该对称结构,提出一种快速搜索方法,为深度学习应用提供参考。

1 算法设计

1.1 对称网络结构及其优化选择模型

根据文献[2]中图像数据压缩方法,参考深度神经网络最常用的 5 层结构,本文改进并提出一种 5+1 层网络结构设计方法:第一层节点数等于样本维数;第一、第二隐层是上一层节点数的 1/3~2/3,形成一个正金字塔结构;然后将金字塔结构倒过来,让塔尖对接,即将第一、第二隐层结构倒置过来,其权重 W 为第一、第二隐层 W 的转置;第四隐层节点数等于输入节点数。5 层 DBN 结构示意图如图 1 所示。

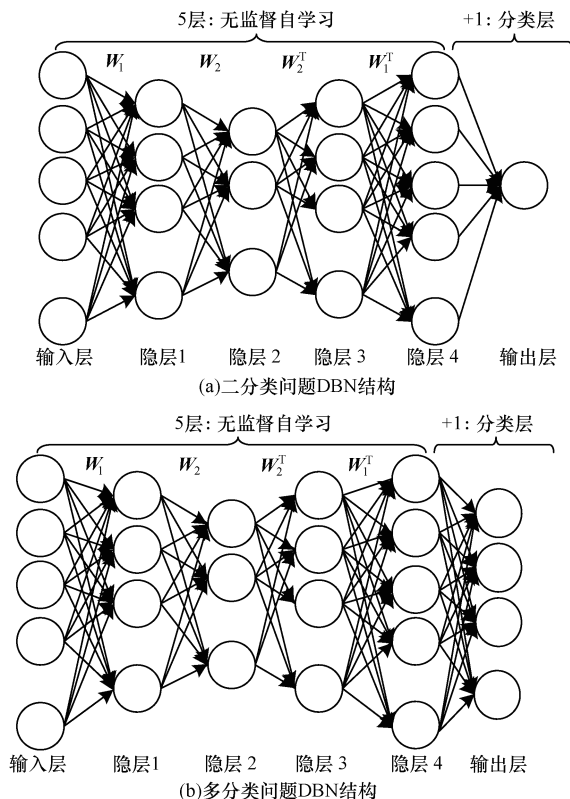


图 1 对称 5 层 DBN 结构示意图

只需无代入标签样本数据,第四隐层重建第一层输入数据,使输入数据经过网络变换后尽可能少地损失信息;输入层到第二隐层,逐层降维,强制丢失大部分信息;第二隐层到第四隐层,再从低维数据逐层重建原始数据,要求重建误差尽可能小,只有重建原数据所必需的特征信息才会在降维过程中保存下来,从而进行特征自动提取;最后一层为输出层,如果是二分类问题,则节点数为 1,如图 1(a)所示;如果是多分类问题,节点数等于类数,如图 1(b)所示。输入层和 4 个隐层为 5+1 层结构中的前 5 层,为无监督学习,代入无标签数据即可进行特征自动提取,最后一层为分类输出层。

上述对称结构符合特征自动提取原理。如果第 5 层等于输入层第一层,即通过调整系统中的参数,使输入数据经过系统变化后没有任何信息损失,使得任何一层都是原有信息的另外一种表示,也即自动地获得了输入数据的一系列层次特征。根据信息处理不等式,大部分的处理过程会逐层丢失信息。因此,在逐层降维与重建信息的处理过程中,丢失的是非特征噪声信息,无标签数据自身的特征信息会自动保留下来。

从预训练的角度来说,DBN 是由多个 RBM 叠加而成的深度前馈神经网络,如图 2 所示,由于本文所述网络结构的前 5 层是一种正反金字塔式的对称结构,因此只需采用 RBM 无监督逐层预训练 DBN 前 2 层的权值 W_1 、 W_2 即可,然后直接用 W_1 、 W_2 的转置作为权值,输出层的权值则直接随机生成。因此,网络结构中每层节点的数量只需优化第一层和第二层的神经元数量;第三层、第四层神经元数量分别与第二层、第一层对应相等;输出层神经元数量等于类数(如果是二分类问题,则可简化为一个神经元)。

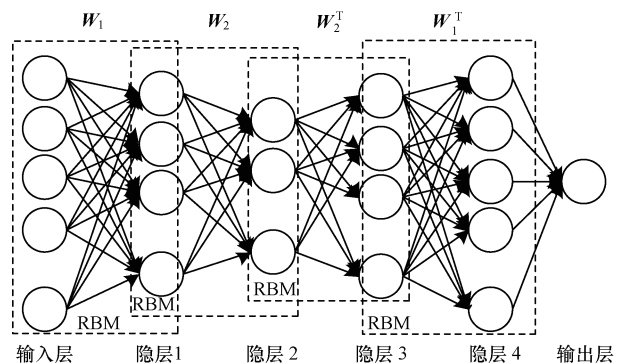


图 2 RBM 预训练 DBN 示意图

将第一隐层神经元数量初步设置为显式层神经元数量的 1/3~2/3 间的某值,依次平均选取 M 个数作为神经元数量,第二隐层神经元数量在第一隐层取定的情况下,分别按同样的方法各取 N 个数,共有 $M \times N$ 种组合。其他参数相同的情况下,对这 $M \times N$ 种网络结构进行训练,分别得到测试集上的错误率。

然后对第一层每个固定神经元做单独的样条插值:以第二层神经元数量为横轴、与其对应的错误率为纵轴,做自然样条插值,共得到 5 条插值曲线。观察所有曲线中的最小值,选出最值点对应的曲线,并估计出该曲线最小值点的位置,从而选出第一、第二层神经元的数量。

1.2 网络结构选择算法

本文模型设计中需要用到感知器 (Perceptron)、自编码器 (Auto-coder) 和反向传播法。RBM 模型及其快速对比散度学习方法理论参见文献[16]。传统 RBM 预训练和反向传播法都指定训练周期数,本文将改进为目标损失函数不再变小时停止训练,从而提高训练效率。

算法 1 随机小批量 RBM 快速对比散度方法

输入 训练数据集 X , 小批量样本数量 m , 隐层节点数 N

输出 RBM 的 W, a, b

1) 随机正态生成 W, a, b , 再乘以 0.001, 指定学习率 r 。

2) 将 X 随机分成 s 组, 每组 m 个数据的小样本集为 d_1, d_2, \dots, d_s , 损失率 $\text{err} = 0$ 。

3) 依次对第 i 组小样本集 $d_i, v_0 \leftarrow d_i, h_0 \leftarrow \text{sigmoid}(Wv_0 + b)$ 做 $\{0, 1\}$ 采样, $v_1 \leftarrow \text{sigmoid}(Wh_0 + a)$ 做 $\{0, 1\}$ 采样; $h_1 \leftarrow \text{sigmoid}(Wv_1 + b)$ 做 $\{0, 1\}$ 采样。

4) $\Delta w \leftarrow v_1' h_1 - v_0' h_0, \Delta b = v_1 - v_0, \Delta a = h_1 - h_0; \text{err} \leftarrow \text{累加 } v_1, v_0$ 残差范数。

5) $W \leftarrow W - r\Delta w, a \leftarrow a - r\Delta a, b \leftarrow b - r\Delta b$, 当 $i < s$ 时, 转至步骤 3); 当 $i = s$ 时, 若相邻 2 个周期的 err 不再变小, 则转至步骤 6), 否则转至步骤 2)。

6) 算法终止, 输出 W, a, b 。

算法 2 DBN 结构选择算法

输入 训练数据集 X , 输出分类数 c

输出 DBN 网络结构

1) 以 (X, c) 建立感知器, 直接采用标签数据训练感知器。

2) 平均选定第一隐层的 s 个神经元节点数 (N_1, N_2, \dots, N_s)。

3) 对第一层每个节点数 N_i , 平均选 1 个第二隐层节点数 ($M_{i1}, M_{i2}, \dots, M_{is}$)。

4) 逐一取第一隐层节点为 $H1 = N_i, i = 1, 2, \dots, s$ 。

(1) 逐一取第二隐层节点为 $H2 = M_{ij}, j = 1, 2, \dots, l$ 。

① 以 $(H1, H2)$ 为第一、第二隐层数量, 建立 DBN, 无监督训练 DBN。

② 以 $(X, H1, H2, H1, X, c)$ 建立 NN。

③ 以 DBN 中的 W_1, W_2 以及其转置 W_2^T, W_1^T 初始化 NN 的 W_1, W_2, W_3, W_4 , 以感知器的权重初始化 NN 输出层的权重。

④ 代入有标签训练集和验证集数据, 对 NN 进行微调。

⑤ 在测试集上测试 NN, 得到错误率 E_{ij} 。

(2) 以 $(M_{i1}, M_{i2}, \dots, M_{is})$ 为横坐标、 $(E_{i1}, E_{i2}, \dots, E_{is})$ 为纵坐标, 对错误率进行自然样本插值, 绘出样条曲线。

5) 从图中 s 个样条曲线中找出错率全局最小值点对应的 $(H1, H2)$, 其中, $H1$ 为最小值点样条曲线对应的第一层节点数, $H2$ 为该曲线对应的横坐标取值。

6) 以 $(H1, H2)$ 为参数建立对称 DBN, 即 $(X-H1-H2-H1-X-c)$, 输出此网络结构。

得到优化后的网络结构 $(X-H1-H2-H1-X-c)$ 后, 重复步骤①~步骤④训练 DBN, 然后在测试集上测试 NN, 作为最终的应用模型。

2 实验结果与分析

本文实验采用文献[17]中的一个手写数字图像数据库 MNIST, 其训练集数据库有 60 000 张, 测试集数据库有 10 000 张。从训练集数据库中随机选取 5 000 张作为验证集, 其余的 55 000 张作为训练集, 测试集数据库中的全部图像作为测试集。验证集只在训练期间度量训练效果使用, 最终结果以测试集上的识别错误率为评估标准。训练采用 mathwork 深度学习工具箱^[18], 采用随机小批量梯度下降法。将数据集随机分成指定大小的小批量组, 全部依次迭代训练一次, 该过程称为一个周期。完成一个周期后, 再重新随机分组, 进行下一个周期的训练, mathwork 深度学习工具箱对周期的处理方式为人工指定周期大小。对此, 本文做了简单的优化处理: 首先, 将周期设置为一个很大的数值, 每周期完成后, 计算验证集上的损失函数值, 并与上一周期进行比较, 如果损失函数在验证集上不再变小, 则终止训练。

由于 MNIST 数字为 $28 \times 28 = 784$ 的灰度图像, 因此对称的 DNN 网络结构的数据输入层为 784 个神经元。第一隐层在数据输入层的 $1/3 \sim 2/3$ 间取 5 个, 如表 1 中第 1 列所示, 对第一隐层每取定一个神经元节点数目后, 第二隐层按同样的比例取 5 个节点数, 即表 1 中的第 2 隐层对应的 5 个数据。

表 1 DNN 第一、第二隐层神经元节点数量

第一隐层	第二隐层				
260	80	105	130	155	180
330	110	140	165	195	220
395	130	165	200	235	270
465	150	190	230	270	310
530	170	220	265	315	360

MNIST 数据库为手写字形识别, 为十分类问题, 因此, DNN 输出层神经元节点数为 10, 表 2 所示为表 1 中选定了第一隐层、第二隐层节点数后的整个 DNN 网络结构 (每行取一个为例)。图 3 所示为网络结构优化搜索后再按自然样本插值的结果, 图中每条曲线代表第一隐层节点数量固定的情况下第二隐层节点数量与错误率间的变化关系。从图 3 可以看出, 节点数取 330 的曲线在 $(100, 150)$ 之间时, 错误率较小, 观察曲线变化趋势, 估选出 125 为第二隐层节点数, 即实验取 $(330, 125)$ 为第一、第二隐层节点数, 因此, 最终网络结构为 $(784-330-125-330-784-10)$ 。

表 2 第一,第二隐层节点数对应的网络结构

第一,第二隐层节点数	网络结构
260,80	(784-260-80-260-784-10)
330,110	(784-330-110-330-784-10)
395,130	(784-395-130-395-784-10)
465,150	(784-465-150-465-784-10)
530,170	(784-530-170-530-784-10)

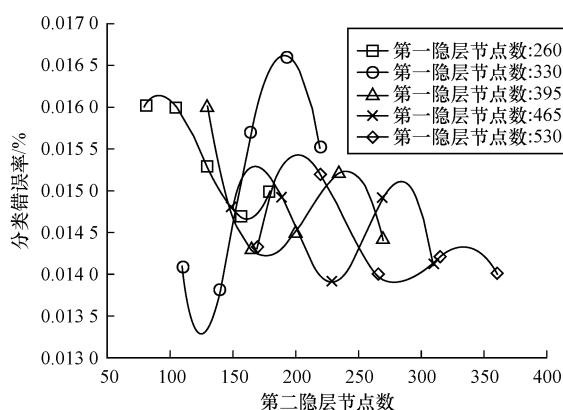


图 3 隐层节点数对分类错误率的影响

优化后的 5 层深度网络结构为 (784-330-125-330-784-10), 先采用 RBM 无监督预训练 DBN 的自学习层, 再监督预训练感知器输出层, 最后展开深度神经网络进行微调, 此时测试集错误率为 1.33%。(784-1 000-1 000-1 000-1 000-10) 为一种常见的网络结构, 隐层节点数大于数据维数, 各层节点数目相同。按类似的训练方式, 该网络结构的测试集错误率为 3.90%。2 种网络结构的性能对比如表 3 所示。

表 3 2 种网络结构性能对比

网络结构	周期数	权重参数数量	错误率/%
(784-1 000-1 000-1 000-1 000-10)	72	3.758e+6	3.90
(784-330-125-330-784-10)	46	5.836 6e+5	1.33

由表 3 可知, 在其他参数相同的条件下, 相对 (784-1 000-1 000-1 000-1 000-10) 网络结构, 本文 5 层对称网络结构微调周期小, 收敛速度快, 权重参数个数小一个数量级, 分类错误率低。

3 结束语

本文提出一种 5 层网络结构设计方法与基于样条插值的节点数量优化方法。对称的网络结构可以简化参数数量, 在提高训练效率的同时降低分类错误率。MNIST 数据集上的实验结果验证了该网络结构的高效性和高准确率。基于 RBM 和自编码器预训练的 5 层对称 DBN 结构适用于标签数据集较少的分类问题。下一步将优化网络结构, 调整其他技术参数, 以继续提升其分类准确率。

参考文献

- [1] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527-1554.
- [2] 陈宇, 郑德权, 赵铁军. 基于 Deep Belief Nets 的中文名实体关系抽取[J]. 软件学报, 2012, 23(10): 2572-2585.
- [3] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504-507.
- [4] HINTON G E. A practical guide to training restricted Boltzmann machines [M]//MONTAVON G, ORR G B, MÜLLER K R. Neural Networks: Tricks of the Trade. Berlin, Germany: Springer, 2012.
- [5] 刘建伟, 刘媛, 罗雄麟. 玻尔兹曼机研究进展[J]. 计算机研究与发展, 2014, 51(1): 1-16.
- [6] BENGIO Y. Learning deep architectures for AI [J]. Foundations and Trends® in Machine Learning, 2009, 2(1): 1-127.
- [7] 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.
- [8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [9] MAAS A L, HANNUN A Y, NG A Y. Rectifier nonlinearities improve neural network acoustic models [C]//Proceedings of the 30th International Conference on Machine Learning. Washington D. C., USA: IEEE Press, 2013: 15-19.
- [10] HINTON G E, DENG L, YU D, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups [J]. IEEE Signal Processing Magazine, 2012, 29(6): 82-97.
- [11] DAHL G E, YU D, DENG L, et al. Context-dependent pre-trained deep neural networks for large vocabulary speech recognition [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(1): 33-42.
- [12] JÜRGEN S. Deep learning in neural networks: an overview [J]. Neural Networks, 2015, 61: 85-117.
- [13] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553): 436-444.
- [14] BENGIO Y. Practical recommendations for gradient-based training of deep architectures [M]//MONTAVON G, ORR G B, MÜLLER K R. Neural Networks: Tricks of the Trade. Berlin, Germany: Springer, 2012.
- [15] SERMANET P, KAVUKCUOGLU K, CHINTALA S, et al. Pedestrian detection with unsupervised multi-stage feature learning [C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2013: 3626-3633.
- [16] WELLING M, HINTON G E. A new learning algorithm for mean field Boltzmann machines [C]//Proceedings of International Conference on Artificial Neural Networks. Berlin, Germany: Springer, 2002: 351-357.
- [17] LECUN Y, CORTES C. The mnist database of handwritten digits [EB/OL]. [2017-04-05]. https://www.iri.fr/~marc/Master2/MNIST_doc.pdf.
- [18] PALM R M. Prediction as a candidate for learning deep hierarchical models of data [EB/OL]. [2017-04-01]. <https://github.com/rasmusbergpalm/DeepLearnToolbox>.

编辑 吴云芳