

文章编号: 2096-1618(2017)03-0259-05

一种改进深度学习网络结构的英文字符识别

任波¹, 王录涛¹, 邓旭², 李孝杰¹

(1. 成都信息工程大学计算机学院, 四川 成都 610225; 2. 成都七中万达学校, 四川 成都 610100)

摘要:自 Geoffrey Hinton 于 2006 年在《Reducing the dimensionality of data with neural networks》一文中首次提出深度学习 (Deep Learning) 的概念, 深度学习就受到了研究人员的持续关注。深度学习利用多层的神经网络模拟人类大脑的多层抽象学习过程。其中网络结构设计和特征提取是数据挖掘和模式识别应用中的关键问题。而深度学习对手写体数字识别的准确率一直是衡量一个深度学习算法或网络结构优劣的重要标准。提出一种改进的深度学习网络结构, 通过对手写体英文数据库 Letter Recognition 的识别实验结果表明, 该深度学习网络结构的识别正确率相比传统的深度学习网络有了明显的提高。

关键词:深度学习; 网络结构; 手写体; Letter Recognition

中图分类号: TP183

文献标志码: A

doi: 10.16836/j.cnki.jcui.2017.03.005

0 引言

人具有识别不同的图像、语音、文字等事物的能力, 这种识别能力与人的生活息息相关, 称之为“模式识别”。模式识别在处理图像识别问题上的步骤如下, 先预处理图像, 从图像中提取出有利于分类的特征, 再将提取出的特征输入分类器, 利用有效的分类决策函数进行分类运算。特征提取和分类决策是模式识别中两个最重要的方面。学术界在研究模式识别时, 其重点也同样放在这两个方面, 而其中特征提取更是模式识别的重中之重。手写体识别 (handwritten recognition, HR) 是模式识别领域中的一个典型的应用。手写体识别分为联机手写体识别和脱机手写体识别, 联机手写体识别是通过硬件设备将手写体的字符实时输入计算机进行识别, 这种识别方式包含了笔画顺序等信息, 其识别准确率较高, 算法已较为成熟。而脱机手写体识别是通过光电设备将手写的字符扫描或拍照等方式传入计算机, 其缺少了笔画顺序等信息, 目前的识别准确率较差, 仍具有重要的研究意义。

目前, 针对脱机手写体识别的深度学习算法为通过 2 层稀疏自编码器进行学习提取原始图像信息的高层特征, 然后将提取出的特征作为分类器的输入进行分类识别。该方法在学习阶段使用无监督的学习策略, 得到的特征为重构的特征, 未包含分类的思想。提出一种新的更改网络结构的思路, 在深度学习的重构阶段融入分类的思想, 期望学习阶段提取出的特征包含部分分类信息, 从而提升手写体识别的准确率。并

利用手写体英文数据集 Letter Recognition 进行试验验证。

1 算法

1.1 深度学习

Geoffrey Hinton 教授首次提出了深度学习 (deep learning)^[1] 的概念以来, 深度学习技术在图像中的应用愈加广泛^[2], 深度学习可以通过学习一种深层非线性网络结构, 实现复杂函数逼近, 表征输入数据的分布式表达, 并展现出了从少数样本数据集中学习数据本质特征的能力^[3]。

典型的深度学习的网络结果如图 1 所示, 包含两个隐藏层。

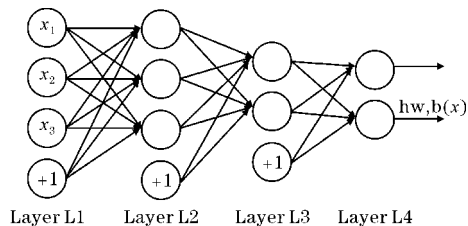


图1 典型深度网络结构模型

对深度学习网络中的符号给予说明:

圆圈表示网络中的神经单元, “+1”表示偏置节点, 最左一层 (列) 表示输入层, 最右一层表示输出层, 中间节点称为隐藏层。 n_l 表示网络层数, L_l 表示第 l 层, $W_{ij}^{(l)}$ 表示第 l 层第 j 单元与第 $l+1$ 层与第 i 单元的连接参数。 $z_i^{(l)}$ 表示第 l 层第 i 单元的加权和, $a_i^{(l)}$ 表示

第 l 层第 i 单元的激活值。

在深度学习中,第 $l+1$ 层第 i 单元的激活值由公式(1)、(2)计算:

$$z^{(l+1)} = W^{(l)} a^{(l)} + b^{(l)} \quad (1)$$

$$\alpha^{(l+1)} = f(z^{(l+1)}) \quad (2)$$

式(2)中 $f(\cdot)$ 为激活函数,一般选用式(3)的形式:

$$f(z) = \frac{1}{1 + \exp(-z)} \quad (3)$$

函数的曲线图像如图2所示,取值范围为 $[0,1]$ 。

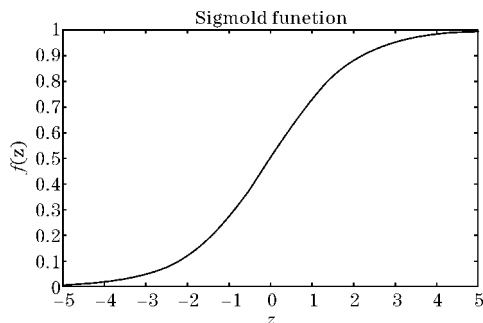


图2 sigmoid 函数曲线图

1.2 稀疏自编码器

自编码算法是一种“无监督”的学习算法,通过对无标签的数据的学习,期望学习得到输出等于输入的函数: $h_{w,b}(x) \approx x$ 。如包含 m 个样本的样本集 $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$, 自编码算法针对单个样本的代价函数为

$$J(W, b; x, y) = \frac{1}{2} \|h_{w,b}(x) - y\|^2 \quad (4)$$

式中 $J(W, b; x, y)$ 表示单个样本的代价函数。根据自编码算法的要求,公式中的 $y=x$ 。

对于包含 m 个样本的输入,其整体代价函数可根据单个样本的代价函数在各个样本集上作算术平均得出:

$$\begin{aligned} J(W, b) &= \left[\frac{1}{m} \sum_{i=1}^m J(W, b; x^{(i)}, y^{(i)}) \right] \\ &= \left[\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \|h_{w,b}(x)^{(i)} - y^{(i)}\|^2 \right) \right] \end{aligned} \quad (5)$$

式(5)中, $J(W, b)$ 为整个样本的代价函数,其中 $x^{(i)} = y^{(i)}$ 。

为防止过度拟合,通常在式(5)加上权重衰减项。

$$\begin{aligned} J(W, b) &= \left[\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \|h_{w,b}(x)^{(i)} - y^{(i)}\|^2 \right) \right] + \\ &\quad \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \end{aligned} \quad (6)$$

该自编码网络虽然输出等于输入,但可以从无标签的数据中提取数据的内在特征^[4]。举例说明,在图

像识别中,灰度图像为 10×10 共 100 个像素点,对应神经网络输入层应有 100 个节点。同样,输出数据也应为 100 维,对应网络输出层包含 100 个节点。如果设置隐藏层包含 50 个节点。要达到输出等于输入,则要求该网络的输出能够从 50 维的隐藏层节点中重构出 100 维的数据。由于输入的数据为有意义的图像,因此输入数据是彼此相关联的,则该算法能够很好的发现这些输入数据(像素点)之间的相关性。这些相关性,即为用于图像识别的特征。

更一般的,不要求隐藏层的数据维度一定小于输入维度,可以通过在自编码神经网络中添加限制因素达到发现输入数据之间相关性的目的——称之为稀疏性限制。

针对实验,神经元的激活函数使用 sigmoid 函数,当神经元的输出接近 1 的时候称为激活状态,神经元的输出接近 0 的时候称为抑制状态。当神经网络中的神经元大部分时间处于抑制状态的限制称为稀疏性限制。这里以图 1 中第 2 层的稀疏限制为例,神经元 j 的平均激活度为(在训练集上取平均):

$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^m [\alpha_j^{(2)}(x^{(i)})]$$

进一步限制:

$$\hat{\rho}_j = \rho$$

ρ 为稀疏性参数,通常设置为一个接近于 0 的数,以限制神经元的激活度接近于 0。

为在深度学习网络中实现该限制,在优化目标函数中引入一个额外的惩罚因子,用于惩罚那些神经元的激活值与 ρ 相差较大的神经元。对于惩罚因子的选择有多种形式,本试验中选择如下^[5]:

$$\sum_{j=1}^{s_2} \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j}$$

式中 s_2 表示隐藏层(第 2 层)的神经元个数, j 依次表示该层的每一个神经元。该函数具有以下性质,当 $\hat{\rho}_j$ 趋近于 ρ 的时候,其值趋近于 0, $\hat{\rho}_j$ 与 ρ 相差越大,其值越大。如 $\rho = 0.2$, 则该函数与 ρ 的关系见图 3。

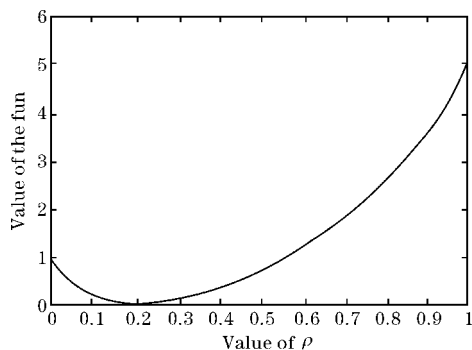


图3 惩罚函数值与 ρ 取值曲线

1.3 softmax 分类器

深度学习提取出的特征,将输入 softmax 分类器^[6]进行分类。

Softmax 分类器是 logistic 回归模型在多分类问题上的推广,属于有监督学习范畴。Softmax 分类器的输出可以包含两个以上的值。针对实验使用的 Letter Recognition 手写体英文字符集,类标签分别对应字符 a-b 的 26 个类。

在 softmax 分类器中,仍然使用同 logistic 回归相同的假设函数(hypothesis function):

$$h_{\theta}(x) = \frac{1}{1 + \exp(-\theta^T x)}$$

式中, θ 为模型参数, x 为输入。在 softmax 分类器中,通过训练模型参数 θ , 达到最小化代价函数的目的,其代价函数如下:

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_{\theta}(x^{(i)})) \right]$$

对于英文字母的识别,包含 m 个样本的数据集 $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$, 其中类别标签 $y^{(i)} \in \{1, 2, \dots, k\}$, $k=26$ 。由于输出包含 26 个标签,假设针对每个输入 x , 其输出应有 26 个值,即假设函数 $h_{\theta}(x^{(i)})$ 有下面的形式:

$$h_{\theta}(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 | x^{(i)}; \theta) \\ p(y^{(i)} = 2 | x^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = k | x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix}$$

式中 $\theta_1, \theta_2, \dots, \theta_k \in R^{n+1}$ 为模型参数。通过对已标记数据的训练,可以优化整个网络的参数,然后利用已优化的参数对未标记的图像进行分类。

1.4 反向传播算法

深度学习利用梯度下降算法,通过调整深度学习网络中的参数去最小化代价函数。

梯度下降法过程中,参数 W, b 在每次迭代中按照如下公式更新:

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b)$$

$$b_i^{(l)} = b_i^{(l)} - \alpha \frac{\partial}{\partial b_i^{(l)}} J(W, b)$$

式中 α 为算法的学习速率,即步长。

由于在深度学习中,整体代价函数 $J(W, b)$ 和单个样例的代价函数 $J(W, b; x, y)$ 存在式(6)所示的关系,可以从单个样例的的代价函数的偏导数计算整体代价函数的偏导数:

$$\frac{\partial}{\partial w_{ij}^{(l)}} J(W, b) = \left[\frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b; x^{(i)}, y^{(i)}) \right] + \lambda W_{ij}^{(l)} \quad (7)$$

$$\frac{\partial}{\partial b_i^{(l)}} J(W, b) = \frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial b_i^{(l)}} J(W, b; x^{(i)}, y^{(i)}) \quad (8)$$

因此,首先求解单个样例的偏导数。单个样例的偏导数由反向传播算法(back propagation, BP)^[7]求解。反向传播算法的步骤如下:

步骤1 利用前向传导式(1)、(2),计算 L_2, L_3, \dots, L_{nl} 各层的激活值;

步骤2 求出第 n_l 层,即输出层的每个单元的 i 的“残差”

$$\delta_i^{(n_l)} = \frac{\partial}{\partial z_i^{(n_l)}} \frac{1}{2} \|y - h_{W, b}(x)\|^2 = -(y_i - a_i^{(n_l)}) \cdot f'(z_i^{(n_l)})$$

步骤3 利用公式(19)依次求 $l = n_l - 1, n_l - 2, \dots, 2$ 层的第 i 个节点的残差

$$\delta_i^{(l)} = \left(\sum_{j=1}^{sl+1} W_{ji}^{(l)} \delta_j^{(l+1)} \right) f'(z_i^{(l)})$$

步骤4 计算各层偏导数

$$\frac{\partial}{\partial W_{ij}^{(l)}} J(W, b; x, y) = \alpha_j^{(l)} \delta_i^{(l+1)}$$

$$\frac{\partial}{\partial b_i^{(l)}} J(W, b; x, y) = \delta_i^{(l+1)}$$

计算得出了单个样例的偏导数之后,由式(7)和式(8)可以计算出整体代价函数的偏导数。另外,BP 算法不适用于层数过多的深度学习^[8]。

2 改进深度学习的网络结构

2.1 提出一个新的网络

稀疏自编码器的学习过程中,令输出等于输入,通过学习,隐藏层得到输入数据的特征,整个过程是无监督的学习。虽然得到的特征可以作为分类器的输入进行分类训练,但是该特征未明确包含分类的信息。考虑到字符识别的主要目的是分类,希望在特征提取阶段即包含分类的思想。文献[9]指出深度学习的网络结构可以很好的表示复杂函数。好的结构模型在深度学习中至关重要。于是在典型的深度学习网络结构模

型(如图 1)上进行改进,在第二个隐藏层上添加 Softmax 分类器,如图 4 所示。

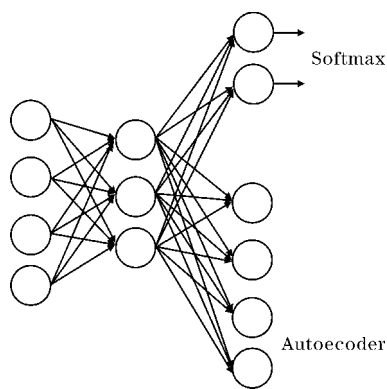


图 4 改进的网络结构模型

在每一次迭代中,同时进行 Softmax 和 Autoencoder 训练。然后将该部分提取出的特征作为 softmax 分类器的输入进行分类训练。

改进的深度学习网络,由于包含 Softmax 有监督训练,一次提取出的特征,同时包含了有监督的分类思想,提取出的特征将更加有利于分类。

改进后的学习网络(improvement autoencoder,简称 IAE)代价函数为

$$J_{IAE} = J_{(W,b)} + \gamma J(\theta) \tag{9}$$

公式中 $J_{(W,b)}$ 和 $J(\theta)$ 分别对应 AE 网络和 softmax 分类器的代价函数; γ 为权重,控制两者的相对关系。

第二个隐藏层残差为 AE 网络部分和 Softmax 分类器部分梯度误差的加权和:

$$\delta^2 = [(\frac{\rho}{1-\rho})^T \delta^{(3)} + \beta(-\frac{\rho}{1-\rho} + \frac{1-\rho}{1-\rho}) + \gamma \theta^T(I-P)] \cdot f'(z^{(2)}) \tag{10}$$

改进后网络的第一层特征提取使用 AE 网络,分类决策选择 softmax 分类器。

2.2 实验参数设计

为验证改进的网络结构能够有效增加对实验数据的识别准确率,数据集为 Letter Recognition 手写体英文字母,包含 1200 个训练样本和 240 个测试样本。输入层节点个数为 784 个,第一个隐藏层节点个数为 200 个,第二个隐藏层层节点个数为 200 个。使用正态分布 $Normal(0, \varepsilon^2)$, 其中 $\varepsilon = 0.01$ 进行参数初始化。激活函数使用 S 型函数,优化算法使用梯度下降算法。训练过程使用逐层贪婪训练方法^[10]。

2.3 实验过程

首先随机初始化参数,然后调节各参数使得改进前的网络结构模型对 Letter Recognition 的识别准确率达

到理想值。再利用改进后的网络结构,通过调剂式(9)和式(10)中的 γ 值,得到多组实验结果,从实验结果中找出最符合条件的一组,将该 γ 值保留,在 Letter Recognition 上进行迭代次数不同的实验,记录结果,对比分析(试验中记录的数据均为 10 实验结果的平均值)。

2.4 结果分析

(1)改进的网络结构对 Letter Recognition 数据进行训练及测试,得到的在测试集上的结果总结如下。迭代次数 100。

表 1 γ 与识别准确率的关系

γ	分类准确率/%
0.001	84.475
0.01	84.376
0.1	84.401
0	84.107
1	84.550
2	84.150
3	83.751
4	83.714
5	83.400
6	84.017
7	81.914
8	83.177
9	83.701
10	83.918
20	84.773
30	85.114
40	83.890
50	83.617
60	84.451
70	84.701
80	85.105
90	82.910
100	83.711

将试验的数据用曲线图表示如图 5 所示。

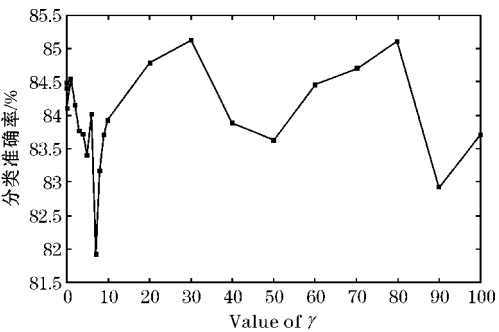


图 5 γ 与识别准确率关系的曲线表示

从图 5 可以方便地看出,当 γ 取值为 30 时,识别

准确率在所有实验的参数数据中最高。

(2)改进后的网络(IAE)与传统网络(AE)对 Letter Recognition 数据集识别准确率的对比(γ 取值为 30,其他参数同实验 1)

表 2 AE 网络与 IAE 网络的识别准确率的关系

迭代次数	100	200	400
AE	84.107	90.527	92.117
IAE	85.114	92.275	93.517

从对比数据中发现,改进后的网络结构,即在训练阶段加入 softmax 分类的思想可以达到提升对英文字符的识别准确率的期望。

3 结束语

文中提出一种改进的深度学习网络结构,在对 Letter Recognition 数据集的分类识别问题中,相对于传统的深度学习网络在对数据集的识别准确率方面有明显提高。由于手写体英文字符本身的差异很大,以及识别率还有待进一步提升,下一步的工作重点在于进一步的提高对该数据集的识别率以及将工作放在实际应用中。

参考文献:

[1] HintonG E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science,2006,313:504.

[2] Dan C, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification[J]. 2012,157(10):3642-3649.

[3] 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.

[4] Cheriyyadat A M. Unsupervised feature learning for aerial scene classification[J]. IEEE Transactions on Geoscience & Remote Sensing,2014,52(1):439-451.

[5] Zhu H. On information and sufficiency[J]. Working Papers,1997,157(1):1-7.

[6] Friedman J, Jerome T, Hastie R. Regularization Paths for Generalized Linear Models Via Coordinate Descent[J]. Journal of Statistical Software, 2010,33(1):1.

[7] The10 Breakthrough Technologies 2013[N]. MIT Technology Review,2013-04-23.

[8] Erhan D, Bengio Y, Couville A, et al. Why does unsupervised pre-training help deep learning[J]. Journal of Machine Learning Research, 2010, 11(3):625-660.

[9] Bengio Y, Lecun Y. Scaling learning algorithms towards AI[C]. Large-Scale Kernel Machines. 2007: 321-358.

[10] 杜敏,赵全友.基于动态权值集成的手写数字识别方法[J]. 计算机工程与应用,2010,46(27):182-184.

An Improved Deep Learning Network Structure for English Character Recognition

REN Bo¹, WANG Lu-tao¹, DENG Xu², LI Xiao-jie¹

(1. College of Computer Sciences, Chengdu University of Information Technology, Chengdu 610225, China;2. Chengdu No. 7 Wanda High School, Chengdu 610100, China)

Abstract: Since Geoffrey Hinton proposed the concept of "Deep Learning" in a paper, "Reducing the dimensionality of data with neural networks", in 2006, in the first time. Deep learning has received sustained attention from researchers. Deep learning use multi-layer neural network to simulation of the multi-layer abstract learning process of human brains. The design of network structure and feature extraction are the key problems in data mining and applications of pattern recognition. The accuracy of handwritten numeral recognition in deep learning has always been an important criterion for measuring the deep learning algorithm or network structure. In this paper, an improved network of deep learning structure is proposed. The experimental results show that the recognition rate of the deep learning network structure is significantly higher than the traditional network structure deep learning.

Keywords: deep learning; network structure; handwriting; letter recognition