

# 基于深度学习的盲文识别方法

李 婷

(河海大学计算机与信息学院, 江苏 南京 211100)

**摘要:** 提出一种基于深度学习的盲文点字识别方法, 利用深度模型——堆叠去噪自动编码器( Stack Denoising AutoEncoder, SDAE) 解决盲文识别中特征的自动提取与降维等问题。在构建深度模型过程中, 采用非监督贪婪逐层训练算法( Greedy Layer-Wise Unsupervised Learning Algorithm) 初始化网络权重, 使用反向传播算法优化网络参数。利用 SDAE 自动学习盲文点字图片特征, 使用 Softmax 分类器进行识别。实验结果表明, 本文所提方法较之传统方法, 可以有效解决样本特征的自动学习与特征降维等问题, 操作更为简易, 并能获得满意的识别结果。

**关键词:** 盲文识别; 深度学习; 特征提取; SDAE; 神经网络

**中图分类号:** TP391      **文献标识码:** A      **doi:** 10.3969/j.issn.1006-2475.2015.06.008

## A Deep Learning Method for Braille Recognition

LI Ting

(College of Computer and Information, Hohai University, Nanjing 211100, China)

**Abstract:** This paper mainly proposes a deep learning method, using Stacked Denoising AutoEncoder (SDAE) to solve the problems of automatic feature extraction and dimension reduction in Braille recognition. In the construction of a network with deep architecture, a feature extractor was trained with unsupervised greedy layer-wise training algorithm to initialize the weights for extracting features from Braille images, and then a following classifier was set up for recognition. The experimental results show that comparing to traditional methods, the constructed network based on the deep learning method can easily recognize Braille images with satisfied performance. The deep learning model can effectively solve the Braille recognition problem in automatic feature extraction and dimension reduction with a reduced preprocessing.

**Key words:** Braille recognition; deep learning; feature extraction; SDAE; neural network

## 0 引言

视障人士是社会中的弱势群体, 按目前美国、日本等发达国家中受教育的盲人人口不到其盲人总数的10%来计算, 我国视力残疾人中至少有800万盲人不能顺利阅读盲文读物, 这对他们的信息获取、融入社会极为不利<sup>[1]</sup>。与此同时, 盲文图书馆书籍的珍贵古旧文献备份、保存等工作, 是由视力正常的工作人员完成。其中, 对文献内容进行备份的操作过程是由懂得盲文的工作人员一边通过触摸阅读盲文内容, 一边翻译成文字并手动输入电脑。由于图书馆文献数目巨大, 文献备份对工作人员来说是一项艰巨的任务<sup>[2-4]</sup>。由此可见, 从盲文到文字的转换在生活中有着实用意义。

本课题主要针对目前盲文识别系统识别率不高

和图片预处理较为复杂等问题进行研究。利用深度模型堆叠去噪编码器<sup>[5]</sup>自动、全面学习样本深层次特征, 避免人为手工选取特征存在的多种弊端, 并用习得的特征作为传统神经网络的输入, 更大程度地避免传统神经网络由于随机选取初值而导致结果陷入局部极值的问题。

## 1 深度模型简介

### 1.1 自动编码器

假设一个未带类别标签的训练样本集合  $X$  ( $X \in x \in [0, 1]^d$ ) 通过变换公式(1)将输入  $x$  变换为激活值  $y$  ( $y \in [0, 1]^d$ )。  $x$  与  $y$  满足公式(1):

$$y = f_{\theta}(x) + s(Wx + b) \quad (1)$$

其中  $s$  是一个非线性函数, 例如 sigmoid 函数;  $\theta = \{W, b\}$  为前向映射过程中的参数集合, 其中  $W$  为  $m$

收稿日期: 2015-03-10

作者简介: 李婷(1989-), 女, 广东广州人, 河海大学计算机与信息学院硕士研究生, 研究方向: 神经网络、模式识别。

$\times n$  的权重矩阵  $b$  为  $m \times 1$  的列向量。然后通过变换公式 (2) 将激活值  $y$  反向变换为对原始输入  $x$  的一种重构表示  $z$ ,  $y$  与  $z$  满足公式 (2):

$$z = f_{\theta'}(y) + s(W'y + b') \quad (2)$$

其中  $\theta' = \{W', b'\}$  为反向映射过程中的参数集合, 其中  $W'$  为  $n \times m$  的权重矩阵  $b'$  为矩阵  $n \times 1$  的列向量。

这里, 公式 (1) 和 (2) 中的 2 个参数  $W'$  和  $W^T$  被限制满足关系  $W' = W^T$  ( $W^T$  表示对  $W$  进行转置), 最后通过不断地修改  $\theta$  和  $\theta'$  使得平均重构误差  $L$  达到最小化。 $L$  定义为:

$$L = \|x - z\|^2 \quad (3)$$

直观上讲, 如果激活值  $y$  能够对它的原始输入  $x$  进行良好重构, 那么认为它保留了原始数据所具有的大部分信息<sup>[6-7]</sup>。

## 1.2 去噪自动编码器

如果只是简单地保留原始输入数据  $x$  的信息, 并不足以让自动编码器<sup>[8]</sup> 习得一种有用的特征表示。为了让它能够学习到一个更复杂的非线性函数, 需要给予自动编码器一定的约束使它学习到一种更好的特征表示。去噪自动编码器 (Denoising AutoEncoder, DAE)<sup>[8-9]</sup> 所要实现的功能是对叠加噪声的原始数据进行学习, 而它学到的特征与从未叠加噪声的数据学到的特征几乎一样<sup>[10]</sup>。但去噪编码器从叠加噪声的输入中习得的特征更具鲁棒性, 同时由去噪编码器构成的网络具有了去噪功能<sup>[11]</sup>。此时对输入  $x$  叠加噪声  $N$  使其变成含噪声输入  $x'$ 。具体算法如下:

1) 对原始数据  $x$  叠加噪声  $N$  变成  $x'$ 。

$$x' \sim qD(x|N) \quad (4)$$

2) 和自动编码器一样通过公式 (5) 将  $x'$  变换为激活值  $y$ ,

$$y = f_{\theta}(x') + s(Wx' + b) \quad (5)$$

3) 通过公式 (6) 将  $y$  重构为数据  $z$ 。

$$z = g_{\theta'}(y) + s(W'y + b') \quad (6)$$

4) 通过训练参数  $\theta = \{W, b\}$  和  $\theta' = \{W', b'\}$  来最小化平均重构误差  $L$ <sup>[12]</sup>。

## 1.3 堆叠去噪自动编码器

去噪自动编码器的堆叠可以形成堆叠去噪自动编码器。利用叠加噪声输入训练每层网络的去噪能力, 这样训练出来的每一层编码器可以作为一个具有容错性能的特征提取器<sup>[13-14]</sup>。

堆叠去噪编码器网络本身不具有分类功能, 只是一个特征提取器, 所以要实现分类功能还需要在网络最后添加分类器, 如多层感知机 (MLP)、Softmax 分类器等。

假设带有 Softmax 分类器的 SDAE 网络共有  $n$

层, 网络训练步骤如下<sup>[5, 15]</sup>:

1) 使用叠加噪声输入数据作为输入, 利用训练去噪编码器方法训练出第一个隐含层结构的网络参数, 并用训练好的参数算出第 1 个隐含层的输出。

2) 将步骤 1) 的输出叠加噪声作为后一个网络的输入, 用同样的方法训练后一个隐含层网络的参数。如此下去直至  $n-1$  层。

3) 使用预训练完成的堆叠去噪自动编码器网络重新学习未叠加噪声数据特征, 将此特征和标签作为 Softmax 分类器的输入, 训练分类器的模型参数。

4) 用步骤 1)、步骤 2) 和步骤 3) 的网络参数作为整个深度网络 ( $n-1$  个隐含层, 1 个 Softmax 分类层) 参数初始化的值, 其后使用 L-BFGS<sup>[16]</sup> 算法迭代求出模型损失函数取最小值时网络的参数值, 此时得到的参数值即为最优参数值。

## 2 基于深度学习的盲文点字识别

### 2.1 盲文样本采集

因为没有现成的盲文数据库, 所以本文在盲文照片中截取盲文获取样本, 自行构造数据集进行实验。

在样本的采集过程中, 考虑了几种将会影响识别结果的因素, 比如: 光照不足、倾斜、点阵偏移、点阵部分缺失或者是多余噪声点, 如图 1 所示。在 63 种盲文中选择了 10 种用做实验, 如图 2 所示, 每一类测试的盲文包括叠加噪声的训练样本和未叠加噪声的训练样本各 4000 个, 测试样本 1000 个。实验中, 样本 RGB 图像转换成灰度图像, 统一缩放成  $40 \times 24$  的矩阵, 再将其转换成  $960 \times 1$  的列向量, 构造出  $40000 \times 960$  的未叠加噪声训练集  $D1$ ,  $40000 \times 960$  的叠加噪声训练集  $D1'$  和  $10000 \times 960$  的测试集  $T1$ 。

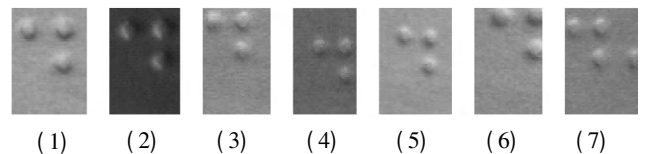


图 1 带有噪声的样本示例

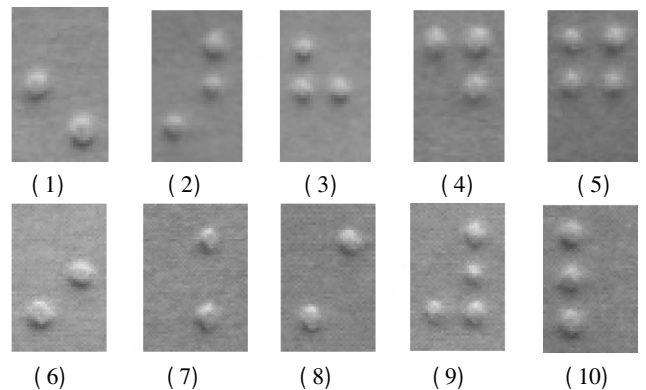


图 2 参加测试的 10 类盲文

## 2.2 SDAE 的训练

SDAE 模型虽以 MNIST 数字(阿拉伯数字手写字体)识别任务作为样例,获得不错的识别效果,且目前已经开源。但源码网络结构、各项参数等,均是以特定任务图像的特点为研究基础设定的,而且将 SDAE 引入盲文识别的具体应用,还未有前人做过此项尝试。为了使 SDAE 在盲文识别中应用,也能够自动学习盲文图片良好特征,获得较好的识别结果,还需要重新通过实验确定模型内部参数。

实验中,根据经验,将 SDAE 设置为 2 层,第一层有 200 个节点,第二层有 50 个节点,使用公式(7) sig-mod 函数作为激活函数:

$$f(x) = (1 + \exp(-x))^{-1} \quad (7)$$

在这一步的自编码过程中,为了迫使神经元可以学习到有用的特征并对高维数据有效降维,稀疏度参数  $\rho$  和稀疏惩罚项  $\beta$  这 2 项参数对不同类型样本以及特征敏感,起到了至关重要的作用。其中  $\rho$  的作用是控制隐层神经元的激活概率,迫使每个神经元只对某类数据有响应,通常是一个接近于 0 的较小的值; $\beta$  则是保证稀疏性的要求。设有  $m$  个无标签训练样本集  $\{x_1, x_2, \dots, x_m\}$  在模型隐层中,  $\rho$  为目标稀疏度,  $\hat{\rho}_j$  为训练样本对隐层第  $j$  个神经元的平均稀疏度,有:

$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^m [a_j(x_i)] \quad (8)$$

其中  $a_j(x_i)$  表示单个样本  $x_i$  对第  $j$  个神经元的激活量。令  $\rho = \hat{\rho}_j$  并引入惩罚因子,使之成立。惩罚因子为:

$$\sum_{j=1}^{s_2} \left( \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \right) \quad (9)$$

其中  $s_2$  为隐层神经元数。这一惩罚因子有如下性质:当  $\hat{\rho}_j = \rho$  时,惩罚因子为 0;当  $\hat{\rho}_j \neq \rho$  时,随着  $|\hat{\rho}_j - \rho|$  的值增大而激增。同时,引入稀疏惩罚项  $\beta$  ( $\beta > 1$ ) 放大惩罚因子对模型的影响。最后,得到关于稀疏性的代价函数  $J_s$ ,如式(10):

$$J_s = \beta \sum_{j=1}^{s_2} \left( \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \right) \quad (10)$$

$J_s$  会在自编码的反向传播中,与误差方差代价函数共同改变网络参数,使每层重构的表达在尽可能接近输入的同时保持稀疏特性。

为了找到适合盲文识别的参数  $\rho$  与  $\beta$ ,本文进行了一系列的组合实验,择优选取。由于  $\rho$  与  $\beta$  只对样本类型敏感,对样本数量敏感性低,所以在兼顾实验准确性与提高训练效率的前提下,在训练集 D1 与 D1' 中的 10 类样本各随机选择了 1000 组数据进行训练,并在测试集 T1 中随机选取 100 组进行测试。不

同的参数组合,测试结果见表 1。

表 1 模型参数选取识别率实验结果

单位: %

$\rho \backslash \beta$	1	1.5	2	2.5	3	3.5	4
0.06	70.2	74.5	77.4	76	74.9	72	69
0.08	83	83.3	84	81.8	75	74.4	73.8
0.10	86.1	86.2	87.1	88	86	85.7	83
0.12	87	88.8	89	91.4	90.2	87	86.3
0.14	84.8	87.7	87.9	89	90.3	88.2	88
0.16	83	86.5	87.1	88	90.4	90.1	88.6

源码中,针对 MNIST 数字识别的参数设置  $\rho = 0.1$ ,  $\beta = 3$ 。由表 1 可见,此时识别率仅为 86%,可知当 SDAE 应用于盲文识别时,并不是最优的参数组合。根据对比择优,本文最终选择的组合为  $\rho = 0.12$ ,  $\beta = 2.5$ 。

得到针对盲文识别的参数  $\rho$  与  $\beta$  后,将叠加噪声训练集 D1' 作为输入,利用 1.3 节中的算法,利用无监督的学习方法,逐层训练 SDAE。SDAE 每个神经元都类似于一个特征提取算子,随机选择第一层 4 个神经元将其网络权重  $W$  可视化(如图 3 所示),第一层神经元可以大致定位凸点所在的位置。

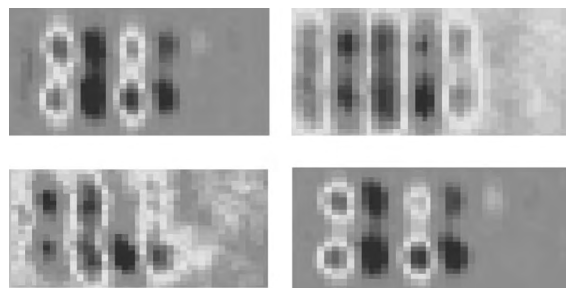


图 3 SDAE 的权重可视化

## 2.3 训练 SDAE 与分类器

为了实现模型的分类功能,需要在 SDAE 的最后添加一个分类器。将未叠加噪声训练集 D1 与数据集对应的标签输入 SDAE 和分类器中,使用有监督的学习方法训练整个网络,然后再用 BP 算法微调网络参数以增强识别效果,最后使用测试集 T1 测试网络的识别率。当选用 Softmax 作为分类器时,用 BP 算法微调前,该模型识别率为 80%;用 BP 算法微调后,该模型识别率增加至 94%。由表 2 可知,当使用其他分类器时,识别率在微调后均有一定上升。由此可见,微调网络参数对识别率有着重要的、积极的影响。

表 2 微调前后分类效果对比

分类器类型	MLP	RBF	Softmax
微调前识别率/%	73	69	80
微调后识别率/%	86	80	94

## 2.4 实验结果与对比

本文实验基于 MATLAB 平台实现,运行环境是配备 3.00 GHz CPU,4 G 内存的服务器,操作系统是 Windows 7。表 3 中,采用多层感知机 (Multi-layer Perceptron, MLP)、径向基函数 (Radial Basis Function, RBF) 和 Softmax 3 种分类器,在使用传统方法,提取了图像的七阶不变矩和灰度共生矩阵特征作为特征向量后,进行分类的对比结果;此外在同样的分类器上,本文使用 SDAE 自动学习获得的特征进行分类测试。值得注意的是,盲文样本在输入 SDAE 之前,除了进行灰度变换之外,没有经过更多的图像处理步骤。

表 3 3 种分类器性能对比

单位: %

分类器类型	MLP	RBF	Softmax
传统特征	64	55	65
SDAE 特征	86	80	94

## 3 结束语

本文探讨了自动编码器包括稀疏自动编码器<sup>[17-18]</sup>、去噪编码器以及堆叠去噪编码器在内的几种变体,并将堆叠去噪编码器应用在盲文识别中。实验证明,与传统神经网络和机器学习办法相比,堆叠去噪编码器在简化图像预处理的条件下,减少人为经验主观因素的干扰,自动提取样本特征,并对盲文样本中的各类噪声有一定的鲁棒性,提高分类性能。

此外,本文还存在诸多不足,如未能深入研究网络层数、神经元数对特征学习的影响,只考虑了单面盲文的识别情况。笔者今后将对这些问题进行更深入的研究。

### 参考文献:

- [1] 张贵建. 盲人高等中医教育——中诊教程的改革与思考 [EB/OL]. <http://www.zgmx.org.cn/ViewInfo.asp?id=1958>, 2007-04-02.
- [2] Zhang Shanjun, Yoshino Kazuyoshi. A braille recognition system by the mobile phone with embedded camera [C]// 2007 the 2nd International Conference Innovation Computing, Information and Control (ICICIC). Japan, 2007: 223.
- [3] 尹佳, 李杰, 王丽荣. 盲文自动识别方法研究 [J]. 长春大学学报, 2010, 20(8): 54-56.
- [4] 李念峰, 董迎红, 肖志国. 基于图像处理的盲文自动识别系统研究 [J]. 制造业自动化, 2012, 34(2): 63-67.
- [5] Poirson P, Idrees H. Multimodal Stacked Denoising Autoencoders [EB/OL]. [http://crev.ucf.edu/REU/reu2013/patrick\\_poirson/report.pdf](http://crev.ucf.edu/REU/reu2013/patrick_poirson/report.pdf), 2014-11-05.
- [6] Baldi P. Autoencoders, unsupervised learning, and deep architectures [J]. JMLR: Workshop and Conference Proceedings, 2012, 27: 37-50.
- [7] Glüge S, Böck R, Wendemuth A. Auto-Encoder Pre-Training of Segmented-Memory recurrent neural networks [C]// Proceedings of the 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2013). 2013: 29-34.
- [8] Lu Xugang, Tsao Y, Matsuda S, et al. Speech enhancement based on deep denoising autoencoder [C]// The 14th Annual Conference of the International Speech Communication Association. 2013: 436-440.
- [9] Vincent P. A connection between score matching and denoising autoencoders [J]. Neural Computation, 2011, 23(7): 1661-1674.
- [10] Chen Minmin, Xu Zhixiang, Weinberger K, et al. Marginalized denoising autoencoders for domain adaptation [C]// Proceedings of the 29th International Conference on Machine Learning. 2012: 1-8.
- [11] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, et al. Extracting and composing robust features with denoising autoencoders [C]// Proceedings of the 25th International Conference (ICML 2008) on Machine Learning. 2008: 1096-1103.
- [12] Lu Xugang, Tsao Y, Matsuda S, et al. Ensemble modeling of denoising autoencoder for speech spectrum restoration [C]// Proc. Interspeech 2014. 2014, 14: 885-889.
- [13] Wu Daoxi, Pan Wei, Xie Lidang, et al. An adaptive stacked denoising auto-encoder architecture for human action recognition [J]. Applied Mechanics and Materials. 2014, 631-632: 403-409.
- [14] Sang Ruoxin, Jin Peiquan, Wan Shouhong. Discriminative feature learning for action recognition using a stacked denoising autoencoder [J]. Intelligent Data Analysis and its Applications, 2014, 297: 521-531.
- [15] Vincent P, Larochelle H, Lajoie I, et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion [J]. Journal of Machine Learning Research, 2010, 11: 3371-3408.
- [16] Zhu Ciyu, Byrd R H, Lu Peihuang, et al. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization [J]. ACM Transactions on Mathematical Software, 1997, 23(4): 550-560.
- [17] Deng L, Seltzer M, Yu Dong, et al. Binary coding of speech spectrograms using a deep auto-encoder [C]// Interspeech 2010. 2010: 1692-1695.
- [18] Lange S, Riedmiller M. Deep auto-encoder neural networks in reinforcement learning [C]// International Joint Conference on Neural Networks. 2010: 1-8.