

DBN 融合模型对脱机手写汉字识别

刘 露, 孙巍巍, 丁 博

(哈尔滨理工大学 计算机科学与技术学院, 黑龙江 哈尔滨 150080)

摘 要: 针对脱机手写汉字识别问题, 提出一种新的分类器级联识别模型。新模型将修正的二次判别函数(modified quadratic discriminant function, MQDF)与深度置信网络(deep belief network, DBN)相融合, 利用MQDF先进行识别并得出结果, 同时计算一个该识别结果的可信度, 通过这个可信度对识别结果进行判别, 若可信度符合要求, 则MQDF的识别结果可作为最终结果直接输出, 否则再与DBN结合进行二次识别, 得到最终的识别结果。实验结果表明, 在ETL-9B手写汉字数据集上进行的脱机手写汉字识别任务中, 使用MQDF与DBN融合模型, 可以取得比单独使用MQDF和DBN更好的准确率。

关键词: 脱机手写字; 二次判别函数; 深度置信网; 汉字识别

DOI: 10.15938/j.jhust.2017.06.016

中图分类号: TP391.412 文献标志码: A 文章编号: 1007-2683(2017)06-0082-05

DBN Fusion Model for Offline Handwritten Chinese Characters Recognition

LIU Lu, SUN Wei-wei, DING Bo

(School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China)

Abstract: The requirement of the recognition result is also increasing in practical applications. In this paper, a new classifier cascade recognition model is proposed for the problem of offline handwritten Chinese character recognition. New model is the fusion of modified quadratic discriminant function (MQDF) and deep belief network (DBN). First to recognize and get result using MQDF, and calculate the reliability of the recognition result. If the reliability can meet the requirement, MQDF recognition result can be as the final result directly output. Otherwise using the DBN to make recognition again and getting the final recognition result. Experiments show that the MQDF and DBN fusion model proposed in this paper can achieve better accuracy than the single use of MQDF and DBN in the offline handwritten Chinese character recognition task, which is performed on the ETL-9B handwritten Chinese character dataset.

Keywords: offline handwritten character; quadratic discriminant function; deep belief network; Chinese character recognition

收稿日期: 2016-03-22

基金项目: 黑龙江省自然科学基金(F201208).

作者简介: 刘 露(1972—), 男, 博士, 教授, E-mail: 445817068@qq.com;

孙巍巍(1991—), 女, 硕士研究生;

丁 博(1983—), 女, 副教授, 硕士研究生导师.

0 引言

文字识别作为模式识别中的一个重要领域,一直受到研究人员的广泛关注。对于手写数字,手写英文字符已经取得了较好的识别结果,但是在手写汉字的识别任务当中仍然存在着不小的挑战。由于汉字在中国传统文化发展中的重要地位,对于汉字识别的研究自然具有非常重要的意义和价值^[1]。最早的汉字识别研究始于 IBM 的 Casey 等对印刷汉字的识别。到了 20 世纪 80 年代,手写汉字识别成为研究的重点并发展至今,许多方法相继被提出,并且取得了不错的效果。手写汉字识别主要分为联机和脱机两种,联机手写汉字通过手机或者其它电子设备人工输入手写字,因此可直接得到汉字的笔画顺序、书写方向以及提笔、落笔等信息,更容易获得汉字的结构。而脱机手写汉字是通过扫描技术将手写字扫描入计算机,得到扫描图像,再对图像进行识别的过程,其特点是数据量大、结构复杂,不同的人具有不同的书写风格,并且在不同的环境中手写字会不同程度的发生形变,所以脱机手写字识别难度远远大于联机手写汉字。

目前,对于手写汉字识别,研究人员提出了许多种方法,并且都取得了不错的识别效果。文 [2] 提出了一种代价敏感的修正二次判别函数 (CMQDF) 分类器,主要针对手写中文地址进行识别;文 [3] 提出了级联修正二次判别函数分类器来进行手写汉字识别,其是一种基于串行结构的分类器集成算法,在度量的层次上将分类结果进行结合,但是对于手写图像的扭曲变形等复杂情况识别效果不是很理想。文 [4] 利用 MQDF 与支持向量机 (support vector machine, SVM) 结合的方式对相似手写汉字进行识别,该方法相比之前的方式无需提取额外特征,并利用分类混淆矩阵自动得到相似汉字特征集。文 [5] 引入深度学习技术,将卷积神经网络 (convolutional neural network, CNN) 融入到手写汉字识别中,并与 MQDF 相结合,取得了不错的识别结果。文 [6] 提出自适应判别局部对齐方法 (adaptive discriminative locality alignment, ADLA),很大程度上改进了不同群体间的最优参数选择问题,进而提升了对相似汉字识别的准确率。文 [7] 在当今大数据情况下,采用深度神经网络 (DNN) 对相似手写汉字进行识别,对大样本取得了好的识别效果。本文针对脱机手写汉字识别问题,提出了一种新的融合模型,该模型将

MQDF 和 DBN 相融合,通过定义可信度来协调两个分类器在识别任务中的具体分工情况,从而达到理想的识别效果。通过实验证明,所提出的融合模型在识别任务中,取得了比两种分类器模型单独使用时更好的识别准确率。

1 修正的二次判别函数

MQDF 是对二次判别函数 (quadratic discriminant function, QDF) 进行 K - L 变换和平滑部分特征值改进得到的^[8],是一个贝叶斯分类器。其具体形式如下:

$$h_i(x) = \frac{1}{\sigma^2} \left\{ \|x - \mu_i\|^2 - \sum_{j=1}^k \left(1 - \frac{\sigma^2}{\lambda_j}\right) [x - \mu_i]^T p_{ij} \right\} + \sum_{j=1}^k \ln \lambda_j + (n - k) \ln \sigma^2 \quad (1)$$

式中: x 为输入的特征向量; n 是特征维数; μ_i 代表第 i 类的均值; p_{ij} 是 i 类协方差矩阵的第 j 个特征向量; 而 λ_j 则是特征向量所对应的特征值。 k 是截断参数,意味着 k 个最大的特征值没有正规化,其余小的特征值被 σ^2 所替代。在 MQDF 中,假设每个类别的先验概率都服从高斯分布,则 $\min_{x \in i} h_i(x)$ 。

MQDF 基于统计方法,通过用常量代替小的特征值,补偿了有限样本的估计误差。其次,由于只需要输入特征向量进行计算,简化了复杂度。由于结构简单、便于多类问题的分类、易于实现以及准确率高等优点,被广泛应用于手写文字识别中。但是,在对于有些稍微复杂的手写字图像时,识别能力仍非常有限。

2 深度置信网

深度置信网 (deep belief network, DBN) 是 Hinton 等在 2006 年提出的一个具有层次特征的概率生成模型^[9],通过训练神经元之间的权重,可以让整个神经网络按照最大概率来生成训练数据。同时 DBN 是一种深层神经网络,采用自底向上的传递,底层的神经元接收原始的特征向量,不断向更高层次抽象,顶层的神经网络形成易于组合的特征向量。通过增加层次就能够使特征向量更高的抽象化,而且,每一层的网络会弱化上一层的错误信息和次要信息,以确保深层网络的精度。DBN 结构如图 1 所示。其是由一系列叠加的受限玻尔兹曼机 (restricted boltzmann machine, RBM)^[10] 和顶层的反向传播

(back propagation, BP) 网络^[1]构成。自提出以来受到了越来越多的关注,并陆续应用于计算机视觉、语音识别、自然语言处理等领域,有着广阔的发展和应用空间。

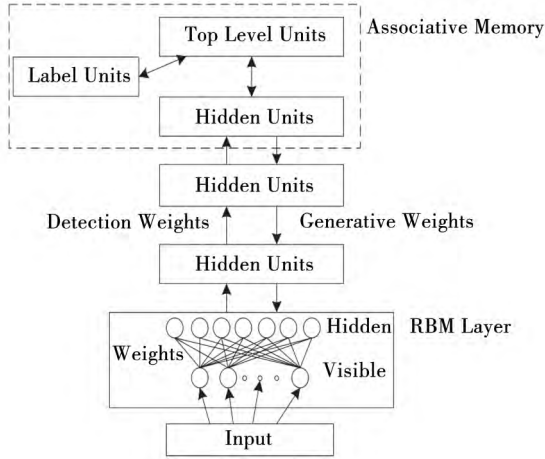


图1 DBN 网络结构

DBN 的训练包含预训练和微调两个步骤。首先,预训练阶段利用大量无类标信息数据,无监督地训练每层 RBM,将下层 RBM 的隐层输出作为上一层 RBM 可见层的输入。微调阶段则采用有监督学习方式对顶层的 BP 网络进行训练,将实际输出与预期输出的误差逐层反向传播,目的是调整网络的权值。DBN 这样做克服了 BP 网络因随机初始化权值而导致的局部最优问题。

2.1 受限玻尔兹曼机介绍

受限玻尔兹曼机(RBM)由隐层 v 和可见层 h 两种神经元节点构成。可见层的作用是输入数据,隐层的作用是特征检测。区别于玻尔兹曼机,RBM 的特点是可见层和隐层之间采用全连接的方式,而在可见层内神经元之间和隐层内各神经元之间是无连接的,以二分图的形式存在,这样相比玻尔兹曼机更为高效。

RBM 是一种基于能量的模型。假设可见层与隐层的神经元数目为 I, J, v_i 表示第 i 个可见层神经元状态, h_j 表示第 j 个隐层神经元状态,那么特定一组 (v, h) 所构成 RBM 的能量函数可以表示为

$$E(v, h | \theta) = - \sum_{i=1}^I a_i v_i - \sum_{j=1}^J b_j h_j - \sum_{i=1}^I \sum_{j=1}^J w_{ij} v_i h_j \quad (2)$$

式(2)中, $\theta = \{w_{ij}, a_i, b_j\}$ 是 RBM 的参数, w_{ij} 表示 v_i 节点和 h_j 节点间的连接权值, a_i 和 b_j 分别表示 v_i 和 h_j 的偏置值。通过能量函数可以得到 (v, h) 的联合概率分布:

$$p(v, h | \theta) = e^{-E(v, h | \theta)} / Z(\theta) \quad (3)$$

其中 $Z(\theta)$ 为归一项。实际问题中最为关心的是 $p(v, h | o)$ 的边缘分布 $p(v | o)$, 即观测变量的分布。

RBM 一般通过迭代的方式训练,目的在于得到 $\theta = \{w_{ij}, a_i, b_j\}$ 参数值,以拟合训练数据。在训练数据集(样本数 T)上求极大对数似然函数:

$$\theta^* = \arg_{\theta} \max L(\theta) = \arg_{\theta} \max \sum_{t=1}^T \ln p(v^{(t)} | \theta) \quad (4)$$

对于各参数值的更新,采用 Hinton 提出的对比散度 (Contrastive Divergence) 算法^[2],具体规则如下:

$$\left. \begin{aligned} \Delta w_{ij} &= \varepsilon (\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{recon}) \\ \Delta a_i &= \varepsilon (\langle v_i \rangle_{data} - \langle v_i \rangle_{recon}) \\ \Delta b_j &= \varepsilon (\langle h_j \rangle_{data} - \langle h_j \rangle_{recon}) \end{aligned} \right\} \quad (5)$$

其中 ε 为预训练学习率, $\langle \cdot \rangle_{data}$ 为训练数据集定义分布上的数学期望, $\langle \cdot \rangle_{recon}$ 为重构后模型所定义分布上的数学期望。

2.2 BP 网络

BP 网络主要是对 RBM 预训练阶段所得特征向量分类,依据分类结果微调 DBN 每层参数。训练 BP 网络包括前向和后向两个过程。将特征向量向前传播,由输出层得到分类结果,将该分类结果与预期结果相比较得到误差值^[3]。后向传播是将所得误差值逐层回传,从而调整 DBN 参数。传播过程中每层需计算一个灵敏度 δ ,然后将 δ 向下传递调整网络权值。

输出层中,第 i 各节点输出为 o_i ,预期输出为 d_i ,则计算 δ 值可以表示为

$$\delta_i = o_i (1 - o_i) (d_i - o_i) \quad (6)$$

对于第 m 个隐层,计算 δ 的表达式为

$$\delta_i^l = y_i^l (1 - y_i^l) \sum_j w_{ij}^l \delta_j^{l+1} \quad (7)$$

在得到各层敏感度值之后,对 DBN 的更新如下。

$$\left. \begin{aligned} w_{ij}^l &= w_{ij}^l + \varepsilon_{fine-tuning} \times y_i^l \delta_j^{l+1} \\ b_j^l &= b_j^l + \varepsilon_{fine-tuning} \times \delta_j^{l+1} \end{aligned} \right\} \quad (8)$$

3 MQDF 与 DBN 融合模型

修正的二次判别函数 MQDF 应用于脱机手写汉字识别任务中,通常可以识别出大部分汉字图像,但对一些较复杂汉字图像识别率有所下降。针对这类汉字图像,使用深度模型 DBN 进行再次识别,即为本文提出的 MQDF-DBN 融合模型。

MQDF-DBN 整体分成训练、识别和融合三部分,如图2所示。先将汉字图像经过预处理,归一化到统一分辨率,此外还包括去噪、二值化等相关操作,再利用训练数据同时对 MQDF 和 DBN 同时训

练。在训练阶段, MQDF 提取梯度特征, DBN 按照之前所介绍方法进行训练。

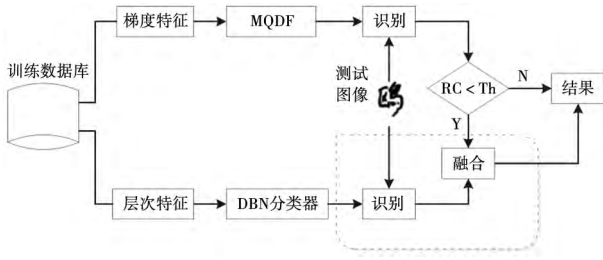


图2 MQDF与DBN融合模型

识别阶段, MQDF 输出结果要经过识别可信度 (recognition confidence, RC) 来进行评估。MQDF 输出前 P 个可能识别差由 (小到大排列)。若 d_1, d_2 ($d_1 < d_2$) 为前两个识别差, 则 RC 的定义如下

$$RC = 1 - d_1/d_2 \quad (9)$$

当 d_1 与 d_2 很接近时 (RC 趋近于零), 就很容易出现错分的情况。因此, 一个样本经 MQDF 分类, 结果的 RC 值越大, 说明识别越准确。所以规定一个门限值 T , 如果 MQDF 输出结果的 RC 大于 T , 就可将该识别结果作为最终识别结果输出, 否则, 样本被输入到 DBN 中进行融合识别。

在进行融合识别之前, 首先要将两个分类器的识别得分规整到同一个可比较的域中。用 w_i 代表第 i 类, x_g 和 x_h 表示待识别汉字图像分别对应 MQDF 和 DBN 的梯度特征和分层特征。条件概率 $p(x_g|w_i)$ 和 MQDF 的识别差 $h_i(x_g)$ 满足如下公式:

$$p(x_g|w_i) \propto e^{-h_i(x_g)/2} \quad (10)$$

根据式 (10), MQDF 中的后验概率 $p(w_i|x_g)$ 可以通过 [4] 中方法, 在前 P 个可能类别上计算得到。DBN 可以输出类标和对应的概率, 所以可以在前 P 个可能类别上的正规化, 重新计算输出层每个神经元的后验概率 $p(w_i|x_h)$ 。

根据 $p(w_i|x_g)$ 和 $p(w_i|x_h)$, 通过特定的融合规则就可以得到 MQDF-DBN 融合模型概率 $p(w_i|x_g, x_h)$ 。具体方法采用线性可信度累积 (Linear confidence accumulation, LCA) 算法 [5], 通过对 $p(w_i|x_g)$ 和 $p(w_i|x_h)$ 线性加权求和得到:

$$p(w_i|x_g, x_h) = \alpha p(w_i|x_g) + \beta p(w_i|x_h) \quad (11)$$

式中 α 和 β 为权衡因子, 可以根据 MQDF 和 DBN 各自的表现来指定。那么融合模型对当前汉字图像的识别结果则为:

$$w_i = \arg \max_{i=1 \dots Q} p(w_i|x_g, x_h) \quad (12)$$

由于 α 和 β 的值不影响 (12) 的结果, 因此可以

假设 $\alpha + \beta = 1$ 。

综上, 本文提出的 MQDF-DBN 融合模型对汉字图像的最终识别结果表示如下

$$w(x) = \begin{cases} w_{MQDF} & RC > T \\ \arg \max_{i=1 \dots Q} p(w_i|x_g, x_h) & \text{其它} \end{cases} \quad (13)$$

其中 w_{MQDF} 为 MQDF 的最佳候选结果。

4 实验与结果分析

为了验证本文提出方法, 选择 ETL-9B 手写汉字数据集进行实验。ETL-9B 数据集由日本电子技术实验室建立的, 其中包括 120 万个数字以及日文、中文、拉丁文的印刷体和手写体样本, 其中汉字类别 2965 个, 每个类别包含 160 或 200 个样本, 部分样本如图 3 所示。实验选用 1500 个类别, 每个类别包含 200 个样本, 前 20 个和后 20 个用来测试, 剩余 160 个用来训练。DBN 选择 Theano 深度学习工具包实现。实验环境为 Intel i5 3.2 GHz 处理器, 8 GB 内存。



图3 ETL-9B 的部分样本

在实验过程中, 将汉字图像线性归一化到 64×64 , 训练 MQDF 时, 提取 588 维的梯度特征。深度置信网设置为两层 RBM 结构, 中间隐层点数 350 个, 迭代次数设置为 10。实验整体分成两部分。首先通过实验确定最优的阈值 T 和加权系数 α ; 然后把识别率作为评价标准, 将 DBN 的识别效果与不同截断参数 k 下的 MQDF 和 MQDF-DBN 的识别效果进行比较。

在 ETL-9B 数据集中, 应用实验所选的汉字图像样本, 在不同的阈值 T 和加权系数 α 情况下对 MQDF-DBN 模型进行实验, 记录两个值不同组合情况下的识别准确率信息。设置 T 值分为 0.08, 0.10, 0.12, 0.15, 0.20, 0.30, 1.00, 然后在 α 值由 0.1 到 0.9 时记录识别准确率, 实验结果如图 4 所

示。可以看出 MQDF-DBN 融合模型在 T 值和 α 值不同时,识别效果有明显不同,当 T 值在 0.20 跟 0.30 之间, α 落在 $[0.4, 0.6]$ 区间中时识别效果最好。

表 1 给出了在实验所用数据集下 DBN 分类器的识别准确率,以及截断参数不同时,单独的 MQDF 分类器,和融合分类器 MQDF-DBN 的识别表现。从实验结果可以看出,随着截断参数 k 值的增加, MQDF 和 MQDF-DBN 的识别准确率都有所提高,而后者要明显高于前者,同样 MQDF-DBN 的识别准确率也高于 DBN 单独使用时的识别准确率。

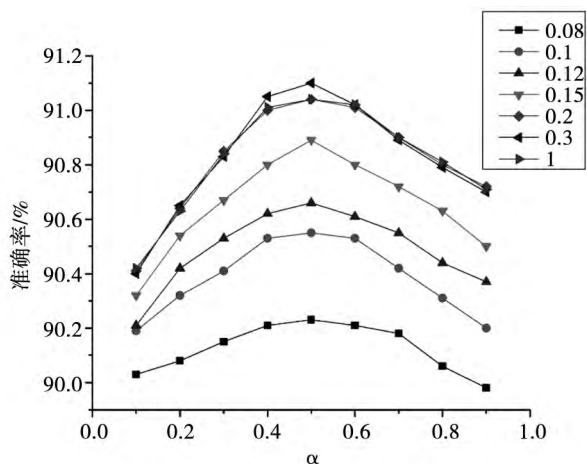


图 4 T 和 α 不同时的识别情况

表 1 不同截断参数下的识别结果

k	5	10	15	20	30	40
MQDF	95.91	96.30	96.39	96.42	96.50	96.56
MQDF + DBN	96.10	97.39	97.54	97.58	97.62	97.64
DBN	96.90	-	-	-	-	-

5 结 语

本文针对脱机手写汉字识别问题,提出了一种修正二次判别函数和深度置信网融合模型,为该方面研究提供了新的思路。二次修正判别函数分类器识别较简单的汉字图像,而较为复杂的汉字图像则交给深度模型 DBN 来处理识别,并且在得出最后结果时结合了 MQDF 的中间识别结果。最后的实验表明,两种模型结合使用能很好地形成互补,取得了不错的识别效果。在接下来的工作中,我们会将工作重心放在实际应用价值更大的相似手写汉字识别

研究当中,探索并改进 MQDF-DBN 模型,力求取得新的突破。

参 考 文 献:

- [1] 丁晓青. 汉字识别研究的回顾 [J]. 电子学报, 2002, 30(9): 1364-1368.
- [2] LU S, WEI X, LU Y. Cost-sensitive MQDF Classifier for Handwritten Chinese Address Recognition. ICDAR [C] // 2015 13th International Conference on. IEEE, 2015: 76-80.
- [3] 付强, 丁晓青, 刘长松. 用于手写汉字识别的级联 MQDF 分类器 [J]. 清华大学学报(自然科学版), 2008, 48(10): 1609-1612.
- [4] 姚超, 卢朝阳, 李静, 等. 一种分类器级联的手写相似汉字识别方法 [J]. 西安电子科技大学学报, 2015, 42(5): 26-32.
- [5] WANG Y, LI X, LIU C, et al. An MQDF-CNN Hybrid Model for Offline Handwritten Chinese Character Recognition. Frontiers in Handwriting Recognition. (ICFHR) [C] // 2014 14th International Conference on. IEEE, 2014: 246-249.
- [6] QU X, XU N, WANG W, et al. Similar Handwritten Chinese Character Recognition Based on Adaptive Discriminative Locality Alignment. Machine Vision Applications (MVA) [C] // 2015 14th IAPR International Conference on. IEEE, 2015: 130-133.
- [7] 杨钊, 陶大鹏, 张树业, 等. 大数据下的基于深度神经网络的相似汉字识别 [J]. 通信学报, 2014, 35(9): 184-189.
- [8] 周明辉, 刘辉, 曹刚. 基于 MQDF 的车牌字符识别 [J]. 信息技术, 2013(7): 121-123.
- [9] HINTON G. E., OSINDERO, S., Teh, Y. A Fast Learning Algorithm for Deep Belief Nets [J]. Neural Computation, 2006(18): 1527-1554.
- [10] SALAKHUTDINOV R, MNH A, HINTON G. Restricted Boltzmann Machines for Collaborative Filtering [C] // Proceedings of the 24th international conference on Machine learning. ACM, 2007: 791-798.
- [11] 王洋, 曾雪琴, 范剑英. 汽车牌照字符识别系统设计 [J]. 哈尔滨理工大学学报, 2012, 17(1): 90-95.
- [12] HINTON G E. Training products of experts by minimizing contrastive divergence [J]. Neural Computation, 2002, 14(8): 1771-1800.
- [13] 何春香, 刘泊. 小波与 BP 神经网络在零件识别应用中的研究 [J]. 哈尔滨理工大学学报, 2008, 13(5): 50-53.
- [14] LIU C. L., MASAKI N. Precise Candidate Selection for Large Character Set Recognition by Confidence Evaluation [C] // IEEE Trans. on Pattern Analysis and Machine Intelligence, 2000: 636-642.
- [15] HUANG Y. S., SUEN C. Y. A Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals [J]. IEEE Trans Pattern Recognition and Machine Intelligence, 1995(17): 90-94.

(编辑: 关 毅)