

基于无监督特征学习的手势识别方法

陶美平,马 力,黄文静,吴雨隆

(西安邮电大学 数字艺术学院, 陕西 西安 710061)

摘 要: 针对静态手势图像的分类识别问题,提出了一种将无监督的特征学习和有监督的分类识别相结合的静态手势图像识别方法,通过无监督的稀疏自编码神经网络训练图像小块提取手势图像的边缘特征,并将此边缘特征作为训练分类器的输入,最后提出对训练好的分类器的参数进行调优从而提高准确率。

关键词: 无监督的特征学习;稀疏自编码神经网络;边缘特征;调优

中图分类号: TP18

文献标识码: A

文章编号: 1000—7180(2016)01—0100—04

A Gesture Recognition Research Based on Unsupervised Feature Learning

TAO Mei-ping, MA Li, HUANG Wen-jing, WU Yu-long

(School of Digital Art, Xi'an University of Posts and Telecommunications, Xi'an 710061, China)

Abstract: In order to solve static gesture image classification problems, the paper presents a unsupervised learning process combining the static supervised classification. According to the unsupervised sparse auto-encoder neural network, the image patches were trained to extract the edge feature of the image, and these edge features are the input of a classifier, and finally. And the classifier parameter was used to improve the classification accuracy.

Key words: unsupervised learning; sparse auto-encoder neural network; edge feature; tuning

1 引言

近年来,随着机器学习的发展,许多无监督特征学习的方法被研究者提出,如受限玻尔兹曼机(RBM)^[1]、稀疏自编码器^[2]、深度信念网络^[3]、卷积神经网络(CNN)^[4]等。无监督特征学习^[5]不仅可以在学习后得到输入数据的良好表示,而且可以挖掘输入数据之间的相关性,发现数据中隐藏的结构,无监督学习的目的就是学习比原始输入数据更适合作为有监督学习的输入。而许多无监督学习算法的思路都是在隐层神经元输出表达中施加一些限制性因素,如稀疏性^[6]。通过这个稀疏性的表示^[7]可以让输出层重构出原始的输入数据。这样训练得到的稀疏表示不仅可以代表原始的输入数据,还是原始的输入数据的一个低维表示的矢量,把这些低维的矢量特征作为有监督分类器的输入进行分类识别。因

此,本文提出了一种将无监督的特征提取过程和有监督的分类识别过程相结合的静态手势识别方法。

2 手势识别流程

针对手势图像的分类和识别应用,本文所提出的基于无监督特征学习方法结合有监督分类识别流程如图 1 所示。

2.1 数据集及数据预处理

如表 1 所示,无标签图片集用于无监督的稀疏自编码深度神经网络的特征学习,学习输入数据间存在的某种关联,有标签训练集用于训练分类器,而有标签的测试集则用于测试该深度神经网络的图像分类的准确率,这些有标签的训练集和测试集均是采用 Simple Sign Language 手势数据库得到的,该数据库总共包含 10 个手势,分别对应着 pinch, scale, buckle, hold, grab, rotate, crawl, ok, cut,

shoot. 每个手势分别是两个人在不同光照,不同背景下完成的. 每个手势共有 500 张图片,总共 10 000 张图片,对于每个手势选取 250 张作为训练数据,另

外 250 张作为测试数据.,故训练样本集和测试样本集各自有 5 000 个数据.

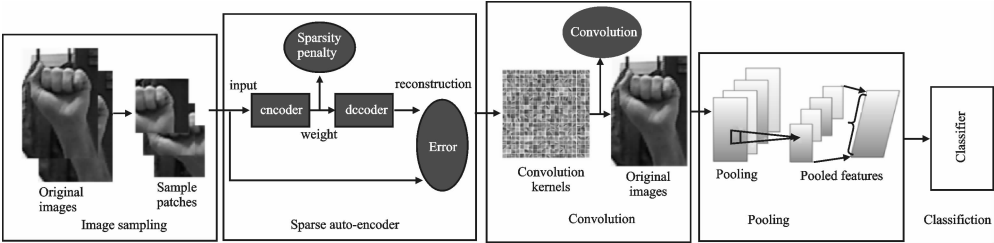


图 1 有监督和无监督相结合的静态手势识别流程图

由于给定的图片大小不统一,所以只能取个大概的中间尺寸值,将图片的尺度归一化到 32×32 大小,然后随机在每张无标签图片中采样 1 000 个 10×10 的图像小块,再利用 ZCA 白化技术对这些图像小块进行预处理,以在尽可能保持原始数据所有维度不变的情况下减少数据之间的冗余性.

表 1 三个数据集

数据集	数量
无标签图片集	100 000
有标签训练集	5 000
有标签测试集	5 000

2.2 特征学习算法

自编码神经网络是深度学习中的一种无监督学习算法,它尝试着去学习一个恒等函数,这个恒等函数能使输入等于输出,而稀疏自编码神经网络就是对自编码神经网络的隐藏单元施加一些限制,从而找出输入数据之间隐藏着的某种关联结构,如果输入数据的个数大于神经网络的隐藏层神经单元的个数,则使自编码神经网络学习输入数据的压缩表达;相反,若输入数据的个数小于或等于隐藏层神经单元的个数,这时候就对隐藏层神经单元加入稀疏性限制,使得隐藏层大部分神经元处于抑制状态,只有极少数能够表示输入数据之间关联的那些神经元处于激活状态,这样自编码神经网络就能寻找出输入数据之间的相关性,如果神经网络的输入数据是完全随机的,那么要学习这些数据内的某种关联性表示则是十分困难的,但如果这些输入数据之间隐含着某些有联系的结构,则自编码算法就能够寻找出这些数据之间的相关性,并且能够在输出层重构出输入数据.

假设神经元的激活函数是 sigmoid 函数,当神经元的输出接近 1 时认为该神经元处于激活状

态,反之,当其输出接近 0 时认为该神经元处于抑制状态,因此,对自编码神经网络加入稀疏性限制是指限制神经元绝大多数处于抑制状态,只有少数可表示输入数据之间关系的神经元处于激活状态. 如图 2 所表示的就是一种典型的自编码神经网络,它包含输入层,隐藏层和输出层,其中 +1 为偏置项系数.

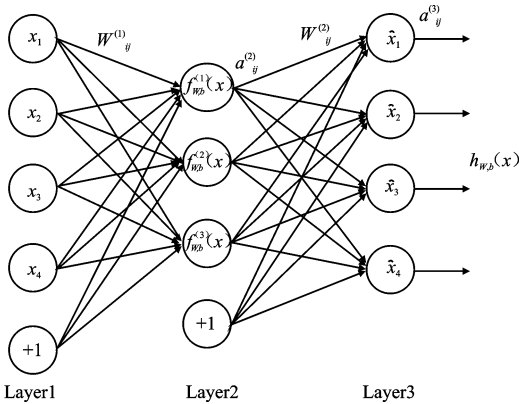


图 2 自编码神经网络示意图

假设有 m 个训练样本 $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$,对于整个训练样本集其损失函数可以被定义为

$$J(W;b) = [\frac{1}{m} \sum_{i=1}^m (\frac{1}{2} \| y - h_{W,b} \|^2)] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W^{(l)}_{ji})^2.$$

式中,第一项 $J(W,b)$ 是一个误差平方项,第二项是一个规则化项,目的是为了减小权重衰减的幅度,防止网络产生过拟合. 使用梯度下降法使 $J(W,b)$ 达到最小,梯度下降法中每次迭代过程都按照如下公式对 W 和 b 进行更新:

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

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

因此,求所有训练样本的损失函数的偏导数:

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

$$\frac{\partial J(W; b)}{\partial b_{ij}^{(l)}} = \frac{1}{m} \frac{\partial J(W, b; x^{(i)}, y^{(j)})}{\partial b_{ij}^{(l)}}.$$

对于单个样例我们使用反向传播算法求解其偏导数.

(1) 首先通过前向传播算法计算隐层 l 中所有神经元的激活值,这里假设隐层的输出函数 $f(z)$ 采用 sigmoid 函数,故

$$f'(z_i^{(l)}) = f(z_i^{(l)})(1 - z_i^{(l)}).$$

$$\text{令 } \alpha_i^{(l)} = f(z_i^{(l)}), \text{ 故}$$

$$f'(z_i^{(l)}) = \alpha_i^{(l)}(1 - \alpha_i^{(l)}).$$

(2) 对于输出层 n_i 的每个神经元 i ,根据以下公式计算残差:

$$\begin{aligned} \sigma_i^{n_1} &= \frac{\partial \frac{1}{2} \|y - h_{w,b}(x)\|^2}{\partial z_i^{(n_1)}} \\ &= -(y_i - \alpha_i^{(n_1)}) \cdot f'(z_i^{(n_1)}). \end{aligned}$$

(3) 对于隐层 l ,通过下面公式计算隐层的第 i 个节点的残差:

$$\sigma_i^{(l)} = \left(\sum_{j=1}^{s_{l+1}} W_{ji}^{(l)} \sigma_j^{l+1} f'(z_i^{(l)}) \right).$$

(4) 按照下式计算单个样例的偏导数:

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

最终根据以上迭代步骤重复利用梯度下降法使损失函数 $J(W; b)$ 的值达到最小值,求解出自编码神经网络.

而稀疏自编码神经网络就是在以上损失函数中施加一些稀疏性限制项,如下:

假设 $\alpha_j^{(l)}$ 表示自编码神经网络在给定输入 x 时,隐层 l 的第 j 个神经元的响应,定义隐层神经元 j 在整个训练集合上的平均活跃程度为

$$\bar{\rho}_j = \frac{1}{m} \sum_{i=1}^m [\alpha_j^{(l)}(x^{(i)})].$$

令 $\rho_j = \rho$,其中 ρ 为稀疏性参数,在误差函数中施加一个惩罚项得到:

$$\begin{aligned} J_{\text{sparse}}(W; b) &= J(W; b) + \beta \sum_{j=1}^{s_l} \rho \log \frac{\rho}{\bar{\rho}_j} \\ &\quad + (1 - \rho) \log \frac{1 - \rho}{1 - \bar{\rho}_j}. \end{aligned}$$

本文提取出来的特征可视化后如图 3 所示(其中隐层神经元的个数为 400 个).

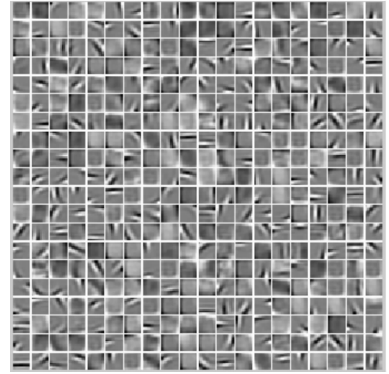


图 3 稀疏自编码神经网络学习得到的特征

2.3 卷积和池化

稀疏自编码神经网络训练完成后会得到一个权值矩阵,这个权值矩阵即为训练学习到了图像的不同位置对图像进行不同方向上的边缘特征检测,而图像具有稳定性,即图像中某个部分的统计特征和其他部位的特征相似,因此学习得到图像小块上的特征也适合于原始图像的其他部位. 本文将学习得到的特征矩阵作为一个卷积核和原始图像作一个卷积过程,卷积后即可得到原始大图片上不同位置所对应的不同特征的激活值,假设原始图片大小为 $m \times n$,而图像小块的大小为 $n \times n$,则卷积后每张图片的激活值的大小为 $(m - n + 1) \times (m - n + 1)$,由于得到的特征维数太高,因此,采用池化的方法对这些激活值进行一个聚合统计,池化后不仅能将最终的手势的全局边缘特征提取出来,还能达到一个降维的作用,大大减少了计算量.

2.4 分类及调优

将池化后的特征作为分类器的输入来训练分类器,分类器训练好后,对训练好的网络参数进行调优,调优的方法采用求出网络的输出值与训练样本实际值之间的误差,通过反向传播和梯度下降法逐层更新网络参数,实现对整个网络的有监督调优,最后利用有标签的测试集来测试该分类器的分类效果,本文分类器选用的是 Softmax 分类器.

3 实验结果及分析

文中将预处理后的图像小块数据输入到一个隐层含有 400 个神经元的稀疏自编码神经网络,训练出来的特征作为卷积核提取原始图片的边缘特征,卷积核的大小为图像小块的大小 $\text{PatchDim}=10$,池化的区域大小 $\text{PoolingDim}=11$. 实验结果的混淆矩

阵图如图 4 所示,从图中可以看出错误率主要是由于手势 buckle 有 10%被错分成了手势 rotate,而手势 shoot 有 18%被错分成了手势 cut,从数据库的样例图可以看出,错分的主要原因是由于这两对手势的边缘特征相似度很高,故导致最终结果产生混淆.

由表 2 所示的实验结果可以看出:有调优的分类准确率高于无调优的分类准确率,且识别的准确率随着隐层节点数的增加而提高,原因是增加节点数可以减少自编码神经网络在无监督特征学习阶段信息的信息损失,有利于提高分类准确率.

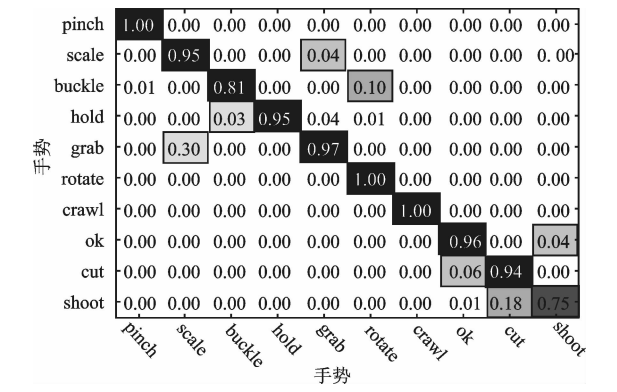


图 4 Simple Sign Language 数据库识别结果混淆矩阵图

表 2 UFL-DNN 在 Simple Sign Language 数据库上的实验结果

实验编号	隐层网络配置	准确率(无调优)/%	准确率(有调优)/%
a	400	80.41	85.72
b	600	80.25	85.81

4 结束语

本文提出了一种将无监督的特征学习和有监督的分类识别相结合的手势图像识别方法,通过无监督的稀疏自编码神经网络训练提取手势图像的边缘特征,并将此边缘特征作为训练分类器的输入.实验证明,将无监督的特征学习方法和深度神经网络相结合的模型可以有效地提取手势图像的边缘特征,并且,对训练好的分类器进行参数调优可以提高分类的准确率.

参考文献:

[1] Liwicki S, Everingham M. Automatic recognition of fingerspelled words in british sign language [C] // Computer Vision and Pattern Recognition Workshops, IEEE Compute Society Conference on IEEE. [s. l.]: IEEE, 2009:50-57.

[2] Geetha M, Rohit Menon, Suranya Jayan, et al. Gesture recognition for american sign language with polygon approximation [C] // 2011 IEEE International Conference on Technology for Education, Chengal, Tamil Nadu:IEEE,2011:241-245.

[3] Isaacs J, Foo S. Hand pose estimation for american sign language recognition [C] // InSystem Theory, 2004. Proceedings of the Thirty-Sixth Southeastern Symposium on. [s. l.]: IEEE, 2004:132-136.

[4] Uebersax D, Gall J, Van den Bergh M, et al. Real-time sign language letter and word recognition from depth data132 - 1362011 [C] // IEEE International Conference on Computer Vision Workshops (ICCV Workshops). Chennai, Tamil Nada: IEEE, 2011: 383-390.

[5] 敖道敢. 无监督特征学习结合神经网络应用于图像识别 [D]. 广州:华南理工大学, 2014.

[6] 王勇, 赵俭辉, 章登义, 等. 基于稀疏自编码深度神经网络的林火图像分类 [J]. Computer Engineering and Applications, 2014, 50(24):173-177.

[7] 张勤, 赵健, 孙道达, 等. 一种基于稀疏表示的手势识别算法 [J]. 西北大学学报: 自然科学版, 2013, 43 (6): 881-884.

作者简介:

陶美平 女,(1991-),硕士.研究方向为数字多媒体技术.
E-mail:tmpaqzy@163.com

马 力 男,(1961-)汉族,博士,教授,研究方向为计算机网络与信息系统集成研究,网络信息智能化处理方法研究、网络环境下教育与学习理论应用实践研究.

黄文静 女,(1992-),汉族,硕士.数字多媒体技术.

吴雨隆 男,(1987-),汉族,硕士.图形图像处理技术.