

文章编号:1005-0523(2020)04-0041-07

DOI:10.16749/j.cnki.jecjtu.2020.04.007

# 基于 CNN 和粒子群优化 SVM 的手写数字识别研究

杨刚,贺冬葛,戴丽珍

(华东交通大学电气与自动化工程学院,江西 南昌 330013)

摘要:针对传统卷积神经网络手写体数字识别中 Softmax 因指数函数运算而易产生计算溢出以及较高的计算机硬件需求问题,提出了基于卷积神经网络特征提取的支持向量机手写体数字识别方法。同时,为了提高手写体数字的识别精度,设计了基于 K-CV 意义下适应度函数的粒子群优化 SVM 参数方法。基于 Semeion 及 MNIST 手写体数字集的实验仿真表明,文章所设计的方法与传统方法相比能够获得更高的识别率。

关键词:卷积神经网络;支持向量机;粒子群算法;手写识别

中图分类号:TP391.9

文献标志码:A

本文引用格式:杨刚,贺冬葛,戴丽珍.基于 CNN 和粒子群优化 SVM 的手写数字识别研究[J].华东交通大学学报,2020,37(4):41-47.

Citation format: YANG G, HE D G, DAI L Z. Improved handwritten digit recognition based on CNN and PSO-SVM[J]. Journal of East China Jiaotong University, 2020, 37(4): 41-47.

手写体数字识别在银行、税务和邮政系统等领域有着重要的应用和广阔的发展前景,尤其在脱机工作环境下有很大的提升空间<sup>[1]</sup>。传统的手写体数字识别方法主要有支持向量机<sup>[2-4]</sup>,神经网络算法<sup>[5-6]</sup>等。但由于手写数字本身所包含的特征较少,加上不同人书写数字字符差异较大,在识别率方面具有较大的提升空间。以深度学习为代表的机器学习方法的出现<sup>[7]</sup>,大大降低了图片识别的难度,为手写体数字识别提供了有效工具。卷积神经网络由于其“深层结构”的优势,常被用于图片特征提取,并通过 Softmax 对相关特征进行分类。尽管该方法具有较理想的处理效果,但对电脑要求过高,且计算复杂、耗时较长。为了提高手写体数字的识别率,本文拟结合卷积神经网络的特征提取能力、支持向量机的分类能力、粒子群优化的寻优能力,通过 CNN 对手写体数字图片进行特征提取,采用粒子群优化 SVM 关键参数对数字特征进行识别,从而实现手写体数字的识别。

## 1 卷积神经网络

卷积神经网络(convolution neural network, CNN)<sup>[8]</sup>是一种前馈神经网络,它的人工神经元可以响应一部分覆盖范围内的周围单元,对于大型图像处理有出色表现。

CNN 是一个受生物视觉启发、以最简化预处理操作为目的的多层感知器的变形,它结构的可拓展性很强,可以采用较深的层数,构成的深度模型具有更强的表达能力<sup>[9]</sup>。卷积神经网络由一个或多个卷积层(convolution layer)和末端的全连接层组成,同时也包括关联权重和池化层(pooling layer)。这一结构使得卷积神经网络能够利用输入数据的二维结构,并且也可以使用反向传播算法进行训练。相比较其他深度、前馈神经网络,卷积神经网络需要考量的参数更少,使之成为一种颇具吸引力的深度学习结构<sup>[10]</sup>。

### 1.1 卷积层

卷积层的主要作用是通过对卷积核大小的设置,由浅入深不断对前一层传输的数据进行特征提取。由于设置了共享权值,在同一特征图中神经元使用同一组卷积核,可以减少训练参数。其中,卷积核数值在初始化后由后续的网络训练确定<sup>[11]</sup>。

收稿日期:2020-02-19

基金项目:国家自然科学基金项目(61663012,61673172,61733005,61863014)

作者简介:杨刚(1983—),男,博士,讲师,研究领域为复杂系统建模,神经计算理论及其应用等。

通信作者:戴丽珍(1983—),女,博士,讲师,研究领域为轨道交通自动化与运行优化。

图片各像素点的值与卷积核的乘积加入偏置后经过激活函数的运算即可得到图片的一个特征映射:

$$a_j^L = f\left(\sum_i a_j^{L-1} \times w_{ij}^L + b_j^L\right) \quad (1)$$

其中:  $a_j^L$  表示  $L$  层卷积后第  $j$  个神经元的输出;  $w_{ij}^L$  表示卷积核;  $b_j^L$  表示偏置。  $f(\cdot)$  为神经元激活函数, 这里我们采用 Sigmoid 函数, 即

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

### 1.2 池化层

池化是卷积神经网络中的一个重要操作, 能够减少图片中冗余特征, 同时保持特征的局部不变性<sup>[12]</sup>。

图片经过卷积处理后, 每个  $n \times n$  邻域内的像素点采用最大池化(MaxPooling)的方法变为一个像素

$$a_j^{L+1} = \text{down}(a_j^L) \quad (3)$$

其中:  $\text{down}(\cdot)$  表示下采样函数, 该层运算不包含可学习的权值和阈值。

### 1.3 全连接层

全连接层可以整合前向传来的具有类别区分性的局部信息; 同时, 可以增强网络的非线性映射能力; 限制网络规模的大小。

## 2 支持向量机

支持向量机是建立在统计学理论基础上的—种数据挖掘算法, 其工作机理是寻找一个满足分类要求的最优分类超平面, 使得该超平面在保证分类精度的同时, 能够使超平面两侧的空白区域最大化。理论上, 支持向量机能够对线性可分数据的最优分类<sup>[13]</sup>。其原理示意图如图 1 所示。

本文使用 LIBSVM 工具箱通过一对一的方法来实现分类器的设计构造。当要对一个未知类别样本分类时, 输出结果值最大的即为该样本类别<sup>[14]</sup>。

经过多年来国内外的研究表明, SVM 以 RBF

核(径向基函数)为核函数具有很强的学习能力和分类效果。文中选用 RBF 作为 SVM 的核函数

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma}\right) \quad (4)$$

### 2.1 基于的 SVM 的参数优化

在用 SVM 做分类预测时若想要提高结果的准确率往往需要对 SVM 参数进行优化以选取最优的参数<sup>[15]</sup> (主要是惩罚参数  $c$  和核函数参数  $g$ )。

SVM 中惩罚系数  $c$  ( $c$  为正数) 用来表征对误差的宽容度;  $c$  值过高或过低均会影响 SVM 的泛化能力。参数  $g$  是选择 RBF 函数作为 Kernel 后该函数自带的一个参数, 隐含地决定了数据映射到新的特征空间后的分布。

PSO 首先是在可行解中初始化一群粒子, 每个粒子都有自身的速度、位置跟适应值, 通过不断的迭代搜索最优解。在每一次迭代中粒子通过个体最优值跟全局最优值来更新自己。

对于随机初始化产生的粒子, 将其视为第一代初始种群, 通过目标函数  $Q(X_i)$  计算得出的适应度值来衡量种群  $X_i$  的优劣。种群中粒子  $i$  当前最优位置为

$$X_{best_i} = \begin{cases} X_{best_i}(t) & \text{若 } Q[X_i(t+1)] > Q_{best_i} \\ X_i(t+1) & \text{若 } Q[X_i(t+1)] > Q_{best_i} \end{cases} \quad (5)$$

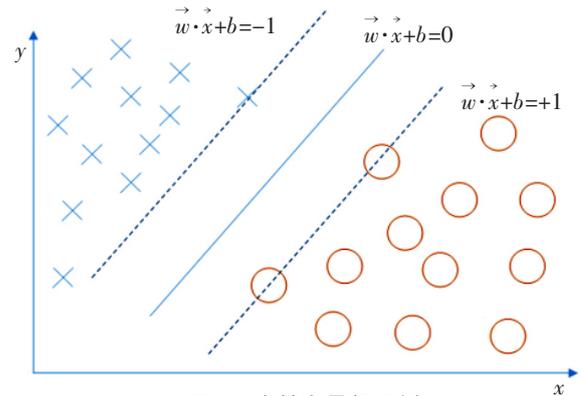


图 1 支持向量机示例

Fig.1 Example of support vector machines

其中: $Xbest_i$  为粒子  $i$  所经历的最优位置标记; $Qbest_i$  为粒子  $i$  位置最优时所对应的适应值。寻优过程中,粒子不断更新自己的位置跟速度,速度更新公式为

$$v_{in}(t+1)=wv_{in}(t)+c_1r_1[Xbest_{in}-x_{in}(t)]+c_2r_2[Xbest_{in}-x_{in}(t)] \quad (i=1,2,\dots,m,n=1,2,\dots,N) \quad V_i \in [-V_{max},V_{max}] \quad (6)$$

其中: $Xbest_g$  为粒子  $g$  所经历的最优位置标记(对应的适应度值应为  $Qbest_g$  此处没有体现), $V_{max}$  为粒子单步更新的速度最大值; $c_1,c_2$  为加速度常数,表示 PSO 参数局部搜索能力; $r_1,r_2$  是两个相互独立的随机参数; $w$  为惯性权重,在优化过程中能够起到提高全局搜索跟局部搜索的作用。

位置更新公式为

$$\begin{cases} x_{in}(t+1)=x_{in}(t)+v_{in}(t+1) \\ x_{in}(t1)=x_{i_{min}}+v_{i_{min}} \end{cases} \quad (7)$$

其中  $x_i=(x_{i1},x_{i2},\dots,x_{in})$  表示粒子  $i$  位置向量中的值。

适应度函数的确定是实现 PSO 算法的先决条件,更是寻优的依据。传统方法是寻找或者设计一个函数用以计算适应度值,以显示寻优效果的优劣<sup>[16]</sup>,但计算复杂,使用困难。本文使用 CV 算法,将对训练集在 CV 意义下的准确率作为粒子群中的适应度函数值。

### 2.2 交叉验证

交叉验证(CV)是一种常用的验证分类器性能的统计分析方法<sup>[17]</sup>。关于对 SVM 参数的优化选取是将  $c$  和  $g$  在一定的取值范围内取值,然后将训练集作为原始数据采用 K-CV 的方法对 SVM 进行训练,训练后得到的模型再通过验证集验证后得出分类准确率,准确率最高的模型即为最佳参数。对于可能出现的过学习状态的发生我们选择最佳参数中惩罚参数  $c$  最小的一组( $c,g$ )参数为最优参数。

这里将对训练集进行 CV 意义下结果中最好的模型识别准确率作为 PSO 的适应度值。基于 PSO 的 SVM 全局参数寻优的整体算法过程如图 2 所示。

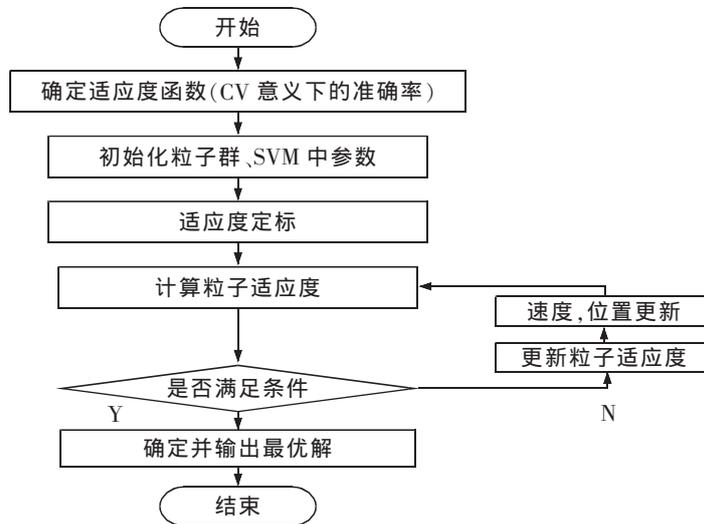


图 2 PSO 优化 SVM 参数流程图  
Fig.2 Flow chart of PSO optimized SVM parameters

### 3 基于 CNN 和 PSO-SVM 的手写数字识别

为提高网络对手写数字图像的识别精度,本文将卷积神经网络和支持向量机结合,并通过 PSO 优化算法优化 SVM;PSO 是基于个体间的协作实现搜索空间中的寻优,能很好的提高 SVM 的识别性能。

在使用粒子群优化 SVM 参数的实验中,相比于随机初始化粒子确定适应度值,本文使用 CV 意义下的准确率将其代替,以达到快速收敛的效果。整体结构如图 3 所示

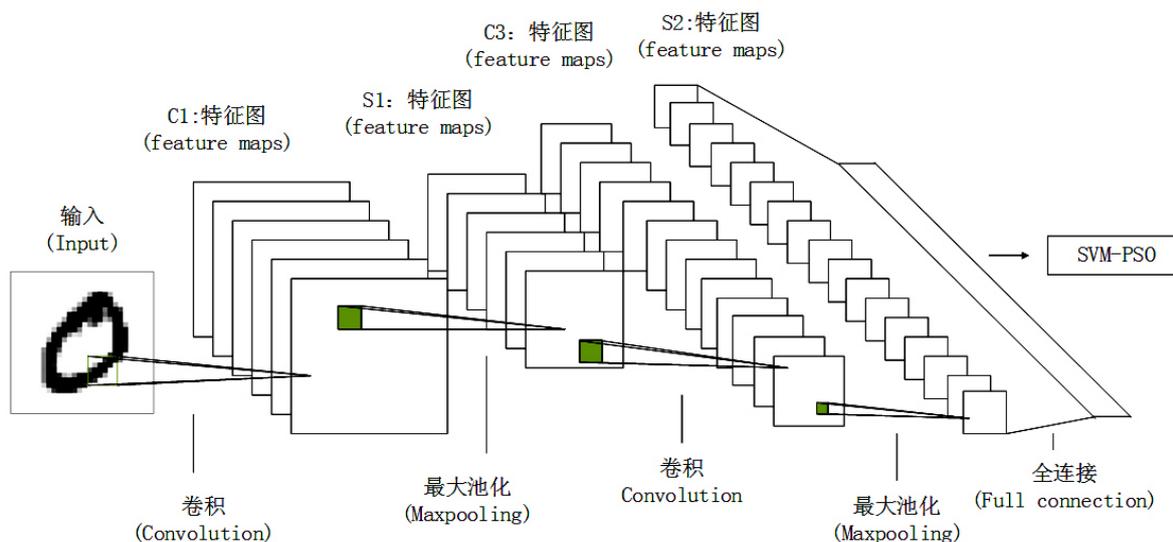


图3 CNN-SVM 整体结构图

Fig.3 CNN-SVM overall structure diagram

基于 CNN 和 PSO-SVM 的手写数字识别流程主要步骤如下:

Step1: 将手写数据集分为训练集和测试集;

Step2: 通过 Python 程序将二维图片数据传入卷积网络输入层;

Step3: 图片经过卷积层、池化层得以特征提取跟降维, 得到含有特征信息较高的数据集, 经全连接层后, 数据由二维转换为一维;

Step4: 将全连接后的一维数据通过 Matlab 传入 SVM;

Step5: 使用 K-CV 的方法将一维的训练数据作为训练数据得到当前组  $(c, g)$  的分类准确率, 并将其作为 PSO 中粒子的适应度函数;

Step6: 根据粒子群对 SVM 参数的不断寻优, 最终确定出最佳的  $c, g$  参数(寻优过程如图 4 所示);

Step7: 训练完成后将一维的训练集送入最终的模型中进行分类验证, 以最终对手写数字图片的分类准确率作为最后的结果。

## 4 实验与分析

本文设计构建的卷积神经网络包含两个卷积层、两个池化层和一个全连接层, 分类层采用优化后的 SVM 分类器。

通过使用 UCI 提供的 Semeion 手写数字集及经典 MNIST 手写数字集分别对网络进行训练。两数据集均将图片分为 0~9 十类, 图片显示的手写阿拉伯数字即为该图片的属性类别。实验中卷积核大小设置为  $5 \times 5$ , 池化层中采样大小为  $2 \times 2$ 。

### 4.1 Semeion 手写数字数据集

UCI 提供的 Semeion 手写数字数据集共有 80 人参与, 每人在非刻意的前提下在纸上快速将数字 0~9 手写 2 次。数据集共有 1 593 张图片, 从中随机抽取 1 500 张图片进行实验并将其分为训练集跟测试集, 训练集共有 700 张图片, 测试集有 800 张图片; 图片为  $16 \times 16$  像素的一维灰度图片。每张图片均采用固定阈值将像素点缩放为二进制(1/0)值。图 4 为该数据集示例图片。



图4 部分手写数字图像

Fig.4 Partial handwritten digital images

实验主要是通过使用粒子群对 SVM 中的参数不断寻优,以提高分类识别层的识别准确率。参数  $c, g$  均在一定范围内取值,于是我们根据实验经验设定  $c \in [0, 100], g \in [-100, 100]$ ,通过初始化选取 20 个种群,经过 200 次迭代寻优。

实验中我们随机将测试数据集均分为 8 组,每组 100 个样本进行实验仿真,实验结果如图 5~图 6 所示。

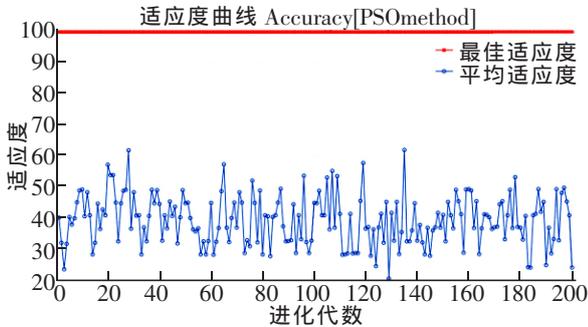


图 5 PSO 参数寻优适应度曲线

Fig.5 PSO parameter optimization fitness curve

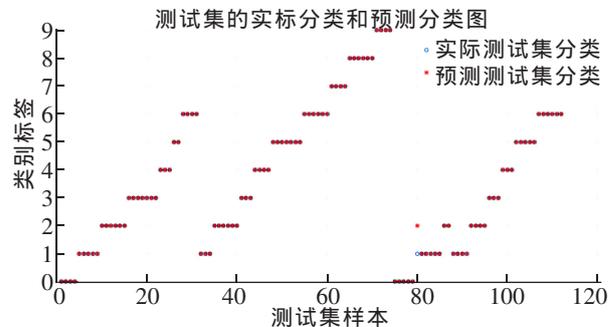


图 6 测试集分类图

Fig.6 Test set classification diagram

可以看到 100 组图片中只有一张未能识别正确;通过测试验证可知该模型对数字的识别率达 99.11%, PSO 寻优最佳参数为  $c=2.297, g=0.01$ 。

在手写数字识别的实验中,对于每组能够正确识别的图片个数  $n$  将其除以每组图片总数  $m$ ,得出该组下的识别正确率  $p$ ,即  $p = n/m \times 100\%$ 。至此,将相同的训练测试数据试验在不同的模型中进行对比,结果如表 1~表 2,图 7 所示。

表 1 UCI 不同算法的识别率

Tab.1 Recognition rate of different algorithms for UCI

| UCI 数据 | 识别率/% |         |       |          |
|--------|-------|---------|-------|----------|
|        | 本文算法  | CNN-SVM | SVM   | SVM(PSO) |
| Data 1 | 95.05 | 94.83   | 80.19 | 89.11    |
| Data 2 | 97.03 | 95.67   | 81.18 | 87.12    |
| Data 3 | 95.15 | 95.84   | 77.23 | 90.09    |
| Data 4 | 96.04 | 94.83   | 83.13 | 88.12    |
| Data 5 | 99.01 | 95.83   | 80.19 | 89.11    |
| Data 6 | 99.11 | 94.06   | 79.21 | 90.10    |
| Data 7 | 98.02 | 95.97   | 82.18 | 92.08    |
| Data 8 | 96.04 | 95.82   | 82.11 | 89.11    |

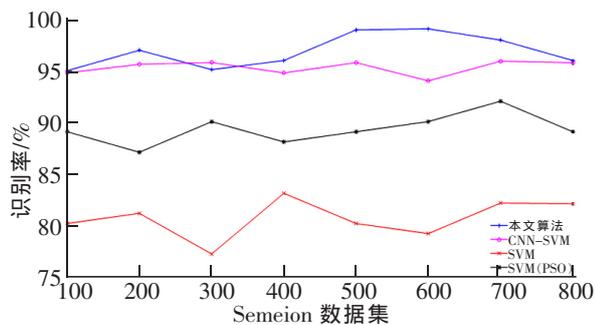


图 7 UCI-Semeion 数据集实验结果

Fig.7 Experimental results of UCI-Semeion dataset

表 2 UCI 数据集识别率

Tab.2 UCI data set recognition rate

| 识别方法                      | 识别率/% (错误个数) |
|---------------------------|--------------|
| CNN-SVM(PSO)              | 99.11(1)     |
| CNN-SVM                   | 95.97(4)     |
| SVM                       | 83.13(18)    |
| SVM(PSO)                  | 92.08(8)     |
| SVM(HSS1) <sup>[18]</sup> | 97.80(*)     |
| FS-CSVM <sup>T19]</sup>   | 98.90(*)     |

可以看出经过 PSO 参数优化后的卷积神经网络结合支持向量机算法精度最高,均保持在 95%以上,对处理图片分析上具有一定的优势。

文献[18]显示为 SVM 在不同特征情况下在该数据集中的识别精度(文献中实验未显示识别错误个数),可以看到在对 SVM 加入不同特征后,本文算法识别正确率仍然高于文献[18]中对应的识别算法。文献[19]中提出了一种基于边界对特征的敏感度值进行特征选择的支持向量机树混合学习模型,其识别率为 98.9%。

#### 4.2 MNIST 数据集

MNIST 手写数据集包含 70 000 个样本,每个样本为  $28 \times 28$  像素的灰度图片,其中训练集有 60 000 张图片,测试集有 10 000 张。图 8 为该数据集的部分样例。

由于数据集数据样本较多,此次实验随机从数据库中选取 600 张训练图片,1 000 张测试图片,并将测试集随机分为十等份(每份 100 张图片)进行测试验证。表 3 显示十份验证集中最优识别情况。

由实验结果可知,本文中采用的方法识别率最高为 96%,即 100 个随机选取的数据样本中只有 4 个未能识别正确。文献[20]使用 SOM 简化算法及并行电路架构,采用 1 000 组数据进行测试,得到 84.13%的准确率。文献[21]中设计了分数阶梯度下降学习机制,在分数阶 BP 神经网络中自适应更新连接权值,最终提出的 PEO-FOBP 在 MNIST 数据集中识别精度有 96.54%其对应的未加入极值优化的自适应的 PEOBP 算法识别度为 95.67%。

对以上实验进行分析可看到,CNN 结合 PSOSVM 的算法对分类识别手写数字图片有良好的效果,在一定程度上提高了对图片的识别率。

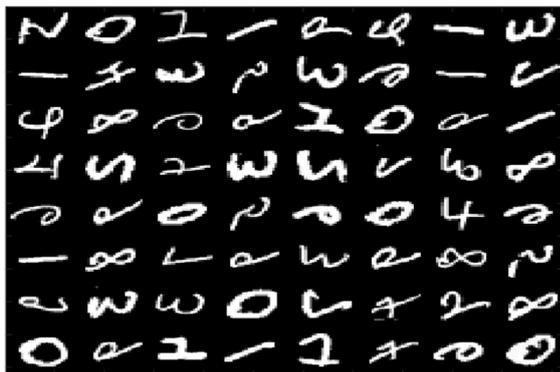


图 8 MNIST 数据集部分样例  
Fig.8 Partial sample of MNIST dataset

表 3 MNIST 数据集识别率  
Tab.3 MNIST dataset recognition rate

| 识别方法                            | 识别率/% (错误个数)   |
|---------------------------------|----------------|
| CNN-SVM(PSO)                    | 96(4)          |
| CNN-SVM                         | 94(6)          |
| SVM                             | 86(13)         |
| SVM(PSO)                        | 83(17)         |
| 自组织映射 <sup>[20]</sup>           | 84.13(*)       |
| PEO-BP/PEO-FOBP <sup>[21]</sup> | 95.67/96.54(*) |

## 5 结束语

本文提出使用卷积结合支持向量机的方法对图片进行识别,并采用基于 K-CV 作为适应度函数的 PSO 对 SVM 参数进行优化,以此提高支持向量机的识别性能。

在图像识别的问题中,卷积神经网络通过自身深度学习的能力获取图像特征,对图像有较高的识别率;传统卷积神经网络使用 Softmax 进行分类识别,Softmax 属于线性模型,本身具有分类速度快、模型占用空间小等优势,但使用较强的线性假设,在分类问题上仍有一定的提升空间;进行计算时涉及指数函数的运算,当函数增长时过高的函数值将影响计算机的输出结果,因此对计算机硬件设备有一定的要求。

实验采用 Python3.5 与 Matlab2015b 作为实验平台,对 UCI 提供的手写图片和 MNIST 手写数据集进行分类识别。在确定最佳参数  $c$ 、 $g$  的值后,最终对两款测试集的识别率分别达到 99.11%和 96%。然而,在实验过程中也遇到一些问题,比如实验设备比较老旧,对于更复杂的程序难以顺利运行(更深层次的卷积神经网络对硬件设备要求较高),像 MNIST 这样文件较大的数据集也很难实现。但对于该次实验设计的网络结构较为简单且实验结果表明该算法识别性能较好,对手写数字正确识别率有一定的提高,因此该算法在研究领域有深入研究的潜在价值。

## 参考文献:

- [1] 苗夺谦,张红云,李道国,等. 基于主曲线的脱机手写数字识别[J]. 电子学报,2005(9):1639-1643.
- [2] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报,2000(1):36-46.
- [3] TANG F, ADAM L, SI B. Group feature selection with multiclass support vector machine[J]. Neurocomputing, 2018, 317:42-49.
- [4] 丁世飞,齐丙娟,谭红艳. 支持向量机理论与算法研究综述[J]. 电子科技大学学报,2011,40(1):2-10.
- [5] 韩力群. 神经网络教程[M]. 北京:北京邮电大学出版社,2006.
- [6] LECUM Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
- [7] 周俊,陈剑云. 基于DSSD的接触网鸟窝识别检测研究[J]. 华东交通大学学报,2019,36(6):70-78.
- [8] 常亮,邓小明,周明全,等. 图像理解中的卷积神经网络[J]. 自动化学报,2016,42(9):1300-1312.
- [9] 周飞燕,金林鹏,董军. 卷积神经网络研究综述[J]. 计算机学报,2017,40(6):1229-1251.
- [10] ZHANG W, JIN L, SONG E, et al. Removal of impulse noise in color images based on convolutional neural network[J]. Applied Soft Computing, 2019:105558.
- [11] ABDEL HAMID O, MOHAMED A R, JIANG H, et al. Convolutional neural networks for speech recognition[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2014, 22(10):1533-1545.
- [12] GUO G, WANG H, YAN Y, et al. A fast face detection method via convolutional neural network [J]. Neurocomputing, 2020, 395:128-137.
- [13] 郭明玮,赵宇宙,项俊平,等. 基于支持向量机的目标检测算法综述[J]. 控制与决策,2014,29(2):193-200.
- [14] FAN R E, CHEN P H, LIN C J. Working set selection using second order information for training SVM[J]. Journal of Machine Learning Research, 2005, 6(4):1889-1918.
- [15] 傅军栋,邹欢,康水华. PSO-SVM算法在智能建筑环境监控系统中的应用[J]. 华东交通大学学报,2016,33(1):121-127.
- [16] GUO CHU C, JIN SHOU YU. Particle swarm optimization algorithm[J]. Information & Control, 2005, 186(3):454-458.
- [17] JIANG G, WANG W. Error estimation based on variance analysis of k-fold cross-validation[J]. Pattern Recognition, 2017, 69:94-106.
- [18] HAFIZ A M, BHAT G M. Handwritten digit recognition using Slope Detail Features[J]. International Journal of Computer Applications, 2014, 93(5):14-19.
- [19] 徐琴珍,杨绿溪. 基于改进的混合学习模型的手写阿拉伯数字识别方法[J]. 电子与信息学报,2010,32(2):433-438.
- [20] 王一木,潘赞,龙彦辰,等. 基于自组织映射的手写数字识别的并行实现[J]. 浙江大学学报(工学版),2014,48(4):742-747.
- [21] MINRONG CHEN, BIPENG CHEN, GUOQIANG ZENG, et al. An adaptive fractional-order BP neural network based on external optimization for handwritten digits recognition[J]. Neurocomputing, 2020, 391:260-272.

## Improved Handwritten Digit Recognition Based on CNN and PSO-SVM

Yang Gang, He Dongge, Dai Lizhen

(School of Electrical and Automation Engineering, East China Jiaotong University, Nanchang 330013, China)

**Abstract:** In order to solve the problems in traditional convolutional neural network-based handwritten digits recognition, such as the computation overflow and high demand of computer hardware caused by the exponential function operation of Softmax, a support vector machine (SVM) handwritten digital recognition method based on the feature extraction of convolutional neural network is proposed. And, in order to improve the recognition accuracy of handwritten digits, a particle swarm optimization method based on the fitness function in the sense of K-CV is designed. Experimental simulation based on Semeion and MNIST handwritten digits shows that the method proposed in this paper can achieve higher recognition rate than the traditional methods.

**Key words:** convolutional neural network (CNN); support vector machine (SVM); particle swarm optimization (PSO); handwritten recognition