

基于深度学习的数字识别方法研究

唐子清, 姚 俭

(上海理工大学 管理学院, 上海 200093)

摘要:近年来,手写数字识别是计算机视觉与模式识别中一个广受关注的问题。该问题的主要挑战是如何设计一种有效方法,以识别用户通过数字设备提交的手写数字。目前,深度学习算法在计算机视觉领域非常流行,被用于处理诸如图像分类、自然语言处理及语音识别等问题。以几种深度学习常见算法,包括线性感知器、卷积神经网络、循环神经网络、长短时记忆网络等为研究对象,分析其在水写数字识别方面的优缺点,并引入 Google 第二代人工智能系统 TensorFlow,对比相同算法在不同框架下的识别速度及准确率。实验结果表明,几类深度学习算法都能明显提高识别准确率,且在训练数据集时不会损耗过多计算资源。

关键词:深度学习;手写数字识别;TensorFlow

DOI:10.11907/rjdk.192629

中图分类号:TP317.4

文献标识码:A

开放科学(资源服务)标识码(OSID):

文章编号:1672-7800(2020)009-0228-05



Research on Digital Recognition Method Based on Deep Learning

TANG Zi-qing, YAO Jian

(School of Management, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: In recent years, handwritten numeral recognition is a widely concerned problem in computer vision and pattern recognition. The main challenge of this problem is to design an effective method to recognize handwritten numbers submitted by users through digital devices. At present, deep learning algorithm is very popular in computer vision, which is used to deal with important problems such as image classification, natural language processing and speech recognition. In this paper, several common algorithms of deep learning, including linear perceptron, convolutional neural network, cyclic neural network and short-term memory network are studied. Their advantages and disadvantages in handwritten numeral recognition are analyzed, and Tensorflow, the second generation of Google artificial intelligence system, is introduced. Then the recognition speed and accuracy of the same algorithm in different frameworks are compared. The results show that these deep learning algorithms can obviously improve the recognition rate, and do not consume too much computing resources when training data sets.

Key Words: deep learning; handwritten digit recognition; TensorFlow

0 引言

图像识别^[1]是模式识别领域的主要应用技术之一,图像识别过程包括训练与测试两个阶段,主要由图像预处理、特征提取与分类3个环节组成。首先对采集到的图像进行预处理,去除图像中的噪声,并将其转化为一个清晰的点矩阵^[2],然后提取所需的目标图像特征,最后对图像特征进行分类。目前主要采用机器学习方法进行图像识别,尤其在信息化、智能化的时代背景下,人们需要准确率与

效率更高的算法以满足当前需求。

1 深度学习概述

深度学习采用隶属于机器学习的层次抽象思想,可对数据进行分层,并从数据每一层学习到有用特征。其优点是可以大大减少人工监督的工作量,从而节省大量人工成本。

1950年,图灵设计一个测试机器能否像人类一样思考的实验;20世纪80年代,BP算法(Back Propagation)开始应用于神经网络,与过去人为建立的规则相比,神经网络

收稿日期:2019-11-19

作者简介:唐子清(1996-),男,上海理工大学管理学院硕士研究生,研究方向为系统控制与优化;姚俭(1961-),男,博士,上海理工大

学管理学院教授,硕士生导师,研究方向为系统工程

模型可通过大量数据推导出规则^[4];90年代,最大熵和支持向量机SVM等浅层机器学习算法在理论分析与应用中取得的成功使得这些算法成为当前最常见的机器学习算法,但人工神经网络研究进入低谷。

2006年,深度学习概念被提出,世界许多顶级大学在深度学习研究中发挥着重要作用,一些组织和公司也专门组建深度学习研究部门以研究相关技术在实际中的应用;2011年,微软和谷歌研究小组彻底改变了原有语音识别技术框架,使语音识别错误率下降了20%~30%;2012年,深度学习在图像识别领域取得显著成果,图像识别错误率下降了11%^[5];2013年,百度创立深度学习研究院,并于当年5月推出针对百度搜索广告的学习系统^[6]。

2 图像识别技术

现代许多领域都需要应用图像识别技术,如许多企业需要准确识别图像中的数字信息。因此,通过提高识别准确率以及缩短识别时间,可大大提高企业工作效率。传统数字识别方法存在需要手动提取特征、效率低以及算法鲁棒性差等缺点,而利用深度学习技术,在特征提取过程中只需前期编码以及环境参数配置正确,即可大大提高在线运算效率^[8]。

2.1 基于神经网络的图像识别方法

基于神经网络的图像识别系统^[12]是一种新型模式识别系统,其识别过程包括以下步骤:首先规范化相关信息,并对信息进行删除、平滑、二值化等处理,然后进行特征提取,最后使用BP神经网络进行分类。分类时需要选择不同类型的样本进行训练,并且每种类型样本数量大致相同,一方面可以防止BP网络对某一类型样本过于敏感,另一方面可以极大地提升训练速度^[13]。

2.2 基于小波矩的图像识别方法

基于小波矩的图像识别方法具有以下特点:

(1)小波矩特征对于只经过平面变换(如平移、旋转缩放等)的样本具有良好的分辨率,而且可以正确区分无噪声情况下的测试样本,识别正确率最高可达98%。

(2)小波矩识别率会随着噪声的增加而下降,但下降速度较为缓慢。

(3)小波矩特征相对稳定,但随着特征数的增加,分辨率会发生一定变化,但变化范围不大,小波矩精度仍能够稳步提高^[14]。

2.3 基于分形特征的图像识别方法

一般认为自然纹理图像满足分形特征,绘制分形特征可实现纹理分离与目标识别。但人为制作的图像诸如画作之类无法进行分形特征提取,因为其并不具有自然的纹理图像。

基于分形特征的红外图像识别首先由红外探测器获取图像进行预处理,然后提取分形特征,最后进行图像识别。该识别方式多应用于生物领域,例如骨纹理识别等。

3 深层神经网络训练方法

3.1 线性感知器算法

感知器^[17]由输入权重、激活函数与输出组成,其存在的问题是,当感知数据集不能线性分离时,感知器规则不会收敛,则无法完成训练。为解决该问题,本文使用微分线性函数而不是感知器的阶跃函数。该感知器被称为线性单位,其特点是在遇到无法线性分离的数据集时,可将已有线性元素收敛到最佳逼近值上。

将权重项与偏置项初始化为0,然后利用以下感知器规则迭代地修改 ω 和 b ,直到训练完成。

$$\omega_i \leftarrow \omega_i + \Delta\omega$$

$$b \leftarrow b + \Delta b$$

其中, $\Delta\omega_i = \eta(t - y)x_i$, $\Delta b = \eta(t - y)$ 。 ω_i 是与输入 x_i 对应的权重项, b 是偏置项, η 是一个用于表示学习速率的常数, t 是训练样本实际值,一般称为label,可根据公式计算得出感知器输出值 y 。

3.2 卷积神经网络算法

卷积神经网络(Convolutional Neural Network, CNN)是最重要的一种神经网络,几乎所有图像识别及语音识别领域的重要突破都应用了卷积神经网络,如谷歌的GoogleNet、微软的ResNet,以及打败李世石的AlphaGo等^[18]。图1为卷积神经网络示意图。

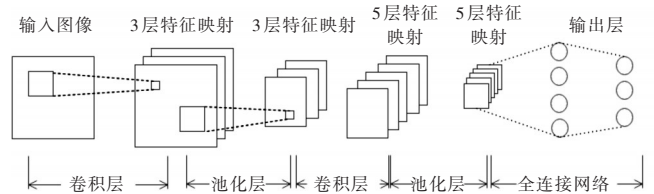
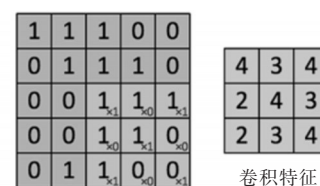


图1 卷积神经网络

不同于传统神经网络,卷积神经网络引入感受野(receptive field)、权值共享(shared weights)、下采样(sub-sampling)3个概念,有助于更好地对对象进行分类或识别。尤其在类似自编码的深层模型框架中,输入层与隐藏层完全连接,如果输入采样数据值很小(如8*8或28*28),计算效果较好。尽管有更大的图像数据(例如96*96,256*256或更多自然图像),通过完全连接的网络计算特征将非常耗时。与28*28补丁图像相比,96*96图像计算过程是倍增功率(=100)的10倍。

卷积是图像的滤波,图2为卷积原理图。图像统计特征在所有区域都是相似的,可使用相同的学习特征,即相同的滤波器或卷积核。



图像

卷积特征

图2 卷积原理

假设给定 $\gamma \times c$ 的大尺寸图像, 将其定义为 X_{large} 。根据从大尺寸图像抽取的 $a \times b$ 小尺寸图像样本 X_{small} , 计算 $f = \sigma(W^{(1)}X_{small} + b^{(1)})$ (σ 为激活函数) 可得到 k 个特征, 其中 $W^{(1)}$ 、 $b^{(1)}$ 是可视单元与隐含单元之间的权重和偏置。对于每一个 $a \times b$ 大小的小图像 X_s , 计算对应的 $f_s = \sigma(W^{(1)}X_s + b^{(1)})$, 对该 $f_{convolved}$ 值作卷积, 可得到 $k \times (\gamma - a + 1) \times (c - b + 1)$ 个卷积后的特征矩阵^[19]。

通过卷积过程可提取本地原始图像关键信息, 然后通过训练过程将该信息传递到下一层。使用卷积函数是因图像具有“静态”属性, 因此需要统计汇总不同位置的特征。将不同地点特征(例如某一区域的平均或最大值)的统计数据汇总起来再进行下一步操作, 使计算出来的特征不仅可以代表原来的高维度特征, 而且结果不易过拟合。下采样(池化)^[20]不仅具有上述功能, 其图像变换还具有特征不变性。因此, 池化后的特征集合可以使后期分类更加容易, 图 3 为池化原理图。

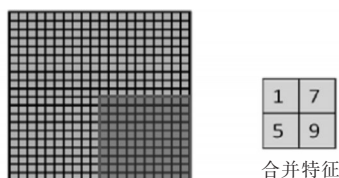


图 3 池化原理

3.3 循环神经网络算法

循环神经网络(Recurrent Neural Network, RNN)种类繁多, 受篇幅所限, 本文只考虑最基本的循环神经网络, 全连接网络如图 4 所示。

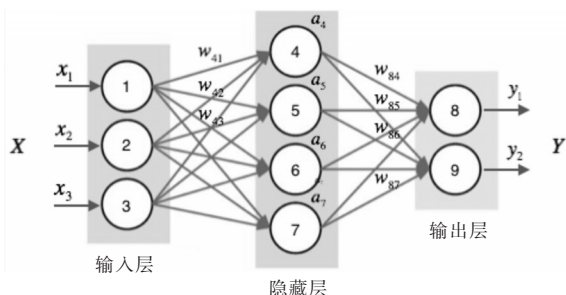


图 4 全连接网络

隐藏层的值只取决于输入的 x , 而 RNN 隐藏层的值 s 不仅取决于当前的输入 x , 还取决于隐藏层上一次的值, 如图 5 所示。其中, t 是时刻, x 是输入层, s 是隐藏层, o 是输出层, 矩阵 W 即为隐藏层上一次的输入值作为这一次的权重。

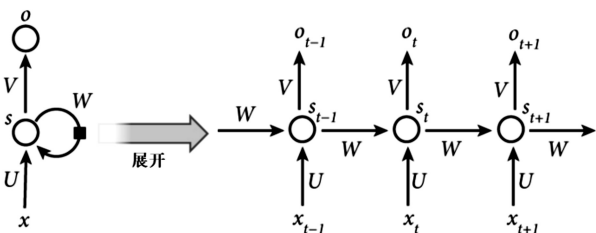


图 5 基本循环神经网络

RNN^[22]输出层 o 和隐藏层 s 计算方法如下:

$$o_t = g(Vs_t) \quad (1)$$

$$S_t = f(Ux_t + Ws_{t-1}) \quad (2)$$

将数据代入式(1)、式(2)后得到:

$$\begin{aligned} o_t &= g(Vs_t) = Vf(Ux_t + Ws_{t-1}) = Vf(Ux_t + Wf(Ux_{t-1} + Ws_{t-2})) \\ &= Vf(Ux_t + Wf(Ux_{t-1} + Wf(Ux_{t-2} + Ws_{t-3}))) \\ &= Vf(Ux_t + Wf(Ux_{t-1} + Wf(Ux_{t-2} + Wf(Ux_{t-3} + \dots)))) \quad (3) \end{aligned}$$

由此可知, 在循环神经网络中可以看到前面所有层的输入值^[21], 从而通过循环网络让信息具有记忆持久性。

3.4 长短期记忆网络算法

长短期记忆网络算法(Long Short Term Memory Network, LSTM)是一种特殊类型的 RNN, 也是一种时间递归神经网络, 多用于处理复杂且需要长期学习的数据, 如图 6 所示。

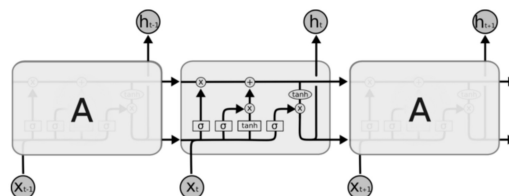


图 6 长短期记忆网络

以图 6 中输入为 h_{t-1} 与 x_t 的模块为例, 其具有比 RNN 更复杂的 4 层重复结构, 需要经过以下 3 个步骤的处理: 首先通过遗忘门控制经过的信息是否为需要的信息, 如果是, 则使全部或需要的部分信息通过, 如果不是, 则进入网络中下一模块继续训练; 然后剔除一部分不需要的旧信息, 输入一部分新信息取而代之, 更新其状态后, 遗忘无用信息, 补充新信息; 最后, 过滤完成, 根据具体需求输出信息。

以上步骤看起来简单, 但可以解决长期存在的复发神经网络问题, 使 LSTM 能够处理不断变化的垂直问题。

3.5 手写数字数据库 MNIST 数据集

MNIST(Mixed National Institute of Standards and Technology Database)是 Python 中提供的一个数据集^[23], 其中包含 70 000 张已经过预处理的手写数字灰度图片, 每张图片都由 28×28 个像素点组成, 数字图片示例如图 7 所示。



图 7 数据集内 0~9 单个数字样本图片

MNIST 数据集包含 4 部分:

Training set images: train-images-idx3-ubyte.gz (包含 60 000 个样本)

Training set labels: train-labels-idx1-ubyte.gz (包含

60 000 个标签)

Test set images: t10k-images-idx3-ubyte.gz(包含 10 000 个样本)

Test set labels: t10k-labels-idx1-ubyte.gz(包含 10 000 个标签)

其训练原理如下:利用深度学习算法进行训练,训练结束后对结果进行匹配,一个样本对应一个标签进行学习。训练完毕后使用测试集内的图片进行测试,识别当前的测试图片得出一个值,并与其对应标签进行匹配。如果该值与对应标签相符合,则此次识别准确,反之,错误次数叠加一次。

4 数字图像识别实现

4.1 基于 Python 的数字图像识别

实验硬件配置如下:实验环境为 Microsoft Windows 10 Enterprise_64bit+Python 3.6.5+TensorFlow 1.8.0;CPU(Central Processing Unit,中央处理器)为 Intel Xeon E3-1231 v3;GPU(Graphics Processing Unit,图形处理器)为 GTX 970;RAM(Random Access Memory,随机存储器)为 16G。

本节使用 Python 实现数字图像识别功能,分别采用线性感知器算法、CNN 算法、RNN 算法以及 LSTM 算法,并附上相应的实验结果截图。

4.1.1 线性感知器算法实现

对 1 000 个样本进行测试,每 10 个数据进行一次准确率输出。其中某次线性感知器算法实现过程截图如图 8 所示。

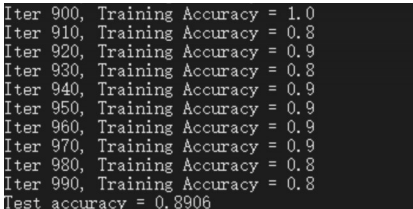


图 8 线性感知器算法实现过程

线性感知器算法实现结果如表 1 所示,平均准确率为 0.892 2。

表 1 线性感知器算法实现结果					
实验序号	1	2	3	4	5
准确率	0.892 8	0.890 6	0.892 1	0.888 4	0.897 1
平均准确率	0.892 2(保留 4 位有效数字)				

4.1.2 CNN 算法实现

利用 CNN 算法进行 5 次迭代,在每次迭代中输出 20 次准确率结果,某次算法实现过程截图如图 9 所示。

CNN 算法实现结果如表 2 所示,平均准确率为 0.896 9。

4.1.3 RNN 算法实现

以 10 为维度,对 2 000 个样本进行训练与测试,通过两层循环层训练得出结果。某次 RNN 算法实现过程截图如图 10 所示。



图 9 CNN 算法实现过程

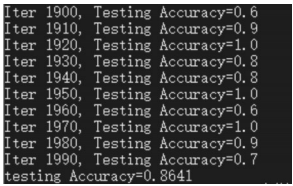


图 10 RNN 算法实现过程

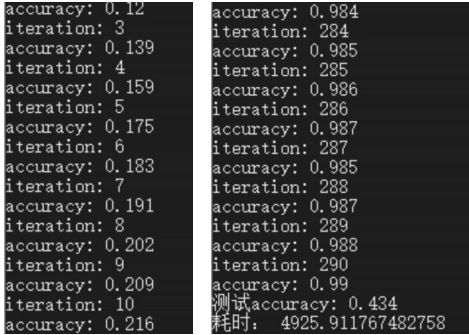
表 2 CNN 算法实现结果					
实验序号	1	2	3	4	5
准确率	0.896 3	0.897 4	0.899 4	0.893 7	0.897 7
平均准确率	0.896 9(保留 4 位有效数字)				

RNN 算法实现结果如表 3 所示,测试平均准确率为 0.432 7。根据测试结果,发现精确率是逐渐上升,但在第 5 次实验时却大幅下降。这是因为在反向传播时出现梯度消失现象,RNN 算法无法更新参数,导致长时记忆效果消失,所以准确率较低。

表 3 RNN 算法实现结果					
实验序号	1	2	3	4	5
准确率	0.110 1	0.228 4	0.856 4	0.864 1	0.099 1
平均准确率	0.432 7(保留 4 位有效数字)				

4.1.4 LSTM 算法实现

LSTM 算法实现过程截图如图 11 所示。由图可知,通过对算法的不断训练,准确率从训练前期的不到 0.2(见图 11(a)),到训练后期可达到 0.99(见图 11(b))。



(a)训练前期 (b)训练后期

图 11 LSTM 算法实现过程

通过多次迭代测试,LSTM 算法实现结果如表 4 所示。实际上识别平均准确率只有 0.448,训练时长达到 4 382.934s。

表 4 LSTM 算法实现结果					
实验序号	1	2	3	4	5
准确率	0.434	0.456	0.447	0.453	0.448
耗费时长/s	4 925.912	4 206.385	4 093.116	4 396.245	4 293.013
平均准确率	0.448(保留 3 位有效数字)				
平均耗时/s	4 382.934(保留 3 位有效数字)				

4.2 基于 TensorFlow 的数字图像识别

4.2.1 线性感知器算法实现

初始化变量后,程序将自动使用 BP 算法,以确定变量如何影响最小化成本值,并随机获取训练数据中的 50 个数据点开始数据集训练。由图 12 可知,在 TensorFlow 框

架下,使用线性感知器算法进行数字识别准确率可达 0.932 6,远高于基于 Python 的数字识别结果。

```
Please use alternatives such as official mnist/data
2018-05-12 05:39:18.457034: I T:\src\github\tensor
rts instructions that this TensorFlow binary was n
WARNING:tensorflow:From C:\Users\Leon_Ree\AppData\
l\tf_should_use.py:118: initialize_all_variables (
d after 2017-03-02.
Instructions for updating:
Use `tf.global_variables_initializer` instead.
0.9326
```

图 12 线性感知器算法准确率

4.2.2 CNN 算法实现

根据图 13 可明显看出一个系统学习的过程,最初识别准确率仅有 0.102 3,经过重复训练后,识别精度不断上升。采用 1 000 个样本(以 50 个样本为一个单位)训练后算法的识别准确率可达到 0.969 8。

```
0.1023
0.7773
0.8854
0.9093
0.9279
0.9361
0.946
0.9453
0.9527
0.9571
0.9576
0.958
0.9616
0.9636
0.9659
0.9648
0.9672
0.9673
0.9698
0.9717
```

图 13 CNN 算法准确率

在卷积神经网络层数较少的情况下,CNN 算法提取样本特征的能力有限,因此可在不改变网络深度的情况下适当增加卷积图层中的滤镜数量。但对于分类数量较多的识别问题,通过自动网络学习到的样品特征不足以区分所有类别。通过研究证明,相对简单的卷积神经网络能更好地识别手写体。当分类数量相对较少时,卷积神经网络可获得更好的识别效果。

4.2.3 RNN 算法实现

图 14 为 TensorFlow 框架下 RNN 算法训练准确率,图 15 为训练完成后实际测试准确率。针对 1 000 个样本(以 100 个样本为一个单位),得到的平均识别准确率为 0.512 1。

```
Iter 0, Testing Accuracy=0.08
Iter 10, Testing Accuracy=0.07
Iter 20, Testing Accuracy=0.2
Iter 30, Testing Accuracy=0.09
Iter 40, Testing Accuracy=0.26
Iter 50, Testing Accuracy=0.16
Iter 60, Testing Accuracy=0.22
Iter 70, Testing Accuracy=0.25
Iter 80, Testing Accuracy=0.3
Iter 90, Testing Accuracy=0.29
Iter 100, Testing Accuracy=0.31
Iter 110, Testing Accuracy=0.39
Iter 120, Testing Accuracy=0.4
Iter 130, Testing Accuracy=0.31
Iter 140, Testing Accuracy=0.31
Iter 150, Testing Accuracy=0.28
Iter 160, Testing Accuracy=0.32
Iter 170, Testing Accuracy=0.39
Iter 180, Testing Accuracy=0.44
Iter 190, Testing Accuracy=0.45
```

图 14 RNN 算法训练准确率

```
Testing Iter 80, Testing Accuracy=0.46
Testing Iter 81, Testing Accuracy=0.59
Testing Iter 82, Testing Accuracy=0.53
Testing Iter 83, Testing Accuracy=0.54
Testing Iter 84, Testing Accuracy=0.51
Testing Iter 85, Testing Accuracy=0.46
Testing Iter 86, Testing Accuracy=0.46
Testing Iter 87, Testing Accuracy=0.5
Testing Iter 88, Testing Accuracy=0.48
Testing Iter 89, Testing Accuracy=0.47
Testing Iter 90, Testing Accuracy=0.48
Testing Iter 91, Testing Accuracy=0.47
Testing Iter 92, Testing Accuracy=0.53
Testing Iter 93, Testing Accuracy=0.48
Testing Iter 94, Testing Accuracy=0.61
Testing Iter 95, Testing Accuracy=0.51
Testing Iter 96, Testing Accuracy=0.49
Testing Iter 97, Testing Accuracy=0.4
Testing Iter 98, Testing Accuracy=0.45
Testing Iter 99, Testing Accuracy=0.44
```

图 15 RNN 算法测试准确率

4.2.4 LSTM 算法实现

图 16 为 TensorFlow 框架下 LSTM 算法识别准确率。经过重复训练,识别精度不断上升,针对 1000 个样本(以 50 个样本为一个单位),训练后的 LSTM 算法最终识别准确率可达到 0.9。

```
Iter 800, Testing Accuracy=0.9
Iter 810, Testing Accuracy=0.9
Iter 820, Testing Accuracy=1.0
Iter 830, Testing Accuracy=0.7
Iter 840, Testing Accuracy=0.9
Iter 850, Testing Accuracy=1.0
Iter 860, Testing Accuracy=0.9
Iter 870, Testing Accuracy=0.8
Iter 880, Testing Accuracy=0.9
Iter 890, Testing Accuracy=0.9
Iter 900, Testing Accuracy=0.9
Iter 910, Testing Accuracy=0.9
Iter 920, Testing Accuracy=0.9
Iter 930, Testing Accuracy=0.9
Iter 940, Testing Accuracy=0.9
Iter 950, Testing Accuracy=0.8
Iter 960, Testing Accuracy=0.6
Iter 970, Testing Accuracy=0.9
Iter 980, Testing Accuracy=1.0
Iter 990, Testing Accuracy=0.9
accuracy: 0.9
```

图 16 LSTM 算法识别准确率

5 结语

研究手写数字识别对于人机交互发展十分重要,本文以线性感知器、卷积神经网络等几种深度学习常见算法为研究对象,分析其在手写数字识别方面的优缺点,并引入 Google 第二代人工智能系统 TensorFlow,对比相同算法在不同框架下的识别速度及准确率。实验结果表明,几类深度学习算法都能明显提高图像识别准确率,又不会过多占用计算资源。未来可综合心理学及人体工程学等相关学科知识进一步提升算法性能。

参考文献:

- [1] 官思发,朝乐门.大数据时代信息分析的关键问题、挑战与对策[J].图书情报工作,2015,59(3):12-18,34.
- [2] 王洪波.单分类支持向量机的学习方法研究[D].杭州:浙江大学,2012.
- [3] 柳长源.相关向量机多分类算法的研究与应用[D].哈尔滨:哈尔滨工程大学,2013.
- [4] 王奕森,夏树涛.集成学习之随机森林算法综述[J].信息通信技术,2018,12(1):49-55.
- [5] 申彦.大规模数据集高效数据挖掘算法研究[D].镇江:江苏大学,2013.
- [6] 嵩天,黄天羽,礼欣.Python 语言:程序设计课程教学改革理想选择[J].中国大学教学,2016(2):42-47.
- [7] 张棣,曹健.面向大数据分析的决策树算法[J].计算机科学,2016,43(S1):374-379,383.
- [8] 孙东联,王宗皓.介绍一种新的客观分析法——递归滤波法[J].气象,1993(2):49-53.
- [9] 李学汇.二义有向图及其路径遍历算法[J].武汉化工学院学报,1997(3):52-55.
- [10] 曹正凤.随机森林算法优化研究[D].北京:首都经济贸易大学,2014.
- [11] 余玉梅,陈克林,白鸿.基于 Rough Set 的 CLS 算法研究[J].云南民族大学学报(自然科学版),2004(1):65-67.
- [12] 张亮,宁芊.CART 决策树的两种改进及应用[J].计算机工程与设计,2015,36(5):1209-1213.
- [13] 汪小龙,袁志发,郭满才,等.最大信息熵原理与群体遗传平衡[J].遗传学报,2002(6):562-564.
- [14] 李俭川,秦国军,温熙森,等.神经网络学习算法的过拟合问题及解决方法[J].振动、测试与诊断,2002(4):16-20,76.
- [15] 龙强.一种求解带等式约束非线性规划问题全局最优解的方法[J].重庆文理学院学报(自然科学版),2010,29(1):20-23.
- [16] 孔英会,景美丽.基于混淆矩阵和集成学习的分类方法研究[J].计算机工程与科学,2012,34(6):111-117.

(责任编辑:黄健)