

基于改进激活函数的用于台风等级分类的深度学习模型

郑宗生 刘兆荣 黄冬梅 宋巍 邹国良 侯倩 郝剑波

(上海海洋大学信息学院 上海 201306)

摘要 针对特定任务中深度学习模型的激活函数不易选取的问题,在分析传统激活函数和现阶段运用比较广泛的激活函数的优缺点的基础上,将 Tanh 激活函数与广泛使用的 ReLU 激活函数相结合,构造了一种能够弥补 Tanh 函数和 ReLU 函数缺点的激活函数 T-ReLU。通过构建台风等级分类的深度学习模型 Typ-CNNs,将日本气象厅发布的台风卫星云图作为自建样本数据集,采用几种不同的激活函数进行对比实验,结果显示使用 T-ReLU 函数得到的台风等级分类的测试精度比使用 ReLU 激活函数的测试精度高出 1.124%,比使用 Tanh 函数的测试精度高出 2.102%;为了进一步验证结果的可靠性,采用 MNIST 通用数据集进行激活函数的对比实验,最终使用 T-ReLU 函数得到 99.855% 的训练精度和 98.620% 的测试精度,其优于其他激活函数的效果。

关键词 深度学习,卷积神经网络,激活函数,台风等级,MNIST 数据集

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.12.028

Deep Learning Model for Typhoon Grade Classification Based on Improved Activation Function

ZHENG Zong-sheng LIU Zhao-rong HUANG Dong-mei SONG Wei ZOU Guo-liang HOU Qian HAO Jian-bo

(College of Information Technology, Shanghai Ocean University, Shanghai 201306, China)

Abstract Aiming at the issue that it is difficult to select the activation function in deep learning model for specific task, on the basis of analyzing the advantages and disadvantages of traditional activation function and the popular activation function at the present stage, this paper constructed an activation function T-ReLU which can make up for the shortcomings of Tanh function and ReLU function by combining the Tanh activation function with the widely used ReLU function. By constructing the deep learning model Typ-CNNs for typhoon grade classification, using the Typhoon satellite image published by the Japan Meteorological Agency as the self-built sample data, this paper made use of several different activation functions to conduct comparison experiments. The results show that the test accuracy of typhoon grade classification using the T-ReLU function is 1.124% higher than that of using ReLU activation function, which is 2.102% higher than that of using Tanh function. In order to further verify the reliability of the results, the MNIST general data set was utilized to carry out the comparison experiment of activation function. The final results show that 99.855% training accuracy and 98.620% test accuracy can be obtained by using T-ReLU function, and it performs better than other activation functions.

Keywords Deep learning, Convolution neural network, Activation function, Typhoon grade, MNIST dataset

1 引言

近年来,深度学习在图像识别与分类、视频处理、自然语义等研究领域得到了广泛应用^[1]。深度学习方法来源于人工神经网络,有一定的生物理论基础,是模拟动物大脑神经来逐层学习事物特征,进而得到事物特征的抽象、多级表达的学习方法^[2-3]。本质上,深度学习就是多层的非线性网络结构算

法,能够学习目标中多层复杂的非线性函数关系,从而使得人们能够更好地对目标的抽象信息进行建模分析。国内外研究者在深度学习模型的构建、参数的优化、激活函数的使用、优化器的选择、误差函数的选取以及过拟合等方面进行了大量研究^[4],例如,文献[5]构建了 CNN+SVM 模型结构,在车辆识别中达到 97.00% 的高准确率,优于传统 CNN 模型;文献[6]研究了多标签损失函数,构建了 CNN 模型,实现了对

到稿日期:2017-11-23 返修日期:2018-02-16 本文受国家自然科学基金项目:基于多模态深度学习的弱特征多源海洋遥感影像协同分类模型研究(41671431),上海市科委地方院校能力建设项目:基于海洋视频时空交叉分析的近岸灾害性海浪预测研究及其应用(17050501900)资助。

郑宗生(1979-),男,博士,副教授,主要研究方向为海洋信息化、深度学习应用;刘兆荣(1992-),女,硕士生,主要研究方向为深度学习应用, E-mail:2415932685@qq.com(通信作者);黄冬梅(1964-),女,博士,教授,主要研究方向为计算机应用研究;宋巍(1977-),女,博士,教授,主要研究方向为移动多媒体、人机交互、海洋遥感图像分析、海洋大数据认知和理解等;邹国良(1961-),男,博士,教授,主要研究方向为海洋信息处理及应用;侯倩(1992-),女,硕士生,主要研究方向为深度学习;郝剑波(1992-),男,硕士生,主要研究方向为海洋信息化, E-mail: lzl.liya2018@outlook.com。

图像的自动标注;文献[7]采用结构风险最小化的损失函数,提出了一种信息融合算法来对遥感图像进行融合,获得了良好的效果。针对特定任务,如何使模型达到最佳性能,激活函数的选用也一直是研究人员关注的重点^[8-10]。

在深度学习中,传统的激活函数是非线性饱和激活函数,如 Sigmoid, Tanh^[11]等。神经科学证明,人脑在感知视觉信号时各个神经元是稀疏响应的^[12]。同时,文献[13]中的研究也表明,同一时刻大脑神经元激活率仅有 1%~4%,与文献[12]中的相应观点相符。文献[10]从统计学的角度解释了传统的饱和非线性函数能同时激活近 50%的神经元。传统激活函数由于其两端饱和性出现的梯度消失问题,会造成神经网络无法继续学习的现象^[14]。

目前,比较成熟的激活函数是修正线性单元 ReLU(Rectified Linear Units)函数^[15-17],ReLU 函数以其计算简单和稀疏特性成功代替了传统的饱和非线性函数。ReLU 函数以其单边抑制和稀疏激活的特点,有效地解决了梯度消失问题,并且更符合生物神经活动的特性。在 ReLU 函数的作用下,当神经元接收到负信号时,神经元处于抑制状态;当接收到正信号时,神经元被激活,并且能够保持梯度不被衰减。Krizhevsky 等^[18]对 ReLU 函数和传统的 Sigmoid 函数做了对比测试,在使用 ReLU 函数时,能够加快模型的收敛速度、缩短模型的训练周期。但是,ReLU 函数会使神经网络出现“神经元坏死”现象。出现该情况的原因是,一般情况下神经网络使用梯度下降的优化算法,这种算法对最初没有被激活的神经元的权值不起作用,而 ReLU 函数使某些处于非激活状态的神经元梯度为 0,从而使这个神经元出现永不被激活的现象。这种现象使 ReLU 函数表现出强制性的稀疏特性,而且这种强烈的稀疏处理会使学习到的有效信息大大缩减,从而降低模型的实用性。

针对“神经元坏死”问题,很多基于 ReLU 函数的改进函数相继出现,如 Hannun 等提出了 Leaky ReLU(LReLU)函数^[19],通过增加固定的微小负梯度改进了 ReLU 函数,有效地避免了“神经元坏死”的情况;Szegedy 等也提出了对 ReLU 函数的改进函数 PReLU^[20],即携带参数的 ReLU 函数,该函数与 LReLU 的思想相近,区别在于将 LReLU 中的固定负梯度改为由训练得到的负梯度;He 等指出,使用 PReLU 函数能将训练效果提高 1.03%^[21]。RReLU(Randomized Leaky ReLU)使模型在训练过程中,将 LReLU 函数中固定的负梯度值用从均匀分布中随机抽取的随机值代替,而在测试阶段使用这些负梯度值的均值^[19]。文献[22]在标准图像分类任务中对 ReLU, LReLU, PReLU 和 RReLU 进行了测试,得出在 ReLU 函数中加入负零部分的非零导数可以提高模型的性能。但是 LReLU 中的系数因子需要先验知识赋值,若选取不当,则对准确率的影响不大;PReLU 和 RReLU 在网络中实现困难,计算复杂。

本文通过对不同激活函数在神经网络中表现出的优缺点进行分析,将 Tanh 函数和 ReLU 函数进行融合,提出一种新的改进激活函数,使其同时具有左侧软饱和性、右侧线性激活及稀疏特性,并将此激活函数用于台风卫星云图自建数据集

和手写数字库 MNIST 通用数据集,通过在卷积神经网络(Convolutional Neural Networks, CNNs)框架下的对比实验,进一步证明了改进的激活函数的有效性。

2 改进的激活函数

2.1 网络残差的理论推导

在神经网络中,每一层网络所学习到的特征表现为权值的形式^[23]。CNNs 网络在训练中通常使用反向传播算法(Back Propagation Algorithm, BP)来更新网络权值,通过将网络中的残差逐层向底层网络反向传播进行网络权值的调整^[24]。残差是网络实际输出 a_N 与期望输出 y 之间的误差信号,是损失函数 $L(W, b)$ 对第 N 层网络输入 z_N 的偏导。假设神经网络输出层的残差为:

$$\delta_N = \frac{\partial L(W, b)}{\partial z_N} = \nabla_{a_N} L \cdot \sigma'(z_N) \quad (1)$$

其中, N 为神经网络的深度, $\sigma'(z_N)$ 为第 N 层网络的激活函数的导数。则隐层网络中第 n 层的残差计算为:

$$\delta_n = (W_{n, n-1} \delta_{n+1}) \sigma'(z_n) \quad (2)$$

其中, $W_{n, n-1}$ 表示第 n 层和第 $n-1$ 层的连接权值, δ_{n+1} 是第 $n+1$ 层的残差, $\sigma'(z_n)$ 是第 n 层网络的激活函数导数。在残差的反传中,经过激活函数的导数 $\sigma'(z_n)$ 和权矩阵 $W_{n, n-1}$ 作用后得到下一个残差 δ_n 。残差作用于网络权值的更新中,第 n 层网络的权值修改量为:

$$\frac{\partial L(W, b)}{\partial W_{n, n-1}} = \delta_n \cdot \frac{\partial z_n}{\partial W_{n, n-1}} = \delta_n \cdot a_{n-1} \quad (3)$$

其中, a_{n-1} 为第 $n-1$ 层网络的实际输出。

权值更新迭代为:

$$W_{n, n-1} = W_{n, n-1} - \eta \frac{\partial L(W, b)}{\partial W_{n, n-1}} \quad (4)$$

其中, η 为学习率, b 为偏移常量。网络的权值是由多次迭代更新得到的,当网络达到最大迭代次数或者得到最小损失值(即随着迭代的进行损失值不再变化)时停止迭代。

2.2 改进的 T-ReLU 函数

对于常见的激活函数 Tanh 和 ReLU,有:

$$\lim_{x \rightarrow \infty} \text{Tanh}'(x) = 0 \text{ 且 } \text{Tanh}'(x) \leq 1 \quad (5)$$

$$\text{ReLU}'(x) = 0 (x \leq 0) + 1 (x > 0) \quad (6)$$

式(5)中,仅当输入为 0 时取“=”。由式(2)、式(5)可知, x 过大或过小都会使 $\text{Tanh}'(x)$ 趋于 0,此时梯度逐渐消失,导致残差 δ 趋于 0 而无法继续向下传播,使网络的学习能力大大降低, Sigmoid 函数与 Tanh 函数的情况类似。对于 ReLU 函数,一方面会出现“神经元坏死”现象;另一方面,由式(2)、式(6)可知,在输入 $x < 0$ 时, $\text{ReLU}'(x) = 0$,此时残差 $\delta = 0$ 使得网络权值得不到更新,并且负值信息完全无法参与训练,造成过于强烈的稀疏性。

针对以上问题,将 ReLU 函数激活区域的线性特点和 Tanh 函数非激活区域的软饱和性进行融合,提出的改进的激活函数为 T-ReLU,其表达式为:

$$\text{T-ReLU}(x) = \begin{cases} x, & x \geq 0 \\ k \cdot \left(\frac{2}{1 + e^{-2x}} - 1 \right), & x < 0 \end{cases} \quad (7)$$

其中, x 表示非线性激活函数 T-ReLU 的输入; k 为常量系

数,取值范围为 $[0.1, 0.5]$ 。T-ReLU 函数图像和其导函数图像如图 1 所示。

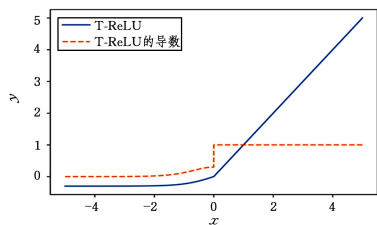


图 1 T-ReLU 激活函数

Fig. 1 T-ReLU activation function

由图 1 可以看出, T-ReLU 函数具有左侧软饱和性和右侧线性的特点, 这使 T-ReLU 函数具备了修正 ReLU 函数和 Tanh 函数缺点的优势。

从残差角度来看, T-ReLU 函数使得神经元接收到正值信号($x > 0$)时处于兴奋状态, 并且 $T-ReLU'(x) = 1$, 即梯度恒定不变, 如图 1 所示。此时:

$$\delta_n = (W_{n,n-1} \delta_{n+1}) \quad (8)$$

即每一个残差值都只由前一个残差和权矩阵决定, 不会因激活函数导数的影响而衰弱, 保证残差反传能够有效地向底层网络传播; 当神经元接收到负值信号($x < 0$)时, T-ReLU 函数发挥左侧的软饱和特性, 保证非激活部分存在连续变化的非零梯度, 即 $T-ReLU'(x) \neq 0 (x < 0)$, 残差值仍然只由前一个残差和权矩阵求得, 并且对部分网络权值进行调整, 即部分负值信息参与到权值更新中。这个过程相当于在网络中加入了少量的噪声, 使得函数不会像 ReLU 激活函数一样强制地增加神经网络稀疏特性, 能够避免丢失过多的信息, 从而提高网络学习能力并且增加网络的鲁棒性。

3 实验

3.1 Typ-CNNs 模型

在深度学习中, 传统的 CNNs 网络使用的是有监督的学习方式, 其拓扑结构是由卷积层和采样层交替使用构成隐藏层^[25], 最后由全连接层得出分类结果, 是 Fikushima 等学者提出的具有高准确率图像识别模型。

本文构建的深度学习模型为 Typ-CNNs, 相对于传统的 CNN 模型, 该模型采用卷积核分解策略将一个大的卷积核分解为两个小的卷积核, 即隐藏层的拓扑结构为两个卷积层和一个采样层的交替使用。通过减小单个卷积核尺寸, 增加网络层数, 从而降低卷积过程的计算量并增加网络模型深度, 以提高模型性能。Typ-CNNs 全模型结构如图 2 所示。

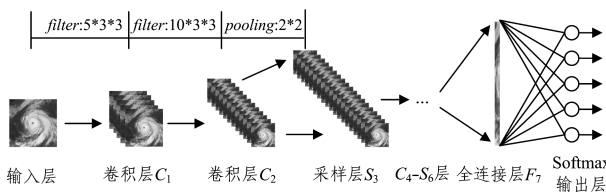


图 2 Typ-CNNs 全模型结构

Fig. 2 Full model structure of Typ-CNNs

图中包括 4 层卷积层($C_1 - C_2, C_4 - C_5$)、2 层采样层(S_3 ,

S_6)、1 层全连接层(F_7)和输出层。输出层采用 Softmax 分类函数, 其输出结果为属于每个类别的概率。

3.2 模型的评价指标

本文通过分类准确率来评判模型性能的好坏。在多分类问题中, 预测得到的类标签与真正的标签集合越吻合, 分类准确率越接近 1。对于长度为 n 的样本集合中的第 i 个样本, 若其真实的类标签是 y_i , 预测的类标签是 o_i , 则分类准确率为:

$$acc(y_i, o_i) = \frac{1}{n} \sum_{i=0}^n 1(y_i = o_i) \quad (9)$$

由式(9)可以看出, 分类准确率是模型性能的外在表征, $acc(y_i, o_i)$ 值越高, 模型性能越好。

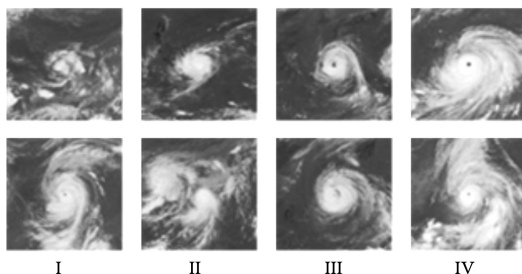
4 实验结果及分析

本文所有对比实验均基于 Debian Linux 操作系统, CPU Intel core i5-4210M 2.60 GHz, 内存 4 GB, 软件采用 python 2.7.9 版本, 使用基于 Tensorflow 的 Keras 深度学习框架搭建 CNN 模型。本文实验分为两组, 第一组实验采用 Typ-CNNs 模型, 训练数据使用自建数据集——预处理后的台风卫星云图, 使用不同的常见激活函数和本文提出的改进的激活函数进行对比实验, 其中 k 的取值为 0.2; 在第二组实验中, 为验证 T-ReLU 函数的普适性, 数据集采用手写数字库中的通用数据集 MNIST 数据作为实验数据, 其他设置与第一组实验相同。

4.1 基于台风云图数据集的对比实验

4.1.1 数据集构建

本文实验自建数据集来自于日本气象厅发布的 1981 年至 2016 年间由静止气象卫星 1~5 (Geostationary Meteorological Satellites, GMS)、GOE9、MTSAT-1R、MTSAT-2、HIMAWARI-8 在北西太平洋地域用红外光拍摄的气象卫星图像, 部分台风样本如图 3 所示。将其发布的分类等级作为等级数据训练标签, 以日本气象厅的台风等级为标准, 按照台风中心不同风速分为热带低气压、台风、强台风、非常强台风, 将其分别记为 1~4 个等级(见表 1)。共采集了 4800 张 128×128 的图像作为实验数据。先用最近邻插值算法对采集来的数据图像进行缩放处理以降低实验训练周期成本, 得到大小为 28×28 的样本数据; 再对得到的图像进行去均值归一化处理以降低冗余信息; 最后随机分为 3 个数据样本集, 其中 800 张作为测试集, 剩余 4000 张作为训练集, 训练数据集的 20% 作为十折交叉验证集进行分类实验。



注: I, II, III, IIII 分别代表 4 个等级的台风

图 3 自建数据集部分样本

Fig. 3 Partial samples of self-built dataset

表1 日本气象厅发布的台风等级

Table 1 Typhoon level released by Japan Meteorological Agency

日本气象厅 分类	最大风速(10 min 平均值)			国际分类
	kt	m/s	km/h	
热带气旋	~33	~17	~62	热带低压(TD)
台风	34~47	18~24	63~88	热带风暴(TS)
	48~63	25~32	89~118	强烈热带风暴(STS)
强台风	64~84	33~43	119~156	台风(TY)或飓风
非常强台风	85~104	45~54	157~192	

4.1.2 结果分析

第一组实验中,将本文提出的 T-ReLU 激活函数与其他几个激活函数应用到 Typ-CNNs 模型中做对比实验,准确率如表 2 所列。

表2 自建数据集上不同激活函数作用下的模型准确率

Table 2 Model accuracy under different activation functions on Typhoon dataset

激活函数	(单位:%)	
	验证准确率	测试准确率
Tanh	83.572	83.489
ReLU	84.216	84.067
Elu	85.137	83.543
T-ReLU	85.574	85.591

不同激活函数使得模型在实验数据集上的表现不同。Tanh 函数在测试集和验证集上的准确率分别为 83.572% 和 83.489%;ReLU 函数在测试集上的精度为 84.067%;Elu 函数在测试集上的准确率为 83.543%,比验证准确率低 1.594%,说明模型可靠性不高;改进后的 T-ReLU 函数在验证集和测试集上的准确率分别为 85.574% 和 85.591%,表现基本一致,且高于其他函数。Elu 函数也是在激活函数抑制区添加了一个非零梯度,使负值信息参与网络的训练,但是其在抑制区域内将所有的负值全部压缩在 $-1 \sim 0$ 之间,而这个范围内的 Elu 函数的导数接近 1,几乎能够影响网络每层的权值更新,即所有负值信息参与了整个网络的训练,而 T-ReLU 函数将负值信息压缩到 $-0.2 \sim 0$ 之间,此区域内函数的导数远小于 1,能够保证负值信息只参与少数网络权值的更新,因此理论上 T-ReLU 的效果优于 Elu 函数。训练集上的准确率的变化曲线如图 4 所示,随着训练迭代次数的增加,准确率不断提高,当网络达到稳定状态后,准确率基本不再提升。

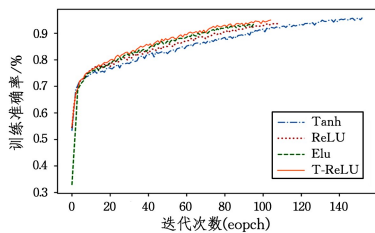


图4 不同激活函数在训练集上的准确率变化曲线

Fig. 4 Accuracy curves of different activation functions on training set

模型训练稳定后,使用 T-ReLU 激活函数的训练准确率为 94.157%,高于使用 ReLU 和 Elu 激活函数时的训练精度;使用 T-ReLU 激活函数的训练准确率低于使用 Tanh 激活函数的训练准确率,但网络达到稳定状态时所需要的迭代次数

远少于 Tanh 函数,这就说明模型训练时间被大大缩短,网络训练的效率和得到了有效提高。

4.2 基于 MNIST 数据集的对比实验

本节将通过 MNIST 数据对比实验的方式验证改进激活函数的有效性。将 MNIST 数据集作为对比实验的数据,共 42000 个 28×28 维的阿拉伯数字手写样本,其中 10000 个样本作为测试集,剩余 32000 个样本作为训练集,将训练数据集的 20% 作为十折交叉验证集进行分类实验,通用 MNIST 数据集的部分样本数据如图 5 所示。

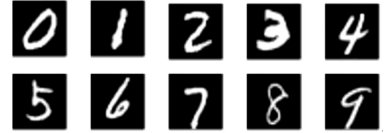


图5 MNIST 通用数据集的部分样本

Fig. 5 Partial samples of MNIST common dataset

本实验使用与第一组实验相同的 Typ-CNNs 网络结构,运用不同的激活函数进行实验对比。将 MNIST 图片数据转化为 784 维的向量输入到 CNN 网络中进行训练,设置最大循环次数为 200,当训练精度不再变化或循环达到最大次数时停止训练,再对实验结果进行分析。

验证数据集上的准确率变化曲线如图 6 所示。使用 T-ReLU 函数时的训练效果明显优于 Elu 及 Tanh,且训练效果较 ReLU 更为稳定。

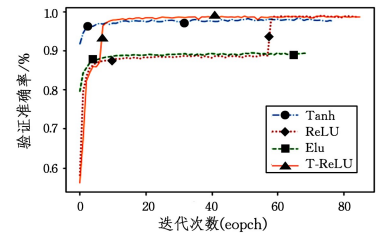


图6 MNIST 数据集上的验证集准确率变化曲线

Fig. 6 Verification set accuracy curve on MNIST dataset

验证数据集、测试数据集及训练数据集上的最终准确率如表 3 所列,可以看出 T-ReLU 在 MNIST 上的分类准确率在验证集、测试集、训练集上分别达到 98.609%,98.620%,99.855%,表现能力优于其他激活函数。

表3 不同激活函数在 MNIST 通用数据集上的模型准确率

Table 3 Model accuracy of different activation functions on MNIST common datasets

激活函数	(单位:%)		
	验证准确率	测试准确率	训练准确率
Tanh	97.328	97.870	99.441
ReLU	98.515	98.580	99.804
Elu	89.234	88.380	89.199
T-ReLU	98.609	98.620	99.855

4.3 激活函数不同组合的对比实验

融合两种激活函数的方式众多,不同的函数组合对模型性能的影响各有差异,本文改进的激活函数就是对 Tanh 函数和 ReLU 函数的优点进行了融合。为了增强这种结合方式的理论依据,本节构造了另一种激活函数——R-T 函数,将

ReLU 函数的左侧与 Tanh 函数的右侧相结合,分别在自建数据集和通用数据集上进行实验对比,结果如表 4 所列。

表 4 R-T 激活函数在两种数据集上的模型准确率

Table 4 Model accuracy of R-T activation function on two datasets

(单位:%)

数据集	验证准确率	测试准确率	训练准确率
MNIST 数据集	97.234	97.609	99.421
自建数据集	64.875	62.500	89.719

由表 4 可以看出,在特征明显的 MNIST 数据集上,R-T 激活函数对模型的影响较小,测试集上的准确率仍能达到 97.609%,仅比使用 T-ReLU 函数时低 0.261%;但在特征复杂度较高的自建数据集上测试准确率只有 62.500%,训练集上的准确率达到了 89.719%,精度远低于 T-ReLU 函数的作用效果;图 7 直观地反映出 R-T 函数使网络模型出现了严重的过拟合现象,并且准确率变化曲线出现了明显的震荡,可见网络模型不稳定。通过对比可知,在特征复杂度较高的样本数据上,T-ReLU 函数的表现能力仍然很强,由此证明了 T-ReLU 函数的结合方式的可行性和可靠性。

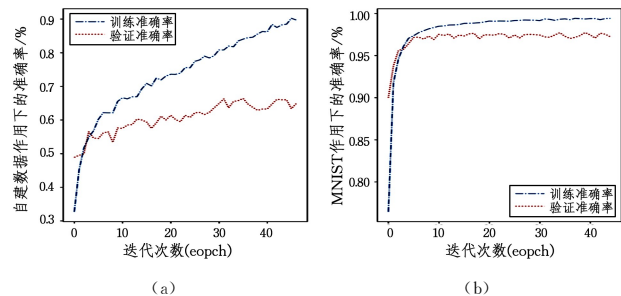


图 7 两种数据集上的训练准确率和测试准确率

Fig. 7 Training accuracy and test accuracy on two datasets

结束语 在不同领域的不同任务中没有统一的深度学习模型,需要不同的模型参数和结构。深度学习模型构建的难度之一是激活函数的选取和模型结构的搭建,不同激活函数和模型的拓展结构关乎着任务完成的准确度和效率。本文针对常用激活函数在神经网络中存在的问题,提出了改进的激活函数;基于台风卫星云图数据构建 Typ-CNNs 分类模型,并将改进后的激活函数应用于该模型中,取得了较好的结果;通过 MNIST 数据集验证了 T-ReLU 激活函数的有效性,达到了预期效果;最后通过两种函数不同的结合方式构造出 R-T 函数进行对比实验,证明了本文改进的 T-ReLU 激活函数能够适用于特征高度复杂的台风样本数据,也进一步证明了构造 T-ReLU 函数的可行性和可靠性。通过本文的研究,为今后研究学者提供了一种新思路,同时为台风等级预测提供了一种新方法。

对于 CNN 分类模型而言,模型泛化能力和数据集构建是两个非常重要的关注点。本文虽然取得了一些成就,但是还存在一些不足:

(1)针对深度学习模型普遍存在的泛化能力弱的问题,本文改进的激活函数仅在自建数据集和 MNIST 数据集上进行了实验,因此本文构建的 Typ-CNN 模型也有可能存在同样的问题,有待进一步研究。

(2)由于在实际情况中,台风卫星云图受很多物理因素的影响,图像特征更为复杂,数据规模巨大,本文自建的数据集可能未包含台风卫星云图图中所隐含的全部特征。因此,下一步将扩增数据集规模,为满足实际应用做进一步研究。

参考文献

- [1] GUO L L, DING S F. Research Progress in Deep Learning [J]. Computer Science, 2015, 42(5): 28-33. (in Chinese)
郭丽丽, 丁世飞. 深度学习研究进展 [J]. 计算机科学, 2015, 42(5): 28-33.
- [2] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553): 436-444.
- [3] SCHULZ H, BEHNKE S. Deep Learning [J]. KI -Künstliche Intelligenz, 2012, 26(4): 357-363.
- [4] CIRSTEAN B I, LIKFORMANSULEM L. Improving a deep convolutional neural network architecture for character recognition [J]. Electronic Imaging, 2016, 2016(17): 1-7.
- [5] PENG Q, JI G S, XIE L J, et al. Application of Convolution Neural Network in Vehicle Identification [J/OL]. <http://kns.cnki.net/kcms/detail/11.5602.TP.20170807.1008.002.html>. (in Chinese)
彭清, 季桂树, 谢林江, 等. 卷积神经网络在车辆识别中的应用 [J/OL]. <http://kns.cnki.net/kcms/detail/11.5602.TP.20170807.1008.002.html>.
- [6] LI J C, YUANG C, SONG Y. Automatic Labeling of Multi-label Images Based on Convolutional Neural Network [J]. Computer Science, 2016, 43(7): 41-45. (in Chinese)
黎健成, 袁春, 宋友. 基于卷积神经网络的多标签图像自动标注 [J]. 计算机科学, 2016, 43(7): 41-45.
- [7] LI H, LIU F, YANG S Y, et al. Remote sensing image fusion based on deep supportive value learning network [J]. Acta Automatica Sinica, 2016, 39(8): 1583-1596. (in Chinese)
李红, 刘芳, 杨淑媛, 等. 基于深度支撑值学习网络的遥感图像融合 [J]. 计算机学报, 2016, 39(8): 1583-1596.
- [8] SHAFIE A S, MOHTAR I A, MASROM S, et al. Backpropagation neural network with new improved error function and activation function for classification problem [C]// Humanities, Science and Engineering Research. IEEE, 2012: 1359-1364.
- [9] GONG Z T, CHEN G X, CAO J S. Application of Convolutional Neural Network in Image Classification of Cerebrospinal Fluid [J]. Computer Engineering and Design, 2017, 38(4): 1056-1061. (in Chinese)
龚震霆, 陈光喜, 曹建收. 卷积神经网络在脑脊液图像分类上的应用 [J]. 计算机工程与设计, 2017, 38(4): 1056-1061.
- [10] WANG F F. Research and application of improved convolution neural network algorithm [D]. Nanjing: Nanjing University of Posts and Telecommunications, 2016. (in Chinese)
王飞飞. 基于改进卷积神经网络算法的研究与应用 [D]. 南京: 南京邮电大学, 2016.
- [11] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines [C]// International Conference on International Conference on Machine Learning. Omnipress, 2010: 807-814.

- [5] ZHANG Y N,XIAO Z L,GUO D S. Singularity-conquering tracking control of a class of chaotic systems using Zhang-gradient dynamics[J]. *IEEE Transactions on Control Theory & Applications*,2015,9(6):871-881.
- [6] ROSSI R,SANTAMARIA-NAVARRO A. Trajectory generation for unmanned aerial manipulators through quadratic programming[J]. *IEEE Robotics & Automation Letters*,2016,2(2):389-396.
- [7] ZHANG H. A finite iterative algorithm for solving the complex generalized coupled Sylvester matrix equations by using the linear operators[J]. *Journal of the Franklin Institute*,2017,354(4):1856-1874.
- [8] JIN L,LI S. Distributed task allocation of multiple robots:a control perspective[J]. *IEEE Transactions on Systems, Man, and Cybernetics*,2016,PP(99):1-10.
- [9] MIN K,FREEMAN C,KANG H. The regulation by phenolic compounds of soil organic matter dynamics under a changing environment[J]. *Journal of Biomedicine and Biotechnology*,2015,2015(6849):433-458.
- [10] DUGULEANA M,BARBUCEANU F G. Obstacle avoidance of redundant manipulators using neural networks based reinforcement learning [J]. *Robotics and Computer-Integrated Manufacturing*,2012,28(2):132-146.
- [11] WHITNEY D E. Resolved motion rate control of manipulators and human prostheses[J]. *IEEE Transactions on Man Machine Systems*,1969,10(2):47-53.
- [12] TCHON K,JANIAC M. Repeatable approximation of the Jacobian pseudo-inverse [J]. *Systems and Control Letters*,2009,58(12):849-856.
- [13] CHENG F,CHEN T,SUN Y. Resolving manipulator redundancy under inequality constraints[J]. *IEEE Transactions on Robotics Automation*,1994,10(1):65-71.
- [14] ZHANG Y,LI W. Physical-limits-constrained minimum velocity norm coordinating scheme for wheeled mobile redundant manipulators[J]. *Robotica*,2015,33(2):1325-1350.
- [15] LI S. Accelerating a recurrent neural network to finite-time convergence for solving time-varying sylvester equation by using a sign-bi-power activation function[J]. *Neural Processing Letters*,2013,37(2):189-205.
- [16] LIN X,LIAO B. A convergence-accelerated Zhang neural network and its solution application to Lyapunov equation[J]. *Neurocomputing*,2016,193(2):213-218.
- [17] ZHANG Y, YOU Q X, LUO Y, et al. Robot arm remote control based on trajectory planning in joint space [J]. *Journal of Chongqing University of Posts and Telecommunications(Natural Science Edition)*,2012,24(1):104-108. (in Chinese)
张毅,游群霞,罗元,等. 基于关节空间轨迹规划的机械臂远程控制[J]. *重庆邮电大学学报(自然科学版)*,2012,24(1):104-108.

(上接第 181 页)

- [12] REHN M,SOMMER F T. A network that uses few active neurons to code visual input predicts the diverse shapes of cortical receptive fields [J]. *Journal of Computational Neuroscience*,2007,22(2):135-146.
- [13] LENNIE P. Supplemental Data The Cost of Cortical Computation[J]. *Current Biology*,2003,13(6):493-497.
- [14] HUANG Y,DUAN X S,SUN S Y, et al. Research on training algorithm of deep neural networks based on improved sigmoid activation function [J]. *Computer Measurement and Control*,2017,25(2):126-129. (in Chinese)
黄毅,段修生,孙世宇,等. 基于改进 sigmoid 激活函数的深度神经网络训练算法研究[J]. *计算机测量与控制*,2017,25(2):126-129.
- [15] GLOROT X,BORDES A,BENGIO Y. Deep sparse rectifier neural networks[C]//*International Conference on Artificial Intelligence and Statistics*,2012:315-323.
- [16] JARRETT K,KAVUKCUOGLU K,RANZATO M, et al. What is the Best Multi-Stage Architecture for Object Recognition?[C]//*IEEE International Conference on Computer Vision*,2009:2146-2153.
- [17] OLSHAUSEN B A,FIELD D J. Sparse coding with an overcomplete basis set:a strategy employed by V1[J]. *Vision Research*,1997,37(23):3311.
- [18] KRIZHEVSKY A,SUTSKEVER I,HINTON G E. ImageNet classification with deep convolutional neural networks[C]//*International Conference on Neural Information Processing Systems*. Curran Associates Inc.,2012:1097-1105.
- [19] MAAS A L,QI P,HANNUN A Y, et al. Building DNN acoustic models for large vocabulary speech recognition [J]. *Computer Speech & Language*,2017,41(C):195-213.
- [20] SZEGEDY C,LIU W,JIA Y, et al. Going deeper with convolutions[C]//*IEEE Conference on Computer Vision and Pattern Recognition*. IEEE,2015:1-9.
- [21] HE K,ZHANG X,REN S, et al. Delving Deep into Rectifiers: Surpassing Human—Level Performance on ImageNet Classification[C]//*2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE,2015.
- [22] XU B,WANG N,CHEN T, et al. Empirical Evaluation of Rectified Activations in Convolutional Network[J/OL]. <https://arxiv.org/abs/1505.00853>.
- [23] POGGIO T,GIROSI F. Networks for approximation and learning[J]. *Proceedings of the IEEE*,1990,78(9):1481-1497.
- [24] SU H,LI G,YU D, et al. Error back propagation for sequence training of Context-Dependent Deep Networks for conversational speech transcription[C]//*IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE,2013:6664-6668.
- [25] VEDALDI A,LENC K. MatConvNet:Convolutional Neural Networks for MATLAB[C]//*ACM International Conference on Multimedia*. ACM,2015:689-692.