

深度学习神经网络的新型自适应激活函数

刘宇晴, 王天昊, 徐旭
(吉林大学 数学学院, 长春 130012)

摘要: 构造一个含有参数的光滑激活函数用于深度学习神经网络, 通过基于误差反向传播算法建立参数的在线修正公式, 避免了梯度弥散、不光滑及过拟合等问题. 与一些常用的激活函数进行对比实验结果表明, 新的激活函数在多个数据集上效果均较好.

关键词: 激活函数; 卷积神经网络; 机器学习

中图分类号: O175.1 **文献标志码:** A **文章编号:** 1671-5489(2019)04-0857-03

New Adaptive Activation Function for Deep Learning Neural Networks

LIU Yuqing, WANG Tianhao, XU Xu
(College of Mathematics, Jilin University, Changchun 130012, China)

Abstract: A smooth activation function with a parameter was constructed for the deep learning neural networks. The online correction formula for this parameter was established based on the error back propagation algorithm, which avoided the problems of gradients losing, non-smooth and over-fitting. Compared with some popular activation functions, the results show that the new activation function works well on many data sets.

Keywords: activation function; convolutional neural network; machine learning

近年来, 深度神经网络在计算机视觉和模式识别领域取得了重大突破, 这一方面归功于深度结构与卷积运算等优势, 另一方面是新型激活函数的深入研究. 在早期的激活函数中(如双曲正切函数 Tanh), 当输入非常大或非常小时神经元的梯度即接近于零, 这是早期神经网络无法深度化的原因之一^[1-7]. ReLU(rectified linear unit)激活函数 $f(x) = \max\{0, x\}$ 在输入值远大于零时梯度不饱和, 计算速度得到了明显提升. 但当神经元的输入为负值时, 激活函数的梯度为零, 导致发生“神经元死亡”现象. 为了克服该问题, 研究者们提出了许多激活函数如 PReLU(parametric rectified linear unit), ELU(exponential linear unit), LReLU(Leaky-ReLU)等, 这些函数可用线性(拟线性)函数替代原来的零值^[8], 但选择负值部分的斜率较难, 同时上述大部分函数在零点处非光滑且效率不高^[9-13]. 基于此, 本文提出一种新型的自适应激活函数, 为深度学习模型激活函数提供一种有效的解决方案.

1 新型激活函数

新型激活函数 Xwish 定义如下:

$$f(x) = x \cdot [\arctan(\beta x) + 0.5\pi] / \pi,$$

收稿日期: 2019-01-17. 网络首发日期: 2019-05-06.

第一作者简介: 刘宇晴(1994—), 女, 汉族, 硕士研究生, 从事人工智能理论的研究, E-mail: 964706874@qq.com. **通信作者简介:** 徐旭(1972—), 男, 汉族, 博士, 教授, 从事动力系统与控制的研究, E-mail: xuxu@jlu.edu.cn.

基金项目: 国家自然科学基金(批准号: 11372117; J1310022)和吉林省科技发展计划项目(批准号: 2017C028-1).

网络首发地址: <http://kns.cnki.net/kcms/detail/22.1340.O.20190505.1537.001.html>.

这里 β 是可以修正的参数. 图 1 为 β 取不同值时 Xwish 函数的图像, 在全体实数范围内是光滑的. 如果 $\beta=0$, Xwish 成为线性函数 $f(x)=x$; 当 $\beta \rightarrow \infty$, Xswish 趋近于 $0 \sim 1$ 函数值, 类似于 ReLU 函数. 因此 Xwish 可视为线性函数与 ReLU 函数的非线性插值. Xwish 的一阶导数为

$$f'(x)=\frac{1}{\pi}[\arctan(\beta x)+0.5\pi]+\frac{1}{\pi}\frac{\beta x}{1+(\beta x)^2},$$

其一阶导数的图像如图 2 所示.

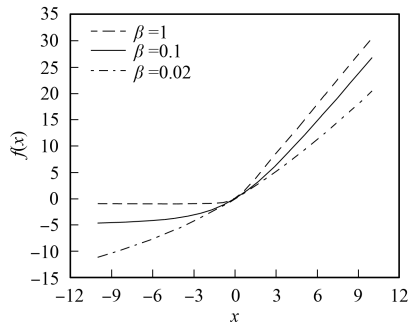


图 1 Xwish 激活函数

Fig. 1 Xwish activation function

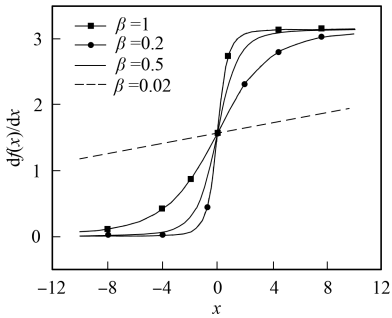


图 2 Xwish 函数的一阶导数

Fig. 2 First derivatives of Xwish function

参数 β 控制着导数趋于 0 或 1 的速度, 可以通过对参数 β 进行在线修正获得更好的收敛性. 下面以一个三层前向网络为例说明具体的修正过程, 其中 x 为输入, z 为隐层的输出, y 为输出层的输出, W 是输入层单元与隐层单元之间的权值, W' 为隐层单元与输出层单元之间的权值, t 为教师信号, 则有

$$\begin{aligned} z &= f(\text{net}) = f(W^T x), \\ y &= f(\text{net } z) = f(W'^T z), \\ \frac{\partial f(x)}{\partial \beta} &= \frac{x}{\pi} \frac{\beta}{1+(\beta x)^2}. \end{aligned}$$

定义误差函数

$$E=0.5(y-t)^T(y-t),$$

根据误差反向传播算法, β 修正如下:

$$\Delta \beta = -\frac{\partial E}{\partial \beta} = (y-t)^T \frac{\partial y}{\partial \beta} = (y-t)^T \left[\frac{\partial f(\text{net } z)}{\partial \beta} + f'(\text{net } z) W_z^T \frac{\partial f(\text{net})}{\partial \beta} \right].$$

在实际计算过程中, 可以在大多数的深度学习库中通过将激活函数替换成 Xwish 函数实现.

2 数值实验

数值验证基于深度学习框架 Tensorflow, 并分别在 MINST 和 Cifar10 数据集上进行, 比较激活函数为 Xwish, ReLU, LReLU, Tanh 时的运行精度. 所有实验都采用相同的卷积神经网络结构, 对于每个实验, 只更改激活函数而保持其他所有设置不变. 首先, 采用 Cifar-10 数据集, 图 3 为不同激活函数下的训练精度随迭代次数的变化曲线. 由图 3 可见, 运行 5 000 次迭代后, Xwish 的结果比 LReLU(取系统默认值 $\alpha=0.2$) 的训练精度提高了 13.2%, 比 ReLU 提高了 13.7%, 比 Tanh 提高了 28.3%, 表明 Xwish 具有很好的收敛精度. 下面用 MINST 数据集进一步验证, 对比实验结果列于表 1. 由表 1 可见, Xwish 训练 5 000 代的精度达 0.982 5, 比其他激活函数训练 15 000 代的精度高, 表明了 Xwish 函数的高效性.

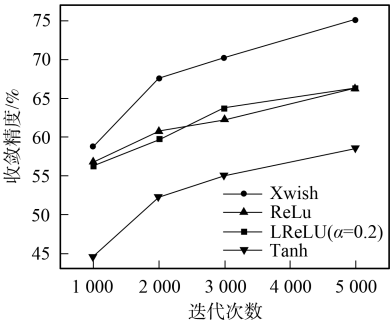


图 3 不同激活函数的收敛精度比较

Fig. 3 Comparison of convergence accuracy for different activation functions

表 1 MINST 数据集对比实验
Table 1 Contrast experiments in MINST dataset

实验标号	训练轮数	激活函数				
		Xwish	ReLU	Leaky-ReLU	Tanh	Sigmoid
1	1 000	0.977 0	0.976 7	0.973 9	0.966 8	0.942 3
2	5 000	0.982 5	0.980 1	0.981 1	0.979 7	0.969 6
3	10 000	0.986 1	0.981 7	0.980 5	0.980 7	0.974 6
4	15 000	0.988 8	0.981 8	0.982 3	0.981 1	0.976 8

综上所述, 本文提出的含有参数的激活函数 Xwish 结构简单、光滑、上下方有界, 且单调递增. 参数 β 可以在线修正, 避免了根据经验选取参数值的困难. 数值实验验证了激活函数 Xwish 的高效性.

参 考 文 献

[1] BALDI P, SADOWSKI P, LU Zhiqin. Learning in the Machine: Random Backpropagation and the Deep Learning Channel [J]. Artificial Intelligence, 2018, 260: 1-35.

[2] NAIR V, HINTON G E. Rectified Linear Units Improve Restrictcd Boltzmann Machines [C]//Proceedings of the 27th International Conference on Machine Learning. New York: ACM, 2010: 807-814.

[3] HOCHREITER S. The Vanishing Gradient Problem during Learning Recurrent Neural Nets and Problem Solutions [J]. International Journal of Uncertainty Fuzziness and Knowledge-Based Systems, 1998, 6(2): 107-116.

[4] LECUN Y, BENGIO Y, HINTON G. Deep Learning [J]. Nature, 2015, 521: 436-444.

[5] GOODFELLOW I, BENGIO Y, COURVILLE A, et al. Deep Learning [M]. Cambridge: MIT Press, 2016.

[6] CHATTERJEE A, GUPTA U, CHINNAKOTLA M K, et al. Understanding Emotions in Text Using Deep Learning and Big Data [J]. Computers in Human Behavior, 2019, 93: 309-317.

[7] LIU Hui, XU Jun, WU Yan, et al. Learning Deconvolutional Deep Neural Network for High Resolution Medical Image Reconstruction [J]. Information Sciences, 2018, 468: 142-154.

[8] RAMACHANDRAN P, ZOPH B, LE Q V. Searching for an Activation Functions [C/OL]//ICLR 2018 Conference. 2017-10-27. <https://arxiv.org/pdf/1710.05941.pdf>.

[9] LI Zechao, TANG Jinhui. Weakly Supervised Deep Metric Learning for Community-Contributed Image Retrieval [J]. IEEE Transactions on Multimedia, 2015, 17(11): 1989-1999.

[10] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Developing Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification [EB/OL]. 2015-02-06. <https://arxiv.org/abs/1502.01852>.

[11] CLEVERT D A, UNTERTHINER T, HOCHREITER S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs) [C/OL]//ICLR 2016. 2016-02-22. <https://arxiv.org/abs/1511.07289>.

[12] 涂淑琴, 薛月菊, 梁云, 等. RGB-D 图像分类方法研究综述 [J]. 激光与光电子学进展, 2016, 53(6): 060003-1-060003-14. (TU Shuqin, XUE Yueju, LIANG Yun, et al. Review on RGB-D Image Classification [J]. Laser & Optoelectronic Progress, 2016, 53(6): 060003-1-060003-14.)

[13] 卢宏涛, 张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述 [J]. 数据采集与处理, 2016, 31(1): 1-17. (LU Hongtao, ZHANG Qinchuan. Applications of Deep Convolutional Neural Network in Computer Vision [J]. Journal of Data Acquisition and Processing, 2016, 31(1): 1-17.)

(责任编辑: 赵立芹)