

基于改进激活函数的卷积神经网络研究

曲之琳, 胡晓飞

(南京邮电大学 通信与信息工程学院, 江苏 南京 210003)

摘要:卷积神经网络是对于人脑的高度抽象,它是深度学习的重要组成部分。对于卷积神经网络的研究,一方面有助于更准确地进行图像的分类与识别,另一方面,有助于人类更真实地模拟人脑,为人工智能的发展指明了方向。分析比较了 Sigmoid、Tanh、ReLU、Softplus 4 种激活函数的优缺点。结合 ReLU 和 Softplus 两种激活函数的优点,设计并构造了一种分段激活函数。最后,基于 Theano 框架和这 5 种激活函数,分别构建了 5 种卷积神经网络,并对 Cifar-10 数据集进行了分类识别。实验结果表明,基于改进后的激活函数所构造的卷积神经网络,不仅收敛速度更快,而且可以更加有效地提高分类的准确率。

关键词:卷积神经网络;深度学习;人工智能;激活函数

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2017)12-0077-04

doi:10.3969/j.issn.1673-629X.2017.12.017

Research on Convolutional Neural Network Based on Improved Activation Function

QU Zhi-lin, HU Xiao-fei

(School of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract:Convolutional neural network is a high degree of abstraction to the human brain and an important part of deep learning. For research on it, on the one hand, it is helpful for a more accurate image classification and recognition. On the other hand, the human brain can be more truly simulated, which points out the direction for the development of artificial intelligence. First the advantages and disadvantages of four kinds of activation functions such as Sigmoid, Tanh, ReLU and Softplus are analyzed and compared. Then, combined with the advantages of ReLU and Softplus, a piecewise activation function is designed and constructed. Finally, based on Theano framework and these activation functions, five convolutional neural networks are established respectively for classification recognition on the Cifar-10 data sets. The experimental results show that the convolution neural network based on the improved activation function not only converges faster, but also improves the classification accuracy more effectively.

Key words:convolutional neural network; deep learning; artificial intelligence; activation function

0 引言

深度学习 (Deep Learning, DL)^[1-2]是近年来机器学习领域最令人瞩目的方向,它通过模仿人脑的工作机制来解释图像、语音和文本等数据。其概念源于人工神经网络 (Artificial Neural Network, ANN)^[3]的研究与发展。反向传播算法 (Back Propagation, BP)^[4-5]的提出,使得机器学习变得不再遥不可及,并最终带来了基于统计模型的机器学习研究的浪潮。

2006 年,多伦多大学的教授 Hinton 及其学生提出

了深度学习这一概念并受到了广泛关注^[6]。在 2012 年 NIPS 会议上,Google 的 Jeff Dean 和斯坦福大学的 Andrew Ng 针对深度学习的 CPU 集群框架提出了 GPU 并行框架^[7]。百度搭建了 Paddle (Parallel asynchronous distributed deep learning) 多机 GPU 训练平台^[8]。将数据分布到不同机器,通过 Parameter Server 协调各机器训练。Paddle 支持数据并行和模型并行。深度学习技术为实现人工智能提供了可能,包括 Google、微软、Facebook、百度、腾讯等公司都纷纷组建深度学习

收稿日期:2016-05-30

修回日期:2016-08-24

网络出版时间:2017-08-01

基金项目:国家自然科学基金资助项目(61271082);江苏省重点研发计划(BE2015700);江苏省自然科学基金(BK20141432)

作者简介:曲之琳(1991-),女,硕士研究生,研究方向为图像处理与图像通信;胡晓飞,副教授,研究生导师,研究方向为视频信号编码与处理、医学图像处理等。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20170801.1549.004.html>

团队,在语音识别、人像识别、自动驾驶等领域取得了长足进展^[9]。2006 年,Hinton 提出的深度置信网(Deep Belief Networks,DBN)开启了深度学习新纪元。在 MNIST 数据库^[10]上取得了 98.8% 的识别率。

2007 年,Bengio 等^[10]借鉴了 DBN 无监督预训练对优化深度网络的优点,将受限玻尔兹曼机替换成自动编码器(Auto-Encoder,AE),进行无监督训练多个 AE,然后再级联起来形成深度结构。这种新型的深度结构被称为层叠自动编码器(Stacked Auto-Encoders,SAE)。早在 1980 年,Fukushima 提出了卷积神经网络(CNN)的概念^[11]。1998 年 Lecun 等对卷积神经网络进行了进一步的实现和优化^[12]。2003 年 Simard 等对其进行了简化^[13]。上面介绍的三种典型的深度学习模型分别来自于不同的团队,DBN 来自于多伦多大学的 Hinton 团队^[14],SAE 来自于蒙特利尔大学的 Bengio 团队^[15],CNN 来自于纽约大学的 Lecun 团队^[16]。由此可知,基于深度卷积神经网络图像分类系统的使用越来越广泛,卷积神经网络的研究工作一直受到研究者的高度重视。但其中一些令人头痛的问题仍没有较好的解决方案。例如,深度学习模型本身比较复杂,实现起来有一定的难度;深度模型的训练算法决定了模型容易梯度弥散,模型也不容易收敛,需要耗费大量时间进行调试;还没有形成完整的通用理论,所以设计网络结构和训练模型需要掌握很多实用技巧并需要不断探索最佳的参数和优化算法等。

为解决模型梯度弥散,分析了激活函数 ReLu、Softplus 的优点和缺陷,并且基于 ReLu 和 Softplus 构造了一种分段函数作为激励函数。通过设计一个卷积神经网络,在公开的数据集上进行实验,分析各种神经元激励函数对网络收敛速度和图像识别准确率的影响。

1 神经网络激活函数的作用

激活函数是指如何把“激活的神经元的特征”通过非线性函数保留并映射出来,这就是神经网络能解决非线性问题的关键所在。当激活函数为线性时,线性方程的线性组合也只有线性表达的能力,就算网络有多层,也只相当于有一个隐藏层的线性网络。这种输入的线性表示网络,还只相当于多层感知机。这样也就根本无法用非线性来逼近任意函数。由于性能远未达到要求,所以尝试使用非线性的组合。使用激活函数增加了神经网络模型的非线性,使得深度神经网络真正具有了意义。同时传统的激活函数会把输入值归约到一个区间内,因为当激活函数的输出值有限时,基于梯度的优化方法会更加稳定。新型的脑神经稀疏激活函数具有高频的训练效率,不过在这种情况下,一

般需要更小的学习率。

2 经典的激活函数

2.1 Sigmoid

Sigmoid 是一种常用的 S 型非线性激活函数。其功能是把一个实数压缩至 0 到 1 之间,对中部区的信号增益较大,对两侧区的信号增益较小。它的输出有界,为神经网络带来了非线性,曾一度作为深度学习的激活函数而广泛使用。数学形式为: $f(x) = 1/(1 + e^{-x})$ 。

虽然 Sigmoid 的函数性质和神经学的神经元的突触一致,而且便于求导,但现在 Sigmoid 却很少使用。原因是它存在一个巨大的缺陷:对于深层网络,Sigmoid 函数反向传播时,很容易出现梯度弥散的情况,在输入特别大或者特别小的地方即 Sigmoid 接近饱和和区时,导数趋于零,从而无法完成深层网络的训练。另外一个缺点就是其输出是非 0 均值的。非 0 均值产生的不利结果就是直接影响梯度下降。比如当输入是正的,反向传播计算出的梯度也会始终是正的。

2.2 Tanh

Tanh 函数也是一种常用的 S 型非线性激活函数,是 Sigmoid 的变种。其功能是把一个实数压缩至 -1 到 +1 之间,对中部区的信号增益较大,对两侧区的信号增益较小。它的输出有界,为神经网络带来了非线性。数学形式为: $f(x) = (e^x - e^{-x})/(e^x + e^{-x})$ 。同时 Tanh 克服了 Sigmoid 非 0 均值输出的缺点,延迟了饱和期,拥有更好的容错能力,性能上要优于 Sigmoid。但它也存在梯度弥散的问题,这种缺点是致命的。这也表明不管是 Sigmoid 还是 Tanh 都存在极大的局限性。

2.3 Relu

当前在神经网络模型中激活函数使用的趋势就是非饱和修正线性函数。修正线性单元(Rectified Linear unit, ReLu)是这样一个分段函数,假如输入值小于等于零时,就强制使其等于 0,输入大于零时就保持原来的值不变。定义式为: $f(x) = \max(0, x)$ 。ReLu 这种直接地强制某些数据为 0 的做法,在一定程度上,为训练后的网络带来了适度的稀疏特性。它不但减少了参数之间的相互依存关系,而且缓解了过拟合问题的发生。相比较传统的 S 型激活函数,ReLu 不含除法和指数运算,计算速度更快。由于它的分段线性的性质,使得计算以后始终保持着分段线性。而 S 型激活函数,由于两端饱和的缺点,容易在传播过程中丢弃有用的特征。所以 ReLu 也更接近于生物神经元激活的本质,即更加符合神经元信号激励原理。

总之,由于 ReLu 是非饱和的,不会像 Sigmoid 和 Tanh 有梯度弥散的问题。这也导致传统的激活函数

很难完成更深层网络的训练。在进行反向传播求误差梯度时,ReLU 有更大的优势,收敛速度很快。由于 ReLu 将部分神经元的输出置为零,这也在一定程度上造成了网络的稀疏特性。

2.4 Softplus

Softplus 函数也是非饱和修正线性函数,可以看作是 ReLu 函数的近似光滑表示。它的定义式为: $f(x) = \log_e(1 + e^x)$ 。它对全部数据进行了非线性映射。但是 Softplus 并不具备很好的稀疏性。对比传统的激活函数,它有单边抑制特性,能做到有限的稀疏特性。同时它的兴奋边界也比较宽阔,能够更好地近似生物神经系统。

3 改进的激活函数

由于 ReLu 是分段线性的非饱和和激活函数,相比传统的 S 型激活,具有更快的随机梯度下降收敛速度。在深度神经网络中使用 Sigmoid 激活函数,必须进行复杂的预处理才能达到 ReLu 相似的结果。往往 ReLu 只需要一个阈值就能获得激活值,无需进行大量繁杂的计算。

相比 Softplus 函数,ReLU 更具稀疏性。然而,粗暴地强制稀疏处理,也会屏蔽很多有用的特征,导致模型学习的效果变差。过分的稀疏性会带来更高的错误率,减少模型的有效容量。虽然 Softplus 相比 Sigmoid 和 Tanh 收敛更快,效果更好,但是它不具备很好的稀疏性,同时比 ReLu 收敛要慢。所以结合 ReLu 和 Softplus 的优点,构造一个新的非饱和修正线性激活函数用于深度神经网络。在数据小于零时,使用 Softplus 函数,并将其函数曲线向下平移 \lg_2 个单位大小。在数据大于零时,使用 ReLu 函数。将该函数记为 Relus_Softplus,其公式如下所示: $f(x) = \max(\lg_e(1 + e^x) - \lg_2, x)$ 。

函数图如图 1 所示。

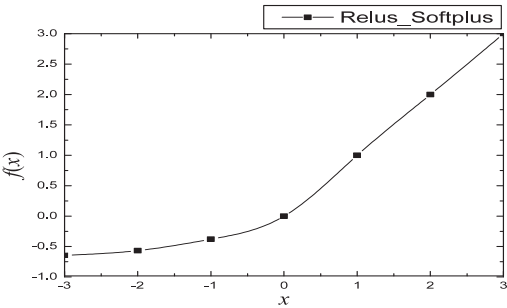


图 1 改进的激活函数

这样做不但修正了数据的分布,而且保留了 ReLu 快速收敛的好处。最重要的是,一些负轴的值得以保存,使得负轴信息不至于全部丢掉。ReLU 容易“死掉”的问题也得到了较好的解决。

4 基于改进激活函数的深度学习实验

4.1 运行环境以及 Cifar-10 数据集

在实验中,使用的 CPU 是 Intel(R)Core(TM)2Duo CPU E7500 @ 2.93 GHz 2.94 GHz, GPU 是 NVIDIA GTX750 Ti, 操作系统是 64 位 Ubuntu14.03, 基于 theano 框架,在 spyder 上进行开发与调试,同时使用了 CUDA 加速。

分别采用 Sigmoid、Tanh、Softplus、ReLU 和 Relus_Softplus 进行训练和交叉验证。共进行 60 次训练。由于激活函数不同,网络的收敛时间也不同。同时应注意,使用 Sigmoid 和 Tanh 时,学习率一般使用 0.01;而当使用修正的激活函数时,应将学习率调低一些,一般使用 0.001。

实验中使用 Cifar-10 数据集。Cifar-10 数据集是一个用于普适物体图像识别的计算机视觉数据集,如图 2 所示。它包含十类物体,分别为飞机、汽车、鸟、猫、鹿、狗、青蛙、马、轮船和卡车,每类 6 000 个,共 60 000 个 32 * 32 的 RGB 彩色图像。其中包括 50 000 个训练图像和 10 000 个测试图像。每个训练批次中所包含的图像的秩序是随机的,且一些训练批次可能包含的某类图像比另一类更多,但所有批次每个种类的图像的总量是不变的。



图 2 Cifar-10 数据集

4.2 实验结果

实验结果如图 3 所示。

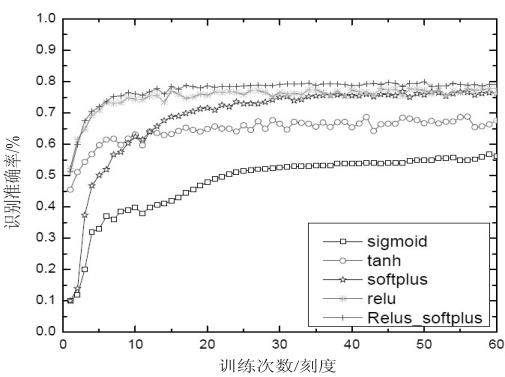


图 3 不同激活函数的网络分类正
确率随训练次数的变化曲线

从图 3 可以看出,使用 Sigmoid 函数作为激活函数

的识别率是最低的,仅为 52.14%。不但如此,Sigmoid 函数作为激活函数的网络很难收敛,要耗费大量精力对网络进行微调;使用 Tanh 函数作为激活函数的网络较 Sigmoid 函数作为激活函数的网络的收敛速度有了明显提升,但是它的最高识别率仅有 67.48%;使用 Softplus 函数作为激活函数的网络虽然收敛速度不够快,但是识别率较之前的激活函数有很大提高,最大识别率为 77.43%;使用 ReLu 函数作为激活函数的网络的识别率较高,而且网络的收敛速度也较快,最大识别率为 76.82%;而使用改进的激活函数的网络的识别率是最高的,最大识别率为 78.59%,网络收敛速度也很快。因此,相较于常见的激活函数,使用改进后的激活函数不但可以提高网络收敛速度,而且可以提高识别准确率。

5 结束语

激活函数是卷积神经网络的重要部分,可以将非线性特征映射出来,从而使卷积神经网络成为可能。针对使用反向传播算法计算传播梯度时可能引起梯度弥散的问题,为了有效缓解该问题,增强网络性能,提高卷积神经网络识别图像的准确率,在研究激活函数的作用并分析各种激活函数优缺点的基础上,提出了一种改进的激活函数。实验结果表明,该函数对图像分类结果有较好的效果,有效缓解了模型的梯度弥散问题,提高了卷积网络图像识别准确率。

参考文献:

- [1] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]//International conference on neural information processing systems. [s. l.]: [s. n.], 2012:1097–1105.
- [2] Arel I, Rose D C, Karnowski T P. Deep machine learning a new frontier in artificial intelligence research [research frontier] [J]. IEEE Computational Intelligence Magazine, 2010, 5 (4):13–18.
- [3] Hsu K, Gupta H V, Sorooshian S. Artificial neural network modeling of the rainfall–runoff process [J]. Water Resources Research, 1995, 31 (31):2517–2530.
- [4] Rumelhart D, Hinton G E, Williams R J. Learning internal representations by error propagation [M]//Parallel distributed processing: exploration of the microstructure of cognition. USA: MIT Press, 1986:318–362.
- [5] Williams D, Hinton G E. Learning representations by back-propagating errors [J]. Nature, 1986, 323 (6088):533–536.
- [6] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313 (5786):504–507.
- [7] Dean J, Corrado G S, Monga R, et al. Large scale distributed deep networks [C]//Proceedings of the neural information processing systems. [s. l.]: [s. n.], 2013:1223–1232.
- [8] Yu K. Large-scale deep learning at Baidu [C]//International conference on information & knowledge management. [s. l.]: ACM, 2013:2211–2212.
- [9] Lecun Y, Jackel L D, Boser B, et al. Handwritten digit recognition: applications of neural net chips and automatic learning [J]. IEEE Communications Magazine, 1989, 27 (11):41–46.
- [10] Bengio Y, Lamblin P, Popovici D, et al. Greedy layer-wise training of deep networks [C]//Advances in neural information processing systems. Cambridge, MA: MIT Press, 2007:153–160.
- [11] Fukushima K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position [J]. Biological Cybernetics, 1980, 36 (4):193–202.
- [12] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86 (11):2278–2324.
- [13] Simard P Y, Steinkraus D, Platt J C. Best practices for convolutional neural networks applied to visual document analysis [C]//International conference on document analysis and recognition. [s. l.]: IEEE, 2003:958.
- [14] Hinton G E. Learning multiple layers of representation [J]. Trends in Cognitive Sciences, 2007, 11 (11):428–434.
- [15] Bengio Y, Yao L, Alain G, et al. Generalized denoising auto-encoders as generative models [J]. Advances in Neural Information Processing Systems, 2013, 26:899–907.
- [16] Dahl J V, Koch K C, Kleinhan E, et al. Convolutional networks and applications in vision [C]//International symposium on circuits and systems. [s. l.]: IEEE, 2010:253–256.