

DOI:10.16644/j.cnki.cn33-1094/tp.2018.11.005

卷积神经网络综述*

刘 健, 袁 谦, 吴 广, 喻 晓

(浙江省电子信息产品检验所, 浙江 杭州 310007)

摘要: 卷积神经网络作为深度学习的一种经典而广泛应用的结构, 克服了过去人工智能中被认为难以解决的一些问题。卷积神经网络的局部连接、权值共享及下采样操作等特性使之可以有效地降低网络的复杂度, 减少训练参数的数目, 使模型对平移、扭曲、缩放具有一定程度的不变性, 并具有强鲁棒性和容错能力, 也易于训练和优化。文章介绍了卷积神经网络的训练方法, 开源工具, 及在图像分类领域中的一些应用, 给出了卷积神经网络待解决的问题及展望。

关键词: 深度学习; 卷积神经网络; 网络结构; 训练方法

中图分类号: TP183

文献标志码: A

文章编号: 1006-8228(2018)11-19-04

Review of convolutional neural networks

Liu Jian, Yuan Qian, Wu Guang, Yu Xiao

(Zhejiang Provincial Testing Institute of Electronic Information Products, Hangzhou, Zhejiang 310007, China)

Abstract: As a classical kind of widely used network structure in deep learning, convolutional neural networks has successfully solved some problems which were considered difficult to solve in artificial intelligence in the past. The characteristics such as local connections, shared weights, under-sampling etc. can reduce the complexity of networks and the number of training parameters, and can also make the model creating invariance to translation, distortion and zoom and having strong robustness and fault tolerance. So it is easy to train and optimize its network structure. This paper introduces the training methods, open source tools of convolutional neural networks and its applications in the field of image classification, the problems and prospects of the convolutional neural network to be solved are given also.

Key words: deep learning; convolutional neural networks; network structure; training method

0 引言

人工神经网络(Artificial Neural Network, ANN)是对生物神经网络的一种近似模拟,是由大量神经元通过相互连接而构成的自适应非线性动态网络系统。1943年,心理学家McCulloch和数理逻辑学家Pitts提出了第一个神经元数学模型,称为MP模型^[1]。他们通过MP模型提出了神经元的形式化数学描述和网络结构方法,证明了单个神经元能执行逻辑功能,从而开创了人工神经网络研究的时代。60年代, Rosenblatt^[2]在MP模型基础上提出了单层感知器模型,但是单层感知器网络模型不能够处理线性不可分

问题。直到1986年, Rumelhart^[3]等人提出了BP算法解决了单层感知器不能解决的问题,掀起了神经网络在机器学习中的研究热潮。神经网络中的大量参数,容易发生局部最优、过拟合、训练时间长的缺陷。2006年, Hinton^[4]等在《Science》上提出了深度学习,这篇文章的两个主要观点是:①多隐层的人工神经网络具有优异的特征学习能力,学习到的数据更能反映数据的本质特征,有利于可视化或分类;②深度神经网络在训练上的难度,可以通过逐层无监督训练有效克服。Bengio系统地介绍了深度学习所包含的网络结构和学习方法。目前,常用的深度学习模型有深度置信网

收稿日期:2018-07-24

*基金项目:浙江省科技计划项目“移动应用软件自动化检测平台”(2017F10031)

作者简介:刘健(1984-),男,浙江杭州人,硕士研究生,主要研究方向:软件开发与测试。

通讯作者:喻晓(1982-),女,浙江湖州人,硕士研究生,主要研究方向:软件开发与测试。

络(Deep Belief Network, DBN)^[5-6]、层叠自动去噪编码器(Stacked Denoising Autoencoders, SDA)^[7-8]、卷积神经网络(Convolutional neural network, CNN)^[9-10]等。

1 CNN概述

神经网络包括输入层、隐藏层和输出层,如图1所示。

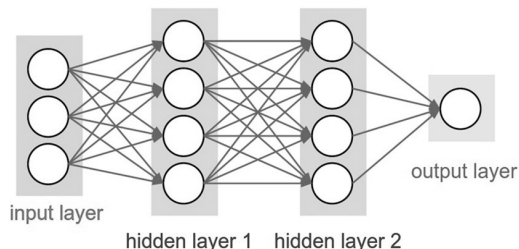


图1 神经网络结构图

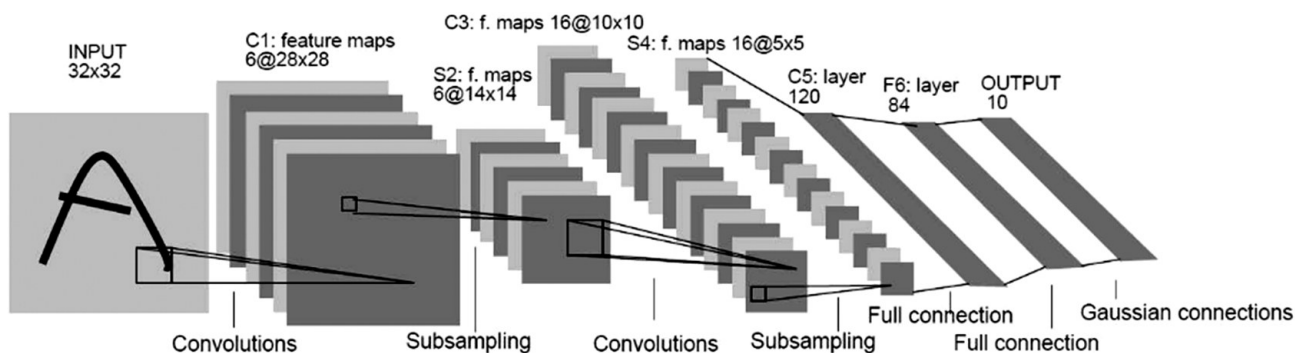


图2 LeNet-5模型

1.1 卷积层

卷积核是卷积层的重要组成部分。特征提取器是卷积核的本质,其主要作用是自动提取输入信号的深层信息。卷积层由多个特征面组成,每个特征面由多个神经元组成,它的每一个神经元通过卷积核与上一层特征面的局部区域相连。上一层的特征图被一个可学习的卷积核进行卷积,然后通过一个激活函数,就可以得到输出特征图。每个输出特征图可以组合卷积多个特征图的值^[11]:

$$\begin{aligned} x_j^l &= f(u_j^l) \\ u_j^l &= \sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l \end{aligned} \quad (1)$$

其中, u_j^l 称为卷积层 l 的第 j 个通道的净激活,它通过对前一层输出特征图 x_i^{l-1} 进行卷积求和与偏置后得到, x_j^l 是卷积层 l 的第 j 个通道的输出。 $f(\cdot)$ 称为激活函数,通常可使用 sigmoid 和 tanh 等函数。 M_j 表示用于计算 u_j^l 的输入特征图子集, k_{ij}^l 是卷积核矩阵, b_j^l 是对卷积后特征图的偏置。对于一个输出特征图 x_j^l ,每个输入特

征图 x_i^{l-1} 对应的卷积核 k_{ij}^l 可能不同,“*”是卷积符号。卷积神经网络是一种带有卷积结构的深度神经网络,卷积结构可以减少深层网络占用的内存量,也可以减少网络的参数个数,缓解模型的过拟合问题。卷积神经网络中隐藏层是重要组成部分。经典的 CNN 由输入层、卷积层、下采样层(也称池化层)、全连接层及输出层组成。典型的 LeNet-5^[9] 是经典的 CNN 结构,如图2所示。卷积层和下采样层一般会取若干个,采用卷积层和下采样层交替设置,即一个卷积层连接一个下采样层,下采样层后再连接一个卷积层,依此类推。由于卷积层中输出特征面的每个神经元与其输入进行局部连接,并通过对应的连接权值与局部输入进行加权求和再加上偏置值,得到该神经元输入值,该过程等同于卷积过程, CNN 也由此而得名^[9]。

征图 x_i^{l-1} 对应的卷积核 k_{ij}^l 可能不同,“*”是卷积符号。

影响 CNN 性能有 3 个因素:层数、特征面的数目及网络组织。使用 9 种结构的 CNN 进行中文手写体识别实验,通过统计测试结果得到具有较小卷积核的 CNN 结构的一些结论:①增加网络的深度能够提升准确率;②增加特征面的数目也可以提升准确率;③增加一个卷积层比增加一个全连接层更能获得一个更高的准确率。Bengio 等人^[12]指出深度网络结构具有两个优点:①可以促进特征的重复利用;②能够获取高层表达中更抽象的特征,由于更抽象的概念可根据抽象性更弱的概念来构造,因此深度结构能够获取更抽象的表达。He 等人^[13]探讨了在限定计算复杂度和时间上如何平衡 CNN 网络结构中深度、特征面数目、卷积核大小等因素的问题。该文献首先研究了深度与卷积核大小间的关系,通过实验结果表明网络深度比卷积核大小更重要;当时间复杂度大致相同时,具有更小卷积核且深度更深的 CNN 结构,比具有更大卷积核同时深度更浅的 CNN 结构能够获得更好的实验结果。其次,该文献也研究了网络深度和特征面数目间

的关系,实验结果表明,深度越深,网络性能越好;然而随着深度的增加,网络性能也逐渐达到饱和。此外,该文献还通过固定网络深度研究了特征面数目和卷积核大小间的关系,通过实验对比发现特征面数目和卷积核大小的优先级差不多,其发挥的作用均没有网络深度大。

1.2 下采样层

下采样层用来实现对特征图的采样处理,在减少数据量的同时保留有用的信息,使CNN具有抗畸变的能力。下采样层将每个输入特征图通过下面的公式下采样输出特征图:

$$\begin{aligned} x_j^l &= f(u_j^l) \\ u_j^l &= \beta_j^l \text{down}(x_j^{l-1}) + b_j^l \end{aligned} \quad (2)$$

其中, u_j^l 称为下采样层 l 的第 j 通道的净激活,它由前一层输出特征图 x_j^{l-1} 进行下采样加权、偏置后得到, β 是下采样层的权重系数, b_j^l 是下采样层的偏置项。符号 $\text{down}(\cdot)$ 表示下采样函数,它通过对输入特征图 x_j^{l-1} 通过滑动窗口方法划分为多个不重叠的 $n \times n$ 图像块,然后对每个图像块内的像素求和、求均值或最大值,于是输出图像在两个维度上都缩小了 n 倍。

下采样层旨在通过降低特征面的分辨率来获得具有空间不变性的特征。下采样层起到二次提取特征的作用,它的每个神经元对局部接受域进行采样操作。常用的方法有最大下采样即取局部接受域中值最大的点、均值下采样即对局部接受域中的所有值求均值、随机下采样^[14-15]。

1.3 全连接层

全连接层一般位于网络尾端,对前面逐层变换和映射提取的特征进行回归分类等处理。将所有二维图像的特征图拼接为一维特征作为全连接网络的输入。全连接层 l 的输出可通过对输入加权求和并通过激活函数的响应得到:

$$\begin{aligned} x^l &= f(u^l) \\ u^l &= w^l x^{l-1} + b^l \end{aligned} \quad (3)$$

其中, u^l 称为全连接层 l 的净激活,它由前一层输出特征图 x^{l-1} 进行加权和偏置后得到的。 w^l 是全连接网络的权重系数, b^l 是全连接层 l 的偏置项。

为了提升CNN网络性能,全连接层每个神经元的激励函数一般采用ReLU函数^[16]。最后一层全连接层的输出值被传递给一个输出层,可以采用softmax逻辑回归进行分类,该层也可称为softmax层。

2 CNN训练方法及开源工具

2.1 训练方法

神经网络用于模式识别的主流是有监督训练^[17],无监督训练更多的是用于聚类分析。对于有监督训练,根据同类样本在空间的分布及不同类样本之间的分离程度找一种适当的空间划分方法,使得不同类样本分别位于不同的区域内。

卷积网络在本质上是一种输入到输出的映射,它能够学习大量的输入与输出之间的映射关系,而不需要任何输入和输出之间的精确的数学表达式,只要用已知的模式对卷积网络加以训练,网络就具有输入输出对之间的映射能力。卷积网络通过BP算法进行有监督训练。在开始训练前,采用一些不同的小随机数对网络中的所有的权值和偏置值进行初始化,“小随机数”用来保证网络不会因权值过大而进入饱和状态,从而导致训练失败;“不同”用来保证网络可以正常地学习。随机初始化的权值和偏置值的范围可为 $[-0.5, 0.5]$ 或者 $[-1, 1]$ (或者是其他合适的区间)^[18]。

在实际应用中,无标注的数据远多于有标注的数据,同时对数据进行人工标注也需要耗费较大的人力。但是为了使有监督CNN得到充分的训练并获得较好的泛化能力,又需要大量有标注的训练样本,这一定程度上制约了CNN在实际中的应用。

2.2 开源工具

深度学习能够广泛应用于众多研究领域,离不开许多优秀的开源深度学习仿真工具。目前常用的深度学习仿真工具有Caffe^[19]、Torch、Theano^[20]、以及TensorFlow等。

Caffe是一个基于C++语言且关于CNN相关算法的架构。Caffe可以在CPU及GPU上运行,它支持MATLAB和Python接口。Caffe提供了一个完整的工具包,用于训练、测试、微调及部署模型。Caffe允许用户对新数据格式、网络层和损失函数进行拓展;它的运行速度也很快,但是灵活性较差。

Torch是一个支持机器学习算法的科学计算框架。它是采用Lua脚本语言和C语言编写的。Torch为设计和训练机器学习模型提供了一个灵活的环境,它还可支持iOS、Android等嵌入式平台,但Torch没有Python接口。

Theano是一个允许用户定义、优化并评价数学表达式的Python库。提供了NumPy的大部分功能,可在

GPU上运行。此外,Theano能够自动求微分,它尤其适用于基于梯度的方法。Theano能够很容易且高效地实现递归神经网络(Recurrent Neural Network, RNN)。然而Theano的编译过程很慢,导入Theano也需要消耗时间。

TensorFlow是Google开源的一款深度学习框架,可以在CPU和GPU上运行,支持C++和Python编程语言,因此可以在不同的计算机上自由运行代码,程序比Theano编译时间短。TensorFlow不仅支持深度学习,还有支持强化学习和其他算法的工具。TensorFlow的计算图是纯Python的,因此速度较慢。

3 CNN应用

近年来,CNN已被广泛应用于图像处理领域中。Krizhevsky等人^[21]第1次将CNN用于LSVRC-12竞赛中,通过加深CNN模型的深度并采用ReLU+dropout技术,取得了当时最好的分类结果(该网络结构也被称为AlexNet)。相比于AlexNet,Szegedy等人^[22]大大增加了CNN的深度,提出了一个超过20层的CNN结构(称为GoogleNet)。在GoogleNet结构中采用了3种类型的卷积操作,该结构的主要特点是提升了计算资源的利用率,它的参数比文献[21]少了12倍,而且GoogleNet的准确率更高,在LSVRC-14中获得了图像分类“指定数据”组的第1名。Simonyan等人^[23]在其发表的文章中探讨了“深度”对于CNN网络的重要性。该文通过在现有的网络结构中不断增加具有 3×3 卷积核的卷积层来增加网络的深度,实验表明,当权值层数达到16~19时,模型的性能能够得到有效提升(也称为VGG模型)。VGG与GoogLeNet的深度都比较深,所以网络结构比较复杂,训练时间长,而且VGG还需要多次微调网络的参数。

4 总结


CNN通过权值共享减少了需要训练的权值个数、降低了网络的计算复杂度,同时通过下采样操作使得网络对输入的局部变换具有一定的不变性,提升了网络的泛化能力。卷积神经网络虽然在一些数据上取得了成功,仍有许多工作需要进一步去做:①很难确定CNN使用哪种网络结构,使用多少层,每一层使用多少个神经元才是合适的;②训练数据集与测试数据集的分布不一样,则CNN很难获得一个好的识别结果;③如何针对更大规模数据、更深结构网络设计高

效的数值优化、并行计算方法和平台。

总之,CNN虽然还有许多有待解决的问题,但是这并不影响今后它在模式识别与人工智能等领域中的发展与应用,它在未来很长的一段时间内仍然会是人们研究的一个热点。

参考文献(References):

- [1] McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 1943.5(4):115-133
- [2] Rosenblatt F. *Principles of Neurodynamics: Preceptron and Theory of Brain Mechanisms*. Washington, USA: Spartan Books, 1962.
- [3] Rumelhart D E, Hinton G, Williams R J. Learning representations by back-propagating errors. *Nature*, 1986.323(6088):533-536
- [4] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, 2006.313(5786):504-507
- [5] Hinton G, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006.18(7): 1527-1554
- [6] Salakhutdinov R, Hinton G. Deep Boltzmann machines. *Journal of Machine Learning Research-Proceedings Track*, 2009.9(1):448-455
- [7] Vincent P, Larochelle H, Bengio Y, et al. Extracting and composing robust features with denoising autoencoders//*Proceedings of the 25th international Conference on Machine Learning*. Helsinki, Finland, 2008:1096-1103
- [8] Vincent P, Larochelle H, Bengio Y, et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 2010.11(12):3371-3408
- [9] LeCun Y, Bottou L, Bengio Y, et al. Gradient-base learning applied to document recognition. *Proceedings of the IEEE*, 1998.86(11):2278-2324
- [10] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1989.11(4):541-551
- [11] Bouvrie J. *Notes On Convolutional Neural Networks*, MIT CBCL Tech Report, Cambridge, MA, 2006.
- [12] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013.35 (8):1798-1828
- [13] He Kai-Ming, Sun Jian. Convolutional neural networks at

- constrained time cost//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Boston,USA,2015:5353-5360
- [14] Boureau Y-L, Roux N L, Bach F, et al. Ask the locals: Multi-way local pooling for image recognition//Proceedings of the 2011 International Conference on Computer Vision. Barcelona,Spain,2011:2651-2658
- [15] Zeiler M D,Fergus R. Stochastic pooling for regularization of deep convolutional neural networks. arXiv:1301.3557v1,2013.
- [16] O'Shea K,Nash R. An introduction to convolutional neural networks. arXiv:1511.08458v2,2015.
- [17] Bengio Y. Learning deep architectures for AI. Foundations and Trends in Machine Learning,2009.2(1):1-127
- [18] Fausett L. Fundamentals of Neural Networks:Architectures, Algorithms, and Applications. London:Pren-tice-Hall,1994.
- [19] Jia Yang-Qing,Shelhamer E,Donahue J,et al. Caffe: Convolutional architecture for fast feature embedding//Proceedings of the ACM International Conference on Multimedia. Orlando,USA,2014:675-678
- [20] Al-Rfou R,Alain G,Almahairi A,et al. Theano:A python framework for fast computation of mathematical expressions. arXiv:1605.02688v1,2016.
- [21] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks//Proceedings of the Advances in Neural Information Processing Systems. Lake Tahoe,USA, 2012:1097-1105
- [22] Szegedy C,Liu Wei,Jia Yang-Qing,et al. Going deeper with convolutions//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Boston,USA,2015:1-9
- [23] Simonyan K,Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556v6,2014. 

(上接第18页)

参考文献(References):

- [1] 王婷婷,韩满,王宇.LDA模型的优化及其主题数量选择研究——以科技文献为例[J].数据分析与知识发现,2018.1:29-39
- [2] 廖列法,勒厚刚,朱亚兰.LDA模型在专利文本分类中的应用[J].现代情报,2017.37(3):35-39
- [3] 逯万辉,谭宗颖.学术成果主题新颖性测度方法研究——基于Doc2Vec和HMM算法[J].数据分析与知识发现,2018.3:22-29
- [4] 颜端武,陶志恒,李兰彬.一种基于HDP模型的主题文献自动推荐方法及应用研究[J].情报理论与实践,2016.39(1):128-132
- [5] 曾立梅.基于文本数据挖掘的硕士论文分类技术[J].重庆邮电大学学报:自然科学版,2010.22(5):669-672
- [6] 王昊,叶鹏,郑三鸿.机器学习在中文期刊论文自动分类研究中的应用[J].现代图书情报技术,2014.3:80-87
- [7] 刘浏,王东波.基于论文自动分类的社科类学科跨学科性研究[J].数据分析与知识发现,2018.3:30-38
- [8] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research,2003.3:993-1022
- [9] 王树义,廖群涛,吴查科.基于情感分类的竞争企业新闻文本主题挖掘[J].数据分析与知识发现,2018.3:70-78
- [10] 苏金树,张博群,徐昕.基于机器学习的文本分类技术研究进展[J].软件学报,2006.17(9):1848-1859
- [11] 周庆平,谭长庚,王宏君,等.基于聚类改进的KNN文本分类算法[J].计算机应用研究,2016.33(11):3374-3377
- [12] 张晨逸,孙建伶,丁铁群.基于MB-LDA模型的微博主题挖掘[J].计算机研究与发展,2011.48(10):1795-1802
- [13] Shi Z M,Lee G,Whinston A B. Toward a Better Measure of Business Proximity: Topic Modeling for Industry Intelligence[J].MIS Quarterly,2016.40(4):1035-1056 