

一种神经网络模型剪枝后泛化能力的验证方法

刘崇阳, 刘勤让

(国家数字交换系统工程技术研究中心, 郑州 450002)

摘要: 针对神经网络模型在剪枝操作中 Dropout 率下调造成的过拟合问题, 提出一种剪枝模型泛化能力的验证方法。研究人为遮挡数据集模拟图像范围的变化情况, 分析不同 Dropout 值和剪枝比例对模型准确率的影响, 进而得到剪枝操作后模型泛化能力变化的原因。在卷积神经网络模型 lenet-5 上进行实验, 结果表明, 剪枝模型泛化能力减弱是因为 Dropout 率下调和剪枝操作时参数数量的变化。

关键词: 深度神经网络; 模型剪枝; 深度学习; 泛化能力; 遮挡数据集

中文引用格式: 刘崇阳, 刘勤让. 一种神经网络模型剪枝后泛化能力的验证方法[J]. 计算机工程, 2019, 45(10): 234-238.

英文引用格式: LIU Chongyang, LIU Qinrang. A verification method on post-pruning generalization ability of Neural Network model [J]. Computer Engineering, 2019, 45(10): 234-238.

A Verification Method on Post-Pruning Generalization Ability of Neural Network Model

LIU Chongyang, LIU Qinrang

(China National Digital Switching System Engineering and Technological R&D Center, Zhengzhou 450002, China)

[Abstract] To address the over-fitting problem caused by the down-regulation of the Dropout rate in the pruning operation of the neural network model, a verification method for the generalization ability of the pruning model is proposed. By artificially occluding the dataset to simulate the change of the image range, the effects of different Dropout values and pruning ratios on the accuracy of the model are analyzed, and the reasons for the change of the generalization ability of the model after pruning operation are obtained. Experiments on the convolutional neural network model lenet-5 show that the generalization ability of the pruning model is weakened because of the drop of the Dropout rate and the change of the parameter amount during the pruning operation.

[Key words] Deep Neural Networks (DNN); model pruning; Deep Learning (DL); generalization ability; occlusion datasets

DOI: 10.19678/j.issn.1000-3428.0052165

0 概述

随着人工智能技术的不断发展, 未来将出现大量廉价、便携以及功耗低的智能设备。机器通过不断学习会变得更加智能, 而深度神经网络 (Deep Neural Networks, DNN) 是机器学习的主要技术之一^[1-2]。虽然深度学习 (Deep Learning, DL) 在学术界的发展速度较快, 但在产业界并未取得重大进展。例如, 深度学习技术在自动驾驶中应用发展多年, 但并未达到商业化要求^[3]。这是因为深度学习算法及模型复杂度高、计算量大, 较难部署在硬件资源有限的嵌入式设备中^[4]。工艺的发展进步是解决该问题的一个途径, 但减小模型参数数量和计算量是更为有效的方法。

文献[5]提出深度压缩 (Deep Compression, DC) 技术, 通过对神经网络模型进行剪枝操作, 将模型权重参

数丢弃, 在剪枝操作完成后, 这些参数固定为 0。由于在剪枝过程中减小了模型容量, 因此需重新调整 Dropout 率。

剪枝操作能够降低模型容量, 防止过拟合^[6-7], 但 Dropout 率降低会导致多模型融合效率降低, 使模型的泛化能力和可拓展性能变差^[8-9]。为此, 本文建立一种基于剪枝的神经网络模型。描述神经网络剪枝操作过程, 并探究其泛化能力, 在此基础上, 以 MNIST 数据集上的 lenet-5^[10] 为模型, 设计 Dropout 和剪枝操作的对比实验。

1 模型剪枝后泛化能力

1.1 模型剪枝操作

随着数据的快速增长, 神经网络训练数据成倍增加。研究人员需要数天甚至数周的时间来训练大型神

基金项目: 国家科技重大专项 (2016ZX01012101); 国家自然科学基金面上项目 (61572520); 国家自然科学基金创新研究群体项目 (61521003)。

作者简介: 刘崇阳 (1994—), 男, 硕士研究生, 主研方向为人工智能、深度学习; 刘勤让, 研究员、博士。

收稿日期: 2018-07-19 **修回日期:** 2018-08-21 **E-mail:** zmylmh1@163.com

神经网络模型,如 Alexnet 的训练时间长达 6 天(基于 2 块 GTX 580 3 GB GPU)^[11]。根据模型的移值性要求,在不影响模型精度的情况下,通过减少参数数量来提高训练效率。文献[5]的神经网络剪枝过程如图 1 所示。

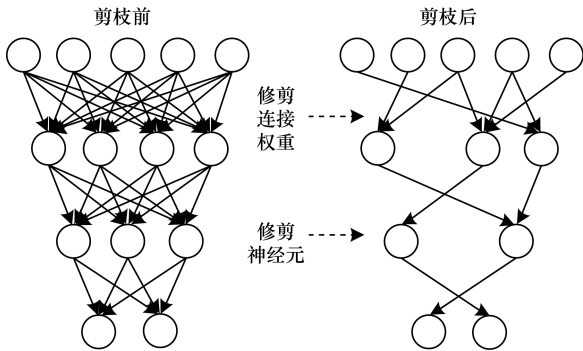


图 1 神经网络剪枝过程

剪枝操作不仅减少了连接权重(图 1 箭头所示),而且神经元节点也会减少(图 1 圆圈所示)。剪枝操作使网络变稀疏,能够简化网络的复杂性。模型复杂性降低会减少过拟合情况的发生,提升模型泛化能力。同时,网络中需要确定的参数量减少。根据训练数据确定模型参数,以进一步降低训练数据的数量。在剪枝后,对稀疏网络进行重新训练,否则模型会有精度损失。重新训练和模型最开始的训练过程类似,已经裁剪掉的不用训练,根据文献[5],剪枝比例一般超过 50%,重新训练的剩余部分只占少数,若继续保持最开始训练时的 Dropout 率,训练出来的模型精度会下降,因此可减少 Dropout 率。

Dropout 层一般放置在全连接层后。在训练过程中,Dropout 层会丢弃一定数量的信息,只有部分数据发挥作用。在预测过程中,Dropout 层将使用所有参数。具体来说,Dropout 将隐含层输出以一定的概率置 0,此神经元不参与前向和后向传播。由于采用随机丢弃的方式,每一次进行前向后向计算时,丢弃掉的数据不同,模型每一次的前向后向计算的表现也不同。Dropout 可以看成是一种组合模型,每个样本训练出不一样的网络结构。Dropout 减少了神经元之间的共适应关系,一个神经元不会依赖另一个神经元,使网络学习到更加鲁棒的特征表示^[12]。改变 Dropout 率具有如下优点:在一定程度上减轻过拟合的情况;使模型具有多模型融合的效果,即提高模型的拓展性和泛化能力。

1.2 泛化能力

神经网络的泛化能力是训练好的网络对于不在训练集内数据的拟合能力^[13-14]。目前,神经网络被广泛应用于计算机视觉,本文以视觉领域最基础的分类任务来探究其泛化能力。分类问题中存在一个十分经典的问题,即遮挡问题^[15]。本文研究基于遮挡问题的神经网络模型泛化能力。人类具有利用部分信息进行分析推断的能力。对于人类来说,如果遮挡物体的一部分,只要不是遮挡住最关键的信息,就可以通过局部信息识别一个完整的物体。计算机

如果有人类的分析推理能力,其应具有较好的泛化能力。神经网络中最简单的手写数字体识别是一个分类问题,本文结合遮挡问题来探究其泛化能力。

手写数字体识别的数据集是 0~9 的数字,以灰度图形式表示。灰度图的每个像素点用 0~255 的数字表示灰度值,值越大越接近纯白色。图 2 所示为数字 1 的灰度图和像素点矩阵形式。



图 2 数字 1 的灰度图和矩阵形式

如果将图 2(a)中数字 1 的上半部分全置 0,就达到了遮挡的效果,如图 3 所示。

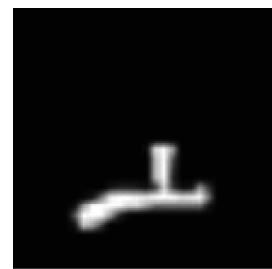


图 3 遮挡上半部分的数字 1

在图 3 中,数字 1 较容易确认。图 4 所示为另外一种形式的数字 1,可以看出,若遮挡上半部分,无法确认是数字 1 还是数字 7。基于图 2~图 4 所描述的遮挡过程,本文研究训练集和测试集有差异时神经网络表现出的泛化能力。

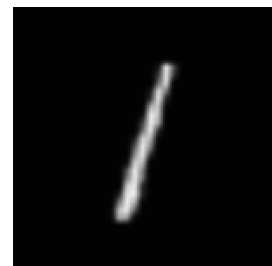


图 4 数字 1 的另一种形式

为探究剪枝后泛化能力的变化,本文在未剪枝模型上测试 Dropout 率变化对泛化能力的影响,对重新训练好的模型进行泛化能力实验,通过对比分析验证剪枝对泛化能力的影响。

2 实验结果与分析

2.1 实验设置

本文实验基于 Caffe 深度学习框架,使用 MNIST 数据集,包含 60 000 张训练数据和 10 000 张测试数据,采用 lenet-5 手写数字体识别卷积神经网络。Caffe 官网文件库(Model Zoo)没有经过训练的 lenet-5 模型权重文件,本文根据下载的 MNIST 数据集训练得到模型权重文件(Caffe 中以 caffemodel 形式存储)。实验计算机配置为: Intel® Core i5-4210M 双核处理器,内存 8 GB RAM, Ubuntu16.04, 利用 Python 脚本实现数据集的处理和剪枝操作。从官网下载的 MNIST 数据集转换为图片格式,然后对图片进行遮挡操作,再构建 txt 标签,生成可供训练(mnist_train_lmdb)和测试(mnist_test_lmdb)的 lmdb 格式的数据集。遮挡操作遮挡数字图上半部分, Dropout 层加在 lenet-5 的第一个全连接层后面, Dropout 率从 0.0 ~ 0.9 间隔 0.1 变化,剪枝操作逐层进行,剪枝极限依据文献[5]提供的数据。所有实验都进行 10 次,计算平均值作为最终结果。

2.2 结果分析

实验 1 Caffe 自带的 lenet-5 网络模型中没有 Dropout 层,在完整的 MNIST 训练集和测试集上 lenet-5 模型的准确率 $acc = 0.9911$ ($Dropout = 0.0$),在第一个全连接层加 Dropout 层,调节 Dropout 率大小,实验结果如表 1 所示。

表 1 不同 Dropout 下的准确率对比 %

Dropout	准确率
0.0	0.9911
0.1	0.9915
0.2	0.9910
0.3	0.9909
0.4	0.9920
0.5	0.9919
0.6	0.9921
0.7	0.9913
0.8	0.9917
0.9	0.9892

从表 1 可以看出,不同 Dropout 的准确率变化程度较小,因此,调整 Dropout 对准确率影响较小。这可能由于 Caffe 自带改进的 lenet-5 模型文件中不含 Dropout 层。

实验 2 训练集保持完整,测试集遮挡上半部分,未加 Dropout 层的准确率 $acc = 0.4593$ 。将

Dropout 层加在第一个全连接层后,以 3 种模式进行训练,具体分析如下:

1) 以随机初始化的方式进行训练,由于分类结果有 10 个,准确率会从 10% 逐渐提升。

2) 以完整训练测试集训练出一个 caffemodel,再在 caffemodel 的基础上进行微调,即将预先训练好的模型参数作为初始化参数。

3) 在训练测试集遮挡上半部分的数据集上训练出一个 caffemodel,然后进行微调。前 2 种模式效果一样,因为训练过程是基于训练集进行调参,参数的确定由训练集决定,测试集给出模型准确率,模式 1 和模式 2 根据同一个训练集而来,模式 2 在模式 1 的基础上多训了一遍,相当于同一种类型。改变 Dropout 率的结果如表 2 和表 3 所示。

表 2 完整训练集不同 Dropout 下的准确率对比 1

Dropout	准确率
0.0	0.4593
0.1	0.4609
0.2	0.4797
0.3	0.4859
0.4	0.4949
0.5	0.5034
0.6	0.5011
0.7	0.5125
0.8	0.5189
0.9	0.5206

表 3 遮挡训练集不同 Dropout 下的准确率对比 1

Dropout	准确率
0.0	0.8573
0.1	0.8290
0.2	0.7875
0.3	0.7592
0.4	0.7483
0.5	0.7292
0.6	0.7054
0.7	0.6731
0.8	0.6258
0.9	0.5641

从表 2 可以看出,未加 Dropout 层的准确率 $acc = 0.4593$,比实验 1 的准确率低。在训练的过程中,根据终端的快照,某些轮次的测试集准确率比 0.4593 要高一些,但是会出现越训练效果不升反降的情况。这是由于训练集和测试集的数据分布和信息容量不同。在训练时,使用了全部的数据信息,在识别时每个位置都会被当作识别的特征加以提取。在测试时,被遮挡后只有一半的数据,有确定的特征突然消失,导致神经网络模型准确率下降。此外,表 2

中准确率呈上升趋势,说明 Dropout 能够使模型具有较高的泛化能力。

实验 3 训练集遮挡上半部分,测试集完整,Dropout 层加在第一个全连接层后。在 2 种不同的 caffemodel 基础上进行微调,结果如表 4 和表 5 所示。

表 4 完整训练集不同 Dropout 下的准确率对比 2

Dropout	准确率
0.0	0.968 0
0.1	0.957 4
0.2	0.952 3
0.3	0.952 1
0.4	0.950 6
0.5	0.949 7
0.6	0.947 0
0.7	0.939 3
0.8	0.930 6
0.9	0.901 0

表 5 遮挡训练集不同 Dropout 下的准确率对比 2

Dropout	准确率
0.0	0.851 9
0.1	0.849 8
0.2	0.849 3
0.3	0.845 8
0.4	0.845 5
0.5	0.845 2
0.6	0.840 8
0.7	0.839 4
0.8	0.833 9
0.9	0.829 4

从表 5 可以看出,随着 Dropout 率上升,准确率逐渐下降。表 5 和表 2 表示结果相反,2 个实验的测试集和训练集都不同,实验 3 的测试数据比实验 2 测试数据多 1 倍。可以看出,测试数据较多时,提高 Dropout 率会放弃更多的特征,使模型的准确率变差。如果实验 3 对应的测试集遮挡 3/4,则 Dropout 率越高准确率越高。表 5 和表 2 对比也可以说明 Dropout 的泛化能力提升仅针对测试集变小的情况。

表 2~表 5 实验具体情况如图 5 所示。表 3 和表 4 两者初始的 caffemodel 和测试集保持一致,训练过程以增大(减少)一半的训练集进行调参,调参的越多就会和测试集偏差越多,因此增大 Dropout 率会放弃更多和测试集相同的特征,导致准确率一直下降。

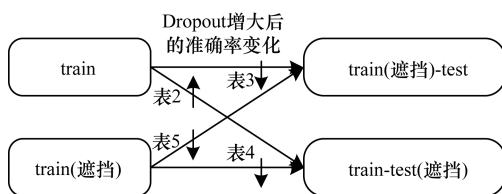


图 5 表 2~表 5 对应的 4 种情况示意图

实验 4 不加 Dropout 层,同时遮挡训练和测试集上半部分后的模型准确率 $acc = 0.912 4$ 。加入 Dropout 并改变其大小,准确率基本维持不变。测试集被遮挡一半,如果把训练数据也变成只有一半,与表 2 中 Dropout 为 0 的情况相比,不加 Dropout 层,训练测试集同时遮挡上半部分后的模型准确率 $acc = 0.912 4$ 。

实验 5 对于剪枝后的模型,本文实验均未加 Dropout 层。剪枝极限参考文献[5],本文稍有改动,剪枝率的参数值设置如下:第 1 个卷积层为 33%,第 2 个卷积层为 80%,第 1 个全连接层为 90% 及第 2 个全连接层为 80%。根据完整的 MNIST 训练集和测试集出来的 lenet-5 剪枝模型准确率 $acc = 0.988 9$ 。向下调整剪枝比例后准确率变化不大,保持在 0.99 附近。加入 Dropout 层准确率基本没有变化。实验结果表明,适当剪枝不会影响准确率。

实验 6 训练集保持完整,测试集遮挡上半部分,在 2 种不同的 caffemodel 基础上进行微调,并调整剪枝比例,得到的结果如表 6 和表 7 所示。剪枝模型在一个固定剪枝比例的基础上改变 Dropout 后的准确率和表 2、表 3 情况类似,不再赘述(以 33%、80%、90%、80% 比例剪枝的模型逐渐增大 Dropout 率,模型准确率呈上升趋势)。

表 6 完整训练集不同剪枝比例下的准确率对比 1

剪枝比例/%	准确率
33,80,90,80	0.413 2
0,80,90,80	0.413 8
33,0,90,80	0.423 7
33,80,0,80	0.426 8
33,80,80,0	0.450 3
16,40,90,80	0.421 3
33,80,45,40	0.452 4

表 7 遮挡训练集不同剪枝比例下的准确率对比 1

剪枝比例/%	准确率
33,80,90,80	0.837 3
0,80,90,80	0.839 3
33,0,90,80	0.840 8
33,80,0,80	0.830 2
33,80,80,0	0.846 0
16,40,90,80	0.838 5
33,80,45,40	0.837 8

从表 6 可以看出,经过剪枝的准确率小于未剪枝时的准确率(表 2 中的 0.459 3),说明剪枝过程本身会降低模型的泛化能力。这是由于剪枝模型权重参数减少,放弃了更多的细微特征,虽然不是主要特征,但在识别阶段提取特征时,很多微小特征也有可能具有明显的相同特征,导致模型准确率降低。加上 Dropout 后和表 2 情况一样,则对于本来有 Dropout 层的网络模型,剪枝必然伴随 Dropout 率的下调,剪枝后模型的泛化能力还会进一步降低。另外,各层剪枝后

的泛化能力稍有区别,例如对单个层不进行剪枝,模型准确率有较大区别,对比表6中第2行~第7行数据,可以得出全连接层比卷积层敏感的结论。

实验7 训练集遮挡上半部分,测试集完整,在2种不同的caffemodel基础上进行微调,并调整剪枝比例,得到的结果如表8和表9所示。剪枝模型固定剪枝率后调节Dropout后情况和表4、表5情况类似,不再赘述。

表8 完整训练集不同剪枝比例下的准确率对比2

剪枝比例/%	准确率
33,80,90,80	0.959 8
0,80,90,80	0.949 5
33,0,90,80	0.958 3
33,80,0,80	0.955 9
33,80,80,0	0.960 5
16,40,90,80	0.958 8
33,80,45,40	0.960 8

表9 遮挡训练集不同剪枝比例下的准确率对比2

剪枝比例/%	准确率
33,80,90,80	0.825 1
0,80,90,80	0.821 2
33,0,90,80	0.830 2
33,80,0,80	0.837 3
33,80,80,0	0.832 2
16,40,90,80	0.834 0
33,80,45,40	0.849 7

从表9可以看出,经过剪枝的准确率小于未剪枝时的准确率(表5中的0.8519)。说明模型剪枝后泛化能力减弱不仅发生在测试集缩小的情况,放大测试集后其泛化能力也会减弱。另外,各层剪枝后的泛化能力稍有区别,与表6结论一致,不再赘述。

实验8 训练集和测试集都遮挡上半部分,模型准确率 $acc = 0.9102$ 。往下调整剪枝比例后准确率变化不大,保持在0.91附近,加入Dropout层准确率也基本没有变化。实验8的结论同实验5的结论。

表7和表8与表3和表4中Dropout为0的情况相比,准确率都降低。准确率下降的原因同实验6,不再赘述。

综上,通过对比表2、表5、表6和表9,可以得到模型的泛化能力:Dropout提升泛化能力仅对测试集变小的情况有效;对于模型剪枝,所有情况的准确率比未剪枝的低,泛化能力比未剪枝的弱。因此,如果模型有Dropout层,在剪枝操作中下调Dropout率,能够进一步降低泛化能力。

3 结束语

剪枝操作中Dropout率下调会给模型带来负面影响。为此,本文提出一种神经网络模型泛化能力的验证方法。在lenet-5模型上的实验结果表明,剪

枝后的模型泛化能力减弱与Dropout率下调和参数量减少有关。下一步将在神经网络模型上研究Dropout层的位置对剪枝模型泛化能力的影响。

参考文献

- [1] 周飞燕,金林鹏,董军. 卷积神经网络研究综述[J]. 计算机学报,2017,40(6):1229-1251.
- [2] HAN Song. Efficient methods and hardware for deep learning[D]. San Francisco, USA: Stanford University, 2017.
- [3] 张新钰,高洪波,赵建辉,等. 基于深度学习的自动驾驶技术综述[J]. 清华大学学报(自然科学版),2018,58(4):438-444
- [4] HAN Song,JEFF P,JOHN T, et al. Learning both weights and connections for efficient neural networks [C]// Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press,2015:1135-1143.
- [5] HAN Song,MAO Huizi,DALLY W J. Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding [EB/OL]. [2018-06-18]. <https://arxiv.org/pdf/1510.00149.pdf>.
- [6] ROSA G H D,PAPA J P,YANG X S. Handling dropout probability estimation in convolution neural networks using meta-heuristics[J]. Soft Computing,2017,22(18):6147-6156.
- [7] POERNOMO A,KANG D K. Biased dropout and crossmap dropout: learning towards effective dropout regularization in convolutional neural network [J]. Neural Networks,2018,104:60-67.
- [8] MESHGI K,MAEDA S I,OBA S, et al. Constructing a meta-tracker using Dropout to imitate the behavior of an arbitrary black-box tracker. [J]. Neural Networks,2017,87:132-148.
- [9] SRIVASTAVA N,HINTON G,KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. The Journal of Machine Learning Research,2014,15(1):1929-1958.
- [10] 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.
- [11] KRIZHEVSKY A,SUTSKEVER I,HINTON G E. ImageNet classification with deep convolutional neural networks [C]//Proceedings of the 25th International Conference on Neural Information Processing Systems. [S.l.]:Curran Associates Inc.,2012:1097-1105.
- [12] HINTON G E,SRIVASTAVA N,KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors [J]. Computer Science,2012,3(4):212-223.
- [13] ZHANG Chiyuan,BENGIO S,HARDT M, et al. Understanding deep learning requires rethinking generalization[EB/OL]. [2018-06-18]. <https://arxiv.org/pdf/1611.03530.pdf>.
- [14] CHEN Binghui,DENG Weihong,DU Junping. Noisy softmax: improving the generalization ability of DCNN via postponing the early softmax saturation [EB/OL]. [2018-06-18]. <https://arxiv.org/pdf/1708.03769.pdf>.
- [15] 李小薪,梁荣华. 有遮挡人脸识别综述:从子空间回归到深度学习[J]. 计算机学报,2018,41(1):177-207.