

深度学习模型训练过程



吴建明wu...

人工智能芯片与计算机视觉专家

60 人赞同了该文章

深度学习模型训练过程

一. 数据准备

1. 基本原则：

- 1) 数据标注前的标签体系设定要合理
- 2) 用于标注的数据集需要无偏、全面、尽可能均衡
- 3) 标注过程要审核

2. 整理数据集

- 1) 将各个标签的数据放于不同的文件夹中，并统计各个标签的数目
- 2) 样本均衡，样本不会绝对均衡，差不多就行了

3) 切分样本集

如：90%用于训练，10%留着测试，比例自己定。训练集合，对于弱势类要重采样，最后的图片列表要shuffle；测试集合就不用重采样了。

训练中要保证样本均衡，学习到弱势类的特征，测试过程要反应真实的数据集分布。

4) 按需要的格式生成tfrecord

按照train.list和validation.list生成需要的格式。生成和解析tfrecord的代码要根据具体情况编写。

二. 基本步骤

1. 定义算法公式，也就是神经网络的前向算法。我们一般使用现成的网络，如inceptionV4，mobilenet等。
2. 定义loss，选择优化器，来让loss最小
3. 对数据进行迭代训练，使loss到达最小



三. 训练

1. 预处理, 根据自己的喜好, 编写预处理策略。

preprocessing的方法, 变换方案诸如: 随机裁剪、随机变换框、添加光照饱和度、修改压缩系数、各种缩放方案、多尺度等。进而, 减均值除方差或归一化到 $[-1, 1]$, 将float类型的Tensor送入网络。

这一步的目的是: 让网络接受的训练样本尽可能多样, 不要最后出现原图没问题, 改改分辨率或宽高比就跪了的情况。

2. 网络设计, 基础网络的选择和Loss的设计。

基础网络的选择和问题的复杂程度息息相关, 用ResNet18可以解决的没必要用101; 还有一些SE、GN等模块加上有没有提升也可以去尝试。

Loss的设计, 一般问题的抽象就是设计Loss数据公式的过程。比如多任务中的各个任务权重配比, center Loss可以让特征分布更紧凑, SmoothL1Loss更平滑避免梯度爆炸等。

3. 优化算法

一般来说, 只要时间足够, Adam和SGD+Momentum可以达到的效果差异不大。用框架提供的理论上最好的优化策略就是了。

4. 训练过程

finetune网络, 我习惯分两步: 首先训练fc层, 迭代几个epoch后保存模型; 然后基于得到的模型, 训练整个网络, 一般迭代40-60个epoch可以得到稳定的结果。total_loss会一直下降的, 过程中可以评测下模型在测试集上的表现。真正的loss往往包括两部分。后面total_loss的下降主要是正则项的功劳了。

四. 难点

1. 训练速度慢, 收敛慢

深度学习其实就是一个反复调整模型参数的过程, 得力于GPU等硬件性能的提升, 使得复杂的深度学习训练成为了可能。收敛速度过慢, 训练时间过长, 一方面使得相同总训练时间内的迭代次数变少, 从而影响准确率, 另一方面使得训练次数变少, 从而减少了尝试不同超参数的机会。

1. 设置合理的初始化权重 w 和偏置 b

$$f(x) = w * x + b$$

2. 优化学习率

型训练就是不断尝试和调整不同的 w 和 b , 那么每次调整的幅度是多少呢, 这个就是学习率。



神经网络训练时，每一层的输入分布都在变化。不论输入值大还是小，我们的学习率都是相同的，这显然是很浪费效率的。而且当输入值很小时，为了保证对它的精细调整，学习率不能设置太大。那有没有办法让输入值标准化得落到某一个范围内，比如[0, 1]之间呢，这样我们就再也不必为太小的输入值而发愁了。

办法当然是有的，那就是正则化！由于我们学习的是输入的特征分布，而不是它的绝对值，故可以对每一个mini-batch数据内部进行标准化，使他们规范化到[0, 1]内。这就是Batch Normalization，简称BN。由大名鼎鼎的inception V2提出。它在每个卷积层后，使用一个BN层，从而使得学习率可以设定为一个较大的值。使用了BN的inceptionV2，只需要以前的1/14的迭代次数就可以达到之前的准确率，大大加快了收敛速度。

4. 采用更先进的网络结构，减少参数量

1) 使用小卷积核来代替大卷积核。

VGGNet全部使用3x3的小卷积核，来代替AlexNet中11x11和5x5等大卷积核。小卷积核虽然参数量较少，但也会带来特征面积捕获过小的问题。inception net认为越往后的卷积层，应该捕获更多更高阶的抽象特征。因此它在靠后的卷积层中使用的5x5等大面积的卷积核的比率较高，而在前面几层卷积中，更多使用的是1x1和3x3的卷积核。

2) 使用两个串联小卷积核来代替一个大卷积核。

inceptionV2中创造性的提出了两个3x3的卷积核代替一个5x5的卷积核。在效果相同的情况下，参数量仅为原先的 $3 \times 3 \times 2 / 5 \times 5 = 18/25$

3) 1x1卷积核的使用。

1x1的卷积核可以说是性价比最高的卷积了，没有之一。它在参数量为1的情况下，同样能够提供线性变换，relu激活，输入输出channel变换等功能。VGGNet创造性的提出了1x1的卷积核

3) 非对称卷积核的使用。

inceptionV3中将一个7x7的卷积拆分成了一个1x7和一个7x1, 卷积效果相同的情况下，大大减少了参数量，同时还提高了卷积的多样性。

4) depthwise卷积的使用。

mobileNet中将一个3x3的卷积拆分成了串联的一个3x3 depthwise卷积和一个1x1正常卷积。对于输入channel为M，输出为N的卷积，正常情况下，每个输出channel均需要M个卷积核对输入的每个channel进行卷积，并叠加。也就是需要MxN个卷积核。而在depthwise卷积中，输出chanr



5) 全局平均池化代替全连接层。

这个才是大杀器！AlexNet和VGGNet中，全连接层几乎占据了90%的参数量。inceptionV1创造性的使用全局平均池化来代替最后的全连接层，使得其在网络结构更深的情况下（22层，AlexNet仅8层），参数量只有500万，仅为AlexNet的1/12。

发布于 2020-05-03 19:29

深度学习 (Deep Learning)

统计学习

机器学习

▲ 赞同 60 ▼

💬 1 条评论

➦ 分享

♥ 喜欢

★ 收藏

📄 申请转载

...

文章被以下专栏收录



自动驾驶工坊

自动驾驶中的感知、定位以及路径规划相关算法

推荐阅读

30分钟了解PAC学习理论——计算学习理论第一讲

马瘦毛长蹄子肥，搞机器学习的谁也不服谁，自从二狗学了PAC，见谁都拿拳往别人小胸脯上捶。啪！今天给大家说一段PAC理论 先放PAC学习相关理论的一个总结：同等条件下，模型越复杂泛化...

小宝剑

CNN是靠什么线索学习到深度信息的？——一个经验性探索

今天给大家带来一篇我很喜欢的风格的paper，详细分析了基于深度学习的单目深度估计算法的局限性。虽然这篇文章没有提出新的方法，但是对于我们理解CNN是如何学习到一些appearance cue来处理需...

Naiya...

发表于Winst...



《神经
- 期望

你说的

1 条评论

⇌ 切换为时间排序

写下你的评论...



问一下，vgg19模型也是先训练全连接层吗

 赞

