



深度学习网络调参技巧



萧瑟 · 1 年前

转载请注明：[炼丹实验室](#)

之前曾经写过一篇文章，讲了一些深度学习训练的技巧，其中包含了部分调参心得：[深度学习训练心得](#)。不过由于一般深度学习实验，相比普通机器学习任务，时间较长，因此调参技巧就显得尤为重要。同时个人实践中，又有一些新的调参心得，因此这里单独写一篇文章，谈一下自己对深度学习调参的理解，大家如果有其他技巧，也欢迎多多交流。

好的实验环境是成功的一半

由于深度学习实验超参众多，代码风格良好的实验环境，可以让你的人工或者自动调参更加省力，有以下几点可能需要注意：

- 将各个参数的设置部分集中在一起。如果参数的设置分布在代码的各个地方，那么修改的过程想必会非常痛苦。
- 可以输出模型的损失函数值以及训练集和验证集上的准确率。
- 可以考虑设计一个子程序，可以根据给定的参数，启动训练并监控和周期性保存评估结果。再由一个主程序，分配参数以及并行启动一系列子程序。



画图是一个很好的习惯，一般是训练数据遍历一轮以后，就输出一下训练集和验证集准确率。同时画到一张图上。这样训练一段时间以后，如果模型一直没有收敛，那么就可以停止训练，尝试其他参数了，以节省时间。

如果训练到最后，训练集，测试集准确率都很低，那么说明模型有可能欠拟合。那么后续调节参数方向，就是增强模型的拟合能力。例如增加网络层数，增加节点数，减少dropout值，减少L2正则值等等。

如果训练集准确率较高，测试集准确率比较低，那么模型有可能过拟合，这个时候就需要向提高模型泛化能力的方向，调节参数。

从粗到细分阶段调参

实践中，一般先进行初步范围搜索，然后根据好结果出现的地方，再缩小范围进行更精细的搜索。

1. 建议先参考相关论文，以论文中给出的参数作为初始参数。至少论文中的参数，是个不差的结果。
2. 如果找不到参考，那么只能自己尝试了。可以先从比较重要，对实验结果影响比较大的参数开始，同时固定其他参数，得到一个差不多的结果以后，在这个结果的基础上，再调其他参数。例如学习率一般就比正则值，dropout值重要的话，学习率设置的不合适，不仅结果可能变差，模型甚至会无法收敛。
3. 如果实在找不到一组参数，可以让模型收敛。那么就需要检查，是不是其他地方出了问题，例如模型实现，数据等等。可以参考我写的[深度学习网络调试技巧](https://zhuanlan.zhihu.com/p/24720954)



调参只是为了寻找合适的参数，而不是产出最终模型。一般在小数据集上合适的参数，在大数据集上效果也不会太差。因此可以尝试对数据进行精简，以提高速度，在有限的时间内可以尝试更多参数。

- 对训练数据进行采样。例如原来100W条数据，先采样成1W，进行实验看看。
- 减少训练类别。例如手写数字识别任务，原来是10个类别，那么我们可以先在2个类别上训练，看看结果如何。

超参数范围

建议优先在对数尺度上进行超参数搜索。比较典型的是学习率和正则化项，我们可以从诸如 0.001 0.01 0.1 1 10，以10为阶数进行尝试。因为他们对训练的影响是相乘的效果。不过有些参数，还是建议在原始尺度上进行搜索，例如dropout值: 0.3 0.5 0.7)。

经验参数

这里给出一些参数的经验值，避免大家调参的时候，毫无头绪。

- learning rate: 1 0.1 0.01 0.001, 一般从1开始尝试。很少见learning rate大于10的。学习率一般要随着训练进行衰减。衰减系数一般是0.5。 衰减时机，可以是验证集准确率不再上升时，或固定训练多少个周期以后。
不过更建议使用自适应梯度的办法，例如adam,adadelta,rmsprop等，这些一般使用相关论文提供的默认值即可，可以避免再费劲调节学习率。对RNN来说，有个经验，如果



- 网络层数：先从1层开始。
- 每层结点数：16 32 128，超过1000的情况比较少见。超过1W的从来没有见过。
- batch size: 128上下开始。batch size值增加，的确能提高训练速度。但是有可能收敛结果变差。如果显存大小允许，可以考虑从一个比较大的值开始尝试。因为batch size太大，一般不会对结果有太大的影响，而batch size太小的话，结果有可能很差。
- clip c(梯度裁剪): 限制最大梯度,其实是 $value = \sqrt{w_1^2 + w_2^2 + \dots}$,如果value超过了阈值，就算一个衰减系数,让value的值等于阈值: 5,10,15
- dropout：0.5
- L2正则：1.0，超过10的很少见。
- 词向量embedding大小：128，256
- 正负样本比例：这个是非常忽视，但是在很多分类问题上，又非常重要的参数。很多人往往习惯使用训练数据中默认的正负类别比例，当训练数据非常不平衡的时候，模型有可能会偏向数目较大的类别，从而影响最终训练结果。除了尝试训练数据默认的正负类别比例之外，建议对数目较小的样本做过采样，例如进行复制。提高他们的比例，看看效果如何，这个对多分类问题同样适用。
在使用mini-batch方法进行训练的时候，尽量让一个batch内，各类别的比例平衡，这个在图像识别等多分类任务上非常重要。

自动调参

人工一直盯着实验，毕竟太累。自动调参当前也有不少研究。下面介绍几种比较实用的办法：



的话，结果比较可靠。缺点是太费时间了，特别像神经网络，一般尝试不了太多的参数组合。


- Random Search. Bengio在[Random Search for Hyper-Parameter Optimization](#)中指出，Random Search比Grid Search更有效。实际操作的时候，一般也是先用Grid Search的方法，得到所有候选参数，然后每次从中随机选择进行训练。
- Bayesian Optimization. 贝叶斯优化，考虑到了不同参数对应的实验结果值，因此更节省时间。和网络搜索相比简直就是老牛和跑车的区别。具体原理可以参考这个论文：[Practical Bayesian Optimization of Machine Learning Algorithms](#)，这里同时推荐两个实现了贝叶斯调参的Python库，可以上手即用：
 - [jaberg/hyperopt](#), 比较简单。
 - [fmfn/BayesianOptimization](#)，比较复杂，支持并行调参。

总结

- 合理性检查，确定模型，数据和其他地方没有问题。
- 训练时跟踪损失函数值，训练集和验证集准确率。
- 使用Random Search来搜索最优超参数，分阶段从粗（较大超参数范围训练较少周期）到细（较小超参数范围训练较长周期）进行搜索。

参考资料

知

首发于
炼丹实验室 写文章[登录](#)

io (2012)

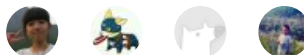
Efficient BackProp, by Yann LeCun, Léon Bottou, Genevieve Orr and Klaus-Robert Müller

Neural Networks: Tricks of the Trade, edited by Grégoire Montavon, Geneviève Orr, and Klaus-Robert Müller.

「真诚赞赏，手留余香」

[赞赏](#)

4 人赞赏



深度学习 (Deep Learning)

机器学习

数据挖掘

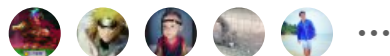


380

收藏

分享

举报



知



首发于
炼丹实验室

写文章

登录



炼丹实验室
深度学习实战经验介绍

进入专栏

20 条评论

写下你的评论...



安宁 回复 马克弗里曼

查看对话

有可能是某些技术只是training部分实现了没有在validation部分实现。我之前碰到这种情况的时候是standardize只在training的时候实现了，validation没做，所以training正常进行，validation的结果没什么变化。

1 年前

以上为精选评论



math love

请问学NLP 需要怎样的机器配置(软硬件)

1 年前



萧瑟（作者） 回复 math love

查看对话

知

首发于
炼丹实验室 写文章[登录](#)

马克弗里曼

请问如果从训练一开始，train loss持续下降，但是validation loss一直都不降，会是什么原因

1 年前



七月

厉害了.....另外请问下，RNN 用什么 Optimizer 比较好？我以前做一个实验，用 RMSProp（默认参数）有时候会 Nan，是因为默认的学习率太大吗？还是我的序列太长了（200 个时间步）？或者是其他原因？谢谢！

1 年前



萧瑟（作者） 回复 七月

 查看对话

分类问题可以考虑adadelata,生成文本可以考虑adam. Nan问题，可以参考我写的这个文章：

[知乎专栏](#)

1 年前



萧瑟（作者） 回复 马克弗里曼

 查看对话

有可能是你的validation数据有问题，和训练集差异太大了。要么就是代码有bug

1 年前

知



首发于
炼丹实验室

写文章

登录

1 年前



七月 回复 萧瑟 (作者)

查看对话

好的，非常感谢！

1 年前



七月 回复 甘小楚

查看对话

半年前的 Keras。现在用 tf 还没试过 RNN.....

1 年前

1

2

下一页

推荐阅读

Theano调试技巧

转载请注明：炼丹实验室Theano是最老牌的深度学习库之一。它灵活的特点使其非常适合学术研究和快速实验，但是它难以调试的问题也遭到过无数吐槽。其实Theano本身提供了很多



深度学习模型使用word2vec向量的方法总结

转载请注明：炼丹实验室使用word2vec工具在大规模外部文本语料上训练得到的向量，可以比较精确的衡量词之间的相关程度。一个比较简单的应用，就是利用词之间的向量的cos得分，来找相关词。... [查看全文](#) >

萧瑟 · 1 年前 · 发表于 炼丹实验室

机器学习和深度学习资源汇总

之前写过一个机器学习入门的书单，这段时间又陆陆续续看了一些书，也接触到了深度学习等热门研究方向，因此将所有读过和没读过的书都列出来，供日后查阅。网络公开课：麻省理工公开课 单变... [查看全文](#) >

ocsponge · 5 个月前



「机器学习」和「深度学习」的 Cheat Sheets


导读：机器学习和深度学习对于新手来说是有点难度的，深度学习类库也同样难以理解。作者在... [查看全文](#) >

Helperhaps · 7 个月前

知



首发于
炼丹实验室

 写文章

登录