# 我在2017年学到了8个深入学习的最佳实践

我很高兴在2017年完成的一件事是更实际地参与到现代人工智能中。我学了很多数学，这当然很有趣，但没有做任何实际的项目，因此我的努力没有任何表现。为了弥补这一点，我在4月份申请了一个，目的是为肯尼亚的语音建立跳过语法模型。我在第一轮就进入了决赛，但未能赢得一笔奖金。

然后，今年9月，我申请了国际奖学金，这是一个现在已经完成的课程，由杰里米·霍华德教授的第1部分第2版《程序员的实践深度学习》。它将在2018年1月的前两周左右作为MOOC公开发布。在7周的时间里，我学会了如何使用8种建筑技术：

1. 来自预训练模型的世界级图像分类器，
2. 情绪分析工具，通过从调查数据集构建语言模型，
3. 如何深入学习结构化数据集，以及
4. 如何利用深度学习通过协同过滤构建推荐引擎。

所有这些都是通过被称为Juyter笔记本的非常有生产力的接口完成的，它由深学习库支持，它本身由Py火炬支持。

下面是8个技巧。在每种情况下，我都会用简短的fastai代码来概括总体思路，并指出（在括号中）每种技术是否普遍适用（无论您是在进行图像识别和分类、NLP、建模结构化数据还是协作过滤方面的DL都很有用），或者更具体地说是针对您尝试的数据类型深入学习。在课堂上，使用Kaggle挑战完成了图像识别课程。

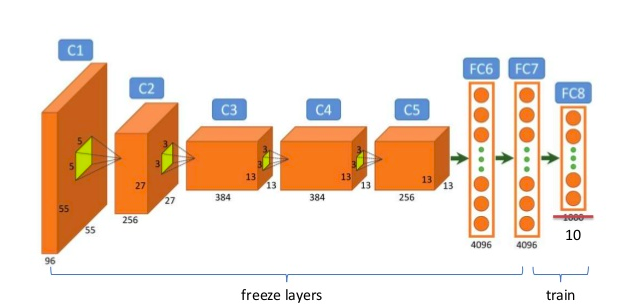
因为这是互联网，我自己做了一个挑战，那就是克隆狗和猫，但用蜘蛛和蝎子代替。蜘蛛和蝎子。我通过搜索“蜘蛛”和“沙漠蝎子”并下载大约1500张图片，从谷歌图片中获取了这些数据。很明显，我把这个自动化了。我不想自动成为深奥蜘蛛和右击的专家。我通过删除任何非jpg图像、任何非图像积垢和没有扩展名的图像来清理数据。当我过滤掉这些垃圾的时候，我有815张图片要处理。每个班[蜘蛛，蝎子]在训练集中有290只蜘蛛，在测试/验证集中有118只蜘蛛和117只蝎子。令人惊讶的是（对我来说，当我学会这一点的时候），它是有效的！我的模型准确率达到了95%。

#### 通过微调VGG-16和ResNext50进行转移学习。（计算机视觉和图像分类）

对于图像分类工作，您可以通过微调获得很多里程数，对于您的特定问题，神经网络架构在更一般的挑战上做得很好。一个例子是残差网络，一个50层的卷积神经网络。它接受了1000类ImageNet挑战的训练，由于表现非常出色，它能够从图像数据中提取的特征足够通用，可以重用。为了让它对我的问题域起作用，我需要做的是将最后一层（输出1000维ImageNet预测向量）替换为输出二维向量的层。这两个输出类在上面代码段中名为PATH的文件夹中指定。对于蜘蛛与蝎子的较量，我有以下几点：

注意，train文件夹的两个内容本身就是文件夹，每个文件夹包含290个图像。

这里显示了微调过程的示例图，它将返回一个10维的最终层：



#### 周期性学习率（一般适用）

学习率可能是。在非适应性环境（即不使用Adam、AdaDelta或其变体）中，通常是由一名DL实践者/研究者进行的，同时进行多个实验，每个实验的学习率之间有一个很小的delta差异。如果你有一个大的数据集，而且很容易出错，如果你没有经验用随机矩阵建立直觉，这需要非常长的时间。然而，2015年，美国海军研究实验室的莱斯利•N•史密斯（Leslie N.Smith）从一个非常小的值开始，通过网络运行几个小批量，在跟踪损失变化的同时调整学习率，直到损失开始减少。fast.ai同学的两篇博文解释了周期性学习率方法是和。

在fastai中，您可以利用学习率退火的优势，在learner对象上运行lr\_find（）和sched.plot（）来确定与最佳学习率一致的点。截图：

周期性学习率论文的一个摘录表明，它表现得更好，达到了最高的精度，并且比让学习率指数衰减快两倍多。

#### 带重启的随机梯度下降（一般适用）

另一种加速随机梯度下降的方法是随着训练的进行逐渐降低学习率。这有助于发现学习率的变化与损失的改善是一致的。当你接近最优权重时，你需要采取较小的步骤，因为如果你采取较大的步骤，你可能会跳过误差面的最佳区域。如果学习率和损失之间的关系是不稳定的，也就是说，如果学习率的小变化导致损失的大变化，那么我们就不在一个稳定的区域（上图2）。然后策略变成周期性地提高学习率。这里的“周期”是一个数字，它决定了提高学习率的次数。这是周期性学习率表。在fastai中，这是使用cycle\_len和cycle\_mult params to learner.fit设置的。在上面的图2中，学习速率被重置3次。当使用正常的学习速率计划时，通常需要更长的时间才能找到最佳损失，在正常的学习速率计划中，开发人员要等到所有时间段都完成后，才能使用不同的学习速率手动重试。

#### 数据增强（目前为计算机视觉和图像分类）

数据扩充是增加训练和测试数据量的一种简单方法。对于图像，这取决于手头的学习问题，因此也取决于数据集中图像的对称性数量。例如蜘蛛与蝎子的玩具挑战赛。这个数据集中的许多图片可以垂直反射，并且仍然显示动物，没有奇怪的扭曲。这就是所谓的变换。例子：

#### 测试时间增加（目前为计算机视觉和图像分类）

我们还可以在推断时（或测试时，因此命名）使用数据扩充。在推断阶段，你所做的只是做预测。您可以对测试集中的各个图像执行此操作，但如果随机生成对所访问的测试集中的每个图像的一些增强，则此过程将变得更加健壮。在fastai中，每个测试图像使用4个随机增强进行预测，这些预测的平均值用作该图像的预测。

#### 用预训练递归神经网络代替词向量

一种不用词向量就可以得到世界级的情感分析框架的方法是，获取你想要分析的整个训练数据集，并从中建立一个深度递归神经网络语言模型。当模型具有高精度时保存模型的编码器，并使用从编码器获得的嵌入来构建情感分析模型。这比从字向量得到的嵌入矩阵更有效，因为RNNs比字向量更好地跟踪远程依赖关系。

#### 时间反向传播（BPTT）（NLP）

深度递归神经网络中的隐藏状态在经过一定时间步的反向传播后，如果不重新设置，则会增长到一个难以处理的大小。例如，在字符级RNN中，如果有一百万个字符，那么还有一百万个隐藏状态向量，每个状态向量都有其历史记录为了调整神经网络的梯度，我们需要对每个字符、每个批执行100万个链规则计算。这会消耗太多的内存。因此，为了降低内存需求，我们设置了最大数量的字符以反向传播。由于递归神经网络中的每个回路都称为时间步，因此限制反向传播保留隐藏状态历史的层数的任务称为时间反向传播。这个数字的值决定了模型计算的时间和内存需求，但是它提高了模型处理长句或动作序列的能力。

#### 类别变量的实体嵌入。（结构化数据和NLP）

在对结构化数据集进行深入学习时，它有助于区分包含连续数据（如在线商店中的价格信息）的列和包含分类数据（如日期和取货地点）的列。然后，可以将这些分类列的一个热编码过程转换为指向神经网络的完全连接嵌入层的查找表。因此，您的神经网络有机会了解那些如果忽略列的分类性质就会被忽略的分类变量/列。它可以学习周期性事件，例如一周中哪几天的销售额最高，公共假日之前和之后发生了什么，对于多年数据集。研究结果为预测产品最优定价和协同过滤提供了一种非常有效的方法。这应该是所有拥有表格数据的公司的标准数据分析和预测方法。这就是全部。所有公司都应该用这个。

这一方法在郭和伯克汉的Kaggle竞赛中得到了应用，即使他们只使用了最小特征工程的深度学习，也获得了第三名。他们在中概述了他们的方法。

#### 鳍。

随着图书馆的发展，人工智能的深度学习子领域越来越容易进入。感觉像是研究者和实践者已经爬山了，他们通过艰苦的工作，编译了大量的数据集和有能力的gpu，取得了越来越大的进步，终于在公开场合发布了一套有望颠覆人类历史进程的工具。在我看来，最大的潜力在于教育和医学，特别是复兴生物技术。甚至在我们采取正确的政策、洞察力、动力和全球协调之前，我们将变得更聪明、更富有，而且应该期望到本世纪末，由于这些工具，我们能活得更长、更健康。