

机器学习纳米学位

林行众

2018年7月5日

I. 问题的定义

项目概述

人工智能技术的快速发展，以人类的经济、社会与文化产生了重要影响。特别是2017年 Google AlphaGO 与人类的围棋大战^[1]，最终 AlphaGO 以 3:1 的比分战胜人类。这场人机大战，激发了大众对人工智能的关注，也让人工智能概念飞入寻常百姓家。而深度学习作为目前人工智能算法中的一个流行的技术，其技术特点在于模拟人脑的神经网络结构进行分析学习。而卷积神经网络作为深度学习结构的一种，以其他深度学习结构相比，其在图像和语音识别方面能够给出很好的结果。

2012年，Krizhevsky等^[2]提出的AlexNet在大型图像数据库ImageNet的图像分类竞赛中以准确度超越第二名11%的巨大优势夺得了冠军，使得卷积神经网络成为了学术界的焦点。卷积神经网络在图像分类上的巨大优势，使得其在相关领域，比如医学影像诊断上取得了许多令人瞩目的成绩^[3]。本文介绍的项目也以卷积神经网络技术，进行图像分类的学习任务的尝试。

本次项目采用的图像数据来源于Kaggle Dogs vs. Dogs^[3]。项目提供了相应的训练集和测试集数据，其中训练集数据中包含猫和狗图像，属于一个二分类的问题。

问题陈述

Dogs vs. Cats 是一个典型的二元图片分类问题，按照Kaggle上提供了训练集和测试集数据，可以看到图片尺寸大小各异，猫狗品种不一，导致图片上动物的外形、毛色、体型大小差异很大，对后面模型的搭建增加了难度。在本项目中，采用卷积神经网络（CNN）来解释猫狗图片分类问题。

评价指标

项目模型的评价指标，可以参考Kaggle上结果评价的指标：log loss^[4]

$$LogLoss = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

其中

- n 为测试集中图片数目
- \hat{y}_i 为图片预测为狗的概率
- y_i ，若图片为狗则为1，为猫为0

- \log 以常数 e 为底的对数

\log loss 值越小，模型效果越好。

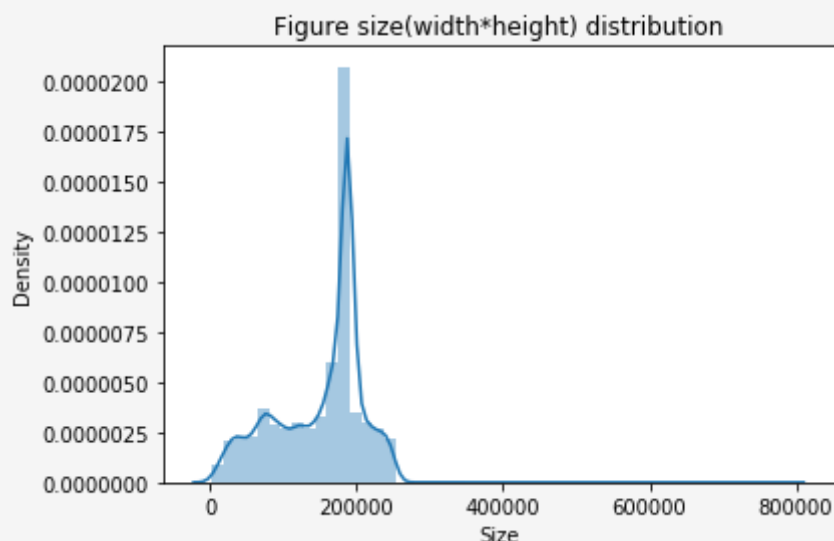
II. 分析

数据的探索

Kaggle 上提供的数据主要有训练集数据和测试集数据。其中训练集数据一共有25000张图片，猫狗图片各为12500张，对应的分类标签保存在图片的文件名中。测试集中的图片有12500张，以数字进行命名。所有图片大小不统一，可能在后续的分析中需要进行处理。

探索性可视化

利用 python 的PIL模块，简单地查看了训练用的25000张图片大小分布情况



从上图可以看到，部分图片尺寸很大，长宽的积最大接近800000，该图片是一个宽1023像素，高768像素的高清图片，这类图片占比很少，对模型的影响可以忽略。大部分图片大小分布在200000附近，这类图片大小应该接近于400X500的大小。

算法和技术

卷积神经网络（Convolutional Neural Network，CNN）一类特殊的神经网络，其特殊性主要在于卷积运算操作。CNN的基本结构为数据输入-卷积操作-汇合（pooling）操作-非线性激活函数-数据输出。同一操作类型在卷积神经网络中称作层，比如卷积层，汇合层等。上述从输入到输出这一过程便是前馈计算（feed-forward）过程。通过计算预测值与真实值之间的误差，凭借反射传播算法（back-propagation algorithm）将误差由最后一层逐层向前反馈，更新每层参数，并在更新参数后再次前馈，如此往复，直到模型收敛，这样就达到了模型训练的目的。

项目中数据的处理和模型搭建使用Python中的Tensorflow模块。Python作为流程的解释型语言，被广泛地运用于机器学习当中。而Tensorflow最初由Google Brain团队开发，在Google内部使用。在2015年，Google将Tensorflow开源，目前Tensorflow开发非常活跃，被用于各种感知和语言理解任务的机器学习^[5]。

基准模型

不是很明白基准模型是需要提供些什么 在这一部分，你需要提供一个可以用于衡量解决方案性能的基准结果/阈值。这个基准模型要能够和你的解决方案的性能进行比较。你也应该讨论你为什么使用这个基准模型。一些需要考虑的问题：

- 你是否提供了作为基准的结果或数值，它们能够衡量模型的性能吗？
- 该基准是如何得到的（是靠数据还是假设）？

III. 方法

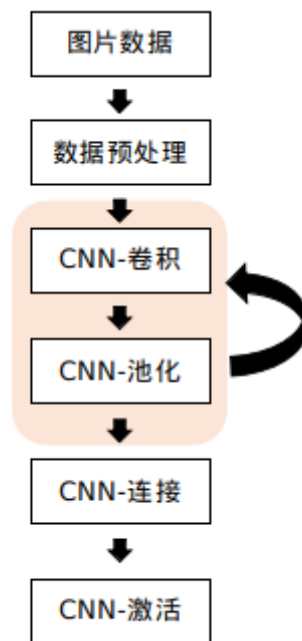
数据预处理

根据提供的数据集特点，采用了下列预处理方法：

1. 对训练集提供的图片文件名，提取对应的分类标签。图片命名格式按照“cat/dog.数字.jpg”格式，根据`cat=0`和`dog=1`提取分类标签。
2. 训练集提供的图片按照文件名顺序存放，可以在读取时将文件顺序随机打乱，以免在后续模型训练时造成影响
3. 训练集中的图片大小不一，而模型要求的训练图片需要固定的大小，所以这里可以利用`tf.image.resize_images`将图片缩放到规定大小

执行过程

卷积神经网络模型的搭建，参考了cifar10^[6]和mnist^[7]代码结构，如下图所示：



其中卷积和池化进行了2步

完善

在这一部分，你需要描述你对原有的算法和技术完善的过程。例如调整模型的参数以达到更好的结果的过程应该有所记录。你需要记录最初和最终的模型，以及过程中有代表性意义的结果。你需要考虑的问题：

- 初始结果是否清晰记录了？
- 完善的过程是否清晰记录了，其中使用了什么技术？
- 完善过程中的结果以及最终结果是否清晰记录了？

IV. 结果

(大概 2-3 页)

模型的评价与验证

在这一部分，你需要对你得出的最终模型的各种技术质量进行详尽的评价。最终模型是怎么得出来的，为什么它会被选为最佳需要清晰地描述。你也需要对模型和结果可靠性作出验证分析，譬如对输入数据或环境的一些操控是否会对结果产生影响（敏感性分析sensitivity analysis）。一些需要考虑的问题：

- 最终的模型是否合理，跟期待的结果是否一致？最后的各种参数是否合理？
- 模型是否对于这个问题是否足够稳健可靠？训练数据或输入的一些微小的改变是否会极大影响结果？（鲁棒性）
- 这个模型得出的结果是否可信？

合理性分析

在这个部分，你需要利用一些统计分析，把你的最终模型得到的结果与你的前面设定的基准模型进行对比。你也分析你的最终模型和结果是否确实解决了你在这个项目里设定的问题。你需要考虑：

- 最终结果对比你的基准模型表现得更好还是有所逊色？
- 你是否详尽地分析和讨论了最终结果？
- 最终结果是不是确实解决了问题？

V. 项目结论

(大概 1-2 页)

结果可视化

在这一部分，你需要用可视化的方式展示项目中需要强调的重要技术特性。至于什么形式，你可以自由把握，但需要表达出一个关于这个项目重要的结论和特点，并对此作出讨论。一些需要考虑的：

- 你是否对一个与问题，数据集，输入数据，或结果相关的，重要的技术特性进行了可视化？
- 可视化结果是否详尽的分析讨论了？
- 绘图的坐标轴，标题，基准面是不是清晰定义了？

对项目的思考

在这一部分，你需要从头到尾总结一下整个问题的解决方案，讨论其中你认为有趣或困难的地方。从整体来反思一下整个项目，确保自己对整个流程是明确掌握的。需要考虑：

- 你是否详尽总结了项目的整个流程？
- 项目里有哪些比较有意思的地方？
- 项目里有哪些比较困难的地方？
- 最终模型和结果是否符合你对这个问题的期望？它可以在通用的场景下解决这些类型的问题吗？

需要作出的改进

在这一部分，你需要讨论你可以怎么样去完善你执行流程中的某一方面。例如考虑一下你的操作的方法是否可以进一步推广，泛化，有没有需要作出变更的地方。你并不需要确实作出这些改进，不过你应能够讨论这些改进可能对结果的影响，并与现有结果进行比较。一些需要考虑的问题：

- 是否可以有算法和技术层面的进一步的完善？
- 是否有一些你了解到，但是你还没有能够实践的算法和技术？

- 如果将你最终模型作为新的基准，你认为还能有更好的解决方案吗？
-

**** 在提交之前，问一下自己... ****

- 你所写的项目报告结构对比于这个模板而言足够清晰了没有？
- 每一个部分（尤其**分析**和**方法**）是否清晰，简洁，明了？有没有存在歧义的术语和用语需要进一步说明的？
- 你的目标读者是不是能够明白你的分析，方法和结果？
- 报告里面是否有语法错误或拼写错误？
- 报告里提到的一些外部资料及来源是不是都正确引述或引用了？
- 代码可读性是否良好？必要的注释是否加上了？
- 代码是否可以顺利运行并重现跟报告相似的结果？

参考文献

1. Google's AlphaGo AI wins three-match series against the world's best Go player. <https://techcrunch.com/2017/05/24/alphago-beats-planets-best-human-go-player-ke-jie/> ↩
2. KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[G]//PEREIRA F, BURGESS C J C, BOTTOU L等. Advances in Neural Information Processing Systems 25. Curran Associates, Inc., 2012: 1097–1105. ↩
3. <https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition> ↩
4. <https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition#evaluation> ↩
5. Tensorflow. <https://zh.wikipedia.org/wiki/TensorFlow> ↩
6. tensorflow-cifar-10. <https://github.com/exelban/tensorflow-cifar-10> ↩
7. MNIST For ML Beginners . https://www.tensorflow.org/versions/r1.0/get_started/mnist/beginners ↩