

Keras作者、谷歌研究员Chollet：深度学习的理论局限

2017年07月19日 13:30:10 量子位

0| | | |

王小新 编译自 Keras Blog

量子位 出品 | 公众号 QbitAI

从图像处理，到自然语言处理，再到语音识别等多个领域，深度学习都取得了很好的成绩，但是仍存在一些领域，还等待深度学习去攻坚克难。

Keras作者Francois Chollet昨天在博客上发布了一篇文章，讲述了当下深度学习理论的限制性及发展方向。这篇文章是他的深度学习书《Deep Learning with Python》第9章第2节的修改版。

阅读本文需要一定的深度学习基础知识，如了解卷积神经网络、长短期记忆模型和生成对抗网络的基本原理等。

你也可以通过Deep Learning with Python的前八章，来学习一定的先验知识。

以几何视角看深度学习

现如今，使用深度学习是如此简单，这也是最令人惊讶的一点。

十年前，没人能想到用梯度下降算法训练得到的简单参数模型，能在机器感知问题上取得这么好的效果。事实证明，我们需要做的，只是构建规模足够大的训练样本，通过梯度下降算法学习得到一个参数足够多的网络模型。

物理学家费曼曾经这样描述宇宙：“它并不复杂，只是很多而已。”

这句话也适用于上述模型。

在深度学习中，一切数据都可看作为一个向量，即一切都是几何空间中的一个点。

模型输入（可以是文本、图像和信号等）和目标标签首先会被“向量化”，即转换成对应的初始输入向量空间和目标向量空间。

随着输入数据的经过，深度学习模型的每一层会进行一个简单的几何变换。这些层加起来，形成了一个非常复杂的几何变换，也可看作是一系列简单几何变换的组合。这种复杂的几何变换尝试将输入空间逐点映射到目标空间，变换的参数取决于各网络层的权重，而权重根据模型的当前训练效果进行迭代更新。

模型的可微分性是这种几何变换的一个关键特点，这也是能通过梯度下降来训练网络参数的必要条件。直观上说，从输入到输出的几何变换必须是平滑且连续的，这是一个重要的约束条件。

将这种复杂的几何变换应用到输入数据中的整个过程，能够在三维空间中可视化。



量子位
关注前沿科技资讯，追踪人工智能动态

热文排行

- 日榜周榜月榜
- 1

细思极恐！社会即将分层，你将会在第几...
- 2

买套300万的房子，我们给国家创造了多...
- 3

孙宏斌“中枪”了，身为美国人的他这么看...
- 4

光说女性靠颜值，其实男性收入受长相的..
- 5

90后创业之神被曝是恋童癖，公司说不知..
- 6

贾跃亭走了！王健林你也要走么？
- 7

私募奇葩真不少：基金经理无资格、700...
- 8

楼市进入关键时期，买房千万不要有这五..
- 9

昨天！楼市变天了，此政策一出以后将不..
- 10

潘石屹：“听明白没有，房价会降。”



想象一个试图展开纸团的人：皱巴巴的纸团是模型开始时输入数据的流形，人对纸团所做的每个动作相当于每个网络层进行的简单几何变换，所有展开操作的手势序列可以看作是模型进行的一系列复杂变换。

深度学习模型就是一种用于“展平”高维数据复杂流形的数学工具。

这就是深度学习的魅力所在：将实际值转化为向量，进而转换到几何空间，然后逐步学习一种能将某个空间映射到另一个空间的复杂几何变换。你只需要确定构建的空间具有足够高的维度，其范围能覆盖在原始数据中挖掘到的所有关联。

深度学习理论局限性

这种简单策略，有着几乎无限的应用空间。然而，即使有大量人工标注的数据集，目前的深度学习技术仍然无法实现很多应用程序所需的功能。

比如说，我们可以构建一个数据集，里面包含成千上万、甚至数百万条产品经理写的软件功能描述，以及程序员们编写的满足这些要求的相应源代码。就算有这样一个数据集，也无法训练一个深度学习模型来简单地阅读产品描述并生成合适的代码库。

这只是许许多多难题中的一个例子。

一般来说，无论构建大规模的数据集，任何像编程这样需要推理、长期规划和类似算法数据操作的任务，都无法使用深度学习模型来解决。甚至连最简单的排序算法，深度神经网络学起来也是相当困难的。

这是因为深度学习模型只是为了将一个向量空间映射到另一个向量空间，而构建了一系列简单且连续的几何变换操作。

假设从X到Y存在一种可学习的连续变换，深度学习能做的只是将一个数据空间X映射到另一个空间Y，并可以将X和Y配对数据的密集采样作为训练集。

因此，我们可以将一个深度学习模型解释为一种应用程序，反之，大多数程序不能用深度学习模型来表示。

对于大多数任务，可能不存在一定规模且能解决问题的深层神经网络。即使存在对应模型，其可能不具备学习能力，即相应的几何变换太过复杂，或者可能没有可以用来学习的合适数据集。

通过堆叠更多网络层和使用更大规模的训练集来扩展当前的深度学习技术，只能一定程度上解决其中的若干问题。但但程序不能表达为一种数据流的连续几何变换时，深度学习技术能做的十分有限。

拟人化机器学习模型的风险

当前，人工智能领域中存在着一个很严重的问题：人们误解了深度学习模型的工作机理，并高估了网络模型的能力。

人类思想的根本特征在于“心智理论”（theory of mind），即个体理解自己与他人的心理状态，包括情绪意图、期望、思考和信念等，并借此信息预测和解释他人行为的一种能力。在岩石上画出一个笑脸，意味着我们心里是开心的。

结合深度学习后，这意味着，我们在一定程度上能成功训练出一个可以根据图片内容生成文字描述的模型。因此，我们会认为这个模型“理解”了图片中和它产生出来的文字。然而，当



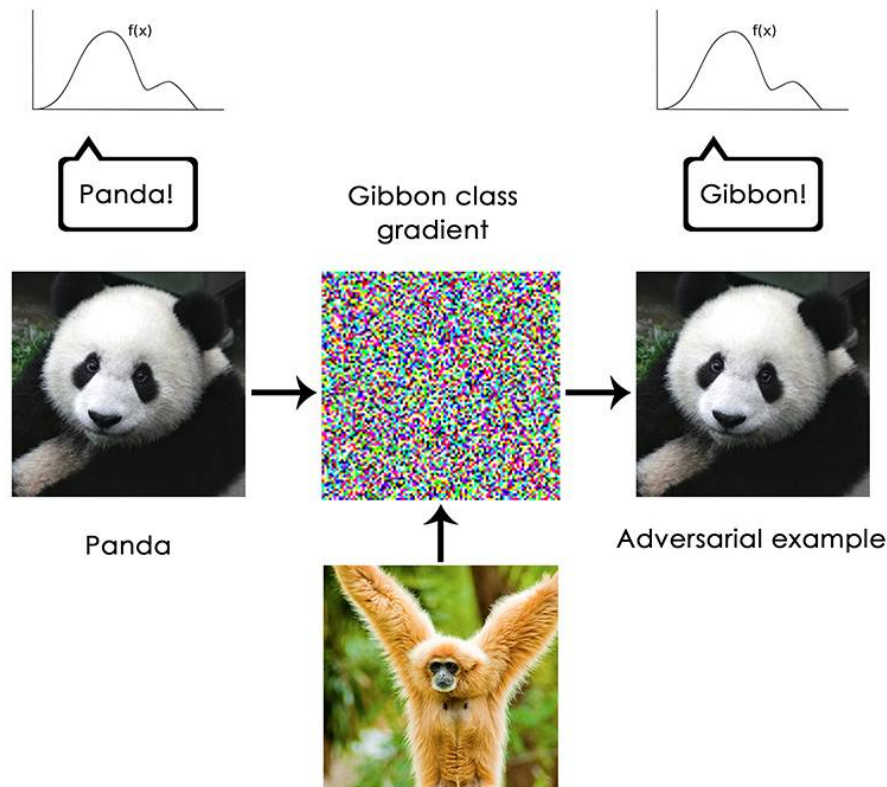
The boy is holding a baseball bat.

△ 深度学习模型把刷牙的小女孩，认成了“一个小男孩拿着棒球棒”。

“对抗样本”——也就是那些经过特别处理，诱导模型将它错认成其他类别的样本——突出了这种问题。

在深度学习中，我们可以在输入空间进行梯度上升，以生成能最大化某些卷积过滤器激活值的相应输入，这是在第5章中卷积核可视化技巧和第8章中Deep Dream算法的共同基础。类似地，根据梯度上升方法，小幅修改图像可以最大化给定类别的预测值。如果我们给定一只熊猫的图片，并设置相应的“长臂猿”梯度值，则得到的神经网络将这个熊猫图片分类到长臂猿类。

这证明了模型的不稳定性，以及网络中输入到输出的映射关系与人类感知之间的巨大差异。



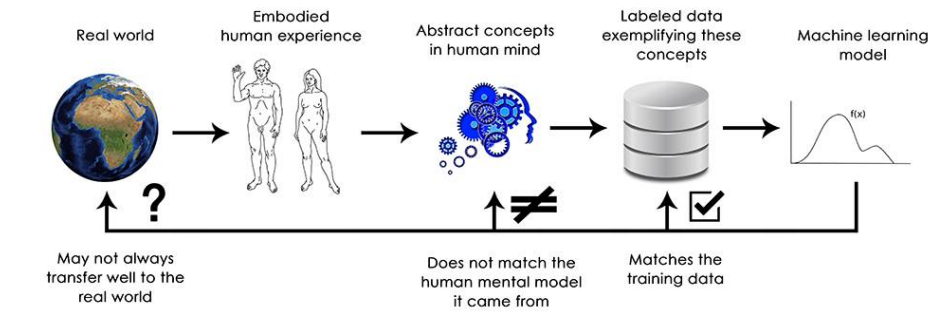
△ “对抗样本”示例：小幅修改图像可能影响模型的输出类别。

总之，深度学习模型对输入数据没有任何理解，至少不是人类那种“理解”。

人类是基于感觉-运动经验来理解图像、声音和语言的，而机器学习模型无法获得这种感知经验，因此不能通过人类的基本方法来“理解”输入数据。

我们把大量标注后的训练样本输入到网络模型中，让它们学习一种特殊的几何变换，能将数据映射到与人类意识相符合的特定类别上。

但是，这种映射只是人类思想中真实模型的简单草图，就像是一面镜子中昏暗的图像。



△ 当前的机器学习模型，就像镜子中的一个昏暗图像。

作为一名机器学习研究人员，始终要记住这一点，不要误以为神经网络理解它所执行的任务内容：不是的，如果非要说不神经网络理解了任务，那也是以一种人类无法理解的方式。

神经网络以极其局限的任务目标来进行训练，仅仅是将训练样本与训练目标进行比较。一旦向网络输入偏离训练数据的任何内容，可能会输出极其荒谬的信息。

局部泛化与极端泛化

深度学习模型中从输入到输出的简单几何变换，与人类思考和学习的方式似乎有着根本性的不同。

人类不是通过大量明确的训练样本来学习，而是通过具体的亲身体验。而除了学习过程不同之外，人类和机器学习的底层表征性质也存在着根本性的不同。

将直接刺激映射到对应的直接反应中，是深度神经网络或者昆虫都有的能力，人类的能力远远超过了这些。

人类对自身所处的环境、自己和他人都建立着多种复杂且抽象的模型，且可以使用这些模型来预测多种不同可能的后果，同时实现长期规划。

人类有能力将已有的概念融合在一起，用来表征先前从未见过的事物，如画出一幅穿着牛仔褲的马图像，或是想象如果突然赢得了彩票大奖，我们会怎么做。这种处理假想、将心理模型空间扩展到直接体验之外的能力，或者说进行抽象和推理的能力，大概就是人类认知行为中的关键特征。我将这种能力称为“极端泛化”（extreme generalization）：在相关数据非常少、甚至是没有相关数据的情况下，适应在先前情境中从未经历过的新奇事物的能力。

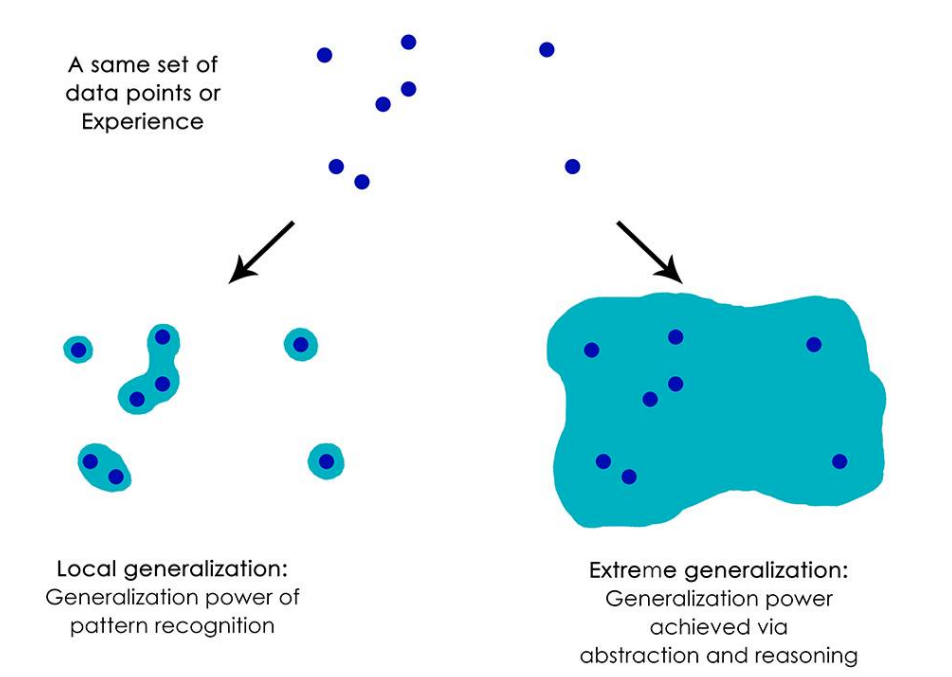
这与深度神经网络形成了鲜明的对比，我将深度神经网络的能力称为“局部泛化”（local generalization）：如果新的测试数据与训练时使用的数据集略有差异，则深度神经网络中从输入到输出的映射会很快地停止工作。

例如学习合适的发射参数来解决火箭登陆月球的问题。如果你准备使用深度神经网络来完成此项任务，在训练中不管是使用监督学习还是强化学习，都需要进行成千上万次的登陆试验来获取训练集。理论上，这解释为将训练数据转换为输入空间的密集采样，以便模型能学习从输入空间到输出空间的可靠映射。

相比之下，人类也可以利用抽象能力来提出一种针对火箭发射的物理模型，并只需一次或几次试验即可得出一个用于火箭登月的精确解决方案。

同样地，如果你开发出一种能控制人类身体的深度神经网络，并让它学会驾车安全无事故地通过一座城市，那么这个网络需要在多种情境下发生过事故，直到该网络能推断出这辆车何时发生危险并作出适当的回避行为。当行驶到一个新的城市，这个网络必须重新学习大部分先验知识。

而人类并非只在发生事故后，才能学习到安全行为知识，这要归功于人类具有能假设情境实现抽象建模的能力。



△ 局部泛化与极端泛化

简言之，尽管我们在机器感知领域取得了不错的进展，但是离拟人化人工智能还很远：当前的深度学习模型只能实现局部泛化，适应与先前数据保持非常接近的新样本，而人类认知属于极端泛化，能迅速适应大胆新奇的新环境，或对未来发展进行长期规划。

总结与讨论

到目前为止，深度学习的唯一真正的成功点在于，当给定大量人工标注后的数据集，它具有使用连续几何变换将空间X映射到空间Y的能力。

在每个行业中，要做好这些，基本上就可以改变这个行业的游戏规则。

但是从拟人化人工智能的角度来看，仍然还有很长的路要走。

为了解决深度学习的这些局限性并开始与人类大脑进行竞争，我们不能满足于直接构建从输入到输出简单映射关系的现状，要将研究重点转移到推理和抽象领域。对各种情境和概念进行抽象建模也应该属于某种计算机程序。

在“Deep Learning with Python”一书中已经提到，机器学习模型可以被定义为“可学习的程序”。

目前，技术上可训练的程序，是在所有可能程序中一个非常狭窄且特定的子集。但是，如果能以模块化和重复利用的方法来学习任何程序，会发生什么现象？

相关阅读

原文地址：

<https://blog.keras.io/the-limitations-of-deep-learning.html>

Deep Learning with Python电子书：

<https://www.manning.com/books/deep-learning-with-python>

Chollet后来又写了篇《深度学习的未来》，同样发表在Keras博客，欢迎阅读英文版，也欢迎明天再来量子位读中文版。

【完】

交流沟通

量子位读者6群开启，对人工智能感兴趣的朋友，欢迎加量子位小助手的微信qbitbot2，申请入群，一起探讨AI。

想要更深一步的交流？

量子位还有自动驾驶、NLP、CV三个专业讨论群，仅接纳相应领域的一线工程师、研究人员等。

同样需要添加qbitbot2为微信好友，提交相应说明，符合条件后将被邀请入群。（审核较严，敬请谅解）

诚挚招聘

量子位正在招募编辑/记者等岗位，工作地点在北京中关村。期待有才气、有热情的同学加入我们！相关细节，请在量子位公众号(QbitAI)对话界面，回复“招聘”两个字。


△ 扫码强行关注『量子位』

追踪人工智能领域最劲内容



作者历史文章

俄罗斯最大搜索引擎Yandex开源梯度上升机器学习库，背后雄心满满



安妮 编译自 Tech Crunch量子位出品 | 公众号 QbitAI昨天，俄罗斯搜索巨头Yandex开源了Gradient boosting机器学习库Cat[详细]

微软将在海外市场为百度Apollo平台提供云服务



唐旭 编译整理量子位出品 | 公众号 QbitAI本月5日，百度在其首届AI开发者大会上公布Apollo自动驾驶生态的盟友时，微软赫然在列；今天，双方共同公布了[详细]

2017年 07月19日 13:30

这家人工智能AR明星制作公司获得腾讯500万美元投资



李杉 编译整理量子位 出品 | 公众号 QbitAI虚拟现实（VR）和增强现实（AR）的世界可能会深刻改变我们与电脑的互动方式，但也会改变我们关注明星的方式。O[详细]

2017年 07月19日 13:30

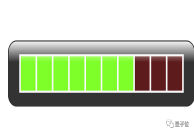
独家 | 他曾是百度少帅，乐视眼中无人车No.1，刚刚加入了A股公司



李根 发自 洛杉矶玛格丽特大道隔壁量子位报道 | 公众号 QbitAI关于乐视的消息已让人疲惫。但猎头和创投仍在保持神经敏锐，特别是涉及乐视汽车动态，可能每一[详细]

2017年 07月19日 13:30

众筹项目能否成功？用机器学习预测可以早知道



安妮 编译自 Shrikar Archak量子位出品 | 公众号 QbitAIKickstarter是一家美国的众筹平台。自2009年成立至今，已经有36万余个[详细]

2017年 07月17日 13:30

用算法做服装设计：从30亿款女式衬衫中找出最受欢迎的9种



李杉 编译自 QZ量子位 报道 | 公众号 QbitAI传统服装设计师的创意流程或许是从草图开始的，但Hybrid Designs却是从人工智能开始的。Hybr[详细]

2017年 07月17日 13:30

纽约大学《机器学习入门》课程讲义（附PDF下载）



允中 编译整理量子位 出品 | 公众号 QbitAI今年春天，Kyunghyun Cho开始在纽约大学教授本科生《机器学习入门》课程。今天，他把这门课程的讲义和[详细]

2017年 07月17日 13:30

独家 | 李开复：谈无人车安卓为时尚早，投自动驾驶有三个标准



唐旭 发自 凹非寺量子位报道 | 公众号 QbitAI今年一月，创新工场人工智能工程院成立之时，李开复曾经公开表示：“中国不再有可以投资的无人驾驶团队了。”这句[详细]

2017年 07月17日 13:30

马斯克：20年内方向盘将消失，AI是人类存在的最大风险



问耕 允中 发自 LZYY量子位 出品 | 公众号 QbitAI这是一次在美国全国州长协会（NGA）大会上的发言。昨天，特斯拉（以及SpaceX）CEO伊隆·马[详细]

2017年 07月16日 14:00

- 1
- 2
- 3
- 4
- 5
-
-