

转自雷锋字幕组双语文章

原文: [深度学习未来发展的三种学习范式：混合学习、成分学习和简化学习](#)

英文原文: [The Future of Deep Learning Can Be Broken Down Into These 3 Learning Paradigms](#)

翻译: 雷锋字幕组(小哲)

深度学习未来发展的三种学习范式：混合学习、成分学习和简化学习

深度学习是一个很大的领域，其核心是一个神经网络的算法，神经网络的尺寸由数百万甚至数十亿个不断改变的参数决定。似乎每隔几天就有大量的新方法提出。

然而，一般来说，现在的深度学习算法可以分为三个基础的学习范式。每一种学习方法和信念都为提高当前深度学习的能力和范围提供了巨大的潜力和兴趣。

- 混合学习——现代深度学习方法如何跨越有监督和无监督学习的界限，以适应大量未使用的未标记数据？
- 成分学习——如何采用一种创造性的方法将不同的组件链接起来生成一个混合的模型，这个模型的效果比各个部分简单的加和效果要好
- 简化学习——如何在保持相同或更大的预测能力的同时，减少模型的大小和信息流，以达到性能和部署的目的？

深度学习的未来主要在于这三种学习范式，每一种都紧密链接。

混合学习

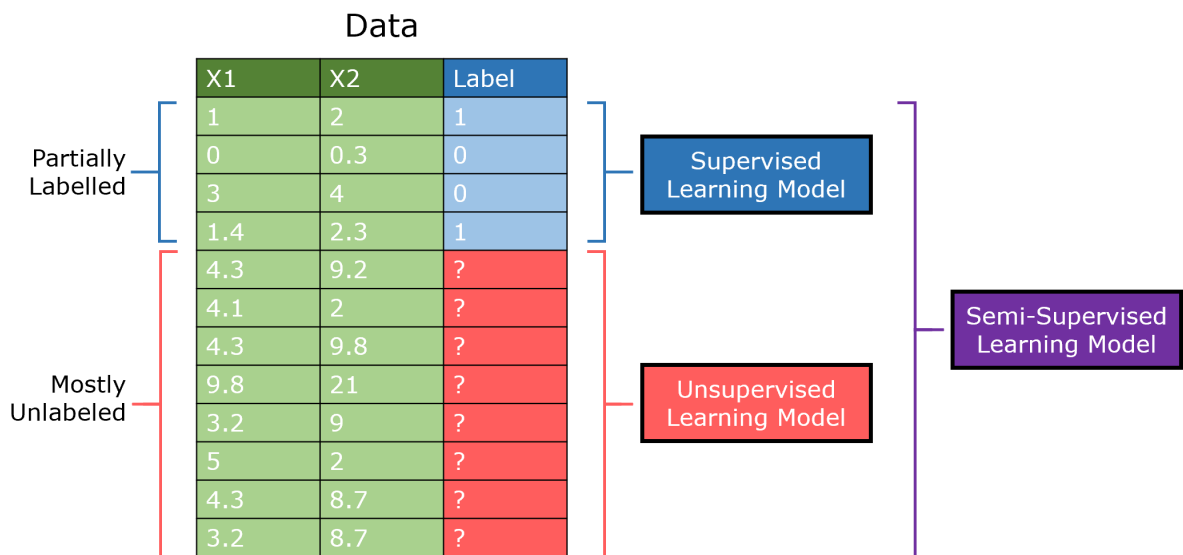
这种学习范式试图去跨越监督学习与无监督学习边界。由于标签数据的匮乏和收集有标注数据集的高昂成本，它经常被用于商业环境中。从本质上讲，混合学习是这个问题的答案。

我们如何才能使用监督学习方法来解决或者链接无监督学习问题？

例如这样一个例子，半监督学习在机器学习领域正日益流行，因为它能够在很少标注数据的情况下对有监督的问题表现得异常出色。例如，一个设计良好的半监督生成对抗网络（Generative antimarial Network）在MNIST数据集上仅使用25个训练样本，就达到了90%以上的准确率。

半监督学习专门为了那些有大量无标注样本和少量有标注样本的数据集。传统来说，监督学习是使用有标注的那一部分数据集，而无监督学习则采用另外无标注的一部分数据集，半监督学习模型可以将有标注数据和从无标注数据集中提取的信息结合起来。

学习



作者创建的图像学习

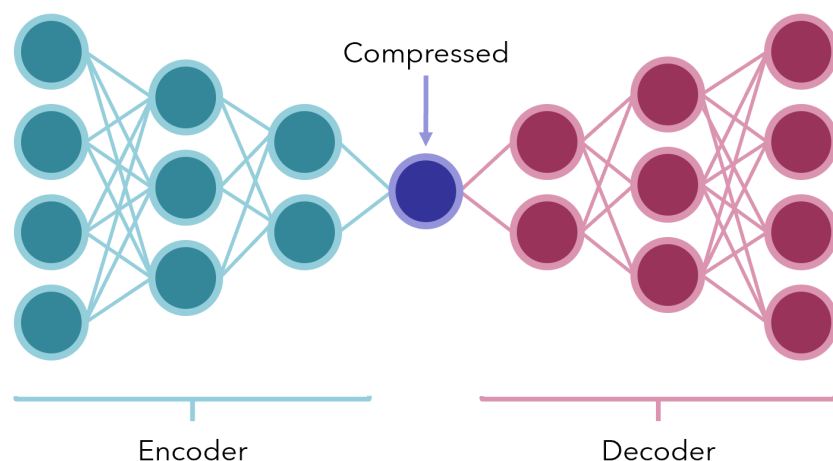
半监督生成对抗网络（简称SGAN），是[标准的生成对抗网络](#)的一种改进。判别器不仅输出0和1去判别是否为生成的图像，而且输出样本的类别（多输出学习）。

这是基于这样的想法，通过判别器学习区分真实和生成的图像，能够在没有标签的情况下学得具体的结构。通过从少量的标记数据中进行额外的增强，半监督模型可以在最少的监督数据量下获得最佳性能。

你可以[在这儿](#)阅读更多关于SGAN和半监督学习的信息。

GAN也涉及了其他的混合学习的领域——自监督学习，在自监督学习中无监督问题被明确地定义为有监督的问题。GANs通过引入生成器来人工创建监督数据；创建的标签被用来来识别真实/生成的图像。在无监督的前提下，创建了一个有监督的任务。

另外，考虑使用进行压缩的编码器-解码器模型。在它们最简单的形式中，它们是中间有少量节点的神经网络，用来表示某种bottleneck与压缩形式，两边的两个部分是编码器和解码器。



作者创建的图像

训练这个网络生成与输入向量相同的输入（一个无监督数据手工设计的有监督任务）。由于中间有一个故意设计的bottleneck，因此网络不能被动地传输信息，相反，为了解码器能够更好的解码，它一定要找到最好的方式将输入的信息保留至一个非常小的单元中。

训练之后，编码器与解码器分离，编码器用在压缩数据的接收端或编码数据用来传输，利用极少的数据格式来传输信息同时保证丢失最少的数据信息。也可以用来降低数据的维度。

另一个例子是，考虑大量的文本集合（也许是来自数字平台的评论）。通过某种聚类或[流形学习](#)方法，我们可以为文本集合生成聚类标签，然后将其作为标签处理（前提是聚类工作做得很好）。

在对每个聚类簇进行解释后（例如，聚类A代表抱怨产品的评论，聚类B代表积极反馈等），然后可以使用[BERT](#)这样的深层NLP架构将新文本分类到这些聚类簇中，所有这些都是完全未标记的数据和最少的人工参与。

这又是一个将无监督任务转换为有监督任务的有趣应用程序。在一个绝大多数数据都是无监督数据的时代，通过混合学习建立创造性的桥梁，跨越有监督和无监督学习之间的界限，具有巨大的价值和潜力。

成分学习

成分学习不仅使用一个模型的知识，而且使用多个模型的知识。人们相信，通过独特的信息组合或注入（包括静态和动态的），深度学习可以比单一的模型在理解和性能上不断深入。

迁移学习是一个非常明显的成分学习的例子，基于这样的想法，在相似问题上预训练的模型权重可以用来在一个特定的问题上进行微调。构建像Inception或者VGG-16这样的预训练模型来区分不同类别的图像。

如果我打算训练一个识别动物（例如猫和狗）的模型，我不会从头训练一个卷积神经网络，因为这样会消耗太多的时间才能够达到很好的结果。相反，我会采用一个像Inception的预训练模型，这个模型已经存储了图像识别的基本信息，然后在这个数据集（猫狗数据集）上训练额外的迭代次数即可。

类似地，在NLP神经网络中的词嵌入模型，它根据单词之间的关系将单词映射到嵌入空间中更接近其他单词的位置（例如，苹果和句子的距离比苹果和卡车的距离要小）。像GloVe这样的预训练embedding可以被放入神经网络中，从已经有效地将单词映射到数值的，有意义的实体开始。

不那么明显的是，竞争也能刺激知识增长。其一，生成性对抗性网络借用了复合学习范式从根本上使两个神经网络相互对立。生成器的目标是欺骗鉴别器，而鉴别器的目标是不被欺骗。

模型之间的竞争将被称为“对抗性学习”，不要与另一种类型的对抗性学习相混淆，那是[设计恶意输入并发现模型中的弱决策边界](#)。

对抗性学习可以刺激模型，通常是不同类型的模型，其中模型的性能可以表示为与其他模型的性能相关。在对抗性学习领域还有很多研究要做，生成性对抗性网络是对抗性学习的唯一突出创举。

另一方面，竞争学习类似于对抗性学习，但是在逐节点规模上进行的：节点竞争对输入数据子集的响应权。竞争学习是在一个“竞争层”中实现的，在竞争层中，除了一些随机分布的权值外，一组神经元都是相同的。

将每个神经元的权重向量与输入向量进行比较，并激活相似度最高的神经元也就是“赢家通吃”神经元（输出=1）。其他的被“停用”（输出=0）。这种无监督技术是自组织映射和特征发现的核心部分。

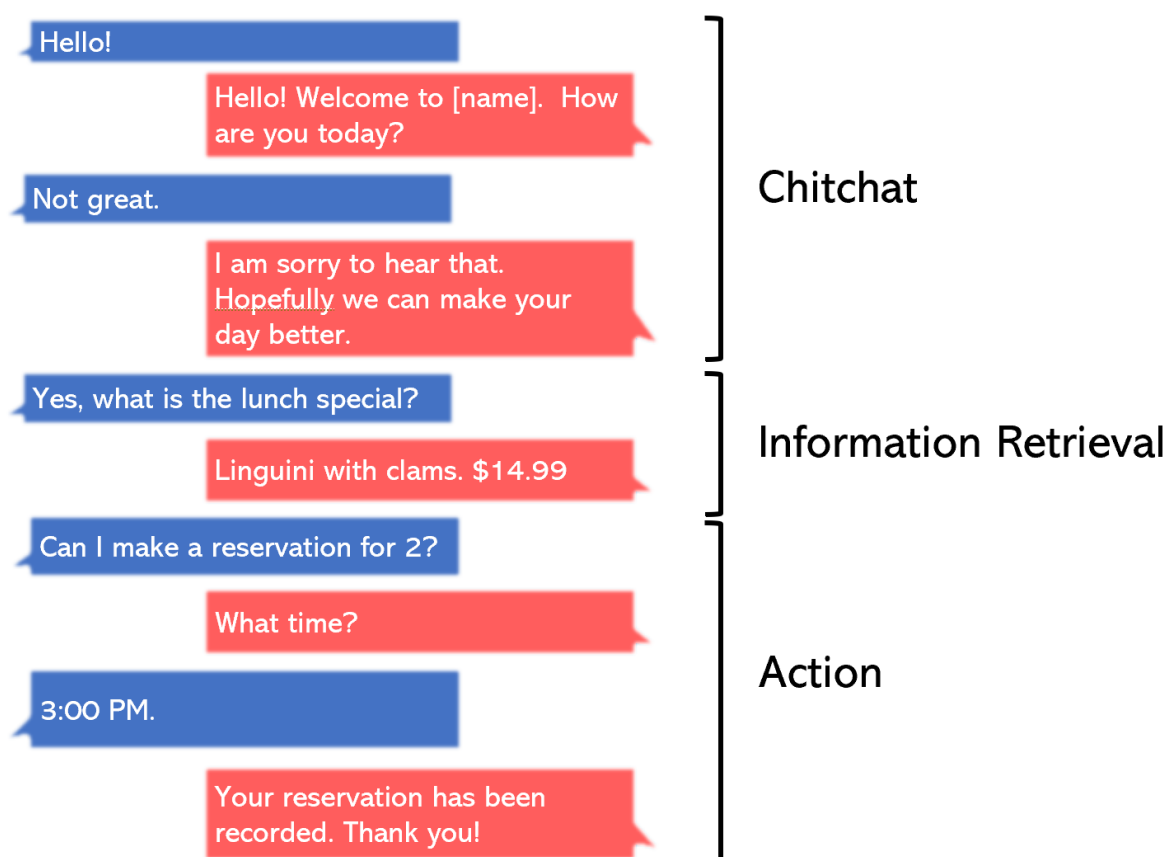
另一个成分学习的又去例子时神经架构搜索。简单来说，在强化学习环境中，一个神经网络（通常时递归神经网络）学习生成对于这个数据集来说最好的网络架构——算法为你找到最好的架构，你可以[在这](#)读到更多的关于这个理论的知识，并且[在这](#)应用python代码实现。

集成的方法在成分学习中也是主要的，深度集成的方法已经展示出了其[高效性](#)。并且模型端到端的堆叠，例如编码器与解码器已经变得非常受欢迎。

许多成分学习都在寻找在不同模型之间建立联系的独特方法。它们都基于这个想法：

单一的模型甚至一个非常大的模型，通常也比几个小模型/组件表现的差，这些小模型每一个都被分配专门处理任务中的一部分

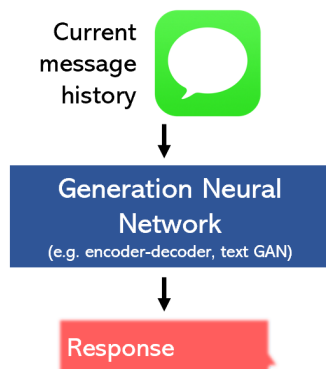
例如，考虑构建餐厅聊天机器人的任务。



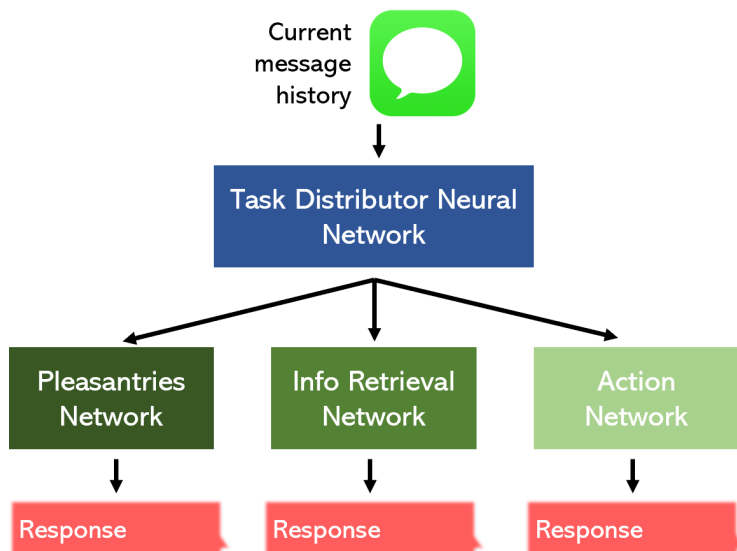
作者创建的图像

我们可以将这个机器人分割为三个分离的部分：寒暄/闲聊，信息检索和行动机器人，并为每一部分专门设计一个模型。或者，我们可以委托一个单一的模型来执行这三个任务。

Singular Model



Compositional Model



作者创建的图像

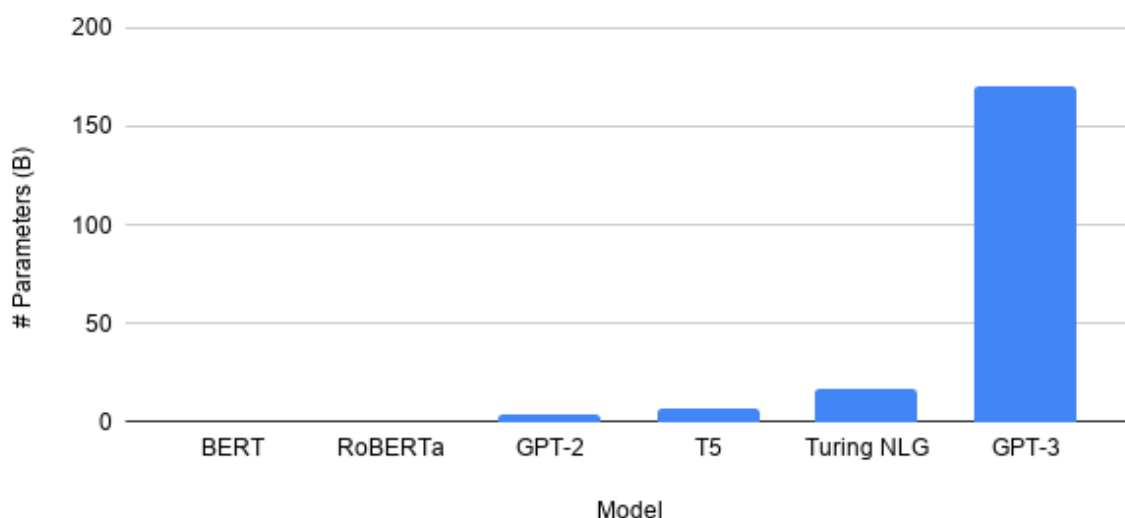
组合模型可以在占用较少空间的同时表现更好，这一点也不奇怪。此外，这些类型的非线性拓扑可以用 [Keras functional API](#) 等工具轻松构建。

为了处理像视频和三维数据等形式日益多样化的数据类型，研究人员必须构建创造性的组合模型

[在这里](#) 阅读更多关于成分学习和未来的信息。

简化学习

在深度学习领域，特别是在NLP（深度学习领域研究最热潮激动人心的领域）中，模型的规模正在不断增长。最新的GPT-3模型有1750亿个参数。把它和[BERT](#)比较就像把木星比作蚊子一样（好吧，不是字面意思）。深度学习的未来会更大吗？



资源：[IDS](#). 可免费分享的图片

按理来说，不会，GPT-3是非常有说服力的，但它在过去反复表明，“成功的科学”是对人类影响最大的科学。学术界总是离现实太远，太过模糊。在19世纪末，由于可用数据太少，神经网络被遗忘了很短一段时间，所以这个想法，无论多么巧妙，都毫无用处。

GPT-3是另一种语言模型，它可以编写令人信服的文本。它的应用在哪里？是的，例如，它可以生成查询的答案。然而，有更有效的方法来做这一点（例如，遍历一个知识图谱并使用一个更小的模型，如BERT来输出答案）。

在计算能力枯竭的情况下，GPT-3的巨大尺寸（更不用说更大的模型）是不可行的或不必要的。

“摩尔定律有点没用了。”

Satya Nadella, 微软首席执行官

取而代之的是，我们正在走向一个人工智能嵌入式世界，智能冰箱可以自动订购食品，而无人机可以自动导航整个城市。强大的机器学习方法应该能够下载到个人电脑、手机和小芯片上。

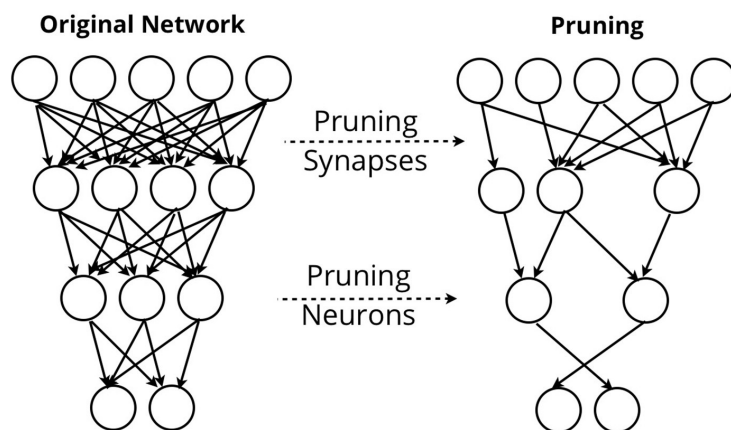
这就需要轻量级人工智能：在保持性能的同时使神经网络更小。

这直接或间接地表明，在深度学习研究中，几乎所有的事情都与减少必要的参数量有关，这与提高泛化能力和性能密切相关。例如，卷积层的引入大大减少了神经网络处理图像所需的参数数量。递归层融合了时间的思想，同时使用相同的权值，使得神经网络能够更好地处理序列，并且参数更少。

嵌入层显式地将实体映射到具有物理意义的数值，这样就不会给附加参数增加负担。在一种解释中，[Dropout](#)层显式地阻止参数对输入的某些部分进行操作。[L1/L2正则化](#)通过确保所有参数都不会增长过大来确保网络利用了所有参数，并且每个参数都能使其信息价值最大化。

随着这种特殊专用层的创建，网络对更复杂和更大的数据所需的参数越来越少。其他较新的方法显式地寻求压缩网络。

[神经网络修剪](#)试图去除那些对网络输出没有价值的突触和神经元。通过修剪，网络可以保持其性能，同时删除几乎所有的自身。



资源：[O'Reilly](#). 免费分享的图片

其他的方法像[Patient Knowledge Distillation](#)找到一些压缩语言模型的方法，使其可以下载到例如用户的手机的格式。对于谷歌神经机器翻译系统来说这是必要的考虑，这个系统支持谷歌翻译，谷歌翻译公司需要创建一个可以离线访问的高性能翻译服务。

本质上，简化学习集中在以部署为中心的设计上。这就是为什么大多数简化学习的研究来自公司的研究部门。以部署为中心的设计的一个方面不是盲目地遵循数据集的性能指标，而是在部署模型时关注潜在的问题。

例如，前面提到的[对抗输入](#)是设计用来欺骗网络的恶意输入。在标牌上喷漆或贴上标签，会诱使自动驾驶汽车加速超过限速。负责任的简化学习的一部分不仅使模型足够轻量级以供使用，而且确保它能够适应数据集中没有出现过的角落情况。

在深度学习的研究中，简化学习可能是最不受关注的，因为“我们通过一个可行的架构尺寸实现了良好的性能”并不像“我们通过由数千千万万个参数组成的体系结构实现了最先进的性能”一样吸引人。

不可避免地，当追求更高比例的宣传消失时，正如创新的历史所示，简化学习——实际上是真正的实践性学习——将得到更多应有的关注。

总结

- 混合学习目标是去跨越监督学习与无监督学习的边界，类似半监督与自监督的方法能够去从无标注的数据中提取信息，当无监督的数据量呈指数级增长时，这是非常有价值的东西。
- 随着任务变得越来越复杂，成分学习将一个任务解构为数个简单的组件。当这些组件一起联合工作或者对抗工作时，结果会是一个更加优良的模型。
- 简化学习没有受到过多的关注因为深度学习正在经历一个大肆宣传阶段，单是很快足够的实践和以部署为中心的设计将会出现。

感谢阅读!