机器学习与深度学习面试系列二十（迁移学习）

迁移学习(Transfer Learning)的动机是什么?

人类大脑具有跨任务传递知识的内在能力。我们在学习一项任务时获得的知识，可能会对我们解决相关的任务有所帮助。任务越相关，我们就越容易转移或交叉利用我们的知识。例如：

- 知道如何骑自行车，学习如何骑摩托车
- 知道如何弹奏古典钢琴，学习如何弹奏爵士钢琴
- 了解数学和统计学，学习机器学习

在这些例子中人类不会从头开始学习所有内容，而是会利用之前学习的领域的知识并将其转移到新的领域和任务中，这就是迁移学习的灵感来源。

标准机器学习的前提假设是训练数据和测试数据的分布是相同的。如果不满足这个假设，在训练集上学习到的模型在测试集上的表现会比较差，而在很多实际场景中，经常碰到的问题是标注数据的成本十分高，无法为一个目标任务准备足够多相同分布的训练数据。因此，如果有一个相关任务已经有了大量的训练数据，虽然这些训练数据的分布和目标任务不同，但是由于训练数据的规模比较大，我们假设可以从中学习某些可以泛化的知识，那么这些知识对目标任务会有一定的帮助。将相关任务的训练数据中的可泛化知识迁移到目标任务上，就是迁移学习。

迁移学习有哪几种类型?

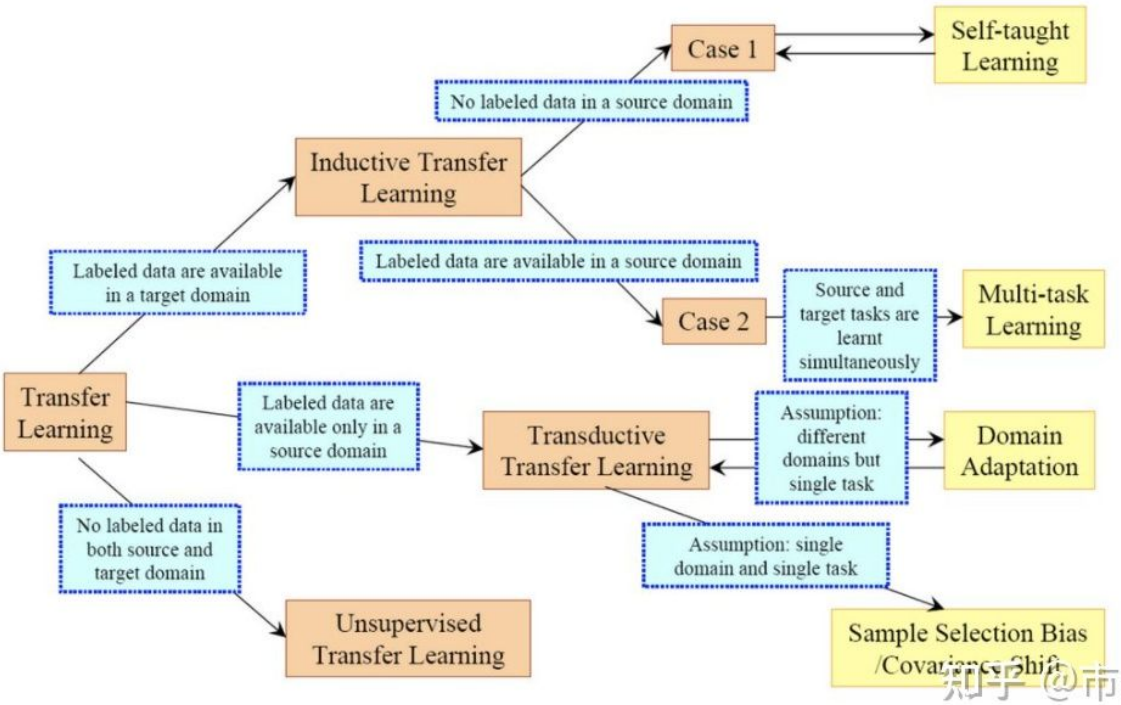
假设一个机器学习任务 $\mathcal{T}$ 的样本空间为 $\mathcal{X} * \mathcal{Y}$ ，其中 $\mathcal{X}$ 为输入空间， $\mathcal{Y}$ 为输出空间，其概率密度函数为 $p(x, y)$ 。为简单起见，这里设 $\mathcal{X}$ 为D维实数空间的一个子集， $\mathcal{Y}$ 为一个离散的集合。一个样本空间及其分布可以称为一个领域(Domain)： $\mathcal{D}=(\mathscr{X}, \mathscr{Y}, p(\boldsymbol{x}, y))$ 。给定两个领域，如果它们的输入空间、输出空间或概率分布中至少一个不同，那么这两个领域就被认为是不同的。从统计学习的观点来看，一个机器学习任务 $\mathcal{T}$ 定义为在一个领域 $\mathcal{D}$ 上的条件概率 $p(y|x)$ 的建模问题。迁移学习就是两个不同领域的知识迁移过程，利用源领域(Source Domain) $\mathcal{D}_S$ 中学到的知识来帮助目标领域(Target Domain) $\mathcal{D}_T$ 上的学习任务。源领域的训练样本数量一般远大于目标领域。

迁移学习根据不同的迁移方式又分为三个类型：

- 归纳迁移学习(Inductive Transfer Learning)
- 转导迁移学习(Transductive Transfer Learning)
- 无监督迁移学习(Unsupervised Transfer Learning)

传统机器学习和各种各种迁移学习方法之间的关系

学习方法		源和目标领域	源和目标任务
传统机器学习		相同	相同
迁移学习	归纳迁移学习/无监督迁移学习	相同	不同但有关联
		不同但有关联	不同但有关联
	转导迁移学习	不同但有关联	相同



什么是归纳迁移学习?

在归纳迁移学习中，源领域和目标领域有相同的输入空间 $\mathcal{X}_S = \mathcal{X}_T$ ，输出空间可以相同也可以不同，源任务和目标任务一般不相同 $\mathcal{T}_S \neq \mathcal{T}_T$ ，即 $p_S(y|x) \neq p_T(y|x)$ 。一般而言，归纳迁移学习要求源领域和目标领域是相关的，并且源领域 $\mathcal{D}_S$ 有大量的训练样本，这些样本可以是有标注的样本，也可以是无标注样本。根据源领域是样本是否有标注，这可以进一步分为两个子类别：分别类似于多任务学习(Multi-task Learning)和自我学习(Self-Taught Learning)。

多任务学习

当源领域有大量的标注数据时，可以直接将源领域上训练的模型迁移到目标领域上。比如在计算机视觉领域有大规模的图像分类数据集ImageNet 。由于在 ImageNet 数据集上有很多预训练的图像分类模型，比如 AlexNet、VGG 和 ResNet 等，我们可以将这些预训练模型迁移到目标任务上。

当源领域只有大量无标注数据时，源任务可以转换为无监督学习任务，比如自编码任务。通过这些无监督任务学习一种可迁移的表示，然后再将这种表示迁移到目标任务上。这种学习方式和自学习(Self-Taught Learning)比较类似。比如在自然语言处理领域，由于语言相关任务的标注成本比较高，很多自然语言处理任务的标注数据都比较少，这导致了在这些自然语言处理任务上经常会受限于训练样本数量而无法充分发挥深度学习模型的能力。同时，由于我们可以低成本地获取大规模的无标注自然语言文本，因此一种自然的迁移学习方式是将大规模文本上的无监督学习(比如语言模型)中学到的知识迁移到一个新的目标任务上。从早期的预训练词向量(比如 word2vec 和 GloVe 等)到句子级表示(比如 ELMO、OpenAI GPT 以及 BERT 等)都对自然语言处理任务有很大的促进作用。

在归纳迁移学习中，由于源领域的训练数据规模非常大，这些预训练模型通常有比较好的泛化性，其学习到的表示通常也适用于目标任务。归纳迁移学习一般有下面两种迁移方式:

- 基于特征的方式：将预训练模型的输出或者是中间隐藏层的输出作为特征直接加入到目标任务的学习模型中。目标任务的学习模型可以是一般的浅层分类器(比如支持向量机等)或一个新的神经网络模型.
- 精调的方式：在目标任务上复用预训练模型的部分组件，并对其参数进行精调(fine-tuning)。

什么时候应该做精调(fine-tuning)?

对于什么时候需要进行精调，做什么程度的精调，斯坦福的课程讲义[1]中给了一个参考。以一个例子来说明，假设我们在庞大的ImageNet数据集上得到了一个ConvNet预训练模型，现在要将这个预训练模型用来做我们自己的目标任务，一些有用的经验是：

1. 新数据集很小并且与原始数据集相似。由于数据很小，如果对ConvNet预训练模型进行精调可能会导致过拟合。由于数据与原始数据相似，我们认为 ConvNet 中的较高级别的特征也与该数据集相关。因此，最好的想法可能是直接训练ConvNet顶部的线性分类器。
2. 新数据集很大并且与原始数据集相似。由于我们拥有更多数据，因此我们可以更有信心的尝试对整个网络进行精调而不会过拟合。
3. 新数据集很小但与原始数据集有很大不同。由于数据很小，最好只训练一个线性分类器。由于数据集非常不同，从网络顶部训练分类器可能不是最好的，因为它包含更多特定于数据集的特征。所以，从网络早期某个地方开始训练一个 SVM 分类器可能会更好。
4. 新数据集很大并且与原始数据集有很大不同。由于数据集非常大，我们可能期望我们有能力从头开始训练 ConvNet。在实践中，使用来自预训练模型的权重进行初始化通常仍然是有益的。

归纳迁移学习和多任务学习有什么区别?

源领域包含的是有标注的样本时，归纳迁移学习和多任务学习很相似。但有下面两点区别:

2. 归纳迁移学习是单向的知识迁移， 希望提高模型在目标任务上的性能，而多任务学习是希望提高所有任务的性能。

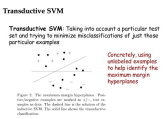
什么是转导迁移学习?

转导迁移学习是一种从样本到样本的迁移，直接利用源领域和目标领域的样本进行迁移学习。转导迁移学习可以看作一种特殊的转导学习(Transductive Learning)。转导迁移学习通常假设源领域有大量的标注数据，而目标领域没有(或只有少量)标注数据，但是有大量的无标注数据，目标领域的数据在训练阶段是可见的。

转导迁移学习我也没太理解，感兴趣的同学可以阅读这两篇博文：

https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-...
towardsdatascience.com

A small and easy introduction to Transductive Learning
codesachin.wordpress.com



什么是无监督迁移学习?

无监督迁移学习类似于归纳迁移学习，但是侧重于目标域中的无监督任务，如聚类降维等。 源领域和目标领域相似，但任务不同。 无监督迁移学习总，源领域和目标领域中均无有标记样本。

参考

1. ^ https://cs231n.github.io/transfer-learning/