我们将采用问答的方式，阐述我们对模型迁移的思考（注：由于原题分类的类别数为 9 和 6，不方便从简阐述模型迁移的思考，这里使用鸡鸭分类迁移到猫狗分类问题来阐述）

1. 什么是模型迁移

模型迁移是指将在A领域（source domain）学习到的知识迁移到B领域（target domain），以提高模型在 B 领域中的能力。而在深度学习下，知识是通过参数的方式隐式的储存的（如模型的.pth 文件中），所以通常将在 A 领域训练得到的模型参数迁移到 B 领域的模型中（如用于模型初始），便是起到知识迁移的效果。特别指出，知识迁移并不完全等于模型迁移的，但这并不妨碍我们从知识迁移的角度，参数迁移的角度理解模型迁移。

2. 为什么需要模型迁移

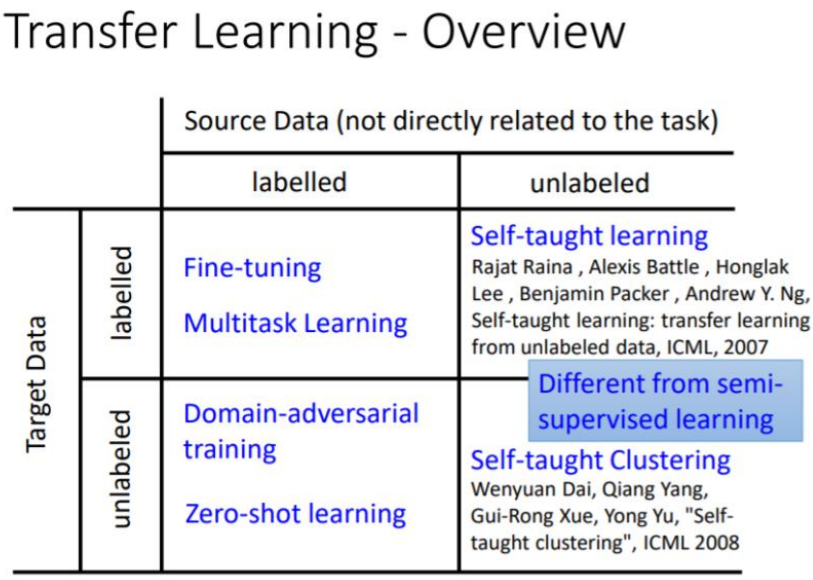
关于为什么要做模型迁移，我们觉得这和深度学习的特点有关，因为当前的深度学习为代表的模型智能是来源于数据。深度学习模型依赖于从大量的数据中感知到知识，而在实际的场景中，在我们感兴趣的领域（target domain）经常没有办法或不能够及时的拿到足够的训练数据以支撑起一个深度模型的训练，也就是所不能够让模型在 target domain 的数据中感知到足够的有效知识以完成目标任务（如图片分类），为此，从一个相关的，已经有足够多有效数据的领域（source domain）中借鉴相关知识（模型迁移，知识迁移）是弥补这一问题的重要途径。当然当感兴趣的领域（target domain）有大量的标签数据的时候可以不需要模型迁移，但从多多益善的角度，借鉴 source domain 的知识也是提升模型效果的不错选择。

3. 为什么模型迁移是有效

下一个问题是，target domain 和 source domain 是不同的领域，甚至所关注的任务都是不一样的（domain shift），比如 source domain 的数据是做鸡鸭分类，而 target domain 需要做的是猫狗分类的情况下 source domain 的知识真的可以帮助到 target domain？其实不同的领域间总是会有共用的知识，在学术上称为 domain share，如鸡鸭分类和猫狗分类虽然任务不一样，但是它们都是图片分类，在图片的 RGB 表示，线条运用等是相通的。也就是说，不同的领域中总会有共用的 domain share 知识（或高层特征，或底层特征），模型迁移有效的基础就是，source domain 和 target domain 间有 domain share 作为连接两个领域的桥梁，为 source domain 的知识在 target domain 中的运用提供了空间。用个形象的例子就是并不会因为你高中学得是理科，到大学选了文科专业，你高中的知识就白学了。

4. 有什么模型迁移的方法

在题目中给出的是两个领域的 label 数据集，同时提供了参数拷贝的迁移途径，但是实际上模型迁移的方法是多种多样的，它们的选择和当前的 source domain 和 target domain 的数据资源情况有非常直接的关系。基于过去跨领域学习（Cross Domain Learning）的工作总结，大体上的分类如下：



（图需要重新做，不要黑色的字体）

同时这里从知识迁移的角度分析一下各类方法的精神和联系：

4.1 Multitask Learning

正如 3 中谈及，source domain 和 target domain share 的特征更可能是一些基础性的特征，自然的，只有当模型在 source domain 中学习到这些 domain share 的知识的时候，迁移到 target domain 才是有效的。同时，换个角度思考，基础性的知识是任务无关的，它可以服务于各个任务，就好像学校开设的基础性课程，将帮助我们更好的完成专业课程的学习。“Multitask Learning”就是基于这样的精神，通过在 source domain 的数据上同时学习不同的任务来期待模型多“基础性“的 domain share 知识进行抽取。总体是通过不同的任务代替不同的领域，进而抽取到基础性的知识（如图片中的线，点， 块关系）

缺个图

4.2 Fine-tuning：

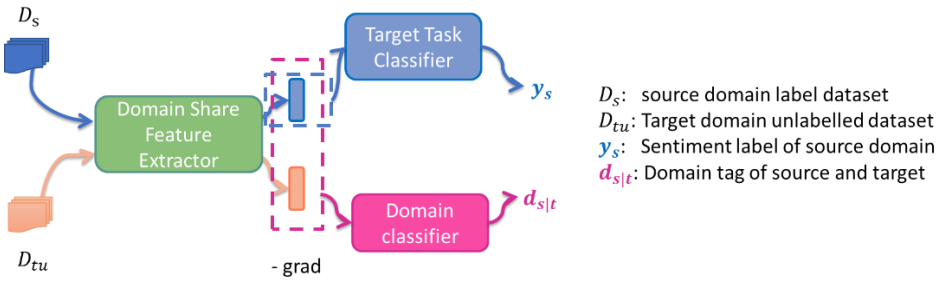
“Multitask Learning”抽取到的特征存在的一个问题是，这些特征是任务无关的，而我们关注的是 target domain 的中某个具体任务的能力。同时针对具体的任务，还需要更加

具体的专业知识的参与（如我们的专业课知识，学术上叫 domain specific feature）才能更加全面的做出正确的决策。为了让模型关注到具体的任务和学习到具体任务的“专业“知识，通常需要讲在 source domain 学习到的模型（知识），进一步在 target domain 的具体任务上进行学习（微调）。这就是 Finetune 思想的精神。

4.3 Domain adversarial Training

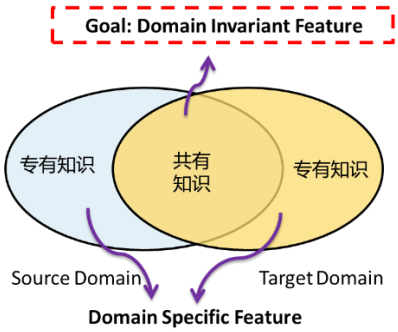
Fine-tuning 通过微调的方式引入领域特定的，具体任务的知识（特征表示）。这种方式实际上依然需要依赖于 target domain 的 label data，而正如 2 所阐述，在我们关注的 target domain 中往往没有足够的 label data 训练一个深度模型，也就是所 Fine-tuning 的做法有 overfitting 的问题在。

相对的 Domain adversarial Training 是利用 target domain 的大量 unlabelled data 进行领域共有（domain share）和任务的专有知识的学习。它的核心思想是在 source domain 的 label data 上学习和 target domain 相同或者相近的任务上学习（训练），同时为这个学习过程添加一条约束--只使用 domain share 的知识（特征）来完成指定任务。这个约束主要有对抗模块（Domain Classifier）提供，其实现框架图如下：



5. 总结和思考

从知识迁移的角度出发，模型迁移的核心动机是将在 source domain 学习到的知识通过参数的方式传递到 target domain 中（欸，知识图谱是不是另一种传递方式呢？x）。同时，关于基础性知识，domain share 知识，亦或是专业性知识，在学术上可以归类为 domain invariant feature 和 domain specific feature，它们的关系如下图

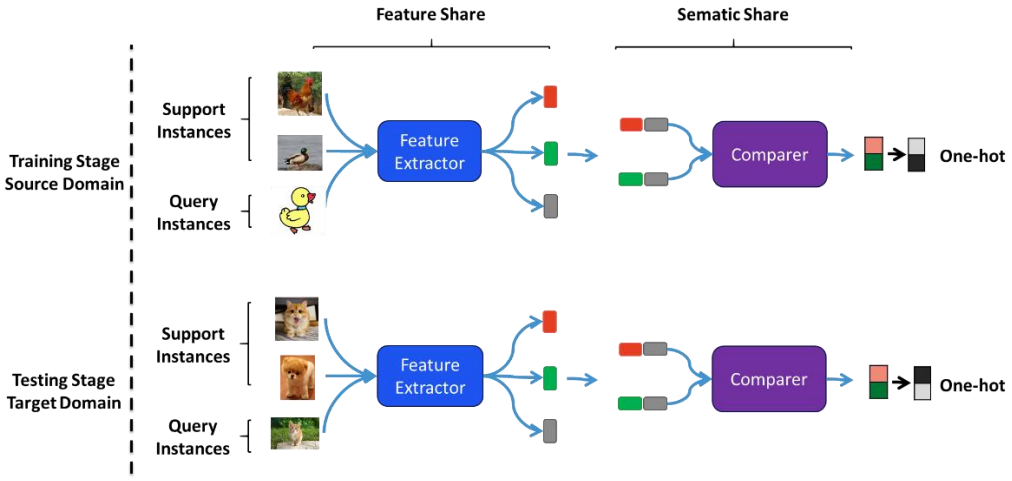


模型迁移的目标是希望模型在 source domain 中学习到 domain Invariant Feature 并在 target domain 使用它们。而对于确实的 target domain specific feature 通过在 target domain fine-tuning 来实现（容易 overfitting）。居于这样的模式，我们认为一下的挑战：

1. 依然没有解决的挑战：语义偏差问题
   通过 Multitask Learning ， Domain adversarial Training 或其他方式获取到的 domain invariant feature 知识保证了特征的 domain share，并不一定保证 Semantic 的 domain share。如从鸡鸭分类迁移到猫狗分类中，毛色是鸡，鸭，猫，狗和都具有的特征（domain share），但在鸡鸭分类中，毛色是否为黄是一个重要的分类特征（因为大体上黄色是鸡的常见色，而鸭的常见色是黑色）， 而在猫，狗中，黄色都是常见色，并不能作为猫狗分类的重要特征，这时如果在鸡鸭分类（source domain）中到了以毛色为重要特征来分类，将这个知识迁移到猫狗分类（target domain），将会成为巨大的“噪音“。

分离特征差异和语义差异：对比学习学习

针对鸡鸭分类迁移到猫狗分类上，分类器会直接判断特征对应的类别，但是在毛色这样的领域共有特征在不同领域的含义会有差异，在猫狗类被上并不能作为分类的主要特征，由此产生了语义偏差问题，影响模型性能。而换个角度，我们能否让模型学习到如何抽取图片的基础性特征（domain share feature），同时，避免直接对这些特征赋予具体的语义，而是通过对比不同一张图片和其他类别的相似性，将该图片预测为与之最相近图片的所属类别。这样做的精神是，“基础性“特征是 domain share 的但是他们的语义却不一定，而对比两张图片是不适用特征语义信息的，同时”对比“是可以 domain share 的，同时也是 domain share sematic 的知识。该模型的预计模型架构如下：



关联领域特有特征到领域不变特征

关于特征，通常认为 source domain 和 target domain 的之间有 domain invariant feature 和 domain specific feature，能够迁移的是 domain invariant feature。但是当两个领域的差异性较大（domain shift）的时候，将会因为忽略了大量的 domain specific

feature 限制了模型的能力。基于这一点，我们有一个提出通过 self-supervised 的方式，进行 domain specific feature 和 domain invariant feature 的关联。具体的预训练方式会 will be explained lately。