

中得到的预测模型（诸如汽车自动驾驶中的躲避障碍物模型）适应不可预见的未来情况。

1.2 迁移学习：定义

首先，我们遵循文献（Pan 和 Yang, 2010）中的符号定义“域”“任务”以及“迁移学习”。域 \mathbb{D} 由两部分组成：特征空间 \mathcal{X} 和边缘概率分布 $P^{\mathcal{X}}$ ，其中每一个输入样本为 $x \in \mathcal{X}$ 。一般来说，两个不同的域具有不同的特征空间或者边缘概率分布。在一个特定的域 $\mathbb{D} = \{\mathcal{X}, P^{\mathcal{X}}\}$ 中，任务 $\mathbb{T} = \{\mathcal{Y}, f(\cdot)\}$ 由标签空间 \mathcal{Y} 和函数 $f(\cdot)$ 组成。函数 $f(\cdot)$ 是预测函数，能够对未知的样本 $\{x^*\}$ 进行标签预测。从概率的角度看， $f(x)$ 等价于 $P(y|x)$ 。在分类问题中，标签可以是二值的，即 $\mathcal{Y} = \{-1, +1\}$ ，也可以是离散值，即多个类。在回归问题中，标签具有连续值。

为简单起见，我们现在只关注只有一个源域（source domain） \mathbb{D}_s 和一个目标域（target domain） \mathbb{D}_t 的情况，两个域的场景是目前为止在文献中最普遍的研究对象。具体地说，我们定义 $\mathcal{D}_s = \{(x_i, y_i)\}_{i=1}^{n_s}$ 为源域有标签数据，其中 $x_i \in \mathcal{X}_s$ 和 $y_i \in \mathcal{Y}_s$ 分别为数据样本和对应的类别标签。类似地， $\mathcal{D}_t = \{(x_i, y_i)\}_{i=1}^{n_t}$ 为目标域有标签数据，其中 $x_i \in \mathcal{X}_t$ 和 $y_i \in \mathcal{Y}_t$ 分别为目标域中的输入和输出。在多数情况下， $0 \leq n_t \ll n_s$ 。基于以上定义的符号，迁移学习定义如下（Pan 和 Yang, 2010）。

定义 1.1 (迁移学习) 给定源域 \mathbb{D}_s 和学习任务 \mathbb{T}_s 、目标域 \mathbb{D}_t 和学习任务 \mathbb{T}_t ，迁移学习的目的是获取源域 \mathbb{D}_s 和学习任务 \mathbb{T}_s 中的知识以帮助提升目标域中的预测函数 $f_t(\cdot)$ 的学习，其中 $\mathbb{D}_s \neq \mathbb{D}_t$ 或者 $\mathbb{T}_s \neq \mathbb{T}_t$ 。

整个迁移学习过程如图 1.1 所示。图中左侧的过程是传统的机器学习过程，右侧为迁移学习过程。我们可以发现，迁移学习不仅利用目标任务中的数据作为学习算法的输入，还利用源域中的所有学习过程（包括训练数据、模型和任务）作为输入。该图显示了迁移学习的一个关键概念：通过从源域获得更多知识来解决目标域中缺少训练数据的问题。

因为每个域由两部分组成，即 $\mathbb{D} = \{\mathcal{X}, P^{\mathcal{X}}\}$ ，所以条件 $\mathbb{D}_s \neq \mathbb{D}_t$ 意味着 $\mathcal{X}_s \neq \mathcal{X}_t$ 或者 $P^{\mathcal{X}_s} \neq P^{\mathcal{X}_t}$ 。类似地，任务也被定义为一对分量 $\mathbb{T} = \{\mathcal{Y}, P^{\mathcal{Y}|X}\}$ ，所以条件 $\mathbb{T}_s \neq \mathbb{T}_t$ 意味着 $\mathcal{Y}_s \neq \mathcal{Y}_t$ 或者 $P^{\mathcal{Y}|X_s} \neq P^{\mathcal{Y}|X_t}$ 。当目标域和源域相同时，即 $\mathbb{D}_s = \mathbb{D}_t$ ，并且其学习任务也相同时，即 $\mathbb{T}_s = \mathbb{T}_t$ ，学习问题就成为一个传统的机器学习问题。

基于上述定义，我们可以制定不同的方式将现有的迁移学习方法进行分类归纳。例如，基于特征空间和/或标签空间是否同构，可以将迁移学习分为两类：同构迁移学习（homogeneous