**Netwrok Struct Learning**

**一、神经结构搜索**

问题引入：当前应用的神经网络结构大多是由人类专家设计形成的，这既耗时又容易出错，因此人们正在研究如何根据需求自动地对网络结构进行搜索，并给出最合理的网络结构。

该问题包含以下三个维度：

**（一）搜索空间**

搜索空间从原则上定义了可以被表示的网络构造。

结合先验知识可以确定适合特定任务的网络结构的某些特性，这将大大简化搜索过程，当然这可能也会引入人的误差，导致最终搜索的网络结构无法超出人的认知，从而无法做到网络结构的创新。

可以分为global search spaces和cell-based search space两类。

**（二）搜索策略**

搜索策略决定了如何从搜索空间中寻找合适的结果。

一方面，我们希望能快速找到表现良好的网络结构；另一方面，我们希望算法不要过早地陷入局部最优。

许多不同的搜索策略可以被用于搜索神经结构的空间，包括随机搜索、贝叶斯优化、进化方法、强化学习（RL）和基于梯度的方法。

**（3）性能评估策略**：

NAS的目标是找到在未见数据上性能表现最佳的结构。

最简单的性能评估方法是应用标准的训练-验证过程，但通常这会耗费大量时间，近期许多研究集中于减少性能评估的成本。

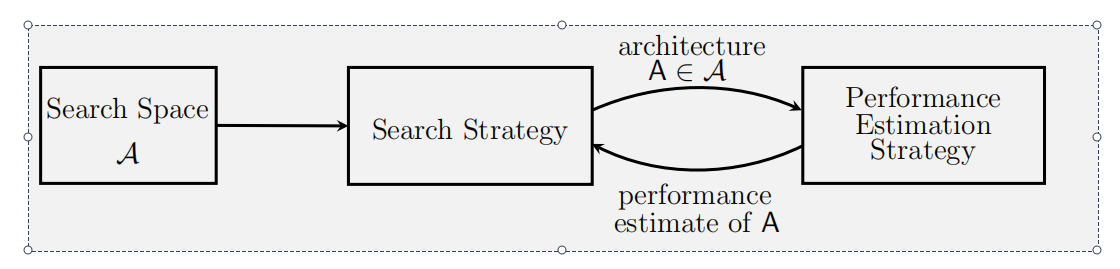


图1 NAS的简单图示

**二、贝叶斯网络结构学习**

贝叶斯网络（BN）是珀尔提出的一种概率图形模型，用于支持不确定性下的推理。概率论提供了一套阐明和操纵人类推理中的关键关系的方法概念：可能性、条件性、相关性和因果性。贝叶斯网络将概率论与图论和因果关系相结合，提供了一种机制，用以回答诸如“这种干预的可能影响是什么？”或“哪些因素与这种影响相关？”之类的问题。

贝叶斯网络B由包含有向无环图G和一组参数θ的元组定义，表征着参数间关系的强度和形状：

有向无环图G包含一组节点（nodes）X和一组有向边（arcs）E：

一条有向边，如，代表了A和B之间的直接条件关系，在因果假设下，意味着A是B的直接原因。

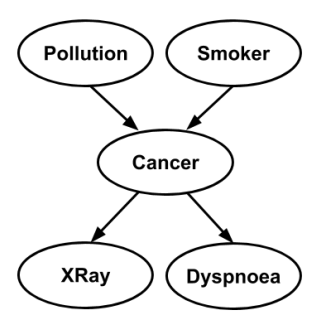


图2 简单的DAG模型，表示癌症的两种原因和癌症的两种影响

**三、知识蒸馏**

问题引入：由于移动设备的计算能力和内存有限，发展移动、嵌入式设备的深度模型具有挑战性，为了开发高效的深度模型，近期的研究工作集中在1）深度模型的高效模块构建，包括深度可分离卷积，如MobileNets， 2）模型压缩。

将较大的深度神经网络模型的知识提取到较小的网络中去，这就是知识蒸馏，它主要包含三个关键部分：知识、蒸馏算法和师生体系结构。

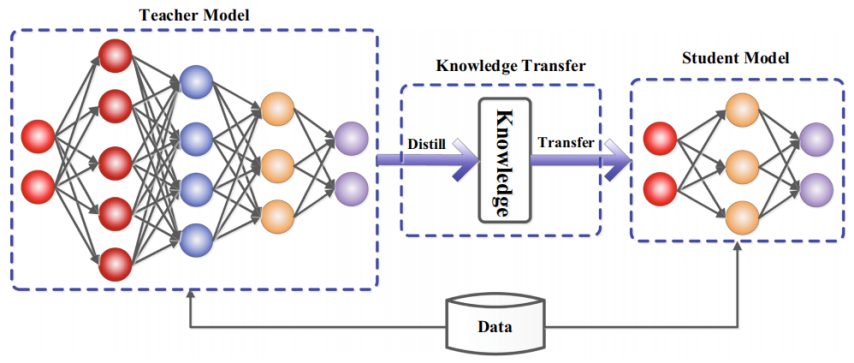


图3 知识蒸馏的师生体系结构的通用表示

**（一）知识**

通常的知识蒸馏使用深度学习模型的Logits作为教师网络的知识，可以用sigmoid的反函数logistic计算该值。

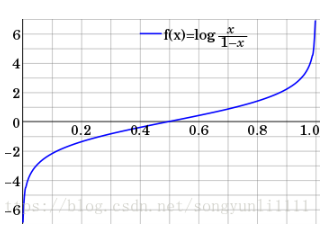


图4 sigmoid函数的反函数图像

也可以使用中间层的激活、神经元、特征作为知识。

知识可以分为基于响应的、基于特征的和基于关系的三类。

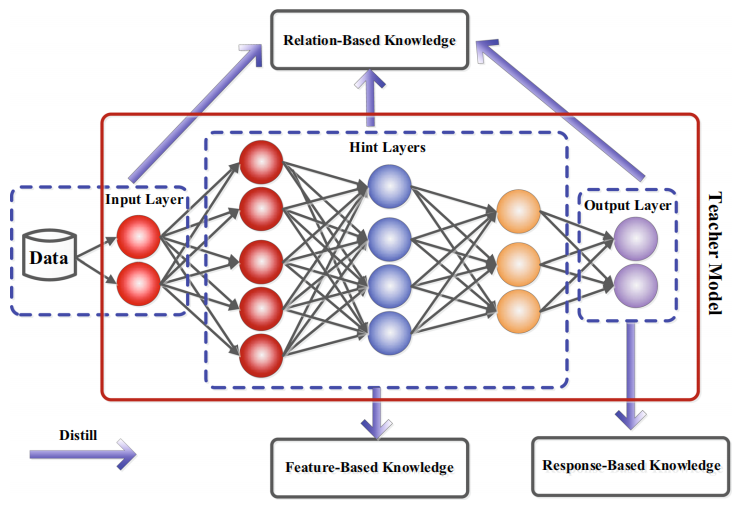


图5 知识的三种种类

**（二）蒸馏算法**

根据教师模型与学生模型是否同时更新，知识蒸馏的学习方案可分为三类：离线蒸馏、在线蒸馏和自蒸馏：

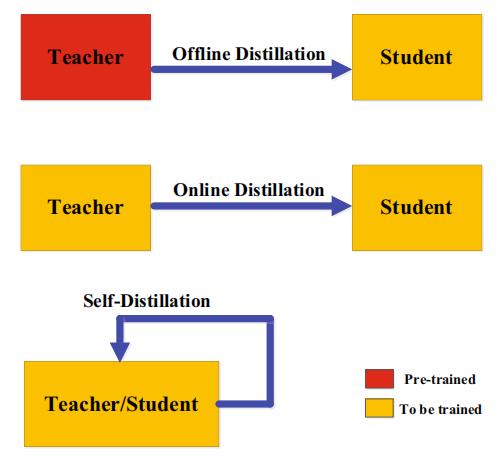


图6 蒸馏算法的三种类型

离线蒸馏：整个训练过程分为两个阶段：1）首先在一组训练样本上训练教师模型，然后进行蒸馏；2） 教师模型被用于以Logits或中间特征的形式提取知识，然后用于在蒸馏过程中指导学生模型的训练。

在线蒸馏：教师模型和学生模型同时更新，整个知识蒸馏框架被训练为端到端。

自蒸馏：教师模型和学生模型使用相同的网络，可以视为在线蒸馏的一种特殊情况。

**（三）师生体系架构**

知识蒸馏最终的结果性能取决于如何设计教师和学生网络。

深度神经网络的复杂性主要来自两个维度：深度和宽度，我们需要通过知识来将更深更宽的神经网络转化为更浅更窄的神经网络，其中学生网络通常被选择为：1）教师网络的简化版本，具有更少的层数和通道数；2） 保留网络结构的教师网络的定量版本；3） 具有高效基本操作的小型网络；4） 具有优化后的全局网络结构的小型网络；5） 与教师网络相同的网络。

需要注意，学生网络和教师网络过大的模型容量差距将不利于知识的蒸馏。

**四、基于优化目标的网络压缩**

不同于网络训练后的模型简化，基于优化目标的网络压缩将网络简化纳入原始的学习目标（如交叉熵），通过增加正则化以简化网络。

经过学习后，我们可以获取一个稀疏的网络结构。