

该方法是对样本间距离分布进行建模,使得属于同类样本靠近,异类样本远离。简单地,我们可以采用无参估计的方法,如 KNN。KNN 虽然不需要训练,但效果依赖距离度量的选取,一般采用的是一个比较随意的距离计算 (L2)。另一种,也是目前比较好的方法,即通过学习一个端到端的最近邻分类器,它同时受益于带参数和无参数的优点,使得不但能快速的学习到新的样本,而且能对已知样本有很好的泛化性。下面介绍 3 个相关的方法。

孪生网络 (Siamese Neural Networks) [1]

这个方法对输入的结构进行限制并自动发现可以从新样本上泛化的特征。通过一个有监督的基于孪生网络的度量学习来训练,然后重用那个网络所提取的特征进行 one/few-shot 学习。

它是一个双路的神经网络,训练时,通过组合不同类的样本成对,同时输入网络进行训练,在最上层通过一个距离的交叉熵进行 loss 的计算,如图 4。在预测的时候,以 5way-5shot 为例,从 5 个类中随机抽取 5 个样本,把这个 mini-batch=25 的数据输入网络,最后获得 25 个值,取分数最高对应的类别作为预测结果,

meta learning 是机器学习的一个子领域,它自动学习一些应用于机器学习实验的元数据,主要目的是使用这些元数据来自动学习如何在解决不同类型的学习问题时变得灵活,从而提高现有的学习算法。灵活性是非常重要的,因为每个学习算法都是基于一组有关数据的假设,即它是归纳偏(bias)的。这意味着如果 bias 与学习问题中的数据相匹配,那么学习就会很好。学习算法在一个学习问题上表现得非常好,但在下一个学习问题上表现得非常糟糕。这对机器学习或数据挖掘技术的使用造成了很大的限制,因为学习问题与不同学习算法的有效性之间的关系尚不清楚。

通过使用不同类型的元数据,如学习问题的属性,算法属性(如性能测量)或从之前数据推导出的模式,可

以选择、更改或组合不同的学习算法,以有效地解决给定的学习问题。

元学习一般有两级,第一级是快速地获得每个任务中的知识,第二级是较慢地提取所有任务中学到的信息。下面从不同角度解释了元学习的方法

优化器学习 (meta-learning LSTM) [6]

文章研究了在少量数据下,基于梯度的优化算法失败的原因,即无法直接用于 meta learning。首先,这些梯度优化算法包括 momentum, adagrad, adadelta, ADAM 等,无法在几步内完成优化,特别是在非凸的问题上,多种超参的选取无法保证收敛的速度。其次,不同任务分别随机初始化会影响任务收敛到好的解上。虽然 finetune 这种迁移学习能缓解这个问题,但当新数据相对原始数据偏差比较大时,迁移学习的性能会大大下降。我们需要一个系统的学习通用初始化,使得训练从一个好的点开始,它和迁移学习不同的是,它能保证该初始化能让 finetune 从一个好的点开始。

文章学习的是一个模新参数的更新函数或更新规则。它不是在多轮的 episodes 学习一个单模型,而是在每个 episode 学习特定的模型。具体地,学习基于梯度下降的参数更新算法,采用 LSTM 表达 meta learner,用其状态表达目标分类器的参数的更新,最终学会如何在新的分类任务上,对分类器网络(learner)进行初始化和参数更新。这个优化算法同时考虑一个任务的短时知识和跨多个任务的长时知识。文章设定目标为通过少量的迭代步骤捕获优化算法的泛化能力,由此 meta learner 可以训练让 learner 在每个任务上收敛到一个好的解。另外,通过捕获所有任务之前共享的基础知识,进而更好地初始化 learner。