

分布式目标识别

问题

假设存在 m 个目标，有 ℓ 种工作模式，且数据分别由 t 个设备采集。对于每条采集到的数据 x_i 可以统一表示为一条长度为 L 的时序序列，且其具备的特性可以表示为

$$(x_i, u_i, v_i, w_i) \in \mathbb{R}^{L+3}$$

其中 $u_i \in \{1, 2, \dots, m\}$ 表示其所属目标的标签， $v_i \in \{1, 2, \dots, \ell\}$ 表示其所属工作模式的标签，而 $w_i \in \{1, 2, \dots, t\}$ 表示其对应采集设备的标签。整个数据集具有 n 条数据，所以可以表示为

$$\{(x_i, u_i, v_i, w_i)\}_{i=1}^n$$

分布式场景

假设存在 K 个观测站，通常情况下 $K < m$ ，观测站 k 所能观测到的目标集合为 \mathcal{N}_k ，假设目标集合满足

$$\cup_{k=1}^K \mathcal{N}_k = \{1, 2, \dots, m\}$$

理解为每个观测站未必可以观测到所有目标，但是 K 个观测站综合在一起可以覆盖所有 m 个目标。最理想的情况为每个观测站都能观测到 m 个目标，而最坏情况为 \mathcal{N}_k 之间无交集，即恰好分配到 K 个观测站。

分布式建模及算法

分布式目标识别可以从两个角度来理解：

1. 中心化训练，分布式执行：统一的识别模型 $f(\theta)$ ，且该模型已中心化方式训练，训练完成后不属于观测站进行识别。此时需要讲观测站数据传回计算中心，并将训练好的模型回传工作站。从而导致不适应于现实场景（数据回传代价大）。此时训练模型为

$$\min_{\theta} \sum_{i=1}^N \text{loss}(f(x_i : \theta), u_i) + h(\theta)$$

2. 分布式训练，分布式执行：分布式训练旨在利用分布式计算技巧分布式训练目标识别模型，此时可以假设第 k 个观测站可以观测到的目标集合为 \mathcal{N}_k ，所观测到的数据集为

$$\{x_j^k, u_j^k, v_j^k, w_j^k\}_{j=1}^{n_i}, u_j^k \in \mathcal{N}_k$$

因此观测站 k 上所对应的目标函数为

$$\sum_{j=1}^{n_i} \text{loss}(f(x_j^k : \theta), u_j^k)$$

所以这个识别任务所对应的最优化问题可以表示为

$$\min_{\theta} \sum_{k=1}^K \left[\sum_{j=1}^{n_i} \text{loss} (f(x_j^k : \theta), u_j^k) \right] + h(\theta)$$

从而可以利用分布式随机梯度法对上述最优化问题进行求解计算。训练过程中，可以允许且只需要在观测站之间传递 θ 在本观测站的最新信息。进一步在分布式训练的过程中探索并回答下面的几个问题：

- 不同的观测站间拓扑网络对算法训练的影响
 - 最优模型的区别
 - 分布式计算训练加速的区别
 - 不同拓扑网络下观测站间信息传递频率的影响（与通信代价相关）
 - 可异步传递信息，及延迟边界的控制
 - 传递频率的探索，少通信的影响
 - 该模式与部分中心化训练模式的比较，尤其是通信代价的比较，及与计算代价的平衡
3. 部分中心化训练，分布式执行：部分中心化训练方式旨在避免传递原始数据造成过度的通信代价，因此引入数据压缩机制在每个观测站分布式训练数据压缩为维度更小的数据表征，再传递到中心计算节点进行中心化目标识别算法训练。训练完成后，分别将数据压缩表征模块和目标识别算法部署于观测站进行目标识别。值得注意的是，数据压缩表征模块的核心为将数据映射到其他低维空间的同时能够保证数据类别的可分行，因此可以在不同的观测站分别训练数据压缩表征模块，也可以使用类似分布式训练的方式进行分布式学习。**具体的训练模型后续讨论。**

多任务下的分布式建模及算法

旨在探索不同观测站承担不同目标识别任务时，利用分布式技巧提升多任务识别能力，借鉴部分中心化训练分布式执行技巧，形成更强的信号表征能力，并互相作为无监督数据提升各自观测站的识别能力。