

Algorithm: VF_TwoStep Self-Training 过程

Algorithm 1 VF_TwoStep Self-Training 过程

Require:

- $\mathcal{D}_L = \{(\mathbf{x}_i^A, \mathbf{x}_i^B, y_i)\}_{i=1}^n$ (已标注数据)
- $\mathcal{D}_U = \{(\mathbf{x}_j^A, \mathbf{x}_j^B)\}_{j=1}^m$ (未标注数据)
- $f_{\text{clf}}, f_{\text{reg}}$ (分类器与回归器)
- $k \in (0, 1]$ (每轮选取的比例)
- $\text{min_confidence} \geq 0$
- $\text{convergence_threshold} > 0$
- $\text{max_iter} \in \mathbb{N}^+$

Ensure: 未标注数据在最终模型上的预测结果

- 1: 根据 \mathbf{y}_L 判断任务类型: 分类或回归
- 2: **if** 分类 **then**
- 3: 初始化分类器 f_{clf} 并训练
- 4: **else**
- 5: 初始化回归器 f_{reg} 并训练
- 6: **end if**
- 7: $\text{epoch} \leftarrow 1$
- 8: **while** $\text{epoch} \leq \text{max_iter}$ **do**
- 9: 对 \mathcal{D}_U 执行预测并估计置信度 $\{\alpha_j\}_{j=1}^m$
- 10: 从 \mathcal{D}_U 中选取置信度最高的前 $k\%$ 或 $\alpha_j \geq \text{min_confidence}$
- 11: **if** 选中样本数 $< \text{convergence_threshold}$ **then**
- 12: **break**
- 13: **end if**
- 14: 将选中样本 \mathcal{S} 的伪标签加入 \mathcal{D}_L , 从 \mathcal{D}_U 中移除
- 15: 利用更新后的 \mathcal{D}_L 重新训练相应模型
- 16: $\text{epoch} \leftarrow \text{epoch} + 1$
- 17: **end while**
- 18: 对剩余的未标注数据执行最终预测, 得到最终预测结果
- 19: **return** $\{\hat{y}_j\}_{j=1}^m$
