LaTeX

March 6, 2025

1 Algorithm

Algorithm 1 Process of Generating Data Using Vertical Federated Semi-supervised Method

Require: X_{al}^{A}, X_{nl}^{A}, $\mathcal{L}_B$, τ,n_{al}, n_{nl}
Ensure: $X^{B_{predict}}$

1: Initialization: $X^{B_{predict}} = \emptyset$, $\mathcal{L}_B^{\text{predict}} = \{(\mu_q, \phi_q^B) \in \mathcal{L}_B \mid u_q > \tau\}$
2: **for** $(\mu_q, \phi_q^B) \in \mathcal{L}_B^{\text{predict}}$ **do**
3: $\quad X_{al}^{B_{predict}} = \{x_i^{B_{predict}}\}_{i=1}^{n_{al}}$
4: $\quad X_{nl}^{B_{predict}} = \{x_i^{B_{predict}}\}_{i=n_{al}+1}^{n_{al}+n_{nl}}$
5: $\quad p = \text{VFPU-M}(X_{al}^A, X_{nl}^A, X_{al}^{B_{predict}}, X_{nl}^{B_{predict}}, \phi_q^B)$
6: $\quad X^{B_{predict}} = X^{B_{predict}} \cup \{p\}$
7: **end for**
8: **return** $X^{B_{predict}}$

Algorithm 2 VFPU-M Algorithm

Require: X_{al}^{A}: Aligned data of party A, X_{nl}^{A}: Non-aligned data of party A, X_{al}^{B}: Aligned data of party B, X_{nl}^{B}: Non-aligned data of party B, y: Labels for the aligned data, α: Confidence threshold, k: Proportion of data selected per iteration,T: Number of iterations.

Ensure: $\mathbf{y}^{\text{pseudo}}$

Initialize the federated model f_θ
$\mathcal{D}_L^{(0)} = \{X_{al}^{A}, X_{al}^{B}, y\}$
$\mathcal{D}_U^{(0)} = \{X_{nl}^{A}, X_{nl}^{B}\}$
for $t = 1$ **to** T **do**
  Federated Model Training:
  $\theta^{(t)} \leftarrow \arg\min_\theta \sum_{(\mathbf{x}^A, \mathbf{x}^B, y) \in \mathcal{D}_L^{(t)}} \ell(f_\theta(\mathbf{x}^A, \mathbf{x}^B), y)$
  Confidence Calculation:
  For $\forall(\mathbf{x}_j^A, \mathbf{x}_j^B) \in \mathcal{D}_U^{(t)}$, calculate the confidence:
  $s_j = \begin{cases} \max_c \mathbb{P}(y = c | f_{\theta^{(t)}}(\mathbf{x}_j^A, \mathbf{x}_j^B)) & \text{Classification task} \\ 1 - \frac{|\hat{y}_j - \mu_t|}{\sigma_t} & \text{Regression task} \end{cases}$
  Where μ_t, σ_t are the current predicted mean and standard deviation
  Sample Selection:
  $\mathcal{C}^{(t)} = \{(\mathbf{x}_j^A, \mathbf{x}_j^B) \in \mathcal{D}_U^{(t)} | s_j \geq \alpha\}$
  Select the top k proportion of samples after sorting by confidence:
  $\mathcal{S}^{(t)} = \text{TopK}(\mathcal{C}^{(t)}, k)$
  Pseudo-label Generation:
  $\forall(\mathbf{x}_j^A, \mathbf{x}_j^B) \in \mathcal{S}^{(t)}, \hat{y}_j = \begin{cases} \arg\max_c P(y = c | f_{\theta^{(t)}}(\mathbf{x}_j^A, \mathbf{x}_j^B)) & \text{Classification} \\ f_{\theta^{(t)}}(\mathbf{x}_j^A, \mathbf{x}_j^B) & \text{Regression} \end{cases}$
  Dataset Update:
  $\mathcal{D}_L^{(t+1)} \leftarrow \mathcal{D}_L^{(t)} \cup \{(\mathbf{x}_j^A, \mathbf{x}_j^B, \hat{y}_j)\}_{j \in \mathcal{S}^{(t)}}$
  $\mathcal{D}_U^{(t+1)} \leftarrow \mathcal{D}_U^{(t)} \setminus \mathcal{S}^{(t)}$
end for
$\mathbf{y}^{\text{pseudo}} = \left[y, \bigcup_{t=1}^{T} \{\hat{y}_j\}_{j \in \mathcal{S}^{(t)}}\right]$ **return** $\mathbf{y}^{\text{pseudo}}$