联邦学习

2023 年 8 月 3 日

1 收敛性

定义相似度矩阵 $\mathbf{L} \in \mathbb{R}^{N\times N}$，$\mathbf{L}_{i,j}$ 表示 client i 和 client j 的相似度，k-DPP 采样：给出子集 $\mathbf{Y}$ 的采样概率, 其中 $\mathbf{Y}$ 表示 k 个 clients 的集合。

$$\mathbf{P}(\mathbf{Y}) = \frac{det(\mathbf{L}_Y)}{\sum_{|\mathbf{Y}'|=k} det(\mathbf{L}_{\mathbf{Y}'})} \tag{1}$$

目标函数

$$\mathbf{J}(\theta) = \mathbb{E}_{x\backsim p_{data}}[l(x;\theta)] \tag{2}$$

其中 p_{data} 是真实分布。

联邦学习的优化目标

$$\mathbf{F}(\theta) = \sum_{i=1}^{n} p_i f_i(\theta) = \sum_{i=1}^{n} \frac{D_i}{\sum_{i=1}^{n} D_i} \frac{1}{D_i} \sum_{x\in\mathcal{D}_i} l(x;\theta) \tag{3}$$

定义 m_i 表示在 dpp 采样时 client i 是否被选中，$m_i \in \{0,1\}$,$b_i = \mathbb{E}(m_i)$, 当使用 dpp 采样方法时，重新定义优化目标为:

$$\widehat{\mathbf{F}}(\theta) = \frac{1}{k}\sum_{i=1}^{n} b_i f_i(\theta) = \frac{1}{k}\sum_{i=1}^{n} b_i \frac{1}{D_i} \sum_{x\in\mathcal{D}_i} l(x;\theta) \tag{4}$$

更新策略:

$$\theta = \theta - \frac{1}{k}\sum_{i=1}^{n} m_i \nabla \widehat{f}_i(\theta) \tag{5}$$

其中 $\nabla \widehat{f}_i(\theta) = \sum_{t=1}^{\tau} \frac{1}{|\mathcal{B}_{i,t}|} \sum_{x\in\mathcal{B}_{i,t}} \nabla l(x;\theta_t)$

根据更新策略, 求参数更新值的期望:

$$\mathbb{E}(g) = \mathbb{E}(\frac{1}{k}\sum_{i=1}^{n} m_i \nabla \widehat{f}_i(\theta)) = \frac{1}{k}\sum_{i=1}^{n} b_i \nabla \widehat{f}_i(\theta) \tag{6}$$

这个期望和目标函数 $\widehat{F}(\theta)$ 的更新值相等，所以是策略是无偏的。

定义协方差

$$\mathbf{C}_{i,j} = \frac{\mathbb{E}[(m_i - b_i)(m_j - b_j)]}{\mathbb{E}(m_i)\mathbb{E}[m_j]} = \frac{\mathbb{E}(m_i m_j)}{b_i b_j} - 1 \tag{7}$$

计算参数更新值的方差:

$$Var(g) = \mathbb{E}[(g - \mathbb{E}(g))^2] = \mathbb{E}[(\frac{1}{k}\sum_{i=1}^{n} m_i \nabla \widehat{f}_i(\theta) - \frac{1}{k}\sum_{i=1}^{n} b_i \nabla \widehat{f}_i(\theta))^2] = \frac{1}{k^2}\mathbb{E}[(\sum_{i=1}^{n}(m_i - b_i)\nabla \widehat{f}_i(\theta))^2] \tag{8}$$

$$= \frac{1}{k^2}\mathbb{E}[(\sum_{i,j=1}^{n}(m_i - b_i)(m_j - b_j)\nabla \widehat{f}_i^T(\theta)\nabla \widehat{f}_j(\theta))] = \frac{1}{k^2}(\sum_{i,j=1}^{n}\mathbb{E}[(m_i - b_i)(m_j - b_j)]\nabla \widehat{f}_i^T(\theta)\nabla \widehat{f}_j(\theta)) \tag{9}$$

$$\begin{aligned} \mathbb{E}[m_i m_j] &= \mathbb{E}[m_i^2]\sigma_{ij} + \mathbb{E}[m_i m_j](1 - \sigma_{i,}) \\ &= \mathbb{E}[m_i]\sigma_{ij} + (\mathbf{C}_{ij} + 1)b_i b_j(1 - \sigma_{ij}) \end{aligned} \tag{10}$$

其中 $\sigma_{ij} = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases}$

$$\mathbb{E}[(m_i - b_i)(m_j - b_j)] = \mathbb{E}[m_j m_i] - b_i b_j \tag{11}$$

$$\begin{aligned} Var(g) &= \frac{1}{k^2}\sum_{i=1}^{n}(b_i - b_i^2)\|\nabla \widehat{f}_i(\theta)\|^2 \\ &+ \frac{1}{k^2}\sum_{i \neq j}\mathbf{C}_{i,j} b_i b_j \nabla \widehat{f}_i^T(\theta)\nabla \widehat{f}_j(\theta) \end{aligned} \tag{12}$$

假设 $\mathbf{C}_{ij}\nabla \widehat{f}_i^T(\theta)\nabla \widehat{f}_j(\theta) < 0$ ，因此 $Var(g)$ 减小了。