

第 5 章：联邦学习效率

思考题 5.1

定义效率指标为准确率提升与通信开销之比：

$$\text{Efficiency} = \frac{\Delta \text{Accuracy}}{\text{Communication Cost}}, \quad \Delta \text{Accuracy} = \text{当前准确率} - \text{初始准确率}$$

各节点计算如下：

$$A: \frac{0.85 - 0.75}{100} = 0.001, \quad B: \frac{0.84 - 0.70}{200} = 0.0007, \\ C: \frac{0.80 - 0.65}{150} = 0.001, \quad D: \frac{0.78 - 0.60}{300} = 0.0006$$

因此，节点 A 和节点 C 的效率最高（均为 0.001），为最优节点。

思考题 5.2

由于模型参数由 32 位浮点数量化为 8 位整数，压缩比例为

$$\frac{8}{32} = \frac{1}{4}$$

因此，量化后模型大小变为原来的 $\frac{1}{4}$ 。

(a) 量化前每个客户端上传的模型大小为

$$20 \text{ MB}$$

量化后每个客户端上传的模型大小为

$$20 \times \frac{1}{4} = 5 \text{ MB}$$

已知

$$1 \text{ MB} = 8,000,000 \text{ bit}, \quad 10 \text{ Mbps} = 10,000,000 \text{ bit/s}$$

所以量化前每轮每个客户端上传的数据量为

$$20 \times 8,000,000 = 160,000,000 \text{ bit}$$

对应通信时间为

$$\frac{160,000,000}{10,000,000} = 16 \text{ s}$$

量化后每轮每个客户端上传的数据量为

$$5 \times 8,000,000 = 40,000,000 \text{ bit}$$

对应通信时间为

$$\frac{40,000,000}{10,000,000} = 4 \text{ s}$$

因此，量化前后每轮每个客户端上传通信时间分别为

$$16 \text{ s}, \quad 4 \text{ s}$$

(b) 若总训练轮数为 100 轮，则未量化时 5 个客户端总通信时间为

$$5 \times 100 \times 16 = 8000 \text{ s}$$

量化后 5 个客户端总通信时间为

$$5 \times 100 \times 4 = 2000 \text{ s}$$

因此，采用量化技术后总共节省的通信时间为

$$8000 - 2000 = 6000 \text{ s}$$

(c) 若量化后收敛轮数增加到 120 轮，则量化情况下 5 个客户端总通信时间为

$$5 \times 120 \times 4 = 2400 \text{ s}$$

而未量化情况下总通信时间仍为

$$5 \times 100 \times 16 = 8000 \text{ s}$$

比较可得

$$2400 < 8000$$

因此，即使量化后训练轮数从 100 轮增加到 120 轮，总通信时间仍然小于未量化的情况。

进一步地，二者相差

$$8000 - 2400 = 5600 \text{ s}$$

说明量化后总体通信效率仍然更优。

思考题 5.3

(a) 在横向联邦学习中，一个全局训练周期包括两部分时间开销：一是各客户端进行本地训练的时间，二是客户端与服务器之间进行一次模型聚合的通信时间。由于服务器通常需要等待所有客户端完成本地训练后再进行聚合，因此一个全局训练周期的总时间由最慢客户端决定，可写为

$$T_{\text{total}} = \max_{i=1, \dots, N} (mT_i) + C$$

若题目中假设所有客户端的本地迭代时间相同，即

$$T_i = 2 \text{ s}, \quad m = 10, \quad C = 3 \text{ s}$$

则每个客户端完成 10 次本地迭代所需时间为

$$mT_i = 10 \times 2 = 20 \text{ s}$$

因此，一个全局训练周期的总时间成本为

$$T_{\text{total}} = 20 + 3 = 23 \text{ s}$$

当 $N = 5$ 时，由于所有客户端速度相同，结果不变，因此具体时间成本为

$$T_{\text{total}} = 23 \text{ s}$$

(b) 若不同客户端的本地迭代时间不同，即第 i 个客户端单次本地迭代耗时为 T_i ，并令其本地迭代次数为 m_i ，则一个全局训练周期的总时间应写为

$$T_{\text{total}} = \max_{i=1, \dots, N} (m_i T_i) + C$$

为了最小化全局训练周期的时间成本，需要合理选择各客户端的本地迭代次数 m_i ，使得各客户端的本地训练时间尽量均衡，即希望

$$m_1 T_1 \approx m_2 T_2 \approx \dots \approx m_N T_N$$

因此可将优化问题写为

$$\min_{m_1, \dots, m_N} \left(\max_{i=1, \dots, N} (m_i T_i) + C \right)$$

若还要求所有客户端在一个全局周期内完成大致相当的本地训练工作量，可加入总迭代预算约束，例如

$$\sum_{i=1}^N m_i = M$$

于是优化问题可进一步写为

$$\begin{aligned} \min_{m_1, \dots, m_N} \quad & \max_{i=1, \dots, N} (m_i T_i) + C \\ \text{s.t.} \quad & \sum_{i=1}^N m_i = M, \\ & m_i \in \mathbb{Z}_{>0}, \quad i = 1, \dots, N \end{aligned}$$

由该目标可以看出，较慢的客户端应分配较少的本地迭代次数，而较快的客户端可分配较多的本地迭代次数。一个直接的均衡原则是令

$$m_i \propto \frac{1}{T_i}$$

即本地迭代次数与单次迭代耗时成反比。若总预算为 M ，则可取

$$m_i = \frac{\frac{1}{T_i}}{\sum_{j=1}^N \frac{1}{T_j}} M$$

再根据实际需要取整。这样可以使各客户端的本地训练耗时尽量接近，从而减小最慢客户端带来的等待时间，降低全局训练周期的总时间成本。

因此，最优思想是：根据各客户端的计算速度自适应地分配本地迭代次数，使得各客户端的本地训练时间尽可能平衡，从而最小化

$$\max_i (m_i T_i) + C$$

这一全局训练周期时间开销。

思考题 5.4

在纵向联邦学习中，每一轮训练通常包括各参与者的本地梯度计算，以及参与者之间为完成联合训练所进行的交互。题目中给定每个参与者一次本地梯度计算时间为

$$L_i = 3 \text{ s}$$

共有

$$N = 4$$

个参与者，总迭代次数为

$$K = 50$$

(a) 不考虑安全多方计算开销时，每一轮训练的本地计算总时间为

$$\sum_{i=1}^N L_i = 4 \times 3 = 12 \text{ s}$$

因此，完整训练的总时间为

$$T_{\text{train}} = K \sum_{i=1}^N L_i$$

代入数值得

$$T_{\text{train}} = 50 \times 12 = 600 \text{ s}$$

所以，在不考虑安全多方计算开销的情况下，完成一次完整训练所需总时间为

$$600 \text{ s}$$

(b) 若进一步考虑安全多方计算开销，并用常数

$$C_{\text{SMC}} = 10 \text{ s}$$

表示每轮额外的安全多方计算通信开销，则每一轮训练总时间变为

$$\sum_{i=1}^N L_i + C_{\text{SMC}}$$

因此，完整训练的总时间更新为

$$T_{\text{train}} = K \left(\sum_{i=1}^N L_i + C_{\text{SMC}} \right)$$

代入数值得

$$T_{\text{train}} = 50 \times (12 + 10) = 50 \times 22 = 1100 \text{ s}$$

因此，在考虑安全多方计算开销后，完成一次完整训练所需总时间为

$$1100 \text{ s}$$

可见，引入安全多方计算后，总训练时间相比原来增加了

$$1100 - 600 = 500 \text{ s}$$