Adaptive Federated Learning in Resource Constrained Edge Computing Systems

主要分享一下我对文章相关数学证明的想法

辅助函数

用$[k]$代表一个训练区间$[(k-1)\tau, k\tau]$，表示第$k-1$次全局更新和第k次全局更新之间的训练区间，在这个开区间之中只有局部模型的更新

一个辅助函数如下：

$v_{[k]}(t) = v_{[k]}(t-1) - \eta \nabla F(v_{[k]}(t-1))$

这个辅助函数的参数更新代表着与局部模型的更新无关，训练的方法是将所有数据集中在同一个设备上进行梯度的更新（也就是不经过联邦学习的框架直接训练，只是一种理想的状况），由于这样子没有进行联合平均的信息损失，因此往往在一个训练周期内会有更好的训练效果（损失函数的值更小），并且这个辅助函数的值会定期更新，也就是每经过一次全局更新时会被重新设置为全局模型的参数值，如下图所示：

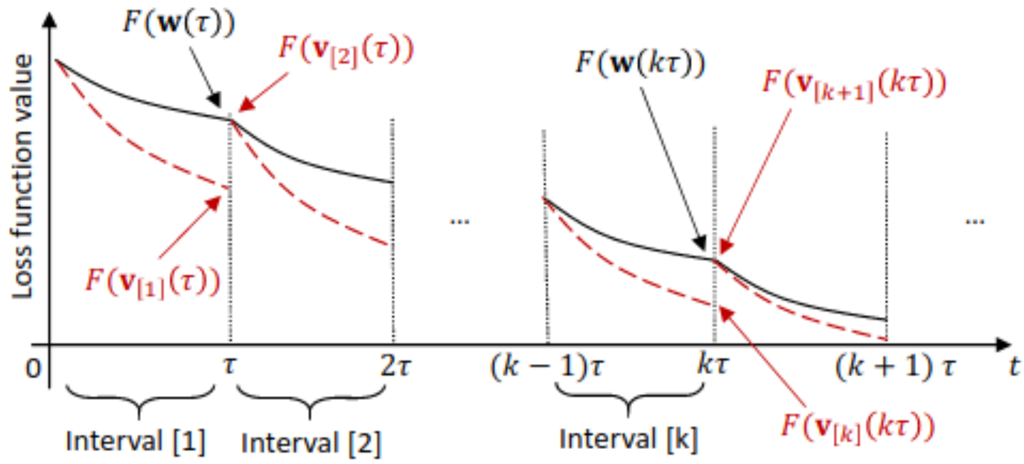

Fig. 3: Illustration of definitions in different intervals.

假设1

- $F_i(w)$ is convex
- $F_i(w)$ is $\rho - Lipshitz$
- $F_i(w)$ is $\beta - smooth$

这里可以轻易得出$F(w)$也是符合上面三个条件的

定理1

Theorem 1. *For any interval* $[k]$ *and* $t \in [k]$, *we have*

$$\left\|\mathbf{w}(t) - \mathbf{v}_{[k]}(t)\right\| \leq h(t - (k-1)\tau) \tag{10}$$

where

$$h(x) \triangleq \frac{\delta}{\beta}\left((\eta\beta + 1)^x - 1\right) - \eta\delta x \tag{11}$$

for any $x = 0, 1, 2, \ldots$.

Furthermore, as $F(\cdot)$ *is* ρ*-Lipschitz, we have* $F(\mathbf{w}(t)) - F(\mathbf{v}_{[k]}(t)) \leq \rho h(t - (k-1)\tau)$.

证明过程：

- 对于每一个设备i：

  Lemma 3. *For any interval* $[k]$, *and* $t \in [(k-1)\tau, k\tau)$, *we have*

  $$\left\|\widetilde{\mathbf{w}}_i(t) - \mathbf{v}_{[k]}(t)\right\| \leq g_i(t - (k-1)\tau)$$

  where we define the function $g_i(x)$ *as*

  $$g_i(x) \triangleq \frac{\delta_i}{\beta}\left((\eta\beta + 1)^x - 1\right)$$

- 再利用联合平均的等式将每个$g_i(x)$关联起来，证明

  $$\begin{aligned}&\left\|\mathbf{w}(t) - \mathbf{v}_{[k]}(t)\right\| - \left\|\mathbf{w}(t-1) - \mathbf{v}_{[k]}(t-1)\right\| \\ &\leq \eta\delta\left((\eta\beta + 1)^{t-1-(k-1)\tau} - 1\right)\end{aligned} \tag{24}$$

- 最后利用加和消项计算出$||w(t) - v_{[k]}(t)||$ =

$$
\begin{aligned}
&= \sum_{y=(k-1)\tau+1}^{t} \left\|\mathbf{w}(y) - \mathbf{v}_{[k]}(y)\right\| - \left\|\mathbf{w}(y-1) - \mathbf{v}_{[k]}(y-1)\right\| \\
&\le \eta\delta \sum_{y=(k-1)\tau+1}^{t} \left((\eta\beta+1)^{y-1-(k-1)\tau} - 1\right) \\
&= \eta\delta \sum_{z=1}^{t-(k-1)\tau} \left((\eta\beta+1)^{z-1} - 1\right) \\
&= \eta\delta \sum_{z=1}^{t-(k-1)\tau} (\eta\beta+1)^{z-1} - \eta\delta(t-(k-1)\tau) \\
&= \eta\delta \frac{(1-(\eta\beta+1)^{t-(k-1)\tau})}{-\eta\beta} - \eta\delta(t-(k-1)\tau) \\
&= \eta\delta \frac{(\eta\beta+1)^{t-(k-1)\tau} - 1}{\eta\beta} - \eta\delta(t-(k-1)\tau) \\
&= \frac{\delta}{\beta}\left((\eta\beta+1)^{t-(k-1)\tau} - 1\right) - \eta\delta(t-(k-1)\tau) \\
&= h(t-(k-1)\tau)
\end{aligned}
$$

问题：如何得出$g_i(x)$这个函数?

解：先用用训练时间间隔作为自变量，将$g_i(x)$代入归纳证明的框架中，一直到以下这一步：

$$
\begin{aligned}
& \qquad\qquad\qquad \text{(from the } \beta\text{-smoothness of } F_i(\cdot) \text{ and } (\boxed{9})) \\
&\le (\eta\beta+1) g_i(t-1-(k-1)\tau) + \eta\delta_i \\
& \qquad\qquad\qquad \text{(from the induction assumption in } (\boxed{22})) \\
&= (\eta\beta+1)\left(\frac{\delta_i}{\beta}\left((\eta\beta+1)^{t-1-(k-1)\tau} - 1\right)\right) + \eta\delta_i
\end{aligned}
$$

在这里将$g_i(t-1-(k-1)\tau)$凑出来，让展开之后化简的表达式可以表示为$g_i(t-(k-1)\tau)$

引理2

Lemma 2. *When all the following conditions are satisfied:*

1) $\eta \leq \frac{1}{\beta}$

2) $\eta\varphi - \frac{\rho h(\tau)}{\tau\varepsilon^2} > 0$

3) $F\left(\mathbf{v}_{[k]}(k\tau)\right) - F(\mathbf{w}^*) \geq \varepsilon$ *for all* k

4) $F\left(\mathbf{w}(T)\right) - F(\mathbf{w}^*) \geq \varepsilon$

for some $\varepsilon > 0$, *where we define* $\varphi \triangleq \omega\left(1 - \frac{\beta\eta}{2}\right)$ *and* $\omega \triangleq \min_k \frac{1}{\left\|\mathbf{v}_{[k]}((k-1)\tau) - \mathbf{w}^*\right\|^2}$, *then the convergence upper bound of Algorithm 1 after* T *iterations is given by*

$$F(\mathbf{w}(T)) - F(\mathbf{w}^*) \leq \frac{1}{T\left(\eta\varphi - \frac{\rho h(\tau)}{\tau\varepsilon^2}\right)}. \qquad (12)$$

为了证明这个引理，提出了一个辅助定义引理4-6来辅助证明

定义2

Definition 2. *For an interval* $[k]$, *we define* $\theta_{[k]}(t) = F(\mathbf{v}_{[k]}(t)) - F(\mathbf{w}^*)$, *for a fixed* k, t *is defined between* $(k-1)\tau \leq t \leq k\tau$.

)

According to the convergence lower bound of gradient descent given in [39, Theorem 3.14], we always have

$$\theta_{[k]}(t) > 0 \qquad (25)$$

for any finite t and k.

)

引理4

Lemma 4. *When* $\eta \leq \frac{1}{\beta}$, *for any* k, *and* $t \in [(k-1)\tau, k\tau]$, *we have that* $\left\|\mathbf{v}_{[k]}(t) - \mathbf{w}^*\right\|$ *does not increase with* t, *where* $\mathbf{w}^*$ *is the optimal parameter defined in (3).*

引理5

Lemma 5. *For any k, when $\eta \leq \frac{1}{\beta}$ and $t \in [(k-1)\tau, k\tau)$, we have*

$$F(\mathbf{v}_{[k]}(t+1)) - F(\mathbf{v}_{[k]}(t)) \leq -\eta\left(1-\frac{\beta\eta}{2}\right)\left\|\nabla F(\mathbf{v}_{[k]}(t))\right\|^2 \tag{26}$$

引理6

Lemma 6. *For any k, when $\eta \leq \frac{1}{\beta}$ and $t \in [(k-1)\tau, k\tau)$, we have*

$$\frac{1}{\theta_{[k]}(t+1)} - \frac{1}{\theta_{[k]}(t)} \geq \omega\eta\left(1-\frac{\beta\eta}{2}\right) \tag{27}$$

where $\omega = \min_k \frac{1}{\left\|\mathbf{v}_{[k]}((k-1)\tau) - \mathbf{w}^*\right\|^2}$

然后开始证明引理2：

同样，先用加和消项：

$$\begin{aligned}\frac{1}{\theta_{[k]}(k\tau)} - \frac{1}{\theta_{[k]}((k-1)\tau)} &= \sum_{z=(k-1)\tau}^{k\tau-1}\left(\frac{1}{\theta_{[k]}(t+1)} - \frac{1}{\theta_{[k]}(t)}\right) \\ &\geq \tau\omega\eta\left(1-\frac{\beta\eta}{2}\right)\end{aligned}$$

进一步：

$$\begin{aligned}\sum_{k=1}^{K}\left(\frac{1}{\theta_{[k]}(k\tau)} - \frac{1}{\theta_{[k]}((k-1)\tau)}\right) &\geq \sum_{k=1}^{K}\tau\omega\eta\left(1-\frac{\beta\eta}{2}\right) \\ &= K\tau\omega\eta\left(1-\frac{\beta\eta}{2}\right)\end{aligned}$$

将等式的加和式左边重写，用T替换$k\tau$：

$$\frac{1}{\theta_{[K]}(T)} - \frac{1}{\theta_{[1]}(0)} - \sum_{k=1}^{K-1}\left(\frac{1}{\theta_{[k+1]}(k\tau)} - \frac{1}{\theta_{[k]}(k\tau)}\right)$$
$$\geq T\omega\eta\left(1 - \frac{\beta\eta}{2}\right)$$

移项：

$$\frac{1}{\theta_{[K]}(T)} - \frac{1}{\theta_{[1]}(0)}$$
$$\geq T\omega\eta\left(1 - \frac{\beta\eta}{2}\right) + \sum_{k=1}^{K-1}\left(\frac{1}{\theta_{[k+1]}(k\tau)} - \frac{1}{\theta_{[k]}(k\tau)}\right) \quad (30)$$

将右边的项化简：

$$\begin{aligned}\frac{1}{\theta_{[k+1]}(k\tau)} - \frac{1}{\theta_{[k]}(k\tau)} &= \frac{\theta_{[k]}(k\tau) - \theta_{[k+1]}(k\tau)}{\theta_{[k]}(k\tau)\theta_{[k+1]}(k\tau)} \\ &= \frac{F(\mathbf{v}_{[k]}(k\tau)) - F(\mathbf{v}_{[k+1]}(k\tau))}{\theta_{[k]}(k\tau)\theta_{[k+1]}(k\tau)} \\ &\geq \frac{-\rho h(\tau)}{\theta_{[k]}(k\tau)\theta_{[k+1]}(k\tau)} \end{aligned} \quad (31)$$

根据引理5,有

$$\begin{aligned}\theta_{[k]}(k\tau)\theta_{[k+1]}(k\tau) &\geq \varepsilon^2 \\ \frac{-1}{\theta_{[k]}(k\tau)\theta_{[k+1]}(k\tau)} &\geq -\frac{1}{\varepsilon^2}\end{aligned} \quad (32)$$

然后就可以缩放了：

$$\begin{aligned}\sum_{k=1}^{K-1}\left(\frac{1}{\theta_{[k+1]}(k\tau)} - \frac{1}{\theta_{[k]}(k\tau)}\right) &\geq -\sum_{k=1}^{K-1}\frac{\rho h(\tau)}{\varepsilon^2} \\ &= -(K-1)\frac{\rho h(\tau)}{\varepsilon^2}\end{aligned}$$

根据假设$F(W(T)) - F(w^*) >= \epsilon$，有：

$$\frac{-1}{(F(\mathbf{w}(T)) - F(\mathbf{w}^*))\,\theta_{[K]}(T)} \geq -\frac{1}{\varepsilon^2} \quad (34)$$

则：

$$
\begin{aligned}
\frac{1}{F(\mathbf{w}(T))-F(\mathbf{w}^*)}-\frac{1}{\theta_{[K]}(T)} &= \frac{\theta_{[K]}(T)-(F(\mathbf{w}(T))-F(\mathbf{w}^*))}{(F(\mathbf{w}(T))-F(\mathbf{w}^*))\,\theta_{[K]}(T)} \\
&= \frac{F(\mathbf{v}_{[K]}(T))-F(\mathbf{w}(T))}{(F(\mathbf{w}(T))-F(\mathbf{w}^*))\,\theta_{[K]}(T)} \\
&\geq \frac{-\rho h(\tau)}{(F(\mathbf{w}(T))-F(\mathbf{w}^*))\,\theta_{[K]}(T)} \\
&\geq -\frac{\rho h(\tau)}{\varepsilon^2} \qquad (35)
\end{aligned}
$$

然后综合上式：

$$
\begin{aligned}
\frac{1}{F(\mathbf{w}(T))-F(\mathbf{w}^*)}-\frac{1}{\theta_{[1]}(0)} &\geq T\omega\eta\left(1-\frac{\beta\eta}{2}\right)-K\frac{\rho h(\tau)}{\varepsilon^2} \\
&= T\omega\eta\left(1-\frac{\beta\eta}{2}\right)-T\frac{\rho h(\tau)}{\tau\varepsilon^2} \\
&= T\left(\omega\eta\left(1-\frac{\beta\eta}{2}\right)-\frac{\rho h(\tau)}{\tau\varepsilon^2}\right)
\end{aligned}
$$

显然：

$$
\begin{aligned}
\frac{1}{F(\mathbf{w}(T))-F(\mathbf{w}^*)} &\geq \frac{1}{F(\mathbf{w}(T))-F(\mathbf{w}^*)}-\frac{1}{\theta_{[1]}(0)} \\
&\geq T\left(\omega\eta\left(1-\frac{\beta\eta}{2}\right)-\frac{\rho h(\tau)}{\tau\varepsilon^2}\right) > 0
\end{aligned}
$$

倒一下项：

$$
\begin{aligned}
F(\mathbf{w}(T))-F(\mathbf{w}^*) &\leq \frac{1}{T\left(\omega\eta\left(1-\frac{\beta\eta}{2}\right)-\frac{\rho h(\tau)}{\tau\varepsilon^2}\right)} \\
&= \frac{1}{T\left(\eta\varphi-\frac{\rho h(\tau)}{\tau\varepsilon^2}\right)}
\end{aligned}
$$

得证

如果倒着看呢？

我们要求解$F(w(T))-F(w^*)$的上界，可以直接用加和消项来求解，但作者应该是试过了这种方法却发现有些地方很难处理，因此通过求解$\frac{1}{F(w(T))-F(w^*)}$的下界来求解$F(w(T))-F(w^*)$的上界，这两者是等价的

$$
\begin{aligned}
\frac{1}{F(\mathbf{w}(T))-F(\mathbf{w}^*)} &\geq \frac{1}{F(\mathbf{w}(T))-F(\mathbf{w}^*)}-\frac{1}{\theta_{[1]}(0)} \\
&\geq T\left(\omega\eta\left(1-\frac{\beta\eta}{2}\right)-\frac{\rho h(\tau)}{\tau\varepsilon^2}\right)>0
\end{aligned}
$$

然后我们看如何得到右边这个式子的，左边放缩之前的式子可以改写为

$$
\frac{1}{F(w(T))-F(w^*)}-\frac{1}{\theta_{[K]}(T)}+\frac{1}{\theta_{[K]}(T)}-\frac{1}{\theta_{[1]}(0)}
$$

然后就分别求解$\frac{1}{F(w(T))-F(w^*)}-\frac{1}{\theta_{[K]}(T)}$和$\frac{1}{\theta_{[K]}(T)}-\frac{1}{\theta_{[1]}(0)}$，后面这项很明显可以用加和消项来处理，首先算出一个训练周期内的差值，再算出从训练开始到训练结束($t=T$)的差值

$$
\begin{aligned}
\frac{1}{\theta_{[k]}(k\tau)}-\frac{1}{\theta_{[k]}((k-1)\tau)} &= \sum_{z=(k-1)\tau}^{k\tau-1}\left(\frac{1}{\theta_{[k]}(t+1)}-\frac{1}{\theta_{[k]}(t)}\right) \\
&\geq \tau\omega\eta\left(1-\frac{\beta\eta}{2}\right)
\end{aligned}
$$

Summing up the above for all $k=1,2...,K$ yields

$$
\begin{aligned}
\sum_{k=1}^{K}\left(\frac{1}{\theta_{[k]}(k\tau)}-\frac{1}{\theta_{[k]}((k-1)\tau)}\right) &\geq \sum_{k=1}^{K}\tau\omega\eta\left(1-\frac{\beta\eta}{2}\right) \\
&= K\tau\omega\eta\left(1-\frac{\beta\eta}{2}\right)
\end{aligned}
$$

Rewriting the left-hand side and noting that $T=K\tau$ yields

$$
\begin{aligned}
&\frac{1}{\theta_{[K]}(T)}-\frac{1}{\theta_{[1]}(0)}-\sum_{k=1}^{K-1}\left(\frac{1}{\theta_{[k+1]}(k\tau)}-\frac{1}{\theta_{[k]}(k\tau)}\right) \\
&\geq T\omega\eta\left(1-\frac{\beta\eta}{2}\right)
\end{aligned}
$$

which is equivalent to i

$$
\frac{1}{\theta_{[K]}(T)}-\frac{1}{\theta_{[1]}(0)}
$$

$$
\geq T\omega\eta\left(1-\frac{\beta\eta}{2}\right)+\sum_{k=1}^{K-1}\left(\frac{1}{\theta_{[k+1]}(k\tau)}-\frac{1}{\theta_{[k]}(k\tau)}\right) \tag{30}
$$

Each term in the sum in right-hand side of (30) can be further expressed as

$$
\begin{aligned}
\frac{1}{\theta_{[k+1]}(k\tau)}-\frac{1}{\theta_{[k]}(k\tau)} &= \frac{\theta_{[k]}(k\tau)-\theta_{[k+1]}(k\tau)}{\theta_{[k]}(k\tau)\theta_{[k+1]}(k\tau)} \\
&= \frac{F(\mathbf{v}_{[k]}(k\tau))-F(\mathbf{v}_{[k+1]}(k\tau))}{\theta_{[k]}(k\tau)\theta_{[k+1]}(k\tau)} \\
&\geq \frac{-\rho h(\tau)}{\theta_{[k]}(k\tau)\theta_{[k+1]}(k\tau)} \qquad (31)
\end{aligned}
$$

where the last inequality is obtained using Theorem 1 and noting that, according to the definition, $\mathbf{v}_{[k+1]}(k\tau)=\mathbf{w}(k\tau)$, thus $F(\mathbf{v}_{[k+1]}(k\tau))=F(\mathbf{w}(k\tau))$.

It is assumed that $F(\mathbf{v}_{[k]}(k\tau))-F(\mathbf{w}^*)\geq\varepsilon$ for all k. According to Lemma 5, $F(\mathbf{v}_{[k]}(t))\geq F(\mathbf{v}_{[k]}(t+1))$ for any $t\in[(k-1)\tau,k\tau)$. Therefore, we have $\theta_{[k]}(t)=F(\mathbf{v}_{[k]}(t))-F(\mathbf{w}^*)\geq\varepsilon$ for all t and k for which $\mathbf{v}_{[k]}(t)$ is defined. Consequently,

$$
\begin{aligned}
\theta_{[k]}(k\tau)\theta_{[k+1]}(k\tau) &\geq \varepsilon^2 \\
\frac{-1}{\theta_{[k]}(k\tau)\theta_{[k+1]}(k\tau)} &\geq -\frac{1}{\varepsilon^2} \qquad (32)
\end{aligned}
$$

Combining (32) with (31), the sum in the right-hand side of (30) can be bounded by

$$
\begin{aligned}
\sum_{k=1}^{K-1}\left(\frac{1}{\theta_{[k+1]}(k\tau)}-\frac{1}{\theta_{[k]}(k\tau)}\right) &\geq -\sum_{k=1}^{K-1}\frac{\rho h(\tau)}{\varepsilon^2} \\
&= -(K-1)\frac{\rho h(\tau)}{\varepsilon^2}
\end{aligned}
$$

Substituting the above into (30), we get

$$
\frac{1}{\theta_{[K]}(T)}-\frac{1}{\theta_{[1]}(0)}\geq T\omega\eta\left(1-\frac{\beta\eta}{2}\right)-(K-1)\frac{\rho h(\tau)}{\varepsilon^2} \tag{33}
$$

(33)

前面这一项可以从已证的下式中变换而来：

$$\frac{-1}{\theta_{[k]}(k\tau)\theta_{[k+1]}(k\tau)} \geq -\frac{1}{\varepsilon^2}$$

We then have

$$\begin{aligned}\frac{1}{F(\mathbf{w}(T))-F(\mathbf{w}^*)}-\frac{1}{\theta_{[K]}(T)} &= \frac{\theta_{[K]}(T)-(F(\mathbf{w}(T))-F(\mathbf{w}^*))}{(F(\mathbf{w}(T))-F(\mathbf{w}^*))\,\theta_{[K]}(T)} \\ &= \frac{F(\mathbf{v}_{[K]}(T))-F(\mathbf{w}(T))}{(F(\mathbf{w}(T))-F(\mathbf{w}^*))\,\theta_{[K]}(T)} \\ &\geq \frac{-\rho h(\tau)}{(F(\mathbf{w}(T))-F(\mathbf{w}^*))\,\theta_{[K]}(T)} \\ &\geq -\frac{\rho h(\tau)}{\varepsilon^2} \qquad (35)\end{aligned}$$

还是那个问题，如何得到这个上界的？（论文只给出了证明，并没有讲如何求解）

顺着推，就能求得这个上界（逆着推就是证明）