Fair Resource Allocation in Federated Learning

Tian Li et.   CMU   ICLR2020

Power by 丸一口

CONTENT

Part 01

Fairness definition

Two reasonable definition of fairness.

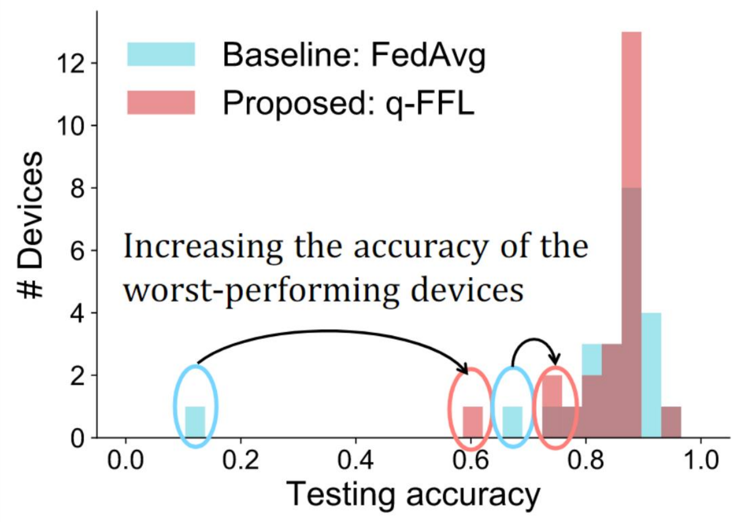
Fairness

More Average: Emphasize that everyone **has equal opportunities** and care about **poor** clients;

均衡公平性[1]：强调 **人人平等有机会** ，关心 **表现差** 的客户；

Contribution: emphasizing distribution according to work, where those who **work more receive more.**

贡献公平性[1]：强调 **按劳分配，多劳多得，优胜劣汰** 。



Definition 1 (*Fairness of performance distribution*)**.** For trained models w and $\tilde{w}$, we informally say that model w provides a more *fair* solution to the federated learning objective (1) than model $\tilde{w}$ if the performance of model w on the m devices, $\{a_1, \dots a_m\}$, is more *uniform* than the performance of model $\tilde{w}$ on the m devices.

Uniformity： In this work, we mainly use the *variance* of the *performance* distribution as a measure of uniformity

Performance: In this work, we take 'performance', a_k, to be the *testing accuracy* of applying the trained model w on the test data for device k.

[1] https://zhuanlan.zhihu.com/p/601227861

Part 02

Algorithm

How to improve the fairness of Federated Learning?

Objection function

FedAvg

$$\min_{w} f(w) = \sum_{k=1}^{m} p_k F_k(w)$$

$\boxed{m \text{ is the total number of devices, } p_k \geq 0, \text{ and } \sum_k p_k = 1}$

q-Fair Federated Learning (q-FFL)

$$\min_{w} \; f_q(w) = \sum_{k=1}^{m} \frac{p_k}{q+1} F_k^{q+1}(w)$$

q = 0 → FedAvg
q increase... → Focus on the big one of F_k (performace bad)
q = ∞ → Agnostic Federated Learning (AFL)[2]

[2] Mohri, Mehryar, Gary Sivek, and Ananda Theertha Suresh. "Agnostic federated learning." International Conference on Machine Learning. PMLR, 2019. 不可知论的联邦学习

Extra parameter: q

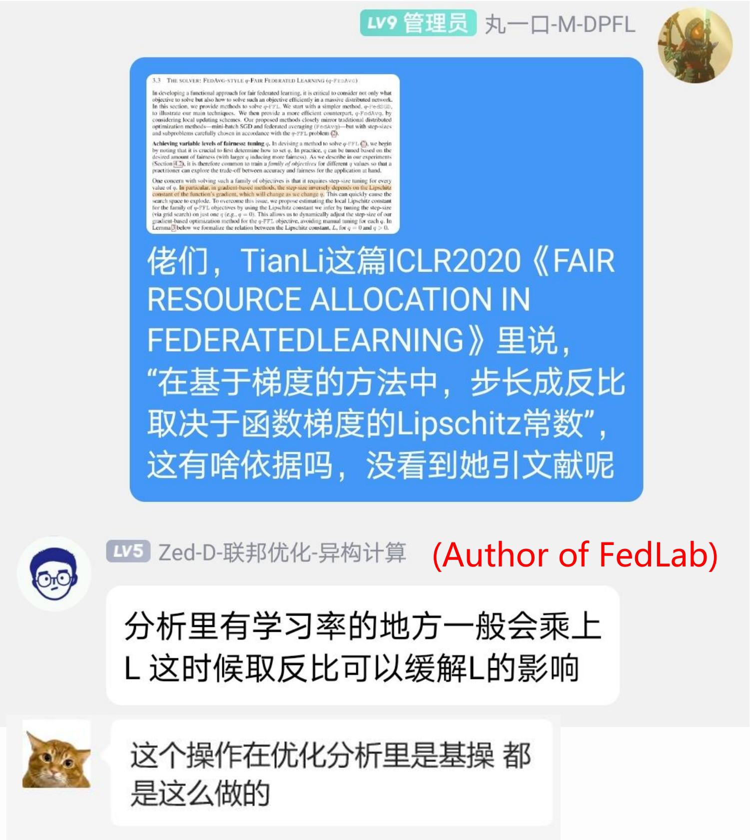
$$\min_{w} \; f_q(w) = \sum_{k=1}^{m} \frac{p_k}{q+1} F_k^{q+1}(w)$$

Effectiveness of q: larger q inducing more fairness

How to get a proper q: grid search

Reduce the parameter search: auto choice learning rate

Auto choice learning rate



One concern with solving such a family of objectives is that it requires step-size tuning for every value of q. In particular, in gradient-based methods, the step-size inversely depends on the Lipschitz constant of the function's gradient, which will change as we change q. This can quickly cause the

$$q \rightarrow L(q) \rightarrow \eta \propto 1/L(q)$$

实际上是二阶导的上界作为L

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta \nabla F(\mathbf{w}_t)$$

$$\rightarrow \eta \nabla F(\mathbf{w}_t) = \mathbf{w}_t - \mathbf{w}_{t+1}$$

$$\rightarrow \frac{1}{\eta} = \frac{\nabla F(\mathbf{w}_t)}{\mathbf{w}_t - \mathbf{w}_{t+1}}$$

$$\rightarrow \frac{\nabla F(\mathbf{w}_{t+1}) - \nabla F(\mathbf{w}_t)}{\mathbf{w}_{t+1} - \mathbf{w}_t} = \frac{\nabla F(\mathbf{w}_t) - \nabla F(\mathbf{w}_{t+1})}{\mathbf{w}_t - \mathbf{w}_{t+1}} \leq \frac{\nabla F(\mathbf{w}_t)}{\mathbf{w}_t - \mathbf{w}_{t+1}} = \frac{1}{\eta} = L$$

Compute L

q → L(q) → η∝ 1/L(q)

所以 Lipschitz continuous **gradient**意味着: [3]

$$\|f'(x) - f'(y)\| \leqslant L\|x - y\|$$

$$\min_{w} f_q(w) = \sum_{k=1}^{m} \frac{p_k}{q+1} F_k^{q+1}(w)$$

Lemma 3. *If the non-negative function $f(\cdot)$ has a Lipschitz gradient with constant L, then for any $q \geq 0$ and at any point w,*

$$L_q(w) = Lf(w)^q + qf(w)^{q-1}\|\nabla f(w)\|^2 \tag{3}$$

is an upper-bound for the local Lipschitz constant of the gradient of $\frac{1}{q+1}f^{q+1}(\cdot)$ at point w.

Proof. At any point w, we can compute the Hessian $\nabla^2\left(\frac{1}{q+1}f^{q+1}(w)\right)$ as: $\boxed{\nabla\left(\frac{1}{q+1}f^{q+1}(w)\right) = f^q(w)\cdot\nabla f(w)}$

$$\nabla^2\left(\frac{1}{q+1}f^{q+1}(w)\right) = qf^{q-1}(w)\underbrace{\nabla f(w)\nabla^T f(w)}_{\preceq\|\nabla f(w)\|^2\times I} + f^q(w)\underbrace{\nabla^2 f(w)}_{\preceq L\times I}. \tag{4}$$

As a result, $\|\nabla^2\frac{1}{q+1}f^{q+1}(w)\|_2 \leq L_q(w) = Lf(w)^q + qf(w)^{q-1}\|\nabla f(w)\|^2$. □

[3] 非凸优化基石：Lipschitz Condition - Zeap的文章 - 知乎 https://zhuanlan.zhihu.com/p/27554191

q-FedSGD

$$\min_{w} f_q(w) = \sum_{k=1}^{m} \frac{p_k}{q+1} F_k^{q+1}(w)\,, \qquad (2)$$

$$L_q(w) = Lf(w)^q + qf(w)^{q-1}\|\nabla f(w)\|^2 \qquad (3)$$

Algorithm 1 q-`FedSGD`

1: **Input:** $K, T, q, 1/L, w^0, p_k, k = 1, \cdots, m$
2: **for** $t = 0, \cdots, T-1$ **do**
3: Server selects a subset S_t of K devices at random (each device k is chosen with prob. p_k)
4: Server sends w^t to all selected devices
5: Each selected device k computes:

$$\Delta_k^t = F_k^q(w^t)\nabla F_k(w^t)$$ grad of (2)

$$h_k^t = qF_k^{q-1}(w^t)\|\nabla F_k(w^t)\|^2 + LF_k^q(w^t)$$ compute (3)

6: Each selected device k sends Δ_k^t and h_k^t back to the server
7: Server updates w^{t+1} as:

$$w^{t+1} = w^t - \frac{\sum_{k \in S_t} \Delta_k^t}{\sum_{k \in S_t} h_k^t}$$ use 1/L as η

8: **end for**

q-FedAvg

$$\min_{w} f_q(w) = \sum_{k=1}^{m} \frac{p_k}{q+1} F_k^{q+1}(w)\,, \tag{2}$$

$$L_q(w) = Lf(w)^q + qf(w)^{q-1}\|\nabla f(w)\|^2 \tag{3}$$

E次梯度变化值，L是当做1/η在用

Algorithm 2 q-FedAvg

Input: K, E, T, q, $1/L$, η, w^0, p_k, $k = 1, \cdots, m$

for $t = 0, \cdots, T-1$ **do**

Server selects a subset S_t of K devices at random (each device k is chosen with prob. p_k)

Server sends w^t to all selected devices

Each selected device k updates w^t for E epochs of SGD on F_k with step-size η to obtain $\bar{w}_k^{t+1}$

Each selected device k computes:

$$\boxed{\Delta w_k^t = L(w^t - \bar{w}_k^{t+1})}$$?

$$\Delta_k^t = F_k^q(w^t)\Delta w_k^t$$ grad of (2)

$$h_k^t = qF_k^{q-1}(w^t)\|\Delta w_k^t\|^2 + LF_k^q(w^t)$$ compute (3)

$$\Delta_k^t = F_k^q(w^t)\nabla F_k(w^t)$$ FedSGD

$$h_k^t = qF_k^{q-1}(w^t)\|\nabla F_k(w^t)\|^2 + LF_k^q(w^t)$$

Each selected device k sends Δ_k^t and h_k^t back to the server

Server updates w^{t+1} as:

$$w^{t+1} = w^t - \frac{\sum_{k \in S_t} \Delta_k^t}{\sum_{k \in S_t} h_k^t}$$ use 1/L as η

end for

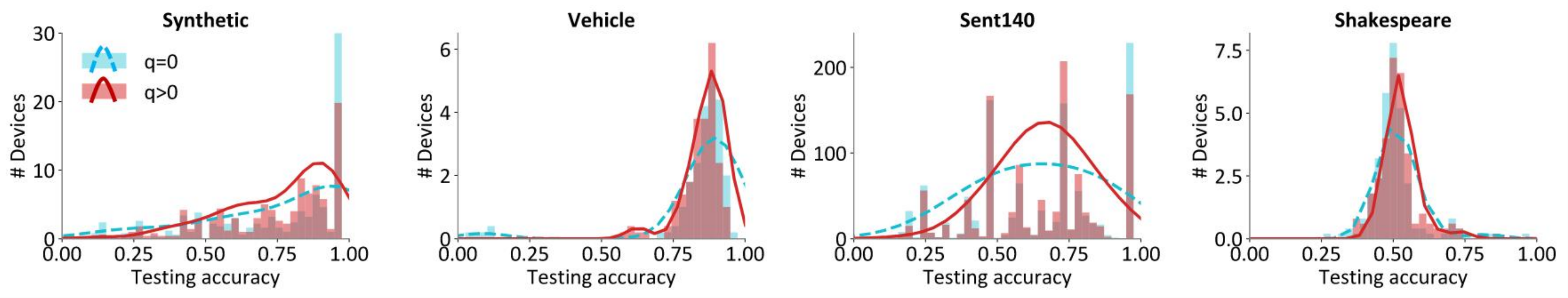
[3] 非凸优化基石：Lipschitz Condition - Zeap的文章 - 知乎 https://zhuanlan.zhihu.com/p/27554191

Part 03

Experiment

How to construct a wonderful paper

Extra parameter: q



Dataset	Objective	Average (%)	Worst 10% (%)	Best 10% (%)	Variance
Synthetic	$q = 0$	80.8 ± .9	18.8 ± 5.0	100.0 ± 0.0	724 ± 72
	$q = 1$	79.0 ± 1.2	**31.1** ± 1.8	100.0 ± 0.0	**472** ± 14
Vehicle	$q = 0$	87.3 ± .5	43.0 ± 1.0	**95.7** ± 1.0	291 ± 18
	$q = 5$	87.7 ± .7	**69.9** ± .6	94.0 ± .9	**48** ± 5
Sent140	$q = 0$	65.1 ± 4.8	15.9 ± 4.9	100.0 ± 0.0	697 ± 132
	$q = 1$	66.5 ± .2	**23.0** ± 1.4	100.0 ± 0.0	**509** ± 30
Shakespeare	$q = 0$	51.1 ± .3	39.7 ± 2.8	**72.9** ± 6.7	82 ± 41
	$q = .001$	52.1 ± .3	**42.1** ± 2.1	69.0 ± 4.4	**54** ± 27

Thanks~

Power by 丸一口