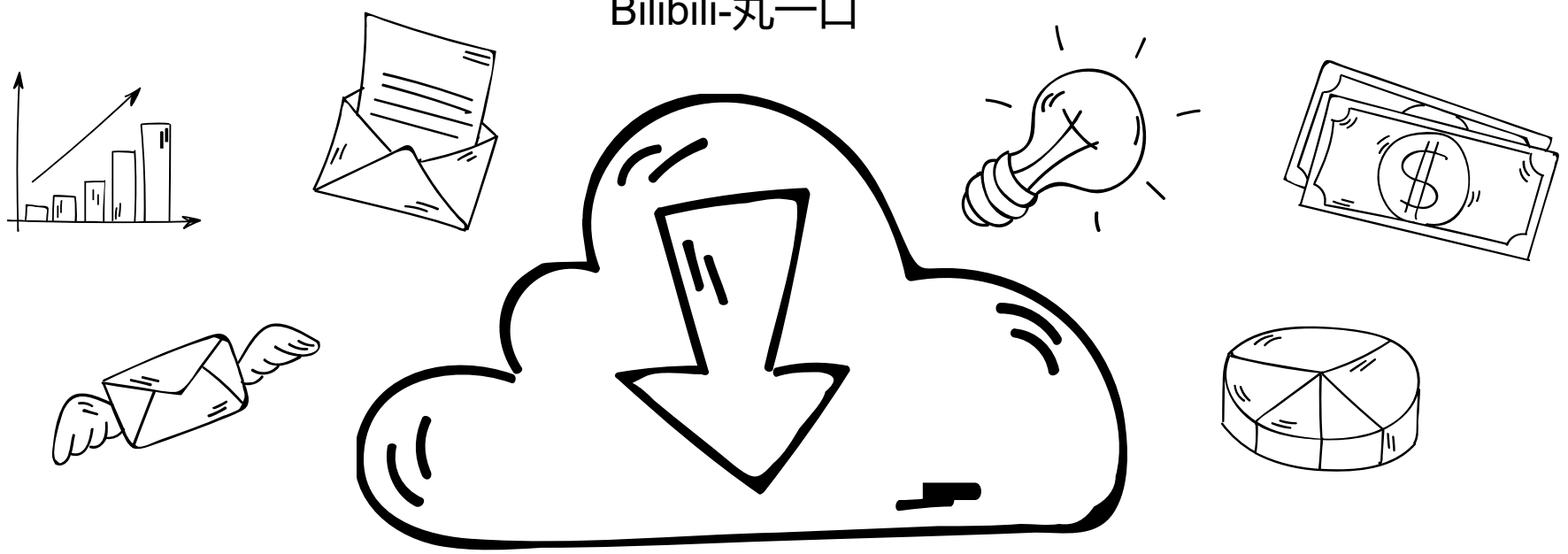
Bilibili-丸一口



Ditto: Fair and Robust Federated Learning Through Personalization

Ditto：通过个性化实现公平和鲁棒的联邦学习

[1] Li, Tian, et al. "Ditto: Fair and robust federated learning through personalization." International Conference on Machine Learning. PMLR, 2021. 谷歌学术引用399

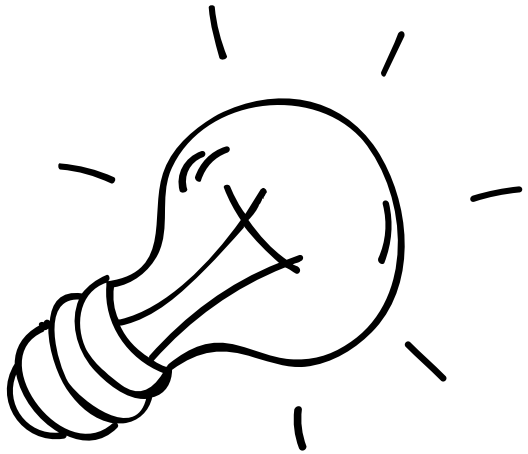
CONTENT

背景介绍

算法介绍

λ取值

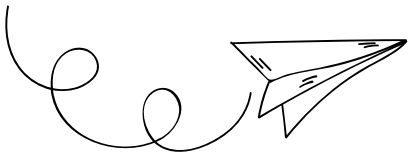
实验



PART 01

背景介绍

“个性化” 及 Fair and Robust



两种类型的“个性化”

Personalization

之前读的一些文章：在某些场景下，提升全局模型的准确率

- 资源受限[2]：自适应epoch，在资源受限的情况下得到acc高的全局模型
- FedNova[4]：对异构的local epoch做归一化，使得算力异构的client能够训练出一个acc高的全局模型

Ditto：在某些场景下，既考虑本地模型的准确率，又考虑全局模型的准确率

- 因为在数据异质的情况下，即使全局模型的准确率高，把训练好的模型拿过来给本地客户端用，效果(acc)不一定很好。
- pFedLA[3]：引入超网络对权重进行管理，实现分层聚合，让重要的层得到更高的权重，在数据异构的场景下得到个性化模型

[2] Wang, Shiqiang, et al. "Adaptive federated learning in resource constrained edge computing systems." IEEE journal on selected areas in communications 37.6 (2019): 1205-1221.
[3] Ma, Xiaosong, et al. "Layer-wised model aggregation for personalized federated learning." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.
[4] Wang, Jianyu, et al. "Tackling the objective inconsistency problem in heterogeneous federated optimization." Advances in neural information processing systems 33 (2020): 7611-7623.

公平和鲁棒的定义

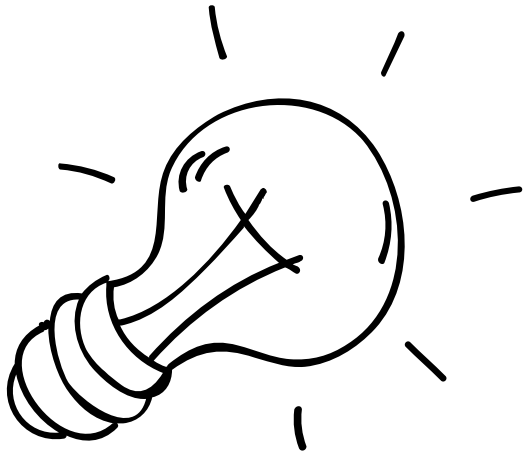
Fair and Robust

Fair（评价指标：良性节点测试准确率方差）

- 如果w_1在网络上的测试性能分布比w_2的更均匀，即$\mathrm{std}\{F_k(w_1)\}_{k\in[K]} < \mathrm{std}\{F_k(w_2)\}_{k\in[K]}$，则模型$w_1$比$w_2$更公平
- 其中，F_k指设备k上的测试损失
- std指标准差: $\mathrm{std}\{F_k(w)\}_{k\in[K]} = \sqrt{\frac{1}{|K|}\sum_{k=1}^{K}(F_k(w)-\bar{F})^2}$
- 有敌手的时候，只在善意的节点上衡量Fair

Robust（评价指标：良性节点的准确率）

- 我们在概念上对**拜占庭鲁棒性**感兴趣（Lamport et al.，2019）
  - 在拜占庭鲁棒性中，恶意设备可以向服务器发送任意更新，以破坏训练。
  - 为了测量鲁棒性，我们评估**良性设备上的平均测试性能**，即，如果在使用攻击进行训练后，良性设备上模型w1的平均测试表现高于w2，则我们认为模型w1对特定攻击的鲁棒性高于w2
- 攻击种类
  - 标签中毒
  - 随机更新
  - 模型替换



PART 02

算法介绍

简单得一批，就只有一个正则化项

联邦多任务学习

这篇论文原来的标题就是叫Federated Multi-Task Learning for Competing Constraints

联邦学习有三大约束：公平性、鲁棒性、隐私性。目前的方法很难同时满足这三个约束，本文主要关注联邦学习的公平性和鲁棒性。

而统计异构性(即不相同的分布数据，Non-IID)是限制这三个约束不能同时上升的根本原因。作者认为**个性化**联邦学习的方法（通过学习每个设备的不同模型来适应联邦设置中的异构性）可能有助于提升公平性和鲁棒性。而**多任务学习**本身提供了内在的鲁棒性和公平性好处，并探索了试图同时满足这两个约束时存在的挑战。

$$\begin{aligned} \min_{v_k} \quad & h_k\left(v_k ; w^*\right) := F_k\left(v_k\right) + \frac{\lambda}{2}\left\|v_k - w^*\right\|^2 \\ \text{s.t.} \quad & w^* \in \arg\min_{w} G\left(F_1(w), \ldots F_K(w)\right) \end{aligned} \qquad (Ditto)$$

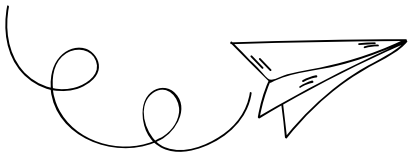
第一个目标：min hk，就是去找使得local model好的模型
第二个目标：min G，就是去找使得global model好的模型

λ的取值

怎样就变成无中心化的场景，怎样就变成FedAVG

$$\begin{aligned} \min_{v_k} \quad & h_k\left(v_k; w^*\right) := F_k\left(v_k\right) + \frac{\lambda}{2}\left\|v_k - w^*\right\|^2 \\ \text{s.t.} \quad & w^* \in \arg\min_{w} G\left(F_1(w), \ldots F_K(w)\right) \end{aligned} \qquad (Ditto)$$

- ✓ 如果λ=0，则hk的优化问题等价于Fk的优化问题，hk就是纯在优化本地model vk，vk的优化只跟自己的梯度下降有关系，跟全局模型w没关系（w的优化不受vk的影响）
- ✓ 如果λ=∞，则vk=w*，则两个目标都变成了去找全局最优模型w*



Ditto

Ditto for Personalized FL

$$\begin{aligned} \min_{v_k} \quad & h_k\left(v_k; w^*\right) := F_k\left(v_k\right) + \frac{\lambda}{2}\left\|v_k - w^*\right\|^2 \\ \text{s.t.} \quad & w^* \in \arg\min_{w} G\left(F_1(w), \ldots F_K(w)\right) \end{aligned} \qquad (Ditto)$$

Algorithm 1: `Ditto` for Personalized FL

1 **Input:** $K, T, s, \lambda, \eta, w^0, \{v_k^0\}_{k\in[K]}$

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

3 Server randomly selects a subset of devices S_t, and sends w^t to them

4 **for** *device* $k \in S_t$ *in parallel* **do**

5 Solve the local sub-problem of $G(\cdot)$ inexactly starting from w^t to obtain w_k^t:

$w_k^t \leftarrow \text{UPDATE\_GLOBAL}(w^t, \nabla F_k(w^t))$

/* Solve $h_k(v_k; w^t)$ */

6 Update v_k for s local iterations:

$v_k = v_k - \eta(\nabla F_k(v_k) + \lambda(v_k - w^t))$

Send $\Delta_k^t := w_k^t - w^t$ back

7 Server aggregates $\{\Delta_k^t\}$:

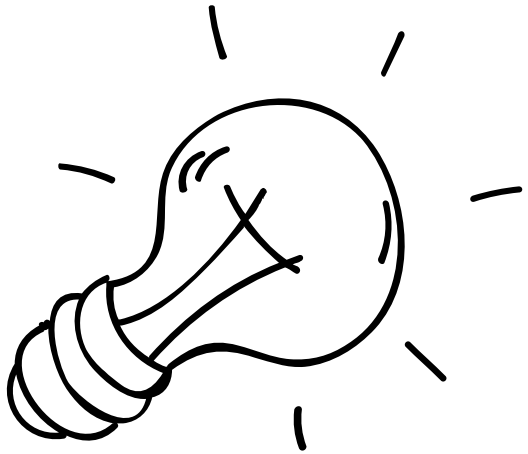
$w^{t+1} \leftarrow \text{AGGREGATE}\left(w^t, \{\Delta_k^t\}_{k\in\{S_t\}}\right)$

8 **return** $\{v_k\}_{k\in[K]}$ *(personalized)*, w^T *(global)*

①注意他返回的是俩，一个是vk，一个是wT

②这里是本地更新，如果是FedAvg的话，就是 $w_k^t = w_k^t - \eta_g \nabla F_k(w_k^t)$

③第6行在解hk，梯度由hk求导得到，算法十分简单



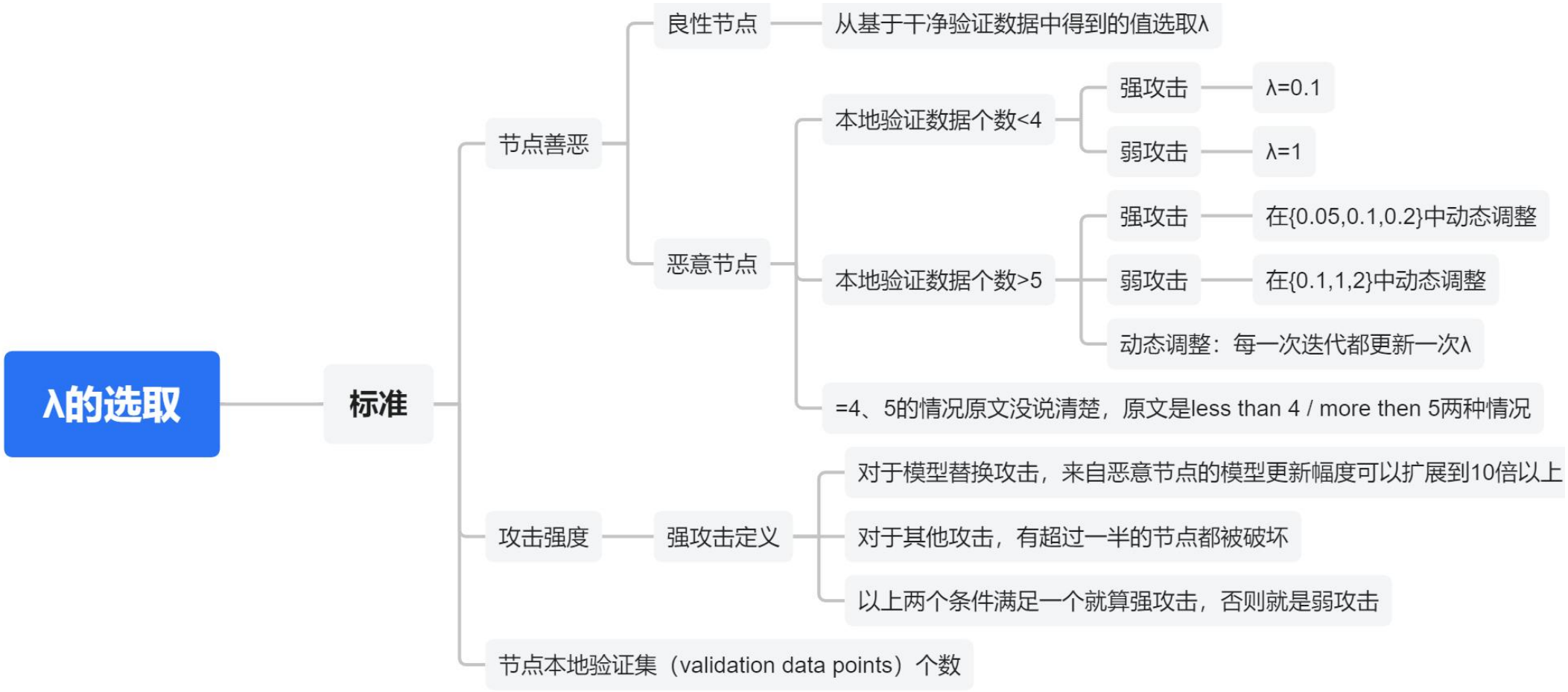
PART 03

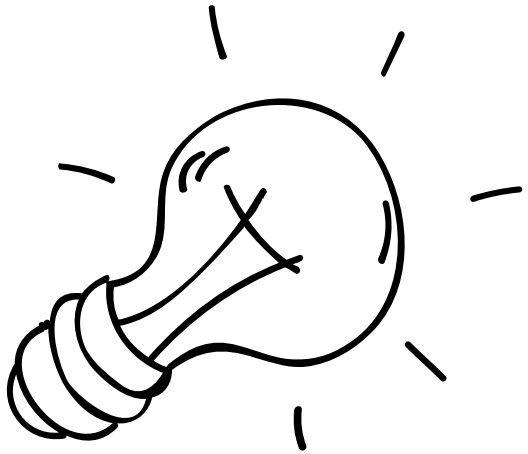
λ的取值

怎么通过改变λ的取值，做到Fair和Robust的trade-off呢?

λ的取值

实验里怎么取


λ的选取
标准
节点善恶
良性节点
从基于干净验证数据中得到的值选取λ
恶意节点
本地验证数据个数<4
强攻击
λ=0.1
弱攻击
λ=1
本地验证数据个数>5
强攻击
在{0.05,0.1,0.2}中动态调整
弱攻击
在{0.1,1,2}中动态调整
动态调整：每一次迭代都更新一次λ
=4、5的情况原文没说清楚，原文是less than 4 / more then 5两种情况
攻击强度
强攻击定义
对于模型替换攻击，来自恶意节点的模型更新幅度可以扩展到10倍以上
对于其他攻击，有超过一半的节点都被破坏
以上两个条件满足一个就算强攻击，否则就是弱攻击
节点本地验证集（validation data points）个数

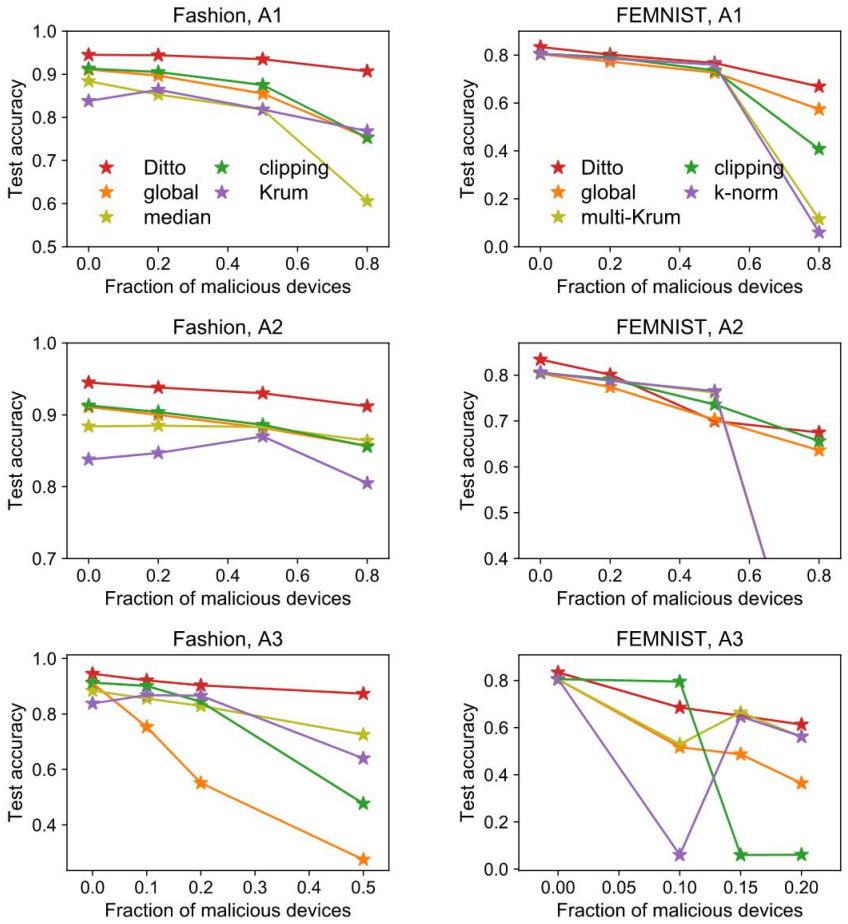


PART 04

实验

鲁棒性实验、公平性实验

鲁棒性实验



- global：learning a global model，按我的理解就是FedAvg
- clipping：“弱”差分隐私[5]，加噪没有严格按照DP来，加了“少量足以抵御(后门)攻击的噪声量”
- 其他的是一些strong defense mechanisms
- A1、A2、A3是攻击种类（标签中毒、随机更新、模型替换）

这里Robust实验的结果就是，Ditto在大多数情况都是best

[5] Sun, Ziteng, et al. "Can you really backdoor federated learning?." arXiv preprint arXiv:1911.07963 (2019).谷歌学术393引用

公平性实验

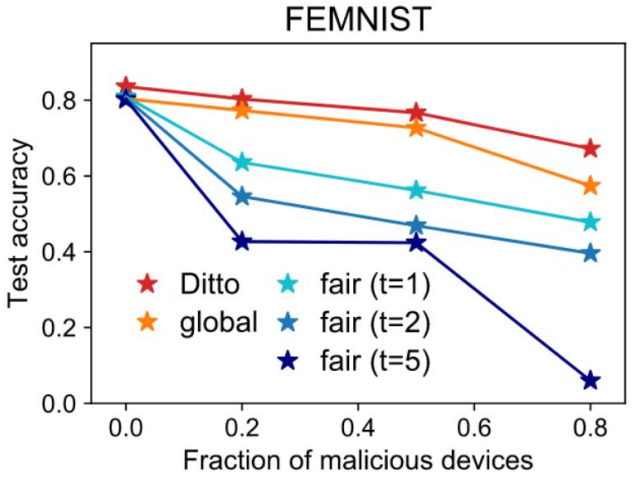
Fashion		**A1** (ratio of adversaries)			**A2** (ratio of adversaries)			**A3** (ratio of adversaries)		
Methods	clean	20%	50%	80%	20%	50%	80%	10%	20%	50%
global	.911 (.08)	.897 (.08)	.855 (.10)	.753 (.13)	.900 (.08)	.882 (.09)	.857 (.10)	.753 (.10)	.551 (.13)	.275 (.12)
local	.876 (.10)	.874 (.10)	.876 (.11)	.879 (.10)	.874 (.10)	.876 (.11)	.879 (.10)	.877 (.10)	.874 (.10)	**.876 (.11)**
fair (TERM, t=1)	.909 (.07)	.751 (.12)	.637 (.13)	.547 (.11)	.731 (.13)	.637 (.14)	.635 (.14)	.653 (.13)	.601 (.12)	.131 (.16)
Ditto	**.943 (.06)**	**.944 (.07)**	**.937 (.07)**	**.907 (.10)**	**.938 (.07)**	**.930 (.08)**	**.913 (.09)**	**.921 (.09)**	**.902 (.09)**	.873 (.11)
FEMNIST		**A1** (ratio of adversaries)			**A2** (ratio of adversaries)			**A3** (ratio of adversaries)		
Methods	clean	20%	50%	80%	20%	50%	80%	10%	15%	20%
global	.804 (.11)	.773 (.11)	.727 (.12)	.574 (.15)	.774 (.11)	**.703 (.14)**	.636 (.15)	.517 (.14)	.487 (.14)	.314 (.13)
local	.628 (.15)	.620 (.14)	.627 (.14)	.607 (.14)	.620 (.14)	.627 (.14)	.607 (.14)	.622 (.14)	.621 (.14)	**.620 (.14)**
fair (TERM, t=1)	.809 (.11)	.636 (.15)	.562 (.13)	.478 (.12)	.440 (.15)	.336 (.12)	.363 (.12)	.353 (.12)	.316 (.12)	.299 (.11)
Ditto	**.834 (.09)**	**.802 (.10)**	**.762 (.11)**	**.672 (.13)**	**.801 (.09)**	.700 (.15)	**.675 (.14)**	**.685 (.15)**	**.650 (.14)**	.613 (.13)

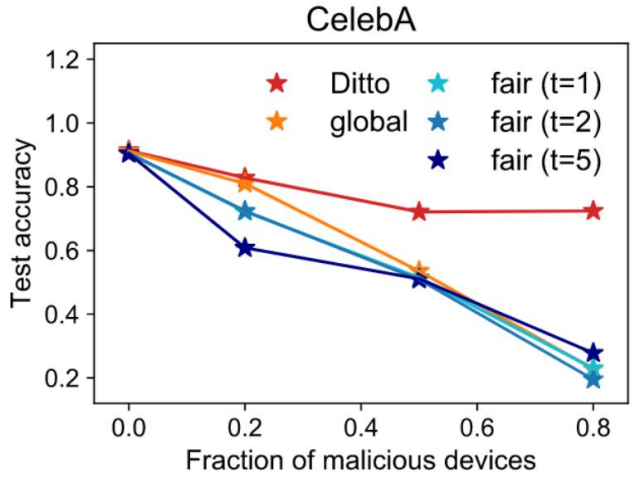
两个浮点数的含义是：准确率、准确率的标准差

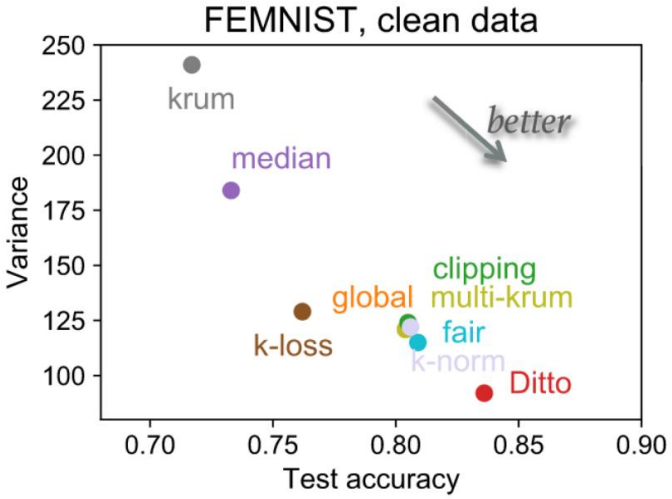
这里t是TERM(倾斜经验风险最小化)算法里的一个超参数，用来调控解决经验风险最小化问题
这个TERM下面也叫fair 方法

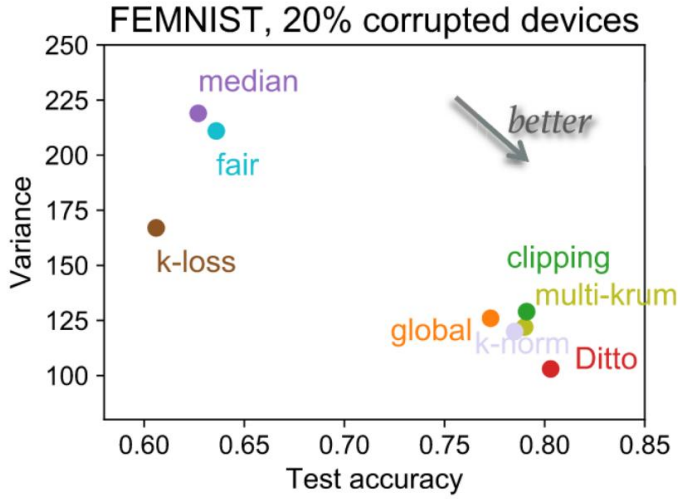
[6] Li, Tian, et al. "Tilted empirical risk minimization." arXiv preprint arXiv:2007.01162 (2020). 谷歌学术引用92，作者自己的paper

跟公平性方法TERM对比

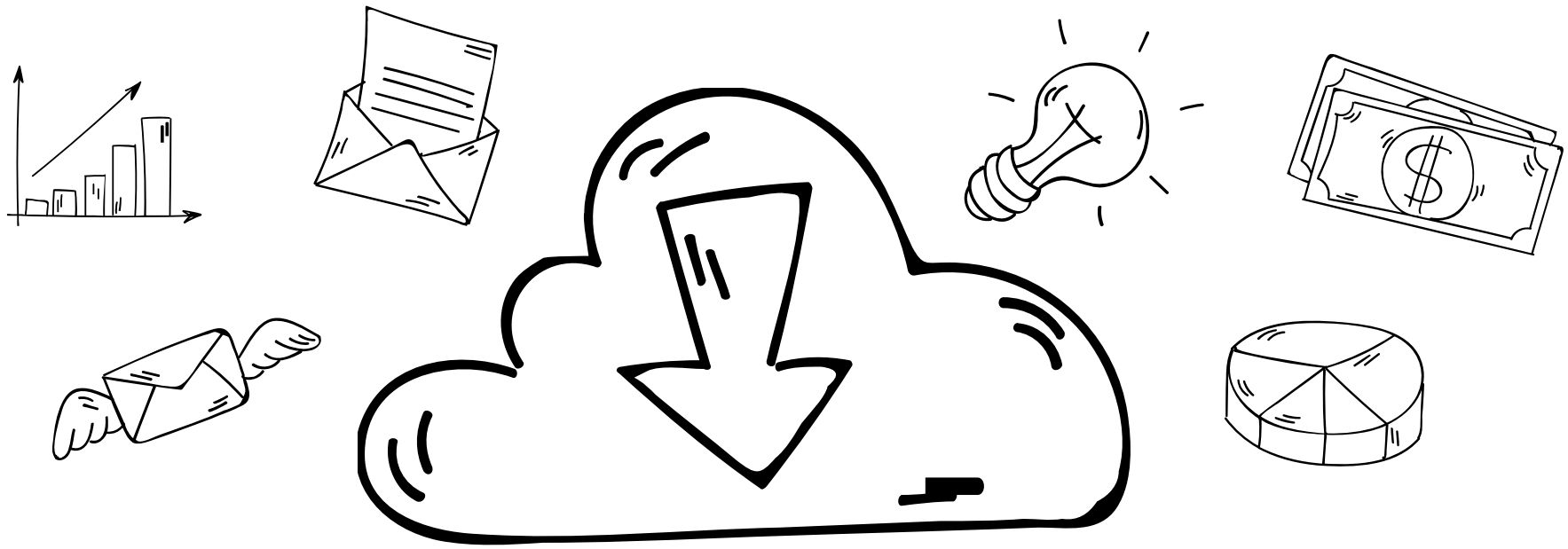








fair方法的鲁棒性不行，恶意节点增加，则准确率下降得很多



感谢您的观看

Bilibili-丸一口