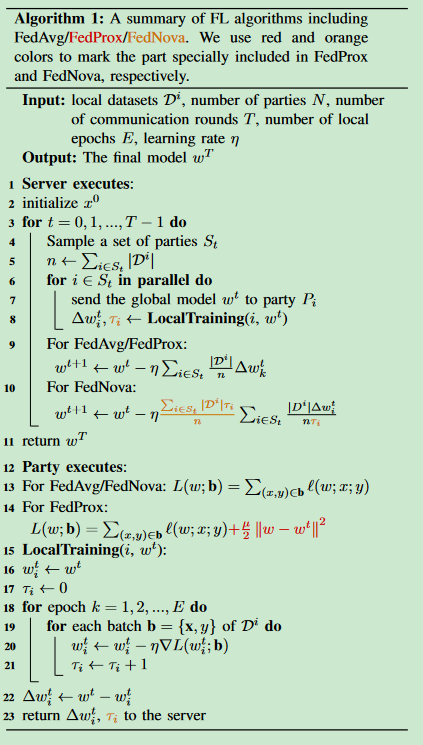
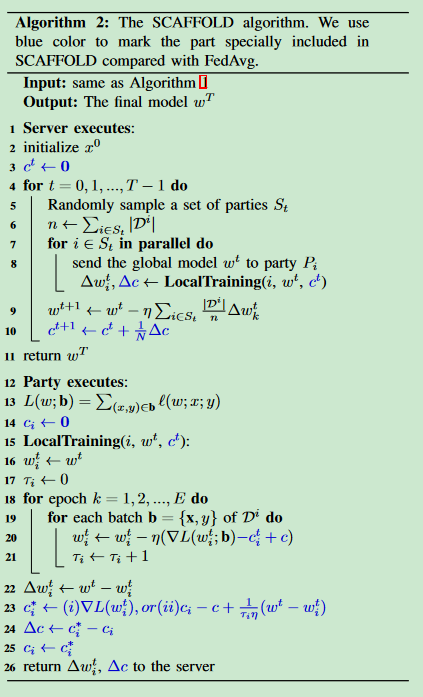
## INTRODUCTION

本文只考虑在横向联邦学习的环境下，通过构造六种不同的None-IID数据分布，测试四种联邦学习算法的效果。基于大量的实验，给出了没有一个联邦学习算法在所有的测试环境下都比其他算法好和Lable Distribution Skew的None-IID数据分布是最影响准确率的结论。

## PRELIMINARIES

FedProx：基于FedAvg，local training模块在loss函数里添加了全局模型和局部模型的距离。同时引入参数u来调整距离对loss函数的影响程度。选择合适的u确保局部最优模型更加接近全局最优模型。

FedNova：基于FedAvg，local traing模块需要计算computation次数ti，并返回给server。server根据ti来调整不同party的局部模型权重。ti越大的权重越小，反之亦然。通过此方法来削弱大数据量的party模型对全局模型的影响力，防止全局模型跑偏。

SCAFFOLD：通过计算party数据在全局模型上的梯度或者使用全局模型和新计算出来的局部模型差值来影响下一轮训练的梯度。使得局部最优模型向全局最优模型靠近。

## SIMULATING NON-IID DATA SETTING

[33]总结了五种None-IID分布：

* Label Distribution Skew：p(y)在party之间有差异。
* Feature Distribution Skew：p(x)在party之间有差异。
* Same Label But Different Features：p(x|y)在party之间有差异。
* Same Features But Different Labels：p(y|x)在party之间有差异
* Quantity Skew：数据量在party之间有差异。

### Label Distribution Skew

* Quantity-based label imbalance：给每个party尽可能均匀地分配k个标签，然后统计每个party所拥有的标签，对数据集按标签进行切分成子集，然后将这些子集平均分配给对应的party。
* Distribution-based label imbalance：对数据集按标签划分成子集，对每个子集按照Dirchlet分布给所有party分配相应的数据集（每分配一次子集重新计算Dirchlet）。

### Feature Distribution Skew

* Noise-based feature imbalance：将数据集平均划分给每个party，对每个party加不同程度的Gaussian noise。
* Synthetic feature imbalance：FCUBE规定数据集是一个三维（x1，x2，x3）的点，且根据x1=0将数据标签划为0或者1。然后再通过x1=0，x2=0，x3=0将Cube切成八块，随机生成一些点形成数据集。将中心对称的块分配给party。
* Real-world feature imbalance：EMNIST数据集是手写数据集，可以返回数据的writer。因为writer之间字迹天生不同，可以据此分配数据。

### Quantity Skew

根据Dirchlet函数对数据集进行划分，然后分配给party。

## EXPERIMENTS

### Overall Accuracy Comparison

* Comparison among different non-IID settings: The label distribution skew case where each party only has samples of a single class is the most challenging setting, while the feature distribution skew and quantity skew setting have little influence on the accuracy of FedAvg.
* Comparison among different algorithms: No algorithm consistently outperforms the other algorithms in all settings. The state-of-the-art algorithms significantly outperform FedAvg only in several cases.
* Comparison among different tasks: CIFAR-10 and tabular datasets are challenging tasks under non-IID settings. MNIST is a simple task under most non-IID settings where the studied algorithms perform similarly well.

### Communication Efficiency

FedProx has almost the same convergence speed compared with FedAvg, while SCAFFOLD and FedNova are more unstable in training.

### Robustness to Local Updates

The number of local epochs can have a large effect on the accuracy of existing algorithms. The optimal value of the number of local epochs is very sensitive to nonIID distributions.

### Party Sampling

In the partial participation setting, SCAFFOLD cannot work effectively, while the other FL algorithms have a very unstable accuracy during training.

### Scalability

The accuracy of all approaches decrease when increasing the number of parties.

### Efficiency

The computation overhead of FedProx is large compared with FedAvg. Moreover, the communication cost of SCAFFOLD is twice of that of FedAvg.

### Mixed Types of Skew

FL is more challenging when there exists mixed types of skew among the local data.