# Towards Addressing Label Skews in One-Shot Federated Learning

Yiqun Diao, Qinbin Li, Bingsheng He

https://github.com/Xtra-Computing/FedOV

NUS
National University of Singapore

# Background

- **Machine learning is data hungry**

- **Data are spread like islands**
  - Data privacy regulations: no sharing raw data

- **Federated learning (FL)**
  - Only sharing model
  - Non-IID challenge
  - Our problem: label skews

# Label skews

- **Defintion: non-IID label distribution among different clients**
  - E.g., disease types can differ among different hospitals

- **Challenge:**
  - Very different local optimas among clients
  - Prediction biased towards seen classes
  - NIID-Bench: label skews cause significant accuracy decay

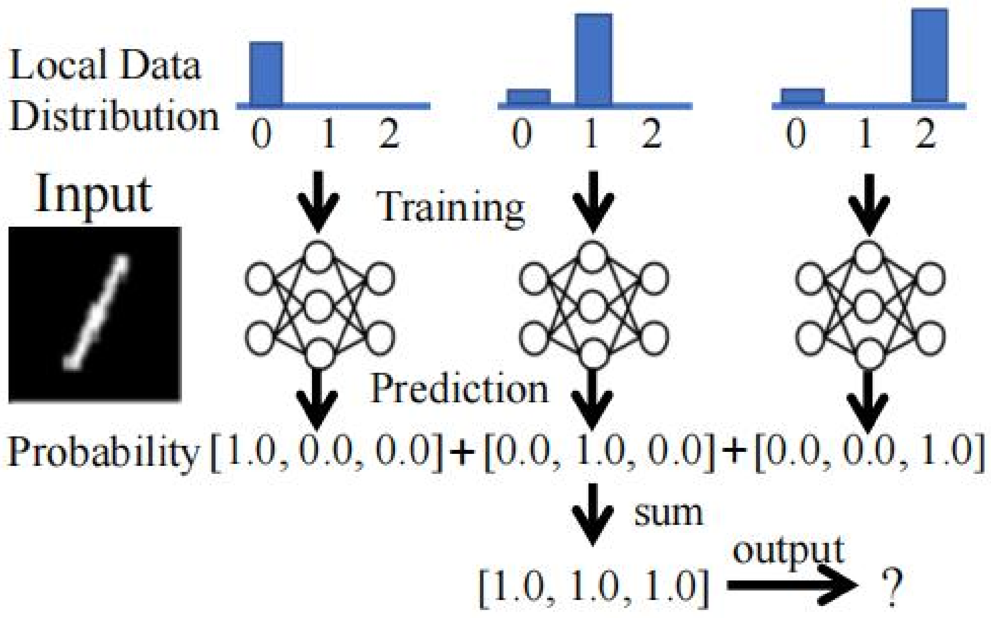
Our prior benchmark work

Fork 71 Starred 300

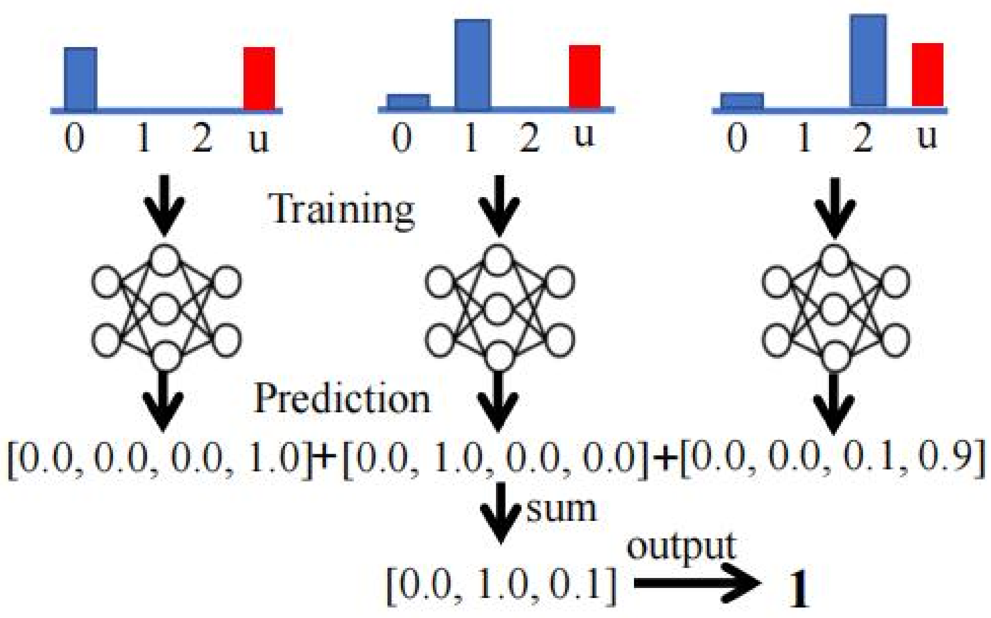
https://github.com/Xtra-Computing/NIID-Bench

# One-Shot FL: model uploaded only once

- **Ensemble: majority voting!**
  - Label skew: prediction biased to seen classes
  - Solution: introduce uncertainty! Open-set Recognition (OSR)
    - **We use a recent OSR algorithm PROSER as a naive baseline**



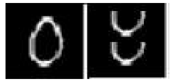
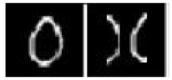
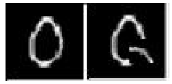
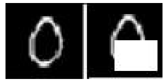
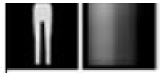
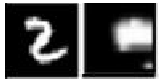

(a) Traditional close-set voting



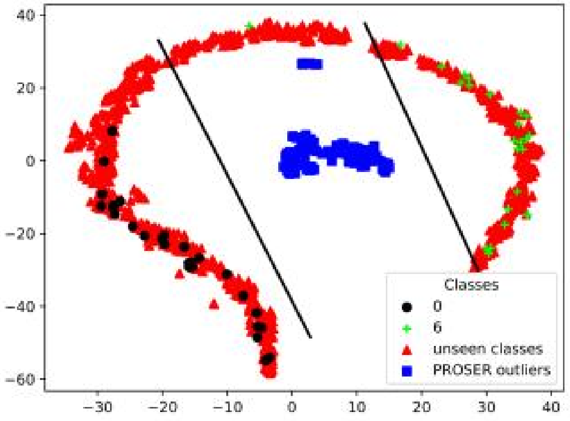

(b) Open-set voting framework (‘u’ denotes unknown class)

# To generate more diverse outliers

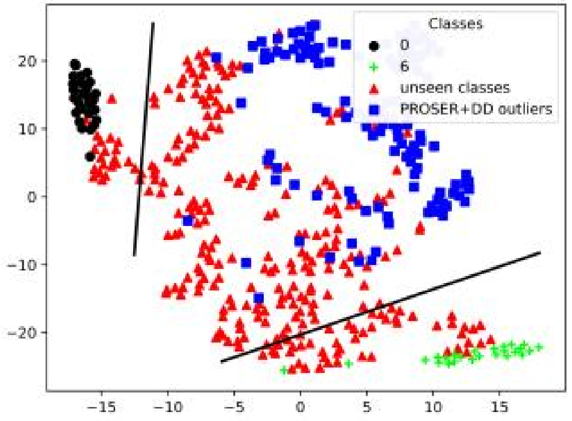


- **Data destruction (DD)**
  - RandomCopyPaste 
  - RandomSwap 
  - RandomRotation 
  - RandomErasing 
  - GaussianBlur 
  - RandomResizedCrop 

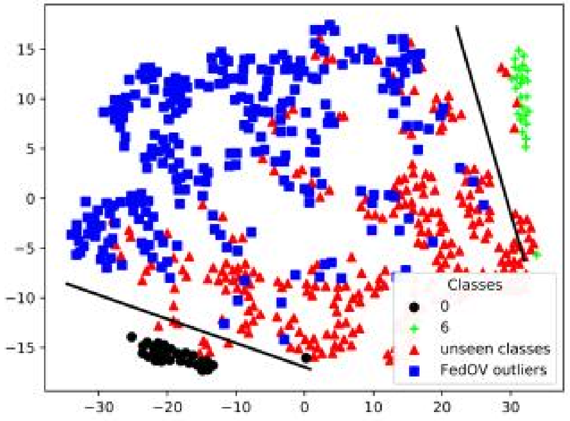
- **Adversarial Outlier Enhancement (AOE)**
  - Generate outliers even closer to true samples
  - Adversarial learning:
    - **Generated outliers x’ from data destruction**
    - **Transform x’ to x”, s.t. the model classifies x” as inliers**
    - **Put x” into training with label “unknown”**




(a) PROSER




(b) PROSER+DD




(c) FedOV

# Experiment

- **FedOV achieves SOTA accuracy in a round under various label skews**

| Dataset | Partition | FedOV | Close-set voting | FedAvg | FedProx | FedNova | SCAFFOLD | FedDF | FedKT |
|---|---|---|---|---|---|---|---|---|---|
| CIFAR-10 | $\#C=1$ | **40.0%±1.7%** | 10.2%±0.2% | 10.5%±1.0% | 10.6%±1.3% | 10.5%±1.0% | 10.5%±1.0% | 10.2%±0.5% | 9.8%±0.2% |
| | $\#C=2$ | **42.0%±2.4%** | 37.2%±2.5% | 11.1%±1.9% | 10.9%±1.6% | 10.5%±0.7% | 11.1%±1.8% | 18.8%±1.1% | 25.7%±2.9% |
| | $\#C=3$ | **55.6%±6.3%** | 43.2%±2.7% | 15.7%±5.1% | 15.9%±5.3% | 14.5%±3.9% | 16.1%±5.0% | 27.5%±4.0% | 31.8%±2.5% |
| | $p_k \sim Dir(0.5)$ | **65.7%±0.7%** | 65.0%±0.1% | 18.4%±7.2% | 18.7%±5.3% | 19.8%±7.2% | 18.6%±5.1% | 35.3%±0.9% | 42.1%±2.5% |
| | $p_k \sim Dir(0.1)$ | **61.7%±1.1%** | 55.9%±1.3% | 10.4%±0.4% | 11.1%±0.9% | 13.1%±3.3% | 13.0%±4.6% | 26.3%±3.0% | 35.0%±1.7% |

- **Both techniques (DD & AOE) are effective**

| Dataset | Partition | Close-set | Open-set (PROSER) | Open-set (PROSER + DD) | FedOV |
|---|---|---|---|---|---|
| CIFAR-10 | $\#C=1$ | 10.2%±0.2% | 10.6%±0.2% | 33.5%±2.3% | **40.0%±1.7%** |
| | $\#C=2$ | 37.2%±2.5% | 34.8%±4.5% | 41.3%±7.7% | **42.0%±2.4%** |
| | $\#C=3$ | 43.2%±2.7% | 50.2%±4.7% | 54.3%±2.1% | **55.6%±6.3%** |
| | $p_k \sim Dir(0.5)$ | 65.0%±0.1% | 66.6%±0.1% | **67.6%±0.3%** | 65.7%±0.7% |
| | $p_k \sim Dir(0.1)$ | 55.9%±1.3% | 58.0%±0.9% | 61.3%±1.0% | **61.7%±1.1%** |

# Conclusion

- **Propose an one-shot FL algorithm: FedOV**
  - Introducing uncertainty in prediction (OSR)
  - Generate diverse outliers during local training

- **FedOV achieves SOTA accuracy in a round under various label skews**

Follow our code!

https://github.com/Xtra-Computing/FedOV