

FEDERATED LEARNING WITH NON-IID DATA

PAPER REPORT

INTRODUCTION

Bài báo đề cập đến trường hợp mất cân bằng dữ liệu trong FL (mỗi máy client chỉ sở hữu dataset chứa một lượng label nhỏ, không bao quát) cũng như đề xuất một phương án giúp làm tăng hiệu suất cho mô hình khi rơi vào tình huống này.

Tác giả đưa ra cách chấm điểm mô hình là so sánh bộ trọng số của mô hình học phân tán lúc này với một mô hình chạy SGD với bộ dữ liệu đầy đủ, hay chính xác là đánh giá mức hiệu quả của phương án đề xuất dựa trên Weight divergence : $weight\ divergence = ||\mathbf{w}^{FedAvg} - \mathbf{w}^{SGD}|| / ||\mathbf{w}^{SGD}||$

PROBLEM

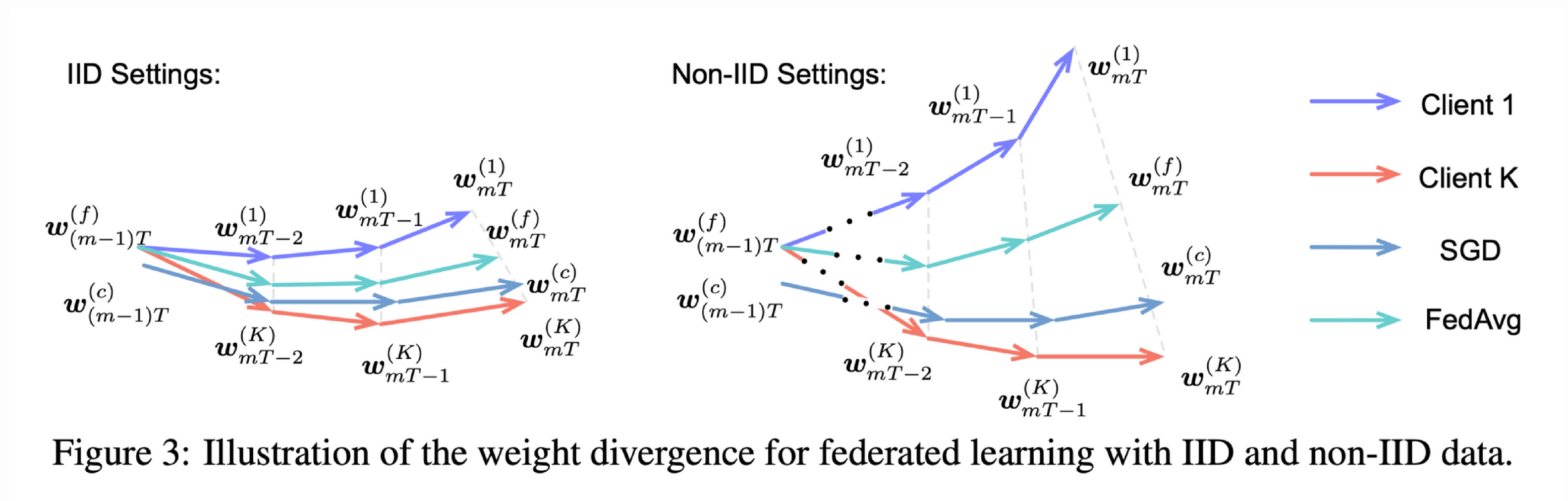


Figure 3: Illustration of the weight divergence for federated learning with IID and non-IID data.

DATA-SHARING STRATEGY

G-set: Global dataset consist the uniform distribution over classes

The warm-up, model trained on **G** and a random **a** portion of **G** are distributed to each client. The local model of each cilent is trained on the shared data from G together with the provate data frome each client. The cloud then aggregates the local models from the clients to train a global model with **FedAvg**.

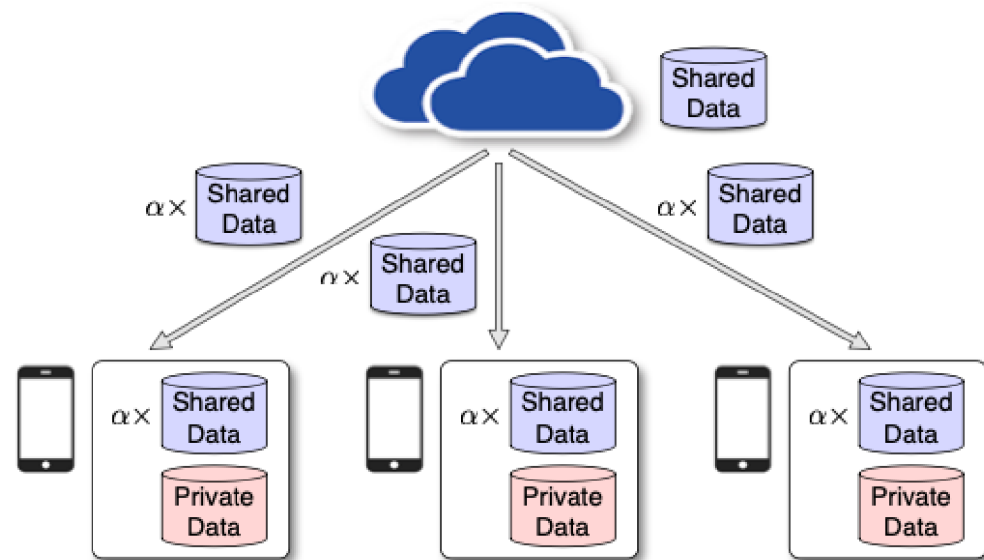


Figure 6: Illustration of the data-sharing strategy.

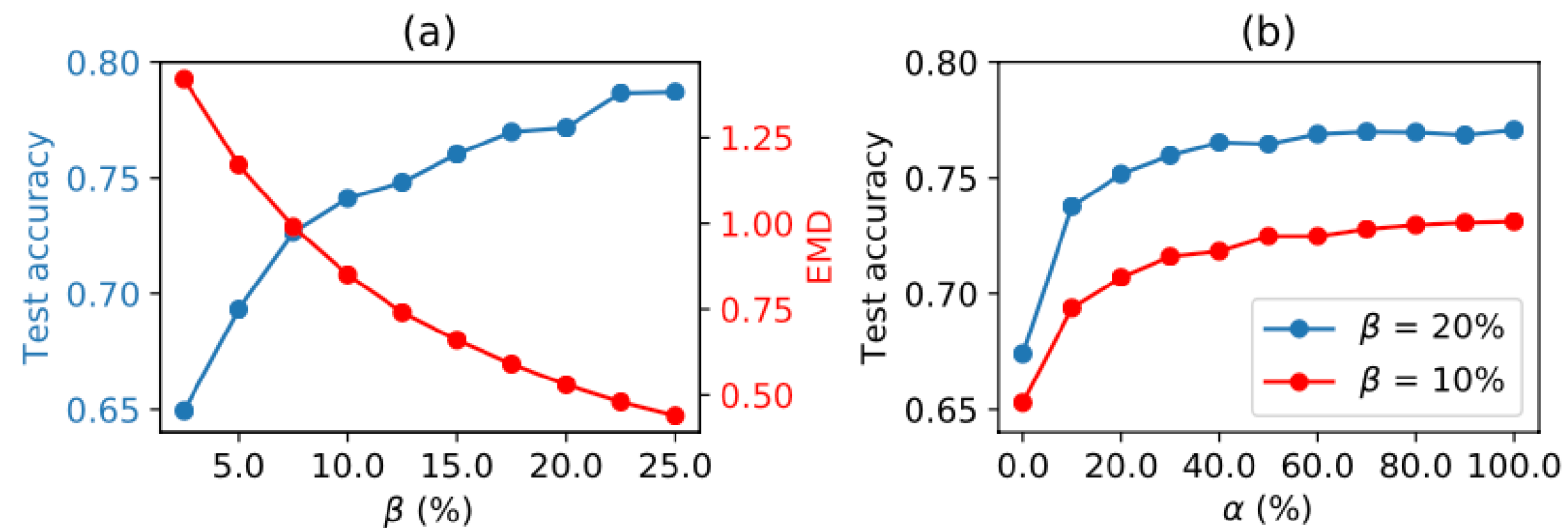


Figure 7: (a) Test accuracy and EMD vs. β (b) Test accuracy vs. the distributed fraction α

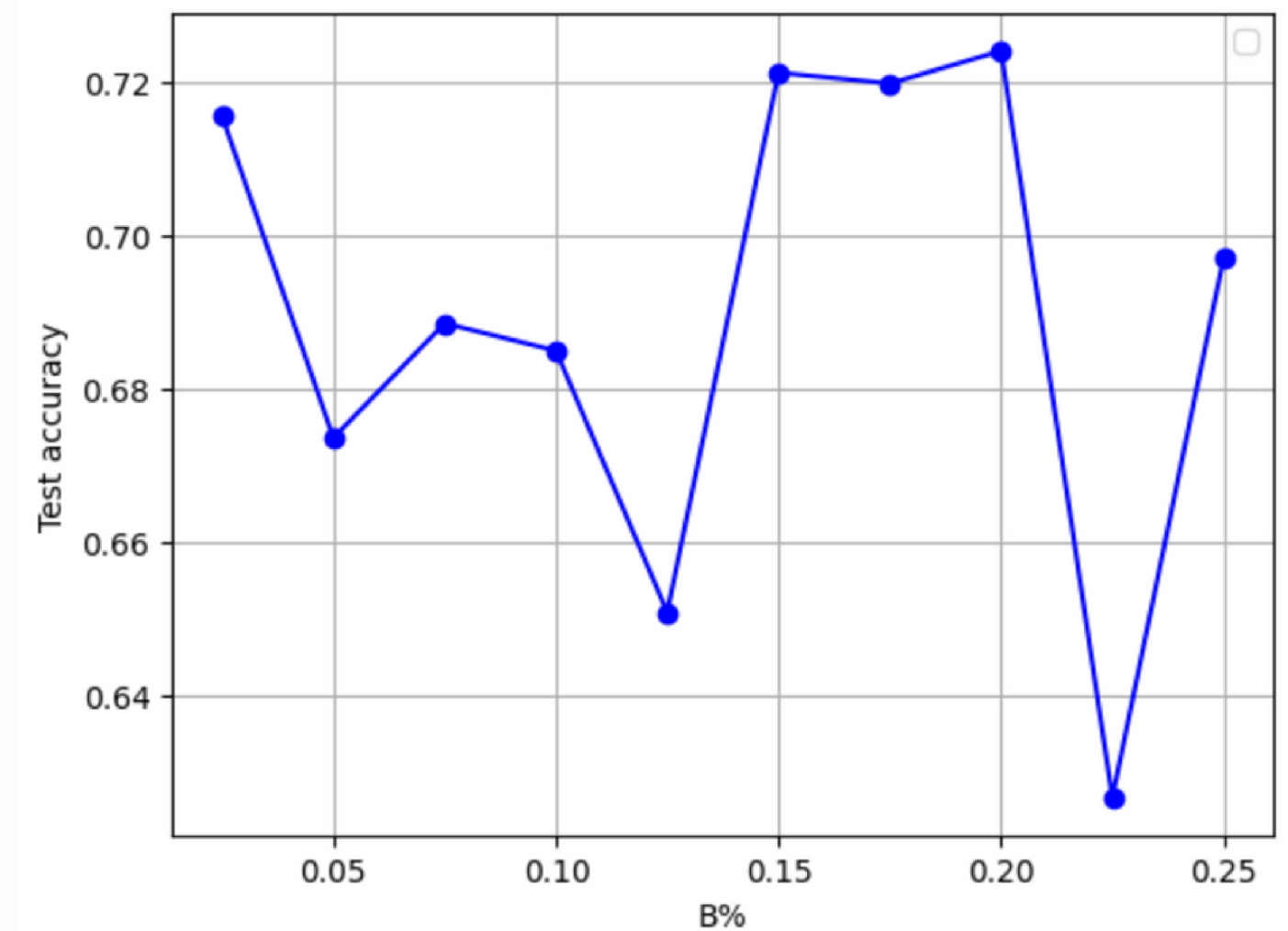
IMPLEMENT ON IOT-BOT

Compare accuracy between IID to nonIID:

score of NonIID dataset : 0.5078236130867709

score of IID dataset : 0.9893314366998578

trade of accuracy and G size:



IMPLEMENT ON IOT-BOT

the red once is plot for B=10% and the blue is 20%:

