Federated Learning mediante consenso

M. Rebollo[1], C. Carrascosa[2], and J.A. Rincón[3]
Inst. Valenciano de Investigación para la IA
Universitat Politècnica de València, Camino de Vera s/n 46022 Valencia
[1]mrebollo@upv.es, [2]carrasco@dsic.upv.es, [3]jrincon@dsic.upv.es

Se conoce por *federated learning* (FL) una técnica de aprendizaje automático por la que un sistema distribuido de nodos entrena una red neuronal (RN) con subconjuntos del conjunto de entrenamiento. Una vez entrenados los modelos individuales, se envían a un nodo central que obtiene el modelo promedio y lo reenvía de vuelta a los nodos. Tras varias rondas, se ha observado que los modelos obtenidos reduciendo el número de iteraciones de entrenamiento tienen una precisión equivalente a haber entrenado una sola red [1].

Hay ocasiones en las que, por el tamaño de los conjuntos de entrenamiento, no es posible centralizar el cálculo. Otros motivos pueden ser restricciones físicas que impiden la conexión con un nodo central. En esos casos es necesario un método realmente distribuido. Este trabajo propone el uso de un protocolo de consenso [2] en el que los nodos de una red llegan al valor promedio intercambiando información con sus vecinos directos.

Supongamos que tenemos n RN idénticas. El objetivo es aprender un modelo global (W, tr) donde W son los pesos entrenados para un conjunto de entrenamiento tr. El conjunto tr se divide en n subconjuntos, uno para cada nodo $tr = \cup_i tr_i$. En cada ronda:

1. cada nodo construye su propio modelo (W_i, tr_i),
2. intercambia la matriz de pesos W_i con sus vecinos hasta que converge usando Eq 1

$$W_i(t+1) = W_i(t) + \varepsilon \sum_{j \in N_i} [W_j(t) - W_i(t)] \qquad (1)$$

Los resultados presentados corresponden con el *dataset* MNIST [3], formado por 60.000 ejemplos de dígitos manuscritos para entrenamiento y 10.000 para test. Se ha utilizado una red convolucional de cinco capas: dos conv2d, una dropout y dos lineales. Se ha estudiado el rendimiento del proceso con distintas topologias y tamaños de red: (i) grid 2d, (ii) grid triangular, (iii) una red navegable de Kleinberg, (iv) un random geometric graph (RGG), (v) una triangulación de Delaunay, y (vi) un grafo de Gabriel (una reducción de la triangulación), y con n variando entre 10 y 600 (Figura 1, superior). Teniendo en cuenta el rendimiento del proceso de consenso en cada tipo de red y otros parámetros, como sus distribuciones de grado o su eficiencia, las RGG resultan la topología más eficiente.

Con ellas se han realizado los experimentos que aparecen en la Figura 1. Se han generado 10 redes de cada tamaño y en cada una de ellas se han realizado 10 repeticiones variando las muestras asignadas a cada nodo tr_i. Se observa que la precisión aumenta cuando disminuye el tamaño de la red. Es un resultado esperable, pues el tamaño del subconjunto de entrenamiento es mayor y, por lo tanto, los pesos quese obtienen están mejor ajustados. Sin embargo, hay que balancearlo con la capacidad de cómputo de cada nodo. La precisión obtenida tras cuatro rondas del FL con consenso es equivalente a la precisión obtenida en una sola RN con diez rondas de ajuste de pesos en el escenario escogido.

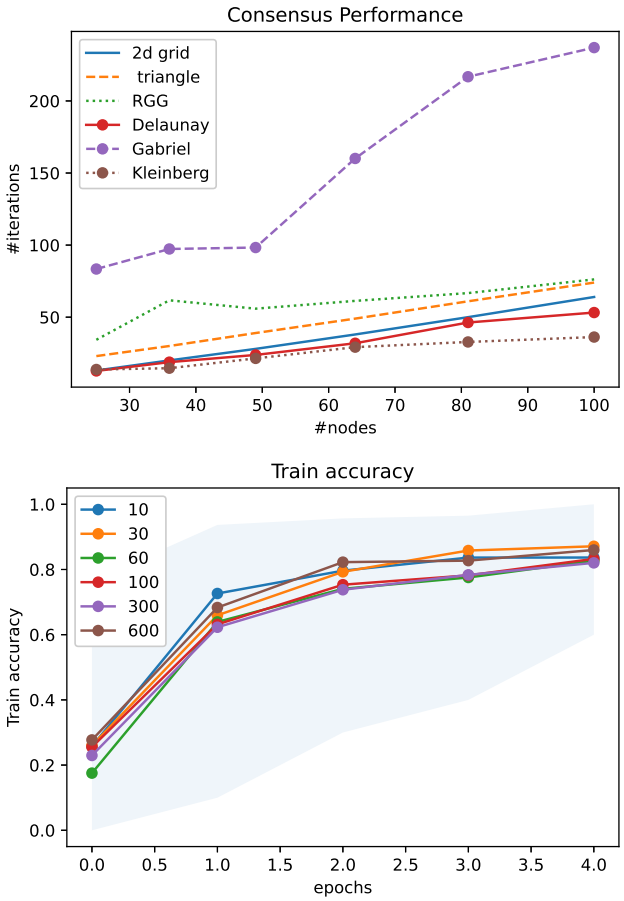

Fig. 1. (arriba) Rendimiento del proceso de consenso según la topología y el tamaño de la red (abajo) Precisión de la red entrenada para el dataset MINST [3] según su tamaño (RGG). El área sombreada señala el rango de mayor a menor precisión obtenidas

Agradecimientos Este trabajo ha sido financiado por el Ministerio de Ciencia e Innovacin bajo el proyecto de cdigo PID2021-123673OB-C31

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, & B.A. Arcas *Communication-Efficient Learning of Deep Networks from Decentralized Data*. In PMLR 54:1273-1282 (2017)

[2] R. Olfati-Saber and R.M.Murray. *Consensus problems in networks of agents with switching topology and time-delays*. IEEE TAC, 49(9):15201533, (2004)

[3] D. Li. *The MNIST database of handwrittendigit images for machine learning research*. IEEE SPM, 29(6):141142, (2012)