Federated Learning

Aprendizaje anónimo

Marta Fuentes

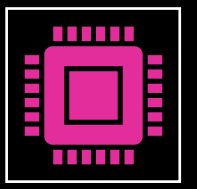
[FIDESOL]
CalT Centro de apoyo a la Innovación Tecnológica

¿Qué es el *Federated Learning*?

Enfoque de aprendizaje automático distribuido

Propuesto por Google en 2016

Intercambia parámetros para preservar la privacidad



Funciona con:

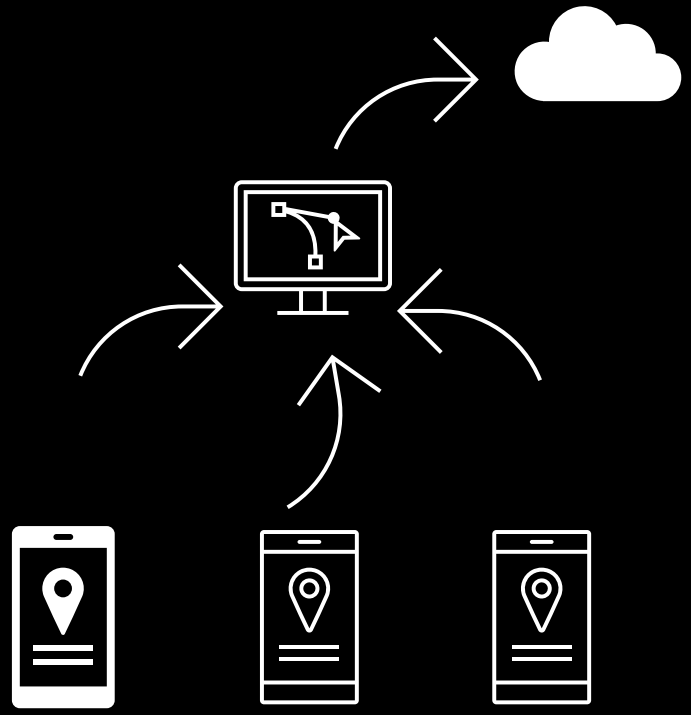
Aprendizaje automático “tradicional”

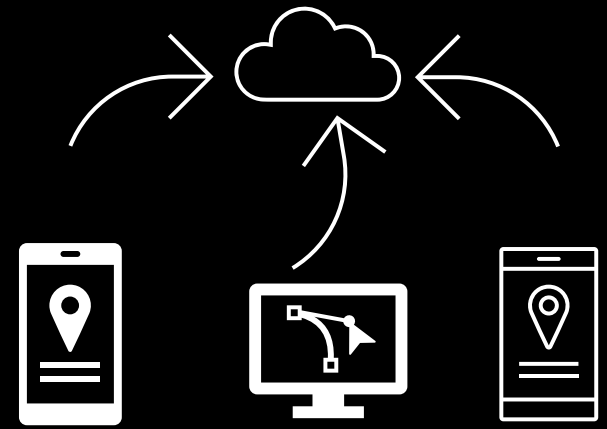
Redes neuronales

Aprendizaje profundo

¿Cómo funciona el *Federated Learning*?

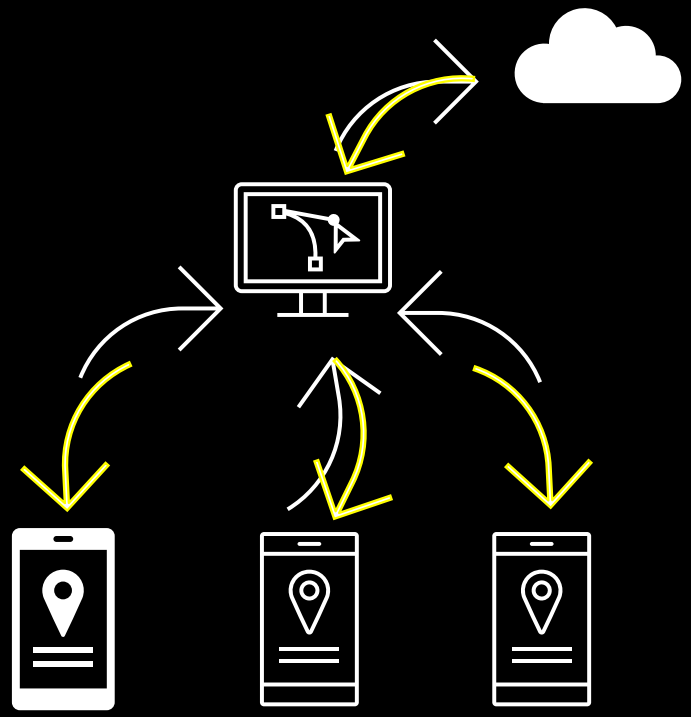
Comparte parámetros, NO datos



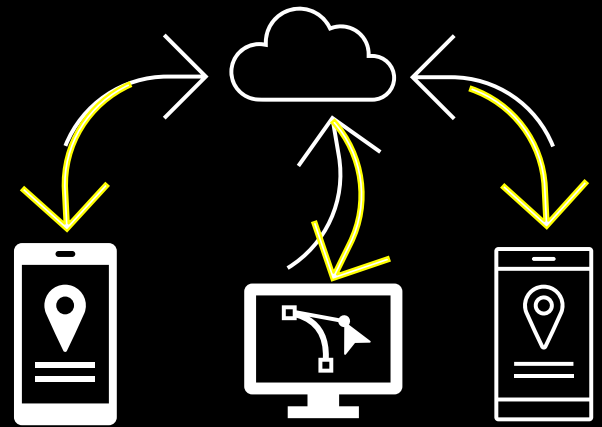


¿Cómo funciona el *Federated Learning*?

Comparte parámetros, NO datos



Distintos enfoques (nube, *Edge*, mixto)
Distinta agregación (centralizada, distribuida, jerárquica)



Un poco de teoría

Características

- Objetivo: minimizar función de pérdida global

$$\min_{\omega} \frac{1}{K} \sum_{n=1}^{N} \sum_{k=1}^{k_n} f_k(\omega)$$

  - K → número de dispositivos
  - w → pesos globales
  - fk → función de pérdida -depende del problema. Ejemplos:
    - Error de predicción ←series temporales
    - Error de clasificación ← clasificación
- Entrenamiento global. Ejemplo: *Stochastic Gradient Descent (SGD)*

$$w_k^{t,\tau} := w_k^{t,\tau-1} - \eta_k \nabla l(w_k^{t,\tau-1}, X_k, Y_k).$$

- Fusión. Ejemplo: *FedAvg*

$$w_{t+1} := \sum_{k \in K} \frac{n_k}{n} w_{k,t}.$$

Consideraciones

👍 Privacidad

👎 Ataques:
- Reconstrucción
- Captura de datos

👍 Comunicación/latencia

👎 Actualización modelo/Convergencia

👍 Disponibilidad clientes

👎 Tipo de clientes → ¿sesgos?

```
(clase_master)                                    master_ciencia_datos>python server.py
INFO flwr 2023-02-21 11:51:16,437 | app.py:142 | Starting Flower server, config: ServerConfig(num_rounds=3, round_tim
eout=None)
INFO flwr 2023-02-21 11:51:16,446 | app.py:156 | Flower ECE: gRPC server running (3 rounds), SSL is disabled
INFO flwr 2023-02-21 11:51:16,446 | server.py:86 | Initializing global parameters
INFO flwr 2023-02-21 11:51:16,447 | server.py:270 | Requesting initial parameters from one random client
INFO flwr 2023-02-21 11:53:56,991 | server.py:274 | Received initial parameters from one random client
INFO flwr 2023-02-21 11:53:56,991 | server.py:88 | Evaluating initial parameters
INFO flwr 2023-02-21 11:53:57,001 | server.py:101 | FL starting
DEBUG flwr 2023-02-21 11:54:22,502 | server.py:220 | fit_round 1: strategy sampled 2 clients (out of 2)
DEBUG flwr 2023-02-21 12:00:20,929 | server.py:234 | fit_round 1 received 2 results and 0 failures
WARNING flwr 2023-02-21 12:00:21,194 | fedavg.py:242 | No fit_metrics_aggregation_fn provided
DEBUG flwr 2023-02-21 12:00:21,196 | server.py:170 | evaluate_round 1: strategy sampled 2 clients (out of 2)
DEBUG flwr 2023-02-21 12:00:35,721 | server.py:184 | evaluate_round 1 received 2 results and 0 failures
WARNING flwr 2023-02-21 12:00:35,721 | fedavg.py:273 | No evaluate_metrics_aggregation_fn provided
DEBUG flwr 2023-02-21 12:00:35,722 | server.py:220 | fit_round 2: strategy sampled 2 clients (out of 2)
DEBUG flwr 2023-02-21 12:05:51,279 | server.py:234 | fit_round 2 received 2 results and 0 failures
DEBUG flwr 2023-02-21 12:05:51,533 | server.py:170 | evaluate_round 2: strategy sampled 2 clients (out of 2)
DEBUG flwr 2023-02-21 12:06:03,729 | server.py:184 | evaluate_round 2 received 2 results and 0 failures
DEBUG flwr 2023-02-21 12:06:03,732 | server.py:220 | fit_round 3: strategy sampled 2 clients (out of 2)
```

- **python server.py**

Símbolo del sistema - python client.py

```
ase_master)                                    master_ciencia_datos>python client.py
3-02-21 11:53:40.148302: I tensorflow/core/platform/cpu_feature_guard.cc:193] This TensorFlow binary is optimized wit
neAPI Deep Neural Network Library (oneDNN) to use the following CPU instructions in performance-critical operations:
X AVX2
enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.
nloading data from https://www.cs.toronto.edu/~kriz/cifar-10-python.tar.gz
498071/170498071 [==============================] - 10s 0us/step
O flwr 2023-02-21 11:53:56,769 | grpc.py:50 | Opened insecure gRPC connection (no certificates were passed)
UG flwr 2023-02-21 11:53:56,769 | connection.py:38 | ChannelConnectivity.IDLE
UG flwr 2023-02-21 11:53:56,772 | connection.py:38 | ChannelConnectivity.CONNECTING
UG flwr 2023-02-21 11:53:56,772 | connection.py:38 | ChannelConnectivity.READY
3/1563 [==============================] - 353s 214ms/step - loss: 1.9296 - accuracy: 0.2974
/313 [==============================] - 14s 39ms/step - loss: 2.3108 - accuracy: 0.1000
3/1563 [==============================] - 314s 201ms/step - loss: 1.7185 - accuracy: 0.3733
/313 [==============================] - 12s 37ms/step - loss: 2.3349 - accuracy: 0.1000
3/1563 [====================>.........] - ETA: 1:30 - loss: 1.5418 - accuracy: 0.4434
```

Símbolo del sistema - python client.py

```
(clase_master)                                    master_ciencia_datos>python client.py
2023-02-21 11:54:20.143007: I tensorflow/core/platform/cpu_feature_guard.cc:193] This TensorFlow binary is optimized wit
h oneAPI Deep Neural Network Library (oneDNN) to use the following CPU instructions in performance-critical operations:
 AVX AVX2
To enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.
INFO flwr 2023-02-21 11:54:22,482 | grpc.py:50 | Opened insecure gRPC connection (no certificates were passed)
DEBUG flwr 2023-02-21 11:54:22,499 | connection.py:38 | ChannelConnectivity.IDLE
DEBUG flwr 2023-02-21 11:54:22,500 | connection.py:38 | ChannelConnectivity.CONNECTING
DEBUG flwr 2023-02-21 11:54:22,502 | connection.py:38 | ChannelConnectivity.READY
1563/1563 [==============================] - 357s 217ms/step - loss: 1.9560 - accuracy: 0.2905
313/313 [==============================] - 14s 39ms/step - loss: 2.3108 - accuracy: 0.1000
1563/1563 [==============================] - 314s 201ms/step - loss: 1.6698 - accuracy: 0.3888
313/313 [==============================] - 12s 37ms/step - loss: 2.3349 - accuracy: 0.1000
1126/1563 [====================>.........] - ETA: 1:29 - loss: 1.5109 - accuracy: 0.4491
```

Un poco de práctica

Referencias de interés

Teoría

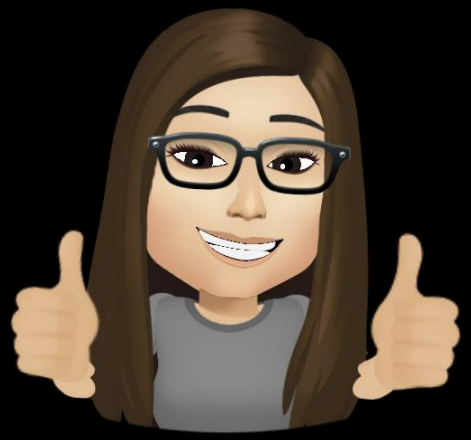
- Martineau. "*What is Federated Learning?*"
- Ludwig y Baracaldo. "*Federated Learning. A Comprehensive Overview of Methods and Applications*". Springer, 2022
- Khan et al. "*Federated Learning for Internet of Things: Recent Advances, Taxonomy, and Open Challenges*". IEEE Access, 2021

Práctica

- "*Quick start TensorFlow*"
- Tijani. "*Federated Learning: A Step by Step Implementation in Tensorflow*"
- Ingerman y Ostrowski "*Introducing TensorFlow Federated*"

Mix

- "Creación de un experimento de aprendizaje federado"
- "¿Cómo aplicar aprendizaje federado y no morir en el intento? Aplicación en fraude de tarjetas de crédito"



Marta Fuentes

[FIDESOL]
CaIT Centro de apoyo a la Innovación Tecnológica