Aprendizaje Federado: aún más retos que soluciones

Senen.barro@usc.es
@SenenBarro

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

AEPIA ASOCIACIÓN ESPAÑOLA PARA LA INTELIGENCIA ARTIFICIAL

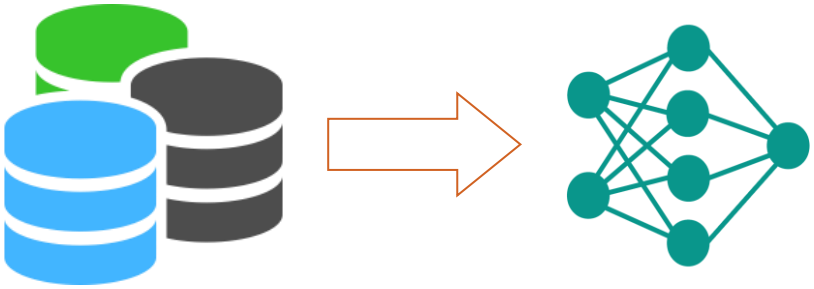
Escuela de Verano de Inteligencia Artificial

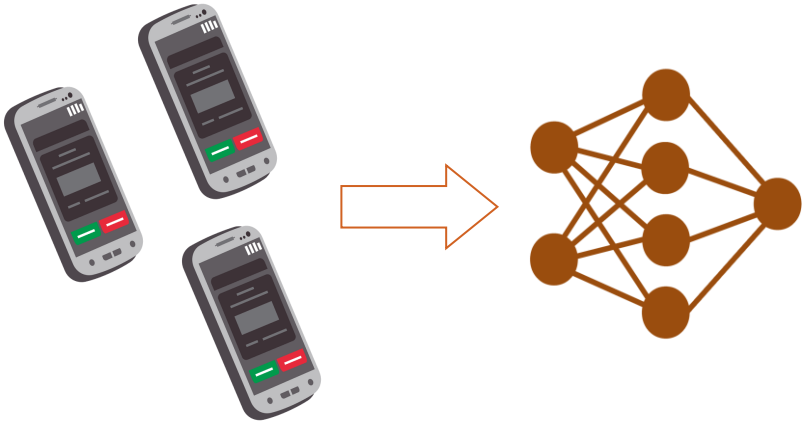
FEDER – Fondo Europeo de Desenvolvemento Rexional
Unha maneira de facer Europa

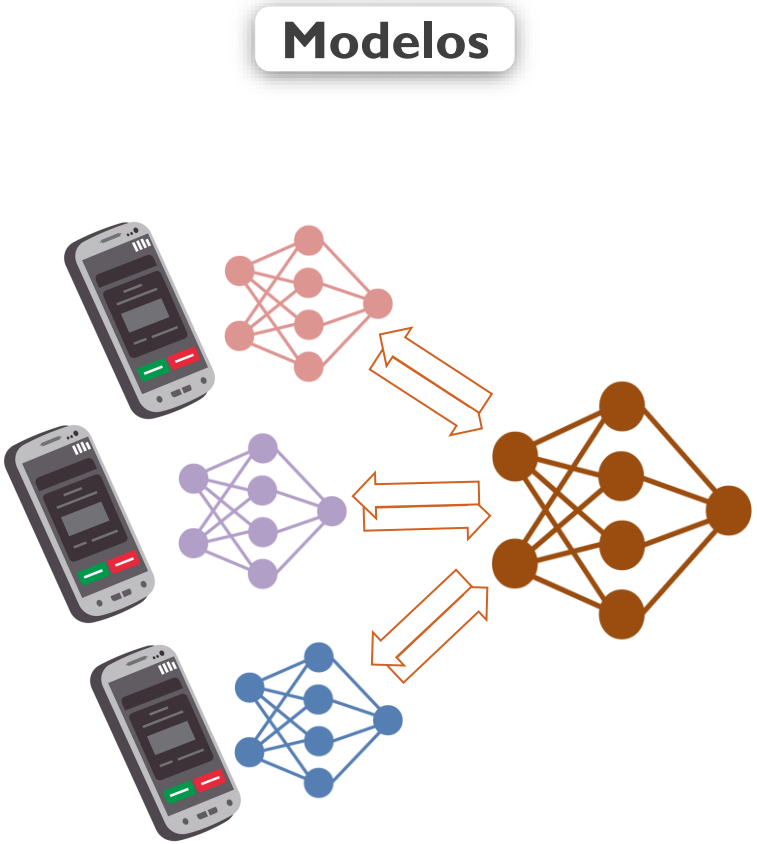
galicia USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA CAMPUS VIDA CAMPUS DE EXCELENCIA INTERNACIONAL UNIÓN EUROPEA XUNTA DE GALICIA

CiTUS

Centro Singular de Investigación en **Tecnoloxías Intelixentes**



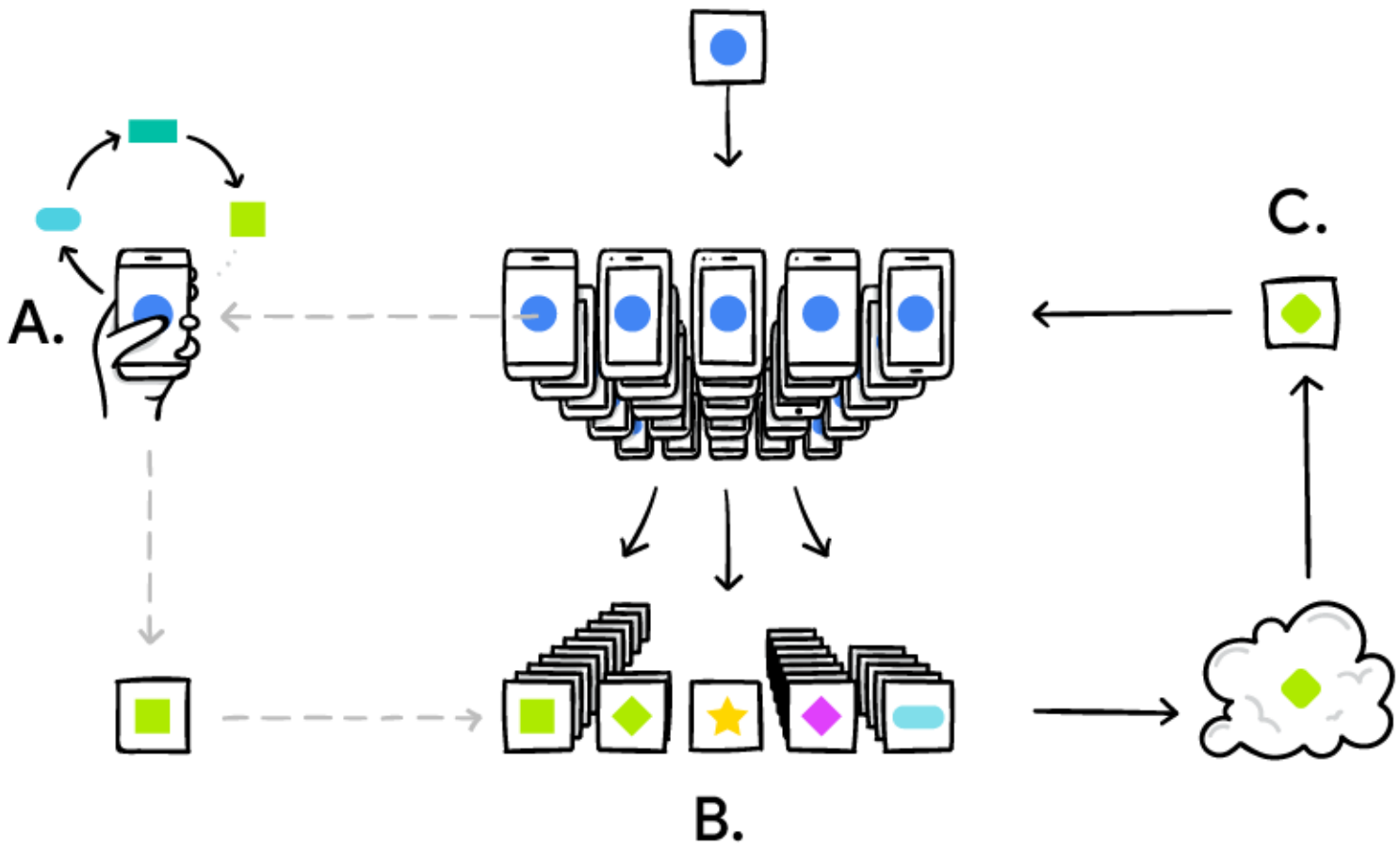




Aprendizaje automático

¿Cómo aprender cuando hay múltiples fuentes de datos?

Aprendizaje Federado: así empezó todo



Google, 2016-2017

https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

Pero no todo comenzó con el Aprendizaje Federado

Título del proyecto: APRENDIZAJE MAQUINA "GLOCAL" Y CONTINUO PARA UNA SOCIEDAD DE DISPOSITIVOS INTELIGENTES (TIN2017-90135-R)

Entidad financiadora: Ministerio de Economía, Industria y Competitividad. Convocatoria: Programa Estatal de I+D+i Orientada a los Retos de la Sociedad

Duración: desde: 01/01/2018 hasta 31/03/2021

Título del proyecto: Aprendizaje federado y continuo a partir de datos heterogéneos en dispositivos y robots. PID2020-119367RB-I00

Entidad financiadora: Ministerio de Ciencia e innovación. Convocatoria: Proyecto de I+D+I de los Programas Estatales de Generación de Conocimiento y Fortalecimiento Científico y Tecnológico del Sistema de I+D+i y de I+D+i Orientada a los Retos de la Sociedad

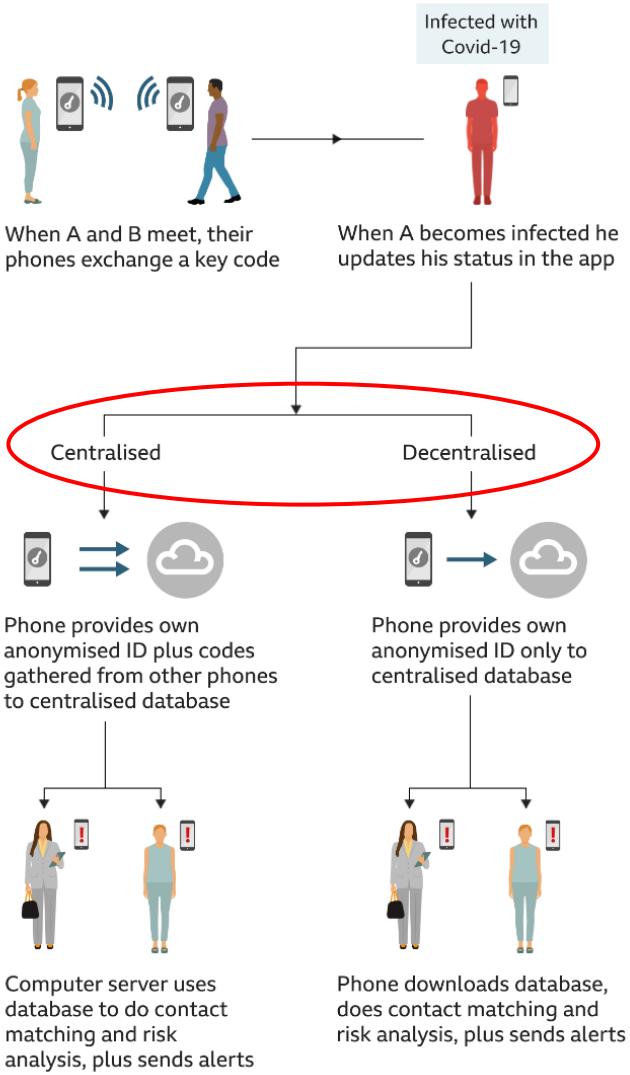
Duración: 3 años. 01/09/2021 - 31/08/2024

Aprendizaje Federado: algunas matizaciones

***Federated Learning** is a distributed machine learning approach which enables training on a large corpus of **decentralized data residing on devices** like mobile phones* [McMahan et al., 2017].

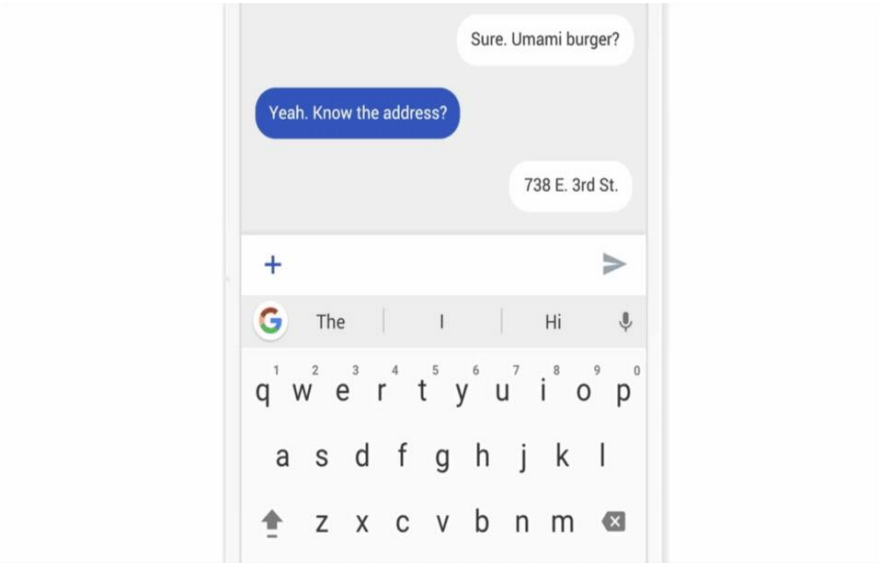
***Federated learning** (also known as collaborative learning) is a machine learning technique that **trains an algorithm via multiple independent sessions,** each using its own dataset* [Wikipedia].

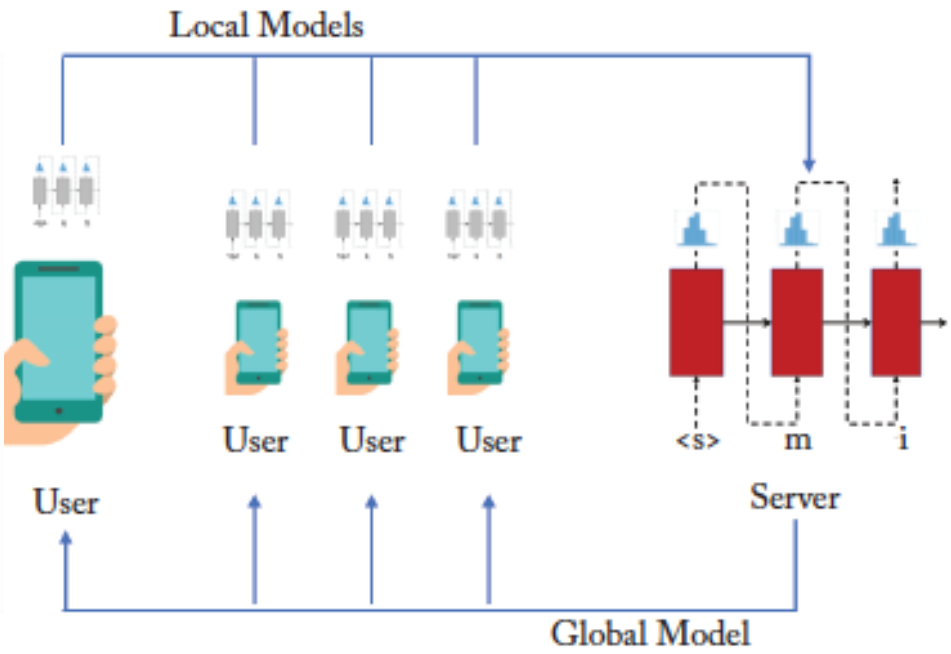
Rastreo de contactos centralizado



Fuente: "Coronavirus contact-tracing: World split between two types of app", BBC News, 7 mayo 2020, Cristina Criddle y Leo Kelion

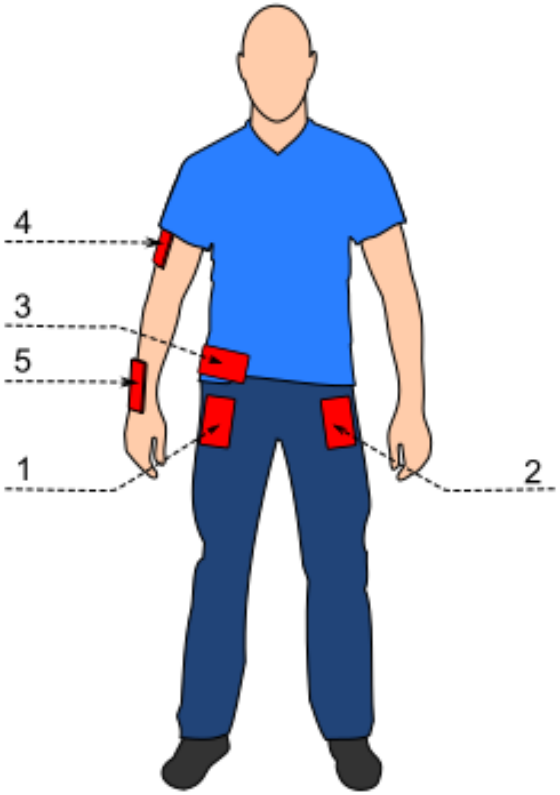
Aprendizaje Federado: un ejemplo





Aprendizaje federado para escritura predictiva (De: Yang *et al.* [2019])

Reconocimiento de la actividad mediante móvil



MATT SIMON SCIENCE 09.07.2020 07:00 AM

Your Smartphone Can Tell If You're Drunk-Walking

In a recent study, scientists could tell if people were intoxicated just by looking at their phones' motion data.

Conducción

Un ejemplo real y de divulgación

Búsquedas de:
“federated learning”

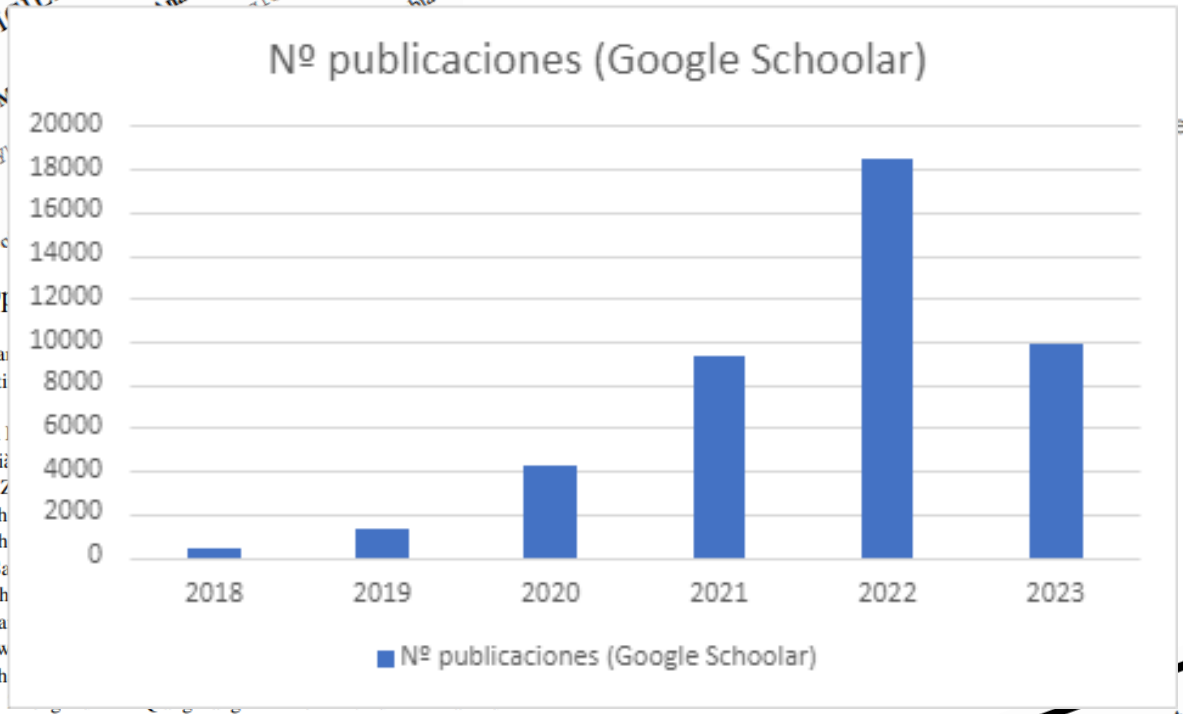
Google Trends, junio 2023

FEDERATED LEARNING: STRATEGIES FOR IMPROVING COMMUNICATION EFFICIENCY

Jakub Konečný, H. Brendan McMahan, Felix X. Yu, Peter Richtárik, Ananda Theertha Suresh & Dave Bacon

FEDERATED LEARNING FOR MOBILE KEYBOARD PREDICTION

Andrew Hard, Kanishka Rao, Rajiv Mathews, Françoise Beaufays
Sean Augenstein, Hubert Eichner, Chloé Kiddon, Daniel Ramage



Advances and Op

Peter Kairouz[7*] H. Brendan
Mehdi Bennis[19] Arjun Nitin
Graham Cormode[23]
Hubert Eichner[7] Salim
Zachary Garrett[7] Adria
Marco Gruteser[7,14]
Zhouyuan Huo [20] Ben Hutch
Gauri Joshi[2] Mikhail Kh
Farinaz Koushanfar[17]
Prateek Mittal[13] Meh
Rasmus Pagh[7,10] Ha
Mariana Raykova[7] Daw
Ziteng Sun[3] Ananda Theerth
Jianyu Wang[2] Li Xiong[5]
Sen Zhao[7]

[1]Australian National University, [2]Carnegie Mellon University, [3]Cornell University,
[4]École Polytechnique Fédérale de Lausanne, [5]Emory University, [6]Georgia Institute of Technology,
[7]Google Research, [8]Hong Kong University of Science and Technology, [9]INRIA, [10]IT University of Copenhagen,
[11]Massachusetts Institute of Technology, [12]Nanyang Technological University, [13]Princeton University,
[14]Rutgers University, [15]Stanford University, [16]University of California Berkeley,
[17] University of California San Diego, [18]University of Illinois Urbana-Champaign, [19]University of Oulu,
[20]University of Pittsburgh, [21]University of Southern California, [22]University of Virginia,
[23]University of Warwick, [24]University of Washington, [25]University of Wisconsin–Madison

saugenst
e.com

Communication-Efficient Learning of Deep Networks from Decentralized Data

H. Brendan McMahan Eider Moore Daniel Ramage Seth Hampson Blaise
Google, Inc., 651 N 34th St., Seattle, WA 98103 USA

Sociedad de dispositivos

Más de 30.000 millones de dispositivos conectados a Internet en el mundo

Aprendizaje y trabajo humano **cooperativo**

¿Qué nos dice la neurociencia?

Aprendizaje Automático

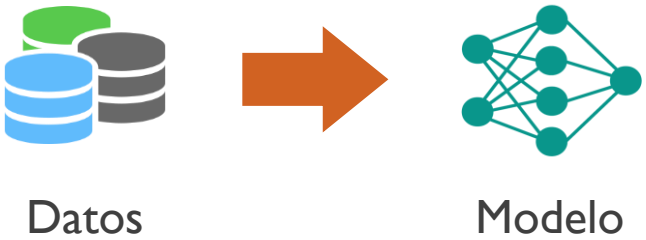
Aprendizaje Automático *es el estudio de algoritmos que mejoran automáticamente a través de la experiencia.*

"Un sistema aprende de la **experiencia** *X* con respecto a una **tarea** *T* y con una medida de **rendimiento** *P,* si su rendimiento para la tarea *T*, medido mediante *P*, mejora con la experiencia *X*"

Mitchell, T. (1997). Machine Learning. McGraw Hill. p. 2. ISBN 978-0-07-042807-2.

Aprendizaje local o individual

Consideramos que una **entidad** *E*, que opera en un **mundo** *W*, en el que realiza una **tarea** *T*, a través de un **modelo** M_L, aprendido localmente a partir de su **experiencia** *X*, con una **medida de rendimiento** P: $A_L(X) \rightarrow M_L$, con $P \geq \theta$



El **modelo** para realizar *T* puede ser **dinámico**, $M_L(t)$, buscando mejorar el rendimiento, $P_L(t)$, a través de **nuevas** experiencias, $X(t)$:

$$A_L(M_L(t), X(t)) \rightarrow M_L(t+\Delta), \text{ con } P_L(t+\Delta) \geq P_L(t)$$

Aprendizaje cooperativo o colectivo

Dado un conjunto de entidades $ES = \{E_1, E_2 \ldots E_S\}$ que resuelven la tarea T mediante modelos propios $MS = \{M_1, M_2 \ldots M_S\}$, obtenidos a partir de sus experiencias particulares $XS = \{X_1, X_2 \ldots X_S\}$, nos planteamos obtener mediante una estrategia de **aprendizaje cooperativo** un **modelo global** MGc, que tenga un rendimiento igual o mejor que cualquier modelo local:

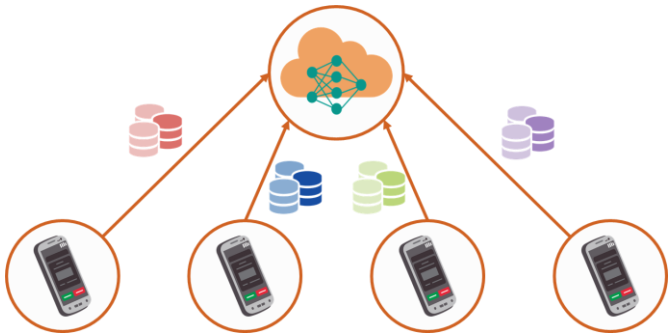
$$Ac(\{(M_s(t), X_s(t)), s=1\ldots S\}) \rightarrow MGc(t+\Delta),$$

tal que $PGc(t) \geq P_s(t)$, $\forall$ s=1…S, si t es suficientemente grande.

Aprendizaje Cooperativo Centralizado

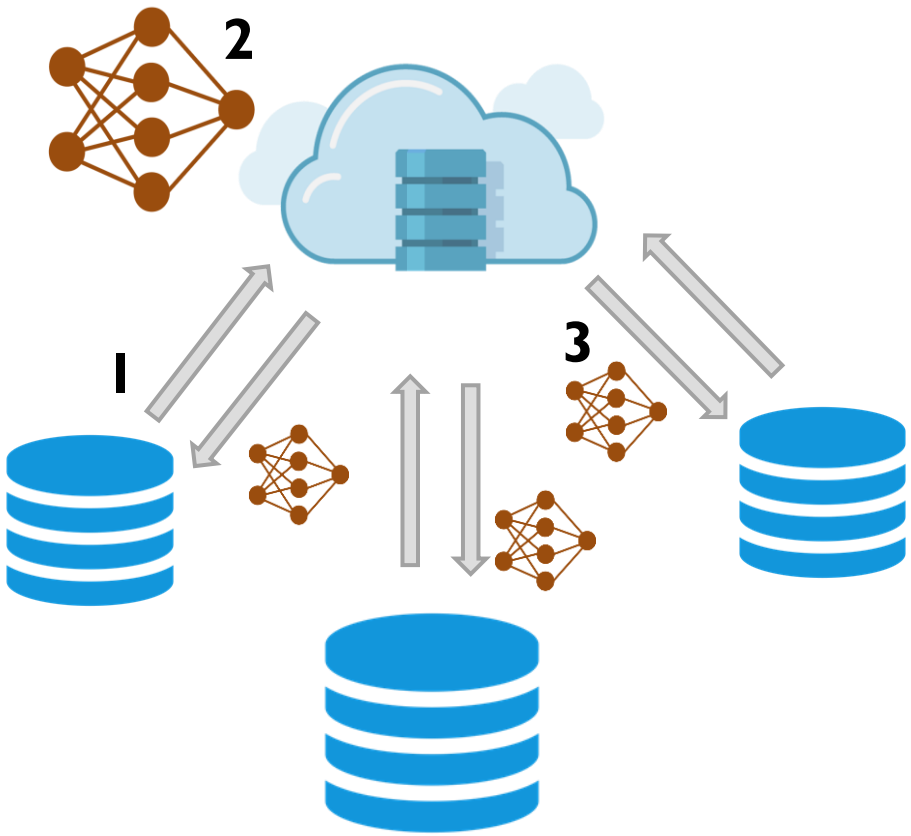
En el **aprendizaje cooperativo centralizado** un conjunto de entidades, $ES = \{E_1, E_2 \ldots E_S\}$, comparte sus experiencias $XS = \{X_1, X_2 \ldots X_S\}$, para calcular directamente el modelo global MG_{CC}, de resolución de la tarea T, a partir de las mismas:

$$Acc(\{X_s(t), s=1\ldots S\}) \rightarrow MGcc(t+\Delta)$$



tal que $PGcc(t) \geq P_s(t), \forall\ s=1\ldots S$.

Aprendizaje Cooperativo Centralizado



Aprendizaje cooperative centralizado, mediante una arquitectura cliente-servidor: 1) Envío de datos [que pueden estar cifrados]; 2) Aprendizaje; 3) Envío de la solución aprendida a los clientes.

Aprendizaje Federado

En el **aprendizaje federado** un conjunto de entidades, $ES = \{E_1, E_2 \ldots E_S\}$, comparte sus **modelos** $MS = \{M_1, M_2 \ldots M_S\}$, de resolución de la tarea T, con el objetivo de obtener un modelo global, MG_F para dicha tarea, buscando que su rendimiento sea igual o mejor que el de cualquier modelo local:

$$A_F(\{M_s(t), s=1\ldots S\}) \rightarrow MG_F(t+\Delta)$$

$$\text{tal que } PG_F(t) \geq P_s(t), \forall\ s=1\ldots S,$$

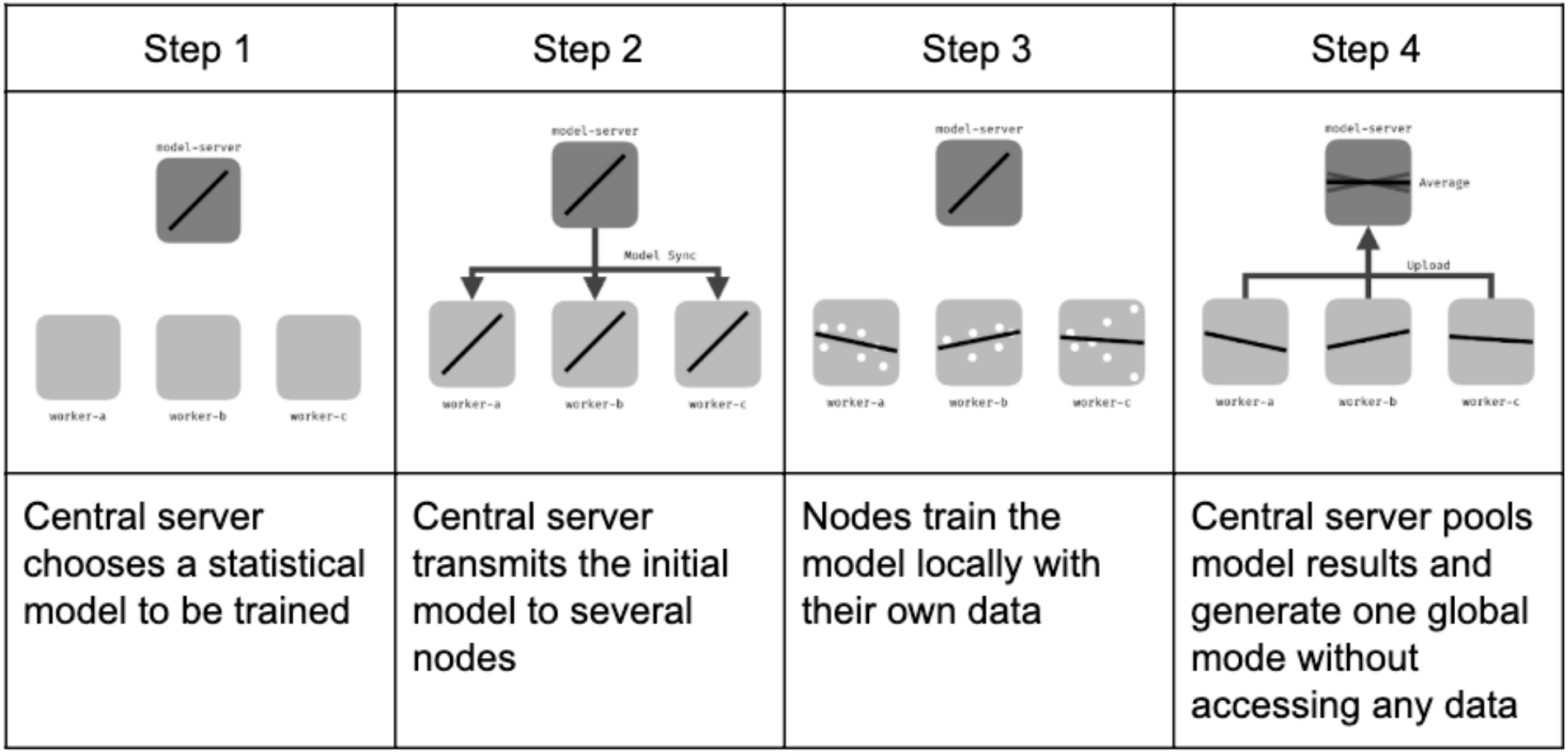
y que, además, se aproxime suficientemente al rendimiento de un modelo centralizado:

$$|PG_{CC}(t) - PG_F(t)| < \varepsilon, \text{ con } \varepsilon \text{ muy pequeño, si t es suficientemente grande.}$$

Aprendizaje Federado: arquitectura cliente-servidor

1) Aprendizaje de modelos locales; 2) Envío al servidor de modelos locales; 3) Obtención de un modelo mejorado; 4) Envío a los clientes del modelo mejorado; 5) Uso local del nuevo modelo (y, en su caso, repetir el proceso)

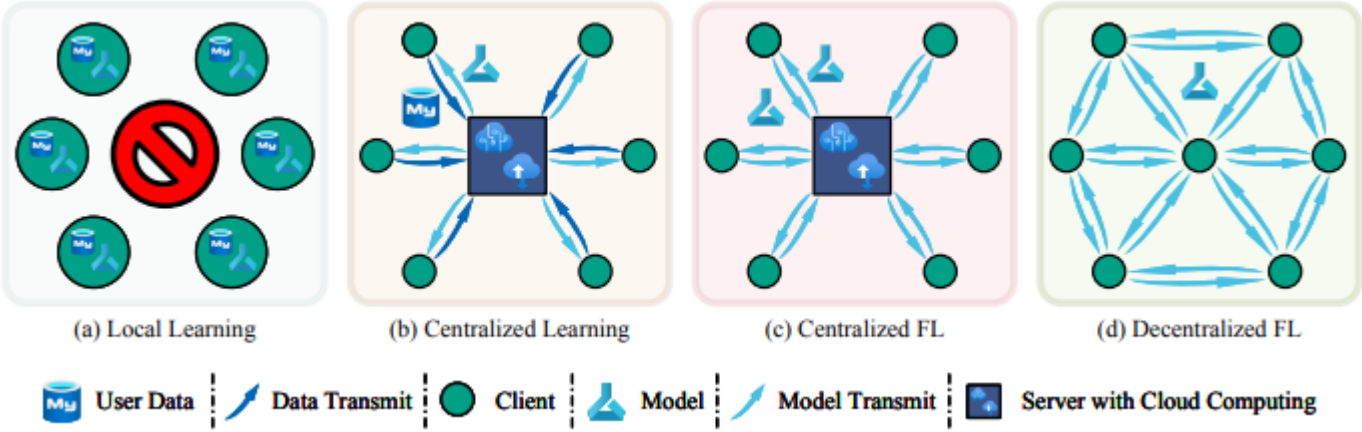
Aprendizaje Federado: combinación de modelos locales



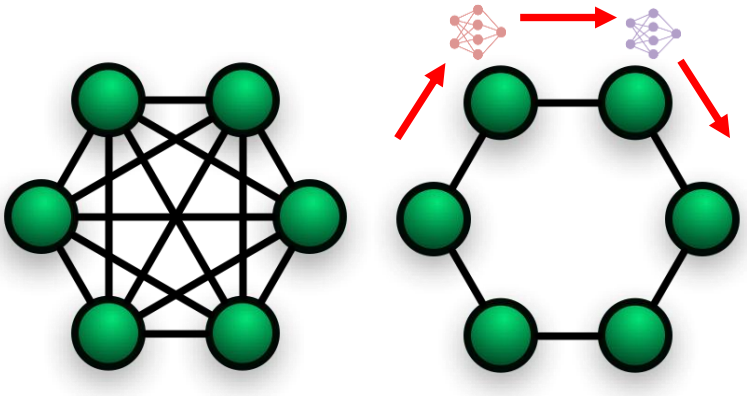
Ejemplo de composición de soluciones locales

Wikipedia (mayo 2023)

Aprendizaje sin servidor (*peer to peer*)

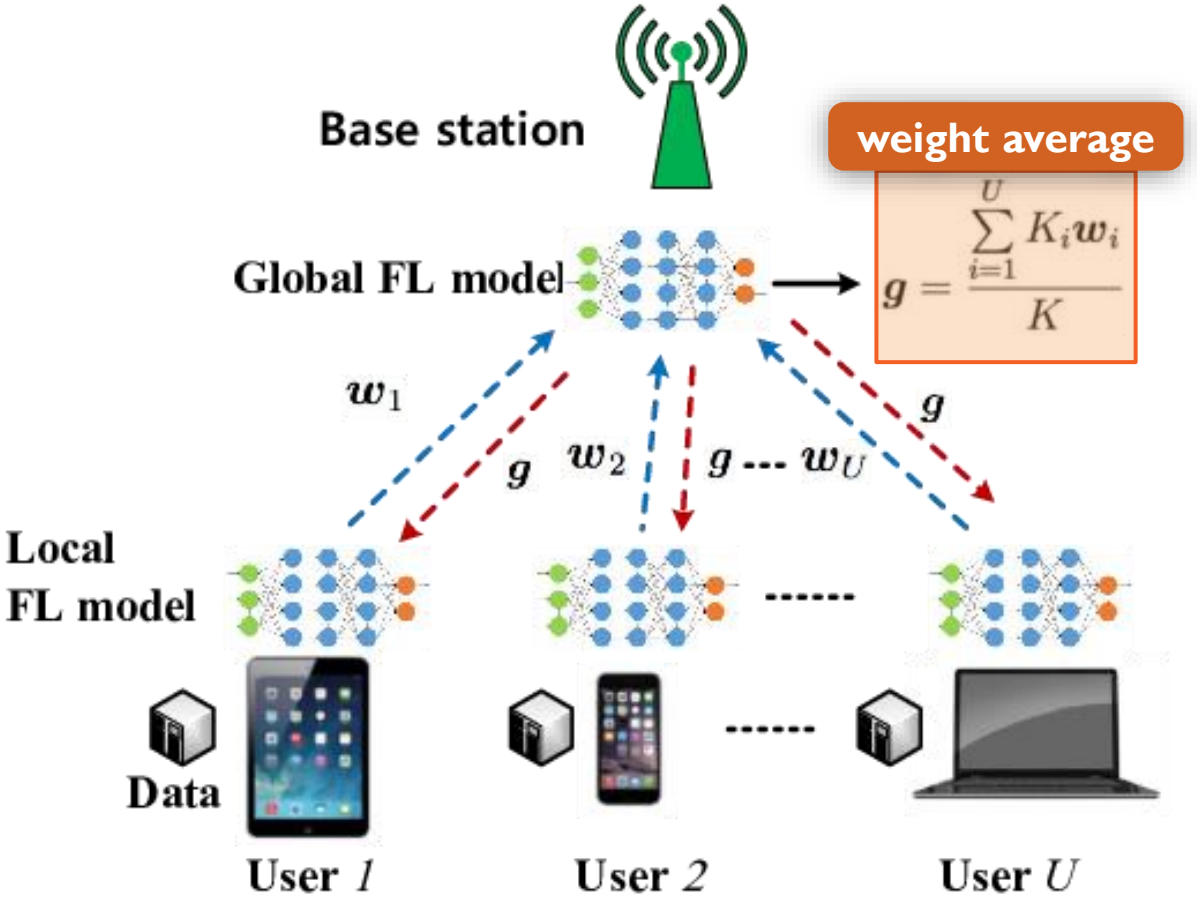


Decentralized Federated Learning: A Survey and Perspective
L Yuan, L Sun, PS Yu, Z Wang - arXiv preprint arXiv:2306.01603, 2023



Arquitectura totalmente conectada (izda.) y cíclica

Aprendizaje Federado: combinación de modelos



McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B.A. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Artificial Intelligence and Statistics* (pp. 1273-1282). PMLR.

Aprendizaje Federado: FedAvg

Algorithm 1 `FederatedAveraging`. The K clients are indexed by k; B is the local minibatch size, E is the number of local epochs, and η is the learning rate.

Server executes:
  initialize w_0
  for each round $t = 1, 2, \ldots$ **do**
    $m \leftarrow \max(C \cdot K, 1)$
    $S_t \leftarrow$ (random set of m clients)
    for each client $k \in S_t$ **in parallel do**
      $w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$
    $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$

ClientUpdate(k, w)**:** *// Run on client k*
  $\mathcal{B} \leftarrow$ (split $\mathcal{P}_k$ into batches of size B)
  for each local epoch i from 1 to E **do**
    for batch $b \in \mathcal{B}$ **do**
      $w \leftarrow w - \eta \nabla \ell(w; b)$
  return w to server

Fuente: Keith Bonawitz, Hubert Eichner et al, “Towards Federated Learning at Scale: System Design”, Proceedings of the 20 th International Conference on Artificial Intelligence and Statistics (AISTATS) 2017. JMLR: W&CP volume 54.

Aprendizaje Federado: ventajas (potenciales)

Aprendizaje más **rápido** y **mejor**, cuando no es posible aprender centralizadamente

Privacidad de datos por diseño

Reducción de sesgos y mayor diversidad de **datos**

Incremento de la **seguridad** (de modo general)

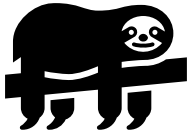
Menor tensión en las **comunicaciones**

Escalabilidad (almacenamiento, comunicaciones, velocidad…)

Aprendizaje Federado: desventajas (potenciales)

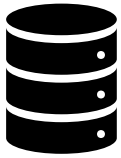
Visión parcial del mundo (espacio problema)

Posibles **problemas de convergencia** en el aprendizaje y/o en el rendimiento del mismo



Lentitud en las **comunicaciones** y posible inestabilidad de las mismas

Dificultad para garantizar la **fiabilidad** de los datos



Posibles problemas de **seguridad** por comportamiento anómalos de clientes

¿El Aprendizaje Federado puede con todo?

Escalabilidad

- Almacenamiento
- Comunicación
- Computación

Razonablemente resueltos con FL

Privacidad:

- Datos personales, como nombre, teléfono, dirección. imágenes, geolocalización, dirección IP…

Adaptabilidad:

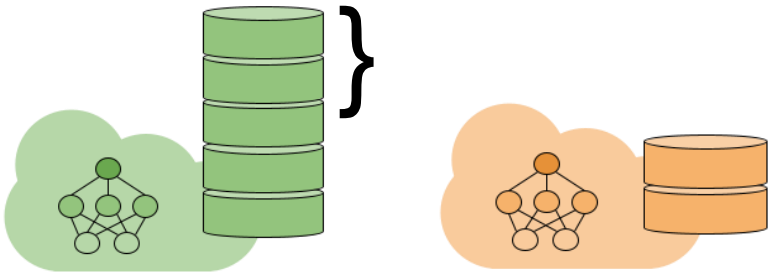
- Diferentes usuarios
- Diferentes dispositivos
- Contextos espacio-temporales

Suele asumirse que los datos son **IID**

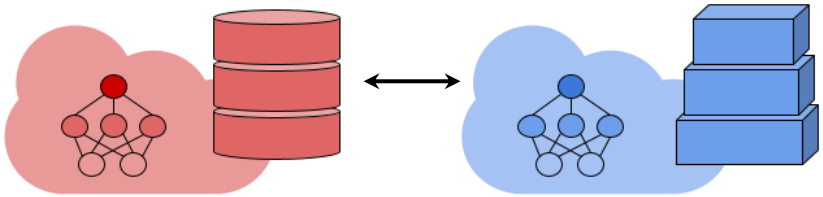
Datos heterogéneos o no-IID

Nos complican mucho la vida

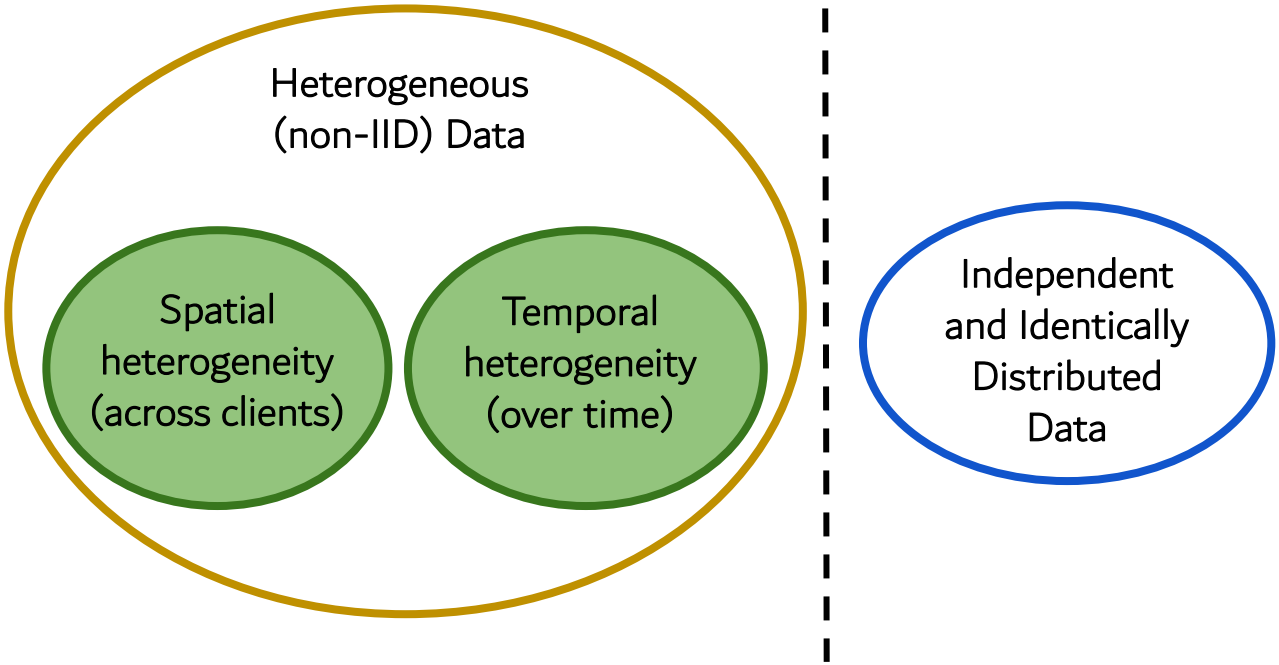
Desequilibrio de datos entre clientes



Datos dispares entre clientes



Datos heterogéneos o no-IID

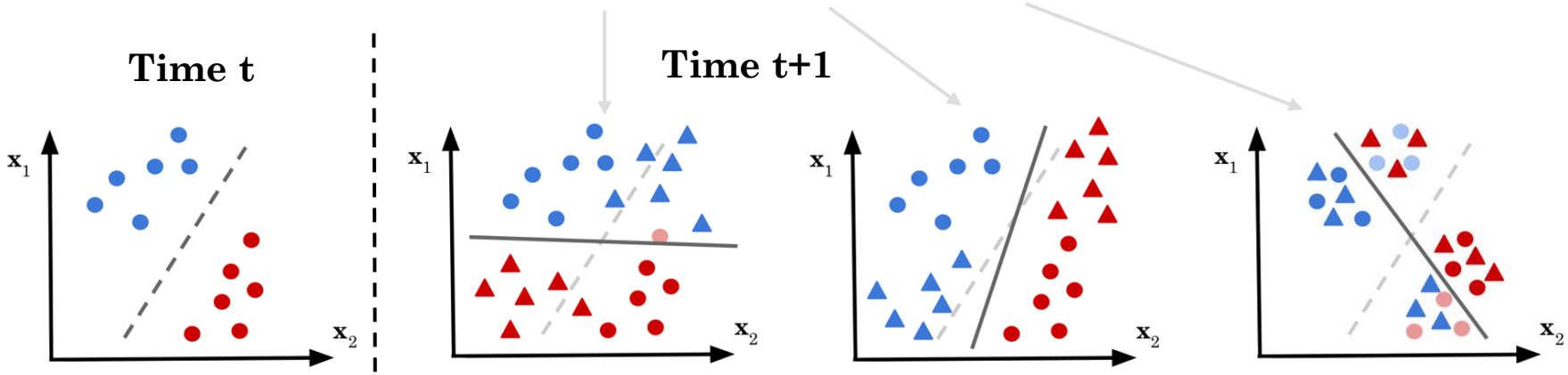


Heterogeneidad espacial: deriva de concepto

La función de densidad de probabilidad conjunta estimada puede descomponerse así:

$$P(\mathbf{x}, \mathbf{y}) = P(\mathbf{x}) \cdot P(\mathbf{y}|\mathbf{x})$$

De modo que:



Datos en tiempo t

Deriva total

$P^t(\mathbf{x}) \neq P^{t+1}(\mathbf{x})$
$P^t(\mathbf{y}\,|\,\mathbf{x}) \neq P^{t+1}(\mathbf{y}\,|\,\mathbf{x})$

Deriva virtual

$P^t(\mathbf{x}) \neq P^{t+1}(\mathbf{x})$
$P^t(\mathbf{y}\,|\,\mathbf{x}) = P^{t+1}(\mathbf{y}\,|\,\mathbf{x})$

Deriva real

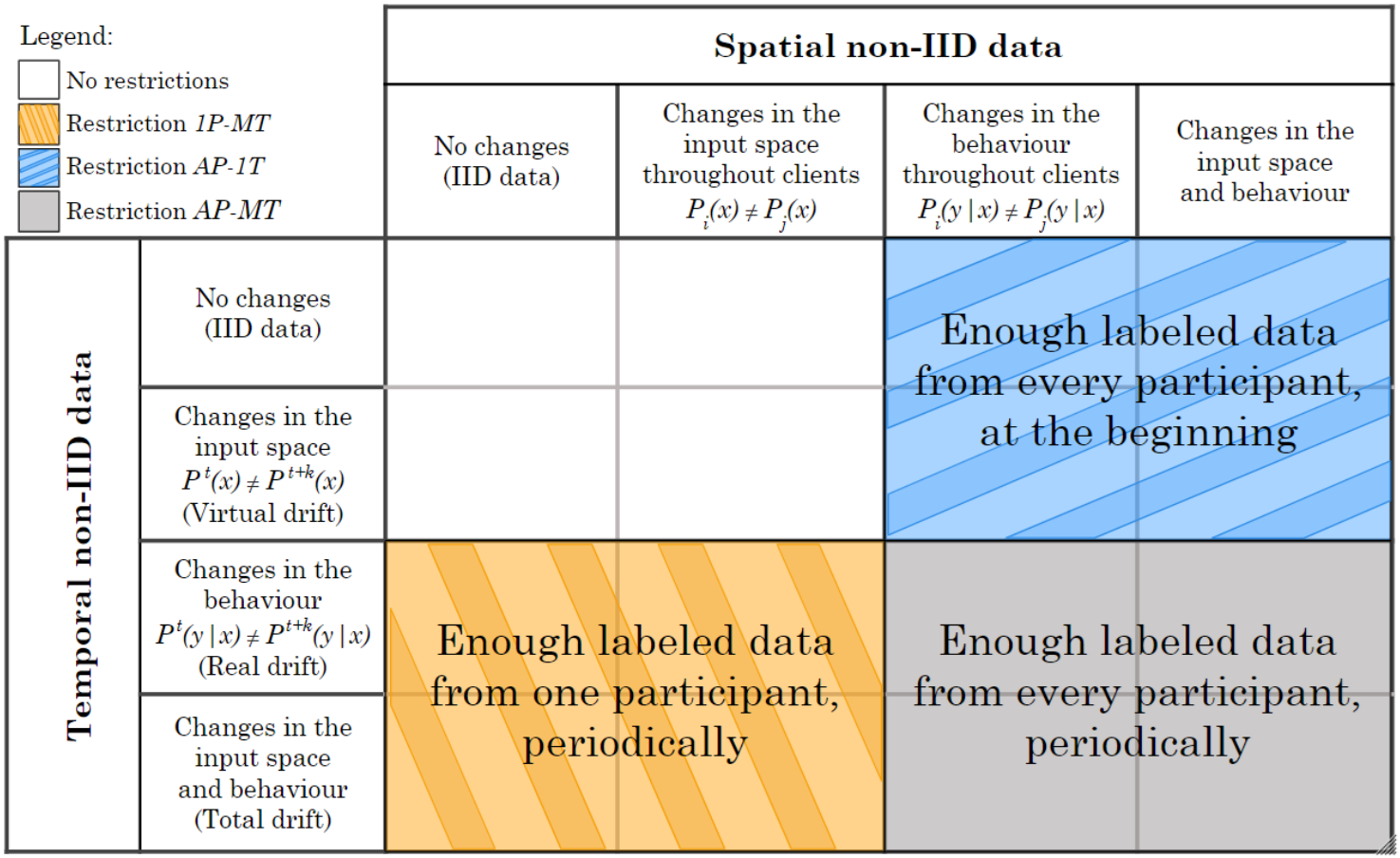
$P^t(\mathbf{x}) = P^{t+1}(\mathbf{x})$
$P^t(\mathbf{y}\,|\,\mathbf{x}) \neq P^{t+1}(\mathbf{y}\,|\,\mathbf{x})$

Manejar la heterogeneidad

		Spatial heterogeneity			
		No changes (IID data)	Changes in the input space throughout clients $P_i(x) \neq P_j(x)$	Changes in the behaviour throughout clients $P_i(y \mid x) \neq P_j(y \mid x)$	Changes in the input space and behaviour throughout clients
Temporal heterogeneity	No changes (IID data)				
	Changes in the input space over time, $P^t(x) \neq P^{t+k}(x)$ (Virtual Concept Drift)		NOT		
	Changes in the behaviour over time, $P^t(y \mid x) \neq P^{t+k}(y \mid x)$ (Real Concept Drift)		ADDRESSED		
	Changes in the input space and behaviour over time (Total Concept Drift)		SO FAR		

Marcos F. Criado, Fernando E. Casado, Roberto Iglesias, Carlos V. Regueiro, Senén Barro, Non-IID data and Continual Learning processes in Federated Learning: A long road ahead, Information Fusion, Volume 88, 2022, Pages 263-280,

Manejar la heterogeneidad



Marcos F. Criado, Fernando E. Casado, Roberto Iglesias, Carlos V. Regueiro, Senén Barro, Non-IID data and Continual Learning processes in Federated Learning: A long road ahead, Information Fusion, Volume 88, 2022, Pages 263-280,

Deriva de concepto y aprendizaje continuo

Arquitectura CDA-FedAvg

- Capaz de detectar y adaptarse a la deriva de concepto
  - Detección:
    - Algoritmo basado en cambios en la distribución de los datos mediante la comparación de datos actuales y pasados
  - Adaptación:
    - Reentreno del modelo combinando datos actuales y pasados (memoria a largo plazo), para evitar el olvido catastrófico

- Responde a dos cuestiones fundamentales:
  - **Qué aprender**
  - **Cuándo aprender**

Casado, F.E., Lema, D., Criado, M.F., Iglesias, R., Regueiro, C.V., and Barro, S. Concept drift detection and adaptation for federated and continual learning. *Multimed Tools Appl* **81**, 3397–3419 (2022).

Arquitectura CDA-FedAvg: desde el servidor

Algorithm 3 Concept-drift-aware federated averaging, server side.

Input : List of participant clients $\mathcal{C} = \{1, 2, \ldots, C\}$
Output: Global model $\mathbf{w}^G$.

Initialize $\mathbf{w}_0^G$
broadcast($\mathbf{w}_0^G, \mathcal{C}$); // Send the model to all clients
while *true** **do**
  Listen for client updates $\forall j \in \mathcal{C}$
  if $\exists j \in \mathcal{C}$: *new update* $\mathbf{w}_t^j$ *is received* **then**
    $\mathbf{w}_t^G \leftarrow \sum_{j=1}^{C} \frac{n_t^j}{N} \mathbf{w}_t^j$; // Averaging aggregation
    broadcast($\mathbf{w}_t^G, \mathcal{C}$)
  end
end

* Given the continual nature, we leave the choice of the stop criteria as a matter of implementation.

Casado, F.E., Lema, D., Criado, M.F., Iglesias, R., Regueiro, C.V., and Barro, S. Concept drift detection and adaptation for federated and continual learning. *Multimed Tools Appl* **81**, 3397–3419 (2022).

Arquitectura CDA-FedAvg: desde los clientes

Algorithm 4 Concept-drift-aware federated averaging, client side.

Input : Minimum amount of data to train L, local minibatch size B, number of rounds R per change, number of local epochs E per round, learning rate η, sensitivity to change λ, padding Δ, and maximum size N_{max} for the sliding window.

Output: None.

```
Q ← {∅};          // Initialize the sliding window (short-term memory)
L ← {∅};                    // Initialize the long-term memory
L ←driftAdaptation(L, L, B, R, E, η);     // Learn first concept (Algorithm 6)
while true* do
    if new data instance, X_i, is observed then
        [ŷ_i, q_i] ← predict(w_t^G, X_i);          // Classify the pattern
        Q ← Q ∪ q_i;                    // Add the confidence into Q
        if |Q| >= N_max then
            Q ← Q \ {q_1};        // Remove the oldest element in Q
        end
        r ← random(0, 1);        // Generate random number in the interval [0,1]
        if e^(−2q_i) ≥ r then
            d ← driftDetection(Q, λ, Δ, N_max);  // Check for drift (Algorithm 5)
            if d is true then
                Q ← {∅}
                L ←driftAdaptation(L, L, B, R, E, η);        // Update (Algorithm 6)
            end
        end
    end
end
```

* Given the continual nature, we leave the choice of the stop criteria as a matter of implementation.

Casado, F.E., Lema, D., Criado, M.F., Iglesias, R., Regueiro, C.V., and Barro, S. Concept drift detection and adaptation for federated and continual learning. *Multimed Tools Appl* **81**, 3397–3419 (2022).

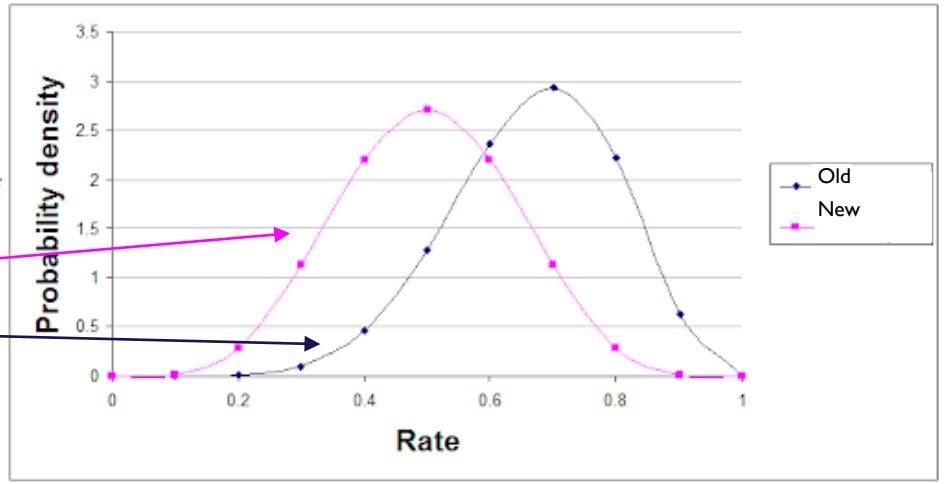
Arquitectura CDA-FedAvg: detección de deriva

Algorithm 5: Drift detection method (`driftDetection`).

Input : Sliding window $\mathcal{Q}$, sensitivity to change λ, and padding Δ and maximum size N_{max} for the sliding window.

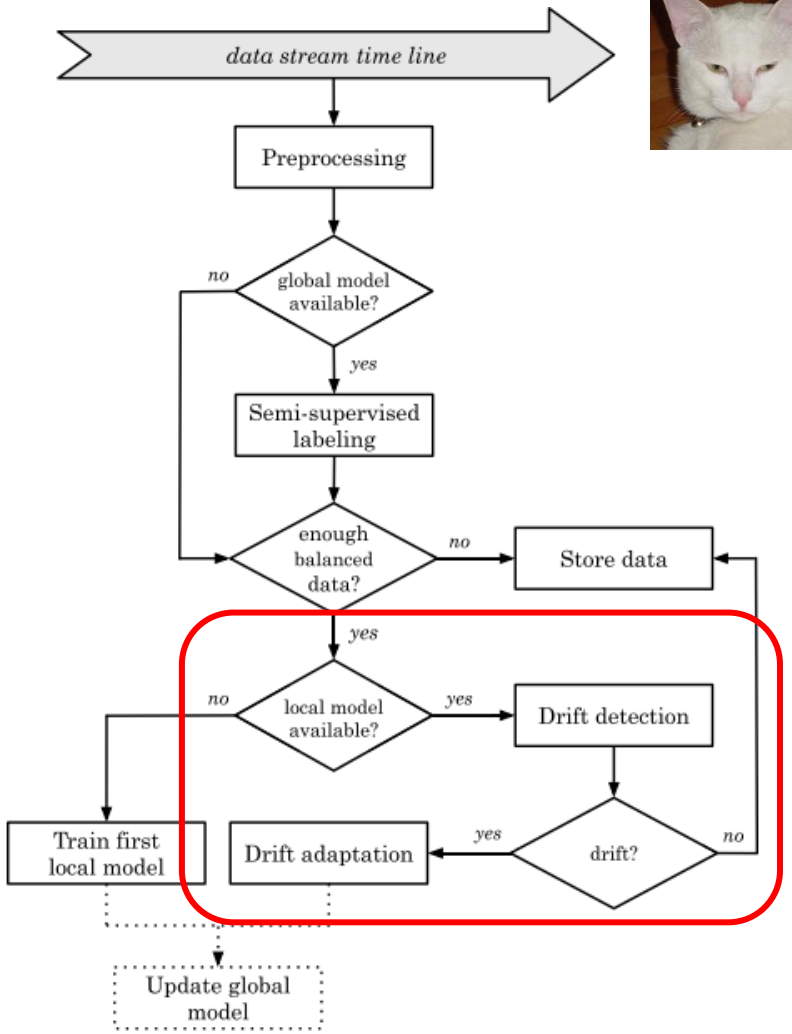
Output : Boolean indicating whether a drift is detected or not.

$s_f \leftarrow 0$
$T_h \leftarrow -\log(\lambda)$
$N \leftarrow |\mathcal{Q}|$
for $k \leftarrow \Delta$ **to** $N - \Delta$ **do**
  $m_b \leftarrow \texttt{mean}(q_1 : q_k \in \mathcal{Q})$
  $m_a \leftarrow \texttt{mean}(q_{k+1} : q_N \in \mathcal{Q})$
  if $m_a \leq (1-\lambda) \cdot m_b$ **then**
    $s_k \leftarrow 0$
    $[\hat{\alpha}_b, \hat{\beta}_b] \leftarrow \texttt{estimateParams}(q_1 : q_k)$ // Get parameters of
    $[\hat{\alpha}_a, \hat{\beta}_a] \leftarrow \texttt{estimateParams}(q_{k+1} : q_N)$
    for $i \leftarrow k+1$ **to** N **do**
      $s_k \leftarrow s_k + \log\left(\frac{f(q_i \mid \hat{\alpha}_a, \hat{\beta}_a)}{f(q_i \mid \hat{\alpha}_b, \hat{\beta}_b)}\right)$
    end
    $s_f \leftarrow \max(s_f, s_k)$
  end
end
if $s_f > T_h$ **then**
  return *true*
else
  return *false*
end



Casado, F.E., Lema, D., Criado, M.F., Iglesias, R., Regueiro, C.V., and Barro, S. Concept drift detection and adaptation for federated and continual learning. *Multimed Tools Appl* **81**, 3397–3419 (2022).

Detección de deriva de concepto



Detección de deriva de concepto:

Varias **aproximaciones**:

- Basados en el ratio de errores
- Basados en la distribución de datos
- Mixtos

Nosotros usamos un **método basado en la distribución de datos**, comparando las funciones de densidad de probabilidad de los datos actual y pasadas

Arquitectura CDA-FedAvg: adaptación a la deriva

Algorithm 6: Drift adaptation method (`driftAdaptation`).

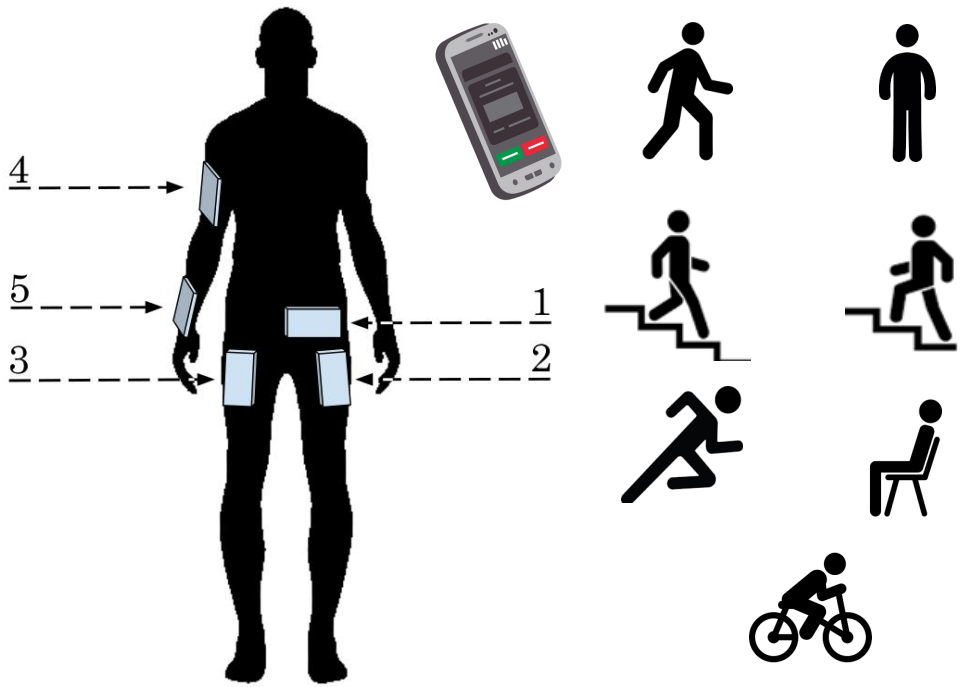
Input : Long-term memory $\mathcal{L}$, minimum amount of data to train L, minibatch size B, number of rounds R per change, number of local epochs E per round, and learning rate η.

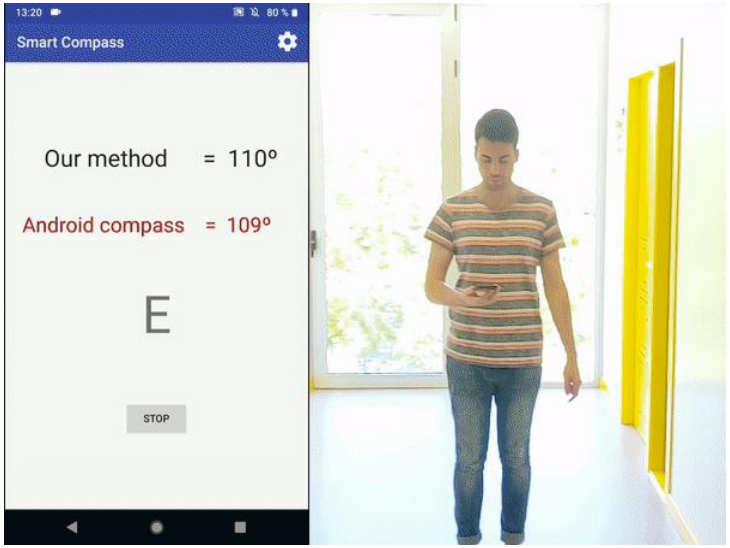
Output : Updated long-term memory $\mathcal{L}$.

```
1  L_new ← collectData()                    // Collect enough new data on the new concept
2  L ← L ∪ L_new                            // Expand the long-term memory with the new data
   /* Perform local training just in this client                                        */
3  for r ← 1 to R do
4      Split L into a set B of batches of size B
5      for e ← 1 to E do
6          for batch b ∈ B do
7              w_{t+1} ← w_t^G − η∇l(w_t^G; b)      // ∇l is the gradient of l on b
8          end
9      end
10     Send w_{t+1} to server to perform global aggregation
11 end
```

Casado, F.E., Lema, D., Criado, M.F., Iglesias, R., Regueiro, C.V., and Barro, S. Concept drift detection and adaptation for federated and continual learning. *Multimed Tools Appl* **81**, 3397–3419 (2022).

Reconocimiento de la actividad

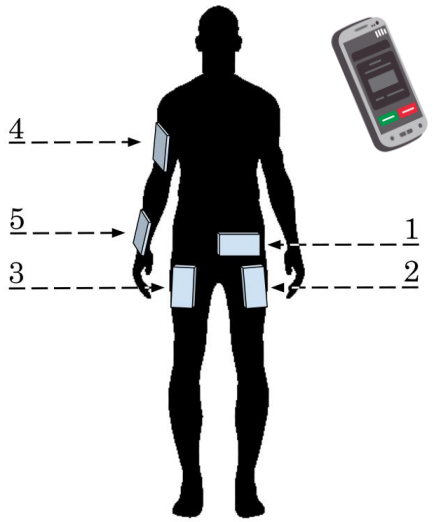


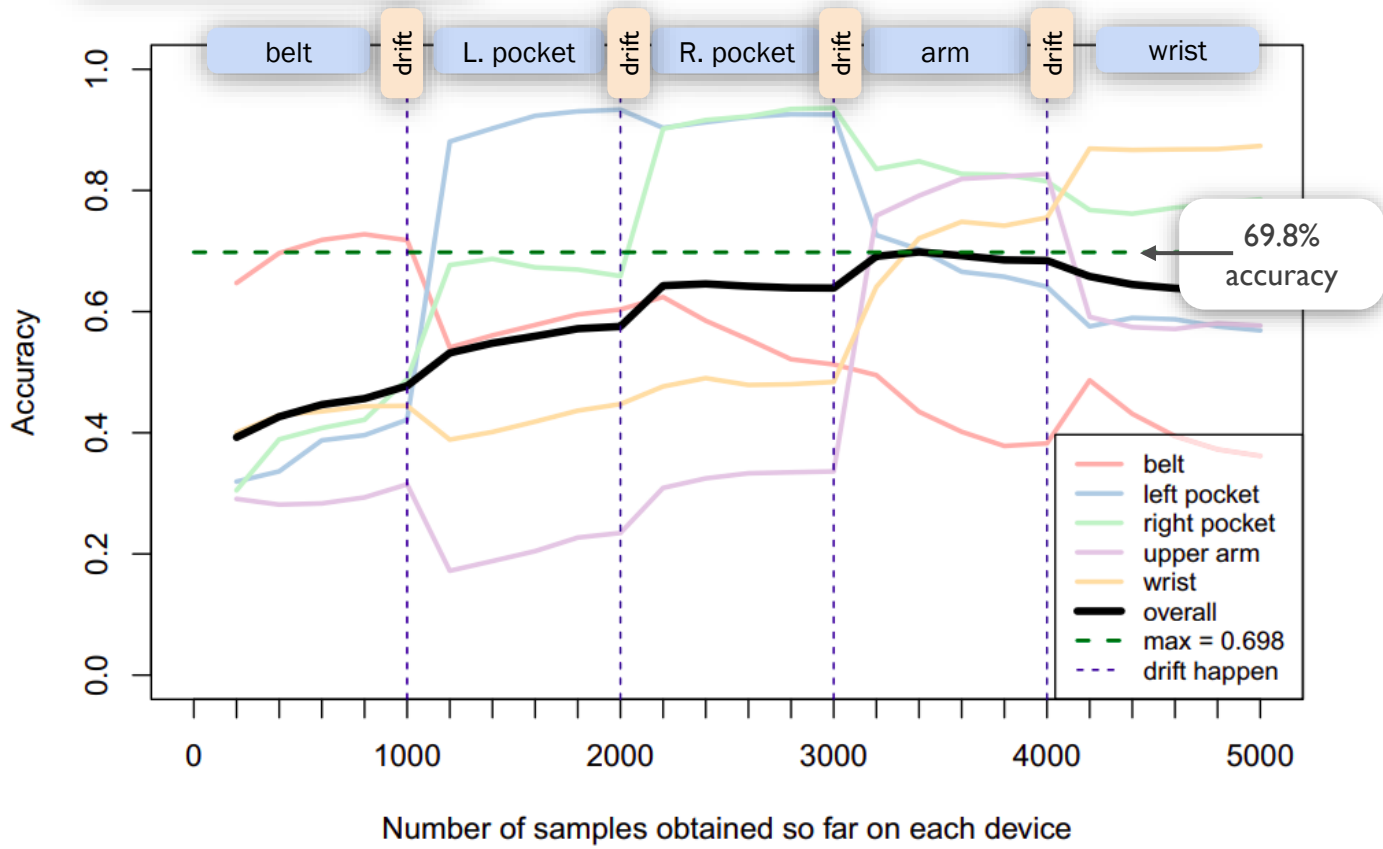


Casado, F.E., Lema, D., Criado, M.F., Iglesias, R., Regueiro, C.V., and Barro, S. Concept drift detection and adaptation for federated and continual learning. *Multimed Tools Appl* **81**, 3397–3419 (2022).

Reconocimiento de la actividad

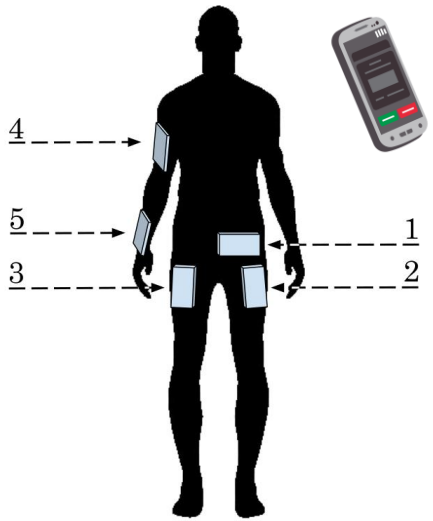
Mediante FedAvg



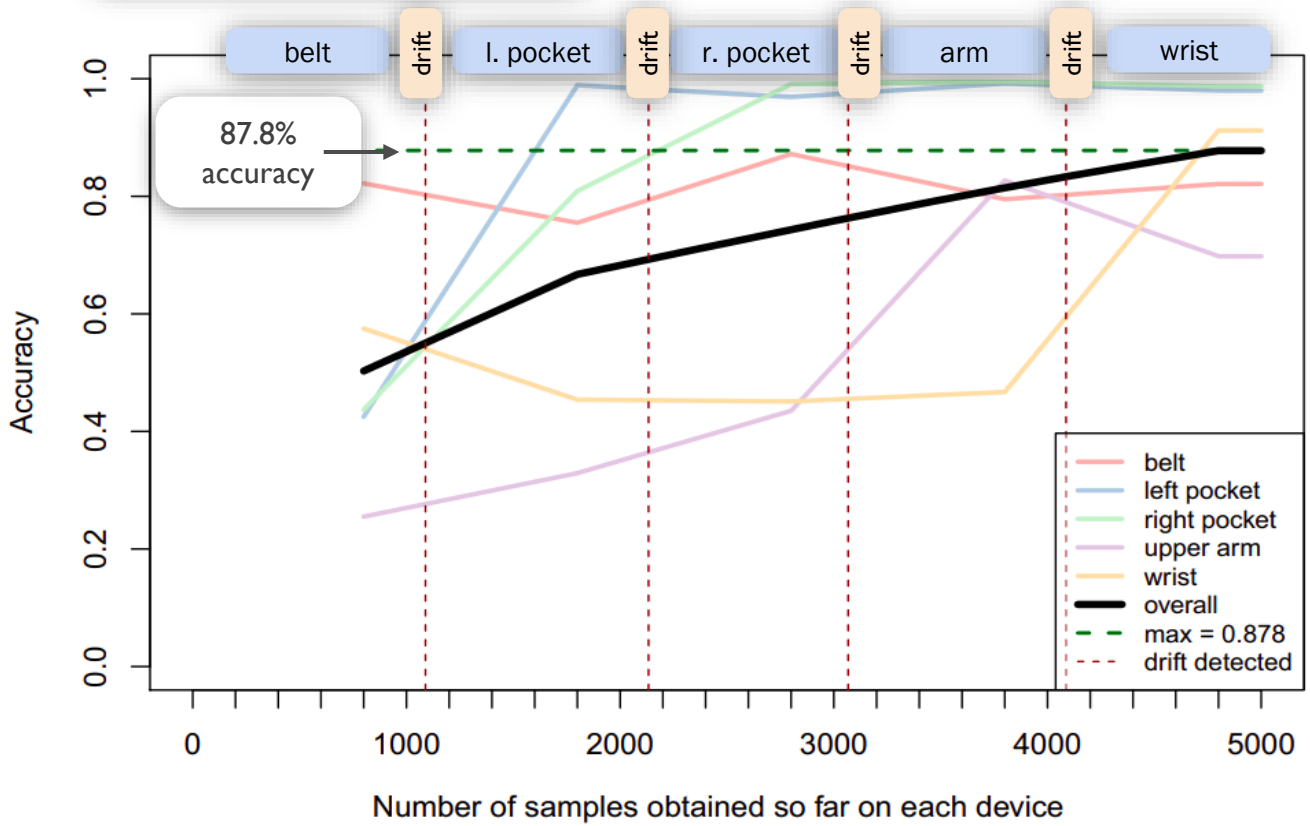


Casado, F.E., Lema, D., Criado, M.F., Iglesias, R., Regueiro, C.V., and Barro, S. Concept drift detection and adaptation for federated and continual learning. *Multimed Tools Appl* **81**, 3397–3419 (2022).

Reconocimiento de la actividad



Mediante CDA-FedAvg



Casado, F.E., Lema, D., Criado, M.F., Iglesias, R., Regueiro, C.V., and Barro, S. Concept drift detection and adaptation for federated and continual learning. *Multimed Tools Appl* **81**, 3397–3419 (2022).

¿Hay vida más allá del FedAvg y sus variaciones?

La mayor parte de las arquitecturas de FL utilizan **modelos de redes neuronales profundas** (DNNs)

- Tienen algunos **inconvenientes**:
  - Complejidad computacional
  - Opacidad
  - Problemas de generalización
  - ...
- **Alternativas**:
  - Uso de otros modelos de aprendizaje automático, como árboles de decisión, SVMs...

Arquitectura de comités ECFL

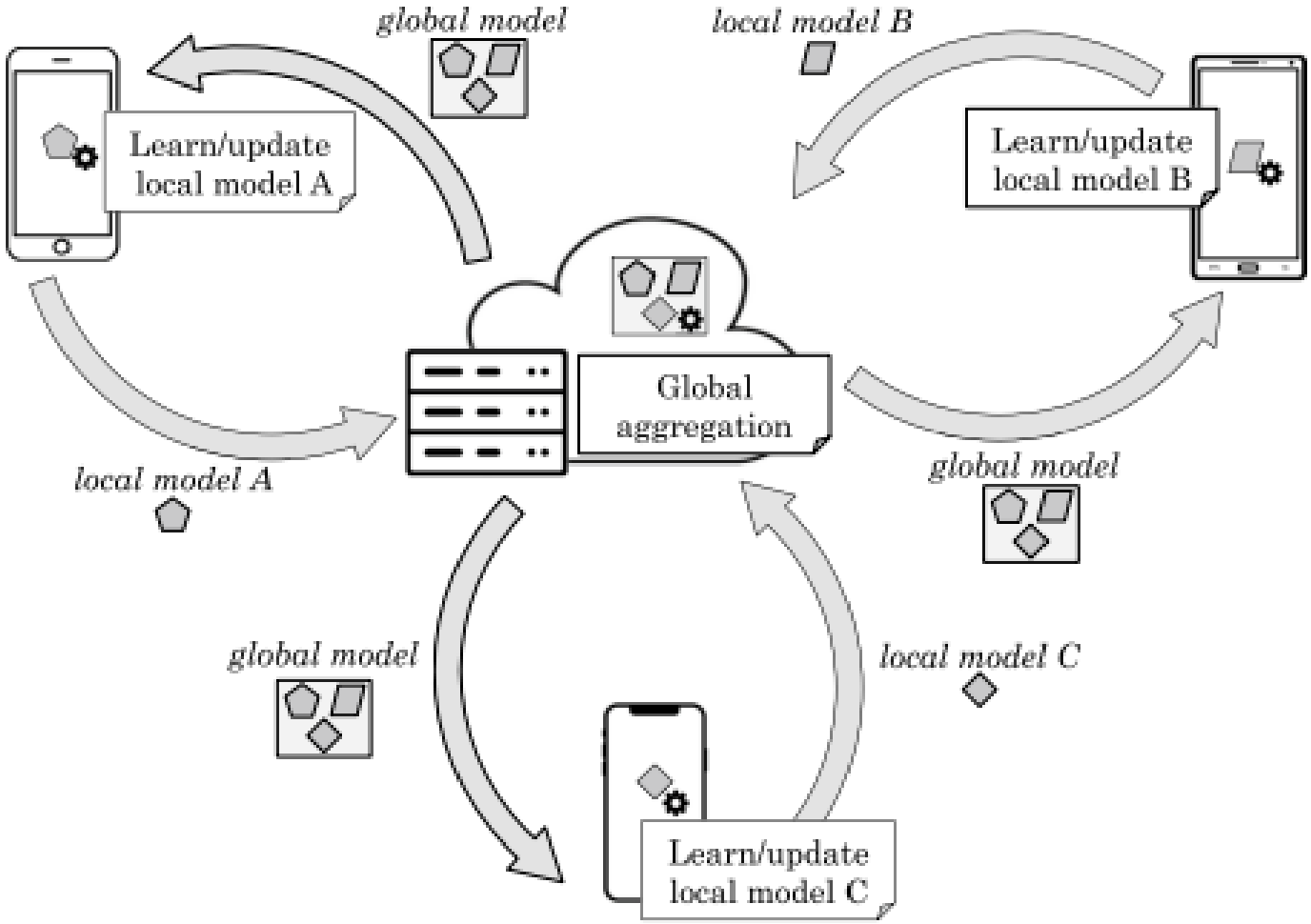
- Arquitectura basada en **comités** (*Ensemble Learning*)

- **Encaje natural** en el marco del **FL:**
  - Cada dispositivo (cliente) adiestra a su propio **modelo local**
  - Los modelos locales contribuyen al **comité**
  - El **modelo global** es utilizado por los clientes para para tomar decisiones futuras, al mismo tiempo que enriquece los conjuntos de datos locales de cara a próximas rondas de entrenamiento

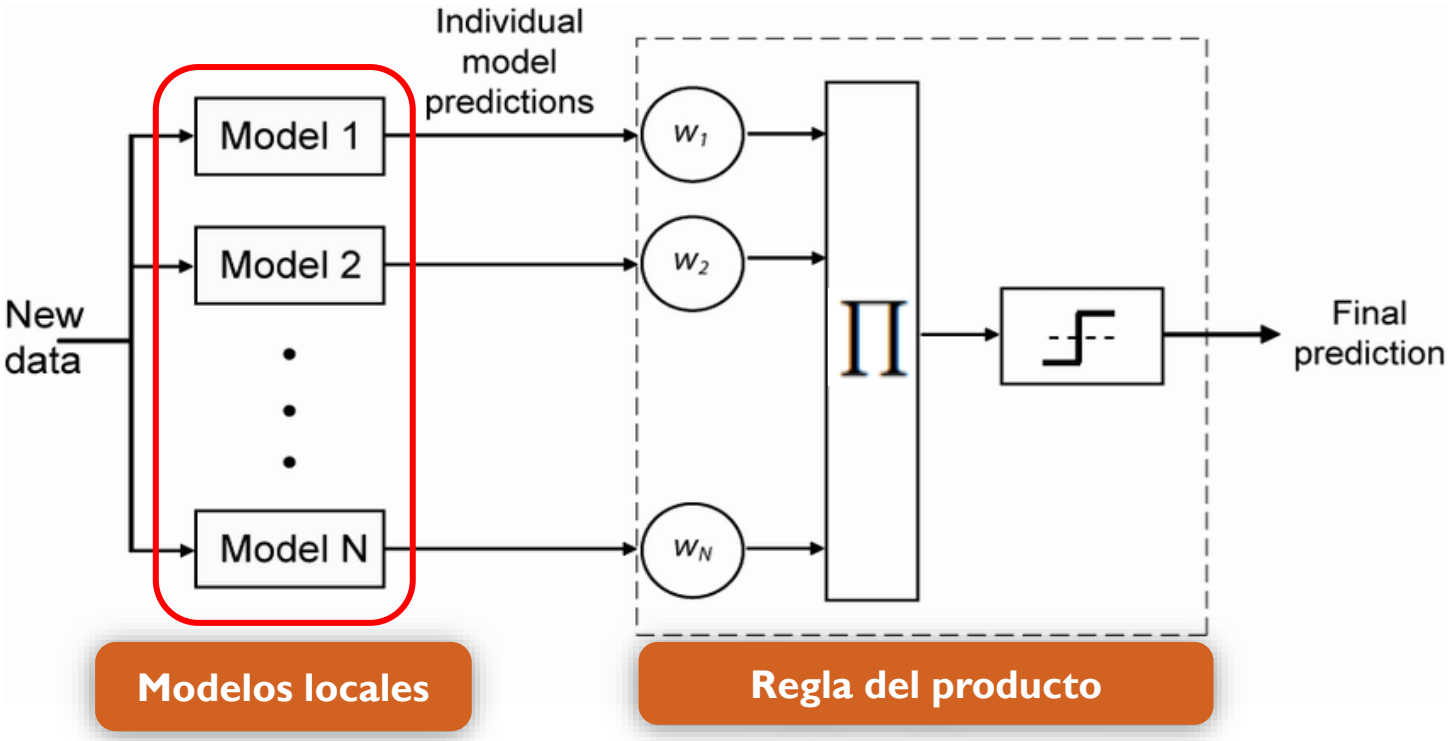
- **Beneficios:**
  - **Simplicidad** (*Green ML*)
  - **Explicabilidad**
  - **Flexibilidad**
  - **Autonomía**
  - **Adaptabilidad** (entrada / salida de clientes)

Casado, F.E., Lema, D., Iglesias, R., Regueiro, C.V., and Barro, S. Ensemble and continual federated learning for classification tasks. *Mach Learn* (2023). https://doi.org/10.1007/s10994-023-06330-z

Arquitectura de comités ECFL



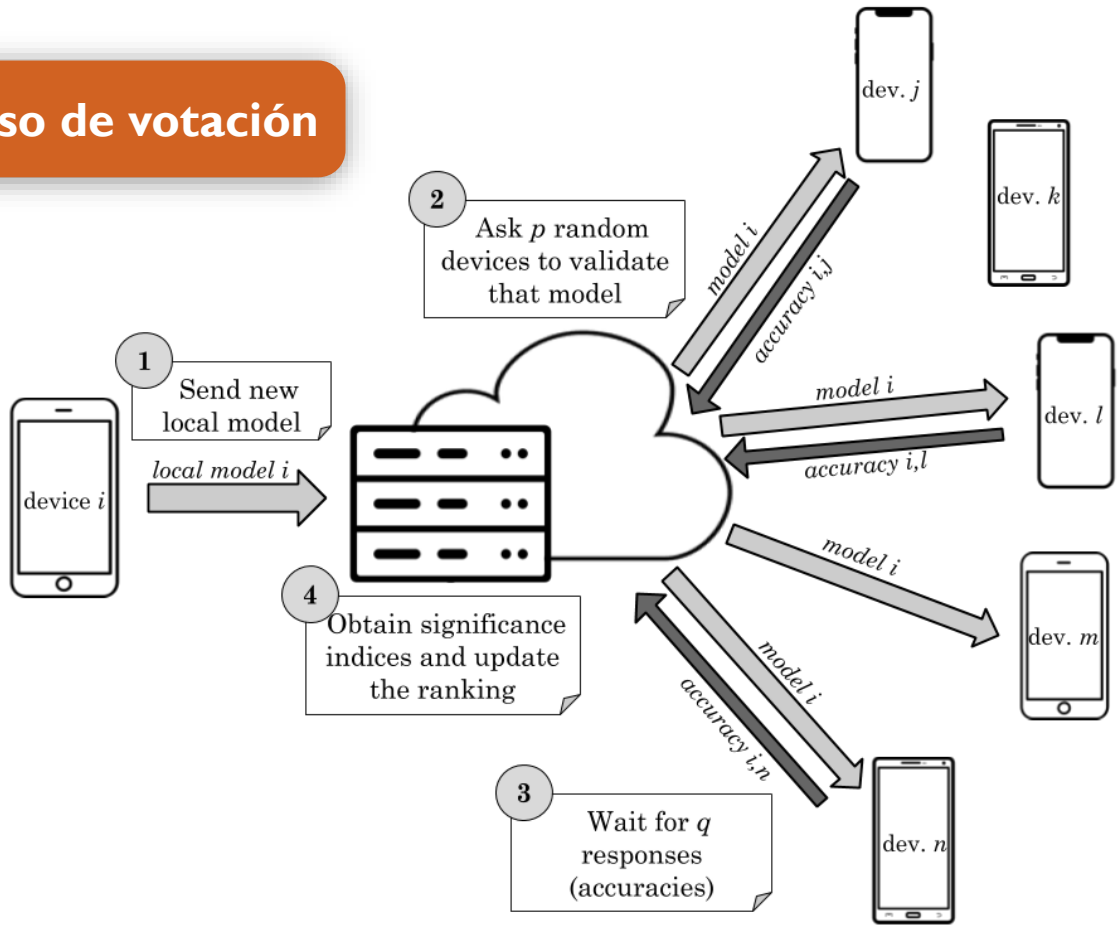
Arquitectura de comités ECFL



$$\prod_{i=1}^{N} y_{i_j}(x) = \max_{k=1}^{C} \prod_{i=1}^{N} y_{i_k}(x)$$

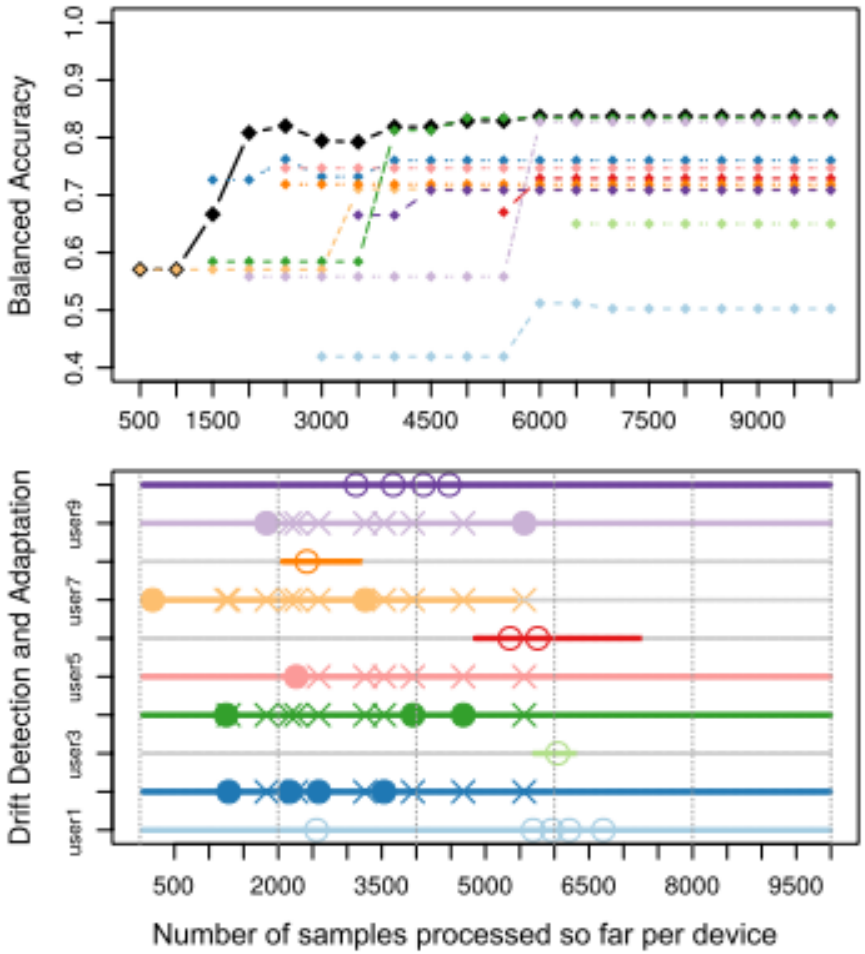
Arquitectura de comités ECFL

Proceso de votación



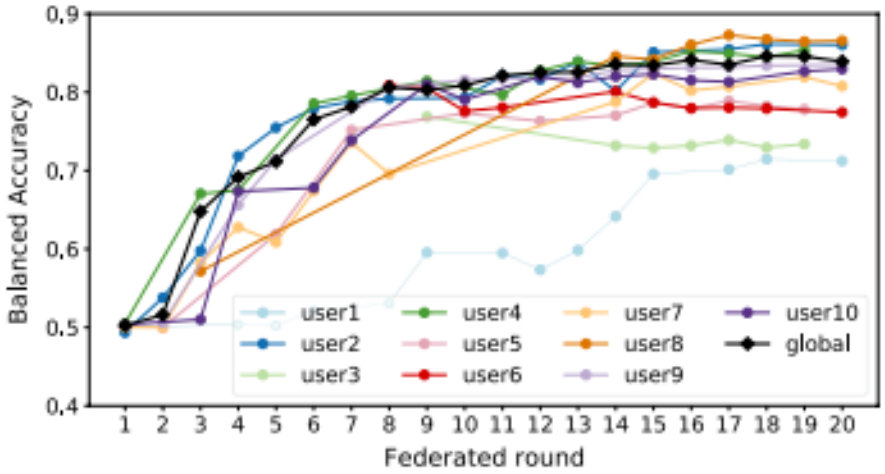
Casado, F.E., Lema, D., Iglesias, R., Regueiro, C.V., and Barro, S. Ensemble and continual federated learning for classification tasks. *Mach Learn* (2023). https://doi.org/10.1007/s10994-023-06330-z

Arquitectura de comités ECFL

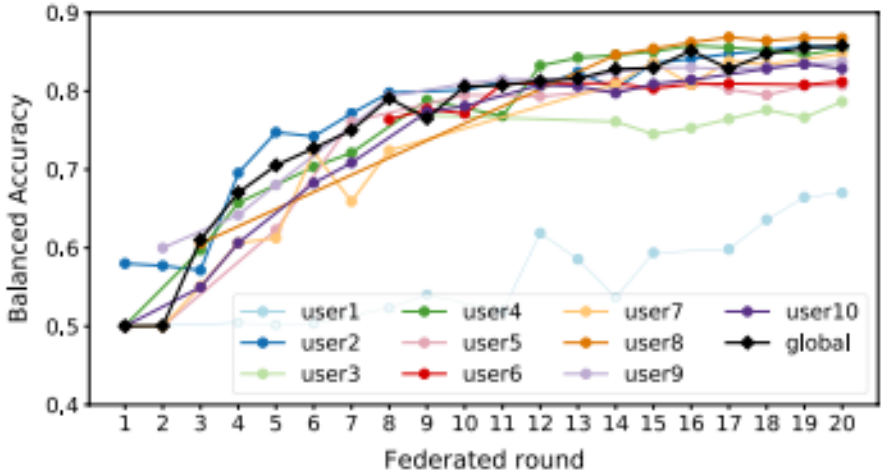


Casado, F.E., Lema, D., Iglesias, R., Regueiro, C.V., and Barro, S. Ensemble and continual federated learning for classification tasks. *Mach Learn* (2023). https://doi.org/10.1007/s10994-023-06330-z

Arquitectura de comités ECFL



(a) FedAvg.



(b) FedProx.

Casado, F.E., Lema, D., Iglesias, R., Regueiro, C.V., and Barro, S. Ensemble and continual federated learning for classification tasks. *Mach Learn* (2023). https://doi.org/10.1007/s10994-023-06330-z

Aprendizaje Federado: muchos retos por delante

Los **clientes** pueden “entrar y salir” del espacio problema

Los clientes pueden procesar los datos a diferentes velocidades (**asincronía**)

El **acceso** a las verdaderas **etiquetas** de los datos puede estar limitado o retrasarse

Datos no-IID (no Independientes e Igualmente Distribuidos)

Los datos pueden no ser estacionarios, lo que en general provoca **derivas de concepto**

Los datos no suelen estar disponibles de antemano, sino que van llegando con el tiempo (**aprendizaje continuo**) y no pueden almacenarse todos

Senen.barro@usc.es

universidade de santiago de compostela

@SenenBarro