Optimización Analítica Exacta para el Aprendizaje Incremental de Clases en Entornos Federados

Alejandro Dopico-Castro[1 [0009−0005−0592−5760]],
Oscar Fontenla-Romero[1 [0000−0003−4203−8720]],
Bertha Guijarro-Berdiñas[1 [0000−0001−8901−5441]], and
Amparo Alonso-Betanzos[1 [0000−0003−0950−0012]]

Universidade da Coruña, CITIC,
Facultade de Informática, Campus de Elviña s/n, A Coruña, Spain
{alejandro.dopico2, oscar.fontenla, berta.guijarro,
amparo.alonso.betanzos}@udc.es

Resumen El Aprendizaje Incremental de Clases Federado (FedCIL) enfrenta el desafío dual de la heterogeneidad espacial de las distribuciones de datos No-IID entre los clientes y la expansión temporal del espacio de etiquetas a medida que se incorporan nuevas clases. Los enfoques actuales, basados en optimización por gradiente, presentan limitaciones críticas: deriva de la representación, elevados costes computacionales y divergencia respecto al óptimo centralizado. En este trabajo, reformulamos FedCIL bajo un paradigma de aprendizaje analítico y proponemos FedIAL (**Fed**erated and **I**ncremental **A**nalytic class **L**earning). Al emplear un extractor de características preentrenado congelado, transformamos el aprendizaje del clasificador en un problema de optimización convexo resuelto mediante Mínimos Cuadrados Recursivos Regularizados (RRLS). FedIAL permite reconstruir el clasificador global mediante una agregación analítica en forma cerrada. Esta formulación produce un modelo global exactamente equivalente a la solución RRLS centralizada, independientemente de la distribución de los datos locales. Para gestionar la expansión de clases sin comprometer la privacidad, se introduce un mecanismo de *replay* latente coordinado que calibra las nuevas clases a partir de representaciones compactas, sin necesidad de almacenar ni intercambiar datos originales. El marco resultante requiere una única ronda de comunicación por tarea, lo que lo hace sustancialmente más eficiente que los métodos existentes. Los resultados en CIFAR-100 demuestran que FedIAL no solo supera el estado del arte en precisión, sino que reduce el consumo energético hasta en un 90 %.

Keywords: Aprendizaje Federado · Aprendizaje Incremental de Clases · Aprendizaje Analítico · Redes Neuronales · Olvido Catastrófico

1. Introducción

El Aprendizaje Federado (FL) permite entrenar modelos con datos distribuidos y privados, mientras que el Aprendizaje Incremental de Clases (CIL)

incorpora nuevas categorías de forma secuencial. En aplicaciones reales (salud, vehículos autónomos, etc.), ambos retos aparecen de forma simultánea. Esta combinación, conocida como FedCIL, es especialmente difícil porque combina heterogeneidad No-IID entre clientes y expansión temporal del espacio de etiquetas, lo que amplifica el olvido catastrófico y la inestabilidad de representación [21].

En este trabajo nos centramos en escenarios de datos no estructurados, donde las muestras se representan mediante embeddings de alta dimensión obtenidos a partir de un extractor de características. La mayoría de propuestas FedCIL siguen el paradigma de aprendizaje por gradientes: ajustan tanto el extractor como el clasificador mediante Descenso de Gradiente Estocástico (SGD) local y agregan los parámetros en el servidor [16]. Aunque este enfoque es efectivo en escenarios simples, puede transmitir señales ruidosas en lugar de una estructura estable, requiriendo múltiples rondas de comunicación sin garantizar la convergencia al óptimo global.

Ante esta problemática, proponemos una alternativa: fijar el extractor y entrenar únicamente el clasificador mediante optimización analítica. Dado un espacio de características estable e invariante en el tiempo, el problema local a resolver admite solución única. En lugar de gradientes, cada cliente comunica estadísticas suficientes de la geometría del espacio latente local, permitiendo reconstruir algebraicamente el clasificador global, lo que resulta equivalente al entrenamiento centralizado.

Sobre esta idea introducimos FedIAL, con tres componentes: (1) un extractor preentrenado congelado; (2) un solucionador basado en Mínimos Cuadrados Recursivos Regularizados (RRLS) local en forma cerrada; y (3) un protocolo federado de agregación analítica.

Nuestras contribuciones principales son:

- Reformulación de FedCIL bajo un paradigma de representación fija, transformando actualizaciones locales en optimización en forma cerrada.
- Esquema federado analítico que agrega estadísticas suficientes (sin gradientes) y una demostración formal de equivalencia con entrenamiento centralizado.
- Protocolo de coordinación global del aprendizaje incremental que reduce el olvido y mejora la calibración de nuevas clases.
- Validación experimental en el conjunto de datos CIFAR-100 con mejoras de precisión y reducción energética de hasta el 90 %.

2. Estado del Arte

El Aprendizaje Federado Incremental de Clases (FedCIL) enfrenta un doble desafío: la heterogeneidad espacial de los datos y la expansión temporal del espacio de etiquetas, lo que deriva en el denominado *olvido compuesto* [18,14]. Mientras que los trabajos fundacionales de FL (FedAvg, FedProx) [16,12] asumen un espacio de clases estático, los métodos de CIL estándar en entornos centralizados abordan el problema mediante tres familias principales de enfoques.

Los métodos basados en *replay* almacenan un subconjunto de ejemplos o representaciones de tareas anteriores para mitigar el olvido mediante la reutilización de conocimiento previo [17]. Los enfoques de regularización penalizan explícitamente la desviación de parámetros importantes para tareas pasadas [13], mientras que los métodos basados en aislamiento arquitectural asignan subconjuntos disjuntos de parámetros a cada tarea para evitar interferencia [15].

Sin embargo, al trasladarse a escenarios federados, estas estrategias pierden eficacia debido al sesgo inducido por la heterogeneidad No-IID, donde las distribuciones locales incompletas amplifican la inestabilidad del aprendizaje y dificultan la consolidación global del conocimiento.

Recientemente, trabajos como GLFC, TARGET o LANDER[2,23,19] han buscado reconciliar las actualizaciones locales con la consistencia global. Sin embargo, estos enfoques dependen de una optimización iterativa costosa por la retropropagación o bien por modelos generativos complejos para evitar el intercambio de datos privados.

Para mitigar esta ineficiencia, surge el Aprendizaje Analítico en CIL. Se ha demostrado que el uso de extractores de características congelados junto a clasificadores analíticos iguala o supera el rendimiento de los métodos basados en gradiente, con la ventaja de un coste computacional ínfimo [24,3,5].

Nuestra propuesta, FedIAL, se integra en este paradigma diferenciándose de aproximaciones del estado del arte, como STSA [6], en dos aspectos fundamentales: (i) opera de forma estrictamente analítica desde la tarea inicial, eliminando la dependencia y sensibilidad a las actualizaciones por gradiente; y (ii) incorpora un mecanismo de coordinación global del *replay* latente que equilibra las contribuciones de todos los clientes antes de la reconstrucción analítica del clasificador. Mediante RRLS [4], esta estrategia garantiza una fusión matemáticamente exacta bajo una representación fija, de modo que el modelo global constituye una agregación sin pérdida de las estadísticas locales, en lugar de una aproximación estocástica.

3. Formalización del Problema

En el escenario de FedCIL, consideramos un modelo global f_θ entrenado de forma colaborativa por N clientes a través de una secuencia de T tareas. Cada tarea t introduce un conjunto disjunto de clases $\mathcal{C}_t$. En un escenario IID, todos los clientes muestrean ejemplos de la misma distribución global y observan potencialmente todas las clases. En contraste, en el escenario *Partial-Class No-IID* considerado en este trabajo, cada cliente i observa únicamente un subconjunto local $\mathcal{C}_i^t \subseteq \mathcal{C}_t$, correspondiente a una fracción ρ de las clases disponibles, cuyas proporciones siguen una distribución de Dirichlet $\mathrm{Dir}(\alpha)$ [8,11]. Esta heterogeneidad induce deriva de cliente (*client drift*) y un sesgo persistente hacia las clases más recientes. El objetivo consiste en hallar los parámetros θ^* que minimicen el riesgo empírico acumulado $\mathcal{L}$ sin necesidad de acceso a los datos originales de tareas previas:

$$\theta^* = \arg\min_{\theta} \sum_{t=1}^{T} \mathbb{E}_{(x,y)\sim\mathcal{D}_t}[\mathcal{L}(f_\theta(x), y)] \tag{1}$$

Mientras que los enfoques convencionales emplean entropía cruzada ($\mathcal{L} = \ell_{ce}$)[24], FedIAL optimiza un residuo cuadrático regularizado ($\mathcal{L} = |\mathbf{Y} - \mathbf{X}\mathbf{W}|^2$) para la capa de clasificación. Esta formulación analítica permite obtener una solución global exacta sin las inestabilidades del gradiente.

4. Método

FedIAL aborda el escenario FedCIL bajo la hipótesis de un espacio latente fijo inducido por un extractor de características preentrenado. En este marco, cada cliente entrena un clasificador analítico sobre representaciones congeladas, mientras que el servidor coordina un mecanismo de *replay* latente y agrega estadísticas suficientes para reconstruir el modelo global. Esta restricción convierte el problema incremental en una optimización convexa, cuya solución puede agregarse exactamente entre clientes.

4.1. Base Arquitectural y Solucionador Local

Para mitigar la deriva de la representación (*representation drift*), se emplea un extractor de características F_θ preentrenado y congelado, que proyecta cada entrada en un espacio latente fijo: $Z_i = F_\theta(X_i) \in \mathbb{R}^{n_i \times d}$ donde X_i denota los datos de entrada locales y Z_i sus correspondientes representaciones en el espacio de embeddings. Al fijar θ, el aprendizaje del clasificador se reduce a un problema convexo, donde sustituimos el SGD por una solución analítica basada en Mínimos Cuadrados Recursivos Regularizados (RRLS)[24,4], garantizando unicidad y optimalidad de la solución.

Cada cliente i realiza una aproximación de primer orden mediante la diagonal de la matriz Jacobiana $\mathbf{G}_i \in \mathbb{R}^{d\times d}$ y calcula la descomposición en valores singulares (SVD) de las características proyectadas:

$$[U_i, S_i, \sim] = \mathrm{SVD}(Z_i\mathbf{G}_i), \tag{2}$$

donde $U_i \in \mathbb{R}^{d\times r}$ contiene los vectores singulares izquierdos, $S_i \in \mathbb{R}^{r\times r}$ los valores singulares y r es el rango de $Z_i\mathbf{G}_i$.

Simultáneamente, se computa la matriz de momentos $M_i \in \mathbb{R}^{d\times|\mathcal{C}_{1:k}|}$, la cual condensa la geometría local necesaria para la síntesis global:

$$M_i = \sum_{n=1}^{m} z_{i,n}^T \big(f'(\bar{d}_{i,n}) \odot f'(\bar{d}_{i,n}) \odot \bar{d}_{i,n}\big) \tag{3}$$

donde $\odot$ denota el producto elemento a elemento y $\bar{d}_{i,n} \in \mathbb{R}^{|C_{1:k}|}$ representa el vector de preactivaciones del clasificador para dicho ejemplo, correspondiente a las clases observadas hasta la tarea k.

La función $f(\cdot)$ denota la activación aplicada a las salidas del clasificador, mientras que $f'(\cdot)$ representa su derivada, que modula la contribución de cada clase en función de la sensibilidad local del modelo.

El triplete (U_i, S_i, M_i) resume la información necesaria para la agregación analítica. En particular, U_i contiene una base ortonormal del subespacio latente local, S_i cuantifica la importancia de cada dirección mediante sus valores singulares y M_i almacena los momentos suficientes para reconstruir el clasificador global sin acceder a los datos originales.

4.2. Replay Coordinado y Síntesis Global

Para calibrar las nuevas clases en entornos No-IID, y evitar la necesidad de intercambiar datos, FedIAL utiliza un mecanismo de sobremuestreo coordinado en dos fases basado únicamente en un subconjunto de *embeddings* que se almacenan en un buffer latente $\mathcal{B}_i$, el cual guarda representaciones de las muestras junto con su etiqueta de clase.

Fase I: Alineamiento. Los clientes envían al coordinador metadatos (etiquetas nuevas y conteos por clase en el *buffer*). El coordinador calcula la frecuencia N_{max} de la clase más representada y asigna a cada cliente una cuota de sobremuestreo para cada una de las clases $q_{i,c} = \left\lceil N_{max} \frac{|B_{i,c}|}{\sum_j |B_{j,c}|} \right\rceil$. Esto garantiza que el entrenamiento global sea balanceado a pesar de la heterogeneidad local. En esta implementación, las muestras a incluir en el *buffer* se seleccionan mediante un muestreo aleatorio uniforme hasta completar la cuota $q_{i,c}$, aunque el método es agnóstico a la técnica de selección empleada.

Fase II: Fusión. Los clientes sobremuestrean su *buffer* según $q_{i,c}$, recalculan sus estadísticas (U_i, S_i, M_i) y las envían de nuevo al coordinador. El coordinador fusiona los momentos y las bases ortogonales mediante un SVD de rango reducido:

$$M_g = \sum_{i=1}^{N} M_i, \quad [U_g, S_g, \sim] = \text{SVD}([U_1 S_1 \parallel U_2 S_2 \parallel \ldots \parallel U_N S_N]) \tag{4}$$

Esta operación es computacionalmente eficiente incluso para un elevado número de clientes N, ya que el coste depende principalmente de la dimensión del *embedding* d y no del número total de muestras originales. Finalmente, calcula los pesos globales exactos como:

$$W_g = U_g (S_g^2 + \lambda I)^{-1} U_g^T M_g \tag{5}$$

donde λ es el parámetro de regularización. El procedimiento completo se resume en el Algoritmo 1.

Algorithm 1 Agregación Federada Analítica

Require: Clientes $c_1 \dots c_N$, tarea k, buffer $\mathcal{B}$
Fase I: Alineamiento estadístico
for cada cliente c_i **en paralelo do**
    $\mathcal{C}_{i,k}, \{n_{i,c}\}, \{|B_{i,c}|\} \leftarrow \text{MetadatosLocales}(c_i)$
    Enviar metadatos al coordinador.
end for
Coordinador: calcular $\mathcal{C}_{global}, N_{max}, \{q_{i,c}\}$ y **difundir**

Fase II: Entrenamiento local y fusión global
for cada cliente c_i **en paralelo do**
    Sobremuestrear $\mathcal{B}_i$ según $q_{i,c}$ para equilibrar la distribución global
    Calcular estadísticas locales (U_i, S_i, M_i) con RRLS
    Enviar (U_i, S_i, M_i) al coordinador.
end for
Coordinador:
    $M_g \leftarrow \sum M_i$
    $[U_g, S_g] \leftarrow \text{SVD}([U_1 S_1 \parallel \dots \parallel U_N S_N])$
    $W \leftarrow U_g (S_g^2 + \lambda I)^{-1} U_g^T M_g$
    Difundir W_g a todos los clientes.

Propiedades y Eficiencia

Teorema 1 (Equivalencia Centralizada) *Bajo la hipótesis de una representación fija compartida por todos los clientes, aritmética exacta y sin truncamiento en la SVD, los pesos globales W_g obtenidos mediante la agregación de FedIAL coinciden con la solución RRLS centralizada W_{batch} calculada sobre la unión de todos los datos.*

En consecuencia, FedIAL es agnóstico a la distribución de datos, al orden de agregación y a la topología de comunicación, preservando exactamente la solución centralizada mientras mantiene una elevada eficiencia computacional.

En términos de eficiencia, asumiendo un escenario de referencia basado en imágenes RGB de 224×224 píxeles frente a *embeddings* latentes de dimensión 512 (como los obtenidos con una ResNet-18), el uso de representaciones latentes reduce el consumo de memoria en aproximadamente un factor de $\sim 75\times$.

Cada tarea incremental requiere una única ronda de comunicación compuesta por dos fases: una fase inicial de intercambio de metadatos para coordinar el *replay* y una segunda fase para la agregación analítica de las estadísticas locales.

5. Experimentos

Evaluamos FedIAL en diversos escenarios de FedCIL, analizando su precisión, estabilidad frente al olvido y eficiencia computacional en comparación con métodos basados en gradiente y alternativas analíticas bajo condiciones equivalentes.

Cuadro 1: Comparación de rendimiento y eficiencia en CIFAR-100 con partición Dirichlet $\alpha = 1{,}0$. Se reportan precisión final (A_T), precisión media ($\bar{A}$), olvido (F_T), energía y tiempo.

Método	$A_T \uparrow$	$\bar{A} \uparrow$	$F_T \downarrow$	Energía (Wh) $\downarrow$	Tiempo (min) $\downarrow$
TARGET	5.20	31.71	42.18	1936.4	375.2
LANDER	28.66	44.75	64.75	2152.1	368.5
STSA	51.95	62.36	**11.62**	**94.5**	**24.9**
FedIAL	**57.38**	**70.13**	13.97	125.4	34.7

5.1. Configuración Experimental

Utilizamos el conjuntos de datos **CIFAR-100**[10] bajo un protocolo de $T = 10$ tareas disjuntas y $N = 30$ clientes y promediamos 3 semillas (42, 1983, 2020). En el escenario No-IID, aplicamos una partición de clases parciales donde cada cliente observa solo una fracción ($\rho = 0{,}6$) de clases, con proporciones regidas por una distribución de Dirichlet $Dir(\alpha)$[11] (por defecto $\alpha = 1{,}0$).

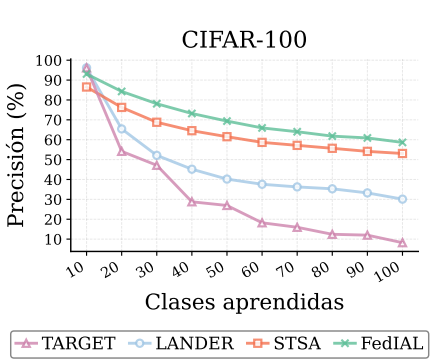
Comparamos nuestra propuesta con los algoritmos de estado del arte TARGET, LANDER y STSA[23,19,6]. Mientras que los dos primeros optimizan la red completa mediante SGD, STSA emplea una solución analítica tras la primera tarea. FedIAL prescinde de planificadores de *learning rate* u optimizadores iterativos, utilizando únicamente un factor de regularización $\lambda = 10^{-3}$. Todos los métodos emplean una arquitectura ResNet-18[7] con un extractor congelado y con preentrenamiento autosupervisado idéntico[22]. En FedIAL, siguiendo la configuración empleada en trabajos previos, cada cliente mantiene un buffer latente de capacidad fija $M = 500$ con muestreo equilibrado por clase. Analizar el impacto de este hiperparámetro queda como trabajo futuro.

Reportamos precisión final A_T, precisión media $\bar{A}$ (promedio de la precisión tras cada tarea) y olvido final F_T (degradación media del rendimiento sobre las clases previamente aprendidas). También medimos tiempo total y energía con CodeCarbon[1] en una NVIDIA 3090.

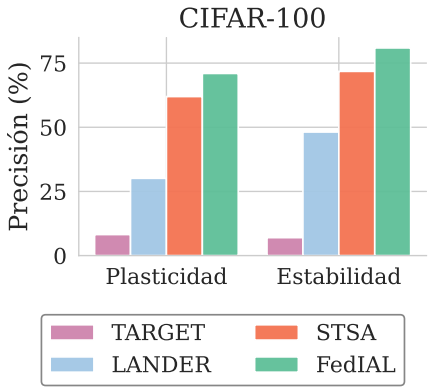
5.2. Análisis de Rendimiento y Eficiencia

La tabla 1 resume los resultados principales en CIFAR-100. FedIAL supera consistentemente a todos los métodos comparados en precisión final (A_T), alcanzando un **57.38 %**. Este comportamiento se mantiene consistente bajo distintos escenarios del mismo conjunto de datos, incluyendo variaciones de heterogeneidad (Tabla 2) y escenarios de largo horizonte temporal (Tabla 3), donde el número de tareas secuenciales aumenta para evaluar la acumulación progresiva de olvido.

Aunque STSA obtiene los mejores resultados en olvido (F_T) y eficiencia computacional, FedIAL mejora de forma consistente la precisión media y final, lo que sugiere una mayor capacidad representacional global. Este comportamiento



(a) Evolución de precisión a lo largo de 10 tareas secuenciales.



(b) Compromiso plasticidad-estabilidad.

Figura 1: Análisis dinámico del rendimiento en CIFAR-100.

evidencia el compromiso entre estabilidad y plasticidad: el olvido mide degradación relativa de conocimiento previo, mientras que la precisión final depende también de la calidad de las representaciones adquiridas en tareas posteriores. Por ello, valores ligeramente superiores de olvido pueden coexistir con mejoras sustanciales en desempeño global.

Además, la reducción de consumo energético superior al 90 % frente a los métodos basados en gradiente (TARGET y LANDER) se mantiene estable en todos los escenarios evaluados, lo que refuerza la ventaja computacional del enfoque analítico.

La evolución temporal de la precisión (Fig. 1a) muestra una degradación más controlada. Al evitar actualizaciones basadas en gradientes y operar directamente sobre estadísticas agregadas, se reduce la interferencia entre clases antiguas y nuevas, estabilizando el aprendizaje a lo largo de las tareas.

En la Figura 1b, se observa el compromiso entre plasticidad y estabilidad. FedIAL evita el colapso de estabilidad típico de métodos puramente iterativos, al mismo tiempo que mantiene una capacidad de adaptación competitiva frente a alternativas analíticas más rígidas.

5.3. Robustez y Escalabilidad

La Tabla 2 demuestra que, mientras los métodos iterativos colapsan ante una alta heterogeneidad ($\alpha = 0{,}5$), FedIAL mantiene un rendimiento casi invariable alrededor de 57 %. Esta estabilidad se deriva de la propiedad de equivalencia centralizada: la fusión de los parámetros de nuestro método preserva la integridad del modelo global independientemente de cómo se distribuyan los datos localmente.

En escenarios de largo plazo (Tabla 3), con 50 tareas secuenciales, la brecha de FedIAL frente a STSA se amplía hasta 6 puntos en A_T y además se reduce el

Cuadro 2: Robustez en CIFAR-100 para distintos niveles de heterogeneidad Dirichlet (α). Menor α implica mayor sesgo de etiquetas.

	IID ($\alpha = \infty$)		$\alpha = 1{,}0$		$\alpha = 0{,}5$	
Método	$A_T \uparrow$	$F_T \downarrow$	$A_T \uparrow$	$F_T \downarrow$	$A_T \uparrow$	$F_T \downarrow$
TARGET	12.54	35.84	5.20	42.18	6.86	34.15
LANDER	33.40	59.11	28.66	64.75	28.27	64.60
STSA	55.30	**11.57**	51.95	**11.62**	53.72	**11.97**
FedIAL	**57.93**	13.37	**57.38**	13.97	**57.40**	13.53

Cuadro 3: Escenario de muchas tareas. Se reportan precisión final (A_T), precisión media ($\bar{A}$) y olvido (F_T).

	CIFAR-100 (50 tareas)		
Método	$A_T \uparrow$	$\bar{A} \uparrow$	$F_T \downarrow$
STSA	51.80	61.79	14.53
FedIAL	**57.57**	**71.81**	**13.77**

olvido. Esto indica que la coordinación global del *replay* latente es crítica para mitigar el error acumulado cuando el horizonte temporal se expande.

6. Conclusión

FedIAL reformula FedCIL como un problema de optimización analítica, estable y con solución en forma cerrada. Al fijar la representación y agregar las estadísticas suficientes del clasificador, mitiga la deriva de cliente y de representación en escenarios No-IID, logrando mejoras sustanciales tanto en precisión como en eficiencia energética. Aunque el método evita el intercambio de imágenes originales, sí requiere almacenar y transmitir representaciones latentes compactas, que constituyen una forma de información derivada de los datos.

Como principal limitación, el rendimiento del sistema depende de la calidad del extractor de características preentrenado, ya que la representación fija condiciona directamente la separabilidad del espacio latente y, por tanto, la capacidad de generalización del modelo global. Este supuesto, aunque clave para la eficiencia del enfoque, restringe su adaptabilidad en dominios donde la representación inicial no es suficientemente expresiva o está desalineada con la distribución objetivo.

Como trabajo futuro, planteamos extender la evaluación a conjuntos de datos más complejos y heterogéneos, incluyendo mayor resolución y dominios distintos a CIFAR-100, con el objetivo de validar la robustez del enfoque en escenarios más realistas. Asimismo, exploramos la adaptación ligera del extractor mediante mecanismos como adaptadores o LoRA [9], relajando el supuesto de representación fija sin comprometer la eficiencia del esquema analítico.

También resulta relevante investigar alternativas al buffer latente basado en muestras, en particular enfoques generativos que permitan sintetizar representaciones de clases pasadas sin almacenamiento explícito de datos [20], ampliando así su aplicabilidad bajo restricciones de privacidad o memoria.

Agradecimientos

Este trabajo es parte del proyecto de I+D+i PID2023-147404OB-I00, financiado por MICIU/AEI/10.13039/501100011033/FEDER y FSE+/UE, y por el Ministerio para la Transformación Digital y de la Función Pública y Next-GenerationEU/PRTR (TSI-100925-2023-1) y Horizon Europe (GA 101070381), así como del proyecto de Grupos de Referencia Competitiva de la Xunta de Galicia ED431C 2026/38. CITIC, como miembro de la Red CIGUS, recibe subvenciones de la Xunta de Galicia y del Programa Operativo FEDER Galicia 2021-2027 (Subvención ED431G 2023/01).

Referencias

1. Courty, B., Schmidt, V., Luccioni, S., Goyal-Kamal, MarionCoutarel, Feld, B., Lecourt, J., LiamConnell, Saboni, A., Inimaz, supatomic, Léval, M., Blanche, L., Cruveiller, A., ouminasara, Zhao, F., Joshi, A., Bogroff, A., de Lavoreille, H., Laskaris, N., Abati, E., Blank, D., Wang, Z., Catovic, A., Alencon, M., Stechly, M., Bauer, C., de Araújo, L.O.N., JPW, MinervaBooks: mlco2/codecarbon: v2.4.1 (May 2024). https://doi.org/10.5281/zenodo.11171501, https://doi.org/10.5281/zenodo.11171501
2. Dong, J., Wang, L., Fang, Z., Sun, G., Xu, S., Wang, X., Zhu, Q.: Federated Class-Incremental Learning. In: CVPR. pp. 10164–10173 (2022)
3. Dopico-Castro, A., Fontenla-Romero, O., Guijarro-Berdiñas, B., Alonso-Betanzos, A.: Efficient single-step framework for incremental class learning in neural networks (2025), https://arxiv.org/abs/2509.11285
4. Fontenla-Romero, O., Guijarro-Berdiñas, B., Hernández-Pereira, E., Pérez-Sánchez, B.: FedHEONN: Federated and homomorphically encrypted learning method for one-layer neural networks. Future Generation Computer Systems **149**, 200–211 (Dec 2023). https://doi.org/10.1016/j.future.2023.07.018
5. Gao, Z., Jia, W., Zhang, X., Zhou, D., Xu, K., Dawei, F., Dou, Y., Mao, X., Wang, H.: Knowledge memorization and rumination for pre-trained model-based class-incremental learning. In: CVPR. pp. 20523–20533 (June 2025)
6. Guan, Z., Zhu, G., Zhou, Y., Liu, W., Wang, W., Luo, J., Gu, X.: Stsa: Federated class-incremental learning via spatial-temporal statistics aggregation. arXiv preprint arXiv:2506.01327 (2025)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. pp. 770–778 (2016)
8. Hsu, T.M.H., Qi, H., Brown, M.: Measuring the effects of non-identical data distribution for federated visual classification. arXiv preprint arXiv:1909.06335 (2019)
9. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. ICLR **1**(2), 3 (2022)

10. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images (2009)
11. Li, Q., Diao, Y., Chen, Q., He, B.: Federated learning on non-iid data silos: An experimental study. In: 2022 IEEE 38th international conference on data engineering (ICDE). pp. 965–978. IEEE (2022)
12. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. Proceedings of Machine learning and systems **2**, 429–450 (2020)
13. Li, Z., Hoiem, D.: Learning without forgetting. IEEE TPAMI **40**(12), 2935–2947 (2017)
14. Ma, Y., Xie, Z., Wang, J., Chen, K., Shou, L.: Continual federated learning based on knowledge distillation. In: IJCAI. pp. 2182–2188 (2022)
15. Mallya, A., Lazebnik, S.: Packnet: Adding multiple tasks to a single network by iterative pruning. In: CVPR. pp. 7765–7773 (2018)
16. McMahan, B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.y.: Communication-Efficient Learning of Deep Networks from Decentralized Data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. pp. 1273–1282. PMLR (Apr 2017), https://proceedings.mlr.press/v54/mcmahan17a.html, iSSN: 2640-3498
17. Rebuffi, S.A., Kolesnikov, A., Sperl, G., Lampert, C.H.: icarl: Incremental classifier and representation learning. CVPR **2017-January**, 5533–5542 (11 2017). https://doi.org/10.1109/CVPR.2017.587
18. Shenaj, D., Toldo, M., Rigon, A., Zanuttigh, P.: Asynchronous federated continual learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5055–5063 (2023)
19. Tran, M.T., Le, T., Le, X.M., Harandi, M., Phung, D.: Text-Enhanced Data-free Approach for Federated Class-Incremental Learning. In: CVPR. pp. 23870–23880 (2024)
20. Van de Ven, G.M., Tolias, A.S.: Generative replay with feedback connections as a general strategy for continual learning. arXiv preprint arXiv:1809.10635 (2018)
21. Van de Ven, G.M., Tuytelaars, T., Tolias, A.S.: Three types of incremental learning. Nature Machine Intelligence **4**(12), 1185–1197 (2022)
22. Yalniz, I.Z., Jégou, H., Chen, K., Paluri, M., Mahajan, D.: Billion-scale semi-supervised learning for image classification. CoRR **abs/1905.00546** (2019), http://arxiv.org/abs/1905.00546
23. Zhang, J., Chen, C., Zhuang, W., Lv, L.: TARGET: Federated class-continual learning via exemplar-free distillation. https://doi.org/10.48550/arXiv.2303.06937, http://arxiv.org/abs/2303.06937
24. Zhuang, H., Weng, Z., Wei, H., Xie, R., Toh, K.A., Lin, Z.: Acil: Analytic class-incremental learning with absolute memorization and privacy protection. In: NeurIPS. vol. 35, pp. 11602–11614 (2022)