



UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO

Métodos de adaptación de dominio profunda no supervisada

Macroentrenamiento en Inteligencia Artificial (MeIA)

Dr. Magdiel Jiménez Guarneros

20 de Junio de 2023

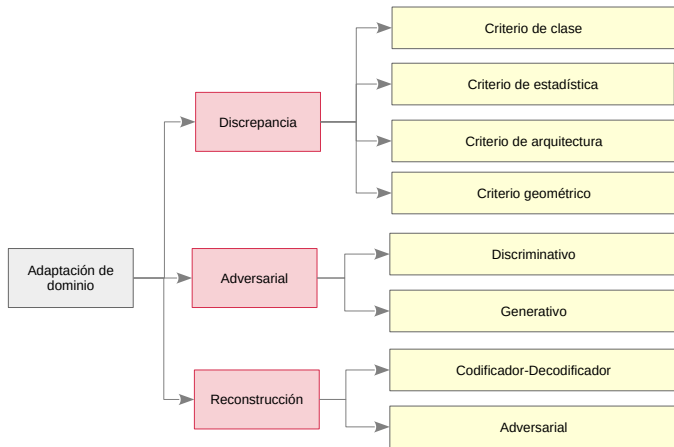




- ▶ Enfoques de adaptación de dominio
- ▶ Reducción de las diferencias marginales

Taxonomía de enfoques de adaptación de dominio

1 Enfoques de adaptación de dominio



[Wang and Deng, 2018]



Este enfoque ajusta una red neuronal profunda para reducir las diferencias de distribución estadística entre los dominios fuente y objetivo.

- **Criterio estadístico:** alinea la diferencia de la distribución estadística entre los dominios de fuente y objetivo utilizando medidas de discrepancia.
- **Criterio de arquitectura:** tiene como objetivo mejorar la capacidad de aprender características más transferibles ajustando las arquitecturas de las redes profundas.
- **Criterio de clase:** utiliza la información de la etiqueta de clase como guía para transferir conocimientos entre diferentes dominios.
- **Criterio geométrico:** supone que la relación de estructuras geométricas puede reducir el cambio de dominio.



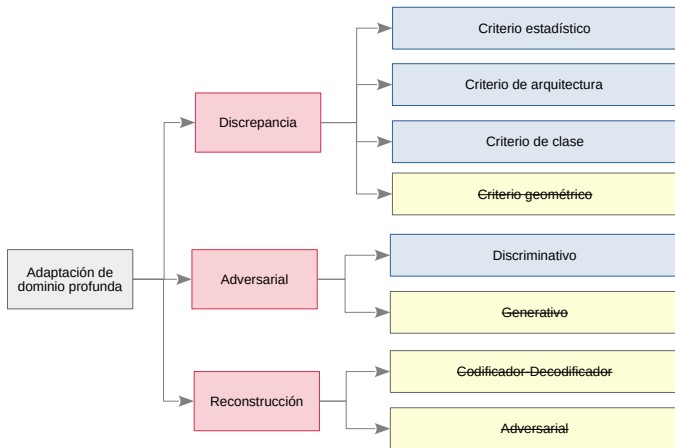
- **Discriminativo:** el extractor de características aprende una representación discriminativa utilizando las etiquetas en el dominio fuente y asigna los datos del dominio objetivo al mismo espacio a través de una función de costo de confusión de dominio, resultando en una representación invariante de dominio.
- **Generativo:** Uno de los casos típicos es utilizar ejemplos con ruido para generar muestras simuladas que sean similares a las muestras de objetivo y conservar la información de anotación del dominio fuente.



- **Codificador-decodificador:** Mediante el uso de Auto-codificadores apilados (SAE), los métodos de reconstrucción de codificador-decodificador combinan la red codificadora para el aprendizaje de representación con una red decodificadora para la reconstrucción de datos.
- **Adversarial:** el error de reconstrucción se mide como la diferencia entre las muestras reconstruidas y originales dentro de cada dominio mediante un mapeo cíclico obtenido a través de un discriminador adversarial.

Taxonomía en la clasificación de señales de EEG

1 Enfoques de adaptación de dominio



[Wang and Deng, 2018]



- Enfoques de adaptación de dominio
- Reducción de las diferencias marginales

Métodos que reducen las distribuciones marginales

2 Reducción de las diferencias marginales



Enfoque	Criterio	Método
Discrepancia	Criterio estadístico	Domain Adaptation Network (DAN) [Long et al., 2015], Correlation Alignment (CORAL) [Sun and Saenko, 2016], Central Moment Discrepancy [Zellinger et al., 2019]
	Arquitectura	Adaptive Batch Normalization (AdaBN) [Li et al., 2018], Domain Specific Batch Normalization (DSBN) [Chang et al., 2019], Feature Normalization [Saito et al., 2019]
Adversarial	Discriminativo	Domain Adversarial Neural Network (DANN) [Ganin et al., 2016], Wasserstein Generative Adversarial Network (WGAN) [Arjovsky et al., 2017]



- Los métodos de criterio estadístico reducen el cambio de la distribución estadística entre los dominios fuente y objetivo mediante el uso de una medida de discrepancia. Para ello se ajustan el conjunto de pesos Θ de una red neuronal, minimizando la siguiente función de pérdida:

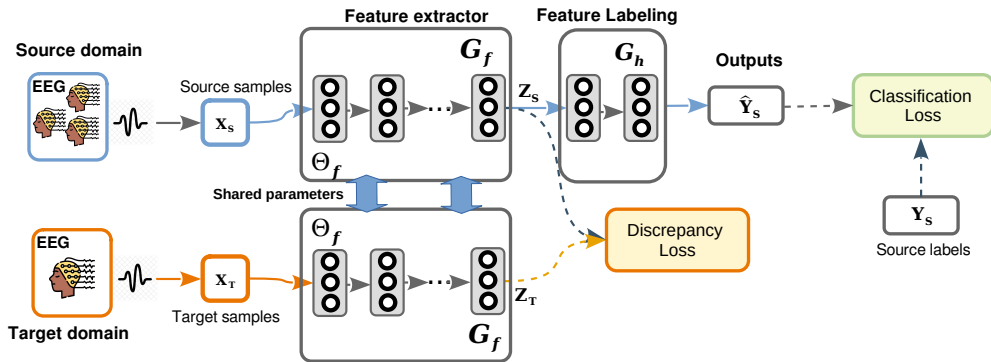
$$\mathcal{L}(\mathbf{X}_S, \mathbf{Y}_S, \mathbf{X}_T, G; \Theta_f, \Theta_h) = \mathcal{L}_{\text{cls}}(\mathbf{X}_S, \mathbf{Y}_S, G; \Theta_f, \Theta_h) + \lambda_{\text{dis}} \cdot \mathcal{L}_{\text{dis}}(\mathbf{X}_S, \mathbf{X}_T, G; \Theta_f), \quad (1)$$

donde el término \mathcal{L}_{cls} es una función de pérdida de clasificación, tal como entropía cruzada.

- \mathcal{L}_{dis} es una función de pérdida basada en una medida de discrepancia para penalizar las diferencias de distribución entre los dominios fuente y objetivo; λ_{dis} es un factor de peso.
- \mathcal{L}_{cls} ajusta los pesos Θ_f, Θ_h para garantizar un espacio separable entre las clases en el dominio fuente. Por otro lado, \mathcal{L}_{dis} ajusta los pesos Θ_f para hacer similares las distribuciones de características de diferentes dominios en una capa oculta de la red neuronal G .

Marco de trabajo: CMD

2 Reducción de las diferencias marginales



Central Moment Discrepancy (CMD)

2 Reducción de las diferencias marginales



CMD propuesto en [Zellinger et al., 2019] reduce los momentos estadísticos de orden más alto entre las representaciones de características de los dominios fuente y objetivo mediante

$$\mathcal{L}_{\text{dis}}(\mathbf{X}_S, \mathbf{X}_T, G_f; \Theta_f) = \|\mathbb{E}(G_f(\mathbf{X}_S)) - \mathbb{E}(G_f(\mathbf{X}_T))\|_2 + \sum_{m=2}^M \|C_m(G_f(\mathbf{X}_S)) - C_m(G_f(\mathbf{X}_T))\|_2 \quad (2)$$

donde M es el orden máximo de los momentos centrales. El término $C_m(\cdot)$ es definido como

$$C_m(\mathbf{A}) = \mathbb{E}[(\mathbf{A} - \mathbb{E}[\mathbf{A}])^m] \quad (3)$$

el cual es el momento central de orden m para el conjunto \mathbf{A} .



Actividades:

- Aplicar Leave-One-Subject-Out Cross-Validation sobre el conjunto de datos de carga cognitiva.
- Configurar los principales hiper-parámetros de la red neuronal: λ_{dis} , tamaño de lote, optimizador, tasa de aprendizaje, capa oculta de la red a alinear.
- Reportar los resultados de clasificación en términos de exactitud.
- Determinar si existe un efecto positivo en la eficacia de clasificación al alinear las distribuciones marginales.



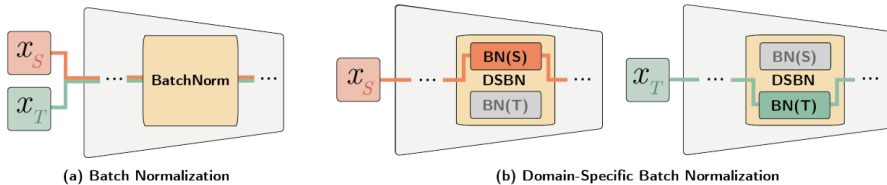
Estos métodos buscan ajustar las arquitecturas de redes profundas para aprender características transferibles.

Normalización de lotes específica por dominio [Chang et al., 2019]: incluye capas de normalización por lotes (*Batch normalization* en inglés) en una red neuronal profunda para realizar una estandarización específica para los dominios de fuente y objetivo.

Normalización de características en un espacio esférico [Saito et al., 2019]: El objetivo es imponer magnitudes equilibradas de vectores de características en todas las clases.

Normalización de lotes específica por dominio

2 Reducción de las diferencias marginales



[Chang et al., 2019]

Una capa de normalización de lotes (BN) por dominio específico (DSBN) [Chang et al., 2019] estandariza cada característica $x^{(j)}$ para cada dominio específico d de la siguiente manera:

$$\tilde{z}_j^{(d)} = \gamma_j^{(d)} \cdot \frac{(z_j^{(d)} - \mu_j^{(d)})}{\sigma_j^{(d)}} + \beta_j^{(d)} \quad (4)$$

donde μ_j y σ_j^2 son la media y la varianza del lote actual para la j -ésima característica; γ_j y β_j son parámetros de escala y traslación.



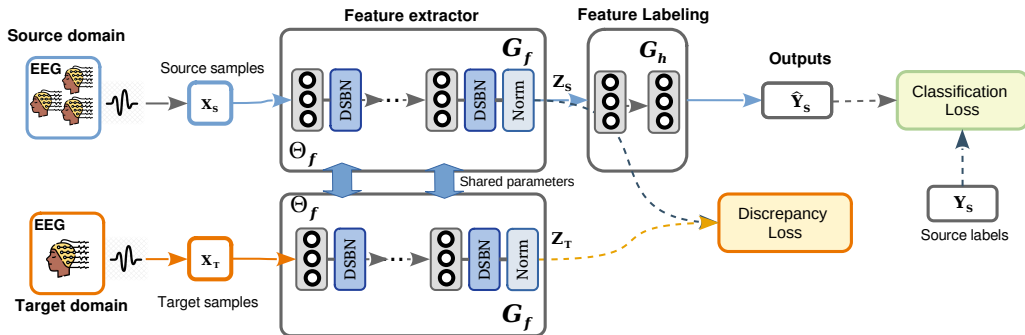
Una normalización L_2 [Saito et al., 2019] es aplicada en la salida del extractor de características con el propósito de normalizar las magnitudes de los vectores de características $\tilde{\mathbf{Z}} = G_f(\mathbf{X})$ como sigue:

$$\bar{\mathbf{Z}} = r \cdot \frac{\tilde{\mathbf{Z}}}{\|\tilde{\mathbf{Z}}\|_2}, \quad (5)$$

donde r es un radio.

Marco de trabajo: Criterio de arquitectura

2 Reducción de las diferencias marginales





Actividades:

- Aplicar Leave-One-Subject-Out Cross-Validation sobre el conjunto de datos de carga cognitiva.
- Configurar los principales hiper-parámetros de los métodos.
- Redefinir la capa de normalización de lotes de 1 dimensión a 3 dimensiones.
- Reportar los resultados de clasificación en términos de exactitud.
- Determinar si existe un efecto positivo en la eficacia de clasificación al alinear las distribuciones marginales por la inclusión de una normalización de lotes específica por dominio o una normalización esférica.



Domain Adversarial Neural Network (DANN) [Ganin et al., 2016] emplea un entrenamiento basado en redes adversarias con el propósito de producir características discriminativas con respecto a sus clases e invariantes a dominio. Para esto, DANN minimiza la pérdida de clasificación \mathcal{L}_{cls} y maximizando la pérdida adversarial \mathcal{L}_{adv} como sigue

$$\mathcal{L}(\mathbf{X}_S, \mathbf{Y}_S, \mathbf{X}_T, G, D; \Theta_f, \Theta_h, \Theta_d) = \mathcal{L}_{\text{cls}}(\mathbf{X}_S, \mathbf{Y}_S, G; \Theta_f, \Theta_h) - \lambda_{\text{adv}} \cdot \mathcal{L}_{\text{adv}}(\mathbf{X}_S, \mathbf{X}_T, G_f, D; \Theta_f, \Theta_d) \quad (6)$$

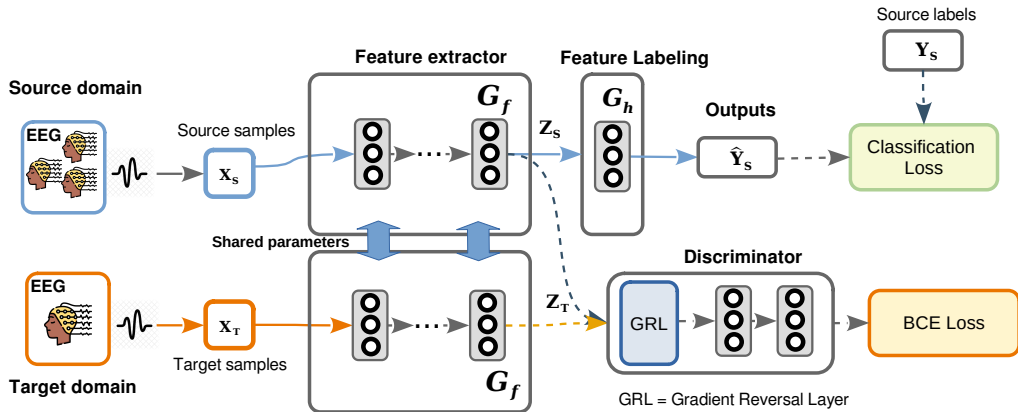
donde λ_{adv} es un factor de peso y la función de pérdida adversarial \mathcal{L}_{adv} es definida como la entropía cruzada binaria (*Binary Cross-entropy*)

$$\mathcal{L}_{\text{adv}}(\mathbf{X}_S, \mathbf{X}_T, G_f, D; \Theta_f, \Theta_d) = \mathbb{E}_{\mathbf{x}_S \sim \mathbf{X}_S} [\log D(G_f(\mathbf{x}_S))] + \mathbb{E}_{\mathbf{x}_T \sim \mathbf{X}_T} [\log (1 - D(G_f(\mathbf{x}_T)))] \quad (7)$$

mientras $D : \mathcal{Z} \rightarrow [0, 1]$ es una red neuronal discriminadora con pesos Θ_d , la cual es enfocada en distinguir si los ejemplos pertenecen al dominio fuente u objetivo.

Marco de trabajo: DANN

2 Reducción de las diferencias marginales











Actividades:

- Aplicar Leave-One-Subject-Out Cross-Validation sobre el conjunto de datos de carga cognitiva.
- Configurar los principales hiper-parámetros de la red neuronal: λ_{adv} , tamaño de lote, optimizador, tasa de aprendizaje, capa oculta de la red a alinear.
- Reportar los resultados de clasificación en términos de exactitud.
- Determinar si existe un efecto positivo en la eficacia de clasificación al alinear las distribuciones marginales por la inclusión de alguno de los componentes.






-  Arjovsky, M., Chintala, S., and Bottou, L. (2017).
Wasserstein generative adversarial networks.
In Precup, D. and Teh, Y. W., editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 214–223. PMLR.
-  Chang, W.-G., You, T., Seo, S., Kwak, S., and Han, B. (2019).
Domain-specific batch normalization for unsupervised domain adaptation.
In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pages 7354–7362.
-  Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., and Lempitsky, V. (2016).
Domain-adversarial training of neural networks.
The journal of machine learning research, 17(1):2096–2030.
-  Li, Y., Wang, N., Shi, J., Hou, X., and Liu, J. (2018).
Adaptive batch normalization for practical domain adaptation.
Pattern Recognition, 80:109–117.
-  Long, M., Cao, Y., Wang, J., and Jordan, M. (2015).
Learning transferable features with deep adaptation networks.
In *International conference on machine learning*, pages 97–105. PMLR.



-  Luo, Y. and Lu, B.-L. (2021).
Wasserstein-distance-based multi-source adversarial domain adaptation for emotion recognition and vigilance estimation.
In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1424–1428.
-  Saito, K., Kim, D., Sclaroff, S., Darrell, T., and Saenko, K. (2019).
Semi-supervised domain adaptation via minimax entropy.
In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8050–8058.
-  Sun, B. and Saenko, K. (2016).
Deep coral: Correlation alignment for deep domain adaptation.
In *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8–10 and 15–16, 2016, Proceedings, Part III 14*, pages 443–450. Springer.
-  Tzeng, E., Hoffman, J., Saenko, K., and Darrell, T. (2017).
Adversarial discriminative domain adaptation.
In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176.
-  Wan, Z., Yang, R., Huang, M., Zeng, N., and Liu, X. (2021).
A review on transfer learning in eeg signal analysis.
Neurocomputing, 421:1–14.



-  Wang, M. and Deng, W. (2018).
Deep visual domain adaptation: A survey.
Neurocomputing, 312:135–153.
-  Zellinger, W., Moser, B. A., Grubinger, T., Lughofer, E., Natschlager, T., and Saminger-Platz, S. (2019).
Robust unsupervised domain adaptation for neural networks via moment alignment.
Information Sciences, 483:174 – 191.
-  Zhu, L., Ding, W., Zhu, J., Xu, P., Liu, Y., Yan, M., and Zhang, J. (2022).
Multisource wasserstein adaptation coding network for eeg emotion recognition.
Biomedical Signal Processing and Control, 76:103687.