Domain Adaptation with Asymmetrically Relaxed Distribution Alignment

Yifan Wu, Ezra Winston, Divyansh Kaushik, Zachary Lipton

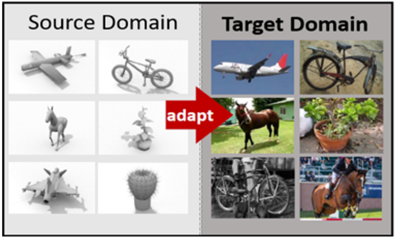
Carnegie Mellon University

ICML 2019

Background - Unsupervised Domain Adaptation

Unsupervised Domain Adaptation:

- Labeled data from **source** domain: $\{(x_i, y_i)\}_{i=1,\ldots,n} \sim p_S \cdot p_{y|x}$.
- Unlabeled data from **target** domain: $\{x_i\}_{i=1,\ldots,m} \sim p_T$
- Goal: learn a good target domain classifier $\hat{y}_x = \mathrm{argmax}_y\, p_{y|x}(y|x)$ for $x \sim p_T$.



Background - Domain Adversarial Training

Domain Adversarial Training (Ganin et al., 2016):

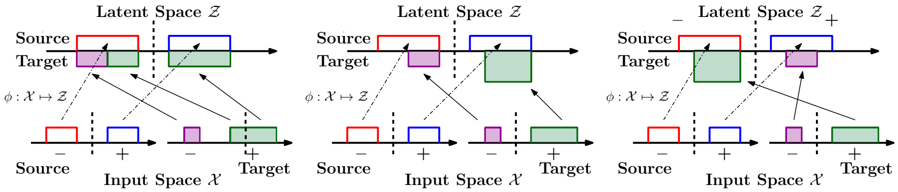
- Learn a predictor $\hat{y}_x = h(\phi(x))$ by optimizing:

$$\min_{\phi,h} \underbrace{\boxed{\mathcal{E}_S(\phi, h)}}_{\text{source domain prediction error}} + \underbrace{\boxed{\lambda D(p_S^\phi, p_T^\phi)}}_{\text{distance between feature distributions in the latent space}} + \Omega(\phi, h)\,.$$

Contribution

Problems with domain adversarial training:

- **Fails under label distribution shift.**
  - **We propose to use relaxed distribution alignment.**
- **Not clear how to prevent cross-label matching.**
  - **We drive a general error bound which explains under what assumptions this CANNOT happen.**



Relaxed Distances between Distributions

Our approach: replace the standard distance between distributions with a relaxed distance:

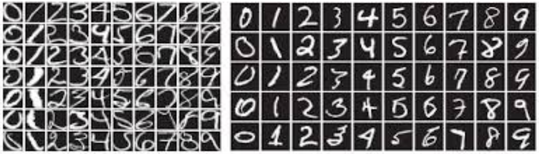
$$\min_{\phi,h} \; \mathcal{E}_S(\phi, h) + \lambda D_\beta(p_S^\phi, p_T^\phi) + \Omega(\phi, h)\,.$$

- Relaxed Jensen-Shannon Divergence:

$$D_{\bar{f}_\beta}(p, q) = \sup_{g:\mathcal{Z}\mapsto(0,1]} \mathbb{E}_{z\sim q}\left[\log \frac{g(z)}{2+\beta}\right] + \mathbb{E}_{z\sim p}\left[\log\left(1 - \frac{g(z)}{2+\beta}\right)\right]\,.$$

- Relaxation for any f-divergence, Wasserstein distance, etc.

Experiments - Handwritten Digits



(a) USPS (b) MNIST

target labels	[0-4] Shift	[5-9] Shift	[0-9] No-Shift
Source	74.3±1.0	59.5±3.0	66.7±2.1
DANN	50.0±1.9	28.2±2.8	78.5±1.6
fDANN-1	71.6±4.0	**67.5±2.3**	73.7±1.5
fDANN-2	74.3±2.5	61.9±2.9	72.6±0.9
fDANN-4	75.9±1.6	64.4±3.6	72.3±1.2
sDANN-1	71.6±3.7	49.1±6.3	81.0±1.3
sDANN-2	76.4±3.1	48.7±9.0	81.7±1.4
sDANN-4	**81.0±1.6**	60.8±7.5	**82.0±0.4**

Table: MNIST $\rightarrow$ USPS

target labels	[0-4] Shift	[5-9] Shift	[0-9] No-Shift
Source	69.4±2.3	30.3±2.8	49.4±2.1
DANN	57.6±1.1	37.1±3.5	**81.9±6.7**
fDANN-1	80.4±2.0	40.1±3.2	75.4±4.5
fDANN-2	**86.6±4.9**	41.7±6.6	70.0±3.3
fDANN-4	77.6±6.8	34.7±7.1	58.5±2.2
sDANN-1	68.2±2.7	**45.4±7.1**	78.8±5.3
sDANN-2	78.6±3.6	36.1±5.2	77.4±5.7
sDANN-4	83.5±2.7	41.1±6.6	75.6±6.9

Table: USPS $\rightarrow$ MNIST

Thank You

Poster 177

References

Ganin, Yaroslav, Ustinova, Evgeniya, Ajakan, Hana, Germain, Pascal, Larochelle, Hugo, Laviolette, François, Marchand, Mario, and Lempitsky, Victor. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030, 2016.