MULTI-SOURCE DOMAIN ADAPTATION WITH VON NEUMANN CONDITIONAL DIVERGENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

The similarity of feature representations plays a pivotal role in the success of domain adaptation and generalization. Feature similarity includes both the invariance of marginal distributions and the closeness of conditional distributions given the desired response y (e.g., class labels). In this work, we introduce the recently proposed von Neumann conditional divergence to improve the transferability across multiple domains. We show that this new divergence is differentiable and eligible to easily quantify the functional dependence between features and y.

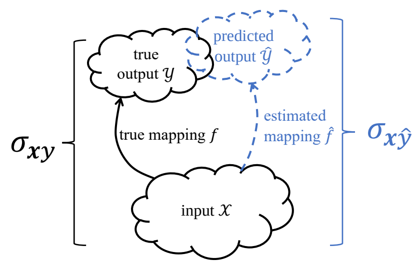

Figure 1: Geometry of loss $\mathcal{L}: \mathbb{S}^p_{++} \times \mathbb{S}^p_{++} \to \mathbb{R}_+$: $\sqrt{J_{vN}}$ searches an "optimal" predictor $\hat{f}$ minimizing the discrepancy between $\sigma_{\mathbf{x},f(\mathbf{x})}$ and $\sigma_{\mathbf{x},\hat{f}(\mathbf{x})}$. $\mathbb{S}^p_{++}$ is the set of $p \times p$ positive definite matrices.

Consider a network consisting of a feature extractor $f_\theta : \mathcal{X} \to \mathcal{T}$ (parametrized by θ) and a predictor $h_\varphi : \mathcal{T} \to \mathcal{Y}$ (parameterized by φ), the similarity of the latent representation $\mathbf{t}$ includes two aspects: the invariance of marginal distributions (i.e., $p(f_\theta(\mathbf{x}))$) across different domains and the functional closeness of using $\mathbf{t}$ to predict y. The predictive power of h_φ can be characterized by the conditional distribution $p(y|\mathbf{t})$. From an information-theoretic perspective, the conditional entropy $H(y|\mathbf{t}) = -\mathbb{E}(\log(p(y|t)))$ also measures the dependence between y and $\mathbf{t}$.

We first introduce the recently proposed von Neumann conditional divergence (Yu *et al.*, 2020) (denote by D_{vN}) and present its properties when used as a loss function to train deep neural networks. We further propose a simple yet effective approach to model $p(y|\mathbf{t})$ by D_{vN} in multi-source domain adaptation (for regression). Our method performs favorably in encouraging positive forward transfer. Our main contributions are summarized as follows: *(i)* We introduce the von Neumann conditional divergence D_{vN} to the problem of domain adaptation and generalization. *(ii)* We show the utility of D_{vN} in a standard unsupervised domain adaptation setup in which multiple source tasks are observed simultaneously (*a.k.a.*, multi-source domain adaptation), and develop a novel learning objective and a generalization bound. The obtained results confirm that the D_{vN}-based learning objective improves the unsupervised robustness when new tasks emerge without labeled data.

REFERENCES

Y. Yu, A. Shaker, F. Alesiani, and J. C. Principe. Measuring the discrepancy between conditional distributions: Methods, properties and applications. In *IJCAI*, pages 2777–2784, 2020.