Quantifying and Improving Transferability in Domain Generalization

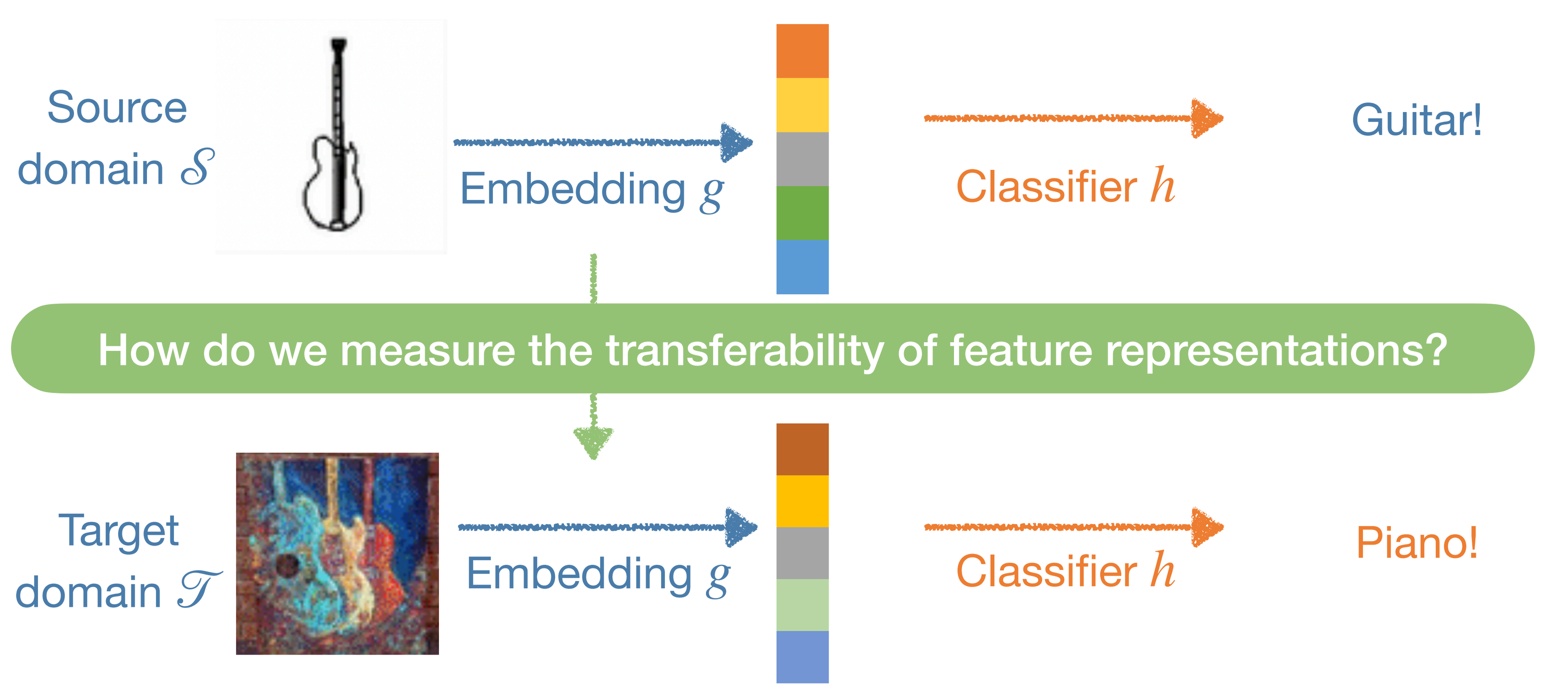
Guojun Zhang, Han Zhao, Yaoliang Yu and Pascal Poupart

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN · VECTOR INSTITUTE · NEURAL INFORMATION PROCESSING SYSTEMS

Main result

We propose an evaluation metric for measuring the transferability in domain generalization. Based on this metric, we design algorithms to further improve out-of-domain generalization over SoTA methods.

Transfer learning

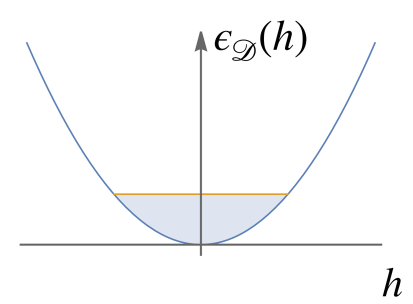


Transferable:
Every near-optimal source classifier is also *near-optimal* on the target

$$\mathrm{argmin}(\epsilon_{\mathscr{D}}, \delta) := \{h \in \mathscr{H}, \epsilon_{\mathscr{D}}(h) \le \inf_{h\in\mathscr{H}} \epsilon_{\mathscr{D}}(h) + \delta\}$$ **δ-minimal set**

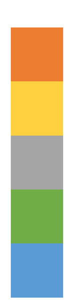
Definition: $\mathscr{S}$ is $(\delta_{\mathscr{S}}, \delta_{\mathscr{T}})$-transferable to $\mathscr{T}$ if

$$\mathrm{argmin}(\epsilon_{\mathscr{S}}, \delta_{\mathscr{S}}) \subseteq \mathrm{argmin}(\epsilon_{\mathscr{T}}, \delta_{\mathscr{T}})$$



Use Transfer Measures to quantify transferability

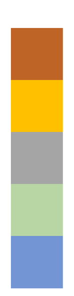
Source domain $\mathscr{S}$



Classifier h Excess risk $\mathrm{exc}_{\mathscr{S}}(h) := \epsilon_{\mathscr{S}}(h) - \inf_h \epsilon_{\mathscr{S}}(h)$

Transfer measure $\mathrm{T}_\Gamma(\mathscr{S}, \mathscr{T}) = \sup_{h\in\Gamma} |\mathrm{exc}_{\mathscr{S}}(h) - \mathrm{exc}_{\mathscr{T}}(h)|$

Target domain $\mathscr{T}$



Classifier h Excess risk $\mathrm{exc}_{\mathscr{T}}(h) := \epsilon_{\mathscr{T}}(h) - \inf_h \epsilon_{\mathscr{T}}(h)$

Transferable $\equiv$ Small transfer measure, if $\Gamma = \mathrm{argmin}(\epsilon_{\mathscr{S}}, \delta)$.

Theoretical results

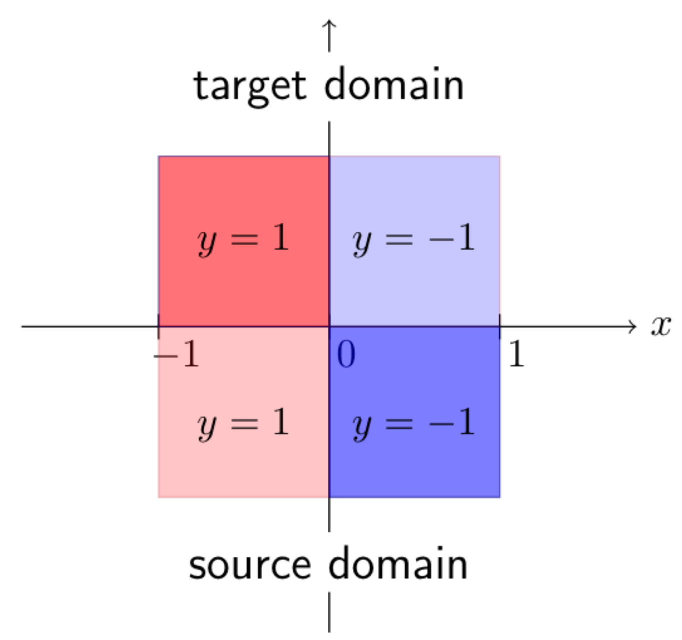
Generalization bound (tighter than the one using $\mathscr{H}$-divergence)

$$\epsilon_{\mathscr{T}}(h) \le \epsilon_{\mathscr{S}}(h) + \mathrm{T}_\Gamma(\mathscr{S}, \mathscr{T}) + \epsilon^*_{\mathscr{T}} - \epsilon^*_{\mathscr{S}}, \quad \epsilon^*_{\mathscr{D}} = \inf_{h\in\mathscr{H}} \epsilon_{\mathscr{D}}(h)$$

If the optimal errors are zero, transfer measure with 0-1 loss is equivalent to total variation when $\Gamma = \mathscr{H}_t$ includes all binary classifiers

Total variation $$d_{\mathrm{TV}}(\mathscr{S}, \mathscr{T}) = \sum_y \int |p_{\mathscr{S}}(x,y) - p_{\mathscr{T}}(x,y)|\, dx$$

Two domains can still be transferable even if the joint distributions are dissimilar



Colors: intensity of the probability mass

There exists a common optimal classifier

Transferability is more general than matching joint distributions

Invariance principles (error depends on both the marginal and conditional distributions)

- Transferability: invariance of excess risks *(joint)*
- H-divergence: invariance of feature distributions $p(z)$ *(marginal)*
- Invariant risk minimization (IRM): invariance of the optimal predictors *(conditional)*
- Generalized label shift: invariance of feature distributions within the same class $p(z\,|\,y)$ *(conditional)*

How to Compute Transfer Measures?

We make the following approximations

- We use surrogate loss to approximate 0-1 loss
- We use the final loss after training to approximate the optimal loss
- We use a neighborhood in the parameter space to approximate the minimal set
- The resulting transfer measure is a *lower bound*, but we can still use it to refute transferability

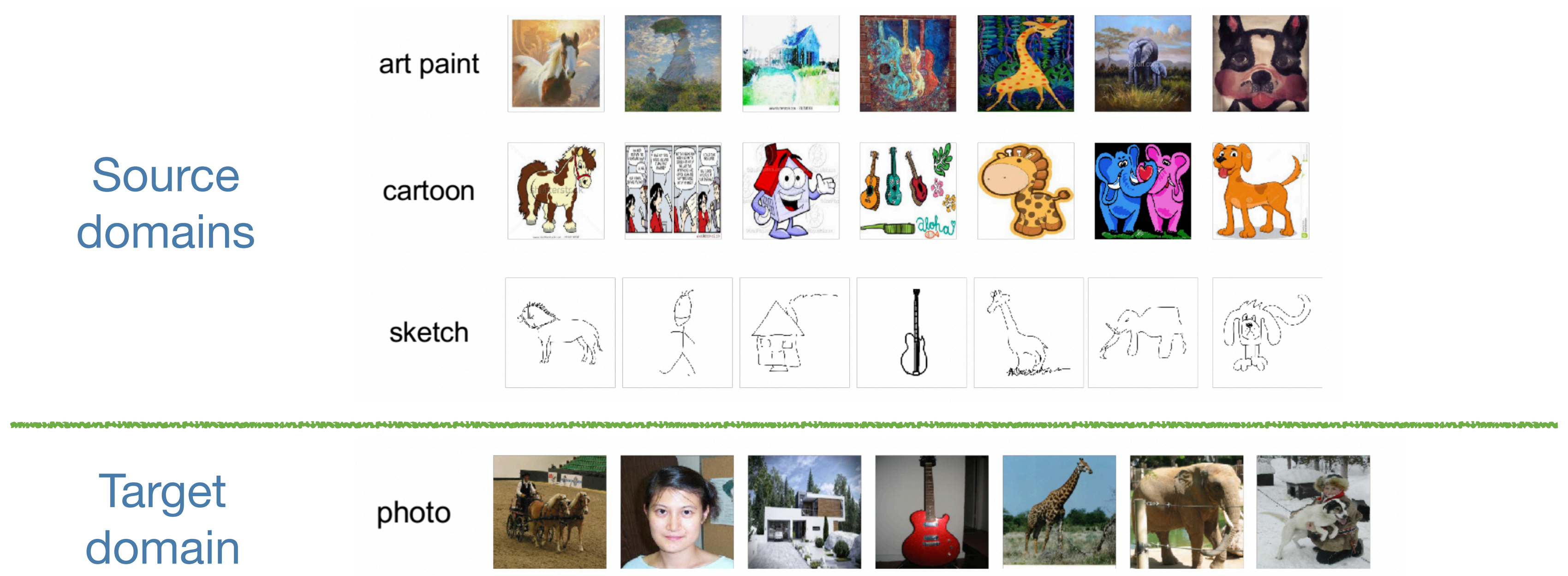
Final result as a lower bound of transfer measure $$\sup_{\|\theta - \widehat{\theta^*}\| \le \delta} \epsilon_{\mathscr{T}}(h) - \epsilon_{\mathscr{S}}(h) - \epsilon_{\mathscr{T}}(\widehat{h^*})$$

θ: parameter of classifier $\widehat{h^*}$ / $\widehat{\theta^*}$: the learned classifier after training

If the lower bound is large, then it is not transferable

Algorithms

Domain Generalization (DG) learns feature embeddings and classifiers from several source domains



Suppose we have n source domains: $\mathscr{S}_1, \ldots, \mathscr{S}_n$ and a target domain $\mathscr{T} = \mathscr{S}_0$

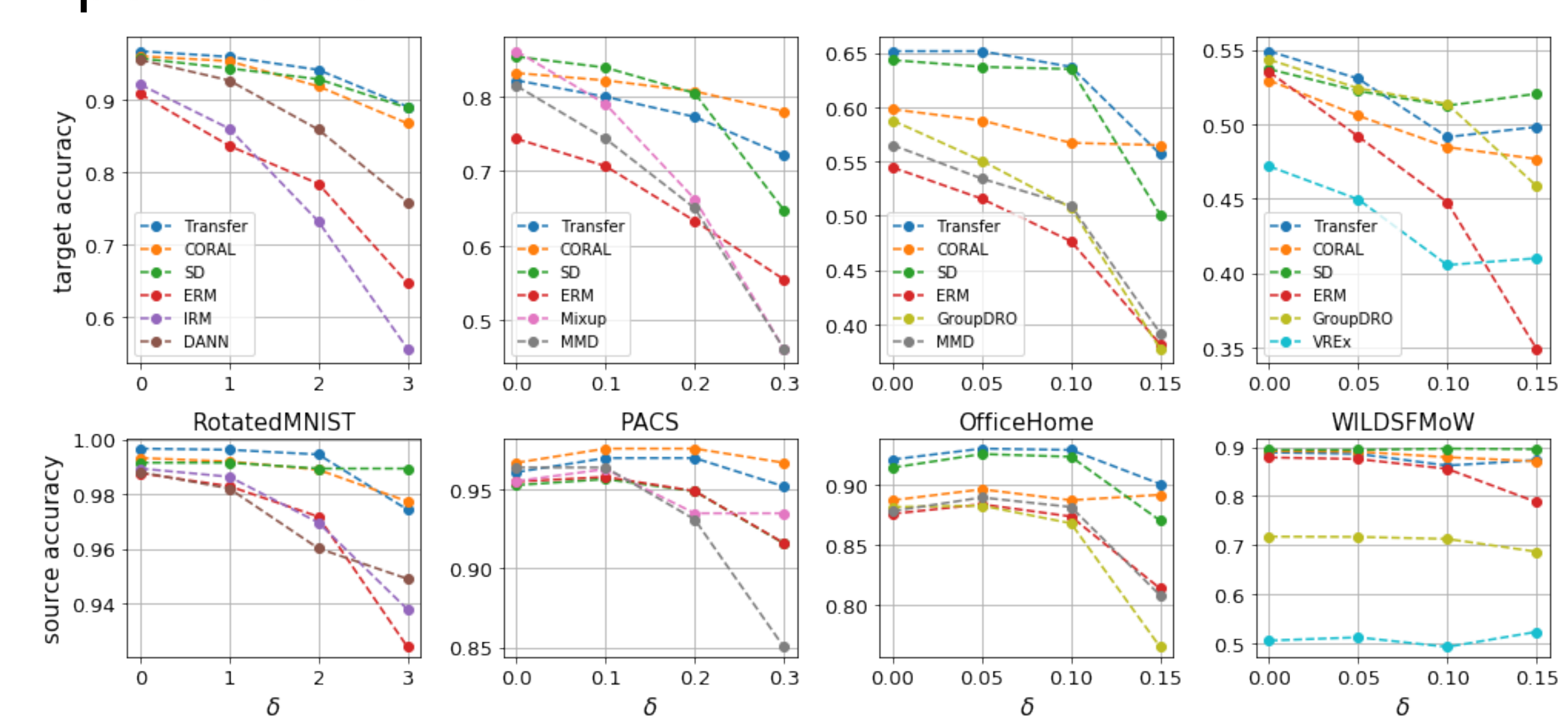
Evaluate transferability $$\max_{i,j\in[0,n]} \max_{\|h - \widehat{h^*}\| \le \delta} \mathrm{exc}_{\mathscr{S}_i}(h \circ g) - \mathrm{exc}_{\mathscr{S}_j}(h \circ g)$$

Improve transferability

$$\min_{g,h} \frac{1}{n}\sum_{i=1}^{n} \epsilon_{\mathscr{S}_i}(h \circ g) + \max_{i,j\in[1,n]} \max_{\|h' - h\| \le \delta} \mathrm{exc}_{\mathscr{S}_i}(h' \circ g) - \mathrm{exc}_{\mathscr{S}_j}(h' \circ g)$$

ERM adversarial training

Experiments



- Many popular algorithms are not learning transferable features: a near-optimal source classifier *is not* near-optimal on the target
- Our Transfer algorithm learns more transferable features than ERM, DANN, IRM, GroupDRO, etc.
- Best performers: Transfer (ours), CORrelation ALignment (CORAL), Spectral Decomposition (SD)

References

- Arjovsky et al. "Invariant risk minimization." *arXiv preprint arXiv:1907.02893* (2019).
- Ben-David et al, "A theory of learning from different domains," Machine Learning, 2010
- Koltchinskii, "Rademacher complexities and bounding the excess risk in active learning," JMLR 2010
- Tachet des Combes et al. "Domain adaptation with conditional distribution matching and generalized label shift." NeurIPS 2020

Full talk: https://www.youtube.com/watch?v=Ce3PyHA54GI

Code: https://github.com/Gordon-Guojun-Zhang/Transferability-NeurIPS2021