Revisiting (ϵ, γ, τ)-similarity learning for domain adaptation

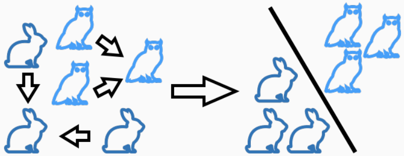
Sofien Dhouib [1] **Ievgen Redko** [2]

[1]CREATIS laboratory, INSA Lyon, University of Lyon

[2]Hubert Curien laboratory, University Jean Monnet of Saint-Etienne

Similarity learning

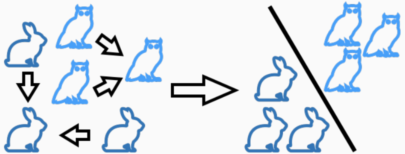
Learn a similarity function tailored to an observed data sample



Context & Goal

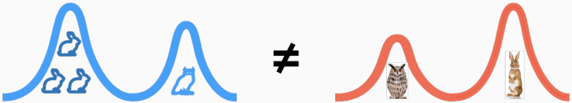
Similarity learning

Learn a similarity function tailored to an observed data sample



Goal

Analyze similarity learning in domain adaptation context



Labeled source sample $S \sim \mathcal{S}$ **Unlabeled target** sample $T \sim \mathcal{T}$

same deterministic labeling function

What we know already (Balcan et al. 2008)

Definition

$\boldsymbol{K}$ is (ϵ, γ, τ)-good similarity for $\mathcal{S}$ if

- $(\mathbf{1} - \epsilon)$ fraction of instances are on average **more similar** to landmarks with the **same label** by a margin γ **at least**
- **fraction of landmark** instances $\geq \boldsymbol{\tau}$

What we know already (Balcan et al. 2008)

Definition

K is (ϵ, γ, τ)-good similarity for $\mathcal{S}$ if

- **$(1-\epsilon)$** fraction of instances are on average **more similar** to landmarks with the **same label** by a margin γ **at least**
- **fraction of landmark** instances $\geq \boldsymbol{\tau}$

Theorem

*If **K** is (ϵ, γ, τ)-good for $\mathcal{S}$ then one can* **draw** $\{x_1, ..., x_L\}$ *from $\mathcal{S}$ and* **build a mapping** $\phi : x \mapsto (K(x, x_1), ..., K(x, x_L))$ *that makes it* **linearly separable** *with a* **large margin**

✓ **Generalization of the kernel trick!**

✓ **Several algorithms that minimize ϵ!**

Our contribution

Idea

Introduce (ϵ, γ)-goodness for $(\mathcal{S}, \mathcal{R})$ with data $\sim \mathcal{S}$ and landmarks $\sim \mathcal{R}$ (potentially $\mathcal{R} \neq \mathcal{S}$)

Our contribution

Idea

Introduce (ϵ, γ)-goodness for $(\mathcal{S}, \mathcal{R})$ with data $\sim \mathcal{S}$ and landmarks $\sim \mathcal{R}$ (potentially $\mathcal{R} \neq \mathcal{S}$)

Theorem

If $\boldsymbol{K}$ is $(\boldsymbol{\epsilon}, \boldsymbol{\gamma})$-good for $(\mathcal{S}, \mathcal{R})$ and μ **dominates** $\mathcal{S}$ and $\mathcal{T}$ then $\boldsymbol{K}$ is $(\underline{\boldsymbol{\epsilon} + \boldsymbol{\epsilon}'}, \gamma)$-good for $(\mathcal{T}, \mathcal{R})$ with

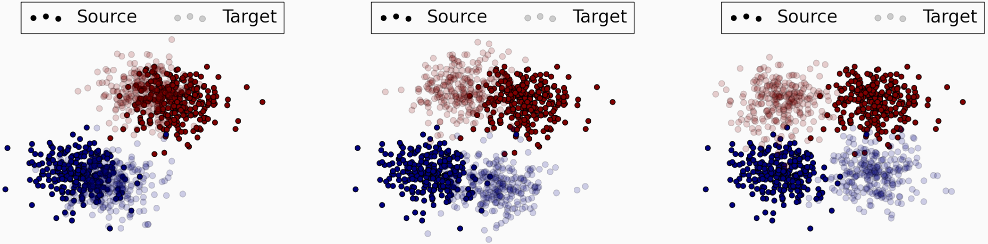
$$\epsilon' = \begin{matrix}\text{L}^1 \textbf{ distance} \\ \text{between } \mathcal{S} \text{ and } \mathcal{T}\end{matrix} \times \begin{matrix}\textbf{Worst margin} \text{ achieved} \\ \text{by } \mathbf{K} \text{ on } x{\sim}\mu\end{matrix}, \text{ if } \mathcal{T} \not\ll \mathcal{S}$$

and

$$\epsilon' = \begin{matrix}\chi^2 \textbf{ distance} \\ \text{between } \mathcal{S} \text{ and } \mathcal{T}\end{matrix} \times \begin{matrix}\textbf{Worst margin} \text{ achieved} \\ \text{by } \mathbf{K} \text{ on } x{\sim}\mathcal{S}\end{matrix} \times \begin{matrix}\epsilon \text{ on} \\ \text{source } \mathcal{S}\end{matrix}, \text{ if } \mathcal{T} \ll \mathcal{S}$$

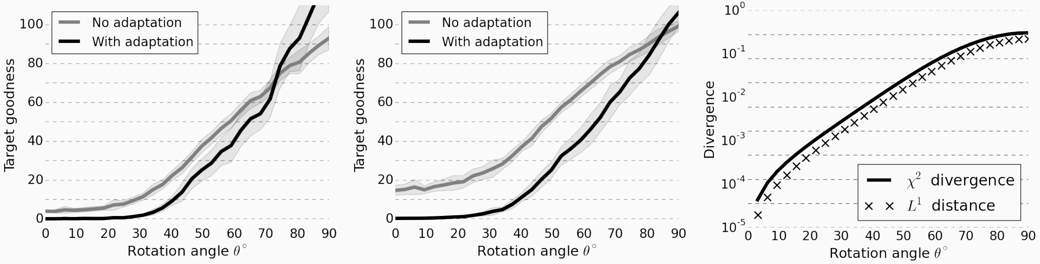
✓ **Multiplicative dependence of the target error on the source one!**

Empirical evaluations



Generated data for (**left**) 30°, (**middle**) 60°, (**right**) 90° degrees rotation

Empirical evaluations



Results for (**left**) $\mathcal{T} \not\ll \mathcal{S}$, (**middle**) $\mathcal{T} \ll \mathcal{S}$ and (**right**) divergence evolution

For more details come visit our poster #152 !

(spoiler: post-doc position available)