Revisiting $(\epsilon, \gamma, \tau)$-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
Context & Goal

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 $S \sim S$
Unlabeled target sample $T \sim T$

same deterministic labeling function
What we know already (Balcan et al. 2008)

Definition

$K$ is $(\epsilon, \gamma, \tau)$-good similarity for $S$ if

- $(1 - \epsilon)$ fraction of instances are on average **more similar** to landmarks with the **same** label by a margin $\gamma$ at least

- fraction of **landmark** instances $\geq \tau$
What we know already (Balcan et al. 2008)

**Definition**

\( K \) is \((\epsilon, \gamma, \tau)\)-good similarity for \( S \) if

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

**Theorem**

If \( K \) is \((\epsilon, \gamma, \tau)\)-good for \( S \) then one can **draw** \( \{x_1, ..., x_L\} \) from \( S \) 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** \( \epsilon \)!
Our contribution

Idea

Introduce $(\epsilon, \gamma)$-goodness for $(S, \mathcal{R})$ with data $\sim S$ and landmarks $\sim \mathcal{R}$ (potentially $\mathcal{R} \neq S$)
Our contribution

Idea

Introduce $(\epsilon, \gamma)$-goodness for $(S, R)$ with data $\sim S$ and landmarks $\sim R$ (potentially $R \neq S$)

Theorem

If $K$ is $(\epsilon, \gamma)$-good for $(S, R)$ and $\mu$ dominates $S$ and $T$ then $K$ is $(\epsilon + \epsilon', \gamma)$-good for $(T, R)$ with

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

and

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

✓ 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) $T \ll S$, (middle) $T \ll S$ and (right) divergence evolution
For more details come visit our poster #152!
(spoiler: post-doc position available)