Large Dimensional Analysis of LS-SVM Transfer Learning: Application to POLSAR Classification

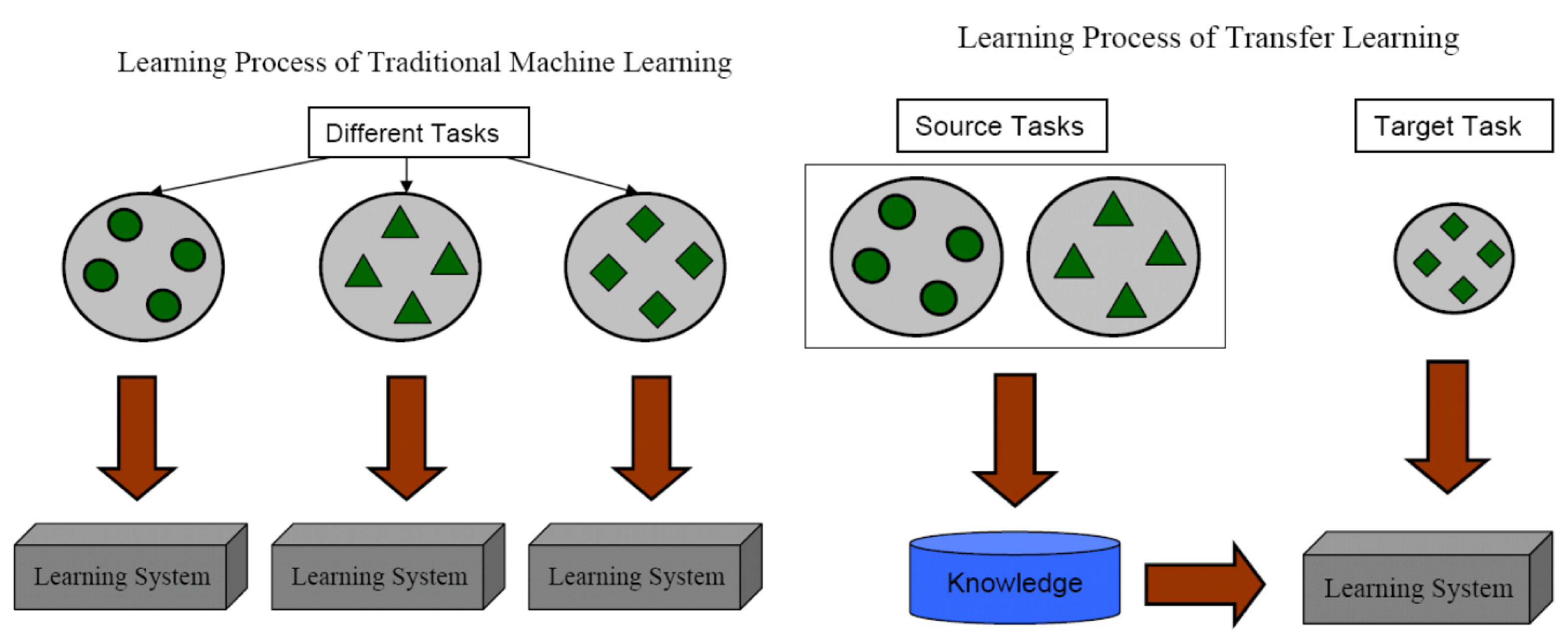
Cyprien DOZ[1], Romain COUILLET[2], Chengfang REN[1], Jean-Philippe OVARLEZ[1,3]

[1]SONDRA, CentraleSupélec, France, [2]GIPSA-Lab, University Grenoble Alpes, France, [3]ONERA, Université Paris-Saclay, France

Abstract

- Analysis, Interpretation and Improvement of transfer learning [1] with Random Matrix Theory [2–4]
- Application to environmental monitoring : label optimization and performance guarantees

Transfer learning framework [1]

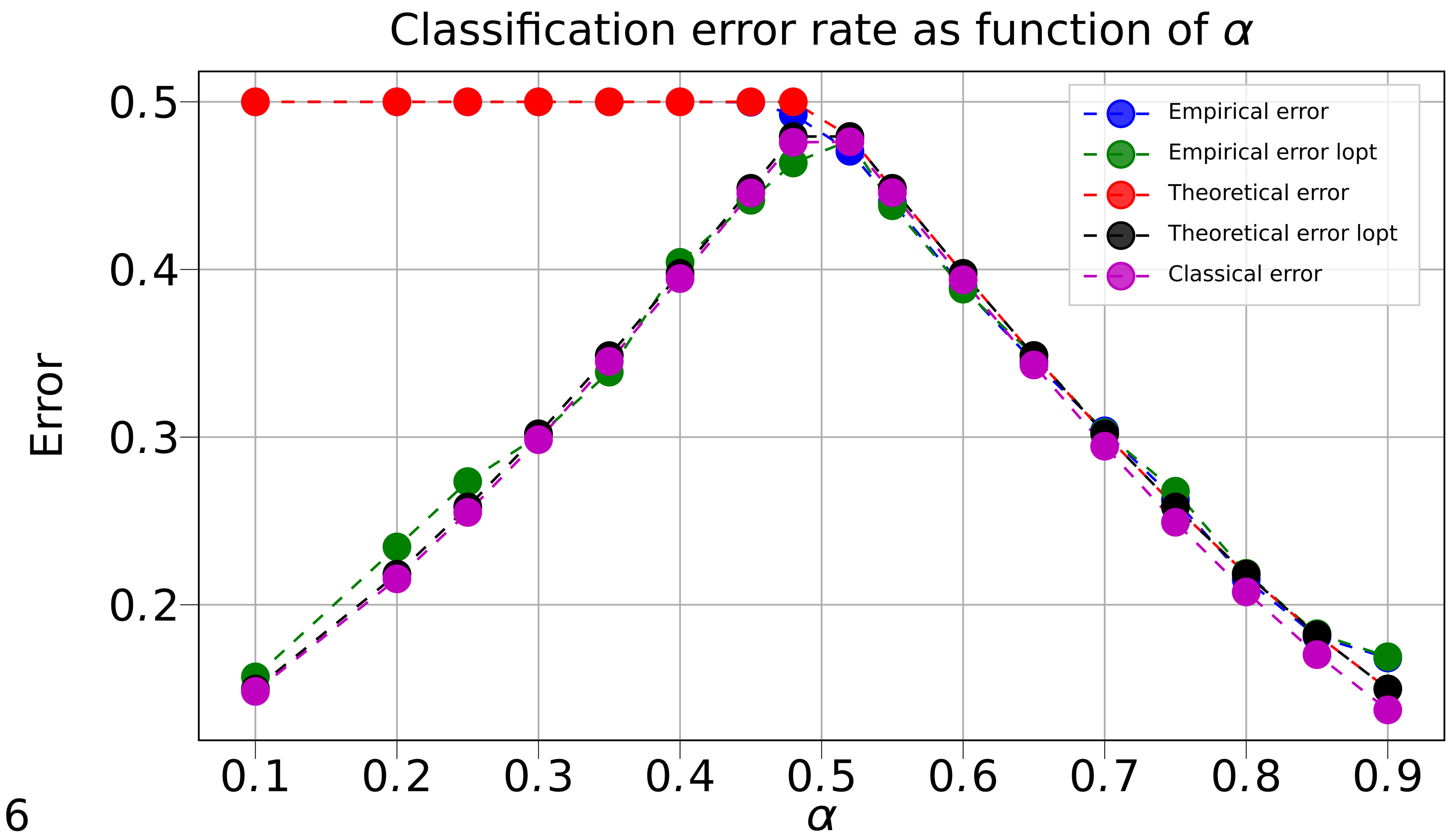


(a) Traditional Machine Learning (b) Transfer Learning

1. $[\mathbf{x}_1^T, \ldots, \mathbf{x}_{n_T}^T]$: target data (annotated) **insufficient**.

   ➥ failing supervised learning
2. $[\mathbf{x}_1^T, \ldots, \mathbf{x}_{n_T}^T] \leftarrow [\mathbf{x}_1^S, \ldots, \mathbf{x}_{n_S}^S]$: source data **similar**
3. new learning set: $[\mathbf{x}_1, \ldots, \ldots \mathbf{x}_n]$, $n = n_S + n_T$

Labels optimization



Classification performance on simulated data, related by parameter α, s.t. $\mathbf{C}_{T_a} = \alpha \mathbf{C}_{S_a} + (1-\alpha)\mathbf{C}_{S_b}$, for various label strategies; $p = 512$, $n_{S_1} = n_{S_2} = 508$, $n_{T_1} = n_{T_2} = 4$, polynomial kernel f.

Avoid negative transfer. Optimal label benefit from sources.

Model

- Gaussian mixture $[\mathbf{x}_1, \ldots, \ldots, \mathbf{x}_n]$ in 4 classes $\mathcal{C}_{S_1}, \mathcal{C}_{T_1}, \mathcal{C}_{S_2}, \mathcal{C}_{T_2}$

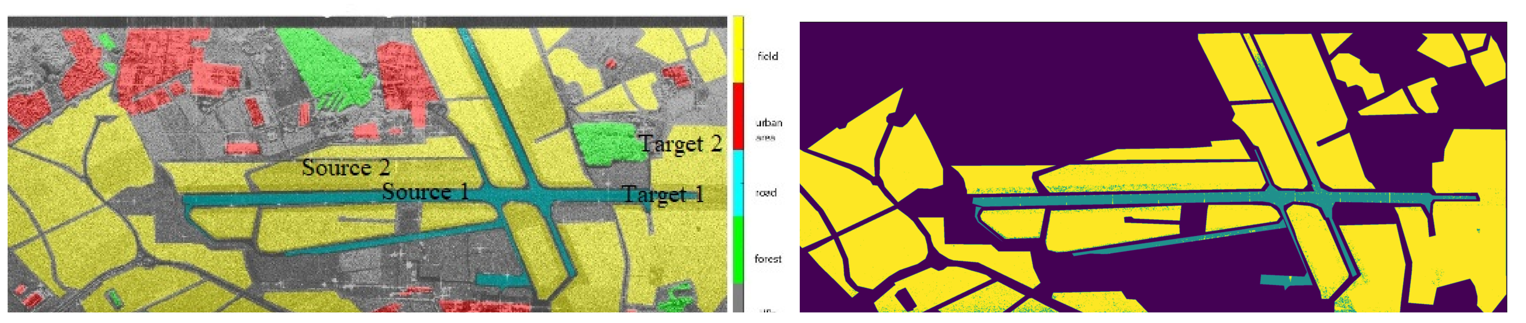
$$\mathbf{x}_i \in \mathcal{C}_a \longleftrightarrow \mathbf{x}_i \sim N(\mathbf{0}, \mathbf{C}_a)\,, \mathbf{C}_a \in \mathbb{R}^{p \times p}$$

for $a \in \{S_1, T_1, T_2, S_2\}$, $n_a = |\mathcal{C}_a|$, $c_a = n_a/n$.

- High dimension hypotheses: $n, p \to \infty$, $p/n < \infty$
- Non trivial classification: $\|\mathbf{C}_a\| = O(1)$ and $\operatorname{tr}(\mathbf{C}_b - \mathbf{C}_a) = O(\sqrt{p})$.

Geographical classification

Classification between two target areas (Target 1 and Target 2) of a polarimetric SAR image of Bretigny provided by ONERA using source areas (Source 1 and Source 2). $p = 54$, $n = 2000$ (1000 per class).



Left : Groundtruth. Right : LS-SVM Classification.

Large Dimensional Asymptotic

Theorem 1 (Gaussian Approximation). *Let $\mathbf{x} \in \mathcal{C}_a$, $a \in \{T_1, T_2\}$ and $g(\mathbf{x})$ LS-SVM decision function. Then,*

$$n \mathrm{V}_a^{-\frac{1}{2}} \left(g(\mathbf{x} \mid \mathbf{x} \in \mathcal{C}_a) - \mathrm{E}_a\right) \xrightarrow{d} \mathcal{N}(0,1), \quad (1)$$

with $$\mathrm{E}_a = \gamma \boldsymbol{\ell}^T \mathbf{P}_c \mathfrak{m}_a + *, \quad (2)$$

$$\mathfrak{m}_a = \frac{f''(\tau)}{p} \mathbf{t}\, t_a + 2 \frac{f''(\tau)}{p^2} \mathbf{t}_{\mathbf{C}_a},$$

and

$$\mathrm{V}_a = 2\gamma^2 \boldsymbol{\ell}^T \left(\mathbf{P}_c \mathbf{W}_1^a \mathbf{P}_c + \mathbf{J}^T \mathbf{P} \mathbf{W}_2^a \mathbf{P} \mathbf{J}\right) \boldsymbol{\ell}, \quad (3)$$

where $$\mathbf{W}_1^a = p^{-3} \left(f''(\tau)\right)^2 \mathbf{t}\mathbf{t}^T \operatorname{tr} \mathbf{C}_a^2,$$

$$\mathbf{W}_2^a = 2 p^{-2} n^{-2} \left(f'(\tau)\right)^2 \operatorname{diag}\left(\mathbf{J} \mathbf{t}_{\mathbf{C}_a}\right),$$

$\mathbf{t} = [t_1, \ldots, t_k]^T$, $\mathbf{t}_{\mathbf{C}_a} \triangleq [\operatorname{tr}(\mathbf{C}_a \mathbf{C}_1), \ldots, \operatorname{tr}(\mathbf{C}_a \mathbf{C}_k)]^T$ $\mathbf{C}^\circ \triangleq \sum_{a=1}^k \frac{n_a}{n} \mathbf{C}_a$ and $\tau \triangleq \frac{2}{p} \operatorname{tr} \mathbf{C}^\circ > 0$, $t_a \triangleq p^{-1/2} \operatorname{tr}(\mathbf{C}_a - \mathbf{C}^\circ)$, $\mathbf{P} \triangleq \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T$, $\mathbf{P}_c \triangleq \frac{1}{n} \mathbf{J}^T \mathbf{P} \mathbf{J} = \operatorname{diag}(\mathbf{c}) - \mathbf{c}\,\mathbf{c}^T$.

Analysis and Improvement tracks

- LS-SVM Gaussian Approximation

  ➥ Probability error of classification, classification threshold
- Adaptability
  - Centered data (e.g. radar, SAR) $\Longrightarrow f'(\tau) = 0$

  ➥ Error minimization (labels)
  - Adaptation criteria on source data
    - ∗ Noise (different value of SNR)
    - ∗ Diversity (Multi-bands and Multi-looks)
    - ∗ Spatial Similarity

References

1. S. J. Pan et al., IEEE Transactions on Knowledge and Data Engineering **22**, 1345–1359 (2010).
2. R. Couillet et al., Electronic Journal of Statistics, 1393–1454 (2016).
3. Z. Liao et al., IEEE Transactions on Signal Processing **67**, 1065–1074 (2019).
4. M. Tiomoko et al., IEEE ICASSP'20) (2020).