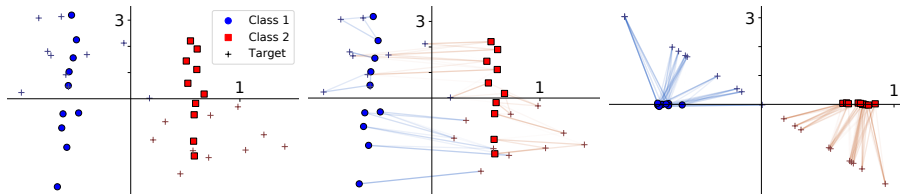


Metric Learning in Optimal Transport for Domain Adaptation

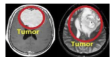
Tanguy Kerdoncuff, Rémi Emonet and Marc Sebban

Univ Lyon, UJM-Saint-Etienne, CNRS, Institut d'Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, SAINT-ETIENNE, France

10 juin 2020



Domain Adaptation (DA) : Intuition



Machine Learning System
to detect tumors

Labeled Data



HOW TO TRANSFER ?



Domain Adaptation

(Hot topic in Machine Learning)



Unlabeled Data



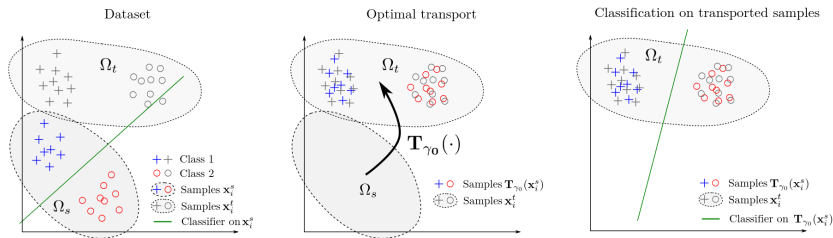
tumor or not tumor

Discrete Optimal Transport (OT)

Discrete Optimal Transport

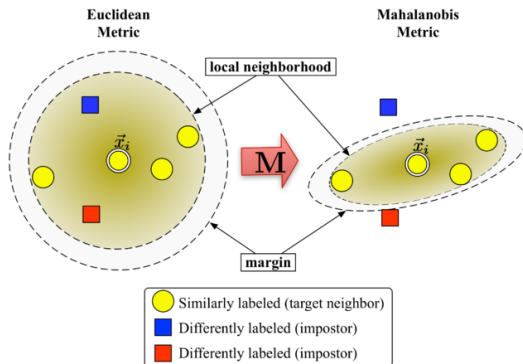
$$\mathcal{W}^2(\hat{\mu}_s, \hat{\mu}_t) = \min_{T \in \hat{\Pi}(\hat{\mu}_s, \hat{\mu}_t)} \sum_{i=1}^{m_s} \sum_{j=1}^{m_t} \|x_i^s - x_j^t\|_2^2 T_{ij} = \min_{T \in \hat{\Pi}(\hat{\mu}_s, \hat{\mu}_t)} \langle T, C^2 \rangle.$$

With $\hat{\Pi}(\hat{\mu}_s, \hat{\mu}_t) = \{T \in \mathbb{R}_+^{m_s \times m_t} \mid T\mathbf{1}_{m_t} = \hat{\mu}_s, T^T\mathbf{1}_{m_s} = \hat{\mu}_t\}$



Optimal Transport for DA [Courty et al., 2017]

Metric learning



LMNN [Weinberger et al., 2006]

Mahalanobis metric

- $C_M^2(x, x') = (x - x')M(x - x')^T = \|Lx - Lx'\|_2^2$ with $L^T L = M$

1 Main contributions

- Bound on the Target Risk
- PCA and Wasserstein Distance
- Algorithm : Metric Learning for Optimal Transport

Classical form of Target Risk Bound [Ben-David et al., 2007]

$$\begin{aligned} \text{target risk} &\leq \text{source risk} \\ &\quad + \textit{distance}(\mu_s, \mu_t) \\ &\quad + \lambda \text{ (error of the best classifier)} \end{aligned} \tag{1}$$

Bound on the Target Risk

Classical form of Target Risk Bound [Ben-David et al., 2007]

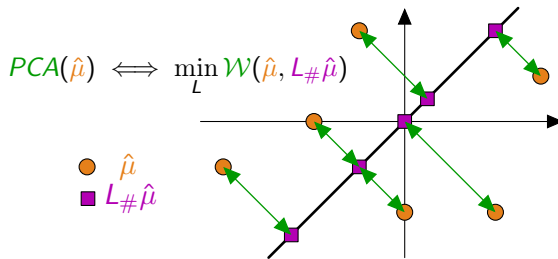
$$\begin{aligned} \text{target risk} &\leq \text{source risk} \\ &\quad + \textit{distance}(\mu_s, \mu_t) \\ &\quad + \lambda (\text{error of the best classifier}) \end{aligned} \tag{1}$$

Our Target Risk Bound

$$\begin{aligned} \text{target risk} &\leq \text{source risk} + \mathcal{W}(L_{s\#}\hat{\mu}_s, L_{t\#}\hat{\mu}_t) \\ &\quad + \mathcal{W}(\hat{\mu}_s, L_{s\#}\hat{\mu}_s) + \mathcal{W}(L_{t\#}\hat{\mu}_t, \hat{\mu}_t) \\ &\quad + \mathcal{W}(\mu_s, \hat{\mu}_s) + \mathcal{W}(\hat{\mu}_t, \mu_t) + \lambda \end{aligned} \tag{2}$$

Our Target Risk Bound

$$\begin{aligned} \text{target risk} \leq & \text{source risk} + \mathcal{W}(L_{s\#}\hat{\mu}_s, L_{t\#}\hat{\mu}_t) \\ & + \mathcal{W}(\hat{\mu}_s, L_{s\#}\hat{\mu}_s) + \mathcal{W}(\hat{\mu}_t, L_{t\#}\hat{\mu}_t) \\ & + \mathcal{W}(\mu_s, \hat{\mu}_s) + \mathcal{W}(\hat{\mu}_t, \mu_t) + \lambda \end{aligned} \quad (3)$$



Metric Learning for Optimal Transport (MLOT)

Input : η (gradient step) X_s X_t Y_s

1: $V_s = PCA(X_s)$, $V_t = PCA(X_t)$

2: $L_s = V_s^T V_s$, $L_t = V_t^T V_t$

3: **for** $i = 1$ **to** N **do**

4: $\gamma = \text{Solve } OT(C(L_s, L_t))$

5: $L_s = L_s - \eta \nabla_{L_s} (\langle \gamma, C^2(L_s, L_t) \rangle + \lambda_l \Omega_l(L_s))$

6: **end for**

7: $\tilde{X}_s = \gamma L_t X_t$

8: **return** classifier((\tilde{X}_s, Y_s) , X_t)

Metric Learning for Optimal Transport (MLOT)

Input : η (gradient step) X_s X_t Y_s

- 1: $V_s = \text{PCA}(X_s)$, $V_t = \text{PCA}(X_t)$
- 2: $L_s = V_s^T V_s$, $L_t = V_t^T V_t$
- 3: **for** $i = 1$ **to** N **do**
- 4: $\gamma = \text{Solve } OT(C(L_s, L_t))$
- 5: $L_s = L_s - \eta \nabla_{L_s} (\langle \gamma, C^2(L_s, L_t) \rangle + \lambda_I \Omega_I(L_s))$
- 6: **end for**
- 7: $\tilde{X}_s = \gamma L_t X_t$
- 8: **return** classifier((\tilde{X}_s, Y_s) , X_t)

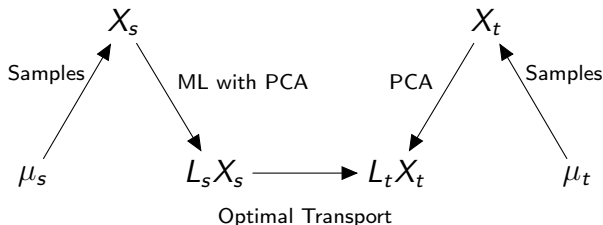


Illustration of the usefulness of Metric Learning

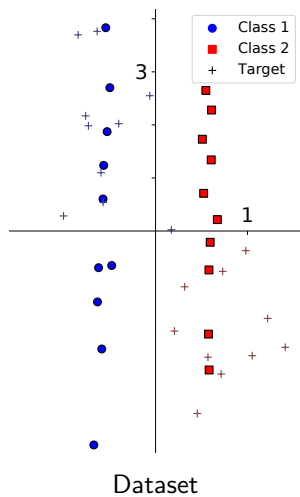


Illustration of the usefulness of Metric Learning

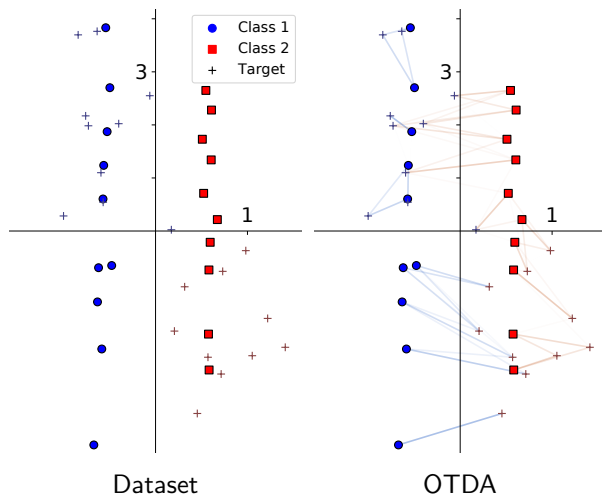
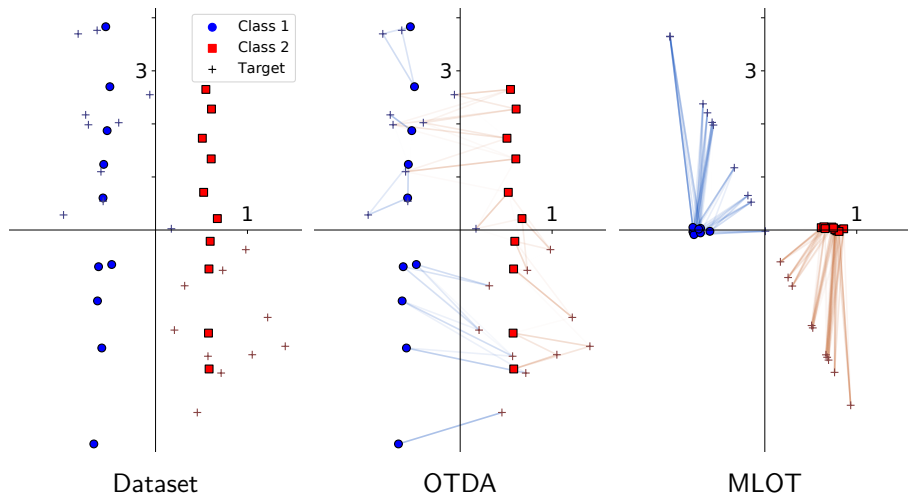


Illustration of the usefulness of Metric Learning





Ben-David, S., Blitzer, J., Crammer, K., and Pereira, F. (2007).

Analysis of representations for domain adaptation.

In *NIPS*.



Courty, N., Flamary, R., Tuia, D., and Rakotomamonjy, A. (2017).

Optimal transport for domain adaptation.

PAMI.



Weinberger, K. Q., Blitzer, J., and Saul, L. K. (2006).

Distance metric learning for large margin nearest neighbor classification.

In *Advances in neural information processing systems*.

- To use the Optimal Transition of the presented work see my github :
<https://github.com/Hv0nnus/TransitionPDF>