Transport Optimal pour l'adaptation de Domaines à travers des mélanges Gaussiens

Eduardo MONTESUMA Fred NGOLÈ Antoine SOULOUMIAC

Université Paris-Saclay, CEA, List, F-91120 Palaiseau France

Sommaire

Introduction

Adaptation de domaines

Domaine source

Données étiquetées: $\mathbf{x}_i^{(P)} \sim P$
et $y_i^{(P)} = h_0(\mathbf{x}_i^{(P)})$

Domaine cible

Données non-étiquetées: $\mathbf{x}_j^{(Q)} \sim Q$

Minimisation du risque empirique,

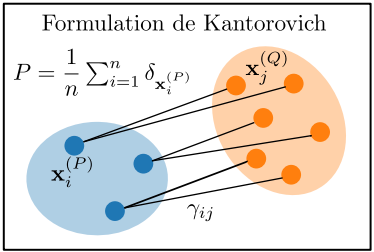
$$h^\star = \underset{h \in \mathcal{H}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(h(\mathbf{x}_i^{(P)}), y_i^{(P)})$$

$h^\star$ généralise pour des échantillons $\mathbf{x} \sim P$

Défi
Comment apprendre un classifieur sur Q
sans ses étiquettes?

⋆ Montesuma, Mboula, Souloumiac, Recent Advances on Optimal Transport for machine learning. arXiv:2306.16156, 2023

Transport optimal Empirique



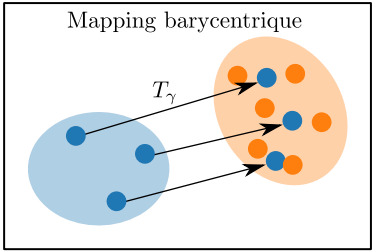
$$\gamma^\star = \underset{\gamma \in \Gamma(P,Q)}{\operatorname{argmin}} \sum_{i=1}^{n}\sum_{j=1}^{m} \gamma_{ij} C_{ij}$$

$$\mathcal{W}_2(P,Q)^2 = \min_{\gamma \in \Gamma(P,Q)} \sum_{i=1}^{n}\sum_{j=1}^{m} \gamma_{ij} C_{ij}$$

$$C_{ij} = \|\mathbf{x}_i^{(P)} - \mathbf{x}_j^{(Q)}\|_2^2$$

Dans le cas discret, il n'est pas evident comment avoir un mapping de Monge. On a, néanmoins, une notion de mapping définit à travers de γ:

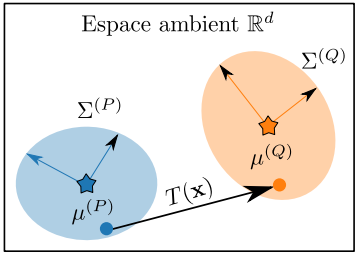
$$T_\gamma(\mathbf{x}_i^{(P)}) = \underset{\mathbf{x}\in\mathbb{R}^d}{\operatorname{argmin}} \sum_{j=1}^{m} \gamma_{ij} c(\mathbf{x}, \mathbf{x}_j^{(Q)})$$



† Peyré, Cuturi, Computational Optimal Transport: with applications to data science. Foundations and Trends in Machine Learning

⋆ Montesuma, Mboula, Souloumiac, Recent Advances on Optimal Transport for machine learning. arXiv:2306.16156, 2023

Transport optimal Gaussien



Si,

$P = \mathcal{N}(\mu^{(P)}, \Sigma^{(P)})$ et $Q = \mathcal{N}(\mu^{(Q)}, \Sigma^{(Q)})$

Alors,

$T(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$ (mapping affine)

où,

$\mathbf{A} = (\Sigma^{(P)})^{-1/2}((\Sigma^{(P)})^{1/2}\Sigma^{(Q)}(\Sigma^{(P)})^{1/2})^{1/2}(\Sigma^{(P)})^{-1/2}$

$\mathbf{b} = \mu^{(Q)} - \mathbf{A}\mu^{(P)}$

Distance de Wasserstein

$$\mathcal{W}_2(P,Q)^2 = \|\mu^{(P)} - \mu^{(Q)}\|_2^2 + \mathcal{B}(\Sigma^{(P)}, \Sigma^{(Q)})^2$$

où $\mathcal{B}(\Sigma^{(P)}, \Sigma^{(Q)})$ est la distance de Bures,

$$\sqrt{\mathrm{Tr}\left(\Sigma^{(P)} + \Sigma^{(Q)} - 2((\Sigma^{(P)})^{1/2}\Sigma^{(Q)}(\Sigma^{(P)})^{1/2})^{1/2}\right)}$$

Avantage:

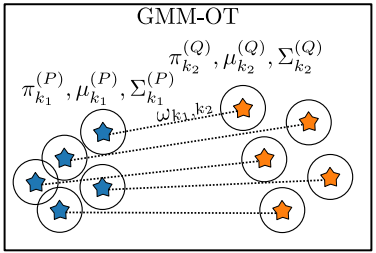
$(\mu^{(P)}, \mu^{(Q)}, \Sigma^{(P)}, \Sigma^{(Q)})$ peuvent être estimées à partir de $\{\mathbf{x}_i^{(P)}\}_{i=1}^n$ et $\{\mathbf{x}_j^{(Q)}\}_{j=1}^m$

ces estimations définissent T et $\mathcal{W}_2$ à partir des échantillons de P et Q

† Peyré, Cuturi, Computational Optimal Transport: with applications to data science. Foundations and Trends in Machine Learning

⋆ Montesuma, Mboula, Souloumiac, Recent Advances on Optimal Transport for machine learning. arXiv:2306.16156, 2023

Transport optimal entre mélanges Gaussiens



Un mélange Gaussien est une mésure sous la forme,

$$P = \sum_{k_1=1}^{K} \pi_{k_1}^{(P)} P_{k_1} \qquad P_{k_1} = \mathcal{N}(\mu_{k_1}^{(P)}, \Sigma_{k_1}^{(P)})$$

Le problème GMM-OT est définit par,

$$\gamma^\star = \underset{\gamma \in \Gamma(P,Q) \cap \mathrm{GMM}_{2d}(+\infty)}{\operatorname{argmin}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} c(\mathbf{x}_1, \mathbf{x}_2) d\gamma(\mathbf{x}_1, \mathbf{x}_2)$$

Problème équivalent discret,

$$\omega^\star = \underset{\omega \in \Gamma(\pi^{(P)}, \pi^{(Q)})}{\operatorname{argmin}} \sum_{k_1=1}^{K_P} \sum_{k_2=1}^{K_Q} \omega_{k_1,k_2} \mathcal{W}_2(P_{k_1}, Q_{k_2})^2$$

$\gamma^\star$ et $\omega^\star$ sont liées:

$$\gamma(\mathbf{x}_1, \mathbf{x}_2) = \sum_{k_1=1}^{K_P} \sum_{k_2=1}^{K_Q} \omega_{ij} \phi(\mathbf{x}_1 | \mu_{k_1}^{(P)}, \Sigma_{k_1}^{(P)}) \delta(\mathbf{x}_2 - T_{k_1,k_2}(\mathbf{x}_1))$$

Remarque 1

$\gamma^\star$ est un plan de transport **entre échantillons**

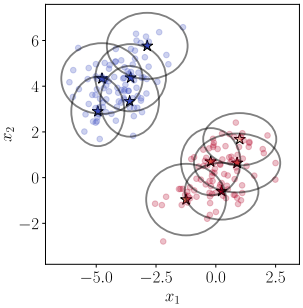
$\omega^\star$ est un plan de transport **entre composantes**

Remarque 2

GMM-OT correspond à un programme linéaire de $K_P \times K_Q$ variables.

⋆ Delon, Desolneux, A wasserstein-type distance in the space of gaussian mixture models. SIAM journal of Imaging Sciences, 2020

Mélanges Gaussiens Étiquettées



Sur un jeu de données labelisé $\{(\mathbf{x}_i^{(P)}, y_i^{(P)})\}_{i=1}^n$, on peut définir des étiquettes pour les composantes d'un GMM par la loi totale des probabilités,

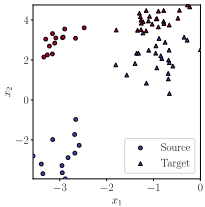
$$P(y|k) = \sum_{i=1}^n P_\theta(\mathbf{x}_i^{(P)}|k) P(y|\mathbf{x}_i^{(P)})$$
$$= \sum_{i=1}^n \left(\frac{p_k P_k(\mathbf{x}_i^{(P)})}{\sum_{k'} p_{k'} P_{k'}(\mathbf{x}_i^{(P)})} \right) P(y|\mathbf{x}_i^{(P)})$$

Basée sur $P(y|k)$, on peut construire un vécteur

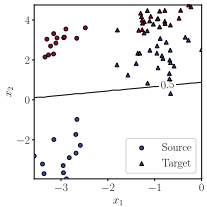
$$\mathbf{y}_k^{(P)} = (P(1|k), \cdots, P(n_c|k)) \in \Delta_{n_c}$$

La GMM $\{(\mu_k^{(P)}, \Sigma_k^{(P)})\}_{k=1}^K$ muni des $\{\tilde{\mathbf{y}}_k^{(P)}\}_{k=1}^K$ nous permet d'avoir un classifieur avec **Maximum a Posteriori** (MAP): $\hat{h}_{MAP}(\mathbf{x}) = \underset{j=1,\cdots,n_c}{\operatorname{argmin}} \sum_{k=1}^K P_\theta(\mathbf{x}|k) P_\theta(j|k)$.

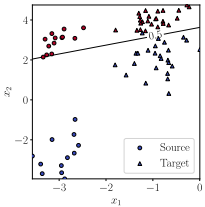
Transport optimal pour l'adaptation de domaines



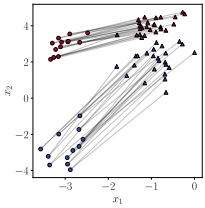
(a) Données de P et Q



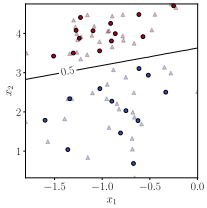
(b) Classifieur (source)



(c) Classifieur (cible)



(d) Plan de transport γ

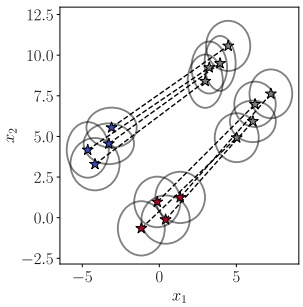


(e) Classifieur appris sur T_γ

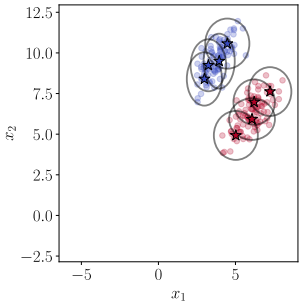
⋆ Courty, Flamary, Tuia, Rakotomamonjy, Optimal transport for domain adaptation, TPAMI'16

Contributions Méthodologiques

Propagation d'étiquettes



(a) Plan de transport ω



(b) Étiquettes propagées

Avec les étiquettes de P et ω, on peut estimer les étiquettes de Q par la formule*:

$$\tilde{\mathbf{y}}_{k_2}^{(Q)} = \frac{1}{\pi_{k_2}^{(Q)}} \sum_{k_1=1}^{K_P} \omega_{k_1,k_2} \tilde{\mathbf{y}}_{k_1}^{(P)}$$

Cela nous permet d'avoir une GMM étiquettée sur la cible, et donc un classifieur (e.g., par MAP)

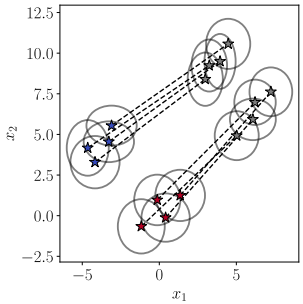
Remarque 3

GMM-OTDA$_M$: $\hat{h}_{MAP}$

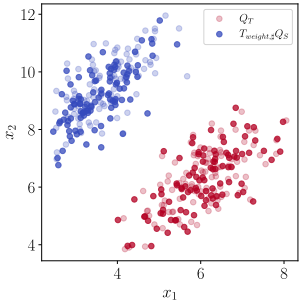
GMM-OTDA$_E$: on échantillone $\mathbf{x} \sim Q$. Si $\mathbf{x}$ vient de Q_k, il hérite $\tilde{\mathbf{y}}_k^{(Q)}$

* Redko, Courty, Flamary, Tuia, Optimal transport for multi-source domain adaptation under target shift. AISTATS'19

Estimation d'une carte de Monge



(a) Plan de transport ω



(b) Points transportées

Basées sur P et ω, on veut définir une carte de monge $T^\star$. Néanmoins, comme γ est une GMM, elle n'est pas sous la forme $(Id, T)_\sharp P$

Nous proposons donc une heuristique,

1. Estimer $k_1 = \text{argmax}_k P_k(\mathbf{x})$
2. Transporter $\mathbf{x}$ avec $T_{k_1,k_2}(\mathbf{x})$ pour tout k_2, avec importance donnée par ω_{k_1,k_2}

Si l'on applique cet heuristique aux échantillons de P, cela nous fait un nouveau jeu de données. Le mapping correspond à

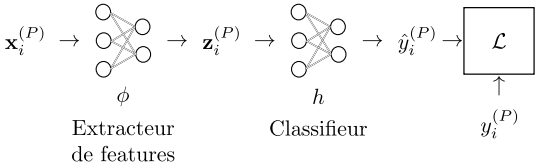
$$\{\mathbf{x}_i^{(P)}, y_i^{(P)}\}_{i=1}^n \mapsto \{\omega_{k_1,k_2}, T_{k_1,k_2}(\mathbf{x}_i^{(P)}), y_i^{(P)}\}_{i=1}^n$$

Cela fait un mapping **affine par morceaux**, en dépendant sur l'assignation $\mathbf{x}_i^{(P)} \mapsto k_1$

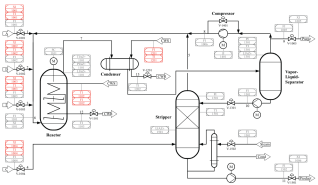
Expérimentations Numériques

Cas d'études

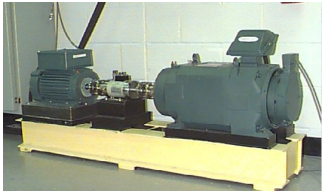
- **Diagnostique de fautes:** à partir des séries temporelles de capteurs, déterminer le type de faute (classe) ou son absence.
- Méthodologie,
  1. Pré-etrainement: on entraîne un réseau de neuronnes (encoder ϕ et classifieur h) sur les données sources
  2. Extraction de features: on applique $\mathbf{z}_i^{(P)} = \phi(\mathbf{x}_i^{(P)})$ (resp. Q).
  3. La classification est faite sur les characteristiques extraites (e.g., SVM)
  4. Adaptation de domaines: on fait l'adaptation de domaines au niveau des features $\mathbf{z}_i^{(P)}$.



Cas d'études



TEP



CWRU

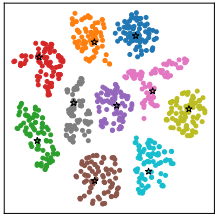
	Tennessee Eastmann Process	Case Western Reserver University
Type	Installation Chimique	Machine méchanique
Domaines	Modes de production	Vitesse de rotation
Type du réseau	CNN	MLP
# Classes	28 fautes + 1	9 fautes + 1
# Features	128	256

Résultats (CWRU)

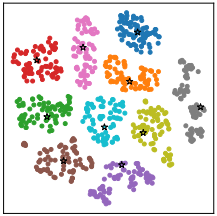
Tâche	SVM	OTDA$_{EMD}$	OTDA$_{Sinkhorn}$	OTDA$_{Linear}$	HOT-DA	GMM-OTDA$_{E}$	GMM-OTDA$_{M}$	GMM-OTDA$_{T}$
A→B	58.3	69.1	70.4	**81.6**	80.0	79.8	79.8	80.0
A→C	47.4	85.7	96.8	94.8	99.9	100.0	100.0	**100.0**
B→A	41.8	67.4	76.0	77.0	79.5	79.6	80.0	**80.0**
B→C	35.2	71.7	75.7	76.6	79.8	79.6	80.0	**80.0**
C→A	58.6	89.1	98.6	93.2	99.0	99.3	99.7	**100.0**
C→B	62.0	70.1	75.0	76.3	80.0	79.8	80.0	**80.0**
Moyenne	50.5	75.5	82.1	83.2	86.4	86.3	86.6	**86.7**

A: 1772rpm, B: 1750rpm, C: 1730rpm

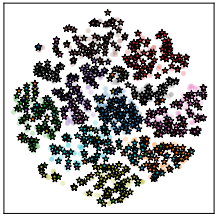
Visualization (CWRU, $A \rightarrow C$)



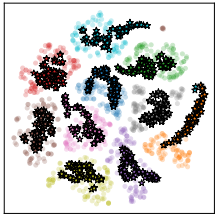
(a) Source



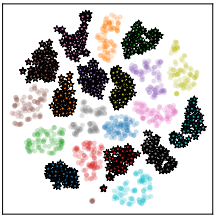
(b) Cible



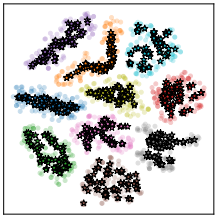
(c) OTDA$_{EMD}$



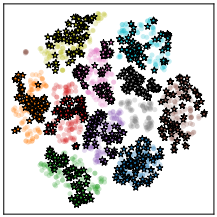
(d) OTDA$_{Sinkhorn}$



(e) OTDA$_{Linear}$



(f) HOT-DA



(g) GMM-OTDA$_T$

Conclusion

Conclusion

- Nous proposons des nouveaux outils pour l'adaptation de domaines à travers une modélisation par mélanges Gaussiens.
- Nos approches sont avantageux par rapport au transport optimal empirique.

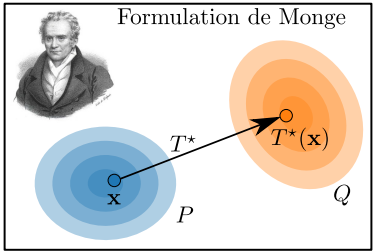
Travaux Futurs

- Formalisation mathématique des concepts proposées par des bornes de l'érreur d'un classifieur appris avec nos méthodes.
- Application en adaptation de domaines multi-sources

Article

Article (MSDA)

Transport optimal

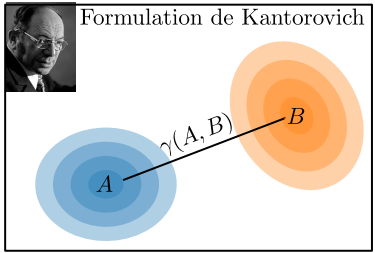


$$T^\star = \underset{T_\sharp P = Q}{\operatorname{argmin}} \int_{\mathbb{R}^d} c(\mathbf{x}, T(\mathbf{x}))dP(\mathbf{x})$$

$$T_\sharp P = Q \iff T^{-1}(B) = A$$

Métrique du terrain,

$$c(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2$$



$$\gamma^\star = \underset{\gamma \in \Gamma(P,Q)}{\operatorname{argmin}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} c(\mathbf{x}_1, \mathbf{x}_2)d\gamma(\mathbf{x}_1, \mathbf{x}_2)$$

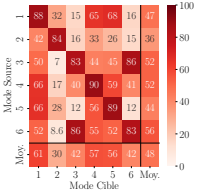
$$\gamma \in \Gamma(P, Q) \iff \begin{array}{l} \int \gamma(A, \mathbf{x}_2)d\mathbf{x}_2 = P(A) \\ \int \gamma(\mathbf{x}_1, B)d\mathbf{x}_1 = Q(B) \end{array}$$

Distance de Wasserstein

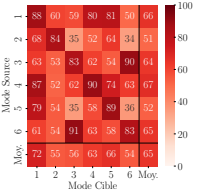
$$\mathcal{W}_2(P, Q)^2 = \min_{\gamma \in \Gamma(P,Q)} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} c(\mathbf{x}_1, \mathbf{x}_2)d\gamma(\mathbf{x}_1, \mathbf{x}_2)$$

⋆ Villani. Optimal Transport: old and new. Springer, 2019

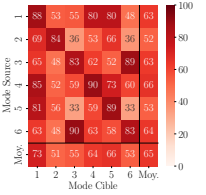
Résultats



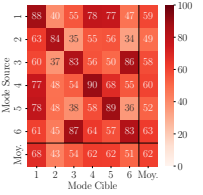
(a) SVM



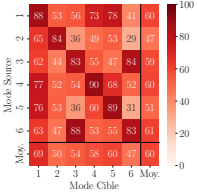
(b) OTDA$_{EMD}$



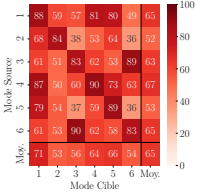
(c) OTDA$_{Sinkhorn}$



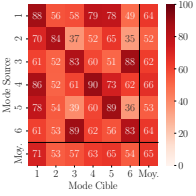
(d) OTDA$_{Linear}$



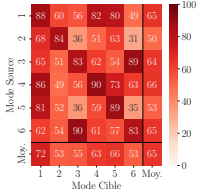
(e) HOT-DA



(f) GMM-OTDA$_E$



(g) GMM-OTDA$_M$



(h) GMM-OTDA$_T$