

Etude des techniques de transfert d'apprentissage appliquées aux données climatiques

Optimisation des performances des transferts d'apprentissage

Florent Crouzet

UQAM, Montréal, Québec, Canada
Ecole Polytechnique, Palaiseau, France

17 juillet 2024



① Intelligence artificielle et données climatiques

② Transfert d'apprentissage : cadre théorique et état de l'art

③ Quantification de la performance des transferts

④ Ouverture

① Intelligence artificielle et données climatiques

Enjeux liés au traitement de données climatiques

Traitements usuels des données climatiques

Intérêt du transfert d'apprentissage

② Transfert d'apprentissage : cadre théorique et état de l'art

③ Quantification de la performance des transferts

④ Ouverture

1 Intelligence artificielle et données climatiques

Enjeux liés au traitement de données climatiques

Traitement usuel des données climatiques

Intérêt du transfert d'apprentissage

4 Ouverture

Enjeux liés au traitement de données climatiques [1]

Enjeux climatiques

Atténuation

- Améliorer l'efficacité des véhicules
- Séquestration du CO₂
- Carburants alternatifs et électrification
- Gestion des feux de forêts
- ...

Adaptation

- Modélisation des impacts
- Prévoir les événements extrêmes
- Calcul de l'empreinte carbone
- Information
- ...

① Intelligence artificielle et données climatiques

Enjeux liés au traitement de données climatiques

Traitements usuels des données climatiques

Intérêt du transfert d'apprentissage

② Transfert d'apprentissage : cadre théorique et état de l'art

③ Quantification de la performance des transferts

④ Ouverture

Exemples de données climatiques

Nom	Alt.	Lat.	Long.	Situation	Débutée
Nantes	26 m	47°30'N	0°11'W	Jaloupi - Nantes	2000-2008
Ajaccio	26 m	42°20'N	0°42'E	Piève	2000-2008
Angers	26 m	47°40'N	0°48'E	Le Pouliguen	2000-2008
Sassan	68 m	47°15'N	0°39'E	Jaïd des Sables - Plages peu élevées Brétignolles-sur-Mer	2005-2008
	106 m	47°25'N	0°40'E	Plages moyennes ou au dessus de la Loire	2005-2008
Tours	138 m	47°30'N	0°04'E	Adrenovane - Plaine	2000-2008
	138 m	47°30'N	0°04'E	Adrenovane - Plaine	2000-2008
Dunkerque	133 m	50°50'N	0°30'E	Adrenovane - Plaine	2000-2008
	133 m	50°50'N	0°30'E	Adrenovane - Plaine	2000-2008
Paris	175 m	48°50'N	0°25'E	Adrenovane - Plaine	2000-2008
	175 m	48°50'N	0°25'E	Adrenovane - Plaine	2000-2008
Le Havre	18 m	49°20'N	0°30'E	Adrenovane - Plaine	2000-2008
	18 m	49°20'N	0°30'E	Adrenovane - Plaine	2000-2008
Rennes	100 m	48°00'N	0°30'E	Ville longtemps envasée	2000-2008
	100 m	48°00'N	0°30'E	Adrenovane - Plaine	2000-2008
Orléans	125 m	47°50'N	0°30'E	Alodrome - Plaine	2000-2008
	125 m	47°50'N	0°30'E	Jonction 50 m au dessus de la Loire	2000-2008

2025 RELEASE UNDER E.O. 14176

Figure 1: Données tabulaires

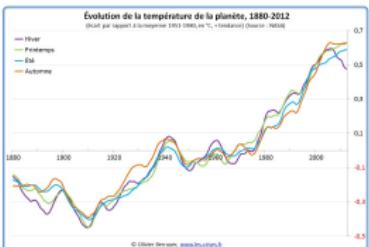


Figure 2: Séries temporelles



Figure 3: Images

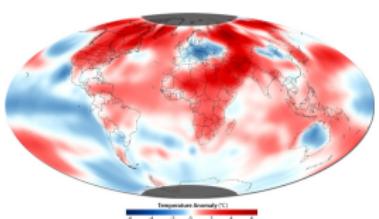


Figure 4: Données géospatiales



Figure 5: Données textuelles

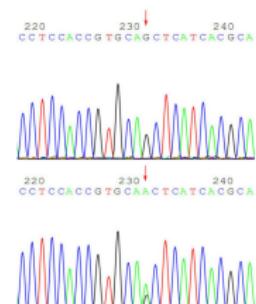


Figure 6: Données génétiques

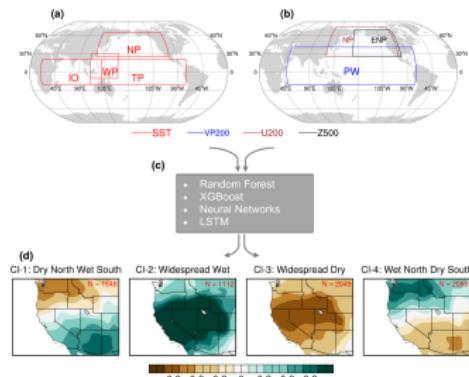
Exemples de méthodes de traitement de données climatiques

Visualisation d'images [2]



- GAN pour créer des images réalistes de lieux sous l'effet des inondations dues au changement climatique.
- Masker : Génère un masque des régions à inonder en utilisant des données simulées.
- Painter : Génère de l'eau réaliste dans les zones définies par le Masker, en s'inspirant du réseau GauGAN.
- Entrainement partiellement suimulé

Prévisions de Précipitations [3]



- Croisement de Random Forest, XGBoost, réseaux de neurones et LSTM.
- Entraînement du modèle sur des simulations climatiques en grand ensemble
- Rivaliser avec ou surpasser les modèles de prévision dynamique
- Améliorer la précision des prévisions sur de longues périodes temporelles.

① Intelligence artificielle et données climatiques

Enjeux liés au traitement de données climatiques

Traitements usuels des données climatiques

Intérêt du transfert d'apprentissage

② Transfert d'apprentissage : cadre théorique et état de l'art

③ Quantification de la performance des transferts

④ Ouverture

Intérêt du transfert d'apprentissage

Rôle du Transfer Learning :

- Efficacité dans certains secteurs spécifiques (habitations, électricité) en raison de la capacité de transférabilité des modèles

Avantages du Transfer Learning :

- Améliorer la précision des prévisions climatiques
- Réduire les coûts associés à leur développement en exploitant les modèles existants

Perspectives :

- Garantir l'efficacité et la fiabilité du TL
- Quantifier les performances prédictives.



Ouvrir



Ouvrir

① Intelligence artificielle et données climatiques

② Transfert d'apprentissage : cadre théorique et état de l'art

Définition et formalisme

Exemples de méthodes algorithmiques usuelles

Exemple d'utilisation du Transfer Learning en climat

③ Quantification de la performance des transferts

④ Ouverture

① Intelligence artificielle et données climatiques

② Transfert d'apprentissage : cadre théorique et état de l'art

Définition et formalisme

Exemples de méthodes algorithmiques usuelles

Exemple d'utilisation du Transfer Learning en climat

③ Quantification de la performance des transferts

④ Ouverture

Problème de Machine Learning

Problème supervisé [4]:

- $X \in \mathcal{X}$: ensemble de caractéristiques
- $Y \in \mathcal{Y}$: ensemble d'étiquettes
- $(X, Y) \sim \mathbb{P}$: distribution de probabilité inconnue
- $\mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$: données d'entraînement i.i.d. selon \mathbb{P}
- $f : \mathcal{X} \rightarrow \mathcal{Y}$: prédicteur
- $\ell : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ la fonction de perte

Objectif : Minimiser le risque

$$\mathcal{R}(f) = \mathbb{E}_{(X, Y) \sim \mathbb{P}}[\ell(Y, f(X))]$$

Méthode : Construire \hat{f} pour minimiser le risque empirique

Cadre Théorique du Transfert d'Apprentissage

Domaines [5]:

- \mathcal{X}_S : Domaine source, \mathcal{X}_T : Domaine cible
- \mathbb{P}_S et \mathbb{P}_T : Lois de probabilité

Tâches:

- (\mathcal{Y}_S, f_S) : Domaine source, (\mathcal{Y}_T, f_T) : Domaine cible

Données d'entraînement:

- Source : $\mathcal{D}_S = \{(X_{S,1}, Y_{S,1}), \dots, (X_{S,n_S}, Y_{S,n_S})\}$
- Cible : $\mathcal{D}_T = \{(X_{T,1}, Y_{T,1}), \dots, (X_{T,n_T}, Y_{T,n_T})\}$, avec $0 \leq n_T \ll n_S$

Objectif du Transfert d'Apprentissage

Utiliser les informations de la source pour améliorer la prédiction dans la cible.

Minimiser le risque :

$$\mathcal{R}_T(f_T) = \mathbb{E}_{(X_T, Y_T) \sim \mathbb{P}_T} [\ell(Y_T, f_T(X_T))]$$

Construire un prédicteur \hat{f}_T minimisant le risque empirique $\hat{\mathcal{R}}_T$.

Cas particuliers du transfert d'apprentissage

Semi-Supervisé :

\mathcal{D}_T partiellement étiqueté. Utilisation de quelques données étiquetées pour capturer la structure cible.

Non Supervisé :

Connaissance de \mathbb{P}_S et \mathbb{P}_T sans données étiquetées. Alignement des distributions marginales $\mathbb{P}_S(X)$ et $\mathbb{P}_T(X)$ et/ou conditionnelles $\mathbb{P}_S(Y|X)$ et $\mathbb{P}_T(Y|X)$.

Cadre Multisources :

k domaines sources X_1, \dots, X_k . Fonctions d'étiquetage $f_k : X_k \rightarrow Y$.

Ensembles $\mathcal{D}_k = \{(x_{ki}, y_{ki})\}_{i=1}^{m_k}$ tirés selon \mathbb{P}_k .

Ensemble cible $\mathcal{D}_T = \{x_i\}_{i=1}^n$ tiré selon \mathbb{P}_T .

Minimisation des $\mathbb{E}_{x \sim \mathbb{P}_T} [L(f_k(x), f_T(x))]$ sur les k .

Méthodologie de transfert

Quoi transférer ? Comment transférer ? Quand transférer ? [5] [6]

Type de transfert	Domaines source et cible ($\mathcal{X}_S, \mathcal{X}_T$)	Tâches source et cible ($\mathcal{Y}_S, \mathcal{Y}_T, f_S, f_T$)
Inductif	Les mêmes	Différents mais liés
Non-supervisé	Différents mais liés	Différents mais liés
Transductif	Différents mais liés	Les mêmes

Table 1: Comparaison des types de transfert d'apprentissage

- **Transfert d'instances** (inductif et transductif) : sélectionne et réutilise des instances spécifiques des données source qui sont pertinentes pour le domaine cible, souvent en utilisant des techniques de pondération, de rééchantillonnage ou de KNN.
- **Transfert de représentation** (tout type de transfert) : transforme les caractéristiques des données source et cible pour qu'elles aient une représentation commune, souvent via des méthodes d'apprentissage adversarial ou de réduction de divergence.

Trouver le bon algorithme de Transfer Learning

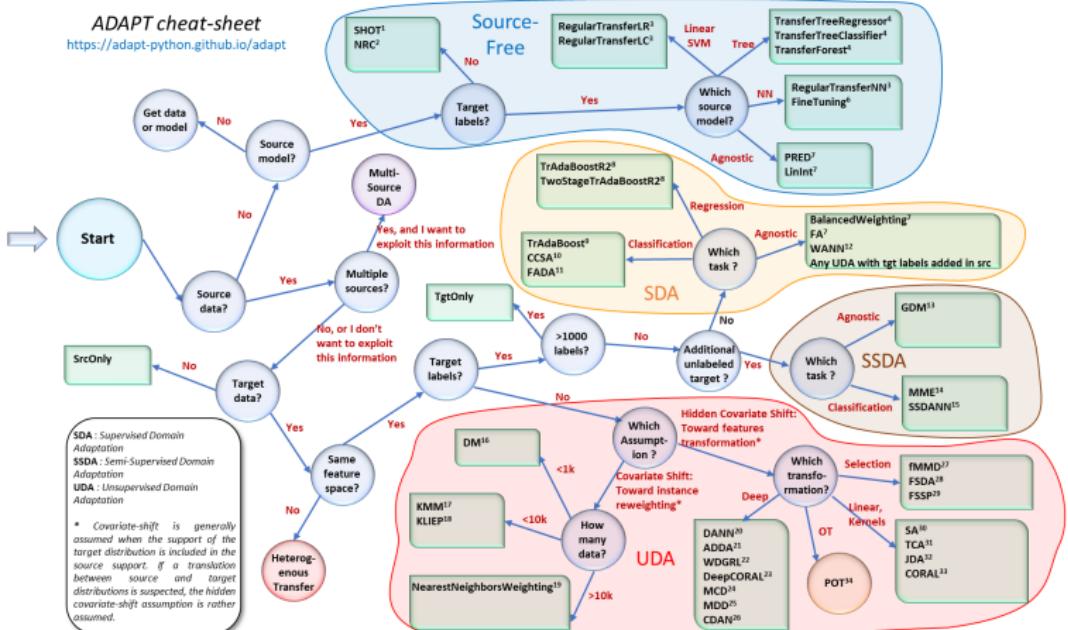


Figure 7: Classification des méthodes de transfert d'apprentissage par [7]

① Intelligence artificielle et données climatiques

② Transfert d'apprentissage : cadre théorique et état de l'art

Définition et formalisme

Exemples de méthodes algorithmiques usuelles

Exemple d'utilisation du Transfer Learning en climat

③ Quantification de la performance des transferts

④ Ouverture

Frustratingly easy domain adaptation (FA) [8]

Principe de l'approche

- Transformer l'adaptation de domaine en un problème standard supervisé.
- Augmentation de l'espace des caractéristiques des données sources et cibles.

Méthode

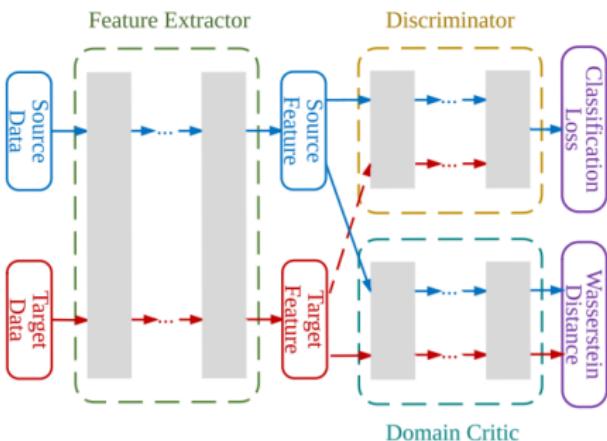
- Augmentation des caractéristiques : $x \rightarrow [x, x, 0]$ pour les sources et $[x, 0, x]$ pour les cibles.
- Intégration des données sources et cibles pour capturer les interactions.

Avantages

- Simple et efficace : amélioration des performances sans modifications complexes des algorithmes.
- Application directe de méthodes de classification ou de régression sur les données augmentées.

Wasserstein Distance Guided Representation Learning (WDGRL) [9]

- Cadre** : transfert d'instances par adaptation de domaine non-supervisé pour des tâches de classification
- Principe** : optimiser des réseaux de neurones adversaires pour résoudre un problème de min-max et utiliser la distance de Wasserstein pour aligner les distributions.
- Extracteur de caractéristiques : réseau de neurones apprenant des représentations de caractéristiques invariantes au domaine.
- Critique de domaine : estimer la distance de Wasserstein entre les distributions de représentations source et cible.



Weighting Adversarial Neural Network (WANN) [10]

- **Cadre** : transfert de représentation par adaptation de domaine (semi-)supervisée pour des tâches de régression
- **Principe** : calculer une pondération des instances sources qui corrige les "décalages" entre la source et la cible en minimisant la distance de disparité.
- **WANN implique trois réseaux :**
 - ① Le réseau de pondération W : apprendre les poids des sources.
 - ② Le réseau de tâche h_t : apprendre la tâche.
 - ③ Le réseau de disparité h_d : estimer une distance de disparité.

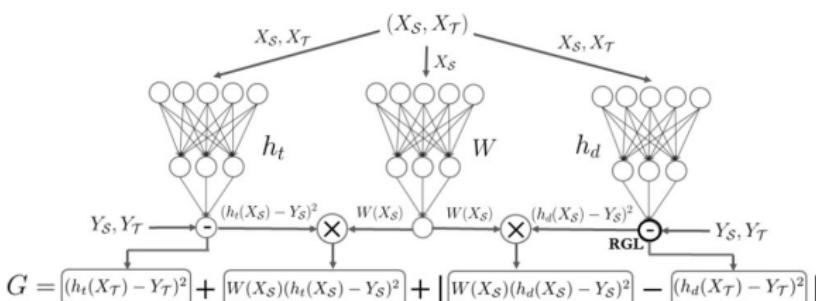


Figure 9: Principe de fonctionnement de WANN d'après [10]

① Intelligence artificielle et données climatiques

② Transfert d'apprentissage : cadre théorique et état de l'art

Définition et formalisme

Exemples de méthodes algorithmiques usuelles

Exemple d'utilisation du Transfer Learning en climat

③ Quantification de la performance des transferts

④ Ouverture

Exemple d'utilisation du Transfer Learning en climat

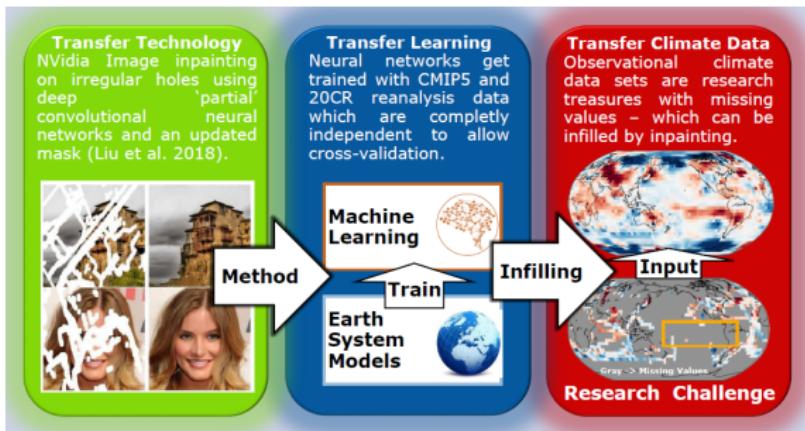


Figure 10: Fonctionnement de la reconstruction d'informations manquantes [11]

- Combinaison des données HadCRUT4 et des modèles CMIP5/20CR pour remplir les lacunes des mesures climatiques historiques.
- Pré-entraînement du Modèle : Les réseaux de neurones convolutifs profonds (CNN) sont d'abord entraînés sur les données sources (CMIP5 et 20CR) qui sont complètes et cohérentes. Cela permet au modèle d'apprendre les caractéristiques et les patterns climatiques de base.
- Application à HadCRUT4

- ① Intelligence artificielle et données climatiques
- ② Transfert d'apprentissage : cadre théorique et état de l'art
- ③ Quantification de la performance des transferts
 - Formalisation du problème
 - Solutions théoriques
 - Discussion
- ④ Ouverture

- ① Intelligence artificielle et données climatiques
- ② Transfert d'apprentissage : cadre théorique et état de l'art
- ③ Quantification de la performance des transferts
 - Formalisation du problème
 - Solutions théoriques
 - Discussion
- ④ Ouverture

Formalisation du problème de la classification 0 – 1

Rappel : objectif du transfert d'apprentissage

Construire un prédicteur \hat{f}_T afin de minimiser le risque empirique $\hat{\mathcal{R}}_T$ correspondant au risque \mathcal{R}_T est défini pour $f : \mathcal{X}_T \rightarrow \mathcal{Y}$ par :

$$\mathcal{R}_T(f) = \mathbb{E}_{(X_T, Y_T) \sim \mathbb{P}_T} [\ell(Y_T, f(X_T))]$$

Objectif : Estimer le risque cible en fonction des paramètres sources du problème ainsi d'une grandeur qui quantifie le transfert entre les distributions.

Cadre de la classification 0 – 1

- $\mathcal{X} = \mathcal{X}_S = \mathcal{X}_T$: Domaine des données.
- $\mathbb{P}_S, \mathbb{P}_T$: Distributions de probabilité sur le domaine X , correspondant aux données source et cible respectivement.
- \mathcal{H} : Classe d'hypothèses $h : X \rightarrow \{0, 1\}$ sur X , où chaque $h \in \mathcal{H}$ est une fonction caractéristique définissant un sous-ensemble de X .
- $VC_{\mathcal{H}}$: la dimension de Vapnik de l'espace \mathcal{H}

Notations et Définitions

- La probabilité selon \mathbb{P}_S qu'une hypothèse $h \in \mathcal{H}$ soit en désaccord avec une fonction d'étiquetage f (qui peut également être une hypothèse) est définie comme :

$$\epsilon_S(h, f) = \mathbb{E}_{x \sim \mathbb{P}_S} [|h(x) - f(x)|].$$

On note $\epsilon_S(h) = \epsilon_S(h, f_S)$, $\hat{\epsilon}_S(h)$ la version empirique. Idem pour $\epsilon_T(h, f)$, $\epsilon_T(h)$, et $\hat{\epsilon}_T(h)$.

- $I(h)$: l'ensemble pour lequel $h \in \mathcal{H}$ est la fonction caractéristique c'est-à-dire $x \in I(h) \Leftrightarrow h(x) = 1$.
- $d_{\mathcal{H}}(\mathbb{P}_S, \mathbb{P}_T)$: Divergence \mathcal{H} entre les distributions \mathbb{P}_S et \mathbb{P}_T , mesurée comme :

$$d_{\mathcal{H}}(\mathbb{P}_S, \mathbb{P}_T) = 2 \sup_{h \in \mathcal{H}} \left| \Pr_{\mathbb{P}_S}[I(h)] - \Pr_{\mathbb{P}_T}[I(h)] \right|$$

avec pour deux échantillons de même taille \mathcal{D}_S et \mathcal{D}_T , $\hat{d}_{\mathcal{H}}(\mathcal{D}_S, \mathcal{D}_T)$ la divergence empirique associée.

- $\mathcal{H}\Delta\mathcal{H}$, la différence symétrique entre \mathcal{H} et lui-même, l'espace des désaccords entre deux hypothèses de \mathcal{H} .
- λ : Erreur combinée idéale, définie comme : $\lambda = \min_{h \in \mathcal{H}} (\epsilon_S(h) + \epsilon_T(h))$.

- ① Intelligence artificielle et données climatiques
- ② Transfert d'apprentissage : cadre théorique et état de l'art
- ③ Quantification de la performance des transferts
 - Formalisation du problème
 - Solutions théoriques
 - Discussion
- ④ Ouverture

Inégalités d'Erreur

Lemme [12]

Soit \mathcal{H} un espace d'hypothèses sur \mathcal{X} de dimension de Vapnik $VC_{\mathcal{H}}$. Si \mathcal{D}_S et \mathcal{D}_T sont des échantillons non labélisés de taille m tirés selon \mathbb{P}_S et \mathbb{P}_T , et si $\hat{d}_{\mathcal{H}}(\mathcal{D}_S, \mathcal{D}_T)$ est la divergence \mathcal{H} empirique entre les échantillons, alors pour tout $\delta \in [0, 1]$, avec une probabilité au moins $1 - \delta$,

$$d_{\mathcal{H}}(\mathbb{P}_S, \mathbb{P}_T) \leq \hat{d}_{\mathcal{H}}(\mathcal{D}_S, \mathcal{D}_T) + 4\sqrt{\frac{2VC_{\mathcal{H}} \log(2m) + \log\left(\frac{2}{\delta}\right)}{m}}.$$

Propriété [12]

Pour tout $\delta \in [0, 1]$, avec probabilité $\geq 1 - \delta$ sur les échantillons non-étiquetés \mathcal{D}_S et \mathcal{D}_T de taille m :

$$\epsilon_T(h) \leq \epsilon_S(h) + \frac{1}{2} \hat{d}_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_S, \mathcal{D}_T) + 4\sqrt{\frac{2VC_{\mathcal{H}} \log(2m) + \log\left(\frac{2}{\delta}\right)}{m}} + \lambda = \epsilon_S(h) + A + \lambda.$$

Combinaison des Erreurs

Théorème [12]

Soit \mathcal{H} un espace d'hypothèses de dimension VC $VC_{\mathcal{H}}$. Soit \mathcal{D}_S et \mathcal{D}_T des échantillons non étiquetés de taille m' chacun, tirés respectivement de \mathbb{P}_S et \mathbb{P}_T . Soit un échantillon étiqueté de taille m généré en tirant $m_T = \beta m$ points selon \mathbb{P}_T et $m_S = (1 - \beta)m$ points selon \mathbb{P}_S et en les étiquetant respectivement selon f_T et f_S . Soit $\epsilon_\alpha = \alpha\epsilon_T + (1 - \alpha)\epsilon_S$. Si $\hat{h} \in \mathcal{H}$ est le minimiseur empirique de $\hat{\epsilon}_\alpha(h)$ sur l'échantillon et $h_T^* = \min_{h \in \mathcal{H}} \epsilon_T(h)$ est le minimiseur de l'erreur cible, alors pour tout $\delta \in [0, 1]$, avec une probabilité au moins $1 - \delta$,

$$\epsilon_T(\hat{h}) \leq \epsilon_T(h_T^*) + 4\sqrt{\frac{\alpha^2}{\beta} + \frac{(1 - \alpha)^2}{1 - \beta}} \sqrt{\frac{2VC_{\mathcal{H}} \log(2(m + 1)) + 2 \log(\frac{8}{\delta})}{m}} + 2(1 - \alpha)A.$$

Avec pour valeur optimale de α en fonction de m_S , m_T et $D^2 = \frac{VC_{\mathcal{H}}}{A^2}$:

$$\alpha^*(m_T, m_S; D) = \begin{cases} 1 & \text{si } m_T \geq D^2, \\ \min\{1, \nu\} & \text{si } m_T \leq D^2, \end{cases}$$

où

$$\nu = \frac{m_T}{m_T + m_S} \left(1 + \frac{m_S}{\sqrt{D^2(m_S + m_T) - m_S m_T}} \right).$$

- ① Intelligence artificielle et données climatiques
- ② Transfert d'apprentissage : cadre théorique et état de l'art
- ③ Quantification de la performance des transferts
 - Formalisation du problème
 - Solutions théoriques
 - Discussion
- ④ Ouverture

Discussion

$$\epsilon_T(\hat{h}) \leq \epsilon_T(h_T^*) + 4 \sqrt{\frac{\alpha^2}{\beta} + \frac{(1-\alpha)^2}{1-\beta}} \sqrt{\frac{2VC_{\mathcal{H}} \log(2(m+1)) + 2 \log\left(\frac{8}{\delta}\right)}{m}} + 2(1-\alpha)A.$$

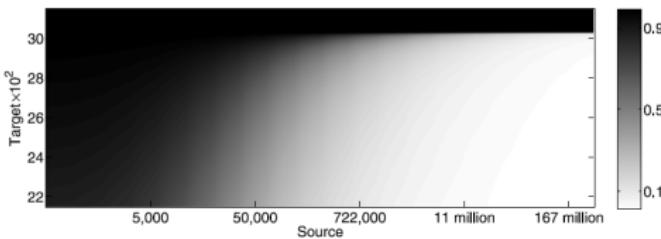
Avec pour valeur optimale de α en fonction de m_S , m_T et $D^2 = \frac{VC_{\mathcal{H}}}{A^2}$:

$$\alpha^*(m_T, m_S; D) = \begin{cases} 1 & \text{si } m_T \geq D^2, \\ \min\{1, \nu\} & \text{si } m_T \leq D^2, \end{cases}$$

où

$$\nu = \frac{m_T}{m_T + m_S} \left(1 + \frac{m_S}{\sqrt{D^2(m_S + m_T) - m_S m_T}} \right).$$

Exemple [12]:



- ① Intelligence artificielle et données climatiques
- ② Transfert d'apprentissage : cadre théorique et état de l'art
- ③ Quantification de la performance des transferts
- ④ Ouverture

Ouverture

Améliorations :

- Y. Mansour, et al. (2009) démontrent une borne point par point sur la fonction de perte sous certaines hypothèses
- de Mathelin et al. (2020) généralisent à la régression multi-sources et au cas supervisé

Perspectives :

- Cas d'application du Transfer Learning en climat
- Application aux changements brutaux
- **Construire une théorie solide des garanties de performance en TL**

Merci pour votre attention !

Amélioration

Définition (Complexité de Rademacher) Soit H un ensemble de fonctions à valeurs réelles définies sur un ensemble X . Étant donné un échantillon $S \in X^m$, la **complexité de Rademacher empirique** de H est définie comme suit :

$$\hat{\mathcal{R}}_S(H) = \frac{2}{m} \mathbb{E}_\sigma \left[\sup_{h \in H} \left| \sum_{i=1}^m \sigma_i h(x_i) \right| \middle| S = (x_1, \dots, x_m) \right].$$

L'espérance est prise sur $\sigma = (\sigma_1, \dots, \sigma_m)$ où les σ_i sont des variables aléatoires uniformes indépendantes prenant des valeurs dans $\{-1, +1\}$. La **complexité de Rademacher** d'un ensemble d'hypothèses H est définie comme l'espérance de $\hat{\mathcal{R}}_S(H)$ sur tous les échantillons de taille m :

$$\mathcal{R}_m(H) = \mathbb{E}_S \left[\hat{\mathcal{R}}_S(H) \middle| |S| = m \right].$$

Définition (Discrepancy Distance) Soit H un ensemble de fonctions de X vers Y et soit $L : Y \times Y \rightarrow \mathbb{R}^+$ une fonction de perte définie sur Y . La discrepancy distance disc_L entre deux distributions Q_1 et Q_2 sur X est définie par

$$\text{disc}_L(Q_1, Q_2) = \max_{h, h' \in H} |\mathcal{L}_{Q_1}(h', h) - \mathcal{L}_{Q_2}(h', h)|.$$

Améliorations

Une fonction de perte L est σ -admissible lorsqu'il existe $\sigma \in \mathbb{R}^+$ tel que pour deux hypothèses $h, h' \in H$ et pour tout $x \in X$, et $y \in Y$,

$$|L(h(x), y) - L(h'(x), y)| \leq \sigma |h(x) - h'(x)|.$$

Théorèmes principaux Soit $K : X \times X \rightarrow \mathbb{R}$ un noyau symétrique défini positif tel que $K(x, x) \leq \kappa^2 < +\infty$ pour tout $x \in X$, et soit \mathcal{H} le RKHS associé à K . Supposons que la fonction de perte L est σ -admissible. Soit h' l'hypothèse retournée par l'algorithme de régularisation basé sur $N(\cdot) = \|\cdot\|_K^2$ pour la distribution empirique \hat{P} , et h celle retournée pour la distribution empirique \hat{Q} . Lorsque L est la perte au carré L_2 et $\delta' = L_{\hat{Q}}(h_P^*(x), f_Q(x))^{1/2} + L_{\hat{P}}(h_P^*(x), f_P(x))^{1/2} \ll 1$, alors, pour tout $x \in X$ et $y \in Y$,

$$|L(h'(x), y) - L(h(x), y)| \leq \frac{2\kappa M}{\lambda} \left(\kappa \delta' + \sqrt{\kappa^2 \delta'^2 + 4\lambda \text{disc}_L(\hat{P}, \hat{Q})} \right).$$

Dans le cas où $f_P \in \mathcal{H}$, alors avec $\delta = L_{\hat{Q}}(f_Q(x), f_P(x))$, on a le même résultat avec δ qui remplace δ' .

Si on suppose de plus que f_P et f_Q coïncident sur $\text{supp}(\hat{Q})$, alors on a le résultat suivant :

$$|L(h'(x), y) - L(h(x), y)| \leq \kappa \sigma \sqrt{\frac{\text{disc}_L(\hat{P}, \hat{Q})}{\lambda}}$$

- [1] D. Rolnick, P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, A. S. Ross, N. Milojevic-Dupont, N. Jaques, A. Waldman-Brown, A. S. Luccioni, T. Maharaj, E. D. Sherwin, S. K. Mukkavilli, K. P. Kording, C. P. Gomes, A. Y. Ng, D. Hassabis, J. C. Platt, F. Creutzig, J. Chayes, and Y. Bengio, "Tackling climate change with machine learning," *ACM Computing Surveys*, vol. 55, p. Article 42, February 2022.
- [2] V. Schmidt, A. Luccioni, S. K. Mukkavilli, N. Balasooriya, K. Sankaran, J. Chayes, and Y. Bengio, "Visualizing the consequences of climate change using cycle-consistent adversarial networks," in *Proceedings of the Conference Name*, 2019.
- [3] P. B. Gibson, W. E. Chapman, A. Altinok, L. Delle Monache, M. J. DeFlorio, and D. E. Waliser, "Training machine learning models on climate model output yields skillful interpretable seasonal precipitation forecasts," *Communications Earth & Environment*, vol. 2, no. 1, pp. 1–11, 2021.
- [4] S. Shalev-Shwartz and S. Ben-David, "Understanding machine learning: From theory to algorithms," in *Proceedings of the Machine Learning Conference*, (Cambridge, UK), Cambridge University Press, 2014. Book.
- [5] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, pp. 1345–1359, October 2010.

- [6] J. Wang and Y. Chen, *Introduction to Transfer Learning: Algorithms and Practice*. Machine Learning: Foundations, Methodologies, and Applications, Springer Singapore, 1 ed., 2023.
- [7] A. de Mathelin, M. Atiq, G. Richard, A. de la Concha, M. Yachouti, and et al., "ADAPT: Awesome domain adaptation python toolbox," 2024.
- [8] H. Daume III, "Frustratingly easy domain adaptation," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, 2007.
- [9] J. Shen, Y. Qu, W. Zhang, and Y. Yu, "Wasserstein distance guided representation learning for domain adaptation," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2018.
- [10] A. de Mathelin, G. Richard, F. Deheeger, M. Mougeot, and N. Vayatis, "Adversarial weighting for domain adaptation in regression," in *2020 IEEE International Conference on Big Data (Big Data)*, IEEE, 2020.
- [11] C. Kadow, D. M. Hall, and U. Ulbrich, "Artificial intelligence reconstructs missing climate information," *Nat. Geosci.*, vol. 13, pp. 408–413, 2020.
- [12] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," in *Proceedings of the 22nd Annual Conference on Learning Theory (COLT)*, Omnipress, 2009.



- [13] S. G. Dobrovolski, *Stochastic Climate Theory: Models and Applications*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, 2000.
- [14] Y. Mansour, M. Mohri, and A. Rostamizadeh, "Domain adaptation: Learning bounds and algorithms," in *Conference on Learning Theory (COLT)*, Springer, 2009.
- [15] G. Richard, A. de Mathelin, G. Hébrail, M. Mougeot, and N. Vayatis, "Unsupervised multi-source domain adaptation for regression," in *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, Springer, 2020.
- [16] O. Catoni, "Pac-bayesian supervised classification: The thermodynamics of statistical learning," in *IMS Lecture Notes Monograph Series*, vol. 56, pp. 1–163, Institute of Mathematical Statistics, 2007.
- [17] L. Tang, C. Xiong, and L. Zhang, "Spatial transferability of neural network models in travel demand modeling," *Journal of Computing in Civil Engineering*, vol. 32, no. 3, 2023.
- [18] J. Jang, Y. Cha, J. Kim, S. Lee, G. Lee, M. Cho, Y. Hwang, and N. Kwak, "Semi-supervised domain adaptation for wildfire detection," in *Proceedings of the Workshop on Tackling Climate Change with Machine Learning, ICLR 2024*, 2024.
- [19] A. Lacoste, A. Luccioni, V. Schmidt, and T. Dandres, "Quantifying the carbon emissions of machine learning," in *Proceedings of the Conference Name*, 2019.



- [20] A. McGovern, K. L. Elmore, D. J. Gagne II, S. E. Haupt, C. D. Karstens, R. Lagerquist, T. Smith, and J. K. Williams, "Using artificial intelligence to improve real-time decision-making for high-impact weather," *Bull. Am. Meteorol. Soc.*, pp. 2073–2090, 2017.
- [21] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," in *Proceedings of the JMLR Workshop and Conference Track*, 2016.
- [22] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [23] B. Sun and K. Saenko, "Deep coral: correlation alignment for deep domain adaptation," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2016.
- [24] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2016.
- [25] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, "Maximum classifier discrepancy for unsupervised domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

- [26] Y. Zhang, T. Liu, M. Long, and M. Jordan, "Bridging theory and algorithm for domain adaptation," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2019.
- [27] D. Berthelot, R. Roelofs, K. Sohn, N. Carlini, and A. Kurakin, "Adamatch: A unified approach to semi-supervised learning and domain adaptation," in *Proceedings of the Conference Name*, 2022.
- [28] J. Hwang, P. Orenstein, J. Cohen, K. Pfeiffer, and L. Mackey, "Improving subseasonal forecasting in the western u.s. with machine learning," in *Proceedings of the Conference Name*, 2019.
- [29] W. Zhang, L. Deng, L. Zhang, and D. Wu, "A survey on negative transfer," in *Proceedings of the Conference Name*, pp. 305–329, 2023.
- [30] S. Ben David, T. Lu, T. Luu, and D. Pal, "Impossibility theorems for domain adaptation," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 9, pp. 129–136, PMLR, 2010.
- [31] Y. Mansour, M. Mohri, and A. Rostamizadeh, "Domain adaptation with multiple sources," in *Advances in Neural Information Processing Systems (NIPS 2008)*, (Vancouver, Canada), 2009.
- [32] C. Wong, "How ai is improving climate forecasts," in *Nature*, March 2024.

- [33] C. W. Team, H. Lee, and J. Romero, "Summary for policymakers," in *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, (Geneva, Switzerland), pp. 1–34, IPCC, 2023.
- [34] V. Antoni and I. Joassard, "Impact du changement climatique," in *Chiffres clés des risques naturels Édition 2023*, pp. 124–126, Ministère de la Transition Écologique et de la Cohésion des Territoires, January 2024.
- [35] M. Blaha, C. Vogel, A. Richard, J. D. Wegner, T. Pock, and K. Schindler, "Large-scale semantic 3D reconstruction: An adaptive multi-resolution model for multi-class volumetric labeling," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3176–3184, 2016.
- [36] I. Antonopoulos, V. Robu, B. Couraud, D. Kirli, S. Norbu, A. Kiprakis, D. Flynn, S. Elizondo-Gonzalez, and S. Wattam, "Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review," in *Renewable and Sustainable Energy Reviews*, vol. 130, 2020.
- [37] A. R. Khan, A. Mahmood, A. Safdar, Z. A. Khan, and N. Ahmed Khan, "Load forecasting, dynamic pricing and dsm in smart grid: A review," in *Renewable and Sustainable Energy Reviews*, vol. 54, 2016.

- [38] P. A. Hancock, I. Nourbakhsh, and J. Stewart, "On the future of transportation in an era of automated and autonomous vehicles," in *Proceedings of the National Academy of Sciences*, vol. 116, 2019.
- [39] E. Mocanu, P. H. Nguyen, W. L. Kling, and M. Gibescu, "Unsupervised energy prediction in a smart grid context using reinforcement cross-building transfer learning," in *Energy and Buildings*, vol. 116, 2016.
- [40] D. Kifer, S. Ben-David, and J. Gehrke, "Detecting change in data streams," in *Proceedings of the Thirtieth International Conference on Very Large Data Bases (VLDB 2004)*, pp. 180–191, Morgan Kaufmann, 2004.