TER - Proposition de Sujet
An Introduction to Domain Adaptation for Binary Classification

Guillaume Metzler

Laboratoire ERIC, Université de Lyon, Université Lyon 2
guillaume.metzler@univ-lyon2.fr

Contexte

Une hypothèse importante en apprentissage automatique est que les données tests suivent la même distribution que les données qui ont servi à apprendre notre modèle. Cette hypothèse, très forte, permet d'établir des résultats importants sur la stabilité (Bousquet and Elisseeff, 2002) ou encore sur les performances en généralisation de notre modèle (Valiant, 1984).

En pratique, cette hypothèse est souvent mise à mal (voir Figure 1) dans de nombreux contextes et notamment la détection de fraudes ou d'anomalies dans lesquels la distribution des données à tendance à varier au cours du temps : **on parle de phénomène de dérive**. Il est donc important de pouvoir apprendre un modèle qui va prendre ce changement dans la distribution. La branche du Machine Learning qui s'intéresse à cela s'appelle l'**apprentissage par transfert** et traite de cas encore plus généraux que celui expliqué ci-dessus.

Dans le cadre de ce TER, on va surtout s'intéresser une branche de l'apprentissage par transfert que l'on appelle l'**Adaptation de domaine**. En quelques mots : il s'agit d'un ensemble de techniques utilisés de façon à ce que le modèle appris sur des données dites **sources** soit également performant sur des données dites **cibles** (ou **target**) et on suppose que **la tâche à effectuer sur les données sources est la même que sur les données cibles** : on parle de *transfert transductif*. Un exemple simple pourrait être la classification de mails *spam/ham* avec un modèle appris sur une boîte mail pro et on voudrait **adapter** ce modèle pour qu'il fonctionne sur une boîte mail perso.

L'adaptation de domaine se rencontre aussi dans le cas où l'on souhaite tirer partie de connaissances sur **des données étiquetées sources** afin d'apprendre un modèle sur **des données non étiquetées cibles** (unsupervised domain adaptation) : c'est surtout le cas lorsque l'annotation de données peut se révéler extrêmement coûteuse !

Objectif et sujets

L'objectif est donc de vous faire travailler sur l'adaptation de domaine et de voir comment on peut adapter des algorithmes existants (par le biais de la loss à optimiser) afin qu'ils puissent performer tout aussi bien sur les ensembles **source** et **cible**.
On se propose de faire cela, par exemple, en repartant d'un papier (Gautheron et al., 2020) qui traite

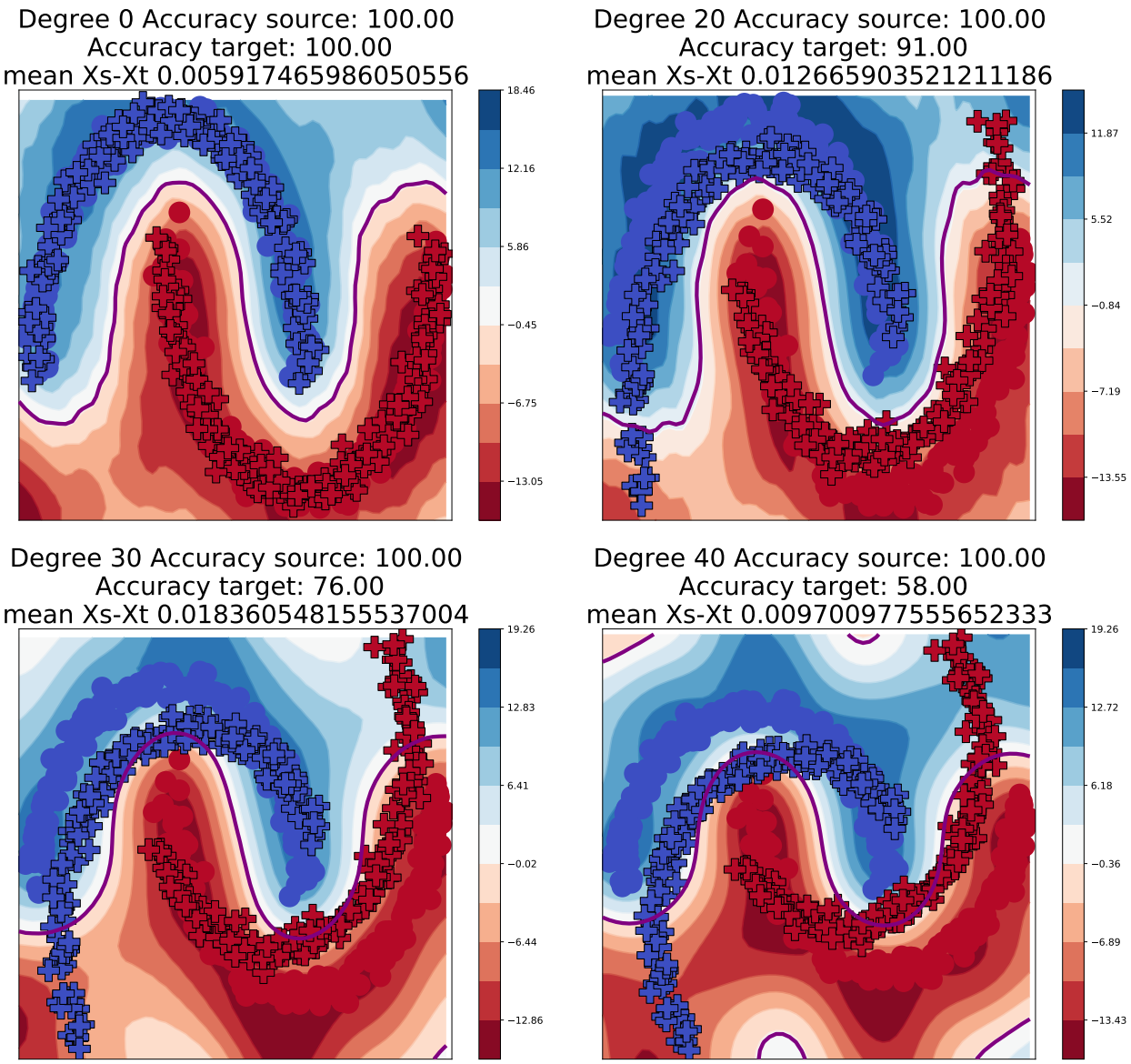

Figure 1: Représentation des performances d'un classifieur standard appris sur des données sources (les ronds) et dont on évalue les performances sur des données cibles (les croix) lorsque l'algorithme n'effectue pas de tâche d'adaptation. On constate qu'il se concentre uniquement sur les données sources sans se préoccuper des données cibles.

de la classification binaire par une méthode d'approximations de noyaux gaussiens dans un contexte de boosting, de le transformer afin de l'employer pour de l'adaptation de domaines.
Ce n'est bien sûr pas une obligation et vous pourrez tout à fait faire le choix de concevoir votre propre algorithme !
Le travail se déroulera en plusieurs étapes :

- comprendre ce qu'est l'apprentissage par transfert et plus précisément ce qu'est l'adaptation de domaine non supervisée (Mansour et al., 2009; Germain et al., 2020; Redko et al., 2020). Vous pourrez également essayer de comprendre quels sont les enjeux théoriques (Ben-David et al., 2010) qui sont derrières ainsi que les applications pratiques.
- on pourra ensuite regarder quelles sont les mesures de divergences entre les domaines en cherchant des articles dans le survey de Redko et al. (2020) ou encore de Gretton et al. (2012).
- étudier quelles sont les deux approches en DA : *Feature Space Remapping* and *Latent Space Representation*.

Après ce travail de compréhension du sujet, on pourra s'attaquer au problème en regardant quelques méthodes de l'état de l'art que l'on peut trouver dans le survey précédent ou en regardant des références qui proposent de s'attaquer à ce problème par différentes approches (Aljundi et al., 2015; Fernando et al., 2014; Germain et al., 2013, 2020; Gong et al., 2012; Geng et al., 2011; Bruzzone and Marconcini, 2009) (on se concentrera sur les approches non deep pour ce travail) on pourra également chercher d'autres références plus récente sur le sujet.

Vous développerez ensuite, en vous basant sur vos lectures, une adaptation de l'algorithme de Gautheron et al. (2020) pour qu'il puisse faire de l'adaptation de domaine et vous comparerez vos résultats à au moins une ou deux méthodes de l'état de l'art.
Donc à eux de voir comment on pourrait reprendre une méthode de l'état de l'art pour cet algorithme pour qu'il fasse du DA puis ils devront se comparer à des d'autres méthodes de DA (papier de Marc, Pascal ou DASVM par exemple).

Rédaction du rapport

Votre rapport, qui ne sera pas forcément très long non plus, environ 8-10 pages (selon ce que vous souhaitez présenter, cela peut être plus aussi) se présentera comme un article scientifique et contiendra les parties suivantes :

- **Abstract :** petit texte de 5-6 phrases dans lequel vous résumez le travail effectuer
- **Introduction :** il vous permet de présenter le contexte de l'étude et quelques généralités. On pourra, par exemple, présenter l' *apprentissage par transfert* puis se focaliser sur l'*adaptation de domaines*, dans quel contexte nous sommes amenés à employer ce type de méthodes mais aussi quelques exemples concrets ou usages pratiques
- **Etat de l'art :** vous devez citer et regarder dans la littérature ce qu'est l'adaptation de domaine ainsi que les méthodes existantes (transférer le source sur le cible ou trouver un espace latent commun). Puis vous présenterez très rapidement le papier sur lequel va se baser votre travail

- **Méthode proposée :** c'est là que vous présentez votre méthode en insistant sur votre contribution
- **Expériences et résultats :** vous présentez les jeux de données utilisées, le protocole expérimental et les résultats ainsi qu'une description des résultats.
- **Conclusion :** synthèse de votre travail et bilan de la méthode proposée et on n'oublie pas d'ouvrir sur une autre piste de réflexion.

References

Rahaf Aljundi, Rémi Emonet, Damien Muselet, and Marc Sebban. Landmarks-based kernelized subspace alignment for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 56–63, 2015.

Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79(1):151–175, 2010.

Olivier Bousquet and André Elisseeff. Stability and generalization. *Journal of Machine Learning Research*, 2:499–526, 2002. ISSN 1532-4435.

Lorenzo Bruzzone and Mattia Marconcini. Domain adaptation problems: A dasvm classification technique and a circular validation strategy. *IEEE transactions on pattern analysis and machine intelligence*, 32(5):770–787, 2009.

Basura Fernando, Amaury Habrard, Marc Sebban, and Tinne Tuytelaars. Subspace alignment for domain adaptation. *arXiv preprint arXiv:1409.5241*, 2014.

Léo Gautheron, Pascal Germain, Amaury Habrard, Guillaume Metzler, Emilie Morvant, Marc Sebban, and Valentina Zantedeschi. Landmark-based ensemble learning with random fourier features and gradient boosting. In *ECML/PKDD (3)*, 2020.

Bo Geng, Dacheng Tao, and Chao Xu. Daml: Domain adaptation metric learning. *IEEE Transactions on Image Processing*, 20(10):2980–2989, 2011.

Pascal Germain, Amaury Habrard, François Laviolette, and Emilie Morvant. A pac-bayesian approach for domain adaptation with specialization to linear classifiers. In *International conference on machine learning*, pages 738–746. PMLR, 2013.

Pascal Germain, Amaury Habrard, François Laviolette, and Emilie Morvant. Pac-bayes and domain adaptation. *Neurocomputing*, 379:379–397, 2020.

Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *2012 IEEE conference on computer vision and pattern recognition*, pages 2066–2073. IEEE, 2012.

Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012.

Yishay Mansour, Mehryar Mohri, and Afshin Rostamizadeh. Domain adaptation: Learning bounds and algorithms. *arXiv preprint arXiv:0902.3430*, 2009.

Ievgen Redko, Emilie Morvant, Amaury Habrard, Marc Sebban, and Younès Bennani. A survey on domain adaptation theory: learning bounds and theoretical guarantees. *arXiv preprint arXiv:2004.11829*, 2020.

Leslie G Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.