

Adaptation de domaine pour la détection automatique du cancer de la prostate en imagerie IRM multiparamétrique

Léo GAUTHERON, Ievgen REDKO, Carole LARTIZIEN

Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69621, Lyon, France
leo_g_autheron@hotmail.fr, Ievgen.Redko, Carole.Lartizien@creatis.insa-lyon.fr

Résumé – Les logiciels d’aide au diagnostic par imagerie médicale (CAD) ont pour objectif d’assister les médecins en pointant les zones suspectes de l’image, généralement à l’aide d’un modèle de classification entraîné sur une base d’apprentissage source étiquetée. Dans ce papier, nous nous intéressons au cas où un CAD est utilisé pour classer des données cible non étiquetées ayant une distribution statistique différente des données source. Nous présentons une étude empirique de différentes méthodes d’adaptation de domaine et montrons le fort potentiel de ces méthodes exprimé par une amélioration importante des résultats de classification par rapport au cas où les données source ne sont pas adaptées.

Abstract – Computer aided diagnostic systems (CAD) are aimed at assisting clinicians by locating suspicious lesions in medical images. These software are usually based on a classification model trained on a labeled source database. In this paper, we focus on the situation where a CAD is used to analyse target data with a different statistical distribution than those of the source training database. An experimental study is conducted to evaluate several unsupervised domain adaptation methods. This study demonstrates the potential of such methods for this application, which allow a significant performance increase compared with the non-adapted classification model.

1 Introduction

Les avancées récentes des techniques d’imagerie médicale conduisent à la génération, pour chaque patient, d’un ensemble de données massives (volume d’images 3D, 4D) que le médecin doit interpréter pour réaliser son diagnostic. Si l’information contenue dans cette masse de données est riche, elle reste cependant difficile à exploiter. La fusion à l’oeil de ces gros volumes de données est une tâche complexe et fastidieuse, en particulier pour des radiologues peu expérimentés, et conduit à une forte inter-intra-variabilité des performances diagnostiques.

Cette problématique clinique alimente aujourd’hui un thème de recherche autour du développement de méthodes d’analyse statistique de l’image permettant d’extraire l’information pertinente pour le diagnostic. Ces méthodes sont fondées sur l’apprentissage statistique et la théorie de la détection. Elles constituent les bases de logiciels d’aide au diagnostic (CAD) dont l’objectif est d’assister le clinicien en identifiant et caractérisant de manière automatique les zones suspectes de l’image. La grande majorité de ces systèmes repose sur des modèles de classification supervisée qui nécessitent la constitution de bases d’apprentissage, contenant des échantillons représentatifs des différentes classes de la pathologie. La constitution de ces bases est un challenge car elle implique un lourd travail d’annotation des images réalisé par des cliniciens experts.

Une question peu adressée dans la communauté scientifique concerne l’estimation de la robustesse de ces systèmes, entraînés sur une base de données annotées, lorsqu’ils sont destinés à être utilisés sur des données test provenant d’une distribution statistique potentiellement différente. Cette situation est très

fréquente en imagerie médicale, en particulier dans le contexte des études multicentriques. Les images acquises sur des systèmes et/ou avec des protocoles différents, présentent en effet des caractéristiques de bruit et de texture hétérogènes. Leur concaténation brute, telle qu’elle est réalisée aujourd’hui dans de nombreuses études, est ainsi potentiellement sous-optimale.

Dans ce travail, nous proposons d’aborder ce problème sous l’angle du nouveau paradigme de l’apprentissage automatique appelé adaptation de domaine. Ce thème de recherche, qui s’inscrit dans le cadre plus général de l’apprentissage par transfert [8], s’intéresse aux problèmes de classification pour lesquels les données de test et d’apprentissage sont générées selon des distributions probabilistes différentes. Dans le cas le plus classique, on dispose d’un modèle de classification appris sur un ensemble d’exemples étiquetés (source) pour une tâche donnée et l’on souhaite adapter ce modèle, pour la même tâche de classification, à un ensemble de données (cible) pas ou partiellement étiquetées et issues d’une distribution différente.

Pour ce travail, nous exploitons les résultats de précédents travaux dans le domaine des CAD pour le cancer de la prostate en imagerie IRM multi-paramétrique [2, 1]. Le contexte méthodologique consiste à adapter le modèle de classification développé à partir d’images IRM annotées acquises sur un scanner 1.5 Tesla (source) à des images acquises sur un scanner 3T (cible). Dans la suite, nous positionnons cette étude par rapport à l’état de l’art, puis présentons les différentes méthodes d’adaptation de domaine considérées et enfin analysons les performances de ces méthodes pour l’application au diagnostic automatique du cancer de prostate en IRM multiparamétrique et multi-sources.

2 État de l’art et originalité

Comme souligné en introduction, les méthodes d’apprentissage par transfert offrent un cadre méthodologique adapté aux problématiques d’apprentissage sur des données d’imagerie provenant de sources différentes. Elles sont actuellement étudiées pour différentes problématiques, par exemple pour la segmentation automatique d’images cérébrales IRM, de données de microscopie électronique ou la classification de données d’imagerie multimodales [9]. Néanmoins, peu de travaux se sont intéressés au potentiel de ces méthodes pour des problèmes d’adaptation de bases de données dans le contexte des systèmes d’aide au diagnostic. Lors d’une étude précédente, nous avons proposé une contribution originale dans le contexte de l’apprentissage semi-supervisé, pour lequel nous disposons d’un faible échantillonnage de données cible annotées [3]. Dans cette étude, nous nous intéressons au contexte non-supervisé et étudions différentes approches, en particulier, une méthode originale basée sur le transport optimal.

3 Adaptation non supervisée

En adaptation de domaine non supervisée, on dispose d’exemples source $\mathbf{S} = \{\mathbf{x}_i^S \in \mathbb{R}^F\}_{i=1}^n$, d’étiquettes connues $\{y_i^S\}_{i=1}^n$ et d’exemples cible $\mathbf{T} = \{\mathbf{x}_i^T \in \mathbb{R}^F\}_{i=1}^m$ d’étiquettes inconnues que l’on veut prédire. On suppose ici que les distributions marginales de \mathbf{S} et \mathbf{T} notées f_S et f_T sont différentes (même modèle de distribution avec des paramètres différents ou modèles différents). Ainsi, l’utilisation d’un modèle entraîné sur \mathbf{S} ne garantit pas d’avoir de bonnes performances sur \mathbf{T} si la divergence entre \mathbf{S} et \mathbf{T} n’est pas minimisée. On peut alors envisager d’adapter les données source aux données cible. Dans la littérature, on distingue 3 familles de méthodes d’adaptation de domaine : (1) les méthodes qui repondèrent les données source pour qu’elles soient les plus proches possible des données cible ; (2) les méthodes qui consistent à rechercher ou à construire un espace de représentation commun aux deux domaines ; (3) les méthodes qui ajustent le modèle appris sur les données source aux données cible de manière itérative.

3.1 Source Scaling

La première approche, appelée Source Scaling (SC), consiste à mettre la distribution de \mathbf{S} à l’échelle de celle de \mathbf{T} et représente donc une méthode de repondération. Cette approche proposée par [3] pour le cas semi-supervisé, est adaptée pour notre cas non supervisé. Soit $\mu(\mathbf{S})$, $\mu(\mathbf{T})$, $\sigma(\mathbf{S})$ et $\sigma(\mathbf{T})$ les vecteurs (pour chaque caractéristique) des moyennes et écart-types de \mathbf{S} et \mathbf{T} . La repondération est alors

$$\mathbf{S} = \frac{\mathbf{S} - \mu(\mathbf{S})}{\sigma(\mathbf{S})} \times \sigma(\mathbf{T}) + \mu(\mathbf{T})$$

Cette méthode a l’avantage d’être simple, mais elle peut échouer à capturer une transformation non linéaire entre les domaines source et cible.

3.2 Subspace Alignment

Cette méthode d’adaptation non supervisée [4] (SA) cherche à projeter \mathbf{S} et \mathbf{T} dans deux espaces où la divergence entre les deux domaines est minimisée. Afin de trouver ces espaces, SA utilise l’analyse en composante principale (ACP) sur les ensembles \mathbf{S} et \mathbf{T} pour obtenir leurs matrices de vecteurs propres X_s et X_t en ne gardant que les d dimensions ayant la plus grande variance. X_s est ensuite alignée à X_t via la matrice M qui minimise la fonction de coût $F(M) = \|X_s M - X_t\|_F^2$ dont la solution est $M = X_s' X_t$. \mathbf{S} et \mathbf{T} sont ensuite projetés dans leur sous-espace via $\mathbf{S} = \mathbf{S} X_s M$ et $\mathbf{T} = \mathbf{T} X_t$. Malgré sa simplicité, cette méthode présente de très bons résultats en pratique et reste une des approches les plus utilisées, son seul inconvénient étant le coût computationnel élevé de l’ACP.

3.3 Transport optimal

La théorie du transport introduite par Gaspard Monge et récemment développée par Cédric Villani [7], fournit un outil mathématiquement fondé pour aligner deux distributions probabilistes arbitraires de manière optimale. Le problème du transport optimal dans le cas discret est de trouver une matrice de couplage γ de deux distributions définie comme une loi de probabilité jointe f sur $\mathbf{S} \times \mathbf{T}$ de lois marginales f_S et f_T telles que pour tout $x \in \mathbf{S}$, $y \in \mathbf{T}$, on minimise le coût de transport de f_S vers f_T pondéré par une fonction $c : \mathbf{S} \times \mathbf{T} \rightarrow \mathbb{R}^+$, i.e. :

$$\gamma_0 = \arg \min_{\gamma \in \Pi(f_S, f_T)} \langle C, \gamma \rangle_F$$

où $\langle \cdot, \cdot \rangle_F$ est le produit matriciel de Frobenius, $\Pi(f_S, f_T) = \{\gamma \in \mathbb{R}_+^{N_S \times N_T} | \gamma \mathbf{1} = f_S, \gamma^T \mathbf{1} = f_T\}$ et $C_{ij} = c(x_i^S, x_j^T)$. La formulation originale du transport optimal (abrégée TO) est un problème d’optimisation qui passe mal à l’échelle à cause de sa complexité computationnelle assez élevée. Ce problème a été résolu par [5] qui a proposé de rajouter un terme de régularisation utilisant l’entropie de γ . Cette régularisation (abrégée TO2) autorise les exemples source à être transportés de manière plus ou moins uniformes suivant un paramètre de régularisation λ à fixer. Cela permet aussi d’optimiser cette variante du transport optimal de manière efficace via l’algorithme de Sinkhorn-Knopp.

L’utilisation du transport optimal pour l’adaptation de domaine a été étudiée pour la première fois dans [6] où C est définie comme la distance euclidienne au carré entre chaque paire d’exemples $(x^S, x^T) \in \mathbf{S} \times \mathbf{T}$. Après avoir trouvé la matrice de couplage optimale γ_0 , les exemples source sont transportés sur les exemples cible via :

$$\mathbf{S} = \text{diag}((\gamma_0 \mathbf{1})^{-1}) \gamma_0 \mathbf{T}$$

Les auteurs présentent également une nouvelle variante de transport (abrégée TO3) avec l’ajout d’une régularisation $\ell_p \ell_1$. Cette variante empêche plus ou moins (suivant un paramètre η à fixer) les exemples source ayant une étiquette de classe différente d’être transportés vers les mêmes exemples cible. Les différences entre ces trois méthodes sont illustrées dans la figure 1.

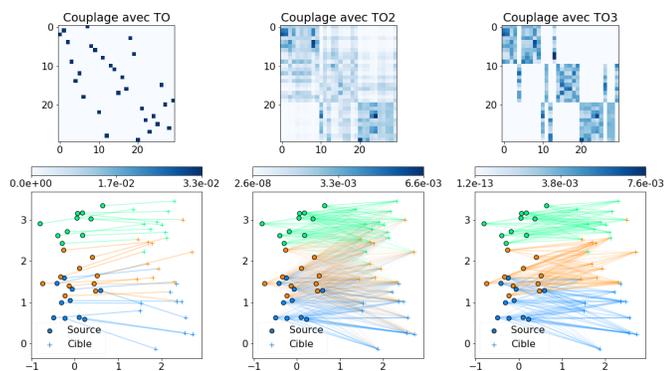


FIGURE 1 – Comparaison des méthodes TO, TO2, et TO3 sur un exemple jouet constitué de 3 classes. Haut : matrice de couplage γ_0 ; le bleu foncé indique une grande valeur de couplage. Bas : transport des points source vers les points cible ; les segments prononcés indiquent un grand couplage.

4 Expériences

Dans cette étude, nous avons évalué les performances des différentes approches d’adaptation non supervisée décrites précédemment dans le contexte applicatif d’un système d’aide au diagnostic du cancer de la prostate en imagerie IRM multiparamétrique [2]. Une description plus détaillée des données et descripteurs utilisés dans cette étude est donnée dans [1, 3].

4.1 Description des données

La base d’images IRM source (S) est constituée de 49 examens réalisés sur des patients atteints d’un cancer de la prostate avec un scanner IRM 1.5 Tesla (Symphony, Siemens) [1]. Chaque examen comprend 3 séquences d’acquisitions (T2, DWI et DCE). La base cible (T) est constituée de 41 examens réalisés selon un protocole similaire sur un scanner 3 Tesla (Discovery MR750, GE). Les IRM des deux domaines ont été annotées (détournement des lésions) par deux radiologues, puis confrontées à la vérité histopathologique (après prostatectomie) afin de caractériser l’agressivité de chaque lésion (score de Gleason).

4.2 Méthodologie

Chaque voxel de la zone périphérique de la prostate est décrit par un vecteur de 95 caractéristiques et par une étiquette de classe. Nous considérons dans cette étude un cas de classification binaire : les voxels caractérisés par un Gleason < 7 (peu agressif) sont considérés comme non cancer (-1), les autres comme cancer (+1). La table 1 indique le nombre de voxels dans chacune des classes des deux domaines. Les 95 caractéristiques [1] incluent des descripteurs d’images des 3 séquences (niveau de gris, texture, gradient) et des caractéristiques pharmacocinétiques. La figure 2 illustre les différences observées sur 2 des 95 caractéristiques entre les voxels source et cible.

Pour le domaine source, nous avons considéré un sous-ensemble de 1500 voxels en prenant 500 voxels de la classe majo-

TABLE 1 – Répartition des voxels

Classe	#voxels S	%	#Voxels T	%
Non Cancer	323 551	77%	786 964	80%
Suspect	31 924	8%	48 604	5%
Gleason 6	7 747	2%	10 988	1%
Gleason 7	33 507	8%	97 099	10%
Gleason 8	15 407	4%	12 372	1%
Gleason 9	7 212	2%	31 369	3%
Total	419 348		987 396	

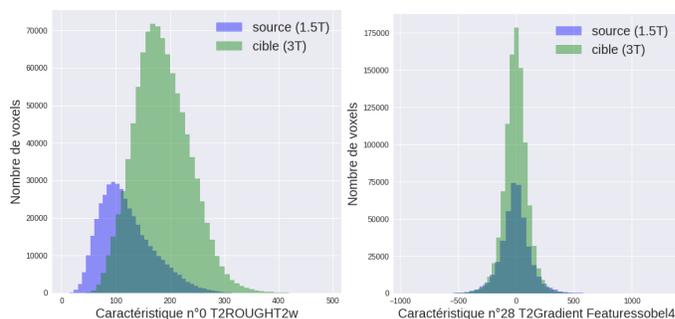


FIGURE 2 – Comparaison de deux caractéristiques entre domaines source (bleu) et cible (vert). À gauche les distributions sont différentes alors qu’elles sont les mêmes à droite.

ritaire (Non cancer) puis 200 pour chacune des 5 autres classes. La sélection aléatoire des voxels d’une classe est réalisée uniformément parmi les patients source présentant des voxels de cette classe. Pour le domaine cible, nous avons sélectionné aléatoirement 50 000 voxels uniformément parmi les 41 patients. L’algorithme TO est utilisé pour transporter les 1500 voxels source vers les 50 000 voxels cible. Pour chaque voxel source, on garde le voxel cible sur lequel il a été transporté. On obtient ainsi 1500 voxels cible dont l’échantillonnage est proche de l’échantillonnage réalisé sur le domaine source tout en n’utilisant aucune étiquette cible. Cette méthode permet ainsi de réaliser une sélection non supervisée des voxels cible utilisés pour l’adaptation.

Pour chaque méthode d’adaptation présentée dans la section 3, on adapte les 1500 voxels source sur les 1500 voxels cible. L’apprentissage se fait ensuite à l’aide d’un SVM linéaire uniquement sur les 1500 voxels source adaptés. Les paramètres des méthodes d’adaptation de domaine et du SVM sont fixés à l’aide d’une recherche sur grille par une méthode de validation croisée de type LOPO (Leave-one-patient-out) [2, 3] sur les 1500 voxels source adaptés. L’aire sous la courbe ROC (AUC) est calculée individuellement pour les 36 patients source (parmi les 49) contenant des voxels des 2 classes. L’AUC moyenne sur les 36 patients est utilisée comme critère de qualité.

Les performances des modèles optimisés (pour l’adaptation et la classification) sont finalement évaluées par LOPO sur les 32 patients cible (parmi les 41) représentant les 2 classes. Les différentes stratégies d’adaptation sont comparées à une méthode de référence (abrégée N) consistant à entraîner le SVM sur les 1500 voxels source sans adaptation puis à tester ce mo-

dèle sur les 32 patients cible.

5 Résultats

Les performances présentées dans la table 2 montrent que la méthode N donne une AUC très faible, de 50.0% ce qui prouve l'utilité de faire une adaptation. Les méthodes de transport optimal avec régularisation entropique (TO2 et TO3) donnent respectivement des performances de 68.3% et 77.5%. L'ajout de la régularisation de classe permet d'améliorer les performances, comme ce qui était attendu. Le transport optimal simple est également légèrement moins performant que les méthodes SA et SC qui produisent une AUC moyenne de 81.5% et 84.6% respectivement. Le test de rang de Wilcoxon met en évidence une différence significative entre les performances de la méthode SC et celles des autres méthodes (p -values $< 10^{-6}$). Les performances de SA et TO sont les seules à ne pas être significativement différentes.

Malgré des valeurs d'AUC moins bonnes pour TO2 et TO3, la figure 3 montre que la lésion cancéreuse (contourée en rouge sur la vérité terrain) est détectée avec une probabilité moyenne d'appartenir à la classe cancer supérieure pour les cartes obtenues avec TO2 et TO3. Ces différences sont quantifiées dans la table 2 par les indices PC et PNC mesurant ces valeurs moyennes de probabilité pour les voxels des classes cancer (+1) et non cancer (-1) respectivement. On s'aperçoit que les écarts entre les deux classes sont plus importants pour les méthodes TO2 et TO3. Or, dans un cas clinique, il peut être intéressant d'avoir une forte différence entre différentes classes. Le transport optimal avec régularisation a donc son utilité.

TABLE 2 – Comparaison des AUC moyennées (sur les patients) et des probabilités moyennes d'être un voxel cancer pour les voxels des classes cancer (PC) et non cancer (PNC).

Méthode	N	SA	SC	TO	TO2	TO3
AUC	50.1	81.5	84.6	80.0	68.3	77.5
PC	00.0	60.0	52.8	51.3	63.6	63.6
PNC	00.0	34.1	25.2	33.4	33.7	23.8

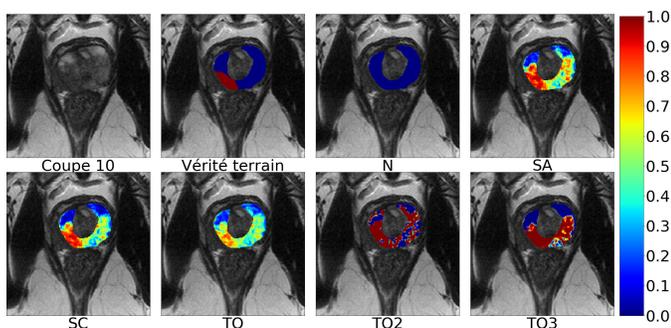


FIGURE 3 – Coupe transversale d'un patient cible ayant une lésion agressive (en rouge sur la vérité terrain) et cartographie probabiliste obtenue à l'aide des différentes stratégies d'adaptation de domaine.

6 Conclusion

Dans ce papier, nous avons présenté une étude empirique des méthodes d'adaptation de domaine appliquées aux données provenant du domaine de l'imagerie médicale. Nous nous sommes placés dans le cadre de l'adaptation de domaine non supervisée : seules les données source, constituées ici de volumes d'images de prostate acquises sur un scanner IRM 1.5T, sont étiquetées ; les données cible composées d'acquisitions obtenues sur un scanner 3T sont non étiquetées. Les résultats de cette étude ont montré un fort potentiel des méthodes d'adaptation de domaine pour la problématique considérée, exprimé par une amélioration importante des résultats de classification par rapport au cas où aucune adaptation n'est faite. Dans l'avenir, nous allons chercher à améliorer les performances des méthodes basées sur le transport optimal en considérant différentes métriques de distance pour définir le coût de transport. Une extension de ces travaux au cas de la classification multi-classe est également prévue.

Références

- [1] E. Niaf, O. Rouvière, F. Mège-Lechevallier, F. Bratan et C. Lartzien. *Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI*. Phys. Med. Biol., 57(12) : 3833-3851, 2012.
- [2] J. Lehaire, R. Flamary, O. Rouvière et C. Lartzien. *Computer-aided diagnostic system for prostate cancer detection and characterization combining learned dictionaries and supervised classification*. IEEE ICIP, 2251-2255, 2014.
- [3] R. Aljundi, J. Lehaire, F. Prost-Boucle, O. Rouvière et C. Lartzien. *Transfer Learning for Prostate Cancer Mapping Based on Multicentric MR Imaging Databases*. In ICML, workshop MLMMI, 2015.
- [4] B. Fernando, A. Habrard, M. Sebban et T. Tuytelaars. *Un-supervised visual domain adaptation using subspace alignment*. Proceedings of ICCV, 2960-2967, 2013.
- [5] M. Cuturi. *Sinkhorn distances : Lightspeed computation of optimal transport*. Proceedings of NIPS, 2292-2300, 2013.
- [6] N. Courty, R. Flamary et D. Tuia. *Domain adaptation with regularized optimal transport*. Proceedings of ECML/PKDD, 1-16, 2014.
- [7] C. Villani. *Optimal transport, old and new*. Springer Science & Business Media, 2008.
- [8] S. J. Pan et Q. Yang. *A survey on transfer learning*. In IEEE Trans. Knowl. Data Eng., 22(10) :1345-1359, 2010.
- [9] A. van Opbroek, M. A. Ikram, M. W. Vernooij et M. de Bruijne. *Transfer learning improves supervised image segmentation across imaging protocols*. In IEEE Trans Med Imaging, 34(5) :1018-30, 2015.