HAL
open science

Sélection de caractéristiques pour l'étude de la maturation cérébrale

Julien Pontabry, François Rousseau

To cite this version:

Julien Pontabry, François Rousseau. Sélection de caractéristiques pour l'étude de la maturation cérébrale. ORASIS - Journées francophones des jeunes chercheurs en vision par ordinateur, Jun 2013, Cluny, France. hal-00829413

HAL Id: hal-00829413

https://hal.science/hal-00829413v1

Submitted on 5 Jun 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sélection de caractéristiques pour l'étude de la maturation cérébrale

J. Pontabry F. Rousseau

ICube, Université de Strasbourg, CNRS

Résumé

La caractérisation des images IRM cérébrales fœtales par la mise en place de biomarqueurs est un outil essentiel pour l'aide au diagnostic précoce. Dans ce contexte, il est intéressant d'extraire des marqueurs fiables de maturation cérébrale in-utero par le biais de méthodes de sélection de caractéristiques permettant de conserver l'interprétation physique des paramètres. Dans ce travail, nous étudions la morphométrie de la surface corticale du lobe frontal chez le fœtus.

Mots Clés

Réduction de dimension, sélection de caractéristiques, maturation cérébrale, morphométrie, surface corticale

Abstract

The study of maturation patterns in MRI fœtal brain images by the establishment of biomarkers are essential tools for early diagnosis. In that context, it is interesting to extract reliable markers of the brain maturation in-utero, within a feature selection framework. Thus, the physical meaning of the parameters is preserved. In this work, we investigate the morphometry of the cortical surface of the frontal lobe in fetal brain.

Keywords

Dimension reduction, feature selection, brain maturation, morphometry, cortical surface

1 Introduction

Les récentes avancées technologiques des systèmes d'imagerie par résonance magnétique couplées à des progrès dans le traitement des images acquises permettent l'étude morphométrique in-utero de la maturation cérébrale chez l'Homme. Cependant, la caractérisation temporelle du développement cérébral demeure délicate par la taille de l'espace à considérer en rapport au nombre d'images acquises. La réduction de la dimension des données est alors un enjeu primordial pour aller plus loin dans leur compréhension.

Un moyen très répandu pour réduire la dimension d'un problème est l'extraction de caractéristiques. Ces méthodes consistent à créer un nouvel espace de caractéristiques dans lequel celles-ci sont obtenues par combinaison ou transformations des caractéristiques originales. Ces techniques sont limitées car les caractéristiques finales ne contiennent pas d'interprétation physique simple. La sélection de caractéristiques est un moyen de réduire la dimension d'un problème tout en gardant le sens physique de l'ensemble réduit de paramètres (voir Guyon et coll. [1] pour une revue des méthodes).

Dans cet article nous étudions la réduction de la dimension d'un problème dans le contexte de l'analyse de la maturation cérébrale en déterminant une méthode de sélection de caractéristiques appropriée. Les contributions sont les suivantes :

- définition d'une modélisation non-paramétrique de la maturation cérébrale ;
- caractérisation d'une méthode de sélection de caractéristiques pour réduire la dimensionnalité du problème.

2 Théorie

La maturation cérébrale, au sens morphométrique, est modélisée par une méthode non-paramétrique à noyaux utilisée pour la création d'atlas longitudinal par régression non-linéaire [2]. Soient p le nombre de paramètres et N le nombre d'échantillons. Le nombre d'échantillons est supposé beaucoup plus petit que le nombre de paramètres du problème, c'est-à-dire $N << p$. L'ensemble des données forme une matrice $\boldsymbol{Y} \in \mathcal{M}_{p\times N}$. L'idée est de trouver le sous-ensemble de paramètres explicatifs tel que l'erreur de reconstruction soit minimale et cet ensemble soit le plus petit possible.

Soient X_i le vecteur de taille $p' \leq p$ et Y_i la i-ème colonne de $\boldsymbol{Y}$ correspondant à l'échantillon i (pour $i = 1 \ldots N$). Les deux vecteurs Y_i et X_i sont liés par une fonction de reconstruction :

$$Y_i = f\left(X_i\right) + \epsilon_i \ , \tag{1}$$

où ϵ_i est l'erreur de reconstruction. Pour $p' = 1$ et $X = t$, l'âge gestationnel, l'équation (1) correspond au problème de la création d'atlas longitudinaux [2]. Dans cette étude préliminaire, nous cherchons X_i un sous-ensemble de Y_i. La fonction de reconstruction f est définie par l'estimateur à noyaux de Nadaraya-Watson :

$$\hat{f}\left(X\right) = \sum_{i=1}^{N} w_i\left(X\right)\ Y_i \ , \tag{2}$$

où la fonction de poids s'écrit

$$w_i\left(X\right) = \frac{K_h\left(X - X_i\right)}{\sum_{j=1}^{N} K_h\left(X - X_j\right)} \ . \tag{3}$$

Dans cette fonction de poids, K_h est un noyau de largeur de bande h.
Nous cherchons à trouver l'ensemble de paramètres le plus petit possible tel que l'erreur de reconstruction soit minimale. La fonction de coût étudiée dans ce travail fait donc apparaître l'erreur de reconstruction ϵ_i. Soit ω un vecteur binaire d'activation de taille p. Ce vecteur permet d'activer ou non certains paramètres. Ainsi, la fonction de coût s'écrit

$$J(\omega) = \sum_{i=1}^{N} \left\| Y_i - \hat{f}\left(X_i^T \cdot \mathrm{diag}(\omega)\right) \right\|_2^2 \quad , \qquad (4)$$

où $\mathrm{diag}(\omega)$ est la matrice diagonale ayant pour coefficients diagonaux le vecteur ω. Une régularisation est incluse dans la modélisation à noyaux, par le paramètre de largeur de bande du noyau.
Dans ce travail, la fonction J est minimisée à l'aide d'un algorithme glouton [3] [1]. Sa complexité moyenne, en nombre d'évaluations de J, est asymptotique en $\mathcal{O}(p^2)$. Les paramètres sont ajoutés itérativement à partir de l'ensemble vide en alternant l'ajout et la suppression d'un paramètre diminuant la valeur de J.
Chaque ajout de paramètre s'accompagne d'une décroissance de J qui donne une certaine énergie au paramètre (différence entre la valeur de J à l'étape précédente et à l'étape courante). Rapporté à la valeur initiale de J ($X_i = \emptyset$), cette énergie fournit un pourcentage de la variance expliquée. La condition d'arrêt de l'algorithme utilisée dans cette étude est atteinte lorsque la variance expliquée de l'ensemble réduit atteint 99 %.

3 Résultats

Les images IRM multi-coupes sont débruitées par un algorithme de moyennes non-locales et reconstruites par des méthodes de super-résolution [4]. Ensuite, les tissus cérébraux sont estimés en utilisant un algorithme de segmentation dédié aux images IRM fœtales reposant sur les c-moyennes floues et des contraintes topologiques [4]. Enfin, l'atlas est construit à partir de cette base d'apprentissage par des méthodes non-paramétriques [2].
Dans ce travail, nous avons appliqué l'algorithme présenté dans la section précédente à un ensemble d'images IRM fœtales en nous concentrant sur la surface corticale du lobe frontal droit. Les données sont représentées par les champs de déformation de chaque image vers l'atlas estimé. Dans cette expérience, $p = 2592$ et $N = 23$. La largeur de bande a été empiriquement fixée à 1. Six minutes ont été nécessaire sur une machine standard pour trouver une solution.
L'application de l'algorithme a permis de sélectionner 5 points d'intérêt représentant 99 % de la variance expliquée, principalement disposés le long de la scissure interhémisphérique et du sillon latéral (voir figure 1). Cette méthode permettrait par exemple de caractériser l'évolution temporelle du cortex au cours du développement fœtal.
Les résultats préliminaires sont encourageants, cependant quelques points restent à étudier. Cela concerne notamment le temps de calcul qui s'accroît fortement avec le nombre de paramètres initiaux et le choix des paramètre de régularisation (largeur de bande h du noyau du modèle).

1. Un algorithme glouton consiste à approcher une solution globalement optimale, de proche en proche, par choix localement optimaux successifs.

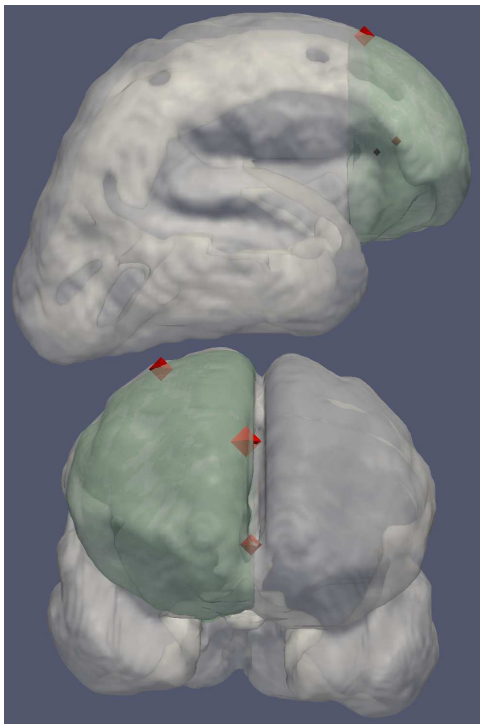

FIGURE 1 – Résultat d'une réduction de 2592 paramètres à 5 paramètres représentant 99 % de la variance expliquée. Les données d'entrée sont des champs de vecteur sur la surface corticale du lobe frontal droit (en transparence). L'ensemble réduit des paramètres est indiqué par des points de couleur (du noir au rouge) et de taille (du plus petit au plus grand) variables selon la part de variance expliquée de chaque paramètre.

Remerciements Les recherches aboutissant à ces résultats sont financées par le Conseil Européen de la Recherche sous le Septième Programme Cadre de la Communauté Européenne (FP7/2007-2013 agrément numéro. 207667)

Références

[1] I. Guyon et A. Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3 :1157–1182, 2003.

[2] J. Pontabry et col. Longitudinal probabilistic atlas of the fetal brain. In *Nouvelles méthodologies en imagerie du vivant*, Lyon, Décembre 2012.

[3] P. Pudil et col. Floating search methods in feature selection. *Pattern Recognition Letters*, 15(11) :1119–1125, 1994.

[4] F. Rousseau et col. BTK : an open-source toolkit for fetal brain mr image processing. *Computer Methods and Programs in Biomedicine*, 109(1) :65–73, 2013.