Fusion tardive en analyse de données fonctionnelles élastique

Julien Ah-Pine, Noé Lebreton

Fusion tardive en analyse de données fonctionnelles élastique

Julien Ah-Pine [1,2,3] & Noé Lebreton [1]

[1] *Univ Lyon, Univ Lyon 2, ERIC UR3083; julien.ah-pine@univ-lyon2.fr*
[2] *Université Clermont Auvergne, LMBP-CMRS, UMR 6620*
[3] *Université Clermont Auvergne, CERDI-CNRS-IRD, UMR 6587*

Résumé. Dans cette communication nous comparons la fusion précoce et la fusion tardive en classification supervisée de données fonctionnelles lorsqu'une séparation des variations d'amplitude et de phase est pertinente. Le cadre méthodologique utilisé est l'analyse de données fonctionnelles élastique qui passe par un alignement des fonctions basé sur la distance de Fisher-Rao. Ensuite, afin de considérer conjointement les variations d'amplitude et de phase, la fusion précoce concatène les fonctions alignées et les fonctions de déformation temporelle dans une fonction composite avant réduction de dimension et apprentissage du modèle de classification. Cette méthode est proposée dans la littérature récente. *A contrario*, nous examinons une fusion tardive dans laquelle, deux réductions de dimension et apprentissages de classifieurs sont appliquées indépendamment sur les deux types de fonctions d'amplitude et de phase, et ce sont les estimations des prédictions des deux modèles qui sont fusionnées par un opérateur d'agrégation.

Mots-clés. Analyse de données fonctionnelles, variabilité d'amplitude et de phase, métrique élastique, fusion de données.

Abstract. In this paper, we compare early and late fusion schemes in classification tasks of functional data when amplitude and phase separation is relevant. We rely on the elastic functional data analysis framework where functions alignment is carried out using the Fisher-Rao distance. Then, in order to jointly consider amplitude and phase variabilities, the early fusion concatenates in a composite function, both the aligned functions and the time warpings, before applying dimension reduction and learning a classifier. This approach is proposed in the recent literature. In contrast, we examine a late fusion method where two dimension reductions and classifiers are independently estimated using the two types of functions and it is the estimated predictions that are combined by means of an aggregation operator.

Keywords. Functional data analysis, amplitude and phase variabilities, elastic metric, data fusion.

1 Introduction

Nous nous intéressons à l'analyse de données fonctionnelles (*FunctionalData Analysis* - FDA par la suite-) définie sur $[0, 1]$ dans les cas où il y a un intérêt à séparer les variations

d'amplitude et de phase comme, par exemple, lorsque les observations présentent des problèmes d'alignement. Dans ce contexte, l'utilisation de la distance de Fisher-Rao, dite métrique élastique (voir par exemple [3]), permet d'aligner les fonctions de façon appropriée et d'isoler en deux vues distinctes, des fonctions mettant en perspective des disparités au niveau des amplitudes (variation sur l'axe vertical des ordonnées) d'une part; et des fonctions de re-paramétrisation encodant des variabilités de phase dans le temps (variation sur l'axe horizontal) d'autre part.

Dans ce cadre, nous nous intéressons au problème de la fusion de ces deux sources d'informations complémentaires pour la tâche de classification supervisée. Nous abordons ce sujet en opposant d'un côté une fusion précoce qui concatène les deux types de fonctions encodant respectivement l'amplitude et la phase; et d'un autre côté, une fusion tardive qui raisonne au niveau des estimations données par deux modèles appris de façon indépendante sur l'un et l'autre des deux types de fonctions.

Nous illustrons sur deux jeux de données réelles, les avantages potentiels de la fusion tardive que nous proposons vis-à-vis de la fusion précoce déjà proposée dans la littérature.

2 Analyse de données fonctionnelles élastique

Dans ce qui suit, f est une fonction de $\mathcal{F}$, l'ensemble des fonctions absolument continues définies sur $[0,1]$; γ est une fonction de Γ, l'ensemble des difféomorphismes de $[0,1]$ avec $\gamma(0)=0$ et $\gamma(1)=1$. La fonction $f\circ\gamma$ correspond à une re-paramétrisation de f. Supposons, sans perte de généralité, que le support $[0,1]$ est un axe temporel, alors $f\circ\gamma$ s'interprète comme un recalage de f par la déformation temporelle γ. Nous notons par q une fonction de $\mathbb{L}^2$, l'espace de Hilbert des fonctions définies sur $[0,1]$. Les notations précédentes sont génériques et lorsque nous travaillons avec un échantillons de n fonctions, nous ajoutons un indice inférieur i (parcourant la suite $1,\ldots,n$) à chacune de ces notations, le cas échéant.

2.1 Alignement des fonctions basée sur la transformation SRVF

Le problème d'alignement en FDA se pose lorsque, par exemple, il existe un décalage dans les observations des différentes fonctions, ce qui produit des imprécisions sur les mesures de distance entre celles-ci. Pour tenir compte de ce problème, les auteurs de [3] proposent d'utiliser la distance de Fisher-Rao, notée d_{FR}, entre fonctions. Cette métrique Riemanienne est adaptée au problème traité en raison de son invariance par rapport aux re-paramétrisations de Γ: $d_{FR}(f_1,f_2)=d_{FR}(f_1\circ\gamma,f_2\circ\gamma)$.

Cette métrique étant Riemanienne, elle est difficile à évaluer dans le cas général. Or, il existe une isométrie transformant les fonctions f de $\mathcal{F}$ en des fonctions q de $\mathbb{L}^2$ avec $d_{FR}(f_1,f_2)=\|q_1-q_2\|_{\mathbb{L}^2}$. Cette propriété permet ainsi de réduire le coût computationnel des calculs. Cette transformation est donnée par $q(t)=\text{sign}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$ où $\dot{f}$ est la

dérivée de f. q est appelée la *Square Root Velocity Function (SRVF)* de f [3] et nous avons par ailleurs la relation suivante: $f(t) = f(0) + \int_0^t q(s)|q(s)|ds$.

Nous renvoyons le lecteur à [3] pour le détail des procédures mettant en oeuvre, en théorie et en pratique, l'alignement dans le cadre précédent. Nous présentons plus dans le détail ce qui nous intéresse pour la suite et qui concerne les fonctions issues de ces procédures.

Si nous avons un échantillon de n fonctions de $\mathcal{F}$, $\{f_i\}_{i=1,\ldots,n}$, ce qui nous intéresse sont d'une part, les fonctions alignées dénotées $\{\tilde{f}_i\}_i$ et leurs SRVFs associées $\{\tilde{q}_i\}_i$ et d'autre part, les fonctions de re-paramétrisation ayant permis le recalage des fonctions et qui sont dénotées $\{\gamma_i\}_i$. En effet, $\{\tilde{f}_i\}_i$ et $\{\tilde{q}_i\}_i$ permettent d'isoler l'étude de la variabilité d'amplitude de celle de la phase qui est elle permise par l'analyse de l'ensemble $\{\gamma_i\}_i$.

2.2 Analyse séparée de l'amplitude et de la phase

Dans [4], un cadre méthodologique basé sur l'analyse en composante principale fonctionnelle (*Functional Principal Component Analysis* -FPCA par la suite-) de SRVFs dans $\mathbb{L}_2$ est proposé, afin d'analyser séparément l'amplitude et la phase.

Dans le premier cas, il s'agit d'étudier les fonctions SRVF alignés $\{\tilde{q}_i\}_i$, augmentées des valeurs $\{f_i(0)\}_i$. L'ajout de ces dernières valeurs fait écho à la relation liant les fonctions et leurs SRVFs rappelée précédemment. Pour analyser l'amplitude, [4] propose de réaliser dans $\mathbb{L}_2 \times \mathbb{R}$ une FPCA sur les fonctions $\{[\tilde{q}_i \quad f_i(0)]\}_i$.

Concernant la phase, l'approche prend comme éléments de départ les fonctions $\{\gamma_i\}_i$ ayant conduit au recalage des $\{f_i\}_i$. La distance de Fisher-Rao est utilisée à nouveau (Γ n'est d'ailleurs pas un espace linéaire) et on calcule ainsi les SRVF des $\{\gamma_i\}_i$ que l'on dénote $\{\psi_i\}_i$. Dans ce cas particulier, nous avons la propriété suivante: $\psi_i = \sqrt{\dot{\gamma}_i}$. Ceci implique que dans $\mathbb{L}_2$, les SRVF $\{\psi_i\}_i$ appartiennent à l'hypersphère de rayon unité. L'espace n'étant pas linéaire, il n'est pas possible d'appliquer une FPCA directement. Pour contourner le problème, [4] propose de projeter (il s'agit d'une bijection) les $\{\psi_i\}_i$ sur l'espace tangent à la moyenne de Karcher des $\{\psi_i\}_i$. Ces *shooting vectors* dénotés $\{v_i\}_i$ sont dans un espace linéaire et il alors possible d'appliquer la FPCA.

3 Fusion des vues amplitude et phase dans la classification supervisée de données fonctionnelles

La plupart des travaux qui étudient des données fonctionnelles (*Functional Data* -FD par la suite-) présentant des variations d'amplitude et de phase, alignent les fonctions puis analysent les fonctions alignées sans tenir compte des déformations temporelles. Or, la variabilité de phase peut apporter une information discriminante dans l'analyse. Dans ce contexte, [2] combine les fonctions alignées aux fonctions de re-paramétrisation en amont d'une FPCA. En termes de fusion d'information, l'approche peut être qualifiée de fusion

précoce (*early fusion*). Nous explorons *a contrario*, une alternative dite de fusion tardive (*late fusion*). Nous étudions ce tème dans le cadre de la classification supervisée de FD. Dans ce cas, une fonction f est associée à un label y, élément d'un espace fini discret $\mathcal{Y}$.

3.1 Fusion précoce

Dans la fusion précoce, on concatène les représentations des différentes vues en une représentation augmentée et composite avant d'appliquer une quelconque analyse. Dans notre contexte, [2] propose la fonction composite g^C définie sur $[0, 2]$ de la façon suivante:

$$g^C(t) = \begin{cases} \tilde{f}(t), & t \in [0, 1[\\ Cv(t-1), & t \in [1, 2] \end{cases}$$

où le paramètre $C > 0$ permet de gérer le déséquilibre d'échelle entre $\tilde{f}$ et v.

Nous nous intéressons aux travaux de [5] qui utilisent la fusion précoce précédente dans la tâche de classification supervisée. Contrairement à [2], [5] préconise d'utiliser $\tilde{q}$ plutôt que $\tilde{f}$ dans la définition de g^C car ces premières sont garanties d'être dans $\mathbb{L}_2$. Nous retenons cette proposition dans ce qui suit et dans nos expériences. Une fois la FPCA appliquée aux fonctions $\{g_i^C\}_i$, [5] utilise la régression logistique sur composantes principales comme modèle de classification (voir également [1]).

3.2 Fusion tardive

Dans l'approche que nous promouvons, la combinaison des informations d'amplitude et de phase se fait tardivement. La démarche consiste à apprendre deux modèles de prédiction, l'un centré sur l'amplitude et l'autre basé sur la phase et la classification finale s'effectue en combinant les estimations de prédiction des deux classifieurs. L'avantage ici est que chaque modèle se focalise sur un seul type d'information ce qui permet de différencier, si besoin, les traitements de chacune des deux vues. De plus, en fusionnant *a posteriori* les prédictions, cela permet d'avoir du recul sur laquelle des deux vues est plus pertinente et si le mélange permet par ailleurs de meilleurs performances.

Dans notre contexte, en pratique, nous réutilisons les résultats des FPCA introduites dans la sous-section 2.2 et qui traitent les fonctions $\{\{[\tilde{q}_i \quad f_i(0)]\}_i\}$ indépendamment des fonctions $\{\gamma_i\}_i$. Nous obtenons deux dimensions de réduction distinctes, l'une pour l'amplitude, l'autre pour la phase. Ces deux espaces peuvent par ailleurs être de dimensions différentes. Ensuite, nous apprenons deux régressions logistiques sur composantes principales dans ces deux espaces de description.

Supposons que le modèle d'alignement et les deux régressions logistiques sur composantes principales aient été estimés à partir des $\{f_i\}$. Étant donné $f \in \mathcal{F}$, voici successivement les étapes pour obtenir sa prédiction: (1) calcul de la SRVF q, alignement et obtention de $\tilde{q}$ et γ; (2) projection de $[\tilde{q} \quad f(0)]$ dans l'espace réduit de l'amplitude et γ dans celui de la phase; (3) calcul des probabilités d'appartenance à chaque classe y pour

chacun des deux modèles de régression logistique; (4) fusion des probabilités de prédiction des deux classifieurs.

Nous détaillons l'étape (4) qui est au coeur de ce travail. Notons $P_V(y|f)$ la probabilité conditionnelle d'observer y, donnée par le classifieur basé sur l'amplitude et $P_H(y|f)$ celle donnée par le classifieur basé sur la phase. L'opérateur de fusion tardive que nous proposons est la moyenne pondérée:

$$P(y|x) = \alpha P_V(y|x) + (1-\alpha)P_H(y|x)$$

où $\alpha \in [0,1]$ est un paramètre permettant de donner des poids différents aux deux vues.

La règle de classification finale est alors $\arg\max_{y\in\mathcal{Y}} P(y|f)$. Afin de proposer une approche simple pour fixer α, nous utilisons également une fusion guidée par les performances respectives de chaque vue. Notons par acc_V et acc_H les taux de reconnaissance obtenus, sur un jeu de données de validation par exemple, par l'amplitude et la phase. Nous proposons d'estimer α et $\beta = 1-\alpha$ par $\text{acc}_V/(\text{acc}_V + \text{acc}_H)$ et $\text{acc}_H/(\text{acc}_V + \text{acc}_H)$.

4 Illustration sur deux jeux de données réels

Nous avons appliqué les méthodes de fusion précoce et tardive décrites précédemment sur deux jeux de données réels connus de la communauté: *Growth* et *Tecator*. Le premier concerne des courbes de croissance et il faut discriminer celles correspondant à des filles et celle correspondant à des garçons. Le deuxième représente des courbes de spectrométrie de viande et il faut indiquer si le taux de graisse est faible ou élevé.

Nous représentons dans la Figure 1, les différentes fonctions initiales, les fonctions de déformation temporelle et les fonctions alignées.

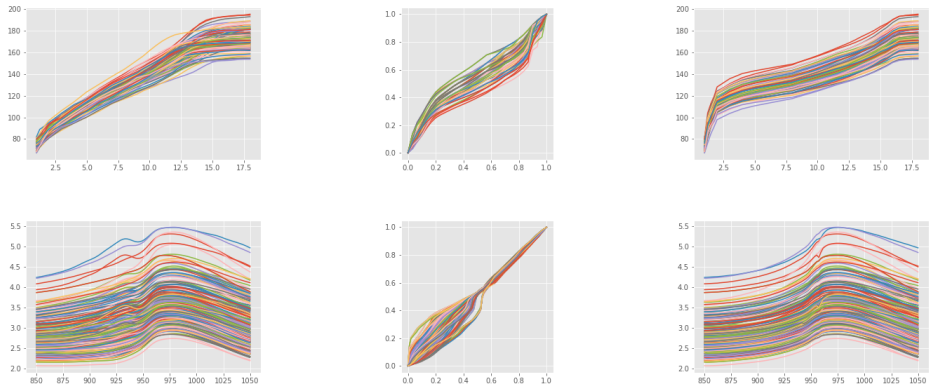

Figure 1: Illustrations des jeux de données: *Growth* 1ère ligne, *Tecator* 2ème ligné. De gauche à droite: $\{f_i\}$, $\{\gamma_i\}$ et $\{\tilde{f}_i\}$

Dans la Figure 2, nous montrons les résultats obtenus pour la fusion précoce avec un paramètre $C = XXX$, la fusion tardive avec $\alpha = \{0, 0.1, \ldots, 1\}$ et la fusion tardive guidée

par les performances deux deux vues. Ces résultats illustrent l'intérêt de la fusion tardive dans la mesure où elle permet d'obtenir des taux de reconnaissance meilleurs que la fusion précoce étant donné que la courbe des points (fusion tardive avec différentes valeurs de α) peut dépasser la ligne de *baseline* (fusion précoce). Les performances reportées pour les données *Tecator* sont particulièrement convaincantes. La fusion tardive guidée par les performances (point jaune) ne donne pas nécessairement les meilleurs résultats et la question de l'estimation du paramètre de mélange α reste donc ouverte.

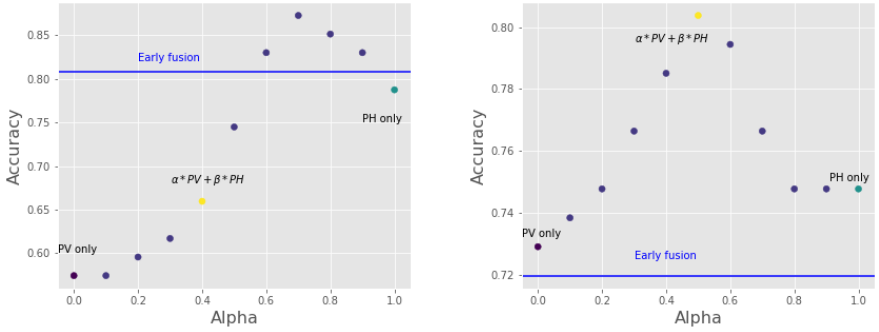

Figure 2: Fusion précoce *vs* fusion tardive: *Growth* à gauche, *Tecator* à droite

References

[1] Ana M Aguilera, Manuel Escabias, and Mariano J Valderrama. Using principal components for estimating logistic regression with high-dimensional multicollinear data. *Computational Statistics & Data Analysis*, 50(8):1905–1924, 2006.

[2] Sungwon Lee and Sungkyu Jung. Combined analysis of amplitude and phase variations in functional data. *arXiv preprint arXiv:1603.01775*, 2016.

[3] Anuj Srivastava, Wei Wu, Sebastian Kurtek, Eric Klassen, and James Stephen Marron. Registration of functional data using fisher-rao metric. *arXiv preprint arXiv:1103.3817*, 2011.

[4] J Derek Tucker. *Functional component analysis and regression using elastic methods*. PhD thesis, The Florida State University, 2014.

[5] J Derek Tucker, John R Lewis, and Anuj Srivastava. Elastic functional principal component regression. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 12(2):101–115, 2019.