

Un critère de sélection de meilleure base d'ondelettes pour la classification

Application aux interfaces cerveau-machines

Denis VAUTRIN¹, Xavier ARTUSI¹, Marie-F. LUCAS¹, Dario FARINA²

¹Institut de Recherche en Communications et Cybernétique de Nantes,
1 rue de la Noë, Nantes, France

²Center for Sensory-Motor Interaction (SMI), Department of Health Science and Technology,
Aalborg University, Aalborg, Denmark

Denis.Vautrin@ircsyn.ec-nantes.fr, Xavier.Artusi@ircsyn.ec-nantes.fr
Marie-Francoise.Lucas@ircsyn.ec-nantes.fr, df@hst.aau.dk

Résumé – Dans le cadre de la reconnaissance de formes, la définition d'un espace de représentation est nécessaire. Nous proposons ici de l'adapter à l'objectif de classification en utilisant la décomposition en paquets d'ondelettes. Notre algorithme s'appuie sur une représentation fréquentielle des signaux (marginales des coefficients des paquets d'ondelettes) qui dépend de la base de décomposition choisie. Nous proposons donc de sélectionner la décomposition fréquentielle la plus pertinente en optimisant la base de décomposition sur une population d'apprentissage. Pour cela, nous avons défini un critère de contraste entre les classes. Il est additif, ce qui nous a permis d'utiliser un algorithme rapide de recherche de meilleure base. Les fonctions de décision sont ensuite apprises par la méthode des SVM. Nous illustrons l'efficacité de cette approche en l'appliquant à des signaux simulés puis à des signaux électroencéphalographiques enregistrés lors de mouvements imaginaires.

Abstract – Any pattern recognition algorithm requires the definition of a representation space. In this study, we propose a method to define a representation space adapted to the objective of classification and based on the wavelet packet decomposition. Our algorithm uses a frequential description of signals (marginals of wavelet packet coefficients). Thus, it depends on the chosen wavelet basis. We propose here to select the most efficient decomposition basis by optimisation on a learning set. With this purpose we defined a contrast criterion between classes. It is additive, so it can be used with a fast best basis search algorithm. The discriminant functions are then learned using the SVM method. In order to illustrate its efficiency, our classification method has been applied on simulated signals and on electroencephalographic signals recorded during imaginary movements.

1 Introduction

Les performances d'un système de reconnaissance de forme dépendent fortement du choix de l'espace de représentation. Celui-ci peut être choisi *a priori*, ou adapté à l'objectif de classification. Les représentations à base d'ondelettes permettent cette adaptation, soit en optimisant l'ondelette mère de la décomposition, soit en optimisant la base de décomposition pour une ondelette mère donnée. La première approche a été étudiée dans [4]. Nous étudions ici la seconde, et proposons un critère de sélection d'une base à partir de la décomposition en paquets d'ondelettes d'une population d'apprentissage.

Nous nous intéressons à la classification de signaux tels que les signaux électroencéphalographiques (EEG) constitués de la superposition d'ondes dont on ne connaît ni la forme, ni les instants d'apparition, et où les classes se distinguent les unes des autres par la répartition de l'activité des signaux dans différentes bandes de fréquence *a priori* inconnues. Dans ce contexte, l'utilisation de la décomposition en paquets d'ondelettes (DWPT) peut s'avérer bénéfique en s'affranchissant de

l'exploration dyadique de l'axe fréquentiel imposée par la DWT. La taille des populations d'apprentissage étant souvent restreinte, il est préférable de ne pas conserver un espace de représentation de grande dimension et de sélectionner une décomposition adaptée à l'objectif et aux signaux.

La méthode proposée dans ce papier est basée sur la définition d'un critère de contraste entre les classes permettant d'obtenir la décomposition fréquentielle la plus pertinente (on définit ainsi le banc de filtres le plus discriminant). Ce critère est additif, ce qui permet l'utilisation d'un algorithme rapide de recherche de l'optimum. Dans la mesure où les signaux considérés ne sont pas nécessairement synchronisés, il est construit à partir des marginales des coefficients des paquets d'ondelettes. Un ensemble d'apprentissage composé de signaux labellisés est utilisé pour sélectionner la base de paquets d'ondelettes la plus discriminante, puis pour apprendre les fonctions de décision (SVM) dans l'espace de représentation défini par cette meilleure base.

Ce papier est organisé de la manière suivante. Nous définis-

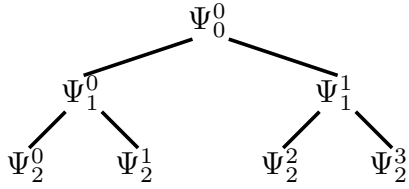


FIGURE 1 – Arbre de paquets d’ondelettes, avec $J = 2$. La sous-base Ψ_j^p explore la bande de fréquence normalisée $2^{-(j+1)}[p, p + 1]$.

sons dans un premier temps l’espace de représentation déduit d’une base particulière de paquets d’ondelettes (§ 2), puis la méthode de sélection de la meilleure base pour optimiser cet espace (§ 3), et les règles de décision dans cet espace (§ 4). Dans un souci de généralité, cette présentation est faite dans le cadre multiclassé. Nous présentons enfin (§ 5) les résultats obtenus, dans un cadre biclasse, d’une part sur des signaux simulés et d’autre part sur des signaux EEG enregistrés.

2 Espace de représentation

La DWPT est une généralisation de la DWT, qui réalise une exploration dichotomique et redondante du contenu fréquentiel d’un signal [5]. Etant donné une fonction échelle $\phi(t)$ et les filtres échelle h et ondelette g , les ondelettes à la résolution $j + 1$ sont définies par :

$$\begin{aligned}\psi_{j+1}^{2p}(t) &= \sum_k h[k] \psi_j^p(t - 2^j k) \\ \psi_{j+1}^{2p+1}(t) &= \sum_k g[k] \psi_j^p(t - 2^j k)\end{aligned}$$

$$p = 0, \dots, 2^j - 1$$

avec $\psi_0^0(t) = \phi(t)$. Ces fonctions sont organisées selon un arbre binaire où chaque noeud (paquet d’ondelettes) est une sous-base $\Psi_j^p = \{\psi_j^p(t - 2^j k)\}_k$ explorant la bande de fréquence normalisée $2^{-(j+1)} \cdot [p, p + 1]$ (Figure 1). Les coefficients de la décomposition d’un signal discret x de longueur N sont obtenus par l’extension de l’algorithme de Mallat [5]. Chaque paquet de coefficients $\Psi_j^p x = \{cx_j^p[k]\}_{k=0, N/2^j-1}$, où $cx_j^p[k] = \langle x, \psi_j^p(t - 2^j k) \rangle$, contient de l’information sur toute la durée du signal dans la bande de fréquence explorée par Ψ_j^p . Dans le cas d’ondelettes orthogonales, une base correspond à un ensemble de paquets $\{\Psi_j^p\}$ qui couvrent l’axe fréquentiel sans redondance (c’est-à-dire sans recouvrement). La décomposition du signal x sur une base B est définie par :

$$Bx = \{\Psi_j^p x | \Psi_j^p \in B\} \quad (1)$$

Un exemple de base est donné Figure 2 avec les bandes de fréquence correspondantes.

Dans de nombreuses applications, les signaux analysés ne sont pas synchronisés les uns par rapport aux autres ; une transformation des coefficients des paquets d’ondelettes est alors

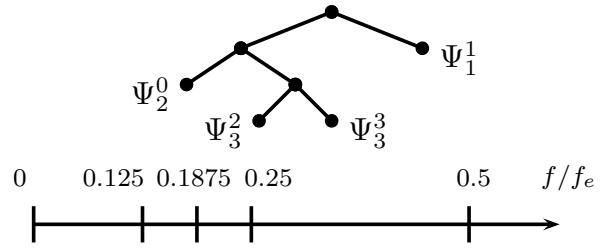


FIGURE 2 – Une base correspond aux feuilles d’un arbre binaire admissible, et explore l’axe fréquentiel sans recouvrement. On présente ici un exemple de base ($J \geq 3$) et les bandes fréquentielles correspondantes (f_e : fréquence d’échantillonnage).

nécessaire afin de ne retenir qu’une description fréquentielle. Dans ce cadre, nous choisissons de représenter un signal par ses marginales ; elles sont définies en chaque noeud par :

$$M\Psi_j^p x = \sum_{k=1}^{N/2^j} |cx_j^p[k]| \quad (2)$$

Pour une base ou une sous-base donnée B , le vecteur des descripteurs est donné par :

$$MBx = [\dots M\Psi_j^p x \dots] \text{ avec } \Psi_j^p \in B \quad (3)$$

3 Optimisation de l’espace de représentation

L’arbre entier de paquets d’ondelettes est très redondant (Ψ_j^p explore le même intervalle fréquentiel que l’union de ses deux fils $\Psi_{j+1}^{2p} \cup \Psi_{j+1}^{2p+1}$). On extrait donc une base, c’est-à-dire un ensemble de paquets d’ondelettes $\{\Psi_j^p\}$ couvrant l’axe des fréquences sans recouvrement ; cela correspond aux feuilles d’un arbre admissible (c’est-à-dire un arbre dont chaque noeud a 0 ou 2 fils), comme illustré dans l’exemple de la figure 2.

Pour sélectionner une base adaptée à la classification, il est nécessaire de définir un critère approprié ainsi qu’une stratégie de recherche pour l’optimiser. Nous utilisons ici la stratégie de Wickerhauser et Coifman, et proposons un nouveau critère additif, que nous présentons dans le cadre général de la décision multiclassé.

La stratégie développée par Wickerhauser et Coifman [1], initialement proposée dans le cadre de la compression de signaux, présente le grand intérêt d’être basée sur une recherche locale, ce qui évite le calcul du critère pour un nombre de bases prohibitif. ($> 2^{2^J-1}$ où J est le niveau de décomposition maximal). Elle consiste à considérer les noeuds de chaque niveau de décomposition, à partir du bas de l’arbre des paquets d’ondelettes jusqu’à la racine. Pour chaque noeud, on compare le coût de deux sous-bases possibles : celle correspondant à un noeud père, et celle résultant de l’union de ses 2 fils ; on conserve la sous-base dont le coût est optimal. L’hypothèse sous-jacente est l’additivité du critère, qui garantit que l’algorithme atteint

l'optimum global, et réduit considérablement le temps de calcul. Un critère \mathcal{C} est additif si, étant données A_1 et A_2 deux sous-bases disjointes, $\mathcal{C}(A_1 \cup A_2) = \mathcal{C}(A_1) + \mathcal{C}(A_2)$.

Dans un contexte de classification, on recherche donc un critère additif qui mesure le pouvoir discriminant d'une sous-base A vis-à-vis d'une population d'apprentissage de c classes. On note cette population $\mathcal{X} = \{\mathcal{X}_1, \dots, \mathcal{X}_i, \dots, \mathcal{X}_c\}$, où \mathcal{X}_i est composé de n_i signaux x appartenant à la classe ω_i . Saïto et Coifman [6] ont proposé un critère additif correspondant à une mesure de contraste calculée à partir d'une distance (euclidienne ou de Kullback-Leibler) entre les représentants des classes. Ce critère, défini initialement pour comparer des représentations temps-fréquence, peut être adapté pour comparer des représentations fréquentielles (marginales). Pour une sous-base A donnée, chaque signal x de \mathcal{X} est représenté par son vecteur de marginales MAx (comme défini dans l'équation 3). Le critère est alors défini par :

$$\mathcal{C}_1(A) = \sum_{i=1}^c \frac{n_i}{n} \|\overline{MA_i} - \overline{MA}\|^2 \quad (4)$$

où $\overline{MA_i} = \frac{1}{n_i} \sum_{x \in \mathcal{X}_i} MAx$ est le représentant de la classe ω_i et $\overline{MA} = \frac{1}{c} \sum_{i=1}^c \overline{MA_i}$. Ce critère est additif, mais ne prend pas en compte la dispersion des classes (inertie intraclasse). Afin d'utiliser cette information, nous proposons d'utiliser un critère de type Fisher. Dans sa forme classique, un critère de ce type est défini comme le rapport de l'inertie interclasse sur l'inertie intraclasse, ce qui correspond pour une sous-base A à :

$$\mathcal{C}_F(A) = \frac{\sum_{i=1}^c \frac{n_i}{n} \|\overline{MA_i} - \overline{MA}\|^2}{\sum_{i=1}^c \frac{n_i}{n} \overline{\overline{MA_i}}} \quad (5)$$

où $\overline{\overline{MA_i}} = \frac{1}{n_i} \sum_{x \in \omega_i} \|MAx - \overline{MA_i}\|^2$ est l'inertie de la classe ω_i .

Maximiser ce critère revient à maximiser la distance entre les centres des classes tout en minimisant la dispersion au sein de chaque classe. Cependant, ce critère n'est pas additif, et peut conduire à une base non optimale s'il est utilisé avec la stratégie de recherche vue précédemment. C'est pourquoi nous proposons le critère défini par :

$$\mathcal{C}_2(A) = K \cdot \sum_{i=1}^c \frac{n_i}{n} \|\overline{MA_i} - \overline{MA}\|^2 - (1 - K) \cdot \left(\sum_{i=1}^c \frac{n_i}{n} \overline{\overline{MA_i}} \right) \quad (6)$$

Si K est choisi entre 0 et 1, on conserve les mêmes propriétés que le critère précédent. Mais sous cette forme le critère est additif et donc approprié à l'utilisation de la stratégie de Wickerhauser et Coifman.

La pertinence des deux critères additifs \mathcal{C}_1 et \mathcal{C}_2 a été évaluée sur des signaux simulés et des signaux EEG réels. La valeur de K a été fixée après des tests effectués sur des signaux simulés et donnant pour K une plage optimale [0.3; 0.7]. Nous avons utilisé $K = 0.5$ dans les exemples présentés en section 5.

4 Classification et évaluation de la méthode

Dans le contexte général multiclasse, on utilise une procédure one-versus-rest (OVR) : pour chaque classe ω_i , on apprend sur la population d'apprentissage une fonction discriminante linéaire f_i définie par la méthode des SVM [2], destinée à séparer ω_i du reste de la population. Alors, pour un signal x décrit par MBx dans l'espace de représentation optimal correspondant à la meilleure base B , la décision finale est prise à partir des c fonctions discriminantes f_i :

$$"x \text{ appartient à } \omega_j" \text{ avec } j = \underset{i=1, \dots, c}{\text{Argmax}} f_i(MBx) \quad (7)$$

Le critère utilisé pour évaluer les performances de la méthode est le pourcentage de mal classés calculé sur une population test. Dans le cas des données EEG, comme on dispose d'un faible nombre de signaux, on utilise une procédure de validation croisée (leave-one-out) [3].

Les résultats présentés dans la suite correspondent à des problèmes biclasses.

5 Résultats sur des signaux simulés et sur des données EEG réelles

Tous les signaux, simulés ou réels, ont été analysés avec l'ondelette de Daubechies 2. Cela correspond à un choix arbitraire parmi des ondelettes orthogonales standards.

Deux classes de signaux ont été simulées pour étudier l'influence de la taille de la population d'apprentissage et du bruit sur les performances relatives aux critères de sélection. Les signaux sont construits par combinaison linéaire d'atomes dont le support fréquentiel diffère d'une classe à l'autre. La localisation temporelle de ces atomes est aléatoire. Une simulation est définie par 2 classes de signaux, constituées chacune de 10 ou 100 signaux d'apprentissage et 1000 signaux de test. Tous les signaux sont de longueur $N = 128$ et bruités par un bruit additif (RSB de 2 ou 10 dB). On présente dans le tableau 1 les moyennes et écart-types des pourcentages de mal classés (notés *moyenne* \pm *écart-type*.) obtenus sur 20 simulations réalisées pour chaque cas considéré. On peut noter que le critère \mathcal{C}_2 fournit toujours un meilleur espace de représentation que le critère \mathcal{C}_1 ou que la base imposée de la DWT et que sa supériorité est d'autant plus nette que la taille de l'apprentissage est petite.

Les données réelles correspondent à des signaux EEG enregistrés sur 5 sujets dans le cadre d'un projet franco-danois sur les interfaces cerveau-machines [4]. Chaque sujet doit imaginer une flexion du pied droit, la classe de chaque signal étant définie par la vitesse du mouvement associé (lent ou rapide). Le nombre de signaux disponibles par classe et par sujet varie de 25 à 48. Les résultats de la classification sont présentés dans le tableau 2. Ils montrent d'une part l'intérêt d'optimiser la base, et d'autre part la pertinence du critère \mathcal{C}_2 proposé.

TABLE 1 – Pourcentage de mal classés, signaux simulés

Méthodes	Critère \mathcal{C}_1 ($J = 3$)	Critère \mathcal{C}_2 ($J = 3$)	Base DWT
10 sig. d'app. RSB = 10 dB	26% \pm 15%	17% \pm 3%	22% \pm 3%
100 sig. d'app. RSB = 10 dB	14% \pm 2%	11% \pm 1%	18% \pm 1%
10 sig. d'app. RSB = 2 dB	32% \pm 13%	26% \pm 5%	29% \pm 5%
100 sig. d'app. RSB = 2 dB	21% \pm 7%	19% \pm 2%	24% \pm 1%

TABLE 2 – Pourcentage de mal classés, signaux EEG

Méthodes	Critère \mathcal{C}_1 ($J = 3$)	Critère \mathcal{C}_2 ($J = 3$)	Base DWT
Sujet 1	48%	32%	34%
Sujet 2	42%	18%	22%
Sujet 3	29%	24%	23%
Sujet 4	45%	23%	44%
Sujet 5	40%	34%	46%
Moyenne	41%	26%	34%
Ecart type	7%	7%	11%

TABLE 3 – Evaluation comparée du critère proposé (\mathcal{C}_2) et du critère de probabilité d'erreur (noté \mathcal{C}_3) sur les données simulées dans les quatre cas caractérisés par le RSB et le nombre de signaux d'apprentissage.

	10db, 10	10db, 100	2db, 10	2db, 100
\mathcal{C}_2	17% \pm 3%	11% \pm 1%	26% \pm 5%	19% \pm 2%
\mathcal{C}_3	17% \pm 3%	11% \pm 1%	27% \pm 6%	18% \pm 1%

6 Discussion

La définition d'un critère pour la sélection d'un espace de représentation doit être cohérente avec la règle de décision utilisée dans cet espace. Dans notre cas (classification par SVM) et dans cette optique, le critère le plus approprié serait la probabilité d'erreur de classification en utilisant la règle de décision définie par l'équation 7. C'est ce critère de probabilité d'erreur qui a été utilisé dans [4] pour optimiser l'ondelette mère. Cependant pour la recherche de la meilleure base, un tel critère, non additif, nécessite d'être calculé pour toutes les bases possibles et est donc extrêmement coûteux. C'est pourquoi nous avons recherché un critère additif, qui permet l'utilisation d'un algorithme rapide de recherche de l'optimum global.

Dans la mesure où il n'est plus directement adapté à la règle de décision (et donc pas justifié *a priori*), le critère proposé \mathcal{C}_2 a été comparé au critère de probabilité d'erreur sur les données simulées : les résultats obtenus (table 3) sont pratiquement identiques quelle que soit la taille de la population d'apprentissage, ce qui met en évidence *a posteriori* la pertinence de ce critère.

En revanche, les tests réalisés avec le critère \mathcal{C}_1 sur des données simulées (table 1) montrent qu'il n'est pas adapté dans le cas de petites populations d'apprentissage, puisqu'il fournit une base moins adaptée que celle de la DWT. Ces résultats se retrouvent sur les données réelles (table 2) et montrent l'importance de la prise en compte de la dispersion des classes dans la robustesse du critère.

Références

- [1] R.R. Coifman, M.V. Wickerhauser. *Entropy based algorithm for best basis selection*. IEEE Transaction on information theory, 38 :713-718, 1992.
- [2] N. Cristianini and J. Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press, 2000.
- [3] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., New York, 2nd edition, 2001.
- [4] D. Farina, O.f. Do Nascimento, M.F. Lucas and C. Doncarli. *Optimization of Wavelets for Classification of Movement-Related Cortical Potentials Generated by Variation of Force-Related Parameters*. Journal of Neuroscience Methods, vol. 162(1), pp. 357-363, May 2007.
- [5] S. Mallat. *A wavelet tour of signal processing*. Academic Press, 1998.
- [6] N. Saito, R.R. Coifman. *Local discriminant bases*. Wavelet applications in Signal and Image Processing II, A.F. Laine and M.A. Unser, Eds. Proc. SPIE vol 2303, 1994.