

# Classification de signaux invariante en translation

Vincent GUIGUE Alain RAKOTOMAMONJY Stéphane CANU

Lab. Perception, Systèmes, Information - CNRS - FRE 2645

Avenue de l'Université, 76801 St Étienne du Rouvray

Tél : 02 32 95 97 03 Fax : 02 32 95 97 08

Vincent.Guigue@insa-rouen.fr

**Résumé** – Cet article présente une méthode de classification de signaux non-stationnaires dans le cas où les signaux contiennent des motifs caractéristiques dont la position temporelle est variable est inconnue. Nous avons montré que la construction d'une représentation graphique des signaux, basée sur une transformée en ondelettes continues permet de s'affranchir d'une référence temporelle absolue et de faire face efficacement à ce problème. Après avoir défini un produit scalaire entre graphes, nous avons comparé les résultats en classification entre les SVM et les kppv.

**Abstract** – This paper addresses the question of the classification of non-stationary signals. We make the hypothesis that each signal includes a pattern, the shape of which is discriminant but the time location of which is random and unknown. We propose to build a graphical representation, based on continuous wavelet transform in order to eliminate the absolute time reference of the description. Then, we define a dot product between graphs and compare the results of SVM and k-nn.

## 1 Introduction

La classification de signaux non-stationnaires est un problème complexe. Nous avons cherché à le résoudre dans le cas où chaque signal contient un motif discriminant dont la position est aléatoire et inconnue. Les descripteurs statistiques classiques comme la moyenne ou les coefficients de Fourier ne donnent pas une représentation satisfaisante pour la classification de signaux non stationnaires. Les deux approches usuelles consistent à traiter directement les signaux ou à utiliser les représentations temps-échelles et temps-fréquences [Dav00], [BJ93]. Dans les deux cas, la description est de grande dimension. Une solution pour traiter l'hypothèse d'un motif discriminant de position inconnue est d'optimiser la forme des ondelettes afin de rendre l'apparition du motif plus nette [LDHD02]. Il est également possible de filtrer les signaux afin de faire ressortir les parties discriminantes.

Cependant, nous souhaitons introduire le minimum de connaissance *a priori* et de paramètres dans le problème. Le but de cet article est de proposer une description originale des données pour caractériser des motifs indépendamment de leurs positions à l'intérieur du signal. Nous montrerons qu'il est possible de créer une telle représentation en décrivant les signaux sous la forme de graphes. Ces graphes sont issus d'une représentation en ondelettes continues, afin de conserver une description temporelle tout en s'affranchissant d'une référence temporelle absolue. Le calcul de produits scalaires entre les graphes [KTI03] nous permet ensuite de classer les signaux avec des *Support Vector Machines* (SVM) [Vap98]. Nous avons comparé les résultats SVM avec un  $k$  plus proches voisins basé sur la distance induite par ce produit scalaire.

## 2 Méthode

### 2.1 Représentation des données

Soit  $\{S_1, \dots, S_n\}$  un ensemble de signaux d'apprentissage de longueur  $N$  étiqueté en  $\{-1, 1\}$ . Nous cherchons une description discriminante, invariante en translation, pour la classification de ces signaux. Le caractère non-stationnaires des signaux nous a conduit à opter pour les représentations temps échelles. Nous noterons  $S(t)$  et  $r(S(t))$  un signal et sa représentation :

$$\begin{aligned} r : \mathcal{X} &\rightarrow \mathcal{F} \subset (\mathbb{R} \times \mathbb{R} \times \mathbb{R})^k \\ S &\rightarrow r(S) = \{t_i, E_i, P_i\}_{i=1}^k \end{aligned} \quad (1)$$

$$\text{avec : } P_{E_i, t_i} = \langle S, \psi_{E_i, t_i} \rangle = \int_{-\infty}^{+\infty} S(t) \psi^* \left( \frac{t - t_i}{E_i} \right) dt$$

où  $k$  est le nombre de coefficients composant le plan, chaque coefficient étant défini comme un triplet temps-échelle-poids.

Pour faire face à l'hypothèse d'un motif dont la position est variable dans le signal, nous avons besoin d'une représentation invariante en translation, c'est à dire :  $r(S(t - \tau)) = r_\tau(S(t))$ , où  $r_\tau$  est une translation de la représentation du signal. Nous nous sommes donc tournés vers les représentations en ondelettes continues. Cette famille analysante  $\{\psi_{E_i, t_i}\}_{i=1, \dots, k}$  ne forme pas une base, elle est redondante mais elle est invariante en translation [Mal97].

Les motifs que nous cherchons à identifier ont une signature dans le plan temps-échelles. Le but est de la faire apparaître en mettant en avant les caractéristiques discriminantes des motifs. Le temps est à la fois un facteur important pour décrire le motif et un facteur pénalisant puisque la position du motif est variable. La solution

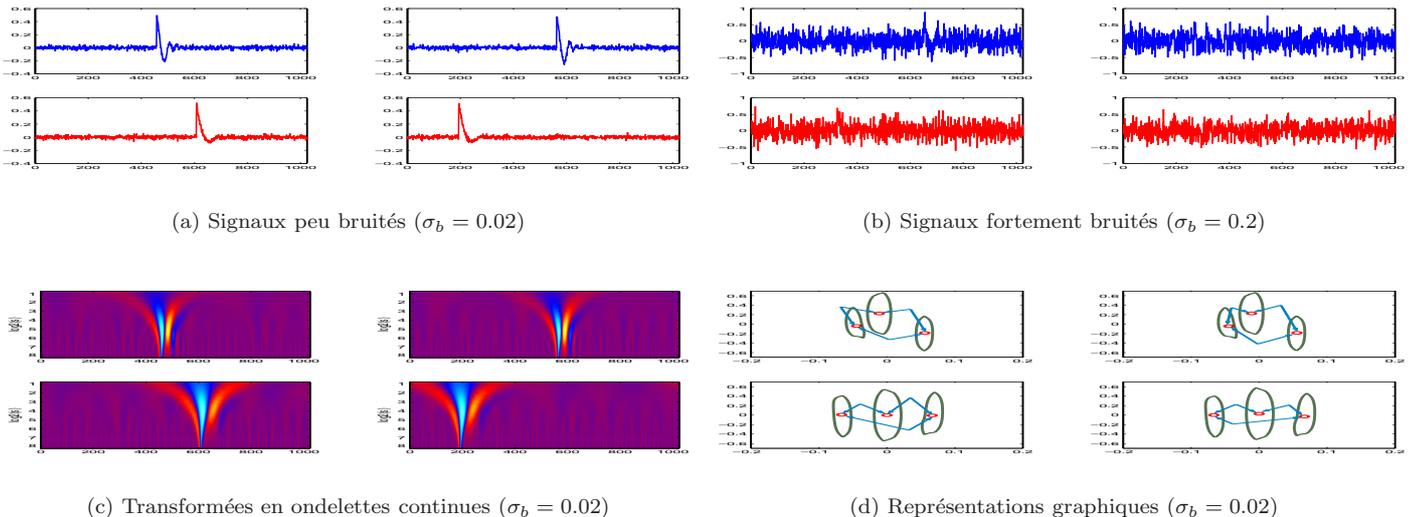


Figure 1: Les 2 représentations graphiques du haut sont de la classe 1 ( $u = 1 \cdot 10^{-3}, v = 2 \cdot 10^{-3}$ ), les 2 du bas sont de la classe -1. ( $u = 5 \cdot 10^{-4}, v = 6 \cdot 10^{-4}$ )

consiste à relativiser les informations temporelles. Cette deuxième représentation  $r_2$  est basée sur une matrice  $A$  des arcs entre les points du plan temps-échelles :

$$r_2 : \begin{array}{l} \mathcal{F} \rightarrow \mathcal{G} \subset (\mathbb{R} \times \mathbb{R})^k \times \mathbb{R}^{k \times k} \\ \{t_i, E_i, P_i\}_{i=1}^k \rightarrow r_2(S) = \{E_i, P_i\}_{i=1}^k, A \end{array} \quad (2)$$

où  $\mathcal{G}$  est l'espace des graphes. Dans cet espace, les signaux prennent la forme de graphes : chaque coefficient de la transformée en ondelettes continues est un nœud lié à tous les autres coefficients du graphe. Les nœuds portent les informations d'échelle et de poids, les arcs portent les informations de temps  $\Delta_t$ .

## 2.2 Réduction de la dimensionnalité

La représentation  $r_2$  est particulièrement volumineuse, elle compte  $k$  nœuds et  $k(k-1)$  arcs. De plus, la complexité du calcul du produit scalaire entre graphes dépend essentiellement du nombre de nœuds des graphes (section 2.4). Il est donc nécessaire de réduire la taille de la représentation pour pouvoir calculer efficacement des produits scalaires dans ce nouvel espace. Cette réduction de la dimensionnalité devra conserver une part significative de l'énergie du signal pour que la représentation ne perde pas son sens.

Nous faisons l'hypothèse que l'information discriminante se trouve dans les zones de fortes énergies du signal. La dimension du problème sera donc réduite tout en conservant une partie significative de la représentation du signal initial. Donoho [Don95] a montré que l'élimination des coefficients les plus faibles dans une transformée en ondelettes orthogonales permet de réduire le bruit dans un signal. Mallat [MZ92] a complété cette étude dans le cas des ondelettes continues, en montrant que les valeurs maximales de la transformée en ondelettes permettent de reconstruire le signal.

La solution consiste à diminuer le nombre de nœuds des graphes tout en augmentant le nombre d'informations présentes dans chacun des nœuds. Nous proposons de prendre un modèle gaussien pour les zones de fortes é-

nergies du plan temps-échelles (cf Fig. 1(d)). Chacun des nœuds est alors une gaussienne, définie par : la matrice de covariance dans le plan, la somme des coefficients de la zone, le nombre de coefficients de la zone et la localisation de la gaussienne en échelle. Les arcs représentent les  $\Delta_t$  entre les positions des gaussiennes dans le temps.

## 2.3 Produit scalaire entre graphes

Le produit scalaire entre ces représentations graphiques que nous utilisons est dérivé de [KTI03]. L'idée est de comparer des chemins de toutes longueurs, partant de chaque nœud dans les deux graphes, pondérés par la probabilité de ces chemins :

Soient :

- $h_i^1$  et  $h_j^2$  les nœuds  $i$  et  $j$  des graphes  $G^1$  et  $G^2$ ,
- $p(h_i|h_{i-1})$  la probabilité de transition du nœud  $h_{i-1}$  au nœud  $h_i$ ,
- $p_q(h_\ell)$  la probabilité de s'arrêter au nœud  $\ell$ ,
- $K_n(h_k^1, h_k^2)$  le produit scalaire entre les nœuds  $h_k^1$  et  $h_k^2$  de  $G^1$  et  $G^2$ ,
- et  $K_a(a_{h_{k-1}^1, h_k^1}^1, a_{h_{k-1}^2, h_k^2}^2)$  le produit scalaire entre 2 arcs  $a^1$  et  $a^2$  de  $G^1$  et  $G^2$ .

$$K(G^1, G^2) = \sum_{\ell=1}^{\infty} \sum_{h^1} \sum_{h^2} p(h_1^1) \prod_{i=2}^{\ell} p(h_i^1|h_{i-1}^1) p_q(h_\ell^1) \times p(h_1^2) \prod_{j=2}^{\ell} p(h_j^2|h_{j-1}^2) p_q(h_\ell^2) \times K_n(h_1^1, h_1^2) \prod_{k=2}^{\ell} K_a(a_{h_{k-1}^1, h_k^1}^1, a_{h_{k-1}^2, h_k^2}^2) K_n(h_k^1, h_k^2) \quad (3)$$

Les détails sont donnés dans [KTI03], les modifications concernant la gestion des nœuds qui contiennent plus d'informations que dans le papier original.

Le calcul de se produit scalaire conduit à la résolution d'un système linéaire de taille  $(|G^1||G^2|)^2$ , où  $|G^i|$  est le nombre de nœuds du graphe  $i$ . Elle nous a forcés à réduire au maximum le nombre de nœuds dans chacun des graphes. Pour conserver une représentation significative des signaux, nous avons choisi d'augmenter le nombre d'informations présentes dans chacun des nœuds à travers une modélisation gaussienne.

## 2.4 Alternatives aux graphes

Afin de valider notre approche, nous avons implémenté quatre descriptions alternatives au noyau de graphes.

- **Les signaux bruts** : la méthode la plus simple pour décrire les données.
- **Les descripteurs statistiques classiques**, généralement utilisés pour faire face à des signaux stationnaires [Hea00, Has00]. Nous avons utilisé 16 descripteurs statistiques (moyenne, écart type, amplitude...) complétés par une analyse en bande de puissances sur 6 bandes de fréquence.
- **La combinaison des translations** possibles entre deux signaux, qui est très couteuse mais bien adaptée au problème d'invariance en translation. L'idée est de trouver la distance tangente [HHB04, SCDV98] entre les deux familles de signaux issues des translations des deux signaux originaux :
$$\langle S_1, S_2 \rangle = \max_{\tau \in \Omega} (k(r(S_1), r_\tau(S_2))) \quad (4)$$
où  $\Omega$  est l'ensemble des translations considérées.
- **Les sacs de vecteurs** : les gaussiennes issues de la modélisation du plan temps-échelles (section 2.2) sont comparées deux à deux, seules les informations temporelles des arcs du graphes sont laissées de côté [WCG03].

$$K_{match} = \frac{1}{2} (K + K^T)$$

avec:  $K(S_1, S_2) = \frac{1}{|G^1|} \sum_{i=1}^{|G^1|} \max_j (k(V_i^1, V_j^2))$  (5)

où  $V_i^k$  est le vecteur décrivant la  $i^e$  gaussienne du signal  $S_k$ .

Les noyaux (4) et (5) ne sont pas définis positifs, du fait de l'utilisation de la fonction max. Nous faisons l'approximation suivante :

$$\max_{y_j \in \mathcal{Y}} (k(x_i, y_j)) \approx \frac{1}{|\mathcal{Y}|} \sum_{j=1}^{|\mathcal{Y}|} \exp\left(-\frac{\|x_i - y_j\|^2}{2\sigma^2}\right)$$

Le noyau (5) est alors équivalent au *Multiple Instance Kernel* [GFKS02].

## 3 Expérience

### 3.1 Données

Dans beaucoup d'applications, les données sont composées de motifs et de bruits. La forme du motif est discriminante

mais sa position est aléatoire et inconnue. Nous travaillons sur des données artificielles présentant de telles caractéristiques. Soit  $m(t)$  un *chirp* à décroissance exponentielle :

$$m(t) = e^{-\alpha t} \cos((u + vt)t + \phi) \quad (6)$$

Pour compliquer ce problème, un bruit blanc gaussien  $b(t)$  de variance  $\sigma_b^2$  est ajouté sur les données. Soit  $\Gamma(t)$  la fonction échelon, le signal  $S(t)$  s'écrit :

$$S(t) = m_{u,v}(t - \tau)\Gamma(t - \tau) + b(t) \quad (7)$$

Le couple  $\{u, v\}$  caractérise le motif et donc la classe du signal, l'instant de déclenchement  $\tau$  n'est pas discriminant. Dans cet article, nous nous sommes limités à l'utilisation d'ondelettes simples de type *Sombrero* pour la transformée en ondelettes continues.

## 3.2 Résultats

Les résultats (Tab. 1) sont présentés suivant deux méthodes de classification : les *Support Vector Machine*, sur les différents produits scalaires, et les  $k$  plus proches voisins, sur les distances induites. Les résultats utilisent 1 ou 400 échantillons en apprentissage.

Les noyaux de base utilisés (pour les nœuds, les arcs, les signaux bruts etc...) sont des noyaux gaussiens de largeur de bande  $\sigma$ . Les paramètres  $\sigma$ ,  $C$  (compromis biais-variance des SVM) et  $k$  (nombre de plus proches voisins) sont optimisés par validation croisée sur l'ensemble d'apprentissage.

Le tableau 1 présente les résultats obtenus en utilisant plusieurs produits scalaires avec différentes méthodes de classification et pour différents niveaux de bruit dans les signaux. La figure 2 montre l'évolution du taux d'erreur de classification en fonction du nombre de points disponibles dans la base d'apprentissage.

Le problème faiblement bruité devient trivial dès lors qu'une représentation adaptée à l'hypothèse du motif de position inconnue est utilisée. Le problème du bruitage des données et celui de la diminution de la taille de la base d'apprentissage mettent en avant les capacités discriminantes de la représentation graphique des données. La comparaison entre représentation graphique et sacs de vecteurs permet de quantifier l'apport lié à la description de la structure des données. En effet, les seules différences entre les deux descriptions sont les arcs qui portent les informations  $\Delta_t$ .

## 4 Conclusion

La description non vectorielle des données offre de nouvelles perspectives dans de nombreux domaines. Dans le cas de la classification de signaux non stationnaires, cette description graphique des données permet de s'affranchir de la référence absolue au temps tout en réduisant la dimension du problème. Le but de l'étude est donc atteint puisque les motifs sont décrits de la même manière quelque soit leur position dans le signal et que cette représentation permet d'améliorer sensiblement les taux de reconnaissance par rapport aux autres descriptions.

Bruit	$\sigma_b = 0.2$			$\sigma_b = 0.02$		
	1/1000	400/1000		1/1000	400/1000	
Classifieurs	1-ppv and SVM	1-ppv	SVM	1-ppv and SVM	1-ppv	SVM
Sig. bruts	49.99% ( $\pm 0.84$ )	48.10% ( $\pm 2.01$ )	47.08% ( $\pm 0.97$ )	48.19% ( $\pm 3.52$ )	41.45% ( $\pm 3.12$ )	38.54% ( $\pm 1.96$ )
Descr. stat.	47.16% ( $\pm 8.22$ )	35.54% ( $\pm 2.12$ )	19.27% ( $\pm 0.98$ )	6.12% ( $\pm 4.30$ )	5.62% ( $\pm 2.12$ )	5.79% ( $\pm 2.22$ )
Translations	49.73% ( $\pm 1.15$ )	43.17% ( $\pm 1.99$ )	42.27% ( $\pm 1.63$ )	28.75% ( $\pm 2.28$ )	15.12% ( $\pm 1.99$ )	14.91% ( $\pm 1.93$ )
Sacs de vect.	29.40% ( $\pm 15.24$ )	11.9% ( $\pm 1.13$ )	8.29% ( $\pm 0.80$ )	0% ( $\pm 0$ )	0% ( $\pm 0$ )	0% ( $\pm 0$ )
Graphes	<b>13.98% (<math>\pm 13.26</math>)</b>	6.66% ( $\pm 0.64$ )	<b>5.25% (<math>\pm 0.35</math>)</b>	0% ( $\pm 0$ )	0% ( $\pm 0$ )	0% ( $\pm 0$ )

Table 1: Taux de mauvaise classification, moyennes et écart-types sur 30 itérations avec des données faiblement bruitées ( $\sigma_b = 0.02$ ) et fortement bruitées ( $\sigma_b = 0.2$ ). Les résultats sont donnés avec 1 (cas extrême) et 400 échantillons en apprentissage.

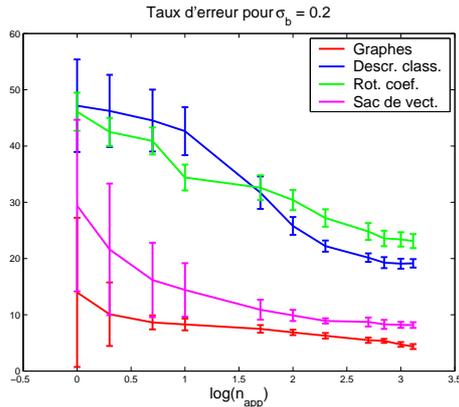


Figure 2: Évolution des taux d'erreur de classification en fonction de la taille de l'ensemble d'apprentissage.

Les perspectives de ce travail concernent l'amélioration de la modélisation du plan temps-fréquences, afin de prendre en compte les cas où l'information discriminante se trouve dans des zones de faibles énergies.

## Remerciements

Ce travail est financé en partie par le programme IST de la communauté européenne, avec le réseau d'excellence PAS-CAL, IST-2002-506778. Cette publication reflète uniquement le point de vue des auteurs.

## References

- [BJ93] R.G. Baraniuk and D.L. Jones. A signal dependant time-frequency representation : Optimal kernel design. In *IEEE Transactions on Signal Processing*, volume 41, pages 1589–1601, 1993.
- [Dav00] M. Davy. *Noyaux optimisés pour la classification dans le plan temps-fréquence - Proposition d'un algorithme constructif et d'une référence bayésienne basée sur les méthodes MCMC - Application au diagnostic d'enceintes acoustiques*. PhD thesis, Université de Nantes, 2000.
- [Don95] D.L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41:613–627, 1995.
- [GFKS02] T. Gärtner, P.A. Flach, A. Kowalczyk, and A.J. Smola. Multi-instance kernels. In *ICML*, pages 179–186, 2002.
- [Has00] E. Haselsteiner. *Neural Based Methods for Time Series Classification*. PhD thesis, Technischen Universität Graz, 2000.
- [Hea00] J. Healey. *Wearable and Automotive Systems for the Recognition of Affect from Physiology*. PhD thesis, MIT, 2000.
- [HHB04] B. Haasdonk, A. Halawani, and H. Burkhardt. Adjustable invariant features by partial haar-integration. In *ICPR 2004*, volume 2, pages 769–774, 2004.
- [KTI03] H. Kashima, K. Tsuda, and A. Inokuchi. Marginalized kernels between labeled graphs. In *20th International Conference on Machine Learning*, pages 321–328. AAAI Press, 2003.
- [LDHD02] M.F. Lucas, C. Doncarli, E. Hitti, and N. Dechamps. Wavelet optimization for classification. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, Orlando, Florida, USA, May 2002.
- [Mal97] S. Mallat. *A Wavelet Tour Of Signal Processing*. Academic Press, 1997.
- [MZ92] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Transactions on Pattern Anal. and Mach. Intell.*, 14(7):710–732, 1992.
- [SCDV98] P. Simard, Y. Le Cun, J. Denker, and B. Victorri. Transformation invariance in pattern recognition, tangent distance and tangent propagation. *Lecture Notes in Computer Science*, 1524:239–274, 1998.
- [Vap98] V. N. Vapnik. *The Statistical Learning Theory*. Springer, 1998.
- [WCG03] C. Wallraven, B. Caputo, and A.B.A. Graf. Recognition with local features : the kernel recipe. In *ICCV 2003 Proceedings*, volume 2, pages 257–264. IEEE Press, 2003.