

Filtrage vaste marge pour l'étiquetage séquentiel de signaux

Rémi Flamary, Benjamin Labbé, Alain Rakotomamonjy

LITIS EA 4108, INSA-Université de Rouen
76800 Saint Etienne du Rouvray, France

Mercredi 19 Mai 2010



Table des Matières

Introduction

Étiquetage d'échantillons

- Définitions

- Discrimination d'échantillons filtrés

Filtrage vaste marge

- Méthode

- Régularisation

Résultats

- Données Jouet

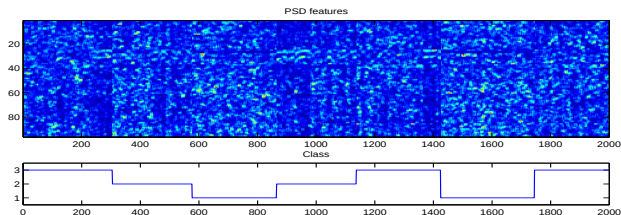
- Données BCI

Conclusion

Étiquetage de séquence (1)

Définition

Obtenir une étiquette pour chaque échantillon temporel d'un signal en prenant en compte leur séquentialité.



Exemple

Décodage continu d'état mental en BCI

- ▶ Le sujet pense au mouvement de sa main droite, gauche ou à un mot.
- ▶ Caractéristiques utilisées: puissance spectrale (PSD).

Étiquetage de séquence (2)

Méthodes existantes

- ▶ Hidden Markov Models [CMR05]
- ▶ Structural SVM [TJHA05]
- ▶ Discriminative HMM [ATH⁺03, SB08]

Limites

- ▶ Pas de décodage “en ligne”.
- ▶ Pré-traitement des signaux par filtrage (*a priori*).
- ▶ Non adaptées au bruit convolutionnel (retard, ...).

⇒ Discrimination d'échantillons temporels filtrés.

Définitions

Définitions

- ▶ $X \in \mathbb{R}^{N \times d}$ matrice des signaux, d canaux N échantillons.
- ▶ y contient les étiquettes.

Filtrage de X par F

$$\tilde{X}_{i,j} = \sum_{m=1}^f F_{m,j} X_{i+1-m+n_0,j} \quad (1)$$

$F \in \mathbb{R}^{f \times d}$ matrice des filtres, d filtres de taille f avec un retard n_0 tel que:

- ▶ $n_0 = 0$: filtre causal
- ▶ $n_0 = f/2$: filtre centré

Noyau: $\tilde{K}_{i,j}^F = k(\tilde{X}_{i,.}, \tilde{X}_{j,.}) = \exp\left(-\frac{\|\tilde{X}_{i,.} - \tilde{X}_{j,.}\|^2}{2\sigma_k^2}\right)$

Discrimination d'échantillons filtrés

Problème avec F fixé

$$\min_g \quad \frac{1}{2} \|g\|^2 + \frac{C}{n} \sum_{i=1}^n H(\mathbf{y}_i, \tilde{X}_{i,\cdot}, g) \quad (2)$$

avec C le paramètre de régularisation, $g(\cdot)$ la fonction de décision et $H(y, x, g) = \max(0, 1 - y \cdot g(x))$ le cout hinge.

Formulation duale:

$$\begin{aligned} \max_{\alpha} J_{SVM}(\alpha, F) &= \max_{\alpha} -\frac{1}{2} \sum_{i,j}^{n,n} \mathbf{y}_i \mathbf{y}_j \alpha_i \alpha_j \tilde{K}_{i,j}^F + \sum_i^N \alpha_i \\ \text{s.t.} \quad \frac{C}{n} &\geq \alpha_i \geq 0 \quad \forall i \quad \text{and} \quad \sum_i^N \alpha_i \mathbf{y}_i = 0 \end{aligned} \quad (3)$$

Filtrage vaste marge

Principe

Apprendre le filtrage temporel et le classifieur de manière jointe.

Problème

$$\min_{g, F} \quad \frac{1}{2} \|g\|^2 + \frac{C}{n} \sum_{i=1}^n H(\mathbf{y}_i, \tilde{X}_{i,\cdot}, g) + \lambda \Omega(F) \quad (4)$$

avec λ un paramètre de régularisation et $\Omega(\cdot)$ une fonction de régularisation pour F

Problème non convexe, mais convexe par rapport à g pour un F fixe (SVM).

$$\min_F J(F) = \min_F J'(F) + \lambda \Omega(F) \quad (5)$$

avec:

$$J'(F) = \min_g \quad \frac{1}{2} \|g\|^2 + \frac{C}{n} \sum_{i=1}^n H(\mathbf{y}_i, \tilde{X}_{i,\cdot}, g) \quad (6)$$

$$= \max_{C/n \geq \alpha \geq 0, \sum_i \alpha_i \mathbf{y}_i = 0} J_{SVM}(\alpha, F) \quad (7)$$

Résolution du problème

Algorithme KF-SVM

```

 $F_{u,v} = 1/f$  pour  $v = 1 \cdots d$  et  $u = 1 \cdots f$ 
 $i=0, D_F^0 = 0$ 
repeat
   $i=i+1$ 
   $G_F^i \leftarrow$  gradient de  $J'(F) + \lambda\Omega(F)$  par rapport à  $F$ 
   $\beta \leftarrow \frac{\|G_F^i\|^2}{\|G_F^{i-1}\|^2}$ : param. du gradient conjugué (Fletcher & Reeves)
   $D_F^i \leftarrow -G_F^i + \beta D_F^{i-1}$ : Direction de recherche
   $(F^i, g^*) \leftarrow$  Line-Search le long de  $D_F^i$  avec calcul de  $g^*$  pour chaque  $F$ .
until Critère d'arrêt
  
```

Complexité:

- Gradient $\mathcal{O}(n_s^2 \cdot f^2)$
- Calcul de la valeur objective: SVM de taille n et un filtrage $\mathcal{O}(n \cdot f \cdot d)$

Régularisation

Frobenius

$$\Omega_2(F) = \sum_{u,v}^{f,d} F_{u,v}^2 \quad (8)$$

Minimisation de l'énergie des filtres et de l'amplitude des canaux non discriminants.

Norme mixte $\ell_1 - \ell_2$

$$\Omega_{1-2}(F) = \sum_v^d \left(\sum_u^f F_{u,v}^2 \right)^{\frac{1}{2}} = \sum_v^d h \left(\|F_{\cdot,v}\|^2 \right) \quad (9)$$

permet la sélection de canaux mais non différentiable en 0.

Résolution de problème par un algorithme Majorization-Minimization: [HL04]

$$\min_{F^{(k+1)}} J'(F) + \lambda \Omega_w(F) \quad (10)$$

$$\text{avec } \Omega_w(F) = \sum_v^d \mathbf{w}_v \sum_u^f F_{u,v}^2 \text{ Frobenius pondérée et } \mathbf{w}_v = \frac{1}{\|F_{\cdot,v}^{(k)}\|}$$

Travaux connexes et extension

Travaux connexes

Apprentissage de filtre temporel:

- ▶ Nos travaux avec un SVM Linéaire [FLR10].
- ▶ Common Sparse Spatio Spectrale Pattern (CSSSP)[DBK⁺06].

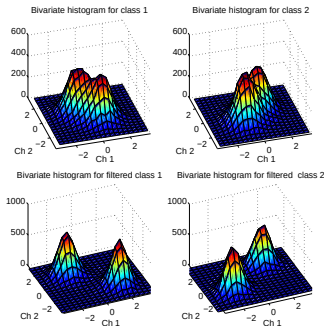
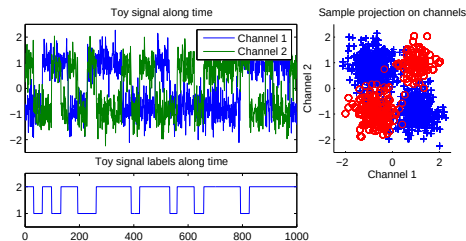
Apprentissage de noyaux:

- ▶ Apprentissage des paramètres du noyau proposé par [CVBM02, GC03].
- ▶ Algorithme similaire à [RBGC08, VB09].

Décodage de Viterbi [GHP04]

- ▶ Normalisation des sorties du classifieur selon Platt [0,1].
- ▶ Décodage hors-ligne de la séquence avec l'algorithme de Viterbi.

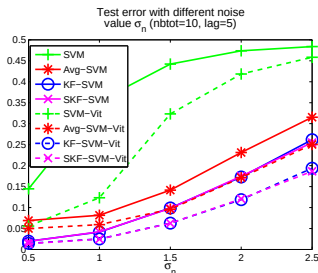
Données Jouet



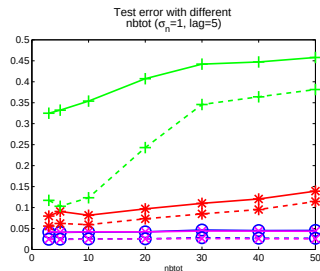
Données

- ▶ *nbtot* canaux dont 2 discriminants, problème non-linéaire.
- ▶ Bruit Gaussien σ et bruit convolutionnel (délai).
- ▶ $f = 11$ et $n_0 = 6$ pour un bon filtre moyennneur.
- ▶ 1000 échantillons d'apprentissage, 10000 en test.

Performances



(a) Bruits σ_n



(b) Nombre de canaux $nbtot$

Figure: Erreur de test pour différents problèmes (pointillés: Décodage de Viterbi).

Performances

- Comparaison de (SVM/ SVM avec moyeneur / KF-SVM).
- Avec et sans décodage de Viterbi.

Sensibilité aux paramètres

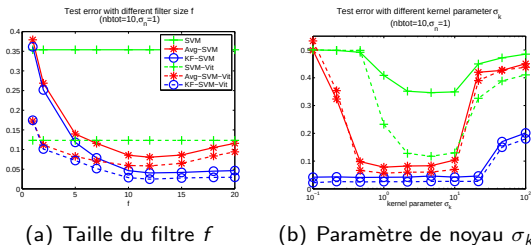


Figure: Erreur de test pour différents problèmes (pointillés: Décodage de Viterbi).

Paramètres

- Méthode dépendant de 4 paramètres (C, σ_k, λ, f).
- Visualisation de la sensibilité aux paramètres f et σ_k .

Données BCI

METHOD	SUB 1	SUB 2	SUB 3	AVG
BCI COMP.	0.2040	0.2969	0.4398	0.3135
SVM	0.2368	0.4207	0.5265	0.3947
KF-SVM				
$f = 2$	0.2140	0.3732	0.4978	0.3617
$f = 5$	0.1840	0.3444	0.4677	0.3320
$f = 10$	0.1598	0.2450	0.4562	0.2870

Table: Erreurs de test pour le Dataset BCI (BCI Compet. III).

Données

- ▶ 3 sujets, 96 canaux, densités spectrales de fréquence (PSD).
- ▶ 3 sessions d'apprentissage, 1 session de test.
- ▶ Sélection des paramètres par validation.

Visualisations BCI

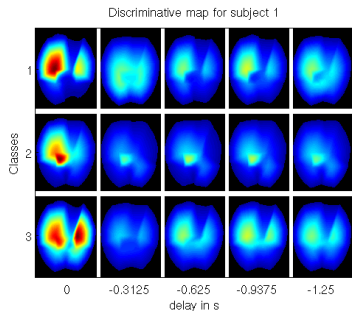
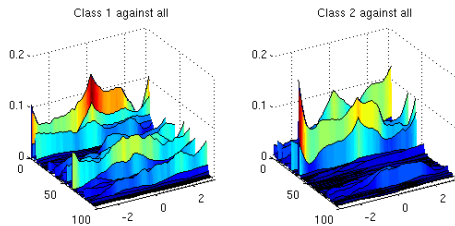


Figure: Filtres F pour différentes tâches de discrimination.

Visualisation des filtres

- ▶ Un filtre par canal.
- ▶ Carte de discrimination spatio/temporelle.

Conclusion

Conclusion

- ▶ Apprentissage d'un filtrage temporel vaste marge.
- ▶ Discrimination d'échantillons temporels.
- ▶ Sélection/pondération automatique des canaux.
- ▶ Visualisation des filtres.

Travaux futurs et en cours

- ▶ Extension 2D de l'apprentissage de filtre.
- ▶ Gérer les données avec beaucoup de points d'apprentissage.

Bibliographie

- [ATH⁺03] Y. Altun, I. Tsochantaridis, T. Hofmann, et al.
Hidden markov support vector machines.
In *International Conference in Machine Learning*, volume 20, page 3, 2003.
- [CMR05] O. Cappé, E. Moulines, and T. Rydén.
Inference in Hidden Markov Models.
Springer, 2005.
- [CVBM02] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukerjee.
Choosing multiple parameters for SVM.
Machine Learning, 46(1-3):131–159, 2002.
- [DBK⁺06] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K. Müller.
Optimizing spatio-temporal filters for improving brain-computer interfacing.
Advances in Neural Information Processing Systems, 18:315, 2006.
- [FLR10] R. Flamary, B. Labbé, and A. Rakotomamonjy.
Large margin filtering for signal sequence labeling.
In *International Conference on Acoustic, Speech and Signal Processing 2010*, 2010.
- [GC03] Y. Grandvalet and S. Canu.
Adaptive scaling for feature selection in svms.
In *Advances in Neural Information Processing Systems*, volume 15. MIT Press, 2003.
- [GHP04] A. Ganapathiraju, J. Hamaker, and J. Picone.
Applications of support vector machines to speech recognition.
IEEE Transactions on Signal Processing, 52(8):2348–2355, 2004.
- [HL04] D.R. Hunter and K. Lange.
A Tutorial on MM Algorithms.
The American Statistician, 58(1):30–38, 2004.
- [RBGC08] A. Rakotomamonjy, F. Bach, Y. Grandvalet, and S. Canu.
SimpleMKL.
Journal of Machine Learning Research, 9:2491–2521, 2008.