

De l'estimation à la classification des activités corticales. Une approche par sélection de variables pour les Interfaces Cerveau Machine

Michel BESSERVE^{1,2}, Line GARNERO¹, Jacques MARTINERIE¹

¹Laboratoire neurosciences cognitives et imagerie cérébrale,
CNRS UPR 640-LENA, Université Pierre et Marie Curie-Paris6, Paris, FRANCE.

²Université Paris-Sud 11, Orsay, FRANCE.

michel.besserve@chups.jussieu.fr, line.garnero@chups.jussieu.fr
jacques.martinerie@chups.jussieu.fr

Résumé – Nous proposons la mise en oeuvre d'une méthodologie de classification des signaux électro-encéphalographiques (EEG) exploitant l'estimation des activités électriques du cortex par résolution de problème inverse. Les difficultés dues au grand nombre de sources corticales estimées (environ 10000) sont contournées en utilisant un algorithme de sélection de variables multivarié: le Séparateur à Vaste Marge pour la norme zéro (L0-SVM). Cette méthode détermine un petit sous ensemble de sources les plus utiles pour différencier les états mentaux.

Nous appliquons l'ensemble de cette approche à une expérience d'Interface Cerveau Machine asynchrone enregistrée dans notre laboratoire. Celle-ci permet d'améliorer le taux de bonne classification par rapport à la mesure directe de l'activité des électrodes EEG.

Abstract – We propose to carry out a classification method for electro-encephalographic signals (EEG), using the activities of cortical sources estimated with an EEG inverse problem. To overcome the difficulties caused by the high number of sources (approximately 10000), we use a multivariate variable selection algorithm: the zero norm Support Vector Machine (L0-SVM). This technique allows to extract a small subset of sources, which are the most useful to allow for the discrimination of the mental states. The whole approach is applied to an asynchronous Brain Computer Interface (BCI) experiment from our lab. It outperforms a method based on the direct measurement of EEG electrodes' activities.

1 Introduction

Les Interfaces Cerveau Machines (ICM) reposent sur la possibilité de résoudre un problème difficile : prédire en essai unique l'état mental d'un sujet à partir des signaux électro-encéphalographiques (EEG) enregistrés à la surface du cuir chevelu. Les informations susceptibles de caractériser ses états mentaux se manifestent dans l'EEG principalement sous forme d'ondes lentes (potentiels évoqués) ou d'oscillations (par exemple dans la bande alpha : 8-12Hz). Or ces phénomènes ne sont que des résidus très atténués et diffus des activités électromagnétiques générées par des assemblées de neurones formant des dipôles de courant distribués au sein du cortex. Comme les différentes régions du cortex présentent des propriétés de spécialisation fonctionnelles, la connaissance "parfaite" des activités en tout point du cortex permettrait de bien meilleures prédictions que l'EEG de surface.

En pratique, on ne peut obtenir qu'une estimation des activités corticales en inversant le modèle physique de génération de l'EEG à partir des dipôles corticaux. On distingue deux classes de méthodes de résolution de ce problème inverse.

1) Les méthodes dipolaires, qui ont été les premières utilisées et qui sont les plus employées encore à l'heure actuelle,

considèrent que l'activité électrique cérébrale est concentrée dans un petit nombre de dipôles. Cependant, les méthodes dipolaires ne peuvent pas expliquer des données qui traduisent l'activité de réseaux étendus de neurones.

2) Des méthodes utilisant des modèles de sources distribuées ont été développées pour s'affranchir des limitations des méthodes dipolaires. Ces approches considèrent un grand nombre de dipôles répartis régulièrement dans une partie ou dans la totalité du volume cérébral. Des contraintes anatomiques peuvent alors être introduites en restreignant le volume de recherche à la surface corticale, extraite à partir de l'IRM anatomique et 'pavée' d'un grand nombre de dipôles. De plus, comme les courants sont émis par les dendrites de neurones disposés perpendiculairement au cortex, l'orientation de chaque dipôle peut être contrainte à être perpendiculaire à cette surface et seule l'amplitude des dipôles reste à déterminer [1]. Une difficulté réside dans le fait que le nombre de dipôles est très grand et que la reconstruction de l'amplitude est alors un problème inverse sous-déterminé qu'il convient de régulariser.

Bien que les méthodes distribuées soient de plus en plus utilisées en analyse des signaux EEG ou MEG, elles ont peu été exploitées dans le domaine des ICM [2, 3]. Pourtant les états cognitifs étudiés en ICM sont des imageries

mentales qui sont susceptibles d'activer des réseaux corticaux largement distribués donc bien adaptés à ces méthodes.

Nous nous proposons ici d'utiliser une des méthodes à sources distribuées usuelles, le problème inverse avec contrainte de norme minimale [4], pour estimer les activités des sources corticales et utiliser ces estimations pour catégoriser l'état mental d'un sujet utilisant un dispositif de ICM asynchrone. Plus particulièrement l'activité de chaque source corticale sera estimée par des mesures de puissance spectrale dans différentes bandes de fréquences et les sources les plus discriminantes seront sélectionnées par un algorithme de sélection de variables multivarié.

2 Méthodologie

2.1 Principe général

La méthode proposée est destinée à la classification en temps réel d'états mentaux à partir d'activités EEG pour un système de ICM asynchrone. Dans ce dispositif, le signal EEG d'un sujet est acquis en temps réel sur des fenêtres glissantes, puis traité par des algorithmes de quantification et de classification du signal EEG. Le système associe ainsi à chaque fenêtre glissante une estimation de l'état mental du sujet, permettant ainsi un transfert d'information. Plus particulièrement, il s'agit dans cet article d'inclure l'estimation de l'activité corticale par problème inverse pour améliorer la quantification (et donc la classification) du signal EEG. Les différentes étapes de traitement de ce dispositif sont décrites ci-dessous.

2.2 Problème inverse en EEG

À l'aide d'un maillage du cortex du sujet et de la position des électrodes EEG par rapport à celui-ci, une matrice dite de problème direct M est calculée avec un modèle à sphères multiples [5], pour lier instantanément l'activité des sources de courant corticales s (vecteur à 10000 éléments) à l'activité EEG e (vecteur à 60 éléments) vérifiant alors $e = Ms$. Le problème inverse avec contrainte de norme minimale permet d'estimer s à partir de e en résolvant l'équation

$$\hat{s} = \arg \min_s (e - Ms)^T (e - Ms) + \lambda s^T s$$

où λ est un paramètre de régularisation contrôlant la sensibilité au bruit de la solution. La solution de ce problème est $\hat{s} = (M^T M + \lambda I_n)^{-1} M^T e = Ce$. L'estimation de l'activité des sources peut ainsi être déduite du signal EEG par une simple multiplication par la matrice constante C . Cette tâche peut donc être facilement exécutée en ligne dans les applications en temps réel d'ICM envisagées.

2.3 Quantification

Les puissances spectrales de chaque source dans différentes bandes de fréquences (8-12Hz, 15-20Hz, 20-30Hz et 30-40Hz) sont alors calculées pour chaque fenêtre i analysée et constituent un nuage de points $\{x_i, i = 1, \dots, N\}$ dans un espace de variables de grande dimension (40000)

quantifiant l'activité EEG. Avec des données pour lesquelles l'état mental y_i de chaque fenêtre i est connu, on peut constituer une base d'apprentissage $\{(x_i, y_i)_{i=1, \dots, N}\}$, qui permet à un algorithme de classification de calculer une fonction $f(x_i)$ qui réalise une estimation de la classe y_i . Le grand nombre de variables de quantification pose un problème pour la classification des signaux ; en effet, il génère un trop grand nombre de degrés de liberté pour les méthodes d'apprentissage statistique, conduisant à un phénomène d'overfitting qui détériore leurs performances sur des données tests (non apprises).

2.4 Classification par L0-SVM

Afin de réduire le nombre de variables, plusieurs méthodes de sélection de variables peuvent être envisagées [6]. Ce type d'algorithme calcule à partir de la base d'apprentissage un sous ensemble de variables optimal pour la classification. Nous proposons ici une technique de sélection de variables multivariée qui réalise simultanément le calcul du classifieur optimal.

Le classifieur utilisé est un Séparateur à Vaste Marge (SVM) linéaire [7] pour lequel on remplace la norme 2 par une norme 0 dans la fonction objectif. La méthode détermine donc un hyperplan séparateur des données d'apprentissage d'équation $\langle w^*, x \rangle + b^* = 0$ comme solution du problème suivant [8] :

$$(w^*, b^*) = \arg \min_{w, b} \|w\|_0$$

sous la contrainte

$$\forall i, y_i (\langle w, x_i \rangle + b) \geq 1$$

La norme 0 utilisée est égale au nombre de coefficients non nuls dans le vecteur w . On recherche donc à calculer une séparatrice affine des données dont l'équation possède un nombre de variables minimales. Une telle méthode de sélection de variables présente l'intérêt de ne pas sélectionner les variables redondantes comme pourrait le faire une méthode univariée telle qu'un test T de Fisher. Ce problème est complexe à résoudre de façon exacte. Cependant, l'algorithme récursif L2-AROM présenté ci-dessous permet de le résoudre de façon approchée [8] :

1. On pose $z = [1, 1, \dots, 1]^T$
2. Résoudre

$$\hat{\alpha} = \arg \max_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \langle z * x_i, z * x_j \rangle$$

sous les contraintes

$$\sum \alpha_i y_i = 0 \quad \text{et} \quad \forall i, \alpha_i \geq 0$$

3. Soit $\hat{w} = \sum \hat{\alpha}_i (x_i * z)$ issu de la solution de 2. On pose $z \leftarrow z * \hat{w}$
4. Revenir à 2 jusqu'à convergence de z . La solution est alors $w^* = z$

Cet algorithme revient donc à implémenter récursivement un Séparateur à Vaste Marge (SVM) classique [7] sur les données $z * w$. Le vecteur w^* ainsi obtenu contient seulement un très faible nombre de coefficients non nuls qui correspondent aux variables sélectionnées. Cette technique

à la particularité exceptionnelle d'avoir un temps de calcul qui ne dépend pas du nombre total de variables car le problème est formulé dans un espace dual. Cette caractéristique est importante car le problème inverse utilisé génère un nombre de variables très important. L'estimation de l'état mental associé à la fenêtre i est alors $\hat{y}_i = \text{sgn}(\langle \mathbf{w}^*, \mathbf{x}_i \rangle + b^*)$. Cette quantité peut alors être restituée en temps réel par exemple pour déplacer un curseur sur un écran.

3 Application à des données d'ICM

3.1 Protocole expérimental

Cette approche est appliquée 'offline' à une expérience de ICM asynchrone réalisée au sein du laboratoire sur 3 sujets pendant trois journées chacun. Les sujets devaient réaliser en continu une tâche mentale correspondant à une consigne affichée sur un écran pendant 20 secondes. Les tâches successives étaient séparées par une période de repos de 3 secondes. Il y avait 6 tâches mentales différentes, incluant trois tâches d'imagerie motrice (préhension, mouvement de l'index droit, mouvement de la langue) et trois tâches cognitives (imagination d'un air de musique, calcul et navigation visuo-spatiale). Ces tâches apparaissaient successivement dans un ordre aléatoire sur l'écran. Durant les expériences, l'EEG était acquis à l'aide d'un casque BrainCap de 60 électrodes et d'un amplificateur Brain Amps 64 voies échantillonné à 500Hz. Les signaux sont analysés suivant la méthode décrite ci-dessus en utilisant une fenêtre glissante d'analyse de 2 secondes.

3.2 Résultats

Pour évaluer la performance de la méthode, un validation croisée a été effectuée sur ces données de ICM pour chaque sujet pour chaque classification binaire possible. L'apprentissage du classifieur a été réalisé sur un jour d'enregistrement par sujet puis le classifieur a été testé sur les deux autres jours d'enregistrement. Pour comparer le taux de classification obtenu sur les sources avec celui obtenu sur les électrodes, le même classifieur a été appliqué en utilisant directement comme quantification les puissances spectrales des électrodes EEG dans les 4 bandes de fréquences considérées. On représente Tab. 1 le taux de bien classés moyenné à travers les 15 couples de discrimination binaires possible entre les 6 tâches pour les deux approches. Le taux moyen de bonne classification de la méthode basée sur les sources corticales atteint 75.01% pour le meilleur des trois sujets et apparaît significativement supérieur à la performance obtenue par la méthode basée sur les électrodes EEG ($p < .05$) par un test non paramétrique réalisé sur l'ensemble des trois sujets. Les résultats sont encourageants compte tenu qu'il s'agit de résultats de prédiction d'un jour à l'autre sur des fenêtres temporelles de 2s. Dans la figure Fig. 1 nous illustrons les propriétés de convergence de l'algorithme en traçant l'évolution du vecteur \mathbf{z} associé à chaque variable de quantification. On observe au fil des itérations qu'un nombre de variables de plus en plus faible se voient attribuer des coefficients

TAB. 1 – Taux de bonne classification de l'algorithme L2AROM à partir des activités mesurées sur les électrodes et à partir des estimations de l'activité corticale.

Mesures	sources	électrodes
sujet 1	75.01% \pm 2.0	72.76% \pm 1.9
sujet 2	65.91% \pm 1.2	63.12% \pm 1.5
sujet 3	60.67% \pm 1.3	61.84% \pm 1.0

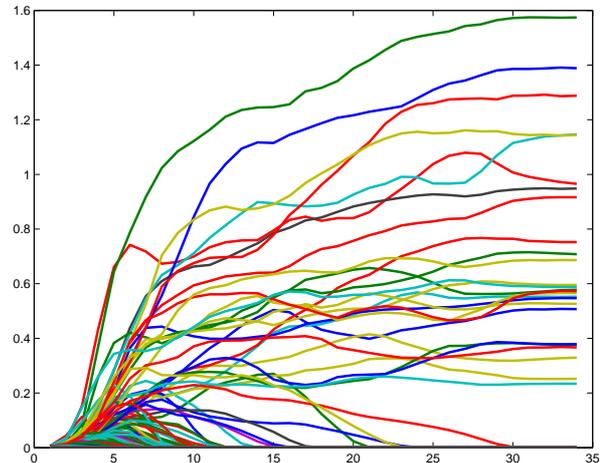


FIG. 1 – Exemple d'évolution de la valeur absolue des coefficients du vecteur \mathbf{z} correspondant à chaque variables en fonction des itérations successives de l'algorithme L2AROM. Seules 25 variables parmi 40000 atteignent une valeur significative lorsque l'algorithme a convergé et sont donc sélectionnées.

importants relativement aux autres. Une vingtaine de variables sont finalement sélectionnées en 34 itération, ce qui illustre la convergence rapide de l'algorithme malgré le grand nombre de variables initial. La répartition topographique des variables choisies est illustrée Fig. 2 pour la discrimination entre les tâches "index droit" et "visuo-spatial". Les activations significativement différentes (mesurées par un test T de Fisher univarié sur chaque source, corrigé pour comparaison multiples par test de permutation) entre les deux tâches dans la bande 15-20 Hz sont représentées sur un cortex gonflé où les régions sombres représentent le fond des sillons corticaux. On observe une activation plus importante du cortex frontal gauche, contralatéral au mouvement, lorsque le sujet imagine bouger l'index droit. La navigation visuo-spatiale active quant à elle des régions occipitales et pariétales droites. Ces localisations correspondent bien à des régions impliquées dans les tâches demandées et confirment que le sujet les réalise correctement. L'algorithme de sélection de variables a sélectionné une variable dans chacune de ces régions ainsi qu'une variable supplémentaire n'apparaissant pas significative par test univarié. L'algorithme semble se comporter ici de manière optimale en sélectionnant des variables non redondantes pour extraire le maximum d'information des activités corticales en vue de la classification.

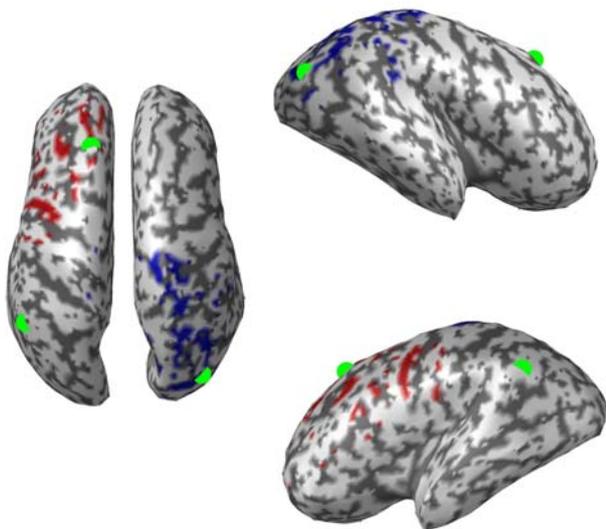


FIG. 2 – Topographie sur un cortex gonflé des activations corticales présentant des différences significatives dans la bande 15-20Hz entre les tâches "index droit" et "navigation visuo-spatiale" pour un sujet (rouge : "index droit" > "navigation visuo-spatiale", bleu pour le contraire). Les points verts correspondent aux variables sélectionnées par l'algorithme L2-AROM

4 Conclusion

Les résultats obtenus montrent que l'utilisation du problème inverse, combiné avec une méthode de sélection de variables adaptée, permet d'apporter des informations supplémentaires en vue de la classification des signaux EEG. De futurs travaux chercheront à améliorer ces résultats en incorporant des mesures d'interaction entre aires corticales dans la méthode.

Références

- [1] S. Baillet and L. Garnero, "A bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem." *IEEE Trans Biomed Eng*, vol. 44, no. 5, pp. 374–385, 1997.
- [2] M. Congedo, F. Lotte, and A. Lécuyer, "Classification of movement intention by spatially filtered electromagnetic inverse solutions," *Phys. Med. Biol.*, vol. 51, pp. 1971–1989, 2006.
- [3] F. Lotte, A. Lécuyer, and B. Arnaldi, "FuRIA : A novel feature extraction algorithm for brain-computer interfaces using inverse models and fuzzy region of interest," *Proc. of the 3rd IEEE-EMBS international Conference on Neural Engineering*, pp. 175–178, 2007.
- [4] S. Baillet, J. Mosher, and R. Leahy, "Electromagnetic brain mapping," *IEEE Signal Processing Magazine*, vol. 18(6), pp. 14–30, 2001.
- [5] J. J. Ermer, J. Mosher, S. Baillet, and R. Leahy, "Rapidly recomputable EEG forward models for realistic head shapes." *Phys Med Biol*, vol. 46, no. 4, pp. 1265–1281, Apr 2001.
- [6] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.
- [7] V. Vapnik, *Statistical Learning Theory*. Wiley-Interscience, New York, 1998.
- [8] J. Weston, A. Elisseeff, B. Schölkopf, and M. Tipping, "Use of the zero norm with linear models and kernel methods," *J. Mach. Learn. Res.*, vol. 3, pp. 1439–1461, 2003.