

P300-speller : Géométrie Riemannienne pour la détection multi-sujets de potentiels d'erreur

Alexandre BARACHANT¹, Rafał CYCÓN², Cédric GOUY-PAILLER³

¹Burke-Cornell Medical Research Institute
785 Mamaroneck Ave, White Plains, NY 10605, États-Unis

²FORNAX sp. z o.o., Czarnowiejska 50B, 30-054 Krakow, Pologne

³CEA, LIST, 91191 Gif-sur-Yvette, France
prenom.nom@{gmail.com, fornax.co, cea.fr}

Résumé – Cette communication présente une méthode de détection de potentiels d'erreur dans le cadre d'une interface cerveau-machine P300-speller. Cette méthode a été employée avec succès lors d'une compétition ayant rassemblé 260 équipes (311 personnes) sur kaggle.com. L'approche repose sur trois piliers principaux : un filtrage spatial concentrant le signal d'intérêt, une classification basée sur des caractéristiques issues d'une interprétation Riemannienne des matrices de covariance du signal électroencéphalographique, et enfin une méthode d'ensemble (bagging) apportant la robustesse nécessaire au transfert des performances pour la détection des potentiels d'erreur sur de nouveaux sujets.

Abstract – A methodology is presented to address the issue of detecting Error Potentials in the context of a P300-speller brain-computer interface. The proposed approach proved highly efficient in a recent competition organized on kaggle.com: authors finished first out of 260 competing teams (311 participants). The approach consists of three main foundations: an ad-hoc spatial filtering focuses the energy of the evoked potential, covariance matrices of the EEG signals are then interpreted in a Riemannian fashion to build features fed to the classification algorithm, and lastly, a bagging procedure increases robustness when detecting Error Potentials on unseen subjects.

1 Introduction

Entre les années 1990 et 2010, les interfaces cerveau-machines (ICMs) ont connu une croissance très importante, mesurable par le nombre de publications dans les journaux et conférences scientifiques ainsi que le nombre d'équipes impliquées dans celles-ci [1]. Depuis 2010, la phase de croissance a débouché sur une phase de maturité. L'abondance des travaux de recherche, et la communication qui y a été associée, a généré une attente sociétale pour le transfert de ces technologies dans des produits grand public. Dans le domaine médical, ces attentes se sont focalisées sur la mise à disposition, auprès de patients souffrant de problèmes moteurs, d'interfaces de contrôle et de communication capables de les assister dans la vie de tous les jours. Sur le marché du grand public, des compagnies se sont lancées dans le développement de systèmes de commandes mentales à partir d'équipements EEG les plus simples possibles (*e.g.* emotiv) avec l'objectif de transformer les paradigmes fonctionnant en laboratoire en produits robustes utilisables dans des cadres moins contraints. Dans ces deux cas les résultats sont pour l'instant mitigés. Les systèmes manquent de robustesse et la charge cognitive qu'induit leur utilisation les rendent difficiles à utiliser pendant de longues périodes [2].

La maturité du domaine se ressent également dans les articles scientifiques qui y sont consacrés. Les processus neurophysiologiques utilisés dans les paradigmes sont relativement

stables, laissant la prédominance au P300 et à l'activation des aires motrices induite par l'imagination de mouvements. Les algorithmes de décodage proposés n'amènent que rarement des améliorations substantielles des performances. Enfin les paradigmes expérimentaux reproduisent souvent les expériences classiques du domaine.

Néanmoins, deux pistes de recherche font figure d'exceptions en raison des sauts technologiques qu'elles sont susceptibles d'amener. D'une part, des équipes de recherche ont mis en évidence un potentiel évoqué appelé « potentiel d'erreur » [6]. Ce potentiel est très intéressant dans le contexte des ICMs car il apparaît dans le cerveau du sujet lorsqu'il y a discordance entre la tâche qu'il souhaite effectuer et le retour qui lui est apporté par le système après la phase de décodage. Ainsi dans le cas du P300-speller, la détection du potentiel d'erreur permet une phase de correction de la lettre détectée et les performances du système global peuvent ainsi être améliorées significativement. Le deuxième verrou s'intéresse à la possibilité d'utiliser une ICM sans avoir à le calibrer pour chaque sujet.

Cet article relate l'approche qui a été mise en place par les auteurs lors d'une compétition¹, ayant eu lieu entre le 19 nov. 2014 et le 24 fév. 2015, et mêlant les deux problèmes évoqués précédemment : détecter des potentiels d'erreur dans une expérience de P300-speller, et faire en sorte que l'algorithme de

1. <http://www.kaggle.com/c/inria-bci-challenge>

détection généralise à de nouveaux sujets. L’approche proposée s’appuie sur trois piliers : d’une part le filtrage spatial permet d’isoler un sous-espace signal concentrant l’énergie des potentiels d’erreur ; d’autre part la tâche de discrimination se fait à partir des matrices de covariance dans un formalisme Riemannien ; enfin la robustesse pour le transfert entre sujets est amenée par la combinaison de rééchantillonnages successifs à partir des données d’entraînement. L’approche proposée a permis aux auteurs de terminer en tête de la compétition.

Après avoir décrit le jeu de données relatif à cette compétition et le critère d’évaluation dans la partie 2, l’approche employée sera décrite dans la partie 3. Les résultats détaillés du concours, ainsi qu’une analyse approfondie de l’importance des différentes étapes du traitement, seront ensuite présentés dans la partie 4. Nous concluons alors la communication par une discussion sur l’approche proposée en partie 5.

2 Description des données

Le jeu de données consiste en l’enregistrement électroencéphalographique (EEG) de 26 sujets ayant participé à une expérience de type P300-speller. Ce paradigme consiste à provoquer la réponse électrophysiologique P300 chez un sujet ayant choisi un symbole (lettre ou chiffre) à épeler parmi une matrice affichée à l’écran. Le P300 est induit chez le sujet lorsque la lettre qu’il a choisie s’illumine. Les symboles s’illuminent par sous-groupe de 6 dans une matrice de taille 6×6 , et plusieurs répétitions sont nécessaires pour que la détection des P300 ait des performances satisfaisantes. Dans ce jeu de données, l’objet n’est pas la détection des P300 mais plutôt la réponse cérébrale du sujet à l’apparition du retour visuel. En effet, lorsque les illuminations de symboles sont terminées, l’algorithme de détection identifie le symbole choisi par l’utilisateur et l’affiche à l’écran. L’objectif de ce jeu de données est de détecter un potentiel électrophysiologique marqueur d’une discordance entre la volonté du sujet et la lettre affichée par l’algorithme à l’écran. Ce marqueur est appelé « potentiel d’erreur ».

Chacun des 26 sujets a participé à cinq sessions d’enregistrement. Pour les quatre premières, douze mots de cinq lettres étaient épelés, tandis que vingt mots de cinq lettres étaient épelés lors de la cinquième session. Afin d’altérer les performances des sujets, deux modes d’illumination ont été utilisés, un mode « long » pour lequel chaque symbole est illuminé 8 fois, et un mode « court » pour lequel seulement 4 illuminations sont faites pour chacune des lettres. Les enregistrements EEG ont été réalisés à l’aide de 56 électrodes positionnées selon une version étendue d’un système appelé 10/20. Pour chacune des sessions, une variable indicatrice indiquant le moment auquel apparaît un retour visuel est ajouté aux signaux temporels EEG.

Dans le cadre de la compétition, le jeu de données est séparé en trois parties. La première partie constitue un jeu de données d’entraînement. Chacun des retours visuels est étiqueté par une variable binaire indiquant si le feedback est en concordance avec la volonté de l’utilisateur (variable à 0) ou bien s’il y a dis-

cordance (variable à 1). Cette variable indique donc pour le jeu de données d’entraînement, la présence supposée d’un potentiel d’erreur. Ce jeu de données est constitué de 16 sujets. Les 10 sujets restants, pour lesquels aucune étiquette n’est donnée aux compétiteurs quant à la présence d’un potentiel d’erreur en réponse aux feedbacks, sont alors séparés en deux groupes. Le premier groupe, constitué de 2 sujets, permet aux concurrents d’obtenir une évaluation de leur performance pendant le déroulement de la compétition. Ce score donne lieu à un classement provisoire, qui n’a aucune valeur pour le classement final des équipes pour des raisons évidentes de sensibilité au sur-apprentissage. Le classement final est quant à lui calculé sur les 8 sujets restants, il n’est visible qu’à la fin de la compétition et constitue le seul classement officiel. Par la suite, nous appelons « ensemble d’entraînement » le jeu de données étiqueté (16 sujets) ; le jeu de données servant au classement pendant la compétition (2 sujets) est appelé « ensemble de validation » ; et enfin le jeu de données utilisé pour le classement final de la compétition (8 sujets) est dénommé « ensemble de test ».

En pratique les équipes envoient un fichier texte au travers de l’interface fournie par kaggle. Ce fichier texte contient deux colonnes, la première identifie de manière unique le sujet, la session et le numéro de feedback ; la deuxième est la probabilité que ce feedback ait donné lieu à un potentiel d’erreur. L’évaluation pour cette compétition est l’aire sous la courbe COR calculée à partir des probabilités fournies dans le fichier.

3 Méthodologie

Pre-traitement. Les signaux EEG sont filtrés par un filtre passe-bande (Butterworth d’ordre 5) entre 1 et 40 Hz. Il sont ensuite découpés selon une fenêtre de 1.3 secondes consécutive à la présentation du retour visuel.

Matrices de covariance. Soit $\mathbf{X}_i \in \mathbb{R}^{C \times N}$ une portion de signal EEG d’indice i , correspondant à N échantillons de signal consécutif au i ème stimulus, observé avec C électrodes. Soit $\mathcal{I}^{(e)}$ et $\mathcal{I}^{(c)}$ l’ensemble des indices correspondant respectivement à la classe « erreur » et à la classe « correct ». On définit les réponses prototypes des potentiels évoqués par :

$$\mathbf{P}_m = \frac{1}{|\mathcal{I}^{(m)}|} \sum_{i \in \mathcal{I}^{(m)}} \mathbf{X}_i, \quad (1)$$

avec $m \in \{e, c\}$. Pour chaque classe, on construit un ensemble de $R \leq C$ filtres spatiaux $\mathbf{V}_m \in \mathbb{R}^{C \times R}$ suivant la méthode XDawn décrite dans [7]. Cette méthode permet de réduire la dimension du prototype et d’isoler un sous-espace signal concentrant l’énergie des potentiels évoqués de chaque classe. Ces filtres spatiaux sont obtenus par diagonalisation jointe de la matrice de covariance de la réponse prototype et de la matrice de covariance de l’ensemble du signal EEG. Dans ces travaux, le nombre de filtres spatiaux est fixé à $R = 5$.

Ces filtres spatiaux nous permettent ensuite de définir la matrice $\mathbf{P} \in \mathbb{R}^{2R \times N}$ telle que :

$$\mathbf{P} = \begin{bmatrix} \mathbf{V}_e^T \mathbf{P}_e \\ \mathbf{V}_c^T \mathbf{P}_c \end{bmatrix} \quad (2)$$

Pour chaque essai \mathbf{X}_i , on construit ensuite un essai $\tilde{\mathbf{X}}_i \in \mathbb{R}^{2R+C \times N}$ par la concaténation de \mathbf{P} et \mathbf{X}_i :

$$\tilde{\mathbf{X}}_i = \begin{bmatrix} \mathbf{P} \\ \mathbf{X}_i \end{bmatrix} \quad (3)$$

Ces nouveaux essais sont utilisés pour estimer les matrices de covariance $\tilde{\Sigma}_i \in \mathbb{R}^{(2R+C) \times (2R+C)}$ sur lesquelles sera appliquée la classification :

$$\tilde{\Sigma}_i = \frac{1}{N-1} \tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_i^T = \begin{bmatrix} \Sigma_P & \mathbf{C}_{P, X_i}^T \\ \mathbf{C}_{P, X_i} & \Sigma_i \end{bmatrix} \quad (4)$$

Ce processus est une généralisation de la méthode employée pour la classification des potentiels évoqués P300 décrite dans [3].

Sélection d'électrode. Une procédure de sélection d'électrode est ensuite appliquée sur ces matrices de covariance. L'algorithme employé est décrit dans [4]. Il s'agit d'une procédure de sélection descendante permettant de trouver le sous-ensemble d'électrodes maximisant la distance Riemannienne entre les matrices covariance moyenne de chaque classe. Si cet algorithme a fait ses preuves pour les ICMs fondées sur l'imagination motrice, elle n'a jamais été employée dans le contexte des potentiels évoqués.

Espace tangent & classification. Les matrices réduites sont ensuite projetées dans l'espace tangent où elles seront classifiées par un algorithme de classification linéaire Elastic-Net. La projection dans l'espace tangent est une opération permettant de projeter les matrices de covariance appartenant à une variété Riemannienne dans un espace euclidien de même dimension. Cette opération est similaire à un noyau, et est décrite en détails dans [5].

Bagging. L'ensemble de cette chaîne de classification est appliqué sur un sous-ensemble de sujets tirés au hasard parmi l'ensemble des sujets de la base d'apprentissage. En répétant cette procédure un certain nombre de fois, on obtient différents modèles dont les prédictions seront moyennées pour obtenir les résultats de classification finaux. Cette procédure augmente le pouvoir de généralisation du modèle global, et est une arme efficace contre le sur-apprentissage.

4 Résultats

La méthode sera évaluée de deux façons différentes. D'une part via les résultats obtenus lors de la compétition, ce qui permet de positionner la méthode face à la concurrence, et d'autre part via une procédure de validation croisée, afin de pouvoir évaluer l'influence de chaque paramètre de la chaîne de classification de manière indépendante.

La procédure de validation croisée est la suivante : chaque sujet est tour à tour exclu de la base d'apprentissage et est utilisé comme sujet de test. La chaîne de classification est donc successivement entraînée sur un sous-ensemble de 15 sujets, et

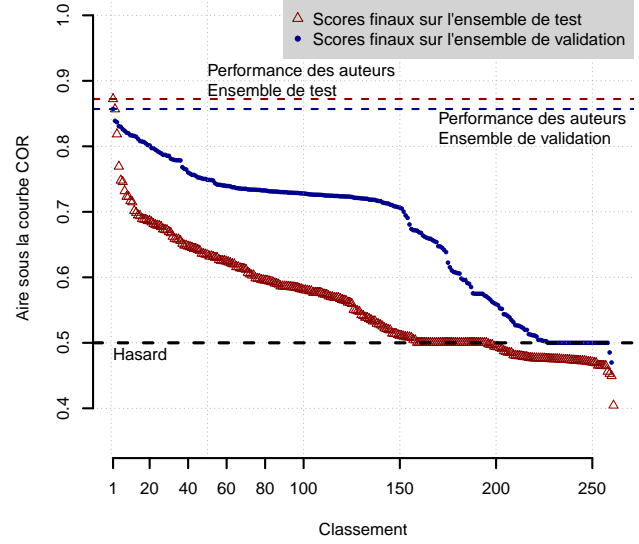


FIGURE 1 – Performances finales sur les ensemble de validation et de test des 260 équipes en lice, et positions des auteurs de l'article dans ces classements (resp. 1^{er} et 1^{er}).

les performances sont évaluées sur le sujet restant. Les résultats finaux sont obtenus par moyennage sur l'ensemble des sujets.

Évaluation compétitive. L'approche présentée ci-dessus a été utilisée dans le cadre d'une compétition organisée sur la plateforme de data science kaggle.com. Ce mode d'évaluation est hautement compétitif en raison du nombre de participants inscrits sur cette plateforme (330000 comptes environ le 29 juin 2015) et du nombre de participants à la compétition. En effet 260 équipes (311 individus) figurent dans le classement final. Nous ne pouvons décrire ici la diversité des approches employées, mais les descriptions qui ont été faites sur le forum de la compétition confirme l'excellence des différents participants et assure que les approches de l'état de l'art ont été largement explorées par les différentes équipes². La figure 1 illustre de manière visuelle les excellentes performances obtenues à l'aide de la méthode présentée dans cet article. L'approche a en effet permis d'occuper les premières places des classements sur les ensembles de validation et de test, respectivement en bleu et rouge. On peut noter la robustesse de la méthode dont les performances n'ont pas été dégradées entre l'ensemble de validation et l'ensemble de test.

Effet du nombre d'électrodes sur les performances. Le paramètre ayant le plus d'impact sur les performances de la chaîne de classification est le nombre d'électrodes utilisées par la procédure de sélection d'électrodes. En effet, un nombre trop faible d'électrodes représente une perte d'information, et un nombre trop élevé conduit généralement à des problèmes d'estimation et de convergence pour les algorithmes utilisant la géométrie Riemannienne. La réduction du nombre d'électrodes diminue

2. <http://www.kaggle.com/c/inria-bci-challenge/forums>

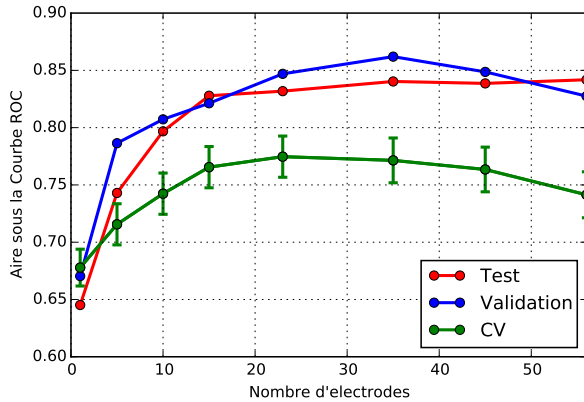


FIGURE 2 – Évolution des performances en fonction du nombre d'électrodes. Les barres d'erreurs représentent l'erreur type.

également la taille de l'espace des caractéristiques, ce qui diminue le risque de sur-apprentissage de la part du classifieur. L'évolution des performances en fonction du nombre d'électrodes est reportée figure 2.

Un nombre de 35 électrodes a été sélectionné pour le modèle final. Ce choix fut motivé principalement par les résultats en validation croisée, mais également par ceux obtenus sur la partition de validation. Néanmoins, il s'est avéré que pour la partition de test, ce paramètre avait finalement peu d'importance et qu'un nombre de 15 électrodes était suffisant pour obtenir des résultats sensiblement similaires aux performances optimales.

Effet du nombre de modèles sur les performances. Le second paramètre par ordre d'importance est le nombre de modèles, c'est-à-dire le nombre de sous-ensembles de sujets, utilisé dans la procédure de bagging. Lors de l'emploi d'une telle procédure, la règle générale veut que les performances augmentent avec le nombre de modèles jusqu'à saturation. Un nombre trop élevé de modèles ne dégrade généralement pas les performances, le choix de ce paramètre est donc fonction du coup de calcul que l'on juge acceptable. Les performances en fonction du nombre de modèles sont données figure 3.

Pour cette compétition, un nombre volontairement élevés de modèles a été choisis (500). Toutefois, la valeur optimale se situe entre 10 et 100 modèles. En revanche, l'introduction de cette procédure diminue la taille apparente de la base d'apprentissage. Un nombre de modèles suffisant doit être choisi pour compenser la possible perte de performance. Dans le cas présent, l'utilisation 3 modèles est suffisant pour améliorer les performances en validation croisée. En revanche, l'amélioration observée sur l'ensemble de test reste très modérée.

5 Discussion et conclusion

En raison du format court de cette communication, certains détails d'implémentation ont volontairement été mis de côté. L'ensemble du code python nécessaire est disponible³.

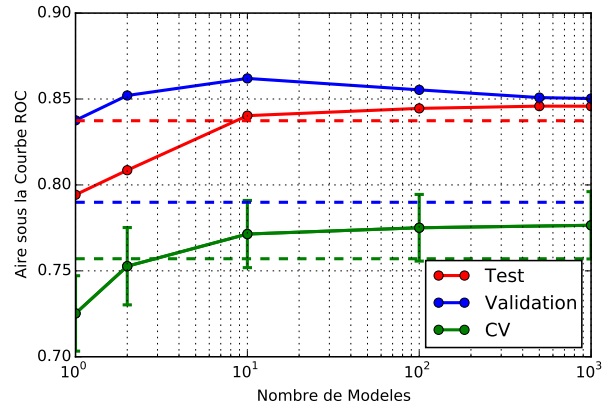


FIGURE 3 – Évolution des performances en fonction du nombre de modèles pour l'ensemble de validation, de test et pour la validation croisée (CV). Les lignes en pointillés représentent les performances obtenues sans l'utilisation de la procédure de bagging. Les barres d'erreurs représentent l'erreur type.

La méthode proposée a démontré d'excellentes performances pour la détection de potentiels d'erreur dans un contexte compétitif inédit. Bien que la sélection d'électrodes et les méthodes d'ensemble n'ont pas apporté d'améliorations significatives du score final sur l'ensemble de test, ils ont permis de limiter au maximum le sur-apprentissage et leur intérêt a été démontré par les scores obtenus en validation croisée.

Références

- [1] J.R. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller et T. Vaughan. Brain-computer interfaces for communication and control. *Clin. neurophysiology*, vol. 113, p. 767–791, 2002.
- [2] E. Pasqualotto, T. Matuz, S. Federici, C.A. Ruf, M. Bartl, M. Olivetti Belardinelli, N. Birbaumer et S. Halder. Usability and workload of access technology for people with severe motor impairment : a comparison of brain-computer interfacing and eye tracking. *Neurorehabil. Neural Repair*, in press, 2015.
- [3] A. Barachant, M. Congedo, G. Van Veen et C. Jutten. Classification de potentiels évoqués P300 par géométrie riemannienne pour les interfaces cerveau-machine EEG. *24ème colloque GRETSI*, 2013.
- [4] A. Barachant et S. Bonnet. Channel selection procedure using riemannian distance for BCI applications. *5th International IEEE/EMBS Conf. on Neural Eng. (NER)*, p. 348–351, 2011.
- [5] A. Barachant, S. Bonnet, M. Congedo et C. Jutten. Classification of covariance matrices using a Riemannian-based kernel for BCI applications, *NeuroComputing*, vol. 112, p. 172–178, 2013.
- [6] M. Perrin, E. Maby, S. Daligault, O. Bertrand et J. Mattout. Objective and subjective evaluation of online error correction during P300-based spelling. *Advances in Human-Computer Interaction*, vol. 2012, Article ID 578295, 2012.
- [7] B. Rivet, A. Souloumiac, V. Attina et G. Gibert. xDAWN algorithm to enhance evoked potentials : application to brain-computer interface. *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, p. 2035–2043, 2009.

3. <https://github.com/alexandrebarachant/bci-challenge-ner-2015>