

Interpréter les données EEG

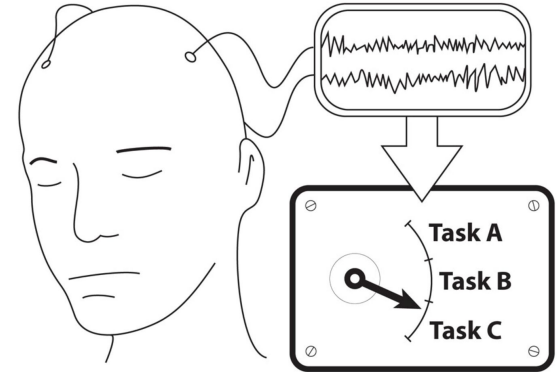
Amin Shahab
Thomas George
René Doumbouya
Marc-Antoine Sayn-Urpar
Vincent Antaki



<https://github.com/tfjgeorge/ml-project>

Présentation du sujet : Classification EEG

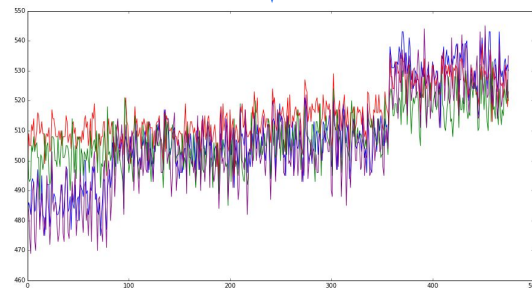
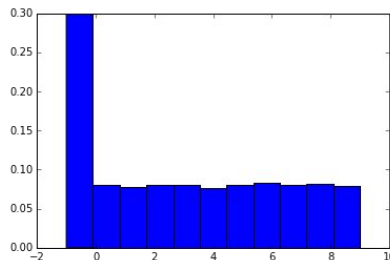
- Plusieurs capteurs disposés sur la tête du sujet
- Un stimulus (image d'un chiffre, prise d'un objet)
- Une réponse par capteur récupérée sous forme d'un signal
- Chaque signal est capturé pendant 2 secondes
- Classification des signaux correspondant chacun à un stimulus
- Le but est de prédire le type de stimulus en entrée en fonction d'un jeu de signaux



Sujet 1 : Mind big data

- Données récente : 14 octobre 2015
- 4 capteurs : Casque Muse
- Stimulus : Image d'un chiffre
- 40.983 extraits de 2 secondes
 - entraînement 30.000 extraits
 - test 10.983 extraits

répartition des classes



But :

- Prédire si le sujet regardait un chiffre (classification binaire)
- Prédire quel chiffre le sujet regardait (classification multiclasse)

Sujet 2 : Grasp-and-Lift EEG*

- Défi : Identifier les mouvements (saisir, lever, replacer) de la main à partir des EEG.
- Données: Enregistrements EEG de sujets (n=12) effectuant des essais de Grasp-And-Lift (GAL). 10 séries d'essais pour chaque sujets.
 - L'ensemble d'entraînement : contient les 8 premières séries.

L'ensemble de test : contient les 9ème et 10ème

id,HandStart,FirstDigitTouch,BothStartLoadPhase,LiftOff,Replace,BothReleased

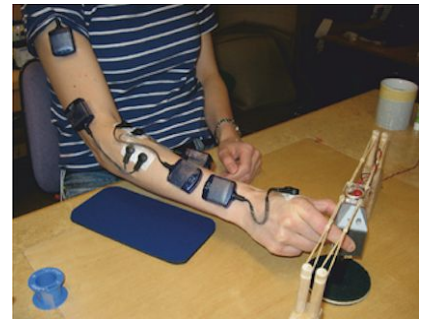
subj1_series9_0,0,0,0,0,0,0

subj1_series9_1,0,0,0,0,0,0

subj1_series9_2,0,0,0,0,0,0

subj1_series9_3,0,0,0,0,0,0

subj1_series9_4,0,0,0,0,0,0



#	Arank	Team Name	* in the money	Score ?
1	—	Cat & Dog	👤 *	0.98109
2	↑2	daheimao	*	0.98029
3	↓1	HEDJ	👤 *	0.97996

*<https://www.kaggle.com/c/grasp-and-lift-eeeg-detection/>

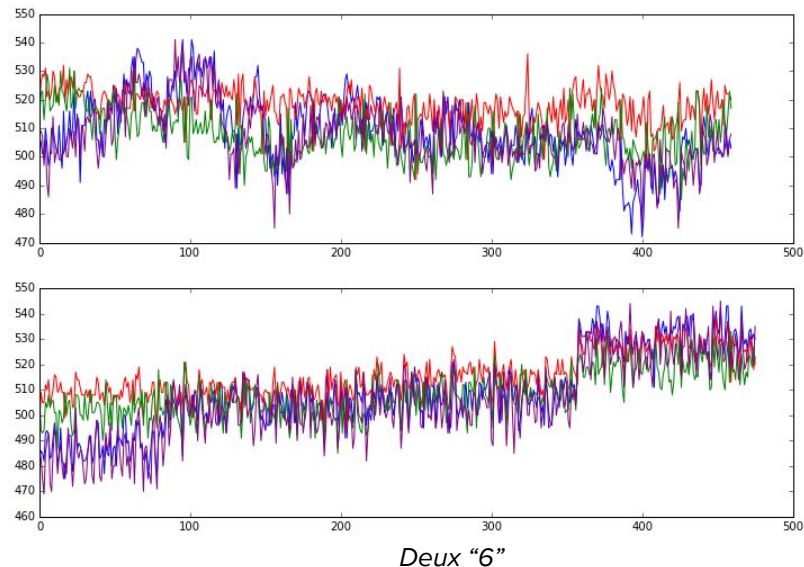
Besoin de traits caractéristiques

Pourquoi utiliser des traits caractéristiques ?

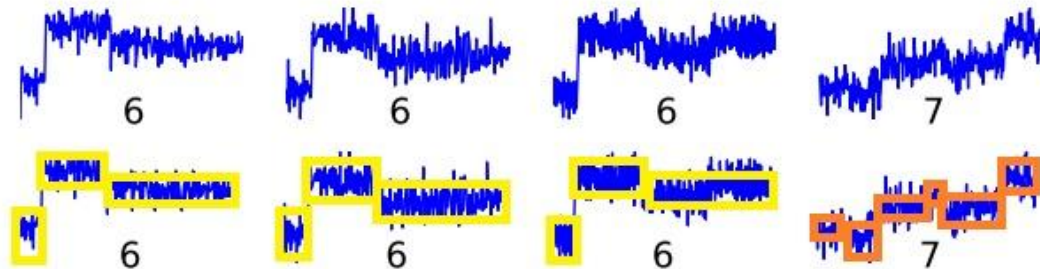
- Test k-ppv avec les données brutes
- Mesure de distance entre séries temporelles ? DTW
- Données bruitées
 - éléments extérieurs non connus
 - bruit des capteurs
 - partie “intéressante” qui ne représente qu’une petite partie de la série
- Séries de longueurs différentes

Traits caractéristiques:

- Méthode visuelle des blocs
- Mentor
- Butterworth
- Mel-frequency cepstral coefficients
- Extraction automatique de traits caractéristiques par réseau de neurones à convolution



Caractéristiques : Méthode “blocs”



De nombreux points dispersés

Réduction de la Précision

Perdre des blocs



Caractéristiques : Mentor^[1]

- Mentor : un système tuteur intelligent qui organise automatiquement le contenu d'apprentissage en fonction de l'état mental de l'apprenant.

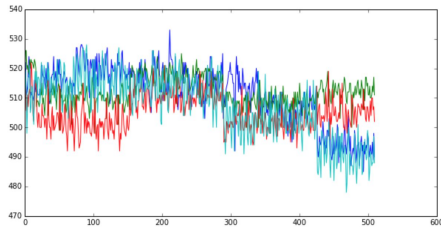


- Le point pertinent est l'**extraction de caractéristiques** pour la construction d'un modèle prédictif du 'workload' d'un apprenant.

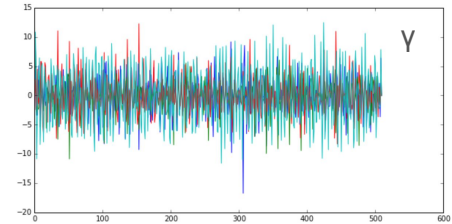
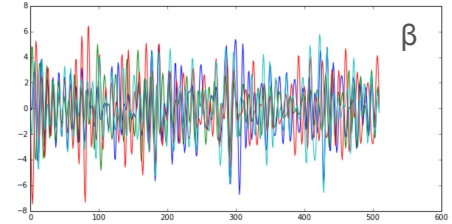
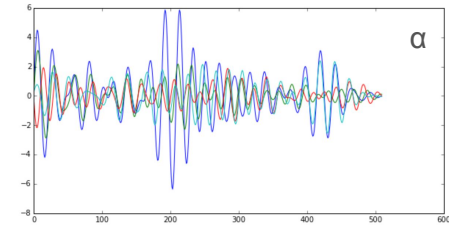
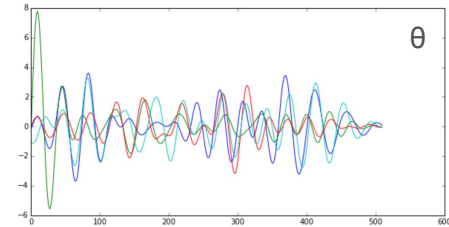
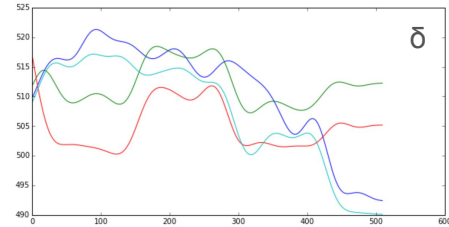


Caractéristiques : Butterworth filter → Puissance^[2]

Données brutes

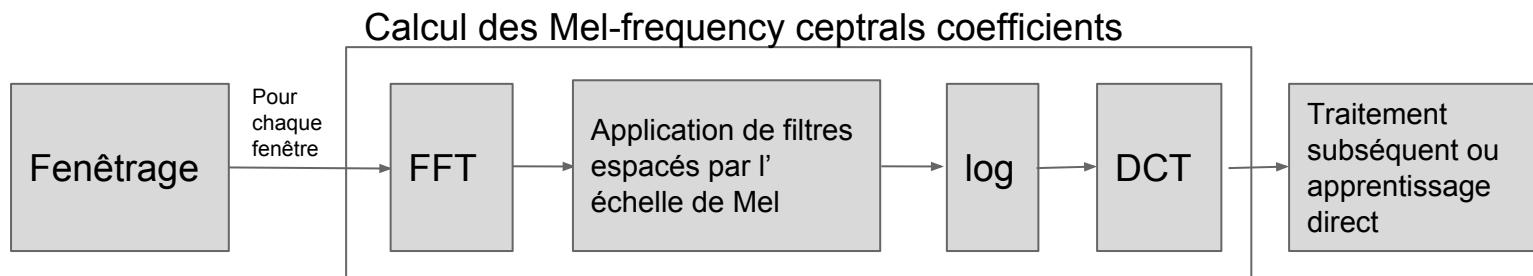


Butterworth filter



Caractéristiques : Mel-frequency cepstral coefficients

- Technique courante en apprentissage de la parole / de l'audio.
- Représente la “résonance des fréquences” sur le spectre d'un (extrait de) signal
- Paramétrable par, entre autre, le nombre de fenêtre et le nombre coefficient



Traitements subséquents possibles :

- Les X résonances les plus fréquentes
- Moyennes des cepstrums pour chaque fréquences de résonnance
- Calcul des dérivées première et secondes entre chaque frame
- Calcul de l'énergie associé à une frame ou une résonnance

Extraction automatique de caractéristiques^[3]

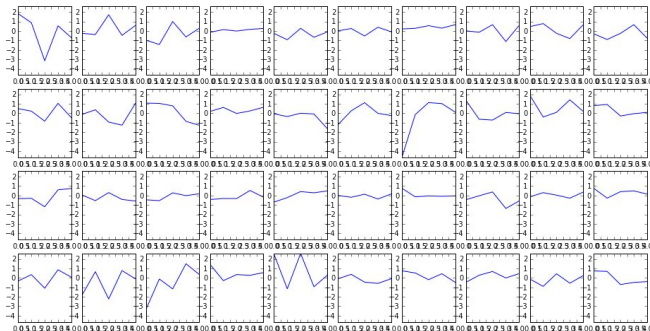
Réseau de neurones à convolutions

Idée :

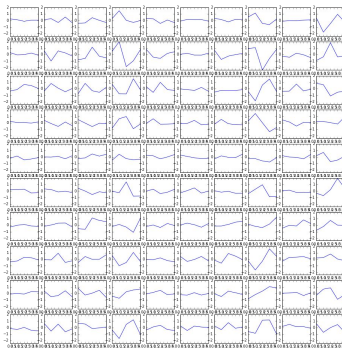
- Exploitation des corrélations temporelles et entre les capteurs
- Prend directement les données brutes en entrée et crée ses features (couches les plus basses)

Exploitation :

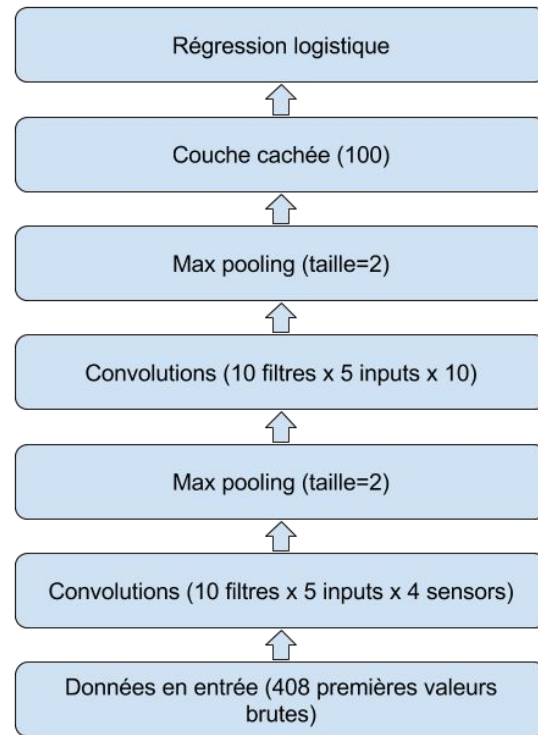
- Adaptation du tutorial lenet (deeplearning.org) pour des conv 1D
- Utilisation de Theano
- Entraînement sur gpu (Kepler GK104 sur aws)



Première couche de convolution



Deuxième couche de convolution



Résultats des expériences

Caractéristiques	Modèle	Problème	Taux correct sur ensemble de test	Baseline (classifieur constant)
moyenne, écart type, nombre de “jumps” positifs et négatifs	k-ppv	Quel chiffre ?	10%	10%
Données brutes	k-ppv (distance DTW)	Quel chiffre ?	10%	10%
Données brutes	k-ppv (distance DTW)	Chiffre vs pas de chiffre	73%	73%
“Mentor”	Gaussian Process	Chiffre vs pas de chiffre	73%	73%
Filtres de Butterworth + puissance	k-ppv	Chiffre vs pas de chiffre	87%	73%
FFT + PCA	k-ppv	Chiffre vs pas de chiffre	97%	73%
Données brutes	CNN	Chiffre vs pas de chiffre	98%	73%

Organisation du travail

Ce qu'on a fait

1. Prétraitement des données EEG et affichages (Thomas)
2. Travail de bibliographie (tout le monde)
3. Extraction de traits caractéristiques
 - a. Amin (Caractéristiques “blocs”)
 - b. Marc (Butterworth + puissance)
 - c. René (FFT + PCA + LR, tiré de ‘Mentor’)
 - d. Thomas (Deep learning)
 - e. Vincent (Mel frequency cepstrum)
4. Tests de classification binaire (-1 vs ≥ 0)
 - a. Thomas (k-ppv caractéristiques simples, k-ppv FFT+PCA, k-ppv distance DTW, réseau de neurones à convolutions)
 - b. Marc (k-ppv Butterworth + puissance)
 - c. René (k-ppv FFT + PCA)
5. Tests de classification multiclasse
 - a. Amin (Bloc + k-ppv)
 - b. Thomas (k-ppv caractéristiques simples, k-ppv distance DTW)

Suite du projet

1. Suite implémentation des extracteurs de caractéristiques
2. Suite des tests de classification binaire (d’autres modèles)
3. Tests de classification multiclasse
4. Utilisation des caractéristiques NN avec d’autres modèles vu en cours
5. Si il reste du temps...
 - a. Regarder les autres jeux de données de Mind big data
 - b. Regarder les liens avec Grasp-and-lift



Questions

Avez vous des questions?

Bibliographie

[1] M. Chaouchi, I Jraidi, C Frasson « MENTOR: A Physiologically Controlled Tutoring System » 2015.

[2] E. Parvinnia, M. Sabeti, M. Zolghadri Jahromi, R. Boostani, «Classification of EEG Signals using adaptive weighted distance nearest neighbor algorithm», 2013.

[3] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, « Gradient-based learning applied to document recognition » , Proceedings of the IEEE, november 1998.

M. Teplan, « Fundamentals of EEG measurement », Measurement science review, 2002.

M. Salama, L. ElSherif, H. Lashin, T. Gamal, « Recognition of Unspoken Words Using Electrode Electroencephalographic Signals » 2014.