A Comparison of Ensemble Methods for Motor Imagery Brain-Computer Interfaces

Davide Valeriani, Ana Matran-Fernandez, Diego Perez-Liebana, Javier Asensio-Cubero, Christian O'Connell and Andrei Iacob

University of Essex

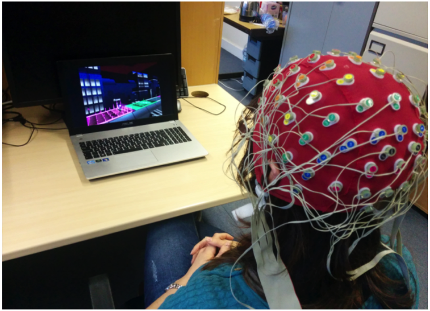
September 3, 2015

Outline

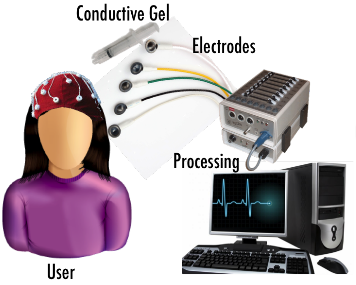
- What is a Brain-Computer Interface (BCI)?
- Motor Imagery BCI
- Data Acquisition
- Data Preprocessing
- Feature Extraction
- Multiclass Classifiers
- Multilayer Ensemble
- Results
- Conclusions

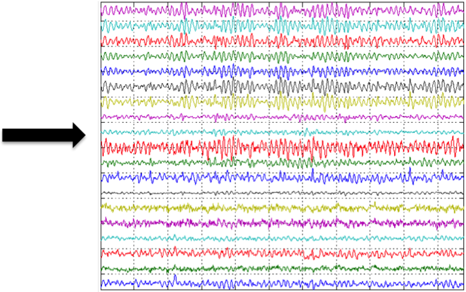
What is a Brain-Computer Interface (BCI)?

- System that converts neural signals into commands for multiple devices.
- BCIs allow people to act on the world without moving any muscle.



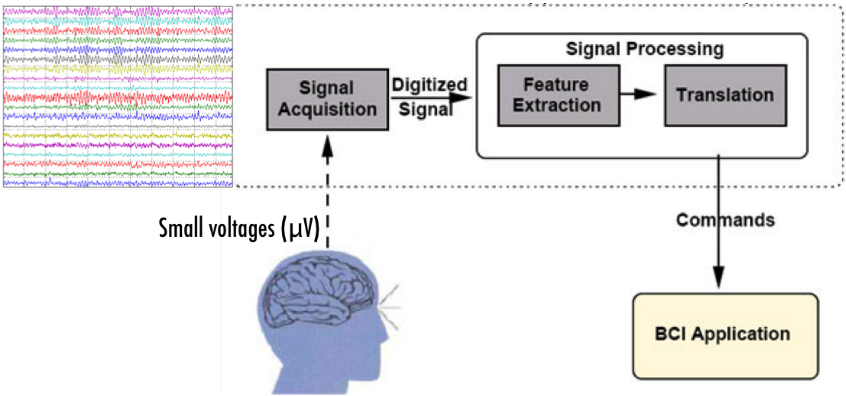
Brain Signals





BCI Architecture

Identify patterns associated with a specific mental action.



Motor Imagery BCI

Imagination of movements of different parts of the body.

Motor Imagery BCI

Imagination of movements of different parts of the body.

Pros

- One of the most used approaches for BCI applications
- Freedom to the user
- Intuitive task to imagine

Motor Imagery BCI

Imagination of movements of different parts of the body.

Pros

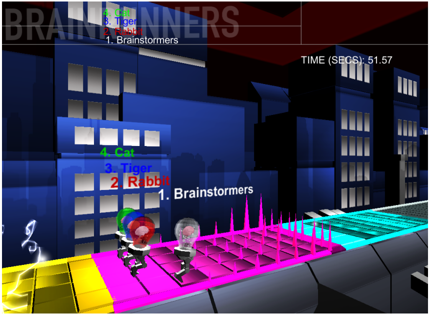
- One of the most used approaches for BCI applications
- Freedom to the user
- Intuitive task to imagine

Cons

- Training required (user and classifier)
- Performance varies across users
- Accuracy drops with multiple classes

Our Application: Gaming

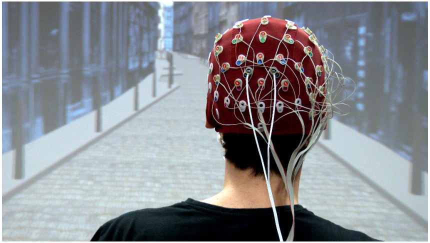
- Racing game where the user controls an avatar.
- Four commands to be sent at different landmarks to speed up the avatar: “run”, “jump”, “roll” and “idle”.
- Penalisation for sending the wrong command.



Paradigm Adopted

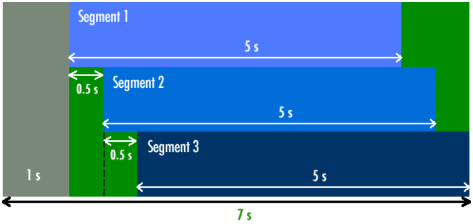
- 4-class motor imagery BCI:

Command	Mental task
Run	Left hand
Jump	Feet
Roll	Right hand
Idle	No movement



Data Acquisition

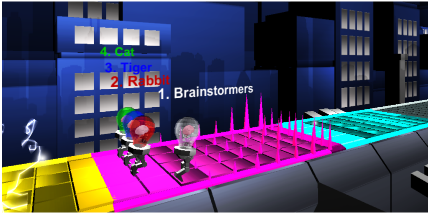
- BioSemi ActiveTwo EEG system with 64 channels.
- Mental task performed for the length of the platform.
- First second of recording discarded for each platform.
- Up to 3 trials extracted per platform for training.
- One trial extracted per platform for testing.



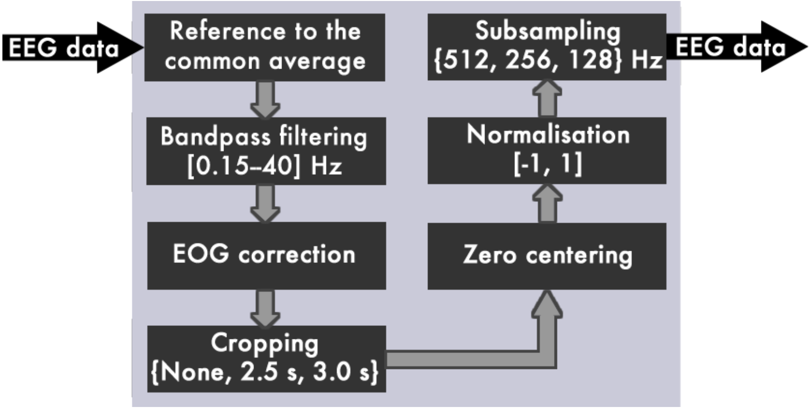
Data Acquisition (cont.)

Number of trials extracted:

Command	Training set	Test set
Run	144	54
Jump	152	45
Roll	126	39
Idle	159	39



Data Preprocessing



Feature Extraction

For each channel:

- Calculate autoregressive model (order 4)

$$x_n = \sum_{k=1}^{4} \alpha_k \cdot x_{n-k} + w(n)$$

- Extract the reflection coefficients (Burg method)
- Calculate the variance to obtain the feature

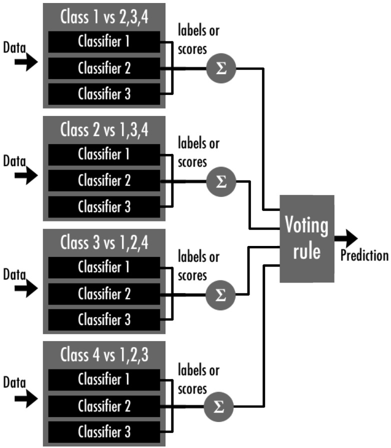
$\longrightarrow$ 64 features in total

Multiclass Classifiers

Three classifiers have been tested:

- Multiclass Linear Discriminant Analysis (LDA)
  $\longrightarrow$ 9 combinations
- Multiclass Least Angle Regression (LARS)
  $\longrightarrow$ 9 combinations
- Multiclass Support Vector Machine (SVM) with linear kernel and $C = 10^i$ where $i \in [-5, 5] \cap \mathbb{Z}$
  $\longrightarrow$ 99 combinations

Multilayer Ensemble



- Dedicated classifier for IDLE class vs default class.
- Labels vs scores as inner-classifiers outputs.
- Use all features for training vs split them in three subsets.
- Use all trials for training vs three different subsets (75%).
- Majority, weighted majority or linear classifier for voting.

$\longrightarrow$ 432 combinations

Metrics

- Accuracy

$$a = \frac{TP + TN}{N}$$

TP $\rightarrow$ true positive   TN $\rightarrow$ true negative   N $\rightarrow$ number of trials

- Cohen's Kappa

$$k = \frac{p_0 - p_e}{1 - p_e}$$

$p_0 \rightarrow$ observed agreement $p_e \rightarrow$ chance agreement

- F1 score

$$F_1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

TP $\rightarrow$ true positive   FP $\rightarrow$ false positive   FN $\rightarrow$ false negative

Results

- 10-fold cross-validation on training set.
- Mean value of Cohen's Kappa on cross-validation used for ranking the multiclass classifiers and ensembles.
- Best five and worst five combinations reported.
- Metrics also computed on the unseen test set.

Multiclass LARS

	Configuration		Accuracy		Cohen's Kappa		F_1 Score	
#	SR	W	CV	Test	CV	Test	CV	Test
1	128 Hz	5.0 s	0.527 ± 0.144	0.322	0.352 ± 0.202	0.116	0.499 ± 0.153	0.254
2	512 Hz	3.0 s	0.516 ± 0.122	0.237	0.340 ± 0.167	-0.005	0.482 ± 0.133	0.211
3	512 Hz	5.0 s	0.492 ± 0.137	0.203	0.305 ± 0.190	-0.048	0.447 ± 0.144	0.139
4	128 Hz	3.0 s	0.481 ± 0.175	0.322	0.296 ± 0.234	0.106	0.453 ± 0.192	0.278
5	512 Hz	2.5 s	0.457 ± 0.072	0.220	0.262 ± 0.101	-0.025	0.420 ± 0.085	0.165
6	256 Hz	5.0 s	0.459 ± 0.082	0.237	0.262 ± 0.108	-0.010	0.411 ± 0.074	0.220
7	256 Hz	3.0 s	0.457 ± 0.176	0.169	0.261 ± 0.239	-0.100	0.424 ± 0.176	0.154
8	128 Hz	2.5 s	0.442 ± 0.119	0.288	0.242 ± 0.158	0.069	0.406 ± 0.105	0.211
9	256 Hz	2.5 s	0.404 ± 0.119	0.254	0.182 ± 0.168	0.010	0.368 ± 0.123	0.218

Legend:

SR $\rightarrow$ Sampling rate

W $\rightarrow$ Window length

Multiclass SVM

	Configuration			Accuracy		Cohen's Kappa		F_1 Score	
#	SR	W	C	CV	Test	CV	Test	CV	Test
1	128 Hz	5.0 s	1e+04	0.751 ± 0.158	0.390	0.670 ± 0.207	0.181	0.742 ± 0.168	0.380
2	128 Hz	5.0 s	1e+02	0.702 ± 0.119	0.390	0.603 ± 0.157	0.190	0.687 ± 0.133	0.330
3	128 Hz	5.0 s	1e+03	0.700 ± 0.151	0.407	0.599 ± 0.202	0.199	0.695 ± 0.152	0.401
4	128 Hz	2.5 s	1e+02	0.695 ± 0.162	0.373	0.591 ± 0.218	0.168	0.683 ± 0.177	0.346
5	128 Hz	2.5 s	1e+04	0.685 ± 0.090	0.356	0.574 ± 0.121	0.154	0.655 ± 0.088	0.272
...									
95	512 Hz	3.0 s	1e-03	0.302 ± 0.022	0.220	0.000 ± 0.000	0.000	0.140 ± 0.018	0.080
96	512 Hz	3.0 s	1e-04	0.302 ± 0.022	0.220	0.000 ± 0.000	0.000	0.140 ± 0.018	0.080
97	128 Hz	2.5 s	1e-04	0.302 ± 0.022	0.220	0.000 ± 0.000	0.000	0.140 ± 0.018	0.080
98	256 Hz	2.5 s	1e-04	0.302 ± 0.022	0.220	0.000 ± 0.000	0.000	0.140 ± 0.018	0.080
99	512 Hz	5.0 s	1e-05	0.302 ± 0.022	0.220	0.000 ± 0.000	0.000	0.140 ± 0.018	0.080

Legend:

SR $\rightarrow$ Sampling rate
W $\rightarrow$ Window length

C $\rightarrow$ Cost of misclassification

Multiclass LDA

	Configuration		Accuracy		Cohen's Kappa		F_1 Score	
#	SR	W	CV	Test	CV	Test	CV	Test
1	512 Hz	2.5 s	0.782 ± 0.166	0.322	0.711 ± 0.219	0.094	0.778 ± 0.163	0.314
2	512 Hz	5.0 s	0.776 ± 0.138	0.339	0.702 ± 0.182	0.117	0.761 ± 0.153	0.325
3	128 Hz	2.5 s	0.775 ± 0.154	0.441	0.701 ± 0.201	0.255	0.772 ± 0.151	0.426
4	128 Hz	5.0 s	0.769 ± 0.192	0.373	0.693 ± 0.251	0.163	0.762 ± 0.197	0.359
5	256 Hz	5.0 s	0.755 ± 0.167	0.339	0.674 ± 0.222	0.123	0.737 ± 0.180	0.314
6	512 Hz	3.0 s	0.755 ± 0.125	0.288	0.670 ± 0.169	0.055	0.748 ± 0.123	0.255
7	256 Hz	2.5 s	0.746 ± 0.116	0.305	0.660 ± 0.156	0.071	0.736 ± 0.121	0.301
8	256 Hz	3.0 s	0.746 ± 0.137	0.305	0.659 ± 0.186	0.077	0.731 ± 0.143	0.286
9	128 Hz	3.0 s	0.727 ± 0.136	0.424	0.634 ± 0.184	0.232	0.716 ± 0.153	0.406

Legend:

SR $\rightarrow$ Sampling rate

W $\rightarrow$ Window length

Ensemble

	Configuration							Accuracy		Cohen's Kappa		F_1 Score	
#	SR	W	I	Vot	TS	FS	Out	CV	Test	CV	Test	CV	Test
1	128 Hz	3.0 s	✓	cla	✓		sco	0.808 ± 0.151	0.339	0.743 ± 0.199	0.117	0.803 ± 0.147	0.313
2	512 Hz	5.0 s	✓	wei			sco	0.788 ± 0.101	0.305	0.718 ± 0.134	0.073	0.779 ± 0.112	0.299
3	256 Hz	3.0 s	✓	cla			lab	0.785 ± 0.143	0.271	0.714 ± 0.186	0.022	0.768 ± 0.149	0.249
4	512 Hz	5.0 s	✓	maj			sco	0.782 ± 0.167	0.424	0.708 ± 0.225	0.237	0.778 ± 0.165	0.397
5	128 Hz	2.5 s	✓	maj	✓		sco	0.778 ± 0.136	0.390	0.704 ± 0.180	0.186	0.768 ± 0.146	0.364
								...					
428	128 Hz	5.0 s		cla		✓	lab	0.252 ± 0.122	0.424	0.012 ± 0.160	0.192	0.184 ± 0.127	0.360
429	128 Hz	2.5 s		cla	✓	✓	sco	0.298 ± 0.041	0.237	0.001 ± 0.055	0.019	0.152 ± 0.042	0.114
430	128 Hz	2.5 s		cla		✓	sco	0.302 ± 0.022	0.220	0.000 ± 0.000	0.000	0.140 ± 0.018	0.080
431	128 Hz	3.0 s		cla		✓	sco	0.302 ± 0.022	0.237	0.000 ± 0.000	0.020	0.140 ± 0.018	0.113
432	128 Hz	5.0 s		cla		✓	sco	0.302 ± 0.022	0.220	0.000 ± 0.000	0.000	0.140 ± 0.018	0.080

Legend:

SR $\rightarrow$ Sampling rate
W $\rightarrow$ Window length
I $\rightarrow$ Separate classifier for Idle
Vot $\rightarrow$ Voting system (majority, weighted majority or classifier)

TS $\rightarrow$ Split trials
FS $\rightarrow$ Split features
Out $\rightarrow$ Output type (labels or scores)

Best Ensemble vs Best Multiclass

Pairwise comparisons using the Wilcoxon rank-sum test between the best ensemble and the best configurations of multiclass classifiers.

Is ↓ better than →	**Ensemble**	**LDA**	**SVM**	**LARS**
Ensemble	—	0.425	0.1278	**0.001244**
LDA #9	0.6044	—	0.988	**0.001408**
SVM #6	0.8874	0.9137	—	**0.006223**
LARS #4	0.999	0.9989	0.995	—

Best Ensemble vs Same Multiclass

Pairwise comparisons using the Wilcoxon rank-sum test between the best ensemble and the same data configuration on multiclass classifiers.

Is ↓ better than →	Ensemble	LDA	SVM	LARS
Ensemble	—	0.08673	**0.04809**	**0.000525**
LDA	0.9246	—	0.2028	**0.001943**
SVM	0.959	0.8179	—	**0.01415**
LARS	0.9996	0.9986	0.9884	—

Conclusions

- Framework for comparing 2-layer ensembles with multiclass classifiers.

Conclusions

- Framework for comparing 2-layer ensembles with multiclass classifiers.
- Multiclass LDA is the only competitive alternative to ensembles.

Conclusions

- Framework for comparing 2-layer ensembles with multiclass classifiers.
- Multiclass LDA is the only competitive alternative to ensembles.
- 4-class ensembles are better than 3-vs-rest.

Conclusions

- Framework for comparing 2-layer ensembles with multiclass classifiers.
- Multiclass LDA is the only competitive alternative to ensembles.
- 4-class ensembles are better than 3-vs-rest.
- Using different subsets of features for each inner classifier reduces performance.

Future Work

- Use data from multiple participants.

Future Work

- Use data from multiple participants.
- Test the winning combinations in an online experiment.

Future Work

- Use data from multiple participants.
- Test the winning combinations in an online experiment.
- Use different features.

Future Work

- Use data from multiple participants.
- Test the winning combinations in an online experiment.
- Use different features.
- Include timing as a metric of evaluation.

Future Work

- Use data from multiple participants.
- Test the winning combinations in an online experiment.
- Use different features.
- Include timing as a metric of evaluation.
- Use a combination of metrics for ranking.

Questions?