Learning from Label Proportions in BCI -- A Symbiotic Design for Stimulus Presentation and Signal Decoding

David Hübner¹, Thibault Verhoeven², Klaus-Robert Müller³, Michael Tangermann¹, Pieter-Jan Kindermans³

[1] Albert-Ludwigs-Universität Freiburg, Germany
[2] Ghent University, Belgium
[3] Technische Universität Berlin
[4] Korea University, Seoul
BCI: We have seen many improvements
Improvement of coding schemes

BCI: We have seen many improvements

- Improved signal quality
- Smaller latencies
- Improved classification algorithms
- Spatial filter
- Transfer learning
- Language models

Hardware

Paradigm

Decoder

Improve stimulus saliency
- Using novel sensory modalities
- Improved signal quality
- More channels

Early stopping

Improved signal quality
Smaller latencies
Improved classification algorithms
Spatial filter
Transfer learning
Language models

Hardware

Paradigm

Decoder
What is the objective?

- Derive an **unsupervised** classifier for ERP signals with **guarantees**
- Use a version of a linear discriminant analysis (LDA) classifier:

\[ w = \sum_{C}^{-1} (\mu_T - \mu_N) \]

 Projection

Class-wise covariance matrix

Class means

*Blankertz et al., NeuroImage, 2010*

*Bishop, Springer, 2006*
What is the objective?

- Derive an **unsupervised** classifier for ERP signals with **guarantees**
- Use a version of a linear discriminant analysis (LDA) classifier:

$$\mathbf{w} = \Sigma^{-1} (\mu_T - \mu_N)$$

- **Unsupervised learning problem boils down to estimating class means**

$$(\mu_T, \mu_N)$$

**Projection**

**Class means**

**Global covariance matrix**

**Class-wise covariance matrix**

*Blankertz et al., NeuroImage, 2010*

*Bishop, Springer, 2006*
IDEA: Jointly improve decoder and paradigm

- Introduction of Learning from Label Proportions (LLP) for ERP-based BCIs
- Original method by Quadrianto et al., 2009
- No explicit label information are necessary, use known proportions in subgroups of the data instead
- In BCI: Modification of the paradigm is necessary
IDEA: Jointly improve decoder and paradigm

- Introduction of Learning from Label Proportions (LLP) for ERP-based BCIs
- Original method by Quadrianto et al., 2009
- No explicit label information are necessary, use known proportions in subgroups of the data instead
- In BCI: Modification of the paradigm is necessary
LLP: Theoretical Background

(A) Spelling a letter
(B) Stimuli and Mix Ratios

\[
\begin{align*}
\mu_1 &= \frac{3}{8} \mu_T + \frac{5}{8} \mu_N \\
\mu_2 &= \frac{2}{18} \mu_T + \frac{16}{18} \mu_N
\end{align*}
\]

Hübner et al., PloS One, 2017
LLP: Theoretical Background

(C) Average

(D) Solving Linear System

(E) Estimation of Class Means

\[ \mu_T = 3.37\mu_1 - 2.37\mu_2 \]

\[ \mu_N = -0.42\mu_1 + 1.42\mu_2 \]
**LLP: Theoretical Background**

- Remember, we started from:

  \[ w = \Sigma^{-1} (\mu_T - \mu_N) \]

- With LLP, we obtained:

  \[ \mu_T = 3.37\mu_1 - 2.37\mu_2 \]
  \[ \mu_N = -0.42\mu_1 + 1.42\mu_2 \]

- Central limit theorem implies for \( N \to \infty \):

  \[ \hat{\mu}_1 \to \mu_1, \hat{\mu}_2 \to \mu_2 \]

\[ \Rightarrow \text{LLP is guaranteed to recover } w \]

*Hubner et al., PloS One, 2017*
**LLP: Online Study**

- **N=13** healthy young subjects performed 32-channel EEG study
- Modified visual ERP speller
- Copy-spelling sentence three times
- Classifier started from scratch, was retrained after each character
- Labels were used only to assess performance

---

"Franzy jagt im komplett verwahrlosten Taxi quer durch Freiburg."

---

**Sentences**
3x63 chars

**Trial**
25s

**Get Ready (4s)**

**Stimulation (17s)**

**Feedback (4s)**

---

*Hubner et al., PloS One, 2017*
LLP works!

Hübner et al., PloS One, 2017
Comparison to other unsupervised methods

- The expectation-maximization (EM) algorithm by Kindermans et al. is probably the only algorithm for ERP which was successfully applied online without using any labeled data.

For EM in ERP-speller: Use selected character as latent variable!

Kindermans et al., PloS One, 2012

Bishop, Springer, 2006
LLP vs Expectation-Maximization

Verhoeven et al. 2017, Journal of Neural Engineering
New Idea: Combine EM and LLP

- Both EM and LLP yield a class-wise mean estimation
- MIX mean estimation is a linear combination of LLP and EM

\[ \hat{\mu}(\gamma) = (1 - \gamma)\hat{\mu}_1 + \gamma\hat{\mu}_2 \]

- Explicit solution can be found:

\[ \gamma^* = \frac{1}{2} \left( \frac{\sum_d Var[\hat{\mu}_{1,d}] - \sum_d Var[\hat{\mu}_{2,d}]}{||\hat{\mu}_1 - \hat{\mu}_2||^2} \right) + 1 \]

- Optimal weight is computed by considering the variance of the estimators (calculated on the data)

Verhoeven et al. 2017, Journal of Neural Engineering

New method is called MIX
How does MIX work in simulations?

Verhoeven et al. 2017, Journal of Neural Engineering
How does MIX work in simulations?

Verhoeven et al. 2017, Journal of Neural Engineering
How does MIX work in simulations?

Verhoeven et al. 2017, Journal of Neural Engineering
Comparison MIX vs supervised

Verhoeven et al. 2017, Journal of Neural Engineering
Mixing unsupervised estimators (online)

~3mins

Subject#

# Characters

EM

LLP

MIX

AUC (%)

# Characters

Hübner et al. 2017, in Preparation
Conclusion

• New unsupervised learning method „Learning from Label Proportions“

• LLP + EM → MIX

• MIX combines unsupervised and supervised properties:
  ✓ No calibration
  ✓ Continuous learning
  ✓ High decoding performance
  ✓ Guaranteed convergence
  ✗ Modification of the spelling paradigm for LLP/MIX is required

• Transfer learning and language models, etc. can easily be incorporated
Acknowledgement

Brain-State Decoding Lab, Freiburg
headed by Michael Tangermann

Pieter-Jan Kindermans
Thibault Verhoeven
Klaus-Robert Müller

Contact:
david.huebner@blbt.uni-freiburg.de
References


