 DOI: 10.3217/978-3-85125-962-9-122

Towards including covariates in EEG classification - a preliminary study on simulated data

David Trocellier[1]*, Bernard N'Kaoua[2], Fabien Lotte[1]

[1] Inria center at the university of Bordeaux/ LaBRI, Talence, France

[2] Univ. Bordeaux, Bordeaux, France

*200 avenue de la Vieille Tour, 33405 Talence, France. E-mail: david.trocellier@inria.fr

Introduction: The classification models commonly used in BCIs do not take into account covariates such as changes in users' mental states, such as motivation or fatigue, which can influence EEG signal dynamic [1] and affect classification performances. We propose a modification of linear discriminant analysis (LDA) [2] to account for these interfering covariates in order to improve performance. Our modification aims to make the LDA linear projection independent of such covariates for improved performance.

Material, Methods and Results: LDA aims to reduce the dimension of the input variables while separating them into classes using a linear combination of the input variables. To do so, it computes the input variable means and covariance matrices for each class. They are used to determine the linear projection that maximizes the distance between class means and minimizes variance within each class. Our method, named independant LDA (iLDA), takes as input a vector of x_i, which is an original variable x to which we subtracted the linear influence that the covariate z has on x, using linear regression (here we used the mean square error regression method), see equation 1. This allows us to clearly consider that the covariate has an influence on the variable and to reduce such influence by projection.

$$x_i = x - (a * z + b) \qquad (1)$$

Equation 1 . x is the original input variable, x_i is the variable were the linear influence of the covariate was removed, a and b are the regression coefficients and error term of the linear regression predicting x from z, z is the covariate

The algorithm was tested on simulated data where each variable follows a normal distribution, with different distributions for each class. A randomly chosen covariate, with a linear influence on all variables, was generated, and a small Gaussian noise added to the covariate. The data have been generated for a different set of variables, from 2 to 100, and a different number of training instances, from 25 to 1000. Each set of condition is repeated 1000 times. The simulated data was tested with three classifiers: a standard LDA, an LDA with the covariate added as input variable (covLDA), and our proposed iLDA. Results showed that both LDAs including the covariate (i.e., covLDA and iLDA) improved classification accuracy (respectively 92.1% and 92.5%) compared to standard LDA (91.3%) and that our model performed significantly better than covLDA in cases with few training instances and many variables.

Discussion: These preliminary results on simulated data showed the interest of considering covariates in classifiers. Our next steps is to test such algorithms with real EEG datasets and features (e.g., band power or CSP features), and covariates that are known to have an influence on EEG signals such as fatigue, attention, motivation, expertise or artifacts, as measured using, e.g., questionnaires.

References

[1] U. Talukdar, et al. « Adaptation of Common Spatial Patterns based on mental fatigue for motor-imagery BCI », Biomedical Signal Processing and Control, vol. 58, p. 101829, avr. 2020,
[2] P. Xanthopoulos, et al. «Linear Discriminant Analysis», in Robust Data Mining, Éd. New York, NY: Springer, 2013, p. 27-33.

Published by Verlag der Technischen Universität Graz