International Journal of Bioelectromagnetism http://www.ijbem.org
Vol. 10, No. 1, pp. 52 - 55 2008

Better than random? A closer look on BCI results

Gernot R. Müller-Putz, Reinhold Scherer, Clemens Brunner, Robert Leeb, Gert Pfurtscheller

Institute for Knowledge Discovery, BCI-Lab, Graz University of Technology, Graz, Austria

Correspondence: G. R. Müller-Putz, Institute of Knowledge Discovery, BCI-Lab, Graz University of Technology, Krenngasse 37, 8010 Graz, Austria: gernot.mueller@tugraz.at, phone +43-316-873-5313, fax +43-316-873-5349

Abstract. Brain-Computer Interface (BCI) research has become a growing field of interest in the last years. The work presented ranges from machine learning approaches in offline results to the application of a BCI in patients. However, reliable classification of brain activity is a crucial issue in BCI research. In contrast to most articles which present methods to enhance classification accuracies, we investigate the opposite side in this work and provide answers to the question: Does my classifier perform better than random?

Keywords: Brain-Computer Interface, statistical analyses, classification accuracy, pattern recognition methods

1. Introduction

Brain-Computer Interface (BCI) research is a growing field [Pfurtscheller et al., 2006]. As a consequence numerous papers appeared during the last years [e.g., Dornhege et al., 2007]. Most articles introduce new feature extraction, optimization or classification methods. However, to be able to estimate the reliability of a new method and compare the achieved results with results obtained by other algorithms, some standard signal processing stages are necessary. One of these standards, and often recommended by reviewers, is the use of a cross-validation statistic when presenting offline classification results. This procedure prevents the classifier from over fitting the data (curse of dimensionality) [Duda and Hart, 1973]. Related to this, it is not only meaningful to present classification accuracies, but also the number of trials on which the computations are based. Exemplarily, the chance level in a simple 2-class paradigm is not exactly 50%; more precisely, it is 50% with a confidence interval at a certain level α depending on the number of trials.

The aim of this paper is to provide more general knowledge about the relation between the number of trials and the classification accuracy.

2. Material and Methods

We decided to investigate the problem in two ways: (i) for the case of a 2-class BCI we present a theoretical approach, (ii) for BCIs based on more than 2 classes we performed a simulation.

2.1 Theoretical

The probability of a correctly classified trial X_i in a two-class paradigm consisting of n trials follows a binomial distribution with $p = 0.5$ (both classes are equally likely to occur) with a classifier that performs at random level. Thus, a confidence interval around the expected value of p can be calculated as follows [Agresti & Caffo, 2000]:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

This unbiased estimator is replaced by

$$\tilde{p} = \frac{n\bar{X} + 2}{n + 4},$$

and the confidence intervals are then given by

$$\tilde{p} \pm \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} z_{1-\frac{\alpha}{2}}.$$

In the equation above, $z_{1-\frac{\alpha}{2}}$ is the $1-\frac{\alpha}{2}$ quantile of the standard normal distribution $N(0,1)$ and α is the significance (typically, $\alpha = 0.05$ or $\alpha = 0.1$).

For example, in a BCI experiment consisting of $n = 100$ trials (50 per class), the expected chance level would be at exactly 50 correctly classified trials (with equally probable classes). If the reported accuracy of a classifier is 59 correctly classified trials (or alternatively, 59%), it is straightforward to see that this probability does not lie within the theoretical limits of 40.39% and 59.61% (for a confidence of $1 - \alpha = 0.95$). Thus, it can be assumed that the given classifier does not significantly differ from a random one.

For more than two classes, the model can be extended to a multinomial one, where several possible estimators exist [e.g., Genz & Kwong, 2000].

2.2 Simulation

For multi-class BCIs, a simulation was implemented as follows: a random class label vector containing normally distributed class information of either a 2-, 3-, 4- or 8-class BCI was generated. Then a new vector with the same number of labels – but randomized - was generated, compared with the original class labels and an error was calculated. This procedure was repeated 10000 times and a histogram was computed. Finally, the confidence intervals of α=5% and α=1% were applied. The whole procedure resulted in the number of trials (and accuracy in [%]) which are at the border of random classification. The simulation results can be seen in Figure 2. Table 1 gives an overview over the most reported number of trials.

3. Results

Figure 1 shows the theoretical 95% and 99% confidence limits of a chance result for a 2-class BCI – keep in mind that the theoretical level is at 50%, whereas in Figure 2 the simulations results of a 2-, 3-, 4- and 8-class BCI for both confidence limits are presented. It is important to note that the theoretical chance levels are 50%, 33.3% 25% and 12.5% respectively. In Table 1 we present some results of commonly used numbers of trials for those multi-class BCIs.

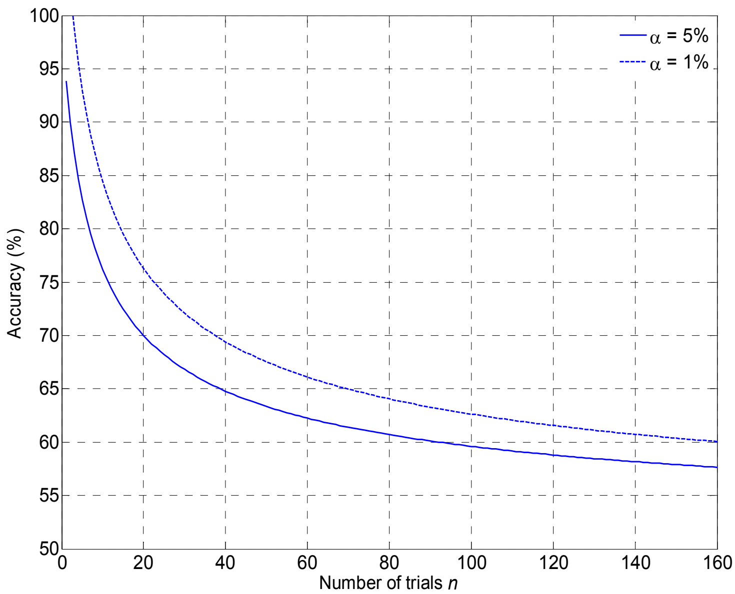

***Figure 1**. Confidence limits of a chance result in a two-class paradigm for a significance level of α=5% (solid line) and α=1% (dashed line) depending on the number of trials per class.*

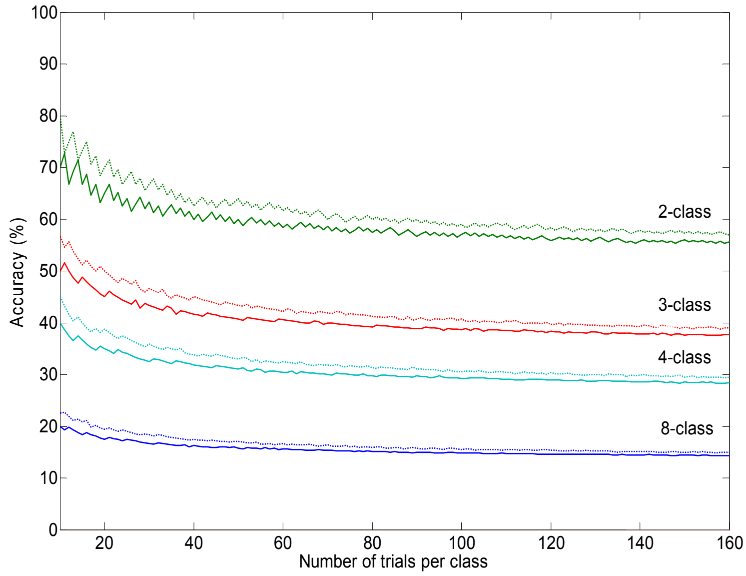

Figure 2. *Confidence limits of chance results in 2-, 3-, 4-, and 8-, class paradigms for a significance level of α=5% (solid line) and α=1% (dotted line) depending of the number of trials per class.*

Table 1*. Simulation results for 2-, 3-, 4-, and 8-class BCI with commonly used number of trials per class. The columns represent the upper confidence limits of a chance result (in %) and in brackets the number of trials. Results are shown for α=5% and α=1%.*

trials/ class	*2-cl α=5%*	*2-cl α=1%*	*3-cl α=5%*	*3-cl α=1%*	*4-cl α=5%*	*4-cl α=1%*	*8-cl α=5%*	*8-cl α=1%*
10	70.0 (14)	80.0 (16)	50.0 (15)	56.7 (17)	40.0 (16)	45.0 (18)	20.0 (16)	22.5 (18)
20	65.0 (26)	70.0 (28)	45.0 (27)	50.0 (30)	35.0 (28)	38.8 (31)	18.1 (29)	19.4 (31)
40	60.0 (48)	65.0 (52)	41.7 (50)	45.0 (54)	31.9 (51)	34.4 (55)	16.3 (52)	17.5 (56)
80	57.5 (92)	60.0 (96)	39.6 (95)	41.3 (99)	29.7 (95)	31.3 (100)	15.2 (97)	15.9 (102)
160	55.6 (178)	56.9 (182)	37.7 (181)	39.0 (187)	28.3 (181)	29.5 (189)	14.3 (183)	14.8 (190)

4. Discussion

Brain-Computer Interface research is a multidisciplinary field and requires a high number of experts, e.g., neurophysiologists, biomedical engineers, computer scientist as well as psychologists. Though having a lot of experts, sometimes important issues are overlooked. In this work we want to direct the researcher's attention to closer look on the number of trials used in their experiments.

The theoretical as well as the simulated results clearly show that the number of trials used have an important impact on the confidence interval of the chance level. When looking on results obtained from 20 trials each class (2-class paradigm, α=1%) the level of chance is at 70%. Comparing weak BCI results with 50%-chance level can lead to a wrong interpretation of the data.

Exemplarily, we are discussing the results reported above by having a closer look on two papers already published by our group.

- In Guger et al. [2001] a 2-class BCI (80 trials/class) training was performed. The classification results presented in this paper showed that for all 3 subjects in the first session (training without feedback) the classification accuracy was random at several time points during the trial. With further training, subjects improved up to 70%-95% accuracy.
- In the 4-class BCI offline study (72 trials/class) of Naeem et al. [2006] different ICA algorithms were investigated. The result of one subject out of eight was at the border to

random with 32.6% with one specific method (session to session transfer, ICA SOBI algorithm). Here the chance level α=1% is at 31.6%.

5. Conclusion

Concluding, we want to encourage researchers to take into account the proposed considerations and to check their results also in relation to the real level of chance and not only to the theoretical one.

Acknowledgements

This work was supported by the EU COST Action BM0601 "NeuroMath" and Wings for Life – The Spinal Cord Foundation project 002/06.

References

Agresti A, Caffo B. Simple and Effective Confidence Intervals for Proportions and Differences of Proportions Result from Adding Two Successes and Two Failures. The American Statistician 54(4):280-288, 2000.

Dornhege G, del R. Millán J, Hinterberger T, McFarland DJ, Müller KR. Toward Brain-Computer Interfacing. The MIT Press, Cambridge, 2007.

Duda RO, Hart PE. Pattern classification and scene analysis, New York: Wiley, 1973.

Genz A, Kwong K-S. Numerical evaluation of singular multivariate normal distributions. Journal of Statistical Computation and Simulation 68(1):1-21, 2000.

Guger C, Schlogl A, Neuper C, Walterspacher D, Strein T, Pfurtscheller G. Rapid prototyping of an EEG-based brain-computer interface (BCI). IEEE Trans Neural Syst Rehabil Eng. 9(1):49-58, 2001.

Naeem M, Brunner C, Leeb R, Graimann B, Pfurtscheller G. Seperability of four-class motor imagery data using independent components analysis. J Neural Eng. 2006 Sep;3(3):208-16, 2006.

Pfurtscheller G, Müller-Putz GR, Schlögl A et al. 15 years of BCI research at Graz University of Technology: current projects. IEEE Trans Neural Syst Rehabil Eng. 14(2):205-10, 2006.