How Much Features in Brain-Computer Interface Are Discriminative?– Quantitative Measure by Relative Entropy

Sangtae Ahn, Sungwook Kang, and Sung Chan Jun

School of Information and Communications,
Gwangju Institute of Science and Technology, South Korea
{stahn,scjun}@gist.ac.kr

Abstract. Brain Computer Interface (BCI) gives opportunities to control a computer or a machine by imagination of limb movement, which activates somatosensory motor region in a discriminative manner. As far as it has been concerned, it has been not well investigated how much the given (extracted) features in BCI are discriminative in the sense of information theory. For this purpose, we cast the feature spaces corresponding to given conditions into probability spaces by yielding corresponding probability distributions. Then the relative entropy (measures to estimate the difference between two probability distributions) is introduced to measure the distance between these probability distributions. Such a distance represents well how two feature spaces are separable. We compare this distance with BCI performance (classification success rate) to see their correlation.

Keywords: Brain Computer Interface, Relative Entropy, Information Theory.

1 Introduction

The development of a brain computer interface (BCI) aims to provide a communication channel from a human to a computer or a machine, that is, it directly translates brain activity into sequences of control commands [1]. BCI classification success rate is a barometer to measure how well the BCI system works; it strongly depends on many factors such as a variety of preprocessing, features to be used, classification techniques. We have an intrinsic question of how much the given (extracted) features in BCI are discriminative. It has been investigated in many ways, but it has not been in the sense of information theory as far as it has been concerned. In this work, we try to investigate how two different conditioned EEG data is discriminative in a theoretical basis and how two feature spaces are well separable. To investigate these issues, the relative entropy as an information-theoretical measure is applied to several motor imagery movement EEG datasets and it is discussed how relative entropy and BCI success rate are related.

2 Data

2.1 Experimental Setup and Preprocessing

Experimental EEG datasets were recorded from five healthy subjects. Each dataset contains only data from the 4 initial sessions without feedback. Visual cues indicated

C. Stephanidis (Ed.): Posters, Part II, HCII 2011, CCIS 174, pp. 274–278, 2011.

for 3.5 seconds which of the following 3 motor imageries the subject should perform: (L) left hand, (R) right hand, (F) right foot. The presentation of target cues was intermitted by periods of random length, 1.75 to 2.25 seconds, in which the subject could relax. There were two types of visual stimulation: (1) where targets were indicated by letters appearing behind a fixation cross, and (2) where a randomly moving object indicated targets (inducing target-uncorrelated eye movements). From subjects ‘al’ and ‘aw’, two sessions of both types were recorded, while from the other subjects, three sessions of type (2) and one session of type (1) were recorded. EEG data were acquired from 118 channels attached on the scalp according to the extended international 10/20-system. Signals were then digitized at 1000 Hz with 16 bit (0.1 μV) accuracy and band-pass filtered between 0.05 and 200 Hz. In this work, we analyzed (R) right hand and (F) right foot datasets.

In order to α or β rhythms, they were band-pass filtered between 8 and 30 Hz and down-sampled at 100 Hz. Also, we extracted temporal window of 1 second long for resting state and 3 seconds long for testing state (imagination period). It is known that right hand and right foot movements generate ERS/ERD signals mainly around C3 and Cz channels. Thus, probability density functions (pdf) within the short-time window (200 ms) corresponding to C3 and Cz channels are estimated through Gaussian Mixture Model (GMM) and EM (Expectation Maximization) algorithm. For each trial per each channel (time series of 4 seconds long), we generated informative samples as follows: EEG time series were averaged over moving short-time window (200 ms) and the window was sliding by 100 ms with overlapping 100 ms. Then for each trial a total of 40 samples were generated (10 samples from the resting state and 30 samples from the imagination period). This procedure was performed over all trials in the same manner.

2.2 Classification Success Rate

We calculated classification success rate from five healthy subjects using CSP and FLDA (Fisher Linear Discriminant Analysis). This classification success rate is compared with an amount of discriminative information later.

Table 1. Classification success rate of BCI Competition III datasets (IVa)

Subjects	*‘aa’*	*‘al’*	*‘av’*	*‘aw’*	*‘ay’*
Success rate (%)	70.5	100	58.6	82.5	84.1

3 Methods

3.1 Gaussian Mixture Model (GMM)

In order to calculate the pdf, we used Gaussian Mixture Model (GMM) [2]. GMM is a parametric probability density function expressed by a weighted sum of Gaussian density functions [3]:

$$p(x \mid \lambda) = \sum_{i=1}^{M} w_i g(x \mid \mu_i, \Sigma_i), \tag{1}$$

where x is a *D*-dimensional vector representing features, $w_i, i = 1,...,M$ are mixture weights, and $g(x \mid \mu_i, \Sigma_i), i = 1,...M,$ are Gaussian density functions, each function is expressed by a *D*-variate Gaussian function of the form:

$$g(x \mid \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left\{ -\frac{1}{2}(x - \mu_i)' \Sigma_i^{-1} (x - \mu_i) \right\} \tag{2}$$

with mean vector μ_i and covariance matrix Σ_i. The mixture weights satisfy the constraint that $\Sigma_{i=1}^{M} w_i = 1$.

3.2 EM (Expectation Maximization) Algorithm

EM (Expectation Maximization) algorithm is a general technique for finding a maximum likelihood solution having latent variables [2]. It is an iterative optimization method to estimate some unknown parameters Θ for the given measurement data U [4]. However, any “hidden” nuisance variables J should be integrated out as follows:

$$P(\Theta \mid U) = \sum_{J \in J^n} P(\Theta, J \mid U) \tag{3}$$

Then the posterior probability density function is maximized:

$$\Theta^* = \arg\max_{\Theta} P(\Theta \mid U) \tag{4}$$

There are two steps in EM algorithm, the first step is called the ‘expectation-step’ or E-step whereas the second step is called the ‘maximization-step’ or M-step. In the E-step, $f^t(J) \cong P(J \mid U, \Theta^t)$ is evaluated using the current guess Θ^t, whereas in the M-step it is optimized as follows: $\Theta^{t+1} = \operatorname{argmax}_{\Theta} [Q^t(t) + \log P(\Theta)$ with respect to the free variable Θ.

3.3 Relative Entropy

The entropy of a random variable is a measure of the uncertainty of the random variable; it is a measure of the amount of information required on the average to describe the random variable [5]. The relative entropy is a measure of the distance between two distributions. In statistics, it arises as an expected logarithm of the likelihood ratio. The relative entropy $D(p \| q)$ is a measure of the inefficiency of assuming that the distribution is q when the true distribution is p. The relative entropy or Kullback-Leiber distance between two probability density functions $p(x)$ and $q(x)$ is defined as

$$D(p \parallel q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} \tag{5}$$

4 Results

Figure 1 shows that relative entropy is higher in magnitude for testing state than for resting state. Distance between C3 and Cz channels increases relatively to resting state after onset of cue in Figure 1, especially subject 'al' or 'ay', but it does not in subject 'av', as shown in Figure 1. Based on these results, we found that the higher success rate has tendency in yielding the relatively higher relative entropy for testing state. After onset of cue, high amplitude of relative entropy means that difference of amount of information appeared noticeably between C3 and Cz channels.

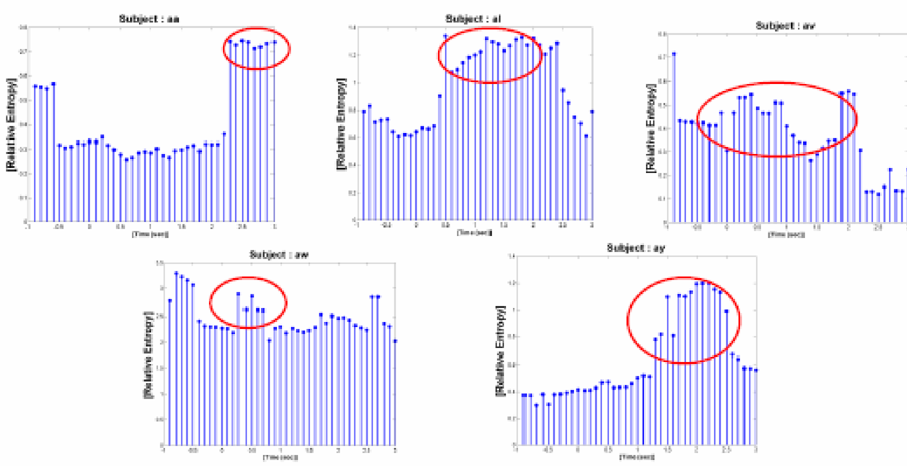

Fig. 1. Time series of relative entropy. Time 0 represents the onset of cue.

5 Conclusion

In general, classification success rate represents that how much the extracted features in BCI are discriminative, but we try to view them in a different manner, namely measuring information amount using relative entropy (measures to estimate the difference between two probability density functions). In relative entropy, we observed how far two different conditioned EEG data are distributed away. For each conditioned EEG data, we modeled a probability density function as a Gaussian Mixture Model and estimated it through EM algorithm. Then the relative entropy was adopted to measure discriminability between them. As a result, we observed visible correlation between classification success rate and difference between averaged relative entropies over resting state and testing state. Thus, relative entropy may be a possible measure of extent of discriminability between two conditioned EEG data, thereby, being applicable to BCI.

Acknowledgements. This work was supported by a NRF grant (NRF-2010-0006135), and a NIPA grant (NIPA-2011-C1090-1131-0006).

References

1. Blankertz, B., et al.: Spatio-Spectral Filters for Improving the Classification of Single Trial EEG. IEEE Transactions on Biomedical Engineering 52(9) (September 2005)
2. Theodoridis, S., Koutroumbas, K.: Pattern Recognition, 4th edn. Elsevier, Oxford (2009)
3. Reynolds, D.: Gaussian Mixture Models, MIT Lincoln Laboratory, 244 Wood St., Lexington, MA 02140, USA
4. Dellaert, F.: The Expectation Maximization Algorithm. College of Computing, Gerorgia Institute of Technology Technical Report number GIT-GVU-02-20 February (2002)
5. Cover, T.M., Thomas, J.A.: Elements Of Information Theory, 2nd edn. Wiley, New Jersey (2006)