

Approximate Entropy for EEG-based Movement Detection

M. Dyson, T. Balli, J. Q. Gan, F. Sepulveda, R. Palaniappan

BCI Group, Dept. of Computing and Electronic Systems, University of Essex, Colchester, UK

<mdyson,tballi>@essex.ac.uk

Abstract

An approximate entropy feature is tested with parameters appropriate for online BCI - a short calculation window and use of the running standard deviation of the EEG signal. Features are extracted from self-paced real movement data, with various values of the embedding dimension and tolerance of comparison. Two alternative features, band power and reflection coefficients, are extracted for comparative purposes. Class separability is measured using classification results from K-means clustering for individual features and linear discriminant analysis for multiple features, as selected by sequential forward floating search. Results show this method of calculating approximate entropy to be a candidate for online movement detection in self-paced BCI systems.

1 Introduction

In many traditional feature extraction methods it is assumed that the fundamental signal characteristics are contained in the amplitude and the frequency spectrum. However for some signals these features are insufficient as signals belonging to different classes have different bandwidths. Such signals can be best distinguished using complexity measures which are independent of the precise frequency content of the signal [1].

In recent years, there have been many research studies on nonlinear complexity measures for analysis of EEG signals [2, 3, 1]. These methods are used for analysis of the electrophysiological condition of subjects, discrimination of mental tasks, and diagnosis of different pathological conditions such as epilepsy, memory impairments and sleep disorders. It is reported that signal complexity is correlated with the mental and physiological condition of subjects.

In this paper the detection of index finger movement using a nonlinear complexity measure, approximate entropy, is investigated. Detection rates using two linear features, band power and reflection coefficients, are included for comparison.

2 Methods

2.1 Approximate Entropy

Approximate entropy (ApEn) is a recently developed method that measures the irregularity of time series data [4]. This measure of irregularity is obtained by comparing the original time series with time shifted versions of itself. For this purpose the original signal is reconstructed in phase space using time delay embedding. The number of previous data points used for making the prediction of the next data point is termed the embedding dimension, m .

Assuming we have EEG data from a single channel;

$$x = [x(1), x(2), \dots, x(N)] \quad (1)$$

with N data points, a sequence of vectors are constructed with time delay embedding as follows;

$$y = [y_1, y_2, y_3, \dots, y_M] \quad (2)$$

where

$$y_i = [x(i), x(i + \tau), x(i + 2\tau), \dots, x(i + (m - 1)\tau)]^T \quad (3)$$

for $i=1,2,\dots,N-(m-1)\tau$. The next step is the calculation of the correlation integral using the reconstructed vectors, which is defined by;

$$C_i^m(r) = \sum_{j=1}^{N-(m-1)} \frac{\Theta(r - \|y_i - y_j\|)}{N - (m - 1)} \quad (4)$$

where N is the length of time series, r is the tolerance of comparison, y_i and y_j are vectors constructed in phase space, $\|\cdot\|$ represents the Euclidean distance between vectors and $\Theta(x)$ is the heaviside function such that $\Theta(x) = 1$ if $x > 0$ and $\Theta(x) = 0$ if $x \leq 0$. The approximate entropy $ApEn(m,r)$ is obtained by;

$$ApEn(m, r) = \Phi^m(r) - \Phi^{m+1}(r), \quad (5)$$

where

$$\Phi^m(r) = \frac{1}{N - (m - 1)} \sum_{i=1}^{N-(m-1)} \ln[C_i^m(r)] \quad (6)$$

2.2 Recording

Signals were acquired at 256Hz. Five bipolar EEG channels were recorded over the motor cortex at locations C3, C1, Cz, C2 and C4 as shown in figure 1. EMG was recorded from the flexors of the left forearm. A right mastoid reference channel was used. Signals were acquired using a Guger Technologies g.BSamp. Data was recorded from eleven right handed subjects, three subjects were female, ages ranged from 23 to 46. Subject 1 was experienced using a BCI system based on self-paced movement, Subjects 7 to 11 had experience in offline BCI experiments, the remaining subjects were naive to BCI use.

As data was un-cued the number of trials performed within each run was variable. Each subject performed three runs in a single session. A run lasted 610 seconds. After a five second pre waiting period a fixation cross appeared on the screen. The fixation cross remained on the screen for 10 minutes during which time data was acquired. A five second post waiting period was used.

Within each run subjects were instructed to perform self paced flexion / extension of the left index finger whilst the fixation cross was visible. Subjects were requested to perform each movement for between 5 and 10 seconds and to rest for at minimum 10 seconds between movements. Instructions were given to concentrate on the fixation cross as much as possible during each run. After each run EMG recordings were assessed to ensure subjects understood requirements and could moderate actions accordingly.

2.3 Feature Extraction and Parameter Initialisation

ApEn was calculated using a window of 32 samples with an overlap of 31 samples; a 1 second averaging window was applied. Prior testing had determined that this window size demonstrated promising results whilst retaining a calculation time appropriate for online use.

The calculation of ApEn involves selection of three parameters namely the embedding dimension m , time delay τ , and the distance within which the neighboring trajectory points must lie (tolerance of comparison) r . There is no fixed manner of determining m and τ values used in the phase space reconstruction of the time series. In this study ApEn values were calculated for m values ranging from 1 to 10, with τ fixed at 1, as suggested in [5].

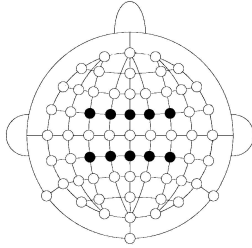


Figure 1: Electrode Layout

In previous studies it has been suggested that r should be calculated via the product of the standard deviation of the original signal [4] with a coefficient value. In order to estimate an appropriate range of coefficient values for r a prior investigation was performed using data recorded from three subjects (2 channel: C3, C4) during use of an online BCI system utilising self paced movement. ApEn values were derived for 10 embedding dimensions over a set of coefficient values, ranging from 1 to 4 incremented by 0.1. Class separation was determined by calculating Bhattacharyya distances [6]. Based on class separation results the maximum coefficient value for testing was increased to 5 and the increment step to 0.2.

As the standard deviation of the entire signal is not an appropriate parameter for use in an online BCI system we substitute the use of the running standard deviation of the relevant EEG channel. The running standard deviation is calculated and updated sequentially based on prior data up to and including the sample point for which features are extracted. As subjects perform three runs the standard deviation value of each channel is recalculated for each run. As the standard deviation of a signal takes time to converge we tested K-means classification accuracy over time during the prior investigation to ensure results would not be detrimentally affected.

2.4 Classification

Class labels were derived through manual markup of the EMG channel. At the point of each class transition 32 samples prior and post were dropped. To ensure equal sample sizes for each class we obtain the number of samples for each class within a run, take the minimum N , and use $n=1 \dots N$ from each class. As we are interested in comparative class separability cross-validation was not applied. Classification results are representative of training classifiers using the entire dataset.

To examine the relationship between class separation, embedding dimension and r values classification results for single electrode-feature pairs were obtained by use of K-means clustering (Mathworks Matlab, default settings). To compare feature separability using multiple electrode-features a sequential forward floating search (SFFS) algorithm [7] was applied using a linear discriminant analysis (LDA) classifier. A maximum of six features were selected, corresponding to lightweight online use.

Classification results were also calculated for ten band power features (BP) corresponding to the delta (0.1-5 Hz), theta (5-8 Hz), alpha (8-12 Hz), sigma (12-15 Hz), beta (15-25 Hz) bands along with five gamma bands (25-35, 35-45, 45-55, 55-65, 65-75 Hz) and autoregressive reflection coefficients (K), orders 1 to 10. To determine if ApEn complexity information could be complementary to these features a SFFS was run using all three features.

3 Results

Maximum K-means classification results for each subject are presented in table 1 along with associated electrode site and related parameter values. Optimal r values all fall within the range 1.2 to 3.8 whilst m values cover all possible embedding dimensions. No significant differences were found between classification accuracies obtained for approximate entropy when compared to

band power and reflection coefficient features ($P > 0.05$, paired t-test, $df = 10$). From K-means classification results we obtained optimal r coefficients for each embedding dimension, figure 2 shows the relationship between these values for subjects demonstrating classification dominance in channel C4.

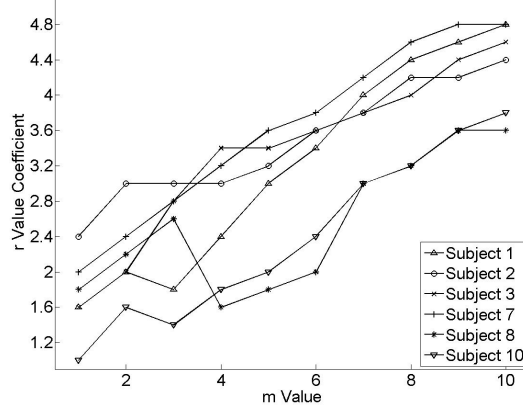


Figure 2: Optimal r Coefficients for m Values, C4 dominant Subjects.

Class separation as measured by multiple feature LDA classification is shown in table 2. Classification accuracy is shown for BP, K, ApEn, the maximum classification result using a single feature (Max) and classification accuracy achieved when all three features are made available to SFFS (Comb). The column ‘Comb Used’ outlines the combination of features used for each subject.

When comparing use of multiple features within feature groups we find significant differences in class separation across subjects for AE and BP ($P < 0.05$, paired t-test, $df = 10$), AE and K ($P < 0.05$, paired t-test, $df = 10$), no significant differences were found between BP and K features. Significant differences were found between accuracy rates obtained using combined features against BP ($P < 0.01$, paired t-test, $df = 10$), K ($P < 0.01$, paired t-test, $df = 10$) and AE ($P < 0.01$, paired t-test, $df = 10$).

| Subject | Band Power | | | Reflection Coefficients | | | Approximate Entropy | | | |
|----------------------------|--------------|------|------|-------------------------|------|-------|---------------------|------|---------|---------|
| | Acc % | Site | Band | Acc % | Site | Order | Acc % | Site | r Coeff | m Value |
| 1 | 81.42 | C4 | 5 | 80.40 | C4 | 3 | 82.56 | C4 | 1.8 | 3 |
| 2 | 66.79 | C4 | 4 | 61.18 | C4 | 1 | 62.37 | C4 | 2.4 | 1 |
| 3 | 58.60 | Cz | 10 | 60.72 | C2 | 2 | 64.37 | C4 | 2.8 | 3 |
| 4 | 63.61 | C4 | 7 | 71.99 | C1 | 5 | 72.25 | C2 | 2.2 | 7 |
| 5 | 65.94 | C3 | 10 | 70.67 | C3 | 5 | 66.11 | C3 | 1.2 | 9 |
| 6 | 56.58 | C2 | 1 | 64.30 | C3 | 5 | 62.18 | Cz | 3.8 | 9 |
| 7 | 61.21 | C4 | 5 | 60.41 | C1 | 3 | 62.02 | C4 | 2.8 | 3 |
| 8 | 65.55 | Cz | 10 | 71.01 | Cz | 6 | 66.76 | C4 | 3.6 | 10 |
| 9 | 58.54 | C1 | 9 | 63.83 | C4 | 6 | 59.55 | Cz | 1.6 | 3 |
| 10 | 57.08 | C4 | 3 | 61.08 | C4 | 3 | 63.03 | C4 | 2.0 | 5 |
| 11 | 59.14 | C4 | 5 | 60.49 | C2 | 4 | 57.33 | C1 | 1.2 | 8 |
| Acc % \bar{x} & σ | 63.13 & 7.09 | | | 66.01 & 6.59 | | | 65.32 & 6.93 | | | |

Table 1: K-Means Classification: Maximum for each Subject.

| Subject | Classification Accuracy % | | | | | Comb Used | | |
|-----------|---------------------------|-------|-------|-------|-------|-----------|------|------|
| | BP | K | ApEn | Max | Comb | BP | K | ApEn |
| 1 | 83.47 | 86.56 | 87.06 | 87.06 | 87.97 | 2 | 0 | 4 |
| 2 | 75.44 | 64.92 | 73.87 | 75.44 | 77.25 | 3 | 0 | 3 |
| 3 | 61.27 | 62.51 | 69.90 | 69.90 | 70.18 | 1 | 0 | 5 |
| 4 | 76.78 | 79.85 | 81.11 | 81.11 | 83.07 | 1 | 1 | 4 |
| 5 | 75.40 | 75.67 | 79.71 | 79.71 | 80.95 | 1 | 1 | 4 |
| 6 | 63.19 | 67.65 | 66.45 | 67.65 | 69.17 | 0 | 3 | 3 |
| 7 | 65.79 | 62.46 | 68.62 | 68.62 | 67.52 | 4 | 1 | 1 |
| 8 | 70.31 | 72.15 | 74.50 | 74.50 | 76.01 | 0 | 1 | 5 |
| 9 | 61.74 | 62.52 | 61.70 | 62.52 | 63.65 | 0 | 2 | 4 |
| 10 | 66.07 | 64.22 | 63.17 | 66.07 | 66.79 | 4 | 0 | 2 |
| 11 | 66.97 | 65.53 | 66.65 | 66.97 | 68.02 | 3 | 0 | 3 |
| \bar{x} | 69.68 | 69.46 | 72.07 | 72.69 | 73.69 | 1.73 | 0.82 | 3.45 |
| σ | 7.20 | 8.11 | 7.99 | 7.53 | 7.84 | | | |

Table 2: LDA Classification: Accuracy using Six Features Selected by SFFS.

4 Discussion

Maximum K-means classification results for each subject, obtained for a single feature, show no significant difference in class separability when using approximate entropy as compared to the more traditional features, band power and autoregressive reflection coefficients. Demonstrating that, given optimal parameters, signal complexity may be comparable to these linear features.

We use the classification rates from K-means to investigate the relationship between parameters varied in the approximate entropy calculation and class separability. As the selection of these parameters is a non-trivial problem, we calculated approximate entropy with a range of m and r coefficients in an exhaustive manner. The results of K-means classification, as shown in figure 2 show parameters m and r are proportionally related to one another; as the distance between embedding vectors increases with m it is necessary to increase r for optimal performance.

Significant differences were found between the degree of class separability, as measured by LDA classification accuracy, when comparing the use of multiple features of a single type. The difference in classification accuracy found between approximate entropy and the linear features are likely to be attributable to the difference in granularity in the search spaces. Parallel work using an increased feature space for band power has failed to find a significant difference in class separability between the features.

When comparing classification accuracy obtained for approximate entropy with the use of all available features, in our case augmenting the approximate entropy feature space with information from band power and reflection coefficients, we find significant differences in classification accuracy with relatively comparative feature spaces. This suggests that the linear features used provide further characterization of EEG time series which is complementary to approximate entropy. Across subjects the use of combined features demonstrates a bias towards the approximate entropy feature, followed by band power which is represented around twice as often as reflection coefficients. The bias towards approximate entropy is again, likely to be influenced by the increased search space for the feature. We did not utilize cross-validation in this study as we are interested in separability of features rather than the generalisation ability of a particular BCI system. An attempt to counter the overfitting problem, to some extent, was made through the use of linear LDA classifiers. We plan on using cross-validation in future tests and during online BCI system design.

The classification results achieved using approximate entropy demonstrate that, although computationally expensive, promising results may be obtained using parameters appropriate for online use. A short window was used to investigate the applicability of approximate entropy for online BCI use where the detection latency and the number of channels to be processed are important

factors. Increasing the window size should lead to a greater accuracy at the cost of latency of onset detection and the feasible number of channels used.

The main constraint to the use of approximate entropy for BCI use is in the estimation of the correct range of parameters. Based on the subjects tested it appears that optimal r coefficient and m values are subject dependant. In this study we used a fixed τ value of 1, this was employed to restrict the search space of the feature. Using methods such as the first local minimum of mutual information and first zero crossing of the autocorrelation function it may be possible to obtain an optimal τ value [8, 9]; we expect this to increase the characterization ability of the approximate entropy method.

5 Conclusion

Based on a limited number of subjects approximate entropy features using a short calculation window appear appropriate for detection of real finger movement. Classification results suggest that the complexity information derived may be complementary to the linear features tested. Further research is necessary to determine if methods used are applicable to imaginary movements.

References

- [1] S. J. Roberts, W. Penny, and I. Rezek. Temporal and spatial complexity measures for EEG-based brain-computer interfacing. *Med. Biol. Eng. Comput.*, 31(1):93–99, 1998.
- [2] I. A. Rezek and S. J. Roberts. Stochastic complexity measures for physiological signal analysis. *IEEE Transactions on Biomedical Engineering*, 45:1186–1191, 1998.
- [3] V. Srinivasan, C. Eswaran, and N. Sriraam. Approximate entropy-based epileptic EEG detection using artificial neural networks. *IEEE Transactions on Information Technology on Biomedicine*, 11:288–295, 2007.
- [4] S. M. Pincus. Approximate entropy as a measure of system complexity. *Prc. Natl. Acad. Sci.*, 88:2297–2307, 1991.
- [5] J. Bhattacharyya. Complexity analysis of spontaneous EEG. *Acta Neurobiol. Exp.*, 60:495–501, 2001.
- [6] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by probability distributions. *Bull. Calcutta Math. Soc.*, 35:99–109, 1943.
- [7] P. Pudil, J. Novovicova, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15:1119–1125, 1994.
- [8] J. C. Sprott. *Chaos and Time-Series Analysis*. Oxford University Press, 2003.
- [9] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, 1997.