Brain Computer Interface Feature Selection using Genetic Algorithm

Timothy B. Mahoney

School of Information Technology and Electrical Engineering

The University of Queensland, Qld., 4072, Australia

Abstract

Poor classification accuracy is among one of the challenges preventing Brain Computer Interface technology from integrating into our daily lives. This paper focuses on optimising EEG data feature selection to provide a classifier with appropriate inputs for high accuracy.

A previous study by Alomari et. al. [1] produced promising results of up to 97% classification accuracy using advanced feature extraction and an SVM classifier. In comparison, we use Genetic Algorithm incorporated into a wrapper feature selection method to determine a heuristically optimal subset of features from a large set of extracted features. Results… Conclude

1 Background

Brain Computer Interfacing (BCI) applications such as robotic prosthesis, or communication via thought have long been confined to the realm of science fiction, however advances in computing performance and machine learning have brought the technology much closer to reality. Organisations across industries including medicine, gaming, lifestyle, social media and information technology have begun to show interest in the field [2] [3] [4] [5]. However, the feasibility of BCI technology as a means to communicate or control devices is currently limited by factors including poor classification accuracy, poor signal strength and quality, and an overall lack of understanding of the brain. This paper focuses on the foremost, by attempting to select features from electroencephalography (EEG) data which provide the most accurate classification.

The method by which EEG data is sampled presents some challenges for classification. Nerual activity is detected as minute variations in electrical potential from numerous locations on the surface of the scalp, these signals are very susceptible to noise originating both from within the body, such as EEG, EMG and EOG, and outside, such as mains coupling [6]. In addition, the neural pathways within the motor cortex resulting in motor function are not entirely understood. Data collected using EEG is often separated into features over space, time and/or frequency [6]. The non-linear relationship between these extracted features and the resulting event is the basis for this classification problem.

Previous studies have shown varying degrees of success in classifying events in EEG data. Alomari *et al*. [1] found that focussing on feature extraction can yield a classification accuracy as high as 97% using a support vector machine (SVM). Alomari *et al*. used phenomena known as event related synchronisation and desynchronization (ERS and ERD). These signals are found by choosing specific EEG channels, filtering to specific frequency bands and sampling long temporal epochs about the motor task. ERS and ERD samples have been shown previously to relate to neural activity involving motor function and were shown by Alomari *et al*. to be very appropriate features for classifying such tasks. One limitation of the paper is the length of the epochs used; each sample ends over five seconds after the beginning of the motor event. This means that there would be a classification lag of over five seconds, which is not ideal for a responsive BCI. The paper also makes assumption regarding the sampling frequency of the BCI. Each sample used to train and test the SVM is partitioned as to line up consistently, in reality a BCI may not sample fast enough to perfectly capture the ERS/ERD.

Rather than selectively segmenting and processing EEG data to extract features as done by Alomari *et al*., this study focuses more on feature selection. EEG data is made up of many channels, which can be filtered into frequency bands and segmented further into time epochs, resulting in a large feature space which can be utilised for classification purposes. However, many of these features will associate strongly with eachother due to their spatial proximity, making them redundant, and many will not associate at all with the event being classified, making them irrelivent. Feature selection aims to reduce both redundance and irrelevance in the feature space [7]. Appropriately selecting a small subset of features from this space would benefit an online BCI system in two ways. First it would allow for fast classification due to the reduced computational expense, resulting in a more intuitive and responsive system. Secondly, there would be a reduction in hardware requirements as only a select few channels may be needed; a system requiring the user to wear only six electrodes is far more viable than one requiring 64. In addition, a reduced feature set generally improves the generalisation of a trained classifier and avoids the curse of dimensionality [7].

Selecting a combination of features from a set can be modelled as an optimisation problem. We wish to optimise the performance of a classifier, given some set of parameters. However, the search space is somewhat unconventional in that it spans *n* binary dimensions, where each dimension denotes either the exculsion (0) or inclusion (1) of one of the *n* possible features. In addition, the number of combinations grows exponentially with *n*, hence an exhaustive search is computationally infeasible. Genetic Algorithm (GA) is a heuristic search method which follows a competitive, survival-of-the-fittest type structure modelled after natural selection [8]. It takes a binary string input and aims to optimise this string given some performance metric, making it an appropriate optimisation technique to apply to feature selection in this case.

2 Aim

The solution presented in this study aims to produce comparable results to those given by Alomari *et al*. with the same dataset, using an approach focused more on feature selection rather than feature extraction. Specifically, the feature subset is selected using GA in a wrapper method with an SVM classifier. The goal of the feature selector is to significantly reduce the number of features in the set and minimise classification error. In addition, the method of sampling from the dataset reflects one which would be feasible should it be implemented as an online BCI system.

3 Dataset

The dataset being used was created in 2004 and was contributed to the PhysioNet Database [9] [10]. It consists of 109 subjects completing a series of tasks while connected to a 64 channel EEG system, sampling at 160Hz. The task being used for this study involves the subject imagioning clenching their left/right fist in response to visual queues. Each subject performs this test numerous times over a two-minute period, on three separate trials, resulting in approximately 650 minutes of data. This dataset was the same used by Alomari *et al*., allowing a meaningful direct comparison between results.

4 Feature Extraction

Despite the focus on feature selection, some small level of feature extraction is required to prepare the data for classification. All data processing is completed using MATLAB. Initially, the data is bandpass filtered between 1-80Hz to reduce baseline drift and high frequency interference, and a notch filter at 60Hz (USA frequency) to eliminate mains coupling. The data also undergos automated artefact removal (AAR), which is a three-step process. First the 64 channels are separated into 15 components using independent component analysis (ICA). Next a neural network trained to recognise EEG artefacts such as EMG, EOG and electrode shift attempts to classify these noise components. Lastly the signal is then reconstructed without the noisy components, thereby removing the artefacts. The signals are then artificially segmented into samples to replicate online data with a sampling frequency of 2Hz. The frequency is forced to be slow to allow for online AAR filtering and classification between samples, as if it were an actual live system. Samples which occur about the left/right hand clench event are labelled as L and R respectively. We perform a surface Laplacian on the EEG channels to improve spatial resolution [11], and we eliminate perimeter channels not ascosiated with motor function resulting in a 41-channel system. The features for each sample are split as such:

1. Each sample’s feature set includes the feature set of five previous samples as well as its own. This is to capture relevant low frequency, temporal neural patterns.
2. Each temporal sample consists of 41 channels as given by the standardised EEG montage [9].
3. Each channel is filtered into three standard EEG frequency ranges commonly ascosiated with motor imagery (alpha, mu and lower beta) [6].

The resulting feature set consists of 738 (5x41x3) features per sample, of which there are over 10129 feature combinations.

5 Feature Selection

Feature selection is implemented as a two-stage wrapper, whereby the metric used to evaluate any given feature set is based on the performance of a classifier trained and tested using that feature set. First, we apply GA to widely explore the global feature space for valid candidate solutions. The algorithm is applied as follows:

1. Eight candidate sets are sampled at random from the feature space to create the *candidate* pool.
2. Evaluate *candidate* pool scores.
3. Automatically include the two highest scoring candidates (elite children) in the new *candidate* pool
4. Create a *mating* pool which includes two copies of the highest scoring candidate and one of each of the next two highest scoreing candidates.
5. Select six sets of two different parents from the *mating* pool at random and perform uniform crossover to produce the remaining six candidates for the *candidate* pool.
6. Apply mutation with a probability of 0.01 to every candidate except for the best.
7. Repeat from step 2 until reaching a stopping criterion of 100 generations.

Candidate scores are calculated using an SVM with tuned MATLAB parameters: boxConstraint (soft margin error) and kernelScale (gamma), to reduce error. Each SVM is trained using the hold out method, whereby 20% of the data is held out for validation.

Apon completing the GA search, the best candidate is chosen for backward elimination to search for a local optimum with fewer features. This algorithm reavaluates the classification error having removed a feature from the list, features which improve or only slightly worsen the error are removed from the feature set permenently. The resulting feature set is the final solution. The performance of this set is evaluated using previously unseen data.

6 Results

7 Discussion

8 Conclusion

# **9 References**

|  |  |
| --- | --- |
| [1] | M. Alomari, A. Samaha and K. AlKamha, "Automated Classification of L/R Hand Movement," *International Journal of Advanced Computer Science and Applications,* vol. 4, no. 6, pp. 207-212, 2013. |
| [2] | "Brain-Computer Interfaces," JHU Neuroengineering & Biomedical Instrumentation Lab, 2008. [Online]. Available: http://neuroengineering.bme.jhu.edu/ecog.htm. [Accessed 22 March 2018]. |
| [3] | "Neuralink," Neuralink Corp, 2018. [Online]. Available: https://www.neuralink.com/. [Accessed 22 March 2018]. |
| [4] | J. Constine, "Facebook is building brain-computer interfaces for typing and skin-hearing," Tech Crunch, 2017. [Online]. Available: https://techcrunch.com/2017/04/19/facebook-brain-interface/. [Accessed October 2018]. |
| [5] | L. Ingham, "Brain computer interfaces: the video game controllers of the future," favour-tech, 27 February 2017. [Online]. Available: http://factor-tech.com/feature/brain-computer-interfaces-the-video-game-controllers-of-the-future/. [Accessed October 2018]. |
| [6] | S. Suily, Y. Li and Y. Zhang, EEG Signal Analysis and Classification: Techniques and Applications, Gewerbestrasse: Springer International Publishing, 2016. |
| [7] | Y. Saeys, I. Inza and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics,* vol. 23, no. 19, pp. 2507-2517, 2007. |
| [8] | K. Deb, "An introduction to genetic algorithms," *Sadhana,* vol. 24, no. 4-5, pp. 293-315, 1999. |
| [9] | A. L. G. L. H. J. I. P. M. R. M. J. M. G. P. C.-K. S. H. Goldberger AL, "EEG Motor Movement / Imagery Dataset, PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220," 13 June 2000. [Online]. Available: https://www.physionet.org/pn4/eegmmidb/. [Accessed March 2018]. |
| [10] | G. M. D. H. T. B. N. W. J. Schalk, "BCI2000: a general-purpose brain-computer interface (BCI) system.," *IEEE Trans Biomed Eng,* vol. 51, no. 6, pp. 1034-1043, 2004. |
| [11] | C. Carvalhaes and J. A. de Barros, "The surface Laplacian technique in EEG: Theory and methods," *International Journal of Psychophysiology,* vol. 97, no. 3, pp. 174-188, 2015. |

*Fig 2. Further and deeper thoughts (10 pt italic Times New Roman)*