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. This is then combined with backward elimination to further reduce the size of the resulting feature set. The presented method yields a classification accuracy of 74.8%. Performance variation between the two approaches is attributed to differences in the quality of features extracted as well as discrepancy in methods of sampling and signal processing.

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. Neural 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 [7] 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 each other due to their spatial proximity, making them redundant, and many will not associate at all with the event being classified, making them irrelevant. Feature selection aims to reduce both redundancy and irrelevance in the feature space [8]. 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 [8].

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 exclusion (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 [9]. 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 [10] [11]. 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 imagining clenching their left/right fist in response to visual cues. 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 undergoes 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 [12], and we eliminate perimeter channels not associated with motor function resulting in a 41-channel system. The features for each sample are split as such:

I. 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.
II. Each temporal sample consists of 41 channels as given by the standardised EEG montage [10].
III. Each channel is filtered into three standard EEG frequency ranges commonly associated 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 10^{129} 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 scoring 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.

Upon completing the GA search, the best candidate is chosen for backward elimination to search for a local optimum with fewer features. This algorithm re-evaluates 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 permanently. The resulting reduced feature set is the final solution. The performance of this set is evaluated using previously unseen data.

6 Results

The feature selection algorithm utilising both GA and backward elimination produced overall positive results, however not as successful as those found by Alomari *et al*. First, we can see the progression of GA as new generations produce candidates with less classification error as shown in Figure 1. The performance function used for GA is not, in this case, concerned with limiting the feature set, it simply searches for the minimal error. This is evident in the figure, the number of features in the set both increases and decreases over generations whereas the error continuously declines. The feature set remaining after the 1000 generations includes 303 of the 738 features and has a classification error of 20.91%.

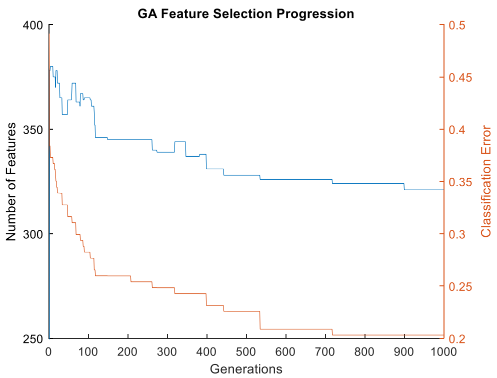

Figure 1 The reduction of classification error over 1000 Genetic Algorithm generations

Once a suitably fit candidate is found, the size of the feature set is reduced considerably with backward elimination. Figure 2 shows the progression of the error and number of features. Interestingly, 160 features were able to be removed and *improve* the classification error, with a minimum of 16.9%. We consider this a poor solution however as the number of included features is still too high. The reduction is continued until the error approaches 50%. Choosing the level of reduction requires balancing solutions with a potentially worse error with fewer features. We chose to eliminate 230 features, increasing the classification error by only 3.1% as further eliminations result in a much faster increase in error.

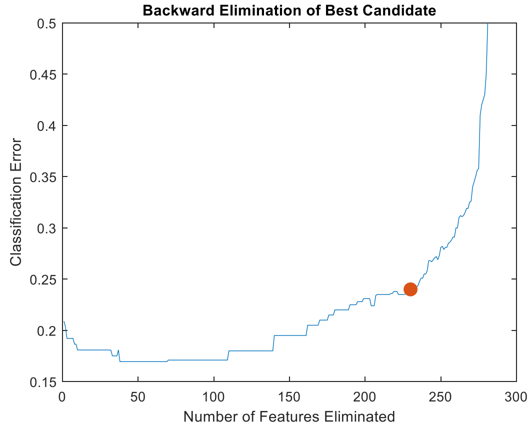

Figure 2 Classification error over iterations of greedy backward elimination. The red marker signifies the chosen number of features to eliminate for the solution.

The classification accuracy given by an SVM trained and tested with previously unseen data using this final feature set was 74.8%.

7 Discussion

The following section will briefly discuss the success of the reported results, before critically analysing the contrasts and limitations of both studies.

The results show that GA was successfully able to reduce the error of the classifier through iteration. Randomised initial candidates would often start with classification error close to 0.5, essentially guessing the outcome. This is quickly reduced through generations by carrying elite children, and performing crossover and mutation. Progress slows in the 0.2-0.3 range however, as better candidates appear to be scarce. Unfortunately, this solution includes a relatively large number of features. Backward eliminating reduced this feature set significantly, from 303 down to just 73 with an increase of error of only 0.31. This is a substantial result as it means there is far less computational load for the process, from filtering all the way to classification. Less features causes a positive snowball effect, whereby our initial estimate of 2Hz sampling for a BCI may be potentially sped up in an online situation due to a lower computational load, resulting in higher quality signal, which can then in turn improve classification performance further. Overall the method proposed appears to be a valid approach to applying feature selection in the context of EEG controlled BCI systems.

The results found in this and Alomari's *et al*. studies are summarised in Table 1.

*Table 1 Comparison between studies. *Best result*

	This Study Feature Selection with GA and Backward Elimination	Automated Classification of L/R Hand Movement [1]
Classification Accuracy	74.8%	97%*

There is an obvious performance difference between the two techniques. There are many plausible justifications for this difference which can be subdivided into two areas: the approach and the signal processing.

First, we will consider variations in the approach. It may be the case that the features extracted using the method proposed by Alomari *et al*. are indeed superior features for motor-imagery classification. The features used include the power, mean and energy of the extracted ERS/ERD phenomena. As stated previously, there is evidence to suggest that these signals are associated with motor-imagery tasks [7]. Utilising knowledge regarding what specific channels and frequencies to examine, and how to process and extract relevant information for these signals may have led to the high performance of Alomari's *et al*. solution. Compared to searching a wide space of generic feature combinations split across position, frequency and time delays, to locate a viable set as was the approach in this study, Alomari's *et al*. approach appears more focussed and informed. However, there are other discrepancies between the approaches which have likely also contributed to the division.

Relative lacking performance in the presented solution may also be attributed to the signal processing techniques used to prepare the data. Part of the aim of this study is to replicate online BCI sampling where the system is required to sample, filter, automatically detect and remove artefacts, extract features and classify events in multichannel EEG data fast enough to provide intuitive and responsive interfacing. To allow for all this processing, a sample rate of 2Hz is assumed as a 'poor scenario' simulation. A slow sampling frequency introduces problems for classification. Due to the slow frequency of the neural signal's EEG is capturing, an event occurring at t=0 will appear quite different from one lagging by 0.25s and captured at the next sample t=0.5s. As a result, the classifier is required to learn a broad selection of potential feature patterns to associate with left or right labels. In comparison, Alomari *et al*. segments samples specifically surrounding events to establish consistent temporal features associated with the ERS/ERD events they attempt to extract, thereby reducing any time delay. Alomari's *et al*. approach would be feasible given the BCI system is able to sample and process data at the rate the data was originally captured, i.e. 160Hz.

In addition to sampling discrepancies, the actual data selected for use from the PhysioNet dataset by each study is different. We chose to use a single recording of all 109 subjects for training and an additional recording of the first 20 subjects for testing. Alomari *et al*. use three recordings of six subjects in total, resulting in 18 two-minute recordings in comparison to the 129 used for this study. These sample sets are inconsistent and could influence the classification accuracy. In general, a larger number of samples provides a more generalised snapshot of the phenomena being classified. In this case however, generalising the classification problem could have made it harder for a classifier to learn the data. This is because there is evidence to show that features for neural event classification may be subject specific [13]. Hence it is possible that Alomari's *et al*. classifier has learnt to classify events with more accuracy due to both the smaller number of subjects and the repeated samples taken from those same subjects reinforcing its training. Concerning filtering the data, both studies use a similar initial bandpass filter and notch for mains, although Alomari *et al*. state using a 50Hz notch rather than 60Hz despite the data being collected in USA. both studies used AAR, however Alomari's *et al*. uses MATLABs EEGLAB software with an inbuilt AAR function, whereas this study trained a neural network to recognise and remove artefacts. Despite these different approaches, both methods are essentially employing ICA analysis therefore any variations in the data post-filtering is likely negligible.

8 Conclusion

Feature selection vs. feature extraction present two different yet valid approaches to dimensionality reduction. This study shows the power of utilising a heuristic optimisation method, Genetic Algorithm, as a feature selection method in the non-trivial problem of EEG motor-event classification. Whilst it has not outperformed methods presented in previous studies, it has still produced a viable alternative to advanced feature extraction in this context. Further areas of research could examine the performance of this method in a more subject specific framework, tailoring each feature set exclusively, or consider other optimisation methods.

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] A. Mohamed, T. Marwala and L. John, "Towards Improved EEG Interpretation for the Control of a Prosthetic/Orthotic Hand using a BCI: Single-trial right and left hand discrimination," in *Conference: Proceedings of the Twenty-Second Annual Symposium of the Pattern Recognition Association of South Africa*, Vanderbijlpark, 2011.

[8] Y. Saeys, I. Inza and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics,* vol. 23, no. 19, pp. 2507-2517, 2007.

[9] K. Deb, "An introduction to genetic algorithms," *Sadhana,* vol. 24, no. 4-5, pp. 293-315, 1999.

[10] 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].

[11] 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.

[12] 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.

[13] M. Schroder, M. Bogdan, T. Hinterberger and N. Birbaumer, "Automated EEG feature selection for brain computer interfaces," in *Conference Proceedings. 1st International IEEE EMBS Conference on Neural Engineering 2003*, Capri Island, Italy, 2003.