33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Particle Swarm Optimization-based Feature Selection for Cognitive State Detection

H. Alexer Firpi, *Member, IEEE*, and R. Jacob Vogelstein, *Member, IEEE*

Abstract—**This manuscript proposes a particle swarm-based feature extraction to monitors brain activity with the goal of identifying correlate cognitive states and intensity of a task. This in turn would allow us to develop a pattern recognition system that will classify such cognitive states and thus to redistribute the workload to other subjects. In this abstract, we present a recognition system that employ multiple features from different domains, a feature selection method using a Particle Swarm Optimization (PSO) search algorithm while the classification is provided using a k-nearest neighbor. Through this approach, we are able to achieve an averaged classification accuracy of 90.25% on held-out, cross-validated data among the eight subjects.**

I. Introduction

HUMAN attention and productivity can be affected by the complexity of a particular task due to factors such as fatigue, stress, etc. causing errors, miscalculations, downturn, and in some cases accidents. A simple solution to address the problem may be to monitor the physiological patterns of the subjects using, for example, a test or medical questionnaire but it will require the implementation of a schedule which makes such an approach unpractical. A more practical approach involve the use of a brain-computer interface-like system, which will monitor the electroencephalographic signals of the brain, among other sensors, and by means of pattern recognition algorithm to detect the cognitive state of the subject. Such an automatized system will allow the redistribution of workload of a particular task to other subjects based on the cognitive state of the subject.

Brain-computer interfaces have become more prominent during the last decade due to multiple factors, including better and less expensive sensors, lower power and faster processors, and the availability of sophisticated machine learning techniques. Their development is largely driven by biomedical applications, such as the need to provide a way for patients suffering from *cerebromedullospinal* disconnection ("locked-in syndrome") to communicate and interact with the environment. However, as BCI technology has matured, military and consumer applications have emerged, as well.

Most current BCI systems are enabled by machine learning algorithms that identify (based on a set of "training data") specific spatiotemporal features of neural activity that are reliably correlated with specific behavioral or cognitive outcomes. There are three main aspects of such a machine learning system: (1) signal conditioning, (2) feature extraction and (3) pattern recognition or classification. For most problems, the preprocessing and classification remains the same from subject to subject. However, the discriminative features are not necessarily conserved across subjects, depending on the specific neural phenomenon under study and the way in which it is encoded in the subjects' brains.

Feature extraction is often considered the most important stage in a pattern recognition system because an optimal configuration of the feature space—feature selection, transformation, and selection of parameters involved—can provide data in a representation that makes the classifier's task straightforward [1]. In the case of cognitive state detection system, we have available 19-channel EEG, VEOG, and HEOG (included for artifact removal) recordings from which we may extract many features (*i.e.*, mathematical metrics). We must choose which EEG channels to use (because some EEG channels are more informative than others), what kind of transformation to apply to the data (*e.g.*, linear, non-linear), whether to compute time- or frequency-based features, and many other factors. It is computationally infeasible to evaluate all combinations of features even once, and to ensure optimal performance, it is desirable to identify the best combination of features for each subject. We therefore employ a Particle Swarm Optimization (PSO) technique to efficiently search the parameter space and identify an effective set of features for each subject.

II. Methodology

A. Data Collection

The data was provided by the Air Force Research Lab and collected while study participants completed the Multi-Attribute Task Battery [2]. There were 8 participants in total, and each participant completed 3 trials on 5 separate days. In each trial, segments of task difficulty intended to produce low, medium and high workload were presented in a random order, with 'transition' time between workload segments. This transition time was 60 seconds between low/high and

Manuscript received April 15, 2011.
H. Alexer Firpi is with the Johns Hopkins University/Applied Physics Laboratory, Laurel, MD 20723-6009 USA is with the Johns Hopkins University/Applied Physics Laboratory, Laurel, MD 20723-6009 USA (phone: 240-228-9921; fax: 443-778-1093; e-mail: Alexer.Firpi@jhuapl.edu).
Jacob Vogelstein, is with the Johns Hopkins University/Applied Physics Laboratory, Laurel, MD 20723-6009 USA (phone: 443-778-2438; fax: 443-778-1093; e-mail: Jacob.Vogelstein@ jhuapl.edu).

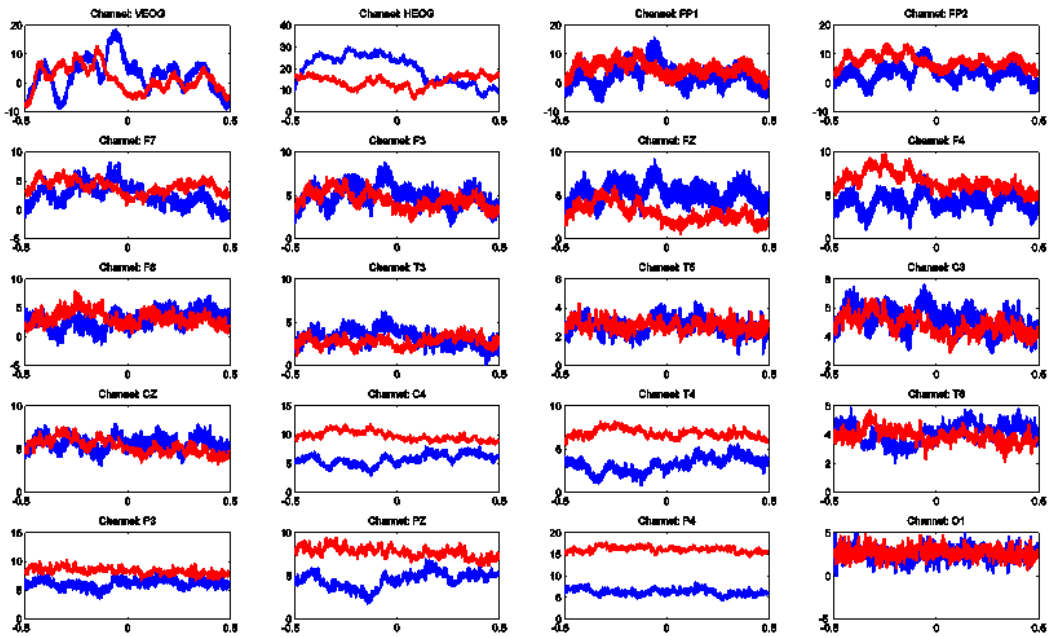

Fig. 1. Average signal for each channel from dataset 1 depicting the high workload (red) and low workload (blue). Channel O2 is not presented.

high/low segments and 30 seconds between low/medium, medium/low, medium/high, and high/medium segments. Each segment was 5 minutes in length. The five days of data collection for each participant were not sequential, but spread out over the course of one month. Starting with the first day ('Day 1'), the next data collection day was the following day ('Day 2'), one week later ('Day 3'), three weeks later ('Day 4') and four weeks (approximately one month) later ('Day 5').

All 21 of these data channels were collected using the MICROAMPS system from SAM Technologies, Inc. MICROAMPS has default high-pass and low-pass filters at 0.05 [Hz] and 100 [Hz], respectively, and a sampling rate of 256 [Hz]. Aside from these filters, no other processing has been done on this dataset. All values are in [μV]. The 19 EEG channels were referenced to a single (left) mastoid. VEOG was a bipolar channel with electrodes placed above and below the left eye. HEOG was also a bipolar channel with electrodes placed outside the outer canthus of each eye. All electrodes were tin cup electrodes (9 [mm]). Impedances for the EEG channels were all below [5 kΩ], and impedances for the VEOG and HEOG channels were all below [15 kΩ].

B. Features and Feature Extraction

After splitting the training set into two-second long non-overlapping windows, we extracted several features for each of the 21 channels. Fig. 1 depicts the average signals for low workload and high workload conditions for one of the subjects. Features were selected from different mathematical and physical domains such as time, frequency statistics, and information theory. The set of features are: *energy*, *nonlinear energy*, *curvelength*, *mean*, *standard deviation*, *skewness*, *kurtosis*, *Renyi entropy*, *Shannon entropy*, *short-term Lyapunov exponent*, *Katz fractal dimension*, and *summation, periodo, peak frequency, and mean frequency*.

Evaluating fifteen features in nineteen channels creates a feature space of 285 dimensions. Most classifiers, however, perform poorly with a large number of features because of the so-called curse of dimensionality [1]. To address this problem, we implemented a feature selection stage using PSO [3]. Table I shows the mathematical expressions for some selected features.

C. Feature Selection and Classification

To train the classifier and assess the performance of each feature set, we calculate the misclassification error (*i.e.* the fitness function) using leave-one-out validation with a *k*-nearest neighbor classifier. This procedure is performed on half of the data for each subject, with the other half held out for testing. The parameters for the PSO were: 100 generations, 100 particle population size, and length of the particle 10 (*i.e.*, maximum number of features). We also implemented a linear descending acceleration coefficient, *w*, that goes from 0.9 to 0.4. The acceleration coefficients were set to $c_1 = c_2 = \sqrt{2}$, X_{max} was set to [1, 285], and V_{max} to [0, 5]. The parameter *k* for the classifier was set to 5.

Classification was performed both on training/testing data (ground-truth is available) and on continuous data (ground-truth not available). For continuous classification, we used a two-second long window (512 pts.) and slid it by 1 second

TABLE I
MATHEMATICAL EQUATIONS FOR FEW SELECTED FEATURES

Feature	Equation
Nonlinear Energy	$f(x_n)=\sum_{i=1+(k-1)(L-D)}^{k(L-D)+D} x_i^2 - x_{i-1}\cdot x_{i+2}$
Curvelength	$f(x_n)=\sum_{i=1+(k-1)(L-D)}^{k(L-D)+D} \lvert x_i - x_{i-1}\rvert$
Lyapunov Exponent	$f(x_n)=\sum_{i=1+(k-1)(L-D)}^{k(L-D)+D} \frac{\lVert I_i^T - I_k^T\rVert}{\lVert I_k^T - I_k\rVert}$
Katz Fractal Dimension	$f(x_n)=\sum_{i=1+(k-1)(L-D)}^{k(L-D)+D} \frac{\log_{10}(p)}{\log_{10}\left(\frac{d}{L}\right)+\log_{10}(p)}$

(256 pts.). Each subject contains eighteen testing segments with duration of 2.5 minutes, half of them (8 records) containing mostly low workload conditions and the other half containing high workload condition

III. RESULTS

To assess the discriminative power of each feature alone, we calculated the area under the curve (AUC) of the receiver operating characteristic (ROC). Table II shows the top five features and channels for each subject (1 to 8), respectively (for these results we also included HEOG and VEOG). Few observations can be made from the table. Although, the AUCs values range from 0.75 to 0.96 (in practice considered satisfactory values), the best feature yielded different AUCs for different subjects. Although some common channels were found among some subjects such as F3, F4, and T4, the subjects seem to have particular signal patterns in different channels. As for the features, some common features could be observed such as peak frequency, Shannon entropy and some features from the statistical domain were selected, however, no global feature was noticed.

TABLE II
AUC VALUES FOR TOP-FIVE FEATURES-CHANNELS FOR EACH SUBJECT

FEATURE	CHANNEL	AUC
	Subject 1	
Lyapunov Exp.	F3	0.9199
Curvelength	O2	0.9032
Standard Dev.	F4	0.8799
Curvelength	O1	0.8799
Peak Frequency	F4	0.8785
	Subject 2	
Peak Frequency	F3	0.9213
Kurtosis	T5	0.7726
Kurtosis	FP1	0.7725
Periodo	CZ	0.7653
Peak Frequency	FP2	0.7552
	Subject 3	
Peak Frequency	T4	0.9624
Periodo	T4	0.9616
Peak Frequency	HEOG	0.9615
Standard Dev.	F4	0.9542
Curvelength	O2	0.9541
	Subject 4	
Peak Frequency	F3	0.8927
Peak Frequency	HEOG	0.8241
Periodo	T4	0.8237
Shannon Entropy	C4	0.8195
Shannon Entropy	F3	0.8179
	Subject 5	
Katz Fractal Dim.	FP1	0.9609
Katz Fractal Dim.	T5	0.9394
Shannon Entropy	F3	0.8816
Skewness	P3	0.8747
Katz Fractal Dim.	HEOG	0.8655
	Subject 6	
Peak Frequency	F3	0.8590
Mean Frequency	C4	0.8489
Mean Frequency	F3	0.8349
Peak Frequency	P3	0.8339
Summation	O1	0.8046
	Subject 7	
Shannon Entropy	C4	0.8804
Shannon Entropy	F7	0.8633
Kurtosis	T5	0.8622
Shannon Entropy	F3	0.8422
Shannon Entropy	FP2	0.8331
	Subject 8	
Skewness	HEOG	0.9309
Standard Dev.	T4	0.9286
Standard Dev.	F4	0.8991
Curvelength	O2	0.8976
Renyi Entropy	F7	0.8374

TABLE III
STATISTICS FOR EACH SUBJECT FOR TESTING DATA

Statistics	Testing
Subject 1	
Sensitivity (High)	90.99%
Specificity (Low)	90.18%
False Positive Rate	9.82%
False Negative Rate	9.01%
Positive Pred. Value	90.18%
Classification Accuracy	90.58 +/- 29.27%
Subject 2	
Sensitivity (High)	87.50%
Specificity (Low)	62.16%
False Positive Rate	37.84%
False Negative Rate	12.50%
Positive Pred. Value	70.00%
Classification Accuracy	74.89 +/- 43.46%
Subject 3	
Sensitivity (High)	99.10%
Specificity (Low)	99.11%
False Positive Rate	0.89%
False Negative Rate	0.90%
Positive Pred. Value	99.10%
Classification Accuracy	99.10 +/- 9.45%

Subject 4	
Sensitivity (High)	85.59%
Specificity (Low)	83.78%
False Positive Rate	16.22%
False Negative Rate	14.41%
Positive Pred. Value	84.07%
Classification Accuracy	84.68 +/- 36.09%
Subject 5	
Sensitivity (High)	97.32%
Specificity (Low)	91.89%
False Positive Rate	8.11%
False Negative Rate	2.68%
Positive Pred. Value	92.37%
Classification Accuracy	94.62 +/- 22.62%
Subject 6	
Sensitivity (High)	90.09%
Specificity (Low)	89.29%
False Positive Rate	10.71%
False Negative Rate	9.91%
Positive Pred. Value	89.29%
Classification Accuracy	89.69 +/- 30.48%
Subject 7	
Sensitivity (High)	82.88%
Specificity (Low)	88.29%
False Positive Rate	11.71%
False Negative Rate	17.12%
Positive Pred. Value	87.62%
Classification Accuracy	85.59 +/- 35.20%
Subject 8	
Sensitivity (High)	94.59%
Specificity (Low)	91.07%
False Positive Rate	8.93%
False Negative Rate	5.41%
Positive Pred. Value	91.30%
Classification Accuracy	92.83 +/- 25.87%

Table II presents the results achieved for each subject using the halt of the data separated for testing purposes. For this study, true positives were associated with detecting high workload. As mentioned before, classification was performed using LOO. It can be noticed that acceptable classification accuracies were achieved for all subjects with subject 2 having the lowest performance which does not come as a surprise given that that particular subject has just one feature with an AUC score over 0.90 (see Table II) and such feature was not selected by PSO during the training stage (among the ten selected, it selected kurtosis on channel T5).

We evaluated our features in continuous data for each of the eighteen testing segments for subject. Table IV shows the average classification accuracy achieved for each subject on each category (*i.e.*, low and high workload). We averaged the results for low and high workload conditions. It can be noticed that while the classifier is achieving an acceptable accuracy for low workload conditions, it performed poorly when confronted with high workload conditions. Although we achieved good classification during the training testing stage for the high workload conditions, we did not include transition periods and medium workload periods in the training stage. It may be beneficial to also include this data into the selection of features and training stage. Such data might be crucial in detecting the high workload activity. Also, we are using conventional metrics that are not necessarily optimized for these signals. Using algorithms to construct features directly from the raw data and combining multiple channels in a way that a multivariable and customized feature can be crafted.

TABLE IV
RESULTS FOR EACH SUBJECT ON CONTINUOUS CLASSIFICATION LOW AND HIGH WORKLOADS

Subject	Low	High
1	73.37 +/- 6.40%	23.21 +/- 7.95%
2	90.94 +/- 4.39%	13.76 +/- 4.50%
3	85.67 +/- 3.10%	14.38 +/- 5.63 %
4	90.91 +/- 5.02%	6.74 +/- 3.41%
5	88.47 +/- 5.13%	10.54 +/- 2.66%
6	79.83 +/- 7.05%	21.35 +/- 5.02%
7	79.88 +/- 5.78%	15.60 +/- 5.86%
8	84.52 +/- 3.99%	15.46+/-3.44%

IV. CONCLUSIONS

As brain-computer interface systems become more pervasive, there is a pressing need for better signal processing and feature extraction methods that can overcome the limitations presented by the current sensor technology. In this paper we present the importance of feature selection for a BCI system to detect cognitive states using EEG signals, among other sensors. Evaluation on continuous EEG data provided an idea of the difficulty of the problem. However, better results may be achieved if we implemented a more non-linear algorithm for feature extraction and construction of features. Algorithms such as PSO-trained neural networks and genetic programming can be implemented to construct nonlinear features—using the features presented in this paper—that provide more discriminative power, reduces the computational cost, and may be personalized to particular subject. To exploit the raw data, a genetic programming can be used to design features which may not have a physical meaning but are optimized for the patterns underlying in the data.

REFERENCES

[1] R. O. Duda, P. E. Hart, D. H. Stork, *Pattern Classification* (2nd ed.). Wiley Interscience, 2000. ISBN 0-471-05669-3.
[2] J. L Comstock ,and R. J. Arnegard, " The Multi-attribute Task Battery for human operator workload and strategic behavior research" *Technical Report 104174.* Hampton, VA: NASA Langley Research Center, 1992.
[3] J. Kennedy, R.C. Eberhart, *Swarm Intelligence*. Morgan Kaufmann, 2001. ISBN 1-55860-595-9.