

© 2007 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Classification of EEG Signals Using a Genetic-Based Machine Learning Classifier

B. T. Skinner, H. T. Nguyen, D. K. Liu

Abstract— This paper investigates the efficacy of the genetic-based learning classifier system XCS, for the classification of noisy, artefact-inclusive human electroencephalogram (EEG) signals represented using large condition strings (108bits). EEG signals from three participants were recorded while they performed four mental tasks designed to elicit hemispheric responses. Autoregressive (AR) models and Fast Fourier Transform (FFT) methods were used to form feature vectors with which mental tasks can be discriminated. XCS achieved a maximum classification accuracy of 99.3% and a best average of 88.9%. The relative classification performance of XCS was then compared against four non-evolutionary classifier systems originating from different learning techniques. The experimental results will be used as part of our larger research effort investigating the feasibility of using EEG signals as an interface to allow paralysed persons to control a powered wheelchair or other devices.

Index Terms— Learning classifier systems (LCSs), XCS, evolutionary computation, genetic-based machine learning (GBML), electroencephalogram.

I. INTRODUCTION

CLASSIFICATION of human electroencephalographic (EEG) signals is a difficult problem due to the high-dimensional and noisy nature of EEG data. Several studies have used a variety of frequency-band representations where features are extracted from the spectral density estimates of the EEG frequency domain signal [1]. Kiern [2] used spectral density estimates, calculated from the FFT to form asymmetry ratios for a specific frequency band. The asymmetry ratios were calculated for all right-to-left combinations of six EEG channels and for each EEG band: delta (0-3Hz), theta (4-7Hz), alpha (8-13Hz) and beta (14-20Hz).

Another representation technique, models the EEG signal as a random process using a zero-mean autoregressive (AR) process of order p [3], where the representation was composed primarily of scalar AR model coefficients. This paper compares the representation of EEG signals using a frequency-band method and three autoregressive methods, namely Burg, Covariance and Modified Covariance.

XCS is a recent and major development in learning classifier systems (LCSs) research that improves on several deficiencies of traditional LCSs, which were introduced by Holland [4]. In

the literature, XCS is commonly applied to data which is non-noisy, artefacts-free and represented using small condition lengths (<70bits). To the best of our knowledge, the classification of noisy and artefact-inclusive EEG signals, represented using condition lengths greater than 100bits is a unique application of XCS to a ‘real-world’ EEG classification problem. The promising experimental results demonstrate the competence of XCS in this difficult classification task.

Section II provides a brief description XCS, Section III describes the methodology for mental tasks and EEG data acquisition. Section IV describes EEG signal representation and feature vector encoding. Section VI presents the results and Section VII summarises conclusions.

II. DESCRIPTION OF XCS

XCS [5, 6] is an accuracy-based classifier system. It exploits a *reinforcement learning method* [7] coupled with the robust and global search capability of a steady-state *genetic algorithm* [4, 8] to produce a population of classifiers. The classifiers tend to form an accurate and maximally general knowledge representation of the target problem in single- and multi-step environments.

XCS evolves a population $[P]$ of classifiers, where each classifier consists of a rule and three main parameters estimating the quality of the rule. Each rule consists of a condition and an action pair $[condition \rightarrow action]$, encoded in binary. The condition part of the rule $C \in \{0,1,\#\}^L$ specifies the input states the classifier is capable of *matching*. The symbol $\#$ is called “don’t-care” and allows formation of generalisations in the condition part of a rule. The action part of the rule specifies the action $a \in A$ or *class* for which a payoff is predicted, that is, when the condition is satisfied.

In XCS, a steady-state genetic algorithm (GA) is responsible for improving the set of rules. The genetic search attempts to discover new classifiers which contribute to the existing knowledge and delete classifiers that do not offer improved contributions. Application of the GA takes place in the action set $[A]$ if and only if the average time since the last GA application in $[A]$ exceeds a constant threshold given by θ_{GA} . Firstly, the GA selects two parental classifiers from the current action set $[A]$ with probability proportional to F . The parental classifiers undergo uniform crossover with probability χ and single-bit *free* mutation with probability μ per allele. Finally, the two resulting offspring classifiers are inserted into the population to compete with their parents.

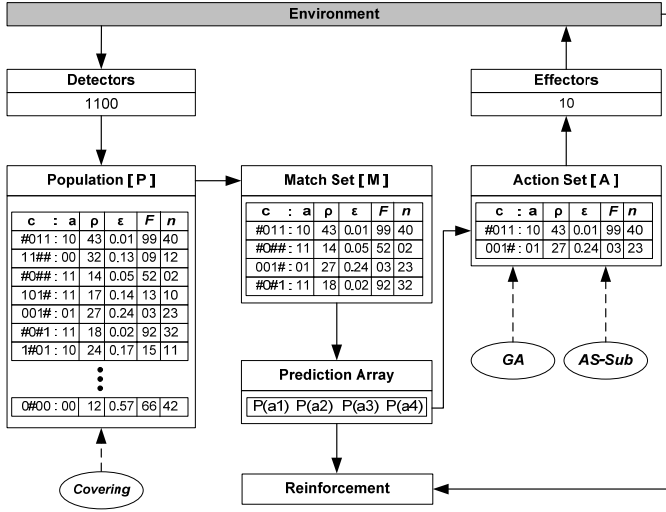


Figure 1: Block Diagram of the XCS Classifier System for single-step environment. It shows the classifier sets of [P], [M], [A] and the classifier condition/action pair ($c:a$). Also shown are the main parameters that estimate the quality of each classifier, namely prediction (p), prediction error (ϵ), fitness (F) and numerosity (n). The parameters (p , ϵ , F) are updated using the Widrow-Hoff Delta Rule.

III. MENTAL TASKS AND DATA ACQUISITION

A. Mental Tasks

Participants performed four mental tasks, chosen to invoke changes in brainwave activity across different hemispheres and across different electrodes. Each task was performed for a continuous period of 10 seconds and repeated for 10 independent trials during a single EEG recording session. Participants were asked to remain still, not make any overt movements or vocalise any task to minimise muscular artefacts. The four tasks studied in this paper were performed with eyes opened and include the following: *Mental Counting (MC)*: The counting of consecutive numbers, *Figure Rotation (FR)*: Visualise a complex object being rotated about one of its axis, *Mental Arithmetic (MA)*: Perform non-trivial and non-repeating multiplication problems, *Letter Composition (LC)*: Mentally compose a letter to a friend.

B. Data Acquisition and Pre-processing

The BiosemiTM Active-Two System (www.biosemi.com) was used in this study for recording EEG signals from each participant. Raw data was acquired using 32-channels at a sampling rate of 1024Hz with 24-bit digital resolution per channel, using the biopotential measurement system and active electrodes. The EEG electrode montage was adapted from the extended 10-20 electrode system[9] and all 32 channels were referenced to electrically-linked electrodes located on the ear lobes of each participant.

The recording of the EEG signals was conducted in a dedicated temperature controlled research laboratory. Artefacts such as eye and muscle movements, which are non-cerebral in origin, were *not* removed from the recorded EEG data. The number of channels was reduced to 6-channels. As in previous research work [2, 3], this study focused on EEG channels (O_2 , P_4 , C_4) for the right hemisphere (R) and (O_1 , P_3 ,

C_3) for the left hemisphere (L). In a preliminary analysis of a single participant, poor classification results were achieved using 2- and 4-channels asymmetrically.

The remaining 6 channels were down-sampled to 256Hz, removing frequencies not typically found in human EEG signals, which can range from 0.1Hz to 100Hz. Then, each 10 second recording was reduced to either 9.0 or 8.0 seconds (depending on the segment size, described next), by removing 0.5 or 1.0 seconds of data from the beginning and end of the EEG data. This provides EEG data that does not contain any transitional effects introduced prior or following each recording experiment as a participant initiates the mental task or concludes early.

Finally, the remaining time-domain EEG data was partitioned into individual *segments* or *blocks* of length 0.5, 1.0 or 2.0 seconds, which corresponds to 128, 256, and 512 samples for each segment respectively. So, for independent ten experiments this produced a total of 40, 90, or 180 time-domain instances based on 2.0, 1.0 and 0.5 second segments respectively, for each mental task.

The study was approved by the institutional research ethics committee and participants were only entered into the study after informed consent.

IV. EEG SIGNAL REPRESENTATION

A. Parametric Methods for EEG Signal representation

Parametric methods assume a description of an EEG signal can be devised from a time-series model of a random process. As such, parametric methods can *model* fixed segments of EEG data as the output of a linear filter of order p driven by a Gaussian white noise sequence with zero-mean Ch6,[10]. The output for such a filter is a p^{th} order *autoregressive* (AR) process or maximum entropy method (MEM), given by

$$x(n) = -\sum_{k=1}^p a(k)x(n-k) + u(n), \text{ where, } x(n) \text{ is the stationary}$$

time-series output sequence that models the fixed segment of EEG data, $a(k)$ are the AR coefficients and $u(n)$ is a Gaussian white noise input driving sequence.

The Burg, Covariance¹ (COV) and Modified Covariance (MCOV) are three AR parameter estimation algorithms used in this study. Each AR method operates on a fixed segment of time samples to recursively yield a p^{th} order AR model of parameter estimates, $a(k)$. It is the AR coefficients that are used to describe the EEG signal opposed to the estimate of the spectral density. Selection of the AR model order (p) for real signals represents a trade-off between increased resolution and decreased prediction variance of AR coefficients. Model order determination was performed for using a random selection of ten 2 second EEG segments from four mental tasks of each participant. The Akaike information criterion (AIC) [10] method was used to determine optimal AR model order. The AIC determines the model order by minimising the information theoretic function of the form:

$$AIC[p] = N \ln(\hat{\rho}_p) + 2p, \text{ where, } p \text{ is the model order, } N \text{ is the}$$

¹ This terminology is based on historical usage in speech processing and is not justified by the statistical definition of the term.

number of data samples given by the EEG segment size multiplied by the sampling frequency, and $\hat{\rho}_p$ is the variance estimate of the white noise input to the AR model for order p . The model order was incrementally increased from $p=1$ to $p=50$ and the AIC curves tended to approach a minimum within a range of orders $5 \leq p \leq 10$. The AR models used for all experiments were implemented with order $p=6$.

B. Frequency-Band Method for EEG Signal representation

Representation of EEG signals based on its frequency content have been commonly applied in EEG classification literature [2, 11]. These approaches are based on early observations that the human EEG spectrum is primarily characterised by four frequency bands: delta (0.1-3Hz), theta (4-7Hz), alpha (8-13Hz) and beta (14-20Hz). In this paper, an extended beta range (14-40Hz)² was also defined providing additional spectral information for the degree of brain wave activity over the cerebral cortex.

An FFT extracted the frequency components from each time-domain instance of the EEG signal and the spectral density relating the magnitude at each integer frequency was calculated for each separate EEG band. A 128-, 256-, 512-point FFT was applied to 0.5, 1.0 and 2.0 second segments respectively. The trapezoidal numerical integration technique was used to calculate the area under the spectral density curve (AUC) for each EEG frequency band ($\delta, \theta, \alpha, \beta$) of each EEG channel ($O_2, P_4, C_4, O_1, P_3, C_3$). In a similar approach used in [2], the *asymmetry ratio* for each EEG frequency band was also determined. The asymmetry ratio is given by the following iterative formula: $(L_i - R_j) / (R_j + L_i)$. Where R_j is the AUC value of the right hemisphere channel, $j=\{O_2, P_4, C_4\}$ and L_i is the AUC value for the left hemisphere channel, $i=\{O_1, P_3, C_3\}$. Consequently, the asymmetry ratio method produces a total combination of $(\# \text{ Channels} / 2)^2$ AUC ratios for each EEG frequency band. So, for the six EEG channels used in this study there are nine corresponding AUC ratios for each EEG band.

C. Feature Vector Encoding

The production of feature sets was the final stage of transforming raw EEG signals into a form that was amenable to the XCS *environment*. Separate training and testing sets, containing multiple feature vectors were created to train XCS and test classification accuracy, specificity and sensitivity.

In the first approach a single feature vector (*instance*) was encoded from the six AR coefficients that represent a single EEG segment (0.5, 1.0, or 2.0 seconds) for each EEG channel ($O_1, P_3, C_3, O_2, P_4, C_4$). The final feature vector for a single EEG segment was created from the concatenation of the six EEG channels containing the six AR coefficients represented as 3-bit binary strings. Thus, a single *instance* was represented by a 108-bit binary string and a class label (L, F, C, A) corresponding to the four mental tasks.

In the second approach, feature vectors were created separately for each EEG frequency band, containing the AUC of the spectral density coupled with the asymmetry ratio. Instances were randomly inserted into training and testing sets

for the seven different representations (Burg, COV, MCOV, $\delta, \theta, \alpha, \beta$). Training and test sets contain 40, 90 or 180 randomly selected instances corresponding to the 0.5, 1 or 2 second time-domain EEG segments respectively. This approach provided 50% of the data for the training phase and the remaining 50% for the testing phase.

V. CLASSIFICATION ENVIRONMENT

This section describes the comparative classifier schemes and XCS parameters used for each classification experiment.

A. Non-Evolutionary Classifier Schemes

The primary aim of this paper was to explore the efficacy of using the genetic-based XCS learning classifier system for the classification of human electroencephalogram signals using AR and FFT representations. A secondary aim was to compare the relative classification performance of XCS with other classifier systems. The non-evolutionary classifier systems selected for the comparison originate from different learning techniques, namely *Naïve Bayes* [12] which is based on Bayes rule of conditional probabilities, *SMO* [13] which implements the sequential minimal optimisation algorithm for training a support vector classifier, *IBk* [14] which uses the k -nearest-neighbour method and *PART* [15] which combines the divide-and-conquer strategy for decision tree learning with the separate-and-conquer scheme for rule learning.

All the non-evolutionary algorithms were obtained from the *Weka* machine learning workbench [16]. They were executed in classification mode using the *default* parameters provided by *Weka*, unless stated otherwise.

B. XCS Parameter Settings

All experiments were performed using the C-code implementation of XCS v1.2 [17]. The XCS parameter settings remained constant during all experiments, except population size ($N=2000, 4000, 8000$). The XCS parameters were set as follows [5, 6]: $\alpha=0.1, \beta=0.2, \delta=0.1, \epsilon_o=0.01, v=5, \theta_{GA}=25, \theta_{mna}=\#actions, \chi=0.9, \mu=0.004, \theta_{del}=100, \theta_{sub}=100, P_{\#}=0.86, doGAsubsumption=1, doASsubsumption=0, P_{explore}=1, reward=1000/0, toursize=0.4, crossover_type=uniform, initpop=0, fitness_reduction=1, force_different_in_tournament=0, elect_tolerance=0.001, do_mam=1, general_mutation=0, niche_mutation=0, p_l=10, \epsilon_l=0.0, f_l=0.01$.

A total of 30 independent experiments were performed for a maximum of 20000 explore iterations each. All remaining parameters were set to default values.

VI. RESULTS

Table I, lists the average classification accuracy and standard deviation for the classification of two mental tasks using XCS for the three participants represented by autoregressive method. The EEG segment size was 2.0seconds and the classifier population size was $N=8000$. From Table I, classification accuracy tends to vary between participants and between mental tasks for the same participant. We hypothesise this was a result of the representation of EEG signals in AR-space, which can produce similar class boundaries for

² Gamma (fast-beta) is usually defined for frequencies above 24Hz.

different mental tasks. This may restrict the ability of XCS to distinguish between different classes using its hyper-rectangle representation. This may be overcome by increasing AR-order or the encoding length of AR feature vectors. There was no statistically significant improvement or degradation between the three AR methods according to a paired *t*-test at a 99.5% confidence level.

TABLE I
CLASSIFICATION RESULTS FOR AR COEFFICIENT METHOD

		PARTICIPANT			
		1	2	3	AVG
FR	Burg	98.3±1.6	77.5±4.4	81.8±5.4	85.9±3.9
	v Cov	98.8±1.1	75.6±4.5	82.0±4.7	85.5±3.5
MC	MCov	99.3±0.7	76.6±4.8	79.2±3.0	85.0±2.7
	MCov	99.3±0.7	76.6±4.8	79.2±3.0	85.0±2.7
FR	Burg	88.8±2.8	66.0±4.4	89.2±4.1	81.3±3.8
	v Cov	90.9±3.2	68.7±5.0	87.3±3.9	82.3±4.0
MA	MCov	90.6±2.7	69.6±4.6	90.8±3.6	83.6±3.7
	MCov	90.6±2.7	69.6±4.6	90.8±3.6	83.6±3.7
FR	Burg	75.3±5.1	88.0±3.2	87.2±4.7	83.5±4.3
	v Cov	72.1±5.2	86.2±4.0	86.1±3.9	81.5±4.4
LC	MCov	74.0±5.2	88.8±3.0	87.8±4.5	83.5±4.3
	MCov	74.0±5.2	88.8±3.0	87.8±4.5	83.5±4.3
MA	Burg	82.1±3.4	85.9±3.1	96.8±1.6	88.3±2.7
	v Cov	82.0±2.6	88.3±3.8	96.3±4.1	88.8±3.5
LC	MCov	81.8±3.0	87.3±5.1	97.2±1.0	88.7±3.0
	MCov	81.8±3.0	87.3±5.1	97.2±1.0	88.7±3.0
MA	Burg	88.7±4.6	70.9±5.4	81.5±6.4	80.4±5.5
	v Cov	88.4±4.5	75.4±6.0	81.6±6.1	81.8±5.5
MC	MCov	88.8±3.6	73.0±5.7	80.5±4.8	80.8±4.7
	MCov	88.8±3.6	73.0±5.7	80.5±4.8	80.8±4.7

The classification accuracy for the frequency-band method (results not shown) was significantly lower for each band ($\delta, \theta, \alpha, \beta$) compared to the parametric methods in Table I.

TABLE II
CLASSIFIER SCHEME COMPARISON - AR METHOD

		NB	SMO	IBk	PART	XCS
FR	Burg	90.0 ○	88.3 ○	85.2	80.0 ●	85.9
	v Cov	88.4 ○	90.0 ○	86.0	83.7 ●	85.5
MC	MCov	88.8 ○	87.2 ○	83.0 ●	82.5 ●	85.0
	MCov	88.8 ○	87.2 ○	83.0 ●	82.5 ●	85.0
FR	Burg	83.3 ○	80.8	79.2 ●	80.0	81.3
	v Cov	80.0 ●	75.0 ●	76.7 ●	78.3 ●	82.3
MA	MCov	82.5	83.3	78.3 ●	78.3 ●	83.6
	MCov	82.5	83.3	78.3 ●	78.3 ●	83.6
FR	Burg	86.7 ○	87.5 ○	84.2	79.2 ●	83.5
	v Cov	85.0 ○	85.8 ○	84.2 ○	66.7 ●	81.5
LC	MCov	85.0	85.0	82.5	76.7 ●	83.5
	MCov	85.0	85.0	82.5	76.7 ●	83.5
MA	Burg	90.8 ○	90.0 ○	84.2 ●	85.8 ●	88.3
	v Cov	91.7 ○	93.3 ○	85.8 ●	84.2 ●	88.8
LC	MCov	91.7 ○	88.3	85.0 ●	86.7 ●	88.7
	MCov	91.7 ○	88.3	85.0 ●	86.7 ●	88.7
MA	Burg	85.0 ○	81.7	73.3 ●	80.8	80.4
	v Cov	85.8 ○	78.3 ●	72.5 ●	77.5 ●	81.8
MC	MCov	85.0 ○	75.8 ●	73.3 ●	80.0	80.8
	MCov	85.0 ○	75.8 ●	73.3 ●	80.0	80.8
Better - Worse		12-1	7-3	1-10	0-12	-

Table II, summarises the average classification accuracy and standard deviation rates of the different learning algorithms. The XCS results are taken from Table I, (AVG) column. Statistical differences are indicated according to a paired *t*-test at a 99.5% confidence level, with the following notation. A significant improvement of XCS with respect to another learning algorithm is denoted by ● and a significant degradation in classification performance by ○.

VII. CONCLUSION AND FUTURE WORK

This paper investigated the efficacy of the GBML classifier XCS for the classification of noisy, artefact-inclusive, multi-channel human electroencephalographic signals, that were

acquired from three participants as they performed four different mental tasks. The study also focused on two very different methods for representing EEG signals, namely three autoregressive parametric methods and a frequency-band method. The results indicate the Burg, Covariance and Modified Covariance AR methods provide higher classification accuracy compared to frequency-band methods.

This paper also compared the average classification accuracy of XCS to four different non-evolutionary algorithms, namely Naïve Bayes, SMO, IBk and PART. In general, XCS performed significantly better than PART and IBk ($k=3$), it was comparable to SMO, but performed worse than Naïve Bayes on the same datasets using the AR representation.

ACKNOWLEDGMENTS

The authors would like to thank Professor Ashley Craig and Dr Yvonne Tran for support in conducting the EEG trials. This research was supported partly by the ARC LIEF grant (LE0668541) and ARC Discovery grant (DP0666942).

REFERENCES

- [1] B. Blankertz, Curio G, and Müller, K, "Classifying single trial EEG: Towards brain computer interfacing," presented at Advances in Neural Inf. Proc. Systems (NIPS 01), British Columbia, 2002.
- [2] Z. A. Keirn and J. J. Aunon, "A new mode of communication between man and his surroundings," *Biomedical Engineering, IEEE Transactions on* vol. 37, pp. 1209-1214, 1990.
- [3] C. W. Anderson, E. A. Stolz, and S. Shamsunder, "Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks," *Biomedical Engineering, IEEE Transactions on* vol. 45, pp. 277-286, 1998.
- [4] J. H. Holland, *Adaptation in natural and artificial systems*, 1st MIT Press ed. ed. Cambridge, Mass: MIT Press, 1992.
- [5] S. Wilson, "Classifier Fitness Based on Accuracy," *Evolutionary Computation*, vol. 3, pp. 149-175, 1995.
- [6] M. Butz and S. Wilson, "An Algorithmic Description of XCS," presented at Advances in Learning Classifier Systems: 3rd Workshop on LCS, Paris, 2000.
- [7] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An introduction*, c1998 ed. Cambridge: MIT Press, 1998.
- [8] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, vol. 1, 16 ed: Addison Wesley Longman, 1997.
- [9] B. J. Fisch and R. Spehlmann, *Fisch and Spehlmann's EEG Primer: Basic Principles of Digital and Analog EEG*, vol. 1, 3rd edition (Dec, 1999) ed. New Orleans: Elsevier, 1999.
- [10] S. L. Marple, *Digital Spectral Analysis with Applications*, vol. 1, 1 ed. Englewood Cliffs, N.J: Prentice-Hall, 1987.
- [11] C. W. Anderson, S. V. Devulapalli, and S. Shamsunder, "EEG signal classification with different signal representations," presented at Neural Networks for Signal Processing [1995] V. Proceedings of the 1995 IEEE Workshop Cambridge, MA 1995.
- [12] G. H. John and P. Langley, "Estimating Continuous Distributions in Bayesian Classifiers," presented at Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, 1995.
- [13] J. C. Platt, *Advances in Kernel Methods - Support Vector Learning*: MIT Press, 1999.
- [14] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-Based Learning Algorithms" *Machine Learning*, vol. 6, pp. 37-66, 1991.
- [15] E. Frank and I. H. Witten, "Generating Accurate Rule Sets Without Global Optimization," presented at 15th International Conference on Machine Learning, Wisconsin, 1998.
- [16] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed, USA, M Kaufmann, 2005.
- [17] M. Butz and S. Wilson, "Documentation of XCS+TS C-Code 1.2," University of Illinois, Urbana-Champaign, Report 2003023, 2003.