Investigation into Machine Learning Algorithms as applied to Motor Cortex Signals for Classification of Movement Stages

Robert L. Hollingshead, *IEEE Member*, David Putrino, Soumya Ghosh, and Tele Tan

Abstract— **Neuroinformatics has recently emerged as a powerful field for the statistical analysis of neural data. This study uses machine learning techniques to analyze neural spiking activities within a population of neurons with the aim of finding spiking patterns associated with different stages of movement. Neural data was recorded during many experimental trials of a cat performing a skilled reach and withdrawal task. Using Weka and the LibSVM classifier, movement stages of the skilled task were identified with a high degree of certainty achieving an area-under-curve (AUC) of the Receiver Operating Characteristic of between 0.900 and 0.997 for the combined data set. Through feature selection, the identification of significant neurons has been made easier. Given this encouraging classification performance, the extension to automatic classification and updating of control models for use with neural prostheses will enable regular adjustments capable of compensating for neural changes.**

I. Introduction

A Brain-Controlled Interface (BCI) is a device that captures brain transmissions involved in a subject's intention to act, with the potential to restore communication and movement to those who are immobilized [1]. Neural signals can be used as an input to the BCI, but need to be decoded, to reveal the subject's intent before use

To determine which neural decoding methods have been applied to neural signal recordings, a review of literature was completed which indicated that statistical analysis methods have existed for more than 50 years. Many of these studies either analyzed single neuron spiking or interactions between pairs of neurons. However, advanced, population-based approaches to studying neural interactions have been developed more recently [2].

Patterns abound in nature and neural spiking is no exception. In the motor system, it has been reported that common, task-related, spatiotemporal spiking patterns only occur at chance levels between trials [3], however, other studies prior to this reported that specific spiking patterns do exist and may have a role in information processing in the brain [4]. A number of studies looked at patterns at the neuron level, either within individual neuron spiking or between pairs of neurons. However, spatiotemporal spiking patterns may also occur at the population level, making it possible to identify neurons, or groups of neurons, that encode the salient information to produce a voluntary movement.

Machine learning is a developing field with many diverse applications. In the field of BCI research, machine learning techniques have only recently been realized as a potentially valuable set of tools. Due to the widespread use of EEG signals in BCI studies, machine learning techniques have been predominantly used and developed for EEG applications. The use of machine learning with cortical neuronal recordings, and its application to BCI is an ongoing area of research.

This study seeks to investigate the value of machine learning algorithms for classification of movement stages based on neural signals from the motor cortex. The specific objectives for this study were:

- Determine a suitable classifier that maximises classification performance,
- Determine the data format that provides the best classification performance, and
- Determine the subset of neural units that maximises classification performance while minimising the number of features required.

II. Methodology

A. Previous Experimental Trials

The data for this project was sourced from experimental trials as described below ([5] and [6] have further detail).

Adult cats were trained to perform reaching and withdrawal movements using either forelimb to retrieve food pellets. The animals were trained in task performance for several months prior to the implantation of the microwires, until the ratio of successful to unsuccessful trials plateaued for at least one month. Once this level stabilized, microwires were implanted into the cortex in forelimb or hindlimb representations of MI (identified by intracortical microstimulation, ICMS). Neural activity was recorded from up to 24 microwires simultaneously on a computer using MC card and MC rack software with a filtering frequency band of 200 Hz–10 kHz.

Activity in each channel was digitized at 25 kHz, and raw digitized data recorded continuously for about 5–10 min. Recording was stopped after a large number of task trials and several minutes quiet sitting. The preamplifiers were then reattached to a different combination of recording sites and a new recording started.

Robert L Hollingshead was with Curtin University, GPO Box U1987, Perth, Western Australia (email: rlhollingshead@iinet.net.au).

David Putrino was with Western Australian Neuroscience Research Institute, Nedlands, Western Australia. He is now with Burke Medical Research Institute, White Plains, New York 10605 USA. (email: dap3002@med.cornell.edu)

Soumya Ghosh is with Western Australian Neuroscience Research Institute, Nedlands, Western Australia (email: soumya.ghosh@anri.uwa.edu.au)

Tele Tan is with Curtin University, GPO Box U1987, Perth, Western Australia (phone: +61 8 92661207; fax: +61 8 92662681; email: t.tan@curtin.edu.au).

An analogue trigger was concurrently recorded with the neural data and was used to separate periods of neural activity occurring during task performance from periods of continuous raw data acquisition. Analogue trigger signals were used to isolate a 3-second epoch of neural data: starting 1.5 second before and finishing 1.5 second after each trigger signal. Video analysis was used to identify five different stages of task performance within these 3-second epochs; "background", "premovement", "reaching", "withdraw", and "feed". After a set of task trials, spike activity was also recorded a 3-5 minute quiet period (control period): where no food was expected or given.

The background stage (ave. length 811 ms/range 700-900 ms) was the period before a food pellet was offered in the task. The premovement stage (307 ms/300-315 ms) occurred between the time the food was offered and the beginning of reach. The reach stage (592 ms/500-600 ms) began when the reaching paw moved up and ended when this paw was placed over the food pellet. The withdrawal stage (567 ms/500-600 ms) began by retraction of the reaching forelimb with the food pellet and ended when the paw was lifted off the pellet in preparation to feed. In the feeding stage (721 ms/700-800 ms), the reaching forepaw was taken off the food pellet and the head bent forward to pick up the pellet. Chewing and swallowing occurred outside the 3-second analysis period.

The recorded data was processed offline, firstly to isolate useful information from that which contained unsuitable data (e.g. failed trials, significant inclusion of noise). This stage was carried out in conjunction with the video recordings, which also enabled timing of each the stages to be determined and to distinguish trials using right and left forelimbs. Data collected during experimental trials was classed as one of three types - Control (no task being performed), LF (trials involving use of the left forelimb) or RF (trials involving the use of the right forelimb). LF trials were not used in this study due to insufficient trials being available. Through MC-Rack, time stamps of spike activity were determined by setting appropriate threshold levels to separate changes in spiking, due to task related actions, from background spiking activity.

B. Feature Extraction and Sampling

1) Feature Extraction

Previous studies ([5]) identified neural units associated with movement related activity (MRA) through ICMS. Using the electrodes associated with MRA, a feature set was built upon the neural units identified at each electrode by [5].

2) Dataset Configurations

Analysis of the spike timing stamp data proceeded using three data configurations; (a) spike time stamp data, (b) frequency domain data (Fast Fourier Transform) and, (c) a concatenation of the spike time stamp and frequency data. The decision to produce a combined data configuration was made to determine if the combination of time and frequency data would improve classification performance.

3) Sampling Routine

During exploratory runs with Weka Data Mining software (Weka) [7], the full size data files could not be processed in a reasonable time, or not at all due to computational limitations. To obviate these processing problems, a sampling routine, based on a normalized binning approach, was developed to reduce the number of attributes. The initial range of bin widths used to produce data sets were: Spike - 5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000 and 10000, and FFT - 1, 2, 5, 10, 20, 50, 100, 200 and 500. The combined data set used combinations of these bin widths for the respective data type.

C. Classifier Evaluation

1) Preliminary Classifier Evaluation

Weka's default configuration includes a large number of classifiers, a total of 43 classifiers were deemed suitable for use in this study. For a classifier to move to the next stage of evaluation, the classification accuracy had to be above 70%. Two data files were used to assess the classification performance of each classifier: Spike data with a bin width of 50, and FFT data with a bin width of 5. Each data file contained only two classes: control and RF. Classification was performed using the default parameters (as set by Weka) for all classifiers. The classifier's ability to predict the class of an instance, was determined by 10-fold cross-validation.

2) Stage 1 Classifier Evaluation

Prior to this stage of evaluation, the properties for each classifier were investigated to determine if classification performance could be improved. To maximize the performance of each classifier, the nominal (i.e. preset choices) properties of each classifier within Weka were explored. To optimize the nominal parameters for each classifier a spike data set with a bin width of 100 (classes: Control vs. RF) was used. The options of each property were cycled through followed by classification with the option producing the best classification performance being retained. In this regard the process of property selection used a greedy selection process, hence the optimal combination of property options may not have been achieved.

Evaluation of the classifiers in the subset was then completed using a range of data sets to identify whether any of the classifiers performed better than others and whether the bin width affected classification performance. All bin widths and all data sets were used during this evaluation stage. Each data set having the Control and RF classes.

Classifier performance was evaluated using the area under the Receiver Operating Characteristic curve (AUC). To ensure that classification performance was repeatable, two classification runs were completed for each classifier on each data set. The second classification run had the instances randomly shuffled using a predetermined seed. This was accomplished using a pre-processing filter available in Weka. 10-fold cross-validation was used to determine the classifier's ability to predict the class of an instance.

3) Stage 2 Classifier Evaluation

For stage two of classifier evaluation, the control class was discarded and the RF class was split into separate vectors for each of the movement stages. The length of each stage vector was based on stage timings from video analysis. The data sets contained two classes: the target class and the remaining stages combined into one class called 'Others'. Due to the variation in stage lengths, instance vectors of equal length were formed by first finding the length of the longest stage, then scaling all other stages to the longest stage.

Before classification, the class imbalance between the two classes needed was addressed. Weka's pre-processing filter, SpreadSubSample, achieved this by randomly selecting instances from the majority class until the number of instances in each class was equal. Classification then proceeded using two classification runs, 10-fold cross-validation with classifier performance evaluated using the AUC value. From this evaluation stage, the classifier and bin width were determined for further analyses of the three data sets (Spike, FFT and Combined).

D. Binary Classification

With the classifier and bin width determined, binary classification of each stage against the others was performed. Classification proceeded as for Stage 2 Classifier Evaluation.

E. Feature Selection

The use of population of neurons while improving decoding performance, also introduces computational issues due to the increased amount of data that is recorded and decoded. By identifying a smaller set of features, while maintaining classification performance, the model developed for classification would contain fewer model parameters and classification time would be reduced.

The sequential forward selection (SFS) method was used to find the minimum number of features. The minimum number of features was achieved when the classification performance (AUC value) first reaches 99% of the classification performance using all features. The SFS method was applied to each data set (with bin widths and classifier as determined during the Classifier Evaluation stage) with performance for each iteration determined by the AUC. As for previous evaluations of classifier performance, 10-fold cross validation was used.

F. Dataset Evaluation

To determine whether the combined data set performed better or worse than either the FFT or Spike data set McNemar's Test was applied. The pairs of data, for this study, are formed between the measures of classification performance (AUC) of two data sets (e.g. Spike vs. combined, FFT vs. Combined, etc).

Three null hypotheses were tested:

i. the FFT data set and combined data set have equivalent classification performance.
ii. the Spike data set and combined data set have equivalent classification performance, and
iii. the Spike data set and FFT data set have equivalent classification performance.

The null hypotheses can be accepted if the Chi-square statistic (χ^2) equals zero or if χ^2 is non-zero H_0 but less than a critical value. For a test with $\alpha = 0.05$, the critical value for the McNemar χ^2 is 3.84. If χ^2 is greater than the critical value then the alternate hypothesis (H_A) can be accepted.

To determine the better performing data set, a series of classifications were performed in which randomizations were introduced. Ten different randomizing seeds were applied to the majority class subset sample, while 50 different randomizing seeds were applied to the instance order randomizations, producing 500 different randomizations for each stage and data set type.

G. Multi-class Classification

To complete multi-class classification, a new data set was constructed containing all five stages based on the combined data set format. The length of the longest stage over all instances determined the length of all instance vectors. The bin widths and classifier determined during previous investigations were used for this analysis.

III. RESULTS AND DISCUSSION

A. Feature Extraction

Using the feature set found previously, the number of instances from each data file (and associated number of features recorded) are shown in Table I. A total of 40 units were identified for inclusion into the feature set.

B. Classifier Evaluation

1) Preliminary Classifier Evaluation

Of the 43 classifier algorithms, only 5 met the acceptance criteria after the preliminary evaluation phase: Bayesian Logistic Regression, Naive Bayes Updateable, LibSVM, SMO, and SPegasos.

2) Stage 1 Classifier Evaluation

The LibSVM, SMO and SPegasos classifiers were found to perform significantly better than the Bayesian Logistic Regression and Naive Bayes (Updateable) classifiers across all three data sets, the Bayesian Logistic Regression and Naive Bayes (Updateable) classifiers were discarded.

Classification performance of the Spike data set showed an improvement with an increase in bin width. The effect of increasing the bin width would have had an increasing smoothing effect on spiking variances. For the evaluations completed during this stage, it appears that spiking variation was reduced sufficiently such that stable classification performance for bin widths greater than 100 (4ms) was possible. Removal of several poor performing bin widths left the remaining bin widths as: 100, 200, 500, 1000, 2000, 5000 and 10000.

Classification performance for the FFT data set indicated that the three classifiers (LibSVM, SMO and SPegasos) performed consistently across all bin widths. A slight drop off in classification performance for bin widths below 100 was noticed, resulted in these bin widths being discarded leaving the remaining bin widths as: 100, 200 and 500.

TABLE I. NO. OF TRIALS AND FEATURES PER DATA FILE

Data File		1	2	3	4	5	6	7	8	9	10	11	12	13	Total
No. of Instances	Control	22	14	18	19	24	18	18	17	30	21	28	7	41	**277**
	RF	9	38	19	35	28	45	54	31	30	30	41	15	35	**410**
No. of Features		7	9	12	14	11	13	11	10	12	11	14	9	11	-

1) Stage 2 Classifier Evaluation

Prior to application of the SpreadSubSample filter the target class had 410 instances, while the 'Others' class had 1230 instances. After application each class contained 410 instances; a total of 820 instances for the data set. Stage two of classifier evaluation indicated that the LibSVM classifier consistently performed better than the SMO and SPegasos classifiers. For subsequent analyses the LibSVM classifier will be used.

Through examination of the LibSVM performance, a bin width of 5000, for the Spike data type, maximizes the classification performance across all stages of the Spike and combined data sets. Similarly, a bin width of 200, for the FFT data type, maximized classification performance across all stages of both the FFT and combined data sets.

Classifier Evaluation established that the LibSVM algorithm as the best performing classifier, and the bin widths producing the best results for each of the data sets is: Spike - 5000, FFT - 200, and combined - 5000/200.

2) Parameter Optimization

Following the guidelines for parameter optimization outlined by [8] a coarse grid search for the best (C,γ) combination followed by a fine grid search (over 0.25 intervals) was performed over the parameter values $-5 \leq \log_2(C,\gamma) \leq 10$. The initial range of parameter values was chosen based on preliminary broad range testing of parameter values. Classification performance was based on the average of two AUC values. The results of the coarse grid search, showed that varying the Cost parameter (C) had no impact on the classification performance, whereas varying γ had a significant impact on classification performance. The fine grid search identified a γ parameter of 0.25 to maximize classification performance across all five stages.

C. Binary Classification

Classification of all stages for all data sets had AUC values above 0.87. Classification performance, for the Premovement stage, with AUC values above 0.99 for all data sets, was consistently higher than that of the other stages.

TABLE II. BINARY CLASSIFICATION PERFORMANCE

	Spike	FFT	Combined
Background	0.970	0.976	0.976
Premovement	0.995	0.998	0.997
Reach	0.958	0.938	0.960
Withdrawal	0.894	0.912	0.924
Feeding	0.876	0.891	0.900

D. Feature Selection

The number of features identified, for each movement stage, was found to be less than half the full feature set for each stage. The number of features required for each stage are: Background - 18, Premovement - 5, Reach - 11, Withdrawal - 16 and Feeding - 16. From these features, a total of 29 unique units are able to provide classification results within 99% of the full feature set performance.

E. Dataset Evaluation

Using the McNemar's Test comparisons, all null hypotheses were rejected. Analyses of the contingency tables (not shown due to space limitations) for the McNemar's Test comparisons indicated that the Combined dataset statistically ($p<<0.05$) produced the better classification results in comparison to the FFT and Spike datasets.

F. Multi-class Classification

Classification of all stages had accuracies above 74% (i.e. better than chance). As for binary classification, multi-class classification performance, for the Premovement stage, with accuracies above 96%, was consistently higher than that of the other stages. Overall accuracy was 83.6%.

IV. CONCLUSION

This study has applied a data driven machine learning approach for the purpose of identifying movement related spiking patterns. Classification of each stage was completed with a high level of certainty in both binary and multi-class classification applications. Dimensionality reduction through feature selection showed that a small subset of features can produce classification results equivalent to the results achieved through use of all features. Given the positive classification results, extension to automatic classification and updating of control models for use with neural prostheses is conceivable.

TABLE III. MULTI-CLASS CLASSIFICATION PERFORMANCE

a	b	c	d	e	<-- classified as
84.6	0.0	1.9	2.7	10.7	**a = Background**
0.0	**96.1**	0.2	3.4	0.2	**b = Premovement**
3.7	0.0	**83.6**	7.8	4.9	**c = Reach**
2.9	1.0	5.6	**79.5**	11.0	**d = Withdrawal**
9.8	0.0	4.4	11.5	**74.4**	**e = Feeding**

REFERENCES

[1] Schwartz, A, Cui, X, Weber, D & Moran, D 2006, 'Brain-Controlled Interfaces: Movement Restoration with Neural Prosthetics', Neuron, vol 52, no 1, pp. 205-220.

[2] Georgopoulos, A, Schwartz, A & Kettner, R 1986, ' Neuronal Population Coding of Movement Direction', Science, vol 233, pp. 1416-1419.

[3] Baker, S & Lemon, R 2000, 'Precise Spatiotemporal Repeating Patterns in Monkey Primary and Supplementary Motor Areas Occur at Chance Levels', Journal of Neurophysiology, vol 84, pp. 1770-1780.

[4] Dayhoff, JE and Gerstein, GL 1983, ' Favored Patterns in nerve spike trains: II Applications', Journal of Neurophysiology, vol 49, pp. 1349–1363.

[5] Ghosh, S, Putrino, D, Burro, B & Ring, A 2009, 'Patterns of spatio-temporal correlations in the neural activity of the cat motor cortex during trained forelimb movements.', Somatosensory and Motor Research, vol 26, no 2, pp 31-49.

[6] Putrino, D, Chen, Z, Ghosh, S & Brown, E 2011, 'Motor Cortical Networks for Skilled Movements Have Dynamic Properties That Are Related to Accurate Reaching', Neural Plasticity, vol 2011, pp. 1-15

[7] Hall, M, Frank, E, Holmes, G , Pfahringer, B , Reutemann, P, Witten, IH 2009, ' The WEKA Data Mining Software: An Update', SIGKDD Explorations, vol 11, issue 1, pp 10-18.

[8] Hsu, C.-W.; Chang, C.-C. & Lin, C.-J. (2003), 'A Practical Guide to Support Vector Classification', Technical report, Department of Computer Science, National Taiwan University.