COLLABORATIVE LEARNING FOR INCREMENTAL CLASSIFICATION OF EMG SIGNALS

Karl Moore, Shamahil Ibunu, Stephen R. Alty, Clive Cheong Took

Royal Holloway University
Department of Electronic Engineering
Egham Hill, Egham, TW20 0EX

ABSTRACT

In this work, we address the problem of online learning for electromyogram (EMG) classification. The main challenge of EMG is its streaming nature, which implies that not all data are available at once. As such, offline and batch methods such as deep learning are generally not appropriate for such small data. To address this problem, we propose an incremental learning technique whereby the algorithm enhances its knowledge about the data progressively, as it receives new data. Moreover, we propose to employ multiple learners (instead of a single learner); each learning and focusing on a specific class. This federated learning strategy in incremental learning shows superior performance over both existing incremental learning and deep learning methods in our simulation studies on small data.

Index Terms— Electromyogram, collaborative learning, incremental learning

1. INTRODUCTION

Electromyogram (EMG) processing often poses a main challenge for machine learning to undertake *online* classification. EMG is inherently a streaming signal, which means not all data are available at once. This limited data can be a problem for learning methods such as deep learning, which cannot cope with small data. Yet, batch and offline learning such as deep learning is often the method of choice due to its high accuracy [1]. A less well-known learning mechanism more suitable for streaming data is incremental learning.

Incremental learning is a method whereby the algorithm learns and improves its knowledge progressively and incrementally, without forgetting previously acquired information. As such, incremental learning is particularly suitable for data streams, when not all data are available at hand. Our work addresses this kind of problem in the form of *online* EMG classification.

A more familiar kind of learning closer to the signal processing community is adaptive learning [2]. Like incremental learning, adaptive learning increments its learning. However, an adaptive learning algorithm such as the least mean square (LMS) algorithm, by design, forgets previously acquired information to tackle the problem of non-stationarity of the data. This is where adaptive learning differs from incremental learning. Our work seeks to leverage the advantage of both incremental learning (which does not lose previous knowledge) and adaptive learning (which is suitable for learning on a sample-by-sample basis).

For classification problems, adaptive learning has been morphed into collaborative learning to undertake such tasks [3]. Collaborative learning occurs when two adaptive learners collaborate to determine the correct class; each learner is assigned to learn about a specific class. For example, if the learner for class 1 can better predict the data than the learner for class 2, then the data is classified as class 1. Collaborative learning has addresssed various problems such as wind regime classification [4], epilepsy detection [5], detection of brain consciousness states [6], as well as voice detection [7].

Some of the advantages of collaborative learning include 1) No need to sanitise the data such as normalisation or whitening of the data; 2) No requirement to segment the data into regions of interest; 3) No need to explicitly label the data since each learner learns the signal that pertains to a particular class. On the other hand, its drawbacks are: 1) Binary classification only; 2) Inability to exploit information beyond two channels per class; and more importantly 3) the learning process forgets previously acquired information as they employed the LMS algorithm. To address these shortcomings of collaborative learning, we propose a novel algorithm called CALM (Collaborative Adaptive Learning for Multi-classes). Prior to its introduction, collaborative adaptive learning (CAL) is next described to provide the background of our work.

2. BACKGROUND

CAL is typically comprised of two adaptive learners $y_1[n]$ and $y_2[n]$, each learning to predict the signal(s) of a particular class. The output of CAL $y[n]$ can be expressed as a convex combination of the two learners:

$$y[n] = \lambda[n]y_1[n] + (1 - \lambda[n])y_2[n] \qquad (1)$$

The evolution of the mixing parameter $\lambda[n] \in [0,1]$ is then tracked to perform online binary classification. For instance, if $\lambda[n] \geq 0.5$, then it is Class 1, otherwise it is Class 2. Based on the steepest descent and on minimising the instantaneous quadratic error, it can be shown that $\lambda[n]$ can be optimised as follows.
For one-channel signal:

$$\lambda[n+1] = \lambda[n] + \mu_\lambda e[n](y_1[n] - y_2[n]) \tag{2}$$

For two-channel (complex-valued) signal:

$$\lambda[n+1] = \lambda[n] + \mu_\lambda \mathcal{R}\{\mathbf{e}[n](\mathbf{y}_1[n] - \mathbf{y}_2[n])^*\} \tag{3}$$

where μ_λ is the learning rate. $\mathcal{R}\{x\}$ and $(\cdot)^*$ denote respectively a function extracting the real-part of x and the complex conjugate operation. Therefore, it is clear that CAL is restricted to two-class classification, and cannot exploit information beyond two channels. Moreover, the consideration of the instantaneous error $e[n]$ in deriving $\lambda[n]$ in (2) and (3) means that CAL forgets previously acquired errors. To address these issues, our work considers a novel algorithm that 1) makes use of previous errors so that incremental learning can be implemented; 2) perform multi-class classification; and 3) can accommodate signals beyond two channels.

3. COLLABORATIVE ADAPTIVE LEARNING FOR MULTI-CLASSES (CALM)

Each of the C adaptive learners, $\mathbf{y}_i[n]$, collaborate to determine the output of CALM algorithm, $\mathbf{y}_\mathrm{o}[n]$, as follows

$$\mathbf{y}_\mathrm{o}[n] = \sum_{i=1}^{C} \lambda_i[n]\mathbf{y}_i[n] \tag{4}$$

Each ith learner learns to predict P channels of the signal corresponding to the ith class. In this work, each learner is considered as the multichannel recursive least square algorithm [8], however, any other incremental learning algorithm can also be considered. Likewise, the overall error $\mathbf{e}_\mathrm{o}[n]$ of CALM can be calculated as

$$\begin{aligned} \mathbf{e}_\mathrm{o}[n] &= \mathbf{d}_\mathrm{o}[n] - \mathbf{y}_\mathrm{o}[n] \\ &= \mathbf{d}_\mathrm{o}[n] - \sum_{i=1}^{C} \lambda_i[n]\mathbf{y}_i[n] \end{aligned} \tag{5}$$

where $\mathbf{d}_\mathrm{o}[n]$ denote the vector for the signals of the desired class. In contrast to (1), the mixing parameter $\lambda_i[n]$ is a multivariable, which is normalised as follows:

$$\lambda_i[n] = \frac{\lambda_i[n]}{\sum_{j=1}^{C} \lambda_j[n]} \tag{6}$$

Tracking the maximum value of the mixing parameters $\{\lambda_1[n], \lambda_2[n], \ldots, \lambda_C[n],\}$ enables us to estimate the correct class at time instant 'n'. The ith mixing parameter is constrained to be non-negative, i.e.

$$\lambda_i[n] = \lambda_i[n]\Big(\lambda_i[n] \geq 0\Big) \tag{7}$$

where $(\lambda_i[n] \geq 0)$ is a logical expression, which takes the value of ± 1. In order to implement incremental learning whereby previous knowledge is not lost, the proposed CALM algorithm is derived based on the recursive least square method. Hence, the gain vector $\mathbf{k}[n]$ can be computed as

$$\mathbf{k}[n] = \frac{\mathbf{P}[n-1]\mathbf{y}[n]}{1 + \mathbf{y}^T[n]\mathbf{P}[n-1]\mathbf{y}[n]} \tag{8}$$

where $\mathbf{P}[n] = E\{\mathbf{y}[n]\mathbf{y}^T[n]\}^{-1}$ is the inverse autocorrelation matrix (of size $PC \times PC$), and $\mathbf{y}[n]$ is the vector containing all the learners' outputs, i.e.

$$\mathbf{y}[n] = \begin{bmatrix} \mathbf{y}_1[n] \\ \mathbf{y}_2[n] \\ \vdots \\ \mathbf{y}_C[n] \end{bmatrix} = \begin{bmatrix} y_{1,1}[n] \\ \vdots \\ y_{1,P}[n] \\ y_{2,1}[n] \\ \vdots \\ y_{2,P}[n] \\ \vdots \\ y_{C,1}[n] \\ \vdots \\ y_{C,P}[n] \end{bmatrix} \tag{9}$$

where the first subscript denote the ith class, and the second subscript denote the pth channel of that class signal. Based on minimising the quadratic errors $\sum_{i=1}^{n}\sum_{p=1}^{P} e_p^2[i]$, the mixing parameter $\lambda_i[n]$ corresponding to the ith class in (6) can be updated as

$$\lambda_i[n] = \lambda_i[n-1] + \sum_{p=1}^{P} k_{i,p}[n] e_{p,\mathrm{o}}[n] \tag{10}$$

where $k_{i,p}[n]$ corresponds to the $\{i,p\}$th entry of the gain vector $\mathbf{k}[n]$ in (8), which has the same structure as in (9). The error corresponding to the pth channel is denoted as $e_{p,\mathrm{o}}[n]$ in (5). Finally, the inverse of the autocorrelation matrix $\boldsymbol{P}[n]$ can be updated as

$$\boldsymbol{P}[n] = \boldsymbol{P}[n-1] - \mathbf{k}[n]\mathbf{y}^T[n]\boldsymbol{P}[n-1] \tag{11}$$

This concludes the introduction of the proposed CALM. For reproducibility purposes, Algorithm 1 offers a practical approach to CALM such as the consideration of the exponential decaying window implemented by variable ϵ as well as the classification of the signal over a segment rather than on a sample-by-sample basis.

3.1. Proposed CALM Algorithm

Algorithm 1 Pseudo-code for CALM algorithm

For all $n \geq 1$

Output Estimate:

$$\mathbf{y}_o[n] = \sum_{i=1}^{C} \lambda_i[n]\mathbf{y}_i[n]$$

Error:

$$\mathbf{e}_o[n] = \mathbf{d}_o[n] - \mathbf{y}_o[n]$$

Gain update:

$$\mathbf{k}[n] = \frac{\mathbf{P}[n-1]\mathbf{y}[n]}{\epsilon + \mathbf{y}^T[n]\mathbf{P}[n-1]\mathbf{y}[n]}$$

Mixing Parameter for ith class update:

$$\lambda_i[n] = \lambda_i[n-1] + \sum_{p=1}^{P} k_{i,p}[n]e_{p,o}[n]$$

Mixing Parameter Non−negativity constraint:

$$\lambda_i[n] = \lambda_i[n]\Big(\lambda_i[n] \geq 0\Big)$$

Mixing Parameter Normalisation:

$$\lambda_i[n] = \frac{\lambda_i[n]}{\sum_{i=1}^{C} \lambda_i[n]}$$

Inverse autocorrelation matrix update:

$$\boldsymbol{P}[n] = \frac{1}{\epsilon}\Big(\boldsymbol{P}[n-1] - \mathbf{k}[n]\mathbf{y}^T[n]\boldsymbol{P}[n-1]\Big)$$

Class detected for nth sample:

$$C[n] = \arg\max_{\lambda_i}(\lambda_1[n], \lambda_2[n], \ldots, \lambda_C[n])$$

For every mod$(n, \alpha) = 0$

For all $0 < \ell \leq \alpha$

Majority vote classification of signal segment of length 'α' samples:

$$C[n-\ell] = \text{Mode}\{C[n-\alpha+1], \ldots, C[n-1], C[n]\}$$

4. SIMULATION ON SMALL DATA

To assess the performance of the proposed method CALM on small data, two EMG datasets were considered. The first dataset was a benchmark EMG dataset on hand and wrist contraction and relaxation [9], whereas the second dataset recorded the contraction and relaxation of individual fingers. 20 examples were randomly selected from both datasets, of which four examples (one example from each class) were used for testing. All examples had a length of 250 samples, which is also the value set for α in Algorithm 1.

For fair performance comparison, our method was assessed against 1) one incremental learning method namely the incremental Naive Bayes [10] and 2) a deep learning method designed specifically for EMG classification, i.e. the Long-Short term Memory (LSTM) [11]. The configuration for the Naive Bayes model comprised of a window-size equal to length 250 with a normal Gaussian distribution. For the LSTM model the following configuration was implemented 80 hidden nodes, fully connected layers set to equal the number of classes for each data set, initial learning rate of 0.001 and a mini-batch size of 32.

4.1. EMG classification of hand and wrist activity

The benchmark data is comprised of four classes [9] : wrist flexion (Class 1), wrist extension (Class 2), hand open (Class 3) and hand closed (Class 4). These gestures are shown in the top plot of Fig. 1 , and were collected from seven sites along the forearm and one on the bicep using Duo-trode Ag-AgCl electrodes (Myotronics, 6140). An Ag-AgCl Red-Dot electrode (3M, 2237) was placed on the wrist for common ground reference. It was inially sampled at 3KHz, but then downsampled to 250Hz to be consistent with the second set of EMG dataset. Table1 summarises the accuracy per class for each algorithm, whereas the middle plot Figure 1 shows how the 'majority vote' classification (see the last line of Algorithm 1) smoothes out the instantaneous classification of the mixing parameters $\lambda_i[n]$ in the bottom plot. Due to the small data, it is clear that a deep learning method like LSTM struggled to provide reasonable performance. The incremental learning Naive Bayes method did much better than LSTM, but the small data also made it difficult to provide a reasonable performance. On the other hand, the strategy of CALM whereby one learner learns specifically about one class, rather than learning all classes at once helps CALM to provide superior performance.

	Classes				
Model	1	2	3	4	Avg (%)
Incremental Bayes [10]	39.2	38.6	38.7	38.1	38.7
LSTM for EMG[11]	0	27.3	18.8	24.0	17.5
Proposed CALM	87.3	87.5	100	100	93.7

Table 1. Test accuracies on hand and wrist classification [9].

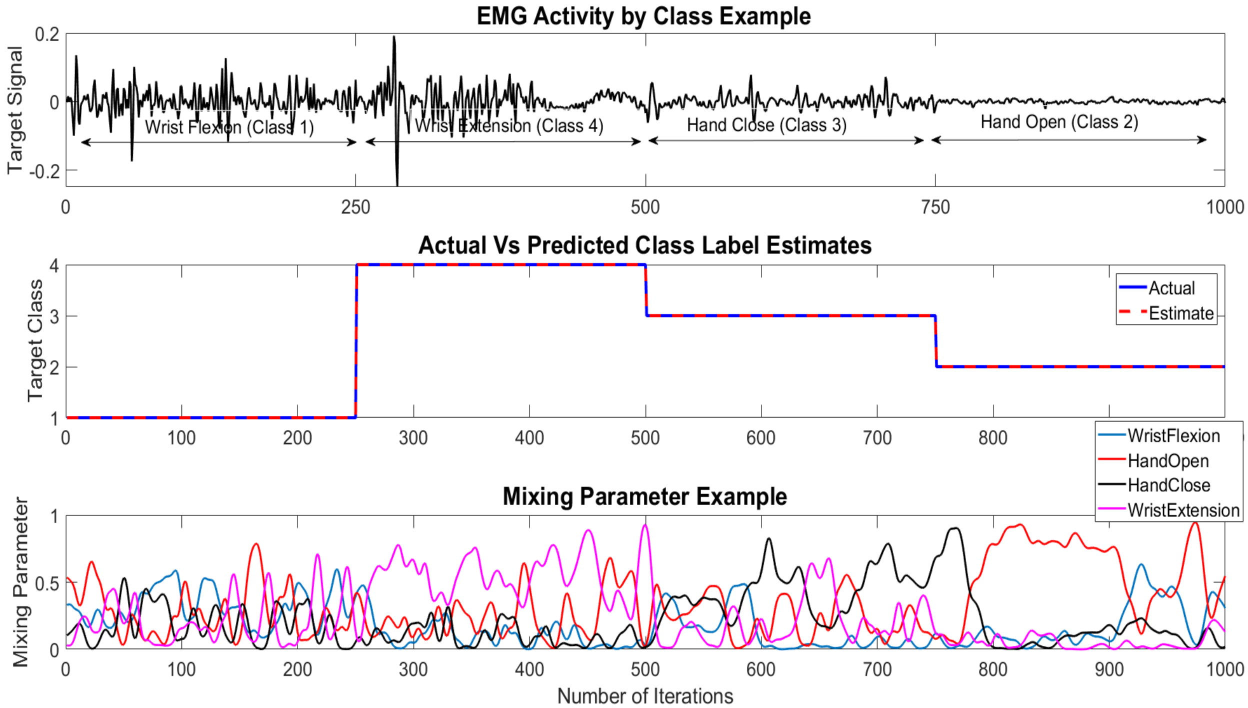

Fig. 1. Results on benchmark dataset [9]. Top to bottom: EMG activity per class; Predicted vs Actual Class Estimate; The evolution of the four mixing parameters for four classes.

4.2. EMG classification of finger activity

The EMG data recorded on finger activities. As such, the five classes were EMGs from the thumb (Class 1), index (Class 2), middle (Class 3), ring (Class 4) and little finger (Class 5). It was was collected using four surface EMG sensors, at a sampling rate of 250 Hz. Table 2 summarises the accuracies of each model.

Table 2 shows the same trend as in previous results: the worst classification results comes from the deep learning method, followed by the incremental Bayes learning method, and our method coped well with small data. However, its poorer performances compared to those in Table 2 can be explained by several factors: 1) the EMG acquired was noisier due to poorer quality of data acquisition equipment (hobbyist 'OpenBCI' compared to medical equipment in [9]); 2) there was an increase in the number of classes; 3) the locations of the EMG activities of the fingers are much closer to each other than those of the hand and wrist activities, causing EMG interference.

	Classes					
Model	1	2	3	4	5	Av (%)
Incremental Bayes [10]	57	52.5	50	51.5	57	53.6
LSTM for EMG[11]	29.8	38.9	96	9.5	8	36.4
Proposed CALM	100	99.8	0	100	37.5	74.9

Table 2. Test accuracies on finger classification.

5. CONCLUSION

An incremental learning framework whereby several learners collaborate to undertake online classification has been introduced. Each learner learns about one specific class rather than all classes at once. It has been shown that our proposed CALM algorithm outperformed both incremental model and deep learning model on small data. The problem of small data is particularly acute when not all data is available in streaming EMG, and is often overlooked in the literature.

6. REFERENCES

[1] Ulysse Côté-Allard, Cheikh Latyr Fall, Alexandre Drouin, Alexandre Campeau-Lecours, Clément Gosselin, Kyrre Glette, François Laviolette, and Benoit Gosselin, "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 4, pp. 760–771, 2019.

[2] Danilo Comminiello and Jose C. Principe, *Adaptive Learning Methods for Nonlinear System Modeling*, Butterworth-Heinemann, 2018.

[3] D. Mandic, P. Vayanos, C. Boukis, B. Jelfs, S.L. Goh, T. Gautama, and T. Rutkowski, "Collaborative adaptive learning using hybrid filters," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, 2007, vol. 3, pp. III–921–III–924.

[4] Beth Jelfs, Danilo P. Mandic, and Scott C. Douglas, "An adaptive approach for the identification of improper complex signals," *Signal Processing*, vol. 92, no. 2, pp. 335–344, 2012.

[5] B. Jelfs, S. Javidi, P. Vayanos, and D. Mandic, "Characterisation of signal modality: Exploiting signal nonlinearity in machine learning and signal processing," *Journal of Signal Processing Systems*, vol. 61, pp. 105–115, 2010.

[6] Ling Li, Yili Xia, Beth Jelfs, Jianting Cao, and Danilo P. Mandic, "Modelling of brain consciousness based on collaborative adaptive filters," *Neurocomputing*, vol. 76, no. 1, pp. 36–43, 2012.

[7] P. Vayanos, M. Chen, B. Jelfs, and D. P. Mandic, "Exploiting nonlinearity in adaptive signal processing," in *In Advances in Nonlinear Speech Processing*. Springer, 2007, pp. 57–77.

[8] Y. Huang and J. Benesty, *Audio Signal Processing for Next Generation Multimedia Communication Systems*, Kluwer Academic Publishers, 2004.

[9] Adrian D.C Chan and Geoffrey C. Green, "Myoelectric control development toolbox," *30th Conference of the Canadian Medical and Biological Engineering Society, Toronto, Canada*, 2007.

[10] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.

[11] Mathworks, "Classify arm motions using EMG signals and deep learning," `https://uk.mathworks.com/help/signal/ug/classify-arm-motions-using-emg-signals-and-dee` `html`, Accessed: 2024-04-26.