Hand Motion Identification of Grasp-and-Lift task from Electroencephalography Recordings using Recurrent Neural Networks

Jinwon An
Department of Industrial Engineering
Seoul National University
Seoul, Republic of Korea
jinwon@dm.snu.ac.kr

Sungzoon Cho
Department of Industrial Engineering
Seoul National University
Seoul, Republic of Korea
zoon@snu.ac.kr

***Abstract*—People who suffer from neuromuscular disorders and amputated limbs require prosthetic devices that are maneuvered through brain computer interfaces. Electroencephalography is a method to record the activity of the brain that is used for inputs for a brain computer interface. In this paper we propose a method for predicting hand motion phases in grasp-and-lift task from electroencephalography recordings using recurrent networks. Various architectures of recurrent neural networks are compared in terms of performance. For consistent prediction, moving average is applied.**

Keywords—Electroencephalography, Recurrent Neural Network, Grasp-and-Lift task

I. INTRODUCTION

There are numerous neuromuscular disorders that damage the channels through which the brain controls our body. Amyotrophic Lateral Sclerosis (ALS) or Lou Gehrig's disease, for example, degenerates motor neurons so that the brain cannot control voluntary muscle activities. There are also people suffering from amputated limbs. Both are deprived of voluntary control of body parts and should be assisted by prosthetic devices. A Brain Computer Interface (BCI) is required for such prosthetic devices to be maneuvered by one's mind.

Electroencephalography (EEG) is a primary method to record the brains activity that can be used as an input to the BCI.

It is a method for detecting the activity of the brain through electrical activities within the neurons of the brain. There has been extensive research on predicting epileptic seizure based on EEG recordings since such events incur aberrations in the EEG recordings. In addition, there also has been a study on the use of EEG as a communication channel between a prosthetic device and the brain [1].

Previous approaches used extensive manual feature engineering, extracting hand-picked features from EEG recordings and applying machine learning afterwards [2]. An obvious drawback is requirement of domain knowledge. However, recent developments in deep learning gave rise to an end-to-end framework, incorporating automatic feature extraction in the process [3]. Deep learning methods, especially recurrent neural networks (RNN), are dominating the field of speech recognition [4]. Since EEG recordings are also multi-channel time series data, it can also take advantage of the recent advancements of RNNs.

In this paper, we propose a method for identifying hand motions in a grasp-and-lift task from EEG recordings using RNN. To the best of authors' knowledge, this is the first application of a recurrent neural network.

II. METHODS

We explain about RNN and the architectures of RNN that are implemented in the experiment.

A. Recurrent Neural Network

A recurrent neural network (RNN) is a neural network that model sequential data via recursive or feedback connections among layer units. Formally, given an input sequence $\mathrm{x} = (x_1, \cdots, x_T)$, the RNN with one hidden layer computes the hidden state h_t and output y_t by

$$h_t = \sigma(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (1)$$

$$y_t = W_{hy}h_t + b_y \quad (2)$$

where W are weight matrices and b are bias vectors. σ is a non-linear function. The connection between h_{t-1} and h_t is where recursion or feedback exists. Hidden states of multiple hidden layers is computed by iterating (1) and (2).

Training of RNNs are done by backpropagation through time and other variants, which is unfolding the RNN in time to form a feed forward neural network to apply backpropagation. The weakness of backpropagation through time is that when the network is unfolded through a considerable amount of time, vanishing gradient problems occur [5].

B. Long Short Term Memory

Long short term memory (LSTM) [6] is an RNN architecture that is composed of memory blocks which uses gating units with a self-connected memory cell. LSTM solves the vanishing gradient problem that traditional RNNs suffer from by the self-connected memory cell that makes a "constant error carousel" during backpropagation. With the advent of deep learning algorithm and graphics processing units (GPU) computing, LSTMs has been widely used to model sequences in fields of Natural language processing, speech recognition and time series prediction. Gated Recurrent Units (GRU) [7] are similar to LSTM in that it has gating units but differs from LSTM in that it does not have a memory cell. Other variants of LSTM and GRU are introduced in [8].

III. Experimental setup and Results

We explain the dataset used for the model and the preprocessing done to the dataset. Problem formulation and network structure are introduced afterwards. Performance evaluation by classification and confusion matrices follows.

A. Dataset

The dataset [9] consists of EEG recordings of grasp-and-lift tasks from 12 participants with information on the time of events that occur in each trial (e.g., object lift-off). The grasp-and-lift tasks consists of reaching for a object, grasping it using their index finger and thumb, lifting it up to a prescribed height, holding it still for a few seconds, and replacing and releasing the object. The beginning of the reach and the beginning of the retraction is cued by turning on and turning off the LED light respectively. We used 294 trials of the 328 trials for each participant that has information on the time of events in each trial.

For each trial, we split it into 6 phases: idle, reach, load and reach, load and hold, load and retract, retract. Table 1 describes each phase. Each trial starts with the idle phase and goes sequentially through the phases of reach, load and reach, load and hold, load and retract, retract and in the end returns to the idle phase.

EEG recordings of each trial are from 32 channels in a standard configuration. The sampling rate is 500Hz. This means 500 time frame of 0.002 seconds are recorded each second.

B. Preprocessing

The cerebral brain signal observed by EEG that does not fall in the range of 1-30Hz are generally assumed to be artifactual, which means that those recordings are not originated from the cerebral brain, under standard clinical recording techniques [10]. Thus to eliminate artifacts we use a band-pass Butterworth filter of 1-30Hz to preprocess the raw EEG data.

TABLE I. Description of phases

Phase	Description
Idle	No action
Reach	Start hand movement and reach for the object
Load and Reach	grasp the object and lift it to a prescribed height
Load and Hold	Hold the object still in the prescribed height
Load and Retract	Retract and replace the object to its original position
Retract	Release the object and return to the original hand position

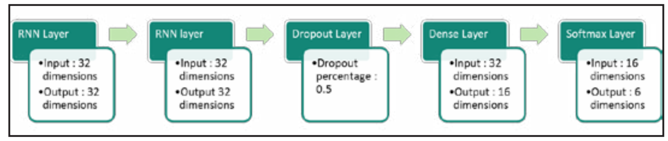

Fig. 1. RNN structure

Typically, Fast Fourier transform is applied to EEG data to analyze it in the frequency domain where brain waves of different frequency bands are separated (delta, theta, alpha, beta and gamma). However, we will use RNNs to automatically extract features from the original EEG data that is in the time domain.

C. Data preparation and problem formulation

Each trial was sliced into an overlapping time window of 256 time frames or 0.512 seconds. The phase of each time window was defined as the phase that makes up the biggest portion of it. Since phase transitions are in a sequential manner, the phase of a given time window is the phase that covers more than 50% of the time window. The time gap between time windows is 1 time frame.

For consistent and smooth predictions of phases during transitions, we used moving averages of predictions instead of using just the prediction for one point. We averaged past predictions, the number of which is decided by validation data, by using simple moving average. We refer to prediction with and without moving average as smoothed prediction and non-smoothed prediction respectively.

D. Recurrent neural network structure

The RNN structure is shown in Fig. 1. The RNN layer is where various RNN architectures such as LSTM, GRU, and its variants are plugged in and compared. The variant architectures we used are from [8], which are named MUT1, MUT2, and MUT3. The softmax layer in the end of the RNN classifies each input to one of six phases.

E. Performance evaluation

A model was trained for each participant, using 200 trials for training data, 44 trials for validation data, and 54 trials for test data.

Table II shows the overall multiclass classification accuracy averaged over participants. The dropout layer improves performance by 4 percentage point in average. By a narrow margin, MUT3 shows the best performance.

Table III is the confusion matrix of non-smoothed predictions that are normalized by class support size. Table V shows the normalized confusion matrix with simple moving average predictions. Columns are predicted phases and rows are true phases. It can be seen that after applying moving average into predictions, erroneous predictions only occur in phases that are adjacent to the true phase.

In both cases, predictions of retracting phases (retract phase and load and retract phase) are less accurate than predictions of reaching phases (reach phase and load and reach phase). This means that there are different mechanisms of the EEG signaling between reach motion and retract motion and that the former is more suited for RNN modelling.

TABLE II. CLASSIFICATION ACCURACY

	RNN architectures				
	LSTM	*GRU*	*MUT1*	*MUT2*	*MUT3*
Before Dropout	82.52%	84.78%	84.26%	85.87%	84.73%
After Dropout	87.98%	88.60%	86.54%	88.21%	88.82%

TABLE III. CONFUSION MATRIX OF NON-SMOOTHED PREDICTION

	Idle	*Reach*	*Load & reach*	*Load & hold*	*Load & retract*	*Retract*
Idle	**91.3%**	0.9%	0.2%	3.2%	1.7%	2.8%
Reach	6.3%	**80.6%**	13.1%	0.0%	0.0%	0.0%
Load & reach	1.8%	3.0%	**90.1%**	4.7%	0.1%	0.3%
Load & hold	1.4%	0.0%	11.8%	**83.2%**	3.5%	0.1%
Load & retract	3.3%	0.0%	2.2%	38.4%	**53.6%**	2.6%
Retract	11.3%	0.0%	0.0%	5.8%	22.9%	**59.9%**

TABLE IV. CONFUSION MATRIX OF SMOOTHED PREDICTION

	Idle	*Reach*	*Load & reach*	*Load & hold*	*Load & retract*	*Retract*
Idle	**96.47%**	1.27%	0.00%	0.74%	0.00%	1.53%
Reach	3.60%	**91.13%**	5.27%	0.00%	0.00%	0.00%
Load & reach	0.00%	9.14%	**82.71%**	8.15%	0.00%	0.00%
Load & hold	0.00%	0.00%	6.38%	**91.98%**	1.64%	0.00%
Load & retract	0.00%	0.00%	0.00%	28.86%	**65.96%**	5.18%
Retract	18.69%	0.00%	0.00%	0.00%	19.36%	**61.95%**

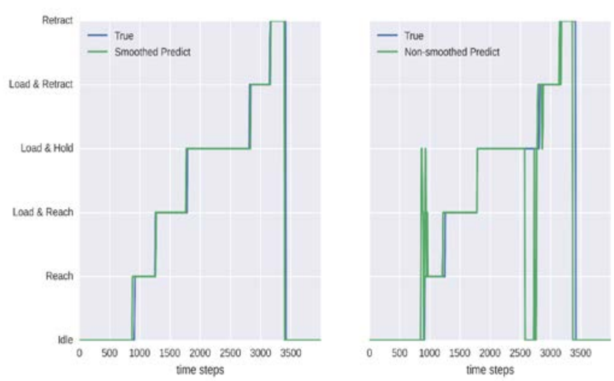

Fig. 2. Prediction comparison of between smoothed and non-smoothed

Fig. 2 shows an example of a prediction for a single trial. Abrupt and inconsistent phase changes in predictions that are present in non-smoothed prediction has disappeared in smoothed predictions.

IV. CONCLUSION

In this paper, we proposed a method for identifying hand motions in grasp-and-lift task from EEG recordings using RNN. Various RNN architectures were compared in performance, and MUT3 showed the best performance. Dropout improved performance of RNNs by an average of 4 percentage point. Smoothing the predictions with moving average helped making consistent predictions, eliminating abrupt and incongruous prediction errors.

ACKNOWLEDGMENT

This work was supported by the BK21 Plus Program(Center for Sustainable and Innovative Industrial Systems, Dept. of Industrial Engineering, Seoul National University) funded by the Ministry of Education, Korea (No. 21A20130012638), the National Research Foundation(NRF) grant funded by the Korea government(MSIP) (No. 2011-0030814), and the Institute for Industrial Systems Innovation of SNU.

REFERENCES

[1] T. Vaughan, J. Wolpaw and E. Donchin, 'EEG-based communication: prospects and problems', *IEEE Transactions on Rehabilitation Engineering*, vol. 4, no. 4, pp. 425-430, 1996.

[2] W. Ting, Y. Guo-zheng, Y. Bang-hua and S. Hong, 'EEG feature extraction based on wavelet packet decomposition for brain computer interface', *Measurement*, vol. 41, no. 6, pp. 618-625, 2008. What is the matter, tjhis

[3] Y. Bengio, A. Courville and P. Vincent, 'Representation Learning: A Review and New Perspectives', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798-1828, 2013.

[4] A. Graves, A. Mohamed and G. Hinton, 'Speech recognition with deep recurrent neural networks', in *Acoustics, Speech and Signal Processing (ICASSP)*, 2013 IEEE International Conference on, pp. 6645-6649, 2013

[5] S. Hochreiter, 'The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions', *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 06, no. 02, pp. 107-116, 1998.

[6] S. Hochreiter, and J. Schmidhuber, 'Long short-term memory', *Neural computation*, vol. *9*, no. 8, pp. 1735-1780, 1997.

[7] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. "Learning phrase representations using rnn encoder-decoder for statistical machine translation."*arXiv preprint arXiv:1406.1078*, 2014.

[8] R. Jozefowicz, W. Zaremba, and I. Sutskever. "An Empirical Exploration of Recurrent Network Architectures." In *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, pp. 2342-2350. 2015.

[9] Luciw, M. D., Jarocka, E. & Edin, B. B. FigShare http://dx.doi.org/10.6084/m9.figshare.988376 (2014).

[10] Z. Hippe, J. Kulikowski and T. Mroczek, *Human - Computer Systems Interaction: Backgrounds and Applications 2*, New York: Springer, p. 296, 2011.