Machine Learning of Spatiotemporal Bursting Behavior in Developing Neural Networks

Jewel YunHsuan Lee, Michael Stiber, *Senior Member, IEEE*, and Dong Si

***Abstract*— As with other modern sciences (and their computational counterparts), neuroscience experiments can now produce data that, in terms of both quantity and complexity, challenge our interpretative abilities. It is relatively common to be faced with datasets containing many millions of neural spikes collected from tens of thousands of neurons. Traditional data analysis methods can, in a relatively straightforward manner, identify large-scale features in such data (such as on the scale of entire networks). What these approaches often cannot do is to connect such macroscopic activity to the relevant small-scale behaviors of individuals cells, especially in the face of ongoing background activity that is not relevant. This communication presents an application of machine learning techniques to bridge the gap between microscopic and macroscopic behaviors and identify the small-scale activity that leads to large-scale behavior, reducing data complexity to a level that can be amenable to further analysis.**

I. Introduction

One of the central goals of neuroscience research is understanding how functional networks form and how the activity of such networks correlates with function. One method for investigating questions in this vein is to culture networks on multi-electrode arrays to allow for stimulation and recording during development [1, 2]. Such experiments, and their simulations, have shed light on a number of behavioral features, including whole network bursting [3, 4], activity wave propagation [5, 6, 7], and neuronal avalanches [8].

This communication addresses one aspect of such behavior: the origination of whole network bursting, manifest as traveling waves that begin among small foci which include one or a few neighboring neurons. In traditional spike train data analysis, point process statistical methods, such as autocorrelation and cross-correlation, are commonly used to characterize patterns formed by several spike trains [9, 10]. These methods work well when applied to small numbers of neurons and restricted time frames. However, our interest lies in analyzing spike data from relatively large neural populations (at least 10^4 cells) over long periods of development (days to weeks), approximating the size and time scale in living preparations. Therefore, we have investigated machine learning methods to study spatiotemporal bursting in such networks. In particular, we are interested in applying such techniques to see if stereotypical brief, localized patterns of activity that trigger network bursts can be identified from among many millions of spikes from many thousands of separate spike trains.

Authors are with the Computing and Software Systems Division, School of STEM, University of Washington Bothell, Bothell, WA 98011 USA. J.Y.L. (jewel87@uw.edu), M.S. (stiber@uw.edu), D.S. (dongsi@uw.edu).

II. Method

Our goal was to perform relatively straightforward analysis of macroscopic, whole-network behavior to provide input to machine learning (ML) algorithms, and then use these algorithms to see if small patterns of the detailed neuron spiking activity were predictive of the macroscopic behavior. If then it would suggest that the microscopic patterns were behaviorally significant. As detailed in the following subsections, we first performed large-scale, long-duration simulations of biological neural networks and collected all of the spiking activity of every neuron. These simulations produced whole-network bursting behavior, and we identified the burst events via macroscopic analysis. We then isolated spike sequences just before bursts, and spike sequences temporally distant from bursts, to provide input to a set of ML algorithms. We applied these algorithms to see if there were features in these sequences that could reliably predict that a burst would occur. We also found the origin location for each burst and applied ML algorithms to determine if predicting such origins could be used to identify relevant spatially localized activity patterns.

A. Data Acquisition

The BrainGrid neural simulator was used to simulate networks of 10,000 neurons in a 100×100 rectangular arrangement for the equivalent of 28 days' development, as described in detail in [4]. These simulations mimicked living preparations in which dissociated cortical cells were cultured on multi-electrode arrays, over a period of weeks forming networks that produced whole-network bursting behaviors. Simulations matching those that produced bursting behavior in that previous study were re-run on a 2.4GHz Intel Xeon E5-2620v3 system with NVIDIA K80 GPUs. This paper presents results from a single simulation (target rate = 1.0 spikes/sec; 90% excitatory neurons) as a representative example.

Each spike produced during a simulation had its time (as an integer time step value, with one step being 0.1ms) and (x, y) neuron position recorded; each dataset was stored as a 30GB HDF5 file.

B. Ground Truth: Burst Identification

Whole-network *bursts* were identified and their start and end times were determined as follows. A burst is defined as an occurrence of a very high rate of neural spiking across the entire network; they are easily distinguished from other activity. In this study, we identified burst events by assimilating all neuron spikes into a single train (i.e., discarding neuron position information) and grouping spikes into *neuronal avalanches*, where an avalanche is a run of sequential spikes all separated by inter-spike intervals (ISIs) smaller than the overall mean ISI [8]. Fig. 1 shows the size

distribution of avalanches. We determined that large avalanches containing more than 10^4 spikes exactly corresponded to the bursts identified in the corresponding simulations from [4]; moreover, all other avalanches had fewer than 10^3 spikes. Thus, bursts are easily distinguishable from other network behaviors by this straightforward, one-dimensional analysis. Out of a total of 15,408,016 avalanches identified, with between 2 and 111500 spikes, there were 4490 bursts in the simulation.

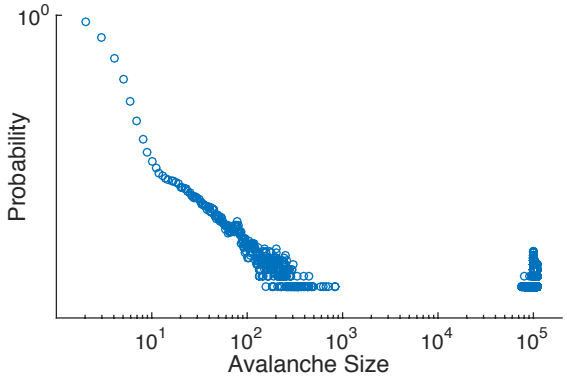

Figure 1. Avalanche size distribution. There are two distinct categories of avalanches that can be easily distinguished by their size.

C. Ground Truth: Burst Origin

We then visualized burst evolution as sequence of images (with each pixel as a neuron and color corresponding to spiking rate) or movies to see their spatiotemporal patterns. Fig. 2 shows the visualization result for burst evolution and burst origin location. This showed that bursts originated at single locations and propagated as waves across the network. We identified the approximate burst origin by calculating the centroid, (x, y), of neurons that spiked the most in the first 100 time steps (0.01s) for each burst.

D. Pre-Burst vs. Non-Burst Precursors

To determine if particular patterns of spatiotemporal activity triggered bursts, we divided the spike data before each burst into *non-burst* and *pre-burst* for pattern recognition. Fig. 3 illustrates how pre-burst and non-burst data were defined relative to a single burst event. With burst start and end times determined as in section II.B, we grouped N consecutive spikes before each burst as a pre-burst data sample. We grouped another N spikes 1000 time steps (0.1s, the "gap" in the figure) earlier as a non-burst data sample. In this study, we chose N values of 50, 100, and 500 for investigation.

For each spike i in a data sample, we retrieved its neuron (x, y) location and its firing time, $\tau_i = t_i - t_0$, relative to the first spike (spike 0) in that data sample. Every data sample was then arranged in the following format to include its spatial and temporal information: $\tau_0, x_0, y_0, \tau_1, x_1, y_1, \ldots, \tau_{N-1}, x_{N-1}, y_{N-1}$.

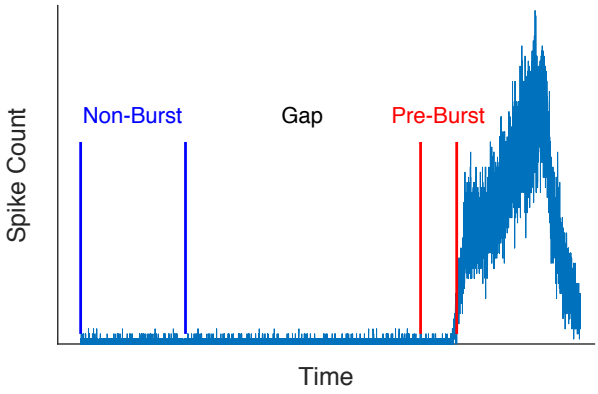

Figure 3. Definition of pre-burst and non-burst data. Blue and red lines indicated the start and end of times for non-burst and pre-burst windows, respectively, with a gap of 1000 time steps in between.

E. Data Analysis with Supervised Learning

We applied machine learning techniques to determine if burst initiation could be predicted from pre-burst activities and, if so, whether the approximate origin of a burst could be predicted as well. We tested both *Decision Tree (DT)* and *Support Vector Machine (SVM)* approaches for binary classification to predict initiation. We labeled each pre-burst data sample as 1 and each non-burst data sample as 0. DT uses a tree-like graph for decision making whereas SVM finds the maximum margin hyper-plane that best separates two classes in a high-dimensional feature space. While SVM is designed to handle high dimensional data, DT can be a cheaper, yet effective solution if the majority of the features (dimensions) don't contribute much information to the classification problem.

In total, we had 8980 samples (4490 pre-burst and 4490 non-burst), where each data sample had $3N$ (τ, x, y) features. Since we had a balanced dataset, model accuracy was used for performance evaluation over F score for its intuitiveness. To prevent prediction results being affected by sequential temporal relationships between consecutive bursts or between consecutive data samples, all data samples were randomly shuffled before they were used for model training. Classification models were evaluated using k-fold cross validation (k=10), in which the data was divided into k subsets and every time only k-1 subsets were used to train the model, leaving the last subset for testing.

For burst origin prediction, each pre-burst data sample (for N=100) was labeled with its subsequent burst origin (x, y) for multivariate regression analysis. One important macroscopic feature of the simulation behavior was that it settled down to a stable burst initiation location during development, with 50% of the bursts originating from the same region at later stages of the network development. We wanted to apply ML techniques to activity patterns that were less easily predictable. Therefore, in this study, we chose

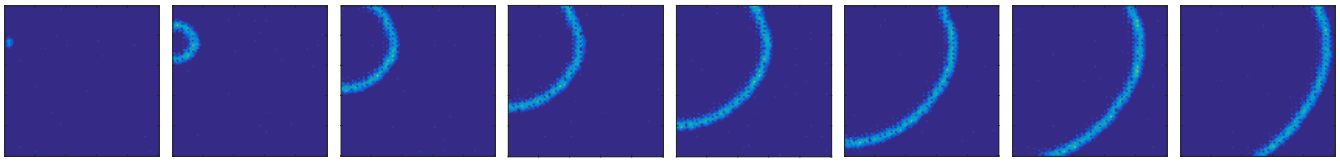

Figure 2. Evolution of a single burst. Images show beginning to the end of a single burst event from left to right, with each image corresponding to network activiy within 10 time steps (1ms) and images are 30 time steps (3ms) apart. Each pixel is a neuron, with firing rate represented by color.

only the first 150 bursts — those with origins that varied across the network — to analyze.

Linear regression, Lasso regression, Ridge regression, and artificial neural networks (ANNs) were used to predict two outcome variables (x, y) using the first 150 pre-burst data samples. We trained the ANN model with grid search for hyperparameter optimization. We experimented with different activation functions (logistic, tanh, ReLU), number of hidden layers (1 to 4), and different number of neurons in each hidden layer (20, 100, 200, 500); we present the best result here. Regression model accuracy were assessed by R-squared score, mean absolute error (MAE), and root mean squared error (RMSE). While R-squared allows us to understand the percentage of variance in the target data explained by the model, MAE and RMSE help quantify the error by averaging the residuals of the model.

All classification work was done in MATLAB using its built-in statistics and machine learning toolboxes. Regression analysis was performed using scikit-learn, a Python machine learning library. All training was done on a 2.4GHz Intel Xeon E5-2620v3 system.

III. Results

Binary classification results are shown in Table 1. We found that DT performed well in all three cases (N=50, 100, 500), with model accuracy all above 98%. Linear SVM performed best when N=50 and 100, also producing smallest k-fold errors. However, SVM model performance plummeted with N=500. The polynomial SVM model did much worse than Linear SVM in this application; hence only the results for a second degree polynomial kernel (d=2) when N = 100 and 50 are presented. As for training time, SVM models generally took longer to train compared to DT models.

TABLE I. Binary classification results for burst initiation

Method	N	Training Time	Model Accuracy	10-fold error
Decision Tree	500	0.7169 s	0.9829	0.0122
	100	0.1053 s	0.9959	0.0041
	50	0.0470 s	0.9922	0.0045
Linear SVM	500	219.8 s	0.6058	0.3848
	100	0.4716 s	0.9978	0.0005
	50	0.3316 s	0.9981	0.0000
Polynomial SVM (d=2)	100	269.2 s	0.8296	0.1013
	50	271.9 s	0.9016	0.0686

Regression results are shown in Table 2. The best performing ANN model had 3 hidden layers, each with 200 neurons with ReLU activation functions.

TABLE II. Regression performance for burst origin

Method	R2	MAE	RMSE
Linear	0.977	3.085	4.212
Lasso	0.9839	2.656	3.775
Ridge	0.9777	3.084	4.210
ANN	0.9406	5.737	8.743

IV. Discussion

From the classification results, we see that the DT model predicts burst initiation despite the large feature size ($3N$). DT constructs a tree by considering information gain as a criterion. It chooses the best feature to split each node so that it produces the "purest" subsets, and stops when data cannot be split further. In other words, a decision tree is built by calculating feature importance. This is one reason that DT is widely used as a feature selection technique. DT results indicate that there was a pattern predictive of burst in all three cases (N=50, 100, 500). We also found that when N=500 (1500 features), there were only 27 predictors being used to construct the DT model with 98% accuracy. In addition, half of the 27 features were from the last 100 spikes. This suggests that such ML techniques can be used as the first stage of analysis for large quantities of biological data to identify regions of interest for subsequent examination by more conventional approaches.

The linear SVM model for N=500 is only 60% accurate while it achieves 99.8% accuracy for N=50 and 100. Since a SVM finds the best classifying hyper-plane in the high-dimensional input space, this indicates that there were many redundant features when we included 500 spikes. This finding helped us narrow down the search window for burst trigger from including 500 spikes to 100 spikes. By comparison, the results for polynomial kernel SVMs were much worse than linear kernels, suggesting that our data's feature space is linearly separable and that the polynomial may have been over-fitting.

The multivariate regression models used here predicted two outcome variables representing the (x, y) location for burst origin based on pre-burst activity. From the R-squared scores, one can see that at least 94% of the data variance can be explained by our models. The fact that the lasso regression model performed best suggests that only a subset of features was important for burst origin prediction, since lasso selects a subset of features by reducing the coefficients of others to zero. ANNs performed worst in this task, however, these models are notoriously sensitive to fine-tuning of hyperparameters and thus continued experimentation might lead to better performance in this case.

For all of our regression models, RMSE was a little higher than MAE. Since RMSE weights large errors more highly than MAE, this indicates that there was a subset of distant outliers in our prediction. Overall, results showed that the predicted (x, y) locations were off by 2 to 5 neurons, which is quite accurate.

V. Conclusion

The ML results allow us to conclude that, even in the absence of detailed analysis of pre-burst spiking patterns, there is commonly a localized spatiotemporal pattern of spikes that provides reliable information not only that a burst will occur but also where that burst will start. This can focus subsequent analysis on the patterns among a small number of spikes and neurons (fewer than 50 for each burst). For the current investigation, that corresponds to a reduction of more than seven orders of magnitude from the full, 600 million spike, data set.

ACKNOWLEDGMENT

MS thanks the Fulbright program of the US Dept. of State and the Czech Republic Fulbright Commission for their support and the Institute of Physiology of the Czech Academy of Sciences for their hospitality and discussions related to this work. JYL thanks the UW Bothell CSS Division/School of STEM for financial support.

REFERENCES

[1] G.W. Gross, "Simultaneous single unit recording in vitro with a photoetched laser deinsulated gold multimicroelectrode surface," *IEEE Trans. Biomedical Engineering*, vol. 5, pp. 273-279, 1979.

[2] D.A. Wagenaar, J. Pine and S.M. Potter, "An extremely rich repertoire of bursting patterns during the development of cortical cultures," *BMC Neuroscience*, vol. 7, no. 11, Feb. 2006.

[3] J. van Pelt, P.S. Wolters, M.A. Corner, W.L.C. Rutten, and G.J.A. Ramakers, "Long-Term Characterization of Firing Dynamics of Spontaneous Bursts in Cultured Neural Networks," *IEEE Trans. Biomedical Engineering*, vol. 51, no. 11, pp. 2051-2062, Nov. 2004.

[4] F. Kawasaki and M. Stiber, "A simple model of cortical culture growth: burst property dependence on network composition and activity," *Biological Cybernetics*, vol. 108, no. 4, pp. 423-443, 2014.

[5] Y. Nishitani, C. Hosokawa, Y. Mizuno-Matsumoto, T. Miyoshi, and S. Tamura, S., "Classification of Spike Wave Propagations in a Cultured Neuronal Network: Investigating a Brain Communication Mechanism," *AIMS Neuroscience*, vol. 4, no. 1, pp. 1-13, Dec. 2016.

[6] S. Sakuma, Y. Mizuno-Matsumoto, Y. Nishitani, and S. Tamura, "Simulation of spike wave propagation and two-to-one communication with dynamic time warping," *AIMS Neuroscience*, vol. 3, no. 4, pp. 474-486, Dec. 2016.

[7] S. Okujeni, S. Kandler, and U. Egert, "Mesoscale architecture shapes initiation and richness of spontaneous network activity," *J. Neuroscience*, vol. 37, no. 14, pp. 3972-3987, 2017.

[8] J.M. Beggs, and D. Plenz, D., "Neuronal avalanches in neocortical circuits," *J. Neuroscience*, vol. 23, no. 35, pp. 11167-11177, Dec. 2003.

[9] I.V. Tetko, and A.E. Villa, "Pattern grouping algorithm and de-convolution filtering of non-stationary correlated Poisson processes," *Neurocomputing*, vol. 38, pp. 1709-1714, 2001.

[10] D.C. Tam, "A spike train analysis for correlating burst firings in neurons," *Neurocomputing*, vol. 38, pp. 951-955, 2001.