

Predicting Neural Progenitor Cell Networks

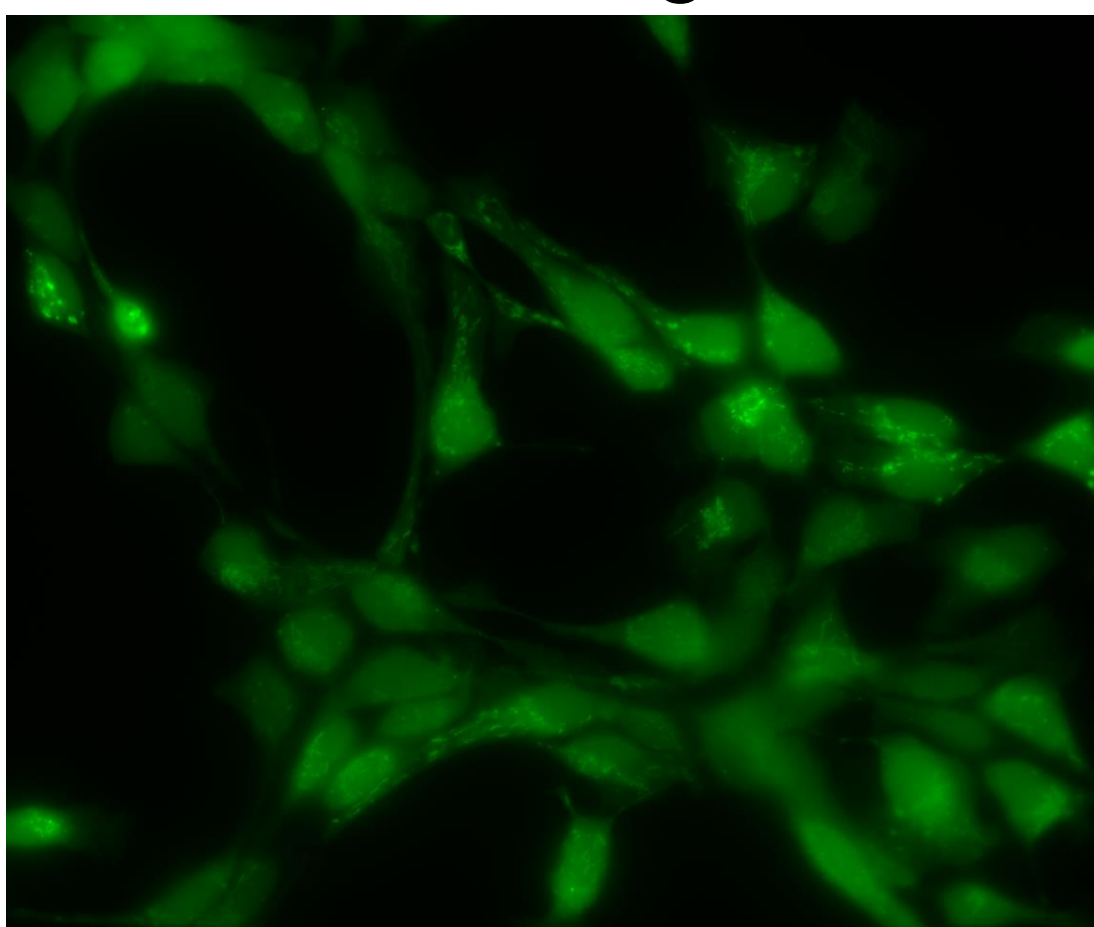
Nicolas Grandel
Stanford University

Overview

The ways in which pre-neurons (known as neural progenitor cells, or NPCs) form communication networks is an active topic of research. Cell communication is often measured using the cross-correlation of their calcium concentration over time. However, this metric is arbitrary, and thus it is difficult to select a cutoff value to differentiate between “talking” and “not-talking” cells. In order to solve this problem, I attempted to use machine learning techniques to predict networks in a more robust manner, hopefully making conclusions based on such data more reliable.

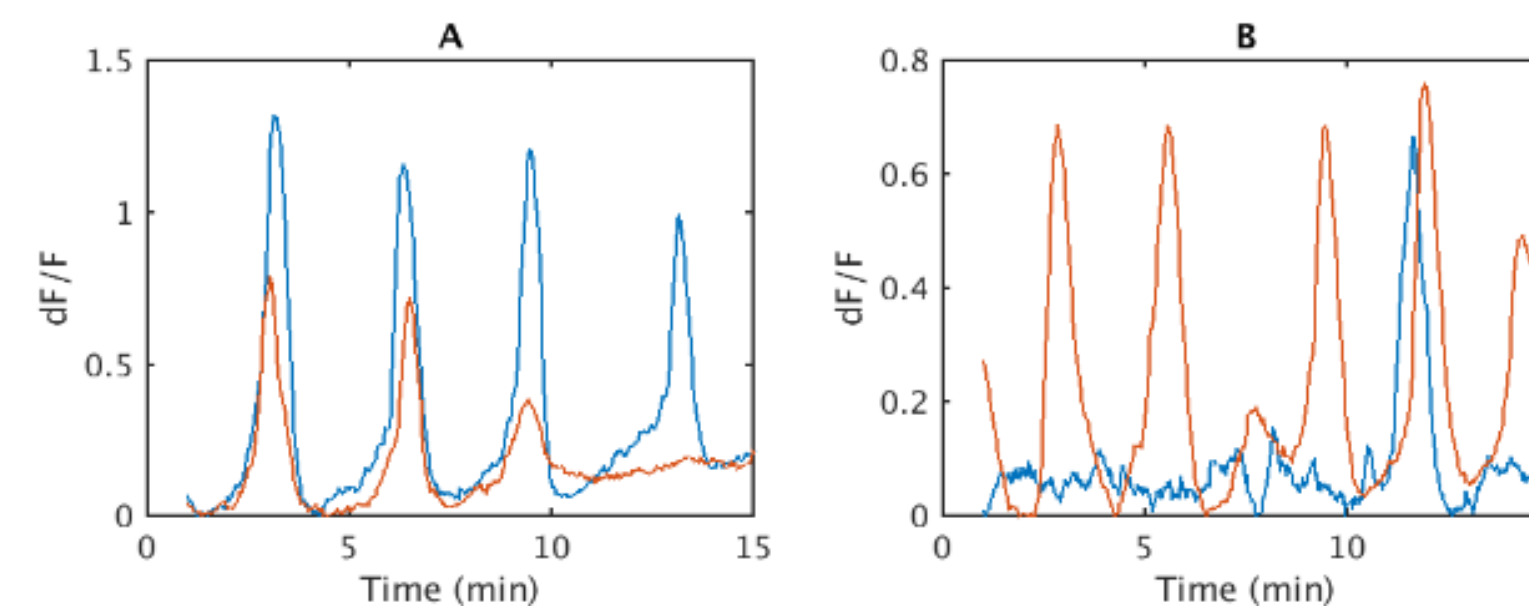
Data

The data used for this project was taken from the Qutub lab, at Rice University. Time-lapse videos were taken of confluent cells over a period of a few hours, where intensity was reflective of calcium concentration. Images were converted to grayscale.



Features

While I attempted to use several derived features, including the period and frequency of calcium “spikes”, I eventually realized that while it was difficult to use cross-correlation as a decision boundary by itself, it could still be used as a metric of similarity, to measure the similarity of two signal sets.



(A) Correlated and (B) uncorrelated spike trains

Models

I initially attempted to model correlated sets by generating artificial spike trains, supervised learning methods failed to accurately predict networks on real data, suggesting that I had insufficient skill to properly model these time series. With this in mind, I turned to using the K-Means clustering algorithm, with the cross-correlation used as a measure of distance, defined below.

$$\operatorname{argmax}_x \frac{1}{n} \sum_x \frac{1}{\sigma_f \sigma_t} (f(x) - \bar{f})(t(x) - \bar{t})$$

Results

Model	Train	Validation	Test
Log. Reg.	.953	.908	.734
SMO	.923	.884	.752
KNN	.896	.851	.728

Compare this to the inter-cluster cross-correlation results of the simple K-Means algorithm.

Model	Train	Test
K-Means	0.885	0.842
Cross-Correlation	0.862	0.837

Discussion and Future

Overall, simple modeling of calcium spike trains, even with gaussian noise, was insufficient to accurately capture the nature of real data. As such, my efforts to use supervised learning to solve this issue requires much more time to be spent on the modeling end, time which I did not have for this project.

However, my efforts to create a simple clustering algorithm to generate NPC networks seems to have worked quite well, attesting to the power cross-correlation has as a relative similarity metric.

References

- Gobel, W., & Helmchen, F. (2007). In Vivo Calcium Imaging of Neural Network Function. *Physiology*, 22(6), 358. <https://doi.org/10.1152/physiol.00032.2007>
- Rahmati V, Kirmse K, Marković D, Holthoff K, Kiebel SJ (2016) Inferring Neuronal Dynamics from Calcium Imaging Data Using Biophysical Models and Bayesian Inference. *PLoS Comput Biol*12(2): e1004736. <https://doi.org/10.1371/journal.pcbi.1004736>