Neuronal Connectivity via Bayesian Tensor Factorization

Chris Mulligan
Columbia University

Statistical Analysis of Neural Data

December 15, 2015

Neuronal Connections

There is significant interest in discovering and modeling connections between neurons.

General goal is to develop a directed graph of neuron connections.

Neuronal Connections

Consider the standard setup:

Time series dataset of y_{it}, for $i = 1, \ldots, N$ neurons over $t = 1, \ldots, T$ time steps.
y_{it} is the number of times neuron i fired at discrete time t

Goal:

For each neuron i, infer the subset of neurons j that either excite or inhibit neuron i.

Neuronal Connections

Lots of work in this space. A few examples:

- Cross-Correlation based techniques
- Granger Causality
- L_1-regularized GLMs
- Factor Analysis
- Network Hawkes Processes

Consider Bayesian Tensor Factorization[1]

We consider using Bayesian tensor factorization as a means of inferring latent components that can be used to understand the relationships between neurons.

Let $\underline{\mathbf{Y}}$ be a three way tensor of size $N \times N \times L$.
y_{ijl} is the count of times that neuron i fired l time steps after neuron j.

[1] Based on Schein, Paisley, Blei, and Wallach (2015)

Tensor Factorization

We use the Canonical Polyadic (CP) decomposition. The K component decomposition decomposes the 3 way tensor into 3 latent factor matrices $\Theta^{(1)}, \Theta^{(3)}, \Theta^{(2)}$, with dimensions $d_m \times K$.

$$y_{ijl} \approx \hat{y}_{ijl} = \sum_{k=1}^{K} \theta^1_{ik} \theta^2_{jk} \theta^3_{lk}$$

- The columns of each of the Θ matrices represent a latent *component*
- $\Theta^{(1)}_k$ vector represents the neurons being “predicted”
- $\Theta^{(3)}_k$ represents the neurons being depended on
- $\Theta^{(2)}_k$ represents the lags of the dependency

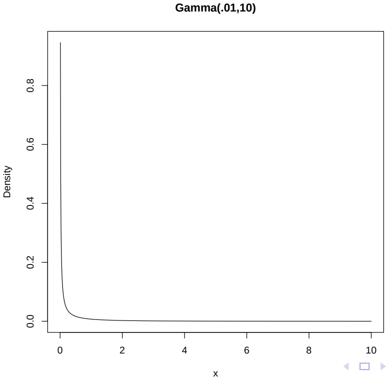
Bayesian Tensor Factorization

View the tensor factorization probabilistically:

$$y_{ijl} \sim Pois(\hat{y}_{ijl})$$

Go Bayesian: impose priors on the latent factors. Gamma is conjugate to Poisson.

Parameters: very small a + small(ish) b. $\implies$ concentrate mass near 0, but heavy tail $\implies$ sparsity in the latent factors!



Approximate Inference

Posterior distribution: $P\left(\Theta^{(1:3)}, b \mid \underline{\mathbf{Y}}, a\right)$
Must be approximated. Use variational inference with a mean field approximation, similar to what's commonly used in Bayesian Poisson Matrix Factorization. Use an independent Gamma for each latent factor, eg:

$$P\left(\theta_{ik}^{(1)}|\cdot\right) \approx Q\left(\theta_{ik}^{(1)}|S_{ik}^{(1)}\right) = Gamma\left(\theta_{ik}^{(1)}|\alpha_{ik}^{(1)}, \beta_{ik}^{(1)}\right)$$

Then choose $S^* = \operatorname{argmin} \mathrm{KL}(P||Q)$ via coordinate ascent, stopping when the ELBO converges.

Interpreting Results

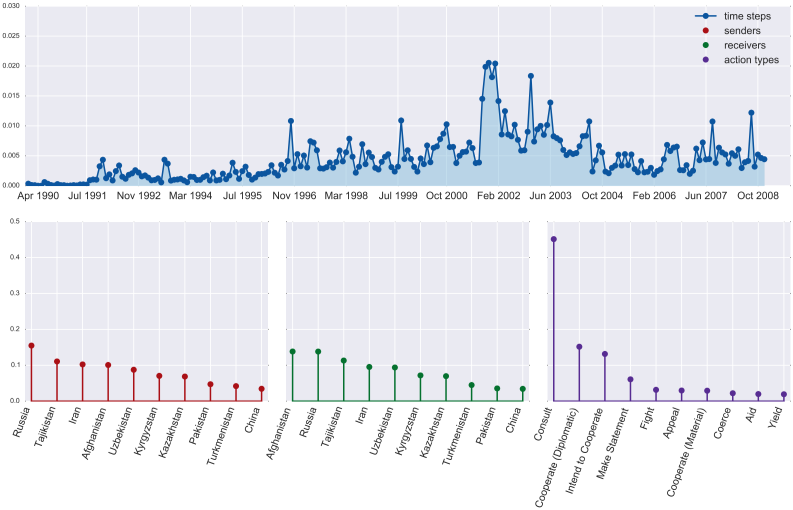


Figure: One K component in Multilateral Relations from Schein et al

Interpreting Results 2

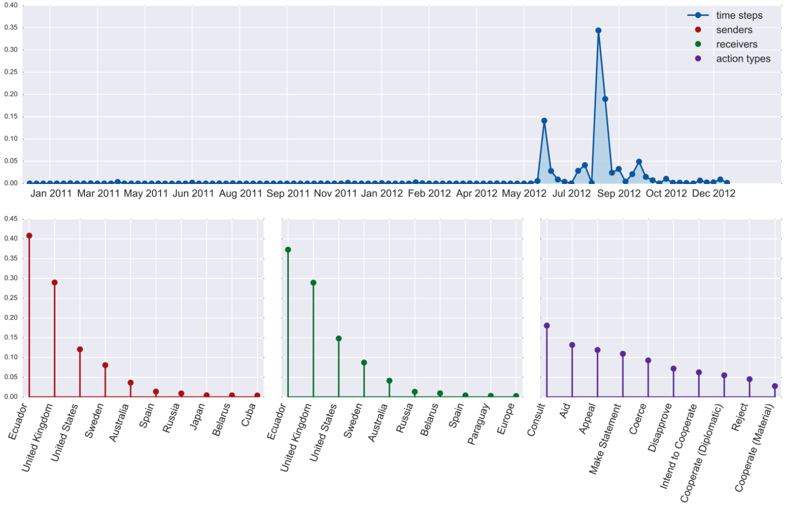


Figure: Another component: Wikileaks

Simple Synthetic Result

A small synthetic dataset consisting of 25 Neurons (with a few very strong connections) over 100,000ms was constructed with EnaS. The dataset was constructed with $L = 20$ maximum lag, and fit with $K = 6$ components.

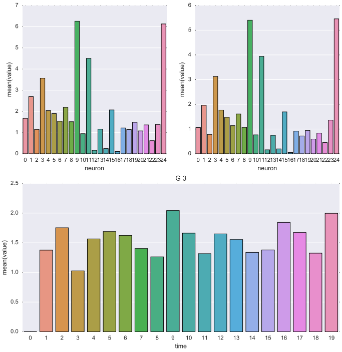


Figure: Component 3, which captures general neuron activity levels

Simple Synthetic Result

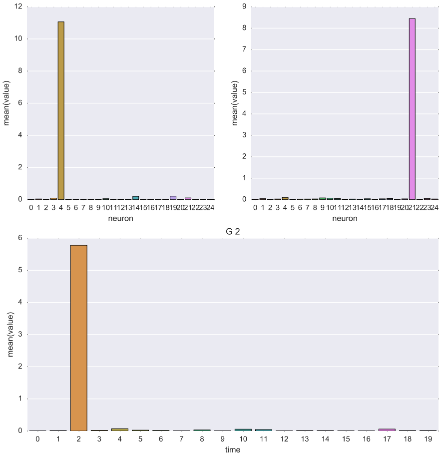


Figure: Component 2: Neuron 5 depends on Neuron 21, with a lag of 2

Simple Synthetic Result

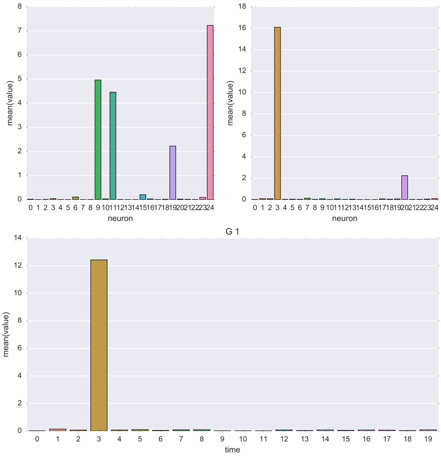
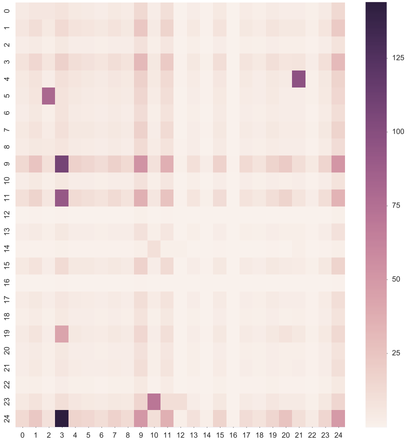


Figure: Component 1: More complicated relationship, with a lag of 3

Simple Synthetic Result

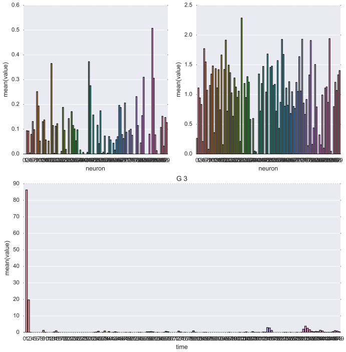
Several sensible ways of looking at the directed graph.
One is $\Theta^{(1)}\Theta^{(2)\mathrm{T}}$, visualized below:



Connectomics Challenges Results

I used the Connectomics Challenge dataset. $N = 100$ neurons, with $T \approx 180,000$. It's a highly clustered network. I deconvolved the fluorescence using PyFNND, which implements Voglestein et al.

The dataset was constructed with $L = 150$ maximum lag, and fit with $K = 50$ components.



Connectomics Challenges Results

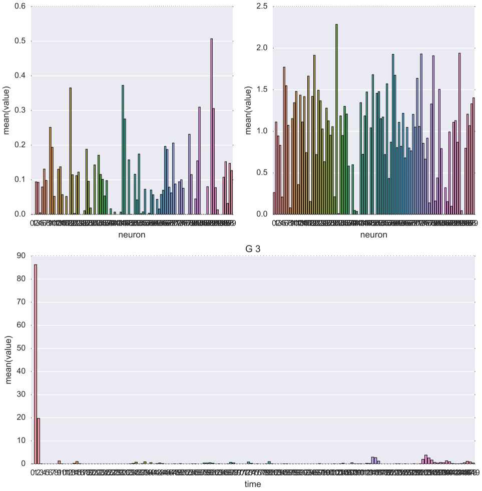


Figure: 3: one of several which captures neuron correlations in time 0

Connectomics Challenges Results

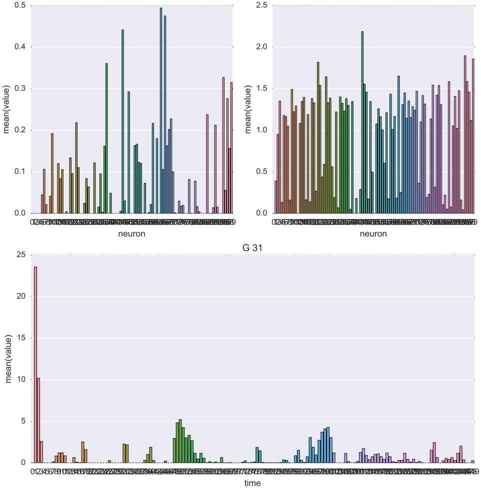


Figure: 31: Interesting temporal pattern, and sparse i vector

Connectomics Challenges Results

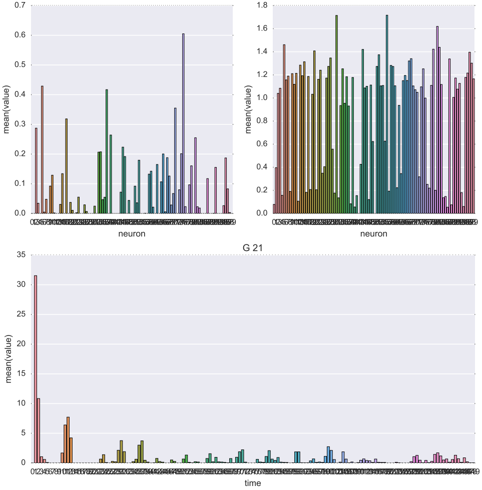


Figure: 21: Note sparsity of neurons, but relative flatness in time

Unfortunate Graph Results

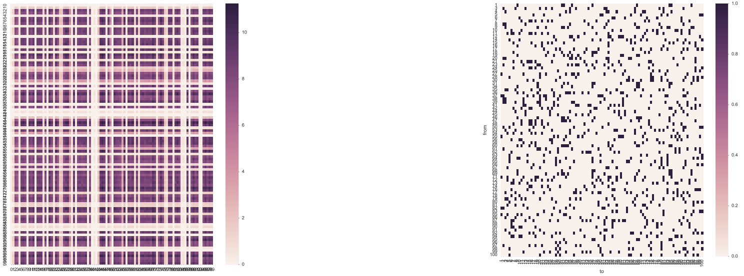


Figure: Left: Inferred with same method. Right: Actual

Thanks!