Inferring Inference

Xaq Pitkow
Rajkumar Vasudeva Raju

part of the MICrONS project with Tolias, Bethge, Patel, Zemel, Urtasun, Xu, Siapas, Paninski, Baraniuk, Reid, Seung

NICE workshop 2017
Hypothesis:
The brain approximates probabilistic inference over a probabilistic graphical model using a message-passing algorithm implicit in population dynamics.
What algorithms can we learn from the brain?

**Architectures?**
cortex, hippocampus, cerebellum, basal ganglia, …

**Transformations?**
nonlinear dynamics from population responses

**Learning rules?**
short and long-term plasticity
**Principles:**

- Probabilistic
- Nonlinear
- Distributed

**Details:**

- Graphical models
- Message-passing inference
- Multiplexed across neurons
Events in the world can cause many neural responses. Neural responses can be caused by many events.

So neural computation is inevitably statistical. This provides us with mathematical predictions.
Why does it matter whether processing is linear or nonlinear?

*If all computation were linear we wouldn’t need a brain.*
Two sources of nonlinearities

Relationships between latent variables

\[ \text{Image} = \text{Light} \times \text{Reflectance} \]

Relationships between uncertainties

*posteriors generally have nonlinear dependencies even for the simplest variables*

*Product rule:*  \[ p(x,y) = p(x) \cdot p(y) \]

*Sum rule:*  \[ L(x) = \log \sum_y \exp L(x,y) \]
Probabilistic Graphical Models:

Simplify joint distribution $p(x|r)$ by specifying how variables interact

$$p(x|r) \propto \prod_{\alpha} \psi_{\alpha}(x_{\alpha})$$
Example: Pairwise Markov Random Field

\[ p(x) = \frac{1}{Z} \prod_{s \in V} e^{J_s(x_s)} \prod_{(s,t) \in E} e^{J_{st}(x_{st})} \]
Approximate inference by *message-passing*:

- Localize information so it is actionable
- Summarize statistics relevant for targets
- Send that information *along graph*
- Iteratively update factors with new information

\[
\theta_{i,t+1} = f(\theta_{it}, \{\theta_{jt}\}_{j \in N_i} | G, J)
\]

*general equation*  
*interactions*  
*posterior for neighbors*  
*message-passing parameters*  
*posterior parameters*
Example message-passing algorithms

- Mean-field (assumes variables are independent)
- Belief propagation (assumes tree graph)
- Expectation propagation (updates parametric posterior)
- ...
- Brain’s clever tricks?
Spatial representation of uncertainty
(e.g. Probabilistic Population Codes, PPCs)

Pattern of activity represents probability.
More spikes generally means more certainty

Ma, Beck, Latham, Pouget 2006, etc
Message-passing updates

\[ \theta_{i,t+1} = f(\theta_{it}, \{\theta_{jt}\}_{j \in N_i} | G, J) \]

embedding

\[ r = U\Theta + \eta \]

Neural dynamics

\[ \dot{r} = F(r_t) \]
linear connections

singleton populations

pairwise populations

nonlinear connections

$$r_{12}$$  $$r_{23}$$
linear connections

singleton populations

pairwise populations

linear connections

nonlinear connections

$x_1 \rightarrow J_{12} \rightarrow x_2 \rightarrow J_{23} \rightarrow x_3$

$J_1 \quad J_2 \quad J_3$

$J_{12} \quad J_{23}$

$J_1 \quad J_2 \quad J_3$
Neural activity
Neural activity
Neural activity

Neural encoding

Information encoded
Neural activity

Neural encoding

Information encoded
Neural interactions

Neural encoding

Information interactions
Neural interactions

Neural encoding

Information interactions

Example: orientation
Network activity can implicitly perform inference

\[
\frac{N_{\text{neurons}}}{N_{\text{params}}} = 1
\]

\[
\frac{N_{\text{neurons}}}{N_{\text{params}}} = 5
\]

Raju and Pitkow 2016
Simulated brain

\[ \dot{b}_{it} = -b_{it} + \sigma \left( \sum_j W [J_{ij}, b_{it}, b_{jt}] b_{jt} + h_{it} \right) \]
\[ W [J_{ij}, b_i, b_j] = 2J_{ij} + 4J_{ij}^2 (1 - 2b_i)(1 - b_j) \]

\[ r_{t+1} = \sigma \left( A r_t + B h_t - \theta \right) \]

Inferring inference

\[ \hat{b} = V r + c \]
\[ \hat{W}[J_{ij}, b_i, b_j] = \sum_{\alpha\beta\gamma} G_{\alpha\beta\gamma} J_{ij}^\alpha b_i^\beta b_j^\gamma \]

Message-passing parameters

Interactions

Infer

Encode

Decode

Fit* *(within family)
Recovery results for simulated brain

Message-passing parameters

$G_{\alpha\beta\gamma}$

Interactions

$J_{ij}$

True

Learnt

$\neq$
Analysis reveals degenerate family of equivalent algorithms.
From *simulated* neural data we have recovered:

- how variables are encoded
- which variables interact
- how they interact
- how the interactions are *used*

<table>
<thead>
<tr>
<th>Representation</th>
<th>Graphical model</th>
<th>Message-Passing algorithm</th>
</tr>
</thead>
</table>
Applying message-passing to novel tasks

- Brain neural network
- Message passing nonlinearity
- Apply to new graphical model structure
- Relax to novel neural network

OR
Next up: applying methods to real brains

stimulus: orientation field

recordings: V1 responses*  

*not to same stimulus recordings from Tolias lab
mementos:

- Neurons can perform inference *implicitly* in a graphical model distributed across a population.

- New method to discover message-passing algorithms by modeling transformations of decoded task variables.
acknowledgements

funding:

collaborators

Alex Pouget
Jeff Beck
Dora Angelaki
Andreas Tolias
Jacob Reimer
Fabian Sinz
Alex Ecker
Ankit Patel

xaqlab.com

Rajkumar Vasudeva Raju

Kaushik Lakshminarasimhan
Qianli Yang
Emin Orhan
Aram Giahi-Saravani
KiJung Yoon
James Bridgewater
Zhengwei Wu
Saurabh Daptardar