Inference, Attention, and Decision in a Bayesian Neural Architecture

Angela J. Yu and Peter Dayan

NIPS 2004

03.12.2009, Seminar?
Lars Büsing
Neural architecture for Bayesian inference

- Posner’s task: Paradigm for studying the influence of spacial attention on visual discrimination
- Neural architecture for modeling experimental data
- Framework: Bayesian inference
- Decision making
Overview

1. Underlying experiment: Posner’s task
2. Theoretical foundation: Bayesian inference in Posner’s task
3. Neural architecture: Modeling Bayesian inference in Posner’s task
4. Comparison of model performance with data
1. Posner’s task
Posner’s task: Setup

- Stimulus has two dimensions: angle $\phi \in [0, 2\pi]$ position on the screen $\mu \in [-1.5, 1.5]$
- Task: estimate $\phi$ independent of $\mu$
- Cue hints where the stimulus is to be expected: valid trial
- Cue gives reliable information about stimulus position with probability $\gamma$ called the validity
- With probability $1 - \gamma$ cue is unrelated to stimulus position: invalid trial
- Paradigm to study spatial attention
- Attention is guided by prior position expectation
Posner’s task: Observations

- “Localization” of spatial attention is controlled by $\gamma$
- For valid vs. invalid trials reaction time, accuracy etc. are better: validity effect
- Hence spacial attention on $\mu$ boosts estimation of variable $\phi$
- This even works if $\mu$ and $\phi$ are statistically independent
- These effects scale with validity $\gamma$
2. Theoretical foundation: Bayesian inference in Posner’s task
Generative model I

- Stimulus has two variables: location $\mu \in [-1.5, 1.5]$ and angle $\phi \in [0, 2\pi]$
- Prior $p(\mu, \phi) = p(\mu)p(\phi)$ with $p(\phi) = 1/(2\pi)$
- Cue for $\mu$ induces no-flat prior $p(\mu)$ over $\mu$: $\mu \sim \gamma N(0, \nu^2) + (1 - \gamma)c$
  ie. wlog cue indicats $\mu = 0$
- Observations $x(n)$
Generative model II

- Grid of input neurons with indices $i, j$ and activities $x_{ij}(t)$ with Gaussian receptive fields with preferred place $\mu_i$ and preffered angle $\phi_j$ equally spaced

- Stimuli/input up to time step $t$: $X_t = (x(1), \ldots, x(t))$ with iid. samples $x(n) \in \mathbb{R}^{N \times N}$

- Average $x_{ij}(t)$ given by $f_{ij} = \exp(-(\mu_i - \mu^*)^2/(2\sigma^2_{\mu}) + k \cos(\phi_j - \phi^*))$ with the true $\mu^*, \phi^* +$ Gaussian noise

- $x_{ij}(t) \sim \mathcal{N}(f_{ij}(\mu^*, \phi^*),$ noise variance$)$

Task: Inference of $p(\phi|X_t)$
Bayesian Inference

Task: Inference of \( p(\phi | X_t) \)

1. Bayesian inference:

\[
p(\phi | X_t) = \frac{1}{p(x(t) | X_{t-1})} \left\{ \begin{array}{l}
\text{likelihood} \\
\text{normalization}
\end{array} \right\} p(x(t) | \phi) p(\phi | X_{t-1})
\]

2. Marginalized likelihood:

\[
p(x(t) | \phi) = \int d\mu \ p(x(t) | \mu, \phi) \ p(\mu | \phi)
\]

\[
= \int d\mu \ p(x(t) | \mu, \phi) \ p(\mu)
\]
Bayesian Inference

Task: Inference of $p(\phi|X_t)$

1. $p(x_t|\mu, \phi)$
2. $p(x_t|\mu, \phi)p(\mu)$
3. $p(x_t|\phi) = \int d\mu \ p(x_t|\mu, \phi) \ p(\mu)$
4. $p(\phi|X_t) = \text{const} \cdot p(x_t|\phi)p(\phi|X_{t-1})$
5. argmax $p(\phi|X_t) \geq \text{threshold}$

These five steps correspond to the five layers of the network
3. Neural architecture: Modeling Bayesian inference in Posner’s task
Neural architecture I

- \[ r^5_j(t) = r^4_j(t) \] if \( r^4_j(t) = \text{argmax}_i(r^4_i(t)) \) and 0 else; WTA step

- \[ r^4_j(t) = r^4_j(t-1) + r^3_j(t) + \text{const} \]
  update of posterior \( p(\phi|X_t) \) via recurrent connections

- \[ r^3_j(t) = \log(\sum_i \exp(r^2_{ij}(t))) \approx \sum_i c_i r^2_{ij}(t) \]
  marginalization over \( \mu \), approximated by sums-of-logs, *marginalized likelihood* \( p(x_t|\phi) \)

- \[ r^2_{ij}(t) = r^1_{ij}(t) + \log P(\mu_i) \]
  likelihood \( \times \mu\)-prior

- \[ r^1_{ij}(t) = \log p(x(t)|\mu_i, \phi_j) \]
  log likelihood
Neural architecture II

- Output neuron $j$ of top layer: $= \arg\max_{\phi_j} p(\phi_j|X_t)$
- Decision making:
  make decision if $\max_j p(\phi_j|X_t) \geq \text{threshold}$
- reaction time $:= t$ number of observations
- Decision $= \arg\max_{\phi_j} p(\phi_j|X_t)$
4. Comparison of model performance with data
Validity effect

- Valid trial: cue predicted position *correctly*: $\mu \sim \mathcal{N}(0, \nu_2)$
- Invalid trial: cue predicted position *incorrectly*: $\mu \sim \text{uniform}$
- Experimental data: Reaction time for valid trials smaller
Understanding the model’s validity effect

Prior information (valid cuing) about over the independent variable $\mu$ can help to make inference $\phi$
Influence of spacial attention

- Modeling assumption: attention = prior $p(\mu)$ about $\mu$
- Experimental observation:
  Attention to parts of visual field increases stimulus-induced responses
- Model can qualitatively reproduce this
Summary

- Neural model of Bayesian inference in Posner’s task
- Prior over spacial variable: spacial attention
- Valid cuing ("attention focused to the right spot") helps estimating independent variable

Issues:
- Realistic input $r_{ij}^1(t) = \log p(x_t|\mu_i, \phi_j)$?
- Equations in the paper are just loosely related to the topic
Input in lowest layer: \( r_{ij}^1(t) = \log p(x_t | \mu_i, \phi_j) \)

\[
r_{ij}^1(t) = \text{const} - (2\nu^2)^{-1} \sum_{kl} \left( x_{kl}(t) - \psi_{kl}^{ij} \right)^2
\]

- \( r_{ij}^1(t) \) quadratic in sensory input \( x_{kl} \)
- \( r_{ij}^1(t) \) has additive weights \( \psi_{kl}^{ij} \)
- Sensory noise level \( \nu \) has to be known