

用于发现大脑状态切换的独热广义线性模型

李成睿¹, Soon Ho Kim², Chris Rodgers³, Hannah Choi², 吴安琪¹

¹ School of Computational Science & Engineering
Georgia Institute of Technology, Atlanta, GA 30305, USA
`{cnlichengrui, anqiwu}@gatech.edu`

² School of Mathematics
Georgia Institute of Technology, Atlanta, GA 30305, USA
`{soonhokim, hannahch}@gatech.edu`

³ Department of Neurosurgery, School of Medicine
Emory University, Atlanta, GA 30322, USA
`christopher.rodgers@emory.edu`

2024 年 3 月 13 日

摘要

探寻有意义、可解释的神经交互对理解大脑神经回路十分重要。由神经信号推断得到的神经交互主要反映了功能性连接。在一个较长的实验中，被试动物可能会经历不同的阶段，取决于实验、刺激或行为状态。因此功能性连接可能会随着时间而改变。为了建模动态改变的功能性连接，之前的工作一般是采用状态可切换的隐马尔可夫模型 (HMM) 加广义线性模型 (GLM) (即, HMM-GLMs)。然而我们认为这样的模型缺乏生物合理性，因为功能性连接是被潜在的大脑解剖连接组塑造和限制的。这里我们提出了两个新的带有先验的状态可切换的广义线性模型，分别为 Gaussian HMM-GLM (Gaussian 先验) 和 one-hot HMM-GLM (Gumbel-Softmax one-hot 先验)。我们学到的先验可以捕捉到不随状态变化的交互模式，这帮我们揭示出潜在的解剖连接组和神经元的实际物理交互。而每个 GLM 建模的随状态变化的交互则能够捕捉大脑不同状态下的功能性变化。我们的方法能够在模拟数据上有效的恢复模型真实的交互结构，达到最高的预测似然，并且在增强了神经交互模式和隐状态在真实神经数据上的可解释性。代码在 <https://github.com/JerrySoybean/onehot-hmmglm>。

1 引言

揭示有意义、可解释的神经交互对理解大脑神经回路十分重要。许多工作都用统计或信息论的方法来研究这些交互，比如互相关图 [Jia et al., 2022]、互信息 [Houghton, 2019]、Granger 因果 [Granger,

1969]、传递熵 [Schreiber, 2000]、动力系统上的广义线性模型 [Linderman et al., 2016, 2017, Glaser et al., 2020] 以及 Hawkes 过程 [Li et al., 2022]. 一般来说, 推断的神经交互反映了神经元之间的动态功能性连接, 其受到神经活动变化的影响. 而直接观察或推断涉及到轴突、树突和突触的解剖连接组通常比较困难. 此外, 和在一段时间内保持相对稳定的解剖连接组不同, 功能性连接会随行为状态变化, 且这种变化是较快的. 因此, 神经元之间的功能性链接能够反映由任务和感觉输入的所塑造的动态计算模式. 现有的实验证据已经表明, 神经系统会表现出多变的激发模式, 这些变化与不同的感觉、感知和行为状态有关 [Sherman, 2001, Haider et al., 2007, Anderson et al., 2000, Sanchez-Vives and McCormick, 2000, Escola et al., 2011].

为了捕捉这种多状态系统中随时间变化的功能性连接, 之前的研究采用了状态可切换的隐马尔可夫模型 (HMM) 加广义线性模型 (GLM), 称为 HMM-GLMs [Escola et al., 2011, Nadagouda and Dav-enport, 2021, Zhou et al., 2021, Morariu-Patrichi and Pakkanen, 2022]. 这些模型引入了离散隐变量, 用于表示每个时间点的状态. 而每个状态都有其自己的 GLM 用于捕捉神经交互. 然而我们认为这样的方法在生物学上不那么合理, 因为由神经信号推断出的两个神经元之间的交互不仅能反映功能性连接, 还能反应两者之间的解剖连接或者说是突触连接. 已有实验证据表明功能网络和解剖网络之间有一定的相关性 [Genç et al., 2016, Siegle et al., 2021]. 因此更合理的做法应该是假设功能性连接在大脑状态动态调节的情况下, 还受限于潜在的解剖连接组.

为了加入这些生物学合理的假设, 我们在 HMM-GLM 的框架下引入了一个新的方法来捕捉多状态神经系统中随时间变化的功能性连接. 不像之前的 HMM-GLM 假设不同状态下的 GLM 是完全独立的, 我们引入了一个可以学的先验于所有状态上, 以限制每个 GLM 的交互权重矩阵的搜索空间. 我们先提供了一个方案, 用一个所有状态下的权重矩阵共享的高斯先验, 记为 Gaussian HMM-GLM. 然而, 高斯先验过于简单, 无法显式的建立功能性连接和解剖连接组之间的关系. 所以随之, 我们提供了第二个解决方案. 我们把 GLM 的权重矩阵分解成了一个邻接矩阵和一个强度矩阵, 而邻接矩阵由独热码的形式表示. 先验则只加在邻接矩阵上而不是整个的权重矩阵. 我们认为, 在有先验规范下的邻接矩阵能够体现潜在的解剖连接组, 揭示神经元之间的实际物理交互. 与此同时, 未受限制的强度矩阵为不同大脑状态下的功能性变化提供了灵活性. 我们的实验结果表明, 与其它方法相比, one-hot HMM-GLM 能够准确地模拟数据上恢复真实的神经元交互结构, 并且在两个真实神经数据集上达到最高的测试集预测似然. 此外, 学习得到的神经元之间的交互结构和隐状态相比于其它方法在真实数据集上更具可解释性.

2 方法

经典 GLM: 把来自 N 个神经元 T 个时间 (桶) 的放电序列数据记为 $\mathbf{X} \in \mathbb{N}^{T \times N}$, $x_{t,n}$ 是第 n 个神经元在第 t 个时间 (桶) 内的放电数, $\mathbf{x}_t \in \mathbb{R}^{N \times 1}$ 是在时间 t 时所有神经元放电数的向量. 当给定 \mathbf{X} , 带有提前给定的基函数的经典 GLM 预测下一个时间 t 神经元的放电率

$$f_{t,n} = \sigma \left(b_n + \sum_{n'=1}^N w_{n \leftarrow n'} \cdot \left(\sum_{k=1}^K x_{t-k,n'} \phi_k \right) \right), \quad \text{放电数 } x_{t,n} \sim \text{Poisson}(f_{t,n}), \quad (1)$$

其中 $\sigma : \mathbb{R} \rightarrow \mathbb{R}_+$ 是一个非线性函数 (比如, Softplus); b_n 是第 n 个神经元的基准强度; $w_{n \leftarrow n'}$ 是第 n' 个神经元给到第 n 个神经元的影响权重, 其矩阵形式为 $\mathbf{W} \in \mathbb{R}^{N \times N}$; $\phi \in \mathbb{R}_+^K$ 是一个提前给定的基函数, 用于合计神经元从 $t-K$ 到 $t-1$ 的放电历史. GLM 通过最大化放电序列 \mathbf{X} 的泊松似然来寻找最优的 \mathbf{W} .

HMM-GLM (HG): 首先我们把 GLM 和隐马尔可夫模型 (HMM) 整合在一起. 假设神经活动中的功能性连接存在 S 个状态. 对于每个时间 t , 我们引入一个离散隐变量 $z_t \in \{1, \dots, S\}$, 其转移概率为 $p(z_{t+1}|z_t) = \pi_{z_t, z_{t+1}}$, 矩阵形式为 $\mathbf{\Pi} \in [0, 1]^{S \times S}$. 给定隐状态 z_t 的条件下, 我们从 S 个权重矩阵 $\{\mathbf{W}_s\}_{s=1}^S$ 中选出 $\mathbf{W}_{z_t} \in \mathbb{R}^{n \times n}$. 那么放电 (条件释放) 模型为 $p(x_{t,n}|z_t, \mathbf{x}_1, \dots, \mathbf{x}_{t-1}) = \text{Poisson}(f_{t,n})$:

$$f_{t,n} = \sigma \left(b_n + \sum_{n'=1}^N w_{z_t, n \leftarrow n'} \cdot \left(\sum_{k=1}^K x_{t-k, n'} \phi_k \right) \right). \quad (2)$$

注意到传统的 HMM 框架假设释放概率分布是时间齐次的, 即释放模型分布不依赖于之前的观测变量. 然而这里我们像之前的 HMM-GLMs [Escola et al., 2011] 一样放宽了这一假设, 引入了对历史观测放电数据的依赖性. 下面我们介绍两种 HG 的变体, 其中加入了一个加在所有状态矩阵 \mathbf{W}_s 上的一个全局共享先验.

Gaussian HMM-GLM (GHG): 为了假设功能性连接在不同状态下应当共享一些共有的结构, 一个最简单的办法就是在权重 $w_{s, n \leftarrow n'}$ 上加带有超参数 σ^2 的高斯先验 $\mathcal{N}(w_{0, n \leftarrow n'}, \sigma^2)$, $\forall s \in \{1, \dots, S\}$. 一个有这样的高斯先验的 HG 称为 Gaussian HMM-GLM (GHG). 那些不同状态下的权重矩阵 \mathbf{W}_s 所共享的信息存在 \mathbf{W}_0 中.

One-hot HMM-GLM (OHG): One-hot HMM-GLM (OHG) 把公式 2 中的权重矩阵分解成了一个表示连接类型的邻接矩阵和一个正值的强度矩阵, 即,

$$w_{s, n \leftarrow n'} = [(-1)a_{s, n \leftarrow n', \text{inh}} + (+1)a_{s, n \leftarrow n', \text{exc}}] \cdot \tilde{w}_{s, n \leftarrow n'}, \quad \forall s \in \{1, \dots, S\}. \quad (3)$$

$\tilde{w}_{s, n \leftarrow n'} \in \mathbb{R}_+$ 状态 s 下权重的强度. 我们定义 $\mathbf{a}_{s, n \leftarrow n'} = [a_{s, n \leftarrow n', \text{inh}}, a_{s, n \leftarrow n', \text{no}}, a_{s, n \leftarrow n', \text{exc}}] \in \Delta^2$ 为状态 s 下从神经元 n' 到神经元 n 的临界 (权重类型), 对应于 {抑制 (inhibitory), 无连接 (no connection), 兴奋 (excitatory)}. $\mathbf{a}_{s, n \leftarrow n'}$ 是一个在单纯形 $\Delta^2 := \{\mathbf{a} \in [0, 1]^3 | \sum_{i=1}^3 a_i = 1\}$ 上的软独热编码向量, 所以其标量表示 $a_{s, n \leftarrow n'} = (-1)a_{s, n \leftarrow n', \text{inh}} + (+1)a_{s, n \leftarrow n', \text{exc}}$ 应当和 -1 、 0 、 $+1$ 中的某一个很接近. 它们的矩阵、张量形式分别记为 $\tilde{\mathbf{W}}_s \in \mathbb{R}_+^{N \times N}$ 和 $\mathbf{A}_s \in [0, 1]^{N \times N \times 3}$, $\forall s \in \{1, \dots, S\}$. 图 1A 给出了一个独热分解的示意图.

为了假设不同状态下的功能性连接要共享一些共有的结构以表示潜在的解剖连接组, 我们在 $\mathbf{a}_{s, n \leftarrow n'}$ 上加 Gumbel-Softmax 先验, 即 $\mathbf{a}_{s, n \leftarrow n'} \sim \text{Gumbel-Softmax}(\mathbf{a}_{0, n \leftarrow n'}, \tau)$, $\forall s \in \{1, \dots, S\}$, 更具体的写法为

$$a_{s, n \leftarrow n', \text{type}} = \frac{\exp[(\ln a_{0, n \leftarrow n', \text{type}} + g_{s, n \leftarrow n', \text{type}})/\tau]}{\sum_{\text{type}' \in \{\text{inh}, \text{no}, \text{exc}\}} \exp[(\ln a_{0, n \leftarrow n', \text{type}'} + g_{s, n \leftarrow n', \text{type}'})/\tau]}, \quad \forall \text{type} \in \{\text{inh}, \text{no}, \text{exc}\} \quad (4)$$

其中 $g_{s, n \leftarrow n', \text{type}} \stackrel{\text{i.i.d.}}{\sim} \text{Gumbel}(0, 1)$. 实际实现中, 我们通过从均匀分布 $\text{Uniform}(0, 1)$ 中采 u 并计算 $g = -\ln(-\ln(u))$ 来实线对 g 的采样. $\tau > 0$ 是温度超参数, 强制 $\mathbf{a}_{s, n \leftarrow n'}$ 为权重类型的一个软独热表示. 对每一个 $a_{s, n \leftarrow n'}$, 三个连接类型中只有一个显著的比另外两个热, 代表了这一连接的类型. $\mathbf{a}_{0, n \leftarrow n'}$

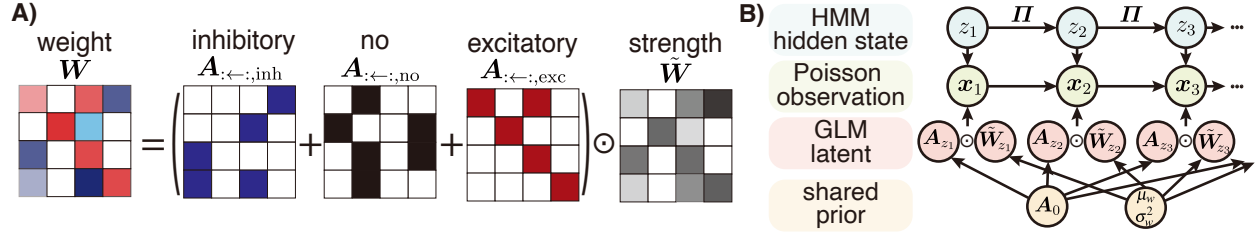


图 1: A) A descriptive schematic of the weight matrix decomposition. B) The graphical model of the one-hot HMM-GLM.

的张量形式为 $\mathbf{A}_0 \in \mathbb{R}^{N \times N \times 3}$, 为一个可学的参数矩阵, 在不同状态上加入了生物结构相似性. 最终, 如果一个连接的解剖 (突触) 类型是兴奋型, 那么其功能性连接类型也有很可能是兴奋型; 其它类型类似. Gumbel-Softmax 分布的对数密度 Jang et al. [2016], Maddison et al. [2016] 为:

$$\begin{aligned} \ln p(\mathbf{a}_{s, n \leftarrow n'} | \mathbf{a}_{0, n \leftarrow n'}) &= \ln 2 + 2\tau - 2 \ln \left(\sum_{\text{type} \in \{\text{inh}, \text{no}, \text{exc}\}} \frac{a_{0, n \leftarrow n', \text{type}}}{(a_{s, n \leftarrow n', \text{type}})^\tau} \right) \\ &+ \sum_{\text{type} \in \{\text{inh}, \text{no}, \text{exc}\}} (\ln a_{0, n \leftarrow n', \text{type}} - (\tau + 1) \ln(a_{s, n \leftarrow n', \text{type}})). \end{aligned} \quad (5)$$

通过在邻接矩阵 \mathbf{A}_s 上引入 Gumbel-Softmax 先验, \mathbf{A}_s 也变成了隐变量. 我们也可以假设 $\tilde{\mathbf{W}}_s$ 以及基准强度 b_n 也为来自某些先验分布的随机变量. 我们在 $\tilde{\mathbf{W}}_s$ 的对数上加一个高斯先验保证其为正, 在 b_n 上也加一个高斯先验, 最终就可以得到一个 OHG 的生成模型:

$$\begin{aligned} z_{t+1} | z_t &\sim \text{Categorical}(\pi_{z_t, 1}, \dots, \pi_{z_t, S}), \quad \forall t \in \{1, \dots, T\} \\ \mathbf{a}_{s, n \leftarrow n'} &\sim \text{Gumbel-Softmax}(\mathbf{a}_{0, n \leftarrow n'}, \tau), \quad \forall s \in \{1, \dots, S\}, \forall n, n' \in \{1, \dots, N\} \\ \ln \tilde{w}_{s, n \leftarrow n'} &\sim \mathcal{N}(\mu_w, \sigma_w^2), \quad \forall s \in \{1, \dots, S\}, \forall n, n' \in \{1, \dots, N\} \\ b_n &\sim \mathcal{N}(\mu_b, \sigma_b^2), \quad \forall n \in \{1, \dots, N\} \\ x_{t, n} &\sim \text{Poisson}(f_{t, n}(\mathbf{x}_1, \dots, \mathbf{x}_{t-1}, \mathbf{A}_{z_t}, \tilde{\mathbf{W}}_{z_t}, b_n)), \quad \forall t \in \{1, \dots, T\}, \forall n, n' \in \{1, \dots, N\}. \end{aligned} \quad (6)$$

这个生成模型的示意图在图 1B 中.

GHG 和 OHG 之间的关系: GHG 和 OHG 很像, 在于它们都假设了不同状态下的 \mathbf{W}_s 共享一些共有的信息 (OHG 中是 \mathbf{A}_0 而 GHG 中是 \mathbf{W}_0). 最主要的区别是 GHG 不区分邻接矩阵和强度矩阵 (公式 3). 因此, 共享的 \mathbf{W}_0 限制了两者. 而在 OHG 中, 由于有分解, \mathbf{A}_0 只在邻接矩阵上加了相似性, 而没有在强度矩阵上. 受约束的邻接矩阵 \mathbf{A}_s 以及它们的先验 \mathbf{A}_0 在某种程度上代表了潜在的解剖连接组. 而不受限制的强度矩阵 $\tilde{\mathbf{W}}_s$ 提供了充足的灵活性以捕捉不同大脑状态下功能性的连接的变化. 相较于 OHG, GHG 是一个过渡模型, 因为先验直接加在了权重矩阵上, 而这样的一个先验也表示了全局共享的一个连接信息, 只不过没有独热分解. 在实验部分, 我们会展示像 OHG 中 \mathbf{A}_0 这种生物学更合理的约束对学到有意义的参数和推断出有意义的隐变量至关重要.

3 推断

生成模型包含四个隐变量 $\{z_t, \mathbf{A}_s, \ln \tilde{\mathbf{W}}_s, b_n\}$. 这些隐变量需要用复杂的完全贝叶斯推断, 这通常非常浪费时间而且计算量巨大. 这里我们给出一种 Baum-Welch 算法来解决推断问题. 在我们这里的 Baum-Welch 算法中, 我们在 E-步中推导 z_t 的后验, 而在 M-步中我们固定 z_t 的后验分布对其它隐变量做极大似然估计. 也就是说, 在 M-步中, 我们同时优化模型的参数和除 z_t 以外的隐变量. 这里的原因是, 通过前向-后向信息传递算法计算 z_t 的后验是比较直接的, 而计算 \mathbf{A}_s 的后验分布是非常麻烦且没有闭式解的. 我们当然可以使用变分分布来估计 \mathbf{A}_s 的后验. 然而, 由于 \mathbf{A}_s 的先验是一个 Gumbel-Softmax 分布, 我们并不清楚该如何参数化其近似后验分布. 鉴于这些挑战, 我们只对 z_t 用前向-后向信息传递算法做 E-步. 在 M-步中我们同时优化模型参数 $\{\boldsymbol{\Pi}, \mathbf{A}_0\}$ 和 $\{\mathbf{A}_s, \ln \tilde{\mathbf{W}}_s, b_n\}$, 共同记为 θ . 超参数集 $\zeta = \{\mu_w, \sigma_w^2, \mu_b, \sigma_b^2, \tau\}$ 为事先给定的 (下面会详细介绍). 我们也固定了基函数 $\phi \in \mathbb{R}_+^K$.

首先, 我们给定上一步的 θ^{old} , 用前向-后向算法 (E-步) 推断隐状态. 这一步中, 为了简洁, 我们略去 θ^{old} . 我们定义 $\gamma_{z_t}(t) := p(z_t | \mathbf{X}; \theta^{\text{old}})$, $\xi_{z_{t-1}, z_t}(t) := p(z_{t-1}, z_t | \mathbf{X}; \theta^{\text{old}})$, 定义 $\alpha_{z_t}(t) := p(\mathbf{x}_1, \dots, \mathbf{x}_t, z_t)$, $\beta_{z_t}(t) := p(z_{t+1}, \dots, z_T | \mathbf{x}_1, \dots, \mathbf{x}_t, z_t)$. 之后我们就可以得到如下关系: $\gamma_{z_t}(t) = \frac{\alpha_{z_t}(t)\beta_{z_t}(t)}{p(\mathbf{X})}$, $\xi_{z_{t-1}, z_t}(t) = \frac{\beta_{z_t}(t)p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_t)\alpha_{z_{t-1}}(t-1)p(z_t | z_{t-1})}{p(\mathbf{X})}$. $\alpha_{z_t}(t)$ 和 $\beta_{z_t}(t)$ 可以通过下面的方式迭代计算:

$$\begin{cases} \alpha_{z_t}(t) = p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_t) \sum_{z_{t-1}=1}^S \alpha_{z_{t-1}}(t) p(z_t | z_{t-1}), & \alpha_{z_1}(1) = p(z_1) p(\mathbf{x}_1 | z_1) \\ \beta_{z_t}(t) = \sum_{z_{t+1}=1}^S \beta_{z_{t+1}}(t+1) p(\mathbf{x}_{t+1} | \mathbf{x}_1, \dots, \mathbf{x}_t, z_{t+1}) p(z_{t+1} | z_t), & \beta_{z_T}(T) = 1 \end{cases}$$

最终得出 $p(\mathbf{X}) = \sum_{z_T=1}^S \alpha_{z_T}(T)$. 有了 z 的后验, 我们就可以在 M-步中最大化, we can update θ in the M-step by maximizing

$$\begin{aligned} Q(\theta, \theta^{\text{old}}) &= \mathbb{E}_{p(z | \mathbf{X}; \theta^{\text{old}})} \ln p(\mathbf{X}, \mathbf{z}; \theta) = \sum p(z | \mathbf{X}; \theta^{\text{old}}) \ln p(\mathbf{X}, \mathbf{z}; \theta) \\ &= \sum_{z_1=1}^S \gamma_{z_1}(1) \ln p(z_1; \theta) + \sum_{t=2}^T \sum_{z_{t-1}=1}^S \sum_{z_t=1}^S \xi_{z_{t-1}, z_t}(t) \ln p(z_t | z_{t-1}; \theta) \\ &\quad + \sum_{t=1}^T \sum_{z_t=1}^S \gamma_{z_t}(t) \ln p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_t; \theta). \end{aligned} \tag{7}$$

以更新 θ . 更多的关于推断的细节在附录 A.1.1 中.

ζ 中有很多关键超参数需要在推断前提前给定. (1) Gumbel-Softmax 温度 τ : Gumbel-Softmax 中的 τ 通常都取自 $[0.1, 1]$. 如果 τ 太大, 那么软独热编码就会过于松弛; 如果 τ 太小, 就会出现数值问题. 在我们的模型中, τ 是用于强制软独热码靠近单纯形的一个角落. 所以我们试了 $\tau \in \{0.1, 0.2, 0.5\}$, 然后发现 one-hot HMM-GLM 的结果对 τ 在这个范围内的选取不是很敏感. 鉴于 τ 的选择对不同的数据集也不敏感, 我们固定 $\tau = 0.2$, 这这也是一个通常采用的适中选择. (2) 生成参数 $\{\mu_w, \sigma_w^2, \mu_b, \sigma_b^2\}$: 我们选择 $\mu_w = -5, \sigma_w = 2$ and $\mu_b = 0, \sigma_b = 2$, 因为这样就是一个对 GLM 中的权重强度和基准强度比较平凡的先验, 使得推断对所有的数据集一视同仁.

4 实验评估

用于比较的其它模型. 我们会将我们的模型和其它最先进的基线模型在一个模拟数据集和两个真实神经数据集上进行比较:

- **GLM** [Pillow et al., 2008]: 最原始的用于研究神经交互的模型, 没有多状态的假设.
- **HMM Corr** [Engel et al., 2016]: 一个用于研究神经放电序列状态切换的 HMM. 由于这个模型只给出了隐状态而没有神经连接, 我们使用基于相关性的方法, 即交互图 (cross-correlogram, CCG) 来求每个状态下的连.
- **HMM Bern** [Ashwood et al., 2022]: 和 HMM Corr 类似, 但是用伯努利而不是播送分布来建模每个时间桶内的放电数.
- **HG** [Escola et al., 2011]: 朴素 HMM-GLM (HG), 其为现有唯一的可以同时推断隐状态并学习神经连接的模型.
- **GHG** (我们的方法): Gaussian HMM-GLM 缩写为 GHG.
- **OHG** (我们的方法): One-hot HMM-GLM 缩写为 OHG.
- **HG-L1** 和 **GHG-L1**: 鉴于独热机制隐式地为权重矩阵加入了稀疏性, 可能大家会怀疑是不是 OHG 的优势仅仅来自于加入的稀疏性. 为解决这一顾虑, 我们又加了两个模型: 一个是给 HG 的权重矩阵加 L1 罚项 (HG-L1), 另一个是给 GHG 的权重矩阵加 L1 罚项 (GHG-L1). L1 罚项的系数通过交叉验证确定.

指标. 我们用下面这些指标来报告不同模型的性能:

- **LL**. 测试集上的对数似然. 更好的模型应当有更强的能力去预测未来的放电事件. 注意这是真实数据集上的唯一可用指标, 因为现实数据集中一般没有真实状态和真实的神经连接可供比较.
- **状态准确率**. 所有时间的推断状态的平均准确率. 这只适用于模拟数据集, 因为我们知道真实的隐状态.
- **权重误差**. 学到的每个状态下的权重矩阵的误差. 注意 HMM Corr 和 HMM Bern 没有权重误差. 由于它们的权重矩阵是用 CCG 得到的, 因此不能和 GLM 的权重矩阵进行比较. 这个指标也只适用于模拟数据集.
- **邻接矩阵准确率**. 学到的所有状态下的邻接矩阵的平衡准确率. 这只适用于模拟数据集. 如果模型不显式的学邻接矩阵, 就用

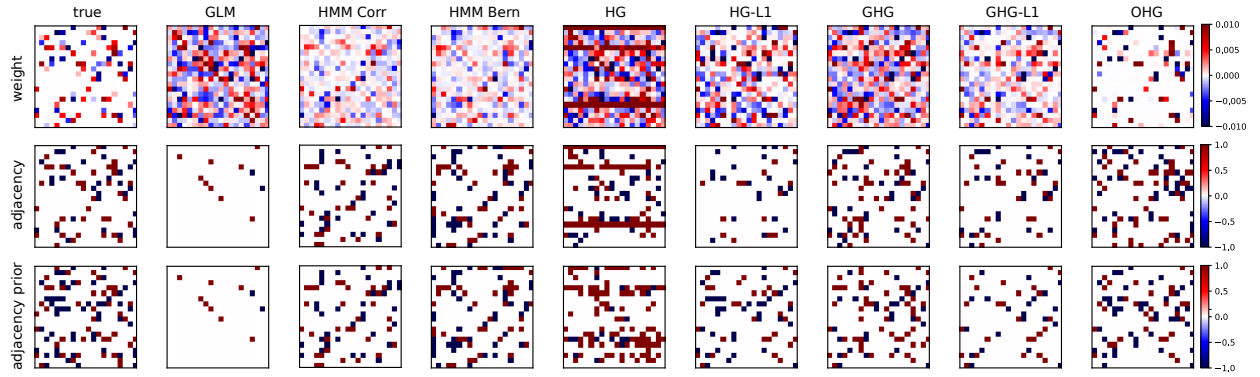
$$\mathbf{a}_{s,n \leftarrow n'} = \begin{cases} \left(0, 1 - \frac{w_{s,n \leftarrow n'}}{\max_{s,n,n'} w_{s,n \leftarrow n'}}, \frac{w_{s,n \leftarrow n'}}{\max_{s,n,n'} w_{s,n \leftarrow n'}}\right), & w_{s,n \leftarrow n'} \geq 0 \\ \left(\frac{w_{s,n \leftarrow n'}}{\min_{s,n,n'} w_{s,n \leftarrow n'}}, 1 - \frac{w_{s,n \leftarrow n'}}{\min_{s,n,n'} w_{s,n \leftarrow n'}}, 0\right), & w_{s,n \leftarrow n'} < 0 \end{cases} \quad (8)$$

从学到的权重矩阵求得邻接矩阵. 用公式 8 是因为这个办法合理且不需要设置阈值. 我们当然也可以用给定的阈值来求邻接矩阵, 但是邻接矩阵的准确率就会受到阈值选取的影响 (见附录 A.2). 在真实神经数据集上, 我们没有真实的邻接矩阵, 因此也无法用准确率来选择最优的阈值. 这更说明了 OHG 显式给出的邻接矩阵是一个非常简洁的表达, 不需要提前设阈值但却能给出满意的估计.

- **邻接矩阵先验准确率**. 这只适用于模拟数据集. 除了 OHG, 我们先把所有状态下的权重矩阵取平均, 然后用公式 8 求邻接矩阵先验.

表 1: 合成数据集上五个指标的量化结.

method	LL \uparrow	state acc \uparrow	weight error \downarrow	adj acc \uparrow	adj prior acc \uparrow
GLM	-8.43(± 0.18)	nan(\pm nan)	24.71(± 0.19)	43.12(± 0.46)	44.81(± 0.61)
HMM Corr	-22.53(± 0.64)	42.84(± 1.47)	nan(\pm nan)	34.04(± 0.12)	15.45(± 2.49)
HMM Bern	-5.68(± 0.23)	87.95(± 0.93)	nan(\pm nan)	36.25(± 0.25)	40.70(± 1.53)
HG	-5.49(± 0.58)	37.73(± 2.80)	109.67(± 2.63)	34.17(± 0.08)	40.91(± 0.48)
HG-L1	9.14(± 0.18)	91.60(± 0.96)	23.14(± 0.08)	37.47(± 0.18)	48.44(± 0.57)
GHG	8.58(± 0.19)	91.80(± 0.92)	21.54(± 0.15)	42.53(± 0.22)	48.93(± 0.54)
GHG-L1	9.77(± 0.20)	92.08(± 0.89)	14.16(± 0.07)	41.08(± 0.22)	46.98(± 0.60)
OHG	14.64 (± 0.23)	92.75 (± 0.87)	10.99 (± 0.21)	73.90 (± 0.52)	80.60 (± 0.59)

图 2: 所有模型在合成数据集的一个试次里学到的状态 2 (共 $S = 5$ 个状态) 的权重矩阵 \mathbf{W}_2 (上)、邻接矩阵 \mathbf{A}_2 (中) 和邻接先验 \mathbf{A}_0 (下).

4.1 应用于模拟数据集

数据集. 我们首先在一个 5-状态-20-神经元的合成数据集上比较不同的模型. 共有 10 个试次. 每个试次生成 20 个长度 $T = 5000$ 的放电序列. 每个放电序列由公式 6 的生成模型生成, 其中 $\pi_{s,s'} = 0.005 + 0.975 \cdot \mathbb{1}[s = s']$, $\tau = 0$, $\mu_w = -5$, $\sigma_w^2 = 1.5$, $\mu_b = 0$, $\sigma_b^2 = 0.0008$. $\mathbf{a}_{0,n' \leftarrow n}$ 采自 $\text{Dirichlet}(0.1, 0.8, 0.1)$, $\forall n, n' \in \{1, \dots, 20\}$. 注意当 $\tau = 0$ 时, 所有的 \mathbf{A}_s 都是独立同分布地采自 \mathbf{A}_0 . 这等价于直接从分类分布采 $a_{s,n \leftarrow n'}$, 即 $a_{s,n' \leftarrow n} \sim \text{Categorical}(a_{0,n' \leftarrow n})$, $\forall s \in \{1, \dots, 5\}$, $\forall n, n' \in \{1, \dots, 20\}$. 这样也是在生成模型和推断模型 (公式 6) 之间引入了一些差异性. 对每一个试次, 我们都在前 10 个序列上训练, 在后 10 个序列上测试.

表 1 的量化结果表明 OHG 在所有五个指标下都是最好的. 两个 GHG 是第二好的, 因为两个 GHG 都在权重矩阵上加了全局先验. 然而, 它们仍然不如 OHG, 说明了独热机制对 OHG 的卓越表现十分重要.

下面我们分析不同模型学到的神经连接 (图 2). 尽管共有 $S = 5$ 个不同的状态, 单状态 GLM 只能捕获五个状态的一个“平均”估计. 对于 HMM Corr 和 HMM Bern 来说, 学习过程被分成了两部,

表 2: PFC-6 数据集上, 不同模型、不同状态数在测试集上的对数似然. 单状态 GLM 的结果是 $-36.35(\pm 0.00)$.

method	2 states	3 states	4 states	5 states
HMM Corr	-37.11(± 0.00)	-36.60(± 0.00)	-36.53(± 0.00)	-36.68(± 0.00)
HMM Bern	-36.89(± 0.00)	-36.57(± 0.00)	-36.38(± 0.00)	-36.38(± 0.00)
HG	-37.30(± 0.05)	-37.61(± 0.17)	-37.22(± 0.14)	-36.98(± 0.19)
HG-L1	-36.91(± 0.01)	-36.90(± 0.02)	-36.73(± 0.09)	-36.63(± 0.13)
GHG	-37.17(± 0.00)	-37.11(± 0.01)	-37.12(± 0.00)	-37.11(± 0.00)
GHG-L1	-36.94(± 0.00)	-36.88(± 0.00)	-36.83(± 0.00)	-36.77(± 0.00)
OHG	-35.92 (± 0.02)	-35.79 (± 0.02)	-35.77 (± 0.03)	-35.71 (± 0.03)

推断隐状态和估计每个状态下的神经连接. 尽管 HMM Bern 推断的隐状态比较好, 每个状态下的邻接矩阵以及邻接矩阵先验估计的不是很好. 对于 HG 来说, 它的结果比较差主要是因为状态转移矩阵的估计不好 (附录 7 中的图 7). 这导致了错误的权重和邻接矩阵估计. 将 HG 和 GHG 以及 OHG 对比可以发现, 给不同状态上加一个限制 (先验) 是得到稳定结果的必要做法, 因为不同状态之间的共享信息可以防止推断的状态和不同状态下的权重走向极端结果或掉入局部最优. 加 L1 罚项可以抑制一些权重噪声, 但仍然对估计每个状态下的邻接矩阵和邻接先验没有帮助, 因为 L1 对区分弱连接和没有连接没有帮助. GHG 和 OHG 之间的主要区别就是它们的权重和邻接矩阵估计. GHG 仍然由很多非零的噪声权重. 而由于 OHG 带有独热机制, 网络的稀疏性很容易就学到了, 并且无连接的权重都被成功地被抑制住了. 这便是 OHG 权重误差低、邻接矩阵准确率高的原因 (OHG 学到的权重矩阵、邻接矩阵和临界先验都和真实参数匹配得最好, 图 2).

4.2 应用于电生理数据

4.2.1 情境任务中的前额叶皮质

数据集. 我们首先把不同模型应用在一个前额叶皮质 (PFC-6) 数据集上 [Peyrache et al., 2018, 2009]¹. 神经放电序列收集于大鼠学习行为情景任务时. 记录数据时, 每个试次大概持续 4 秒, 然后有大约 24 秒的一个休息间隔. 用于训练和测试的放电序列是从一个很长的实验段 (session) 中分割出来的. 每个序列开始于行为开始前 5 秒, 结束于行为开始后 10 秒. 所以每个序列对应一个试次. 我们用 $\frac{2}{3}$ 的序列所谓训练集, 剩下 $\frac{1}{3}$ 为测试集. 每个试次的神经放电序列被划分为 750 个尺寸为 20 ms 的时间桶. 由于我们不知道真实的状态数, 我们尝试 $S \in \{2, 3, 4, 5\}$.

表 2 表明 OHG 的测试对数似然在所有状态数下都比其它模型好. 图 3 展示了不同模型学到的权重和邻接矩阵的一个例子. 对于 HG, 学到的权重矩阵非常稠密且噪声较大, 导致了测试集上较差的对数似然. 对于 GHG, 权重没那么稠密了, 但结果还是噪声很大. 加 L1 罚项有助于 HG 和 GHG 减少噪声, 但对区分弱连接和无连接没有帮助. 用 OHG 可以得到一个很清晰的强度-邻接分解和一个邻接先

¹<https://crcns.org/data-sets/pfc/pfc-6>

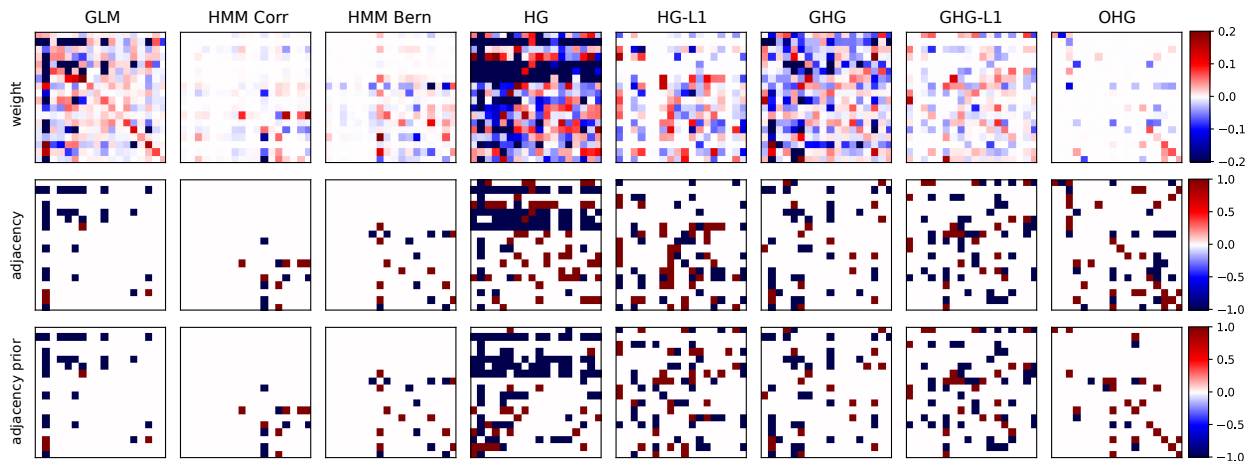


图 3: 所有模型在 PFC-6 数据集的一个试次里学到的状态 2 (共 $S=45$ 个状态) 的权重矩阵 \mathbf{W}_4 (上)、邻接矩阵 \mathbf{A}_4 (中) 和邻接先验 \mathbf{A}_0 (下)。

验。邻接先验给出的全局限制的确像解剖连接组一样影响了功能性连接，这提高了 OHG 的测试对数似然。注意 GLM 其实也得到了比较不错的结果，仅仅比不上 OHG。这说明在这种真实的情况下，不同状态下的功能性连接的确共享一个全局的静态结构（可能反映的是解剖连接组），这种影响是强于不同状态下的功能性差异的，所以要认真处理。对于 $S > 5$ ，性能不再有明显提升而是趋于平缓（附录 A.4 中的图 10）。

尽管没有真实的隐状态可供比较，我们仍然可以结合行为数据分析不同模型推断出的隐状态。以 4 个状态为例。在图 4 中，我们画出了一个正确试次的隐状态推断结果（图 4A）和一个错误试次的隐状态推断结果（图 4B）。我们还在右侧画出了大鼠运动的轨迹，并用 OHG 推断出的隐状态对其进行着色。和之前的结论差不多，HG 还是给出了比较乱的隐状态预测结果，没什么可解释性。尽管假设了 $S = 4$ 个隐状态，GHG 只得到了两个有效的隐状态。从状态 4 到状态 3 的转移一般发生在大鼠在错误目标位置回头的时候。然而，OHG 却能找到四个可解释的有效隐状态。每个试次开始之前，大鼠回到 Y 形迷宫的起点，对应状态 4。然后大鼠转过头去向 Y 形迷宫的分岔点爬去，对应状态 3。在做决策后，大鼠在 Y 形迷宫的一个分支里向顶端爬，对应状态 2。如果大鼠去到了正确的目标位，就在目标位得到了奖励，大鼠就会进入状态 4 并保持一段时间。但如果大鼠去到了错误的目标位，就没有奖励了，大鼠就会立即回来，对应状态 1。这些状态解释反映在图 4 的着色运动轨迹上。注意这样的状态模式并不是筛数据挑出来的。我们在其它的正确、错误试次里也发现了类似的状态切换，见附录 A.4 中的图 8。我们还用推断的隐状态做了一个任务信息解码，进一步验证了这一状态解释的合理性。

4.2.2 胡须挥动时的桶状皮质

数据集 我们下面把不同的模型应用到小鼠躯体感觉桶状皮质的电极记录数据上。小鼠们在做一个形状区分任务 [Rodgers et al., 2021, Rodgers, 2022, Nogueira et al., 2023] (Fig. 5A)。小鼠们训练到能够用它们的胡须来区分形状的凹凸。具体来说，小鼠们要主动用胡须去触碰物体。高速摄像机会计录下胡须的运动，以帮助我们分析胡须是如何主动运动来感觉周围环境的。数据包含来自 5 只小鼠的 27 个实

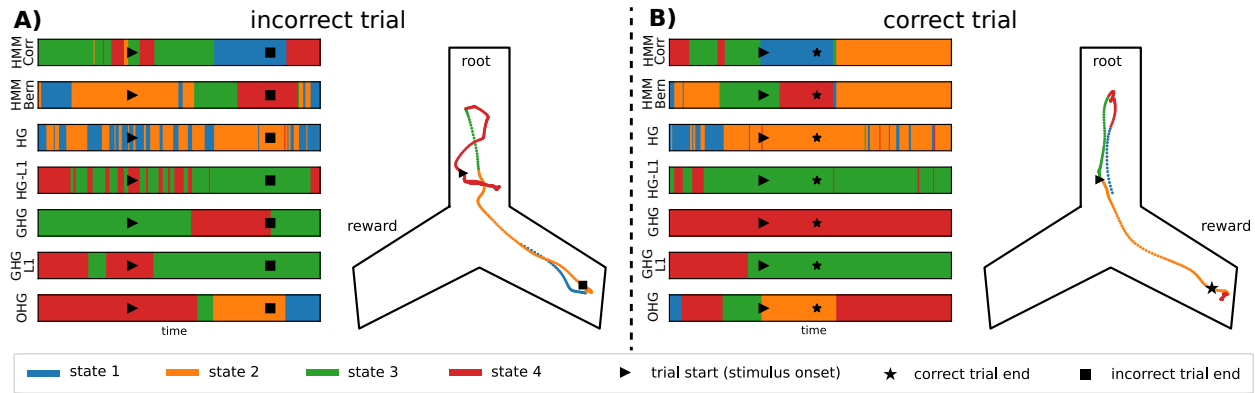


图 4: 错误试次 (A) 和正确试次 (B) 里不同模型 (包括 HG、GHG 和 OHG) 推断的隐状态. 右侧大鼠的运动轨迹是根据 OHG 推断出的隐状态进行着色的.

验段. 记录到的神经元个数在 10 到 44 个之间不等. 每个试次取 6 秒到放电序列, 用 3 ms 宽的时间桶离散化. 每个实验段取前 30 个试次, 其中 10 个随机选取的试次作为测试集, 剩余 20 个试次用于训练模型.

鉴于我们不知道行为本身有多少状态, 我们还是对桶状皮质数据尝试不同的隐状态数, 即 $S = \{2, 3, 4, 5\}$. 桶状皮质数据集上不同模型的对数似然的变化趋势和 PFC-6 数据集类似. OHG 总是有最高的对数似然, GHG 的对数似然在不同隐状态总数下都比其它基线模型强 (见图 5B).

图 5C 画出了每个模型的胡须挥动位置、接触和推断的隐状态. 我们以 $S = 2$ 个隐状态为例可视化. 尽管 OHG 的似然在 S 增加到 5 点过程中不断增大, 但对于 $S > 2$, 有很多实验段都包含很多空状态, 且不同状态之间的差别变得很小. 3-5 个状态的结果在附录 A.5 中. 当假设有两个状态时, 通常 GHG 和 OHG 都会发现有一个状态对应挥动胡须接触物体的事件. 而 HG 的状态切换过于频繁.

尽管 GHG 和 OHG 都能预测到胡须挥动事件, 但这一状态的持续时间不同 (图 5C). OHG 推断出了一个稳定的状态, 在挥动循环 (whisking cycles) 时持续 1 s, 而 GHG 推断的状态以快速地在切换 (< 0.1 s). OHG 因此能更好的捕捉到持续的挥动循环 (Deschênes et al. [2012], Rodgers et al. [2021]). 图 5D 进一步展示了每个模型在图 5C 的实验段上学到的权重矩阵、邻接矩阵和邻接先验. 和 PFC-6 数据集一样, OHG 能够学到稀疏且清晰的权重矩阵, 而 HG 和 GHG 学到的就比较稠密且噪声比较大.

我们进一步用假设检验的方法看 GHG 和 OHG 推断的状态是否和主动挥动胡须事件是相关的. 我们先计算出每一个状态下胡须接触的频率, 并 χ^2 检验在假设状态和胡须接触无关联的情况下胡须接触的频率. 在 11 个实验段中, 三个模型推断出的状态频率都是不偏的 (最不常出现的状态也至少在 5% 的时间内出现了). HG 中有 6 个实验段 (54%) 拒绝了零假设 ($p < 0.001$), 而 GHG 和 OHG 中有 8 个实验段 (54%) 拒绝了零假设. 此外, 我们还对所有实验段求每个状态的权重矩阵 \mathbf{W} 的元素和. 我们发现 OHG 在胡须挥动状态下, 权重矩阵元素和有显著的提升 ($p = 0.008$, 双侧 Wilcoxon 秩和检验), 而 HG 和 GHG 则没有这样的结果 ($p > 0.1$). 这说明了 OHG 能够检测到和行为状态切换相关联的神经元功能性交互的状态切换.

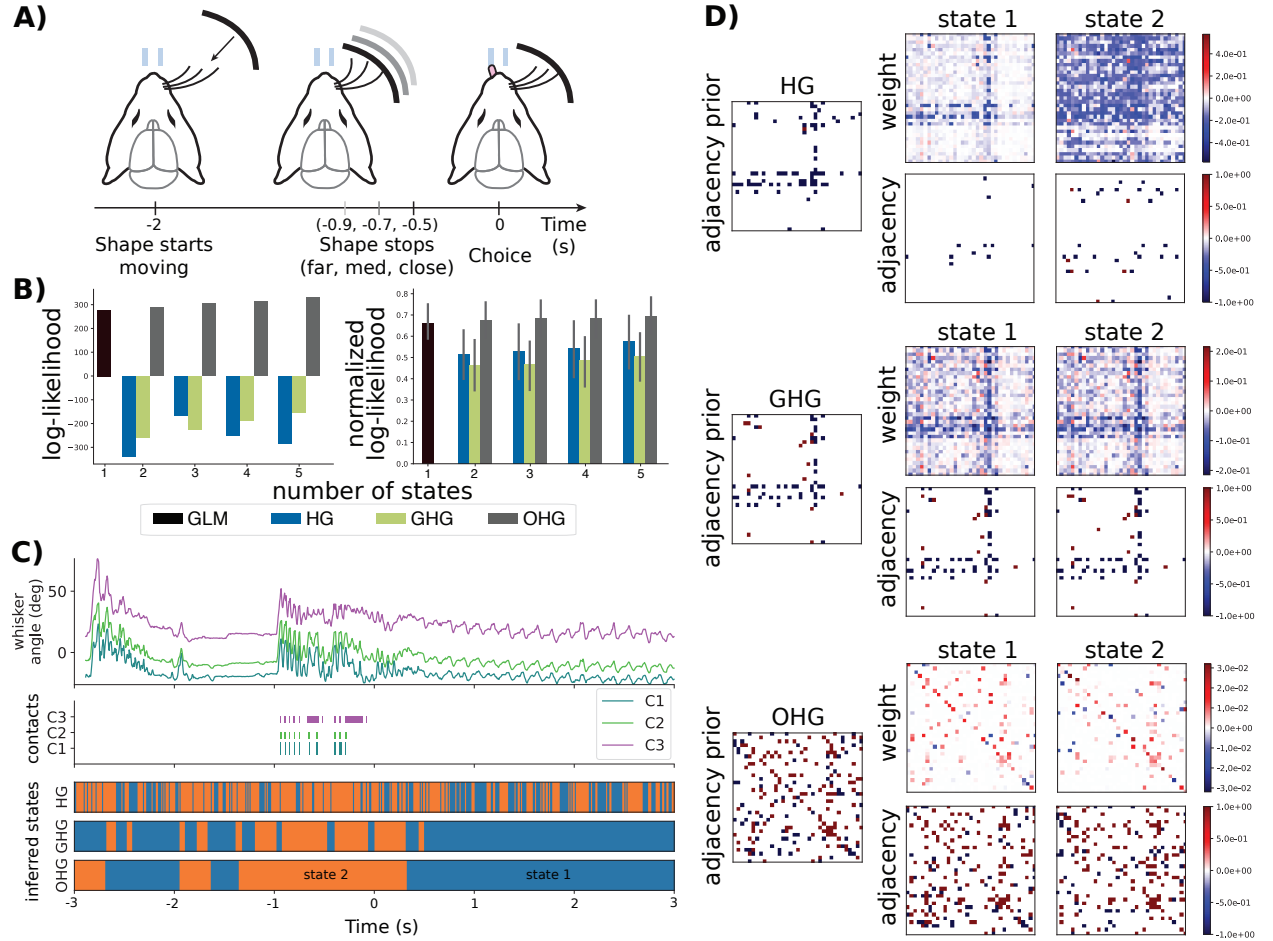


图 5: A) 胡须挥动实验设置 (改编自 Nogueira et al. [2023]). B) 不同模型的测试 (以及规范化后的) 对数似然. 原始的对数似然 (左) 没有误差线, 因为不同实验段之间的差异性极大. 为了去除这一影响, 我们又求了规范化后的对数似然 (右). C) 某一试次中的胡须位置 (上)、胡须接触物体 (中) 和推断的隐状态 (下). $t = 0$ s 对应响应窗口打开 (这之后, 小鼠舔舐的方向被视作做出的决策), 刺激大概出现在 $t = -1$ s. D) C 所示的实验段的三个模型的权重、邻接矩阵和邻接矩阵先验.

5 结论

我们搞出了一个新的 one-hot HMM-GLM (OHG) 来估计多状态神经系统中随时间变化的功能性连接网络. 新提出的 OHG 把传统的 GLM 权重矩阵分解成了一个离散的代表连接类型的邻接矩阵和一个正值的强度矩阵. 在构造 OHG 时, 我们加了一个 Gumbel-Softmax 先验在所有状态的邻接矩阵上, 要求这些邻接矩阵要共享一些信息. 这样受约束的邻接矩阵和它们的先验能够给我们提供一些解剖连接组的信息, 因此我们能从中发现更像神经元之间的真实物理交互. 不怎么受限的强度矩阵可以在不同状态下不受先验约束地自由变化, 因此为我们提供了足够高的自由度以捕捉不同大脑状态之间的功能变化. 鉴于上述优点, 我们认为 OHG 是生物学意义上更可信的. Gaussian HMM-GLM (GHG) 是

一个过渡模型，其共享的先验直接加在了权重矩阵上，而没有（独热）强度-邻接分解，证实了这一分解对成功推断多状态是至关重要的。实验表明，在和其它模型比较时，OHG 得到了更好的连接网络和推断的状态。OHG 不仅成功的在模拟数据上恢复了真实的连接网络，还在一个 PFC 数据集和一个桶状皮质数据集上达到了最好的测试集预测似然。在这两个真实数据集上，OHG 发现了更可解释的连接网络和隐状态序列。

参考文献

- Xiaoxuan Jia, Joshua H Siegle, Séverine Durand, Gregory Heller, Tamina K Ramirez, Christof Koch, and Shawn R Olsen. Multi-regional module-based signal transmission in mouse visual cortex. *Neuron*, 110(9):1585–1598, 2022.
- Conor Houghton. Calculating the mutual information between two spike trains. *Neural Computation*, 31(2):330–343, 2019.
- Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.
- Thomas Schreiber. Measuring information transfer. *Physical Review Letters*, 85(2):461, 2000.
- Scott Linderman, Ryan P Adams, and Jonathan W Pillow. Bayesian latent structure discovery from multi-neuron recordings. *Advances in Neural Information Processing Systems*, 29, 2016.
- Scott Linderman, Matthew Johnson, Andrew Miller, Ryan Adams, David Blei, and Liam Paninski. Bayesian learning and inference in recurrent switching linear dynamical systems. In *Artificial Intelligence and Statistics*, pages 914–922. PMLR, 2017.
- Joshua Glaser, Matthew Whiteway, John P Cunningham, Liam Paninski, and Scott Linderman. Recurrent switching dynamical systems models for multiple interacting neural populations. *Advances in neural information processing systems*, 33:14867–14878, 2020.
- Weihan Li, Yu Qi, and Gang Pan. Online neural sequence detection with hierarchical dirichlet point process. *Advances in Neural Information Processing Systems*, 35:6654–6665, 2022.
- S Murray Sherman. Tonic and burst firing: dual modes of thalamocortical relay. *Trends in neurosciences*, 24(2):122–126, 2001.
- Bilal Haider, Alvaro Duque, Andrea R Hasenstaub, Yuguo Yu, and David A McCormick. Enhancement of visual responsiveness by spontaneous local network activity in vivo. *Journal of Neurophysiology*, 97(6):4186–4202, 2007.
- Jeffrey Anderson, Ilan Lampl, Iva Reichova, Matteo Carandini, and David Ferster. Stimulus dependence of two-state fluctuations of membrane potential in cat visual cortex. *Nature Neuroscience*, 3(6):617–621, 2000.

- Maria V Sanchez-Vives and David A McCormick. Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nature Neuroscience*, 3(10):1027–1034, 2000.
- Sean Escola, Alfredo Fontanini, Don Katz, and Liam Paninski. Hidden markov models for the stimulus-response relationships of multistate neural systems. *Neural computation*, 23(5):1071–1132, 2011.
- Namrata Nadagouda and Mark A Davenport. Switched hawkes processes. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5170–5174. IEEE, 2021.
- Feng Zhou, Quyu Kong, Yixuan Zhang, Cheng Feng, and Jun Zhu. Nonlinear hawkes processes in time-varying system. *arXiv preprint arXiv:2106.04844*, 2021.
- Maxime Morariu-Patrichi and Mikko S Pakkanen. State-dependent hawkes processes and their application to limit order book modelling. *Quantitative Finance*, 22(3):563–583, 2022.
- Erhan Genç, Marieke Louise Schölvinck, Johanna Bergmann, Wolf Singer, and Axel Kohler. Functional connectivity patterns of visual cortex reflect its anatomical organization. *Cerebral Cortex*, 26(9):3719–3731, 2016.
- Joshua H Siegle, Xiaoxuan Jia, Séverine Durand, Sam Gale, Corbett Bennett, Nile Graddis, Gregory Heller, Tamina K Ramirez, Hannah Choi, Jennifer A Luviano, et al. Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*, 592(7852):86–92, 2021.
- Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*, 2016.
- Chris J Maddison, Andriy Mnih, and Yee Whye Teh. The concrete distribution: A continuous relaxation of discrete random variables. *arXiv preprint arXiv:1611.00712*, 2016.
- Jonathan W Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M Litke, EJ Chichilnisky, and Eero P Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.
- Tatiana A Engel, Nicholas A Steinmetz, Marc A Gieselmann, Alexander Thiele, Tirin Moore, and Kwabena Boahen. Selective modulation of cortical state during spatial attention. *Science*, 354(6316):1140–1144, 2016.
- Zoe C Ashwood, Nicholas A Roy, Iris R Stone, International Brain Laboratory, Anne E Urai, Anne K Churchland, Alexandre Pouget, and Jonathan W Pillow. Mice alternate between discrete strategies during perceptual decision-making. *Nature Neuroscience*, 25(2):201–212, 2022.
- A Peyrache, M Khamassi, K Benchenane, SI Wiener, and F Battaglia. Activity of neurons in rat medial prefrontal cortex during learning and sleep. 2018.

- A Peyrache, M Khamassi, K Benchenane, SI Wiener, and F Battaglia. Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nature Neuroscience*, 12(7):919–926, 2009.
- CC Rodgers, R Nogueira, B Christina Pil, EA Greeman, JM Park, YK Hong, S Fusi, and RM Bruno. Sensorimotor strategies and neuronal representations for shape discrimination. *Neuron*, 109:2308–2325, 2021.
- Chris C Rodgers. A detailed behavioral, videographic, and neural dataset on object recognition in mice. *Scientific Data*, 9(1):620, 2022.
- Ramon Nogueira, Chris C Rodgers, Randy M Bruno, and Stefano Fusi. The geometry of cortical representations of touch in rodents. *Nature Neuroscience*, pages 1–12, 2023.
- Martin Deschênes, Jeffrey Moore, and David Kleinfeld. Sniffing and whisking in rodents. *Current Opinion in Neurobiology*, 22(2):243–250, 2012.
- Christopher M Bishop and Nasser M Nasrabadi. *Pattern Recognition and Machine Learning*, volume 4. Springer, 2006.
- Guy Ackerson and K Fu. On state estimation in switching environments. *IEEE transactions on automatic control*, 15(1):10–17, 1970.
- Chaw-Bing Chang and Michael Athans. State estimation for discrete systems with switching parameters. *IEEE Transactions on Aerospace and Electronic Systems*, (3):418–425, 1978.
- James D Hamilton. Analysis of time series subject to changes in regime. *Journal of econometrics*, 45(1-2):39–70, 1990.
- Zoubin Ghahramani and Geoffrey E Hinton. Switching state-space models. Technical report, Technical Report CRG-TR-96-3 DRAFT, Dept. of Computer Science, University of ..., 1996.
- Zoubin Ghahramani and Geoffrey E Hinton. Variational learning for switching state-space models. *Neural computation*, 12(4):831–864, 2000.
- Kevin P Murphy. Switching kalman filters. 1998.
- Emily Fox, Erik Sudderth, Michael Jordan, and Alan Willsky. Nonparametric bayesian learning of switching linear dynamical systems. *Advances in neural information processing systems*, 21, 2008.

附录 A

A.1 HMM-GLM 的推断和学习算法

A.1.1 前向后向推断

这一部分我们计算在给定旧参数时，隐变量的后验分布，对应于 EM 算法的 E-步。定义

$$\begin{cases} \gamma_{z_t}(t) := p(z_t | \mathbf{X}; \theta^{\text{old}}) \\ \xi_{z_{t-1}, z_t}(t) := p(z_{t-1}, z_t | \mathbf{X}; \theta^{\text{old}}) \end{cases}, \quad (9)$$

其中 z_t 索引 S 个不同状态中的某一个。

定义

$$\begin{cases} \alpha_{z_t}(t) := p(\mathbf{x}_1, \dots, \mathbf{x}_t, z_t) \\ \beta_{z_t}(t) := p(z_{t+1}, \dots, z_T | \mathbf{x}_1, \dots, \mathbf{x}_t, z_t) \end{cases}, \quad (10)$$

有

$$\begin{aligned} \gamma_{z_t}(t) &= \frac{p(\mathbf{X}, z_t)}{p(\mathbf{X})} \\ &= \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_t, z_t) p(\mathbf{x}_{t+1}, \dots, \mathbf{x}_T | \mathbf{x}_1, \dots, \mathbf{x}_t, z_t)}{p(\mathbf{X})} \\ &= \frac{\alpha_{z_t}(t) \beta_{z_t}(t)}{p(\mathbf{X})}. \end{aligned} \quad (11)$$

$$\begin{aligned} \alpha_{z_t}(t) &= p(\mathbf{x}_1, \dots, \mathbf{x}_t, z_t) \\ &= p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_t) p(\mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_t) \\ &= p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_t) \sum_{z_{t-1}} p(\mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_{t-1}, z_t) \\ &= p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_t) \sum_{z_{t-1}} p(\mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_t | z_{t-1}) p(z_{t-1}) \\ &= p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_t) \sum_{z_{t-1}} p(\mathbf{x}_1, \dots, \mathbf{x}_{t-1} | z_t, z_{t-1}) p(z_t | z_{t-1}) p(z_{t-1}) \\ &= p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_t) \sum_{z_{t-1}} p(\mathbf{x}_1, \dots, \mathbf{x}_{t-1} | z_{t-1}) p(z_{t-1}) p(z_t | z_{t-1}) \\ &= p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_t) \sum_{z_{t-1}} \alpha_{z_{t-1}}(t) p(z_t | z_{t-1}) \end{aligned} \quad (12)$$

以及初始条件

$$\alpha_{z_1}(1) = p(\mathbf{x}_1, z_1) = p(z_1) p(\mathbf{x}_1 | z_1). \quad (13)$$

$$\begin{aligned}
\beta_{z_t}(t) &= p(\mathbf{x}_{t+1}, \dots, \mathbf{x}_T | \mathbf{x}_1, \dots, \mathbf{x}_t, z_t) \\
&= \sum_{z_{t+1}} p(\mathbf{x}_{t+1}, \dots, \mathbf{x}_T, z_{t+1} | \mathbf{x}_1, \dots, \mathbf{x}_t, z_t) \\
&= \sum_{z_{t+1}} p(\mathbf{x}_{t+1}, \dots, \mathbf{x}_T | \mathbf{x}_1, \dots, \mathbf{x}_t, z_{t+1}, z_t) p(z_{t+1} | \mathbf{x}_1, \dots, \mathbf{x}_t, z_t) \\
&= \sum_{z_{t+1}} p(\mathbf{x}_{t+1}, \dots, \mathbf{x}_T | \mathbf{x}_1, \dots, \mathbf{x}_t, z_{t+1}) p(z_{t+1} | z_t) \\
&= \sum_{z_{t+1}} p(\mathbf{x}_{t+2}, \dots, \mathbf{x}_T | \mathbf{x}_1, \dots, \mathbf{x}_t, \mathbf{x}_{t+1}, z_{t+1}) p(\mathbf{x}_{t+1} | \mathbf{x}_1, \dots, \mathbf{x}_t, z_{t+1}) p(z_{t+1} | z_t) \\
&= \sum_{z_{t+1}} \beta_{z_{t+1}}(t+1) p(\mathbf{x}_{t+1} | \mathbf{x}_1, \dots, \mathbf{x}_t, z_{t+1}) p(z_{t+1} | z_t)
\end{aligned} \tag{14}$$

以及初始条件 $\beta_{z_T}(T) = 1$, 因为公式 11 有

$$\gamma_{z_T}(T) = p(z_T | \mathbf{X}) = \frac{p(\mathbf{X}, z_T) \beta_{z_T}(T)}{p(\mathbf{X})} \equiv \frac{p(\mathbf{X}, z_T)}{p(\mathbf{X})}. \tag{15}$$

如果我们对公式 11 两边求和,

$$1 = \sum_{z_t} \gamma_{z_t}(t) = \frac{\sum_{z_t} \alpha_{z_t}(t) \beta_{z_t}(t)}{p(\mathbf{X})} \implies p(\mathbf{X}) = \sum_{z_t} \alpha_{z_t}(t) \beta_{z_t}(t), \tag{16}$$

也可以直接用 $t = T$ 时的 $p(\mathbf{X}) = \sum_{z_T} \alpha_{z_T}(T)$.

$$\begin{aligned}
\xi_{z_{t-1}, z_t}(t) &= p(z_{t-1}, z_t | \mathbf{X}) \\
&= \frac{p(\mathbf{X} | z_{t-1}, z_t) p(z_{t-1}, z_t)}{p(\mathbf{X})} \\
&= \frac{p(\mathbf{X} | z_{t-1}, z_t) p(z_t | z_{t-1}) p(z_{t-1})}{p(\mathbf{X})} \\
&= \frac{p(\mathbf{x}_t, \dots, \mathbf{x}_T | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_{t-1}, z_t) p(\mathbf{x}_1, \dots, \mathbf{x}_{t-1} | z_{t-1}, z_t) p(z_t | z_{t-1}) p(z_{t-1})}{p(\mathbf{X})} \\
&= \frac{p(\mathbf{x}_{t+1}, \dots, \mathbf{x}_T | \mathbf{x}_1, \dots, \mathbf{x}_t, z_{t-1}, z_t) p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_{t-1}, z_t) \alpha_{z_{t-1}}(t-1) p(z_t | z_{t-1})}{p(\mathbf{X})} \\
&= \frac{p(\mathbf{x}_{t+1}, \dots, \mathbf{x}_T | \mathbf{x}_1, \dots, \mathbf{x}_t, z_t) p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_t) \alpha_{z_{t-1}}(t-1) p(z_t | z_{t-1})}{p(\mathbf{X})} \\
&= \frac{\beta_{z_t}(t) p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_t) \alpha_{z_{t-1}}(t-1) p(z_t | z_{t-1})}{p(\mathbf{X})}.
\end{aligned} \tag{17}$$

A.1.2 Baum-Welch 算法

现在我们已经有了后验, 就可以进行 EM 算法的 M-步了.

$$p(\mathbf{X}, \mathbf{z}; \theta) = p(z_1; \theta) \left[\prod_{t=2}^T p(z_t | z_{t-1}; \theta) \right] \prod_{t=1}^T p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_t; \theta). \tag{18}$$

$$\ln p(\mathbf{X}, \mathbf{z}; \theta) = \ln p(z_1; \theta) + \sum_{t=2}^T \ln p(z_t | z_{t-1}; \theta) + \sum_{t=1}^T \ln p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_t; \theta). \quad (19)$$

注意到

$$\begin{aligned} Q(\theta, \theta^{\text{old}}) &= \sum_{z_1=1}^S \gamma_{z_1}(1) \ln p(z_1; \theta) + \sum_{t=2}^T \sum_{z_{t-1}=1}^S \sum_{z_t=1}^S \xi_{z_{t-1}, z_t}(t) \ln p(z_t | z_{t-1}; \theta) \\ &\quad + \sum_{t=1}^T \sum_{z_t=1}^S \gamma_{z_t}(t) \ln p(\mathbf{x}_t | \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, z_t; \theta). \end{aligned} \quad (20)$$

关于前向-后向算法中为了数值稳定性的放缩因子以及用于预测最可能的隐状态序列的 Viterbi 算法都和普通的 HMM 一样，具体可以在 [Bishop and Nasrabadi, 2006] 中找到。

A.2 阈值

图~6 展示了当阈值从 0 变到 0.5 时，用阈值求得的邻接矩阵和邻接矩阵先验的平衡准确率。这两张图说明了，一般情况下，准确率对阈值的选取是非常敏感的。

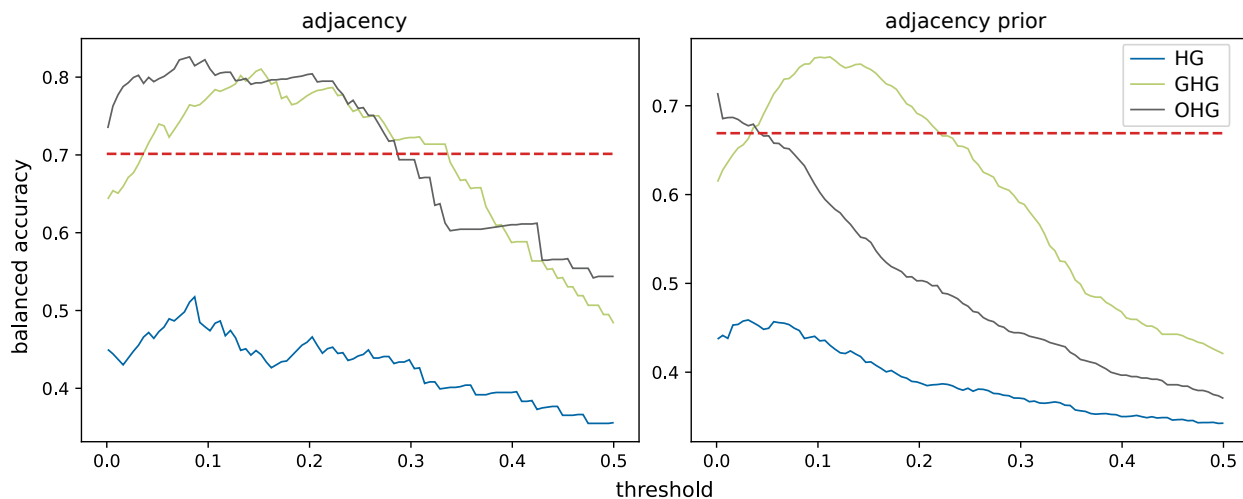


图 6: 用不同阈值离散化权重矩阵来得到邻接矩阵和邻接先验。红色的段划线是直接用 OHG 得到的 A_0 的平衡准确率。

A.3 Synthetic dataset

图 7 展示了所有模型在一个合成放电序列上的状态推断。

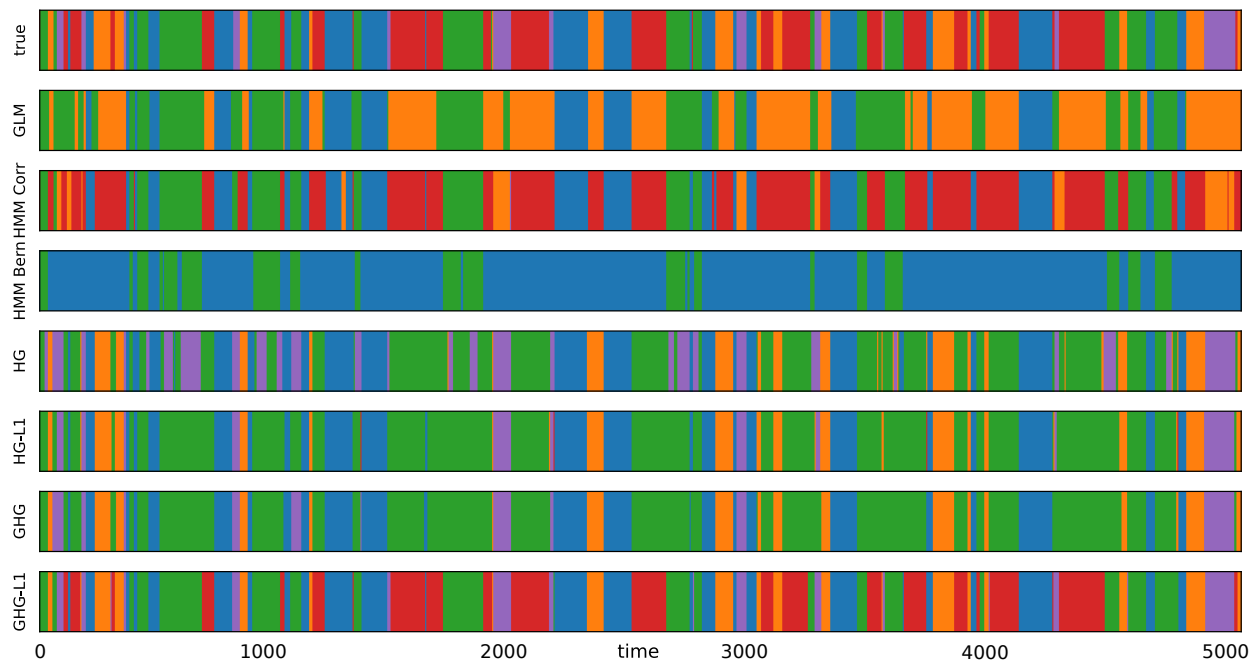


图 7: 所有模型在合成放电序列数据的一个试次上的推断的状态序列。不同颜色表示不同状态。

A.4 PFC-6 数据集

A.4.1 连续试次的状态可视化

图 8 展示了所有模型在试次 15-24 上的推断状态序列。

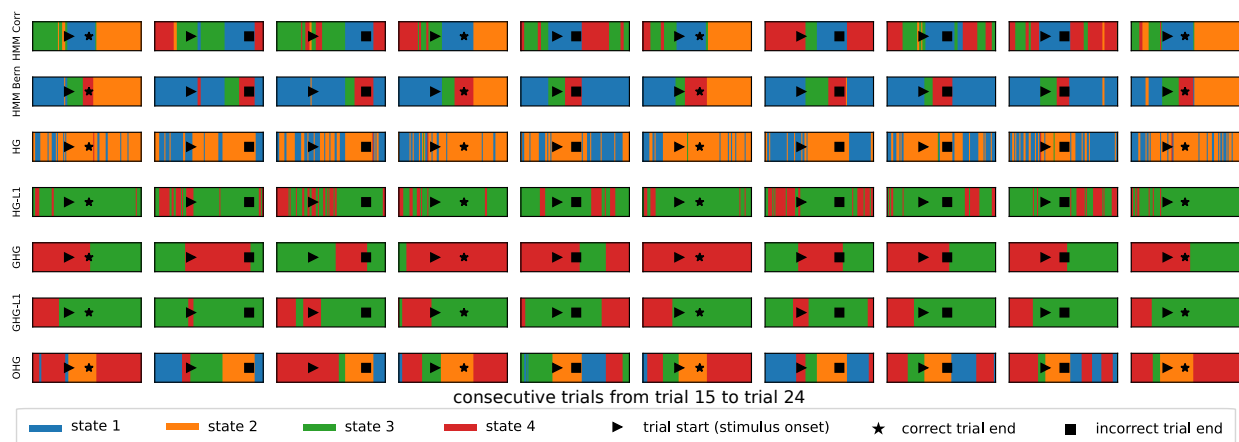


图 8: 所有模型在 PFC-6 数据集的连续试次上的推断状态序列。

A.4.2 由推断的隐状态解码任务信息

为了进一步确认我们对 OHG 推断状态的解释，我们用 logistic 回归从推断的状态解码每个试次的正确性。具体来说，每个试次被视为 logistic 回归的一个样本点。输入变量是推断的隐状态的独热表示，尺寸为 $S \times T$ 。我们用独热表示，因为每个时间桶内的状态是一个分类变量。输出是个二值变量，代表试次的正确性。我们用 2/3 试次来训练 logistic regression，用剩下的 1/3 测试。

图 9 用 OHG 推断的状态拟合 logistic 回归能得到最高的解码准确率。这意味着 OHG 推断的状态的确包含足够多与试次正确性相关的信息。在正文中，我们提到了如果是一个正确的试次，动物就会因为在正确目标位置获得了奖励而进入状态 4 这一解释。这对应了试次结束前状态 4 对应的正系数，即图 9 中黑框圈起的部分。

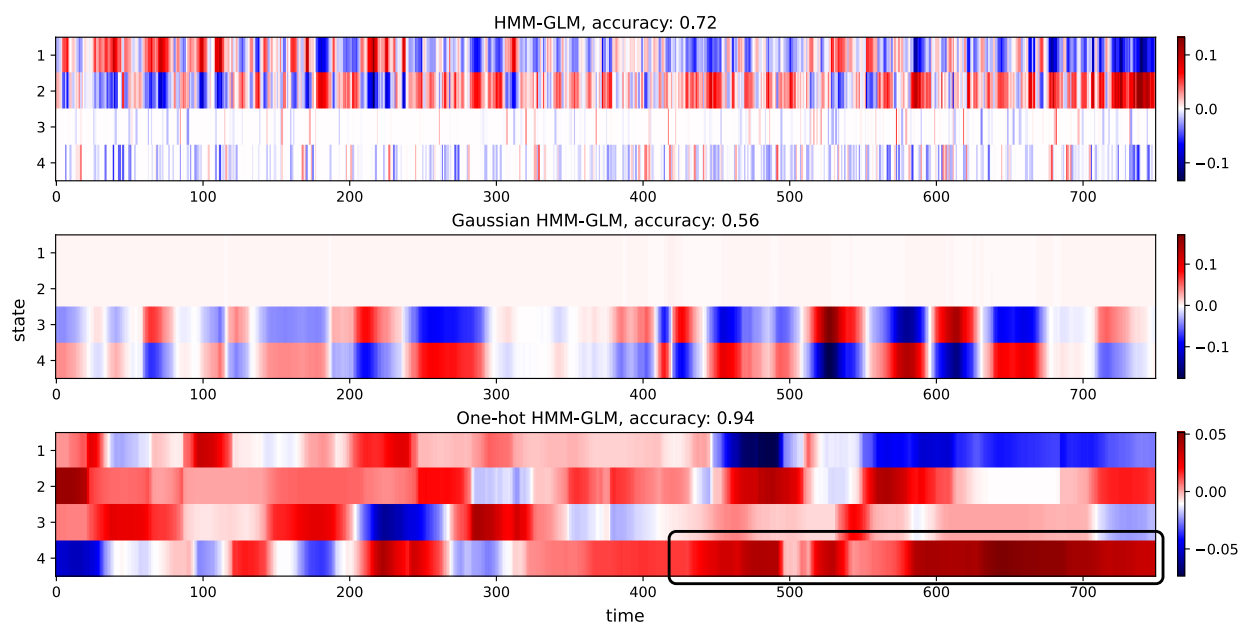


图 9: Logistic 模型 正确性 = LogisticRegression(推断状态) 的系数矩阵以及正确率。

为进一步确认推断得到的隐状态对预测试次正确性的不可替代的角色，我们训练了一个多层感知机 (MLP) (隐藏层有 100 个节点，这里我们试了很多种配置，并选出了最好的)，直接用神经放电序列作为输入预测试次的正确性。测试集上的准确率只有 0.72，明显不如 OHG。这意味着 OHG 在从神经放电序列中总结任务信息上扮演了重要角色，就像 CNN 里的卷积层一样，起着不可替代的作用。

A.4.3 OHG with up to 20 states

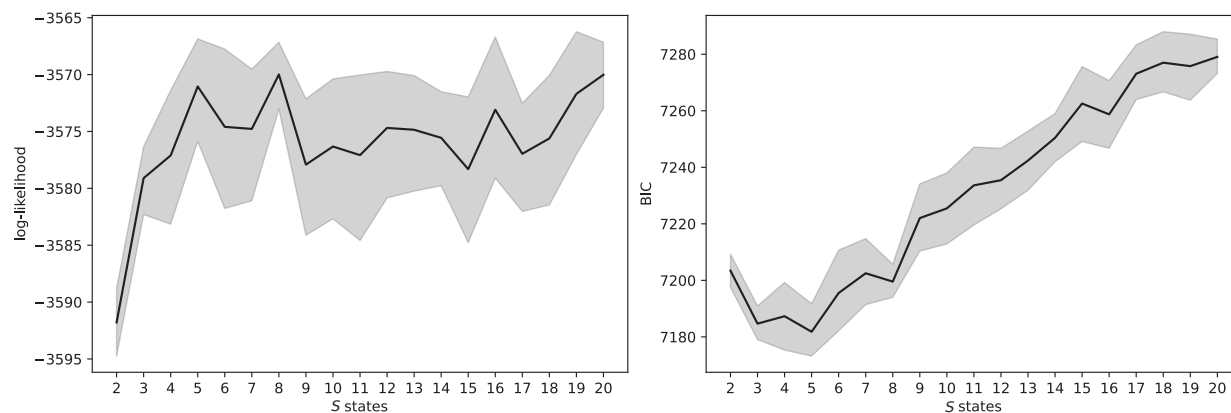


图 10: OHG 的对数似然 (左) 和贝叶斯信息量准则 (Bayesian information criterion, BIC) (右) 关于最多到 20 个状态的曲线。性能在 $S \geq 5$ 时就不再显著提升, 而是变得平缓了。



图 11: 不同状态数的 OHG 推断的隐状态. 尽管假设了更多的状态, 但只学到了大约 4 个有效状态, 且对其的解释都和正文中对 4 状态 OHG 的解释类似.

A.5 桶状皮质数据, 5 个隐状态

正如正文所指出的, OHG 的测试对数似然随状态数 S 的增加而提升. 当 $S = 5$ 时, OHG 推断出了大概 2 或 3 个主要状态, 而其它状态在实验段内很少出现. 图 12 展示了 $S = 5$ 的某一试次. OHG 有一个主要状态 (蓝色), 而其它状态的持续时长都比较短, 大约都是 0.1-0.3 s, 且在胡须接触附近表现出复杂的切换模式. 对应的权重、邻接和邻接先验在图 13 中. 这些状态的意义需要进一步研究.

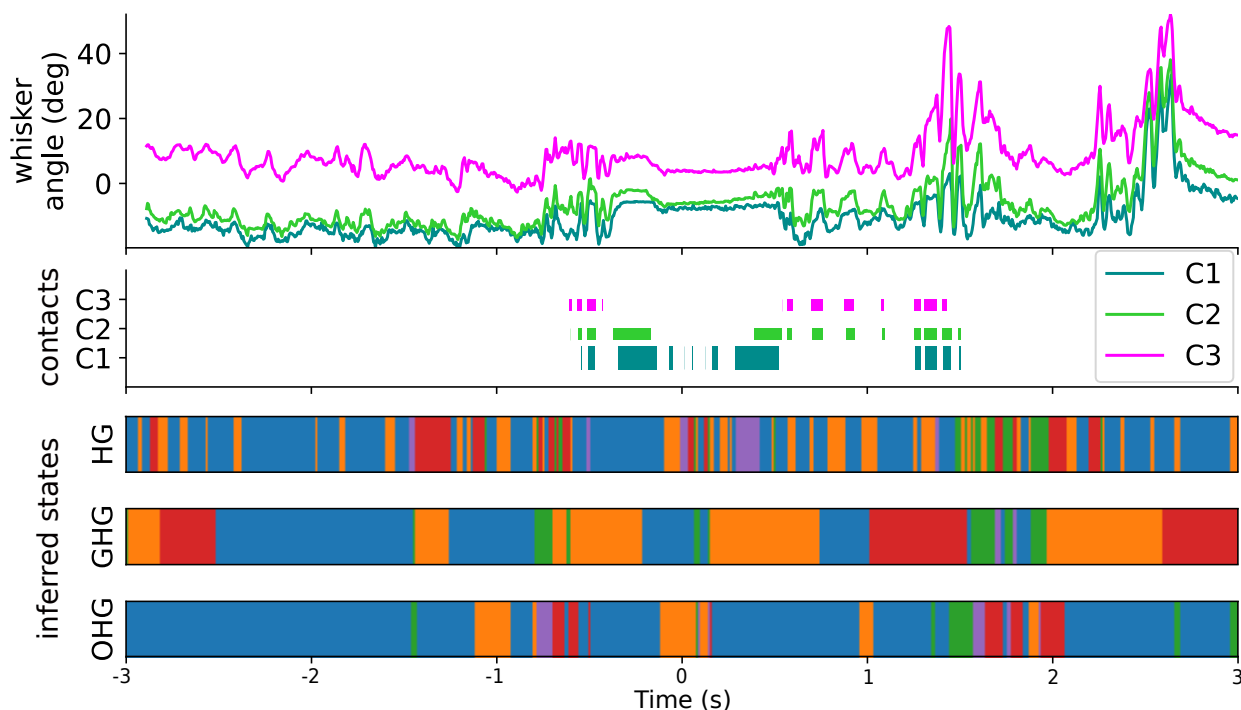


图 12: 桶状皮质数据集的一个试次, 假设 $S = 5$ 个隐状态.

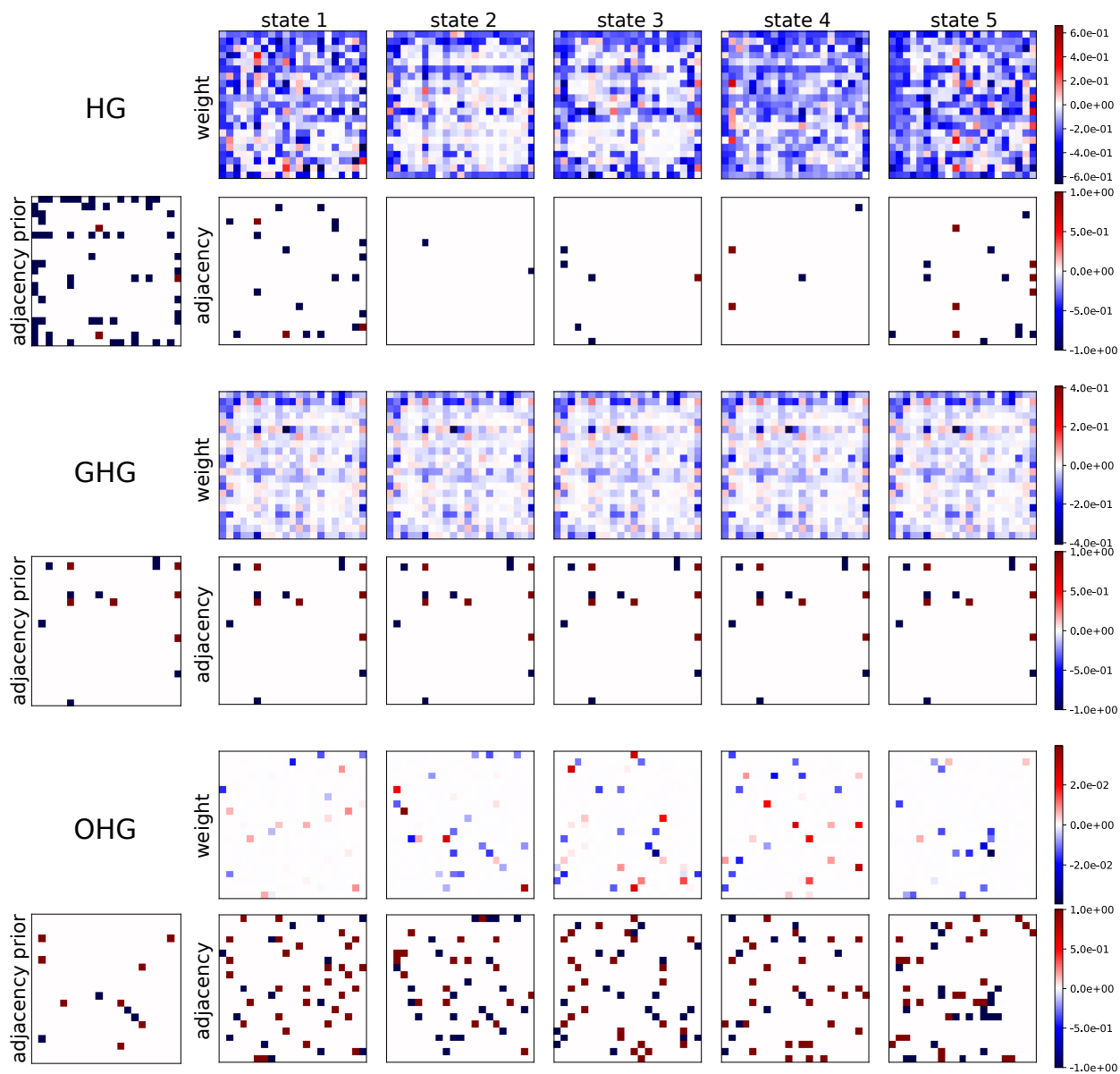


图 13: HG、GHG 和 OHG 应用到图 12 所示的实验段后，每个状态下的权重、邻接以及邻接先验。

A.6 子总体（多区域）问题

尽管本文中所有的 HMM-GLM（包括我们提出的 GHG 和 OHG）仅适用于一个全局先验，我们还是有可能把模型用在多区域数据上。比如， C 组神经元 N_1, N_2, \dots, N_C 基本上独立的运作，但它们之间也偶尔交流。此外，假设每组神经元有它们自己的全局先验和它们自己的状态数 S_1, S_2, \dots, S_C 。那么我们可以把这样的问题规约成一个单全局先验结构 $\text{diag}(\mathbf{A}_{1,0}, \mathbf{A}_{2,0}, \dots, \mathbf{A}_{C,0})$ 的模型。那么所有神经元组总共就会有 $S = S_1 \times S_2 \times \dots \times S_C$ 个状态。

为了更好的解释这一方法，我们跑了一个简单的实验，有 3 个子总体（组），每个组有 5 个神经元和 2 个状态。因此我们需要一个包含 15 个神经元、8 个状态的模型来匹配这一配置。

图 14A 表明 OHG 在 LL 和权重误差上比 HG 和 GHG 好。GHG 和 OHG 的状态准确率、邻接准确率和邻接先验准确率比 HG 好。然而，由于我们的模型的目标不涵盖子总体问题，所有模型的状态准确率基本都比较差。因为规约后的问题的状态复杂度随组数指数增长。图 14B 展示了 8 个状态组合（用二进制编码）下的真实的权重矩阵和 HG、GHG 以及 OHG 学的的权重矩阵。和图 14A 中报告的权重误差结果一致，OHG 优于 GHG 优于 HG。特别的，OHG 学到了每一组中有意义的权重矩阵。比如，OHG 的状态 1、2、3 的左上块对应第 $c = 1$ 组状态 0 的权重矩阵。OHG 的状态 5、6 的左上块对应 $c = 1$ 组状态 1 的权重矩阵。图 14C 可以看出，GHG 和 OHG 推断出的状态和真实的状态转移比，还是可以接受的。

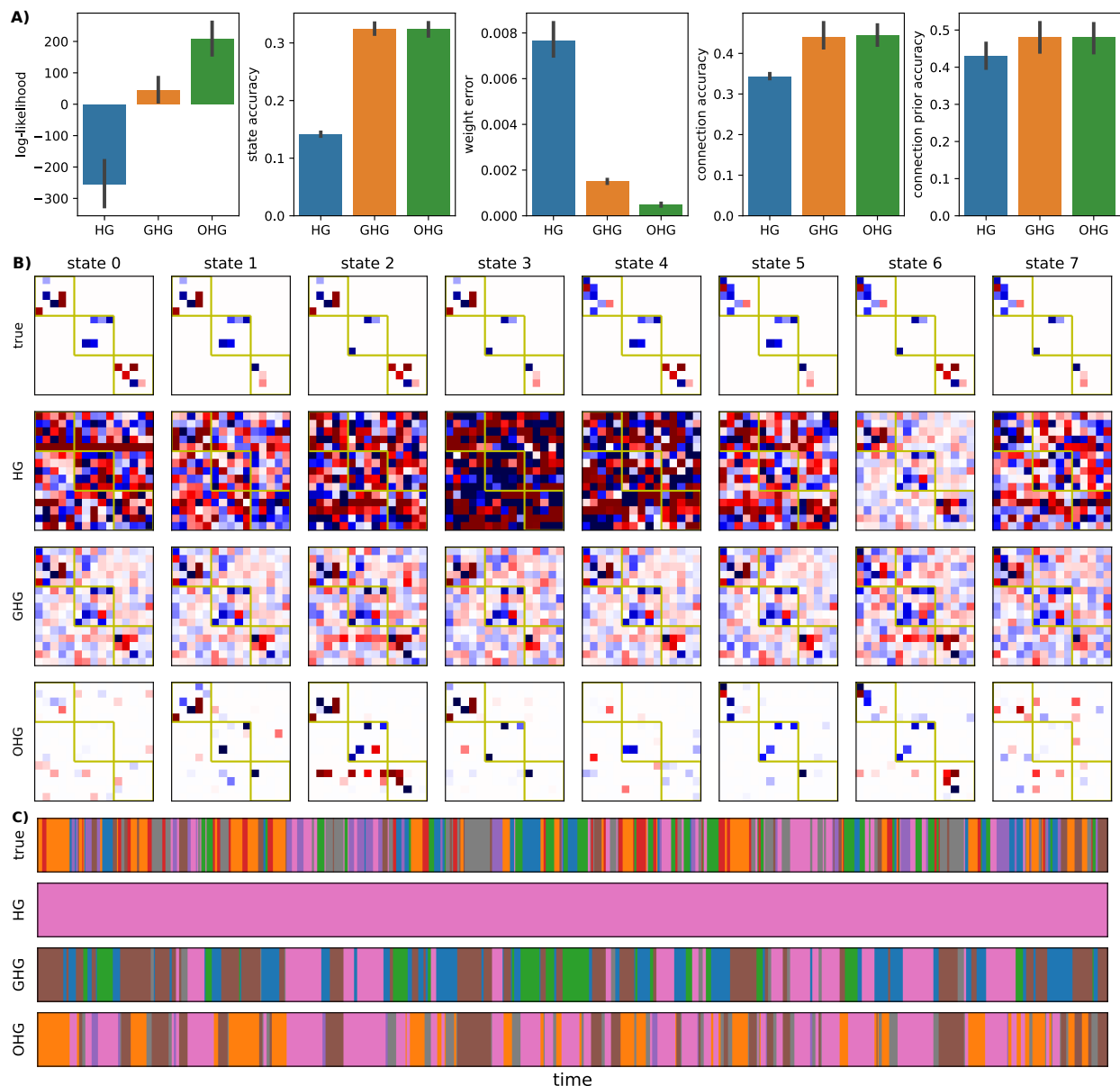


图 14: A) HG、GHG 和 OHG 五个指标的量化结果. B) 学到的 8 个状态的权重矩阵. 权重矩阵的左上块对应第 1 组; 中央块对应第 2 组; 右下块对应第 3 组. C) 真实的状态和 HG、GHG 以及 OHG 推断的状态. 8 个颜色代表 8 个状态. 特别的, 每个时间桶 t 内, 如果状态是 (s_1, s_2, s_3) , 那么综合状态为 $s = 2^2 \times s_1 + 2^1 \times s_2 + 2^0 \times s_3$. 即, (s_1, s_2, s_3) 为综合状态 s 的二进制编码.

A.7 Relationships to SLDS, rSLDS, and mp-rSLDS

与我们这种基于 HMM-GLM 的模型考虑状态切换发生在神经连接上不同, SLDS [Ackerson and Fu, 1970, Chang and Athans, 1978, Hamilton, 1990, Ghahramani and Hinton, 1996, 2000, Murphy, 1998, Fox et al., 2008], rSLDS [Linderman et al., 2016, 2017] 和 mp-rSLDS [Glaser et al., 2020] 考虑状态切换发生在驱动观测放电序列的潜在线性动力系统中. 此外, mp-rSLDS 也考虑了引入先验信息 (比如解剖信息) 到不同状态下的线性映射上. 然而, 其共享的先验是一个给定的超参数, 而我们 GHG 和 OHG 的共享全局先验是一个可以学习的参数.

尽管 SLDS、rSLDS 和 mp-rSLDS 不显式学不同状态下的神经连接, 但基于 SLDS 的模型和基于 HMM-GLM 的模型都从放电序列中学离散状态的切换. 由于本篇文章不考虑多区域或子总体问题, 我们就比较 OHG、SLDS 和 rSLDS 在 PFC-6 数据集上推断的状态.

图 15 表明 SLDS 和 rSLDS 推断的状态切换的非常快, 这妨碍了它们的可解释性. 这样的结果可能意味着, 状态切换是发生在神经连接上这一假设, 在处理多阶段实验得到的神经数据要被考虑进来.

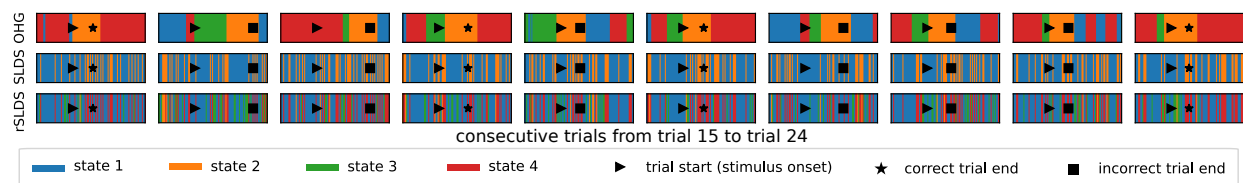


图 15: OHG、SLDS 和 rSLDS 在 PFC-6 放电序列数据的连续试次上, 推断出的状态.

A.8 大量神经元的可扩展性

从合成数据集中，我们看到所有基于 HMM-GLM 的模型，其状态准确率都在 40-50 个神经元时开始下降。在这个尺度下，网络中将会有超过 $S \times [1600, 2500]$ 个边（连接）需要学，这也为确定正确的状态切换带来了挑战。在这种情况下，EM 算法的 E-步和 M-步会互相影响，导致它们两个的效果都非常差。然而这并不意味着基于 HMM-GLM 的模型不能用在真实数据集上。尽管可能有很多神经元，还是可以尝试基于 HMM-GLM 的方法，因为图 16 表明平均到每个神经元的似然并没有随神经元数量的上升而下降。这意味着基于 HMM-GLM 的模型（尤其是 OHG）仍然能有效的预测放电率，其可用以重建放电序列。我们将这一问题视为一个重要的未来研究方向。

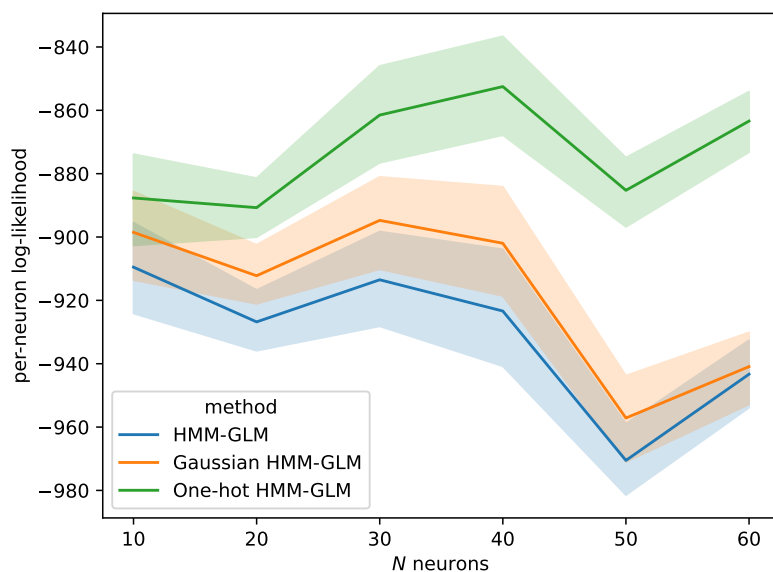


图 16: HG、GHG 和 OHG 平均到每个神经元的似然关于神经元数量 N 的变化曲线。

A.9 加先验在状态转移矩阵上

抑制 HG 快速切换的一个直观的想法就是加 L2 正则项到转移矩阵上。然而，图 17 表明当状态转移次数被抑制时，推断的状态也是没有意义的。具体来说，绝大多数状态切换都发生在一个很短的时间段内，看上去完全是噪声类型的状态切换。只有一个主要的状态占据了整个试次。

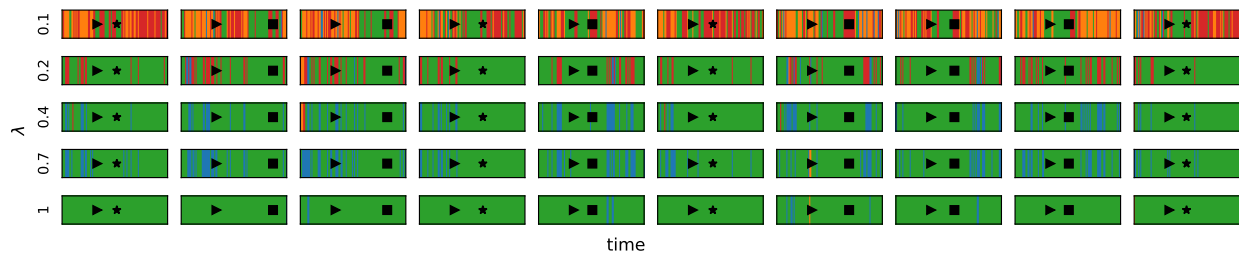


图 17: 带不同正则项的 HG 在 PFC-6 放电序列数据的连续试次上，推断出的状态。