

Atividade 2 - Bioestatística

Alunos: Marlon Valmórbida Cendron e Arthur Antunes Pereira Costa

1. Contextualização e justificativa do estudo (até uma página)

A neurogênese adulta no giro denteado (GD) do hipocampo desempenha um papel crucial na codificação e consolidação de memórias, como demonstrado por estudos experimentais que apontam uma piora em tarefas de memória quando esse processo é inibido. Porém, o processo ou computação pelo qual a neurogênese desempenha esse papel não é entendido. Entre as principais hipóteses, acredita-se que a neurogênese melhore uma função já estabelecida do GD: a separação de padrões, processo essencial para distinguir informações similares.

Estudos de modelagem computacional da neurogênese adulta no GD reportam resultados conflitantes acerca de seu papel na separação de padrões, e seus efeitos em outras áreas do hipocampo, como CA3, permanecem muito pouco explorados. Este estudo busca preencher essas lacunas, analisando um modelo do circuito GD-CA3 que leva em consideração as observações anatômicas mais recentes dessa área e a maturação temporal dos neurônios gerados pela neurogênese.

Para analisar a separação de padrões, será conduzido um experimento em três modelos computacionais do circuito GD-CA3: um modelo controle sem neurogênese (C), um modelo com neurogênese e uma população estática de neurônios jovens (NG) e um modelo com neurogênese e maturação através do tempo dos neurônios jovens (NGt). O experimento consistirá em apresentar ao circuito pares de sinais de entrada com níveis de similaridade variáveis, desde 0% de similaridade a 100%, e medir o nível de similaridade de saída do GD e do CA3. Para validar se o circuito desempenhou separação de padrões, é analisada a similaridade de ativação da área de interesse entre um sinal e outro, chamada a similaridade de saída; quando a similaridade de saída for menor que a similaridade de entrada, diz-se que o circuito desempenhou separação de padrões, pois ele foi capaz de gerar padrões muito diferentes entre si embora tenha recebido padrões muito similares, por exemplo, uma similaridade de entrada de 80% e uma similaridade de saída de 20%.

2. Hipóteses do estudo

1. A separação de padrões será maior no GD do que no CA3 independente do modelo utilizado.
2. O modelo NG apresentará maior separação de padrões que o modelo C, assumindo que a neurogênese irá melhorar a separação de padrões.
3. O modelo NGt apresentará maior separação de padrões que o modelo NG, visto o grau de realismo maior desse modelo.

3. Hipóteses nula e alternativa do teste estatístico.

$$H_0 : S_{NG}^s = S_{NGt}^s = S_C^s$$
$$H_1 : \exists i, j \mid S_i^s \neq S_j^s$$

Onde S_k^s corresponde à similaridade de saída do modelo k .

4. Apresentação do cálculo amostral. (Mostrar artigo de referência, dados utilizados, detalhamento completo de como foi calculado.)

Por se tratar de um modelo computacional, o tamanho da amostra, ou número de vezes que a simulação é executada pode ser definido arbitrariamente. Isso ocorre porque o ruído inerente às simulações é controlável e reproduzível. Logo, optamos por utilizar 10 simulações por condição, tomando como referência Kassad & Alexandre (2018) .

5. Características da amostra.

Como dito anteriormente, a amostra é composta por 3 modelos do circuito GD-CA3: Circuito sem neurogênese (C); com neurogênese estática, ou seja, sem maturação temporal das unidades (NG); e com neurogênese e maturação temporal (NGt). Serão realizadas, para cada modelo, 10 simulações. Cada uma dessas consiste de dois estímulos com um certo nível de similaridade entre si que o modelo deverá diferenciar (separar). Cada simulação terá 11 níveis de similaridade (de 0 a 100%, com 10 de diferença entre eles). A saída do modelo é similaridade entre as ativações das unidades das duas áreas para os dois estímulos.

6. Descrição do tratamento estatístico planejado

Todas as análises serão realizadas utilizando diferentes bibliotecas (NumPy, SciPy, statsmodels, Seaborn) escritas na linguagem de programação Python (versão 3.12).

Inicialmente, serão computadas médias e desvios padrões (através de biblioteca NumPy) das similaridades de saída no GD e CA3 para cada modelo e cada nível de similaridade de entrada. Também será investigado o ajuste dos dados à distribuição normal (SciPy).

Para averiguar se os modelos são capazes de separar padrões, será comparada a similaridade de saída e a similaridade de entrada para cada modelo utilizando teste-t em caso de normalidade, e o teste de Wilcoxon em caso de desvio da normalidade (statsmodels).

Em caso de normalidade dos dados, a possibilidade de diferença entre os modelos será investigada utilizando uma ANOVA de medidas repetidas (statsmodel) tendo como fatores: Modelo (G, NG, NGt); Área (GD, CA3); Similaridade de entrada (0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100). Caso sejam encontradas diferenças estatisticamente significativas, será realizado post-hoc de Tukey entre os pares para avaliar a direção da diferença. Caso haja desvio da normalidade, serão aplicados Modelos Lineares Generalizados (statsmodel), utilizando a fórmula:

$$\textit{similaridade de saída} \sim \textit{modelo} + \textit{regiao} + \textit{modelo} : \textit{regiao}$$

Neste caso, modelo:regiao é o fator de interação entre as duas variáveis. Visualizações dos dados serão geradas através da biblioteca Seaborn.

7. Arquivo dos dados a serem analisados

Os dados foram gerados artificialmente e, portanto, não correspondem a valores reais gerados pelos modelos. O arquivo CSV com os dados íntegros será enviado em anexo.

8. Referências Bibliográficas

Kassab, R., & Alexandre, F. (2018). Pattern separation in the hippocampus: distinct circuits under different conditions. *Brain Structure and Function*, 223(6), 2785–2808. <https://doi.org/10.1007/S00429-018-1659-4/METRICS>