

基于 BDM-SNN 的 Mario 避障

摘要：本次实验基于 BrainCog 框架展开，调用框架中的 BDM-SNN 网络进行实验，基本实现了 Mario 游戏。实验过程中，首先完成了实验环境的配置，安装 BrainCog 及实验所用的库函数；其次，通过 BrainCog 提供的案例完成对配置环境的检验；最后，结合 BDM-SNN 的生物机理及其网络模型，实现了 Mario 游戏。其实验结果表明，Mario 可以通过环境自主的学习与决策，及时的跳过怪兽，从而完成避障。

关键词：BrainCog；BDM-SNN；生物机理；Mario

一、实验背景

决策（Decision-Making）是人类高级认知科学领域中的热门研究课题之一，它在经济学、社会学和人工智能等方面具有潜在的应用价值。决策作为高级认知过程,它涉及多个大脑区域间的信息交互,而且决策信息的加工一般在毫秒级时间尺度内完成。但是,当前研究对决策过程的大脑神经机制还缺乏深入的了解^[1]。

脉冲神经网络（Spiking Neural Network, SNN）包含时序动力学特性的神经元节点、稳态-可塑性平衡的突触结构、功能特异性的网络环路等，高度借鉴了生物启发的局部非监督（如脉冲时序依赖可塑性）和全局弱监督（如多巴胺奖赏学习）的生物优化方法，因此有强大的时空信息表征、异步事件信息处理、网络自组织学习等能力^[2]。

全脉冲神经网络的类脑认知智能引擎“智脉”(Brain-inspired Cognitive Engine, BrainCog),以多尺度生物可塑性原理为基础,支持全脉冲神经网络(SNN)建模,具备脑启发的人工智能模型以及脑功能和结构模拟能力^[3],为开展类脑人工智能研究和计算神经研究提供了平台。

为了进一步理解类人决策的计算机制，本实验根据前人的研究成果^[4]，基于 BrainCog 的开源平台，结合 python-game 设计了 Mario 避障游戏，实现了基于大脑决策脉冲神经网络（Brain-Inspired Decision-Making Spiking Neural Network, BDM-SNN）的 Mario 避障功能。

二、BDM-SNN 的生物机理

在日常生活中，我们常常需要根据每个选项的价值在不同选项之间进行选择。决策(decision-making)这个认知过程要求我们在一系列不同的刺激中采取适应性行为：确定选择，抑制直接或间接的诱惑，综合不同的变量(如概率和价值)，并预测选择所产生的后果。认知控制的神经基础主要集中在前额叶皮层(prefrontal cortex, PFC)，尤其是其背外侧和内侧部分。PFC 脑区被发现在灵活决策中发挥关键作用，通过调节情感和奖赏系统影响认知功能。

在许多情况下，对每个选项的评价是一个复杂的过程，在这个过程中必须结合多个信息，眶额叶皮层(OFC)被认为在这个过程中发挥主要作用。背外侧前额叶皮层(dorsolateral prefrontal cortex, DLPFC)接收来自 OFC 的输入，将基于刺激的决策转化为动作。

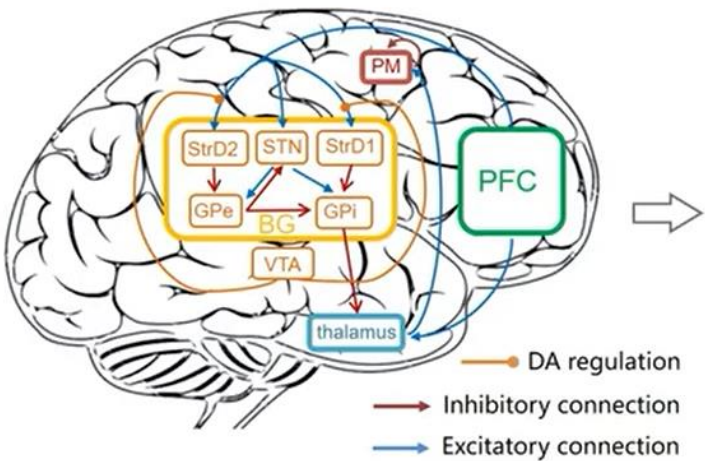


图 1 为 PFC 脑区图

BDM-SNN 是基于哺乳动物脑自主决策的神经环路来构建的多脑区协同脉冲神经网络。在哺乳动物自主决策神经环路由前额叶皮层→基底神经节→丘脑→运动皮层四个主要脑区之间的连接组成。其中前额叶皮层到基底神经节分为三条通路：直接通路(StrD1)、间接通路(StrD2)和超直接通路(STN)。其主要作用的环路机制为：中脑腹侧被盖区(VTA)中的多巴胺能神经元会以相反(兴奋/抑制)的方式分别调控直接通路和间接通路，进而放大两条神经通路之间的差异，从而实现快速决策。BDM-SNN 通过脉冲神经网络构建自主决策神经环路多个脑区之间兴奋和抑制性的连接，并通过直接和间接通路来优化网络的权重。

三、BDM-SNN 的网络模型

BDM-SNN 网络架构如图所示，包含 11 个模块，它们对应于皮质基底节丘脑皮质环上的关键脑区。BDM-SNN 模型的模块包括 MOFC、LOFC、DLPFC、PM、SNc/VTA、StrD1、StrD2、STN、GPe、SNr/GPi 和丘脑。StrD1 和 StrD2 对应于纹状体中的“Go”和“No Go”细胞。其他模块与皮质基底节丘脑皮质环上的大脑区域功能对应。不同模块之间的兴奋性和抑制性连接受到皮质基底节丘脑皮质环上的连接的启发。DA 调节 DLPFC 和纹状体之间的连接。

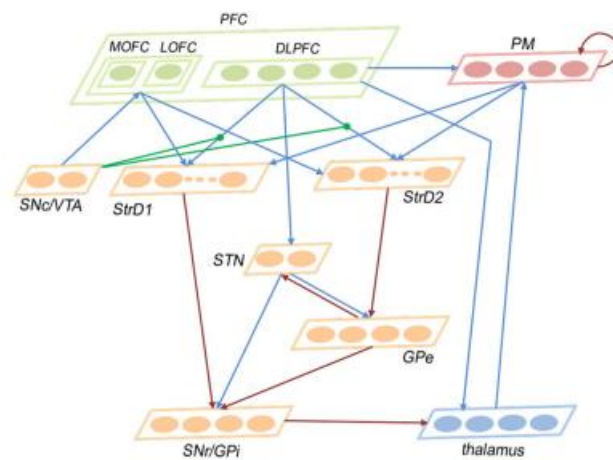


图 2 为 BDM-SNN 模型的网络架构

图 2 中橙色大脑区域代表 BG 的子区域。绿色大脑区域代表 PFC 的子区域。蓝色线表示兴奋性连接，而红色代表抑制性连接。绿色连接表示 DA 调制器连接。

不同区域之间的连接方式基于其功能。这里，完全连接意味着所有连接。特定连接是指特定状态或动作与特定状态-动作对之间的连接。来自 DLPFC 和 PM 的状态和执行的动作被传输到 StrD1 和 StrD2。StrD1 和 StrD2 显示所有状态动作对。因此，特定连接意味着从特定状态（DLPFC）和动作（PM）到状态动作对（StrD1 和 StrD2）的连接。SNc/VTA 中的；两个神经元分别对应于正反馈和负反馈。正面的（奖励相关的 DA）与 MOFC 相连，负面的（惩罚相关的 DA）与 LOFC 相连。

下面介绍 BDM-SNN 模型的具体设计和实现，SNN 以尖峰序列而非传统人工神经网络中的尖峰速率来编码信息。SNN 考虑了尖峰放电的时间因素，与其他替代网络相比，生物神经元模型和突触可塑性模型在生物学上更合理。我们使

用 SNN 对大脑决策回路进行建模。在这个模型中，DLPFC 中的每个神经元都代表一种状态。视觉输入首先被预处理并分配给状态。然后 DLPFC 中相应的神经元接收恒定的输入以激励该神经元。该模型产生的动作基于 PM 区域中的第一个尖峰神经元。每个动作对应于 PM 中的一个神经元。在这个模型中使用了延迟编码方法，其中更强的刺激比较弱的刺激使神经元更早激发。

3.1 神经元模型

为了在生物学现实的考虑和计算效率之间取得平衡，在这个模型中应用了 Izhikevich 神经元模型来构建大脑启发的 SNN。它比带泄漏整合发放模型 (LIF) 具有更多的离子动力学，并且在计算上比霍金-赫胥黎模型 (H-H) 更有效。Izhikevich 神经元模型介绍了一种能够产生多种生物模式的神经元模型，这在生物学上与 H-H 模型一样合理，但在计算上却与整合发放模型一样有效。这种神经元模型如公式 1-4 所示。其中 v 表示尖峰神经元的膜电位， u 表示膜恢复变量， a 、 b 、 c 、 d 是控制尖峰动力学类型的参数。 I 是输入，每个神经元接收来自突触前神经元的加权输入，如方程 3 所示， w_{ji} 是第 j 个神经元到第 i 个神经元的连接强度， o_j 是突触前神经元的输出：如果 $v_j \geq 30\text{mV}$ 则为 1，否则为 0。当膜电位 v 超过其峰值 30mV 时，发生动作电位（尖峰），膜电位重置为其初始值 c ，恢复变量增加 d 。Izhikevich 神经元模型主要分为两类：（1）兴奋性神经元：规则尖峰（RS），固有爆发（IB）和震颤（CH）。（2）抑制性神经元：低阈值尖峰（LTS）和快速尖峰（FS）。不同的神经元模型对应于参数 a 、 b 、 c 、 d 的不同值。本模型中使用 RS 神经元，因为它在短的间隔周期内发射一些尖峰，然后周期增加。RS 的参数为 $a = 0.02$, $b = 0.2$, $c = -65$, $d = 8$ 。

$$v' = 0.04v^2 + 5v + 140 - u + I \quad (1)$$

$$u' = a(bv - u) \quad (2)$$

$$I_i = \sum_{j=1}^N w_{ji} o_j \quad (3)$$

$$\text{if } v \geq 30\text{mV, then } \begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} \quad (4)$$

3.2 突触可塑性

尖峰时间依赖可塑性 (STDP) 是生物大脑最重要的学习原则之一。STDP 假

设突触的强度取决于前神经元和后神经元的尖峰时间差。在这里，模型使用 STDP 根据突触前和突触后神经元尖峰之间的相对时间来学习突触权重。调制原理是：如果突触后神经元在突触前神经元后几毫秒激发，神经元之间的连接将加强；否则，连接将减弱。更新功能如公式 5 所示，其中 A_+ 和 A_- 是学习速率， τ_- 和 τ_+ 是时间常数， Δt_i 是突触前尖峰到突触后尖峰的延迟时间。这里， $A_+ = 0.925$, $A_- = 0.9$, $\tau_- = \tau_+ = 20$ 。

$$\Delta w_j = \begin{cases} A_+ e^{\left(\frac{\Delta t_i}{\tau_+}\right)} \Delta t_i < 0 \\ -A_- e^{\left(\frac{\Delta t_i}{\tau_-}\right)} \Delta t_i > 0 \end{cases} \quad (5)$$

四、BDM-SNN 的实现

本项目的代码中，定义了 Mario 类和 Monster 类，其中 Mario 类定义了 mario 对象的各种参数，如跳跃时的上升速度和越过障碍物的下降速度等，而 Monster 类定义的对象是 Mario 需要进行躲避的障碍对象，在其中定义了怪物的移动速度等参数；

chooseAct()函数根据输入的当前网络的状态，返回下一时刻 Mario 需要采取的动作状态以及网络的权重，Mario 可采取的行为包括地面奔跑状态和遇到 monster 进行跳跃两种；

judgeState()函数则是根据 Mario 和 monster 的距离判断 Mario 当前所处的状态，如未遇到 monster、遇到 monster 并跳跃、是否越过 monster 等；

getReward()函数是根据执行所选择的行为后的当前状态，并判断 Mario 避障所得的分数，如成功越过障碍则分数增加，反之则减少；

updateNet()函数则是根据所得分数、执行的行为等参数更新网络权重。在主函数部分首先调用 braincog.model_zoo.bdmsnn 中的 BDMSNN()函数，构建 BDM-SNN 网络，接着创建 Mario 和 monster 对象的初始化，之后便调用上述定义的函数进行 Mario 躲避障碍物的实验。monster 出现的高度是随机变化的，而 Mario 会根据自身与 monster 的距离调整提前跳跃的时机，从而越过不同高度的 monster，完成避障任务。在实验进行的同时记录避障结果并计算得分，根据所选取的行为和得分结果更新网络权重。

五、基于 BDM-SNN 的 Mario 避障的实验结果

通过实验发现，Mario 可以通过环境自主的学习与决策，及时的跳过怪兽，从而完成避障。在 4 种不同状态下，建立不同的奖励机制，从而完成对 BDM-SNN 网络权重的更新。通过对多代的实验结果进行观察可以发现，Mario 可以很好的避开不同高度的怪物，具有良好的学习与决策能力。

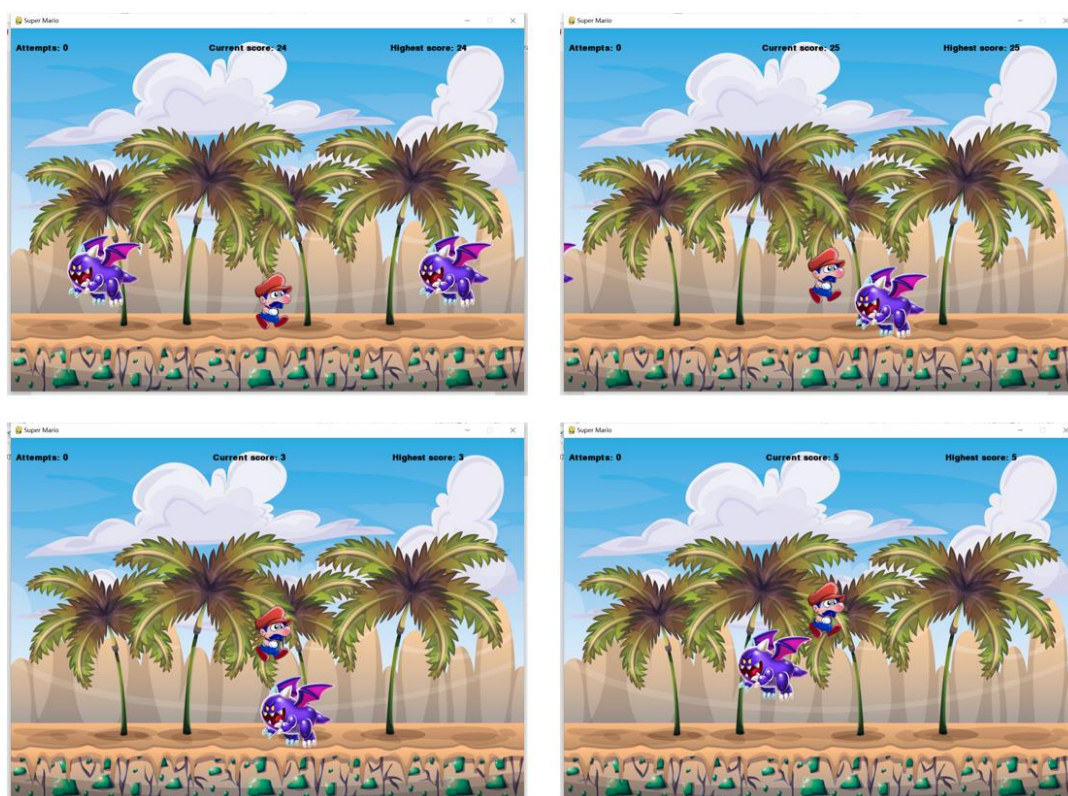


图 3 为 Mario 在决策过程中的状态

六、总结

通过本次实验，初步掌握了 BrainCog 的使用。基于平台提供的 SNN 网络，其帮助实验人员进一步学习并了解了 SNN 网络的基本架构与实现机理，同时完善的网络库，调用起来非常便利，为实验过程提供了极大的便利。本次实验较好的完成了 Mario 的避障过程，结合最终结果，可以观察到 Mario 具有了一定的自主学习与决策能力。

表 1 为各位组员实验过程参与情况

姓名	学号	工作
段竣然	202228014825021	实验编程、报告撰写、统筹规划
陈奥新	2022E8014682034	实验编程、报告撰写
伍宇锋	202228015926039	实验编程、报告撰写
王明川	202228019426022	实验编程、报告撰写
孙煜茗	202228010415012	实验编程、报告撰写

实验过程中，小组成员均积极认真参与了编程与报告撰写，故具体贡献不作区分。同时排名不分先后，具有同等地位。

参考文献

- [1] 司亚静. 基于脑电的决策网络研究[D].电子科技大学,2020.2020.000273.
- [2] 张铁林,徐波.脉冲神经网络研究现状及展望[J].计算机学报,2021,44(09):1767-1785.
- [3] Zeng Y, Zhao D, et al. BrainCog: A Spiking Neural Network based Brain-inspired Cognitive Intelligence Engine for Brain-inspired AI and Brain Simulation[J]. arXiv preprint arXiv:2207.08533
- [4] Zhao F , Zeng Y , et al. A Brain-Inspired Decision Making Spiking Neural Network and Its Application in Unmanned Aerial Vehicle[J]. frontiers in Neurorobotics , 2018.