

# 基于学习的多目标脑决策模型研究\*

叶伟杰<sup>1</sup> 刘深泉<sup>2†</sup>

(1.广东财经大学统计与数学学院,广州 510320) (2.华南理工大学数学学院,广州 510640)

**摘要** 多目标决策在大脑的认知功能中起着关键的作用. 在本研究中,我们将一个额叶视区网络模型扩展为一个基于学习的模型,并训练这个模型使其完成一个认知决策任务——non-choice 任务,然后用模拟结果解释大脑在进行多目标选择时的认知过程. 经过上千次训练后,额叶视区模型从随机选择决策目标转变为选择与最大奖励相关联的决策. 在训练过程中,模型的多目标决策顺序也与目标关联的奖励梯度相关. 此外,改变不同决策间的奖励差对模型的决策速度有重要的影响,可以使模型进入两种学习阶段:快速学习阶段和慢速学习阶段.

**关键词** 多目标决策, 学习, 眼跳, 额叶视区

DOI: 10.6052/1672-6553-2017-51

## 引言

在多个目标中进行决策是一项重要的脑认知活动.这类多目标脑决策过程非常复杂,不仅包括感觉信号输入的处理和运动信号的输出,还需要其它信息的处理与积累<sup>[1,2]</sup>.大量的实验研究已经揭示了许多类型的信息如过去的经验,新的规则,奖励与惩罚等都会影响大脑在目标选择中的决策<sup>[3-7]</sup>.这些信息的处理使得大脑能够根据不同情况做出相应的决策.正是基于这种灵活性,人类与其它生物才能以不同的方式对某一特定刺激做出反应<sup>[8,9]</sup>.

目前,脑决策研究主要利用一些眼跳运动相关的选择任务来进行,如反向眼跳任务, no-go 任务和 non-choice 任务等<sup>[10-12]</sup>.这是因为一次眼跳运动可以看作一次决策的结果,且眼球从一个点到另一个点的转动能够较容易地捕捉.眼跳相关的选择任务的目的是让测试对象根据给定的视觉刺激进行反应,并执行计划的眼跳运动.若测试对象做出正确的决策则会收到奖励.这些奖励引导测试对象学习正确的视觉-运动映射并抑制错误的决策<sup>[13,14]</sup>.换言之,大脑会在奖励的指导下重新建立感觉信号与行为动作之间的联系<sup>[15,16]</sup>.在与眼跳运动有关的脑

区中,额叶视区(frontal eye field, FEF)接收和处理来自视觉通路的信息,并将处理后的神经信号传递到额叶与上丘等脑区,其作用非常关键.神经元生理实验和病变研究均证实额叶视区在眼跳运动过程中的视觉信息输入处理,视觉运动的计划以及视觉注意力分配等功能中起着极其关键的作用<sup>[17,18]</sup>.此外,近年来的实验研究证实了眼跳运动的选择与额叶视区的突触可塑性有关,使得视觉选择任务中的多目标决策过程可以用大脑学习的相关机制来解析<sup>[19-23]</sup>.这些证据为额叶视区通过学习来控制多目标决策任务中的视觉-运动映射重建提供了实验和理论的支持.尽管生理实验和行为实验在脑决策机制的研究中有显著的进展,但实验的困难使得模型理论研究成为实验之外最有效的脑决策机制研究手段.本文主要研究学习在多目标脑决策中的作用机制,基于一个改进的额叶视区模型,对多目标决策任务 non-choice 任务进行仿真研究.

## 1 模型的改进与任务仿真

### 1.1 额叶视区网络模型的改进

本研究将 Heinzle 等<sup>[24]</sup>提出的额叶视区层级模型修改为一个基于学习的模型,主要包括以下三方面:

2016-12-28 收到第1稿,2017-04-01 收到修改稿.

\* 国家自然科学基金资助项目(11572127)

† 通讯作者 E-mail: mashqliu@scut.edu.cn

(1)模型中的识别模块和第 6 层级被移除,修改后的模型只能产生正向眼跳。

(2)除了注视输入层,第 4 层级,第 2/3 层级和第 5 层级中的每个神经元群体都加入了方向偏好特性,即神经元群体对来自特定方向的刺激所产生的神经元活动会比其他群体的更强烈<sup>[25-30]</sup>。

(3)增加了从第 4 层级的兴奋性神经元群体(E4)到第 2/3 层级的抑制性神经元群体(I23)的连接。此外,从第 4 层级到第 2/3 层级的所有突触连接都具有突出可塑性,可以通过学习来改变对应的连接情况<sup>[31-38]</sup>。

改进后的额叶视区网络模型的结构如图 1(a)所示。模型有四个层级,它们有着不同的作用:第 4 层级(L4)处理视觉输入,第 2/3 层级(L23)分配注意力,注视输入层(FIX)处理注视信息和第 5 层级(L5)输出运动信号。

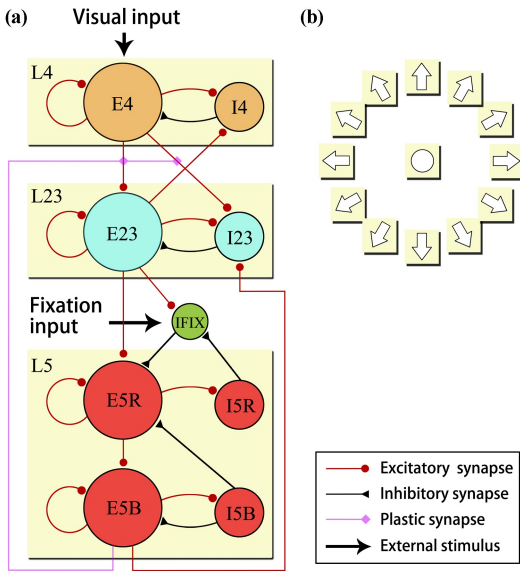


图 1 额叶视区网络模型的结构示意图

Fig.1 Structure of FEF neural network model

除 FIX 外,每一层都有 13 个有各自方向偏好的视网膜位点。在不同的神经元层级中,每一个视网膜位点由不同数量的神经元组成。L4 和 L23 的每一个视网膜位点包含 100 个兴奋性神经元和 25 个抑制性神经元,而 L5R 和 L5B 的视网膜位点则由 40 个兴奋性神经元和 25 个抑制性神经元组成。FIX 有 100 个抑制性神经元。

## 1.2 改进后的额叶视区网络的数学模型

改进的额叶视区网络模型的神经元都采用整合发放模型,其描述如下:

$$\tau_m \frac{dV}{dt} = -V - g_{exc}(V - V_e) - g_{inh}(V - V_i) + I_{ext} \quad (1)$$

其中, $V$ 是膜电位, $\tau_m$ 代表时间常数,在兴奋性神经元中 $\tau_m = 20$  ms,在抑制性神经元中 $\tau_m = 12$  ms,而 $V_e = 74$  mV, $V_i = -10$  mV 分别定义了兴奋性和抑制性突触的反转电位。电位阈值 $V_{th}$ 是 20 mV,而重设电位 $V_{reset}$ 是 10mV。兴奋性和抑制性神经元的不应期分别是 1.8 ms 和 1.2 ms。

在公式(1)中,电导 $g_{exc}$ 和 $g_{inh}$ 均由三部分组成: $g_{e,i}$ , $g_{noise}$ 和 $g_{plastic}$ ,分别是突触电导( $e$ 代表兴奋性, $i$ 代表抑制性),噪声输入和可塑性电导。

(1)突触电导 $g_{e,i}$ 满足如下方程:

$$g_{e,i}^{k-j} = \sum_l g_{kj} s_{e,i}^{k-j} \quad (2)$$

$$\tau_{e,i} \frac{ds_{e,i}}{dt} = -s_{e,i}$$

其中 $s_{e,i}$ 是激活变量, $\tau_{e,i}$ 是时间常数。 $\tau_{e,i}$ 在不同的连接中有不同的值:在连接E5R→E5R中 $\tau_{e,i} = 50$  ms,在连接E23→E23,I5B→E5R中 $\tau_{e,i} = 10$  ms,而在其他连接中 $\tau_{e,i} = 5$  ms。 $g_{kj}$ 表示对方向 $\theta_k$ 和 $\theta_j$ 有偏好的神经元之间的方向偏好因子<sup>[39]</sup>:

$$g_{kj}(\theta_k - \theta_j) = g_a e^{-\frac{(\theta_k - \theta_j)^2}{2\sigma^2}} \quad (3)$$

其中 $\sigma = 18^\circ$ 。E4 与 I23 之间的 $g_a$ 等于 0.01,表示其连接是弱连接,其他连接中 $g_a = 1.0$ 。对于视网膜中心点,与其他视网膜位点间的 $g_{kj}$ 为 $g_{kj}(30^\circ)$ 。

(2)噪声输入 $g_{noise}$ 模拟大脑的背景噪声<sup>[40-44]</sup>,其满足:

$$\tau_{noise} \frac{dg_{noise}}{dt} = -(g_{noise} - \mu_{e,i}) + \sqrt{\frac{\mu_{e,i} w_{e,i}}{\tau_{noise}}} \chi(t) \quad (4)$$

其中 $\chi(t)$ 满足 $N(0, 1)$ 正态分布。 $\mu_{e,i}$ 是平均外部输入。时间常数 $\tau_{noise} = 3$  ms,而外部权重在兴奋性突触中为 $w_e = 0.02$ ,在抑制性突触中为 $w_i = 0.02$ 。

(3)可塑性连接可能处于增强状态和减弱状态。为了方便叙述,我们将E4→E23和E4→I23的可塑性突触分别命名为兴奋性可塑性突触和抑制性可塑性突触。它们满足以下方程<sup>[31-33]</sup>:

$$g_{plastic}^k = g_p \sum_j c_{kj} \quad (5)$$

其中 $g_p = 1.0$ , $c_{kj}$ 是可塑性输入强度。在每次训练之后,突触前和突触后的神经元活性将决定可塑性突触会增强或减弱,并根据一个基于奖励的 Hebbian 学习机制来更新可塑性输入强度。突触的增强由以

下方程决定:

$$c \rightarrow c + qQ(S_{\max})(1-c) \quad (6)$$

而突触的减弱要经历以下过程:

$$c \rightarrow c - qQ(S_{\max})c \quad (7)$$

其中  $q$  是奖励因子. 可塑性率  $Q(S_{\max})$  依赖于突触前神经元的最大发放率  $S_{\max}$ :

$$Q(S_{\max}) = \left(1 + e^{\left(\frac{S_0 - S_{\max}}{\sigma_q}\right)}\right)^{-1}$$

其中群体发放率  $S(t)$  在下文描述,  $S_0 = 60\text{Hz}$  和  $\sigma_q = 5\text{Hz}$ .

$I_{\text{ext}}$  是输入到 L4 神经元的外部刺激, 其同样有方向偏好特性的:

$$I_{\text{ext}}(\theta) = I_m e^{\left(-\frac{(\theta - \theta_p)^2}{2\sigma_p^2}\right)} \quad (8)$$

其中  $I_m = 0.056\text{nA}$ ,  $\sigma_p = 15^\circ$ , 而  $\theta_p$  表示偏好的方向.

### 1.3 Non-choice 任务过程

Non-choice 任务并没有预先设定某一个特定的眼跳目标, 测试对象需要根据最终决策所得到的不同奖励来调节它的决策. 在本研究中, Non-choice 任务需要对神经网络进行 5000 次训练. 在每一次训练中, 神经网络模型在注视中心点一段时间后会同时接收四种不同的方向偏好刺激, 模型需要在这些方向中做出选择并输出眼跳运动信号. 这四个方向的眼跳分别对应不同大小的奖励, 当网络模型产生某一方向的眼跳输出后会得到对应的奖励<sup>[45]</sup>, 并根据得到的奖励更新可塑性连接的状态. 这个状态会一直保持直到下一次更新. 在任务的最后, 我们期望网络模型能做出与最大奖励相关联的决策.

在 Non-choice 任务中, 我们用两类发放率来描述网络模型的表现: 神经元发放率以及群体发放率<sup>[46,47]</sup>. 神经元发放率由单个神经元的膜电位决定:

$$r(V) = \left(1 + e^{\left(\frac{\gamma - V}{\beta}\right)}\right)^{-1} \quad (9)$$

其中  $\beta = 2.0$ ,  $\gamma = 16.5$ . 群体发放率是通过数 1 ms 内单个神经元群体产生峰发放的个数来计算且由  $S(t)$  平滑:

$$S(t) = \frac{\left(1 - e^{\left(-\frac{t}{\tau_{\text{rise}}}\right)}\right) e^{\left(-\frac{t}{\tau_{\text{decay}}}\right)}}{\int_0^\infty \left(1 - e^{\left(-\frac{t'}{\tau_{\text{rise}}}\right)}\right) e^{\left(-\frac{t'}{\tau_{\text{decay}}}\right)} dt'} \quad (10)$$

其中  $\tau_{\text{rise}} = 1\text{ms}$ ,  $\tau_{\text{decay}} = 10\text{ms}$ .

此外, 根据 non-choice 任务的特点, 我们设计了调节率指标来定量地表示两种可塑性突触的平

均调节作用:

$$M_j^{e,i} = \left\langle \frac{\sum_k (g_{\text{plastic}}^{e,i}(k) + g_{\text{syn}}^{e,i}(k))}{\sum_k (g_{\text{syn}}^{e,i}(kj))} \right\rangle \quad (11)$$

其中  $k$  和  $j$  分别代表突触连接的起始和目标位点.  $M_e$  和  $M_i$  分别是兴奋性和抑制性调节率. 运算  $\langle a \rangle$  表示任务中  $a$  的 5 次重复试验的平均值. 调节率的应用是为了估计突触强度的变化并且避免一些突触强度接近零而出现的差异较大的情况.

模型的数值模拟是用步长为 0.1ms 的欧拉方法在 Matlab 中进行.

## 2 结果

### 2.1 额叶视区网络模型的初始状态

额叶视区网络模型在初始状态下的决策行为有两个特点: (1) 只产生正向眼跳, 即当视觉目标出现在某一方向时, 眼睛眼跳到该目标的方向; (2) 整体上不偏好于产生某一方向的眼跳. 图 2 中的两个子图分别展示了起始状态的两个特点.

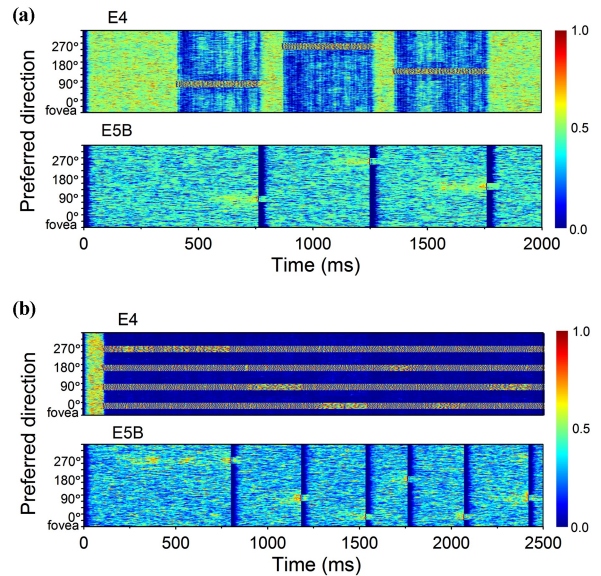


图 2 额叶视区网络模型的起始状态示例  
图中颜色代表神经元发放率

Fig.2 Two example the FEF neural network he colors represent the neuronal firing rate

在图 2(a) 中, 模型首先注视中心点 400 ms, 然后视觉目标先后在  $90^\circ$ ,  $270^\circ$  和  $150^\circ$  方向出现, 它们都能引起 E4 中相应的方向偏好神经元的持续活动, 并将神经冲动经由其他层级传到 E5B. 当 E5B 中方向偏好神经元群体的发放率达到了阈值



时,模型就会产生眼跳. 最后的结果显示这三个方向偏好刺激全都引起了相应的正向眼跳.

图 2(b) 展示了当视觉输入同时出现在  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  四个方向时模型的活动. E4 中  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  这四类方向偏好神经元都产生了持续的感觉信息处理活动,表明其同时处理来自四个方向的视觉输入. 从图 2(b) 的下图可以观察到模型先是眼跳到  $270^\circ$  方向,然后是  $90^\circ$ ,之后分别是  $0^\circ$ ,  $180^\circ$ ,  $0^\circ$  和  $90^\circ$ . 眼跳的位置可以落在四个视觉目标中的任意一个,没有任何一个目标有足够的吸引力让模型一直注视着它. 结果表明模型整体上不偏好于产生某一方向的眼跳.

### 2.2 额叶视区网络模型的训练过程

Non-choice 任务并没有一个特定的眼跳目标,影响训练结果的主要因素是奖励的大小. 事实上,近年来的一些实验成果已经表明了不同的决策-奖励联合会引导动物根据不同的奖励差异来做出相应决策<sup>[45,48-51]</sup>. 目前的问题是改进后的额叶视区网络模型是否能够根据奖励的差异来做出最佳的选择. 本次训练任务仍用  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  四个方向的目标同时刺激网络模型,并将  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  方向的眼跳分别以从多到少地与不同的奖励相关联. 为了更好地捕捉训练过程中的一些性质,我们设定一个包含 50 次试验的窗口,将 5000 次训练分成 100 个试验窗口. 首先,我们计算每个实验窗口中 E5R 的平均群体发放率,其代表神经网络在这段时期所做出的决策分布. 图 3 显示了 E5R 中的  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  偏好群体的平均群体发放率变化. 通过对比可以看出,在早期的试验窗口中,这四个群体的活动分布比较平均,并没有哪个群体的活动占主要优势,表明模型并没有选定最终的决策目标. 随着实验窗口的增加, $0^\circ$  偏好群体的平均发放率逐渐增强,表现出较其他群体更强的活动,并在训练的后后期变为主要的活动,而  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  偏好群体的响应下降到一个相对较低的水平,表明神经网络在这时期主要的眼跳决策集中在与最大的奖励相关的  $0^\circ$  方向上. 其次,更精细的训练过程可以用眼跳百分比 (saccade percent probability, SPP). SPP 的定义是一个窗口中正确试验的次数和全部试验次数的比值. 图 4(a) 中的结果为这四个方向的决策之间的竞争提供了一个更直观的说明. 根据 SPP 的变化可以把训练过程分为三个

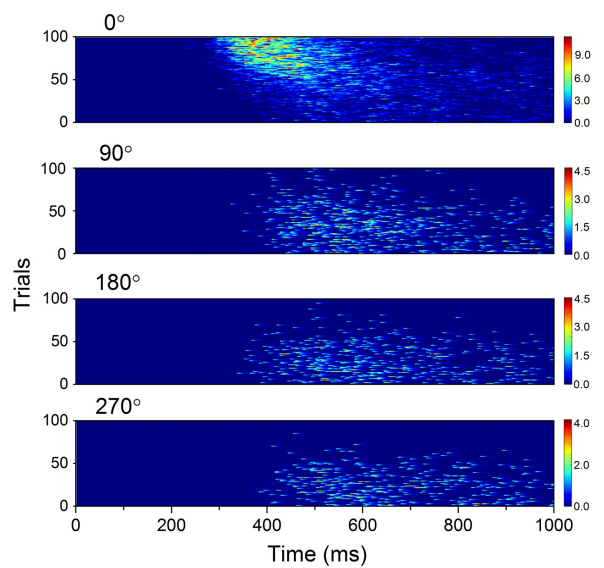


图 3 神经网络模型在训练过程中的平均群体发放率变化. 从上到下分别指的是 E5R 中的  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  偏好群体. Fig. 3 Average population firing rate of different populations in the training process top to down are  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  preferred populations of E5R

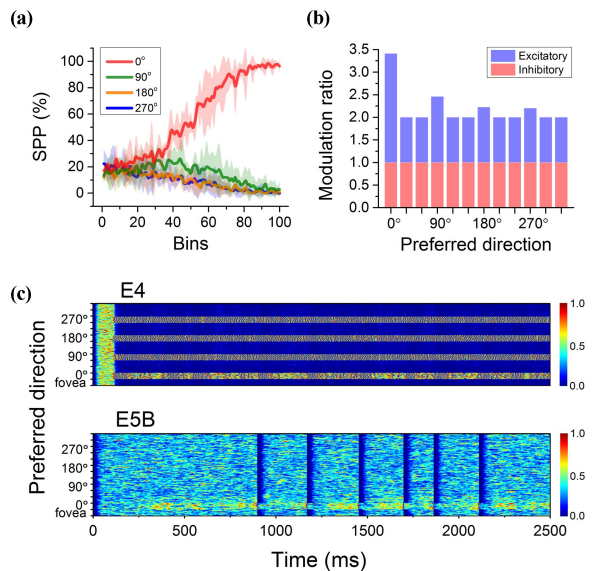


图 4 (a)  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  方向的眼跳百分比随着试验窗口的变化, (b) 连接着两类可塑性突触的不同方向偏好神经元群体的调节率, (c) 一个例子显示训练后的网络模型已经形成了与最大奖励相关的决策偏好.

Fig.4 (a) Variation of SPP in  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  directions, (b) Modulation rate of different plastic synapses, (c) An example the activities of the trained network

阶段. 第一个阶段是试验窗口 0-20, 四个方向偏好群体的 SPP 保持在 15% 到 25% 之间, 四类决策相互竞争, 但是没有其中一个成为胜者. 第二个阶段是 20-35 试验窗口. 此时,  $180^\circ$ ,  $270^\circ$  偏好群体的

SPP 呈下降趋势,并少于  $0^\circ, 90^\circ$  偏好群体的 SPP. 这个阶段表现为  $0^\circ$  与  $90^\circ$  眼跳决策的竞争,而  $180^\circ, 270^\circ$  决策已经成为败者. 第三个阶段是 35–100 试验窗口,  $0^\circ$  偏好群体的 SPP 迅速上升,而  $90^\circ$  偏好群体的 SPP 开始逐渐下降,  $0^\circ$  眼跳决策成为最终胜者. 这三个阶段的形成是由于与决策相关联的梯度奖励对 L4 和 L23 之间的连接产生了不同的兴奋性可塑性调节(图 4(b)). 通过积累这些不平衡的兴奋性调节使得与最大奖励相关的决策渐渐地赢得了竞争,变为主要的决策,而相对较低的奖励关联的决策最终输掉了竞争. 值得注意的是,  $90^\circ, 180^\circ, 270^\circ$  眼跳决策在竞争中落败的顺序也与关联奖励的大小相关. 一个决策关联的奖励越少,在竞争中输的越快. 最后,我们用与图 2(b) 相同的试验流程来验证训练后的模型的任务表现. 从图 4(c) 可以看出,当  $0^\circ, 90^\circ, 180^\circ, 270^\circ$  方向同时出现视觉刺激时,模型不再随意选择眼跳方向,而是只眼跳到  $0^\circ$  方向的目标,证明训练成功地将眼跳集中在与最大奖励关联的目标.

### 2.3 奖励差别对任务决策过程的影响

在 non-choice 任务中,一个非常关键的问题是决策之间的奖励差别是如何影响决策过程的. 为了定量地回答这个问题,我们将奖励因子设为  $q_i = q_0 + k_i RD$ . 其中,  $i$  为受刺激目标  $0^\circ, 90^\circ, 180^\circ, 270^\circ$ , 对应的  $k_i$  分别为 3, 2, 1, 0,  $q_0 = 0.0002$  而  $RD$  代表奖励差. 我们共进行 20 次任务训练,每次训练测试不同的  $RD$  下网络模型的决策情况. 在这些训练中,我们仍然将实验分为 100 个窗口,并计算每个窗口中不同眼跳的 SPP. 一旦其中一个决策的 SPP 超过 80%, 该决策就成为胜者,网络就被认定为偏好于做出这一决策;当 SPP 低于 5%, 相应的决策就认为是败者. SPP 超过 80% 或低于 5% 时所处于的试验窗口就是决策窗口(decision bins, DB). 奖励差与 DB 之间的关系描述在图 5 中,值得注意的是 20 次训练中胜者都是与最大奖励关联的  $0^\circ$  方向决策. 图 5 的结果显示,不管是决策中的胜者或是败者,它们的决策窗口都随着奖励差的增加而逐渐提前,表明决策之间的奖励差形成一种驱动力并控制着学习的速度,奖励差越大学习速度越快. 根据图 5 中的曲线,我们能从中总结出由不同奖励差所形成的两个学习阶段:慢速学习阶段( $0.00005 \leq RD < 0.00025$ )和快速学习阶段( $0.00025 \leq RD \leq 0.001$ ).

胜者决策在快速学习阶段的平均决策窗口要比慢速学习阶段的少 69.8%. 此外,网络模型在这些训练中的表现显示 non-choice 任务中还有两方面是值得注意的. 第一,当  $RD$  较少时, DB 的误差非常大,说明较低的  $RD$  并没有形成足够强大的驱动力使得模型能稳定地做出决策. 第二,  $0^\circ$  和  $90^\circ$  方向的决策窗口非常地接近,表明多个选择中的胜者和最后一位败者在较接近的训练阶段中出现,此后模型已经掌握了通过奖励差别引导的决策规则.

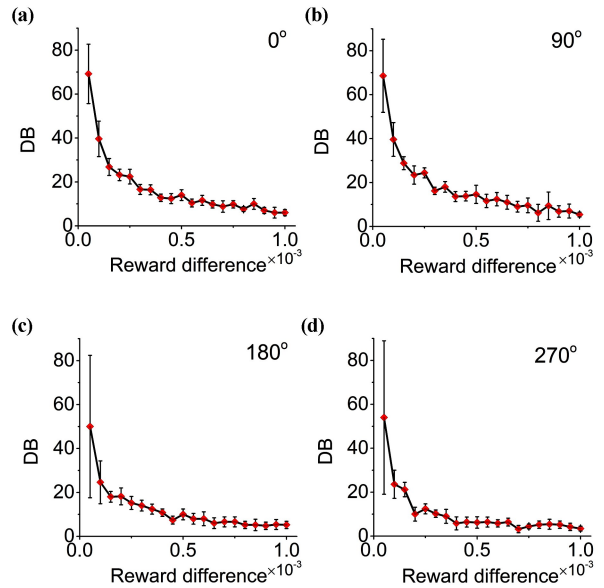


图 5  $0^\circ, 90^\circ, 180^\circ, 270^\circ$  决策在不同的奖励差下对应的决策试验窗口 (DB)

Fig.5 Corresponding decision bins at different amount of reward in  $0^\circ, 90^\circ, 180^\circ, 270^\circ$  decision

### 3 总结

本研究将一个经典的额叶视区神经网络模型改进为基于学习的模型,并用于研究大脑中的多目标决策机制. 模型一开始只能产生反射性眼跳,且对任何一个方向都没有偏好. 经过训练后,模型能成功地完成一个经典的认知决策任务——non-choice 任务,将眼跳目标集中在与最大奖励关联的眼跳决策中. 这个过程是由于训练过程中奖励对网络模型的兴奋性调节作用所决定的. 此外,不同决策间的奖励差对决策过程也有非常重要的影响. 但这种作用并不是线性的,不同关联奖励差能使模型进入两种学习阶段:快速学习阶段和慢速学习阶段. 这些结果仍需在真实的生物实验中作进一步的验证.

## 参 考 文 献

- 1 Salinas E. Context-dependent selection of visuomotor maps. *BMC Neuroscience*, 2004,5:47~69
- 2 Noorani I. LATER models of neural decision behavior in choice tasks. *Frontiers in Integrative Neuroscience*, 2014, 8:67~76
- 3 Drugowitsch J, Moreno-Bote R, Churchland A K, et al. The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience*, 2012,32(11):3612~3628
- 4 Cutsuridis V, Kumari V, Ettinger U. Antisaccade performance in schizophrenia: a neural model of decision making in the superior colliculus. *Frontiers in Neuroscience*, 2014, 8:13~26
- 5 Pleger B, Ruff C C, Blankenburg F, et al. Neural coding of tactile decisions in the human prefrontal cortex. *Journal of Neuroscience*, 2006,26(48):12596~12601
- 6 Chaumon M, Kveraga K, Barrett L F, et al. Visual predictions in the orbitofrontal cortex rely on associative content. *Cerebral Cortex*, 2014,24(11):2899~2907
- 7 Kan J Y, Niel U, Dorris M C. Evidence for a link between the experiential allocation of saccade preparation and visuospatial attention. *Journal of Neurophysiology*, 2012, 107(5):1413~1420
- 8 Drea C M, Wallen K. Low-status monkeys “play dumb” when learning in mixed social groups. *Proceedings of the National Academy of Sciences of the United States of America*, 1999,96(22):12965~12969
- 9 Platt M L, Glimcher P W. Neural correlates of decision variables in parietal cortex. *Nature*, 1999, 400(6741):233~238
- 10 Schlag Rey M, Amador N, Sanchez H, et al. Antisaccade performance predicted by neuronal activity in the supplementary eye field. *Nature*, 1997,390(6658):398~401
- 11 Hutton S B. Cognitive control of saccadic eye movements. *Brain Cogn*, 2008,68(3):327~340
- 12 Bell A, Everling S, Munoz D. Influence of stimulus eccentricity and direction on characteristics of pro- and antisaccades in non-human primates. *Journal of Neurophysiology*, 2000,84(5):2595~2604
- 13 Munoz D P, Everling S. Look away: the anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, 2004,5(3):218~228
- 14 Baldassarre G, Mannella F, Fiore V G, et al. Intrinsically motivated action-outcome learning and goal-based action recall: a system-level bio-constrained computational model. *Neural Networks*, 2013,41:168~187
- 15 Brown J W, Bullock D, Grossberg S. How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks*, 2004,17(4):471~510
- 16 Luhmann C C, Chun M M, Yi D J, et al. Neural Dissociation of Delay and Uncertainty in Intertemporal Choice. *Journal of Neuroscience*, 2008,28(53):14459~14466
- 17 Bizzi E. Discharge of frontal eye field neurons during eye movements in unanesthetized monkeys. *Science*, 1967,157(3796):1588~1590
- 18 Dias E C, Segraves M A. Muscimol-induced inactivation of monkey frontal eye field: effects on visually and memory-guided saccades. *Journal of Neurophysiology*, 1999, 81(5):2191~2214
- 19 Bichot N P, Schall J D, Thompson K G. Visual feature selectivity in frontal eye fields induced by experience in mature macaques. *Nature*, 1996,381:697~699
- 20 Chen L L, Wise S P. Neuronal activity in the supplementary eye field during acquisition of conditional oculomotor associations. *Journal of Neurophysiology*, 1995, 73(3):1101~1121
- 21 Tseng P, Chang C F, Chiau H Y, et al. The dorsal attentional system in oculomotor learning of predictive information. *Frontiers in Human Neuroscience*, 2013,7:404~412
- 22 Lewis C M, Baldassarre A, Comitteri G, et al. Learning sculpts the spontaneous activity of the resting human brain. *Proceedings of the National Academy of Sciences*, 2009,106(41):17558~17563
- 23 Lee K M, Keller E L. Neural activity in the frontal eye fields modulated by the number of alternatives in target choice. *Journal of Neuroscience*, 2008, 28(9):2242~2251
- 24 Heinze J, Hepp K, Martin K A. A microcircuit model of the frontal eye fields. *Journal of Neuroscience*, 2007, 27(35):9341~9353
- 25 Douglas R J, Martin K A, Whitteridge D. An intracellular analysis of the visual responses of neurones in cat visual cortex. *The Journal of Physiology*, 1991,440:659~696
- 26 Li C Y, Creutzfeldt O. The representation of contrast and other stimulus parameters by single neurons in area 17 of the cat. *Pflugers Archiv: European journal of Physiology*,

- 1984,401(3):304~314
- 27 Schiller P H, Finlay B L, Volman S F. Quantitative studies of single-cell properties in monkey striate cortex. I. Spatio-temporal organization of receptive fields. *Journal of Neurophysiology*, 1976,39(6):1288~1319
- 28 Ringach D L, Shapley R M, Hawken M J. Orientation selectivity in macaque V1: Diversity and Laminar dependence. *Journal of Neuroscience*, 2002,22(13):5639~5651
- 29 Nowak L G, Sanchez-Vives M V, McCormick D A. Lack of orientation and direction selectivity in a subgroup of fast-spiking inhibitory interneurons: cellular and synaptic mechanisms and comparison with other electrophysiological cell types. *Cerebral Cortex*, 2008,18(5):1058~1078
- 30 Hansel D, van Vreeswijk C. The mechanism of orientation selectivity in primary visual cortex without a functional map. *Journal of Neuroscience*, 2012,32(12):4049~4064
- 31 Pfeiffer M, Nessler B, Douglas R J, et al. Reward-Modulated Hebbian learning of decision making. *Neural Computation*, 2010,22(6):1399~1444
- 32 Engel T A, Wang X J. Same or different? A neural circuit mechanism of similarity-based pattern match decision making. *Journal of Neuroscience*, 2011,31(19):6982~6996
- 33 Ardid S, Wang X J. A tweaking principle for executive control: neuronal circuit mechanism for rule-based task switching and conflict resolution. *Journal of Neuroscience*, 2013,33(50):19504~19517
- 34 Gottlieb J, Hayhoe M, Hikosaka O, et al. Attention, reward, and information seeking. *Journal of Neuroscience*, 2014,34(46):15497~15504
- 35 Gold J I, Shadlen M N. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, 2002,36(2):299~308
- 36 Gold J I, Shadlen M N. The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. *The Journal of Neuroscience*, 2003,23(2):632~651
- 37 Law C T, Gold J I. Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nature Neuroscience*, 2009,12(5):655~663
- 38 Connolly P M, Bennur S, Gold J I. Correlates of perceptual learning in an oculomotor decision variable. *Journal of Neuroscience*, 2009,29(7):2136~2150
- 39 Ardid S, Wang X J, Compte A. An integrated microcircuit model of attentional processing in the neocortex. *Journal of Neuroscience*, 2007,27(32):8486~8495
- 40 Destexhe A, Rudolph M, Fellous J M, et al. Fluctuating synaptic conductances recreate in vivo-like activity in neocortical neurons. *Neuroscience*, 2001,107(1):13~24
- 41 Salinas E. Background synaptic activity as a switch between dynamical states in a network. *Neural Computation*, 2003,15(7):1439~1475
- 42 Moreno-Bote R, Parga N. Role of synaptic filtering on the firing response of simple model neurons. *Physical Review Letters*, 2004,92(2):028102
- 43 Moreno-Bote R, Parga N. Response of integrate-and-fire neurons to noisy inputs filtered by synapses with arbitrary timescales: Firing rate and correlations. *Neural Computation*, 2010,22(6):1528~1572
- 44 Ye W, Liu S, Liu X, et al. A neural model of the frontal eye fields with reward-based learning. *Neural Networks*, 2016,81:39~51
- 45 Levy I, Lazzaro S C, Rutledge R B, et al. Choice from non-choice: predicting consumer preferences from blood oxygenation level-dependent signals obtained during passive viewing. *Journal of Neuroscience*, 2011,31(1):118~125
- 46 王如彬, 张志康, 谢智刚等. 多个神经振荡器网络的相位动力学编码. *动力学与控制学报*, 2009,7(3):217~225 (Wang R B, Zhang Z K, Xie Z G, et al. Dynamic coding of phase on multi populations of neural oscillators. *Journal of Dynamics and Control*, 2009,7(3):217~225 (in Chinese))
- 47 徐旭颖, 王如彬. 神经元高低状态切换发放的神经动力学. *动力学与控制学报*, 2015,13(1):62~67 (Xu X Y, Wang R B. Neurodynamics of up and down transitions. *Journal of Dynamics and Control*, 2015,13(1):62~67 (in Chinese))
- 48 So N Y, Stuphorn V. Supplementary eye field encodes option and action value for saccades with variable reward. *Journal of Neurophysiology*, 2010,104(5):2634~2653
- 49 van der Togt C, Stănişor L, Pooresmaeili A, et al. Learning a new selection rule in visual and frontal cortex. *Cerebral Cortex*, 2016,26(8):3611~3626
- 50 Costello M G, Zhu D, May P J, et al. Task dependence of decision-and choice-related activity in monkey oculomotor thalamus. *Journal of Neurophysiology*, 2016,115(1):581~601
- 51 Schiffer A M, Muller T, Yeung N, et al. Reward activates stimulus-specific and task-dependent representations in visual association cortices. *Journal of Neuroscience*, 2014,



34(47):15610~15620

# RESEARCH ON LEARNING-BASED MULTIPLE CHOICE DECISION-MAKING MODEL OF BRAIN<sup>\*</sup>

Ye Weijie<sup>1</sup> Liu Shenquan<sup>2†</sup>*(1.School of Statistics and Mathematics,Guangdong University of Finance and Economics,Guangzhou 510320, China)**(2.School of Mathematics,South China University of Technology,Guangzhou 510640, China)*

**Abstract** Decision-making with multiple choices plays a key role on cognition. In this work, we extended a network model of Frontal eye field to a learning based model, and trained it to complete a cognitive task; non-choice task then used the simulation results to explain the cognitive process of multiple choice decision-making. After thousands of trainings, the network model changed from selecting target randomly into choosing the largest reward-related decision. In the training process, the sequence of multiple choice decision was related to the reward gradient. In addition, the reward differences between distinct decisions played an important role on the learning speed of the network model, making the model exhibit two learning phases: the fast learning phase and the slow learning phase.

**Key words** multiple choice decision-making, learning, saccade, frontal eye field

---

Received 28 December 2016, revised 1 April 2017.

<sup>\*</sup> The project supported by the National Natural Science Foundation of China(11572127).

<sup>†</sup> Corresponding author E-mail; mashqliu@scut.edu.cn