中国科学院大学
University of Chinese Academy of Sciences

系统与计算神经科学小组作业

基于 RSNN 的模拟小鼠决策算法

学生姓名	马国庆	**学号**	202218020428003
学生姓名	林楚儿	**学号**	202218020428002
学生姓名	张予涵	**学号**	202218014628014
学生姓名	魏雅轩	**学号**	202218020415011
学生姓名	江德杨	**学号**	202218020428001

中国科学院大学制

【人员分工】

马国庆：代码编写，实验数据生成；

林楚儿、张予涵：网络和优化算法设计；

魏雅轩：小鼠实验环境设计建模；

江德杨：实验数据整理和分析。

【摘要】

本项目基于 RSNN 模拟小鼠决策任务。程序模拟了小鼠的感知、工作记忆和决策实验。实验设定小鼠在丁字形通道内，经过带有提示信息的通道之后继续前进一段距离，最终在路口处决定左转或右转。在小鼠完成决策后，通过奖励信息对其行为进行训练。本程序通过计算建模对此实验进行模拟：搭建 RSNN 循环脉冲神经网络模拟小鼠大脑活动，通过监督训练模拟小鼠的学习行为，最终使得网络具有感知、工作记忆和决策能力。

一、背景

人工智能广泛使用的最大障碍之一是人工神经网络的学习活动会产生巨大的能耗。而解决这一问题的一种方法是从大脑中获得灵感。大脑神经元之间可以通过短的电脉冲或尖峰进行有效传输，因而极大地节省了能量。此外，脉冲神经网络的信息传递方式与生物脑相当接近，用其进行建模有利于探究大脑的运行机制。因此本次课程项目中，我们使用循环脉冲神经网络模拟大脑的感知、记忆和决策过程。

如图 1 所示，一个简单的循环神经网络（Recurrent Neural Network，RNN）由输入层 x、隐含层 s 和输出层 y 构成，至少包含一个反馈连接。其中 U 为输入层到隐含层的连接权重矩阵，V 为隐含层到输出层的连接权重矩阵，W 为隐含层神经元之间的连接权重矩阵。因此，网络的激励可以沿着一个 loop 进行流动。这种网络结构的处理单元之间既有内部的反馈连接又有前向连接，比前馈神经网络具有更强的动态行为和计算能力，可以利用内部记忆来处理任意时序的输入序列。这种特点使它可以更容易地处理如不分段的手写识别、语音识别等。

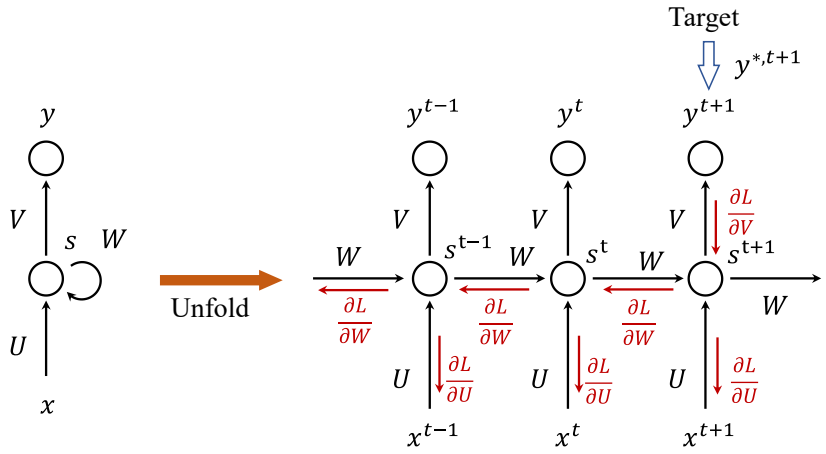

图 1 RNN 的网络结构

RNN 网络可以采用通过时间反向传播（Back Propagation Through Time，BPTT）算法按照以下步骤进行训练：①前向计算每个神经元的输出值；②反向计算每个神经元的误差项值，它是损失函数 L 对各个神经元的加权和的偏导数；③计算每个权重的梯度；④利用随机梯度下降算法沿时间轴更新权重。BPTT 算法的核心任务是计算目标函数对各参数的导数：$\frac{\partial L}{\partial W}$、$\frac{\partial L}{\partial U}$和$\frac{\partial L}{\partial V}$。

然而，RNN 网络和 BPTT 算法难以在生物学上实现：RNN 网络是以数值的形式进行传播，而大脑是通过短的电脉冲或尖峰进行传播。在此基础上，RNN 网络采用脉冲神经元模型得到循环脉冲神经网络（Recurrent Spiking Neural Network，RSNN）。RSNN 网络是在 BPTT 算法的基础上采用了梯度替代进行训练。但由于 BPTT 需要不停追溯之前的时间点直到初始时间点才更新权值，层数会随着时序的展开而加深，这将会出现梯度消失和训练时间太长的问题，给网络训练带来了困难，也无助于我们理解大脑中的学习过程。因此，如何优化 BPTT，使其在缩短训练时长的基础上，更具有类脑性能成为了本次课程项目的目标。

二、建模小鼠及实验环境设计

我们建立了一个 T 形迷宫的虚拟环境进行决策任务。如图 2 所示，当小鼠在 T 形迷宫中沿着主干臂移动时，将接收到白底黑点或者黑底白点的视觉线索（前者对应左转，后者对应右转）。当它到达 T 形路口时，它将根据之前看到的视觉线索决定转向左臂还是右臂，当其转向正确对应的方向时，将得到糖水奖励。

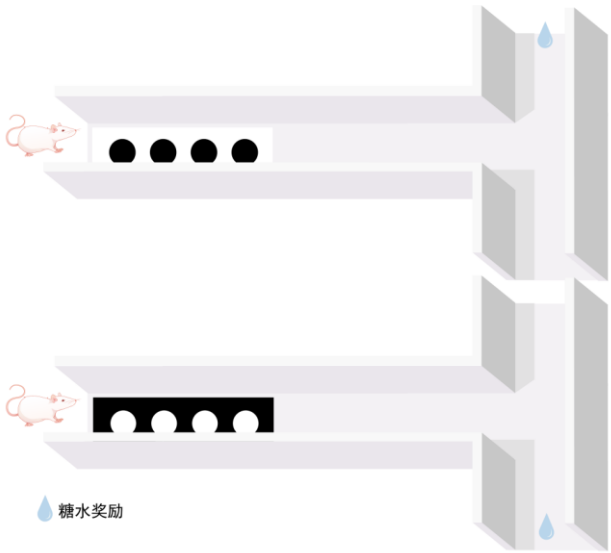

图 2 小鼠的决策任务

而后我们将此决策任务转化为监督学习的 SNN，模拟这一过程的认知计算。如图 3 所示，试验的前 156ms，计算机接收模拟视觉线索的输入，经过 121ms 的延迟期（Delay Period）进行认知计算，最后决策转向左臂或右臂，不同于小鼠正确决策后会被给予奖励，计算机在每次试验结束之后会被告知选择是否正确。该网络模拟了小鼠的感知、工作记忆和决策行为。

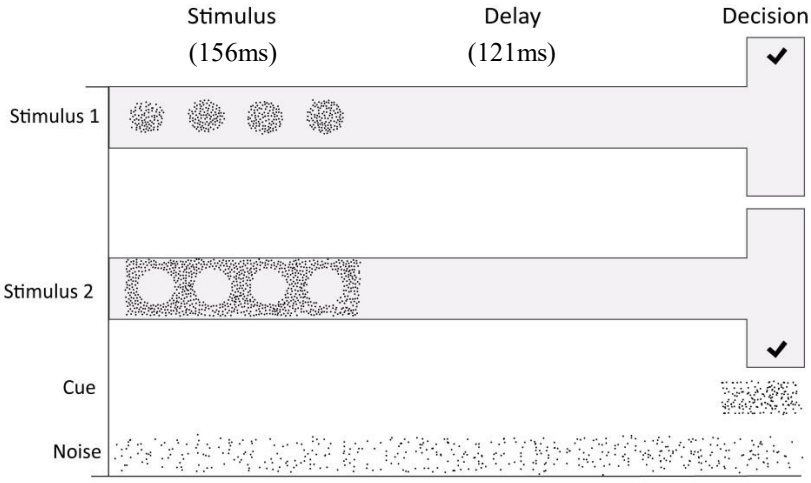

图 3 实验环境设计

图 4 为网络的模型构建。我们由小鼠大脑信息决策过程建模了一个监督 SNN 任务。视网膜结构接收视觉信息，传递到初级视觉皮层，通过背腹侧通路进行信息处理和决策，决策信息通过脊髓下行支配肌肉，做出运动行为。神经网络的输入层模拟视觉信息的接收，隐藏层对应大脑的视觉信息处理过程，输出层则体现决策结果。

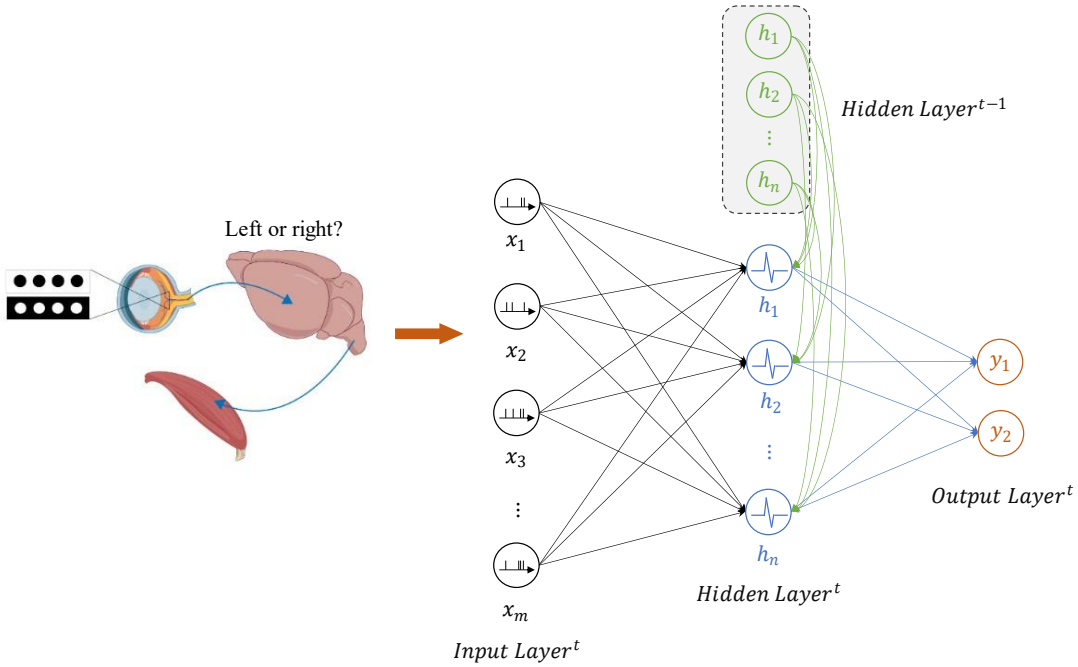

图 4 网络模型构建

三、e-prop 算法

来自神经科学的两个实验数据流提供了大脑中在线网络学习的线索：①大脑中的神经元在分子水平上保持先前活动的痕迹，例如，以钙离子或激活的 CaMKII 酶的形式。值得注意的是，它们对突触前神经元先于突触后神经元放电的事件保持着衰退的记忆，如果随后出现自上而下的学习信号，会诱导突触可塑性。此类痕迹通常被称为资格迹。②在大脑中，存在大量自上而下的信号，如多巴胺、乙酰胆碱和与错误相关的消极性相关的神经放电，这些信号将行为结果告知局部的神经元群体。此外，研究者们已经发现多巴胺信号对不同的神经元目标群体具有特异性，而不是全局性的。在我们的学习模型中，我们将这种自上而下的信号称为学习信号。

通过查阅文献，我们在参考文献[1]中学习了一种 e-prop 算法，其将局部资格迹和自上而下的学习信号进行组合，而不需要花费一定时间对信号进行反向传播。如图 5 所示，用 e-prop 算法按以下步骤训练 RSNN 网络：①前向计算资格迹e^{t+1}；②根据输出神经元在时间 t+1 的实际输出y^{t+1}与其给定目标值$y^{*,t+1}$的偏差计算近似的学习信号L^{t+1}；③根据资格迹e^{t+1}和近似学习信号L^{t+1}产生瞬时权重变化ΔW^{t+1}并更新权重 W。接下来我们将详细介绍 e-prop 算法。

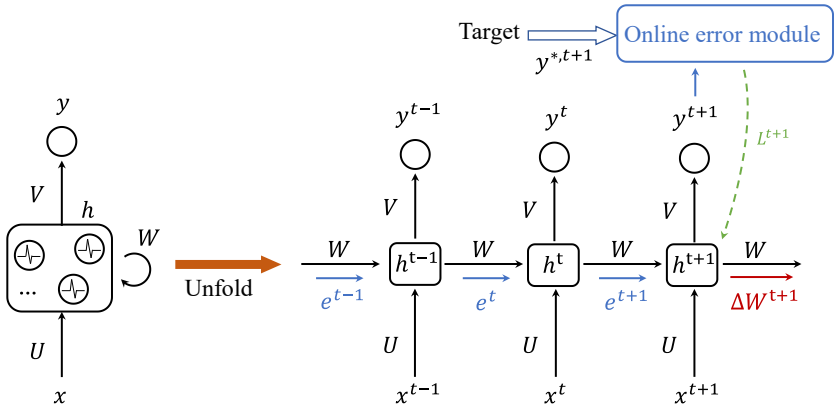

图 5 用 e-prop 算法训练 RSNN 网络结构

在 RSNN 网络结构中，用一个二进制变量z_j^t表示神经元发放的尖峰信号，如果神经元 j 在时间 t 激发，则z_j^t为 1，否则为 0。在模型中，通常让 t 在小的离散时间步长变化，例如 1ms。网络学习的目标是找到使给定损失函数 E 最小化的突触权重 W。损失函数 E 在回归或分类学习的情况下测量每个输出神经元 k 在时间 t 的实际输出y_k^t与其给定目标值$y_k^{*,t}$的偏差；在强化学习（Reinforcement Learning，RL）中则衡量当前决策行为的不足，以实现获取最多奖励（Reward）的目的。

从神经元 i 到神经元 j 的突触的权重W_{ji}的梯度$\frac{dE}{dW_{ji}}$告诉我们应该如何改变该权重以减少 E。尽管隐式离散变量z_j^t不可微，原则上在取尖峰合适伪导数的情况下是可以估计该梯度的，。如公式（1）所示，e-prop 算法的关键是将梯度$\frac{dE}{dW_{ji}}$表示为 RSNN 计算的时间步长 t 上的乘积之和。其中，第二个因子是不依赖于 E 的局部梯度：

$$\frac{dE}{dW_{ji}} = \sum_t \frac{dE}{dz_j^t} \cdot \left[\frac{dz_j^t}{dW_{ji}}\right]_{local} \tag{1}$$

该局部梯度被定义为关于神经元 j 在时间 t 的隐含层状态h_j^t与之前的时间步长的偏导数的乘积之和，可以在 RNN 的正向计算期间通过简单的递归来更新。它不是近似值，可以收集关于网络梯度$\frac{dE}{dW_{ji}}$的最大信息量。对于内部状态只有膜电位的简单神经元模型来说，该局部梯度会是突触可塑性的资格迹：

$$e_{ji}^t \stackrel{\text{def}}{=} \left[\frac{dz_j^t}{dW_{ji}}\right]_{local} \tag{2}$$

但大多数生物神经元都有额外的隐变量，这些变量在较慢的时间尺度上发

生变化，例如神经元的发放脉冲的阈值具有启动阈值适应性。此外，神经元中的这些较慢的过程对于获得与 LSTM 网络类似的强大计算能力至关重要。这种自适应神经元的资格迹e_{ji}^t的形式对于理解 e-prop 至关重要，它是提升 RSNN 计算能力的主要因素，这也可以通过生物学上合理的学习实现。

神经元 j 的学习信号表示为：

$$L_j^t \stackrel{\text{def}}{=} \frac{dE}{dz_j^t} \tag{3}$$

根据公式（1）、（2）和（3）可以推出：

$$\frac{dE}{dW_{ji}} = \sum_t L_j^t e_{ji}^t \tag{4}$$

该公式通过突触可塑性的局部资格迹来近似网络损失梯度：将步骤 t 中的每个权重W_{ji}与$-L_j^t e_{ji}^t$成比例地改变，或者将这些所谓的标签累积在隐变量中，该隐变量偶尔被转换为实际的权重变化。因此，从严格意义上讲，e-prop 是一种在线学习方法，可以在每个时间点更新权重；而不需要像 BPTT 算法，在训练完一段时间后才更新一次权重。所以与 BPTT 算法相比，e-prop 算法在生物学上更有可能实现。

由于学习信号L_j^t的理想值$\frac{dE}{dz_j^t}$还包含了神经元 j 的当前尖峰输出z_j^t可能通过其他神经元的未来尖峰对 E 所产生的影响，因此该精确值通常在时间 t 上不可导。因此，e-prop 算法用近似值$\frac{\partial E}{\partial z_j^t}$来代替学习信号的值，只关注尖峰$z_j^t$对损失函数 E 的直接影响。该近似仅考虑 RSNN 的输出神经元 k 当前产生的损失，并使用神经元特定权重B_{jk}将其与网络神经元 j 联系起来：

$$L_j^t = \sum_k B_{jk} \underbrace{\left(y_k^t - y_k^{*,t}\right)}_{\text{输出神经元}k\text{在时间}t\text{的偏差}} \tag{5}$$

尽管该近似学习信号L_j^t仅计算在当前时间步长 t 处出现的误差，但它在公式（4）中与资格轨迹e_{ji}^t相结合。该资格迹可以追溯到神经元 j 的过去，从而减轻了通过在时间上向后传播信号来解决时间信度分配问题的需要（如 BPTT）。e-prop 算法将在线学习信号的权重B_{jk}设置为从神经元 j 到输出神经元 k 的突触连接的相应权重W_{kj}^{out}。该学习信号将在网络没有重复连接的情况下实现$\frac{dE}{dz_j^t}$。

深度 RL 的在线突触可塑性规则如公式（6）所示，其类似于公式（4），不

同之处在于这里将衰减记忆滤波器$\mathcal{F}_\gamma$应用于术语$L_j^t\bar{e}_{ji}^t$，其中γ是未来奖励的折扣因子，$\bar{e}_{ji}^t$表示经过低通滤波后的资格迹e_{ji}^t。该项在突触可塑性规则中乘以奖励预测误差$\delta^t = r^t + \gamma V^{t+1} - V^t$，其中$r^t$是在时间 t 收到的奖励。这产生了瞬时权重变化：

$$\Delta W_{ji}^t = -\eta\delta^t\mathcal{F}_\gamma\left(L_j^t\bar{e}_{ji}^t\right) \tag{6}$$

之前 RL 的三因素学习规则通常为$\Delta W_{ji}^t = -\eta\delta^t\bar{e}_{ji}^t$，仅通过将网络神经元的输出与奖励预测误差相关来估计策略的梯度。由于所得梯度估计中的高噪声，已知这种方法的学习能力非常有限。相比之下，在基于奖励的 e-prop 的可塑性公式（6）中，低通滤波后的资格迹$\bar{e}_{ji}^t$先与神经元特定反馈L_j^t相结合，然后与奖励预测误差δ^t相乘。这产生了策略梯度和价值梯度的估计，与 BPTT 的深度 RL 中的估计类似。

四、结果分析

我们的网络采用40×1的向量作为每个时刻的输入，隐藏层包括 100 个 LIF 脉冲神经元节点。网络在前 154 个时间节点时获取视觉线索进行感知，然后进入工作记忆阶段，最后在 t=375 时获得决策信号并进行决策。

实验中使用泊松编码的方式生成圆点作为视觉线索，生成了包含 1000 个样本的训练集和 50 个样本的测试集进行训练，每轮训练中从训练集中取出 100 个样本进行训练并测试。

训练过程中的 loss 曲线和测试集准确率的变化情况如图 6 所示。可以看到，在训练了 300 个样本以后，我们的网络已经达到了 100%的准确率，可见我们的网络能够很好的完成所设定的任务。

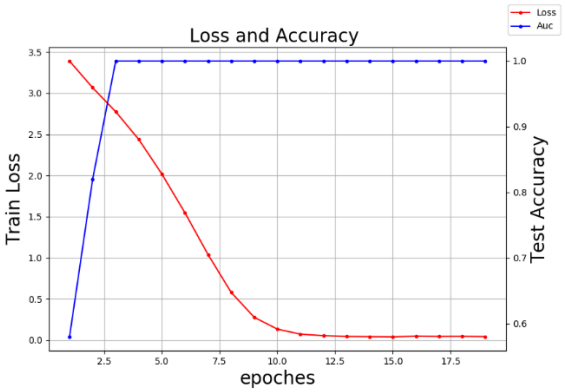

图 6 训练过程中的 loss 和准确率曲线

我们对训练过程中的神经元的脉冲传递过程展开了研究。图 7 展示了在训练过程中，各神经元的激活情况随时间的变化情况；图 8 展示了在训练了前后，各神经元的膜电位随时间的变化情况。从图 7 我们可以看到，激活的神经元数量随着训练的进行而不断增加，说明越来越多的神经元参与到我们的感知与记忆任务当中。同时我们可以发现，在开始的感知阶段，脉冲神经元相比后续的记忆阶段有着更高的活跃程度。

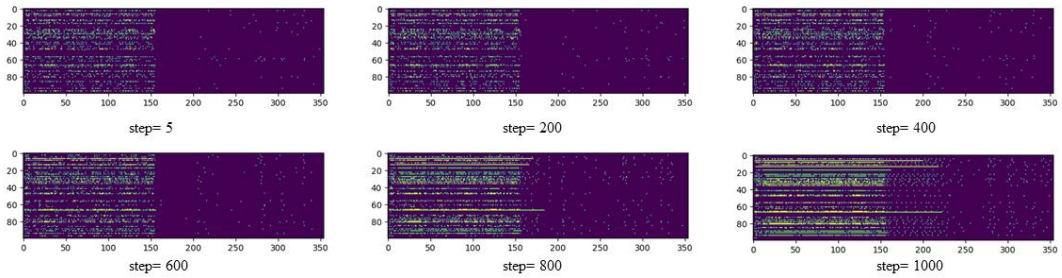

图 7 已训练 5~1000 个样本时的各神经元激活情况

（横坐标代表时间，纵坐标代表 100 个脉冲神经元）

对比图 8(a)(b)我们可以发现，在训练初期的记忆阶段，大部分的神经元膜电位的变化十分缓慢。而随着训练的进行，许多神经元膜电位在记忆阶段表现出了非常明显的持续波动。这说明此时尽管不再有视觉线索输入，但这些神经元依然能够记住先前时刻获取的信息，并将其向后不断传递。可以推断，我们的训练使得这些神经元拥有了记忆功能，能够记住一段时间之前获取的信息。

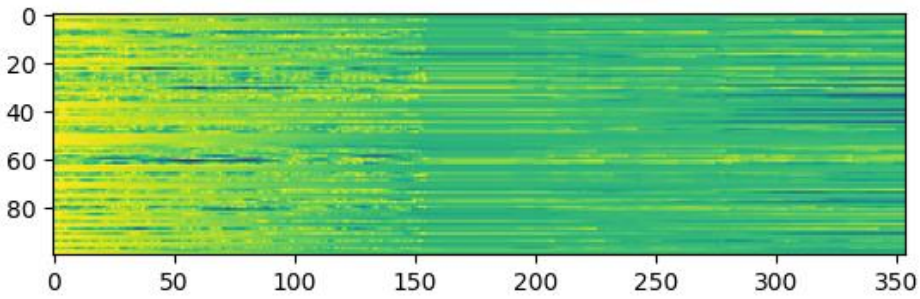

图 8(a) 训练开始时各神经元膜电位随时间的变化

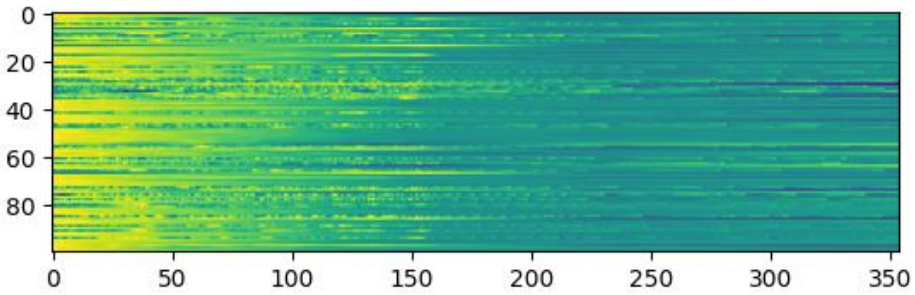

图 8(b) 训练结束时各神经元膜电位随时间的变化

五、未来展望

为了模拟真实的决策环境，我们设置了 121ms 的延时时间，使 RSNN 网络循环 12 次后再做出决策。虽然 RSNN 网络只有三层，但随着时序的展开而层数会加深，这将会出现梯度消失的问题，给网络训练带来了困难。

LSNN 网络是在 RSNN 网络基础上加入 LSTM 单元。LSTM 网络将 RNN 网络中隐含层的隐含单元设计为所谓的 LSTM 细胞单元。如图 9 所示，每个 LSTM 细胞含有与传统的 RNN 细胞相同的输入和输出，但额外包含一个控制信息流动的“门结点系统”。门系统包含三个部分，除了对 LSTM 细胞的输入、输出进行加权控制之外，还对记忆（遗忘）进行加权控制。LSTM 克服了 RNN 在长距离信息传递时的有限性，它会在之前的时间步中保留一些重要信息，遗忘一些不重要的信息。相比之下，加入了门系统的 LSNN 网络可以在一定程度上缓和 RSNN 网络的梯度消失或扩散的问题。因此，我们的进一步研究将着眼于如何结合 LSNN 网络和 e-prop 算法来更好地建模小鼠决策问题。

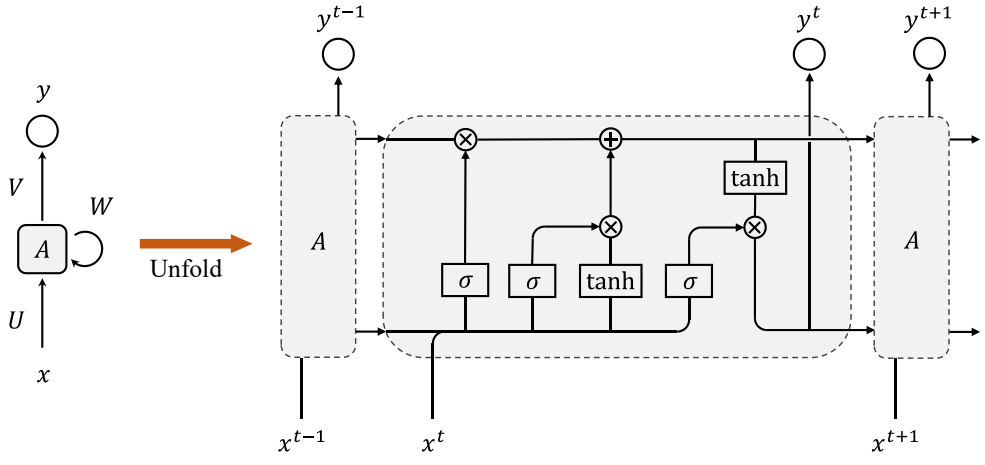

图 9 LSTM 网络结构

六、参考文献

[1] Bellec G, Scherr F, Subramoney A, et al. A solution to the learning dilemma for recurrent networks of spiking neurons[J]. Nature communications, 2020, 11(1): 1-15.

[2] ENGELHARD B, FINKELSTEIN J, COX J, et al. Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons[J]. Nature, 2019, 570(7762): 509-513.

[3] MORCOS A S, HARVEY C D. History-dependent variability in population dynamics during evidence accumulation in cortex[J]. Nature Neuroscience, 2016, 19(12): 1672-1681.