【封面】

【人员分工】

马国庆负责代码编写，实验数据生成，整合小组报告；林楚儿和张予涵负责网络和优化算法设计以及背景和未来展望的撰写；魏雅轩负责小鼠实验环境设计建模；江德杨负责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，先经过带有提示信息的通道，再经过没有信息的通道，最终路口处决定左转或右转。通过奖励信息对小鼠行为进行训练。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，极大节省了能量。此外，脉冲神经网络信息传递方式与生物脑接近，用其进行建模有利于探究大脑运行机制。因此本次大作业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、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。

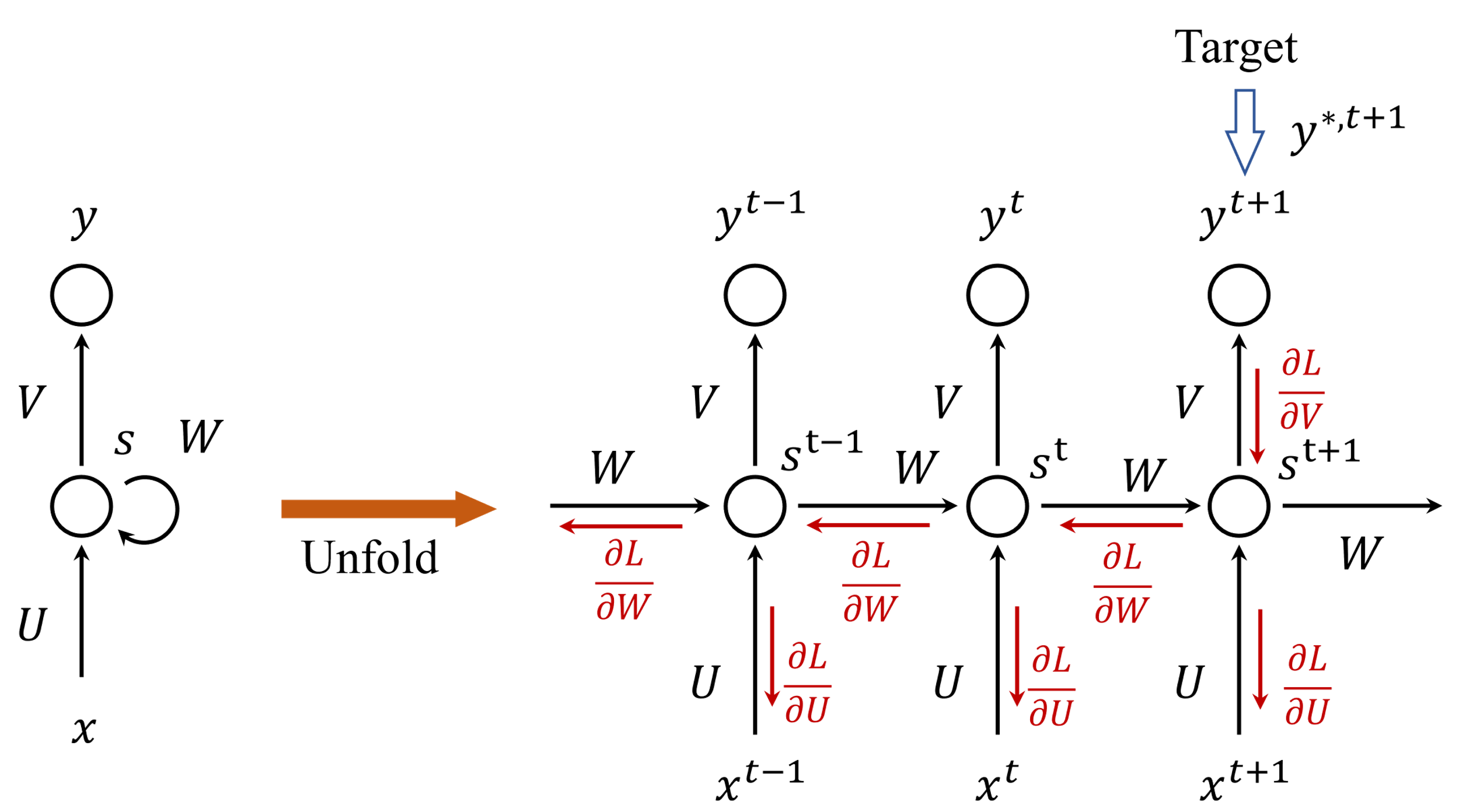


图1

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：、和。

RNN网络和BPTT算法难以在生物学上实现。因为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，所以在RNN网络基础上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。

二、建模小鼠及实验环境设计

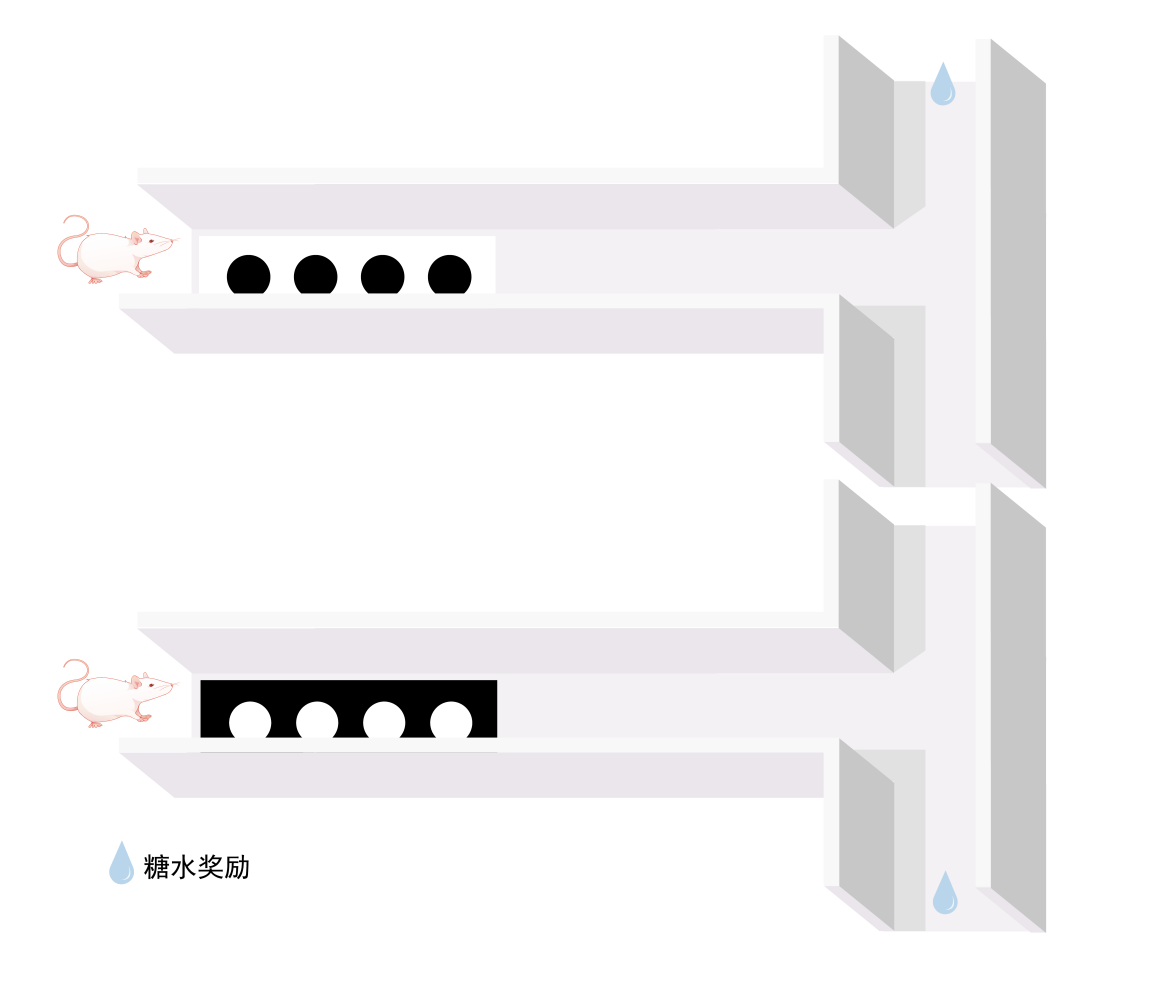


图2

我们建立了一个T形迷宫的虚拟环境进行决策任务。当小鼠在T形迷宫中沿着主干臂移动，将接收到白色背景黑色圆点或者黑色背景白色圆点的视觉线索（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。

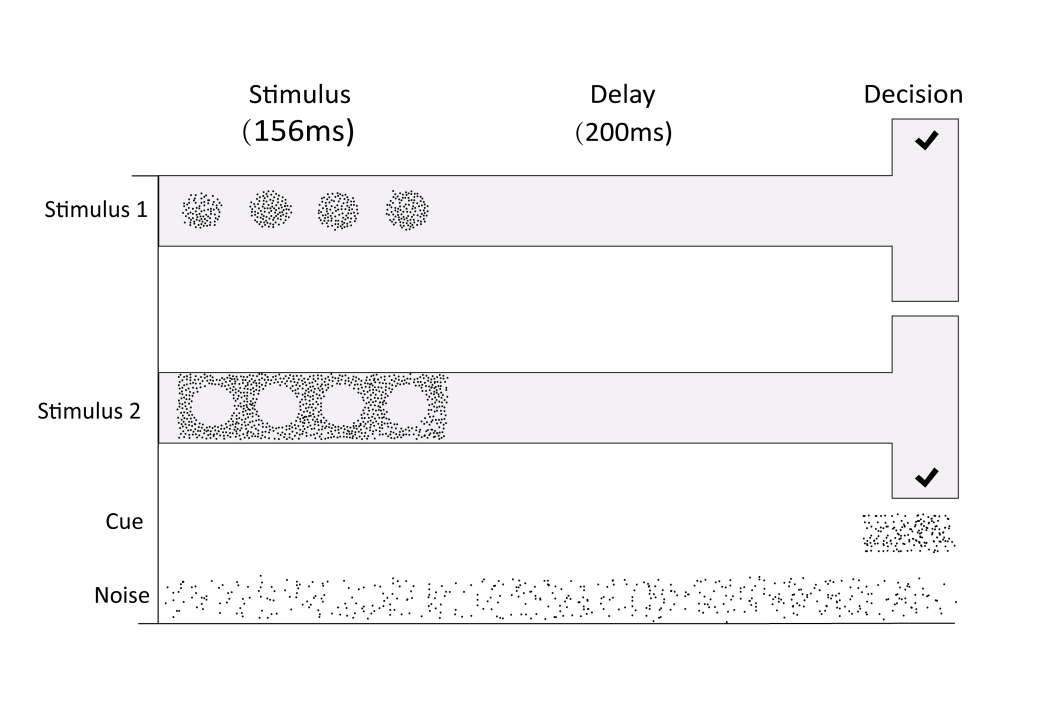
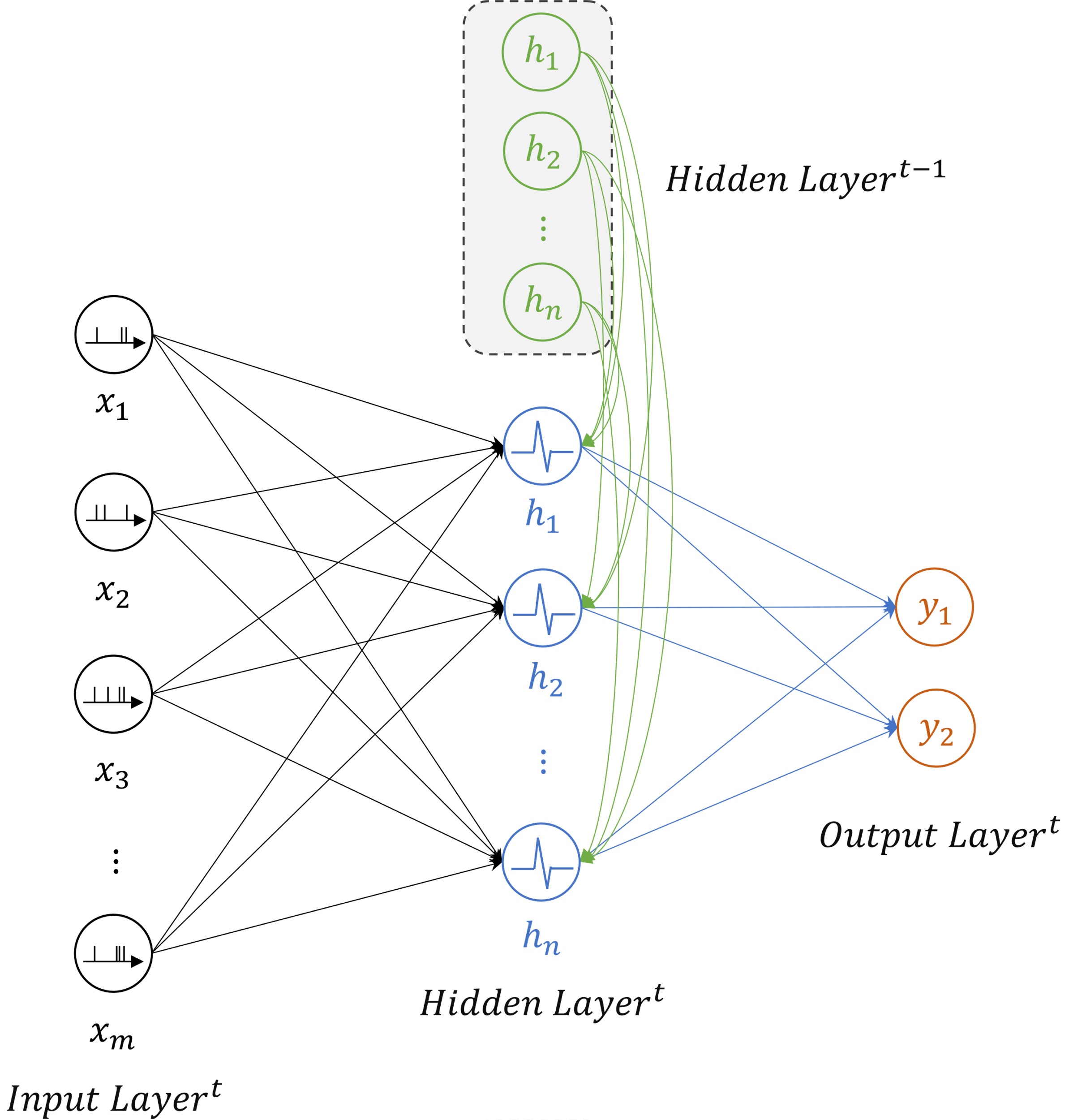
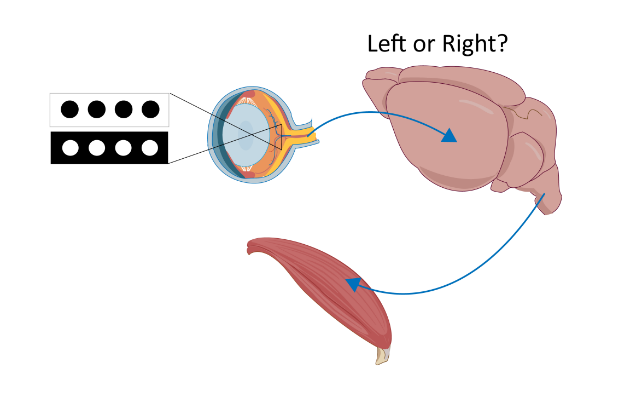


图3

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。试验的前156ms，计算机接收模拟视觉线索的输入，经过200ms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仅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。



4

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。

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。特别是，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，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]上学习了一种e-prop算法，将局部资格迹和自上而下的学习信号进行最佳组合，而不需要通过时间对信号进行反向传播。

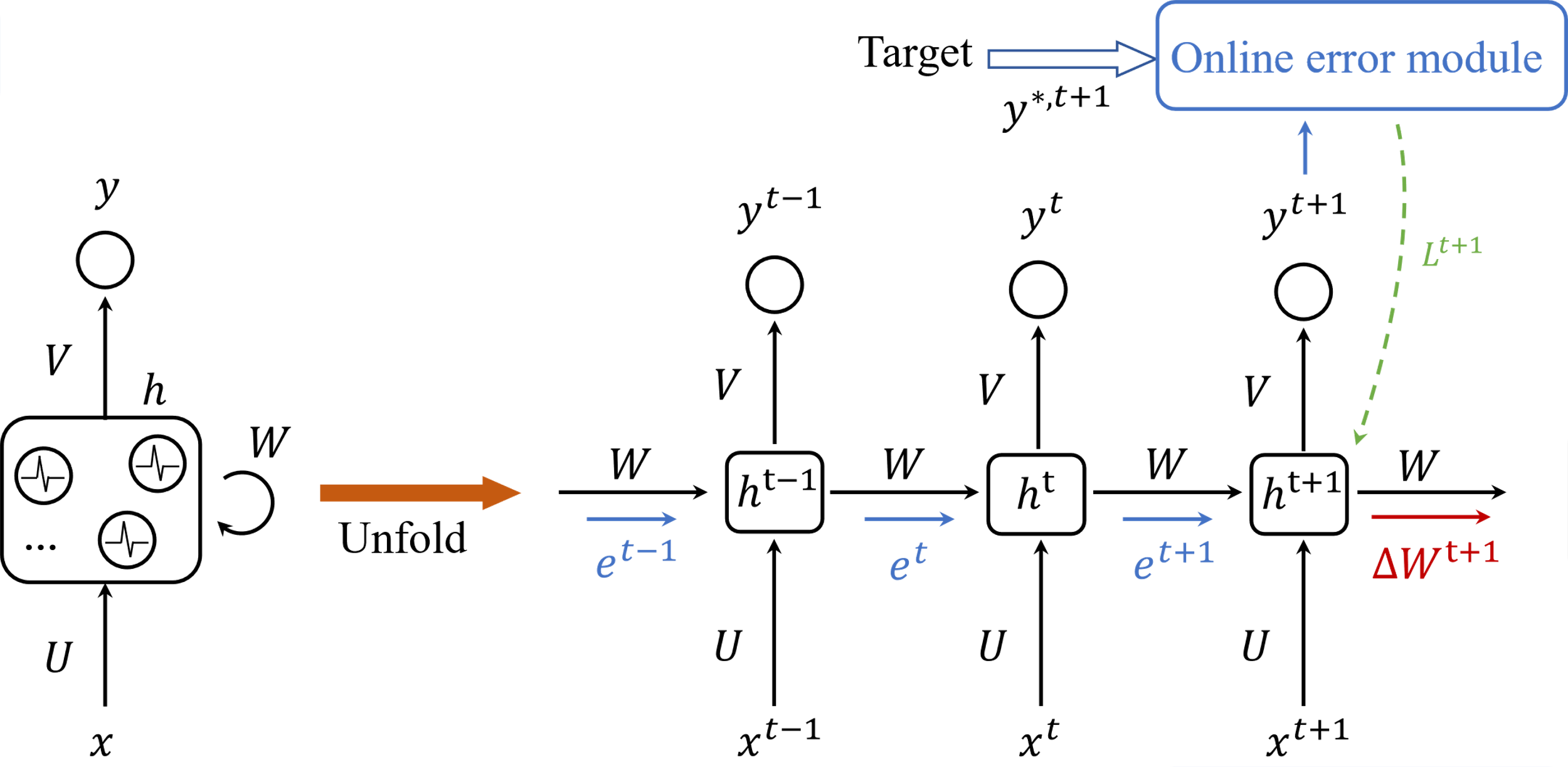


图5

用一个二进制变量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的实际输出与其给定目标值的偏差；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的目的。

从神经元i到神经元j的突触的权重的梯度告诉我们应该如何改变该权重以减少E。原则上，在尖峰合适伪导数时是可以估计该梯度的，尽管隐式离散变量不可微。e-prop算法的关键是梯度可以表示为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：

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的最大信息量。由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，所以定义该局部梯度为资格迹：

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的形式对于理解e-prop至关重要，它是RSNN计算能力质性跃升的主要驱动因素，这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：

根据等式（1）、（2）和（3）可以表示：

该方程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中的每个权重与成比例地改变，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，只有等一段时间训练完后才更新一次权重。所以与BPTT算法相比，e-prop算法在生物学上更有可能实现。

由于学习信号的理想值还包含了神经元j的当前尖峰输出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。e-prop算法用近似值来代替，比如，它忽略了这些间接影响，只关注尖峰对损失函数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将其与网络神经元j联系起来：

尽管该近似学习信号仅计算在当前时间步长t处出现的误差，但它在等式（4）中与资格轨迹相结合，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。该学习信号将在网络没有重复连接的情况下实现。

深度RL的在线突触可塑性规则如等式（6）所示，其类似于等式（4），不同之处在于这里将衰减记忆滤波器应用于术语，其中γ是未来奖励的折扣因子，表示经过低通滤波后的资格迹。该项在突触可塑性规则中乘以奖励预测误差，其中是在时间t收到的奖励。这产生了瞬时权重变化：

之前RL的三因素学习规则通常为，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的可塑性规则（6）中，低通滤波后的资格迹先与神经元特定反馈相结合，然后与奖励预测误差相乘。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，每个epoch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。

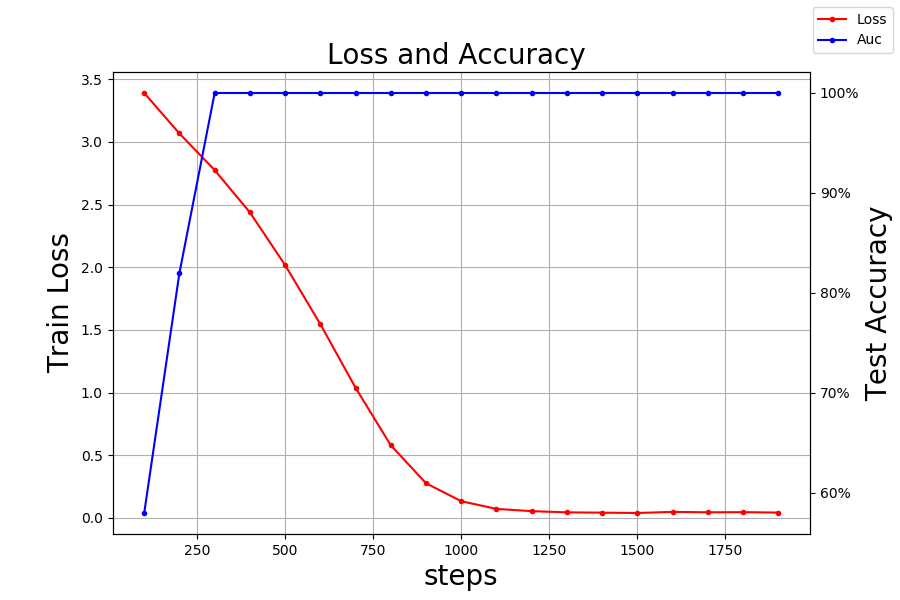


图6 训练过程中的loss和准确率曲线

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。

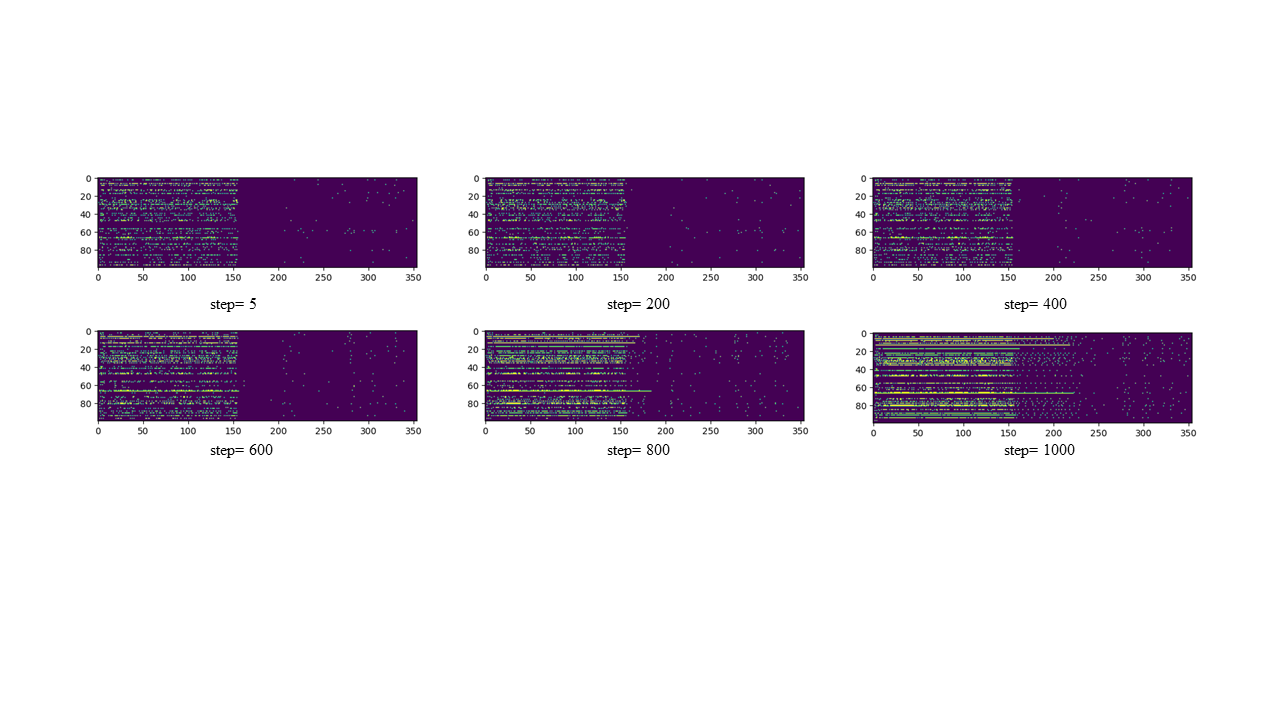


图7已训练5~1000个样本时的各神经元激活情况

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，我们可以发现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。

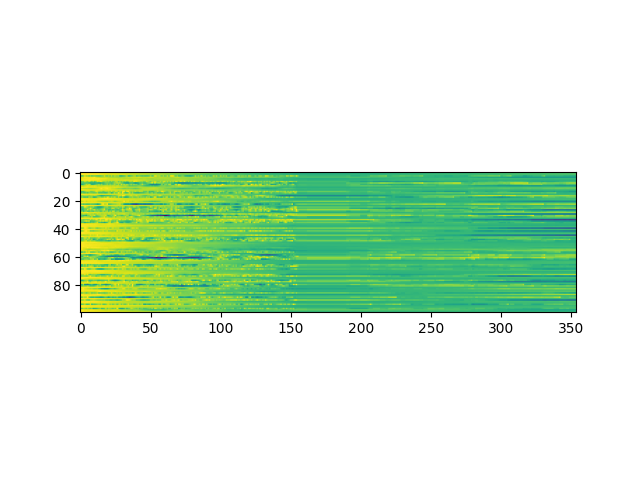


图8(a) 训练开始时各神经元膜电位随时间的变化

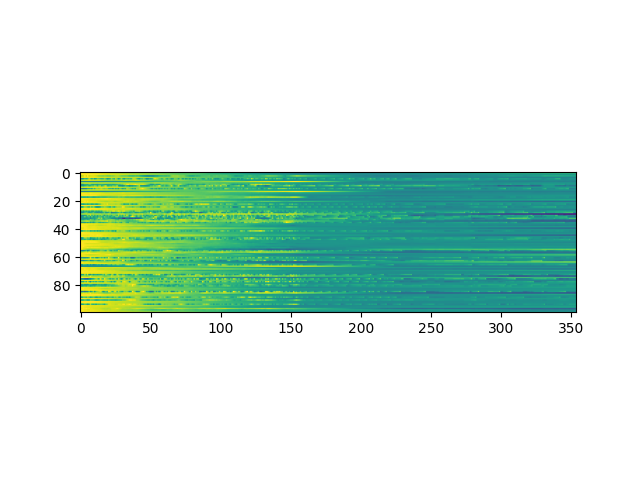


图8(b) 训练结束时各神经元膜电位随时间的变化

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。

通过查阅文献，我们学习了LSNN网络，该网络是在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。

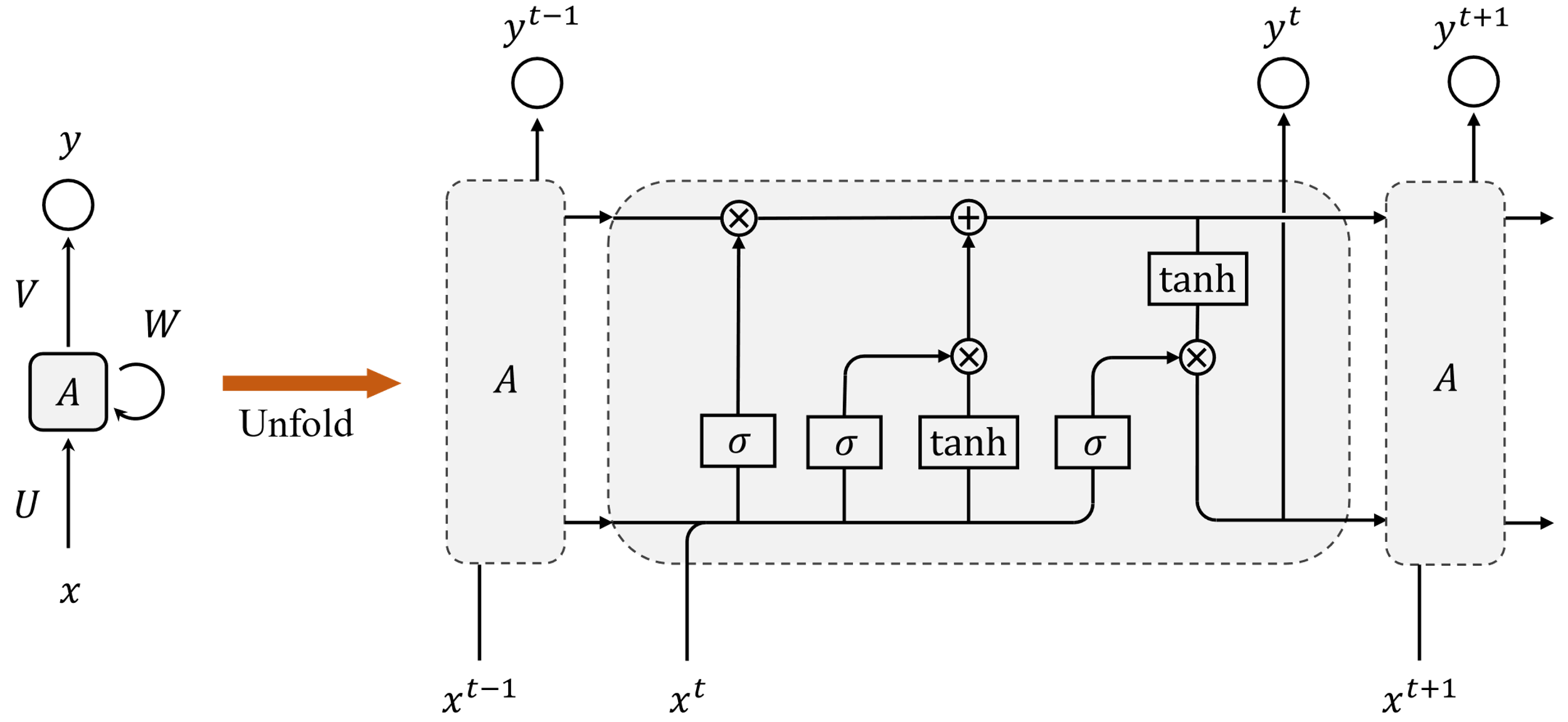


图9

为了缓和RSNN网络的梯度消失或扩散的问题，我们下一步将学习如何结合LSNN网络和e-prop算法来更好地解决小鼠决策问题。

|  |
| --- |
|  |

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.