神经网络中突触可塑性的控制 *

Mohammad Modiri **

摘要

大脑是一个非线性且高度递归的循环神经网络（RNN）。这个循环神经网络具有惊人的可塑性，使我们能够学习并执行复杂任务，展现出非凡的能力。然而，由于大脑具有非线性特征，且决定每个突触对输出误差的贡献机制尚不明朗，学习过程极其复杂。这个问题称为信用分配问题（CAP），是神经科学和人工智能（AI）领域面临的一项根本挑战。尽管如此，按照当前认知神经科学的理解，人们普遍认为，反馈回路系统在突触可塑性中发挥着至关重要的作用。受此启发，我们提出一种结合神经网络（NN）与非线性最优控制理论的计算模型。该框架包含一种基于神经网络的新型演员-评论家方法，用于模拟误差反馈回路系统及其对神经网络突触可塑性的投射，从而确保输出误差最小化。

关键词： 学习规则， 神经网络，非线性控制，储备池计算。

* 本文是一个新学习规则的一部分，该规则已分享出来，以征求研究人员的反馈。**电子邮箱地址：Mohammad.Modiri@Gmail.com

1. 提出的框架

本节提出了一种新颖的脑启发式学习规则。其主要思路是采用基于强化学习的方法来控制突触可塑性。图1展示了这一应用于时间序列分类任务的计算模型背后的主要思想。

在所提出的框架中，突触可塑性被重新表述为一个最优跟踪问题。为此，模仿表示每个时间序列模式真实类别的独热输出向量，使其长度与每个时间序列模式相同，并将其视为参考轨迹，如图1(C)底部子图所示。储备池属于一类特殊的高维非线性动态系统，其状态和时间都是连续的。因此，可应用HJB方程和自适应动态规划的连续时间表述来推导学习规则（最优控制律）。所推导的学习规则促使所提出框架的输出（分类任务中的预测类别）模仿参考轨迹（分类任务中的真实类别），从而隐式地最小化交叉熵。以下各小节将进一步详细阐述网络架构和学习规则。

1.1 网络架构与动力学

所提出的框架由前馈通路和反馈通路组成。前馈模块的主要部分包括：

- 输入数据编码层。
- 循环神经网络（RNN）。
- 线性分类器（输出函数）。

反馈模块由以下部分组成：

- 评论家RC。
- 行动者RC。

图1展示了所提出框架架构的框图。该架构的功能基于以下步骤：

a. 编码层将输入模式 $x(t)$ 投射到循环神经网络中。在脉冲神经网络中，数据向量中的每个实数值都会被转换为离散的脉冲序列，

适合在脉冲网络中处理，因为脉冲神经网络只能处理离散的脉冲序列。b. 在预训练过程中，采用梯度下降、岭回归等监督学习算法，调整与每个训练样本相关联的解码器参数。解码器训练完成后，在储备池训练过程中将其视为常量。c. 在循环神经网络训练过程中，采用第 1.2 小节提出的基于强化学习的算法，通过一种新颖的自适应动态规划方法调整循环神经网络中的参数，使循环神经网络学习输入模式$x(t)$与期望输出 $y(t)$之间的时间关系。在所提出的自适应动态规划机制中，利用两个基于储备池计算的演员和评论家来近似真实的储备池参数。d

. 对新数据的模型回忆 .

前馈通路与传统神经网络的结构相似，但反馈通路的方法，尤其是本研究提出的方法，与经典人工智能方法有显著差异。

注1 为简洁起见，表示动力系统变量时省略时间依赖关系。例如，记号 $x(t)$、$y(t)$、$u(t)$、$v(t)$ 和$e(t)$改写为 x、 y、 u、 v 和 e。

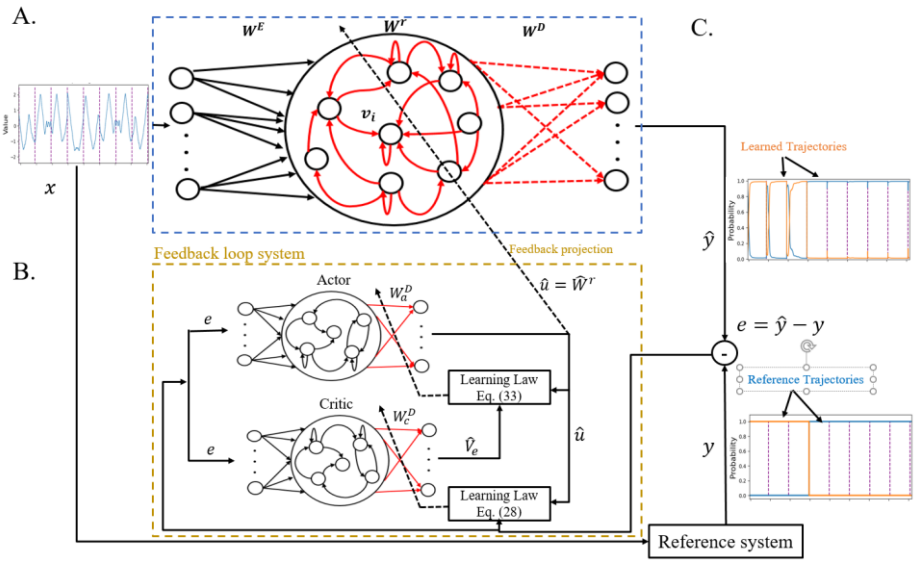

图1. 所提出架构的示意图（红色为可塑性权重）。(A) 前馈通路由一个含固定参数和随机参数的编码层、一个含可训练参数的循环神经网络以及一个解码层组成；这些部分在预训练过程中进行训练，并在循环神经网络训练过程中保持不变。(B) 在反馈通路中，采用非线性最优控制来估计 RNN 参数。该通路由执行器和评论家神经网络组成。(C) 下方子图显示参考轨迹，上方子图显示学习得到的轨迹。

如图1所示，所提出框架的核心是一个循环神经网络（RNN），即一个结构由邻接矩阵 $W^r \in \mathrm{R}^{n^r \times n^r}$ 表示的循环网络。这里，该循环神经网络由 n^r 个神经元组成，其膜电位动力学描述如下：

$$\dot{v}_i = \psi(v_i) + \mathrm{I}_i \tag{1}$$

$$I_i = W_i^E \emptyset(v^E) + W_i^r \emptyset(v^r) \tag{2}$$

，其中 $v = [v_1 \circ v_2 \circ \cdots \circ v_{n^r}] \in \mathrm{R}^{n^r}$ 是循环神经网络神经元的状态或膜电位，$\emptyset(\cdot)$ 是利普希茨非线性树突，例如 $Sigmoid(\cdot)$。$Tanh(\cdot)$，并且，$\psi(v)$: R $^{n^r} \to \mathrm{R}^{n^r}$ 是利普希茨泄漏项，其中我们使用了 $-\alpha_i v$；在非脉冲神经元中，v^E，v^r 直接等于输入 x 和循环神经网络状态 v；在脉冲神经元中，v_i 表示神经元 i 滤波后的脉冲活动。$S_i(t)$ 是神经元 i的脉冲序列，并建模为狄拉克δ函数之和：

$$q_i = \left(S_i * \kappa\right)(t) = \int_{-\infty}^{t} S_i(s)\kappa(t-s)ds \tag{3}$$

$$\kappa(t) = \exp(-t/\tau)/\tau \tag{4}$$

此外，假设线性解码器如下：

$$\hat{y} = h(v) = W^D v \tag{5}$$

1.2 突触可塑性定律

根据第 1.1 小节中提到的方程，循环神经网络的突触可塑性定律可以建模为一个动力系统。因此，可以应用最优控制理论的原理来推导学习规则（最优控制律）。为此，我们将人工智能任务重新表述为一个控制问题。因此，输出误差考虑如下：

$$e = \hat{y} - y \in \mathbb{R}^c \tag{6}$$

，其中 c 表示类别数，$y \in \mathrm{R}^c$ 表示时间 t输入模式的期望输出，$\hat{y} \in \mathrm{R}^c$ 表示相应的预测输出。对式（6）关于 t求导，可将误差动力学描述为：

$$\dot{e} = \dot{\hat{y}} - \dot{y} \in \mathbb{R}^c \tag{7}$$

鉴于假设解码器权重 W^D 在循环神经网络训练过程中保持不变，因此，分类误差动力学仅取决于循环神经网络的神经动力学。将神经空间方程（1）和（2）改写为式（7）的形式，可得：

$$\begin{aligned} \dot{e} = W^D\dot{v} - \dot{y} &= W^D\left(\psi(v) + W^E\emptyset(v^E) + W^r\emptyset(v^r)\right) - \dot{y} \\ &= f(v) + g(v)u^r - f_d(v) - g_d(v)u_d^r \end{aligned} \tag{8}$$

神经空间方程（8）可视为以下仿射状态空间方程：

$$\dot{e} = f(e) + g(e)u \tag{9}$$

其中，控制输入被视为循环神经网络参数的一部分，其数量根据问题和训练数据进行选择：

$$u^r = [u^f; u] = vec(W^r) \in \mathbb{R}^{N^r} \tag{10}$$

，其中 $N^r = n^r \times n^r$ 是 RNN 参数的数量，$vec(\cdot)$ 是矩阵的向量形式，$u^f \in \mathrm{R}^{N^r - N}$ 是固定的 RNN 参数，$u \in \mathrm{R}^N$ 是 RNN 可塑性突触或控制输入，且 $[u^f;\ u]$ 是 u^f与 u的组合。因此，我们得到以下非线性连续时间方程：

$$\begin{aligned} f_1(v) &= W^D\left(W^E\emptyset(v^E) + \psi(v) + W^f\emptyset(v^f)\right) = f^E(v) + f^f(v) \\ f^E(v) &= W^D W^E\emptyset(v^E) \\ f^f(v) &= W^D(\psi(v) + W^f\emptyset(v^f)) \end{aligned} \tag{11}$$

$$g(v) = W^D\emptyset(v) \tag{12}$$

，其中 $f(e)$: $\mathrm{R}^c \to \mathrm{R}^c$ 是系统的漂移动力学，$g(e)$: $\mathrm{R}^c \to \mathrm{R}^{c\times N}$ 是系统的输入动力学。在这方面，一种常见且推荐的做法是向 u 和 $g(e)$ 中添加噪声，以实现持续激励（PE）条件 [1]，从而收集关于未知参数的充分信息，并使 $g(v)$ 满秩。

在人工智能的所有任务中，学习的目标是调整 RNN 参数 W^r，使输入模式在每个时间戳处的误差最小化，并确保 RNN 参数或控制输入保持有界。因此，定义了以下代价函数：

$$\min_{u(\cdot)\in U} \Im\big(u(\cdot).e(\cdot)\big) = \int_t^{t_f} \ell\big(e(\tau).u(\tau)\big)d\tau \tag{13}$$

$$\ell(e.u) = \eta e^T e + u^T R u \tag{14}$$

，其中 l(·.·) 是效用函数，t_f 是输入模式的最后一个元素，R是对称正定矩阵，η 是一个用于

确保代价函数中的误差得到充分有效的处理的常数正超参数。因此，最优价值函数可写为：

$$V(e) = \underset{u(\cdot)\in U}{Inf}\ \Im(u.e) \tag{15}$$

在典型假设下，$f(0) = 0$ 和 $f(v) + g(v)u$ 在包含原点的集合 $\Omega \in \mathrm{R}^n$ 上利普希茨连续。最终，我们希望得到最优输入控制（权重更新律）u^*，使系统公式（11）稳定并使代价函数公式（15）最小化。这种输入控制 u 称为容许控制[2]。

现在，给定公式（11）中的误差动态和公式（15）中的代价函数，学习规则已构建。为求解这一动态优化问题，采用HJB方程，因此定义与控制输入 u 相关的代价函数公式（15）的哈密顿量如下：

$$\mathcal{H}(e.u.V_e) = \ell(e.u) + V_e^T(\dot{e}) \tag{16}$$

，其中 $V_e = \partial V/\partial e$ 是代价函数的偏导数，对于容许控制策略 μ，有：

$$\mathcal{H}(e.\mu.V_e) = 0 \tag{17}$$

本研究假设公式（17）的解是光滑的，从而得到最优代价函数：

$$V^*(e) = \min_u(\int_t^{t_f} \ell\big(u(\tau).e(\tau)\big)d\tau) \tag{18}$$

，它满足HJB方程

$$\min_u \mathcal{H}(e.u.V_e^*) = 0 \tag{19}$$

假设等式（19）左侧的最小值存在，则应用驻值条件 $\partial \mathrm{H}(e.\ u.\ V_e)/\partial u = 0$，可得学习规则（最优控制）：

$$u^*(e) = -\frac{1}{2}R^{-1}g^T(e)\,V_e^* \tag{20}$$

最优价值函数可表示为:

$$V^*(e) = \min_u\left(\int_t^{t_f} \eta e^T e + u^{*T} R u^* d\tau\right) \tag{21}$$

将这一最优学习规则式（20）代入非线性李雅普诺夫方程式（16），可得以 V_e^* 表示的 HJB方程式（19）。

$$0 = \eta e^T e + V_e^{*T}(e) f(v) - \frac{1}{4} V_e^{*T}(e) g(e) R^{-1} g^T(v) V_e^*(e) \tag{22}$$

要为循环神经网络求得学习规则，需要先求解 HJB方程以获得价值函数，再将该解代入以得到所需的学习规则。虽然 HJB方程给出了关于损失函数的学习规则（控制律）达到最优的充要条件，但遗憾的是，由于循环神经网络具有非线性特征，对于实际应用中关注的系统，要以显式形式求解 HJB方程十分困难，甚至无法做到。因此，所提出的框架侧重于采用自适应动态规划方法近似其解。因此，我们将着重采用自适应动态规划方法近似其解。为了解决这一问题，受多巴胺能区域结构的启发，并遵循魏尔斯特拉斯高阶逼近定理 [3], ，我们提出了一种基于 s-RC 的新型演员-评论家方法。简而言之，调整评论家权重的目标是最小化贝尔曼方程误差，而调整演员权重的目标是最小化近似价值。

1.2.1 评价器网络设计

在本小节中，采用 s-RC 作为评论家神经网络，以近似值函数的导数，如下所示:

$$\begin{aligned} &V_e\,(e) = W_c^D\, z_c + \varepsilon_c \in \mathbb{R}^{c\times 1} \\ &z_c \in \mathbb{R}^{n_c\times 1} \\ &W_c^D \in \mathbb{R}^{\,c\times n_c} \end{aligned} \tag{23}$$

其中，z_c 是输入e的新表示，由循环神经网络生成；W_c^D 是解码器权重矩阵。

由于评价器解码器的权重能否为哈密顿函数方程（16）提供最佳近似解尚不确定，因此利用 $\widehat{W}_c^D$ 对其进行估计：

$$\hat{V}_e(e) = \widehat{W}_c^D z_c \tag{24}$$

希望选择评价器网络的权重，使评价器的损失函数最小，其定义如下：

$$E_c\left(\widehat{W}_c^D\right) = \frac{1}{2} e_c^T e_c \tag{25}$$

其中，对于任意给定的容许策略 u，残差误差为：

$$\mathcal{H}(e.u.V_e) = \mathcal{H}\left(e.u.\widehat{W}_c^D\right) = \eta e^T e + u^T R u + (\widehat{W}_c^D z_c)^T(\dot{e}) = e_c \tag{26}$$

因此，$\widehat{W}_c^D \to W_c^D$ 和 $e_c \to \varepsilon_c$。

评论家权重的权重更新律采用梯度下降算法，即：

$$\begin{aligned} \dot{\widehat{W}}_c^D &= -\alpha_c \frac{\partial E_c}{\partial \widehat{W}_c^D} = -\alpha_c (f(e) + g(e)u) {z_c}^T e_c \\ &= -\alpha_c \sigma_c ({\sigma_c}^T \widehat{W}_c^D + \eta e^T e + u^T R u) \end{aligned} \tag{27}$$

其中，$\alpha_c > 0$ 为学习率，$\sigma_c = (f(e) + g(e)u){z_c}^T$。

1.2.2 执行器网络设计

与评论家类似，另一个 s-RC 被用作执行器，以近似突触可塑性规则（反馈控制策略）：

$$u(e) = W_a^D z_a + \varepsilon_a \in \mathbb{R}^{N\times 1} \tag{28}$$

$z_a \in {}^{n_a\times 1}$ R R
$W_a^D \in {}^{N\times n_a}$

令 z_a 为输入 e 的新表示，$\widehat{W}_a^D$ 为基于现有训练数据对未知矩阵 W_a^D 的估计，因此反馈控制策略可表示为：

$$\hat{u}(e) = \widehat{W}_a^D z_a \tag{29}$$

执行器的损失函数定义为：

$$E_a(\widehat{W}_a^D) = \frac{1}{2} e_a^T e_a \tag{30}$$

其中，e_a 定义为式（29）与式（20）之差

$$e_a = \hat{u} - u = \widehat{W}_a^D z_a + \frac{1}{2} R^{-1} g^T(e)(\widehat{W}_c^D z_c) \tag{31}$$

通过应用梯度下降，执行器的权重更新表达式可写为：

$$\dot{\widehat{W}}_a^D = -\alpha_a \frac{\partial E_a}{\partial \widehat{W}_a^D} = -\alpha_a(\widehat{W}_a^D z_a + \frac{1}{2} R^{-1} g^T(e)(\widehat{W}_c^D z_c)) z_a{}^T \tag{32}$$

其中，$\alpha_a > 0$ 为学习率。

最后，RNN 参数更新规则可推导如下：

$$vec(\widehat{W}^r) = \hat{u}(e) \tag{33}$$

至此，RNN 学习规则推导完毕。

在 RNN 学习过程中，执行器-评论家神经网络所构成的反馈通路会促使网络输出（估计类别）跟随参考轨迹（真实类别），这一效应在控制理论中得到了广泛应用。经过足够长时间的学习后，不带反馈的前馈通路能够以可接受的准确率对输入模式进行分类。

1.3 稳定性与收敛性分析

本小节将分析所提出的整个框架的稳定性、收敛性以及分类误差。为此，我们需要证明 $\lim_{t\to\infty} ||\widehat{W}_c^D|| = \lim_{n_e\to\infty} ||\widehat{W}_c^D|| \to 0$、$\lim_{t\to\infty} ||\widehat{W}_a^D|| = \lim_{n_e\to\infty} ||\widehat{W}_a^D|| \to 0$ 和 $\lim_{t\to\infty} ||e|| = \lim ||e|| \to 0$。对于离线问题，$t = n_e n_l(n_s - 1) + i$，其中 n_e 是训练轮次的数量，$n_e\to\infty$是样本数，n_s 是时间序列长度，n_l 是当前时间序列元素的索引。由于在离线问题中，训练数据集 $n_s n_l$是固定的，因此 $t \to \infty$ 随着 $n_e \to \infty$变化。

为了证明所提出的框架渐近收敛，首先需要证明分类误差 e 以及 RNN评价器和执行器解码器的权重估计误差 $\widetilde{W}_c^D$、 $\widetilde{W}_a^D$ 是一致最终有界的（UUB）。因此，我们考虑以下李雅普诺夫函数。为便于进行一致最终有界性分析，作出以下假设；在当前问题设置下，该假设可以合理满足。

假设1 假设满足以下条件：

a. 我们假设 $g(\cdot)$ 有界且为正，即 $g \leq \|g(\chi)\| \leq \overline{g}$ 因此我们有 $\underline{G} \leq G \leq \overline{G}$，其中 $G = gR^{-1}g^T$。 b. 评价器和执行器RNN解码器的理想权重有界，即 $\|W_c^D\| \leq \overline{W_c^D}$ 且 $\|W_a^D\| \leq W_a^D$。 c. 近似误差有界，即 $\|\varepsilon_c\| \leq \overline{\varepsilon_c}$、$\|\varepsilon_a\| \leq \varepsilon_a$。 d. 循环神经网络评价器和执行器的输出有界，即 $z_c \leq \|z_c\| \leq z_c$、 $z_a \leq \|z_a\| \leq \overline{z_a}$。

根据持续激励假设 $\sigma_c = (f + gu)z_c{}^T$ 有界，即 $\sigma_c \leq ||\sigma_c|| \leq \underline{\sigma_c}$。

现在，我们考虑以下李雅普诺夫函数：

$$\begin{aligned} L(t) &= L_{\widetilde{W}_c^D}(t) + L_{\widetilde{W}_a^D}(t) + L_e(t) \\ &= \frac{1}{2\alpha_c} tr\left\{\widetilde{W}_c^{D\,T}\widetilde{W}_c^D\right\} + \frac{\alpha_c}{2\alpha_a} tr\left\{\widetilde{W}_a^{D\,T}\widetilde{W}_a^D\right\} \\ &+ \alpha_c\alpha_a(e^Te + V(e)) \end{aligned} \tag{34}$$

其中 $\alpha > 0$，$\widetilde{W}_c^D$ 是评价器解码器的权重估计误差，其定义如下：

$$\widetilde{W}_c^D = \widehat{W}_c^D - W_c^D \tag{35}$$

此外，$\widetilde{W}_c^D$ 的动力学表示为：

$$\dot{\widetilde{W}}_c^D = -\alpha_c(\sigma_c\sigma_c{}^T\widetilde{W}_c^D + \sigma_c\varepsilon_H) \tag{36}$$

其中

$$\varepsilon_H = -\varepsilon_c(f + gu) \tag{37}$$

此外，执行器权重估计误差定义为：

$$\widetilde{W}_a^D = \widehat{W}_a^D - W_a^D \tag{38}$$

将式（23）和（28）与（32）结合，得到：

$$\begin{aligned}\dot{\widetilde{W}}_a^D &= -\alpha_a\left(\widetilde{W}_a^D z_a + \varepsilon_a + \frac{1}{2}R^{-1}g^T(e)\left(\widetilde{W}_c^D z_c + \varepsilon_c\right)\right){z_a}^T \\ &= -\alpha_a\left(\widetilde{W}_a^D z_a + \frac{1}{2}R^{-1}g^T(v)\left(\widetilde{W}_c^D z_c\right) + {\varepsilon'}_a\right){z_a}^T\end{aligned} \tag{39}$$

其中 ${\varepsilon'}_a$ 变为：

$${\varepsilon'}_a = -(\varepsilon_a + \frac{1}{2}R^{-1}g^T(e)\varepsilon_c) \tag{40}$$

因此，$L(t)$ 的时间导数为：

$$\begin{aligned}\dot{L}(t) &= \dot{L}_{\widetilde{W}_c^D}(t) + \dot{L}_{\widetilde{W}_a^D}(t) + \dot{L}_e(t) \\ &= \frac{1}{\alpha_c}tr\left\{{\widetilde{W}_c^D}^T\dot{\widetilde{W}}_c^D\right\} + \frac{\alpha_c}{\alpha_a}tr\left\{{\widetilde{W}_a^D}^T\dot{\widetilde{W}}_a^D\right\} + \alpha_c\alpha_a(2e^T\dot{e} \\ &\quad + \dot{V}(e))\end{aligned} \tag{41}$$

将式（36）代入式（41）中的 $\dot{\widetilde{W}}_c^D$，并假设 $\|{\sigma_c}^T\widetilde{W}_c\| > \varepsilon_H$，则 $\dot{L}_{\widetilde{W}_c^D}$ 变为：

$$\begin{aligned}\dot{L}_{\widetilde{W}_c^D}(t) &= \frac{1}{\alpha_c}tr\left\{{\widetilde{W}_c^D}^T\dot{\widetilde{W}}_c^D\right\} \\ &= -\frac{1}{\alpha_c}tr\left\{\alpha_c{\widetilde{W}_c^D}^T\sigma_c{\sigma_c}^T\widetilde{W}_c^D\right\} - \frac{1}{\alpha_c}tr\left\{\alpha_c{\widetilde{W}_c^D}^T\sigma_c\varepsilon_H\right\} \\ &\le -(\underline{{\sigma_c}^2})\left\|\widetilde{W}_c^D\right\|^2 + {\varepsilon_H}^2\end{aligned} \tag{42}$$

将式（39）代入$\dot{\widetilde{W}}_a^D$ 式（41）中，第二项变为：

$$\begin{aligned}\dot{L}_{\widetilde{W}_a^D}(t) &= \frac{\alpha_c}{\alpha_a}tr\left\{{\widetilde{W}_a^D}^T\dot{\widetilde{W}}_a^D\right\} \\ &= -\frac{\alpha_c}{\alpha_a}tr\left\{{\widetilde{W}_a^D}^T\left[\alpha_a\left(\widetilde{W}_a^D z_a + \frac{1}{2}R^{-1}g^T(e)\left(\widetilde{W}_c^D z_c\right)\right.\right.\right. \\ &\quad \left.\left.\left. + {\varepsilon'}_a\right){z_a}^T\right]\right\}\end{aligned} \tag{43}$$

根据式（20）、式（29）和$\widetilde{W}_a^D z_a = -\frac{1}{2}R^{-1}g^T$（$e$）（$\widetilde{W}_c^D z_c$），我们有：

$$\dot{L}_{\widetilde{W}_a^D}(t) \leq -(\alpha_c \underline{z_a}^2)\left\|\widetilde{W}_a^D\right\|^2 + \frac{\alpha_c}{4\alpha_a}\|R^{-1}\|^2\overline{g}^2\overline{z_c}^2\left\|\widetilde{W}_c^D\right\|^2 + \frac{\alpha_c}{\alpha_a}{\varepsilon'}_a^2 \tag{44}$$

将公式（29）代入误差动力学公式（11），并根据假设1 a.，公式（41）的第三项变为：

$$\dot{L}_e(t) = \alpha_c\alpha_a(2e^T\dot{e} + \dot{V}(e)) = 2\alpha_c\alpha_a e^T\dot{e} + 2\alpha_c\alpha_a V_e^T\dot{e}$$

将 $\dot{e}$ 代入公式（9），并将 $V_e^T\dot{e}$ 代入 $-(\eta e^T e + u^T R u)$，根据公式（16）和公式（17）：

$$\dot{L}_e(t) = 2\alpha_c\alpha_a e^T(f(e) + g(e)u) - \alpha_c\alpha_a(\eta e^T e + u^T u)$$

将 u 代入公式（20）和公式（24），其结果等于 $u = -\frac{1}{2}R^{-1}g^T(v)\widehat{W}_c^D z_c$：

$$\begin{aligned}\dot{L}_e(t) = {} & 2\alpha_c\alpha_a e^T\left(f(e) - \frac{1}{2}g(e)R^{-1}g^T(e)\,\widehat{W}_c^D z_c\right) \\ & - \alpha_c\alpha_a(\eta e^T e + u^T R u) \\ = {} & 2\alpha_c\alpha_a e^T\left(f(e) - \frac{1}{2}G\widehat{W}_c^D z_c\right) \\ & - \alpha_c\alpha_a(\eta e^T e + u^T R u)\end{aligned} \tag{45}$$

此外，根据杨氏不等式，我们有：

$$e^T f(e) \leq \|e^T f(e)\| \leq \|e^T\|\|f(e)\| \leq \frac{1}{2}\|e\|^2 + \frac{1}{2}\|f(e)\|^2$$

此外，根据关于 $G\widehat{W}_c^D z_c$ 的假设1，我们有：

$$\begin{aligned}\dot{L}_e(t) \le{}& \alpha_c\alpha_a\|e\|^2 + \alpha_c\alpha_a\|f(e)\|^2\|e\| + \alpha_c\alpha_a\overline{G}\overline{z_c}\overline{W_c^D}\,\|e\| \\ &- \alpha_c\alpha_a\eta\|e\|^2 - \alpha_c\alpha_a\lambda_{min}(R)\|u\|^2 \\ &\le -\alpha_c\alpha_a(\eta-1)\|e\|^2 \\ &+ \alpha_c\alpha_a\left(\overline{G}\overline{z_c}\overline{W_c^D} + \|f(e)\|^2\right)\|e\| \\ &- \alpha_c\alpha_a\lambda_{min}(R)\|u\|^2\end{aligned}$$

其中，$\lambda_{min}(R)$ 是 R 的最小特征值。结合式（42）、式（43）和式（45）：

$$\begin{aligned}\dot{L}(t) \le{}& -K_{\widetilde{W}_c^{D^2}}\left\|\widetilde{W}_c^D\right\|^2 - K_{\widetilde{W}_a^{D^2}}\left\|\widetilde{W}_a^D\right\|^2 - K_{u^2}\|u\|^2 - K_{e^2}\|e\|^2 + K_e\|e\| \\ &+ K_1 \\ &= -K_{\widetilde{W}_c^{D^2}}\left\|\widetilde{W}_c^D\right\|^2 - K_{\widetilde{W}_a^{D^2}}\left\|\widetilde{W}_a^D\right\|^2 - K_{u^2}\|u\|^2 \\ &- K_{e^2}\left(\|e\| - \frac{K_e}{2K_{e^2}}\right)^2 + \frac{{K_e}^2}{4K_{e^2}} + K_1 \\ &\le -K_{\widetilde{W}_c^{D^2}}\left\|\widetilde{W}_c^D\right\|^2 - K_{\widetilde{W}_a^{D^2}}\left\|\widetilde{W}_a^D\right\|^2 - K_{u^2}\|u\|^2 \\ &- K_{e^2}\left(\|e\| - \frac{K_e}{2K_{e^2}}\right)^2 + \overline{K}\end{aligned}$$

其中 (46)

$$K_{\widetilde{W}_c^{D^2}} = \underline{\sigma_c}^2 - \frac{\alpha_c}{4\alpha_a}\|R^{-1}\|^2\overline{g}^2\overline{z_c}^2$$

$$K_{\widetilde{W}_a^{D^2}} = \alpha_c\underline{z_a}^2$$

$$K_{e^2} = \alpha_c\alpha_a(\eta-1)$$

$$K_e = \alpha_c\alpha_a\left(\overline{G}\overline{z_c}\overline{W_c^D} + \|f(e)\|^2\right)$$

$$K_{u^2} = \alpha_c\alpha_a\lambda_{min}(R)$$

$$K_1 = {\varepsilon_H}^2 + \frac{{\varepsilon'_a}^2}{2} \le \overline{K_1}$$

$$K = \frac{{K_e}^2}{4K_{e^2}} + K_1 \le \overline{K}$$

因此，$\dot{L}(t) < 0$ 如果所提出的模型超参数包括 $\alpha_{c\circ}$ α_a, η, R, $\psi(v)$ 公式（11）中循环神经网络参数的泄漏项或 $W^f \emptyset(q^f)$ 固定部分被选定，以满足以下三个不等式：

$$0 < \alpha_c < \frac{4\alpha_a \underline{\sigma_c}^2}{\|R^{-1}\|^2 \overline{g}^2 \overline{z_c}^2} \tag{47}$$

And

$$0 < \alpha_a \tag{48}$$

And

$$\eta > 1 \tag{49}$$

由于 K 的所有项均有界，因此根据 [4]，分类误差、评论家误差和执行器误差均为半全局一致最终有界（SGUUB），这对于分类任务而言完全足够。不过，为了达到更稳定的状态，可以满足以下不等式：

$$\|e\| > \frac{K_e}{K_{e^2}} + \sqrt{\frac{K_1}{K_{e^2}}} \tag{50}$$

Or

$$\left\|\widetilde{W}_c^D\right\| > \sqrt{\frac{\overline{K}}{K_{\widetilde{W}_c}^2}} \tag{51}$$

Or

$$\left\|\widetilde{W}_a^D\right\| > \sqrt{\frac{\overline{K}}{K_{\widetilde{W}_a}^2}} \tag{52}$$

Or

$$\|u\| > \sqrt{\frac{\overline{K}}{K_u}} \tag{53}$$

因此，根据李雅普诺夫理论，分类误差 e、RNN评价器和执行器解码器的权重估计误差 $\widetilde{W}_c^D$， $\widetilde{W}_a^D$ 以及 RNN 参数 u 均一致最终有界。

前述基于李雅普诺夫理论的分析表明， $L_e(t)$有界且单调递减。因此，我们有：

$$\begin{aligned} &L_e(t+T) = L_e(t) + \int_t^{t+T} \dot{L}_e(\tau)d\tau \leq L_e(t) - \gamma T \\ &\Longrightarrow L_e(t) \leq L_e(t+T) \end{aligned} \tag{54}$$

其中， $\gamma > 0$。将表示训练数据集长度的 $T = n_l n_s$ 代入，并根据李雅普诺夫函数的性质[5]，对式（54）两边求和，可得：

$$\sum_{t=0}^{n_l n_s} L_e(t) \leq \sum_{t=0}^{n_l n_s} L_e(t+T) \tag{55}$$

因此，式（55）保证了每个训练轮次的分类误差不高于前一训练轮次。由此，随着训练轮次增加，分类训练误差趋于零（即 $\lim_{t\to\infty} e(t) \to 0$）。

同理，我们可以证明，对于 $L_{\widetilde{W}_c^D}(t)$ 和 $L_{\widetilde{W}_a^D}(t)$，结论成立。至此，稳定性与收敛性证明完成。

参考文献

[1] K. Vamvoudakis 和 F. L. Lewis，“求解连续时间无限时域最优控制问题的在线演员–评论家算法”，Automatica，第 46 卷，第 878–888 页，2010年。[2] M. Khalaf 和 F. L. Lewis，“采用神经网络 HJB 方法处理带饱和执行器的非线性系统的近似最优控制律”，Automatica，第 41 卷，第 5 期，第 779–791 页，2005年。[3] B. . A. Finlayson，《加权残值法与变分原理》，纽约：Academic Press，1990年。[4] S. Ge、C. Hang 和 T. Zhang，“通过状态和输出反馈实现非线性系统的自适应神经网络控制”，IEEE《系统、人类与控制论汇刊》，第 29 卷，第 6 期，第 818–828 页，1999年。[5] H. K. Khalil，《非线性控制》，爱丁堡：培生，2014年。[7] H. Ding、G. Trajcevski、P. Scheuermann、X. Wang 和 E. Keogh，“时间序列数据的查询与挖掘：表示与距离度量的实验比较”，《VLDB Endowment 会议录》，第 1 卷，第 2 期，第 1542–1552 页，2008年。[8] P. F. Felzenszwalb 和 R. Zabih，“计算机视觉中的动态规划与图算法”，IEEE《模式分析与机器智能汇刊》，第 33 卷，第 4 期，第 721–740 页，2011年。[9] A. Bagnall、J. Lines、J. Hills 和 A. Bostrom，“采用 COTE 进行时间序列分类：基于变换的集成集合”，IEEE《知识与数据工程汇刊》，第 27 卷，第 9 期，第 2522–2535 页，2015年。[10] J. Grabocka、N. Schilling、M. Wistuba 和 L. Schmidt-Thieme，“时间序列形状片段学习”，载于 ACM SIGKDD 国际知识发现与数据挖掘会议，2014年。[11] N. Srivastava、G. Hinton、A. Krizhevsky、I. Sutskever 和 R. Salakhutdinov，“Dropout：一种防止神经网络过拟合的简单方法”，《机器学习研究杂志》，第 15 卷，第 1 期，第 1929{1958 页，2014年。

[12] Z. Wang、W. Yan 和 T. Oates，“从零开始使用深度神经网络进行时间序列分类：一个强基线”，载于《国际神经网络联合会议》（IJCNN），安克雷奇，2017年。[13] J. Serra、a. Pascual 和 A. Karatzoglou，“迈向通用时间序列神经网络编码器”，《人工智能研究与发展：当前挑战、新趋势与应用》，第308:120页，2018年。[14] I. Sutskever、J. Martens 和 G. Hinton，“使用循环神经网络生成文本”，载于《ICML 11：第28届国际国际会议论文集》，华盛顿，2011年。[15] P. Tanisaro 和 G. Heidemann，“使用时间扭曲不变回声状态网络进行时间序列分类”，载于《IEEE国际机器学习与应用会议》（ICMLA），阿纳海姆，2016年。[16] N. Kasabov，“NeuCube：用于映射、学习和理解时空脑数据的脉冲神经网络架构”，《神经网络》，第52卷，第62-76页，2014年。[17] A. Bagnall、J. Lines、A. Bostrom、J. Large 和 E. Keogh，“时间序列分类大赛：近期算法进展综述与实验评估”，《数据挖掘与知识发现》，第31卷第3期，第606–660页，2016年。[18] H. I. Fawaz、G. Forestier、J. Weber、L. Idoumghar 和 . P.-A. Muller，“时间序列分类的深度学习：综述”，《数据挖掘与知识发现》，第33卷，第917–963页，2019年。[19] J. M. Thérèse，“多巴胺：突触可塑性和记忆机制的潜在基础”，《神经生物学进展》，第69卷，第375–390页，2003年。[20] A. Padmanabhan 和 B. Luna，“发育影像遗传学：将多巴胺功能与青少年行为联系起来”，《脑与认知》，第98卷，第27-38页，2014年。[21] S. . J. Russo 和 E. J. Nestler，“情绪障碍中的大脑奖赏环路”，《自然综述：神经科学》，第14卷，第609–625页，2013年。[22] E. H. Telzer，“多巴胺能奖赏敏感性可促进青少年健康：关于腹侧纹状体激活机制的新视角”，

发展认知神经科学，第17卷，第57–67页，2016年。

[23] N. Kasabov、K. Dhoble、N. Nuntalid 和 G. Indiveri，“用于在线时空与时频模式识别的动态演化脉冲神经网络”，《神经网络》，第41卷，第188-201页，2013年。[24] D. P. Hunt 和 D. Parry，“使用回声状态网络对非预设的真实世界瞬时活动进行分类”，载于《IEEE国际机器学习与应用会议》（ICMLA），阿纳海姆，加利福尼亚州，2016年。[25] H. Jaeger，“分析和训练循环神经网络的回声状态方法”，德国国家信息技术研究中心，波恩，德国，2001年。[26] N. Kasabov、N. M. Scott、E. Tu、S. Marks、N. Sengupta、E. Capecci 和 M. Othman，“基于NeuCube神经形态框架的演化时空数据机器：设计方法及精选应用”，《神经网络》，第78卷，第1-14页，2016年。[27] S. Thorpe 和 G. Jacques，“秩序编码”，《计算神经科学》，第113-118页，1998年。[28] N. Caporale 和 Y. Dan，“脉冲时序依赖可塑性：一种赫布学习规则”，《神经科学年度评论》，第31卷，第25-46页，2008年。[29] J. R. Hollerman 和 W. Schultz，“多巴胺神经元在学习过程中报告奖励时间预测中的误差”，《自然神经科学》，第1卷第4期，第304-309页，1998年。[30] D. Rumelhart、G. E. Hinton 和 R. J. Williams，“通过误差传播学习内部表征”，载于《并行分布式处理：认知微观结构探索》，剑桥，麻省理工学院出版社，1987年。[31] R. J. Williams 和 D. Zipser，“一种用于持续运行的全循环神经网络的学习算法”，《神经计算》，第1卷第2期，第270–280页，1989年。[32] A. Alalshekmubarak 和 L. S. Smith，“一种结合循环神经网络与支持向量机进行时间序列分类的新方法”，载于《信息技术创新》（IIT），2013年。[33] F. Bianchi、S. Scardapane、S. Løkse 和 R. Jenssen，“双向深度读出回声状态网络”，载于《欧洲人工神经研讨会》

网络、计算智能与机器学习，布鲁日，2018年。

[34] M. Lukoševičius，《神经网络：行业诀窍》，载于《回声状态网络应用实用指南》，施普林格，2012年，第659-686页。[35] F. Jiang、H. Berry 和 M. Schoenauer，《回声状态网络的监督学习与进化学习》，载于 PPSN，2008年。[36] A. Rodan 和 P. Tino，《最小复杂度回声状态网络》，IEEE 神经网络汇刊，第200卷，第1期，第131-144页，2010年。[37] F. Lewis 和 V. Syrmos，《最优控制》，纽约：威立，1992年。[38] T. W. Miller、R. S. Sutton 和 P. J. Werbos，《用于控制的神经网络》，剑桥：麻省理工学院出版社，1995年。[39] Y. Chen、E. Keogh、B. Hu、N. Begum、A. Bagnall、A. Mueen 和 G. Batista，《UCR时间序列分类档案》，2015年。[在线]。网址：www.cs.ucr.edu/~eamonn/time_series_data/。[访问日期：5月20日2020]。[40] J. Cristian Borges Gamboa，《时间序列分析的深度学习》，arXiv 1701.01887，2017年。[41] Z. Cui、W. Chen 和 Y. Chen，《用于时间序列分类的多尺度卷积神经网络》，arXiv:1603.06995，2016年。[42] A. L. Guennec、S. Malinowski 和 R. Tavenard，《使用卷积神经网络进行时间序列分类的数据增强》，ECML/PKDD时间数据高级分析与学习研讨会，2016年。[43] V. M. Popov 和 R. Georgescu，《控制系统的超稳定性》，密歇根州：罗马尼亚科学院出版社，1973年。[44] W. Maass、T. Nachtschlaeger 和 H. Markram，《无稳定状态的实时计算：一种基于扰动的神经计算新框架》，《神经计算》，第14卷，第11期，第2531-2560页，2002年。[45] Y. Wang 和 L. Han，《自适应时间序列预测与推荐》，Information Processing & Management，第58卷，第3期，2021年。[46] A. Talaei-Khoei 和 J. M. Wilson，《利用时间序列分析预测具有结构性趋势变化的疾病计数》，Information Processing & Management，第56卷，第3期，第674-686页，2019年。

[47] Y. Zhang、B. J. Jansen 和 A. Spink，《网络搜索引擎交易日志的时间序列分析》，Information Processing & Management，第45卷，第2期，第230-245页，2009年。[48] Z. Wang、W. Yan 和 T. Oates，《从零开始使用深度神经网络进行时间序列分类：一个强有力的基准》，载于国际神经网络联合会议（IJCNN），安克雷奇，2017年。[49] B. Kiumarsi 和 F. L. Lewis，《基于行动者–评论家的部分未知非线性离散时间系统最优跟踪》，IEEE神经网络与学习系统汇刊，第26卷，第1期，第140-151页，2014年。[50] W. Gao 和 Z.-P. Jiang，《线性系统的自适应动态规划与自适应最优输出调节》，IEEE自动控制汇刊，第61卷，第12期，第4164-4169页，2016年。[51] R. Kamalapurkar、H. Dinh、S. Bhasin 和 W. E. Dixon，《连续时间非线性系统的近似最优轨迹跟踪》，Automatica，第15卷，第40-48页，2015年。[52] J. Zhao，《通过强化学习实现多参与者连续时间非线性系统非零和博弈的基于神经网络的最优跟踪控制》，Neurocomputing，第412卷，第167-176页，2020年。[53] Z. Zha、B. Wang 和 X. Tang，《更准确地评估、解释和探索状态：一种面向复杂环境的改进行动者-评论家算法》，Neural Computing for Advanced Applications，2020年。[54] P. Tanisaro 和 G. Heidemann，《使用时间扭曲不变回声状态网络进行时间序列分类》，载于 IEEE国际机器学习与应用会议（ICMLA），阿纳海姆，2016年。[55] E. Trentin、S. Scherer 和 F. Schwenker，《通过概率回声状态网络识别语音信号中的情感》，Pattern Recognition Letters，第66卷，第4-12页，2015年。[56] L. Gao、X. Deng 和 W. Yang，《智慧城市基础设施保护：采用在线储备池计算架构进行实时威胁检测》，Neural Computing and Applications，2021年。[57] D. Bacciu、P. Barsocchi、S. Chessa、C. Gallicchio 和 A. Micheli，《环境辅助生活中储备池计算的实验性表征

应用，Neural Computing and Applications，第 2014 卷，第 24 期，第 1451–1464 页，2013年。[58] H. Jaeger，《循环神经网络训练教程：涵盖 BPPT、实时循环学习、扩展卡尔曼滤波和回声状态网络方法》，德国国家信息研究中心，2002年。[59] H. Zhang、L. Cui、X. Zhang 和 Y. Luo，《基于自适应动态规划方法的未知一般非线性系统数据驱动鲁棒近似最优跟踪控制》，IEEE 神经网络汇刊，第 22 卷，第 12 期，第 2226-2236 页，2011年。[60] M. Modiri、M. M. Homayounpour 和 M. M. Ebadzadeh，《基于自适应动态规划的储备池权重学习及其在时间序列分类中的应用》，神经计算与应用，第 34 卷，第 2 期，第 1-17 页，2022年。