文章编号: 1001-0920(2018)11-1967-08 DOI: 10.13195/j.kzyjc.2017.0791

基于Actor-Critic和神经网络的闭环脑机接口控制器设计

孙京诰[1†], 杨嘉雄[1], 王 硕[1], 薛 瑞[1], 潘红光[2]

(1. 华东理工大学 信息科学与工程学院，上海 200237；2. 西安科技大学 电气与控制工程学院，西安 710054)

摘 要: 在皮层神经元放电活动模型的基础上进行单关节自发运动的研究,从控制理论的角度分析闭环脑机接口的工作原理.使用卡尔曼滤波器和人工神经网络设计系统的解码器替代原系统的脊髓电流,并且比较这两种解码器的优劣.由于在无感知反馈的信号下,解码器的性能下降得比较明显,使用强化学习中Actor-Critic算法结合人工神经网络设计PID控制器,用以产生刺激信号来刺激大脑皮层神经元,使其能够跟踪有感知反馈信号时皮层神经元的放电活动,从而恢复解码器的性能.最后,通过与其他控制算法对比,验证了基于强化学习算法的人工感知反馈信号设计的有效性.

关键词: 大脑皮层放电模型；神经网络；解码器；强化学习；控制器设计

中图分类号: TP273 **文献标志码:** A

Design of closed-loop brain machine interface controller based on Actor-Critic and neural network

SUN Jing-gao[1†], *YANG Jia-xiong*[1], *WANG Shuo* [1], *XUE Rui* [1], *PAN Hong-guang* [2]

(1. College of Information Science and Engineering，East China University of Science and Technology，Shanghai 200237，China；2. College of Electrical and Control Engineering, Xi'an University of Science and Technology，Xi'an 710054，China)

Abstract: In this paper, the spontaneous motion of the single joint is studied on the basis of the cortical neuron firing activity model, and the working principle of the closed-loop brain machine interface is analyzed from the perspective of the control theory. The Kalman filter and artificial neural network are used to design system decoders to replace the original system of spinal cord current, then the advantages and disadvantages of these two decoders are compared. Due to the dramatically decrease of the decoder in the absence of natural proprioception, the reinforcement learning algorithm(Actor-Critic) combined with the artificial neural network is used to design the PID controller, which can generate the stimulus signal to stimulate the neurons of the cerebral cortex, track cortical neuron firing activity with the natural proprioception and restore the performance of the decoder. Finally, the validity of the artificial sensing feedback signal design based on the reinforcement learning algorithm is verified by comparing with other control algorithms.

Keywords: brain cortical neuron firing model；neural network；decoder；reinforcement learning；controller design

0 引 言

脑机接口(BMI)是一种人机结合系统,其在大脑与机器之间提供了用于传递皮层神经元电信号的通道,进而修复一些受损的运动机能,它能够帮助运动障碍患者完成简单的运动任务,从而提高对外交流能力[1].患者的脊髓神经元无法准确控制肌肉运动,同时肌肉不会根据环境产生反馈信号给大脑而导致其本体反馈的缺失,因此将采集到的皮层运动信号用于闭环控制系统研究,不仅能够完善BMI系统的理论基础,推动BMI系统在实际领域中的应用,还拓展了各类控制算法的应用领域,因此具有较高的理论创新价值和实际应用意义.脑机接口主要包括神经元放电活动的测量、运动相关电信号的提取(解码器)和运动相关电信号的反馈(编码器)3部分[2],共同构成闭环控制系统.

近年来,国内外针对大脑皮层运动信息提取的BMI研究已经取得了较大的进展.文献[3]提出了一种基于生理学的数学模型,以表征大脑皮层放电活

收稿日期: 2017-06-20；**修回日期:** 2017-12-06.

基金项目: 国家自然科学基金项目(61603295).

责任编委: 曹进德.

作者简介: 孙京诰(1971−),男,副教授,博士,从事智能优化算法及其应用等研究；杨嘉雄(1993−),男,硕士生,从事闭环脑机接口控制器设计与优化的研究.

†通讯作者. E-mail: sunjinggao@126.com

动与肢体运动的关系. 文献[4]使用最优BMI模型分析皮层神经元放电活动,用于评估单个神经元和皮层区域在产生预期运动时所起的作用. 文献[5]提出了一个生物启发式的皮层网络用于控制机械手的关节移动,效果较好. 文献[6]在开环条件下,使用解码器将EEGs直接转换为控制信号驱动外部设备恢复人体运动,但外部设备运动执行缓慢且不能完全匹配. 文献[7]通过加入基于状态空间模型的控制器将外部设备的位置信息传递给编码器,构成了闭环BMI系统,从而提高了外部设备的执行精度. 文献[8]使用Winner滤波器构造闭环BMI系统,由于输入量与输出量之间的非线性关系限制了解码器模型的应用,导致其控制效果并不理想. 文献[9]在无本体反馈的情况下,引入了尖峰神经元模型和电荷平衡的皮层内微刺激电流(ICMS)模型,用于设计人工感知反馈来刺激大脑皮层,由于使用了PSO参数优化方法使得仿真时间消耗较长,实时性较差.

本文在已有研究的基础上,结合文献[3]提出的皮层生理电路模型,使用人工神经网络设计解码器表征放电信号与肢体运动的非线性关系,同时结合Batch normalization算法[10]加快学习速率并解决模型存在的过拟合问题. 此外,提出一种强化学习的方法[11]用以设计闭环BMI控制器产生刺激电流,以补偿运动障碍患者本体反馈的缺失. 最后,将所提出方法与已有控制算法进行对比,验证了人工感知反馈信号的设计能够恢复系统闭环特性.

1 皮层神经元放电模型

1.1 系统模型描述

Bullock等[3]提出的皮层电路模型已经得到神经生理学实验验证,结构如图1所示.

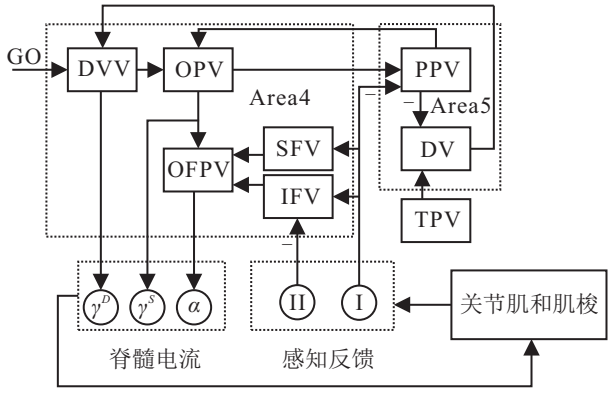

图1 单关节自发运动皮层神经元放电模型

图1中,大脑区域4和区域5主要用于控制肢体关节运动. 差矢量DV用于计算目标向量与肢体感知位置向量之间的差值,该区域的神经元平均放电活动描述为

$$r_i(t)=\max\{T_i-x_i(t)+B^r,0\}. \tag{1}$$

其中:$r_i(t)$满足$0\leqslant r_i(t)\leqslant 1$,下标$i$表示主动肌神经元的平均放电活动,与之对应的下标$j$表示拮抗肌神经元的平均放电活动(主动肌和拮抗肌共同完成单关节的自发运动);B^r为DV神经元的基础放电活动;T_i为目标位置矢量(TPV)中主动肌的目标位置;$x_i(t)$为感知位置矢量(PPV)神经元的平均放电活动,可以用于计算主动肌在运动过程中的实时位置.

DV神经元信号随后传递给区域4中期望速度矢量(DVV),该区域的神经元平均放电活动描述为

$$u_i(t)=\max\{g(t)(r_i(t)-r_j(t))+B^u,0\}. \tag{2}$$

其中:B^u为DVV神经元的基础放电活动;$g(t)$为内部"GO"信号,该运动指令信号由大脑发出,且只与单关节运动速度有关,放电信号描述为

$$\frac{\mathrm{d}g^{(l)}(t)}{\mathrm{d}t}=\varepsilon(-g^{(l)}(t)+(C-g^{(l)}(t))g^{(l-1)}(t)), \tag{3}$$

$$g(t)=g^{(0)}(t)\frac{g^{(2)}(t)}{C}, \tag{4}$$

$l=\{1,2\}, g^{(0)}(t)=g^0, \varepsilon$为低速整合率,$C$为"GO"神经元饱和值. 输出位置矢量(OPV)描述为

$$\begin{aligned}&\frac{\mathrm{d}y_i(t)}{\mathrm{d}t}=\\&(1-y_i(t))(\eta x_i(t)+\max\{u_i(t)-u_j(t),0\})-\\&y_i(t)(\eta x_j(t)+\max\{u_j(t)-u_i(t),0\}).\end{aligned} \tag{5}$$

其中:η为缩放因子,主动肌和拮抗肌的输出位置矢量满足$y_i(t)+y_j(t)=1$. 静态和动态γ运动神经元平均放电活动表示为

$$\gamma_i^S(t)=y_i(t),\ \gamma_i^D=\rho\max\{u_i(t)-u_j(t),0\}, \tag{6}$$

其中ρ为缩放因子. 原发性和继发性肌梭传入描述为

$$\begin{aligned}s_i^1(t)&=S\Big(\theta\max\{\gamma_i^S(t)-p_i(t),0\}+\\&\qquad\phi\max\Big\{\gamma_i^D(t)-\frac{\mathrm{d}p_i(t)}{\mathrm{d}t},0\Big\}\Big),\\s_i^2(t)&=S(\theta\max\{\gamma_i^S(t)-p_i(t),0\}).\end{aligned} \tag{7}$$

其中:$p_i(t)$为主动肌的位置,θ为静态核袋和链纤维的灵敏度,ϕ为动态核袋纤维,肌梭传入饱和函数由$S(\omega)=\omega/(1+100\omega^2)$给出. 感知位置矢量(PPV)的平均放电活动描述为

$$\begin{aligned}&\frac{\mathrm{d}x_i(t)}{\mathrm{d}t}=\\&(1-x_i(t))\max\{\Theta y_i(t)+s_j^1(t-\tau)-s_i^1(t-\tau),0\}-\\&x_i(t)\max\{\Theta y_j(t)+s_i^1(t-\tau)-s_j^1(t-\tau),0\}.\end{aligned} \tag{8}$$

其中:τ为纺锤反馈的延迟时间,Θ为恒定增益,主动

肌和拮抗肌的感知位置矢量满足$x_i(t)+x_j(t)=1$. 区域4中的惯性力矢量(IFV)和静态力矢量(SFV)可描述为

$$q_i(t)=\lambda_i\max\{s_i^1(t-\tau)-s_i^2(t-\tau)-\Lambda,0\},\quad(9)$$

$$\frac{\mathrm{d}f_i(t)}{\mathrm{d}t}=$$
$$(1-f_i(t))hs_i^1(t-\tau)-\psi f_i(t)(f_j(t)+s_j^1(t-\tau)).\quad(10)$$

其中: Λ为恒定阈值, h为一个控制外部负载补偿力度和速度的恒定增益, ψ为抑制缩放的参数. 输出力和位置矢量(OFPV)描述为

$$a_i(t)=y_i(t)+q_i(t)+f_i(t).\quad(11)$$

α运动神经元可以描述为

$$\alpha_i(t)=a_i(t)+\delta s_i^1(t),\quad(12)$$

其中δ表示牵张反射增益.

1.2　肢体运动描述

肢体的运动主要由主动肌和拮抗肌两部分肌群协同完成, 例如肢体完成屈肘动作时, 当主动肌适度拉伸后, 位于它们相反一侧的拮抗肌同时松弛和伸长, 因此肢体运动可以由这两部分肌肉的合力进行驱动. Bullock等[3]将合力近似描述为

$$\Delta M(t)=M_i(c_i(t)-p_i(t))-M_j(c_j(t)-p_j(t)),\quad(13)$$

$$M_i(c_i(t)-p_i(t))=\max\{c_i(t)-p_i(t),0\}.\quad(14)$$

其中: $M_i(c_i(t)-p_i(t))$为主动肌产生的作用力; $c_i(t)$为主动肌收缩活动的力度, 有

$$\frac{\mathrm{d}c_i(t)}{\mathrm{d}t}=v(\alpha_i(t)-c_i(t)),\quad(15)$$

肢体运动状态描述为

$$\frac{d^2p_i(t)}{\mathrm{d}t^2}=\frac{1}{I}\Big(M_i(c_i(t)-p_i(t))-M_j(c_j(t)-p_j(t))+$$
$$E_i-V\frac{\mathrm{d}p_i(t)}{\mathrm{d}t}\Big).\quad(16)$$

其中: I为肢体的惯性力矩; V为关节粘度; E_i为作用到关节的外部力; 主动肌和拮抗肌的位置满足$p_i(t)+p_j(t)=1$, 通过肌肉模型驱动肢体运动进而完成肢体的自发运动.

2　数据集获取以及解码器设计

2.1　数据集获取

在实际的BMI实验中, 受试者能在规定时间内完成指定的运动任务, 那么这次实验可以被认为是成功的. 根据第1节描述的单关节自发运动模型, 由于“GO”信号与运动速度有关, 为了获得充足的数据集, 令g^0满足期望为0.75、方差为0.002 5的高斯随机分布. 这里g^0表示一个常数, 每个g^0对应一次独立的任务.

实验前50 ms内, 大脑中的“GO”信号处于初始启动状态, 在该状态下, 系统部分参数设置为

$$y_i(0)=y_j(0)=0.5,\ x_i(0)=x_j(0)=0.5,$$
$$p_i(0)=p_j(0)=0.5,\ u_i(0)=u_j(0)=B^u,$$
$$r_i(0)=r_j(0)=B^r,$$

其余矢量的初始状态设置为0. 系统中其余参数根据文献[3]分别设为

$$I=200,\ V=10,\ v=0.15,\ B^r=0.1,$$
$$B^u=0.01,\ \Theta=0,\ \theta=0.5,\ \phi=1,$$
$$\eta=0.7,\ \rho=0.04,\ \lambda_i=150,\ \Lambda=0.001,$$
$$\delta=0.1,\ C=25,\ \varepsilon=0.05,\ \psi=4,$$
$$h=0.01,\ T_i=0.7,\ \tau=0.$$

在Matlab仿真软件中共进行1 600次单关节重复伸展任务, 每次运动任务的时间设定为1.46 s, 采样周期为10 ms, 因此共获得233 600组综合数据集用于下面的解码器设计.

2.2　基于卡尔曼滤波的解码器设计

对于一个给定的运动任务, 需要设计一个可以描述该电信号的解码器, 即一个准确的数学模型, 用来替代原模型的脊髓电流. 在过去的研究中, 已有的解码器研究方式主要包括维纳滤波器(VF)[8]、卡尔曼滤波器(KF)[12]和循环神经网络(RNN)[13]等方法. 本节主要使用线性和非线性两种解码器设计方法, 从连续的放电活动中提取主动肌与对应的拮抗肌之间的合力$\Delta M(k)$, 同时比较其优劣性.

卡尔曼滤波器是一种用于时变线性系统的递归滤波器, 该滤波器将过去的测量误差结合新的测量误差用于估计将来误差. 假设系统的状态方程如下:

$$x(k)=Ax(k-1)+w(k),\quad(17)$$

$$z(k)=Hx(k)+v(k).\quad(18)$$

卡尔曼滤波方程为

$$\hat{x}(k|k-1)=A\hat{x}(k-1|k-1),\quad(19)$$

$$\hat{x}(k|k)=\hat{x}(k|k-1)+K_g(z(k)-H\hat{x}(k|k-1)).\quad(20)$$

增益的递推矩阵为

$$K_g=\hat{P}(k|k-1)H^{\mathrm{T}}(H\hat{P}(k|k-1)H^{\mathrm{T}}+V)^{-1},\quad(21)$$

$$\hat{P}(k|k-1)=A\hat{P}(k-1|k-1)A^{\mathrm{T}}+W,\quad(22)$$

$$\hat{P}(k|k)=(I-K_gH)\hat{P}(k|k-1). \tag{23}$$

其中:$\hat{x}(k|k-1)$和$\hat{x}(k|k)$分别为状态向量$x(k)$的先验估计和后验估计,$\hat{P}(k|k-1)$和$\hat{P}(k|k)$为$\hat{x}(k|k-1)$和$\hat{x}(k|k)$对应的协方差矩阵,W和V分别为系统噪声和测量噪声的协方差矩阵,K_g为卡尔曼增益,I为单位矩阵.令$x(k)$=$\Delta M(k)$表示主动肌与拮抗肌之间的合力,提取的放电活动主要来自区域4中“DVV”、“OPV”和“OFPV”神经元,因此观测向量$z(k)=[y_i,y_j,u_i,u_j,a_i,a_j]$是一个$6\times1$维向量.

训练卡尔曼滤波器,需要估计模型的参数矩阵A、H、W和V,对于观测矩阵H,在没有确定模型的情况下能够获得系统的状态和观测数据,可以使用简单的线性回归进行估计,表示如下:

$$H=ZX^{\mathrm{T}}(XX^{\mathrm{T}})^{-1}. \tag{24}$$

其中:Z和X为表示连续放电活动和与之对应的肌肉合力.由于本次卡尔曼滤波器训练使用220 000组数据,$D=220\,000$,矩阵Z和X分别为

$$Z=\begin{bmatrix} z_{1,1} & \cdots & z_{1,D} \\ \vdots & \ddots & \vdots \\ z_{6,1} & \cdots & z_{6,D} \end{bmatrix}, \tag{25}$$

$$X=[x_{1,1},x_{1,2},\cdots,x_{1,D}]. \tag{26}$$

矩阵A、W和V可以通过式(24)~(26)从训练集上进行简单地估计,表示为

$$A=X_2X_1^{\mathrm{T}}(X_1X_1^{\mathrm{T}})^{-1}, \tag{27}$$

$$W=\frac{1}{D-1}(X_2-AX_1)(X_2-AX_1)^{\mathrm{T}}, \tag{28}$$

$$V=\frac{1}{D}(Z-HX)(Z-HX)^{\mathrm{T}}. \tag{29}$$

通过引入一个单位的偏移步长表示矩阵X_1和X_2,有

$$X_1=[x_{1,1},x_{1,2},\cdots,x_{1,D-1}], \tag{30}$$

$$X_2=[x_{1,2},x_{1,3},\cdots,x_{1,D}]. \tag{31}$$

计算解码器参数A,W,H,V,使用剩余的500组数据验证解码器的性能.经过仿真,基于卡尔曼滤波器的解码器性能测试如图2所示.

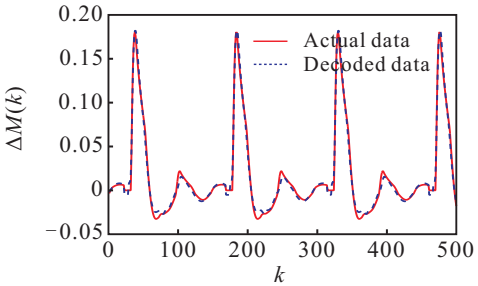

图2 基于卡尔曼滤波器的解码器测试

2.3 基于BN-BP的解码器设计

人工神经网络(ANN)是一种应用类似于大脑神经突触连接结构进行信息处理的数学模型,具有联想记忆功能和非线性映射等特点.但是由于其强大的非线性拟合能力,使用神经网络会带来学习时间过长和模型过拟合等问题.Ioffe等[10]提出了一种通过Batch normalization(BN)加速网络训练的方法,将该方法用于人工神经网络,不仅能有效地解决网络存在的过拟合问题,而且能够使用较高的学习速率,进而加速网络训练,因此本节结合BN方法使用BP神经网络设计解码器.

与网络的隐藏层、函数激活层一样,BN也相当于网络中的一层,根据实验发现,将BN层置于函数激活层前可以使网络具有更好的效果.BN的主要思想是数据在进入网络中间层时,对数据先进行归一化处理,使得每一维的均值为0,方差为1.为了不破坏网络所学习到的特征分布,对归一化后的数据进行变换重构,引入γ和β两个学习参数,这样即可恢复原始层学习到的特征分布.BN网络层的前向传递过程如下:

Step 1: 选取合适的mini batch数量m,x_i表示该数据集中的d维向量.

Step 2: 求输入向量均值$\mu_B=\dfrac{1}{m}\sum\limits_{i=1}^{m}x_i$和方差$\sigma_B^2\leftarrow\dfrac{1}{m}\sum\limits_{i=1}^{m}(x_i-\mu_B)^2$.

Step 3: 对输入向量进行归一化处理后得到$\hat{x}_i=\dfrac{x_i-\mu_B}{\sqrt{\sigma_B^2+\varepsilon}}$,对归一化结果进行变换重构得到BN的结果$y_i\leftarrow\gamma\hat{x}_i+\beta\equiv\mathrm{BN}_{\gamma,\beta}(x_i)$.

在训练过程中需要反向传递损失函数的梯度,同时需要使用链式法则更新网络的参数,参数更新如下:

$$\frac{\partial\ell}{\partial x_i}=\frac{\partial\ell}{\partial\hat{x}_i}\cdot\frac{1}{\sqrt{\sigma_B^2+\varepsilon}}+\frac{\partial\ell}{\partial\sigma_B^2}\cdot\frac{2(x_i-\mu_B)}{m}+\frac{\partial\ell}{\partial\mu_B}\cdot\frac{1}{m}, \tag{32}$$

$$\frac{\partial\ell}{\partial\hat{x}_i}=\frac{\partial\ell}{\partial y_i}\cdot\gamma, \tag{33}$$

$$\frac{\partial\ell}{\partial\sigma_B^2}=-\frac{1}{2}\sum_{i=1}^{m}\frac{\partial\ell}{\partial\hat{x}_i}\cdot(x_i-\mu_B)\cdot(\sigma_B^2+\varepsilon)^{-3/2}, \tag{34}$$

$$\frac{\partial\ell}{\partial\mu_B}=\sum_{i=1}^{m}\frac{\partial\ell}{\partial\hat{x}_i}\cdot\frac{1}{\sqrt{\sigma_B^2+\varepsilon}}. \tag{35}$$

由于γ和β也是网络需要的学习参数,应对其进行更新,偏导为

$$\frac{\partial\ell}{\partial\gamma}=\sum_{i=1}^{m}\frac{\partial\ell}{\partial y_i}\cdot\hat{x}_i,\quad\frac{\partial\ell}{\partial\beta}=\sum_{i=1}^{m}\frac{\partial\ell}{\partial y_i}. \tag{36}$$

结合 BP 神经网络原理, 基于 BN-BP 神经网络的解码器设计如下.

Step 1: 确定网络输入向量. 令 $z_1 = y_i, z_2 = y_j, z_3 = u_i, z_4 = u_j, z_5 = a_i, z_6 = a_j, L = 10$ 为延迟单元, 将 $z(60 \times 1)$ 作为网络输入, 选取数据集前 220 000 组数据 (取 k 为 1 ~ 220 000) 后打乱, mini batch 的数量设为 128.

Step 2: 确定网络的初始参数. 输入层有 60 个神经元, 隐层有 2 个, 每一层均有 512 个神经元, 输出层有 1 个神经元, 网络的权值和阈值初值随机给出, 学习速率设为 0.02, BN 层的参数 $\varepsilon = 0.001$, α 和 β 随机给出, 损失函数定义为 $\ell = \dfrac{1}{2}\sum_{P=1}^{128}(T^P - O^P)^2$.

Step 3: 网络训练. 每经过网络的前向和反向传递后, 网络的参数更新, 数据集训练 50 次后达到规定误差要求 (10^{-3}), 退出网络训练, 将剩余数据集的 500 组数据代入网络以测试解码器的性能, 解码器的性能结果如图 3 所示.

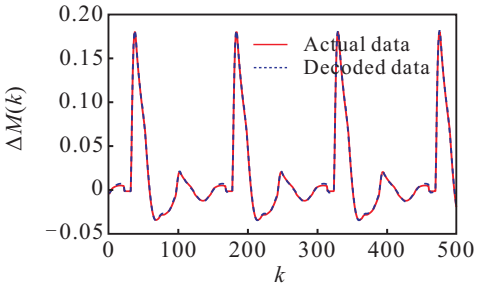

图 3　基于 BN-BP 的解码器测试

均方根误差评价指标 (RMSE), 即观测值与真值的残差均方根, 常用于度量模型或估计器预测值与实际观测值之间的差异. 文献 [14] 利用均方根误差估计了无线传感器分布式卡尔曼一致性滤波精度, 以表明设计算法的估计精度有显著提升. 文献 [15] 分析表征了多输入多输出正交频分复用系统可分辨路径的仰角和方位角的均方根误差, 该误差指标对一组测量值中的特大或特小误差反应非常敏感, 能够很好地反映出路径上仰角和方位角的准确度. 因此, 引入该误差评价指标比较两种解码器的性能, 有

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{k=1}^{500}\left(\Delta\overline{M}(k) - \Delta M(k)\right)^2}{500}}, \tag{37}$$

其中 $\Delta\overline{M}(k)$ 为原系统主动肌和拮抗肌产生的合力. 通过计算发现, 卡尔曼滤波设计的解码器均方根误差值为 4.7×10^{-3}, 使用基于 BN-BP 神经网络的均方根误差值为 1.1×10^{-3}, 性能较卡尔曼滤波器提升显著. 可见, 使用线性卡尔曼滤波器设计解码器替代脊髓电流虽然在计算的便捷性上具有较大的优势, 但是从实际效果看, 测试数据的误差较大, 不能很好地表征皮层神经元放电活动与由主动肌和拮抗肌产生的合力之间的非线性关系, 因此考虑使用 BN-BP 神经网络来设计解码器.

2.4　开环解码器性能

在闭环情况下, 上述设计的解码器已经能很好地替代脊髓电流在原系统中的作用, 下面需要研究的是假如一个受试者没有感知反馈, 即缺少原发性和继发性肌梭传入电信号给 PPV 神经元, 所设计的解码器是否还能起作用, 此时系统处于开环状态. 本次实验中, 只进行一次单关节自发运动, 令 $g^0 = 0.75$, 当不存在感知反馈时, 令式 (6) 和 (7) 中的参数 θ 和 ρ 为 0.

在原系统中, 当系统不存在感知反馈时, 由图 4 可知, 虽然主动肌运动的上升时间变长, 超调增大, 但当实验时间足够长时, 主动肌可以到达期望位置; 当系统用解码器替代后, 如果系统不存在感知反馈, 即当系统处于开环状态时, 则由图 5 可知, 主动肌已经完全不能到达指定运动位置, 肢体运动几乎只能停留在初始状态. 同时, 仿真后发现其他神经元的 (如 DVV、OFPV 和 IFV 等) 放电活动也出现了显著的变化, 即大脑皮层的生理结构发生了改变.

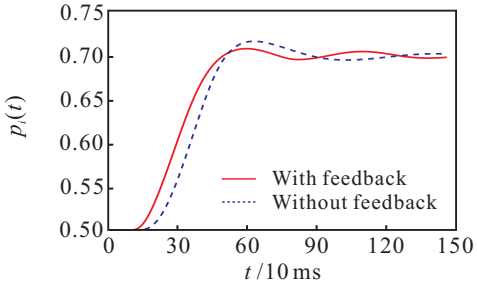

图 4　在有无感知反馈时, 主动肌位置对比 (原系统)

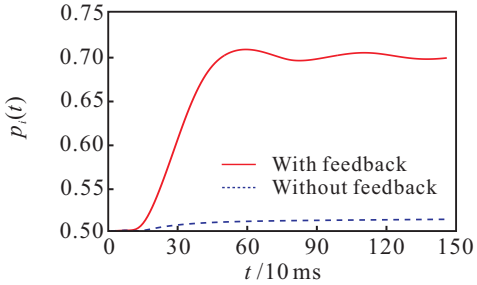

图 5　在有无感知反馈时, 主动肌位置对比 (解码器)

此次设计的解码器权重由于不能适应在缺失感知反馈时所带来的较大偏差, 使得解码器性能大幅度下降, 导致运动任务失败. 为了解决在缺失自然感知反馈后带来的解码器性能的下降问题, 需要设计 PPV 神经元的人工皮层刺激信号作为人工感知反馈信号, 以恢复系统的闭环性能.

3 皮层刺激信号设计

3.1 强化学习与Actor-Critic网络

强化学习是机器学习的一个分支,其主要思想[11]是一个学习系统(Agent)在与外部环境不断试错的过程中,通过优化自己的动作行为(Action)找到合适的策略(Policy)来获得最大的奖励函数回报值(强化信号),进而更新评估函数,在满足整个学习条件后退出.强化学习已经在机器人路径规划[16]、交通信号灯控制[17]、自适应控制[18]和倒立摆平衡控制[19]等非线性系统控制领域取得重要成果.

Actor-Critic(AC)算法是强化学习的一种重要算法,主要由动作网络(Actor)和动作评价网络(Critic)组成.动作网络根据系统状态生成最优策略向环境输出动作,动作评价网络根据环境给出即时回报和系统状态来输出值函数,用于评价动作的好坏,由Critic网络产生的TD误差驱动整个网络的运行.

3.2 皮层刺激信号设计

如果系统不存在本体反馈,则式(8)可以转换为(38),因此目标是设计最优控制律$u(t)$,使得PPV神经元能够恢复到有本体感知反馈的状态,从而恢复主动肌的运动性能,有

$$\frac{\mathrm{d}x_i(t)}{\mathrm{d}t}=(1-x_i(t))\max\{\Theta y_i(t)-u(t),0\}-x_i(t)\max\{\Theta y_i(t)+u(t),0\}. \tag{38}$$

Actor-Critic网络中的动作函数和评价函数由神经网络进行逼近[20],因此本节将BP神经网络与Actor-Critic网络相结合用于设计自适应PID控制器.控制器结构如图6所示.

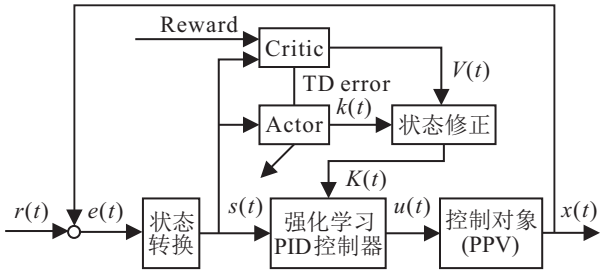

图6 基于Actor-Critic网络的自适应PID控制结构

1) 系统状态确定和回报函数设计.

本次控制器设计使用增量式PID控制器,控制器方程如下:

$$u(t)=u(t-1)+\Delta u(t)=u(t-1)+k_L e(t)+k_P\Delta e(t)+k_D\Delta^2 e(t). \tag{39}$$

其中:$e(t)=x_r^{\mathrm{ref}}(t)-x_i(t)$,$\Delta e(t)=e(t)-e(t-1)$,$\Delta^2 e(t)=e(t)-2e(t-1)+e(t-2)$分别为系统的误差、一次误差和二次误差,可得网络所需要学习的状态向量$s(t)=[e(t),\Delta e(t),\Delta^2 e(t)]$.

控制器从环境中得到的唯一反馈是状态转移时从环境中获得的回报函数,描述为

$$r(t)=\begin{cases}0,\ |x_i^{\mathrm{ref}}(t)-x_i(t)|\leqslant\varepsilon;\\ -1,\ \text{Otherwise.}\end{cases}. \tag{40}$$

2) 动作修正.

为了解决强化学习中探索与利用之间的平衡问题,需要对神经网络的输出k_P、k_I和k_D加上一个高斯随机噪声以强化对动作空间的探索,当值函数较小时,噪声的幅值相应增加,可以对更大的动作空间进行探索,描述公式如下:

$$K(t)=k(t)+\eta_K(0,\sigma_V(t)), \tag{41}$$

$$\sigma_V(t)=\frac{1}{1+2e^{V(t)}}. \tag{42}$$

3) TD误差.

基于AC的BP神经网络权值更新常采用TD误差算法,描述为

$$\delta_{\mathrm{TD}}(t)=r(t)+\gamma V(t+1)-V(t). \tag{43}$$

其中:$r(t)$为控制器从环境中得到的立即回报,$V(t+1)$为下一步的状态值函数,γ为折扣因子. 因此误差函数为$E(t)=\frac{1}{2}\delta_{\mathrm{TD}}^2(t)$.

4) 神经网络设计.

AC强化学习网络使用神经网络来学习PID控制器的3个参数和值函数. 网络参数对控制性能有着直接的影响,例如学习率过大,可能导致算法不稳定,控制曲线出现振荡导致性能下降;学习率过小则导致收敛速度慢,控制器性能变差. 此外,如果隐层神经元数目过大会导致网络计算量增大,因此需要选择合适的网络参数. 网络输入为系统状态向量$s(t)$,输出为$k(t)$和$V(t)$,m为隐藏神经元数目.隐层输入输出分别为

$$\mathrm{net}_i^{(1)}(t)=\sum_{j=1}^{3}w_{ij}(t)s_j(t), \tag{44}$$

$$O_i^{(1)}(t)=f(\mathrm{net}_i^{(1)}(t)),\ i=1,2,\cdots,m. \tag{45}$$

输出层输入输出分别为

$$\mathrm{net}_k^{(2)}(t)=\sum_{i=1}^{m}w_{ki}(t)O_i^{(1)}(t), \tag{46}$$

$$O_k^{(2)}(t)=g(\mathrm{net}_k^{(2)}(t)),\ k=1,2,3,4. \tag{47}$$

按照梯度下降法和链式法则对网络的输出层权值进行更新,并附加一个惯性项,使得整个网络能够快速收敛至全局最小,描述为

$$\Delta w_{ki}(t)=-\alpha\tau_k+\beta\Delta w_{ki}(t-1), \tag{48}$$

$$\tau_k = \delta_{\mathrm{TD}}(t)\cdot c_k \cdot g'(\mathrm{net}_k^{(2)}(t))O_i^{(1)}(t), \tag{49}$$

其中参数c_k表示为

$$c_1 = \frac{\partial \delta_{\mathrm{TD}}(t)}{\partial u(t)}\cdot \Delta e(t)\cdot \frac{K_P(t)-k_P(t)}{\sigma_V(t)},$$

$$c_2 = \frac{\partial \delta_{\mathrm{TD}}(t)}{\partial u(t)}\cdot e(t)\cdot \frac{K_I(t)-k_I(t)}{\sigma_V(t)},$$

$$c_3 = \frac{\partial \delta_{\mathrm{TD}}(t)}{\partial u(t)}\cdot \Delta^2 e(t)\cdot \frac{K_D(t)-k_D(t)}{\sigma_V(t)},$$

$$c_4 = \gamma. \tag{50}$$

同理,隐藏层的权值更新如下:

$$\Delta w_{ij}(t) = -\alpha v_i s_j(t) + \beta \Delta w_{ij}(t-1), \tag{51}$$

$$v_i = f'(\mathrm{net}_i^{(1)}(t))\sum_{k=1}^{4}(\tau_k w_{ki}(t)),$$

$$i = 1, 2, \cdots, m. \tag{52}$$

综上所述,控制器设计过程如下.

Step 1: 初始化网络参数,容许偏差$\varepsilon = 0.01$,折扣因子$\gamma = 0.9$,学习速率$\alpha = 0.1, \beta = 0.1$,神经网络隐层数目$m = 10$.

Step 2: 根据误差$e(t)$计算神经网络需要的状态变量$s(t)$,由式(40),(44)~(47)计算立即回报值$r(t)$、神经网络输出$k(t)$和$V(t)$,经过动作修正后得到$K(t)$.

Step 3: 由式(39)计算下一时刻控制作用$u(t+1)$,将$u(t+1)$作用于被控对象PPV神经元,得到放电率$x_i(t+1)$,并计算下一时刻的状态$s(t+1)$和立即回报$r(t+1)$,将下一时刻状态代入网络后计算得到$k(t+1)$和$V(t+1)$.

Step 4: 由式(43)计算TD误差,通过式(48)~(52)更新神经网络参数,判断是否满足控制条件,若满足条件则退出,否则转Step 2.

3.3　仿真实验

使用上述算法设计PID控制器,使得PPV神经元的放电活动能够恢复到有本体反馈时的状态,并对比实验使用基于BP神经网络的PID算法. PPV神经元放电等实验对比结果如图7~图9所示.

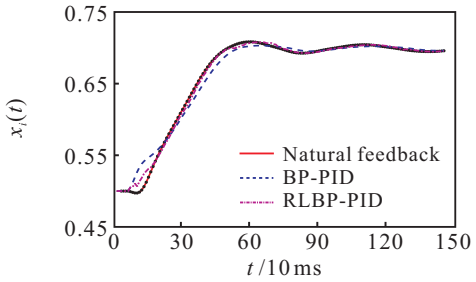

图 7　PPV 神经元放电结果对比

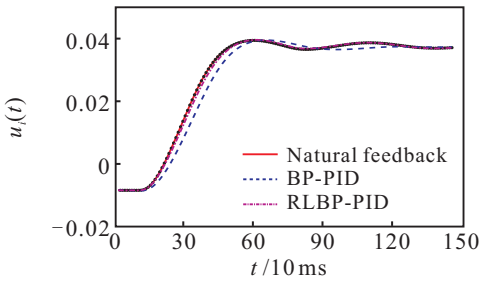

图 8　神经元放电恢复人工反馈与本体反馈对比

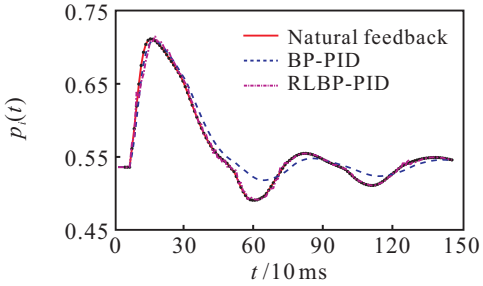

图 9　主动肌位置轨迹跟踪结果对比

对于恢复神经元放电活动任务,如果在给定的1.46 s时间内,跟踪曲线接近实际放电活动,则此次皮层刺激信号达到了设计要求,这里仍使用均方根误差作为评价指标. 由图8可见,使用RLBP-PID设计的控制器由于添加了高斯随机噪声,增强了对动作空间的探索,产生的皮层刺激信号u_i的均方根误差1.3×10^{-3}明显小于使用BP-PID设计控制器产生的误差2.8×10^{-3},精度较高. 将该信号作用于生理皮层模型,由图7可见,使用该信号替代本体反馈后,由于使用BP-PID设计的刺激信号在0.5 s~0.7 s时间内误差较大,使用RLBP-PID设计的控制器能够更好地跟踪PPV神经元的放电活动,跟踪误差降低50 %以上. 最后将设计好的控制信号代入肢体运动模型,由图9可见,在系统超调量接近一致的情况下,使用BP-PID设计的控制器出现了滞后现象,导致峰值时间、调节时间和上升时间均变长,误差较大,系统响应速度变慢,无法在规定时间内跟踪自然反馈条件下的主动肌运动轨迹,实时性变差. 由于本次实验的主动肌位置设定为0.7 m,当肢体运动时间超过设定时间后,肢体位置会在0.7 m处趋向稳定. 综上所述,使用RLBP-PID算法设计皮层刺激信号在跟踪期望运动轨迹上执行效果较好,恢复了主动肌的位置轨迹,也进一步验证了在无本体反馈的情况下设计人工刺激信号能够恢复解码器的闭环性能.

4　结　论

本文首先对大脑皮层神经元和肢体运动模型进行仿真,获取了在无视觉反馈时大脑区域4和区域5

的放电数据;然后针对大脑区域4使用卡尔曼滤波器和BN-BP神经网络进行解码器设计以替代脊髓电流.根据后续实验发现,当系统不存在本体反馈,即系统处于开环状态时,解码器的性能显著下降.针对上述问题,提出了基于强化学习的神经网络PID控制器设计刺激电流以恢复单关节自发运动的闭环性能.通过将基于强化学习的神经网络PID 控制与传统控制算法进行比较可以发现,前者的效果明显优于后者,也进一步验证了外在刺激电流能够恢复大脑皮层神经元的放电活动.

在当前的实验性闭环脑机接口研究中,对于皮层刺激电流的设计,采取的是电荷平衡双相波形的反馈电流形式.此电流形式通常用于在外部刺激脑皮层感觉区神经元,从而在脑机接口的闭环操作期间提供人工感觉反馈.后续的研究可以利用刺激输入电流约束在闭环脑机接口中的双相波形来设计最优人工感觉反馈,实现基于广义刺激的闭环脑机接口框架.

参考文献(References)

[1] Nicolelis M A L, Lebedev M A. Principles of neural ensemble physiology underlying the operation of brain-machine interfaces[J]. Nature Review Neuroscience, 2009, 10(7): 530-540.

[2] Rodolphe H, Karunesh G, Jessica J, et al. Learning in closed-loop brain-machine interfaces: Modeling and experimental validation[J]. IEEE Trans on Systems, Man, and Cybernetics, Part B: Cybernetics, 2010, 40(5): 1387-1397.

[3] Bullock D, Cisek P, Grossberg S. Cortical networks for control of voluntary arm movements under variable force conditions[J]. Cerebral Cortex, 1998, 8(1): 48-62.

[4] Sanchez J C, Erdogmus D, Rao Y, et al. Interpreting neural activity through linear and nonlinear models for brain machine interfaces[C]. American Control Conf. Chicago: IEEE, 2003: 2160-2163.

[5] García-Córdova F. A cortical network for control of voluntary movements in a robot finger[J]. Neurocomputing, 2007, 71(1/2/3): 374-391.

[6] Katona J, Kovari A. EEG-based computer control interface for brain-machine interaction[J]. Int J of Online Engineering, 2015, 11(6): 43-48.

[7] Shanechi M M, Williams Z M, Wornell G W, et al. A real-time brain-machine interface combining motor target and trajectory intent using an optimal feedback control design[J]. Plos One, 2013, 8(4): e59049.

[8] Kumar G, Schieber M H, Thakor N V, et al. Designing closed-loop brain machine interfaces using optimal receding horizon control[C]. American Control Conf. Washington DC: IEEE, 2013: 5029-5034.

[9] Pan H G, Ding B C, Zhong W M, et al. Designing closed-loop brain-machine interfaces with network of spiking neurons using MPC strategy[C]. American Control Conf. Chicago: IEEE, 2015: 2543-2548.

[10] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]. Int Conf on Machine Learning. France: IEEE, 2015: 448-456.

[11] Theodorou E, Buchli J, Schaal S, et al. A generalized path integral control approach to reinforcement learning[J]. J of Machine Learning Research, 2010, 11(11): 3137-3181.

[12] Dangi S, Gowda S, Héliot R, et al. Adaptive Kalman filtering for closed-loop brain-machine interface systems[C]. IEEE Int Conf on Neural Engineering. Cancun: IEEE, 2011: 609-612.

[13] Sussillo D, Nuyujukian P, Joline M F, et al. A recurrent neural network for closed-loop intra cortical brain-machine interface decoders[J]. J of Neural Engineering, 2012, 9(2): 1-10.

[14] Li W L, Jia Y M, Du J P. Distributed Kalman consensus filter with intermittent observations[J]. J of the Franklin Institute, 2015, 352(9): 3764-3781.

[15] Shafin R, Liu L J, Zhang J Z. DoA estimation and RMSE characterization for 3D massive-MIMO/FD-MIMO OFDM system[C]. Global Communications Conf. New York: IEEE, 2015: 1-6.

[16] Yang H Y, Guo Q, Xu X, et al. Self-learning PD algorithms based on approximate dynamic programming for robot motion planning[C]. Int Joint Conf on Neural Networks. Beijing: IEEE, 2014: 3663-3670.

[17] Eltantawy S, Abdulhai B, Abdelgawad H. Design of reinforcement learning parameters for seamless application of adaptive traffic signal control[J]. J of Intelligent Transportation Systems, 2014, 18(3): 227-245.

[18] Khan S G, Herrmann G, Lewis F L, et al. Reinforcement learning and optimal adaptive control: An overview and implementation examples[J]. Annual Reviews in Control, 2012, 36(1): 42-59.

[19] Figueroa R, Faust A, Cruz P, et al. Reinforcement learning for balancing a flying inverted pendulum[C]. Intelligent Control and Automation. Shenyang: IEEE, 2015: 1787-1793.

[20] Lin L G, Xie H B, Shen L C. Application of reinforcement learning to autonomous heading control for bionic underwater robots[C]. IEEE Int Conf on Robotics and Biomimetics. Guilin: IEEE, 2009: 2486-2490.

(责任编辑：郑晓蕾)