

[自控·检测]

DOI:10.3969/j.issn.1005-2895.2014.04.016

基于关系网络的新型学习控制

闫乐, 董宇

(浙江大学电气工程学院, 浙江 杭州 310027)

摘要:设计了一种不同于传统智能算法如自适应控制与模糊控制等的智能控制算法,该算法基于目标输入关系网络,借鉴强化学习的思想,采用输出试探与评价进行学习,是一种模型无关的控制算法,该算法可以与随机优化算法结合,通过实验,验证了其在控制过程中有效性,并通过与遗传算法结合,验证了与遗传算法相结合的该智能算法的控制效果。

关键词:智能算法;强化学习;关系网络;权值矩阵;稳定控制

中图分类号:TP273.2

文献标志码:A

文章编号:1005-2895(2014)04-0065-04

New Learning Control Algorithm Based on Relationship Network

YAN Le, DONG Yu

(College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China)

Abstract: A intelligent control algorithm was presented, which was different from traditional intelligent algorithms like adaptive control and fuzzy control. This intelligent control algorithm was based on the target-input relationship network, which learned from reinforcement learning, and used try and evaluation to learn. It was a model-independent control algorithm. This control algorithm could be combined with stochastic optimization algorithms. Experiments verify the effectiveness of this algorithm by combination with genetic algorithms.

Key words: smart algorithm; reinforcement learning; network of relationships; weight matrix; stability control

近年来,芯片的计算能力不断提升。嵌入式芯片得以大规模普及,智能家居、智能输电不断涌现。此时,人们可以相应地提出新的智能控制算法。传统的控制算法要求人们了解被控对象的数学模型,或通过辨识被控对象的数学模型进行控制,对于模型难以辨识的被控对象,则需要参考专家经验进行控制,如PID控制与模糊控制^[1-2]。

如何在了解被控对象的数学模型,且没有专家经验可以参考的情况下,仅凭算法本身的试探与经验进行学习,从而对被控对象进行适当的控制,是本文提出的控制算法所要达到的目的。

本文的思想源于对生物学习的思考,生物通过试探与记忆学习新事物,这种思想与强化学习^[3]的思想不谋而合。

本算法是一种与模型无关的控制算法,这一点上与预测控制^[4]相近。但是相对于预测控制,本算法不要求被控系统是线性系统,且本算法是一种学习型自

适应算法,控制能力与效果可随着模型变化而变化,随着时间逐渐优化。

1 算法的具体实现

本算法是一种试探性算法,广义的试探包含算法的试探,试验不同的控制算法,根据不同的工况与控制结果从不同的算法中选出最优的算法或整合出最优的算法。狭义的试探指的是输出的直接试探。文中为了设计与实验方便,采用直接试探设计控制算法。

该智能控制算法主要包括状态分割、输出随机选择、输出插值与关系网络更新4个组成部分。

1.1 学习算法的状态分割

被控系统的输入信号都呈连续变化的,构成了连续的输入空间。在连续输入空间内,学习算法的搜索范围太大。因此,在建立关系网络前,需要采取量化的方法,将连续的输入空间映射到离散的输入空间,降低输入空间的复杂度。量化的方法主要有BOX方法、Kohonen神经网络方法及模糊方法^[5]。

收稿日期:2013-12-26;修回日期:2014-02-06

作者简介:闫乐(1988),男,辽宁葫芦岛人,硕士研究生,主要研究方向为智能控制与智能算法。E-mail:yanhappy@zju.edu.cn

本设计算法为了便于仿真,先针对单输入、单输出控制系统进行设计,多输入、多输出的被控系统可以类比设计。

被控系统的输出即环境的输出为控制器的输入,在实际处理中为误差 e ;被控系统的输入为控制器的输出,在实际处理中为输出 μ 。在控制多输入、多输出系统时,可以通过前置神经网络或者其他分类器将输入状态分为有限个输入状态。在单输入单输出时,直接对输入输出做出线性分割,分割情况如表1所示。

表1 状态分割表

Table 1 Status partition table

e 的范围	对应状态	μ 的范围	对应状态
$(-\infty, -0.75]$	1	$(-1.00, -0.75]$	1
$(-0.75, -0.50]$	2	$(-0.75, -0.50]$	2
$(-0.50, -0.25]$	3	$(-0.50, -0.25]$	3
$(-0.25, 0.00]$	4	$(-0.25, 0.00]$	4
$(0.00, 0.25]$	5	$(0.00, 0.25]$	5
$(0.25, 0.50]$	6	$(0.25, 0.50]$	6
$(0.50, 0.75]$	7	$(0.50, 0.75]$	7
$(0.75, +\infty]$	8	$(0.75, 1.00]$	8

状态分割矩阵不是一成不变的,还可以采用非线性分割方式。对于不同的控制状态,可以灵活对控制范围进行缩放。例如在误差范围变化时,用新的误差波动范围与新的输出范围代替旧的波动范围,在误差 e 缩小时,缩小输出 μ 的范围,从而使控制更加平稳,在某些情况下,利于系统的稳定控制。

1.2 关系网络

误差 e 与输出 μ 一一对应的网络,称之为关系网络。在表1中定义的状态分割中,我们可知关系网络的权重矩阵 $W_{n \times n}$ 为 8×8 的矩阵,当然,采用不同的量化方法,可以有不同大小的权重矩阵。

$W_{n \times n}$ 权重矩阵需要的初始化,可以通过离线学习得到,也可直接设为零矩阵。根据 $W_{n \times n}$ 权重矩阵,在出现相应误差 e 时,我们可以选择最有利的输出 μ 。选择不同的权重矩阵学习函数,我们将得到不同的 $W_{n \times n}$ 权重矩阵,从而满足不同的控制要求。

1.3 输出选择策略

选择建立在对输出 μ 的评价上,该评价体现在 $W_{n \times n}$ 权重矩阵中。在线学习时,初始阶段权重网络矩阵 $W_{n \times n}$ 为零矩阵。为防止控制算法陷入局部最优,初始时采用随机算法选择输出。随机动作输出策略有基于连接权重比例轮盘式选择方法^[6],公式如下:

$$p_{i^*j} = w_{i^*j} / \sum_{j=1}^n w_{i^*j} \quad (1)$$

$i = i^*$ 是输入 e 对应状态。 i 表示误差 e 对应状态, j 表示输出 μ 对应状态, p_{i^*j} 表示输入为 i^* 时,输出 j 被选中概率。

排序选择方式:根据连接权重 $w_{i,j}$ 排序赋予相应选中概率 p_{i^*j} 。

此外,还有随机联赛选择方式^[7]等。本文选择基于排序的线性选择方式。

由于输出只有有限状态,为保证控制平稳,需要对输出进行插值,可以选择线性插值,非线性插值等,本文应用插值公式如下:

$$u = u_{jd} + (e - e_{id}) * du/de \quad (2)$$

其中, u 与 e 为实际输出与输入。 u_{jd} 与 e_{id} 分别是状态对应的离散值, du 与 de 分别是 u 与 e 的离散化间隔。

1.4 关系网络的更新运算

在线学习时,初始阶段权重网络矩阵 $W_{n \times n}$ 为零矩阵,通过 $e_t, e_{t+1}, u_t, u_{t-1}$ 修正 $W_{n \times n}$ 权重网络。一般公式为

$$w_{i,j} = \alpha(|e_t - e^*| - |e_{t+1} - e^*|) + \gamma w_{i,j} \quad (3)$$

其中, e^* 为设定目标,一般为0。 e_t 与 e_{t+1} 为本次误差与下次误差, γ 为折扣因子,在 $[0,1]$ 之间, $\alpha(|e_t - e^*| - |e_{t+1} - e^*|)$ 为新息, α 为学习步长。可见,若输出使误差接近目标,则关系网络连接权重加大,反之,减小。

当 $e_t = e_{t+1} = e^*$ 时, $w_{i,j} = \gamma w_{i,j} + \alpha$,以保证在达到目标后,还能得到更新。

在权值矩阵 $W_{n \times n}$ 稳定后,可以直接选择最大 $w_{i,j}$ 对应的 u ,而不必继续采用随机选择策略。

在设计最优控制器时,若需要 u 平稳输出,关系网络修正公式为

$$w_{i,j} = \alpha(|e_t - e^*| - |e_{t+1} - e^*|) + \gamma w_{i,j} - \beta(|u_t - u_{t-1}|) \quad (4)$$

可见,通过修正公式的优化设计,可以灵活控制网络权值矩阵 $W_{n \times n}$ 的更新,使得控制器设计灵活合理,满足不同的控制要求。算法的整体流程图如图1所示。

1.5 收敛性

在线学习时,若被控系统是稳定系统或临界稳定系统,则由于控制器输出有限,所以被控系统的误差 e 必然为有限值,显然此时新息 $\alpha(|e_t - e^*| - |e_{t+1} - e^*|)$ 为有限值,设新息最大值为 M 。

当关系网络更新次数有限时,显然算法的关系网络连接系数 $w_{i,j}$ 收敛。当更新次数无限时, $w_{i,j}$ 最大增长公式为

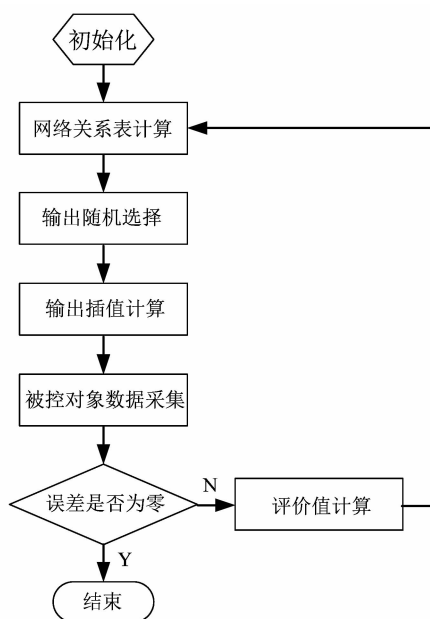


图 1 控制算法流程图

Figure 1 Flow chart of the control algorithm

$$w_{i,j} = \gamma w_{i,j} + M \quad (5)$$

显然, $w_{i,j_{\max}} = M/1 - \gamma$ 是有限值, 关系网络的权值矩阵 $W_{n \times n}$ 收敛。

若根据权值矩阵 $W_{n \times n}$ 选择输出, 根据公式(3), 选择的输出 u 将使误差值 e 逐渐逼近目标 e^* , 系统则一定会收敛。

可知, 若被控系统是临界稳定系统或者稳定系统, 则在本算法的控制下, 系统一定收敛。若被控系统是不稳定系统, 存在不稳定零极点, 误差不一定存在极限, 则本算法不能保障收敛。事实上, 模型无关算法如何控制不稳定系统一直是一个难题。

若为离线学习, 无论被控对象是否稳定, 本算法均可保证收敛。

2 策略空间寻优

对于单输入单输出控制系统而言, 直接将输出离散化、构建关系网络比较合适。在多数情况下, 尤其在现代控制系统中, 多数是多输入多输出系统, 且不同的输入与输出之间存在着耦合关系。此时, 按照上述方法直接构建关系网络会导致权重矩阵维数变得很大, 此时运算将变得非常繁琐以致在线控制难以实现。因此, 需要简化策略寻优空间, 应用智能优化算法^[8] 选择合适的输出, 从而避免繁琐的运算。通过少数初始解在迭代过程中寻找最佳的输出组合。

遗传算法(GA)^[9] 适合离散空间寻优, 此特点是其它智能寻优算法难以比拟。并且 GA 算法具有潜在

并行性以及收敛性, 因此, 选择 GA 算法作为策略寻优算法。

GA 算法可以有效解决控制器多个输出 $U = (u_1, u_2, \dots, u_l)$ 的问题, 对于多个输入 $E = (e_1, e_2, \dots, e_m)$, 一般没有太好的方法, 直接扩展关系网络比较合适。对于控制目标主次分明的还可以用加权法将多个 e 整合为一个, 精度要求不高的可以采用径向基函数神经网络^[10] 映射。

$U = (u_1, u_2, \dots, u_l)$ 量化之后可以仍为 8 组, 随机初始化或均匀初始化, 一般应该选择均匀设置, 防止过早陷入局部最优解。

选择运算时, 输出 $U = (u_1, u_2, \dots, u_l)$ 被选择的概率与 $W_{n \times n}$ 权重网络系数有关, 一般排序选择。如按照排序, 被选择概率依次为 $[1.6 \ 1.4 \ 1.2 \ 1.0 \ 0.8 \ 0.6 \ 0.4 \ 0.2 \ 0.0]$, 被选择后的输出组合作为父代。交叉运算时, 按照排序产生的概率选择父代, 接着在输出 $1 \sim l$ 间选取一点, 进行交叉互换, 交叉后该输出 U 对应的 $W_{n \times n}$ 权重系数为父代的平均值。

为保证收敛, 每次 GA 运算, 变异运算只进行一次, 且最优父代不参与变异, 保证网络的收敛与稳定。因为初始化时, $W_{n \times n}$ 权重网络为零矩阵, 为保证 GA 运算时有完备的 $W_{n \times n}$ 权重网络可供参考, 采用近似线性化方法快速建立 $W_{n \times n}$ 系数矩阵。

3 实验结果与分析

采用双闭环伺服电机系统作为被控对象。在线直接试探的控制效果如图 2 所示, 学习稳定后, 再次进行控制效果如图 3 所示。

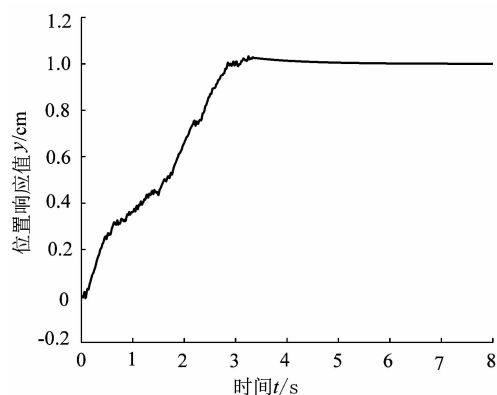


图 2 直接试探控制效果图

Figure 2 Result figure of direct control

对比图像可知, 在线控制由于有试探, 系统存在一定的波动。而在学习之后, 借助已经生成的关系矩阵, 控制算法完全可以稳定地控制被控对象。

为了检验加入 GA 算法进行寻优的控制算法, 设

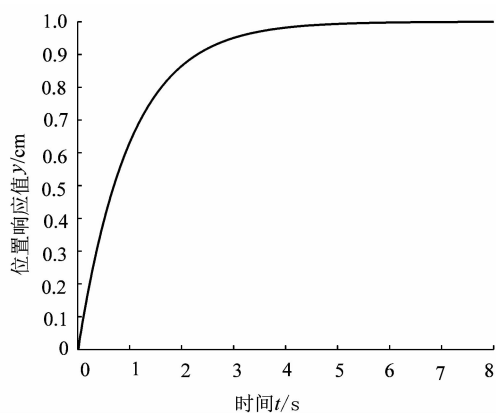


图3 学习稳定后控制效果图

Figure 3 Result figure of control after learning

计了1个临界稳定系统,它需要1个单输入,4输出的控制器。

直接使用PID进行控制的效果如图4所示,结合GA算法的智能算法控制效果如图5所示,未结合GA算法的智能算法控制效果如图6所示。

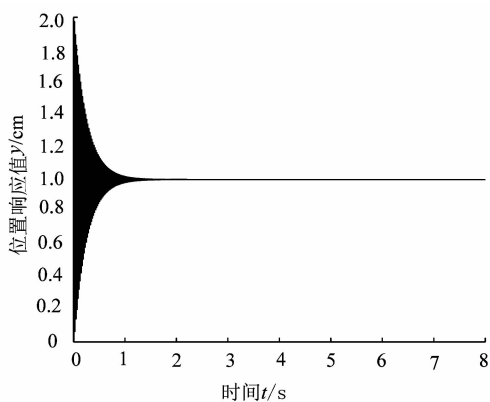


图4 PID控制效果图

Figure 4 Result figure of PID control

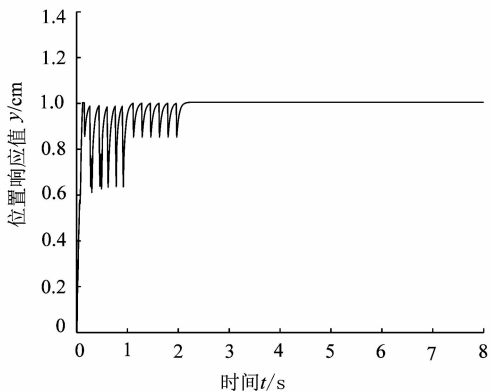


图5 加入GA的智能控制算法控制效果图

Figure 5 Result figure of intelligent control with GA

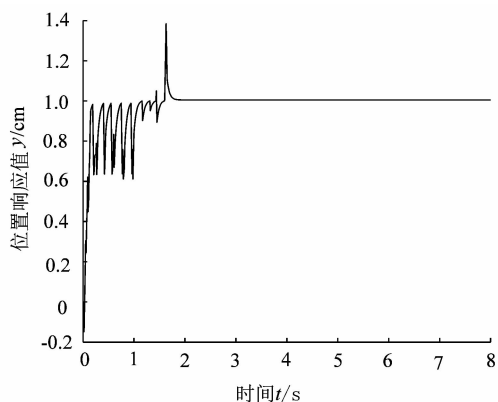


图6 未加入GA的智能控制算法控制效果图

Figure 6 Result figure of intelligent control without GA

可见,GA控制使控制波动平稳。但是,在运行过程中,有时GA算法会陷入局部最优解,从而导致控制出现静态误差。结合GA寻优的智能控制算法还可以进一步完善。

4 结 语

本文采用强化学习的思想,设计一种基于关系网络的模型无关智能控制算法,在实验过程中,取得了不错的效果,并且通过与GA算法相融合,成功应用于多输入系统。针对权重矩阵 $W_{n \times n}$ 不同的改进策略,可以设计满足不同要求的控制器。应该看到,本算法还有很多不足之处,需要在今后进一步深入研究。

参考文献:

- [1] 吴振顺,姚建均,岳东海. 模糊自整定PID控制器的设计及其应用[J]. 哈尔滨工业大学学报,2004,36(11):1578-1580.
- [2] 王庆明,孙怡. 基于模糊PID控制的直流电机同步控制系统[J]. 机电工程,2012,29(5):493-496.
- [3] 高阳,陈世福,陆鑫. 强化学习研究综述[J]. 自动化学报,2004,30(1):86-100.
- [4] 李国勇. 智能预测控制及其MATLAB实现[M]. 北京:电子工业出版社,2010:255-268.
- [5] 张汝波. 强化学习研究及其在AUV导航系统中的应用[D]. 哈尔滨:哈尔滨工程大学,1999.
- [6] 高玮,尹志喜. 现代智能仿生算法及其应用[M]. 北京:科学出版社,2011:54-55.
- [7] 刘文豪,黄宗南. 针对作业调度问题的遗传选择算子性能分析[J]. 机械制造,2011,49(558):79-82.
- [8] 高永超. 智能优化算法的性能及搜索空间研究[D]. 山东:山东大学,2007.
- [9] 史峰,王辉,郁磊,等. Matlab智能算法30个案例分析[M]. 北京:北京航空航天大学出版社,2011:17-26.
- [10] 史忠植. 神经网络[M]. 北京:高等教育出版社,2009:140-164.