



(12) 发明专利

(10) 授权公告号 CN 112632860 B

(45) 授权公告日 2024.06.04

(21) 申请号 202110002104.6

G06N 20/00 (2019.01)

(22) 申请日 2021.01.04

G06F 111/08 (2020.01)

G06F 119/14 (2020.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 112632860 A

(43) 申请公布日 2021.04.09

(73) 专利权人 华中科技大学

地址 430074 湖北省武汉市洪山区珞喻路
1037号

(72) 发明人 丁建完 陈立平 郭超 彭奇

(74) 专利代理机构 华中科技大学专利中心

42201

专利代理师 李智

(51) Int. Cl.

G06F 30/27 (2020.01)

G06F 30/25 (2020.01)

(56) 对比文件

CN 107341294 A, 2017.11.10

CN 109522602 A, 2019.03.26

US 2019347370 A1, 2019.11.14

吴义忠; 蒋占四; 陈立平. 基于Modelica语言的多领域模型仿真优化研究. 系统仿真学报. 2009, (12), 全文.

审查员 徐生芹

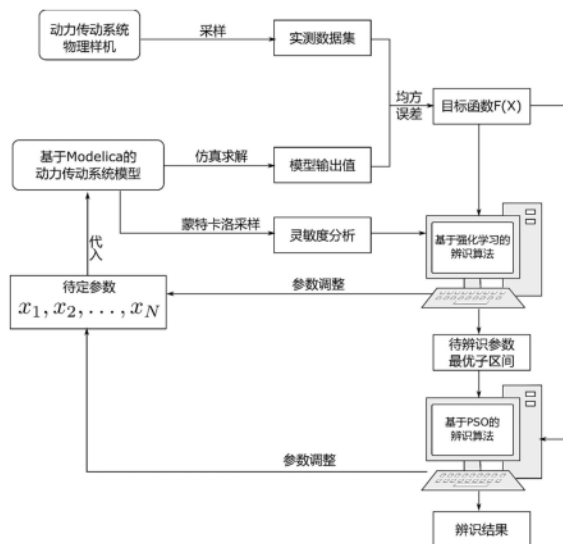
权利要求书3页 说明书10页 附图2页

(54) 发明名称

一种基于强化学习的动力传动系统模型参数辨识方法

(57) 摘要

本发明公开了一种基于强化学习的动力传动系统模型参数辨识方法,属于系统建模仿真领域。本发明针对动力传动系统模型参数敏感度不一致,现有辨识算法收敛速度慢、搜索范围要求较高的问题,构建用于动力传动系统模型参数辨识的强化学习框架,具有防止局部最优、收敛速度快、搜索范围大的优点;本发明采用分阶段的辨识流程,粗调阶段利用强化学习收敛速度快、搜索范围大的特点,快速找到各参数的最优子区间,微调阶段利用启发式算法精度高、全局搜索能力强的特点,在最优子区间中确定最终辨识结果,与只使用单个算法相比,分阶段的辨识效果更好。



1. 一种基于强化学习的动力传动系统模型参数辨识方法,其特征在于,包括:
S1. 构建基于多领域统一建模语言Modelica的动力传动系统动力学模型;
S2. 对模型待辨识参数进行灵敏度分析;
S3. 基于强化学习算法对待辨识参数进行粗调:

构建用于Modelica动力传动系统模型参数辨识的强化学习框架;
利用强化学习框架进行迭代训练,得到每个待辨识参数的最优子区间;
强化学习框架构建过程具体包括:

(1) 将模型估计值 Y_{est} 与实测值 Y_{mea} 的均方误差作为强化学习目标函数 $F(X)$;

(2) 构建单步奖赏:

$$r = \min(1, \max(0, (F(X_{mean}) - F(X_{best})) / (F(X_{mean}) - F(X_{cur}))))$$

其中, r 表示单步奖赏值, $F(X_{cur})$ 表示当前参数下的目标函数值, $F(X_{best})$ 表示最优参数下的目标函数值, $F(X_{mean})$ 表示参数平均值下的目标函数值;

(3) 根据参数最小变化量 G_i ($i=1, 2, \dots, N$) 及每个参数的范围设置动作:

将第 i 个参数的搜索范围拆分为 $D_i = \frac{x_{imax} - x_{imin}}{G_i}$ 个子区间,选中一个子区间后,在该

子区间随机获取一个值作为动作;其中, x_{imax} 为第 i 个参数最大值, x_{imin} 为第 i 个参数最小值;最小变化量 G_i 指第 i 个参数在辨识过程中每步增加或减少的变化量;

(4) 构建动作选择策略:

01. 选择搜索路径:

确定下一轮选择的动作在当前动作的左边还是右边,选择指标为 $L_p(i, i)$,计算公式为:

$$L_p(i, j) = \sum_{n=1}^k \lambda_1^{n-1} q_{i,n}^* \quad p = 1, 2$$

其中, k 表示参数变换组合的数量; $q_{i,n}^*$ 为第1个路径上与当前动作 $a_{i,j}$ 相邻的 k 个动作中第 n 大的 Q 值, λ_1 为路径权重系数;

获取一个 $[0, 1]$ 之间的随机数 ε_1 ,使用如下公式确定搜索路径1:

$$l = \begin{cases} \operatorname{argmax}_p (L_p(i, j)) & \text{if } \varepsilon_1 > 0.3 \\ \operatorname{rand}(1, 2) & \text{else} \end{cases}$$

$\operatorname{rand}(1, 2)$ 表示1-2区间内的随机概率分布;

02. 确定动作:

获取一个 $[0, 1]$ 之间的随机数 ε_2 ,使用如下公式确定动作 a :

$$a = \begin{cases} \operatorname{argmax}_m q(i, m) & m = j - k, \dots, j - 1 \quad \text{if } \varepsilon_2 > 0.3 \\ \operatorname{rand}(j - k, j - k + 1, \dots, j - 1) & \text{else} \end{cases}$$

$q(i, m)$ 表示第 i 个待辨识参数的 Q 值, ε_1 和 ε_2 都是用来保证强化学习探索性的随机数;

(5) 构建 Q 值函数的更新策略:

第*i*个参数对应的Q值函数更新公式如下:

$$Q^{r+1}(i, j) += \alpha (r + (1 - \lambda_2) \max(L_p(i, j)) + \lambda_2 \min(L_p(i, j)) - Q^r(i, j))$$

式中 α 为控制学习速率的超参数, r 为单步奖赏, λ_2 为控制更新幅度的超参数;

S4. 参数微调:

以实测数据与模型估计值的均方误差为目标函数,在待辨识模型参数所形成的解空间中迭代寻优,以得到目标函数值最小时对应的各参数值,作为最终的辨识结果。

2. 根据权利要求1所述的一种基于强化学习的动力传动系统模型参数辨识方法,其特征在于,步骤S2具体为,使用sobel方法对参数待定的动力传动系统模型进行参数灵敏度分析,具体步骤如下:

01. 在*N*个待辨识参数的可能取值区间范围内分别进行蒙特卡洛采样,生成初始样本矩阵一A、初始样本矩阵二B以及交叉样本矩阵 \mathbf{A}_B^i ;其中*i* = {1, 2, ..., *N*};

02. 将样本矩阵A、B、 \mathbf{A}_B^i 作为输入对动力传动系统模型进行仿真求解,分别得到初始样本矩阵A、初始样本矩阵B以及交叉样本矩阵 \mathbf{A}_B^i 的模型仿真结果向量f(A)、f(B)、 $\mathbf{f}(\mathbf{A}_B^i)$;

03. 基于仿真结果和以下公式求出每个参数的全局影响指数 S_{Ti} :

$$S_{Ti} = \frac{\frac{1}{2N} \|\mathbf{f}(\mathbf{A}) - \mathbf{f}(\mathbf{A}_B^i)\|^2}{\text{Var}(\mathbf{Y})}$$

其中, \mathbf{Y} 表示f(A)、f(B)和 $\mathbf{f}(\mathbf{A}_B^i)$ 构成的向量集合;Var(Y)表示动力传动系统模型输出的方差;

04. 根据参数全局影响指数的大小对待辨识参数进行敏感度排序;影响因子越大表示越敏感;

05. 将灵敏度低于设定阈值的待辨识参数进行组合。

3. 根据权利要求1所述的一种基于强化学习的动力传动系统模型参数辨识方法,其特征在于,参数粗调阶段,使用基于Q-Learning算法的强化学习框架,迭代训练的具体步骤如下:

(1) 随机初始化所有待辨识参数,并代入动力传动系统模型中计算,与实测数据相比得到均方误差作为最优目标函数F(X)的初始值;

(2) 多个智能体按照待辨识参数灵敏度从大到小执行串行学习;强化学习框架中根据Q值表调整待辨识参数的环节称为智能体,每个待辨识参数都唯一对应一个智能体;智能体调整待辨识参数的过程称为智能体的学习行为;

学习过程包括:在当前智能体对应参数的可能取值区间范围内随机选取一个动作 $a_{\text{rand}}(i)$,固定其它参数,将此参数应用到动力传动系统模型中,得到当前参数下的目标函数值F(X_{cur})以及单步奖赏值r,根据Q值函数的更新策略更新当前智能体的Q值表;若F(X_{cur}) \leq F(X_{best}),认为当前参数有搜索价值,转入步骤(3),否则重复上述学习过程;

(3) 所述动作选择策略选择一个动作 $a_{\text{iter}}(i)$,固定其它参数,将此参数应用到FMU中,得到目标函数值F(X_{cur})以及奖惩值r,更新Q值表;

(4) 针对第*i*个参数迭代执行步骤(2)与(3) $I_{e_{max}}$ 次,完成第*i*个参数的一次辨识,转入第*i*+1个参数的辨识;

(5) 所有待辨识参数串行执行步骤(2)~(4)后完成一个训练周期,若当前完成周期数小于给定训练次数 $I_{s_{max}}$,转步骤(2)进入下一个训练周期,否则训练结束。

4. 根据权利要求1所述的一种基于强化学习的动力传动系统模型参数辨识方法,其特征在于,步骤S4具体为,使用PSO优化算法对参数进行微调,得到最终的辨识结果。

5. 根据权利要求1所述的一种基于强化学习的动力传动系统模型参数辨识方法,其特征在于,使用PSO优化算法对参数进行微调,具体步骤如下:

将微调范围设置为 $[x_{i_{opt}} - \mu G_i, x_{i_{opt}} + (\mu + 1)G_i]$;其中, μ 值根据粒子群算法的搜索能力设置, $[x_{i_{opt}}, x_{i_{opt}} + G_i]$ 为强化学习得到的最优子区间, $x_{i_{opt}}$ 是第*i*个参数经过强化学习辨识后的最优值;

初始化一个N维空间以及粒子群,创建每个粒子的速度向量 v_i 、位置向量 x_i 、历史最优位置向量 p_i ,以及整个粒子群的历史最优位置向量 p_g ;

以模型预测值与实测值的均方误差建立适应度函数G(X),模型预测值由粒子的当前位置代入FMU中仿真求解得到;整个粒子群的搜索目标是使G(X)最小;

开始迭代搜索:在一轮迭代中,计算每个粒子的适应度值,更新单个粒子的 p_i 值及种群的 p_g 值,再根据下式更新每个粒子的速度和位置向量;

$$v_i^{t+1} = \omega v_i^t + c_1 r_1 (p_i^t - x_i^t) + c_2 r_2 (p_g^t - x_i^t)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1}$$

式中 ω 为惯性权重, c_1 、 c_2 为学习因子, r_1 、 r_2 为[0,1]之间的随机数;

达到最大迭代次数后,粒子群的历史最优位置向量 p_g 即为参数辨识的最终结果。

6. 一种基于强化学习的动力传动系统模型参数辨识系统,其特征在于,包括:计算机可读存储介质和处理器;

所述计算机可读存储介质用于存储可执行指令;

所述处理器用于读取所述计算机可读存储介质中存储的可执行指令,执行权利要求1至5任一项所述的基于强化学习的动力传动系统模型参数辨识方法。

一种基于强化学习的动力传动系统模型参数辨识方法

技术领域

[0001] 本发明属于系统建模仿真领域,更具体地,涉及一种基于强化学习的动力传动系统模型参数辨识方法。

背景技术

[0002] 随着多领域系统建模仿真技术的不断发展,Modelica语言逐渐成为了多领域多学科系统仿真领域发展的行业标准,其目标是定义一种通用的用于复杂系统建模的编程语言。Modelica模型即为使用Modelica语言编写的具有面向对象建模、多领域统一建模、非因果陈述式建模和连续离散混合建模能力的程序,可用于不同行业的工程师搭建自己的仿真模型系统,开展相应的动态仿真。对于Modelica模型而言,想要尽可能地与物理样机保持一致,关键的一点就在于模型参数的设置,搭建模型的过程只是确定了模型的基本形式,想要获得最优的仿真性能,必须设置精确的模型参数。

[0003] 动力传动系统是在控制器控制下从功率源以速度/角速度、力/力矩形式获取能量并传递到系统下一个环节的装置,主要包括电动机、减速器等部件,动力传动过程中存在能量损失、速度改变等状态。

[0004] 现有的模型参数辨识方法有多种,目前最常用的即为最小二乘法,该方法是建模软件中默认的参数优化方法。但是,最小二乘估计对数据的依赖性过高,辨识结果容易受噪声影响,且对参数的初始化范围要求比较苛刻。另外比较常用的辨识方法为进化算法,如粒子群算法(Particle Swarm Optimization,PSO)、遗传算法(Genetic Algorithm,GA)等,这类方法逼近全局最优的能力较强,但收敛速度较慢,精度也有待提高。因此,急需一种收敛速度快、对参数初始范围要求不高的辨识方法来对基于Modelica的动力传动系统模型进行参数辨识。

发明内容

[0005] 针对现有技术的以上缺陷或改进需求,本发明提供了一种基于强化学习的动力传动系统模型参数辨识方法,其目的在于提出一种收敛速度快、对参数初始范围要求不高的辨识方法来对基于Modelica的动力传动系统模型进行参数辨识。

[0006] 为实现上述目的,本发明提供了一种基于强化学习的动力传动系统模型参数辨识方法,包括:

[0007] S1. 构建基于多领域统一建模语言Modelica的动力传动系统动力学模型;

[0008] S2. 对模型待辨识参数进行灵敏度分析;

[0009] S3. 基于强化学习算法对待辨识参数进行粗调;

[0010] 构建用于Modelica动力传动系统模型参数辨识的强化学习框架;

[0011] 利用强化学习框架进行迭代训练,得到每个待辨识参数的最优子区间;

[0012] S4. 参数微调;

[0013] 以实测数据与模型估计值的均方误差为目标函数,在待辨识模型参数所形成的解

空间中迭代寻优,以得到目标函数值最小时对应的各参数值,作为最终的辨识结果。

[0014] 进一步地,步骤S2具体为,使用sobo1方法对参数待定的动力传动系统模型进行参数灵敏度分析,具体步骤如下:

[0015] 01.在N个待辨识参数的可能取值区间范围内分别进行蒙特卡洛采样,生成初始样本矩阵A、初始样本矩阵B以及交叉样本矩阵 \mathbf{A}_B^i ;其中 $i = \{1, 2, \dots, N\}$;

[0016] 02.将样本矩阵A、B、 \mathbf{A}_B^i 作为输入对动力传动系统模型进行仿真求解,分别得到初始样本矩阵A、初始样本矩阵B以及交叉样本矩阵 \mathbf{A}_B^i 的模型仿真结果向量 $f(A)$ 、 $f(B)$ 、 $\mathbf{f}(\mathbf{A}_B^i)$;

[0017] 03.基于仿真结果和以下公式求出每个参数的全局影响指数 ST_i :

$$[0018] \quad ST_i = \frac{\frac{1}{2N} \|\mathbf{f}(\mathbf{A}) - \mathbf{f}(\mathbf{A}_B^i)\|^2}{Var(\mathbf{Y})}$$

[0019] 其中,Y表示 $f(A)$ 、 $f(B)$ 和 $\mathbf{f}(\mathbf{A}_B^i)$ 构成的向量集合;Var(Y)表示动力传动系统模型输出的方差;

[0020] 04.根据参数全局影响指数的大小对待辨识参数进行敏感度排序;影响因子越大表示越敏感;

[0021] 05.将灵敏度低于设定阈值的待辨识参数进行组合。

[0022] 进一步地,强化学习框架构建过程具体包括:

[0023] (1)将模型估计值 Y_{est} 与实测值 Y_{mea} 的均方误差作为强化学习目标函数 $F(X)$;

[0024] (2)构建单步奖赏:

$$[0025] \quad r = \min(1, \max(0, (F(X_{mean}) - F(X_{best})) / (F(X_{mean}) - F(X_{cur}))) \cdot 0.00)$$

[0026] 其中,r表示单步奖赏值, $F(X_{cur})$ 表示当前参数下的目标函数值, $F(X_{best})$ 表示最优参数下的目标函数值, $F(X_{meab})$ 表示参数平均值下的目标函数值;

[0027] (3)根据参数最小变化量 G_i ($i = 1, 2, \dots, N$)及每个参数的范围设置动作:

[0028] 将第i个参数的搜索范围拆分为 $D_i = \frac{x_{imax} - x_{imin}}{G_i}$ 个子区间,选中一个子区间后,

在该子区间随机获取一个值作为动作;其中, x_{imax} 为第i个参数最大值, x_{imin} 为第i个参数最小值;最小变化量 G_i 指第i个参数在辨识过程中每步增加或减少的变化量;

[0029] (4)构建动作选择策略:

[0030] 01.选择搜索路径:

[0031] 确定下一轮选择的动作在当前动作的左边还是右边,选择指标为 $L_p(i, j)$,计算公式为:

$$[0032] \quad L_p(i, j) = \sum_{n=1}^k \lambda_1^{n-1} q_{i,n}^* \quad p = 1, 2$$

[0033] 其中,k表示参数变换组合的数量; $q_{i,n}^*$ 为第1个路径上与当前动作 $a_{i,j}$ 相邻的k个动作中第n大的Q值, λ_1 为路径权重系数;

[0034] 获取一个 $[0,1]$ 之间的随机数 ε_1 ,使用如下公式确定搜索路径1:

$$[0035] \quad l = \begin{cases} \operatorname{argmax}_p (L_p(i, j)) & \text{if } \varepsilon_1 > 0.3 \\ \operatorname{rand}(1, 2) & \text{else} \end{cases}$$

[0036] $\operatorname{rand}(1, 2)$ 表示1-2区间内的随机概率分布;

[0037] 02. 确定动作:

[0038] 获取一个 $[0,1]$ 之间的随机数 ε_2 ,使用如下公式确定动作a:

$$[0039] \quad a = \begin{cases} \operatorname{argmax}_m q(i, m) \quad m = j - k, \dots, j - 1 & \text{if } \varepsilon_2 > 0.3 \\ \operatorname{rand}(j - k, j - k + 1, \dots, j - 1) & \text{else} \end{cases}$$

[0040] $q(i, m)$ 表示第 i 个待辨识参数的Q值, ε_1 和 ε_2 是用来保证强化学习探索性的随机数;

[0041] (5) 构建Q值函数的更新策略:

[0042] 第 i 个参数对应的Q值函数更新公式如下:

$$[0043] \quad Q^{r+1}(i, j) = \alpha (r + (1 - \lambda_2) \max(L_p(i, j)) + \lambda_2 \min(L_p(i, j)) - Q^r(i, j))$$

[0044] 式中 α 为控制学习速率的超参数, r 为单步奖赏, λ_2 为控制更新幅度的超参数。

[0045] 进一步地,参数粗调阶段,使用基于Q-Learning算法的强化学习框架,迭代训练的具体步骤如下:

[0046] (1) 随机初始化所有待辨识参数,并代入动力传动系统模型中计算,与实测数据相比得到均方误差作为最优目标函数 $F(X)$ 的初始值;

[0047] (2) 多个智能体按照待辨识参数灵敏度从大到小执行串行学习;强化学习框架中根据Q值表调整待辨识参数的环节称为智能体,每个待辨识参数都唯一对应一个智能体;智能体调整待辨识参数的过程称为智能体的学习行为;

[0048] 学习过程包括:在当前智能体对应参数的可能取值区间范围内随机选取一个动作 $a_{\operatorname{rand}}(i)$,固定其它参数,将此参数应用到动力传动系统模型中,得到当前参数下的目标函数值 $F(X_{\operatorname{cur}})$ 以及单步奖赏值 r ,根据Q值函数的更新策略更新当前智能体的Q值表;若 $F(X_{\operatorname{cur}}) \leq F(X_{\operatorname{best}})$,认为当前参数有搜索价值,转入步骤(3),否则重复上述学习过程;

[0049] (3) 根据动作选择策略选择一个动作 $a_{\operatorname{iter}}(i)$,固定其它参数,将此参数应用到FMU中,得到目标函数值 $F(X_{\operatorname{cur}})$ 以及奖惩值 r ,更新Q值表;

[0050] (4) 针对第 i 个参数迭代执行步骤(2)与(3) $I_{e_{\max}}$ 次,完成第 i 个参数的一次辨识,转入第 $i+1$ 个参数的辨识;

[0051] (5) 所有待辨识参数串行执行步骤(2)~(4)后完成一个训练周期,若当前完成周期数小于给定训练次数 $I_{s_{\max}}$,转(2)进入下一个训练周期,否则训练结束。

[0052] 进一步地,步骤S4具体为,使用PSO优化算法对参数进行微调,得到最终的辨识结果,具体步骤如下:

[0053] 考虑到强化学习的辨识误差,将微调范围设置为 $[x_{i_{\operatorname{opt}}} - \mu G_i, x_{i_{\operatorname{opt}}} + (\mu + 1)G_i]$;

其中, μ 值根据粒子群算法的搜索能力设置, $[x_{i_{\operatorname{opt}}}, x_{i_{\operatorname{opt}}} + G_i]$ 为强化学习得到的最优子区

间, $x_{i_{opt}}$ 是第 i 个参数经过强化学习辨识后的最优值;

[0054] 初始化一个 N 维空间以及粒子群, 创建每个粒子的速度向量 v_i 、位置向量 x_i 、历史最优位置向量 p_i , 以及整个粒子群的历史最优位置向量 p_g ;

[0055] 以模型预测值与实测值的均方误差建立适应度函数 $G(X)$, 模型预测值由粒子的当前位置代入 FMU 中仿真求解得到; 整个粒子群的搜索目标就是使 $G(X)$ 最小;

[0056] 开始迭代搜索:

[0057] 在一轮迭代中, 计算每个粒子的适应度值, 更新单个粒子的 p_i 值及种群的 p_g 值, 再根据下式更新每个粒子的速度和位置向量;

$$[0058] \quad v_i^{t+1} = \omega v_i^t + c_1 r_1 (p_i^t - x_i^t) + c_2 r_2 (p_g^t - x_i^t)$$

$$[0059] \quad x_i^{t+1} = x_i^t + v_i^{t+1}$$

[0060] 式中 ω 为惯性权重, 默认为 0.6, c_1 、 c_2 为学习因子, r_1 、 r_2 为 $[0, 1]$ 之间的随机数;

[0061] 达到最大迭代次数后, 粒子群的历史最优位置向量 p_g 即为参数辨识的最终结果。

[0062] 总体而言, 通过本发明所构思的以上技术方案与现有技术相比, 能够取得下列有益效果。

[0063] (1) 本发明针对 Modelica 模型参数敏感度不一致, 现有辨识算法收敛速度慢、搜索范围要求较高的问题, 构建用于 Modelica 动力传动系统模型参数辨识的强化学习框架, 具有防止局部最优、收敛速度快、搜索范围大的优点; 本发明采用分阶段的辨识流程, 粗调阶段利用强化学习收敛速度快、搜索范围大的特点, 快速找到各参数的最优子区间, 微调阶段利用启发式算法精度高、全局搜索能力强的特点, 在最优子区间中确定最终辨识结果, 与只使用单个算法相比, 分阶段的辨识效果更好,

[0064] (2) 针对参数数量过多引起的搜索空间庞大的问题, 本发明还对待辨识参数进行了灵敏度分析, 将灵敏度较低的待辨识参数进行组合, 有助于减少待辨识参数的数量, 节约强化学习辨识参数过程中计算资源的开销; 同时利用灵敏度大小对参数辨识优先级进行设定, 能够为强化学习方法辨识参数过程中学习率设定提供依据, 有助于提高参数辨识精度。

[0065] (3) 本发明基于 Sobol 法设计了灵敏度分析方案, Sobol 法基于方差在可能的取值空间同时探索模型的所有参数, 并且对模型的非线性和单调性没有限制, 适合对任意模型进行参数全局灵敏度分析。

附图说明

[0066] 图1是本发明提供的基于强化学习的动力传动系统模型参数辨识方法流程图;

[0067] 图2是基于 Modelica 创建的动力传动系统模型;

[0068] 图3是 inertia2.J/spring.c/damping/和 Jmotor 参数的全局影响指数;

[0069] 图4是针对数控机床进给系统建立基于 Modelica 动力传递模型, 采用本发明方法进行参数辨识, 在相同输入条件下, 物理样机输出结果与模型仿真结果。

具体实施方式

[0070] 为了使本发明的目的、技术方案及优点更加清楚明白, 以下结合附图及实施例, 对本发明进行进一步详细说明。应当理解, 此处所描述的具体实施例仅仅用以解释本发明, 并

不用于限定本发明。此外,下面所描述的本发明各个实施方式中所涉及到的技术特征只要彼此之间未构成冲突就可以相互组合。

[0071] 本发明提供了一种基于强化学习的Modelica模型参数辨识方法,可以有效地辨识出Modelica模型的关键参数以提高模型的仿真性能。本发明为达到收敛速度快、辨识精度高的要求,以模型对应的物理机实测数据作为训练样本,在参数粗调阶段采用基于Q-Learning算法改进的强化学习框架,对每个待辨识参数代表的智能体进行串行的迭代训练;在微调阶段采用PSO/GA算法,确定待辨识参数的最终值。本发明具有收敛速度快、辨识精度高的特点。

[0072] 强化学习是一种基于“试错机制”的机器学习方法,起源于对动物智能领域的研究,随后发展为机器学习领域的一个重要分支。强化学习考虑的是由目标指引的智能体与未知环境之间相互作用的过程,在不断试错的过程中积累经验并不断更新动作选择策略,具有收敛速度快、易于避免局部最优、无需先验知识等优点。目前,强化学习多用于组合优化(序列决策)领域且已有成功案例,而参数辨识所属的函数优化领域还未有其应用。考虑到Modelica模型参数辨识所需的快速收敛、全局最优等要求,因此本发明提供一种基于强化学习的参数辨识方法来解决该问题。

[0073] 本发明采用的整体技术方案为:使用基于Sobol的灵敏度分析方法分析待辨识Modelica模型的参数灵敏度,根据参数的全局影响因子大小对参数进行排序,并确定参数在粗调阶段的训练顺序及离散化程度(即在初始化范围拆分的格子数);使用本发明中基于Q-Learning改进的强化学习框架,以物理机上的实测数据和模型仿真数据之间的均方误差为目标函数及环境状态,参数在可行范围内的调整为动作,当前目标函数值与历史最优/平均值之间的关系为单步奖赏,对Modelica模型的待辨识参数进行粗调并得到每个参数的最优子区间;基于上述所有参数的最优子区间,采用PSO/GA算法,迭代训练得到每个参数的最优辨识值,即为整个技术方案最终的辨识结果。

[0074] S1.基于Sobol方法的Modelica模型参数灵敏度分析,包括以下几个步骤:

[0075] (1) 导出待辨识Modelica模型的FMU(Functional Mock-up,功能模型单元,包含模型的C代码、求解器等,可以被直接调用并独立完成模型的仿真求解),设置为“Co-Simulation”(联合仿真)模式;

[0076] (2) 在N个待辨识参数的范围内分别进行蒙特卡洛采样,生成样本矩阵A、B以及 AB^i ($i=1,2,\dots,N$),将样本矩阵作为输入对FMU进行仿真求解得到Y值矩阵,基于Y值矩阵和影响指数公式得到每个待辨识参数的全局影响因子并排序。根据待辨识参数影响因子的大小确定训练顺序,影响因子越大(即越敏感)的参数在单轮训练中越先被优化;

[0077] (3) 在每个参数的搜索范围内获取一个随机值,代入FMU计算出结果作为基准值,再依次确定每个参数的最小子区间大小 G_i ($i=1,2,\dots,N$),确保第i个参数在增减 G_i 后对仿真结果有比较显著的影响。

[0078] S2.基于Q-Learning算法的强化学习框架,用于参数粗调阶段。其特征在于,构成如下:

[0079] (1) 环境状态设置为当前轮数的模型估计值 Y_{est} 与物理机实测值 Y_{mea} 的均方误差(即强化学习目标函数 $F(X)$),用于表示当前模型参数组合与实际参数组合之间的差别;

[0080] (2) 单步奖赏用于衡量当前状态所选动作的好坏,是目标函数值的表达式,有如下

两种模式:

$$[0081] \quad r = \begin{cases} 1 & \text{if } F(X_{cur}) \leq F(X_{best}) \\ -1 & \text{else} \end{cases}$$

$$[0082] \quad iir = \min(1, \max(0, (F(X_{mean}) - F(X_{best})) / (F(X_{mean}) - f(X_{cur}))))$$

[0083] r 表示奖惩值, $F(X_{mean})$ 表示参数平均值下的目标函数值, $F(X_{best})$ 表示最优参数下的目标函数值, $F(X_{cur})$ 表示当前参数下的目标函数值

[0084] (3) 动作根据参数最小子区间 G_i ($i = 1, 2, \dots, N$) 及每个参数的范围设置, 将第 i 个参数的搜索范围拆分为 $D_i = \frac{x_{i_{max}} - x_{i_{min}}}{G_i}$ 个子区间 ($x_{i_{max}}$ 为参数最大值, $x_{i_{min}}$ 为参数最小值), 选中一个子区间后, 在该区间随机获取一个值作为动作即可;

[0085] (4) 动作选择策略分为两个阶段: 选择搜索路径和确定行为。前者确定下一轮选择的行为在当前行为的左边还是右边, 选择指标为 $L_p(i, j)$, 计算公式如下:

$$[0086] \quad L_p(i, j) = \sum_{n=1}^k \lambda_1^{n-1} q_{i,n}^* \quad p = 1, 2$$

[0087] $L_p(i, j)$ 表示……, k 表示参数变换组合的数量

[0088] $q_{i,n}^*$ 为第 1 个路径上与当前行为 $a_{i,j}$ 相邻的 k 个行为中第 n 大的 Q 值, λ_1 为路径权重系数。

[0089] 获取一个 $[0, 1]$ 之间的随机数 ε_1 , 使用如下公式确定搜索路径:

$$[0090] \quad l = \begin{cases} \operatorname{argmax}_p (L_p(i, j)) & \text{if } \varepsilon_1 > 0.3 \\ \operatorname{rand}(1, 2) & \text{else} \end{cases}$$

[0091] $\operatorname{rand}(1, 2)$ 表示 1-2 区间内的随机概率分布

[0092] 确定行为则是在搜索路径上与当前行为相邻的 k 个行为中选择一个, 获取一个 $[0, 1]$ 之间的随机数 ε_2 , 使用如下公式确定行为:

$$[0093] \quad a = \begin{cases} \operatorname{argmax}_m q(i, m) \quad m = j - k, \dots, j - 1 & \text{if } \varepsilon_2 > 0.3 \\ \operatorname{rand}(j - k, j - k + 1, \dots, j - 1) & \text{else} \end{cases}$$

[0094] ε_1 和 ε_2 都是用来保证强化学习探索性的随机数。

[0095] (5) 基于 Q-Learning 算法设计 Q 值函数的更新策略, 第 i 个参数对应的 Q 值函数更新公式如下:

$$[0096] \quad Q^{r+1}(i, j) += \alpha (r + (1 - \lambda_2) \max(L_p(i, j)) + \lambda_2 \min(L_p(i, j)) - Q^r(i, j))$$

[0097] 式中 α 为控制学习速率的超参数, r 为单步奖赏, λ_2 为控制更新幅度的超参数, 其值越小, Q 值表更新幅度越大。

[0098] S3. 基于强化学习框架的 Modelica 模型参数辨识算法, 用于参数粗调阶段, 具体步骤如下:

[0099] (1) 随机初始化所有待辨识参数, 并代入 FMU 中计算, 与实测数据相比得到均方误差作为最优目标函数的初始值, 并记录下来;

[0100] (2) 进入一个训练周期,在该周期内多个智能体执行串行学习,即每个参数轮流优化,单个智能体的学习过程称之为“一回合”;

[0101] (3) 进入一个回合,在该参数的搜索范围内随机选取一个动作 $a_{\text{rand}}(i)$,固定其它参数,将此参数应用到FMU中,得到当前回合的目标函数值 $F(X_{\text{cur}})$ 以及当前奖惩值 r ,根据Q值函数的更新策略更新当前智能体的Q值表。若 $F(X_{\text{cur}}) \leq F(X_{\text{best}})$,认为当前参数有搜索价值,转(4),否则进入下一回合;

[0102] (4) 进入一个搜索周期,在该周期内执行 $I_{e_{\text{max}}}$ 次迭代,每轮迭代根据动作选择策略选择一个动作 $a_{\text{iter}}(i)$,固定其它参数,将此参数应用到FMU中,得到当前回合的目标函数值 $F(X_{\text{cur}})$ 以及当前奖惩值 r ,更新当前智能体的Q值表;

[0103] (5) 一个智能体的搜索周期结束后,进入下一个智能体的“一回合”,直到所有智能体都经历一个回合,一个训练周期结束,若当前周期数小于 $I_{s_{\text{max}}}$,转(2)进入下一个训练周期,否则训练结束。

[0104] S4. 基于启发式优化算法PSO的参数微调方案,构成如下:

[0105] (1) 根据强化学习算法的运行结果,确定每个待辨识参数的微调范围。若强化学习得到的最优子区间为 $[x_{i_{\text{opt}}} - \mu G_i, x_{i_{\text{opt}}} + (\mu + 1)G_i]$, $i=1, 2, \dots, N$,考虑到强化学习的辨识误差,将微调范围设置为 $[x_{i_{\text{opt}}}, x_{i_{\text{opt}}} + G_i]$, μ 值根据粒子群算法的搜索能力来设置, $x_{i_{\text{opt}}}$ 是第 i 个参数经过强化学习辨识后的最优值;

[0106] (2) 初始化一个N维空间以及粒子群,创建每个粒子的速度向量 v_i 、位置向量 x_i 、历史最优位置向量 p_i ,以及整个粒子群的历史最优位置向量 p_g ;

[0107] (3) 以模型预测值与实测值的均方误差建立适应度函数 $F(X)$,模型预测值由粒子的当前位置代入FMU中仿真求解得到。整个粒子群的搜索目标就是使 $F(X)$ 最小;

[0108] (4) 开始迭代搜索,在一轮迭代中,计算每个粒子的适应度值,更新单个粒子的 p_i 值及种群的 p_g 值,再根据下式更新每个粒子的速度和位置向量,

$$[0109] \quad v_i^{t+1} = \omega v_i^t + c_1 r_1 (p_i^t - x_i^t) + c_2 r_2 (p_g^t - x_i^t)$$

$$[0110] \quad 10. x_i^{t+1} = x_i^t + v_i^{t+1}$$

[0111] 式(1)中 ω 为惯性权重,默认为0.6, c_1 、 c_2 为学习因子,通常 $c_1=c_2=2$, r_1 、 r_2 为[0, 1]之间的随机数;

[0112] 达到最大迭代次数后,粒子群的历史最优位置向量 p_g 即为参数微调的最终结果,亦是整个参数辨识流程的最终结果。

[0113] 本发明实施例以附图2所示的基于Modelica创建的动力传动系统模型(数控机床单轴伺服进给系统)为例对上述方法进行详细介绍,该系统模拟的是正弦信号驱动齿轮减速器,带动负载运转,系统中间还有两个用于模拟惯性的组件以及影响整个系统的弹簧和减震器。整个系统模型的输入为正弦信号,输出为组件inertia3的绝对角速度,待辨识的参数及其附属信息如表1所示。

参数名称	标准值	搜索范围	单位	描述
inertia2.J	2.0	[0.0, 5.0]	$kg \cdot m^2$	减速器负载转动惯量
[0114] spring.c	10000.0	[0.0, 50000.0]	$N \cdot m/rad$	扭簧刚度系数
damping	10.0	[0.0, 20.0]	$N \cdot ms/rad$	阻尼器阻尼系数
Jmotor	0.1	[0.0, 1.0]	$kg \cdot m^2$	电机输出轴转动惯量

[0115] 表1

[0116] 应当解释,此实施方案是为详解本发明技术方案并证明其有效性及可行性,而非真实工程案例,所以此实施方案中不以真实物理机的实测数据作为样本集,而是以标准模型的输出作为样本集,再随机初始化所有待辨识参数,进而利用本发明的技术方案进行参数辨识使得待定参数尽可能地靠近参数标准值,即完成辨识。完整的辨识流程,包括下列步骤:

[0117] 1.导出该传动系统Modelica模型的FMU(功能模型单元),设置好仿真求解的步长、开始时间、结束时间及求解器等仿真参数;

[0118] 2.将FMU中的待辨识参数设置为标准值,仿真求解得到输出作为样本集。输出曲线上以固定周期取样1000个点作为此次参数辨识的数据样本集,后续的强化学习目标函数和PSO算法适应度函数都为这1000个点的纵坐标和模型预测值之间的均方误差;

[0119] 3.对Modelica模型进行参数灵敏度分析。如上所述,4个待辨识的参数分别为减速器负载转动惯量inertia2.J/扭簧刚度系数spring.c/阻尼器阻尼系数damping/和电机输出轴转动惯量Jmotor,灵敏度分析步骤如下:

[0120] (1) 设定样本:维数为4(即4个输入变量),个数为10000(即10000个输入样本),基于每个参数的范围生成一个 10000×8 的Sobol矩阵,并按照Sobol法对其进行处理得到矩阵A、B以及 $A_B^{(i)}$ ($i=1,2,\dots,8$);

[0121] (2) 将上述样本矩阵输入到FMU中仿真求解得到输出,按照Sobol方法将输出整理为Y值矩阵 Y_A 、 Y_B 和 Y_{AB} ,分别用 $f(A)$ 、 $f(B)$ 和 $f(A_B^i)$ 表示;

[0122] (3) 根据以下公式求出每个参数的全局影响指数 S_{Ti} :

$$[0123] \quad S_{Ti} = \frac{E_{X_{-i}} \left(Var_{X_i} (Y | X_{-i}) \right)}{Var(Y)} = \frac{1}{2N} \sum_{j=1}^N \left(f(A)_j - f(A_B^i)_j \right)^2$$

[0124] 结果如附图3所示,就敏感性而言, $Jmotor > spring.c > damping > inertia2.J$;

[0125] (4) 依步骤(3)得到的敏感度排序确定学习顺序为Jmotor、spring.c、damping、inertia2.J,再确定每个参数对应的 G_i ,进而确定每个参数对应的拆分格子数分别为200/1000/100/30。

[0126] 4.对Modelica模型进行参数粗调。使用基于Q-Learning的强化学习,需要为每个参数设定一个搜索范围,本案例中的的4个待辨识参数,其搜索范围由查阅手册或经验得出,在表1亦有体现。同时设置好算法中的参数,令强化学习算法学习率 $\alpha=0.7$, $\lambda_1=0.5$ 为路径权重系数, $\lambda_2=0.25$ 为控制更新幅度的超参数, $k=4$ 为参数变换组合的数量,然后使用如下步骤对参数进行初步辨识:

[0127] (1) 随机初始化所有待辨识参数,代入基于Modelica动力传动系统模型中计算,与数据样本集相比得到均方误差作为最优目标函数的初始值,并记录下来;

[0128] (2) 进入一个训练周期,在该周期内多个智能体执行串行学习,即每个参数轮流优化,单个智能体的学习过程称之为“一回合”;

[0129] (3) 进入一个回合,在该参数的搜索范围内随机选取一个动作 $a_{rand}(i)$,固定其它参数,将此参数应用到FMU中,得到当前回合的目标函数值 $F(X_{cur})$,使用公式

$$r = \begin{cases} 1 & \text{if } F(X_{cur}) \leq F(X_{best}) \\ -1 & \text{else} \end{cases}$$

得到单步奖赏 r ,使用公式 $Q^{r+1}(i, j) = \alpha(r + (1 - \lambda_2) \max(L_p(i, j)) + \lambda_2 \min(L_p(i, j)) - Q^r(i, j))$ 更新当前智能体的Q值表。若 $F(X_{cur}) \leq F(X_{best})$,认为当前参数有搜索价值,转(4),否则进入下一回合;

[0130] (4) 进入一个搜索周期,在该周期内执行 $I_{e_{max}}$ 次迭代,每轮迭代根据动作选择策略选择一个动作 $a_{iter}(i)$,固定其它参数,将此参数应用到FMU中,得到当前回合的目标函数值 $F(X_{cur})$ 以及当前奖惩值 r ,更新当前智能体的Q值表;

[0131] (5) 一个智能体的搜索周期结束后,进入下一个智能体的“一回合”,直到所有智能体都经历一个回合,一个训练周期结束,若当前周期数小于 $I_{s_{max}}$,转(2)进入下一个训练周期,否则训练结束。

[0132] 此阶段获得的初步辨识结果为每个参数的最优子区间,如表2所示,在此基础上,还需确定精确解;

参数名称	初步辨识子区间
inertia2.J	[1.892, 2.201]
spring.c	[10050.000, 10100.000]
damping	[10.400, 10.600]
Jmotor	[0.095, 0.100]

[0134] 表2

[0135] 5. 对Modelica模型进行参数微调。步骤4只是得到了每个待辨识参数所在的最优子区间,还需要在此区间内使用PSO算法确定最终辨识结果,具体步骤如下:

[0136] (1) 步骤4得到的最优子区间为 $[x_{i_{min}}, x_{i_{min}} + G_i]$ $i = 1, 2, \dots, N$,考虑到强化学习的辨识误差,将微调范围设置为 $[x_{i_{min}} - \mu G_i, x_{i_{min}} + (\mu + 1)G_i]$,基于粒子群算法的搜索能力,将 μ 值都设为2;

[0137] (2) 初始化一个4维空间以及粒子群,创建每个粒子的速度向量 v_i 、位置向量 x_i 、历史最优位置向量 p_i ,以及整个粒子群的历史最优位置向量 p_g ;

[0138] (3) 以模型预测值与数据样本集的均方误差建立适应度函数 $F(X)$,模型预测值由粒子的当前位置代入FMU中仿真求解得到。整个粒子群的搜索目标就是使 $F(X)$ 最小;

[0139] (4) 开始迭代搜索,在一轮迭代中,计算每个粒子的适应度值,更新单个粒子的 p_i 值及种群的 p_g 值,再根据下式更新每个粒子的速度和位置向量,

$$[0140] \quad v_i^{t+1} = \omega v_i^t + c_1 r_1 (p_i^t - x_i^t) + c_2 r_2 (p_g^t - x_i^t)$$

$$[0141] \quad x_i^{t+1} = x_i^t + v_i^{t+1}$$

[0142] 式(1)中 ω 为惯性权重,默认为0.6, c_1 、 c_2 为学习因子,通常 $c_1=c_2=2$, r_1 、 r_2 为[0,1]之间的随机数;

[0143] (5)达到最大迭代次数后,粒子群的历史最优位置向量 p_g 即为参数微调的最终结果,亦是整个参数辨识流程的最终结果。

[0144] 最终辨识结果与参数标准值的对比如表3,参数辨识结果与标准值误差均小于5%,满足工程使用要求。

	参数名称	最终辨识结果	标准值
	inertia2.J	2.010850	2.0
[0145]	spring.c	10000.497589	10000.0
	damping	10.53364	10.0
	Jmotor	0.095948	0.1

[0146] 表3

[0147] 针对数控机床进给系统建立基于Modelica动力传递模型,并采用本发明方法进行参数辨识,在相同输入条件下,物理样机输出结果与模型仿真结果如图4所示,可以看出,模型仿真曲线变换趋势、数值大小与物理样机输出曲线基本一致,说明了本发明参数辨识方法的正确性与有效性。

[0148] 本领域的技术人员容易理解,以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等,均应包含在本发明的保护范围之内。

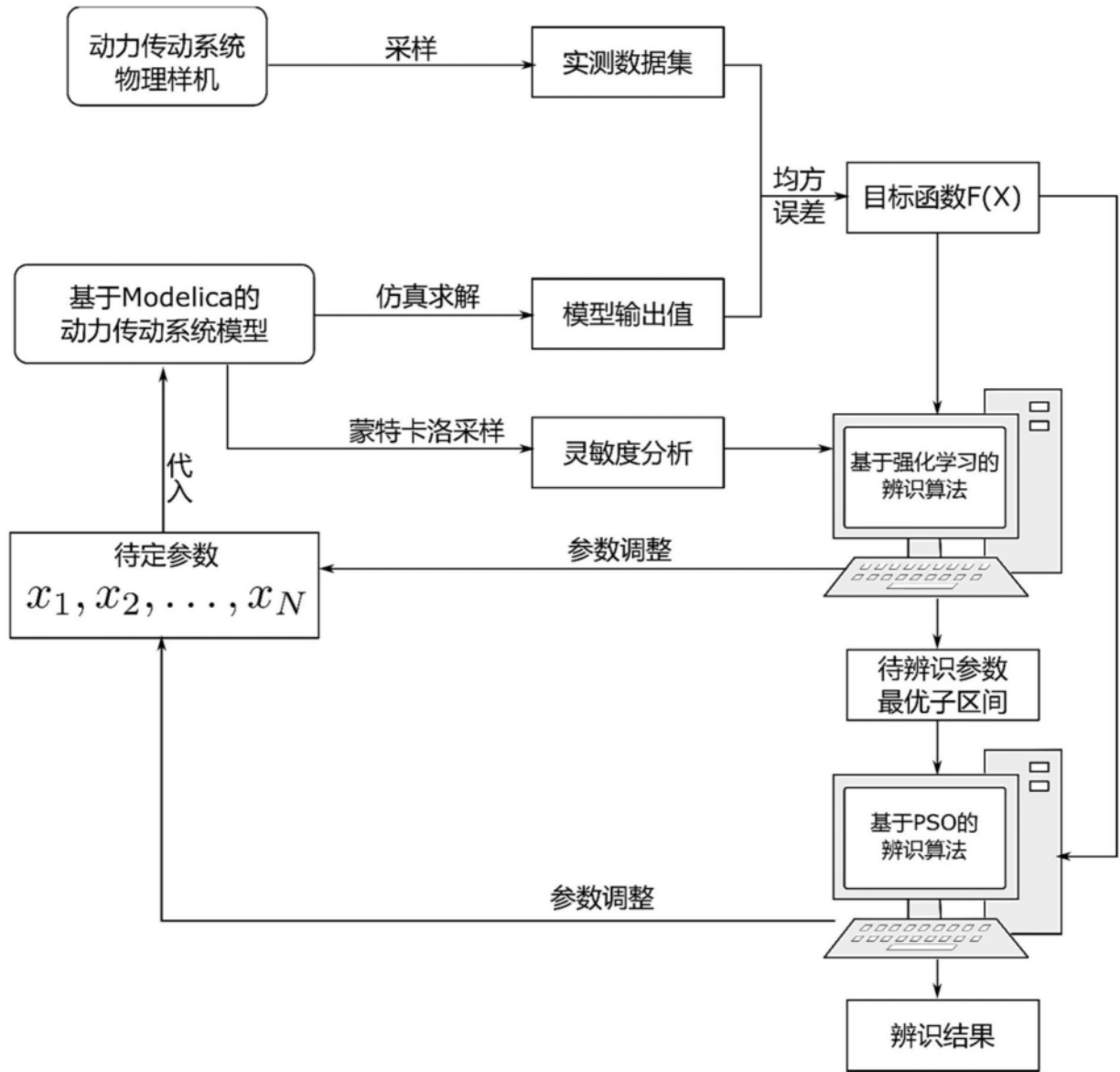


图1

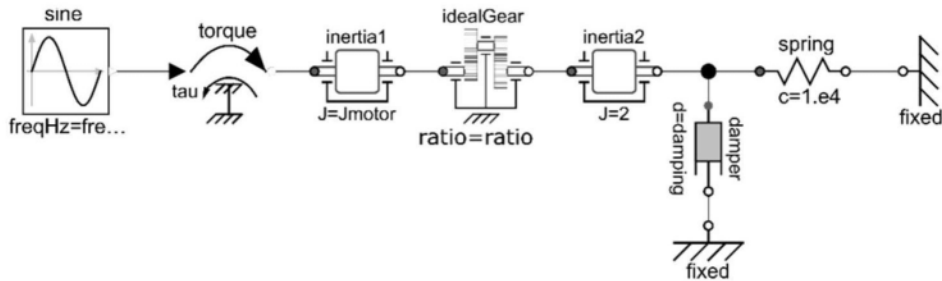


图2

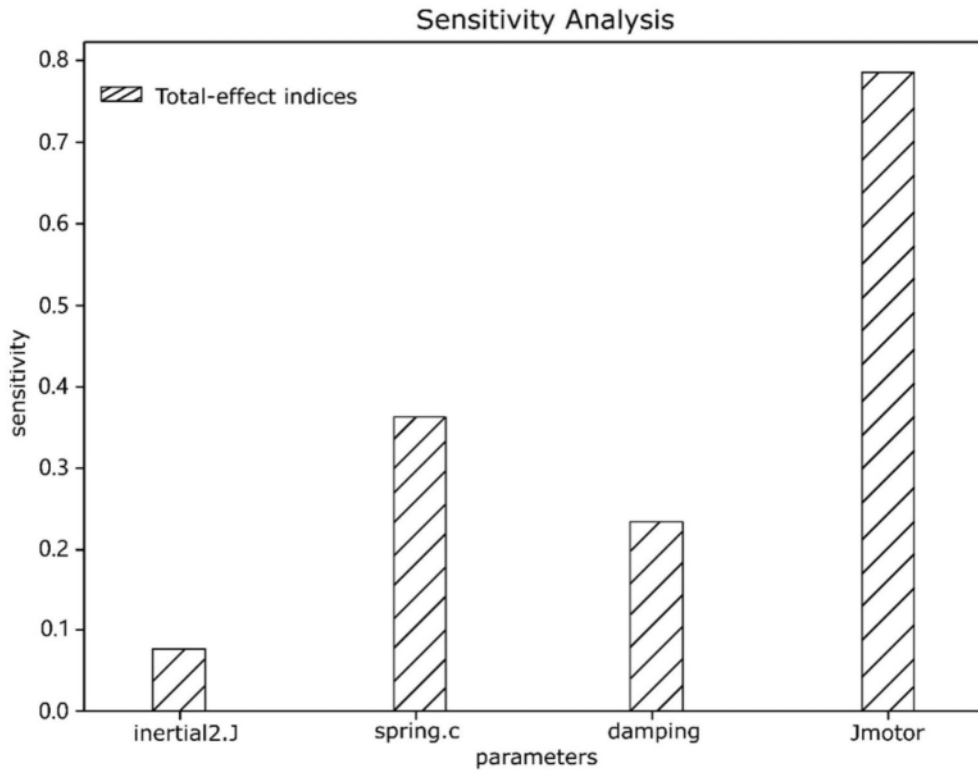


图3

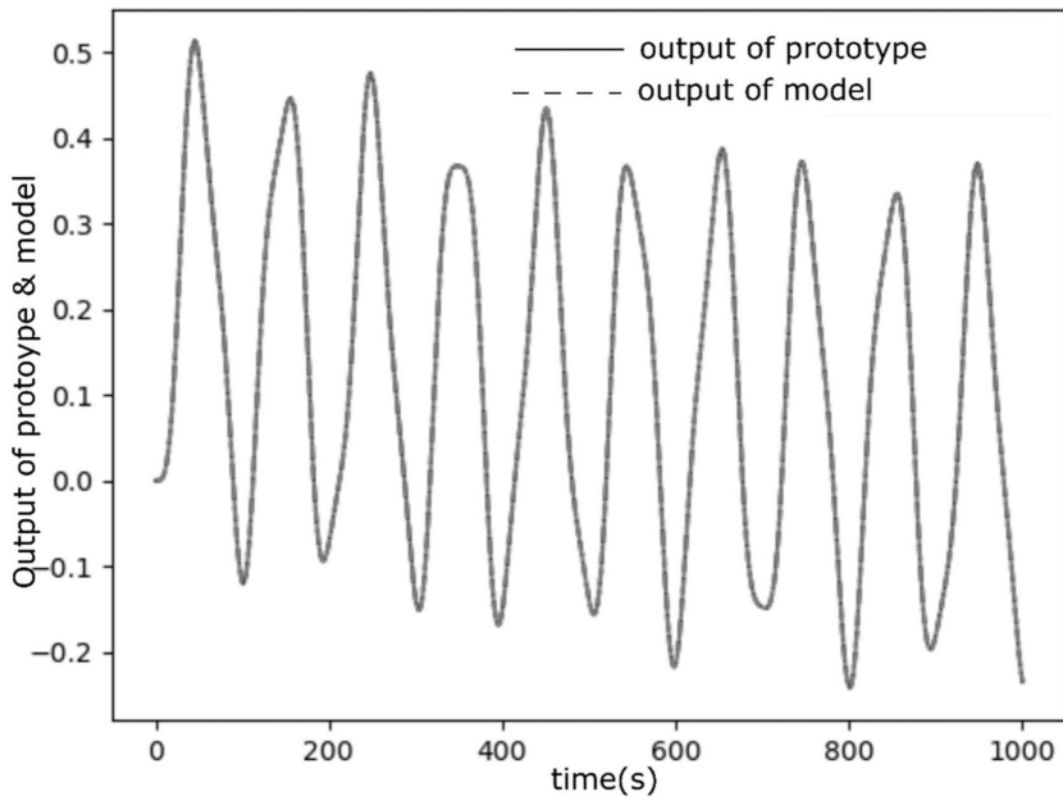


图4