

DOI: 10.19344/j.cnki.issn1671-5276.2025.02.046

# 制造过程中协作机器人意图理解研究

薛锦辉, 苗鸿宾, 贾程翔, 张灿

(中北大学 机械工程学院, 山西 太原 030051)

**摘要:** 协作机器人作为与人共同工作的新一代机器人, 准确理解操作员的意图并作出反应尤为重要。现有的协作机器人意图预测都是与传统机器人一样, 通过识别操作员的语言、动作来进行意图识别。但这种方法并没有发挥协作机器人的优点, 即与人更近距离地接触。该研究在语言、动作的基础上, 增加通过触碰来让协作机器人感知操作员的意图, 设计多层前馈神经网络, 利用 LM 与 GA 算法结合对其进行训练并得到参数最优解。训练完成以后进行对比实验以及意图预测实验。实验结果表明: 训练方均差最终达到 0.098 6, 预测准确度达到 92.3%。相较于其他神经网络具有更优良的预测性能。

**关键词:** 协作机器人; 人机协作; 意图理解; 神经网络

**中图分类号:** TP242.2 **文献标志码:** A **文章编号:** 1671-5276(2025)02-0238-05

## Research on Intent Understanding of Collaborative Robots in Manufacturing Process

XUE Jinhui, MIAO Hongbin, JIA Chengxiang, ZHANG Can

(School of Mechanical Engineering, North University of China, Taiyuan 030051, China)

**Abstract:** As a new generation robots co-working with humans, it is most important for collaborative robots to accurately understand the operator's intentions and respond accordingly. The existing collaborative robot intention prediction, similar to traditional robots in recognizing the operator's language and actions for intention recognition, fails to exploit the advantages of collaborative robots, i. e. having closer contact with humans. This study based on language and action, adds touching to make collaborative robots perceive the operator's intention, designs a multi-layer feedforward neural network, and trains it by LM and GA algorithms obtaining the optimal parameter solution. Comparative experiments and intention prediction experiments are conducted after training completion. The experimental results show that the training mean square error ultimately reaches 0.098 6, and the prediction accuracy is 92.3%, which has better predictive performance compared to other neural networks.

**Keywords:** collaborative robot; human-robot cooperation; intentional understanding; neural network

### 0 引言

人机协作(human-robot cooperation, HRC)是探索如何将机器人的精准、迅速和强大能力与人类的丰富经验、知识、学习能力、适应性和理解控制策略能力相结合的创新领域。随着自然人机交互技术的进步,传统的以人或机器人为中心的设计模式的局限性逐渐显现。无论是哪种协作型机器人,理解操作者的意图对于确保人机协作的安全至关重要<sup>[1]</sup>。理想的协作机器人设计应根据任务需求灵活调整行为,协助人类更高效地适应环境变化,满足人类需求。在日常生活中,人们通过非言语行为如眼神、手势、动作等来表达意图并与他人建立默契。协作机器人的一个显著特点是在操作过程中与操作者保持近距离接触,这为操作者提供了更多与机器人互动的机会,例如在

确保安全的前提下,通过触碰机器人的不同部位来传递信号,从而拓展了人机协作的潜力。

机器人系统中应该有一种高效的操作人员意图检测方法。虽然目前基于意图理解的人机协作已经出现了很多种算法<sup>[2-6]</sup>,但是这些算法所包含的输入都比较单一,无法应对复杂的人机协作环境。为此,本文提出一种基于多模态的人机协作意图理解神经网络。

本研究首先对3种模态特征进行数据收集和分析,再利用遗传算法与LM算法结合来训练MLFFNN。LM算法可以有效避免过拟合,而遗传算法是解决黑盒优化问题最高效可靠的方法<sup>[7-8]</sup>。因此,采用LM算法优化神经网络,遗传算法助力网络结构调整。在网络训练过程中,采用早期停止策略以减少或避免训练过程中发生过拟合现象。为了验证MLFFNN具有更好的预测

**基金项目:** 中北大学研究生科技立项项目(20231912)

**第一作者简介:** 薛锦辉(2000—),男,山西临汾人,硕士研究生,研究方向为工业机器人协作, xjh8237@163.com。

性能,将其训练过程中的预测效果与其他 3 种网络进行比较。在训练完成后进行意图预测实验,进一步验证模型的预测性能。

## 1 神经网络架构设计

### 1.1 用户多模态特征提取

多模态指的是多种模态的信息<sup>[9]</sup>,包括文本、图像、视频、音频等。用户的行为特征被定义为在人机合作过程中做出的自主、自然的动作。

本文中设置了 3 种不同的模态,以 3 种形式实现用户的输入,分别是:语音( $X_g$ )、手势( $X_v$ )和力矩( $X_f$ )。这些数据由连接工业机器人的传感器

和摄像头进行获取,之后利用稳定、高效的数据处理方法来处理。本研究在语音模态方面,将用户的语音转换为文本,结合关键词识别将其分为 5 类。这里定义的语言模态并非指的是特定词汇,而是某些具有相同意思的词汇,例如“向上移动”和“往上移动”二者所代表的意图相同,只是在机器人获得输入时给予的分数或有差别。类别编号与关键词如表 1 所示。

肢体手势模态方面,主要作用为辅助语言模块的输入,为机器人的动作提供方位指引功能。例如通过语言模块输入令机器人运动、通过识别人的肢体手势指定机器人运动的方向等。手势动作与语音模态配合情况如表 1 所示。

表 1 模态的输入及指令

编号	关键词及手势	代表的指令	力矩输入
1	向(某方向)移动	机器人向(某方向)移动	给予(某方向)的力
2	拿起(指向)工具	拿出工具架上的工具	竖直向下及向后的力
3	把工具给我	将工具给予操作员	—
4	将工具放回(工作区)	将工具放回工具区域	竖直向下及向左的力
5	停止	停止运动	同时给予向上及向下的力

力矩模态方面,通过给机器人添加额外的力传感器模块,在机器人工作期间可以迅速有效地感应到来自外部的力,再利用机器人关节动力学方程,就可以得到力矩模块的输入。根据力的作用方向不同分为 4 类。

外部力矩的输入与机器人运动学息息相关, $n$ -连杆机器人的动力学方程如下:

$$M(\theta)\ddot{\theta} + C(\theta, \dot{\theta})\dot{\theta} + G(\theta) = \tau + \tau_{ext} \quad (1)$$

式中:向量  $\theta, \dot{\theta}, \ddot{\theta}$  分别表示机器人  $n$  个关节的位置、速度、加速度;  $M(\theta) \in R^{n \times n}$  表示惯性矩阵;  $C(\theta, \dot{\theta}) \in R^{n \times n}$  表示包含科里奥利力和离心矩阵;  $G(\theta) \in R^n$  是  $n$  连杆机器人的重力矢量。机器人各关节之间的动态耦合就表现在重力以及科里奥利力上。因为重力的存在,其作用于机器人的任何一根连杆上,这会对执行器转矩  $\tau \in R^{n \times n}$  产生影响。

根据式(1)可以推出机器人的逆运动学公式进而得到机器人受到的外部力矩。

$$\tau_{ext} = f(\theta, \dot{\theta}, \ddot{\theta}, \tau) = M(\theta)\ddot{\theta} + C(\theta, \dot{\theta})\dot{\theta} + G(\theta) - \tau \quad (2)$$

### 1.2 神经网络属性

#### 1) MLFFNN

本文采用的是多层前反馈神经网络(MLFFNN)。MLFFNN 由输入层、隐藏层、输出层 3 层网络组成,相较于其他神经网络结构更为简单。但隐藏层的数量不是固定的,可以增加到 2 层或 3 层甚至更多。增加隐藏层的数量可以提升神经网络的函数拟合能力以及泛化能力,即模型对未知数据的预测能力,但同时计算成本以及过拟合风险也随之提高。MLFFNN 虽然简单实用,但其存在训练周期较长的缺点,较其他网络需要大量的输入来进行训练,才能得到最好的参数。

图 1 所示为具有 3 个隐藏层的 MLFFNN。将输入层记为第 0 层,第 1 个隐藏层记为第 1 层,输出层记作第 4 层。

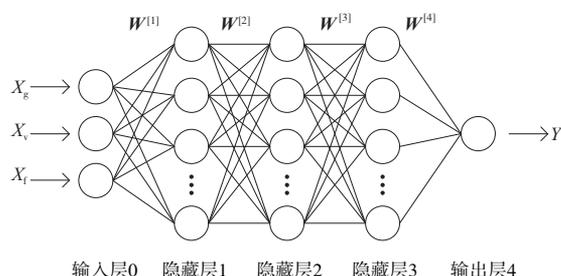


图 1 具有 3 个隐藏层的 MLFFNN 的结构

在本研究中,神经网络隐藏层的激活函数选取的是双曲正切函数,则具有 3 个隐藏层的 MLFFNN 的输出可以由下式得出:

$$Y = \mathbf{W}^{[4]} \mathbf{a}^{[3]} + \mathbf{b}^{[4]} \quad (3)$$

$$\mathbf{a}^{[n]} = \tanh(\mathbf{W}^{[n]} \mathbf{a}^{[n-1]} + \mathbf{b}^{[n]}) \quad (4)$$

式中: $Y$  为网络的标准化输出值; $\mathbf{a}^{[n]}$  为第  $n$  层的输入向量,其中  $\mathbf{a}^{[0]}$  由语言、姿态、力矩 3 个输入组成; $\mathbf{W}^{[n]}$  为对应隐藏层的权重矩阵; $\mathbf{b}^{[n]}$  为对应偏置向量。

## 2) ANN

为体现出本文设计的 3 个隐藏层的 MLFFNN 的预测能力,引入同样常用于数据预测的人工神经网络(ANN)做对比。与 MLFFNN 类似,本次用作对比的 ANN 同样由输入层、隐藏层和输出层 3 部分组成,其结构如图 2 所示。与 MLFFNN 不同,ANN 只有一个隐藏层,同样使用双曲正切函数作为隐藏层的激活函数,则 ANN 输出结果的计算式如下:

$$Y' = \mathbf{W}^{[2]} \tanh(\mathbf{W}^{[1]} \mathbf{Z} + \mathbf{b}^{[1]}) + \mathbf{b}^{[2]} \quad (5)$$

式中: $\mathbf{Z}$  为 ANN 的标准化输入向量; $Y'$  为 ANN 的标准化输出值。 $\mathbf{W}^{[1]}$  和  $\mathbf{W}^{[2]}$  分别为输入层与隐藏层之间和隐藏层与输出层之间的权重矩阵, $\mathbf{b}^{[1]}$  和  $\mathbf{b}^{[2]}$  则为与之对应的偏置向量。

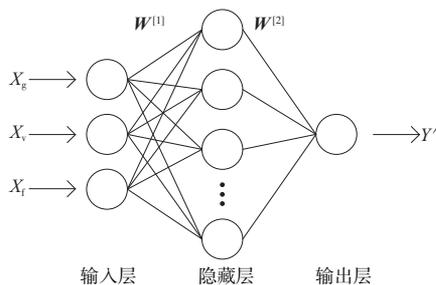


图 2 ANN 结构图

## 2 神经网络训练

本研究采用 LM-GA 算法对 MLFFNN 进行训练。

### 2.1 LM 算法

采用 Levenberg-Marquardt (LM) 算法对设计的神经网络进行训练的优点在于:在面对大量的输入时,其具有很快收敛速度的同时又保证了精准性。这是传统 BP 算法和牛顿法所不具备的。这就解决了前文说到 MLFFNN 需要用大量数据集进行训练导致训练周期长的问题。

利用 LM 算法调整后的权值由下式得到:

$$X_{k+1} = X_k - [\mathbf{H} + \lambda \mathbf{I}]^{-1} \mathbf{g} \quad (6)$$

式中: $\mathbf{H}$  为一个多维向量的矩阵; $\lambda$  为步长; $\mathbf{I}$  为与  $\mathbf{H}$  相同维数的单位矩阵; $k$  为一个正则化或加载参数,它的作用是令矩阵  $[\mathbf{H} + \lambda \mathbf{I}]$  在整个计算过程中保持正定; $\mathbf{g}$  为二阶函数的梯度向量。

在神经网络中,学习率和隐藏层神经元数量是容易修改的超参数。恰当的学习率与神经元数量搭配对 MLFFNN 性能至关重要。理想学习率能提升训练效率,避免局部极小值停滞,少量神经元可能无法应对本研究的预测问题,而过多则可能导致过拟合。同时,最优学习率会随神经元数量变化而变化。因此寻找合适的超参数是一件很费时费力的事情。

### 2.2 遗传算法

遗传算法(genetic algorithm, GA)是一种受自然选择启发的进化算法,适用于无梯度情况下的局部最小值克服。本研究利用 GA 寻找两个变量(学习率及隐藏层神经元数量)的超参数集。通过多重 hold-out 方法,得到 30 个独立模型,并依据测试数据集中的 MSE 评估预测精度。适应度函数为 30 个独立模型平均预测精度。GA 的流程如下:首先随机生成超参数集,计算当前代成员适应度值,并根据得分进行选择、突变和交叉操作。重复此过程,直至满足停止准则。设置种群大小、突变率和交叉率分别为 60、0.01 和 0.8。学习率搜索范围分别为 0~0.5。为获得稳定解,独立运行 GA 10 次,选取最优解作为最终超参数集。遗传算法步骤如下。

1) 初始化种群并编码。用染色体表示神经网络结构,包括结构基因和权重基因。结构基因采用二进制编码,表示隐层节点结构;权重基因  $w_n(x+y)$  表示连接权值和阈值,采用实数编码。

本研究中,初始种群大小设置为 60,初始隐藏节点设为 30。

2) 进行适应度评价,公式如下:

$$\begin{cases} E = \sum (S-P)^2 \\ F = 1/E \end{cases} \quad (7)$$

式中: $P$  表示神经网络经过 LM 训练后的预测值; $S$  表示实际测量值; $E$  表示误差平方和 (SSE); $F$  表示适应度函数。进化过程总体朝着适应度函数  $F$  较高的方向发展,误差逐渐减小,结构逐渐简化。

3) 确定终止进化条件的两种方式,分别为迭

代次数达到 500 和方均差 (MSE) 小于 0.01 且符合早期停止策略,二者满足其一即可终止进化。

4) 遗传算子包括选择、交叉和变异算子。本研究采用特定策略。选择算子适用轮盘赌策略;根据染色体编码结构的不同,采用交叉策略的选择,结构基因用单点交叉,权重基因用算术交叉;变异算子的选择采用结构基因选择基本位变异算子,而权重基因选择非均匀变异算子。初始交叉率和变异率分别为 0.8 和 0.01,可进行自适应调整。

### 2.3 CEP 策略

CEP 策略 (complex event processing) 全称为复杂事件处理框架。CEP 策略的思想是集成分类处理,其输入是一些简单的事件流,接收到输入的事件流后,通过内部核心的判别逻辑,根据事件特点,路由到相应的处理分支进行处理。CEP 是一个不太准确的预测器,它包括经验函数、非昂贵的实验测量、小规模模拟和计算。CEP 虽然不太准确,但可以提供有关目标属性的定性信息。加入 CEP 的集成可能有助于提高机器学习模型的预测精度。

### 2.4 早期停止策略

本研究采用早期停止策略 (early stopping),以防止过拟合。此方法将可用的数据集随机分为 3 个子集:训练、验证和测试数据集。训练数据集用于计算梯度并更新权重和偏差,而验证数据集则用于判断训练过程何时停止。通常在训练初期,训练误差和验证误差都会减小。然而,当网络开始过拟合时验证数据集中的误差会开始上升。一旦验证误差增加到预定的 epoch 数,训练过程便停止。因此认为早期停止策略是一种有效的过拟合避免方法。

## 3 实验验证

### 3.1 模型测试实验

经过 LM-GA 算法的训练,本文得到了经过计算优化后的 MLFFNN 相关超参数,同时在 GA 优化的基础上,将隐藏层的数量从 2 层增加到 4 层,以此来验证具有 3 个隐藏层的 MLFFNN 较其他网络具有更加优秀的预测性能。3 种神经网络的相关超参数如表 2 所示。表中  $N[n]$  代表第  $n$  个隐藏层所含的神经元的数量。ANN 的计算成本相对 MLFFNN 而言较小,因此本次研究在结构为 4-(25)-1 的神经网络中进行。

表 2 训练得出的神经网络参数

参数	ML-3	ML-2	ML-4	ANN
$N[1]$	35	30	35	25
$N[2]$	30	30	30	—
$N[3]$	24	—	24	—
$N[4]$	—	—	20	—
学习率	0.304 9	0.312 6	0.302 1	0.305 4

验证设计的神经网络标准直观地表现在机器人做出的反应与操作人员的主观意愿是否一致,数据上主要体现在神经网络训练得到的 MSE 值大小以及神经网络计算出的分数与操作员实际意图的误差上。在本研究中采用了多重 hold-out 方法,为每个神经网络架构获得了 30 个独立模型,并对 4 种神经网络进行了训练。每个神经网络的学习率也经过了调整,已获得各个网络最佳的预测性能。

图 3 绘制了 4 种神经网络在训练以及测试过程中的最小 MSE 值。从图 3 中可以直观地看出,具有 3 个隐藏层的 MLFFNN 无论在训练还是测试中都具有更低的 MSE 值,也就是更好的预测性能。需要注意,本次所使用的神经网络均属于浅层网络,在目前的数据集中,其预测精度并未达到最优精度。这是因为 CEP 的集成增加了预测问题的复杂性,需要更多的神经元来准确预测。然而,增加隐藏层数会导致计算成本上升,同时预测精度和稳定性下降,也就是过拟合,需注意避免。因此,本文选择 3-(35-30-24)-1 的架构。

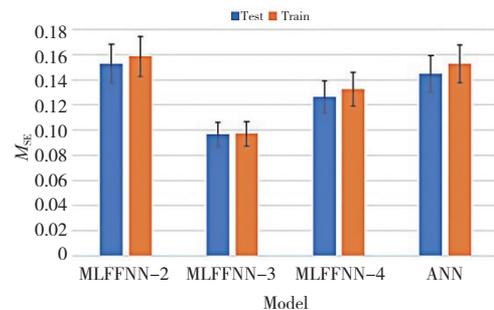


图 3 不同架构神经网络性能

为了进一步分析具有 3 个隐藏层的 MLFFNN,给出了其在训练、验证、测试过程中 MSE 值的变化过程,如图 4 所示。由于使用了提前停止策略,训练过程在 epoch252 处停止,在 epoch246 之后的 6 次 epoch, MSE 值连续增加。因此,将 epoch246 的权重和偏差作为终端参数返回。

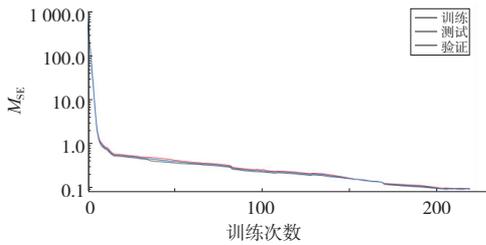


图 4 训练、验证和测试数据集中的 MSE 值

评价神经网络优劣的一个重要标准是其对非训练与测试时所用数据集的准确度。因此,本文设计了一组测试用的数据并进行实验,所得到的误差如图 5 所示。根据图 5 中表示出的误差来看,在面对全新的数据集时,本神经网络依然具有很高的预测精度,符合设计预期。

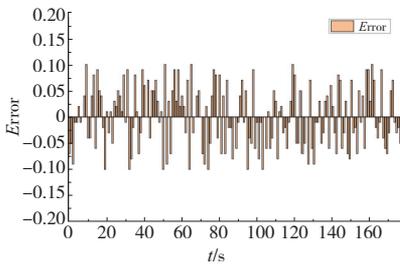


图 5 未见数据集上的误差

### 3.2 意图预测实验

人机协作意图预测的核心是对操作人员的预计动作进行评估分类,从这方面入手,引入准确率、召回率、F1 分数这 3 项指标对设计的模型进行试验评估。4 种模型参数均是评估模型准确率的常用参数。从数据集中选择部分作为实验数据集进行实验验证。

将表 1 中列出的 5 种动作中的方向参数补全之后,从上至下分别以 Act-1、Act-2、Act-3、Act-4 以及 Act-5 命名,之后进行实验。选择实验中预测为 5 种动作的数据各 50 个,其预测效果的混淆矩阵如图 6 所示。

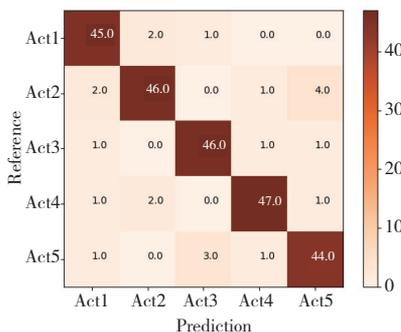


图 6 5 种动作的混淆矩阵

从给出的混淆矩阵中可以看出,5 种动作的预测效果相当优秀,预测成功次数都在 45 次左右,证明了本文设计的 3 个隐藏层 MLFFNN 模型在人机协作意图预测中具有优良的预测性能。之后选取 Act-1 为实验对象进行对比实验,得到不同网络模型在测试集中的检测结果如表 3 所示。

表 3 不同网络的检测结果

模型	动作	指标		
		准确率	召回率	F1 值
ANN	Act-1	0.870	0.883	0.835
ML-2	Act-1	0.634	0.652	0.659
ML-4	Act-1	0.865	0.858	0.853
	Act-1	0.923	0.931	0.927
ML-3	Act-2	0.927	0.929	0.920
	Act-3	0.916	0.921	0.918
	Act-4	0.918	0.922	0.917
	Act-5	0.926	0.932	0.922

在表 3 中,每行代表了 1 个模型的性能指标。从表 3 中可以看到,在 Act-1 的动作预测结果对比中,本文设计的 MLFFNN-3 网络模型在 3 项指标上都具有更高的分数值,3 项指标都在 0.9 以上,在准确率方面,较其他 3 种模型中效果最好的 ANN,也有 5 个百分点的提升。同时从表 3 中可以看出,MLFFNN-3 网络在其他 4 种动作上也有着同样出色的表现。这证明了所设计的模型能够正确地动作进行预测(高准确率),同时还能识别出大多数具有某种意图的样本(高召回率)。模型的 F1 值也证明了在多数预测任务中的良好性能。

## 4 结语

在本研究中,设计了一种基于多模态输入的 MLFFNN 来预测人机协作中操作人员的意图。采用 CEP 策略有效提高了 MLFFNN 的泛化能力。同时设置 LM-GA 算法优化的超参数,让具有 3 个隐藏层的 MLFFNN 与其他几种网络相比具有更好的预测精度以及更好的泛化能力。实验结果表明,在优良的超参数和正确的训练策略基础上,MLFFNN 使用相对较小的参数集也可以获得较高的预测精度,同时有望进一步提高。训练完成后进行人机协作意图预测实验,进一步验证了模型在实际应用中同样具有很高的准确率。

(下转第 247 页)

Lissajou 图形。Lissajou 图中的曲线越接近圆形,表明该方法的信号补偿效果越好,不同方法的 Lissajou 图形如图 5 所示。

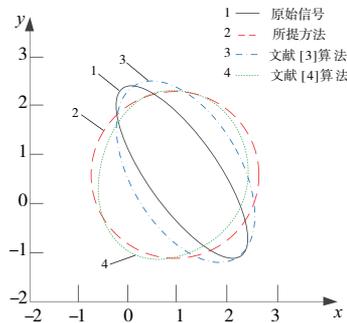


图 5 信号补偿 Lissajou 图

分析图 5 可知,原始图像的正交相位误差较大且呈现椭圆形,经所提方法展开处理后的信号接近于圆形:表明正弦信号的正交相位误差得到了较好的补偿。

### 3 结语

目前的信号补偿方法存在信号处理效果差、补偿效果差和补偿误差大等问题。为解决上述问题,提出了一种煤炭采样机械臂 CAN 控制信号传输畸变补偿技术。经验证得到以下结论:

- 1) 该方法能够有效清除噪声信号,使得有用信号更加清晰可辨,具有较好的信号处理效果;
- 2) 该方法误差波动幅度较小,具有较高的信

号补偿精度和稳定性;

3) 该方法能够有效补偿机械臂系统中信号的幅值偏差,使得补偿后的信号幅值更加一致;

4) 该方法能够较好地补偿正交相位误差,使得信号形成的曲线更加接近圆形。

### 参考文献:

- [1] 姚兴岭,张亮,张淑坤. 基于挖掘力模型的采样机械臂运动位姿节能控制研究[J]. 中国工程机械学报, 2021, 19(4): 318-323.
- [2] 卢俊,张群飞,史文涛,等. 基于通信信号的水下目标主动探测多普勒估计与补偿方法[J]. 西北工业大学学报, 2021, 39(5): 962-970.
- [3] 徐珂,尚自乾,严发宝,等. 毫米波宽带太阳射电观测系统的信号平坦度补偿方法[J]. 中国科学:技术科学, 2021, 51(4): 413-423.
- [4] 陈广森,秦风,高原. 基于维纳滤波的同轴电缆脉冲信号传输畸变补偿研究[J]. 强激光与粒子束, 2021, 33(12): 173-178.
- [5] 何慰,方勇纯,梁潇,等. 一种两自由度飞行机械臂系统的设计与实现[J]. 航空学报, 2021, 42(2): 324280.
- [6] 吴勇,朱建军,邹奔. EEMD 改进算法在异步电机轴承故障诊断中的应用[J]. 中国舰船研究, 2022, 17(6): 111-117.
- [7] 马星河,孔卫东,李自强,等. 一种基于 S\_VMD 与 Sdr\_SampEn 的局部放电信号去噪方法[J]. 电力系统保护与控制, 2022, 50(18): 29-38.

收稿日期:2023-08-10

(上接第 242 页)

### 参考文献:

- [1] 徐伟强. 工业环境下人机协作过程中机器人主动安全策略研究[D]. 南昌:华东交通大学, 2023.
- [2] SHARKAWY A N, MOSTFA A A. Neural networks design and training for safe human-robot cooperation[J]. Journal of King Saud University - Engineering Sciences, 2022, 34(8): 582-596.
- [3] 张堃,刘志诚,刘纪元,等. 面向人机协作系统的上肢姿态精准识别算法研究[J]. 仪器仪表学报, 2023, 44(1): 275-282.
- [4] 张蕊,杨冬,沈永旺,等. 基于 GA-BP 神经网络的接触式人机协作意图理解方法研究[J]. 组合机床与自动化加工技术, 2019(11): 86-91.
- [5] 张蕊. 基于触力觉融合信息的人机交互与柔顺控制方法研究[D]. 天津:河北工业大学, 2019.

- [6] CAI Z Y, FENG Z Q, ZHOU L R, et al. A framework and algorithm for human-robot collaboration based on multimodal reinforcement learning [J]. Computational Intelligence and Neuroscience, 2022: 2341898.
- [7] 姚启, 缪新颖. 基于主成分分析及 GA-LM 的水产养殖环境溶解氧和氨氮含量预测[J]. 大连海洋大学学报, 2021, 36(5): 851-858.
- [8] 汪卫, 陈建政, 吴越. 基于 GA-LSMT 的轮轨力连续测量研究[J]. 机械制造与自动化, 2024, 53(4): 111-118.
- [9] LU Q, LIU S, LI W, et al. Combination of thermodynamic knowledge and multilayer feedforward neural networks for accurate prediction of M S temperature in steels [J]. Materials and Design, 2020, 192: 108696.

收稿日期:2024-10-16