

文章编号: 1674-8190(XXXX)XX-001-12

# 融合注意力机制的航空发动机推力估计方法研究

邹雨杭, 赵永平

(南京航空航天大学 能源与动力学院, 南京 210016)

**摘要:** 准确预测航空发动机推力大小对直接控制发动机推力具有重要意义。为了提升航空发动机推力估计模型的准确性和实用性, 针对时间序列预测构建融合 LSTM 和注意力机制的多任务 LSTM-Attention 模型; 针对不同飞行条件下推力估计的问题, 运用 Fine-tune 和改进 DANN 的迁移学习方法以增强模型对于多工况条件下的适应性。结果表明: LSTM 融合注意力机制可以有效地对长时间序列数据进行建模, 修正了 LSTM 在全局建模能力上不足的问题, 同时也克服了注意力机制难以捕捉相对位置信息的局限; 多任务学习策略能显著提高模型在油门杆突变节点处的预测精度, 进一步提高了模型的准确性; 当目标域数据较少时应当选择 Fine-tune, 而在目标域数据充足的情况下使用改进 DANN 方法将得到准确性更高的模型。

**关键词:** 推力估计; 时间序列预测; LSTM; 注意力机制; 多任务学习; 迁移学习

中图分类号: V239

文献标识码: A

DOI: 10.16615/j.cnki.1674-8190.XXXX.XX.01

## Research on aircraft engine thrust estimation method incorporating attention mechanism

ZOU Yuhang, ZHAO Yongping

(College of Energy and Power Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

**Abstract:** Accurately predicting the thrust of aircraft engines is of great significance for directly controlling engine thrust. This study aims to enhance the accuracy and practicality of thrust estimation models for aero engines. The research first constructs a multi-task LSTM-Attention model that integrates Long Short-Term Memory (LSTM) and attention mechanisms for time series forecasting. Additionally, to address the issue of thrust estimation under different flight conditions, this paper employs Fine-tuning and an improved Domain-Adversarial Neural Network (DANN) transfer learning method to strengthen the model's adaptability to multiple operational conditions. The results demonstrate that LSTM combined with the attention mechanism can effectively model long time series data, rectifying LSTM's insufficiency in global modeling capabilities, while also overcoming the limitation of the attention mechanism in capturing relative position information. The multi-task learning strategy can significantly improve the model's prediction accuracy at the abrupt changes in the throttle levers, further enhancing the model's accuracy. The study of thrust prediction under different conditions based on transfer learning methods indicates that Fine-tuning should be selected when there is limited target domain data, while the modified DANN method will yield a model with higher accuracy when there is sufficient target domain data.

**Key words:** thrust estimation; time series forecasting; LSTM; attention; multi-task learning; transfer learning

收稿日期: 2023-09-12; 修回日期: 2024-02-05

基金项目: 国家科技重大专项(J2019-I-0010-0010); 中央高校基本科研业务费(NS2022027); 航空发动机及燃气轮机基础科学中心项目(P2022-B-V-002-001)

通信作者: 赵永平(1982-), 男, 博士, 研究员。E-mail: y.p.zhao@163.com

引用格式: 邹雨杭, 赵永平. 融合注意力机制的航空发动机推力估计方法研究[J]. 航空工程进展, XXXX, XX(XX): 1-12.

ZOU Yuhang, ZHAO Yongping. Research on aircraft engine thrust estimation method incorporating attention mechanism[J]. Advances in Aeronautical Science and Engineering, XXXX, XX(XX): 1-12. (in Chinese)

## 0 引言

航空发动机是飞机最核心的组件之一,它为飞机提供了飞行过程所需的推力。飞机在执行不同的飞行任务过程中对发动机推力的需求是不同的,这对航空发动机的推力控制系统提出了极高的要求。传统的发动机大多以间接控制的方式控制发动机的推力,但是在控制精确度和响应速度上难以满足发动机对高性能的需求。因此,研究人员对如何直接控制航空发动机的推力方法进行了深入研究<sup>[1]</sup>,准确预测航空发动机推力大小对直接控制发动机推力具有重要的价值和参考意义<sup>[2-3]</sup>。

航空发动机推力估计通常采用数据驱动的建模方法<sup>[4]</sup>,该方法是利用飞行过程中的历史飞行数据,如高度、马赫数、燃油量等,建立发动机推力估计模型。近年来,深度学习因人工神经网络强大的非线性拟合能力,在计算机视觉、自然语言处理、生成式人工智能等领域上得到应用。神经网络依靠其强大的建模能力,成为航空发动机推力估计这一高度非线性多维时间序列预测任务的解决方案之一<sup>[5-7]</sup>。

国内外诸多研究者采用深度学习的建模方法对时间序列建模进行了研究,还有研究人员将其用于航空发动机推力估计的研究之中。循环神经网络(Recurrent Neural Networks,简称RNN)于19世纪80年代被提出,Chung等<sup>[8]</sup>利用其独特的结构实现了时间序列建模,证明了RNN模型拥有依靠过去的数据对未来做预测的能力;然而,Bengio等<sup>[9]</sup>指出,循环神经网络存在着许多缺点,如难以处理长期依赖性和梯度消失/梯度爆炸等问题。为了克服这些问题,Hochreiter等<sup>[10]</sup>提出了改进的RNN变体——长短期记忆网络(Long Short-Term Memory,简称LSTM)。LSTM通过门控机制控制信息的流动,保证了它可以忽略某些时刻的状态或者着重记住某些时刻的状态,从而更有效地处理梯度。李智强<sup>[11]</sup>将LSTM网络用于航空发动机推力估计,并在实验中得到了显著的效果。然而,LSTM并未完全解决RNN的长期依赖问题,在长时间序列估计上表现较差;而且基于LSTM的航空发动机推力估计模型在油门杆角度突变时,推力大小会产生巨大的波动,推力预测曲线的跟随能力较差,极大地降低了模型的性能。

除此之外,仅使用一个高度、马赫数工况下的数据进行的模型训练,由于特征空间偏移的存在,模型无法在全包线内都有良好的表现。

综上所述,如何训练更可靠、实用性更高的模型还需进一步探究。本文针对LSTM网络结构的缺陷进行研究,通过优化网络结构来提高模型的精度;使用一种新的训练框架,解决油门杆突变时推力产生较大误差的问题;针对多工况预测的问题,在不同训练数据量的条件下,实验对比Fine-tune、改进DANN两种迁移学习方法,研究运用迁移学习解决多领域预测问题的方案。

## 1 问题描述

本文的研究对象为某型混排涡扇发动机部件级模型,其主要的部件包括进气道(Inlet)、风扇(Fan)、压气机(High Pressure Compressor,简称HPC)、高压涡轮(High Pressure Turbine,简称HPT)、低压涡轮(Low Pressure Turbine,简称LPT)、尾喷管(Nozzle)等,其结构如图1所示。

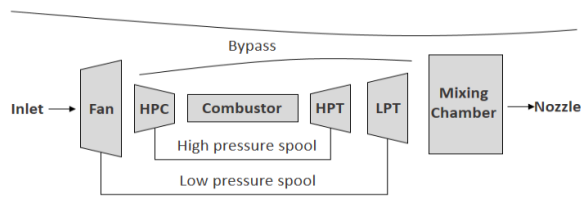


图1 涡扇X发动机结构示意图

Fig. 1 Schematic diagram of turbofan X engine structure

在飞机发动机运行期间,系统主动吸入空气,经过整流、压缩、燃烧等过程,将空气转化为高温高压气体。随后气体高速从发动机排出,为发动机提供了推进动力。

航空发动机飞行历史数据是按照时间顺序排列的一系列观测点的数据集合,其最大的特点是具有时间相关性,即前序时间点的数据可以影响后序时间点的数据。飞机飞行过程中,可采集到的数据中包含多种特征变量,一些变量之间存在较高的相关性,耦合度较高的特征作为训练数据训练模型可能会影响模型的泛化性能,所以需要选取最合适的特征进行推力估计。根据文献<sup>[12]</sup>所述,航空发动机的推力大小与飞机飞行过程中的七个特征变量有密切关系,分别是高度 $H$ 、马赫

数  $Ma$ 、外涵出口总压  $P_{16}$ 、主燃油流量  $W_{FA}$ 、加力燃油流量  $W_{FB}$ 、发动机温比  $E_{TR}$  以及喉道总温  $T_8$ ，本文训练模型所用训练数据采集于上述七个特征变量。

本文模型需要完成的主要任务为航空发动机推力估计，该问题属于多维时间序列回归预测问题，即根据已采集到的历史数据预测未来确定时刻的推力大小，可表示为

$$F_{k+w} = f(D_k, D_{k+1}, \dots, D_{k+w+1}) \quad (1)$$

式中： $w$  为窗口长度； $f(\cdot)$  为本文提出的航空发动机推力估计模型。

在多维时间序列回归预测问题中，多维指的是多维度特征空间，即将长度为  $w$  的时间序列每一个状态点中与推力相关的七个特征变量作为一组输入，输出该序列后一时间点的推力大小。同时，由于发动机推力大小受到飞行状态等因素的影响，每一时刻下的推力与其他时刻下的推力大小和飞行状态具有极强的依赖关系，这一现象也体现了时间序列问题的特点。为此，本文使用类循环神经网络与注意力机制建模，充分解决时间序列的时序依赖问题。

## 2 模型原理与训练方法

### 2.1 Transformer 与自注意力机制

Transformer 模型(如图 2 所示)最初在自然语言处理领域被引入，并在机器翻译等任务中取得了显著的成果。与传统的 RNN 相比，Transformer 模型采用了自注意力机制，使模型能够直接对序列中的所有位置进行建模，并且能够捕捉长期依赖关系<sup>[13]</sup>。这种机制不仅改善了模型的建模能力，还具备了更高的并行计算效率，这对于时间序列预测问题尤为重要。

原始的 Transformer 模型是由多个 Encoder 和 Decoder 模块堆叠构造，编码器负责从序列中提取特征，而解码器用于生成目标序列。Encoder 包含一个 Multi-head Self-attention 层，如图 3 所示，数据经 Multi-head Self-attention 输出后经过与输入数据残差连接，防止网络退化，再通过 LN(Layer Normalization) 标准化层进行归一化处理，加快模型的收敛速度。

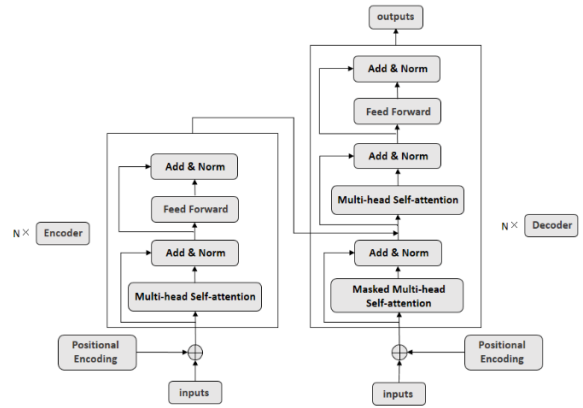


图 2 Transformer 结构示意图

Fig. 2 Transformer structure schematic diagram

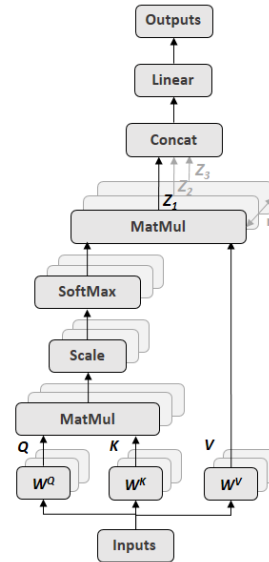


图 3 Multi-head Self-attention 结构示意图

Fig. 3 Multi-head Self-attention schematic diagram

单个自注意力层的输入是由特征向量组成的矩阵，输入矩阵分别经过线性网络层  $W^Q$ 、 $W^K$ 、 $W^V$  得到矩阵  $Q$ 、 $K$ 、 $V$ ，该线性网络层即为训练时需要学习的对象。计算自注意力网络层的输出，如式 (2) 所示。

$$Z_i = Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

Multi-head Self-attention 由  $n$  个自注意力网络构成，每个网络处理不同的特征子空间，得到一系列输出矩阵  $Z_1, Z_2, \dots, Z_n$ ，将  $n$  个输出矩阵拼接后再传入一个全连接层得到最终的输出  $Z$ ，如式 (3) 所示。模型可以同时关注输入时间序列中的多个相关信息，从而更好地捕捉输入之间的相关性，提

高模型的鲁棒性和泛化能力。

$$Z = Fc(\text{Concat}(Z_1, Z_2, \dots, Z_n)) \quad (3)$$

式中:  $\text{Concat}(\cdot)$  为矩阵拼接函数;  $Fc(\cdot)$  为全连接层。

## 2.2 多任务 LSTM-Attention

本文结合 LSTM 的序列建模能力和 Multi-head Self-attention 的获取序列中全局模式和依赖关系的能力, 提出了 LSTM-Attention 模型结构。LSTM 能够学习序列中的长期依赖关系, 而 Attention 机制可以帮助模型关注到序列中的重要部分, 所以将 LSTM 放在前面 attention 放在后面效果更好。如图 4 所示, 共享层 LSTM 从原始时间序列中提取特征, 特有层 LSTM 从上层神经网络中提取该任务特征。

基于 LSTM 和注意力机制的单任务时间序列预测任务中, 仅依靠当前时间点之前的序列数据进行建模, 导致模型难以捕捉在油门杆角度突变时航空发动机推力变化的规律, 使得模型在推力突变点的估计精度降低。如果模型能预知推力大小何时发生突变, 将极大提高模型的推理能力。为了降低单任务模式对模型性能的限制, 进一步提高模型的估计精度, 本文提出使用多任务 LSTM-Attention (Multi-Task-LSTM-Attention, MT-LA) 作为模型框架, 实现过去和未来信息的互补<sup>[14]</sup>, 降低模型在推力突变时的估计误差。

传统的神经网络模型只执行一个预测任务  $P_t$ , 如图 4 所示本文提出的多任务学习框架执行了三个预测任务  $P_{t-1}$ 、 $P_t$ 、 $P_{t+1}$ , 分别估计  $t-1$ 、 $t$ 、 $t+1$  时刻的推力大小。其中  $P_t$  为主任务,  $P_{t-1}$ 、 $P_{t+1}$  为辅助任务, 为主任务  $P_t$  提供更多信息, 提高推力估计精度。MT-LA 采取参数共享的方式, 其结构包含共享层和特有层两部分。共享层包含 LSTM 层和位置编码模块, 共享层参数被所有预测任务共享, 在训练过程中所有任务都参与了该层网络参数的更新, 所以建立了各个任务之间关联。特有层由 LSTM-Attention 模型构成, 特征变量经过共享层的特征提取后, 数据将被送入各任务的特有层中, 基于提取出的通用特征执行单独的训练任务。在模型的推理过程中, 时间序列数据先经过共享层网络, 再通过主任务  $P_t$  的特有层推理得到推力的估计值。详细的算法流程表 1 中所示, 主要介绍了模型的训练流程和模型的推理过程。

表 1 MT-LA 算法具体流程  
Table 1 The specific process of the MT-LA

训练过程算法流程:

$$\text{训练集为 } \left\{ \left( (x_{t-1}^i, y_{t-1}^i), (x_t^i, y_t^i), (x_{t+1}^i, y_{t+1}^i) \right) \right\}_i^M$$

$x_{t-1}^i, x_t^i, x_{t+1}^i$  分别表示  $t-1, t, t+1$  时刻相同窗口大小截取的时间序列特征变量

步骤 1: 模型初始化;

步骤 2: 将训练集输入模型, 通过公式(4)计算多任务模型的损失;

$$L_{\text{MTL}} = \sum_i w_i \cdot L_i \quad (4)$$

其中  $w_i$  为任务权重系数,  $L_i$  为单任务的 MSE 损失

步骤 3: 通过公式(5)更新共享层 LSTM 网络参数

$$W_{sh} = W_{sh} - \gamma \frac{\partial L_{\text{MTL}}}{\partial W_{sh}} \quad (5)$$

其中  $W_{sh}$  为共享层参数矩阵,  $\gamma$  为学习率, 梯度反向传播使每个任务均参与共享层网络参数的更新。

步骤 4: 各任务通过公式(6)更新其特有层网络参数

$$W_{sp_i} = W_{sp_i} - \gamma_i \frac{\partial L_i}{\partial W_{sp_i}} \quad (6)$$

其中  $W_{sp_i}$  为特有层参数矩阵,  $\gamma_i$  为该任务学习率, 特有层网络仅通过计算主任务梯度进行参数的更新。

步骤 5: 经过多次迭代重复步骤 2、3、4, 训练出在测试集性能最优的模型。

模型的推理过程:

步骤 1: 通过固定长度为  $N$  的滑动窗口截取时间序列, 得到模型的输入  $x_i$

步骤 2: 通过公式(7)得到该时刻的推力估计值  $F_{pre}^i$

$$F_{pre}^i = F_{sp}(F_{sh}(x_i)) \quad (7)$$

其中  $F_{sh}(\cdot)$  为共享层网络,  $F_{sp}(\cdot)$  为主任务  $P_t$  的特有层网络

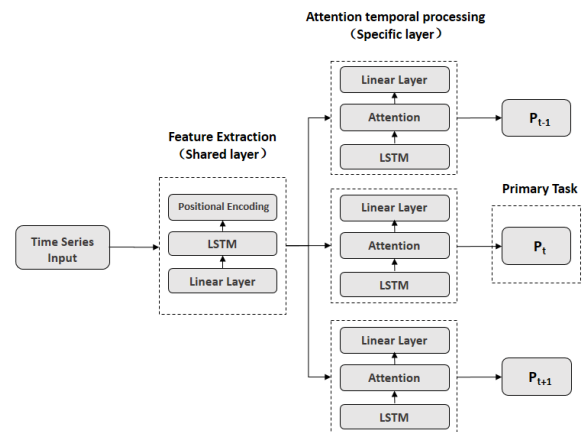


图 4 多任务 LSTM-Attention 结构原理图

Fig. 4 Schematic diagram of multi-task LSTM-Attention structure



### 2.3 Fine-tune 与改进 DANN 的迁移学习策略

本文提出的 MT-LA 模型仅对一个工况进行建模,由于各工况特征分布不同,单一模型难以适用于包线内多领域工况。航空发动机推力估计模型需要具备全工况的预测能力,为了让模型能够估计飞行包线内所有工况的推力大小,为每一个工况训练一个模型是不现实的,如何利用有限的数据集训练一个全包线通用推力估计器是一个巨大的挑战。迁移学习的核心思想是将模型的能力在不同领域上进行迁移,通常是在一个任务中学到的知识迁移到另一个相关的任务上<sup>[15-17]</sup>。飞行包线内不同的工况点训练数据和测试数据存在分布上的差异,迁移学习则可以通过学习源域数据的知识提高在目标领域上的泛化能力。

Fine-tune 是迁移学习的常用方法,该方法一般是通过修改预训练网络模型结构,选择性载入预训练网络权重,使用目标域数据训练新的模型。本文使用的 Fine-tune 策略为小学习率微调 MT-LA 网络中共享层特征提取网络 LSTM,以保留先前学到的知识,重新训练预训练模型特有层中的注意力层。这种迁移学习的训练方法最大的优势在于可以利用已学习到的知识,使用少量有标签样本就可以将预训练模型微调为更适用于目标域的模型,减少了对数据量的依赖,同时也能取得较好的预测效果。模型微调的原理如图 5 所示。

相较于 Fine-tune 的微调方法,为解决迁移学习问题中源域和目标域的适配问题, Muhammad 等<sup>[18]</sup>提出了域自适应神经网络(Domain Adaptive

Neural Network,简称 DANN)。这种神经网络由特征层和分类层组成,通过计算最大均值差异(maximum mean discrepancy,简称 MMD)计算源域和目标域分布差距,源域与目标域之间 MMD 越小表明分布差异越小。在模型的训练过程中,将 MMD 计入损失函数中,影响网络参数的更新。假设源域的随机变量为  $S = \{x_1, x_2, \dots, x_n\}$ ,目标域的随机变量为  $T = \{p_1, p_2, \dots, p_m\}$ ,则 MMD 的计算公式如下:

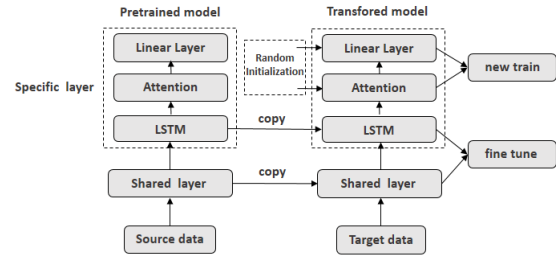


图 5 Fine-tune 原理图

Fig. 5 Schematic diagram of Fine-tune

$$\lambda_{MMD}^2(S, T) = \left\| \frac{1}{n} \sum_{i=1}^n \varphi(x_i) - \frac{1}{m} \sum_{j=1}^m \varphi(p_j) \right\|_H^2 \quad (8)$$

式中:  $\varphi(\cdot)$  为映射函数,将原数据映射至再生核希尔伯特空间(reproducing kernel hilbert space,简称 RKHS);  $\|\cdot\|_H^2$  为源域数据与目标域数据在 RKHS 中的距离,该距离用来衡量分布之间的差异。由于映射函数是未知的,所以通常将式(8)平方展开,并且用高斯核函数代替原来的映射函数。

MMD 的核函数公式如下:

$$\lambda_{MMD}^2(S, T) = \left\| \frac{1}{n^2} \sum_{i=1}^n \sum_{i'=1}^n K(x_i, x_{i'}) - \frac{2}{mn} \sum_{i=1}^n \sum_{j=1}^m K(x_i, p_j) + \frac{1}{m^2} \sum_{j=1}^m \sum_{j'=1}^m K(p_j, p_{j'}) \right\|_H \quad (9)$$

式中:  $K(\cdot)$  为核函数。

在模型迁移训练的过程中,通过减小 MMD 误差,减小源域与目标域的分布偏移问题,从而实现知识从源域向目标域的迁移, DANN 的损失函数式(10)所示:

$$L_{DANN} = L_{MSE} + \gamma \lambda_{MMD}^2(q_s, q_t) \quad (10)$$

式中:  $L_{DANN}$  为 DANN 网络的损失函数;  $L_{MSE}$  为模型 MSE 损失;  $\gamma$  为 MMD 的权重系数,一般设置为 0.25;  $q_s$  与  $q_t$  分别为源域与目标域数据经过特征提取层得到的输出。

对原 DANN 模型进行改进,使得模型可以用于多维时序回归预测问题。去除 MT-LA 中主任任务最后的线性层,其余网络部分作为 DANN 的特征提取层,将原 DANN 模型中的分类层更换为全连接层,全连接层 1、2 为参数共享层,全连接层 3 与全连接层 4 是根据不同特征空间下推力估计任务设置的独立的全连接层,两个连接层的神经元的个数一致。基于 MT-LA 的改进 DANN 模型结构如图 6 所示,具体训练流程如表 2 所示。

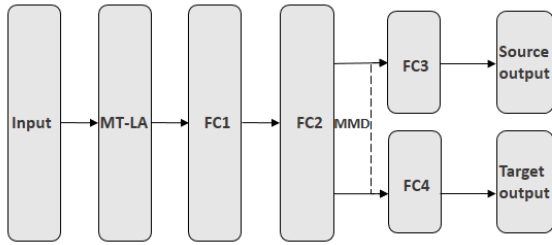


图6 改进DANN模型

Fig. 6 Schematic diagram of DANN

表2 改进DANN模型的训练流程

Table2 The training process of the DANN model

**步骤1:**将所有训练数据归一化并记录最大最小值,以便后续进行反归一化;

**步骤2:**分别将源域数据集与目标域数据集按一定比例划分为训练集与测试集;

**步骤3:**将步骤2中的源域和目标域的训练集输入模型进行训练,公式(10)中的均方误差计算公式如下:

$$L_{MSE} = \omega_1 f_{MSE}(y_s, \hat{y}_s) + \omega_2 f_{MSE}(y_t, \hat{y}_t) \quad (11)$$

式中  $f_{MSE}(\cdot)$  为均方根误差损失函数;  $y_s, y_t$  分别为源域与目标域推力大小的实际值,  $\hat{y}_s, \hat{y}_t$  分别为源域与目标域推力大小的估计值;  $\omega_1, \omega_2$  分别为源域和目标域均方误差的权重因子。计算公式(10)中的  $L_{DANN}$ , 通过梯度反向传播算法更新网络参数。

**步骤4:**将步骤3训练好的模型在目标域数据集上进行测试,得到目标域上推力大小的归一化估计值

### 3 实验与结果分析

#### 3.1 模型验证实验数据集

实验数据选取高度为 10 km, 马赫数为 0.8 的 5 000 条训练样本和 5 000 条测试样本。序列窗口长度设为 20, 即时间步长为 20, 航空发动机在  $t$  时刻的推力由  $t$  时刻的前 20 个时刻点的信息决定。采集发动机数据时, 以两种不同油门杆角度变化规律将 PLA 从  $30^\circ$  调至  $80^\circ$  再调回  $30^\circ$ , 得到两组特征分布相同的训练集与测试集用于实验的模型验证。将数据用于模型训练前需对数据进行归一化, 然后分别对归一化后的数据集单独进行对应的输入输出处理。归一化前每一条数据样本均由七个相关特征与推力大小组成, 一条样本如下所展示(训练前所有特征均单独归一化处理, 单位不做考虑):

{高度  $H$ : 1.403 3, 马赫数  $Ma$ : 0.640 3, 外函出

口总压  $P_{16}$ : 229 330.688 753 239, 主燃油流量  $W_{FA}$ : 0.786 797 195 525 061, 加力燃油流量  $W_{FB}$ : 0, 发动机温比  $E_{TR}$ : 2.030 291 226 928 15, 喉道总温  $T_8$ : 612.965 148 026 226, 推力大小  $F$ : 32 324.251 631 669 5}

表3 实验数据集信息

Table3 Experimental dataset information

高度/马赫	训练集		测试集	
	输入规模	输出规模	输入规模	输出规模
H=10Km Ma=0.8Mach	(5 000, 20, 7)	(4 977, 1)	(5 000, 20, 7)	(4 977, 1)

输入规模(5 000, 20, 7)表示, 共 5 000 条数据样本, 输入序列长度为 20, 7 个特征变量。输出规模(4 977, 1), 由于输入序列长度为 20 以及舍去最后不满足数量大小的批次, 输出推力预测序列长度为 4977。

后续迁移学习方法验证实验中需要使用到的数据集, 将在 3.4.1 迁移学习实验方法中进行补充介绍。

#### 3.2 实验评价指标

将预测结果进行直观比较, 本文采用平均绝对误差(MAE)、平均绝对百分比误差(MAPE)、均方根误差(RMSE)为评价指标。计算公式如下:

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_{act} - y_{pre}| \quad (12)$$

$$MAPE = \frac{1}{m} \sum_{i=1}^m \frac{|(y_{act} - y_{pre})|}{y_{act}} \quad (13)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_{act} - y_{pre})^2} \quad (14)$$

式中:  $y_{act}$  为测试样本的实际值;  $y_{pre}$  为模型的预测值。

#### 3.3 MT-LA 模型实验验证与结果分析

实验中, 模型的输入序列长度为 20, 共享 LSTM 的输入维度设为 200, 输出维度设为 500, 隐藏层单元数设为 2 000, 同时特有层 LSTM 的输出维度设为 20, 隐藏层单元数设为 2 000, 注意力层头数设置为 4。训练时使用 AdamW 作为优化器, 并设置批次大小 batch\_size 为 64。选择 MSE 作为模型的损失函数, 同时模型训练的迭代次数设置为 1000, 初始学习率设置为 0.001, 并随迭代次数

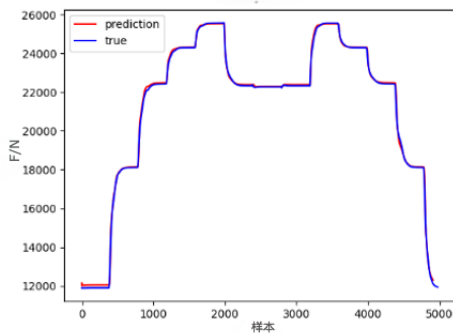
动态调整学习率。

实验首先验证了 LSTM-Attention 相较于其他深度神经网络模型的优越性。以本文提出的 LSTM-Attention 模型为基准模型,分别移除 LSTM 和 Attention 层训练模型,对测试集进行预测。表 4 中展示了各模型的评价指标,通过各模型预测曲线(如图 7 所示)横向对比可以发现 LSTM-Attention 各项性能指标都有明显优势,体现出各网络模块在航空发动机推力估计模型中对多维时间序列数据处理的重要性。通过对预测结果进行统计,LSTM-Attention 模型预测结果与推力的实际值的绝对误差绝大部分在 100 N 以下,相对误差在 1% 以下,基本满足航空发动机的推力估计精度要求。可以看到由于推力的数量级较大,预测曲线上差距不明显,但是通过绝对误差和相对误差曲线可以发现在油门杆角度变化的节点上,极少数预测结果偏差较大,但是模型的调整跟随能力较强,预测误差很快降至 1% 以下。综上结果,验证了 LSTM-Attention 模型结构在性能上的优越性。

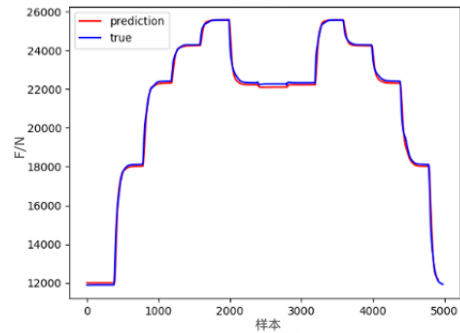
表 4 MT-LA 验证实验结果

Table 4 MT-LA validation experiment results

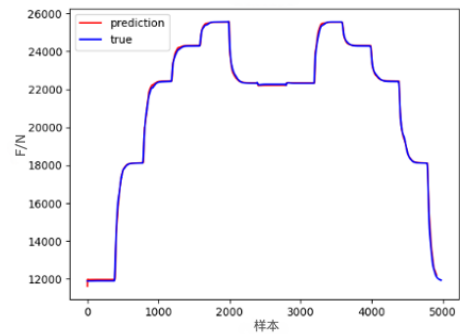
评价指标	LSTM	Transformer	LSTM-Attention	MT-LA
MAE	63.66	121.10	45.81	29.47
MAPE	0.003 4	0.006 1	0.002 4	0.001 2
RMSE	90.14	180.19	73.13	56.22



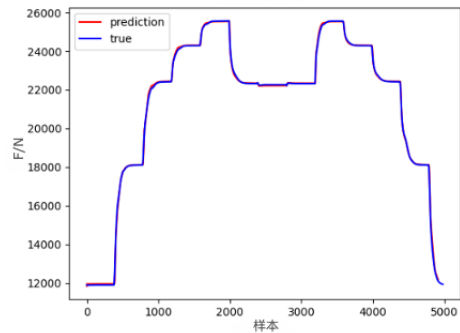
(a) LSTM 预测曲线



(b) Transformer 预测曲线



(c) LSTM-Attention 预测曲线

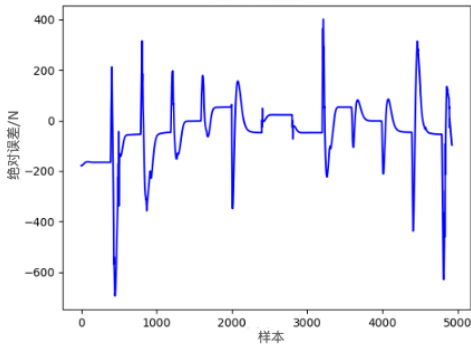


(d) MT-LA 预测曲线

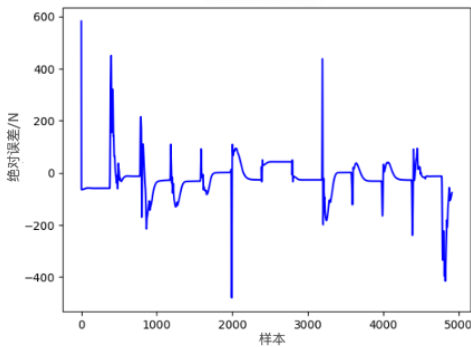
图 7 各模型预测曲线对比  
Fig. 7 Comparison of prediction curves from various models

为了进一步提高模型的预测精度,降低油门杆变化节点处的误差,采取多任务 LSTM-Attention 架构进行训练,使用与单任务训练模式相同数据集对 MT-LA 模型进行实验验证。实验结果表明,在各项评价指标上多任务学习模式均有提优,对比 LSTM-Attention 和 MT-LA 的绝对误差曲线

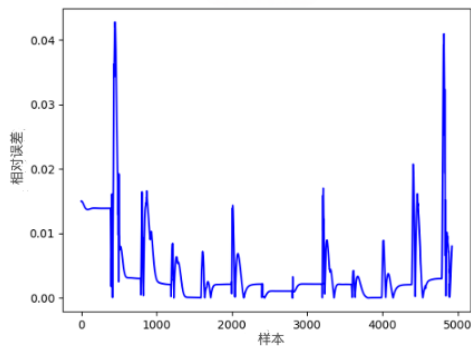
和相对误差曲线(如图8所示),可以发现采用多任务训练框架使得模型在油门杆变化节点处误差降低较为明显。统计计算油门杆突变点前后100个预测值的绝对误差与相对误差的均值,LSTM-Attention的绝对误差均值约为279.53 N、相对误差均值约为1.48%,MT-LA的绝对误差均值约为117.91 N、相对误差均值约为0.42%,MT-LA体现出较为明显的优势。



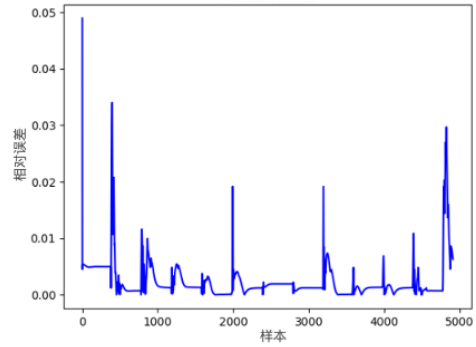
(a) LSTM-Attention 绝对误差曲线



(b) MT-LA 绝对误差曲线



(c) LSTM-Attention 相对误差曲线



(d) MT-LA 相对误差曲线

图8 LSTM-Attention与MT-La误差对比  
Fig.8 Comparison of errors between LSTM-Attention and MT-LA

### 3.4 基于迁移学习的多工况推力估计实验

#### 3.4.1 迁移学习实验方法

在目标域微调数据量充足和数据量较少的两种情况下,对比通过Fine-tune和改进DANN两种迁移学习方法训练的迁移学习模型,验证方法的可行性,得到目标域数据量少时应使用Fine-tune,目标域数据量充足时使用改进DANN模型更为合适的结论。同时由于源域数据集的差异,分为单源域与多源域两种情况分别进行实验。实验过程流程如图9所示。

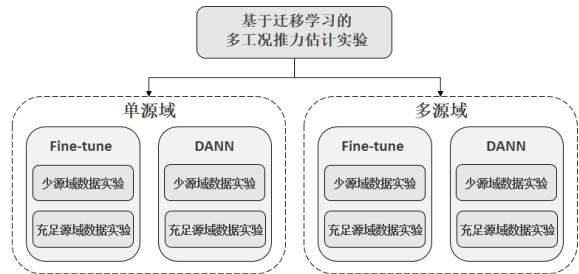


图9 迁移学习验证实验流程  
Fig.9 Validation experiment workflow for transfer learning

#### 3.4.2 单源域迁移实验结果验证与分析

以3.3节中训练得到的MT-LA模型为预训练模型,取工况高度0 km,马赫数0的7500条数据进行实验,数据取样方法与3.1节中相同。实验验证方法为,将7500条数据分为两部分,其中500条用于模拟少样本训练条件,2000条用于模拟样本数据充足的训练条件,另外5000条为微调后模型的测试集。基于上述两种目标域数据集,分别用

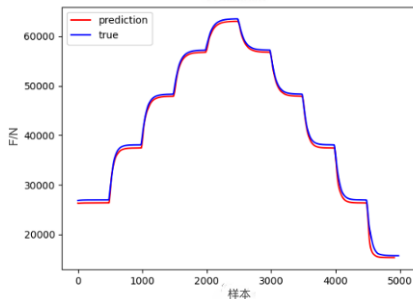


Fine-tune 的微调方法和改进 DANN 网络训练迁移学习模型。单源域下各模型预测曲线对比图(如图 10 所示)直观地展示了使用 Fine-tune 与改进 DANN 的两种迁移学习策略在不同训练条件下学习得到模型的性能差异。实验结果的各项性能指标如表 5 所示,通过对比表中结果,可以发现 Fine-tune 的微调策略对数据样本变化并不敏感,增大目标域样本参与迁移学习评价指标均只有少量提升。相反,使用的微调数据充足的情况下改进 DANN 模型效果提升十分明显,所以目标域样本充足时应选用改进 DANN 模型做迁移学习。

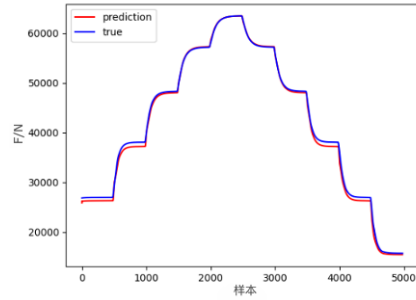
单源域迁移实验的目的主要是验证采用迁移学习方法解决多领域推力估计问题的有效性,通过对比结果可以看到,迁移学习的策略能够在一定程度上解决变工况预测的需求,将 Fine-tune 和改进 DANN 与 MT-LA 模型结合对航空发动机推力估计模型建模是可行的。但是单源域迁移模型预测性能不满足使用要求,如何使用迁移学习的方法训练得到预测精度更高的模型还需要进一步探究。

表 5 单源域迁移模型前后实验结果  
Table 5 Experimental results before and after single-source domain transfer model

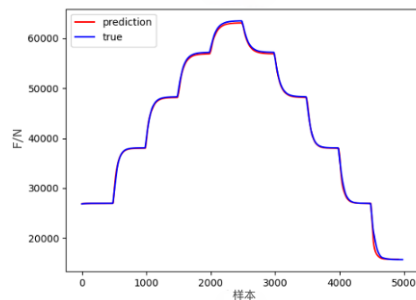
评价指标	迁移学习前	Fine-tune		DANN	
		500 样本	2000 样本	500 样本	2000 样本
MAE	445.66	262.94	234.44	624.11	137.61
MAPE	0.0140	0.0082	0.0071	0.0245	0.0008
RMSE	567.31	327.15	297.87	715.19	163.22



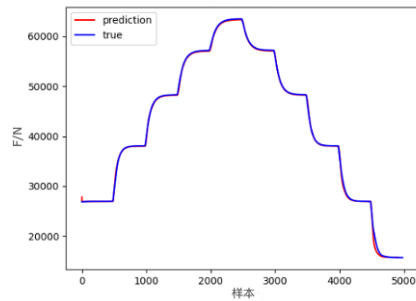
(a) 迁移前预测曲线



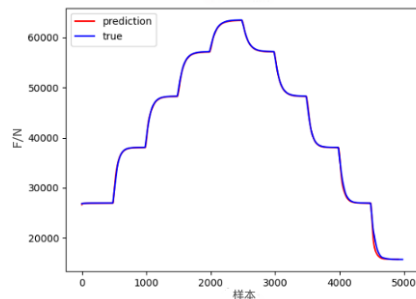
(b) 500 样本 Fine-tune 预测曲线



(c) 2000 样本 Fine-tune 预测曲线



(d) 500 样本改进 DANN 预测曲线



(e) 2000 样本改进 DANN 预测曲线

图 10 单源域迁移模型微调预测曲线对比  
Fig. 10 Comparison of prediction curves for single-source domain transfer models

### 3.4.2 多源域迁移实验结果验证与分析

从上述实验中得出结论,两种迁移学习方法可以从预训练模型中学习到有用的知识,但是仅微调单源域预训练模型,迁移模型的预测能力较差。多源域的数据集包含了更多的领域,使得模型在学习过程中能够更好地适应不同的情境和模式。同时数据量相对较大,有更多的实例用于学习,有助于模型更好地捕捉特征和模式。基于神经网络强大的拟合能力和数据处理能力,将飞行包线内多领域内更多的飞行数据用于训练,将学习得到的多源域时序特征分别用上述两种迁移学习方法进行实验,实验验证过程如下。

实验预训练模型的训练集包含飞行包线内多工况飞行数据共 30 000 条数据,所取训练数据的高度和马赫数如图 11 所示,每 3 000 个采样点作为一组,每组高度均匀变化 5 km,马赫数均匀变化 0.5,以此模拟飞机动态飞行过程。测试数据集选取高度为 0 km,马赫数为 0 的数据样本 7 500 条,数据采样与模型验证方法与 3.4.2 节中实验相同。预训练模型的训练集是多源域的且数据量大,旨在验证多源域多数据量的情况下预训练模型的效果更好。

不同训练条件下各迁移模型预测曲线对比如图 12 所示,实验结果的各项性能指标如表 6 所示,同时对比表 5 与表 6 中的实验结果,表明多源域迁移学习模型学习到的时序特征比单源域迁移学习到的时序特征更充分,利用更多领域的更多数据做迁移学习效果更好。该实验的意义在于,在模型训练条件允许的前提下,可以使用大量的多源域飞行数据进行训练,涵盖更多不同高度和马赫数的工况,这样通过多源域迁移学习方法可以对飞行包线内任一工况下航空发动机推力进行准确估计,大大提高了模型的通用性,解决了因飞行包线内不同工况特征空间不同而限制了预测模型估计精度的问题。同时也能得出结论,无论是单源域还是多源域的情况下,当目标域数据量少的时候应选择 Fine-tune 的策略迁移学习,当目标域数据充足时应选用改进 DANN 网络。

为验证结论的科学性,在高度为 10 km,马赫数为 1.5 的工况下进行采样,采样方式与上述实验一致,并将数据集划分为微调数据集与测试数据集。与上述实验方法相同对两个预训练模型进行

微调,验证微调模型的预测能力,实验结果见表 7 与表 8,表格中数据与高度为 0 km 马赫数为 0 数据集所得结果一致,验证了上述结论。

表 6 多源域迁移模型实验结果  
Table 6 Experimental results for multi-source domain transfer models

评价指标	Fine-tune		DANN	
	500 样本	2 000 样本	500 样本	2 000 样本
MAE	103.11	147.72	369.86	37.19
MAPE	0.0051	0.0009	0.0125	0.0005
RMSE	163.27	208.54	471.69	80.31

表 7  $H=10$  km,  $Ma=1.5$  单源域模型结果  
Table 7  $H=10$  km,  $Ma=1.5$  Results of the single-source domain model

评价指标	Fine-tune		DANN	
	500 样本	2000 样本	500 样本	2000 样本
MAE	231.11	248.29	472.93	106.22
MAPE	0.0130	0.0189	0.0369	0.0011
RMSE	273.68	303.74	631.29	53.080

表 8  $H=10$  km,  $Ma=1.5$  多源域模型结果  
Table 8  $H=10$  km,  $Ma=1.5$  Results of the multi-source domain model

评价指标	Fine-tune		DANN	
	500 样本	2000 样本	500 样本	2000 样本
MAE	158.02	213.42	314.21	16.500
MAPE	0.0072	0.0109	0.0512	0.0027
RMSE	196.60	217.52	519.15	90.122

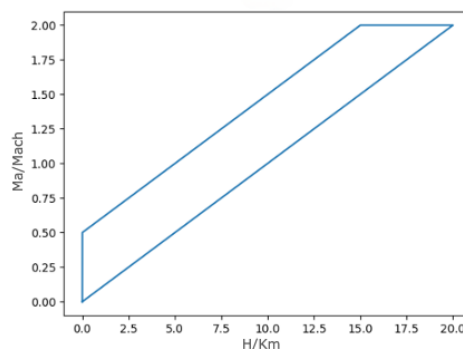
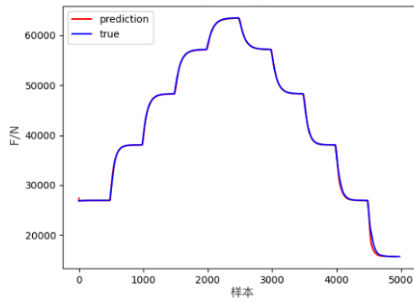
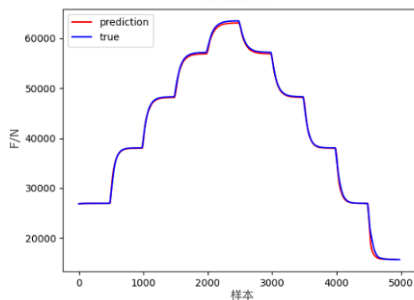


图 11 源域数据集分布图

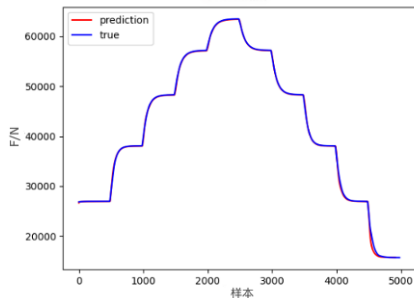
Fig. 11 Distribution of source domain dataset



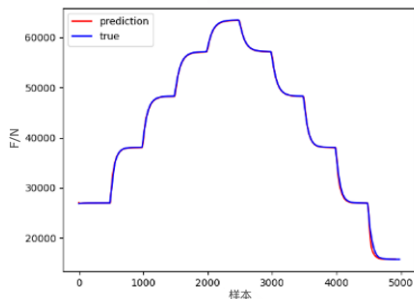
(a) 500 样本 Fine-tune 预测曲线



(b) 2000 样本 Fine-tune 预测曲线



(c) 500 样本改进 DANN 预测曲线



(d) 2000 样本改进 DANN 预测曲线

图 12 多源域迁移模型预测曲线对比

Fig. 12 Comparison of prediction curves for multi-source domain transfer models

## 4 结 论

1) LSTM 的网络结构在长序列上的全局建模和学习能力不足,而 Transformer 中的自注意力机制模块从全局的角度对不同时间步的多维向量进行自适应加权学习,弥补了 LSTM 的缺陷。同时, LSTM 也弥补了自注意力机制难以学习到相对位置信息的缺点。

2) 多任务 LSTM-Attention 模型架构训练模型,可以结合当前预测时间点过去和未来的信息,对模型的预测结果进行补充修正,提高了油门杆突变节点时模型的预测精度。

3) 针对飞行包线多工况推力预测的难题, Fine-tune 和改进 DANN 的迁移学习方法均可得到在目标域上推力估计表现良好的迁移学习模型,当目标域数据集较少时应选用 Fine-tune,反之用改进 DANN 可以训练估计精度更高的模型。

本文仍然存在一定的局限性,未来展望如下:

1) 在模型的实际运用中,目标域的样本可能是不带标签数据的,无法运用本文提出的迁移学习策略。现有的文献主要是针对有监督的领域自适应问题进行了探究,对多维时间序列回归预测的无监督领域自适应的研究较少。如何解决标签样本极少,甚至没有带标签数据的样本的问题,需要进一步探究。

2) 增量学习的学习方式也适用于航空发动机的推力估计问题。增量学习的优势在于能够在保留先前知识的基础上不断更新模型,减少对历史数据的依赖,并能够快速适应新的数据和任务。它可以避免重新训练整个模型,提高学习效率和灵活性。可以通过给模型提供新的飞行数据,让模型得到“在线更新”,从而适应于新的飞行环境。增量学习的学习方式是今后继续探究的一个可行的方向。

## 参 考 文 献

- [1] MAGGIORE M, ORDÓNEZ R, PASSINO K M, et al. Estimator design in jet engine applications[J]. Engineering Applications of Artificial Intelligence, 2003, 16(7/8): 579-593.
- [2] 李晓白,崔秀伶,郎荣玲. 航空发动机性能参数预测方法[J]. 北京航空航天大学学报, 2008, 34(3): 253-256.  
LI Xiaobai, CUI Xiuling, LANG Rongling, et al. Forecasting method for aeroengine performance parameters[J]. Jour-

- nal of Beijing University of Aeronautics and Astronautics, 2008, 34(3): 253-256. (in Chinese)
- [3] 王芳, 吴丹, 樊思齐. 航空发动机性能参数估计方法研究 [C]// 中国航空学会第十二届发动机自动控制学术会议. 北京: 中国航空协会, 2004: 112-115.  
WANG Fang, WU Dan, FAN Siqi. Research on methods for estimating aircraft engine performance parameters [C]// The 12th Academic Conference on Engine Automatic Control of the Chinese Aeronautical Society. Beijing: Chinese Society of Aeronautics and Astronautics, 2014: 112-115. (in Chinese)
- [4] SONG H Q, LI B W, ZHANG Y, et al. Aero-engine thrust estimator design based on clustering and particle swarm optimization extreme learning machine [J]. Tuijin Jishu/Journal of Propulsion Technology, 2017, 38: 1379-1385.
- [5] 陈恬, 孙健国. 基于相关性分析和神经网络的直接推力控制 [J]. 南京航空航天大学学报, 2005, 37(2): 52-56.  
CHEN Tian, SUN Jianguo. Direct control of aeroengine thrust based on correlation analysis and neural networks [J]. Journal of Nanjing University of Aeronautics and Astronautics, 2005, 37(2): 52-56. (in Chinese)
- [6] 姚彦龙, 孙健国. 自适应遗传神经网络算法在推力估计器设计中的应用 [J]. 航空动力学报, 2007, 22(10): 1748-1753.  
YAO Zhaolong, SUN Jianguo. Application of adaptive genetic neural network algorithm in design of thrust estimator [J]. Journal of Aerospace Power, 2007, 22(10): 1748-1753. (in Chinese)
- [7] SCHMIDHUBER J, WIERSTRA D, GOMEZ F J. Evolino: hybrid neuroevolution/optimal linear search for sequence prediction [C]// Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI). 2005: 853-858.
- [8] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [J]. arXiv preprint arXiv:1412.3555, 2014.
- [9] BENGIO Y, FRASCONI P, SIMARD P. The problem of learning long-term dependencies in recurrent networks [C]// IEEE international conference on neural networks. IEEE, 1993: 1183-1188.
- [10] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural computation, 1997, 9(8): 1735-1780.
- [11] 李智强. 基于神经网络的航空发动机推力估计方法研究 [D]. 南京: 南京航空航天大学, 2019.
- LI Zhiqiang. Research on Aeroengine Thrust Estimation Methods Based on Neural Networks [D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2019. (in Chinese)
- [12] ZHAO Y P, SUN J G, WANG Q Y, et al. Thrust estimator design based on K-means clustering and reduced least squares support vector regression [J]. Journal of Aerospace Power, 2010, 25(5): 1177-1183.
- [13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in neural information processing systems, 2017, 30.
- [14] 叶锐. 基于迁移学习和多任务学习的时间序列预测算法 [D]. 南京: 南京航空航天大学, 2019.  
YE Rui. Time series prediction algorithms based on transfer learning and multi-task learning [D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2019. (in Chinese)
- [15] 付松. 民用航空发动机气路性能异常检测与剩余寿命预测方法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2022.  
FU Song. Research on the methods of gas path performance anomaly detection and remain useful life prediction for civil aeroengine [D]. Harbin: Harbin Institute of Technology, 2022. (in Chinese)
- [16] 陶一凡. 融合迁移学习和长短记忆网络的建筑负荷预测研究 [D]. 徐州: 中国矿业大学, 2021.  
TAO Yifan. Research on Building Load Forecasting with Transfer Learning and Long Short Memory Network [D]. Xuzhou: China University of Mining and Technology, 2021. (in Chinese)
- [17] 黄崧琳, 景博, 潘晋新, 等. 基于退化阶段识别与 LSTM-fine-tune 的氧气浓缩器多阶段退化预测 [J]. 电子测量与仪器学报, 2022, 36(7): 136-143.  
HUANG Songlin, JING Bo, PAN Jinxin, et al. Multistage degradation prediction of oxygen concentrator based on degradation pattern recognition and LSTM-fine-tune [J]. Journal of Electronic Measurement and Instrument, 2022, 36(7): 136-143. (in Chinese)
- [18] GHIFARY M, KLEIJN W B, ZHANG M. Domain adaptive neural networks for object recognition [C]// Pacific Rim International Conference on Artificial Intelligence. Gold Coast, QLD, Australia: Springer International Publishing, 2014: 898-904.

(编辑: 马文静)