

# 基于 DAM 和 CNN-LSTM 的辅助动力 装置性能参数预测模型

王力<sup>1</sup>, 马宪<sup>2</sup>

(1. 中国民航大学 职业技术学院, 天津 300300; 2. 中国民航大学 电子信息与自动化学院, 天津 300300)

**摘要:** 针对辅助动力装置 (APU, auxiliary power unit) 性能参数难以准确预测的问题, 提出一种基于特征与时序的双阶段注意力机制 (DAM, dual-stage attention mechanism) 和卷积神经网络 (CNN) - 长短期记忆网络 (LSTM) 的混合模型; 所提的方法在特征提取阶段加入了通道注意力机制 (CAM, channel attention mechanism); 输出阶段加入了时序注意力机制 (TAM, temporal attention mechanism), 加强了 CNN 对重要特征的提取能力和历史关键信息对预测输出的影响, 并利用改进的粒子群算法对模型关键参数寻优, 提高预测精度; 实验结果表明, 所提出的新方法在多变量输入和多步长的 APU 排气温度 (EGT, exhaust gas temperature) 预测中均取得了很好的效果, 预测精度大幅提高。

**关键词:** 辅助动力装置; 性能参数预测; 双阶段注意力机制; 长短期记忆网络; 排气温度; 粒子群算法

## Prediction Model of Auxiliary Power Unit Performance Parameter Based on DAM and CNN-LSTM

WANG Li<sup>1</sup>, MA Xian<sup>2</sup>

(1. Vocation and Technical College, Civil Aviation University of China, Tianjin 300300, China;

2. College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China)

**Abstract:** In order to solve the handicap that the performance parameters of auxiliary power unit (APU) are hard to accurately forecast, a hybrid model of dual-stage attention mechanism (DAM) and CNN (convolutional neural network) - LSTM (long short-term memory) based on characteristic and temporal is proposed. A channel attention mechanism is obtained in the characteristic extraction phase and joined a temporal attention mechanism in the output phase. The ability for CNN is enhanced to extract important characteristics. The influence of key information from historical moments on the prediction output was enhanced simultaneously. In order to improve the accuracy of prediction, the improved pso algorithm was used to optimize the hyper parameters of the LSTM network. The proposed method has extremely low errors in the prediction of APU exhaust gas temperature (EGT) with multi-variable input and multi-step length, and the experimental results show that the accuracy of prediction is significantly improved.

**Keywords:** auxiliary power unit; performance parameter prediction; dual-stage attention mechanism; long and short-term memory network; exhaust gas temperature; particle swarm optimization

## 0 引言

飞机辅助动力装置<sup>[1]</sup> (APU, auxiliary power unit) 本质上是一个小型涡轮燃气发动机, 其结构简单, 重量小, 功率大, 主要用于辅助飞机主发动机启动并为机载用电设备供电<sup>[2]</sup>。APU 作为第二动力系统, 为飞机的安全性与乘客的舒适性提供保障, 是一个不可缺少的重要子系统。

排气温度<sup>[3]</sup> (EGT) 作为一个关键的气路性能参数, 可以直接反应 APU 的健康状态, 因此可以利用历史时刻 EGT 组成的时序数据预测未来一段时间 APU 的状态。APU 是一个复杂非线性系统, 多个部件共同作用使 EGT 呈现出非线性的时变规律, 故对 EGT 的预测要考虑到

APU 内各种性能参数。

国内外学者对排气温度预测展开了大量研究, 文献 [4] 利用排气温度与其他性能参数间的关系对 EGT 进行预测。文献 [5] 采用遗传算法优化支持向量机<sup>[6]</sup> 参数来提高预测精度。文献 [7] 采用优化后的最小二乘支持向量机模型预测 EGT。虽然上述的方法取得了较不错的预测效果, 但是对长时间序列, 进行多变量和多时间步预测时效果不是很好。近些年深度学习发展迅速, 在处理长时间多变量的时序数据有着明显的优势。长短期记忆网络<sup>[8-10]</sup> (LSTM) 是神经网络的一个变种, 解决了序列长时间被遗忘的问题。卷积神经网络<sup>[11-12]</sup> (CNN) 可以挖掘输入

收稿日期: 2021-12-03; 修回日期: 2022-01-06。

基金项目: 国家自然科学基金民航联合基金 (U1733119)。

作者简介: 王力 (1973-), 男, 重庆人, 教授, 硕士, 博士, 主要从事航空电子系统维修技术与方法方向的研究。

通讯作者: 马宪 (1995-), 男, 黑龙江牡丹江人, 硕士研究生, 主要从事 APU 故障预测方向的研究。

引用格式: 王力, 马宪. 基于 DAM 和 CNN-LSTM 的辅助动力装置性能参数预测模型[J]. 计算机测量与控制, 2022, 30(5): 55-61.

数据的特征,提取深层次的变量间局部依赖关系,一定程度上提高预测精度。在时序预测中引入注意力机制<sup>[13-14]</sup>可以为重要特征和时刻加权,加强重要信息对输出的影响,从而提高预测模型精度。粒子群算法<sup>[15]</sup>(PSO)可以对长短期记忆网络的超参数进行优化,提升预测模型的性能,从而提高预测精度。文献[16]提出基于 LSTM 的多变量时间序列预测模型。文献[17]提出了一种基于 CNN 和 LSTM 的混合模型预测短期电力负荷。文献[18]提出了一种基于长短期记忆神经网络和注意力机制的模型,对短期光伏功率进行预测。文献[19]提出一种基于自适应 PSO 的 LSTM 预测模型,对股票价格进行预测。

本文提出一种基于 DAM 和 CNN-LSTM 的模型对 APU 的气路参数 EGT 进行多变量预测<sup>[20]</sup>,首先在 CNN 中引入通道注意力机制,对各特征通道进行加权,突出重要特征的表达,提高了 CNN 提取重要特征的能力;然后在 LSTM 输出侧引入时序特征注意力机制,加强历史时刻重要信息对预测输出的影响。最后利用改进的粒子群算法对该模型的关键参数进行寻优,使用优化后的模型对 EGT 进行预测。此方法旨在使用深度神经网络的不同结构,更有效的处理 EGT 数据特征信息,达到准确预测 EGT 动态变化的目的。

## 1 深度学习理论

### 1.1 CNN 原理结构

CNN 是一种得到深度学习领域广泛使用的前馈神经网络。其卷积层通过局部连接和权值共享的方式有效提取原始数据的局部特征信息,保留了平面结构信息。由于时序序列具有局部相关性,故可以使用 1DCNN 处理其局部特征。CNN 结构如图 1 所示,主要由卷积层、池化层和全连接层组成。卷积层对输入数据进行特征提取,公式如下:

$$F = f(W \times X + b) \quad (1)$$

式中,  $W$  为卷积核权重矩阵,  $X$  为输入参数向量,  $b$  为偏置向量,  $f(\cdot)$  为激活函数。

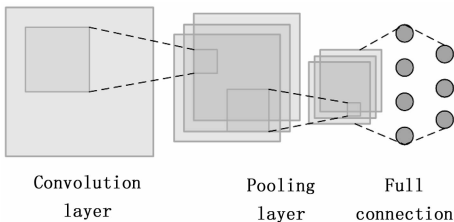


图 1 CNN 结构

### 1.2 LSTM 原理结构

LSTM 网络是在在循环神经网络 (RNN, recurrent neural network) 的基础上改进而来,可以弥补 RNN 中固有的无法处理长距离的依赖的缺陷。LSTM 在 RNN 的基础上增加了细胞状态和 3 个门,控制各个门的状态来对上一时刻旧信息和输入的新信息进行舍弃、保存和更新,最后得到的隐藏层输出由多个函数共同作用,避免了梯度消失问题。这种复杂的网络结构使 LSTM 比 RNN 能长期记住

更多过去的有用信息。LSTM 的网络结构如图 2 所示。

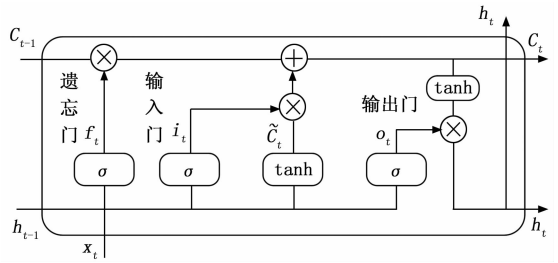


图 2 LSTM 结构图

图 2 中的一个 LSTM 存储单元由一个记忆细胞和输入门 ( $i_t$ )、输出门 ( $o_t$ )、遗忘门 ( $f_t$ ) 组成。遗忘门 ( $f_t$ ) 决定了上一时刻的细胞状态  $C_{t-1}$  中的哪些信息会传递到当前时刻细胞状态  $C_t$ 。输入门 ( $i_t$ ) 中的  $x_t$  分别经过  $\sigma$  和  $\tanh$  生成新的细胞状态候选值  $\tilde{C}_t$  再与遗忘门 ( $f_t$ ) 一起更新细胞状态。输出门 ( $o_t$ ) 与经过  $\tanh$  的更新后细胞状态  $C_t$  进行运算得到当前时刻隐层状态  $h_t$ 。LSTM 细胞状态更新方程为:

$$f_t = \sigma(W_f(h_{t-1}, x_t) + b_f) \quad (2)$$

$$i_t = \sigma(W_i(h_{t-1}, x_t) + b_i) \quad (3)$$

$$\tilde{C}_t = \tanh(W_c(h_{t-1}, x_t) + b_c) \quad (4)$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \quad (5)$$

$$o_t = \sigma(W_o(h_{t-1}, x_t) + b_o) \quad (6)$$

$$h_t = o_t \tanh(C_t) \quad (7)$$

式中,  $x_t$  为  $t$  时刻的输入;  $h_{t-1}$  和  $h_t$  为  $t-1$  时刻和  $t$  时刻隐藏层状态;  $f_t, i_t, o_t$  分别为遗忘门、输入门和输出门;  $C_{t-1}$  和  $C_t$  分别为  $t-1$  和  $t$  时刻的细胞状态;  $W_f, W_i, W_o, W_c$  分别为不同状态门的权重矩阵;  $b_f, b_i, b_o, b_c$  分别是不同状态门的偏移量;  $\sigma$  表示 sigmoid 激活函数。

### 1.3 注意力机制

注意力机制本质上是一种权重分配机制,受到人类视觉注意力机制的启发而提出。深度学习中的注意力机制的核心目标是令神经网络更加关注数据出现的趋势和变化,对关键信息赋予更多的概率权重以提高神经网络对关键信息的提取。在时序预测中,注意力机制既可以加入到 CNN 中,为重要的特征通道信息赋予更大的权重,从而降低无关特征信息对预测结果的影响,也可以作用到 LSTM 网络输出侧,加大历史时刻关键信息对当前输出的影响。

### 1.4 粒子群算法

粒子群算法是通过模拟鸟群觅食行为而发展起来的一种基于群体协作的随机搜索算法,被广泛的应用在各种优化问题。粒子群算法中每个粒子的初始位置都是随机的,其位置代表优化问题的解。所有的粒子都有一个由被优化的函数决定的适应值,适应值是评价粒子位置是否为最优解的一个评价指标。粒子仅具有位置和速度两个属性,进行迭代寻优的过程中会不断更新自身的位置和速度。在搜索空间内,每个粒子单独搜索全局最优解,个体间通过协作和信息共享的方式来调整位置与速度。粒子速度、位置更新公式为:

$$v_{i,j}^{t+1} = \omega v_{i,j}^t + c_1 r_1 (pbest_{i,j}^t - x_{i,j}^t) + c_2 r_2 (gbest_j^t - x_{i,j}^t) \quad (8)$$

$$x_{i,j}^{t+1} = x_{i,j}^t + v_{i,j}^t \quad (9)$$

式中,  $\omega$  为惯性因子;  $t$  为迭代次数;  $v_{i,j}^t, x_{i,j}^t, pbest_{i,j}^t, gbest_j^t$  分别为第  $t$  次迭代时, 第  $i$  个粒子在第  $j$  维上的速度、位置、个体最优值和群体最优值; 学习因子  $c_1$  和  $c_2$  是粒子根据自身经验和群体经验对速度进行调整的步长;  $r_1$  和  $r_2$  相互独立, 随机分布在  $[0, 1]$ 。

## 2 基于双阶段注意力机制的 CNN-LSTM 混合模型

### 2.1 模型描述

长序列数据包含许多潜在的规律信息, 传统的预测模型处理数据时, 无法提取数据特征信息之间的潜在规律, 导致了相对较大的预测误差。CNN 通过局部链接和权重共享的方式对原始数据进行高维特征映射, 能够挖掘出数据的特征信息。但通过卷积进行特征提取时, 产生的冗余特征会对预测精度产生影响, 对含有不同特征的通道使用注意力机制, 在训练过程中学习各个通道的权重, 给重要特征分配更大的权重, 从而获得预测模型更多的关注, 降低无关特征对预测结果的影响。LSTM 网络通过门控状态选择信息的保存与更新, 保存需要长时间记忆的关键信息, 遗忘无关信息。在处理具有长期依赖的时序数据时有较好的表现。由于排气温度的预测值受历史状态影响较大, 为了使预测值能自主处理历史状态信息, 加强对当前时刻输出影响大的历史时刻信息表达, 在 LSTM 网络的输出侧加入时序注意力机制, 能够得到更接近真实值的预测值。因此提出基于双阶段注意力机制的 CNN-LSTM 混合模型, 引入注意力机制加强模型对关键信息的关注, 使网络学习到更多关键信息的特征, 进而提高预测精度。

### 2.2 通道注意力机制

为更好的利用重要的特征通道信息, 在 CNN 卷积层后加入通道注意力模块, 突出重要特征通道的表达。具体细节如图 3 所示。

在通道注意力模块中, 输入  $F$  为经过 CNN 卷积层特征映射得到的特征图, 大小为  $H \times W \times C$ 。为汇总输入特征图  $F$  的空间信息, 对  $F$  分别进行全局最大池化  $P_{max}$  和全局平均池化  $P_{avg}$ , 得到输出最大池化特征  $F_{max}^c \in R^{1 \times 1 \times C}$  和平均池化特征  $F_{avg}^c \in R^{1 \times 1 \times C}$ , 然后将  $F_{max}^c$  和  $F_{avg}^c$  传入一个共享的多层感知机 (MLP, multiple layer perception) 来计算通道注意力矩阵  $M_c(F)$ 。该多层感知机包含两个全连接层和一个 relu 激活函数,  $F_{max}^c$  和  $F_{avg}^c$  先经过第一个全连接层, 其输出通道

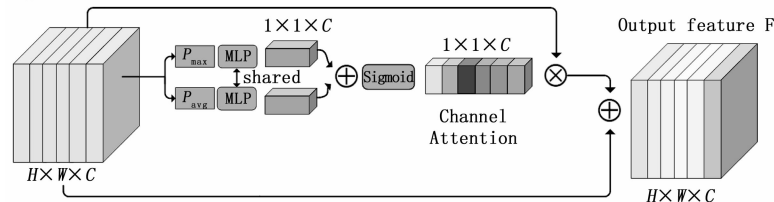


图 3 通道注意力机制

数为  $C/r$  ( $r$  为缩减率), 再经过 relu 激活函数, 后通过第二个全连接层, 输出通道数为  $C$ , 最后将共享的多层感知机的输出相加输入到 sigmoid 激活函数, 从而得到通道注意力矩阵  $M_c(F)$ , 计算公式为:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_1(\delta W_0(F_{avg}^c)) + W_1(\delta W_0(F_{max}^c))) \quad (10)$$

其中:  $\sigma$  代表 sigmoid 激活函数,  $\delta$  代表 relu 激活函数,  $W_1 \in R^{c/r \times c}$  和  $W_0 \in R^{c \times c/r}$  分别为全连接层的两个权重矩阵。

最后, 将通道注意力矩阵  $M_c(F)$  与输入特征图  $F$  相乘, 并且通过残差的方式与原始的输入特征图相加, 得到该通道注意力模块的最后输出。最终, 每个特征通道都乘上了学习到的相应权重, 增强了网络对某些重要通道信息的注意力, 能够更好的提取数据的时空特征。

### 2.3 时序注意力机制

在 EGT 预测中, 当前时刻的预测值与历史时刻预测值有较大联系, 但每个历史时刻信息对当前输出的影响程度不同, 为了突出重要历史时刻信息的影响, 在 LSTM 网络的输出侧加入时序注意力模块为重要历史信息分配更大的权重。具体细节如图 4 所示。

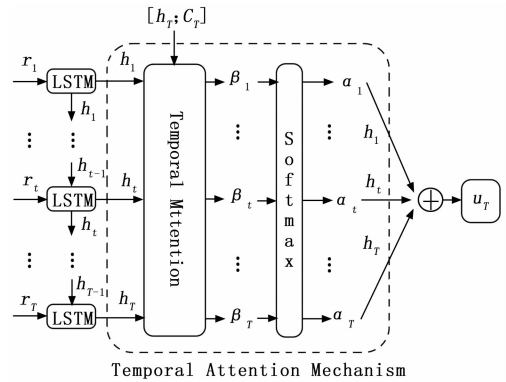


图 4 时序注意力机制

在  $t$  时刻, CNN 提取的特征参数向量  $r_t$  与 LSTM 上一时刻隐层状态  $h_{t-1}$  一同输入到 LSTM 网络中, 得到  $t$  时刻的输出  $h_t$ , 表示为:

$$h_t = f_{lstm}(h_{t-1}, r_t) \quad (11)$$

式中,  $f_{lstm}$  表示 LSTM 网络单元。

时序注意力机制的输入为经过 LSTM 网络处理过的特征矩阵  $H$ , 随后时序注意力机制会对历史时刻隐层状态信息分配权重系数, 其计算公式如下:

$$\beta_t = V_h^T \tanh(W_h [h_T; C_T] + U_h h_t + b_h) \quad (12)$$

$$\alpha_t = \text{softmax}(\beta_t) = \frac{\exp(\beta_t)}{\sum_{i=1}^T \exp(\beta_i)} \quad (13)$$

式中,  $[h_T; C_T]$  是 LSTM 单元隐层状态与细胞状态的拼接;  $\beta_t$  是 LSTM 单元隐层状态  $h_t$  和  $[h_T; C_T]$  的相关性分数;  $V_h, W_h, U_h$  为时序注意力相应的权重矩阵;  $b_h$  为偏差项;  $\alpha_t$  为经过 softmax 函数归一化的注意力权重系数。

$$u_T = \sum_{j=1}^T \alpha_j h_j \tag{14}$$

$\alpha_i$  对  $t$  时刻隐层输出  $h_i$  与当前时刻  $[h_T; C_T]$  的相关性进行了量化, 提高与当前时刻输出关联密切的历史信息的影响, 最后各时刻的隐层输出加权求和得到了时序注意力机制的输出  $u_T$ ,  $u_T$  经过全连接层进行维度变换得到最终的 EGT 预测值。

### 2.4 粒子群算法改进

传统的粒子群算法粒子速度的改变总是保持同一水平, 导致了算法收敛速度慢并且搜索全局性差, 易陷入局部最优值。通过改进惯性因子来调节粒子的全局和局部寻优能力。改进后的惯性因子  $\omega$  公式为:

$$\omega_t = \omega_{start} - (\omega_{start} - \omega_{end}) \times (\frac{t}{T_{max}})^2 \tag{15}$$

式中,  $\omega_{start}$  和  $\omega_{end}$  分别为惯性因子的最大值与最小值, 一般设  $\omega_{start}=0.9$ ,  $\omega_{end}=0.4$ ,  $t$  为当前迭代次数,  $T_{max}$  为最大迭代数。

$$\frac{d\omega}{dt} = -\frac{2t(\omega_{start} - \omega_{end})}{T_{max}^2} \tag{16}$$

惯性因子代表粒子对当前速度的继承程度, 上式为惯性因子的微分, 代表惯性因子的变化率。在迭代次数  $t$  较小时,  $\omega$  接近最大值, 粒子能快速分布到整个搜索空间, 以找出最优值的大概范围; 随着迭代次数的增加,  $\omega$  下降的越来越快, 粒子速度变小确保了局部搜索的精度。使用非线性惯性因子调节粒子速度, 使搜索的全局性与局部性达到动态平衡, 提高了粒子群算法寻优的性能。

改进 PSO 算法对 LSTM 网络超参数寻优的具体步骤如下:

- 1) 输入经过预处理的训练集数据;
- 2) 设置初始粒子种群大小、迭代次数、学习因子等相关参数;
- 3) 利用改进 PSO 算法对 LSTM 网络的学习率 lr, 步长 step 进行寻优, 并确定搜索范围;
- 4) 粒子的适应度计算与对比, 以 LSTM 的预测平均绝对百分比误差作为目标函数, 计算各粒子的适应度值并进行比较, 找到粒子的全局最优值 gbest;
- 5) 利用式 (8) 和式 (9) 对粒子的速度和位置进行更新;
- 6) 判断是否达到最大迭代次数或最小误差。若两个条件均不满足则返回步骤 4);
- 7) 输出 LSTM 网络的最优学习率与步长。

在改进的粒子群算法中设置最大迭代次数为 100, 种群规模为 20, 学习因子  $c_1 = 2, c_2 = 2$ , 惯性因子最大值与最小值分别为 0.9, 0.4。LSTM 网络训练过程中使用 adam 优化器, 损失函数采用均方误差 (MSE), 学习率取值范围为  $[0.001, 0.01]$ , 步长的取值范围为  $[1, 10]$ , 改进 PSO 搜索到的最优组合参数学习率为 0.003, 步长为 6。

### 2.5 基于双阶段注意力机制的 CNN-LSTM 混合模型

整个预测模型结构如图 5 所示, 原始数据经过输入层

进行预处理, 然后输入到 1D 卷积层进行特征提取, 利用通道注意力机制为各特征通道分配权重, 再经过池化层、全连接层的处理, 信息进入 LSTM 网络, 其隐层状态作为时序注意力机制的输入, 时序注意力机制实现了选择不同时刻特征的重要性, 加权后的特征信息经过全连接层维度变换后得到最后的预测值。

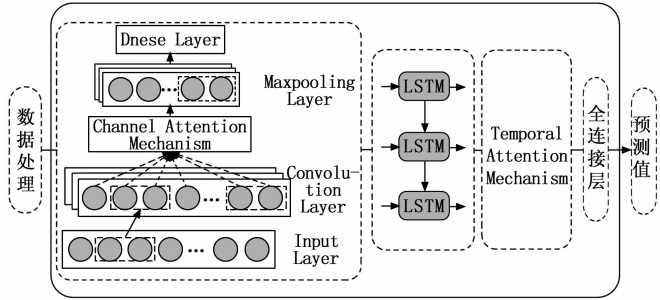


图 5 DAM 和 CNN-LSTM 模型结构

模型中的每层描述如下:

1) 输入层: 对原始输入数据进行预处理, 使输入数据满足预测模型的要求。

2) 引入通道注意力机制的 CNN 层: 对经过预处理的数据进行特征提取。CNN 由卷积层、池化层和全连接层组成, APU 性能参数为时序数据, 具有局部相关性, 故将卷积层设计为一维卷积便于处理其局部特征, 该卷积层包含 64 个  $1 \times 1$  的卷积核, 其激活函数选取 Relu 函数, 相比 sigmoid 函数和 tanh 函数, Relu 函数缓解了梯度消失问题并加快了模型收敛速度。它将小于 0 的输入直接输出为 0, 使得神经网络中的神经元具有了稀疏激活性, 能够一定程度上防止过拟合。通道注意力模块为经过卷积层的特征分配权重, 突出重要特征信息的表达。池化层选择最大池化, 能够保留更多的特征信息。为了避免过拟合, 在池化层和全连接层中增加 Dropout 层, 以一定概率随机断开神经元的连接, 最后经过全连接层转换数据维度将其输出。

3) 引入时序注意力机制的 LSTM 层: LSTM 网络对 CNN 层提取的特征向量进行学习, 挖掘出具有长期依赖性的特征。LSTM 网络学习率为 0.003, 隐层神经元节点数为 128, 步长为 6, batch\_size 为 64。时序注意力机制为 LSTM 网络的输出分配权重。时序注意力机制的输入为 LSTM 输出的特征向量  $H = [h_1, h_2, \dots, h_t]$ 。

4) 输出层: 其输入是时序注意力机制模块的输出。由全连接层输出最后的预测值, 激活函数选取 sigmoid 函数。

## 3 实验结果与分析

### 3.1 数据预处理与评价指标

本文实验数据采用航空公司某型飞机快速存取记录器 (QAR, quick access recorder) 系统采集的 APU 数据, 其中包括滑油温度 ( $T_o$ )、引气压力 ( $P_T$ )、引气流量 ( $W_B$ )、发电机负载 ( $G_L$ )、APU 转速 ( $N_A$ )、EGT, 共 3 300 组数据。

EGT 的变化受多个 APU 性能参数影响, 但过多的输入参数会造成预测模型网络复杂化, 还会使模型学习到与

输出无关的特征, 从而影响预测精度。相关性是衡量不同参数间变化一致程度的指标, 选择与 EGT 相关的性能参数作为输入数据可以保存大量特征信息并降低计算量。因此采用数据分析软件 SPSS 计算各参数间的 Pearson 相关系数, 如表 1 所示。

表 1 不同参数的相关系数比较

Corr	EGT	$W_B$	$N_A$	$G_L$	$P_T$	$T_O$
EGT	1.00	0.93	0.87	0.83	-0.63	0.28
$W_B$	0.93	1.00	0.91	0.83	-0.85	0.30
$N_A$	0.87	0.91	1.00	0.51	-0.64	0.58
$G_L$	0.83	0.83	0.51	1.00	-0.62	-0.20
$P_T$	-0.63	-0.85	-0.64	-0.62	1.00	-0.27
$T_O$	0.28	0.30	0.58	-0.20	-0.27	1.00

通常情况下, 相关系数绝对值大于 0.8, 认为参数间高度线性相关。在 0.5~0.8 之间, 认为参数间具有显著线性相关性。在 0.4 以下, 一般认为有极弱线性相关性或不存在线性相关。在进行多步预测时, 具有高度相关性的输入数据有助于加强神经网络模型的学习过程, 提高预测结果的准确率。

根据表 1 不同参数相关系数选择模型输入参数为 APU 转速  $N_A$ 、引气流量  $W_B$ 、发电机负载  $G_L$ 。由于 APU 排气温度与相关的特征参数在 APU 工作过程中波动范围较大, 且各个特征参数的量纲级别也不同, 为了防止各特征参数数值差异过大造成奇异解, 加快模型收敛速度, 需要对原始数据进行归一化以适应模型训练。

$$x = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (17)$$

式中,  $x_{\max}$  为输入数据  $x_i$  的最大值,  $x_{\min}$  为最小值。

对各特征参数数据进行 min-max 标准化, 线性变换到 [0, 1] 区间, 适应 sigmoid 和 tanh 激活函数。在模型的输出阶段, 需要对数据进行反归一化, 使得输入数据和输出数据保持一致。

本文选取 3 种误差评价指标衡量模型的预测精度, 分别是 MAE (mean absolute error)、MAPE (mean absolute percentage error) 与 RMSE (root mean square error), 计算公式如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \tilde{y}_i| \quad (18)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2} \quad (19)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \tilde{y}_i|}{y_i} \times 100\% \quad (20)$$

式中,  $y_i$  和  $\tilde{y}_i$  分别为第  $i$  个样本的实际值与预测值,  $n$  为数据个数。MAE 可以评估模型的预测精度。RMSE 能够反映预测的稳定性。MAPE 通过百分比直观表现混合模型预测性能优劣。以上 3 个评价指标越小, 模型性能越好。

### 3.2 模型收敛性

设置最优模型参数后, 需要对训练过程中 DAM 和

CNN-LSTM 混合模型的收敛性进行验证。计算该模型训练过程中的损失值, 并与 CNN-CAM-LSTM 模型、CNN-LSTM-TAM 模型、CNN-LSTM 模型进行比较。各模型收敛性如图 6 所示。

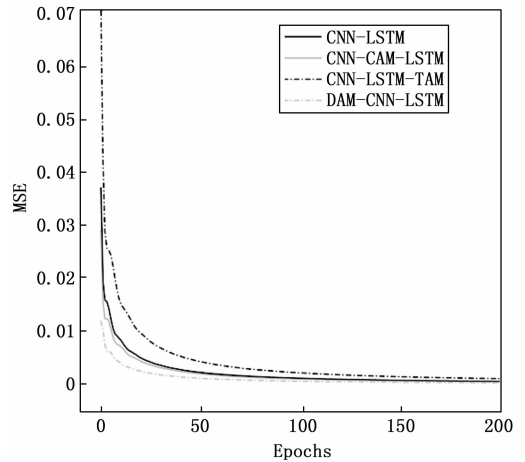


图 6 不同模型进行 EGT 预测的损失收敛

实验中各模型参数均相同, 采用均方误差 MSE (mean square error) 作为模型训练的损失函数。由图 6 可以看出各模型在经过 200 次迭代后其损失函数 MSE 趋于平稳。在整个训练过程中本文所提模型的损失函数 MSE 的值均小于另外 3 种模型, 具有更好的收敛效果。

### 3.3 实验结果分析

为验证双阶段注意力机制对特征提取和时序依赖关系的有效性, 对样本数据采用本文提出的 DAM 和 CNN-LSTM 混合模型、CNN-CAM-LSTM 模型、CNN-LSTM-TAM 模型、CNN-LSTM 模型进行排气温度预测。上述模型中使用的 LSTM 网络均采用相同的超参数 (学习率为 0.003, 隐层神经节点数为 128, 步长为 6, batch\_size 为 64); 1DCNN 的卷积核个数为 64, 卷积核大小为 1, 最大池化层为 2, dropout 为 0.3。数据集前 90% 作为训练集, 后 10% 作为测试集。在模型训练过程中选择 Adam 算法优化模型参数, Adam 算法吸收了 AdaGrad 和 RMSProp 两种梯度下降算法的优点, 既能适应稀疏梯度, 又缓解了梯度震荡问题。文中预测模型的代码都是用 Python 编写, 使用 Pytorch 深度学习框架进行开发。采用本文选取的 3 种误差评价指标来评价不同模型的预测性能与精度, 实验对比结果如表 2 所示。

表 2 不同模型预测精度比较

Models	MAE	RMSE	MAPE
CNN-LSTM	13.36	22.82	3.59
CNN-CAM-LSTM	7.78	15.93	2.04
CNN-LSTM-TAM	4.93	11.14	1.43
DAM-CNN-LSTM	3.87	10.07	1.11

根据实验结果分析, 加入双阶段注意力机制对预测效果的提升显著, MAE、RMSE、MAPE 均有明显下降。为

验证通道注意力机制和时序注意力机制对模型性能的提升效果，将两种注意力机制单独加入到 CNN-LSTM 模型 (CAM 为通道注意力机制；TAM 为时序注意力机制)。由表 2 可以看出，在输出侧加入时序注意力机制对模型预测性能的提升更为明显。

本文所提模型比 CNN-LSTM 模型，CNN-CAM-LSTM 模型，CNN-LSTM-TAM 模型的 MAE 分别降低了 71.03%、50.26%、21.50%；RMSE 分别降低了 55.87%、36.79%、9.61%；MAPE 分别降低了 2.48%、0.93%、0.32%。本文算法在 3 种误差评价指标上都有明显的降低，证明了该模型在多步多变量排气温度预测的可行性。

各模型在测试集上的预测输出曲线对比如图 7 所示。

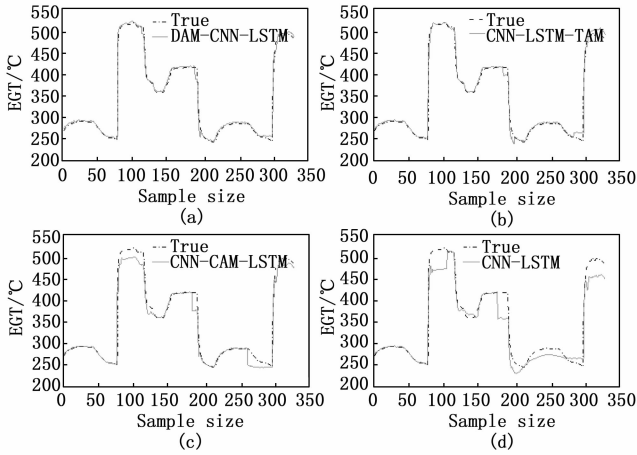


图 7 不同模型的 EGT 预测温度比较

由图 7 (d) 可以看出 CNN-LSTM 模型在高低峰值和大幅波动区间的预测输出值与真实值拟合程度较低，难以反映 EGT 真实值的变化趋势。图 7 (c) 为 CNN-CAM-LSTM 模型的预测值，该模型在特征提取阶段加入通道注意力机制对特征通道分配权重，加强了模型对重要特征通道的关注，在高低峰值区间的预测效果有一定的提升，但在高峰阶段的预测值与真实值仍有较大误差，无法准确的预测出排气温度的峰值，从而忽视因排气温度超温而引起的故障。CNN-LSTM-TAM 模型在 LSTM 输出侧加入时序注意力机制加强了历史关键时刻信息对当前输出的影响，由图 7 (b) 可以看到整个区间的预测精度均有明显提升，在高低峰值和大幅波动区间的预测输出值与真实值拟合度较高。由图 7 (a) 可以看出本文提出的 DAM 和 CNN-LSTM 混合模型在整个区间的预测输出与 EGT 真实值拟合程度很高，准确的捕捉到了 EGT 变化规律。其他几种预测模型在部分区间虽然也有较好的预测效果，但在高低峰值和大幅波动区间的预测精度均低于本文所提模型。

总体而言，提出的 DAM-CNN-LSTM 方法能够有效地学习参数数据的内部变化规律，可以用于未来排气温度预测。该预测方法可为未来 APU 健康状态预测提供预警作用，预测未来 EGT，并结合其他 APU 性能参数，查看是否在合理的范围内，及时排查潜在故障原因保障 APU 正常运行。

### 3.4 不同步长效果验证

为了进一步验证本文方法的泛化性和稳定性，使用不同步长进行 EGT 预测，采用上文提出的 3 种误差评价指标对 4 种模型的预测性能进行评估与分析。

图 8 和图 9 分别为步长为 10 和 15 时，4 种模型的 MAE、RMSE、MAPE 值。从图中可以看出随着步长的增加，3 种误差评价指标值均升高，说明步长的增加导致各模型的泛化性和稳定性变差。与 6 步预测结果相比，各模型在进行 10 步预测时，RMSE 分别增长了 32.11%、35.10%、33.67%、35.02%，在进行 15 步预测时，RMSE 分别增加了 37.33%、39.01%、42.37%、41.31%。本文方法的 RMSE 增幅还是较高，但与其他方法相比，RMSE 增长率最低，说明在步长增加的情况下，预测精度下降的更缓慢，MAE 和 MAPE 也明显低于其他方法。总体而言，本文提出的预测模型在十步和十五步预测中，3 种误差评价指标是最低的，表明其预测性能要优于其他预测模型。

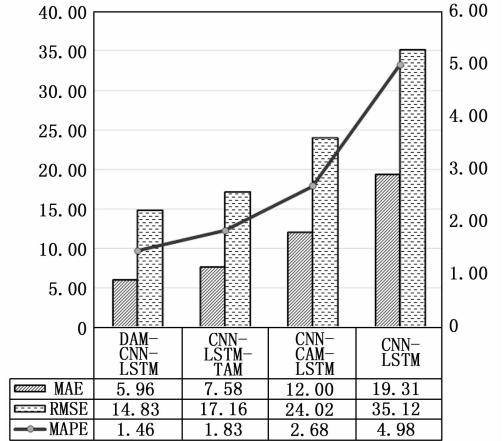


图 8 十步预测不同模型的评价指标值

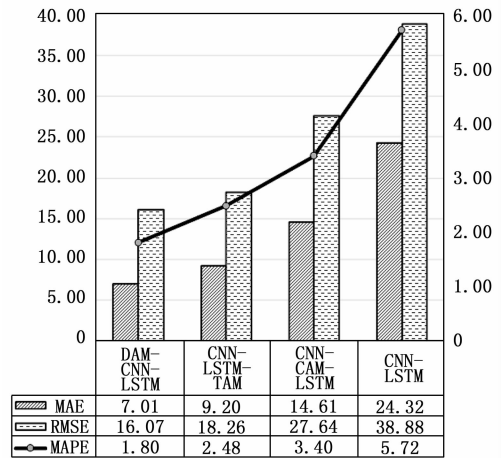


图 9 十五步预测不同模型的评价指标值

步长的增加导致了各预测模型的预测性能下降，计算单位步长增加时各预测模型的误差平均增长率，增长率越低，说明预测模型性能衰减的越慢，各预测模型误差平均增长率如表 3 所示。

表 3 不同预测模型平均误差增长率

	MAE/%	RMSE/%	MAPE/%
DAM-CNN-LSTM	9.02	6.62	6.90
CNN-LSTM-TAM	9.62	7.10	8.15
CNN-CAM-LSTM	9.75	8.17	7.41
CNN-LSTM	9.12	7.82	7.60

由表 3 可以看出, 本文所提预测模型在 MAE、RMSE、MAPE 的平均增长率分别为 9.02%、6.62%、6.90%, 均低于其他预测模型。其中 MAE 是衡量预测模型精度的一个指标; RMSE 主要反映模型的稳定性; MAPE 百分比表现模型预测性能, 该预测模型在引入双阶段注意力机制后, 在多步长预测中精度和稳定性均优于 CNN-LSTM-TAM 模型、CNN-CAM-LSTM 模型、CNN-LSTM 模型, 相比其他模型更适用于多步长预测。

DAM-CNN-LSTM 预测模型利用通道注意力机制和时序注意力机制充分学习了特征参数间的内部相关性, 可较为准确地对未来 APU 的 EGT 进行预测。

#### 4 结束语

本文提出一种基于 DAM 和 CNN-LSTM 的混合模型用来提高 APU 排气温度短期预测的精度与稳定性, 得到了以下结论:

1) 多维冗余的无关输入数据会导致模型结构复杂且预测精度下降。为提升预测模型的效率与性能, 对输入数据进行相关性分析, 选择与 EGT 相关性较强的性能参数参与预测, 提高数据可靠性。

2) 在 CNN 中加入通道注意力机制学习各个特征通道的权重, 为重要特征分配更大的权重, 从而获得预测模型更多的关注, 降低无关特征对预测结果的影响, 提高了 CNN 提取重要特征的能力。在 LSTM 网络输出阶段加入时序注意力机制可以捕捉时序数据间的依赖关系, 加强历史关键信息对当前输出的影响, 提高模型对长时间序列的预测性能。

3) 非线性惯性因子提高了 PSO 的寻优能力, 避免算法陷入局部最优值并加快收敛速度。利用改进的粒子群算法优化 LSTM 网络的学习率和步长, 找到最优组合参数。

4) 双阶段注意力机制的加入使模型对重要特征及重要历史信息投入更多的关注, 通过实验验证了注意力机制对多变量多步 EGT 预测模型性能的提升。当预测步长增加时, 虽然 3 种误差评价指标均有所上升, 但上升幅度较低, 预测精度仍高于其他 3 种模型, 证明了本文所提预测模型的稳定性。

#### 参考文献:

[1] GORINEVSKY D, DITTMAR K, MYLARASWAMY D, et al. Model-based diagnostics for an aircraft auxiliary power unit [C] //Proceedings of the international Conference on control applications. IEEE, 215-220.

[2] 董 平. APU 系统的状态监测与维修决策研究 [D]. 南京: 南京航空航天大学, 2018.

[3] ILBAS M, TURKMEN M. Estimation of exhaust gas temperature using artificial neural network in turbofan engines [J]. Journal of Thermal Science and Technology, 2012, 32 (2): 11-18.

[4] YILMAZ I. Evaluation of the relationship between exhaust gas temperature and operational parameters in CFM56-7B engines [J]. Proceedings of the Institution of Mechanical Engineers, 2009, 223 (4): 433-440.

[5] 李艳军, 张 建, 曹愈远, 等. 基于模糊信息粒化和优化 SVM 的航空发动机性能趋势预测 [J]. 航空动力学报, 2017, 32 (12): 3022-3030.

[6] 蔡坤烽, 蔡 景, 周 迪, 等. 基于 SVM 方法的 APU 故障预测方法 [J]. 南京航空航天大学学报, 2019, 51 (4): 466-473.

[7] 张 弦, 王宏力. 嵌入维数自适应最小二乘支持向量机状态时间序列预测方法 [J]. 航空学报, 2010, 31 (12): 2309-2314.

[8] 王 鑫, 吴 际, 刘 超, 等. 基于 LSTM 循环神经网络的故障时间序列预测 [J]. 北京航空航天大学学报, 2018, 44 (4): 772-784.

[9] KONG W, DONG Z Y, JIA Y, et al. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network [J]. IEEE Trans on Smart Grid (S1949-3053), 2019, 1 (10): 841-851.

[10] ZHAO Z, CHEN W, WU X, et al. LSTM network: a deep learning approach for short-term traffic forecast [J]. IET Intelligent Transport Systems, 2017, 11 (2): 68-75.

[11] KIM T Y, CHO S B. Predicting residential energy consumption using CNN-LSTM neural networks [J]. Energy, 2019, 182: 72-81.

[12] Rahim Barzegar, Mohammad Taghi Aalami, et al. Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model [J]. Stochastic Environmental Research and Risk Assessment, 2020, 34: 415-433.

[13] SHIH S Y, SUN F K, LEE H. Temporal pattern attention for multivariate time series forecasting [J]. Machine Learning, 2019, 108 (8): 1421-1441.

[14] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [C] //Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[15] 康岚兰, 董文永, 宋婉娟, 等. 无惯性自适应精英变异反向粒子群优化算法 [J]. 通信学报, 2017, 38 (8): 66-78.

[16] 杜维仲, 王 硕. 基于 LSTM 的辅助动力装置系统辨识与仿真 [J]. 计算机测量与控制, 2020, 28 (2): 157-161.

[17] 陆继翔, 张琪培, 杨志宏, 等. 基于 CNN-LSTM 混合神经网络模型的短期负荷预测方法 [J]. 电力系统自动化, 2019, 43 (8): 131-137.

[18] ZHOU H, ZHANG Y, YANG L, et al. Short-Term Photovoltaic Power Forecasting Based on Long Short Term Memory Neural Network and Attention Mechanism [J]. IEEE Access, Volume 7, 2019, 78063-78074.

[19] 宋 刚, 张云峰, 包芳勋, 等. 基于粒子群优化 LSTM 的股票预测模型 [J]. 北京航空航天大学学报, 2019, 45 (12): 2533-2542.

[20] 杨 柯, 范世东. 基于长短期记忆网络时序数据趋势预测及应用 [J]. 推进技术, 2021, 42 (3): 675-682.