

# 长短周期记忆网络微调的氧气浓缩器寿命预测模型迁移\*

崔展博, 景博, 焦晓璇, 潘晋新, 王生龙  
(空军工程大学航空工程学院, 陕西西安 710038)

**摘要:**针对寿命预测模型迁移问题,提出了一种长短周期记忆网络微调(long short-term memory fine tune, LSTM-fine-tune)的迁移模型,利用理想条件下的试验数据对模型进行训练。在迁移过程中,对部分LSTM网络层进行冻结,利用实际服役环境下的数据对网络其他部分进行修正。为验证模型的泛化能力,采用不同相位与幅值的正弦函数生成数据,通过学习数据获取正弦函数的经验知识,并应用至其他正弦函数的回归,结果表明LSTM-fine-tune模型能够快速拟合,平均均方误差仅为1.033 5,明显低于直接预测误差1.536 8。为通过实际监测数据检验本方法泛化能力,分别获取了试验条件下与实际服役环境下氧气浓缩器的数据,对模型的泛化能力进行验证。结果表明,迁移后训练集预测精度提高了43.0%,测试集预测精度提高了20.2%。

**关键词:**迁移学习;寿命预测;氧气浓缩器;灰狼优化器

**中图分类号:**TP391.4 **文献标志码:**A **文章编号:**1001-2486(2023)04-243-10

## Transfer of oxygen concentrator life prediction model based on LSTM-fine-tune

CUI Zhanbo, JING Bo, JIAO Xiaoxuan, PAN Jinxin, WANG Shenglong

(Aviation Engineering School, Air Force Engineering University, Xi'an 710038, China)

**Abstract:** To transfer the life prediction model, an LSTM-fine-tune(long short-term memory fine tune) model was proposed. The model was trained by using experimental data under ideal conditions. During the transfer process, part of the LSTM network layer was frozen, and other parts of the network were modified by using data in actual service environment. In order to verify the generalization ability of the model, sinusoidal functions with different phases and amplitudes to generate data were used, obtained the knowledge of the sinusoidal function, and applied it to the regression of other sinusoidal functions. The results show that the LSTM-fine-tune model can be fitted quickly, and the average mean square error is only 1.033 5. It is significantly lower than the direct prediction error 1.536 8. In order to test the generalization ability of this method through actual monitoring data, the data of oxygen concentrators under test conditions and actual service environment respectively is obtained, verifies the generalization ability of the model. The results show that the prediction accuracy of the training set is improved by 43.0% and that of the test set is improved by 20.2%.

**Keywords:** transfer learning; life prediction; oxygen concentrator; grey wolf optimizer

建立装备的退化模型是装备故障预测与健康管理技术的关键<sup>[1-3]</sup>。近年来,大量时间序列预测方法在退化预测模型中得到应用。从相关文献来看,大的研究方向集中在数学模型、随机过程、统计模型、机器学习等的应用与创新,一系列算法及其变体例如受限玻尔兹曼机、循环神经网络、长短周期记忆网络、生成对抗网络等,为寿命预测提供了众多实现途径<sup>[4-5]</sup>。

从信息获取的角度来看,数据驱动方法是从装备历史数据中获取相关经验知识,模型的预测

能力与数据的完备程度紧密相关。然而,在实际服役条件下,存在大量非理想、非平衡数据,存在数据样本少、部分样本缺失、强噪声等特点。对于这样的数据,其中蕴含的装备信息知识有限,如果直接采用传统的数据驱动方法进行建模,预测模型难以达到预期效果。

实际服役环境下的数据建模是工程应用中的一大难题,在研究中常常在试验环境下获取较为理想的数据,基于数据进行建模。但试验环境下的数据与实际服役环境下的数据存在较大差异,

\* 收稿日期:2022-05-10

基金项目:河北省“三三三人才工程”资助项目(B20221011);陕西省自然科学基金资助项目(2017JQ6034)

作者简介:崔展博(1987-),男,陕西西安人,博士研究生,E-mail:1091313541@qq.com;

焦晓璇(通信作者),男,山西运城人,副教授,博士,E-mail:564325155@qq.com

在应用上存在不适配的问题。如何有效开展模型迁移,是制约预测模型实际应用的难题。

迁移学习是解决模型领域泛化的一种途径,众多数据迁移、特征迁移、模型迁移等方法针对不同情况下的迁移取得了一定成果<sup>[6-9]</sup>。Sun 等<sup>[10]</sup>基于稀疏自动编码器,采用权重转移、特征迁移和权值更新的策略对稀疏自编码器生成的模型进行迁移,并将方法运用在刀具寿命预测迁移上;Mao 等<sup>[11]</sup>融合了转移成分分析与压缩去噪自编码器,对轴承开展特征迁移研究,并利用最小二乘支持向量机对轴承剩余寿命进行预测;Ragab 等<sup>[12]</sup>提出了一种无监督对抗域自适应模型开展特征迁移,对信息进行约简并将模型用于涡扇发动机的剩余寿命预测。

然而,对于寿命预测的模型泛化而言,由于装备时间序列退化数据存在连续性、缓慢变化的特点,难以有效提取经验知识,造成模型迁移的困难。因此现有文献中对寿命预测模型迁移研究较少,多数的寿命预测迁移还是基于特征迁移展开的<sup>[13-15]</sup>。然而,装备的退化数据是一种时间序列数据,仅采用特征迁移难以发挥循环神经网络等时间序列处理方法的优点,如何开展寿命预测模型迁移泛化是制约预测模型应用的技术难题。

模型微调 (fine-tune) 是迁移学习的方法之一,通过理想数据条件下获取的数据建立的模型,在数据分布相近的非理想数据条件下对模型进行微调,从而达到知识迁移的目的,以提升下游目标任务<sup>[16-18]</sup>。如文本分类以及机器翻译等相关自然语言处理任务的性能,在基于转换器的双向编码器表征 (bidirectional encoder representation from transformers, BERT)、生成预训练转换器 - 3 (generative pre-trained transformer 3, GPT3)<sup>[19-20]</sup> 等著名预训练模型证明了这一范式的强大。模型微调的知识迁移能力可以从以下三个方面进行解释:一是预训练的过程可以看作是特征学习,在较为完备的数据中学习到的经验知识,对于相近的数据也能起到辅助作用;二是通过预训练得到了网络的初始参数,相比于随机初始化网络参数,模型微调将参数空间限制在了一个更小的范围内,这样可以加快对目标任务的收敛,同时对于小样本数据,模型微调过拟合现象更少、模型泛化性能更强;三是模型微调可被视为小数据梯度方向的限制,相当于是对参数的正则化,可以缓解网络在小数据上过拟合的现象<sup>[21]</sup>。

针对时间序列数据的预测问题,本文提出了一种长短周期记忆网络微调 (long short-term

memory fine tune, LSTM-fine-tune) 迁移模型,利用理想条件下的试验数据对模型进行训练,在迁移过程中,对部分 LSTM 网络层进行冻结,利用实际服役环境下的数据对网络其他部分进行修正。为确定 LSTM 结构层数与冻结层数,本文采用灰狼算法搜寻最优超参数,使目标域数据预测精度趋于最优。为验证模型的有效性,本文分别获取了试验条件下与实际服役环境下氧气浓缩器的数据,对模型的泛化能力进行验证。

### 1 基础方法

#### 1.1 迁移学习

假设针对某些高可靠产品,故障样本少,通过有限的试验仅能得到样本数据集  $\{x_i^t, y_i^t\}_{i=1}^{n_t}$ , 其中包含  $n_t$  个样本,对应于迁移学习概念中的目标域  $D^t = \{x^t, P(x^t)\}$ , 目标任务  $T^t = \{y^t, f^t(\cdot)\}$ 。  $P(x^t)$  表示  $x^t$  的边缘概率分布,  $y^t$  表示目标域的标记空间,  $f^t(\cdot)$  表示将目标域中的样本  $x^t$  映射到标记空间  $y^t$  的函数。

另存在一个具有丰富样本的数据集  $\{x_i^s, y_i^s\}_{i=1}^{n_s}$ , 其中包含  $n_s$  个样本,  $n_s \gg n_t$ , 对应于迁移学习概念中的源域  $D^s = \{x^s, P(x^s)\}$ , 源任务  $T^s = \{y^s, f^s(\cdot)\}$ 。  $P(x^s)$  表示  $x^s$  的边缘概率分布,  $y^s$  表示源域的标记空间,  $f^s(\cdot)$  表示将源域中的样本  $x^s$  映射到标记空间  $y^s$  的函数<sup>[22]</sup>。

迁移学习的目标是通过源域装备的监测数据样本训练,建立由源域监测数据的样本空间至标签的非线性映射关系,作为装备故障诊断及退化经验知识<sup>[23]</sup>。

#### 1.2 长短周期记忆网络

LSTM 神经网络结构的输入输出如图 1 所示,  $CELL_1, CELL_2, \dots, CELL_n$  表示网络不同时间点的网络状态,其中  $n$  为时间索引,因此网络状态随时间不断变化。  $x_1, x_2, \dots, x_n$  表示时间长度为  $n$  的序列输入,  $y_1, y_2, \dots, y_n$  分别为模型在不同时间点的输出。

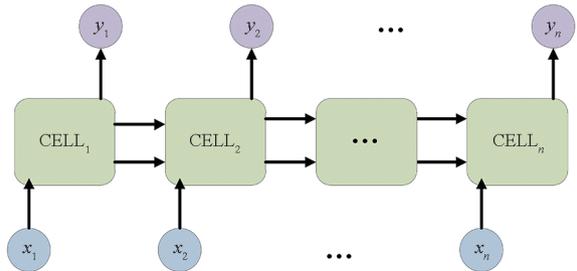


图 1 LSTM 网络结构示意图

Fig. 1 Schematic diagram of LSTM network structure

LSTM 单元的内部结构如图 2 所示, LSTM CELL 包含遗忘门、输入门、输出门 3 个门,在不同的时间点,该状态下的 CELL 接收到上一个时间状态的  $h_{(t-1)}$  与  $C_{(t-1)}$ 。输入后,分为 4 个流向,分别在 4 个全连接网络与激活函数的运算后,形成了 4 路输出,分别为  $f_{(t)}$ 、 $a_{(t)}$ 、 $i_{(t)}$ 、 $o_{(t)}$ 。其中  $f_{(t)}$  与  $C_{(t-1)}$  相乘,其操作称为遗忘门;  $i_{(t)}$  与  $a_{(t)}$  相

乘,其操作称为输入门。遗忘门与输入门的相加,输出的结果作为下一个时间的  $C_{(t)}$  的输入,同时,该结果在激活函数运算后与  $o_{(t)}$  相乘,得到的结果作为这一时间点的输出,并作为下一个时间点的  $h_{(t)}$  [24-25]。

## 2 LSTM-fine-tune 寿命预测领域泛化模型

### 2.1 网络结构

LSTM-fine-tune 整体结构如图 3 所示,对于源域与目标域的时间序列退化数据,采用多层的 LSTM 网络分别进行抽象,再通过全连接层网络对数据进行预测。由于源域的数据是理想的、完备的全寿命周期退化数据,其时间序列预测模型具有较强的参考价值,因此目标域在时间序列预测模型的建立上,将源域网络参数直接移植,并将部分 LSTM 网络参数进行固定,不再更改。目标域的数据通过网络训练对另外一部分可以更改的网络参数进行调整,使网络更加适应目标域的输入输出映射关系。

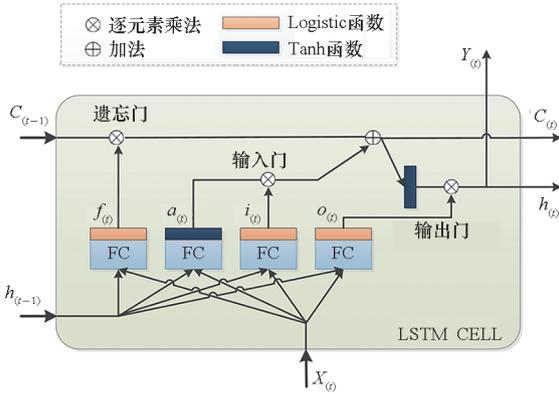


图2 LSTM 单元内部结构示意图

Fig.2 LSTM CELL internal structure diagram

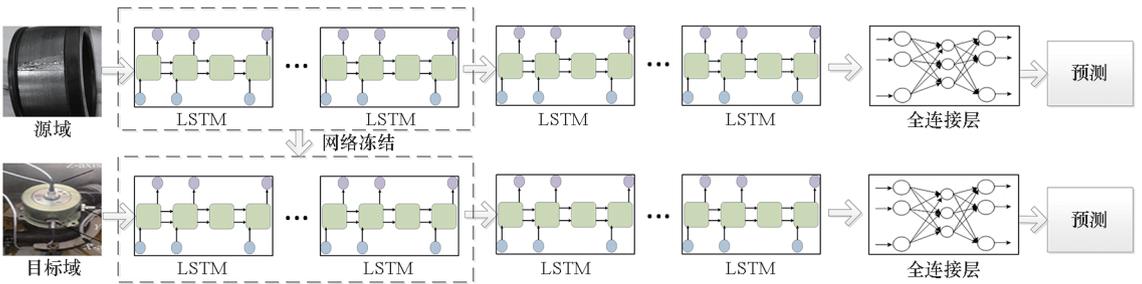


图3 LSTM-fine-tune 整体架构

Fig.3 LSTM-fine-tune overall architecture

### 2.2 基于灰狼算法的网络层数与冻结层数确定

在模型应用过程中,存在两个关键的超参数,即 LSTM 网络层数与冻结网络层数,两个超参数的确定影响着网络的结构,直接关系到迁移的效果。灰狼算法是元启发算法的一种,本质是解决优化问题,通过模拟狼群捕杀猎物过程中的分工合作机制,以寻找给定空间中的参数最优解。灰狼算法由于步骤简单、收敛性强、适用范围广,在参数优化领域得到广泛应用。为研究网络层数与冻结层数对迁移精度的影响,确定网络最佳超参数,采用灰狼算法对网络层数与冻结层数进行寻优[26]。灰狼行为可以表示如下:

1) 灰狼群结构划分。将灰狼群的狩猎空间设为一个  $N \times M$  的欧式空间,其中  $N$  为灰狼的数量,  $M$  为猎物的数量,则每一匹狼的位置表示  $X_i = (x_1, x_2, \dots, x_M)$  ( $i = 1, 2, \dots, N$ ), 依据每一匹灰狼的位

置信息评价该灰狼与猎物的远近,评选出位置最优的三匹狼分别为  $\alpha$  狼、 $\beta$  狼、 $\delta$  狼,其余狼为  $\omega$  狼。

2) 搜索猎物。灰狼群的搜索活动由  $\alpha$  狼、 $\beta$  狼、 $\delta$  狼指引完成,在三匹狼的号召下,对猎物进行搜寻,搜寻模型如式(1)~(4)所示。其中,  $a$  的取值随着迭代次数的增加从 2 线性递减到 0,  $r_1$ 、 $r_2$  在每次迭代过程中在 (0, 1) 之间随机取值,“ $\cdot$ ”表示 Hadamard 乘积。

$$\begin{cases} D_\alpha = |C_1 \cdot X_\alpha - X| \\ D_\beta = |C_2 \cdot X_\beta - X| \\ D_\delta = |C_3 \cdot X_\delta - X| \end{cases} \quad (1)$$

$$\begin{cases} X_1 = X_\alpha - A_1 \cdot D_\alpha \\ X_2 = X_\beta - A_2 \cdot D_\beta \\ X_3 = X_\delta - A_3 \cdot D_\delta \end{cases} \quad (2)$$

$$\begin{cases} A = 2a \cdot r_1 - a \\ C = 2r_2 \end{cases} \quad (3)$$

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \quad (4)$$

将 LSTM 网络层数与冻结层数作为灰狼算法的待寻优参数,即每匹狼的位置是一个二维的坐标,第一维代表网络层数,第二维代表冻结层数,按照式(1)~(4)进行搜索。在每次搜索过程后,进行一次灰狼群结构的重新划分,通过评价指标判断距离猎物最近的三匹狼自动转化为  $\alpha$  狼、 $\beta$  狼、 $\delta$  狼,并组织下一轮搜索。

当搜索迭代次数较少时,  $a$  的取值较大,由于还未掌握网络层数与冻结层数的最优值大致范围,这时的灰狼群趋向于扩大搜索范围,部分灰狼群向  $\alpha$  狼、 $\beta$  狼、 $\delta$  狼方向收缩,另一部分灰狼群向相反方向运动。对于寻优而言,这样的搜索方式能够克服局部最优值的影响。

随着搜索迭代次数的不断增加,  $a$  的值不断减小,这时灰狼群基本已经掌握网络层数与冻结层数的最优值范围。灰狼群逐渐趋于收缩,开始发动包围和猛攻。

$r_1$ 、 $r_2$  随机值的选取,给灰狼群的搜索带来了不定因素,这也符合大自然环境下灰狼群信息传递的失真以及灰狼游走的随机性,对于模型而言,能够有效避免陷入局部最优值。

通过多次搜索行为后,狼群逐渐逼近猎物,从概率上而言,  $\alpha$  狼是最接近于猎物的狼。因此,从概率的角度判断,此时  $\alpha$  狼的位置信息,即网络层数与冻结层数数值,代表了最大概率接近于最优化参数的数值,因此可以选作网络构建的超参数。

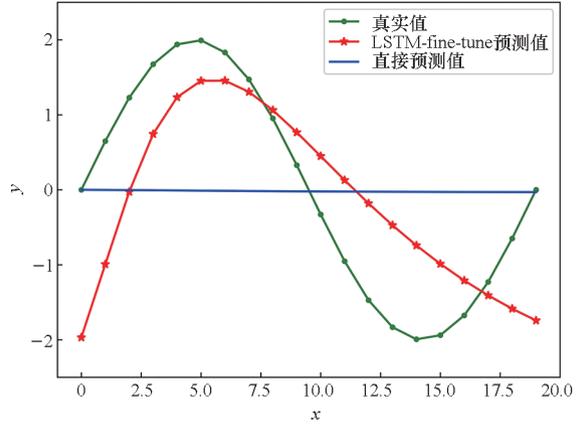
### 3 典型函数验证

为验证 LSTM-fine-tune 对预测模型迁移泛化的能力,采用简单的回归问题,利用 100 层的 LSTM 网络从多种回归函数生成的数据中提取知识,并应用于其他回归任务,多次迭代统计 LSTM-fine-tune 的知识迁移能力。在源域中,数据生成方式是基于正弦函数生成的,但每个任务生成过程中,相位与幅值都是随机产生的,其中振幅在  $[0.1, 3.0]$  范围内变化,相位在  $[0, \pi/2]$  范围内变化,每个函数生成 100 个间距不等的  $x - y$  数组。为防止网络过拟合,生成的数据对模型仅进行两步更新,总共采用 50 个任务,即 50 个相位、振幅各异的函数生成的数据对模型进行更新,使网络从不同数据中学习正弦函数输入输出的通用知识。

模型训练后,将 100 层 LSTM 网络中的 95 层进行冻结,仅利用 5 层网络基于新任务的数据进行微调。在新任务的生成上,也采用正弦函数,振幅

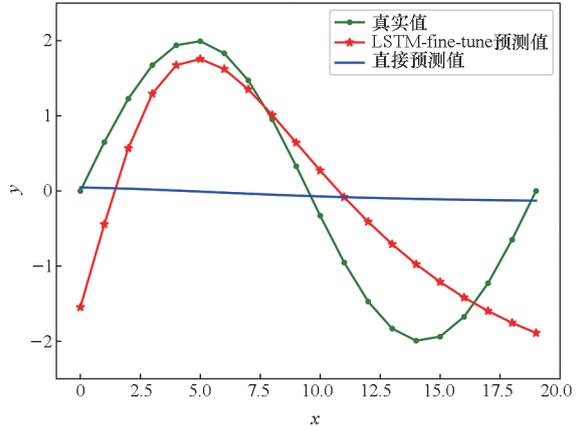
在  $[0.1, 3.0]$  范围内变化,相位在  $[0, \pi/2]$  范围内变化。为测试网络对正弦数据的快速收敛能力,测试任务上仅使用 20 个数据对网络进行微调。

不同更新步数条件下预测数据与真实数据情况如图 4 所示。从不同更新步数条件下预测数据收敛情况来看,LSTM-fine-tune 的预测数据能够在少量数据情况下快速收敛到正弦函数生成的真实



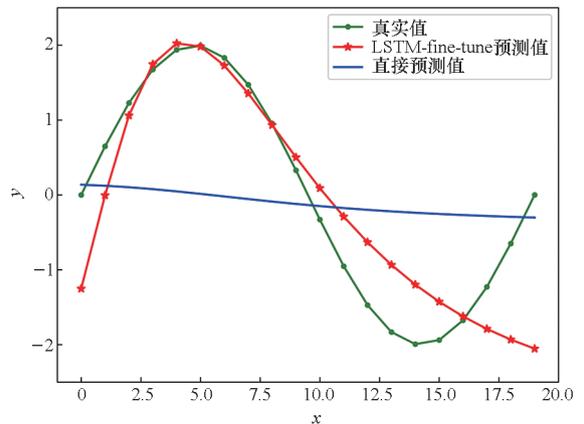
(a) 训练更新步数: 0

(a) Training update steps: 0



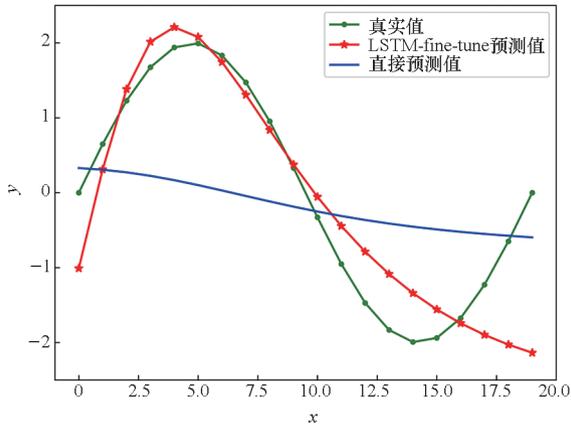
(b) 训练更新步数: 10

(b) Training update steps: 10



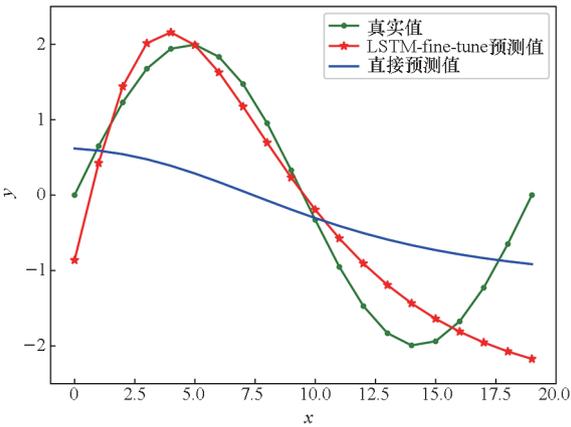
(c) 训练更新步数: 20

(c) Training update steps: 20



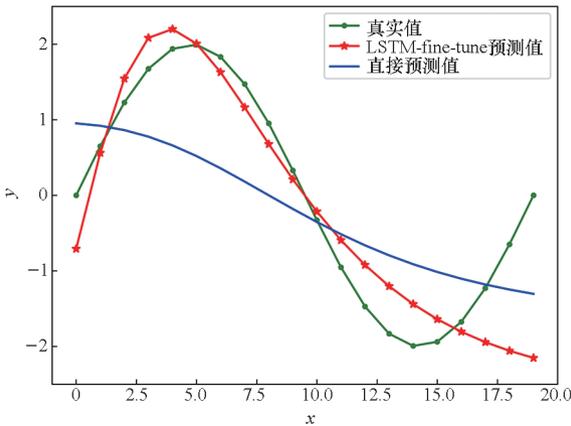
(d) 训练更新步数: 30

(d) Training update steps: 30



(e) 训练更新步数: 40

(e) Training update steps: 40



(f) 训练更新步数: 50

(f) Training update steps: 50

图4 不同训练更新步数下预测效果

Fig.4 Prediction renderings under different training updating steps

数据,这说明在训练过程中,LSTM-fine-tune 模型从大量正弦函数生成数据的任务中学习到了正弦函数的通用表示形式,能够快速适应新的预测任务。

为进一步量化测试 LSTM-fine-tune 模型的预测能力,在新任务上采用随机生成 100 组不同的测试数据,每组数据 20 个数据点。测试过程中,每组数据进行 50 次更新,并采用均方根误差作为结果判定指标。结果表明,LSTM-fine-tune 模型 100 组任务误差平均值为 1.033 5,直接预测误差平均值为 1.536 8。

## 4 氧气浓缩器退化数据获取

为验证提出模型的有效性,以氧气浓缩器为例,获取氧气浓缩器在理想试验条件下与实际服役环境下的退化数据,将试验数据建立的模型迁移至实际服役数据,以测试模型迁移泛化能力。

### 4.1 试验数据

#### 4.1.1 试验方案

对于试验条件下的理想数据,本文作者与中航工业合肥江航飞机装备股份有限公司合作,以氧气浓缩器为研究对象,设计了退化试验方案。通过试验,模拟了 YNQ-8 氧气浓缩器在实际工作状态下的引气压力、含水量、温度等参数的变化,以此观察氧气浓缩器整体及内部零件退化失效规律<sup>[27]</sup>。

氧气浓缩器退化试验硬件连接如图 5 所示。试验台由空气压缩机、过滤器、加湿器、试验箱、真空泵等组成。空气通过空气压缩机进入管道内,形成高压气体,高压气体通过两级过滤器过滤掉杂质及水汽,以控制无关的气体成分;通过加湿器对气体的水汽含量进行控制,以影响氧气浓缩器的退化;气体进入试验箱后,在氧气浓缩器中与分子筛床充分接触,进行吸附、解吸过程,将氮气排除,留下富氧气体,同时,水汽也与分子筛床充分接触,影响着分子筛床的退化;气体离开试验箱后,从管道排出。

试验台设有压力传感器、湿度传感器、氧气分析仪等参数监控装置,以监控试验台工作情况及氧气浓缩器性能情况。在空气压缩机后及低压舱后都设置了压力传感器,以及时对引气压力进行调节;在低压舱前设置湿度传感器,以检测进入氧气浓缩器产品气体中的水汽含量;低压舱后设置测氧仪,检测从氧气浓缩器产品输出气体的氧分压及氧浓度含量,作为评判氧气浓缩器工作性能的指标。

#### 4.1.2 试验数据获取

设定 4 h 为一个工作周期来模拟飞机一个起落架次。由于氧气浓缩器最低工作引气压力为 0.14 MPa,且飞机上引气压力一般不超过 1.0 MPa,因此将引气压力按照 0.14 MPa、

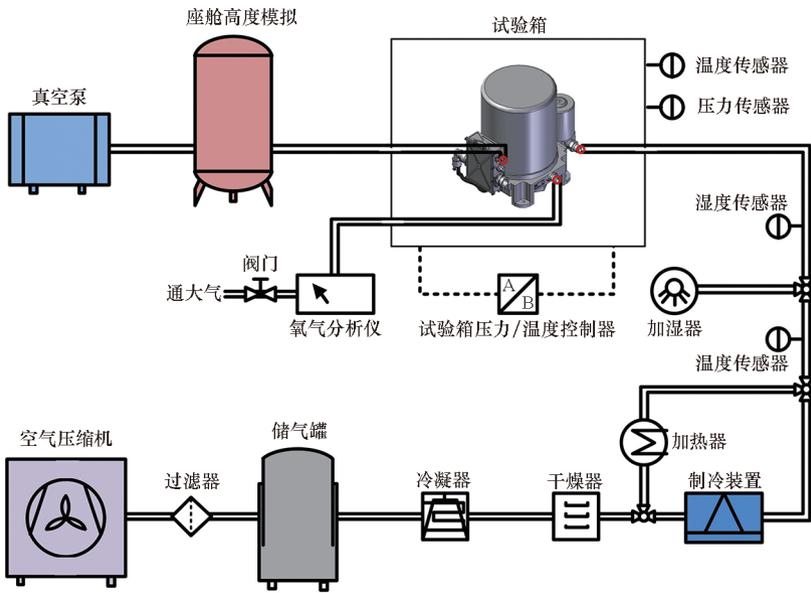


图 5 氧气浓缩器退化试验硬件连接图

Fig. 5 Hardware connection diagram of oxygen concentrator degradation test

0.30 MPa、0.70 MPa、1.00 MPa、0.70 MPa、0.30 MPa、0.14 MPa 的应力谱进行循环试验,并在每一个工作周期后测试氧气浓缩器的性能下降情况,验证引气条件对产品的影响程度,一个试验循环退化数据记录见表 1。由于水汽对分子筛的影响是一个相对缓慢的过程,短时间内分子筛的吸附性能不会出现明显变化,为充分研究退化趋势,设定工作总时长为 250 h。

表 1 氧气浓缩器一个试验循环退化数据记录

Tab. 1 Oxygen concentrator one test cycle degradation data recording

| 引气压力/<br>MPa | 引气湿度/<br>(g/L) | 引气<br>温度/°C | 氧气体积<br>分数/% | 氧分压/<br>kPa |
|--------------|----------------|-------------|--------------|-------------|
| 0.14         | 32.84          | 70          | 38.1         | 34.9        |
| 0.30         | 61.78          | 69          | 59.8         | 57.5        |
| 0.70         | 80.63          | 69          | 63.9         | 61.8        |
| 1.00         | 94.47          | 71          | 64.3         | 63.5        |
| 0.70         | 74.38          | 71          | 65.0         | 63.3        |
| 0.30         | 54.00          | 69          | 64.4         | 62.4        |
| 0.14         | 34.02          | 69          | 38.3         | 37.9        |

### 4.2 实际服役环境数据获取与预处理

#### 4.2.1 数据获取

为对比氧气浓缩器在实际服役环境下的退化规律,收集了某氧气浓缩器实际工作运行中的退

化数据。通过初步分析发现,该部件从装机到最终发生故障的运行时间为 213 h。初步分析发现,该数据存在较多噪点与坏值,通过数据清洗及坏值剔除后得到的退化数据如图 6 所示。

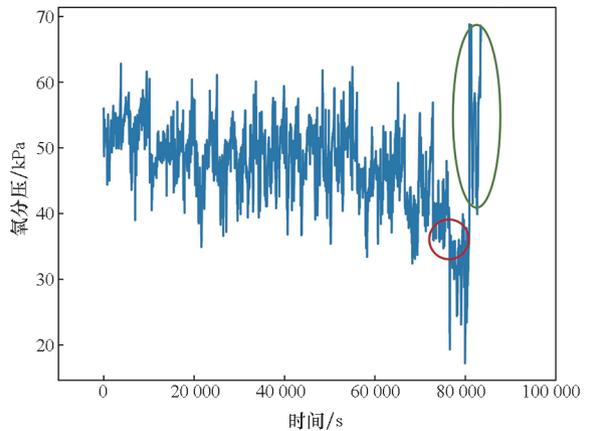


图 6 实际服役环境下氧气浓缩器退化数据

Fig. 6 Oxygen concentrator degradation data under actual service conditions

图 6 中红圈标注部位有大幅度的下降,绿色椭圆标注区域有大幅度的上升。然而,通过原理分析发现,氧气浓缩器的退化过程缓慢,一般不会出现突降与突升的现象。基于分析,判断红圈区域出现的原因可能是膜片破损,这导致指标突降到设定值以下,触发了故障报警。在更换氧气浓缩器后,氧分压回到正常状态,因此出现绿圈的氧分压突升。为印证该假设,调研了厂家中该部件的检修记录,发现故障原因系膜片破损,与分析结果一致。

#### 4.2.2 数据预处理

在实际服役环境下,随着高度的增加,大气压力将会降低,对于解吸附过程而言,环境压力的降低将会有利于氮气的排出,因此,高度的变化可能会引起氧分压的波动。同时,氧气浓缩器引气压力的变化,将会影响到氧气浓缩器的进气量,造成分子筛气体环境的改变,影响最终氧分压的数值。在数据分析中发现,由于实际工况下高度在不断发生变化,引气压力也随之改变,导致了氧分压数据的波动,对退化建模带来困难。针对该问题,通过限制数据中的高度、引气压力两个变量,控制其他变量对于引气压力的影响。图7所示为控制高度为4 000~5 000 m、引气压力为200~250 kPa得到的引气压力数据。由图7可知,该氧气浓缩器退化可以分为四个过程:一个是平稳阶段,氧分压整体趋势平稳,局部数据波动;二是退化阶段,氧分压整体呈下降趋势;三是膜片破损阶段,氧分压急剧降低;四是换件阶段,氧分压恢复正常水平。

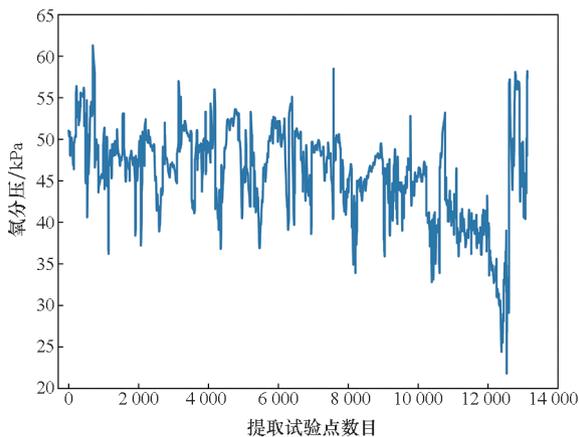


图7 环境因素限制条件下氧气浓缩器退化数据

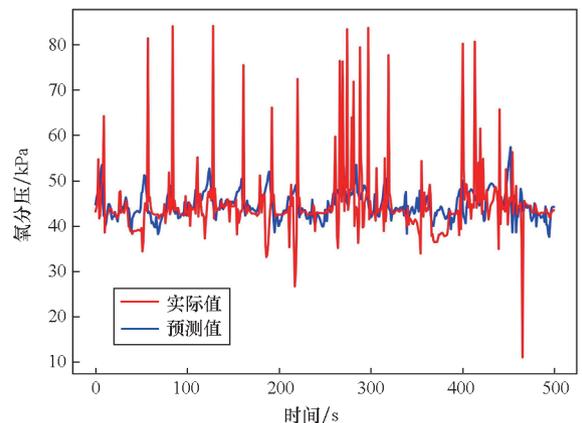
Fig. 7 Degradation data of oxygen concentrator under environmental constraints

## 5 模型验证

为研究网络层数与冻结层数对迁移精度的影响,确定网络最佳超参数,采用灰狼算法对网络层数与冻结层数进行寻优。在数据集划分上,目标域数据来源为实际服役环境下的氧气浓缩器退化数据,截取了1 000个数据点,其中500个作为训练集,500个作为测试集;源域数据来源为实验条件下的退化数据,数据集划分与目标域一致。在灰狼算法对两个超参数进行寻优过程中,LSTM-fine-tune神经网络对训练集进行训练,用测试集进行测试,判定网络对数据的预测精度。随着灰

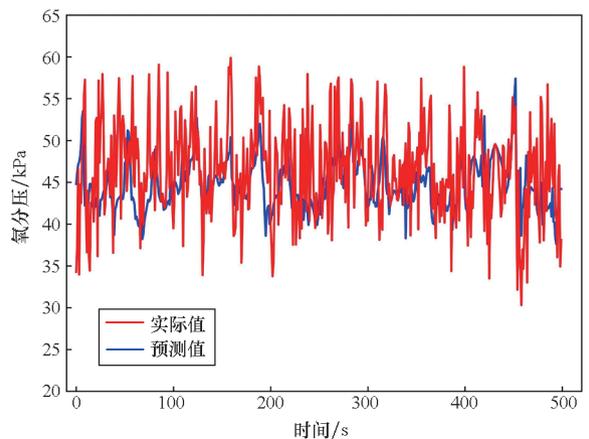
狼群对LSTM的参数不断寻优、迭代,对于验证集的预测精度也在不断提高,不同迭代次数下的模型预测数据与实际数据对比如图8所示。从图中可以发现,在寻优初期,迭代次数为1时,预测均方误差达到62.013 8,随着灰狼群对参数不断迭代寻优,预测均方误差降至36.810 7、12.744、9.798 8,最终随着迭代30次后,均方误差仅为3.214 2,远低于迭代寻优之前<sup>[28]</sup>。

通过网络预训练发现,当网络层数为7、冻结层数为3时,迁移效果最佳。故首先基于实验条件下的源域数据对7层LSTM+全连接层的神经网络进行训练,训练完成后,将模型迁移至实际服役环境下的目标域数据中。将3层LSTM参数进行冻结操作,不再进行调参,其他层训练结果作为目标域网络参数的初始值,并基于实际环境下的退化数据对网络参数进行进一步微调,使网络更加适应实际环境下的退化数据。采用迁移架构的预测结果可视化如图9(a)所示。作为对照,采用7层LSTM+全连接层的神经网络对目标域数据直接进行预测,其结果可视化如图9(b)所示。



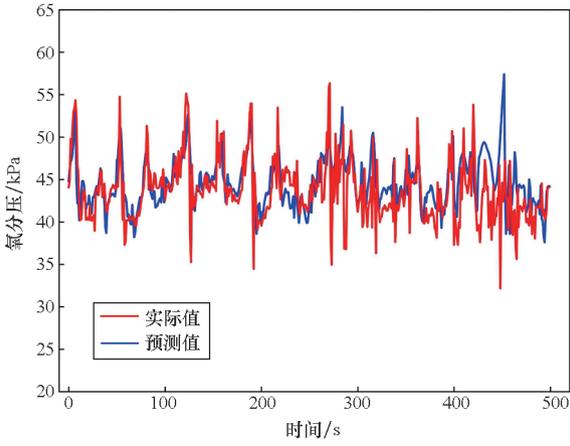
(a) 迭代次数: 1

(a) Iterations: 1

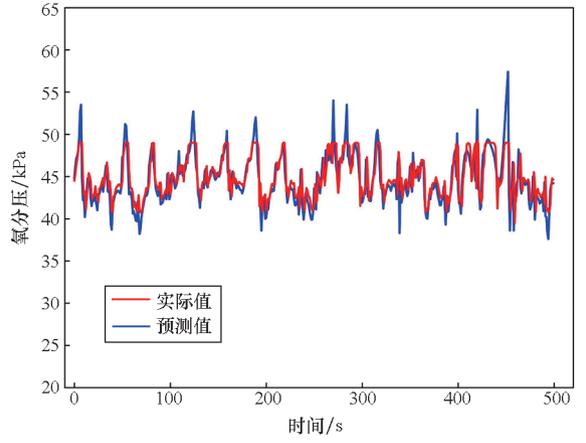


(b) 迭代次数: 5

(b) Iterations: 5



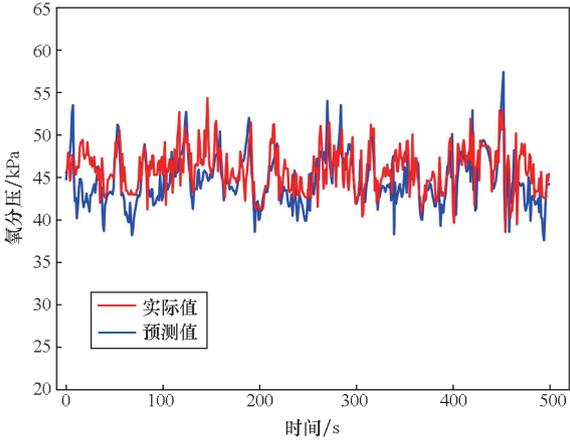
(c) 迭代次数: 10  
(c) Iterations: 10



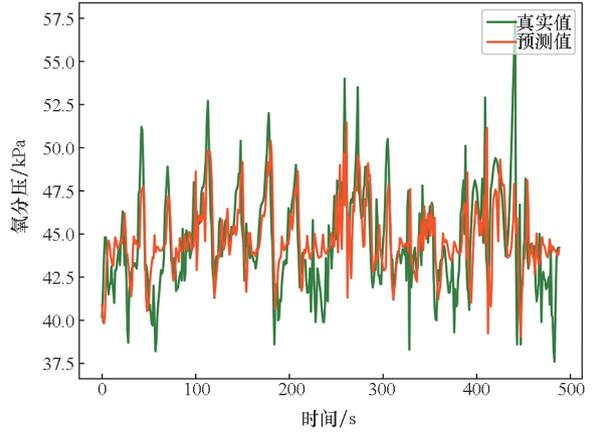
(f) 迭代次数: 30  
(f) Iterations: 30

图 8 不同迭代次数下预测效果图

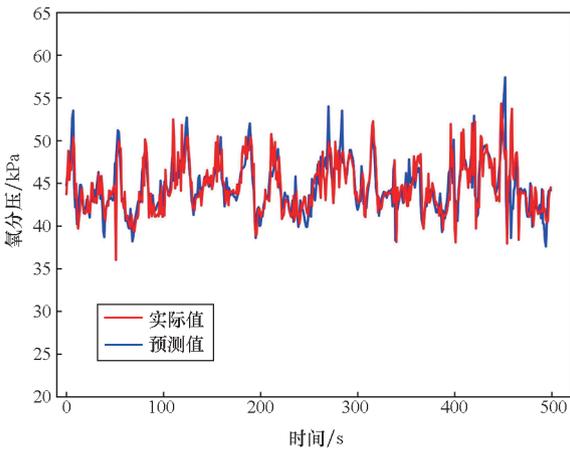
Fig. 8 Prediction renderings under different iterations



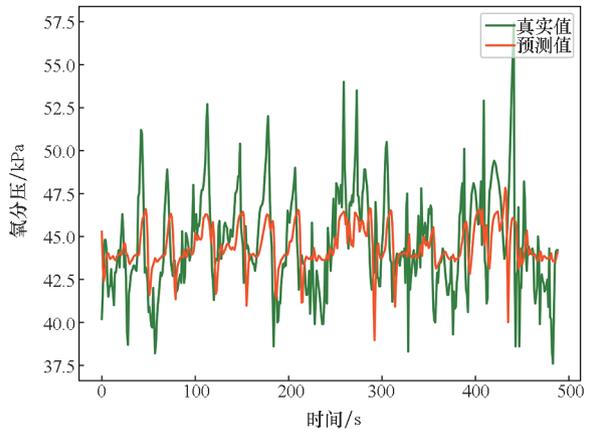
(d) 迭代次数: 15  
(d) Iterations: 15



(a) 采用迁移架构的预测结果  
(a) Result using migration architecture



(e) 迭代次数: 20  
(e) Iterations: 20



(b) 直接预测结果  
(b) Result of predicting directly

图 9 预测结果可视化

Fig. 9 Visualization of prediction results

由于对寿命预测模型迁移相关研究有限,选取了迁移成分分析自编码器支持向量回归(transfer component analysis-autoencoder-support vector regression, TCA-AE-SVR)与小波变换模拟退火(simulate anneal-wavelet transform, WT-SA)两种方法进行对比,分别对氧气浓缩器退化数据进行预测。预测结果对比见表2。可以发现,经过LSTM-fine-tune方法进行预测模型迁移学习的网络,从训练集到验证集,均方误差都小于迁移学习前的结果。对于训练集,预测精度提高了43.0%;对于测试集,预测精度提高了33.5%。这表明LSTM-fine-tune架构能够将试验条件下的退化经验迁移至实际服役环境下的寿命预测中,从而有效提高了实际服役环境下小样本数据的预测精度。LSTM-fine-tune在训练集与预测集的精度上都明显高于表2中其他对比方法,这也从侧面验证了方法的有效性。

表2 预测结果对比

Tab.2 Comparison of prediction results

| 数据集     | LSTM  | LSTM-fine-tune | TCA-AE-SVR | WT-SA |
|---------|-------|----------------|------------|-------|
| 训练集均方误差 | 1.899 | 1.082          | 3.526      | 1.569 |
| 测试集均方误差 | 4.836 | 3.214          | 7.632      | 4.336 |

## 6 结论

针对寿命预测模型迁移问题,本文提出了一种LSTM-fine-tune迁移模型,利用理想条件下的试验数据对模型进行训练,在迁移过程中,对部分LSTM网络层进行冻结,利用实际服役环境下的数据对网络其他部分进行修正。为确定LSTM结构层数与冻结层数,采用灰狼算法搜寻最优超参数,使目标域数据预测精度趋于最优。为验证模型的有效性,分别获取了试验条件下与实际服役环境下氧气浓缩器的数据,对模型的泛化能力进行验证。结果表明,迁移后训练集预测精度提高了43.0%,测试集预测精度提高了20.2%,明显高于对照组的传统方法。灰狼群搜寻过程中,预测精度随网络结构与冻结层数的变化收敛于最优值,这说明源域知识通过网络冻结迁移到目标域,从而使目标域预测精度提高。

在模型应用方面,首先要得到研究对象的理

想数据,理想数据可以通过试验或者仿真等方法获取。之后,利用理想条件下的试验数据对LSTM进行训练,得到研究对象在理想数据条件下的退化模型。然而,这一模型针对的数据与实际服役环境还存在较大差异,因此采用本文方法,对其中一定层数的网络进行冻结,其他未冻结部分利用实际服役环境下的数据对网络其他部分进行修正,从而达到将理想数据条件下提取的知识迁移到实际服役环境的目的。

## 参考文献(References)

- [1] HUANG R Q, XI L F, LI X L, et al. Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods [J]. Mechanical Systems and Signal Processing, 2007, 21(1): 193-207.
- [2] 车畅畅,王华伟,倪晓梅,等.基于1D-CNN和Bi-LSTM的航空发动机剩余寿命预测[J].机械工程学报,2021,57(14):304-312.  
CHE C C, WANG H W, NI X M, et al. Residual life prediction of aeroengine based on 1D-CNN and Bi-LSTM[J]. Journal of Mechanical Engineering, 2021, 57(14): 304-312. (in Chinese)
- [3] LI X, DING Q, SUN J Q. Remaining useful life estimation in prognostics using deep convolution neural networks [J]. Reliability Engineering & System Safety, 2018, 172: 1-11.
- [4] DEUTSCH J, HE D. Using deep learning-based approach to predict remaining useful life of rotating components[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2018, 48(1): 11-20.
- [5] 刘博,王明烁,李永,等.深度学习在时空序列预测中的应用综述[J].北京工业大学学报,2021,47(8):925-941.  
LIU B, WANG M S, LI Y, et al. Deep learning for spatio-temporal sequence forecasting: a survey [J]. Journal of Beijing University of Technology, 2021, 47(8): 925-941. (in Chinese)
- [6] YONG H W, HUANG J Q, MENG D Y, et al. Momentum batch normalization for deep learning with small batch size[C]//Proceedings of Computer Vision-ECCV, 2020: 224-240.
- [7] ZHOU K Y, YANG Y X, HOSPEDALES T, et al. Learning to generate novel domains for domain generalization [C]//Proceedings of Computer Vision-ECCV, 2020: 561-578.
- [8] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks [C]//Proceedings of the 34th International Conference on Machine Learning, 2017.
- [9] VU T, HU B T, MUNKHDALAI T, et al. Sentence simplification with memory-augmented neural networks[C]//Proceedings of the NAACL-HLT, 2018: 79-85.
- [10] SUN C, MA M, ZHAO Z B, et al. Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing[J]. IEEE Transactions on Industrial Informatics, 2019, 15(4): 2416-2425.
- [11] MAO W T, HE J L, ZUO M J. Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning [J]. IEEE Transactions on

- Instrumentation and Measurement, 2020, 69(4): 1594 – 1608.
- [12] RAGAB M, CHEN Z H, WU M, et al. Adversarial transfer learning for machine remaining useful life prediction [C]// Proceedings of IEEE International Conference on Prognostics and Health Management (ICPHM), 2020: 1 – 7.
- [13] WANG J D, CHEN Y Q, FENG W J, et al. Transfer learning with dynamic distribution adaptation [J]. ACM Transactions on Intelligent Systems and Technology, 2020, 11(1): 1 – 25.
- [14] WANG J D, FENG W J, CHEN Y Q, et al. Visual domain adaptation with manifold embedded distribution alignment [C]// Proceedings of Computer Vision and Pattern Recognition, 2018.
- [15] ZHU Y C, ZHUANG F Z, WANG J D, et al. Deep subdomain adaptation network for image classification [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(4): 1713 – 1722.
- [16] CHEN L Z, MOSCHITTI A. Transfer learning for sequence labeling using source model and target data [C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2019: 6260 – 6267.
- [17] CUI W Y, ZHENG G Y, SHEN Z Q, et al. Transfer learning for sequences via learning to collocate [C]// Proceedings of International Conference on Learning Representations, 2019.
- [18] FAWAZ H I, FORESTIER G, WEBER J, et al. Transfer learning for time series classification [C]// Proceedings of IEEE International Conference, 2018.
- [19] DEVLIN J, CHANG M, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]// Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics, 2019: 4171 – 4186.
- [20] WINATA G I, MADOTTO A, LIN Z J, et al. Language models are few-shot multilingual learners [C]// Proceedings of the 1st Workshop on Multilingual Representation Learning, 2021: 1 – 15.
- [21] 彭程. 基于预训练的工业数据软测量 [D]. 杭州: 浙江大学, 2022.
- PENG C. Soft sensor of industrial data based on pre-training [D]. Hangzhou: Zhejiang University, 2022. (in Chinese)
- [22] PAN S J, YANG Q. A survey on transfer learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345 – 1359.
- [23] YAN R Q, SHEN F, SUN C, et al. Knowledge transfer for rotary machine fault diagnosis [J]. IEEE Sensors Journal, 2020, 20(15): 8374 – 8393.
- [24] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735 – 1780.
- [25] ZHAO Z, CHEN W H, WU X M, et al. LSTM network; a deep learning approach for short-term traffic forecast [J]. IET Intelligent Transport Systems, 2017, 11(2): 68 – 75.
- [26] XU L Y, LI Q, YU J, et al. Spatio-temporal predictions of SST time series in China's offshore waters using a regional convolution long short-term memory (RC-LSTM) network [J]. International Journal of Remote Sensing, 2020, 41(9): 3368 – 3389.
- [27] 潘晋新, 景博, 焦晓璇, 等. 多应力耦合条件下氧气浓缩器退化建模 [J]. 北京航空航天大学学报, 2023, 49(2): 472 – 481.
- PAN J X, JING B, JIAO X X, et al. Degradation modeling of oxygen concentrator in multiple stress coupling [J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(2): 472 – 481. (in Chinese)
- [28] MIRJALILI S, MIRJALILI S M, LEWIS A. Grey wolf optimizer [J]. Advances in Engineering Software, 2014, 69: 46 – 61.