DOI : 10.19666/j.rlfd.202501003

基于 TimeGAN 的燃气轮机气路故障数据增强方法

陈启泰[1]，翟春华[2]，黄　庆[2]，牛晨晖[3]，苏新民[3]，焦道顺[2]，金建华[2]，
潘赫男[3]，李　冬[3]，奚新国[2]，袁　源[2]，阴海强[2]，苏　通[2]
（1.西安热工研究院有限公司，陕西　西安　710054；2.华能南京燃机发电有限公司，江苏　南京　210046；
3.华能国际电力江苏能源开发有限公司，江苏　南京　210000）

［摘　　要］为解决燃气轮机（燃机）气路故障机器学习模型训练时存在的故障样本不平衡问题，提出基于时间序列生成对抗网络（TimeGAN）的燃机气路故障数据增强方法。首先，对实际工况下的燃机气路故障瞬态运行数据进行清洗、标准化、特征工程及 PCA 降维，以保留关键故障特征。然后，构建包含嵌入、生成、判别与重构 4 模块的 TimeGAN 模型，并利用 Optuna 超参数优化工具实施贝叶斯优化以获得最优超参数。最后，通过预测指标、分类指标（基于后验 RNN 判别器）及可视化指标（PCA 与 t-SNE）对生成数据质量进行综合评估。研究结果表明，TimeGAN 生成的数据在预测任务中平均绝对误差为 4.47%，与真实数据在统计特征及潜在表示空间上高度相近；可视化分析亦表明，生成数据的聚类形态和分布趋势与真实数据高度近似。

［关　键　词］燃气轮机气路故障；时间序列生成对抗网络；数据增强；故障诊断

［引用本文格式］陈启泰，翟春华，黄庆，等. 基于 TimeGAN 的燃气轮机气路故障数据增强方法[J]. 热力发电，年，卷(期)：起始页码-终止页码.　CHEN Qitai, ZHAI Chunhua, HUANG Qing, et al. Gas turbine gas path fault data augmentation method based on TimeGAN[J]. Thermal Power Generation, 年, 卷(期): 起始页码-终止页码.

Gas turbine gas path fault data augmentation method based on TimeGAN

CHEN Qitai[1], ZHAI Chunhua[2], HUANG Qing[2], NIU Chenhui[3], SU Xinmin[3],
JIAO Daoshun[2], JIN Jianhua[2], PAN Henan[3], LI Dong[3], XI Xinguo[2],
YUAN Yuan[2], YIN Haiqiang[2], SU Tong[2]
（1.Xi'an Thermal Power Research Institute Co., Ltd., Xi'an, Shaanxi 710054, China;
2.Huaneng Nanjing Gas Turbine Power Generation Co., Ltd., Nanjing, Jiangsu 210046, China;
3.Huaneng International Power Jiangsu Energy Development Co., Ltd., Nanjing, Jiangsu 210000, China）

Abstract: To address the issue of imbalanced fault samples in training machine learning models for gas turbine gas path faults, a data augmentation method based on Time-Series Generative Adversarial Networks (TimeGAN) is proposed. First, transient operational data from actual gas turbine gas path faults are cleaned, standardized, processed through feature engineering, and reduced using PCA to preserve critical fault features. Next, a TimeGAN model comprising embedding, generation, discrimination, and reconstruction modules is constructed, and Bayesian optimization is performed using the Optuna hyperparameter optimization tool to determine the optimal hyperparameters. Finally, the quality of the generated data is comprehensively evaluated through prediction metrics, classification metrics (based on a posterior RNN discriminator), and visualization metrics (PCA and t-SNE). The results demonstrate that data generated by TimeGAN achieve a mean absolute error of 4.47% in prediction tasks, exhibiting high similarity to real data in terms of statistical features and latent

收 稿 日 期：2025-01-02　网络首发日期：XXXX-XX-XX
基 金 项 目：中国华能集团有限公司总部科技项目资助（HNKJ22-H16）
Supported by：Science and Technology Project of China Huaneng Group Co., Ltd.(HNKJ22-H16)
第一作者简介：陈启泰（2000），男，硕士研究生，主要研究方向为燃气轮机故障诊断，chenqitai@tpri.com.cn。

representation space, making it difficult for the discriminator to distinguish between the two. Visualization analysis further confirms that the clustering patterns and distribution trends of the generated data closely resemble those of the real data.

Key words: gas-path fault of gas turbine; time-series generative adversarial network (TimeGAN); data augmentation; fault diagnosis

燃气轮机（燃机）是燃气-蒸汽联合循环发电机组的核心设备，具有高效、灵活、低污染的特性[1]。燃机结构复杂，在高温、高转速环境下持续运行，受机械磨损、热应力、腐蚀和疲劳等因素影响，易出现性能衰退或故障[2]。鉴于此，及时有效的燃机故障诊断与预警技术至关重要。

气路故障诊断技术是燃机故障诊断与预警技术的关键组成部分[3]。燃机气路故障指压气机、透平等部件由于腐蚀磨损、热畸变、外/内物损伤等原因造成部件效率及通流能力下降的故障[4-5]。

针对燃机气路故障，学界提出了基于数据驱动和基于物理模型驱动的 2 种主要解决方法[6]。建立具有实际工业应用价值的物理模型，需要燃机设备详细的设计数据，由于燃机制造商对设计数据的严格保密，获取这些数据非常困难，导致基于物理模型的方法较难实现。相较于基于物理模型的方法，基于数据驱动的机器学习方法主要通过利用历史大数据，不需要使用燃机设备的设计数据，所以更易实现并有着更强的泛化能力。

近年来，基于数据驱动的燃机气路故障诊断方法中，基于机器学习的方法逐渐成为主流。机器学习方法通过从大量历史数据中自动学习特征、模式和规律，在故障诊断领域展现了出色的性能[7]。孙守泰等[8]利用高斯过程回归的全局映射能力对长短时记忆（long short-term memory, LSTM）神经网络的结果做二次估计，从而提高模型在线辨识结果的准确度。彭道刚等[9]采用 LSTM 神经网络与支持向量机（support vector machine, SVM）相结合，对压气机的叶片积垢、磨损腐蚀、进气口结冰及喘振等故障进行诊断分类，取得了较高的分类准确率。蒋龙陈等[10]将峰值抑采样（peak-holding data smoothing, PHDS）方法、粒子群优化（particle swarm optimization, PSO）算法与深度信念网络（deep belief network, DBN）相结合，提高了 DBN 燃机故障诊断模型的准确率。闫丽萍等[11]采用遗传算法实现 DBN 初始参数自适应调节，优化后模型取得了较高的故障诊断平均精度。Liu 等人[12]采用卷积神经网络（convolutional neural network, CNN）进行热组件故障检测，利用 CNN 提取关键特征，减少涡流影响，取得较高的诊断准确度。Cao 等人[13]基于优化的深度前馈神经网络（deep feedforward neural network, DFNN）实现了故障分类预测，通过热力学模型与神经网络相结合，使得透平故障分类准确率提高。

然而，燃机运行参数间存在高度耦合及不同故障类型间样本数量不平衡等问题，易使故障表征能力与分类准确率下降[14–16]，也使现有方法提取故障子空间的有效性和故障表征能力受到影响。因此，研究基于数据增强的故障分类方法显得尤为重要。

在故障样本数据增强方面，生成对抗网络（generative adversarial networks, GAN）[17-18]和自编码器（variational autoencoder, VAE）[19-20]等方法已经被广泛应用于图像分类领域。Arjovsky 等人[21]提出 Wasserstein GAN（WGAN），使用 Wasserstein 距离来衡量真实样本和生成样本之间的分布距离，从而提升生成样本质量。随后，Gulrajani 等人[22]在 WGAN 的基础上加入了梯度惩罚项，提出了 WGAN-GP，进一步增强了训练的稳定性。此外，合成少数过采样技术（synthetic minority over-sampling technique，SMOTE）[23]作为一种重采样方法，通过随机线性插值在少数样本及其邻居样本之间生成新的样本，在故障诊断领域也被用以解决样本不平衡的问题。然而，这些方法在捕捉时间序列数据的时间相关性和动态特性方面表现有限。

针对传统数据增强方法难以捕捉时间相关性问题，本文基于时间序列生成对抗网络（time-series generative adversarial networks, TimeGAN）[24-26]构建数据增强模型。模型采用 LSTM 神经网络构建 TimeGAN 各模块，能够有效捕捉燃机运行数据的时间依赖性与动态特性。针对燃机运行数据的非线性和强耦合特征，采用主成分分析（principal component analysis, PCA）对原始故障数据进行预处理，提取故障子空间以增强故障表征能力，同时提升模型训练效率。最终构建的 TimeGAN 模型能够扩充历史故障样本从而缓解数据不平衡问题，为后续故障诊断与预警工作提供可靠数据基础。

1 基础模型和方法

1.1 LSTM 神经网络

LSTM 神经网络是一种改进的时间递归神经网络（recurrent neural network, RNN），其核心创新在于引入了一种门控机制（gating mechanism），包括输入门（input gate）、遗忘门（forget gate）和输出门（output gate）3 种门控单元。该门控结构旨在对信息流进行选择性控制，从而判断输入信息对模型预测的贡献大小。LSTM 神经网络通过保留重要信息并遗忘无关信息，能够在长时间序列建模中有效应对梯度爆炸和梯度消失等问题[27]。相较于传统的 RNN，LSTM 神经网络在处理长时间序列数据时表现出更强的学习能力和记忆能力。图 1 为 LSTM 神经网络单元结构[28]。在时间序列数据处理中，LSTM 神经网络的每个时间步的计算过程可以表示如下。

1)遗忘门　决定遗忘多少过去的信息，计算公式为：

$$f_t = \sigma\left(W_{\mathrm{f}}\left(h_{\{t-1\}}, x_t\right) + b_{\mathrm{f}}\right) \tag{1}$$

式中：f_t为遗忘门的输出；σ为 Sigmoid 激活函数；w_{f}为遗忘门的权重；$h_{t\text{-}1}$为上一时刻的隐藏状态；x_t为当前输入；b_{f}为偏置项。

2)输入门　决定新的信息有多少存入记忆单元，计算公式为：

$$i_t = \sigma(W_{\mathrm{i}}\, x_t + U_{\mathrm{i}}\, h_{t-1} + b_{\mathrm{i}}) \tag{2}$$

$$\widetilde{C}_{\mathrm{t}} = \tanh\left(W_{\mathrm{C}} x_{\mathrm{t}} + U_{\mathrm{C}}\, h_{\mathrm{t-1}} + b_{\mathrm{C}}\right) \tag{3}$$

式中：i_t为输入门的激活值；$\widetilde{C}_{\mathrm{t}}$为候选记忆单元状态；tanh 为双曲正切函数。

3)记忆单元更新　结合遗忘门和输入门的结果更新记忆单元状态：

$$C_t = f_t\, C_{t-1} + i_t\, \widetilde{C}_t \tag{4}$$

式中：C_t为当前时间步的记忆单元状态。

4)输出门　决定隐藏状态的更新，计算公式为：

$$o_{\mathrm{t}} = \sigma(W_{\mathrm{o}}\, x_{\mathrm{t}} + U_{\mathrm{o}}\, h_{\mathrm{t-1}} + b_{\mathrm{o}}) \tag{5}$$

$$h_{\mathrm{t}} = o_{\mathrm{t}} \cdot \tanh\left(C_{\mathrm{t}}\right) \tag{6}$$

式中：h_t为当前时间步的隐藏状态；o_t为输出门的激活值。

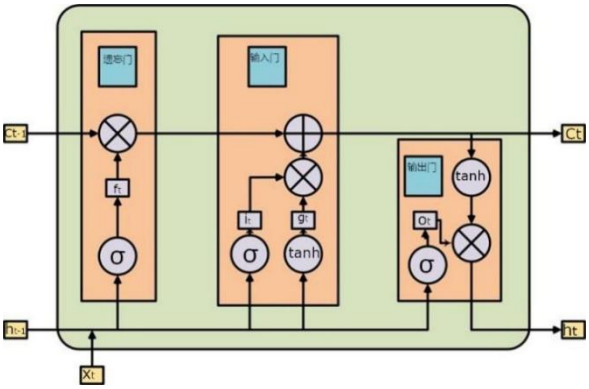


图 1 LSTM 神经网络单元结构

Fig.1 LSTM network unit structure

1.2 TimeGAN 模型

在燃机等复杂工业设备的故障诊断中，时间序列数据具有高度的非线性和复杂的时间依赖性。TimeGAN 是一种专为时间序列数据设计的深度生成模型，其核心创新在于结合了生成对抗网络（GAN）的随机生成能力与序列建模网络（如 RNN、LSTM 神经网络）的时间依赖性学习能力。通过结合自监督学习和生成对抗网络的优势，可以有效捕捉时间序列的动态特性，实现高质量的模拟数据生成。TimeGAN 网络结构如图 2 所示。

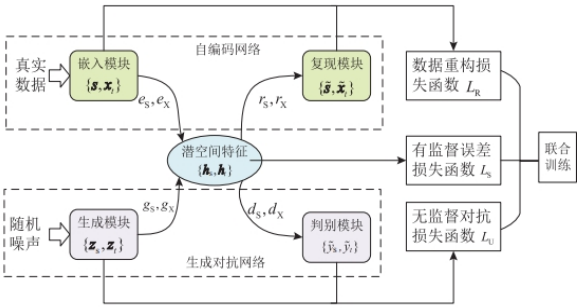


图 2 TimeGAN 网络结构

Fig.2 TimeGAN network architecture

1.2.1 TimeGAN 模型架构

TimeGAN 模型包含嵌入模块（embedder）、生成器模块（generator）、判别器模块（discriminator）和重构模块（recovery）4 个核心模块。各模块的功能及结构如下。

1)嵌入模块　嵌入模块用于提取输入时间序列数据 $X=\{x_1, x_2, \ldots, x_{\mathrm{T}}\}$的特征，并映射到潜在空间，$e_t$可表示为：

$$e_{\mathrm{t}} = \mathrm{RNN}_{\mathrm{embed}(x_{\mathrm{t}})} \tag{7}$$

式中：e_t为时间序列的潜在表示；x_t为时间序列在时刻 t 的观测值。

2)生成器模块　生成器模块接受随机噪声 $Z=\{z_1,z_2,...,z_T\}$和嵌入表示e_t，生成合成时间序列$\hat{X}$：

$$\hat{h}=\mathrm{RNN}_{gen}(z) \tag{8}$$

$$\hat{x}_t=G(\hat{h}) \tag{9}$$

式中：$\hat{h}$为生成的隐藏状态；G 为生成器函数，用于将隐藏状态映射为生成的时间序列值$\hat{x}_t$。

3)判别器模块　判别器模块用于区分真实时间序列 X 和生成的时间序列$\hat{X}$，其输出为输入数据为真实数据的概率 $D(x_t)$，计算公式为：

$$D(x_t)=\sigma(\mathrm{RNN}_{disc}(x_t)) \tag{10}$$

式中：$D(x_t)$为判别器判断的输入x_t为真实数据的概率。

4)重构模块　重构模块通过对嵌入表示e_t进行解码，重构原始时间序列 X。重构误差用于评估重构质量，计算公式为：

$$\hat{x}_t=R(e_t) \tag{11}$$

重构损失函数为：

$$L_{recon}=\| X-\hat{X} \|^2 \tag{12}$$

式中：L_{recon}为重构损失；X 为真实时间序列；$\hat{X}$为重构后的时间序列。

1.2.2 联合损失函数

TimeGAN 通过联合优化多个损失函数来提升生成数据的质量，其联合损失函数定义如下：

$$L_{TimeGAN}=L_{GAN}+\lambda_{recon}\cdot L_{recon}+\lambda_{super}\cdot L_{super} \tag{13}$$

式中：λ_{recon}和λ_{super}为损失函数权重参数，用于平衡各部分损失的贡献。

生成对抗损失L_{GAN}的定义为：

$$L_{GAN}=E[\log(D(X))+E[\log(1-D(\hat{X})) \tag{14}$$

式中：$D(X)$为判别器判断真实时间序列为真的概率；$D(\hat{X})$为判别器判断模拟时间序列为真的概率。

时间监督损失L_{super}的定义为：

$$L_{super}=\| h(X)-h(\hat{X}) \|^2 \tag{15}$$

式中：$h(X)$和$h(\hat{X})$分别表示真实数据和生成数据的隐藏状态。

2 燃机模拟气路故障数据 TimeGAN 模型

2.1 建模步骤

使用 LSTM-TimeGAN 模型结构建立燃机模拟气路故障数据生成模型的步骤如图 3 所示。

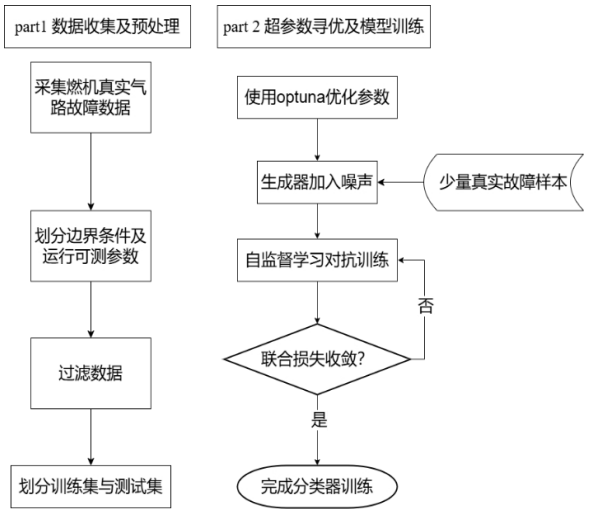


图 3 生成模拟故障数据 TimeGAN 模型建立流程
Fig.3 Process of building the TimeGAN model for generating simulated fault data

由图 3 可见，首先，对原始故障数据进行预处理。原始故障数据采用某型燃机透平动叶损伤的瞬态运行数据，数据采集间隔为 1 s。数据预处理步骤包括数据清洗、数据标准化、特征工程以及输入参数降维。完成对数据的预处理后，为提高模型的生成数据质量，采用超参数优化技术自动化搜索最优参数组合。

本文采用 Optuna 框架作为超参数优化工具。Optuna 是一种高效的超参数优化工具，支持基于树结构的 Parzen 估计器（TPE）算法和网格搜索。通过定义目标函数，Optuna 可在超参数空间中自动搜索，使目标值达到极值（最优参数）。在确定最优参数后，将数据集按比例划分为训练集（80%）与测试集（20%），并在模型性能验证环节中使用测试集对模型进行验证。最后，将确定的最优参数代入模型，并以训练集对模型进行训练。

2.2 输入参数选取

图 4 为重型燃机工作原理。由图 4 可见，气体通过压气机、燃烧室以及透平等部件完成能量转化。首先，压气机将空气压缩至较高压力，然后送入燃烧室进行燃烧，生成高温高压气体，最后气体进入透平膨胀并驱动发电机。在整个过程中，气流的温度、压力、流量等参数对燃机的性能和稳定性有重要影响，这些参数直接反映了气路故障引起的气流不稳定、压缩效率下降和热交换不良等问题。

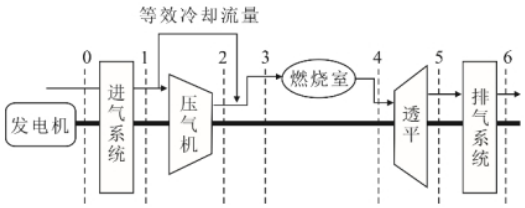


图 4 重型燃机工作原理示意
Fig.4 Heavy-duty gas turbine operating principle

因此，本文选择如表 1 所示的关键参数来构建燃机气路故障数据增强模型，这些参数包括边界条件参数和气路可测参数。所选参数能够较全面地反映燃机气路的运行状态，并有效捕捉气路故障的特征。其中气路可测参数包括压气机进气温度、压气机出口温度、压气机进口压损、压气机出口压力、转速、IGV 开度、天然气质量流量、透平出口压力、透平出口温度。边界条件参数包括大气压力、大气温度、机组负荷、相对湿度。

表 1 气路故障数据增强模型输入参数
Tab.1 Input parameters of gas path fault data augmentation model

项目	参数
边界条件参数	大气压力/Pa 大气温度/℃ 机组负荷/MW 相对湿度/%
气路可测参数	IGV/(°) 压气机进口压损/Pa 压气机进口温度/℃ 压气机出口压力/Pa 压气机出口温度/℃ 透平出口压力/Pa 透平出口温度/℃ 转速/(r•min^{-1}) 天然气质量流量/(kg•s^{-1})

2.3 PCA 降维

为了提高数据的处理效率并增强模型的泛化能力，本文采用了主成分分析（principal component analysis, PCA）对高维输入数据进行降维。PCA 是一种无监督的线性降维方法，通过线性变换将原始数据映射到一个新的坐标系中，在新的坐标系下，各主成分按方差大小排序。通过选择影响最大的主成分组，该方法可以在保留大部分信息的同时，减少数据冗余，提高模型的训练效率。

本文首先对原始的燃机气路故障数据进行标准化处理，以消除各变量之间的量纲差异。随后，应用 PCA 对标准化后的数据进行降维处理。

PCA 降维分析结果如图 5 所示，主成分 1 和主成分 2 的贡献率之和接近 70%，表明这 2 个主成分保留数据的主要信息。进一步分析发现，贡献率最高的前 5 个主成分的累计贡献率达到 93.2%，涵盖了数据中绝大部分的信息。因此，选择这 5 个主成分作为模型的输入参数，将原始的 14 维输入数据有效降维至 5 维。

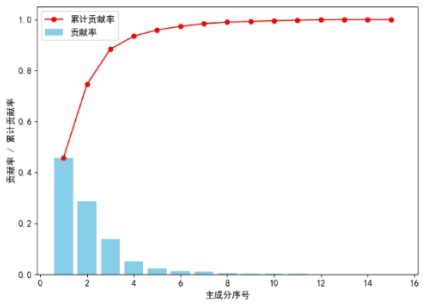


图 5 主成分贡献率及累计贡献率
Fig.5 Principal component contribution rate and cumulative contribution rate

2.4 超参数寻优

本文采用贝叶斯优化方法，并结合 Optuna 框架，对 TimeGAN 模型的超参数进行优化。贝叶斯优化是一种基于概率模型的全局优化方法，尤其适用于高维且计算代价较高的超参数空间。通过该方法能够有效探索不同的超参数组合，从而提升 TimeGAN 模型在时间序列生成任务中的表现。

Optuna 框架基于贝叶斯优化中的 TPE（tree-structured parzen estimator）算法进行超参数搜索。TPE 是一种基于概率模型的优化方法，与传统的随机搜索方法相比，TPE 能够更加高效地探索超参数空间，避免无效区域，从而加速优化过程。

在进行超参数优化时，首先需要为模型的关键超参数定义合适搜索空间。本文动态地定义了多个超参数的搜索范围，具体超参数及搜索范围见表 2。

表 2 TimeGAN 模型超参数优化范围
Tab.2 Hyperparameter optimization range of the TimeGAN model

超参数	描述	取值范围或候选值
隐藏层维度	每层隐藏单元数	{16, 32}
批次大小	每次训练的样本数量	{64, 128, 256}
训练迭代次数	总的训练迭代次数	[400, 2000]（步长 200）
序列长度	时间序列的输入长度	[24, 240]（步长 24）
网络层数	LSTM 的堆叠层数	[2, 5]（步长 1）

在贝叶斯优化过程中，目标函数用于评估每一组超参数组合的性能。本文目标函数的设计综合考虑了 TimeGAN 模型的生成能力和区分能力。具体而言，目标函数通过计算模型在训练数据上的预测能力（predictive）和生成数据与真实数据的区分能

力（discriminative），并将这 2 个指标的加权平均值作为优化目标。在优化过程中，目标函数得分越高，表示该超参数组合的效果越好。

为了确保优化过程的有效性，通过多轮实验评估每一组超参数组合，并根据返回的得分引导搜索。通过这种方式，得到针对本文模型和所用数据集的最佳超参数。最佳超参数见表 3。

表 3 TimeGAN 模型最优超参数
Tab.3 Optimal hyperparameters of the TimeGAN model

超参数名称	最佳超参数值
隐藏层维度	16
批次大小	256
训练迭代次数	800
序列长度	48
网络层数	4

3 模型性能评估指标及分析

为全面评估 TimeGAN 模型生成数据的质量，本文采用 3 种常见的评估指标：预测指标（predictive metrics）、判别指标（discriminative metrics）和可视化指标（visualization metrics）。这些指标从不同的角度考察了生成数据与原始数据之间的分布相似程度，以及生成数据在特定任务中的表现。

3.1 预测指标

预测指标主要考察模型生成数据在下游预测任务中的适用性与有效性。本文选取压气机压比作为关键参数进行一阶预测任务，具体步骤为：将 TimeGAN 生成的模拟故障数据作为训练集输入 LSTM 预测模型，对独立留出的测试集（真实故障数据）进行一阶预测，随后以平均绝对误差（mean absolute error, δ_{MAE}）作为衡量指标。若生成数据与真实数据的统计特性、时间依赖关系及变化趋势相近，则利用生成数据训练的预测模型在测试集上应能取得较低的平均绝对误差，从而间接反映生成数据的“质量”与对下游任务的支持度。预测指标通过式(16)计算：

$$\delta\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}|\hat{y}_{\mathrm{i}}-y_{\mathrm{i}}| \tag{16}$$

式中：n 为样本数量；$\hat{y}_i$为预测值；y_i为真实值。

训练后的预测模型对测试集压气机压比的预测平均绝对误差约为 4.47%。此结果表明，以 TimeGAN 生成的数据进行训练的 LSTM 预测模型能在真实数据上获得较为合理的预测精度，可以为下游的预测与分析环节提供有效支持。

3.2 判别指标

判别指标旨在直接衡量生成数据与真实数据在分布层面上的可区分性。本文利用后验 RNN 模型作为判别器，对真实故障数据与生成数据进行二分类测试：若判别器较难区分二者，即分类准确率接近随机猜测（50%），则说明生成数据与真实数据在特征分布与统计结构上较为相似。为了衡量生成数据与原始数据的相似性，采用式(17)判别指标，判别指标越小，表示生成数据与真实数据越难区分，分布相似度越高。

$$\left|0.5-\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}\right| \tag{17}$$

式中：TP 为分类模型正确预测为正类的样本数量；TN 为分类模型正确预测为负类的样本数量；FP 为分类模型错误预测为正类的样本数量；FN 为分类模型错误预测为负类的样本数量。

判别器模型输出的判别指标约为 0.193，表明模型生成的模拟故障数据与真实故障数据分布较为接近，判别器难以清晰区分二者。

3.3 可视化指标

为更直观地观察生成数据与真实数据在特征空间的分布情况，本文采用 PCA 与 t 分布随机领域嵌入（t-distributed stochastic neighbor embedding，t-SNE）2 种可视化降维技术。PCA 作为线性降维手段，能将高维数据映射到低维主成分空间，以最大方差方向为主轴，考察数据整体分布；t-SNE 则通过非线性映射，将高维数据在低维空间中进行邻域保持，从而更加凸显局部数据点间相对位置关系。

图 6 为原始数据与生成数据 PCA 对比。由图 6 可见，TimeGAN 生成的模拟数据在二维主成分空间中的分布特征与真实数据较为相近，整体趋势变化基本一致。这表明在最大方差方向上，生成数据较好地重现了真实数据的重要结构特征。

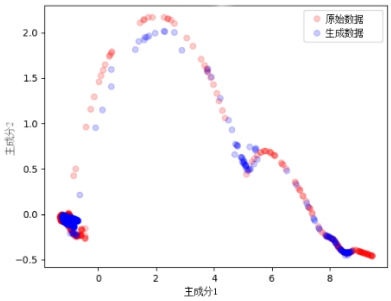


图 6 原始数据与生成数据 PCA 对比
Fig.6 Comparison of PCA results between original and generated data

进一步利用 t-SNE 进行可视化，图 7 为原始数据与生成数据 t-SNE 对比。由图 7 可加，尽管 2 类数据在局部分布上仍有微小差别，但整体上生成数据已较好地分布在与真实数据相近的区域，且在非线性降维空间中依然呈现出类似的群落与结构特点。这为前述预测与判别指标的定量结果提供了直观的佐证。

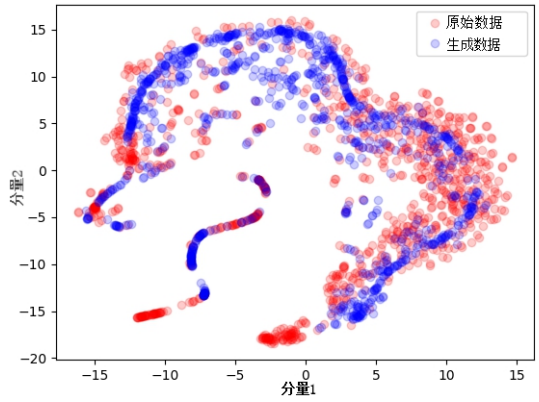


图 7 原始数据与生成数据 t-SNE 对比
Fig.7 Comparison of t-SNE results between original and generated data

综上，TimeGAN 模型在所采用的燃机气路故障数据集上的表现优良。预测指标表明生成数据在时间序列特性上与原始故障数据相近，能够较好地再现其动态变化规律；判别指标表面生成数据在统计分布上与真实数据具有较高相似性；可视化指标则进一步表明生成数据在低维投影空间中与真实数据的聚类和分布模式基本吻合。上述指标相互印证，从多角度验证了本文所构建的 TimeGAN 模型在燃机气路故障模拟数据生成任务上的可行性与有效性。

4 结 论

本文针对燃机气路故障样本不平衡、数据分布复杂且存在多源非线性时间依赖等问题，搭建了基于 TimeGAN 的高质量燃机气路模拟故障数据生成模型，并对模型的关键参数与结构设置进行了优化与验证。在系统性分析和实证研究的基础上，得出以下主要结论：

1）将 TimeGAN 用于燃机气路故障数据模拟生成的关键在于充分利用 LSTM 神经网络等 RNN 类神经网络的时间依赖建模能力与 GAN 框架的对抗训练机制。通过引入嵌入、生成、判别及重构等多模块协同作用，本文实现了对真实故障数据时序特征的有效学习与逼真再现。该方法具备通用性，可为其他复杂工业设备的故障数据增强提供参考。

2）在降维环节引入 PCA 及后续可视化分析（t-SNE），帮助深入理解数据特征，证明前期特征工程与主成分提取能够有效保留故障数据的关键信息。经降维映射后，TimeGAN 生成数据与真实数据在主成分空间中的分布基本吻合，表明所采用的方法对高维数据结构的优异拟合能力。

3）通过由预测指标（LSTM 模型一阶预测实验）、判别指标（基于后验 RNN 判别器的分类准确率）以及可视化指标（PCA 与 t-SNE 可视化）构建的多维度评价体系评价模型生成数据的质量。结果显示，TimeGAN 生成的数据在分布拟合度、时间序列预测性能以及低维特征空间的聚类形态方面均接近真实数据。多重证据相互印证，为生成数据的高质量特征表达与应用价值提供了有力支持。

[参 考 文 献]

[1] 陈蕊，朱博骐，段天宇. 天然气发电在我国能源转型中的作用及发展建议[J]. 天然气工业，2020, 40(1): 120-128.
CHEN Rui, ZHU Boqi, DUAN Tianyu. Role of natural gas power generation in China's energy transformation and suggestions on its development[J]. Natural Gas Industry, 2020, 40(1): 120-128.

[2] 林志文，冯景浩，杨承，等. PG 9351FA 燃气-蒸汽联合循环机组性能劣化研究[J]. 热力发电，2022, 51(2): 132-141.
LIN Zhiwen, FENG Jinghao, YANG Cheng, et al. Investigation on performance degradation of PG 9351FA gas-steam combined cycle unit[J]. Thermal Power Generation, 2022, 51(2): 132-141.

[3] GIBSON N. A prognostic and diagnostic approach to engine health management[C]. ASME Turbo Expo: Power for Land, Sea, & Air, 2006: 673-680.

[4] DIAKUNCHAK I S. Performance deterioration in industrial gas turbines[J]. Journal of Engineering for Gas Turbines and Power, 1992, 114(2):161-168.

[5] KURZ R, BRUN K. Degradation in gas turbine systems[J]. Journal of Engineering for Gas Turbines and Power, 2000, 123(1): 70-77.

[6] 许未晴，冀守虎，安永伟，等. 机器学习算法在重型燃气轮机健康监测的应用现状[J]. 液压与气动，2023, 47(4):71-86.
XU Weiqing, JI Shouhu, AN Yongwei, et al. Application status of machine learning algorithms in health monitoring of heavy-duty gas turbines[J]. Hydraulic and Pneumatic, 2023, 47(4): 71-86.

[7] New findings from institute of networked and embedded systems in the area of machine learning published (topology-based machine learning: predicting power line communication quality in smart grids)[J]. Robotics & Machine Learning Daily News, 2023(4): 22.

[8] 孙守泰，薛亚丽，王明春，等. 基于混合深度学习的燃

气轮机动态过程关键参数在线辨识[J]. 中国舰船研究, 2023, 18(3): 222-230.
SUN, S.; XUE, Y.; WANG, M.; Sun, L. Online identification of key dynamic parameters for gas turbine processes based on hybrid deep learning. Chinese Journal of Ship Research, 2023, 18(3): 222-230.

[9] 彭道刚, 姬传晟, 涂煊, 等. 基于 LSTM-SVM 的燃气轮机压气机故障预警研究 [J]. 动力工程学报, 2021, 41(5): 394-399.
PENG Daogang, JI Chuansheng, TU Xuan, et al. Research on gas turbine compressor fault early warning based on LSTM-SVM[J]. Journal of Chinese Society of Power Engineering, 2021, 41(5): 394-399.

[10] 蒋龙陈, 王红军, 张顺利. 燃气轮机气流激振深度信念网络故障诊断模型[J]. 电子测量与仪器学报, 2021, 35(2): 115-121.
JIANG Longchen, WANG Hongjun, ZHANG Shunli. Gas turbine flow excitation fault diagnosis model based on deep belief network[J]. Journal of Electronic Measure-ment and Instrumentation, 2021, 35(2): 115-121.

[11] 闫丽萍, 董学智, 张永军, 等. 基于深度置信网络的燃气轮机气路故障诊断方法[J]. 工程热物理学报, 2020, 41(4): 840-844.
YAN Liping, DONG Xuezhi, ZHANG Yongjun, et al. Gas path fault diagnosis method for gas turbines based on deep belief network[J]. Journal of Engineering Thermophysics, 2020, 41(4): 840-844.

[12] LIU J, LIU J, YU D, et al. Fault detection for gas turbine hot components based on a convolutional neural network [J]. Energies, 2018, 11(8): 2149.

[13] CAO Q, CHEN S, ZHANG Y, et al. Classification and prediction of gas turbine gas path degradation based on deep neural networks[J]. International Journal of Energy Research, 2021, 45(7): 10513-10526.

[14] JIANG X Y, GE Z Q. Data augmentation classifier for imbalanced fault classification[J]. IEEE Transactions on Automation Science and Engineering, 2021, 18(3): 1206-1217.

[15] 闫啸家, 梁伟阁, 张钢, 等. 非均衡小样本条件下基于 SAE-ACGANs 的复杂供输机构故障诊断方法[J]. 振动与冲击, 2023, 42(2): 89-99.
YAN Xiaojia, LIANG Weige, ZHANG Gang, et al. Fault diagnosis method for complex feeding and ramming mechanisms based on SAE-ACGANs with unbalanced limited training data[J]. Journal of Vibration and Shock, , 2023, 42(2): 89-99.

[16] YI H K, JIANG Q C, YAN X F, et al. Imbalanced classification based on minority clustering synthetic minority oversampling technique with wind turbine fault detection application[J]. IEEE Transactions on Industrial Informatics, 2021, 17(9): 5867-5875.

[17] ZHUO Y, GE Z Q. Auxiliary information-guided industrial data augmentation for any-shot fault learning and diagnosis[J]. IEEE Transactions on Industrial Informatics, 2021, 17(11):7535-7545

[18] ZHANG J X, LI Z G F, CHEN H X, et al. Integrated generative networks embedded with ensemble classifiers for fault detection and diagnosis under small and imbalanced data of building air condition system[J]. Energy & Buildings, 2022.

[19] Zhou X y, Niu S j, Li X h, et al. Spatial–contextual variational autoencoder with attention correction for anomaly detection in retinal OCT images[J]. Computers in Biology and Medicine, 2023.

[20] GUO Y B, ZHOU D M, RUAN X l, et al.Variational gated autoencoder-based feature extraction model for inferring disease-miRNA associations based on multiview features[C]//Neural networks : the official journal of the International Neural Network Society, 2023.

[21] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein GAN[EB/OL]. (2017-12-06) [2025-02- 01]. https://arxiv.org/abs/1701. 07875.

[22] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of Wasserstein GANs. [EB/OL]. (2017-12-25)[2025-02-01]. https://arxiv.org/abs/1704. 00028.

[23] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. J. Artif. Int. Res., 2002, 16(1): 321-357.

[24] 孙晨峰, 吕卫民, 戴洪德, 等. 一种基于 TimeGAN 和 OCSVM 的多元退化设备小子样数据增广方法[J]. 电子学报, 2022, 50(11): 2678-2687.
SUN Chenfeng, LYU Weimin, DAI Hongde, et al. A small sample data augmentation method for multivariate degradation equipment based on TimeGAN and OCSVM[J]. Acta Electronica Sinica, 2022, 50(11): 2678-2687.

[25] SHANGGUAN A, XIE G, FEI R, et al. Train wheel degradation generation and prediction based on the time series generation adversarial network[J]. Reliability Engineering & System Safety, 2023, 229: 108816.

[26] YOON J, JARRETT D, SCHAAR M V D. Time-series generative adversarial networks[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: NeurIPS, 2019: 5508-5518.

[27] 庞传军, 张波, 余建明. 基于 LSTM 循环神经网络的短期电力负荷预测[J]. 电力工程技术, 2021, 40(1): 175-180.

[28] 周锐, 康英伟. 基于 CNN-LSTM 的燃气轮机燃烧室故障预警[J]. 热能动力工程, 2024, 39(1): 191-197.
ZHOU Rui, KANG Yingwei. Fault warning of gas turbine combustion chamber based on CNN-LSTM[J]. Thermal Energy Power Engineering, 2024, 39(1): 191-197.

（责任编辑 杜亚勤）