

# 基于损失优化的DAE降低风电机组 不确定因素的影响

沈俞恒

浙江理工大学机械工程学院, 浙江 杭州

收稿日期: 2023年3月20日; 录用日期: 2023年5月22日; 发布日期: 2023年5月29日

## 摘要

考虑到风电机组运行过程中存在大量的不确定因素, 这些不确定因素会导致监测参数出现质量波动, 进而影响风电机组的异常检测结果或状态评估。为了解决不确定因素对风电机组监测参数的影响, 提出一种基于损失优化的深度自编码器模型, 降低风电机组中不确定因素的影响。该模型结合信息瓶颈理论, 并以最小充分统计变量的存在条件为损失函数来训练。从而实现在对数据压缩重构的过程中, 去掉不确定因素带来的干扰信息, 同时最大程度上地保留有用的信息, 以降低监测参数的波动。通过实际案例表明该方法的有效性, 为后续的高精度的异常状态检测和状态评估奠定良好的基础。

## 关键词

风电机组主传动系统, 不确定因素, 损失函数, 基于损失优化的深度自编码

# Reducing the Impact of Uncertain Factors on Wind Turbine Units Based on Loss-Optimized DAE

Yuheng Shen

School of Mechanical Engineering, Zhejiang Sci-Tech University, Hangzhou Zhejiang

Received: Mar. 20<sup>th</sup>, 2023; accepted: May 22<sup>nd</sup>, 2023; published: May 29<sup>th</sup>, 2023

## Abstract

Considering that there are a large number of uncertain factors in the operation of wind turbines, these uncertain factors will lead to quality fluctuations in monitoring parameters, which will affect

the abnormal detection results or status evaluation of wind turbines. In order to solve the influence of uncertain factors on the monitoring parameters of wind turbines, a deep autoencoder model based on loss optimization is proposed to reduce the influence of uncertain factors in wind turbines. The model incorporates information bottleneck theory and is trained with the existence condition of the minimum sufficient statistical variable as the loss function. In this way, in the process of data compression and reconstruction, the interference information caused by uncertain factors is removed, while the useful information is retained to the greatest extent, so as to reduce the fluctuation of monitoring parameters. The effectiveness of the method is shown through practical cases, which lays a good foundation for subsequent high-precision abnormal state detection and state evaluation.

## Keywords

Wind Turbine Main Drive System, Uncertain Factors, Loss Function, Deep Self-Encoding Based on Loss Optimization

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着环境问题的日益严重，各国对环境的保护意识和重视程度越来越高。在“绿色经济，可持续发展”的大环境下，风能作为一种可再生的清洁能源，成为各国的重点关注的资源。风电机组作为风力发电的主要设备，由于其长期处于恶劣工作环境下，导致故障频发。而风电机组主传动系统不仅故障发生率占比高，而且因其故障导致给风机停机的时长占比也是最高[1]，带来巨大的经济和时间损失。因此，作为风电机组能量传输的重要子部件，如何实现对主传动系统及时有效的监测，提前检测异常，成为维持风电机组稳定运行、降低故障损失以及提高效益的关键问题。

针对风电机组故障频发的问题，国内外不少研究学者已将开展了很多的工作。例如，Gray 和 Watson [2]建立了一个理论损伤模型，用于描述风电机组操作环境和损伤累计率之间的关系，以此来判断风电机组运行状态的劣化情况。胡阳等人[3]为了能够实时了解风电机组的运行状态，建立了风电机组的半数字孪生仿真系统，用于模拟实时运行状态。Huanguo Chen 等人针对[4]提出一种基于 K-Means 算法，对风电机组运行工况多次划分后，实现对风电机组运行状态的监测。Yongchao Zhu [5]提出一种基于 LSTM 框架的模糊综合评价方法，旨在降低风电机组的 O&M 成本。Jialiang Zhang [6]针对浮式海上风电机组(FOWT)，用故障树分析法对 FOWT 的运行状态进行量化分析。Ferguson [7]等人使用标准化的数据温度，结合速度和发电机扭矩，检测齿轮箱中的轴承故障。文献[8] [9] [10]中的学者利用支持向量机及其变体网络，通过 SCADA 数据对风电机组不同部件的状态进行状态评估。

虽然，上述针对风电机组运行状态监测的方法，都取得了不错的效果。但是，少有学者考虑到风电机组的不确定性因素对于风电机组的影响。风电机组的运行过程中，存在许多不确定的因素，如传感器故障或采集精度的误差、外部环境的不确定以及自身耦合性过强等因素。这些不确定因素导致监测参数质量出现不确定性，进而影响对风电机组的监测准确度和可靠性。针对上述研究方法存在的不足，本文以提出一种基于损失优化的 DAE 神经网络模型，通过对数据的编码压缩成最小充分统计变量的过程，削减不确定因素的信息干扰影响，尽可能地保留有用的特征信息。随后将数据重构，以达到削减监测参数

中不确定性因素带来的干扰信息影响。

## 2. 风电机组主传动系统及其不确定因素

由于风电机组运行过程中存在很多的不确定因素，这些不确定因素会主传动系统的正常运行带来影响，接下来我们将介绍对主传动系统运行影响较大的不确定因素，并分析其成因。

### 1) 传感器采集精度和数据传输质量的不确定

由于传感器采集精度的误差和数据传输过程中的质量难以把控，导致从现场采集的状态监测数据的质量受到影响后，进而导致数据波动增大，干扰信息增多。

### 2) 外部环境因素的不确定

外部环境因素主要是指风速、温度、湿度以及噪声。由于环境因素是不可把控的。环境参数的波动会导致风电机组的 SCADA 监测参数的波动，甚至有可能被误判为异常。

### 3) 风电机组自身因素的不确定

风电机组是一个耦合性较强的非线性系统，由于其较强的耦合性，当其中一个部件发生异常时，不仅会对这个部件的监测参数带来影响，还会影响相关性较强的部件的状态监测参数。这种现象称为“传递效应”，可能导致出现“假异常”的误报情况。

上面所述的不确定因素会直接影响风电机组的状态监测参数，导致状态监测参数波动变大、干扰信息增多，与能够反映风电机组和真实运行状态的数据产生偏差。

## 3. 基于损失优化的 DAE-BiGRU 异常检测模型

### 3.1. 信息瓶颈理论

根据信息瓶颈理论[11] (Information Bottleneck)是由 Naftali Tishby 在 2000 年提出的，它是一种解释深度学习工作方式的理论。该理论可以完美地诠释自编码器的工作过程，信息瓶颈理论认为神经网络工作过程可以分为两个步骤：特征压缩和特征重构，如同沙漏工作方式一样。信息瓶颈框架如图 1 所示，在特征压缩阶段，网络将输入  $X$  压缩成中间变量  $Z$ ，然后在特征重构阶段将中间变量  $Z$  重构特征  $Y$ 。

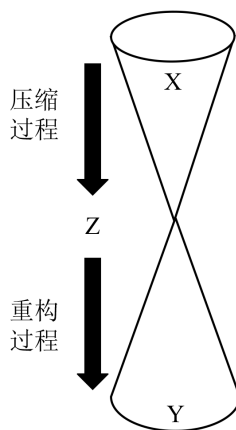


Figure 1. Information bottleneck theory framework

图 1. 信息瓶颈理论框架

其核心理论计算如公式(1):

$$U_{IB}(\sigma) = -\alpha I(Z, X, \sigma) + I(Z, Y, \sigma) \quad (1)$$

式中,  $U_B(\sigma)$  即信息瓶颈,  $\sigma$  为网络参数,  $I(Z, Y, \sigma)$  为输入  $X$  和中间潜在变量  $Z$  的互信息[12],  $I(Z, Y, \sigma)$  是潜在变量  $Z$  和输出信息  $Y$  之间的互信息,  $\alpha$  权值。

信息瓶颈理论的核心在于同时最大化  $Z$  和  $Y$  之间的互信息, 以及最小化  $X$  和  $Z$  之间的互信息。这样可以确保  $Z$  尽可能地保留了  $X$  中的有用特征, 而过滤掉了冗余或无关的信息。 $Z$  被称为“最小充分统计变量”, 因为它在保留最重要的特征信息的同时, 尽可能地减少了信息的冗余和干扰。在特征重构阶段, 基于  $Z$  的特征, 网络可以重构出与  $X$  最相关的特征  $Y$ , 从而实现了原始数据的有效编码和解码。

### 3.1. 最小充分统计变量

尽管信息瓶颈理论中提出了最小充分统计变量存在的概念, 但在实际工程中的可行性仍存在疑问。Steven Ding [13] 基于信息瓶颈理论, 利用非线性系统的图像表示和子空间, 并结合 Hamiltonian Extension 和内部分析的方法, 成功证明了“最小充分统计变量”在实际工程中的非线性系统中的存在, 并提出内循环自编码器 (Inner-Autoencoder, I-AE) 用于非线性系统的故障诊断, 这项工作为信息瓶颈理论的实际应用提供了有力的支持, 并为进一步的研究和应用奠定了基础。

风电机组的主传动系统作为典型非线性系统之一, 本文结合 Seven Ding 提出的“最小充分统计变量”在非线性系统中存在的证明, 推导风电机组主传动系统中, “最小充分统计变量”的存在条件。

风电机组主传动系统的非线性过程模型, 如式(2):

$$\sum: \begin{cases} \dot{u}(t) = f[u(t), x(t)] \\ y(t) = g[u(t), x(t)] \end{cases} \quad (2)$$

式中,  $u$  表示状态向量,  $x$  为过程输入,  $y$  为过程输出,  $f$ 、 $g$  为非线性连续函数。

考虑到风电机组运行过程存在不确定性影响, 因此过程模型中需要加上不确定性扰动  $d$  影响, 如式(3):

$$\sum: \begin{cases} \dot{u}(t) = A \cdot u(t) + B \cdot x(t) + C \cdot d(t) \\ y(t) = D \cdot u(t) + E \cdot x(t) + F \cdot d(t) \end{cases} \quad (3)$$

式中,  $u$  表示状态向量,  $x$  为过程为反映风机运行状态的真实数据,  $y$  为过程输出,  $d$  为风机运行过程中的不确定性因素产生的影响,  $A$ 、 $B$ 、 $C$ 、 $D$ 、 $E$ 、 $F$  为非线性连续函数。

根据信息瓶颈理论, 引入最小充分统计变量  $v$ , 将主传动系统仿射模型改写成式(4):

$$\begin{bmatrix} \widehat{x(t)+d(t)} \\ \widehat{y(t)} \end{bmatrix} = \Pi \begin{bmatrix} x(t)+d(t) \\ y(t) \end{bmatrix} \quad (4)$$

式中,  $\Pi$  表示映射关系。

继续将公式(4)改写为公式(5):

$$\Pi \begin{bmatrix} x(t)+d(t) \\ y(t) \end{bmatrix} = \sum_D \cdot \sum_E \left( \begin{bmatrix} x(t)+d(t) \\ y(t) \end{bmatrix} \right) = \sum_D(v) = \begin{bmatrix} \widehat{x(t)+d(t)} \\ \widehat{y(t)} \end{bmatrix} \quad (5)$$

式中  $\sum_D$  为信息瓶颈理论中所提的特征压缩过程,  $\sum_E$  为特征重构过程,  $v$  为最小充分变量。

根据相关文献的推导和所需条件, 已证明最小充分统计变量在非线性系统中具有重要意义。基于互信息理论, 推导出最小充分变量在风电机组主传动系统中的条件, 如式(6):

$$\left\{ \begin{array}{l} I\left(\left[\begin{array}{c} x(t)+d(t) \\ y(t) \end{array}\right], v^1\right)=0 \\ I\left(v^1, \left[\begin{array}{c} \widehat{x(t)}^1 + \widehat{d(t)}^1 \\ \widehat{y(t)}^1 \end{array}\right]\right)=1 \\ I(v^1, v^2)=1 \\ I\left(\left[\begin{array}{c} \widehat{x(t)}^1 + \widehat{d(t)}^1 \\ \widehat{y(t)}^1 \end{array}\right], \left[\begin{array}{c} \widehat{x(t)}^2 + \widehat{d(t)}^2 \\ \widehat{y(t)}^2 \end{array}\right]\right)=1 \end{array} \right. \quad (6)$$

式中,  $v^1$  指第一次编码后的最小充分统计变量,  $\widehat{x(t)}^1 + \widehat{d(t)}^1$  为第一次重构后的输出,  $v^2$  为第二次编码后

的最小充分统计变量,  $\widehat{x(t)}^2 + \widehat{d(t)}^2$  为第二次重构后的输出。

式(6)同时也说明了, 当存在最小充分变量存在时, 此时系统的输入变量和最小充分统计变量互信息为 0, 而最小充分统计变量和输出变量的互信息值最大。这意味着数据在压缩过程中, 最小充分统计变量最大程度上的保留了输入变量中与输出变量最相关的特征信息, 减少了与输出变量不相关的信息。

### 3.3. 基于损失优化的深度自编码器

自编码器是由编码层和解码层两个部分组成的神经网络结构, 核心思想是通过学习稀疏的高阶特征来重构输入数据。具体而言, 自编码器先利用编码层将高维输入向量压缩成低维潜在变量, 通过解码层将潜在变量映射成原始样本数据形式, 并以最小化重构前后数据的差异作为训练目标。通过这种方式, 自编码器可以有效地提取数据特征, 可以用于数据压缩降维和生成等任务。

深度自编码器是一种深度神经网络模型, 相比普通自编码器, 则是由多层编码层和解码层构成, 相对于单层自编码器, 通过对输入数据不断地压缩和提取特征, 可以更好地提取数据中的特征, 通过多层解码器可以更准确地重构数据, 具有不错地特征提取和数据降维功能。单层自编码器和深度自编码器的结构如图 2 和图 3 所示。

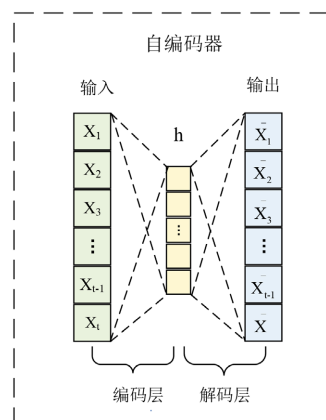


Figure 2. Autoencoder  
图 2. 自编码器

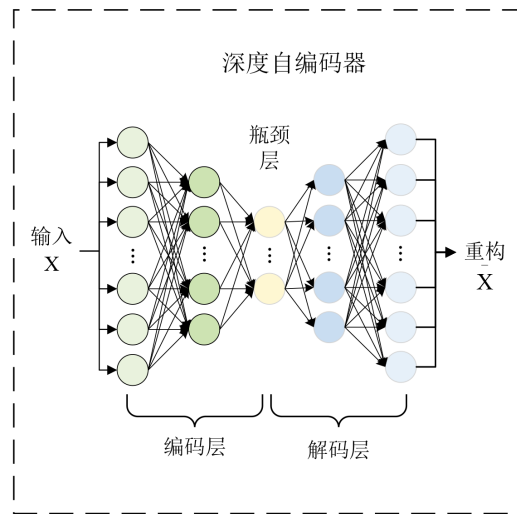


Figure 3. Deep-autoencoder  
图 3. 深度自编码器

式(6)给出了最小充分统计变量等价于自编码器潜在变量的条件,为了更好地应用于自编码器,将等价条件作为自编码器的损失函数,提出一种基于损失函数优化的深度自编码器模型。通过对输入数据进行编码和解码,使得该模型可以从原始数据中学习到具有最小特征损失的表示,实现对数据的有效特征提取,剔除干扰信息。

损失函数一的计算公式如式(7):

$$loss_1(\varepsilon, d(e(\varepsilon))) = \frac{1}{n} \sum_{t=1}^n \left\| \begin{bmatrix} x(t)+d(t) \\ y(t) \end{bmatrix} - d \left( e \left( \begin{bmatrix} x(t)+d(t) \\ y(t) \end{bmatrix} \right) \right) \right\|_2^2 \quad (7)$$

损失函数二的计算公式如式(8):

$$loss_2(\varepsilon, d(e(\varepsilon))) = \frac{1}{n} \sum_{t=1}^n \left\| \begin{bmatrix} \widehat{x(t)} + \widehat{d(t)} \\ \widehat{y(t)} \end{bmatrix} - d \left( e \left( \begin{bmatrix} \widehat{x(t)} + \widehat{d(t)} \\ \widehat{y(t)} \end{bmatrix} \right) \right) \right\|_2^2 \quad (8)$$

损失函数三的计算公式如式(9):

$$loss_3(e, d(e(e))) = \frac{1}{n} \sum_{t=1}^n \left\| e \left( \begin{bmatrix} x(t)+d(t) \\ y(t) \end{bmatrix} \right) - e \left( \begin{bmatrix} \widehat{x(t)} + \widehat{d(t)} \\ \widehat{y(t)} \end{bmatrix} \right) \right\|_2^2 \quad (9)$$

总的损失函数  $Loss$  为公式(10)所示:

$$Loss = loss_1 + loss_2 + loss_3 \quad (10)$$

通过损失函数  $Loss$  来训练深度自编码器模型,由于添加了正则项,可以提高模型的稳定性、减少过拟合,并增强模型的可解释性和可理解性。基于损失优化的深度自编码器模型,结合信息瓶颈理论和最小充分统计量的概念,能够降低风电机组主传动系统中不确定因素带来的负面影响,可以让重构后的数据能更好的反应主传动系统的运行状态,图4展示了基于损失优化的深度自编码器的网络结构图。

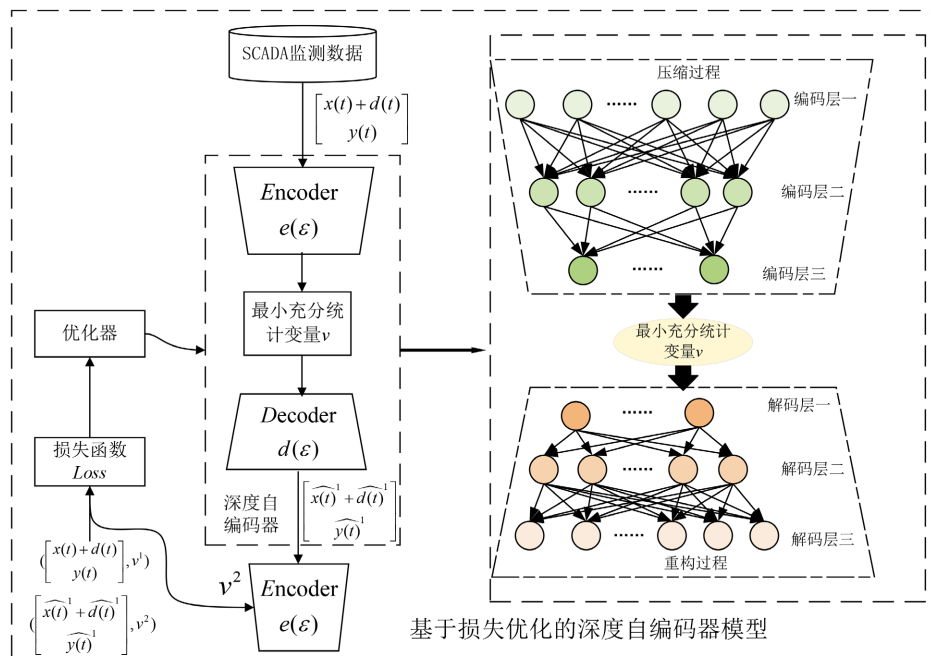


Figure 4. Loss-optimized deep autoencoder model  
图 4. 基于损失优化的深度自编码器模型

## 4. 方法验证

为了验证本文提出的 L-DAE 模型能否减小风电机组主传动系统不确定因素对数据质量的影响, 尤其是那些表面看起来没什么问题, 但在反映风电机组主传动系统真实运行状态会存在一定偏差的数据。本文从湖南某风场采集风电机组再正常运行期间的历史数据, 并经过数据预处理后构造成实例分析的数据集。本文选取 80% 的数据用于训练基于损失优化的深度自编码模型, 并且用剩下的 20% 的数据(共 10,000 个点)来验证该模型。

### 4.1. 训练损失对比

对于基 L-DAE 的网络结构和参数, 是通过经过多次试验对后选出的相对最优选项。首先, 编码层结构与解码层的结构呈镜像对称设置的; 其次, 输入特征参数的个数决定了输入层神经元个数; 最后, 因为 L-DAE 模型的本质, 是对输入特征参数的编码压缩和解码复原, 通过这个过程来降低不确定性因素对监测参数的质量影响。因此输出层的神经元个数应当与输入层神经元个数相同。模型参数如表 1 所示。

Table 1. Parameters of the L-DAE model

表 1. L-DAE 器模型参数

参数	参数设置
编码层结构	16-10-5
解码层结构	5-10-16
输出层神经元个数	16
损失函数	Loss
Epoch	1000
优化器	Adam
激活函数	RELU

为了验证本文提出 L-DAE 模型在实际应用中的优势。比较了 L-DAE 和 I-AE 模型在训练损失方面的表现。如图 5 所示。

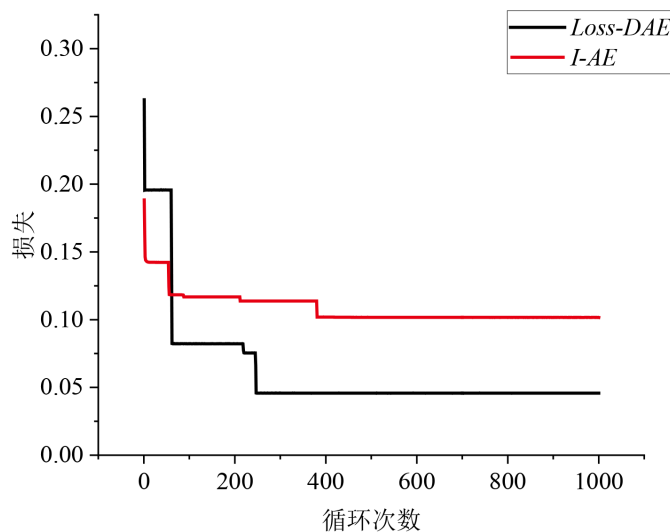
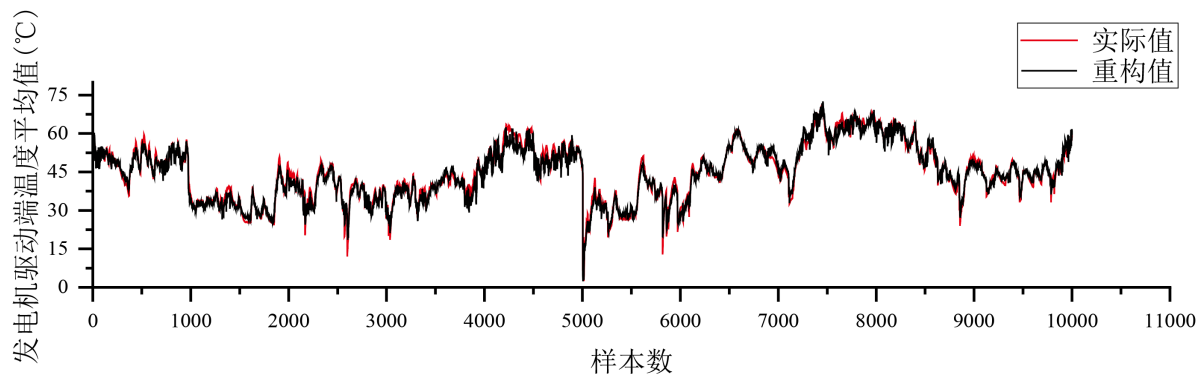


Figure 5. Comparison of training losses  
图 5. 训练损失对比

从图 5 可以看出，本文所提模型在经过多次的损失下降，模型才能完全收敛，大约用了将近 250 个循环，而 I-AE 模型完全收敛却用了将近 400 个循环，本文所提模型在收敛速度上明显要比 I-AE 快，并且模型的收敛精度明显要高于 I-AE 模型。从整体的训练过程来看，本文所提的基于损失优化的深度自编码器模型表现出了不错的效果。

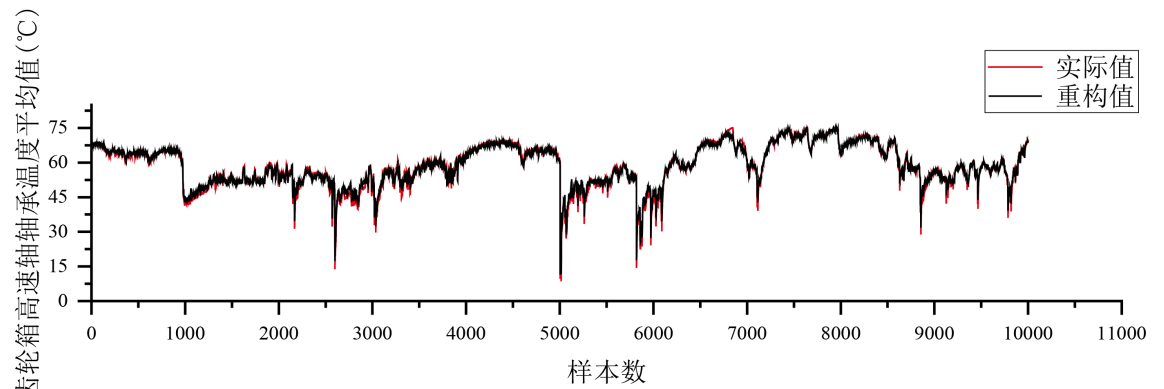
#### 4.2. 效果分析

为了进一步验证模型的有效性，本文将数据集中 20% (共 10,000 个数据) 作为验证集，用于评估模型的泛华性能。提取主传动系统中用于监测发电机驱动端平均温度、齿轮箱高速轴轴承平均温度和有功功率三个参数的重构值和实际值进行对比，并绘制了验证结果图，并对结果进行分析。验证结果如图 6 所呈现的：

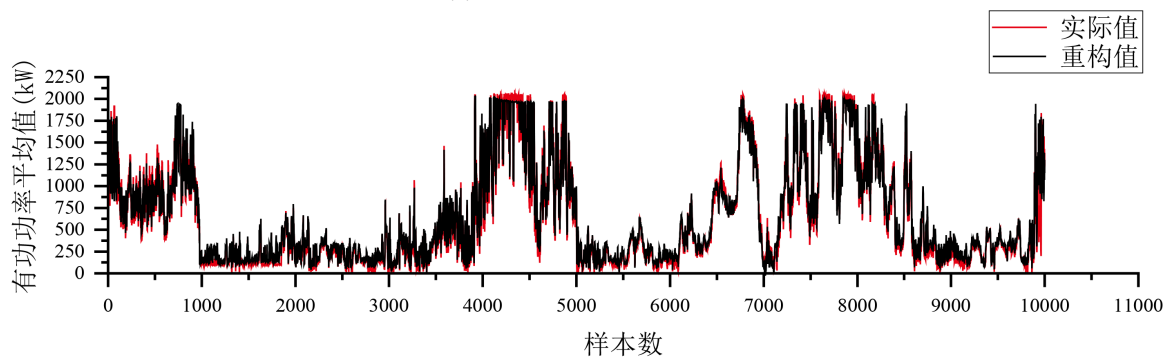


(a) 发电机驱动端温度





(b) 齿轮箱高速轴轴承温度



(c) 有功功率

**Figure 6.** Reduction of the impact of uncertain factors**图 6.** 减少不确定性影响的效果

图 6 中(a)、(b)、(c)分别展示发电机驱动端温度平均值、齿轮箱高速轴轴承温度平均值和有功功率的重构值和实际值的对比。通过比较,我们可以发现,经过 L-DAE 模型重构后的数据和实际数据的变化趋势基本上是一致的。但是相比于实际值,重构的数据明显更加平滑,波动幅度更小。这是由于主传动系统运行过程中存在各种不确定因素,这些不确定因素会影响数据质量,导致数据波动变大。而本文提出的模型,通过根据信息瓶颈理论设计损失函数进行训练,将输入数据压缩成最小充分统计变量,最小化输入数据与最小成分统计变量之间的互信息,最大程度上去除了不确定因素和无用数据的干扰,从而重构出更加平滑集中的数据。这证明本文所提的模型能够在处理主传动系统不确定性问题时,有效地去除不确定因素的干扰信息,提高数据重构的精度和准确性。

### 4.3. 对比分析

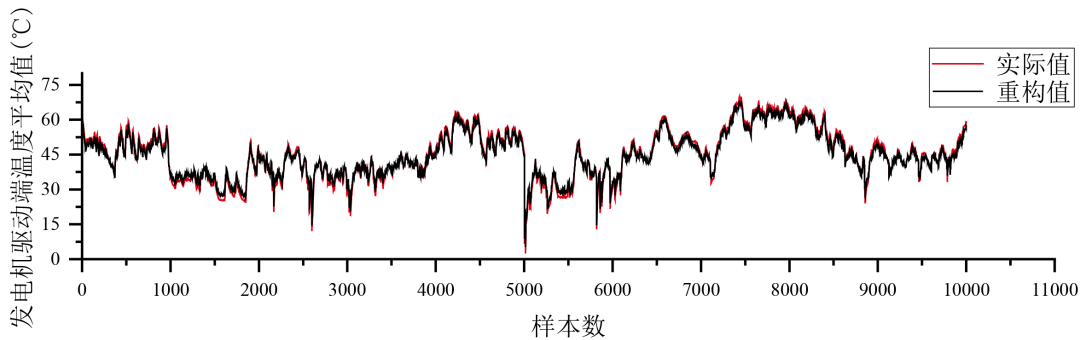
为了进一步验证本文提出模型降低不确定因素干扰的效果,对采集到数据加入高斯噪声,构造成含有不确定因素干扰的数据集,并用 L-DAE 与 I-AE 分别对数据进行重构,对比结果如图 7 所示,残差对比如图 8~10 所示。

采用平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)来比较两种模型在含有不确定因素影响情况下的重构精度,计算公式如(3)~(14)所示:

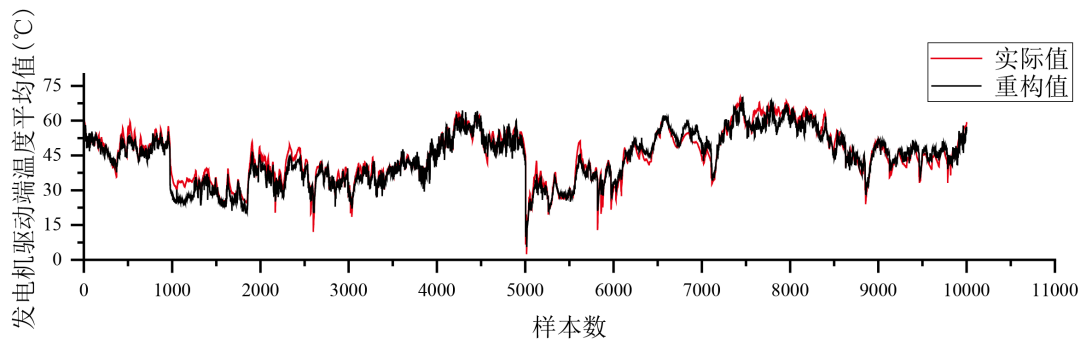
$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{x}_i - x_i}{x_i} \right| \times 100\% \quad (11)$$

式中  $n$  为样本个数,  $\hat{x}_i$  为重构值,  $x_i$  为原始值, MAPE 值越小代表重构效果越好。

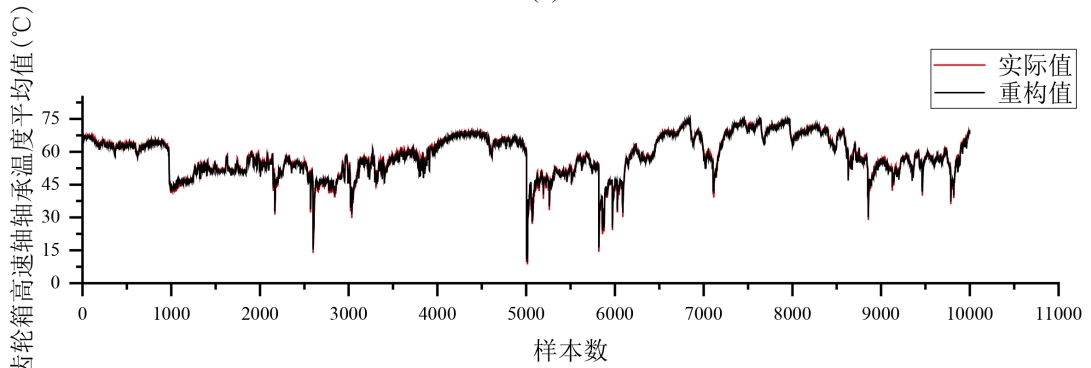
根据式(11)分别计算出 L-DAE 和 I-AE 的 MAPE 值, 如表 2 所示。



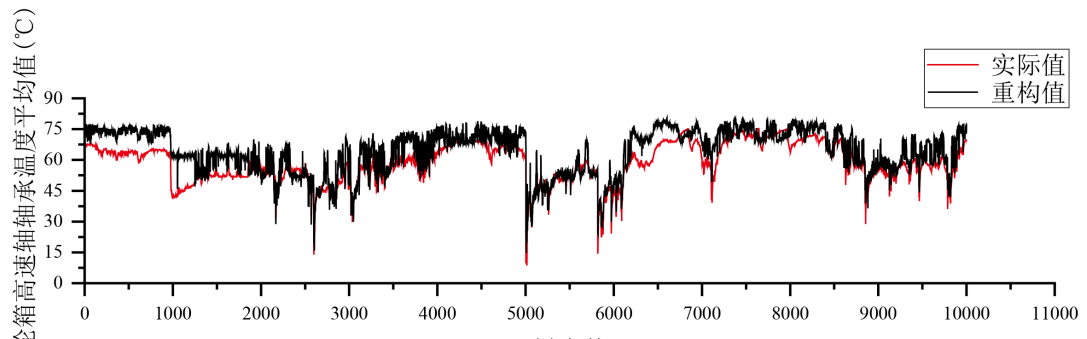
(a) L-DAE



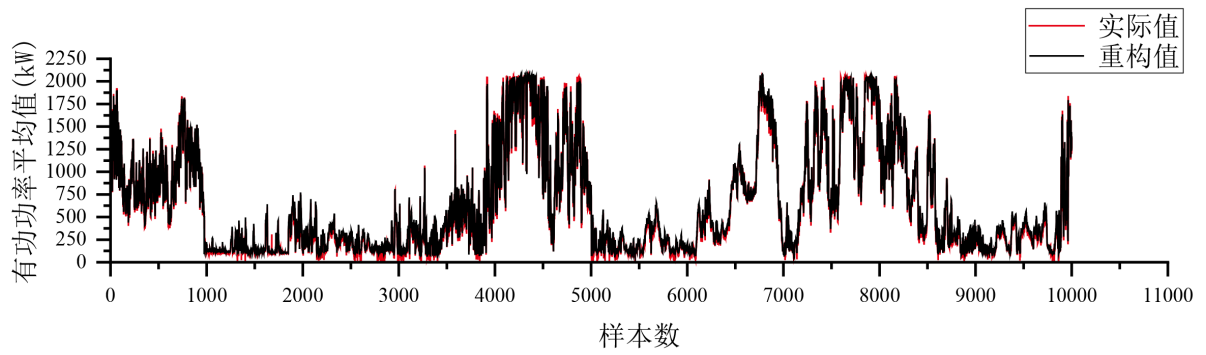
(b) I-AE



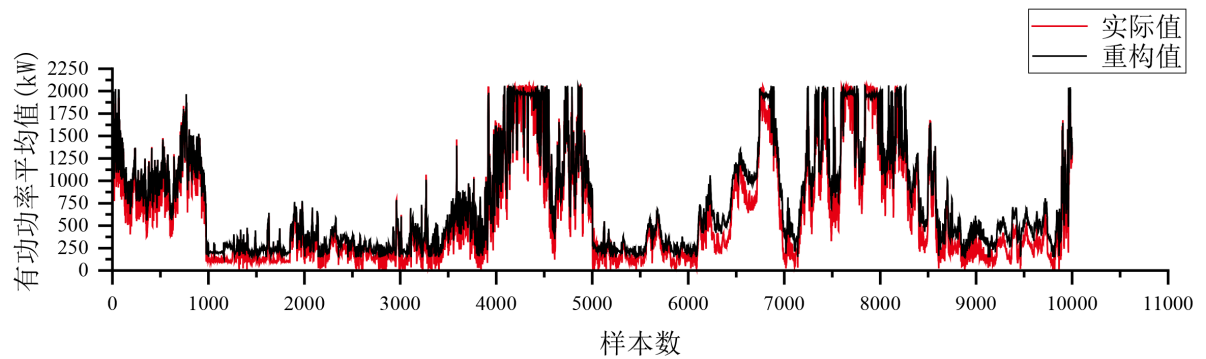
(c) L-DAE



(d) I-AE



(e) L-DAE



(f) I-AE

Figure 7. Effect comparison

图 7. 效果对比

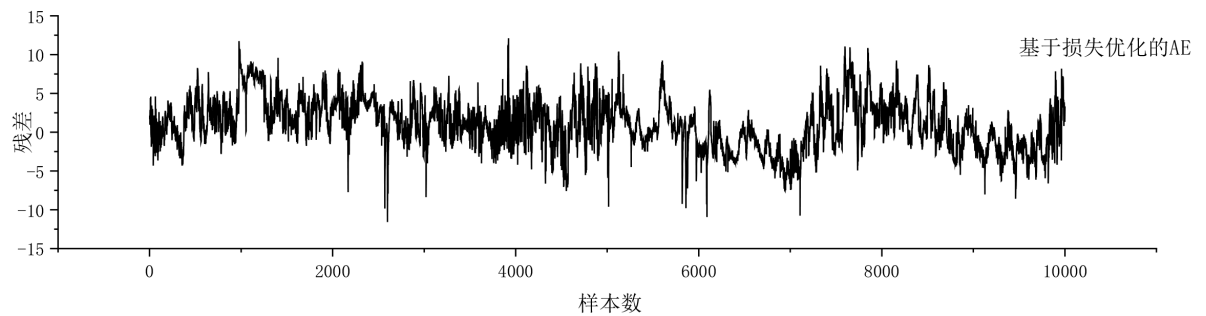
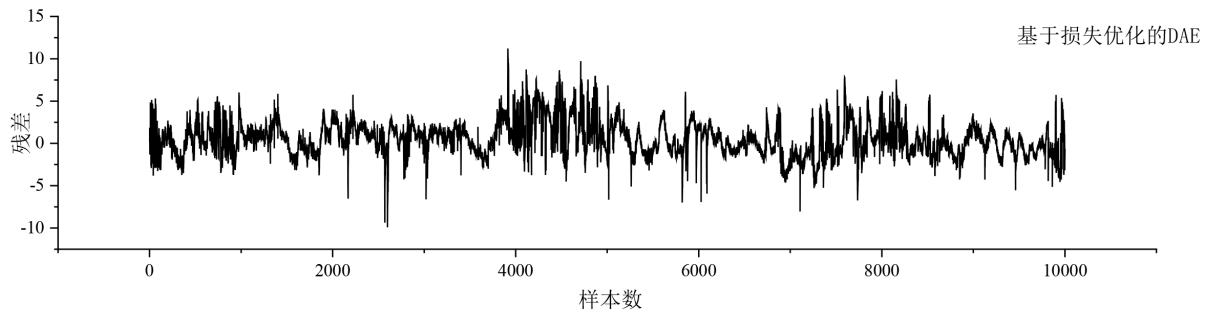


Figure 8. Generator drive-end temperature

图 8. 发电机驱动端温度

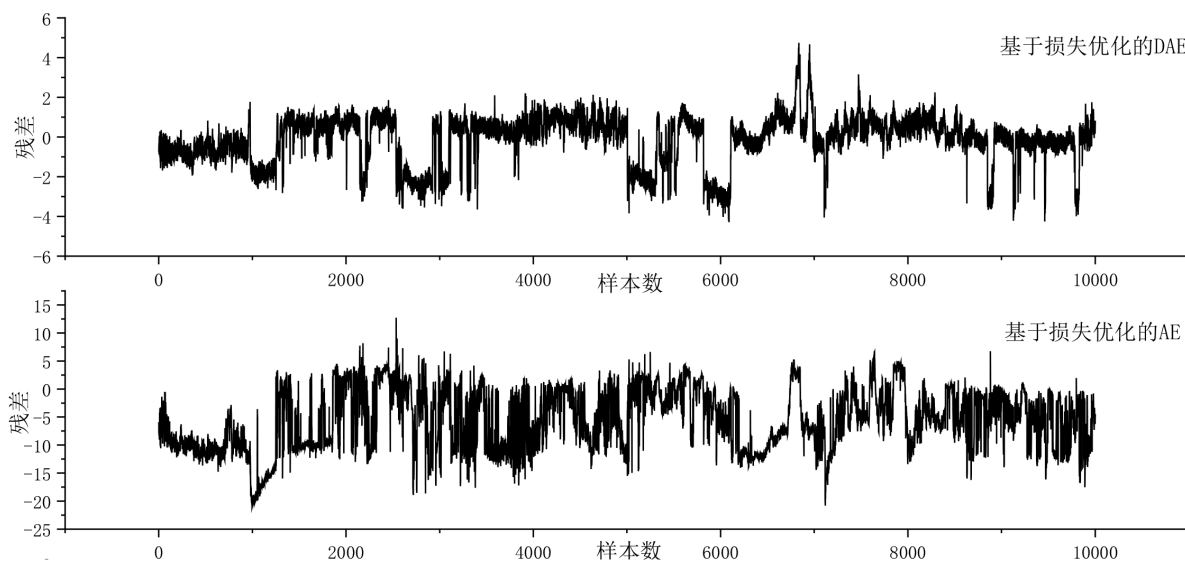


Figure 9. Gearbox high speed shaft bearing temperature

图 9. 齿轮箱高速轴轴承温度

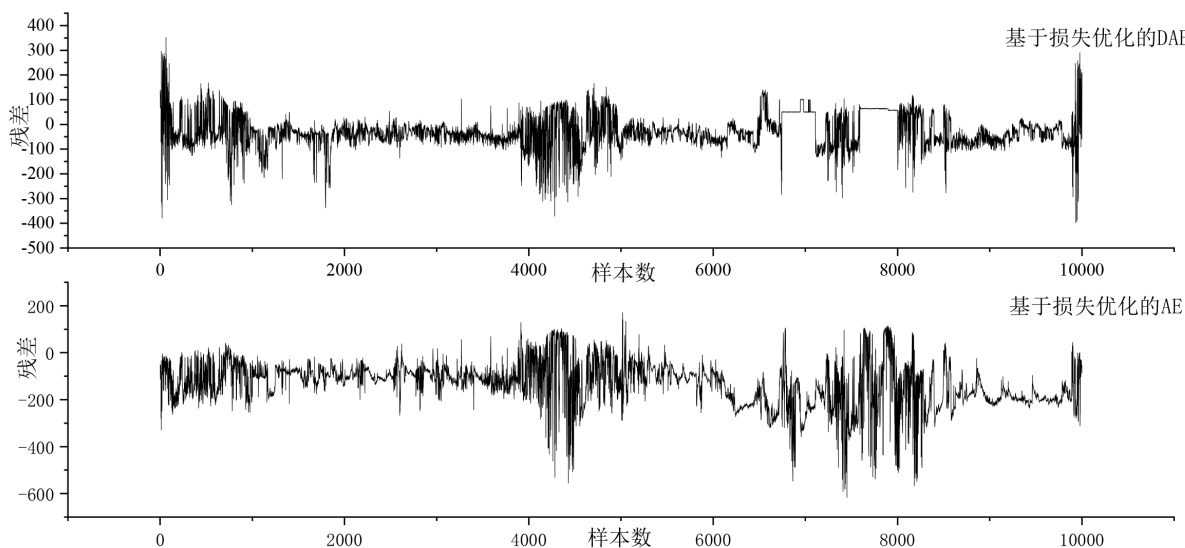


Figure 10. Active power

图 10. 有功功率

Table 2. MAPE of the L-DAE and I-AE

表 2. L-DAE 和 I-AE 的 MAPE

	MAPE		
	发电机驱动端温度	齿轮箱高速轴轴承温度	有功功率
L-DAE	1.375	0.958	60.426
I-AE	2.636	6.269	136.565

图 7~10 和表 2 分别展示了 L-DAE 和 I-AE 模型对发电机驱动端温度、齿轮箱高速轴轴承温度和有功功率三个监测参数重构结果、残差以及 MAPE 值。不难看出，本文提出的 L-DAE 模型相比 I-AE 模型，

在含有高斯噪声干扰的情况下，重构精度明显要比 I-AE 高，I-AE 对齿轮箱高速轴轴承温度和有功功率的重构出现了比较明显的偏移，这是因为 I-AE 提取数据特征时没能减少噪声带来的干扰信息，导致重构结果较差。对比两个模型的重构残差波动，明显看出本文所以模型重构后和数据与原始数据间的残差要小于 I-AE，MAPE 值也更小。

综合上面的结果分析可知，本文提出的 L-DAE 模型，可以通过将数据压缩成最小充分统计变量，来降低不确定因素产生的干扰信息，提取数据中最有效的特征，因此重构后的数据相较于原始数据存在一定损失，但是数据整体上不会偏离实际数据的变化趋势，整体数据波动幅度更小。相比于 I-AE 模型，对数据特征提取并重构的效果明显更好，抗不确定因素影响的能力更强，稳定性更好。

## 5. 总结

本文基于信息瓶颈理论，结合最小充分统计变量存在的条件，并将其改写为一种损失函数，用来训练深度自编码器，从而提出一种基于损失优化的深度自编码器模型，通过试验得出以下结论：本文提出的 L-DAE 模型可以降低风电机组中不确定因素对于机组监测参数数据质量的影响，减少不确定因素产生的干扰信息，尤其是导致波动增大的问题。在含有不确定因素影响的情况下，可以提高对数据特征的提取能力和重构精度。重构后的数据相对于原始数据更加平滑，波动幅度更小，为提高后续状态评估和异常检测的准确度奠定了良好的数据基础，对风电机组运维提供了准确度信息支持，可以降低机组的运维成本，提高风场收益，具有一定实际工程意义。

## 基金项目

国家自然科学基金面上项目(项目号: 51975535、52075164); 浙江省重点研发计划项目(2021C01133)。

## 参考文献

- [1] Kusiak, A. and Center, S. (2008) Wind Energy: Intelligent Manufacturing Perspective. *Proceedings of 6th International Symposium on Intelligent and Manufacturing Systems*, Iowa City, 14-16 October 2008, 42-53.
- [2] Gray, C.S. and Watson, S.J. (2009) Physics of Failure Approach to Wind Turbine Condition Based Maintenance. *Wind Energy*, **13**, 395-405. <https://doi.org/10.1002/we.360>
- [3] 胡阳, 王蔚然, 房方, 宋子秋, 许昱涵, 刘吉臻. 风电机组运行动态数字孪生建模及半物理仿真[J/OL]. 系统仿真学报: 1-12. <https://doi.org/10.16182/j.issn1004731x.joss.22-1307>, 2023-03-15.
- [4] Chen, H.G., Xie, C., Dai, J.C., et al. (2021) SCADA Data-Based Working Condition Classification for Condition Assessment of Wind Turbine Main Transmission System. *Energies*, **14**, 7043. <https://doi.org/10.3390/en14217043>
- [5] Zhu, Y., Zhu, C., Tan, J., et al. (2022) Operational State Assessment of Wind Turbine Gearbox Based on Long Short-Term Memory Networks and Fuzzy Synthesis. *Renewable Energy*, **181**, 1167-1176. <https://doi.org/10.1016/j.renene.2021.09.070>
- [6] Zhang, J., Kang, J., Sun, L., et al. (2021) Risk Assessment of Floating Offshore Wind Turbines Based on Fuzzy Fault Tree Analysis. *Ocean Engineering*, **239**, Article ID: 109859. <https://doi.org/10.1016/j.oceaneng.2021.109859>
- [7] Ferguson, D., McDonald, A., Carroll, J., et al. (2019) Standardisation of Wind Turbine SCADA Data for Gearbox Fault Detection. *The Journal of Engineering*, **2019**, 5147-5151. <https://doi.org/10.1049/joe.2018.9319>
- [8] 段震清. 基于大数据分析的风电机组运行状态评估及故障诊断[D]: [硕士学位论文]. 太原: 山西大学, 2018.
- [9] 朱永超. 基于机器学习的风电齿轮箱运行状态评估研究[D]: [硕士学位论文]. 重庆: 重庆大学, 2019.
- [10] 韩德鹏. 基于 SCADA 运行数据的风电机组健康状况评价[D]: [硕士学位论文]. 保定: 华北电力大学, 2019.
- [11] Kolchinsky, A., Tracey, B.D. and Wolpert, D.H. (2019) Nonlinear Information Bottleneck. *Entropy*, **21**, 1181. <https://doi.org/10.3390/e21121181>
- [12] 胡兵兵, 唐华, 吴幼龙. 基于互信息约束的生成对抗网络分类模型[J]. 中国科学院大学学报, 2022, 39(4): 551-560.
- [13] Li, L., Ding, S.X., Liang, K., et al. (2022) Control Theoretically Explainable Application of Autoencoder Methods to Fault Detection in Nonlinear Dynamic Systems. <https://doi.org/10.48550/arXiv.2208.01291>