

学校代码： 10246

復旦大學

硕士学位论文
(专业学位)

面向港口设备的多维时间序列异常检测
及异常预测技术研究

**Study on Multivariate Time Series Anomaly Detection and
Anomaly Prediction Technology for Port Equipment**

编 号 :

专业学位类别 (领域) :

院 系 :

完 成 日 期 : 年 月 日

目 录

摘 要	I
Abstract	II
第 1 章 绪论	1
1.1 研究背景及意义	1
1.2 国内外研究现状	2
1.2.1 时间序列异常检测	2
1.2.2 时间序列异常预测	7
1.3 本文研究挑战	10
1.4 本文研究内容	13
1.5 本文组织架构	14
第 2 章 相关理论与技术	15
2.1 时间序列相关概念	15
2.1.1 单维时间序列	15
2.1.2 多维时间序列	15
2.2 时间序列异常相关概念	16
2.2.1 异常类型	16
2.2.2 异常检测	17
2.2.3 异常预测	18
2.3 时间序列异常分析相关技术	19
2.4 本章小结	22
第 3 章 基于时空双维特征协同学习的多维时间序列异常检测模型	23
3.1 本章研究挑战	23
3.1.1 时空双维特征协同学习	23
3.1.2 时间维度特征学习	24
3.1.3 空间维度特征学习	25
3.2 本章主要工作	25
3.3 问题定义	26
3.4 MTAD-TSD模型介绍	27
3.4.1 MTAD-TSD整体结构概述	27
3.4.2 时间维度特征学习模块	28
3.4.3 空间维度特征学习模块	33
3.4.4 异常检测模块	35
3.5 实验及分析	35
3.5.1 实验设置	35
3.5.2 对比实验	39
3.5.3 消融实验	41

3.5.4 参数敏感性实验.....	43
3.6 本章小结.....	44
第 4 章 基于主导周期掩码和多尺度特征学习的多维时间序列异常预测模型 ...	47
4.1 本章研究挑战.....	47
4.1.1 稀疏异常前兆数据特征学习.....	47
4.1.2 多类型异常前兆数据特征学习.....	48
4.1.3 基于异常前兆特征的异常预测.....	48
4.2 本章主要工作.....	49
4.3 问题定义.....	50
4.4 MTAP-DM模型介绍.....	51
4.4.1 MTAP-DM整体结构概述.....	51
4.4.2 主导周期掩码序列生成模块.....	52
4.4.3 多尺度特征学习模块.....	53
4.4.4 异常预测模块.....	54
4.5 实验与分析.....	56
4.5.1 实验设置.....	56
4.5.2 对比实验.....	57
4.5.3 消融实验.....	59
4.5.4 参数敏感性实验.....	62
4.5.5 MTAD-TSD和MTAP-DM联合实验	63
4.6 本章小结.....	64
第 5 章 港口设备异常分析系统.....	65
5.1 系统设计需求.....	65
5.2 系统设计关键技术.....	66
5.3 系统结构与展示.....	67
5.4 本章小结.....	69
第 6 章 总结与展望.....	71
6.1 本文工作总结.....	71
6.2 本文工作展望.....	72
参考文献.....	73

摘要

随着港口物流规模的不断扩大与设备系统结构的日益复杂，港口设备的异常管理已成为保障港口运行安全、提升运行效率的重要环节。通过对港口设备生成的多维时间序列数据进行有效的异常分析，可以显著提高设备异常管理的效率。时间序列异常分析主要包括异常检测与异常预测两项任务，近年来，随着深度学习尤其是大模型技术的发展，Transformer和图神经网络等方法被广泛应用于相关研究中。然而港口设备因其复杂的运行工况、作业模式、操作行为使其所产生的时序数据具有高维异质、变量动态相关、局部变化模式复杂、异常前兆数据稀疏等特点，这给现有方法在港口设备异常分析中的应用带来了多方面挑战。

在港口设备异常检测任务中，现有方法存在以下挑战：1. 时空特征建模割裂，缺乏时间与空间特征的协同学习方法；2. 基于Transformer的方法对变量在时间维度的局部特征与异质性特征提取不足；3. 基于图神经网络的方法对变量在空间维度的动态相关性建模不充分。针对上述挑战，本文设计了一种基于时空双维特征协同学习的多维时间序列异常检测模型MTAD-TSD，其主要创新包括：（1）对多维时间序列进行时间和空间特征的联合建模学习；（2）对Transformer的点积注意力进行改进并提出了单变量注意力机制UA，实现对变量异质性特征的有效学习，同时基于单变量注意力机制设计了时间特征学习网络Uaformer，实现对变量局部特征的充分建模；（3）设计了一种基于局部图结构自适应性构建的图结构动态学习方法，实现对变量间关联结构的动态构建。

在港口设备异常预测任务中，现有方法存在以下挑战：1. 难以从大量正常数据中准确学习稀疏异常前兆数据特征；2. 无法充分学习不同异常类型的异常前兆数据特征；3. 缺乏异常前兆数据特征与高置信度异常预测结果之间的关联机制。针对上述挑战，本文设计了一种基于主导周期掩码与多尺度特征学习的多维时间序列异常预测模型MTAP-DM，其主要创新包括：（1）从频域视角对时间序列进行分析并生成主导周期掩码序列，减少正常数据对稀疏异常前兆数据特征学习的干扰；（2）对主导周期掩码序列进行多尺度特征学习及融合，充分建模多类型异常前兆数据特征；（3）基于多尺度特征相似度与序列重构方法，构建异常前兆数据特征与异常预测结果之间高效的关联机制。

本文在多个数据集上设计了对比实验，充分验证了本文提出的异常检测模型MTAD-TSD和异常预测模型MTAP-DM的优秀性能；本文设计了消融实验来验证这两个模型关键设计的有效性；本文设计了参数敏感性实验来分析模型参数对其性能的影响。本文最后基于MTAD-TSD和MTAP-DM搭建了港口设备异常分析系统，将相关理论创新应用到实际场景中。

关键词：港口设备；多维时间序列；异常检测；异常预测

中图分类号：TP3

Abstract

With the continuous expansion of port logistics and the increasing complexity of equipment systems, anomaly management for port equipment has become a critical aspect of ensuring operational safety and improving efficiency. By conducting effective anomaly analysis on multivariate time series data generated by port equipment, the efficiency of anomaly management can be significantly enhanced. Time series anomaly analysis primarily involves two tasks: anomaly detection and anomaly prediction. In recent years, with the advancement of deep learning—particularly large-scale models—methods such as Transformer and Graph Neural Network have been widely adopted in related research. However, due to the complex operating conditions, diverse working modes, and varying operational behaviors of port equipment, the resulting time series data often exhibit characteristics such as high-dimensional heterogeneity, dynamic inter-variable correlations, complex local variation patterns, and sparsity of anomaly precursors. These characteristics pose substantial challenges for applying existing methods to port equipment anomaly analysis.

In the task of port equipment anomaly detection, existing methods suffer from several challenges: 1. the modeling of spatial and temporal features is often decoupled, lacking a unified framework for joint spatiotemporal feature learning; 2. Transformer-based methods struggle to capture local temporal patterns and heterogeneous characteristics across variables; 3. GNN-based methods lack the capability to effectively model dynamic correlations among variables in the spatial dimension. To overcome these challenges, this study proposes MTAD-TSD, a multivariate time series anomaly detection model based on collaborative spatiotemporal feature learning. The primary innovations are as follows: (1) joint modeling of temporal and spatial features in multivariate time series; (2) a reformulation of the dot-product attention mechanism in Transformer, introducing a univariate attention (UA) mechanism to better capture heterogeneous variable features, and the design of the Uaformer network based on this mechanism to enhance local temporal feature extraction; (3) a dynamic graph learning method based on adaptively constructed local graph structures to effectively model inter-variable dependencies.

For the task of port equipment anomaly prediction, existing methods also encounter several challenges: 1. difficulty in accurately learning features of sparse anomaly precursors from a large volume of normal data; 2. inability to sufficiently learn features corresponding to different types of anomaly precursors; 3. lack of an effective association mechanism between anomaly precursors and high-confidence

anomaly predictions. To overcome these challenges, this study proposes MTAP-DM, a multivariate time series anomaly prediction model based on dominant period masking and multi-scale feature learning. The primary innovations are as follows: (1) frequency-domain analysis of time series data to generate a dominant period mask sequence, thereby reducing interference from normal data during precursor learning; (2) multi-scale learning and fusion of masked sequences to effectively capture features of diverse anomaly precursors; (3) construction of a robust association mechanism between precursor features and predicted anomalies via multi-scale feature similarity and sequence reconstruction techniques.

Comprehensive experiments on multiple datasets are conducted to validate the superior performance of the proposed anomaly detection model MTAD-TSD and anomaly prediction model MTAP-DM. Ablation studies are designed to demonstrate the effectiveness of key components in both models, and parameter sensitivity experiments are carried out to assess the impact of different model configurations. Finally, based on MTAD-TSD and MTAP-DM, a port equipment anomaly analysis system is developed, demonstrating the practical applicability of the proposed theoretical innovations.

Keywords: Port Equipment; Multivariate time series; Anomaly detection; Anomaly prediction

CLC code: TP3

第1章 绪论

1.1 研究背景及意义

随着全球经济的不断发展，港口作为重要的物流枢纽，在全球贸易中扮演着至关重要的角色。港口的运营效率直接影响到国际物流的速度和成本，因此，如何提高港口设备的运行效率和保障设备的安全性，越来越成为港口设备管理和运维工作中的重点研究内容，图1-1展示了本文合作港口智能化设备管理系统。近年来，随着传感器技术、自动化技术和数据处理技术的不断发展，港口设备的智能化和数字化水平也在不断提升。然而，港口设备日益复杂的工作环境及高强度的运行负荷，使得设备故障和异常现象频繁发生，一旦发生故障，不仅会导致港口的作业中断，还可能带来巨大的经济损失和安全隐患。因此，及时、准确地检测和预测设备的异常行为，对于提高港口的运营效率、减少故障停机时间、降低维修成本具有重要的现实意义。



图1-1 港口智能化设备管理系统

港口设备的种类繁多，包括起重机、输送带、集装箱堆场管理设备、港口拖船等。这些设备在运行过程中涉及到多个关键参数的监控，如温度、电流、压力、振动、转速等。随着设备数量的增加和运作环境的复杂化，设备的管理面临着越来越大的挑战。技术人员定期巡检设备是传统的设备管理方法，这种方法不仅需要耗费大量人力，而且会存在漏检设备潜在异常的问题。随着传感器技术的发展，越来越多的港口设备配备了传感器，能够实时采集设备运行数据，形成了庞

大的时间序列数据集。这些时序数据包含了设备的各种运行状态以及外部环境信息，从而为设备故障的预警提供了宝贵的依据。然而，如何从这些多维度的时间序列数据中提取有价值的信息，并及时发现潜在的异常或故障，成为当前港口设备管理的核心问题之一。

异常分析作为一种用于发现数据中不符合预期模式的技术，广泛应用于机器故障诊断、网络安全、金融欺诈检测等领域^[1-9]。在港口设备的故障诊断中，异常分析技术也被视为一种有效的手段。传统的异常分析方法通常依赖于单一的数据维度，如基于温度、压力或振动等传感器数据的异常检测，这些方法通过统计学或机器学习算法对单一维度数据进行建模，并根据模型输出的结果判断是否存在异常，然而，港口设备的运行状态通常由多个维度的数据共同决定，各种传感器数据之间可能存在复杂的相互关系，单一维度的异常分析方法无法有效捕捉设备在多维数据空间中的复杂行为模式^[10]。近年来，随着物联网技术的快速发展，越来越多的港口设备配备了传感器，实时采集设备的多维运行数据，这为设备的健康管理和故障诊断提供了新的可能。然而，港口设备的运行环境复杂，数据的维度多样且变化迅速，单一的故障诊断模型难以应对这种高度复杂的多维数据。因此，如何基于港口设备多维时间序列数据构建一个高效、鲁棒的异常分析系统，是港口设备智能化运维的迫切需求。

时间序列异常分析可以分为异常检测和异常预测两大类任务。近年来，随着深度学习技术的不断发展，越来越多的研究开始探索基于多维时间序列数据的异常分析方法^[11-13]。时间序列数据通常具有强烈的时序依赖性特征，传统的机器学习方法往往难以有效捕捉到这些时序信息。基于Transformer模型的异常分析方法通过自注意力机制能够有效捕捉时间序列中长距离的时序依赖关系，基于图神经网络的异常分析方法通过建模数据点之间的关系来有效地捕捉不同变量维度数据间的空间结构信息，相关方法取得了相对较好的时间序列异常分析结果。然而港口设备具有复杂的运行工况、作业模式和操作行为，使其所产生的时序数据具有高维异质、变量动态相关、局部变化模式复杂、异常前兆数据稀疏等特点，这给现有方法在港口设备异常分析中的应用带来了多方面挑战。

1.2 国内外研究现状

1.2.1 时间序列异常检测

时间序列异常检测是时间序列异常分析研究中的一个重要任务，其目标是基于数据的历史行为、时序模式或特征表示，检测出与正常模式显著偏离的数据。当前应用较多的时间序列异常检测方法大致可归类为基于传统统计方法、基于机器学习方法以及基于深度学习方法。

1.2.1.1 基于统计学的异常检测方法

基于统计学的异常检测方法主要运用统计学原理，利用数据的分布特性来识别与正常模式显著偏离的异常数据点。其基本原理是通过假设正常数据遵循某种统计分布，如正态分布、泊松分布等，然后使用相应的统计方法计算数据与正常模式的偏差，从而判断其是否为异常。Yule提出了自回归模型AR（Auto Regressive），Qin-ma等人针对一阶AR模型存在精度偏低的问题，提出了采用高阶AR模型^[14]来替代一阶AR模型的方法，以提升模型精度。滑动平均模型^[15]（Moving Average, MA）、自回归滑动平均模型^[16-17]（Autoregressive Moving Average, ARMA）、自回归差分移动平均模型^[18]（Auto Regressive Integrated Moving Average, ARIMA）以及这些平均模型的变体都被广泛应用在时间序列异常检测任务中。Hardin^[19]设计了一种基于F分布的稳健聚类算法，该算法融入了离群点辨识技术，并具备扩展到多聚类场景的能力，特别是针对大小不一、形态各异的聚类簇，该算法展现出优异的性能，其结果充分表明F分布在该领域的实际应用价值。Yu等人^[20]针对水文数据时间序列提出了一种异常值检测方法，利用历史数据构建预测模型，并据此推导出预测的置信范围，若观测值偏离这一范围，则被视为异常，从而实现了高效的异常检测。基于统计的全局异常检测方法HBOS^[21]在处理大规模数据集时表现出色，具有卓越的性能。尽管基于统计学的异常检测方法在某些场景下表现良好，但在数据分布不明确或数据中存在复杂的时序或空间依赖关系时，它们可能无法提供足够准确的结果。因此，针对这些局限性，许多研究已经开始结合机器学习和深度学习技术来改进异常检测的效果。

1.2.1.2 基于传统机器学习的异常检测方法

在处理复杂、高维、非线性和大规模数据时，传统的统计方法往往无法充分捕捉到数据的潜在模式。基于机器学习的异常检测方法能够从数据中自动提取规律，从而识别隐藏在数据中的异常模式。基于机器学习的异常检测方法通常可以分为有监督学习方法、无监督学习方法^[22]。

（1）有监督学习方法。有监督学习方法通常依赖于标注数据进行训练，即数据集中的每个样本都有一个明确的标签，指示该样本是否为异常，其核心思想是通过训练一个分类器，学习正常样本和异常样本之间的边界，从而判断数据是否是异常数据。支持向量机^[23]（Support Vector Machine, SVM）通过构建最大间隔的超平面，实现对不同类别样本的有效分隔。单类支持向量机^[24]（One Class Support Vector Machine, OCSVM）被广泛应用在异常检测任务中，One-Class SVM通过在高维空间中寻找一个超平面，使得大部分正常数据点位于该超平面的一侧，而异常数据点则位于另一侧，该方法适用于正常样本多而异常样本稀缺的情形。决策树是一种常见的分类模型，它通过一系列的条件判断将数据划分成不同的类别，在异常检测任务中，决策树模型可以根据样本的特征值来判断是否

异常。随机森林^[25] (Random Forest, RF) 是决策树的一种集成方法, 通过构建多个决策树并投票决定最终分类结果^[26-27]。隔离森林^[28] (Isolation Forest, iForest) 也被广泛应用于序列数据的异常检测任务中, Mao等人^[29] 在隔离森林的基础上提出了一种用于异常检测的方法, 该方法利用定长滑动窗口对时间序列进行离散化处理, 然后基于主成分分析和隔离森林来对序列特征空间进行降维和异常检测。K近邻算法^[30] (K-Nearest Neighbor, KNN) 通过计算样本之间的距离来判断样本的类别, 在异常检测中, KNN计算数据点与其邻居之间的距离, 若某个数据点与其邻居的距离异常较大, 则认为该数据点为异常, KNN方法对于数据的密度非常敏感, 适用于处理那些正常数据分布较为密集, 而异常数据离群分布的场景。

(2) 无监督学习方法。无监督学习方法通常在训练过程中无需依赖标签数据, 通过挖掘数据的内在结构模式来识别异常, 无监督学习方法适用于无法获取标注数据的场景。大多数基于无监督学习的异常检测方法试图在正常数据的分布基础上识别出不同于正常模式的样本。K-means聚类算法^[31]将数据分为k个簇, 并且使得簇内的数据点尽量相似, 簇间的数据点尽量不同, 可以根据数据点到其所在簇的中心的距离进行异常检测, 远离中心的点可能是异常点。DBSCAN^[32]是一种基于数据分布密度的聚类算法, 通过数据分布密度连接来发现数据中的簇, 与K-means不同, DBSCAN不需要预先指定簇的数量, 并且能够自动检测出异常点, 在DBSCAN中, 异常点通常是那些分布密度较低且没有足够邻居的数据点。PCA^[33]的目标是将高维数据投影到一个低维空间, 以保留数据中最重要的变化信息, 它通过计算数据的协方差矩阵并对其进行特征值分解, 找到数据中方差最大的方向, 然后将数据投影到这些主成分上进行异常检测。由于机器学习的计算较为复杂, 对于大规模数据集, 经典的机器学习方法可能面临较高的计算复杂度, 影响实时性和效率。深度学习算法能够自动从数据中学习特征表示, 无需繁琐的手动特征工程, 实现了对高维数据的更精准、更高效的异常检测^[34]。

1.2.1.3 基于深度学习的异常检测方法

基于深度学习的异常检测方法充分利用神经网络强大的特征提取能力, 从原始数据中自动提取高层次隐含特征, 并通过学习正常数据的模式来识别偏离这些模式的异常数据, 相比传统方法, 这类方法能够有效处理高维、非线性和大规模数据, 且无需依赖复杂的人工特征工程。

循环神经网络^[35] (Recurrent Neural Networks, RNN) 尤其是其改进版本的长短期记忆网络^[36] (Long Short-Term Memory, LSTM) 和门控循环单元^[37] (Gated Recurrent Unit, GRU) 被广泛应用于时间序列数据的异常检测任务中。RNN通过内部状态的循环连接能够捕捉数据的时序依赖关系, 并建立对正常序列的预测模型。THOC^[38]通过包含多层扩张RNN和分层SVDD^[39] (Support Vector Data

Description) 的复杂网络来在多个尺度上捕捉详细的时间特征, 高效地识别时间序列数据中的复杂模式, 并通过将当前数据与其正常模式数据进行比较来检测异常。LSTM专为解决传统的RNN模型在处理长时间序列数据时容易出现的梯度消失和梯度爆炸问题而设计, 其能够更有效地捕捉数据中的长期依赖关系^[40]。LSTM-NDT^[41]使用长短期时间网络来建模时间序列的时间依赖关系, 并且通过引入一种能够自动调整动态数据的异常阈值确定技术, 来有效解决动态数据中的异常多样性和不稳定性问题。LSTM-VAE^[42]利用长短期时间网络来捕捉时间序列动态特性, 并结合变分自编码器^[43] (Variational Autoencoder, VAE) 进行序列重构, 基于重构误差检测异常。GRU通过引入门控机制来控制信息的传递, 从而有效缓解了传统RNN在长序列数据中容易出现的梯度消失问题, 与长短期记忆网络相比, GRU的结构更为简洁, 参数更少, 但在许多任务中仍能达到较好的效果。OmniAnomaly^[44] 利用GRU模型处理时间序列数据, 并采用随机变量连接技术, 显式地对随机变量之间的时间依赖性进行建模, 使用重建概率作为异常检测依据。尽管RNN类模型在时间序列建模上取得了一定的成功, 但它们仍然面临需要对时序数据进行固定窗口大小划分的问题, 限制了该类模型在处理长时间序列时的性能。

卷积神经网络 (Convolutional Neural Networks, CNN) 通过卷积层、池化层、全连接层来识别数据中的层次化模式, 其局部感受域和权重共享机制能够对数据进行有效地特征提取。CARLA^[45]采用一种自监督对比学习方法, 首先通过在数据中注入异常样本, 并与原始样本进行区分, 从而学习有效的特征表示, 接着利用特征表示空间中样本的邻域信息, 学习正常行为模式及其偏差, 以提高模型异常检测能力。DACAD^[46] 结合了时间卷积网络^[47] (Temporal Convolutional Networks, TCN) 和无监督领域自适应技术, 通过生成异常样本, 提高模型的特征学习能力和跨领域泛化能力, 并且利用TCN进行时序特征提取, 同时结合对比学习方法, 使模型能够适应不同数据分布, 提高异常检测的鲁棒性。

自编码器^[48] (Autoencoder, AE) 通过编码器将输入数据压缩成低维空间表示, 并通过解码器重构输入数据, 在异常检测中, 自编码器通过重构误差来判断异常数据, 若某个数据点的重构误差较大, 则认为该数据点与正常数据存在显著差异, 可能是异常点。AMSL^[49]采用基于卷积网络的自编码器进行端到端训练方法, 从时间序列数据中提取特征, 并使用记忆网络学习数据中常见的特征和异常模式, 其内部的自适应融合模块能够将多种特征融合为新的重构特征, 提升了模型的异常检测性能。ContextDA^[50]将时间序列的上下文采样视为马尔可夫决策过程, 使用判别器来对齐源领域和目标领域, 并且通过引入源领域标签提高正常样本的对齐效果, 从而增强模型异常检测的准确性。

变分自编码器 (Variational Autoencoder, VAE) 与自编码器类似, 但不同与

AE直接将输入数据映射为单个点的潜在表示，VAE学习输入数据在潜在空间的概率分布，从而具有更强的生成能力。InterFusion^[51]使用分层变分自编码器来提升模型的异常检测性能和可解释性。RDSSM^[52]通过采用分布误差缓冲区方法来处理污染数据和无标签数据，并且使用生成模型来处理时间序列中的波动噪声，其设计的具有适应性异常分数的概率检测方法能够有效提高模型异常检测的精确性和鲁棒性。

生成式对抗网络^[53] (Generative Adversarial Networks, GAN) 由生成器和判别器构成，生成器用于通过对抗训练逐步学习正常数据的分布特征，判别器则用于区分真实数据和生成数据，其优势在于能够生成逼真的样本，同时对数据分布建模较为精准，为异常检测提供了一种全新的视角。AnoGAN^[54]利用训练好的GAN模型，通过生成器将输入样本映射到隐藏空间，再利用判别器对样本进行打分，如果样本在重构过程中的表现与正常数据存在较大差异，或者判别器认为该样本与正常数据分布不符，则该样本会被标记为异常。GANomaly^[55]是一种基于生成式对抗网络的异常检测方法，其核心思想是通过生成模型学习正常数据的分布，以便有效地检测异常样本，并且通过判别模型判断数据是否来自正常分布，利用重构误差作为异常检测的依据，适合用于无监督的异常检测任务。DCdetector^[56]通过设计一个基于对比学习的双分支注意力结构来学习数据的潜在模式特征，并且基于对比训练方式来放大正常数据和异常数据之间的特征差异，进而有效识别异常数据。

近年来Transformer^[57]模型凭借其高效的全局依赖建模能力和并行计算优势，逐渐被引入时间序列异常检测任务中，注意力机制能够直接捕捉序列数据中的远距离依赖关系，对于长时间序列的异常检测尤为有效。TranAD^[58]是一种基于Transformer模型和生成式对抗网络的时间序列异常检测模型，其通过对输入序列施加多种变换来生成具有不同时间特性的增强样本，并且通过引入生成式对抗网络学习时间序列更鲁棒的特征表示，其结合局部和全局时间依赖关系来捕捉不同时间尺度上的异常，联合优化重构误差和对抗损失，有效提升了模型异常检测性能；Anomaly Transformer^[59]基于自注意力机制提出异常注意力机制，对每个数据点同时建模两种关联关系：基于可学习高斯核函数构建的先验关联关系，用于刻画时间点之间的局部邻接偏好；基于传统自注意力机制计算的序列关联关系，用于捕捉全局依赖结构，并且设计了一种极小极大化的联合优化策略来放大正常样本与异常样本之间的关联差异，从而构建出一种基于关联差异的异常检测方法。Dual-TF^[60]采用两个并行的Transformer结构分别分析时间序列的时间域和频域信息，从不同角度提取特征并进行融合，能够提高模型在复杂时间序列中的异常检测性能。TS2Vec^[61]基于层次化的Transformer结构在多个时间尺度上捕捉时间序列的上下文信息，并且采用自监督对比学习方法将异常检测视为下游学习任务。MT-RVAE^[62]使用Transformer来建模时间序列长期依

赖关系，并且使用变分自编码器来学习时间序列数据的潜在分布模式，有效提升了模型异常检测的性能和稳定性。基于Transformer模型的异常检测方法主要针对时间序列的时间维度进行建模分析，忽视了多维时间序列中不同变量维度之间的空间相关性对异常检测的重要性，同时传统Transformer模型内部的结构设计使其在捕捉时间序列数据的局部语义信息存在一定局限性。

图神经网络^[63]（Graph Neural Network, GNN）可以有效地建模图结构数据，多维时间序列不同变量维度之间的空间相关性也可以看作是一种空间上的图结构，因此越来越多的研究将图神经网络运用到时间序列异常检测任务中。GDN^[64]通过引入图注意力网络来更新学习不同节点之间的信息以充分建模时间序列的空间依赖性；FuSAGNet^[65]在GDN基础上引入了稀疏自编码器（Sparse-AE），通过在其潜在空间中引入稀疏性来进行重构任务，通过稀疏性约束的联合优化方法来同时优化Sparse-AE和GDN组件，进而同时利用重构误差和预测误差计算异常得分；STGAT-MAD^[66]使用一张完全图结构来建模多维时间序列中任意两个变量维度之间的依赖关系；Series2Graph^[67]将不同的子序列映射到二维空间中，并且基于投影空间的密集区域来为对应的子序列构建图结构，通过计算节点的度和边权重来评估图中不同大小子图的异常程度。GANF^[68]将时间序列建模成贝叶斯网络，从而学习变量之间的条件密度关系，并且利用图神经网络学习时间序列各变量维度之间的依赖关系，提高模型特征表达能力，并且通过优化图的邻接矩阵，动态调整变量之间的关系，使得模型能更准确地检测异常。IVAD^[69]基于图神经网络提出了一种双重监测异常方法，通过局部监测和全局监测提高模型异常检测的精确度和可解释性。EEG-CGS^[70]将多维时间序列转化为属性图，通过基于局部子结构的正负子图采样和改进的对比学习策略来增强时间序列异常检测的对比学习能力。然而现有大多数图神经网络方法在构建时间序列图结构时普遍采用先验知识手动定义或者计算静态邻接矩阵构建节点之间的连接关系，这种连接关系在模型的整个训练和测试过程中保持不变，忽略了在实际系统中变量之间连接关系的动态变化特性。

1.2.2 时间序列异常预测

时间序列异常预测作为时间序列异常分析研究中的另一个重要任务，其目标是在给定的历史时间序列中，利用预测模型推测未来一段时间内的行为模式，并识别其中可能出现的潜在异常事件。与传统的异常检测侧重于事后识别数据中已存在的异常不同，异常预测旨在对未来可能出现的异常提前预警，从而为系统维护、风险控制和决策支持提供更为主动的干预措施，在实际应用中，异常预测任务往往比异常检测任务更具有实用价值。与时间序列异常预测相关的研究方向包括时间序列预测、时间序列极端稀有事件的预测。

时间序列预测研究致力于对时间序列的整体趋势或未来数值进行准确预测

，相关研究方法包括基于统计学的经典方法以及近年来兴起的深度学习方法。生成式对抗网络因其对复杂时序数据的优秀建模能力而被广泛运用在时间序列预测模型中。AEC-GAN^[71]在生成式对抗网络的基础上引入误差校正模块，通过对真实序列加入对抗扰动并进行重构训练，从而在生成过程中动态检测并修正生成数据的偏差，并且设计了一种基于对抗样本的训练增强方法，用于稳定生成式对抗网络的训练过程，有效解决了传统的生成式对抗网络在长时间序列生成过程中存在的数据分布偏移问题，显著提升了模型的预测性能。MAGAN^[72]基于生成式对抗网络设计了一种多注意力机制的植被指数预测方法，其编码器提取驱动序列的潜在表示、生成器建模长期时序模式、判别器提供分类与反馈优化，并支持异构数据的融合预测，有效解决了归一化植被指数在多步预测中因非平稳性及外部因素复杂性所导致的预测不准确问题。Transformer模型因其强大的长时间依赖关系建模能力而被越来越多的时间序列预测模型所使用。Informer^[73]模型在Transformer原始的自注意力机制基础上设计了稀疏概率自注意力机制，通过仅计算对预测贡献最大的注意力得分，显著降低了计算复杂度；同时其自注意力蒸馏机制在编码器中逐层压缩序列长度，只保留主要注意力特征，从而减少冗余信息和内存开销，提高了对长时间序列的建模能力，同时其内部的生成式解码器采用一次性预测整个序列的方式，在提升了模型推理速度的同时避免了误差逐步累积的问题。FEDformer^[74]将傅里叶变换与Transformer模型相结合，利用大多数时序数据在傅里叶变换中具有稀疏表示的特性将时域信息拆解为频域信息，随机采样部分频域信息，然后投影回时域，能够在降低计算复杂度的同时过滤掉多数噪声数据，使得模型能够更好地学习长时间序列中的依赖关系，有效提升了模型的预测性能。iTransformer^[75]在保留Transformer原有模块的基础上直接反转Transformer结构，并且将独立序列作为变量标记，通过注意力机制捕捉多元时序相关性，并利用层归一化和前馈网络学习时间序列特征表示，提升模型的时序预测能力。目前部分时间序列预测研究通过构建级联结构来将多个模型或算法按层次、顺序组合并逐步优化预测结果，进而提升预测的精度。FCPM^[76]设计了一个特征驱动的级联预测结构，其通过构建特征影响树并系统地评估各特征对预测结果的贡献，自动选择关键特征，并将每个特征与预测结果关联，提升了模型的可解释性和预测精度。TimesNet^[77]基于时间序列多周期性的特点，使用卷积神经网络将复杂的时序变化分解为多个周期内和周期间的变化，并且通过将时间序列从一维空间扩展到二维空间中，实现了将周期内和周期间的变化分别嵌入到二维张量的列和行中，从而易于使用二维卷积核来进行特征提取，有效提升了模型的泛化性能。文献^[78]基于跨域信息融合思想设计了一个网络信息级联预测框架CasCIFF，其通过引入全局社交网络结构中的多跳邻居信息，构建更具区分性的用户影响力表示，增强用户嵌入的鲁棒性，并将信息传播的时间编码到图结构中，以捕捉信息传播的动态特征，

同时采用多任务学习策略共享用户分类与级联预测结果的中间表示，从而提升模型的泛化能力。

然而时间序列预测模型在面对具有低发生概率的极端稀有事件时，往往表现出性能瓶颈，因此针对时间序列中极端稀有事件的预测研究也逐渐兴起。文献^[79]针对时间序列稀有事件预测中的数据不平衡和分布偏移问题，提出了一种基于元学习的重加权策略和微调机制的稀有事件预测模型，该模型通过引入动态优化的加权损失函数来增强对稀有事件样本的关注，并在此基础上进行仅基于稀有事件样本的微调，以提升模型对稀有事件的泛化能力。文献^[80]提出一种结合自逻辑回归的时间序列共演建模方法，主要解决稀有事件预测中的精度不足和类别不平衡问题，该方法通过自逻辑回归模型建模多个时间序列历史时间片段中的自相关性与变量间依赖性，提升稀有事件预测的上下文感知能力，并且通过对不同类别预测误差设置权重，在优化模型的同时更关注少数稀有事件的预测精度。文献^[81]设计了一种用于在临床医学时序数据中预测不规则事件的模型XTSFormer，其通过使用基于特征引导的周期感知时间位置编码来有效捕捉临床时间序列中常见的周期性、循环结构和时间不规则性特征，并且使用跨时间尺度的层级注意力机制来提升模型的预测性能。文献^[82]提出了一种用于时间序列极端稀有事件预测的深度混合模型 DEMMA，该模型通过对广义帕累托分布的重参数化，构建了一个适用于重尾分布数据建模的广义混合分布结构，同时引入了基于LSTM的自编码器以提取时间序列中的时序特征，并结合时间注意力机制构建量化预测模块，提升了模型对时间序列中极端事件的预测精度。文献^[83]提出了一种用于预测极端水温事件的元迁移学习方法，该方法使用基于上下文相似度的样本重加权策略，在模型训练阶段提升极端样本的重要性，同时保持对正常样本的预测性能，并且通过引入极值分布约束进一步增强对极端事件的建模能力。

尽管异常预测和极端稀有事件预测都是处理非典型数据行为的任务，但两者在本质上存在差异。极端稀有事件预测针对的数据在时间序列中是极其稀少的，并且具有较大的幅度或变化，其目标是准确预测发生极端稀有事件的时刻或条件；而异常预测关注的是与常规时序数据模式不同的数据，这些数据可能具有不同的变化幅度，也可能是局部的偏离，其目标是预测出时间序列中所有偏离正常模式的事件，无论其是否极端。由于时间序列异常预测任务难度较大，目前相关的研究相对较少。文献^[84-85]通过传统统计方法对地震信号进行分析进而给出对于地震异常信号的预测；文献^[86]设计了一种面向大规模网络日志数据的多维时间序列异常预测模型SNetAD，该模型通过建模时间点之间及设备之间的依赖关系，生成网络行为的表示向量，利用对比中心损失函数引导表示向量形成正常与异常两类聚类结构，并结合少量标签与大量未标记样本优化聚类的可分性，提高异常预测的准确性。文献^[87]设计了一种基于异常前兆数据的时间序列异常预测模型PAD，其通过使用神经控制微分方程来捕捉数据的连续变化模式，提高对非

均匀采样数据的建模能力，同时结合多任务学习和自监督知识蒸馏机制协同优化异常预测目标，有效提升了多维时间序列异常预测的准确性和鲁棒性。文献^[88]针对时间序列异常检测任务中缺少异常延迟时间与异常时间范围研究的问题，设计了一种融合时间动态信息的异常预测方法，该方法通过结合异常检测模型与Wasserstein损失函数来对未来异常事件的概率分布进行建模，进而有效预测异常发生的时间范围与密度分布，并且构建了用于评估异常预测性能的合成数据集和评价指标体系来对异常预测模型性能进行整体评估。文献^[89]以长序列预测模型为基础设计了一种时间序列异常预测模型FCM，通过将未来时序片段的预测结果作为上下文信息，并与当前时间窗口数据进行关联建模，放大异常前兆与正常趋势之间的潜在差异，此外FCM还采用了双视角的正常性建模策略，分别从当前观测窗口及其与未来预测序列之间的对应关系中学习异常前兆模式，同时通过多维注意力机制增强对多维时间序列中跨维度依赖和动态特征的建模能力。目前大多数时间序列异常预测研究都是在未标注的时间序列上进行无监督或弱监督训练，导致模型主要学习了大量正常数据的变化模式，而对于异常发生前的稀疏异常前兆数据特征学习不足，同时没有考虑到不同类型异常之间演变趋势和演变时间的差异，缺乏对不同类型异常前兆数据的精准建模能力。

1.3 本文研究挑战

本文聚焦于港口设备时序数据的异常检测及异常预测任务展开相关研究。相较于传统领域中的时序数据，港口设备时序数据由于受其复杂的工作环境、多样化的作业任务以及高度自动化的操作流程所影响，在结构特征与动态行为方面均呈现出较为显著的复杂性与特殊性，具体特点总结如下：

(1) 高维异质性。为实现对关键部件与运行状态的全面监测，港口设备部署了数量众多的传感器，图1-2展示了本文合作港口部分设备所安装的传感器数量，可以直观体现出其传感器系统的高维复杂性。

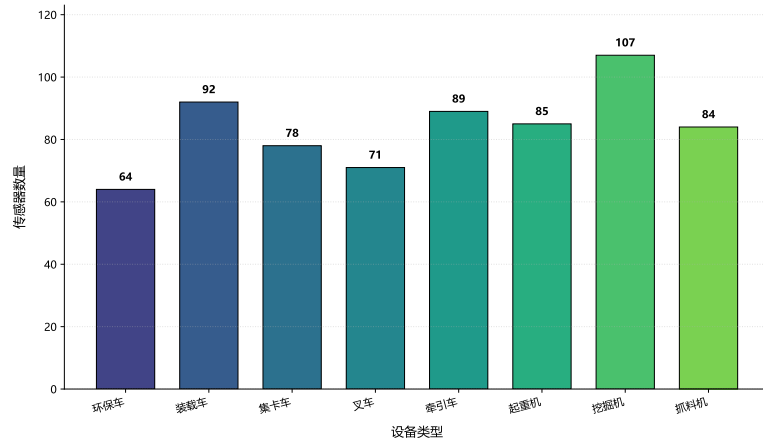


图1-2 港口设备传感器数量

而这些不同类型传感器采集的变量在量纲类型、数值尺度、变化速率与响

应延迟等方面差异显著，构成变量间高度异质化的多维时间序列。如图1-3所示，岸桥液压压力随负载变化呈周期趋势，振动信号则对微扰高度敏感且波动剧烈，而红外温度则变化平缓。同时这种变量间的异质性特征使得港口设备时序数据在不同变量维度上会具有多种类型的异常，包括短时的突发性故障、长期的性能退化等，这些异常的发生方式、表现形式以及影响的时间尺度各不相同，使得传统统一建模策略难以有效提取各维度特征中的异常模式，进而影响模型的准确性与鲁棒性。

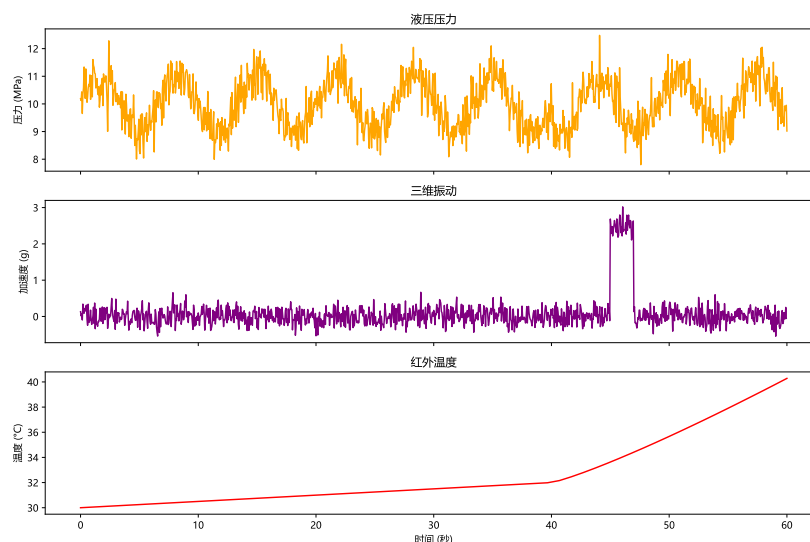


图1-3 岸桥不同传感器数据变化示意

(2) 变量动态相关性。港口设备各子系统间存在高度协同特性，这使得多传感器变量之间表现出显著的变量相关性，同时这种相关性不仅体现为某些维度之间的长期关联，还表现出随工况、操作行为、作业模式变化而发生动态变化的特性。如图1-4所示，在不同工况下传感器之间的相关性会动态变化，这种变量的动态相关性反映了系统内部状态的变化特征，准确建模和学习这种动态关联结构特征，对于识别潜在异常、理解异常传播路径具有重要意义。

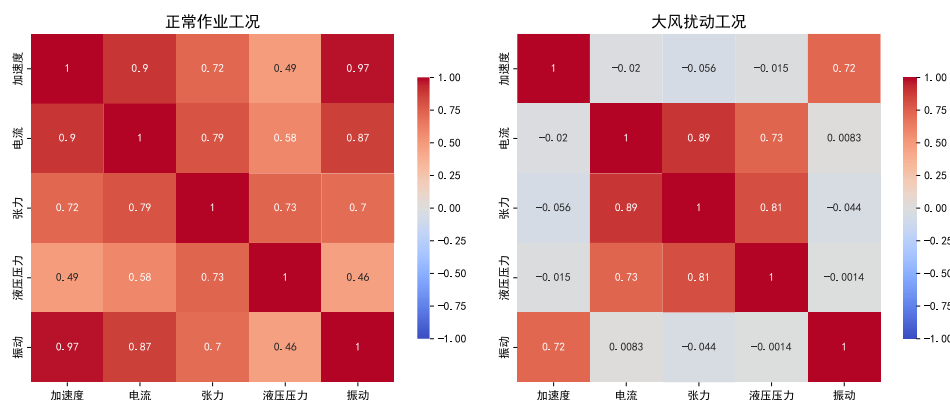


图1-4 不同工况下岸桥传感器相关系数

(3) 局部变化模式复杂。港口设备在不同作业阶段中往往呈现出快速切换的运行状态，导致传感器数据在较短的局部时间内呈现非线性、非平稳的剧烈波动或细微扰动，构成复杂的局部变化模式，如图1-5所示。此类局部变化模式不

仅反映了设备的动态响应特性，还可能隐含早期故障信号，尤其在异常演化的初始阶段，关键特征往往仅出现在极窄的时间窗口中，表现形式可能是轻微偏离、突变或高频噪声嵌套，若仅依赖整体趋势或全局特征将难以有效捕捉与识别异常信号。

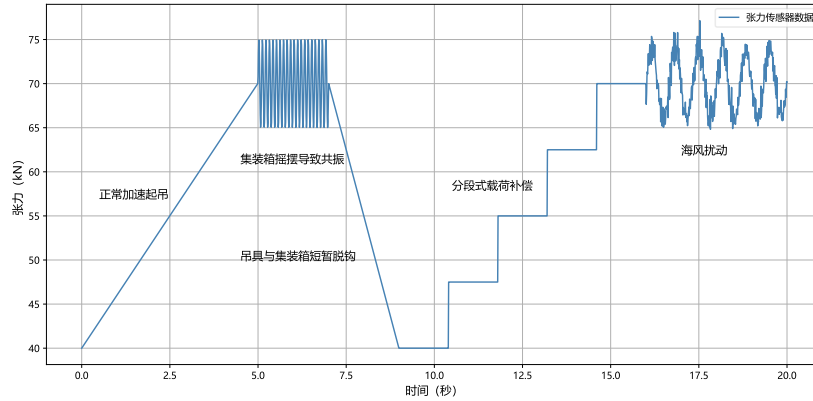


图1-5 岸桥张力传感器数据局部变化示意

(4) 异常前兆数据稀疏。港口设备一些潜在的严重故障如轴承疲劳裂纹、齿轮啮合异常等，在异常演变早期阶段往往不会表现出显著的异常前兆信号，同时这些异常前兆数据在整体时间序列中所占比例极低，数量远少于正常状态下产生的数据，使得异常前兆数据的特征通常被大量正常数据所掩盖，如图1-6所示，在时序数据中未识别出明显的异常前兆数据，但通过对数据进行频域分析识别出了特定频段的异常前兆数据，这种异常前兆数据的稀疏性使得模型在特征学习过程中更易偏向学习正常数据的特征，从而难以有效提取和识别异常前兆数据特征。

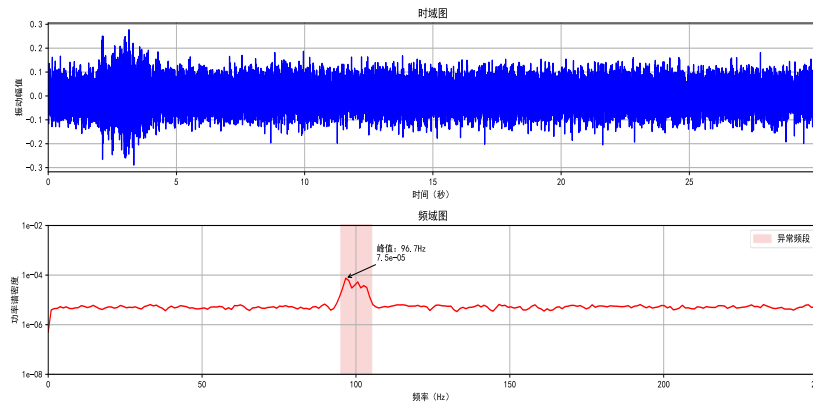


图1-6 自动导引车齿轮振动信号时域及频域图

近年来多维时间序列异常检测相关研究取得了显著进展，但现有研究在处理港口设备在复杂工况下所产生的高维复杂时序数据时，仍存在以下局限性：

(1) 时空特征建模割裂。现有多维时间序列异常检测方法大多基于单一建模视角，或仅从时间维度提取序列动态演化特征，或仅基于空间结构分析变量间关系，缺乏对时空特征联合建模的方法。

(2) 基于Transformer的时间维度特征学习方法难以捕捉局部特征与变量

异质性特征。目前基于Transformer模型的时间维度特征学习方法大多对数据采取基于点级语义颗粒度的建模方式，无法充分提取数据的局部动态特征，进而难以有效捕捉港口设备数据中复杂的短时变化模式。同时这些方法通常采用统一建模策略或共享注意力机制，容易造成多变量间的信息干扰，难以针对不同变量的异质特性进行差异化建模。

(3) 基于图神经网络的空间维度特征学习方法难以捕捉变量动态相关性。

现有空间维度建模方法大多基于静态、预定义的图结构构建变量之间的全局关联关系，缺乏对变量之间局部动态关联关系的自适应建模机制。

目前时间序列异常预测相关研究较少，并且现有时间序列异常预测模型在处理港口设备时序数据时也存在以下局限性：

(1) 难以从大量正常数据中准确学习稀疏异常前兆数据特征。现有时间序列异常预测方法通常依赖于对历史序列的直接建模，以提取潜在的异常前兆特征，然而在异常前兆数据稀疏且被大量正常数据掩盖的时序数据中，此类建模方法往往更容易学习到正常数据的特征模式，导致对异常前兆数据特征学习不足，降低了模型对异常早期信号识别的准确性和鲁棒性。

(2) 多类型异常前兆数据特征建模不足。现有的时间序列异常预测方法普遍在统一的时间尺度下对历史序列进行建模，忽略了不同类型异常在演变过程中具有不同的异常演变趋势以及演变时间，缺乏对多类型异常前兆数据的有效建模能力。

(3) 缺乏异常前兆数据特征与高置信度异常预测结果之间的关联机制。现有时间序列异常预测方法普遍缺乏有效的评分或映射机制来将异常前兆数据特征转化为对未来异常事件的高置信度预测结果，在学习到异常前兆数据特征后难以判断哪些异常前兆数据特征具备真正的异常预测价值。

1.4 本文研究内容

本文针对港口设备在实际运行过程中生成的具有高维异质、变量动态相关、局部变化模式复杂、异常前兆数据稀疏等特性的时序数据，对异常检测与异常预测两个核心任务开展了如下研究工作：

(1) 本文设计了一种基于时空双维特征协同学习的多维时间序列异常检测模型MTAD-TSD，以解决现有方法存在的时空特征建模割裂的问题。在时间维度特征学习方面，本文在Transformer原始的点积注意力基础上进行改进，提出了单变量注意力机制UA，使模型专注于各个变量异质性特征的提取，降低了多变量间的信息干扰，本文基于UA设计了时间维度特征学习网络Uaformer，其通过时间序列分段和Time2Vec编码方法实现了对时间序列长期依赖关系与局部特征的联合建模，提升了模型对高维异质与局部变化模式复杂的时间序列建模能力；在空间维度特征学习方面，本文设计了一种基于局部图结构自适应性构建的图结

构动态学习方法，实现对变量间关联结构的动态构建，并结合图注意力网络，进一步挖掘变量间的局部动态相关性和全局静态相关性特征。

(2) 本文设计了一种基于主导周期掩码和多尺度特征学习的多维时间序列异常预测模型MTAP-DM。模型首先从频域角度对时间序列进行分析，基于快速傅里叶变换计算各个变量的主导周期，并为各个变量自适应地生成不同的主导周期掩码序列，以减少正常数据对稀疏异常前兆数据特征学习的干扰；同时模型使用多尺度特征学习策略，在多个时间尺度下分别进行异常前兆数据的特征学习与融合，捕捉具有不同异常演变趋势和演变时间的异常前兆数据特征，提高对多类型异常前兆数据特征建模的能力；模型最终通过计算多尺度特征相似度和重构误差来建模异常前兆数据特征与异常预测结果之间可靠的关联关系。

(3) 本文在多个数据集上对MTAD-TSD和MTAP-DM进行了大量的对比实验、消融实验和参数敏感性实验。对比实验结果表明本文设计的两个模型分别在多维时间序列异常检测和异常预测任务中具有优秀性能；消融实验结果验证了各个模型内部关键模块和方法设计的有效性；参数敏感性实验探讨了不同超参数对模型性能的影响。

(4) 本文基于所提出的MTAD-TSD与MTAP-DM模型搭建了一个面向港口设备异常管理的多维时间序列异常检测及异常预测系统，能够实现对港口设备运行数据的异常检测及预测。该系统为港口设备运维管理中的故障检测与风险防控提供了有效技术支撑，具有较高的工程应用价值与实践意义。

1.5 本文组织架构

第1章为绪论，本章首先阐述了多维时间序列异常检测与异常预测在港口设备管理中的关键作用，随后综述了国内外在该领域的研究现状，接着分析了港口设备时间序列数据的典型特征及面临的技术挑战，最后明确了本文的研究目标与主要内容，并对全文结构进行了安排。

第2章是相关理论与技术，本章介绍了与本文研究密切相关的理论基础与关键技术，包括异常、异常检测与异常预测的基本概念和分类方法，以及本文所采用的深度学习技术和相关模型架构。

第3章设计了一种基于时空双维特征协同学习的多维时间序列异常检测模型，并在多个数据集上进行实验验证了模型性能。

第4章设计了一种基于自适应主导周期掩码和多尺度学习的多维时间序列异常预测模型，并在多个数据集上进行实验验证了模型性能。

第5章基于前述两个模型搭建了一个面向港口设备的多维时间序列异常检测及预测系统，并介绍了系统架构、模块功能及实际应用效果。

第6章总结了全文的研究内容和工作，并对后续的研究方向进行了展望。

第2章 相关理论与技术

2.1 时间序列相关概念

时间序列是按照时间顺序排列的一系列数据观测值，通常用来描述某一现象或变量随时间的变化情况^[90]。其主要特点包括：

（1）时间依赖性：时间序列数据的每个观测值都有明确的时间戳，前后数据之间存在一定的依赖关系，过去的值可能对未来的值产生影响，这种依赖性使得时间序列分析不同于传统的独立同分布数据分析。

（2）趋势性：很多时间序列数据表现出长期的上升、下降或平稳趋势，这反映了数据在较长时间范围内的基本变化趋势。

（3）季节性：部分时间序列数据在一定周期内会重复出现类似的波动模式，这种周期性的波动称为季节性或周期性现象，例如一年四季的气温变化、每周的消费模式等。

（4）随机性：除了趋势和季节性之外，时间序列数据中还包含随机波动成分，即难以预测的噪声和异常波动。

2.1.1 单维时间序列

单维时间序列（Univariate Time Series）指的是在连续或离散时间点上对单一变量进行的依次观测所构成的数据序列，如公式（2.1）所示，其中 X 为一单维时间序列， T 为时间序列的长度， x_t 表示在 $t \in T$ 时刻的观测值。单维时间序列是描述单一变量随时间变化规律的重要数据形式，其分析方法在经济学、气象、工程和金融等领域均有广泛应用，通过构建合适的模型，不仅可以揭示数据的内在结构，还能实现对时间序列未来值的预测。

$$X = \{x_1, x_2, \dots, x_T\} \quad (2.1)$$

2.1.2 多维时间序列

多维时间序列（Multivariate Time Series）是指在相同时间点上，对多个相互关联的变量进行同步观测所构成的数据序列。与单变量时间序列不同，多维时间序列不仅反映了各个变量随时间的变化趋势，还包含了变量之间的相互依

赖关系。如公式（2.2）所示，其中 X 为一多维时间序列， T 为时间序列的长度， N 为多维时间序列的维度， x_t^N 表示在 $t \in T$ 时刻的观测值。多维时间序列在经济、金融、气象、工程、医疗等领域均有广泛应用。例如，在金融市场中，不同股票价格、交易量等数据构成的多维时间序列能帮助分析市场的整体动态；在工业控制中，多个传感器数据构成的多维时间序列用于监控设备状态并进行故障预测；在气象领域，不同气象变量构成的多维时间序列有助于准确预测天气变化。

$$X = \{x_1^N, x_2^N, \dots, x_T^N\} \quad (2.2)$$

2.2 时间序列异常相关概念

2.2.1 异常类型

在时间序列分析与建模过程中，异常通常指的是与大部分数据样本显著不同的观测值或事件，其行为或特征偏离了数据所遵循的正常模式或概率分布。异常不仅可能反映出数据收集或记录过程中的噪声与错误，更可能揭示出潜在的、具有重要实际意义的稀有事件或突变现象。Hawkins^[91]认为数据集合中明显和其它数据样本不同的数据是异常，这些数据是由不同的机制产生的，并不是随机偏差导致的。时间序列中的异常类型一般可以分为以下三类：点异常（Point Anomaly）、条件异常（Contextual Anomaly）和集体异常（Collective Anomaly）^[92]。

点异常指的是在时间序列中某一单独数据点与其他数据点相比显著偏离正常模式的情况，通常表现为孤立的异常值，其偏离程度可以通过统计指标进行量化，如图2-1所示。检测点异常的方法多依赖于对数据分布的建模和局部统计特性的分析，常见方法包括基于统计阈值的方法、孤立森林以及基于自编码器的重构误差分析等。

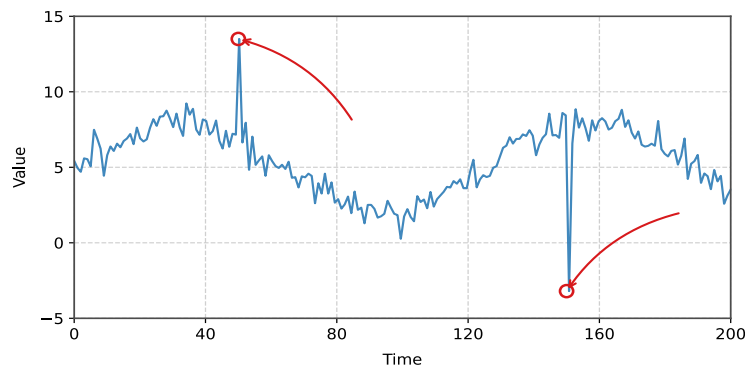


图2-1 时间序列点异常

条件异常指在特定的上下文条件下，一个数据点表现出异常行为，而在其他上下文中则可能是正常的，这里的上下文通常指时间、空间或其他环境变量，如图2-2所示。在进行上下文异常检测时，必须充分考虑时间序列的背景信息或

外部变量，通过构建条件概率模型或上下文敏感的检测方法来实现。

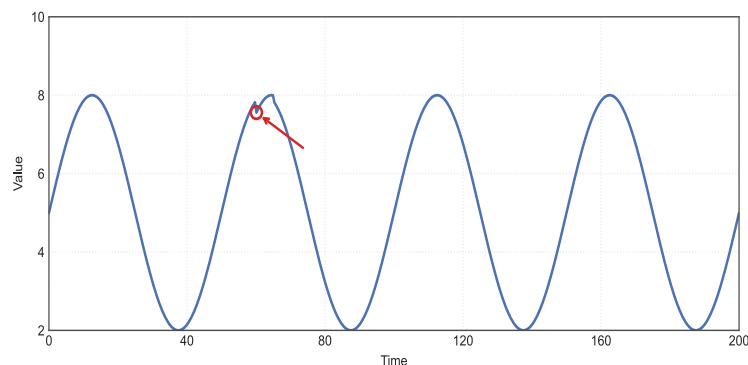


图2-2 时间序列条件异常

集体异常指的是一组数据点整体上偏离了正常模式，尽管单个数据点可能并不显得异常，即个别数据点在孤立情况下看似合理，但当它们以特定的组合出现时，其联合行为与历史数据的正常模式不符，如图2-3所示。集体异常常见于系统状态突变或潜在故障过程的早期信号中，其检测通常依赖于序列建模方法来捕捉数据点间的依赖结构，并通过比较实际序列与预测或模型重构结果的偏差来判断。

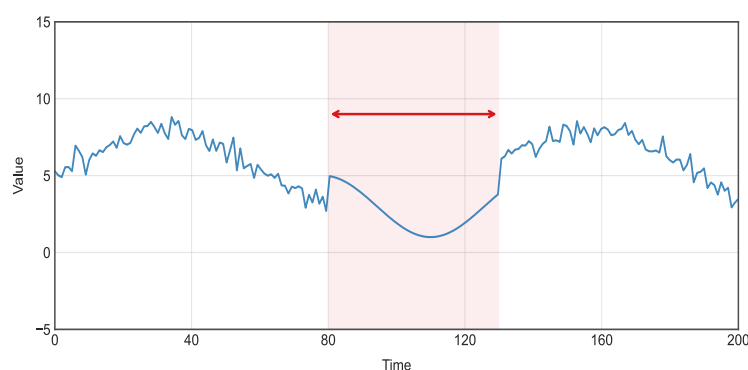


图2-3 时间序列集体异常

2.2.2 异常检测

时间序列异常分析通常可以划分为两个核心任务：异常检测（Anomaly Detection）与异常预测（Anomaly Prediction）。异常检测侧重于识别历史时序数据中已发生的异常行为，通常用于实时监控与回溯分析；异常预测则侧重于基于时间序列的异常演化规律预测未来可能发生的异常，从而实现异常提前预警与风险干预。这两者在研究目标、技术路径及应用场景上各具特点，常被视为互补的两个研究方向。

时间序列异常检测的目标是识别历史数据中的异常模式，即发现不符合正常数据变化模式的异常数据点或异常时间段，如图2-4所示，这些异常可能是设备故障、攻击行为、市场异常波动等。时间序列异常检测方法主要分为基于预测和基于重构两大类，这两类方法分别从不同角度出发，通过对正常数据模式进行建

模，利用检测模型输出与实际观测数据之间的偏差来识别异常。基于预测的时间序列异常检测方法的基本思想在于构建能够捕捉正常数据时序特征和变量间相互关系的预测模型，从而在预测阶段对未来数据进行预估，当模型在正常状态下进行训练后，其对未来数据的预测应具有较高的准确性，而在异常发生时，实际观测值与模型预测值之间往往会出现显著偏差，通过量化预测误差可实现对异常的检测。基于重构的方法则侧重于对正常数据变化模式的建模和学习，在训练阶段，模型通常只接触到正常数据，通过学习其潜在的特征模式，使模型具备较强的重构能力，而在测试阶段模型可以较好地重构正常数据，但由于异常模式未在训练阶段中出现，模型通常无法准确重构异常数据，从而导致重构误差升高，通过设置合适的误差阈值，可以将高误差的数据判定为异常。

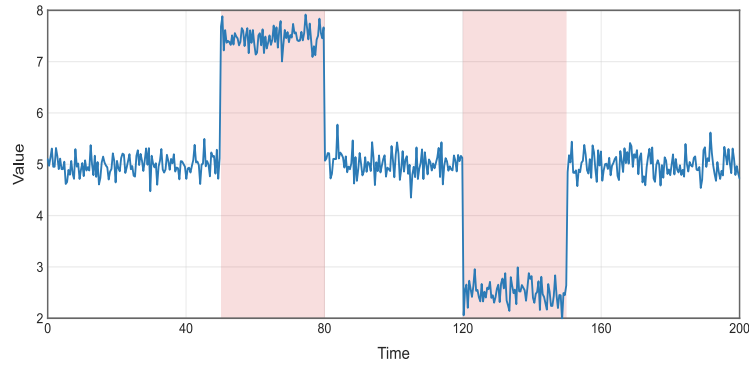


图2-4 时间序列异常检测

2.2.3 异常预测

时间序列异常预测的目标是通过对历史时间序列数据进行分析，预测未来可能发生的异常情况，如图2-5所示，这与异常检测基于历史数据识别已发生的异常不同。

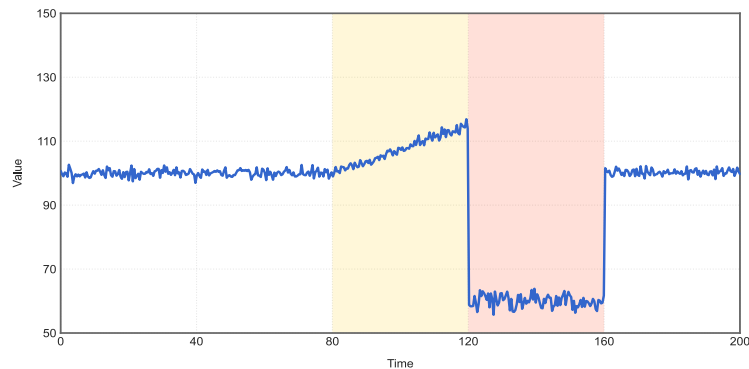


图2-5 时间序列异常预测

尽管时间序列中部分异常行为具有明显的演化规律，但由于突发性异常通常呈现出瞬时爆发和高度不确定性，难以提前准确捕捉其异常特征，因此，当前时间序列异常预测的研究主要集中于具有异常演变趋势的异常类型，即那些在实际异常发生前存在可观察先兆或缓慢演化过程的异常行为。时间序列异常预测任务的核心在于识别和建模异常发生前的异常演变时间与异常前兆数据，通过学习其

演变趋势和异常前兆行为，实现在异常真正发生之前的有效预警。

异常演变时间是指从异常演变初始阶段直至异常真正发生的这一时间区间。在这一时间段内，时间序列已开始呈现出异常的微弱征兆，具备一定的规律性、趋势性或结构性变化，该时间段内的数据常常携带了反映系统状态退化、风险积累或功能扰动的全局趋势信息，具有重要的建模价值，是模型学习异常演变过程的重要依据。

异常前兆数据是指异常演变时间段中能够显著反映异常演变趋势的数据片段，这些数据通常包含早期的异常信号特征。通过对异常前兆数据的深入分析与建模，模型不仅能够捕捉异常的潜在演化轨迹，还能一定程度上解释异常发生的诱因和机制，从而提升预测模型的准确性、鲁棒性与可解释性。

2.3 时间序列异常分析相关技术

(1) 多层感知机

多层感知机^[93] (Multilayer Perceptron, MLP) 是一种前馈神经网络，它是深度学习最基本的神经网络架构，广泛应用于分类、回归、特征提取等任务，其基本结构如图2-6所示。MLP主要由输入层、隐藏层、输出层组成：输入层负责接受原始数据；隐藏层通过线性变换和激活函数来进行特征学习，常见的激活函数有修正线性单元函数ReLU^[94]、sigmoid函数^[95]等，并且每个隐藏层的输出都会作为下一个隐藏层的输入；对于分类任务，输出层会对结果进行归一化操作并以概率的形式输出，对于回归任务，输出层则会输出一个连续值。MLP通过反向传播算法计算损失函数的梯度，并使用优化器如Adam来更新权重。

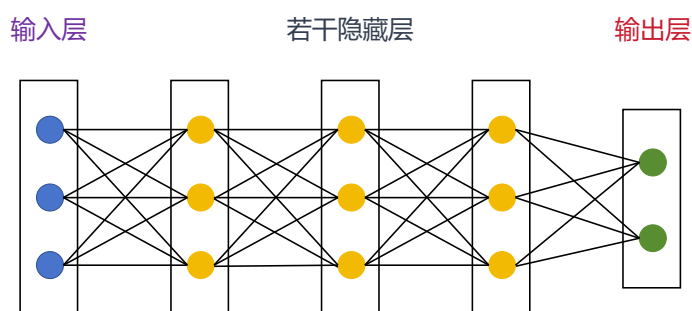


图2-6 多层感知机结构示意图

(2) 注意力机制

注意力机制 (Attention Mechanism) 是一种用于增强神经网络对重要信息的关注并忽略无关信息的技术。注意力机制的核心思想是计算查询 (Query) 与键 (Key) 的匹配度，并对值 (Value) 进行加权求和，其基本过程如图2-7所示。Transformer模型通过缩放点积注意力以及多头注意力机制来有效建模时间序列的长时间的依赖关系，注意力计算过程如公式 (3.3) 所示，其中 Q 是查询向量，代表当前需要关注的内容； K 是键向量，代表所有可参考的信息； V 是值向量，

代表最终加权得到的输出信息； d_k 是键向量的维度， $\sqrt{d_k}$ 作为缩放因子防止梯度过大。

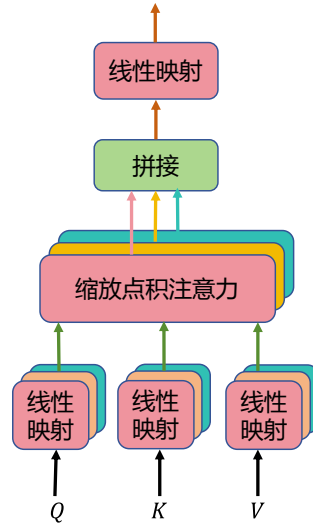


图2-7 多头注意力机制

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3.3)$$

(3) Transformer模型

Transformer^[57]是由Google公司于2017年提出的一种基于多头注意力机制的深度学习模型，其最初主要被用于进行自然语言处理任务，经过近些年的不断发展，如今已被广泛应用于计算机视觉、时间序列预测和异常检测等领域。与传统的RNN模型和LSTM模型比较，Transformer具有更强的并行计算能力，并且能有效捕捉长时间依赖关系，因此特别适用于进行时间序列异常检测与异常预测任务。

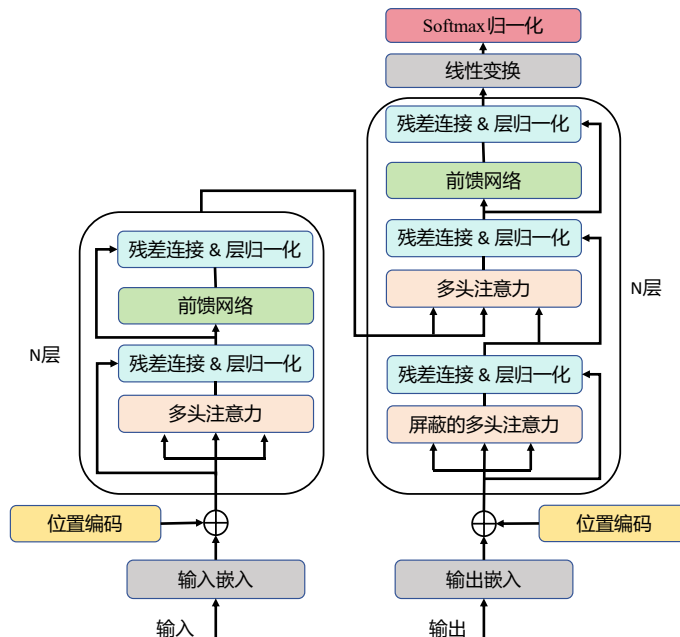


图2-8 Transformer模型结构

Transformer模型的基本结构如图2-8所示，其核心由编码器和解码器组成，输入数据经过位置编码以补充时间位置信息，随后通过自注意力机制生成注意力分数来加权输入数据，从而关注关键时间步信息。多头注意力机制允许模型同时关注多个不同的时间依赖模式，并结合前馈神经网络进一步提取特征，最后通过残差连接和层归一化提高训练稳定性。

（4）图神经网络

图神经网络（Graph Neural Network, GNN）是一类专门用于处理图结构数据的深度学习模型，广泛应用于社交网络分析、推荐系统、生物信息学以及时间序列分析等领域。GNN通过聚合节点及其邻居的信息，学习节点的特征表示，从而有效捕捉图数据中的局部和全局关系，如图2-9所示。其基本框架包括消息传递（Message Passing）机制，在每一层中，节点通过邻居节点的信息进行更新，以增强特征表达能力。典型的GNN变种网络包括图注意力网络^[96]（Graph Attention Network, GAT）、图卷积网络^[97]（Graph Convolutional Network, GCN）、图自编码器^[98-99]（Graph Autoencoder, GAE）等。

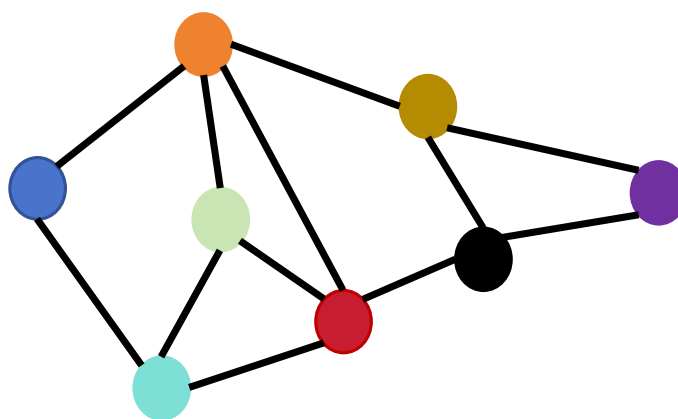


图2-9 图神经网络结构

（5）频域分析技术

频域分析技术作为一种经典而有效的信号处理方法，目前被广泛应用于时间序列分析中的周期性模式识别、噪声抑制与特征提取等任务，常见的技术手段包括快速傅里叶变换、离散傅里叶变换、功率谱密度估计等，主要原理是通过傅里叶变换（Fourier Transform, FT）将原始的时间序列从时域映射到频域，从而揭示时序数据中所包含的各类频率成分及其能量分布特性，如图2-10所示。相比于传统的时域分析，频域分析能够更清晰地识别序列中的周期性波动、高频噪声及低频趋势，相关频域特征也可以和时域特征相结合输入到机器学习或深度学习模型中，以提升时间序列分析结果的准确性和鲁棒性。

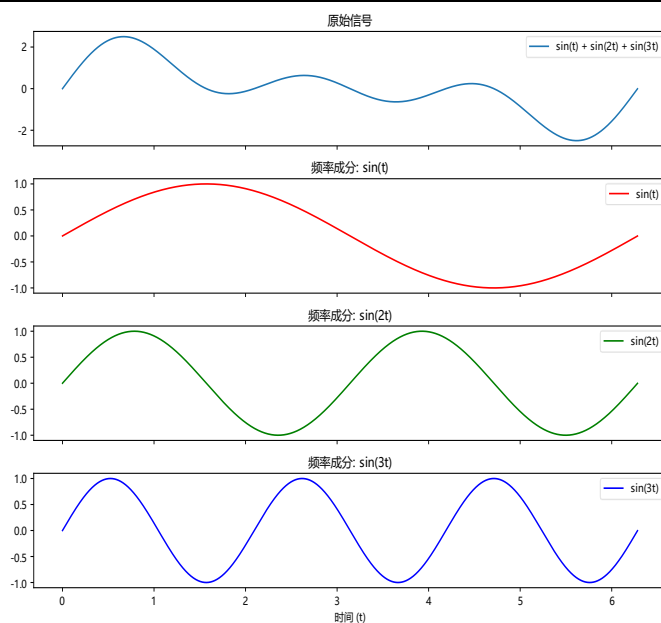


图2-10 使用傅里叶变换提取原始信号中不同频率成分

2.4 本章小结

本章主要介绍本文研究所涉及到的相关理论知识，包括对时间序列、异常、多维时间序列异常检测和异常预测的相关理论知识的介绍，以及本文研究所涉及到的深度学习中一些典型模型和技术的介绍，为本文后续模型的设计奠定理论基础。

第3章 基于时空双维特征协同学习的多维时间序列异常检测模型

多维时间序列通常是由多个互相关联的单变量时间序列组成，其中每个单变量时间序列都对应了一个复杂系统内部的不同特征或者特定属性。港口设备所产生的时序数据是由不同类型的传感器采集形成，本质上是多维时间序列。多维时间序列既和单变量时间序列一样具有时间维度（Temporal Dimension）特征，即各个单变量时间序列中的数据随时间变化的模式特征，这些特征不仅与具体的时间点相关，也与时间的顺序、历史信息以及时间间隔等因素密切相关；除此以外多维时间序列还具有单变量时间序列所不具有的空间维度（Spatial Dimension）特征，即多维时间序列不同变量序列之间的关系、结构或者依赖性。因此多维时间序列每个时刻的每个变量值同时取决于该变量的历史数据值和其它变量此时刻的数据值，要想达到先进的多维时间序列异常检测效果就必须同时考虑到时间维度和空间维度特征，这就引出了多维时间序列异常检测任务中的重要问题和挑战——如何准确、鲁棒地进行时间维度和空间维度的特征学习。

3.1 本章研究挑战

3.1.1 时空双维特征协同学习

在港口设备管理场景中，实际采集的数据通常来源于多个异构传感器，涵盖电气系统、液压系统、结构振动监测、操作行为记录等多个子系统，不同类型变量在采样频率、数值尺度、演化规律、响应延迟等方面存在显著差异，导致时序数据呈现显著的高维异质性，同时港口设备各子系统之间具有复杂的物理结构连接与逻辑控制关系，异常信号往往呈现跨维度传播或多点协同激发的现象，呈现出显著的变量相关性。

而当前多维时间序列异常检测方法大多聚焦于单一建模视角，或仅从时间维度提取序列动态演化特征，或仅基于空间结构分析变量间关系，缺乏对时间依赖性与空间关联关系的联合建模能力，难以有效捕捉港口设备运行过程中关键数据的异常演化路径与传播机制，这会降低模型对多变量间复杂协同行为的建模能力，影响模型异常检测的准确性、鲁棒性与对高风险异常事件的响应效

率。因此，构建融合时间依赖建模与空间交互结构学习的统一建模策略，对于提升港口设备多维时间序列异常检测性能具有重要意义。

为解决现有方法在处理港口设备时序数据时存在的时空特征建模割裂问题，本章希望构建一种面向时间与空间双维特征协同学习框架，联合挖掘时间序列的动态演化特性与变量之间的结构依赖关系，提升模型对复杂时空交互模式的表征能力，从而增强模型异常检测的准确性与鲁棒性。

3.1.2 时间维度特征学习

在多维时间序列的时间维度特征学习任务中，Transformer模型因为其内部的自注意力机制能够有效捕捉时间序列数据中的长期依赖关系而被越来越多的研究人员使用。Transformer模型在设计之初主要用于完成自然语言处理任务（Natural Language Processing, NLP），其内部的多头自注意力机制能够有效捕捉文本数据中的长距离依赖关系和全局上下文信息，克服了传统的循环神经网络和长短期记忆网络模型在这方面的限制，极大地促进了自然语言处理技术的发展。近些年，不少研究人员使用Transformer模型作为骨干网络，将其运用在时间序列的时间维度特征学习任务中，来更有效地捕捉时间序列数据中的长期依赖关系。然而港口设备在实际运行中会产生具有高维异质、变量动态相关、局部变化模式负责等特点的多维时间序列数据，这类数据在结构、分布和动态行为特征上均显著区别于自然语言文本，这些特点导致传统Transformer模型在处理该类数据时会面临以下新的挑战：

（1）面向文本任务的建模机制难以充分捕捉港口设备时序数据中复杂的局部变化模式。自然语言处理任务中的输入数据是词向量序列，其中的每个单词都具有单独的语义信息，而时间序列中每个时刻的数据点并不具备独立的语义表达能力，通常需要结合其邻近时间片段所构成的局部上下文区域来共同建模其动态变化特征。在港口设备的运行监测数据中，设备状态变化常表现为周期短、幅度小但趋势突出的行为模式，如突发性波动、短周期异常抖动等，这类局部行为模式往往不具备全局性特征，需要通过局部时间窗口内的数据特征进行识别。然而传统的Transformer模型会将时间序列的每个时刻的数据点当作一个token进行处理，忽视了时间序列和词向量序列之间的点级语义颗粒度所包含的语义信息的差异，导致对时间序列中的局部语义信息提取不充分，缺乏对时序数据局部语义信息及局部动态变化模式的细粒度建模能力。

（2）共享注意力机制难以准确学习港口设备时序数据各个变量的异质特征。港口设备数据由多个异构类型的传感器采集，不同变量在物理意义、动态变化模式以及异常表现形式上存在显著差异，变量间存在高度异质性。而传统Transformer模型的共享点积注意力机制通过矩阵相乘的方式来直接进行不同变量间的信息交互，对所有变量采取统一建模策略，导致模型在学习各个变量

的异质特征时会引入无关变量的信息干扰，削弱各个变量内部时序依赖关系的学习效果，难以实现对不同异常行为的精准建模，严重影响模型的检测性能与泛化能力。

本章希望通过对传统Transformer模型的内部结构和数据处理的方式进行改进，以使其在处理港口设备时序数据时既能够利用高效的注意力机制建模时间序列长期的依赖关系，又能够有效解决其面向文本任务的建模机制难以充分捕捉港口设备时序数据中复杂的局部变化模式特征问题，同时提升模型对变量异质性特征的学习能力，削弱不同变量语义信息之间的差异性对特定变量内部时序依赖关系学习的干扰。

3.1.3 空间维度特征学习

港口设备运行产生的多维时间序列是通过不同类型的传感器采集，这些传感器数据之间往往存在着某种关联性，例如在一些机械系统中压力和温度往往会相互影响。充分建模多维时间序列不同变量之间的依赖关系可以帮助异常检测模型更好地分析异常在多维时间序列内部不同变量之间的传递情况。不同传感器采集的数据之间的依赖关系在空间上可以看作是一种图结构，每个传感器可以看作是图结构上的一个图节点，图结构中节点与节点之间的边可以看作是不同传感器数据之间的依赖关系。

图神经网络近年来在多维时间序列的异常检测任务中表现出色，许多研究已利用其强大的图结构建模能力来捕捉多变量数据之间的空间依赖关系。然而图神经网络难以有效捕捉港口设备时序数据变量之间的动态相关性，这是因为现有大多数图神经网络方法在构建图结构时普遍采用先验知识手动定义节点之间的连接关系，或者通过相似性度量计算静态邻接矩阵来构建图结构，这些做法忽视了在实际系统中不同变量之间的连接关系并非恒定不变，而是随着时间和外部扰动不断变化，静态的图结构无法有效反映这些随时间动态变化的变量间依赖关系，从而限制了模型对异常模式变化的响应能力和检测准确度。

本章希望设计一种图结构动态学习机制，用于实现多维时间序列图结构的动态构建，充分挖掘不同变量间随时间变化的局部依赖关系，使用图注意力网络对局部动态图中节点特征进行深度更新与聚合，并且结合局部动态图与全局静态图的信息，构建更为精准且具有鲁棒性的空间特征表示。

3.2 本章主要工作

本章设计了一种新颖基于时空双维特征协同学习的多维时间序列异常检测模型MTAD-TSD (Multivariate Time Series Anomaly Detection via Temporal-Spatial Dual-Dimensional Feature Learning)，旨在解决现有方法在对港口设备数据进行异常检测时存在的以下问题：时空特征建模割裂、基于Transformer模型

的时间维度特征学习方法对变量局部特征和异质性特征提取不足、基于图神经网络的空间维度建模方法难以捕捉变量动态相关性。MTAD-TSD能够同时从时间维度和空间维度对多维时间序列进行特征学习，并在融合两种维度特征信息的基础上实现更高效、鲁棒的异常检测性能。在时间维度特征学习方面，本章在Transformer原始的点积注意力基础上进行改进，提出了单变量注意力机制UA（Univariate Attention），使模型专注于各个变量异质性特征的提取，降低了多变量间的信息干扰，本章基于UA设计了时间维度特征学习网络Uaformer，其通过时间序列分段和Time2Vec编码方法实现了对时间序列长期依赖关系与局部特征的联合建模，提升了模型对高维异质与局部变化模式复杂的时间序列建模能力；在空间维度特征学习方面，本章设计了一种基于局部图结构自适应性构建的图结构动态学习方法，实现对变量间关联结构的动态构建，并结合图注意力网络，进一步挖掘变量间的局部动态相关性和全局静态相关性特征，有效解决了传统图神经网络无法捕捉变量动态相关性的问题。

本章在三个公开的数据集和两个港口设备数据集上进行了大量对比实验、消融实验和参数敏感性实验，验证了模型的异常检测性能、内部关键模块和方法设计的有效性以及参数设置的有效性。本章工作的创新点概括如下：

（1）本章设计了一种多维时间序列异常检测模型MTAD-TSD，该模型首次将时空双维特征协同学习方法引入到港口设备运行场景中，有效解决了现有方法在该场景下存在的时空特征建模割裂的问题。

（2）在时间维度特征学习方面，本章设计了一种单变量注意力机制UA，并基于此机制提出了时间维度特征学习网络Uaformer，其能够有效建模时间序列长时间依赖关系，并提升了对数据局部动态特征的提取能力，同时充分解决了在建模变量异质性特征时存在的多变量信息干扰问题。

（3）在空间维度特征学习方面，本章设计了一种基于局部图结构自适应性构建的图结构动态学习方法，实现对变量间关联结构的动态构建，并结合图注意力网络，进一步挖掘变量间的局部动态相关性和全局静态相关性特征，有效解决了传统图神经网络无法捕捉变量动态相关性的问题。

3.3 问题定义

定义3.1（多维时间序列） 给定 $X = \{x_1, x_2, \dots, x_N\} \in R^{N \times M}$ 为一个长度为 N 的多维时间序列， M 为变量维度， $x^t \in R^M$ 是在 t 时刻的多维时间序列观测值，是一个 M 维的向量。

定义3.2（多维时间序列异常检测任务） 给定一个长度为 N 的多维时间序列 $X = \{x_1, x_2, \dots, x_N\} \in R^{N \times M}$ ，变量维度为 M ， $x^t \in R^M$ 是其 t 时刻的值。多维时间序列异常检测任务是给出一个二元标签值用来表示 t 时刻是否存在异常。

3.4 MTAD-TSD模型介绍

3.4.1 MTAD-TSD整体结构概述

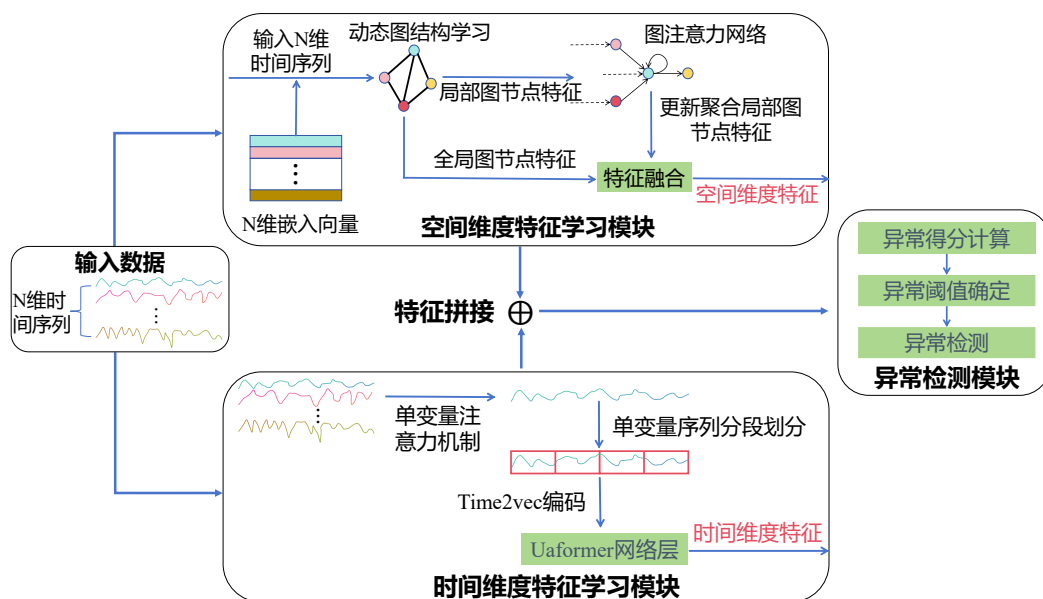


图3-1 MTAD-TSD整体结构

本章设计的多维时间序列异常检测模型MTAD-TSD创新性地从多维时间序列的时间维度和空间维度同时进行特征学习，能够高效、鲁棒地实现多维时间序列异常检测。MTAD-TSD框架的整体结构如图3-1所示，其主要包含以下几个部分：

(1) 时间维度特征学习模块。为了充分提取时间序列内部的语义信息并且减少不同变量之间的依赖关系对于各个变量内部信息提取的干扰，本章在原始Transformer模型内部的缩放点积注意力机制基础上进行了改进，提出了单变量注意力机制，并在单变量注意力机制的基础上设计了Transformer模型的变体Uaformer，用来进行多维时间序列时间维度的特征学习。该模块的输入为待处理的多维时间序列，输出是学习到的时间维度的特征向量。

(2) 空间维度特征学习模块。为了准确学习到不同变量之间的依赖关系并且克服无法自动构建用于图神经网络学习的多维时间序列图结构和传统的图神经网络无法捕捉多维时间序列中每个变量独特特性的缺陷，本章设计了一种多维时间序列图结构自动学习和图节点权重更新机制。该模块的输入为待处理的多维时间序列以及为每个变量随机生成的嵌入向量，在训练过程中，嵌入向量会不断更新，以更好地表征变量的全局信息，从而更有效地捕捉变量之间的依赖关系，该模块会输出学习到的空间维度特征向量。

(3) 异常检测模块。该模块用来计算多维时间序列的异常得分以及进行异常检测，其输入为时间维度特征向量和空间维度特征向量拼接而成的特征向量，输出为一个二元标签值用来表示是否监测到有异常发生。

3.4.2 时间维度特征学习模块

3.4.2.1 单变量注意力机制

为了能够有效学习多维时间序列时间维度特征，MTAD-TSD使用能够有效提取时间序列中长时间依赖关系的Transformer模型作为骨干网络。本章在3.1.1.1节中详细分析了直接将传统的Transformer模型运用在多维时间序列时间维度特征学习任务中的一些局限性，其中一个局限就是Transformer模型原始的点积注意力机制会干扰多维时间序列各个变量内部特征信息的提取，为了有效解决这一问题，本章在Transformer模型原始的点积注意力基础上进行改进，设计了单变量注意力机制UA，接下来本章对此展开详细说明。

首先解释Transformer模型原始的点积注意力机制是如何干扰多维时间序列各个变量内部特征信息的有效学习。设 Q 、 K 、 V 分别代表Transformer模型中的Query、Key、Value矩阵， d_k 为 Q 、 K 的维度， M 、 N 分别为多维时间序列的变量数和序列长度，Transformer模型原始点积注意力机制的计算过程可用如下公式表示：

$$\tilde{R} = QK^T \quad (3.1)$$

$$R = \text{softmax}\left(\frac{\tilde{R}}{\sqrt{d_k}}\right) \quad (3.2)$$

$$H = \text{Attention}(Q, K, V) = RV \quad (3.3)$$

公式（3.1）中每个查询矩阵和键矩阵会进行点积操作从而得到注意力得分矩阵 \tilde{R} ，公式（3.2）中注意力得分矩阵 \tilde{R} 会通过缩放与softmax归一化操作被转换为对应的注意力权重矩阵 R ，公式（3.3）中注意力权重矩阵 R 和值矩阵 V 会进行加权求和操作，从而得到最终的嵌入矩阵 H 。嵌入矩阵 H 中的任一元素 $H_{i,j}$ 的计算过程可以通过如下公式表示：

$$\tilde{R}_{i,j} = \sum_{m=1}^M Q_{i,m} \times K_{m,j}^T \quad (3.4)$$

$$H_{i,j} = \sum_{n=1}^N R_{i,n} \times V_{n,j} \quad (3.5)$$

图3-2进一步揭示了Transformer模型原始的点积注意力机制计算输入序列对应的嵌入表示的过程，其中变量数 $M = 5$ ，序列长度 $N = 4$ ，符号“ Δ ”表示在生成嵌入序列元素 $H_{1,1}$ 时参与注意力计算过程的元素位置。根据公式（3.4）、公式（3.5）和图3-2可知，Transformer模型原始的点积注意力机制在为多维时间序列的每一个变量维度计算嵌入表示时，都会有其它变量维度的信息参与计算

过程，这可能会对当前变量内信息提取造成干扰，导致各个变量内部的时序信息无法被准确学习。

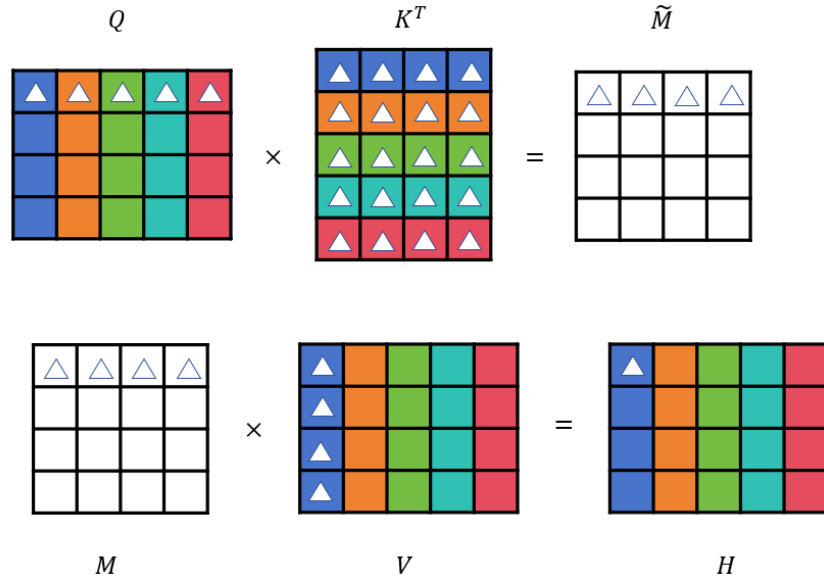


图3-2 点积注意力机制中的信息流动示意图

本章设计的单变量注意力机制UA的结构如图3-3所示。单变量注意力机制首先会对输入的多维时间序列进行变量分离处理，多维时间序列 $X \in R^{N \times M}$ 将会按照变量被分为 M 个长度为 N 的单变量时间序列，随后分离后的单变量时间序列会被并行地进行缩放点积注意力计算得到每个变量的嵌入向量，之后会对嵌入向量进行特征拼接，最终得到整个多维时间序列的嵌入矩阵 H 。

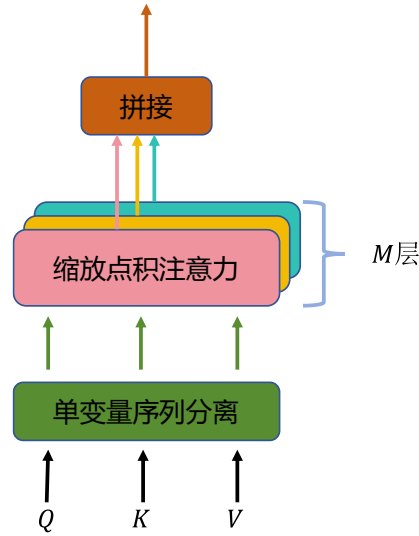


图3-3 单变量注意力机制结构流程示意

单变量注意力机制的计算过程可表示为如下公式：

$$\tilde{R}_{\sigma,ij} = Q_{\sigma,i} \times K_{\sigma,j}^T \quad (3.6)$$

$$H_{\sigma,i} = \sum_{n=1}^N R_{\sigma,in} \times V_{\sigma,n} \quad (3.7)$$

$$H = \text{Concat}(H_{1,:}, \dots, H_{\sigma,:}, \dots, H_{M,:}) \quad (3.8)$$

$R_{\sigma,ij}$ 表示多维时间序列中的第 $\sigma \in [1, M]$ 个单变量序列数据在进行点积注意力计算后得到的注意力权重矩阵 R 中的第 i 行第 j 列元素，公式（3.6）到公式（3.8）表明经过单变量注意力计算后得到的嵌入矩阵 H 中，每个变量数据维度的信息是该变量内部不同时刻上的数据进行信息交互运算得到的，并没有涉及到其它变量维度的信息交互，因此单变量注意力机制能够有效减少其它变量维度时序信息对本变量维度时序信息提取的干扰。可以看到单变量注意力机制并未考虑不同变量之间的信息交互，这并不意味着这部分信息交互不重要，MTAD-TSD模型将这部分工作放到了空间维度信息提取模块中，以更好地建模多维时间序列空间维度特征。

3.4.2.2 时间序列分段与Time2vec编码

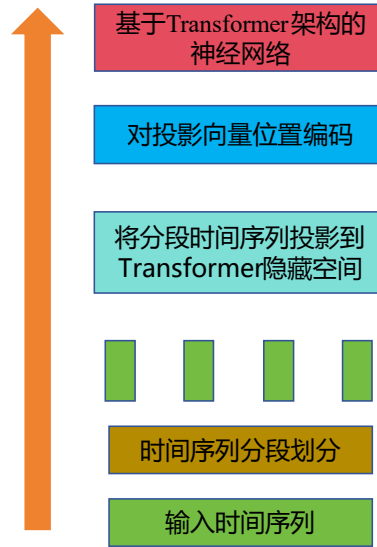


图3-4 时间序列分段处理流程示意

传统的Transformer模型在进行多维时间序列时间维度特征学习时的另一个局限性是其无法有效提取时间序列中的局部语义信息，因为其将时间序列逐点进行输入处理，而时间序列中的单个时间步上的数据并没有明确的语义意义，只有和局部范围内的数据放在一起才具有一定的语义信息。为了能够充分学习到时间序列中的局部语义信息，MTAD-TSD将Transformer原本处理时间序列数据的方式由逐点输入处理改为逐段输入处理，具体来说就是对输入的时间序列进行分段划分，接着使用一个可训练的投影矩阵将每一个分段后的时间序列映射成具有Transformer隐藏层维度的向量空间，并对映射后的向量进行编码生成对应的嵌入向量输入到Transformer结构中。图3-4展示了MTAD-TSD进行时

间序列分段处理的过程。

由于自注意力机制无法直接捕获输入序列数据的顺序关系，因此本章在将时间序列进行分段处理以及投影到Transformer的隐藏层空间后还需要对投影后的向量进行编码以生成带位置信息的嵌入向量，这里本文指出Transformer原始的位置编码方法在对分段时间序列进行编码时存在如下问题。如公式(3.9)和公式(3.10)所示，Transformer模型原始的标准位置编码方法使用正弦和余弦函数编码序列位置，其中 pos 是时间步的位置， i 是变量维度索引， d_{model} 是嵌入向量的维度，并且所使用的正余弦函数具有固定的频率和相位，因此当其对不同分段时间序列进行编码时，会为不同分段时间序列中具有相同位置索引的时间步上的数据生成完全一致的编码，无法捕捉时间序列的周期性和复杂的时间模式。

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \quad (3.9)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \quad (3.10)$$

为了弥补Transformer模型中正余弦位置编码在处理分段时间序列时的不足，本章决定采用Time2Vec编码^[100]方法，这是一种针对时序信息设计的可学习向量表示方法，其能够从输入数据中动态学习正弦函数的权重、频率和相位偏移，以捕获时序数据的周期性模式，同时使用线性函数捕获非周期性模式，其编码过程可以用公式(3.11)表示，其中 \mathcal{F} 是周期性激活函数，这里Time2Vec使用正弦函数作为周期性激活函数，因为正弦函数具有天然的周期性，适合捕捉时间序列数据中的周期性模式。 ω 和 φ 是可学习参数， ω 控制正弦函数的周期，通过学习得到不同的频率，适应不同时间模式的周期性， φ 调整正弦函数的相位，用于捕捉时间序列的偏移特性。Time2Vec首先使用线性函数将时间序列由低维空间映射到高维空间，用于学习时间序列整体趋势或者偏移，接着使用周期性激活函数对时间序列进行映射，用来捕捉时间序列中的周期性变化模式，最终将映射后的线性成分和周期性成分拼接成向量进行输出，可以使得时间序列的每个时间点数据都可以由一个既包括线性变化部分也包括周期性变化部分的向量表示。

$$t2v(X^i)[j] = \begin{cases} \omega_j X^i + \varphi_j & \text{if } j = 0 \\ \mathcal{F}(\omega_j X^i + \varphi_j) & \text{if } 1 \leq j \leq m \end{cases} \quad (3.11)$$

3.4.2.3 Uaformer网络

在前文描述的基础上，本章将Transformer模型内的多头注意力机制变更为单变量注意力机制，并且对输入的时间序列进行分段和Time2Vec编码处理，设计了Uaformer网络结构，其主要结构及工作流程如图3-5所示。

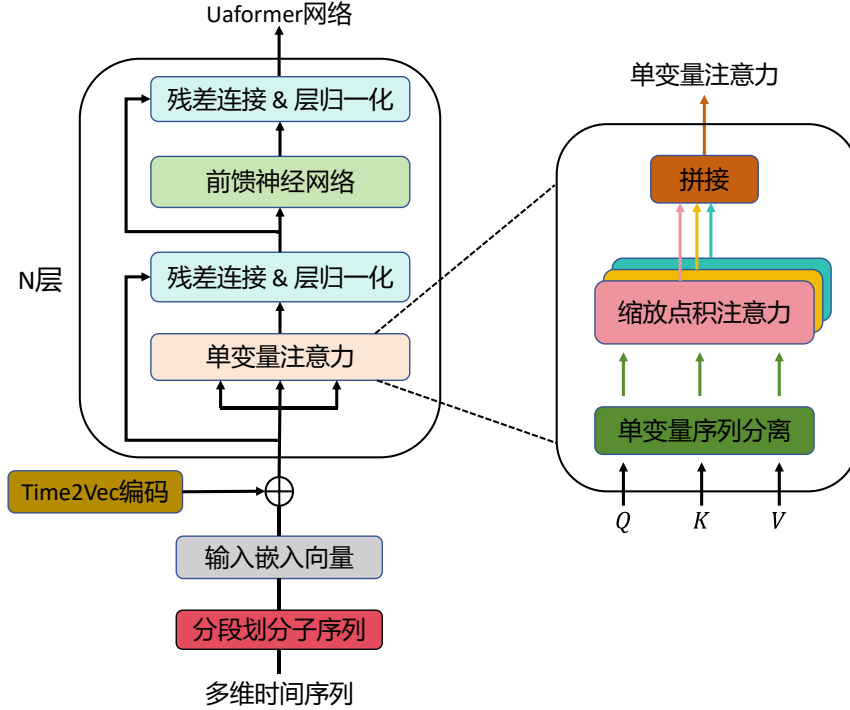


图3-5 基于单变量注意力的Uaformer网络结构示意图

MTAD-TSD在学习多维时间序列时间维度特征时将多个Uaformer网络层堆叠在一起，实现对时间序列各个变量内部时序信息的深度建模学习。Uaformer网络学习时间序列时间维度特征的过程如公式（3.12）和公式（3.13）所示，其中 $z_{t,k}$ 表示第 k 层Uaformer网络在时刻 t 时的输出结果，假设总共有 N 层网络， $LayerNorm$ 表示网络层的归一化操作， UA 表示单变量注意力机制， MLP 表示多层感知机。同时为了避免梯度消失和提高模型的收敛速度，MTAD-TSD对第 N 层Uaformer网络的输出结构进行加权残差连接操作，如公式（3.14）所示，其中 $x_t^d \in R^{d \times M}$ 表示在输入时刻为 t 时，变量维度 M ，长度 d 的时间窗口多维时间序列， μ 为模型超参数，用以动态调整在残差连接中的原始输入的比例，以取得更好的模型泛化性能。

$$\hat{z}_{t,k} = LayerNorm(\hat{z}_{t,k-1} + UA(z_{t,k-1})) \quad (3.12)$$

$$z_{t,k} = LayerNorm(\hat{z}_{t,k} + MLP(\hat{z}_{t,k})) \quad (3.13)$$

$$\tilde{z}_t = \mu x_t^d + z_{t,N} \quad (3.14)$$

3.4.3 空间维度特征学习模块

3.4.3.1 局部动态图结构学习

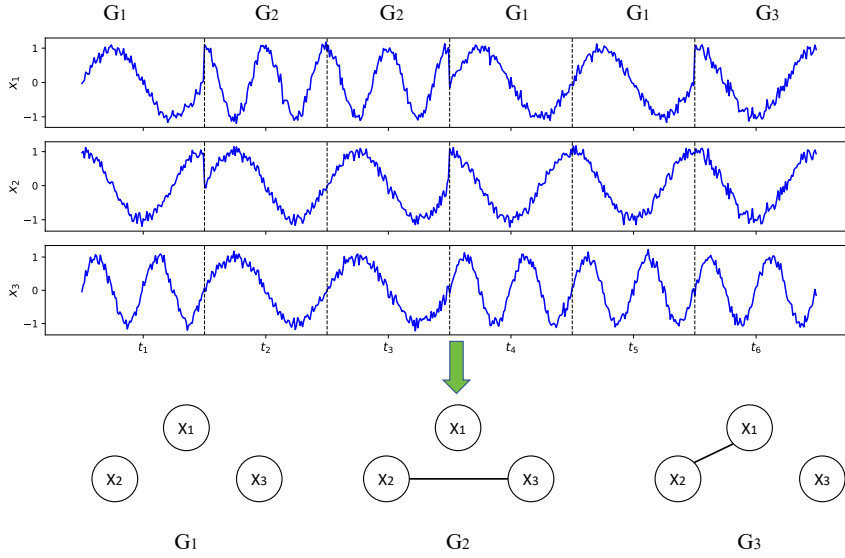


图3-6 局部动态图结构构建

多维时间序列变量间的连接关系不是固定不变的，而是会随着时间变化在不同的时间段内呈现出不同的连接关系，如图3-6所示。为了能够捕捉变量间的动态相关性，MTAD-TSD设计了一种图结构动态学习方法，该方法首先会基于相同的时间窗口大小 d 将输入的多维时间序列 X 划分成 q 个长度相同的时间序列片段 $\{X_1, X_2, \dots, X_q\}$ ，同时在每个时间序列片段 X_k 中使用一个有向图结构 $G_k = (V, E)$ 来表示各个变量维度之间的依赖关系，其中各个变量维度由图节点集合 $V = \{v_m | m \in [1, M]\}$ 来表示，各个变量维度之间的依赖关系由有向图的边集合 $E = \{e_{ij} | e_{ij} \in A_k\}$ 表示，其中 A_k 表示有向图 G_k 的邻接矩阵， $e_{ij} \in \{0, 1\}$ 是邻接矩阵 A_k 中的元素， $e_{ij} = 0$ 表示有向图 G_k 中的节点 i 和节点 j 之间没有边，也就是说多维时间序列中的变量 i 和变量 j 所代表的序列数据在时间片段 X_k 中不存在依赖关系，反之若 $e_{ij} = 1$ 表示有向图 G_k 中的节点 i 和节点 j 之间有边连接，即多维时间序列中的变量 i 和变量 j 所代表的序列数据在时间片段 X_k 中存在依赖关系。因此构建有向图 G_k 的关键就是计算出 G_k 的邻接矩阵 A_k ，本章采用公式(3.15)至公式(3.17)所示的计算方法计算邻接矩阵 A_k ，其中多层感知机 MLP_1 和 MLP_2 分别用于对时间序列片段 X_k 进行线性映射， α 是缩放因子，用于控制非线性函数的饱和度以防止梯度消失，通过使用 $LeakyReLU$ 激活函数和外积差计算非对称的邻接矩阵 A_k 。

$$\tilde{X}_k^1 = \tanh(\alpha MLP_1(X_k)) \quad (3.15)$$

$$\tilde{X}_k^2 = \tanh(\alpha MLP_2(X_k)) \quad (3.16)$$

$$A_k = LeakyReLU\left(\tanh\left(\alpha\left(\tilde{X}_k^1(\tilde{X}_k^2)^T - \tilde{X}_k^2(\tilde{X}_k^1)^T\right)\right)\right) \quad (3.17)$$

通过对所有的时间序列片段 $\{X_1, X_2, \dots, X_q\}$ 分别进行上述计算过程，最终能够得到 q 个不同的有向图邻接矩阵 $A = \{A_1, A_2, \dots, A_q\}$ 。接着每个有向图邻接矩阵 A_k 会与对应的时间片段序列 X_k 一起输入到图神经网络中进行图节点特征提取，进而得到不同时间片段内的图节点结构特征 h_k ，如公式（3.18）所示。

$$h_k = GNN(X_k, A_k) \quad (3.18)$$

3.4.3.2 基于图注意力网络的局部图节点特征学习

为了进一步学习不同变量维度间的局部依赖关系，MTAD-TSD引入了图注意力网络（Graph Attention Networks, GAT）用来深度聚合和更新各个时间序列片段中的图节点特征信息，使得各个节点的特征表示能够包含其相关邻居节点的信息。图注意力网络进行图节点特征学习的计算过程可由公式（3.19）至公式（3.23）表示。其中 h_k^i 是模型通过局部图结构动态学习后得到的节点 i 在时间序列片段 X_k 中的特征向量表示， x_k^i 是节点 i 在时间序列片段为 X_k 中输入特征， $W \in R^{M \times M}$ 是一个可训练的权重矩阵用来对各个节点的输入进行线性变换， g_k^i 由这两者拼接而成， u 是一个用于计算注意力权重的可学习参数矩阵， ρ_{ij} 为节点 i 和节点 j 之间的注意力权重， $N(i) = \{j | e_{ji} > 0\}$ 表示节点 i 的邻居节点集合， Z_k^i 表示在时间序列片段 X_k 中经过图注意力网络计算得到的节点 i 的特征表示， Z^i 为将节点 i 在各个时间序列片段中的特征表示拼接后得到的最终特征。

$$g_k^i = h_k^i \oplus W x_k^i \quad (3.19)$$

$$\tau_{ij} = u_i^T g_k^i + u_j^T g_k^j \quad (3.20)$$

$$\rho_{ij} = \frac{\exp(\tau_{ij})}{\sum_{c \in N(i) \cup \{i\}} \exp(\tau_{ic})} \quad (3.21)$$

$$Z_k^i = \text{ReLU} \left(\rho_{ii} W x_k^i + \sum_{j \in N(i)} \rho_{ij} W x_k^j \right) \quad (3.22)$$

$$Z^i = \text{Concat}(Z_1^i, Z_2^i, \dots, Z_q^i) \quad (3.23)$$

3.4.3.3 全局静态图结构学习

为了在聚焦局部动态图结构学习的同时丰富各个节点的全局特征表示，MTAD-TSD在模型训练开始前会随机生成 N 维嵌入向量 $S = \{s^i | i \in [1, M]\}$ ，这些嵌入向量会随着多维时间序列一起参与模型后续的训练，并且在训练过程中不断更新向量，以学习不同节点的全局特征。在每一个训练轮次结束时，MTAD-TSD会将通过图结构动态学习和图注意力网络得到的节点特征 Z^i 与更新

后的嵌入向量 s_i 进行特征融合，并将融合后的结果作为最终学习到的各个节点的空间特征表示，如公式（3.24）所示，最终得到的节点特征表示 χ^i 既包含了节点 i 在空间结构上的局部动态特征信息，也包含了全局特征信息。

$$\chi^i = s^i \circ Z^i \quad (3.24)$$

3.4.4 异常检测模块

MTAD-TSD采取基于预测的方式进行异常检测，即对于在时刻 t 给定的输入时间窗口大小为 d 的数据 $x_d^t = [x^{t-d}, x^{t-d+1}, \dots, x^{t-1}]$ ，模型通过公式（3.25）预测该时间窗口的下一个时刻数据 \hat{x}^t ，其中 z_{time}^t 和 $z_{spatial}^t$ 分别表示经过时间维度特征提取模块和空间维度特征提取模块学习到的多维时间序列时间维度特征和空间维度特征，“ \oplus ”表示特征拼接操作， f_{MLP} 表示全连接层输出函数。

$$\hat{x}^t = f_{MLP}(z_{time}^t \oplus z_{spatial}^t) \quad (3.25)$$

MTAD-TSD使用均方误差（Mean Square Error, MSE）作为模型的损失函数进行训练，其定义如公式（3.26）所示，其中 T_{train} 训练数据集的数据长度， x^t 为时刻 t 时的实际数据。

$$L_{MSE} = \frac{1}{T_{train} - d} \sum_{t=d+1}^{T_{train}} \|\hat{x}^t - x^t\|_2^2 \quad (3.26)$$

MTAD-TSD将 \hat{x}^t 与 x^t 差的 L_2 范数的平方值作为时刻 t 的异常得分，如公式（3.27）所示，同时为了保证异常时间段能够被准确检测到，MTAD-TSD采用了与文献^[101]相同的异常阈值确定方法，该方法能够通过动态修正，提升模型在实际异常检测任务中的准确性和鲁棒性。

$$Score = \|\hat{x}^t - x^t\|_2^2 \quad (3.27)$$

3.5 实验及分析

3.5.1 实验设置

3.5.1.1 软硬件环境设置

本章实验所使用到的软硬件环境包括Python版本3.9.12、Pytorch版本1.13.0、Pytorch-cuda版本11.6、cuda版本11.6、CPU型号12核i7-12700F、显卡型号RTX 3070Ti、内存64GB。

3.5.1.2 模型参数设置

本章实验相关模型参数设置包括：模型训练批次大小为32、模型学习率为0.01、模型输出层维数为512、Uaformer网络层数量为4、单变量注意力嵌入维数为64、子序列窗口大小为16、优化器为Adam。

3.5.1.3 实验数据集

本章选取了三个在多维时间序列研究中被广泛使用的公开数据集以及两个来自本文合作港口的设备运行数据集作为实验数据集来评估MTAD-TSD的异常检测性能，相关数据集的具体说明如下：

MSL^[102] (Mars Science Laboratory Rover)：MSL数据集由美国国家航天局发布的好奇号火星车遥测数据，共包括55个传感器记录的数据。本实验中MSL的训练集和测试集数据数目分别为78317和53729，异常数据比例为10.72%。

SWaT^[103] (Secure Water Treatment)：SWaT数据集是一个用于工业控制系统安全研究的公开数据集，记录了一个模拟水处理设施在正常运行和遭受攻击时的多种数据，数据集包含了来自51个传感器记录的实时数据。本实验中SWaT的训练集和测试集数据数目分别为47515和44986，异常数据比例为11.97%。

WADI^[104] (Water Distribution System Attack)：WADI数据集是一个用于研究水分配系统安全性的公开数据集，记录了一个模拟的水分配系统在正常操作和遭受攻击时的多种数据，数据集包含了来自127个传感器记录的实时数据。本实验中WADI的训练集和测试集数据数目分别为78795和57275，异常数据比例为5.99%。

港口数据集1：港口数据集1来自一个国内大型港口公司的港口流动机械设备运行实时数据，数据集中包含了来自107个传感器记录的实时数据。本实验中港口数据集1的训练集和测试集数据数目分别为126347和73596，异常数据比例为6.74%。

港口数据集2：港口数据集2来自一个国内大型港口公司的港口固定机械设备运行实时数据，数据集包含了来自85个传感器记录的实时数据。本实验中港口数据集2的训练集和测试集数据数目分别为95317和86375，异常数据比例为8.34%。

3.5.1.4 对比算法

为了评估MTAD-TSD在进行多维时间序列异常检测时的性能表现，本章选取了十种可以进行多维时间序列异常检测任务的算法进行对比实验，这些算法具体说明如下：

PCA^[33]：主成分分析方法 (Principal Component Analysis, PCA) 常用于对

数据进行降维，从而从高维数据中提取最具代表性的特征，其主要通过线性变换将原始数据转化为一个新的坐标系，新的坐标轴即主成分是按照数据方差大小排列的，从而实现数据的压缩和降维。在异常检测任务中，PCA通过对数据进行降维来识别与大部分数据不同的异常点。

K-NN^[30]: K近邻算法 (K-Nearest Neighbors, K-NN) 是一种有监督学习算法，常用于分类和回归任务。其基本原理是基于给定样本的数据特征计算其与训练集中其它样本的距离，选择最近的k个邻居，并依据这些邻居的标签来确定给定样本的标签。

Random Forest^[25]: 随机森林 (Random Forest, RF) 是一种基于决策树的集成学习方法，其主要思路是通过训练一组树模型，并结合每棵树的决策，来识别数据中的异常点或离群点。

LSTM-VAE^[42]: LSTM-VAE结合了长短期记忆网络的序列建模能力和变分自编码器的生成建模能力，适用于处理具有时间依赖性的复杂数据，通常用于处理时序数据的生成和异常检测任务。其中的变分自编码器可以对潜在空间的特征表示进行解码并进行序列重构，可以根据重构误差来检测时间窗口中是否存在异常前兆数据。

DCdetector^[56]: DCdetector是一种基于对比学习的异常检测模型，其通过设计一种具有双重注意力的对比学习结构来从不同角度学习时间序列的特征表示，并且通过对比学习机制来不断放大正常数据和异常数据之间的特征差异性，进而进行数据异常检测。

GDN^[64]: GDN是一种基于图神经网络的异常检测模型，其通过图神经网络构建不同变量维度之间的依赖关系，并且使用图注意力网络对节点间的特征表示进行更新，基于节点的特征表示进行预测和异常检测。

FuSAGNet^[65]: FuSAGNet在GDN基础上引入了稀疏自编码器，通过在其潜在空间中引入稀疏性来执行重构任务，通过稀疏性约束的联合优化方法来同时优化Sparse-AE和GDN组件，进而同时利用重构误差和预测误差计算异常得分。

Anomaly Transformer^[59]: Anomaly Transformer基于自注意力机制提出异常注意力机制，对每个数据点同时建模两种关联关系：基于可学习高斯核函数构建的先验关联关系，用于刻画时间点之间的局部邻接偏好；基于传统自注意力机制计算的序列关联关系，用于捕捉全局依赖结构，并且设计了一种极小极大化的联合优化策略来放大正常样本与异常样本之间的关联差异，从而构建出一种基于关联差异的异常检测方法。

FEDformer^[74]: FEDformer将傅里叶变换与Transformer模型相结合，利用大多数时序数据在傅里叶变换中具有稀疏表示的特性将时域信息拆解为频域信息，随机采样部分频域，然后投影回时域，能够在降低计算复杂度的同时过滤掉多数噪声数据，使得Transformer模型能够更好地学习长时间序列中的依赖关系

，可以有效提升模型的预测及异常检测性能。

TimesNet^[77]: TimesNet基于时间序列多周期性的特点，使用卷积神经网络将复杂的时序变化分解为多个周期内和周期间的变化，并且通过将时间序列从一维空间扩展到二维空间中，实现了将周期内和周期间的变化分别嵌入到二维张量的列和行中，从而易于使用二维卷积核来进行特征提取，有效提升了模型的泛化性能。

Graph-MoE^[105]: Graph-MoE是一种基于图神经网络和混合专家模型（Mixture of Experts, MoE）的多维时间序列异常检测模型，其内部每一层图神经网络中都实现了一个特定的专家网络，此专家网络包含全局引导实体注意力块，用于图节点表示的层内聚合，同时模型内部的记忆增强路由能够挖掘时间序列之间的跨序列相关模式特征，促进实体表示的层间聚合，最终整合不同图神经网络层的层次信息到实体表示中，使得模型能够充分挖掘多层图神经网络之间的中间信息，有效提升了异常检测性能。

ModernTCN^[106]: ModernTCN是一种基于传统时间卷积网络进行改进的时间序列建模模型，其通过残差连接将不同时间卷积网络层的特征信息进行融合，用于解决梯度消失问题，并且通过门控机制来自适应地控制卷积层之间的信息流动，同时模型采用空洞卷积技术扩大了卷积核的感知范围，有助于捕捉时间序列中的长期依赖关系，ModernTCN还使用了自注意力机制来计算序列中不同位置之间的相关性，进一步提高了模型的建模能力。

iTransformer^[75]: iTransformer是一种基于Transformer模型的时间序列分析模型，其在不修改Transformer任何原有模块的情况下反转Transformer结构，并且将独立序列作为变量标记，通过注意力机制捕捉时序多元相关性，并利用层归一化和前馈网络学习时间序列特征表示，提升模型的时序建模能力。

3.5.1.5 评价指标

本章使用精确率（Precision, P）、召回率（Recall, R）、F1得分（F1-Score, F1）来评价模型性能。精确率是指模型预测为正类的样本中，真正为正类的比例，衡量的是模型在预测正类样本时的准确性，其计算公式如公式（3.28）所示，其中TP（True Positive）表示被模型正确判断为正类的样本数量，FP（False Positive）表示被模型错误地将负类判断为正类的样本数量。召回率是指模型预测为正类的样本中，实际为正类的比例，衡量的是模型能够正确识别正类样本的能力，其计算公式如公式（3.29）所示，其中FN（False Negative）表示模型错误地将正类预测为负类的样本数量。F1分数是精确率和召回率的调和平均数，它结合了精确率和召回率的权衡，提供一个综合的性能评估指标，其计算公式如公式（3.30）所示，本章所进行的实验均选择在多次实验结果中最优的F1分数来评价模型性能。

$$Precision = \frac{TP}{TP + FP} \quad (3.28)$$

$$Recall = \frac{TP}{TP + FN} \quad (3.29)$$

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (3.30)$$

3.5.2 对比实验

MTAD-TSD的对比实验结果如表3-1至表3-2所示，每次实验结果中最优和次优的精确率（%）、召回率（%）、F1分数（%）已在表格中分别被加粗和下划线标记。

表3-1 在公开数据集上的对比实验结果

方法	SWaT			MSL			WADI		
	P	R	F1	P	R	F1	P	R	F1
PCA	51.35	65.26	57.48	62.64	70.19	66.20	81.19	63.24	71.10
K-NN	46.52	61.24	52.87	86.57	72.17	78.72	67.30	58.91	62.83
Random Forest	61.37	74.28	67.21	90.16	63.25	74.34	72.18	87.93	79.28
LSTM-VAE	82.49	76.16	78.10	81.57	58.36	68.04	92.13	71.25	80.29
DCdetector	84.18	81.53	82.83	89.87	82.59	86.10	89.63	85.39	87.46
GDN	88.54	80.72	84.45	93.59	78.63	85.46	95.78	58.21	72.41
FuSAGNet	80.21	83.74	81.94	92.43	83.76	87.92	91.72	78.59	84.99
A.Transformer	78.13	84.62	81.25	68.39	92.33	78.58	76.56	87.28	81.57
FEDformer	79.26	85.71	82.36	74.24	81.32	77.62	88.25	83.48	85.80
TimesNet	86.19	91.73	88.87	87.61	86.24	86.92	90.53	<u>92.81</u>	<u>91.66</u>
Graph-MoE	80.52	78.49	79.49	84.76	77.95	81.21	86.74	83.75	85.22
ModernTCN	86.38	<u>90.56</u>	<u>88.42</u>	84.78	81.17	82.94	84.51	87.12	85.80
iTransformer	84.26	85.39	84.82	<u>93.62</u>	<u>89.43</u>	91.45	85.68	89.93	87.75
MTAD-TSD	<u>86.59</u>	88.47	87.52	95.46	85.17	<u>90.02</u>	<u>93.68</u>	95.49	94.58

实验结果表明MTAD-TSD在WADI、港口数据集1和港口数据集2上均取得了最优的结果，在MSL数据集上取得了次优结果，整体表现突出，这是因为MTAD-TSD基于时空双维特征协同学习的方法能够充分、准确地学习到多维时间序列的时间维度和空间维度的依赖关系，其中时间维度特征学习模块能够有效兼顾时间序列长期依赖关系和局部变化特征的提取，并且能够有效减少其它变量时间序列的时序信息带来的干扰，准确学习到各个变量的异质性特征，空间维度特征学习模块有效解决了传统图神经网络无法捕捉变量动态相关性的问题，实现了对变量间关联结构的动态构建。另外从实验结果可以看出，传统的机器学习方法如PCA、K-NN、Random Forest由于没有充分利用时间序列时间

维度的依赖信息而表现较差，与之相比的是LSTM-VAE、Anomaly Transformer、FEDformer和iTransformer由于充分捕捉了时间维度依赖关系而有较好的异常检测效果，而GDN、FuSAGNet和Graph-MoE由于充分建模了时间序列空间维度依赖关系也有较好的表现，但这些模型未能同时学习时间维度和空间维度特征，从而整体表现不如MTAD-TSD。DCdetector由于通过对比学习充分捕捉了时间序列中正常数据和异常数据之间的特征差异性也取得了相对较好的表现，但因缺乏对序列的整体建模分析从而表现不如MTAD-TSD。TimesNet和ModernTCN通过卷积神经网络建模时间序列中的长期依赖关系也取得了较好的检测效果，但未能充分挖掘序列局部语义信息和变量独特特性，从而整体表现不如MTAD-TSD。

表3-2 在港口数据集上的对比实验结果

方法	港口数据集1			港口数据集2		
	P	R	F1	P	R	F1
PCA	68.12	49.32	57.22	64.21	58.79	61.38
K-NN	58.28	76.94	66.32	53.47	62.31	57.55
Random Forest	78.39	72.18	75.16	75.32	69.16	72.11
LSTM-VAE	86.57	62.14	72.35	82.49	77.26	79.79
DCdetector	82.85	78.08	80.40	79.04	75.33	77.14
GDN	84.71	72.43	78.09	85.53	78.24	81.72
FuSAGNet	87.46	80.17	83.56	86.15	82.29	84.18
A.Transformer	75.06	92.86	83.02	80.54	83.71	82.09
FEDformer	82.25	78.46	80.31	83.35	85.24	84.28
TimesNet	87.63	83.74	85.64	<u>87.08</u>	89.43	<u>88.24</u>
Graph-MoE	80.15	81.28	80.71	76.59	83.05	79.69
ModernTCN	<u>89.32</u>	85.83	<u>87.54</u>	83.57	87.34	85.41
iTransformer	88.68	84.59	86.59	85.62	<u>90.47</u>	87.98
MTAD-TSD	92.72	<u>87.38</u>	89.97	89.17	92.65	90.88

图3-7展示了MTAD-TSD在各个数据集上具体的异常检测数据，结果显示MTAD-TSD在大多数数据集中均取得了较高的TP和TN值，说明其对异常样本和正常样本均具有良好的鉴别能力，表明模型在这些场景下具有较高的异常检测准确性与鲁棒性。但模型在部分测试集下也具有一定程度的漏报现象，如在SWaT数据集中未被正确检测出的异常数据占异常数据总数的比例达到了13.42%，表明模型在减少异常数据漏报方面还需进一步优化，以提高其在实际工业环境中的适用性和可靠性。

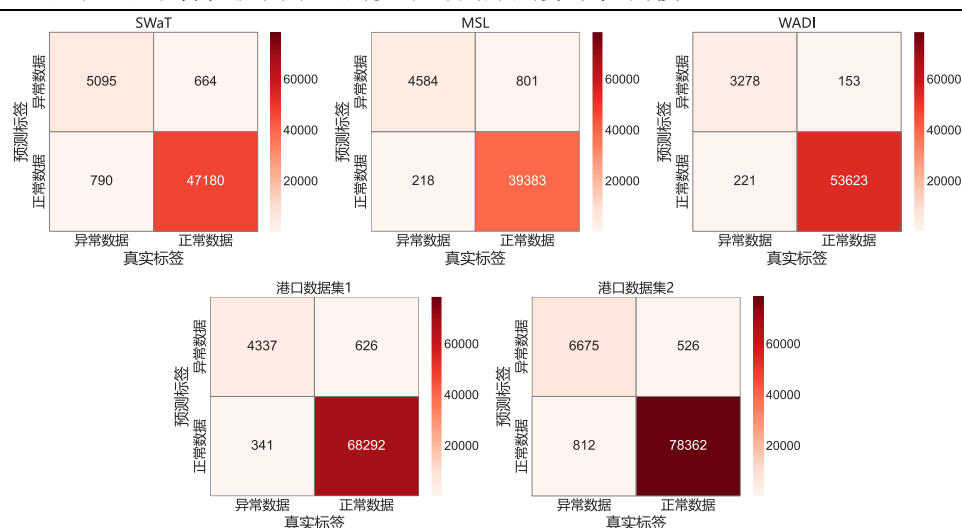


图3-7 MTAD-TSD在各个数据集上的异常检测数据

3.5.3 消融实验

为了验证本章提出的MTAD-TSD模型的时间维度特征学习模块和空间维度特征学习模块对模型异常检测性能提升的有效性，本章在五个数据集上均进行了模块消融实验，消融实验对象分别是无时间维度特征学习模块的MTAD-TSD、无空间维度特征学习模块的MTAD-TSD、完整的MTAD-TSD，实验结果如表3-3至表3-4所示，完整MTAD-TSD模型的实验结果已在表中被加粗表示。

表3-3 在公开数据集上的模块间消融实验结果

消融模块	SWaT			MSL			WADI		
	P	R	F1	P	R	F1	P	R	F1
时间维度特征学习	79.52	81.34	80.43	84.79	80.24	82.44	83.71	85.65	84.67
空间维度特征学习	75.68	72.36	73.97	81.95	69.42	75.47	85.62	78.43	81.79
完整MTAD-TSD	86.59	88.47	87.52	95.46	85.17	90.02	93.68	95.49	94.58

表3-4 在港口数据集上的模块间消融实验结果

消融模块	港口数据集1			港口数据集2		
	P	R	F1	P	R	F1
时间维度特征学习	83.29	78.93	81.05	85.13	76.52	80.56
空间维度特征学习	74.27	80.51	77.26	78.32	82.68	80.44
完整MTAD-TSD	92.72	87.38	89.97	89.17	92.65	90.88

实验结果表明在消融时间维度特征学习模块的情况下，MTAD-TSD的异常检测性能显著下降，这主要是由于时间维度特征学习模块能够有效兼顾时间序列长期语义信息和局部语义信息的提取，充分学习到时间序列的短时变化模式，并且能够有效减少其它变量时间序列的时序信息带来的干扰，准确学习到各个变量时间序列内部的异质性特征。此外，消融空间维度特征学习模块也会

大幅降低MTAD-TSD的异常检测性能，因为空间维度特征学习模块能够有效构建时间序列各个变量间的局部动态结构和全局静态结构，从而高效地进行空间维度特征学习。

为了验证MTAD-TSD的时间维度特征学习模块中一些关键创新方法对模型异常检测性能的影响，本章在所有数据集上均进行了时间维度特征学习模块内部消融实验。这些关键创新方法包括单变量注意力机制、Time2Vec编码、单变量序列分段处理，本次消融实验对象分别包括将单变量注意力机制改为点积注意力机制、将Time2Vec编码改为位置编码、不进行单变量序列分段处理后的MTAD-TSD模型。实验结果如表3-5至3-6所示，其中完整MTAD-TSD模型的实验结果已在表中被加粗表示。实验结果表明使用Transformer传统的点积注意力替换本章提出的单变量注意力后，MTAD-TSD的异常检测性能出现了显著下降，主要原因是单变量注意力机制在对多维时间序列时间维度进行特征学习时能够有效减少不同变量时间序列的时序信息带来的干扰，准确学习到各个变量时间序列内部异质性特征。使用位置编码替换Time2Vec编码后，模型的异常检测性能也有所下降，这是因为在对时间序列进行分段处理后传统的位置编码无法有效捕捉时间序列中的周期性和非周期性变化模式，对时间序列特征学习不充分。不使用MTAD-TSD中的单变量序列分段处理方法也会显著降低模型异常检测性能，因为单变量序列分段处理方法使模型能够有效兼顾时间序列长期语义信息和局部语义信息的同时提取，充分捕捉序列局部变化模式特征。

表3-5 在公开数据集上的时间维度特征学习模块内部消融实验结果

消融方法	SWaT			MSL			WADI		
	P	R	F1	P	R	F1	P	R	F1
单变量注意力	82.31	84.95	83.61	89.97	82.14	85.88	88.67	86.47	87.56
Time2Vec编码	81.79	83.47	82.63	89.06	81.37	85.48	84.57	89.26	86.84
单变量序列分段	82.19	83.53	82.85	86.21	80.79	83.41	85.74	81.31	83.47
完整MTAD-TSD	86.59	88.47	87.52	95.46	85.17	90.02	93.68	95.49	94.58

表3-6 在港口数据集上的时间维度特征学习模块内部消融实验结果

消融方法	港口数据集1			港口数据集2		
	P	R	F1	P	R	F1
单变量注意力	87.34	82.13	84.65	87.20	84.07	85.61
Time2Vec编码	88.16	82.18	85.07	85.49	80.93	83.15
单变量序列分段	84.38	80.25	82.26	85.95	83.06	84.48
完整MTAD-TSD	92.72	87.38	89.97	89.17	92.65	90.88

为了验证MTAD-TSD的空间维度特征学习模块中一些关键创新方法对模型异常检测性能的影响，本章在五个数据集上均进行了空间维度特征学习模块内部消融实验。这些关键方法包括局部图结构学习、图注意力网络、全局图结构

学习，本次消融实验对象分别包括未进行局部图结构学习、未使用图注意力网络、未进行全局图结构学习后的MTAD-TSD。实验结果如表3-7至3-8所示，完整MTAD-TSD模型的实验结果已在表中被加粗表示。实验结果表明未进行局部图结构学习后，模型的异常检测性能显著下降，因为本章提出的局部图结构自适应学习方法实现了对变量之间局部动态结构的自动构建，充分捕捉了时间序列变量之间的动态相关性，提升模型的异常检测性能。未使用图注意力网络后，模型的异常检测性能出现了下降，因为图注意力网络能够深度聚合和更新时间序列图结构中各个节点的嵌入向量信息及其邻居节点的特征表示。未进行全局图结构学习后，模型的异常检测性能有所下降，因为模型此时学习到的空间维度特征只包含了变量间的局部相关性，容易被短时噪声干扰，降低了模型的鲁棒性。

表3-7 在公开数据集上的空间维度特征学习模块内部消融实验结果

消融方法	SWaT			MSL			WADI		
	P	R	F1	P	R	F1	P	R	F1
局部图结构学习	79.62	80.16	79.89	85.16	76.23	80.45	87.18	88.46	87.82
图注意力网络	78.56	81.93	80.21	84.56	80.39	82.42	87.47	85.27	86.36
全局图结构学习	81.85	83.76	82.79	87.39	79.27	83.13	89.32	87.18	88.24
完整MTAD-TSD	86.59	88.47	87.52	95.46	85.17	90.02	93.68	95.49	94.58

表3-8 在港口数据集上的空间维度特征学习模块内部消融实验结果

消融方法	港口数据集1			港口数据集2		
	P	R	F1	P	R	F1
局部图结构学习	82.13	81.76	81.94	83.07	85.51	84.27
图注意力网络	78.56	81.10	79.81	80.76	84.37	82.53
全局图结构学习	84.54	82.46	83.49	85.31	88.07	86.67
完整MTAD-TSD	92.72	87.38	89.97	89.17	92.65	90.88

3.5.4 参数敏感性实验

为了评估模型关键参数对MTAD-TSD异常检测性能的影响，本章设计并进行了参数敏感性实验，重点分析Uaformer网络层数与子序列窗口大小这两个参数的变化对模型性能的影响。本实验过程中仅调整上述两个参数，其余参数均保持与3.5.1.2节中的参数设置一致。

图3-8展示了不同的Uaformer网络层数量对模型性能的影响结果。实验表明，在WADI、港口数据集1与港口数据集2上，当Uaformer网络层数设置为4时，模型的F1-Score达到最优；而在MSL和SWaT数据集上，最优的Uaformer网络层数分别为2和6。这一差异反映出不同数据集在时序特征学习方面的复杂性存在差异。例如SWaT数据集中可能包含更复杂的特征信息，因此需要更多的网络结

构以捕捉时序数据的深度依赖关系；而MSL数据集的特征结构相对简单，较少的网络层即可实现有效建模。

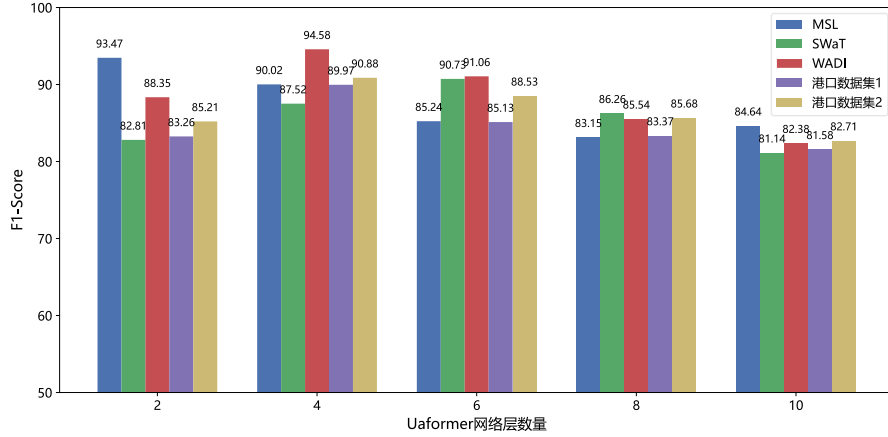


图3-8 不同数据集上Uaformer网络层数量对模型性能的影响

图3-9展示了子序列窗口大小对模型性能的影响。实验结果表明，当子序列窗口大小设置为16时，模型在MSL、港口数据集1及港口数据集2上取得了最优的结果；而对于SWaT和WADI数据集，最优的子序列窗口大小分别为8和32。主要原因在于不同数据集在局部时序结构中的特征密度存在显著差异，SWaT数据集中较短的子序列已包含足够丰富的关键特征，因此较小的子序列窗口即可实现有效的特征学习；而WADI数据集中的局部时序结构更为复杂，需要更长的子序列窗口来进行数据特征学习。此外，子序列窗口大小还影响模型对局部图结构的动态构建能力：子序列窗口过短可能无法覆盖完整的局部图结构，导致模型无法学习到足够的局部图结构信息；而过长的子序列窗口则可能引入多个不同的局部图结构，干扰局部图结构动态构建的准确性。

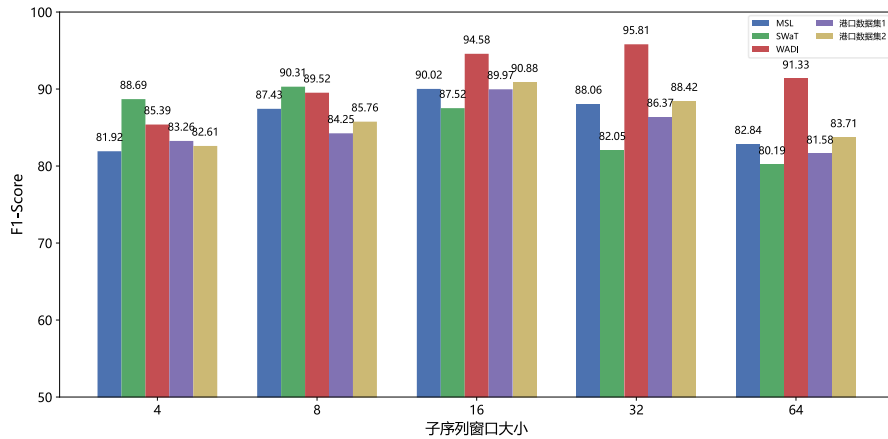


图3-9 不同数据集上子序列窗口大小对模型性能的影响

3.6 本章小结

本章设计了一个基于时空双维特征协同学习的多维时间序列异常检测模型MTAD-TSD。MTAD-TSD能有效解决在港口设备运行场景下多维时间序列异常

检测任务中的以下挑战：1. 时空特征建模割裂，缺乏时间与空间特征的协同学习方法；2. 基于Transformer的方法对变量在时间维度上的局部特征与异质性特征提取不足；3. 基于图神经网络的方法对变量在空间维度上的动态相关性建模不充分。针对第一个挑战，MTAD-TSD设计了一种时空双维特征协同学习方法，能够同时从时间维度和空间维度对多维时间序列进行特征学习，并在融合两种维度特征信息的基础上实现更高效、鲁棒的异常检测性能；针对第二个挑战，MTAD-TSD在传统Transformer模型的共享点积注意力机制基础上设计了单变量注意力机制UA，并基于此设计了时间维度特征学习网络Uaformer，其能够有效建模时间序列长时间依赖关系，并提升了对数据局部动态特征的提取能力，同时有效解决了在建模变量异质性特征时存在的多变量信息干扰问题；针对第三个挑战，MTAD-TSD设计了一种多维时间序列图结构动态学习方法，通过基于局部图结构的自适应性学习实现了对变量之间局部动态结构的自动构建，并且结合图注意力网络和全局图结构特征学习，实现了对多维时间序列更准确、全面地空间特征提取。

本章在三个公开数据集以及两个港口数据集上对MTAD-TSD的异常检测性能进行了大量的对比实验、消融实验和参数敏感性实验，对比实验结果表明MTAD-TSD具有优异的异常检测性能；消融实验结果验证了MTAD-TSD内部关键模块和方法设计的有效性；参数敏感性实验探讨了部分超参数对模型性能的影响。

第4章 基于主导周期掩码和多尺度特征学习的多维时间序列异常预测模型

近年来，多维时间序列异常检测任务在金融风控、网络安全、工业通信等多个关键领域得到了广泛研究与应用。现有研究大多聚焦于异常发生后的检测与定位，即通过对历史观测数据的分析识别已发生的异常点或异常模式，以辅助系统故障的溯源、处置与维护。然而在诸多实际工业生产场景中，异常事件一旦发生往往会造成不可逆的系统损害，因此提前预测潜在异常的发生时间与风险状态，较之事后检测更具现实价值与应用意义。

与异常检测任务不同，异常预测任务旨在基于已有历史数据中潜在的异常演化趋势，判断未来一段时间内是否存在异常发生的可能性，其核心在于提前识别异常前兆数据并进行预警判断。在港口设备等关键基础设施运行场景中，异常预测的重要性尤为突出，港口设备时序数据中会包含一些具有缓慢变化趋势的异常，如性能退化、结构松动等，这些异常通常会伴随着一定程度上的异常前兆波动信号，通过对这些异常前兆信号进行捕捉和特征学习，可以为异常的进行一定程度上的提前预警，降低设备出现实际异常的风险。

然而目前时间序列异常预测研究相对缺乏，并且现有方法大多在统一的时间尺度下对历史序列进行建模，忽略了不同类型异常在演变过程中具有不同的变化趋势以及演变时间，缺乏对多类型异常前兆数据的针对性建模能力，难以有效处理港口设备时序数据中异常前兆数据稀疏且异常类型多样的复杂特点。为此，本章以港口设备运行时序数据为对象，针对具有异常演变趋势而非瞬时突变型的异常开展异常预测相关研究工作，探索适用于港口设备运行场景下的多维时间序列异常预测方法。

4.1 本章研究挑战

4.1.1 稀疏异常前兆数据特征学习

本章所研究的异常类型都是具有异常演变趋势的异常，不包括瞬时突发性异常。具体而言，此类异常在实际发生前，系统状态通常会经历一段时间持续的数据波动、趋势偏移或频率变化，这段时间被定义为异常演变时间，其对应的时间序列数据则称为异常前兆数据。异常前兆数据往往包含着系统潜在故障

或状态异常的早期信号，因而在实现异常的提前识别与预测方面具有重要意义。在港口设备的实际应用中，一些潜在的严重故障如轴承疲劳裂纹、齿轮啮合异常等在其初始演变阶段常常缺乏显著的异常前兆数据特征，并且此类异常前兆数据在完整的时间序列中占比极低，数量远少于正常状态下产生的数据，导致其特征易被海量正常数据所掩盖。这种异常前兆数据的稀疏性使得异常预测模型在特征学习过程中更倾向于捕捉正常数据的特征，进而难以有效建模并识别异常前兆信号。

因此如何在港口设备的多维时间序列中有效挖掘稀疏的异常前兆数据特征，增强模型对异常前兆数据的敏感性，同时降低正常数据对异常前兆数据特征学习的干扰，已成为异常预测任务中的关键挑战。因此本章旨在设计一种能够有效建模稀疏异常前兆数据特征的方法，减少正常数据对特征学习的干扰，提高模型在港口设备场景下异常预测的准确性和鲁棒性。

4.1.2 多类型异常前兆数据特征学习

由于港口设备系统结构复杂、传感器种类繁多，其运行过程中产生的多维时间序列数据中往往包含多种类型的异常，这些异常在不同变量维度上的传播机制、表现形式以及演化路径可能存在显著差异，导致各类型异常在不同变量维度上呈现出不一致的异常演变时间与异常演变趋势。以港口机械设备的监测为例，当设备发生潜在故障时，温度传感器数据可能在短时间内迅速升高，压力传感器数据则可能缓慢上升，而振动传感器数据则可能仅在故障即将发生时才显示出明显异常，这些现象反映出异常在不同变量维度上的时序错位性与模式异构性，即不同变量维度不仅具有不同的异常演变趋势，其异常演变时间也存在差异性，这种差异性对异常前兆数据的建模带来了更高的挑战。

现有时间序列异常预测方法大多在统一的时间尺度下对历史序列进行建模，往往忽略了不同类型的异常在演变过程中所表现出的不同演变趋势与演变时间，缺乏对多类型异常前兆数据的准确建模能力。因此本章希望提出一种能够有效建模多类型异常前兆数据的特征学习方法，以实现具有不同演变趋势及演变时间的异常前兆数据特征的准确学习，从而提升模型对多类型异常的预测能力。

4.1.3 基于异常前兆特征的异常预测

基于异常前兆数据特征进行异常预测，重点在于实现对异常前兆数据的存在性检测及其异常波动程度评估，并在此基础上构建更加准确和可靠的异常预测机制。具体而言，本章研究的异常预测结果被划分为两个关键组成部分：异常前兆数据存在性的定性判断；异常前兆数据的异常波动程度的定量评估。

传统的异常检测专注于发现当前数据中的异常点或者异常模式，从而检测

系统当前是否存在异常行为，这侧重于对已发生数据的分析并进行及时响应和处理。异常前兆数据检测则专注于识别实际的异常发生之前的趋势变化，通过识别异常前兆数据进而提前发现问题、提供预警和演变时间，侧重于分析异常演变过程中的波动、趋势或模式，适用于需要提前干预的场景。传统的异常检测方法如聚类分析、统计分析、深度学习等方法都是通过分析异常数据和正常数据之间的特征区别来识别出实际的异常点或者异常模式，但是异常前兆数据并不是真正的异常数据，只是反映了异常演变过程中的变化趋势，其数据特征可能无法支持区分异常前兆数据和正常数据，因此传统的异常检测方法无法有效地检测出异常前兆数据。

当检测到了异常前兆数据的存在后，还无法仅基于此准确预测出未来一段时间内是否会出现异常，因为检测到的异常前兆数据可能处在异常演化过程中的任一阶段，如果异常前兆数据处于异常演化过程的初始阶段，那么实际异常可能会在未来较远的时间窗口内发生，而如果异常前兆数据处于异常演化过程的末段，那么实际异常可能会在未来较近的时间窗口内发生。因此为了给出更加精确的异常预测，就需要确定异常前兆数据处在异常演化过程的哪一阶段。由于本章研究的异常是存在异常演化过程的异常，异常前兆数据所处的阶段越靠近异常演化末段，其波动程度往往越剧烈，因此可以通过评估异常前兆数据的异常程度来推断其所处的异常演化阶段，并且可以对异常前兆数据波动程度进行定量分析，从而实现更加准确的异常预测。

本章希望构建一种异常前兆数据特征与异常预测结果之间高效的关联机制，通过学习到的异常前兆数据特征来进行异常前兆数据的存在性判断，并根据其异常波动程度推断异常演化阶段，最终给出更准确、全面的异常预测结果。

4.2 本章主要工作

为了应对多维时间序列异常预测任务中稀疏异常前兆数据特征难以有效学习、多类型异常前兆数据特征建模不足的挑战，本章设计了一种基于主导周期掩码和多尺度特征学习方法相结合的多维时间序列异常预测模型MTAP-DM（Multivariate Time Series Anomaly Prediction via Dominant Period Masking and Multi-Scale Feature Learning）。MTAP-DM通过生成主导周期掩码序列和多尺度特征学习方法来对异常前兆数据进行特征学习，并且根据特征学习结果进行异常预测。具体来说，MTAP-DM首先从频域视角对每个单变量时间序列进行周期性分析，借助快速傅里叶变换计算其主导周期，并构造主导周期掩码序列，以减少正常数据对稀疏异常前兆数据特征学习的干扰，强化异常前兆数据特征学习结果；在主导周期掩码序列基础上，MTAP-DM进一步采用多尺度子序列划分策略，构建不同时间窗口下的子序列视图，通过并行的多尺度特征学习模块对不同尺度的时序数据进行特征提取与融合，以捕捉具有不同演变趋势和演

变时间的异常前兆数据特征；最终MTAP-DM根据学习到的不同尺度下的异常前兆数据特征分别提出了基于多尺度特征相似度的异常前兆数据检测方法和基于序列重构的异常前兆数据波动程度评估方法，并综合这两部分计算结果给出最终的异常预测结果。

本章在三个公开的数据集和两个港口设备数据集上进行了大量的对比实验、消融实验、参数敏感性实验，验证了MTAP-DM的异常预测性能、内部关键模块和方法设计的有效性以及参数设置的有效性。本章研究工作中的主要创新点可以概括如下：

(1) 本章设计了一种多维时间序列异常预测模型MTAP-DM，该模型通过生成主导周期掩码序列和多尺度特征学习方法对异常前兆数据进行建模和特征学习，适用于港口设备这种具有稀疏异常前兆数据且异常类型多样的复杂工业场景。

(2) MTAP-DM从频域视角对时序数据进行分析，使用快速傅里叶变换计算时间序列的主导周期，并且通过生成基于主导周期的掩码序列来有效减少时间序列中的正常数据对异常前兆数据特征学习的干扰，从而提升对稀疏异常前兆数据特征学习的准确性。

(3) MTAP-DM通过对主导周期掩码序列进行基于子序列长度的多尺度划分和特征学习，来有效学习具有不同演变趋势和演变时间的异常前兆数据特征，提升了模型对多类型异常前兆数据的特征建模能力。

(4) MTAP-DM设计了基于多尺度特征相似度的异常前兆数据检测方法和基于序列重构的异常前兆数据波动程度评估方法，并基于这两种方法构建异常前兆数据特征与异常预测结果之间高效的关联机制。

4.3 问题定义

定义4.1（异常演变时间和异常前兆信号） 对于一个长度为 N 的多维时间序列 $X = \{x_1, x_2, \dots, x_N\} \in R^{N \times c}$ ， c 为变量维度， $x^t \in R^c$ 是在 t 时刻的多维时间序列观测值。异常演变时间是一个长度为 r 的时间段 $[t-r, t]$ ，在该时间段的数据 $X_{t-r:t}$ 上存在异常波动，并且在未来的数据 $X_{t+1:t+f}$ 中存在真实的异常， f 是未来时间窗口的长度，数据 $X_{t-r:t}$ 被称为异常前兆数据。在异常演变时间内，数据开始从正常变为异常，异常演变时间内数据波动程度越大，未来发生异常的概率越高。需要注意的是，本章研究的异常类型是具有异常演变时间的异常。

定义4.2（多维时间序列异常预测任务） 给定当前时间 t 和长度为 h 的历史子序列 $X_{t-h:t}$ ，模型需要输出一个二元标签值用来表示在未来的序列 $X_{t+1:t+f}$ 中是否会发生异常。

4.4 MTAP-DM模型介绍

4.4.1 MTAP-DM整体结构概述

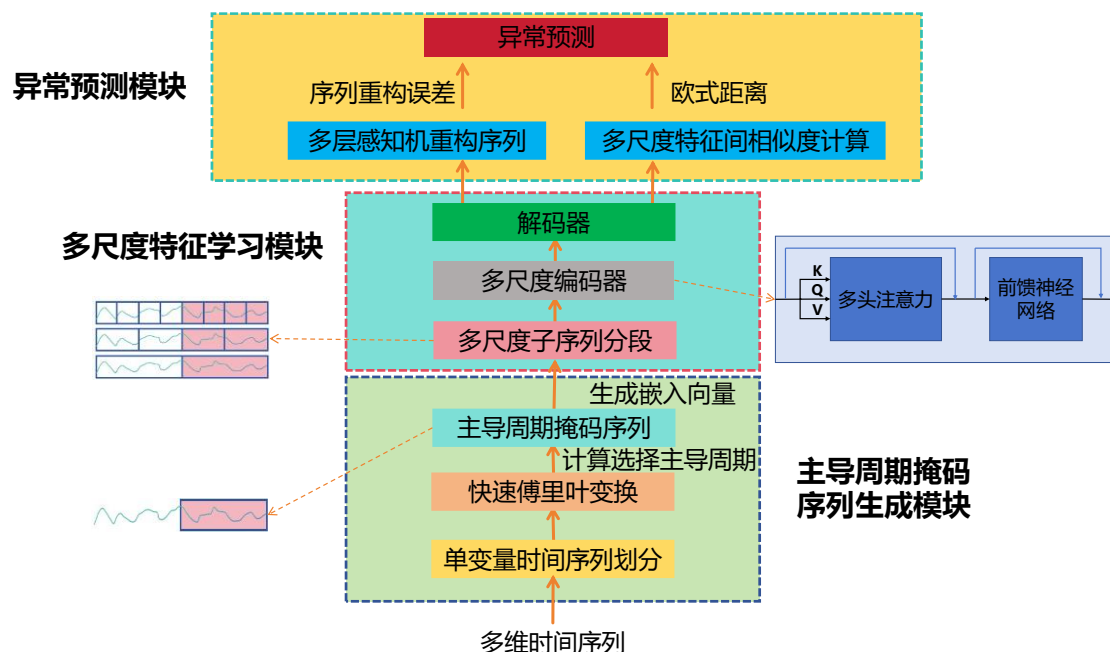


图4-1 MTAP-DM整体结构图

本章设计的多维时间序列异常预测模型MTAP-DM创新性地采用主导周期掩码和多尺度特征学习相结合的方法，能够准确、有效建模异常前兆数据并基于此进行多维时间序列异常预测任务。MTAP-DM的整体结构如图4-1所示，其主要包含以下几个部分：

(1) **主导周期掩码序列生成模块**。MTAP-DM采用快速傅里叶变换方法从时间序列的频域角度对序列进行分析，计算出时间序列的主导周期，并基于主导周期生成对应的主导周期掩码序列，减少时间序列中正常数据对异常前兆数据特征学习的干扰，使模型能够更专注于对稀疏异常前兆数据进行特征学习。

(2) **多尺度特征学习模块**。MTAP-DM对主导周期掩码序列进行了基于子序列长度的多尺度划分，并使用多尺度编码器和解码器进行多尺度特征学习和融合，使模型能够充分、准确地建模并学习具有不同异常演变趋势和演变时间的异常前兆数据特征。

(3) **异常预测模块**。MTAP-DM会对异常前兆数据的多尺度特征进行特征间相似度计算，并根据计算结果定性判断是否存在异常前兆数据；同时MTAP-DM会基于异常前兆数据的多尺度特征进行序列重构，并根据序列重构误差对异常前兆数据的波动程度进行定量计算。最终模型根据这两部分结果建立异常前兆数据特征与异常预测结果之间的关联关系，并给出最终的异常预测结果。

4.4.2 主导周期掩码序列生成模块

时间序列中的主导周期往往反映了系统内在变化的周期性规律，这些周期性的变化代表了序列在正常情况下的演变模式，而异常通常会打破这些规律。当时间序列出现异常时，它的变化模式通常会与正常的周期性模式不一致。异常演变时间可以看作是时间序列从正常状态过渡到异常状态所需要的时间，而这个过渡通常会表现为主导周期的变化或消失，并且主导周期往往和时间序列异常的变化速率有关，主导周期较长的时间序列通常变化较缓慢，而主导周期较短的时间序列则变化较快，如果一个序列的主导周期较短，它可能会较快地对外部扰动或异常产生反应，因而异常演变时间较短；而主导周期较长的序列可能表现出较慢的反应，因而异常演变时间较长。通过分析主导周期，模型可以推断时间序列对异常的反应速度。因此对时间序列的主导周期进行分析计算有助于确定异常演变时间，而异常预测主要基于对异常演变时间内的异常前兆数据进行特征学习来进行的。

为了能够准确学习到异常演变时间内的异常前兆数据特征，减少正常数据对异常前兆数据特征学习的干扰，MTAP-DM从时间序列的频域角度进行分析计算，得到一个主导周期掩码时间序列，并对该掩码序列进行特征学习。MTAP-DM首先利用快速傅里叶变换（Fast Fourier Transformer, FFT）将时间序列的时域信号转化为频域信号，进而提取出时间序列的周期性信息。公式（4.1）是傅里叶变换的计算原理，其中 $x(t)$ 是原始信号随时间 t 变化的函数， $X(f)$ 是信号的频域表示，代表了信号中不同频率成分的强度， $e^{-j2\pi ft}$ 是一个复数指数，代表了一个频率为 f 的旋转因子，傅里叶变换可以计算出信号中每个频率的振幅和相位。

$$X(f) = \int_{-\infty}^{+\infty} x(t)e^{-j2\pi ft} dt \quad (4.1)$$

对于离散信号，傅里叶变换就变为离散傅里叶变换（Discrete Fourier Transform, DFT），如公式（4.2）所示，其中 x_n 是时域信号的第 n 个样本， X_k 是频域信号的第 k 个频率成分， N 是信号的长度， j 是虚数单位。

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi}{N}kn} \quad k = 0, 1, \dots, N-1 \quad (4.2)$$

快速傅里叶变换的思想就是通过递归分治的方法将较大的离散傅里叶变换分解成多个更小的离散傅里叶变换，从而减少计算复杂度、提高计算效率，快速傅里叶变换的计算原理如公式（4.3）所示，对于长度为 N 的信号，FFT首先将信号按照奇偶索引分为两个长度为 $N/2$ 的子序列（分别代表偶数项和奇数项），然后对这两个子序列分别进行离散傅里叶变换，这个过程可以递归进行，直到每个子序列的长度为1为止。

$$X_k = \sum_{n=0}^{\frac{N}{2}-1} x_{2n} e^{-j\frac{2\pi}{N}kn} + \sum_{n=0}^{\frac{N}{2}-1} x_{2n+1} e^{-j\frac{2\pi}{N}kn} + \quad k = 0, 1, \dots, N-1 \quad (4.3)$$

MTAP-DM首先使用快速傅里叶变换计算出时间序列频域中每个频率成分的幅度值，如公式（4.4）所示，其中 $FFT(\cdot)$ 表示快速傅里叶变换， $Amp(\cdot)$ 表示对傅里叶变换结果进行幅度计算， \mathbf{A} 表示频域中每个频率成分的幅度值。

$$\mathbf{A} = Amp(FFT(x)) \quad (4.4)$$

接着MTAP-DM基于频域相似度来度量每个单变量时间序列的子序列之间的相似性，进而选择每个单变量时间序列的主导周期。具体来说，对于当前时间 t ，MTAP-DM选择历史时间步长 h 内的子序列，并计算它们的频率相似度。如公式（4.5）、公式（4.6）和公式（4.7）所示。

$$\mathbf{A}_i = Amp(FFT(x_{t-h:t})) \quad (4.6)$$

$$\mathbf{A}_j = Amp(FFT(x_{t-h-1:t-h-1})) \quad (4.7)$$

$$S_{i,j} = \frac{\mathbf{A}_i \cdot \mathbf{A}_j}{\|\mathbf{A}_i\|_2 \cdot \|\mathbf{A}_j\|_2} \quad (4.8)$$

MTAP-DM会从所有的子序列中选择频率相似度最大的一对子序列作为主导周期候选项，并最终选择相似度最大的 k 个主导周期，如公式（4.8）和公式（4.9）所示。其中 $\mathbf{f} = \{f_1, \dots, f_k\} \in \mathbb{R}^{k \times 1}$ 是在当前时刻之前出现次数最多的 k 个主导周期的频率，与之相对应的 $\mathbf{L} = \{l_1, \dots, l_k\} \in \mathbb{R}^{k \times 1}$ 是这 k 个主导周期的长度。

$$i_{max}, j_{max} = argmax(S_{i,j}) \quad (4.8)$$

$$\mathbf{f} = TopK(\mathbf{A}_{i_{max}}, \mathbf{A}_{j_{max}}) \quad (4.9)$$

为了能够更加准确地学习到异常演变时间内的异常前兆数据特征，MTAP-DM从 \mathbf{L} 中随机挑选一个长度 l ，并对单变量时间序列 x 进行掩码生成对应的掩码序列 x_m ，如公式（4.10）所示，其中 $\|\mathbf{M}\|_1 = l$ 并且 $\mathbf{M} = \{0, \dots, 0, 1, \dots, 1\} \in [0, 1]^T$ 表示在当前时刻之前长度为 l 的主导周期掩码矩阵。这样就得到了一个包含异常演变时间内异常前兆数据的主导周期掩码序列，后续的特征学习都是基于此掩码序列展开的。

$$x_m = \mathbf{M} \odot x \quad (4.10)$$

4.4.3 多尺度特征学习模块

不同类型的异常具有不同长度的异常演变时间，为了能够准确、充分地学习到不同异常类型的异常前兆数据特征，MTAP-DM采用多尺度特征学习方法

对主导周期掩码序列生成模块生成的掩码时间序列 x_m 进行特征学习。具体来说对于掩码时间序列 x_m ，MTAP-DM从细粒度到粗粒度构建 a 个具有不同分段子序列长度的序列，通过在不同尺度下对时间序列序列进行分段划分，得到不同层次的分段序列，从而捕捉到时间序列中不同时间尺度上的信息，如图4-2所示。

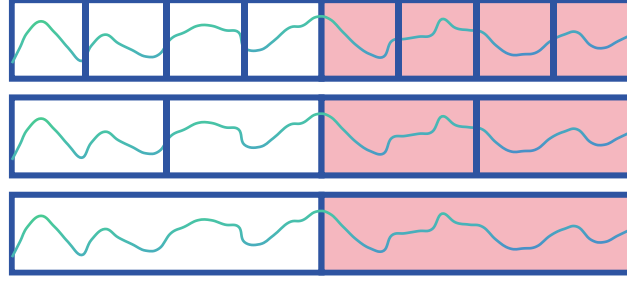


图4-2 对掩码序列进行不同尺度子序列划分

x_m 首先被划分为子序列 $x_1 \in \mathbb{R}^{N_1 \times P_1}$ ，其中 P_1 是最细粒度的子序列尺度大小， N_1 是子序列的数量。第 i 层尺度($1 < i \leq a$)的分段序列由第 $i-1$ 层尺度的分段序列的相邻分段子序列拼接而成，设第 $i-1$ 层的分段序列的子序列大小是 P_{i-1} ，子序列数是 N_{i-1} ，则第 i 层分段序列的生成过程如公式(4.11)所示，其中“ \oplus ”表示相邻子序列拼接操作，这样每个层次序列的子序列分段大小都是上一层次的两倍。

$$x_i = [P_1 \oplus P_2, P_3 \oplus P_4, \dots, P_{N_{i-1}-1} \oplus P_{N_{i-1}}] \quad (4.11)$$

最终经过不断的递进和拼接操作，MTAP-DM会得到不同尺度的分段序列集合 $X_m = \{x_1, x_2, \dots, x_a\}$ ，接着 X_m 会被输入到尺度编码器中，尺度编码器由 a 个编码器组成，每个编码器对应处理不同尺度的分段序列。编码器基于Transformer结构实现，首先通过嵌入层将每个分段序列 x_i 转换为对应的嵌入表示 $x_{i,e}$ ，接着 $x_{i,e}$ 会被输入到多头自注意力层，通过Transformer的多头自注意力机制捕捉跨不同时间区间的时间特征以及不同时间步之间的依赖关系，如公式(4.12)和公式(4.13)所示， Z 是通过尺度编码器学习到的多尺度特征集合。

$$z_p = \text{Transformer}(x_{p,e}) \quad (4.12)$$

$$Z = \{z_1, z_2, \dots, z_a\} \quad (4.13)$$

多尺度特征 Z 之后会通过解码器进行信息融合，解码器由多层感知机实现，如公式(4.14)所示， R 即多尺度学习模块进行多尺度信息融合后的输出特征向量集合。

$$R = \text{Decoder}(Z) \quad (4.14)$$

4.4.4 异常预测模块

异常预测模块主要由两部分组成，一部分用于检测异常前兆数据是否存在

，另一部分主要用于对异常前兆数据的异常波动程度进行评估，最终模型给出的异常预测结果会结合这两部分的计算结果。

首先是异常前兆数据的存在性检测，在多尺度特征学习的背景下，正常的时间序列会有相对一致的模式或趋势，不同尺度样本特征之间的相似性可以用来评估这些样本是否遵循相似的时间序列模式，如果两个样本在特征空间是相似的，那表示它们符合正常模式，反之如果两个样本在特征空间差异较大，那表明样本中可能存在一些异常模式。

因此MTAP-DM通过计算不同尺度特征间的欧几里得距离用来检测是否有异常前兆数据存在，如公式（4.15）所示，其中 z_i 和 z_j 分别是在第 i 和第 j 个尺度下的特征表示， c 为多维时间序列的变量维度数， z_i^k 和 z_j^k 分别表示 z_i 和 z_j 向量中的第 k 个元素。

$$Dist(z_{i,j}) = \sqrt{\sum_{k=1}^c (z_i^k - z_j^k)^2} \quad (4.15)$$

为了准确评估异常前兆数据异常波动的程度，MTAP-DM使用多层感知机来对信息融合后的多尺度特征向量进行重构，并使用重构序列 \hat{x} 和输入序列 x 的均方误差来表示异常前兆数据的波动程度，如公式（4.16）和公式（4.17）所示，其中 N 为输入序列的长度。

$$\hat{x} = MLP(R) \quad (4.16)$$

$$MSE(\hat{x}, x) = \frac{1}{N} \sum_{t=1}^N (\hat{x}_t - x_t)^2 \quad (4.17)$$

最终的异常得分由上述两部分结果组成，如公式（4.18）所示，当前时刻 t 的异常概率 \hat{p}_t 为输入时间窗口内所有数据点的异常得分均值，如公式（4.19）所示。

本章参考文献^[56]中的异常阈值设置方法， \hat{p}_t 若大于设定的异常阈值，则认为时刻 t 之后的一段时间窗口的时间序列会出现异常，这样就将异常概率转换为二分标签值输出。

$$f(x) = Dist(z_{i,j})_{i \neq j, i, j=1, \dots, a} + MSE(\hat{x}, x) \quad (4.18)$$

$$\hat{p}_t = \frac{1}{N} \sum_{t=1}^N f(x_t) \quad (4.19)$$

同时模型的损失函数也由两部分组成，如公式（4.20）、公式（4.21）和公式（4.22）所示，其中 \mathcal{L}_{Rec} 和 \mathcal{L}_{Dist} 分别表示重构操作和多尺度样本间相似度的损失函数， λ_{Rec} 和 λ_{Dist} 是模型权重超参数，用来平衡这两项损失函数的贡献。

$$\mathcal{L} = \lambda_{Rec} \mathcal{L}_{Rec} + \lambda_{Dist} \mathcal{L}_{Dist} \quad (4.20)$$

$$\mathcal{L}_{Rec} = \frac{1}{c} \sum_{i=1}^c \|\hat{x} - x\|_2^2 \quad (4.21)$$

$$\mathcal{L}_{Dist} = \frac{1}{a(a-1)} \sum_{i,j,i \neq j} \|z_i - z_j\|_2^2 \quad (4.22)$$

4.5 实验与分析

4.5.1 实验设置

4.5.1.1 实验环境设置

本章实验所使用的软硬件环境设置与3.5.1.1节中的软硬件环境设置一致，此处不在赘述。

4.5.1.2 模型参数设置

本章实验相关模型参数设置包括：模型训练批次大小为32、模型学习率为0.01、模型输出层维数为512、特征尺度组合为（2,4,8）、注意力嵌入维数为64、历史时间窗口大小为64、预测时间窗口大小为4、优化器为Adam。

4.5.1.3 实验数据集

本章实验所选用的数据集与3.5.1.3节中所使用的数据集一致，数据集的相关信息已在3.5.1.3节中详细介绍，此处不再赘述。

4.5.1.4 对比算法

为了评估MTAP-DM在多维时间序列异常预测任务中的性能，本章选取了一种专门用于多维时间序列异常预测任务的模型PAD和另外七种基于序列重构的多维时间序列异常检测模型进行对比实验，为了使这七种异常检测模型能够适用于异常预测任务，本章的处理方式是把这些异常检测模型的重构序列输入到输出层，同时在这个输出层中拟合一个从当前窗口的重构序列到下一时间窗口异常标签的映射函数。本章实验所选取对比算法中的LSTM-VAE、Anomaly Transformer、DCdetector已经在3.5.1.4节中详细介绍，这里不再赘述，其余对比算法具体说明如下：

PAD^[87]：PAD是一种专门用于对时间序列进行异常预测任务的模型，其主要通过神经控制微分方程对时间序列进行建模，该模型基于“数据增广-知识蒸馏-异常预测”的方法来对数据进行训练，并且在知识蒸馏这一过程中通过使用学生模型拟合教师模型的预测结果来将当前时刻输入时间窗口的数据映射到下

一个时间窗口中，进而进行多维时间序列的异常预测。

D³R^[107]: D³R主要关注数据漂移问题对时间序列异常检测模型性能的影响，该模型提出了一种数据动态分解和扩散重建方法，能够在动态变化的时间序列数据中进行有效的异常检测，以提高在不稳定或非平稳数据上的异常检测效果，其基于重构误差来进行异常检测。

GANomaly^[55]: GANomaly是一种基于生成对抗网络的异常检测方法，该模型的基本思想是利用生成模型来学习正常数据的分布，通过判别模型判断数据是否来自正常分布，并利用重构误差作为异常检测的标准，适合用于无监督的异常检测任务。

CAE-Ensemble^[108]: CAE-Ensemble是一种基于卷积自编码器集成学习的时间序列异常检测模型，其核心思想是利用多样性和集成学习来克服单一模型的局限性，尤其适用于高维、多变量时间序列数据，其通过结合多个卷积自编码器的预测结果来提高异常检测的鲁棒性和准确性。

TranAD^[58]: TranAD是一种基于Transformer模型和生成对抗网络的时间序列异常检测模型，其通过对输入序列施加多种变换来生成具有不同时间特性的增强样本，并且通过引入生成对抗网络学习时间序列更鲁棒的特征表示，其结合局部和全局时间依赖关系来捕捉不同时间尺度上的异常，联合优化重构误差和对抗损失，有效提升了模型异常检测性能。

4.5.1.5 评价指标

本章实验使用精确率（Precision, P）、召回率（Recall, R）、F1分数（F1-Score）来评估模型性能，相关概念和计算公式已在3.4.1.4节中进行说明，这里不再赘述。

4.5.2 对比实验

MTAP-DM的对比实验结果如表4-1和表4-2所示，每次实验中最优和次优的精确率、召回率、F1分数已在表格中分别被加粗和下划线标记。实验结果表明MTAP-DM在除MSL数据集之外的其余数据集上均取得了最优的结果，在MSL数据集上取得了次优的结果，这充分验证了MTAP-DM具有较好的异常预测能力，主要原因在于MTAP-DM从时间序列频域角度进行分析，通过快速傅里叶变换和基于主导周期的掩码序列来减少正常数据对稀疏异常前兆数据特征学习的干扰，准确学习稀疏异常前兆数据特征，同时使用多尺度特征学习方法来进行不同尺度下的异常前兆数据特征学习及融合，进而学习不同异常类型下的异常前兆数据特征，最终通过计算异常前兆数据的多尺度特征相似度和序列重构误差来进行异常前兆数据的存在性判断和异常波动评估，进而给出全面、准确的异常预测结果。同时实验结果表明PAD在所有数据集上也取得了较好的异常

预测结果，主要原因在于PAD通过神经控制微分方程对时间序列进行建模，具有较高的灵活性和特征表达能力，同时其采用基于数据增广、知识蒸馏、异常预测的数据训练方法能够使原始数据集中的正常和异常数据比例更为均衡，有效提升了模型在正常和异常数据不均衡数据集上的泛化能力，并且提升了在异常预测任务中的异常前兆数据特征学习能力。而其余异常检测模型在本次实验中表现一般，主要原因在于这些模型在进行序列建模时未充分考虑到正常数据对异常前兆数据特征学习的干扰，无法充分学习到被大量正常数据掩盖的稀疏异常前兆数据特征，并且这些异常检测模型没有针对异常预测任务设计有效的异常预测评估方法，包括对异常前兆数据的检测和异常程度评估，无法给出全面、准确的异常预测结果。

表4-1 在公开数据集上的对比实验结果

方法	SWaT			MSL			WADI		
	P	R	F1	P	R	F1	P	R	F1
PAD	74.53	86.18	79.93	81.26	89.67	85.26	78.47	<u>81.52</u>	<u>79.97</u>
LSTM-VAE	71.24	73.28	72.25	58.72	64.31	61.39	75.21	67.32	71.05
D ³ R	57.94	48.37	52.72	68.91	53.06	59.96	73.46	63.35	68.03
A.Transformer	65.79	78.30	71.50	74.89	81.24	77.94	<u>81.41</u>	76.28	78.76
GANomaly	81.75	64.23	71.94	76.07	68.55	72.11	65.54	70.18	67.78
DCdetector	78.64	75.28	76.92	71.68	77.39	74.43	76.64	75.45	76.04
CAE-Ensemble	79.31	73.42	76.25	73.08	70.82	71.93	78.07	72.42	75.14
TranAD	<u>82.39</u>	80.53	<u>81.45</u>	80.15	82.41	81.26	75.38	79.24	77.26
MTAP-DM	87.57	<u>84.95</u>	86.24	<u>80.37</u>	<u>88.59</u>	<u>84.28</u>	84.76	82.17	83.44

表4-2 在港口数据集上的对比实验结果

方法	港口数据集1			港口数据集2		
	P	R	F1	P	R	F1
PAD	<u>74.39</u>	78.12	<u>76.21</u>	<u>76.84</u>	<u>80.73</u>	<u>78.74</u>
LSTM-VAE	62.07	71.32	66.37	68.74	65.78	67.23
D ³ R	65.87	74.38	69.87	70.15	76.32	73.11
A.Transformer	71.21	63.56	67.17	74.59	68.25	71.28
GANomaly	53.42	49.35	51.30	62.19	55.31	58.55
DCdetector	68.73	72.54	70.58	73.06	76.55	74.76
CAE-Ensemble	72.87	<u>78.43</u>	75.55	68.23	69.74	68.98
TranAD	70.74	74.61	72.62	73.39	77.23	75.26
MTAP-DM	86.85	90.14	88.46	85.57	88.35	86.94

图4-3统计了MTAP-DM在各个数据集上异常预测的具体结果，结果显示MTAP-DM在所有数据集上均取得了较高的TP和TN值，说明其对异常数据和正常数据均具有较好的预测建模能力。但模型在部分测试集下也具有一定程度的

异常漏报现象，如在MSL数据集和WADI数据集中未被正确预测的异常数据占异常数据总数的比例分别达到了19.63%和15.24%，表明模型在建模异常前兆数据特征及其与异常预测结果之间的关联关系时，仍然存在部分异常前兆数据特征学习不充分和关联关系建立不全面的问题。

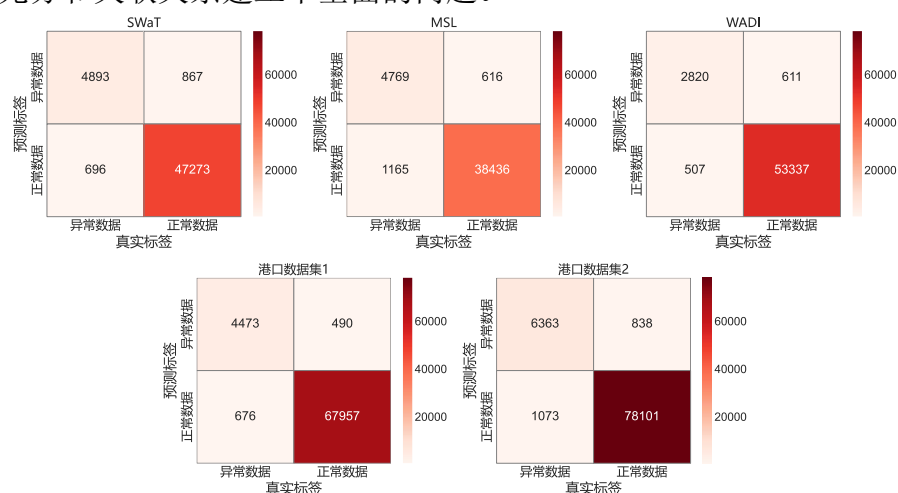


图4-3 MTAP-DM在各个数据集上的异常预测结果

4.5.3 消融实验

为了验证本章提出的MTAP-DM中的主导周期掩码序列生成模块、多尺度特征学习模块和异常预测模块对模型异常预测性能的影响，本章在所有实验数据集上均进行了模块消融实验，实验对象分别包括移除主导周期掩码序列生成模块、移除多尺度特征学习模块、移除异常预测模块改用基于重构序列进行异常预测后的MTAP-DM。实验结果如表4-3和表4-4所示，完整MTAP-DM模型的实验结果已在表中被加粗表示。

表4-3 在公开数据集上的模块消融实验结果

消融模块	SWaT			MSL			WADI		
	P	R	F1	P	R	F1	P	R	F1
频域分析	75.63	72.14	73.84	70.12	72.47	71.28	71.68	70.19	70.93
多尺度特征学习	80.16	78.26	79.20	75.29	71.36	73.25	78.49	73.58	75.88
异常预测	81.27	75.42	78.23	78.56	83.21	80.81	80.18	76.07	78.11
完整MTAP-DM	87.57	84.95	86.24	80.37	88.59	84.28	84.76	82.17	83.44

表4-4 在港口数据集上的模块消融实验结果

消融模块	港口数据集1			港口数据集2		
	P	R	F1	P	R	F1
频域分析	79.02	81.21	80.10	77.18	79.85	78.49
多尺度特征学习	78.24	76.32	77.27	75.61	77.32	76.46
异常预测	83.54	82.56	83.05	81.74	85.65	83.65
完整MTAP-DM	86.85	90.14	88.46	85.57	88.35	86.94

实验结果表明移除主导周期掩码序列生成模块后，模型异常预测性能明显下降，因为在对主导周期掩码序列进行特征学习时可以有效减少正常数据对稀疏异常前兆数据特征学习的干扰，更准确地学习到稀疏异常前兆数据特征。移除多尺度特征学习模块后，模型异常预测性能也显著下降，原因在于多尺度特征学习模块能够进行不同尺度下的异常前兆数据特征学习及融合，能够对处理具有不同异常演变趋势和演变时间的异常前兆数据进行充分地特征学习，能够提高MTAP-DM在处理具有多种异常类型时序数据时的异常预测性能和鲁棒性。移除异常预测模块改用基于重构序列进行异常预测后，MTAP-DM的异常预测性能有所下降，因为该模块可以同时进行异常前兆数据的存在性检测和异常前兆数据的波动程度评估，最终结合这两部分结果给出更准确、全面的异常预测结果。

为了验证MTAP-DM的主导周期掩码序列生成模块中部分关键的创新方法对模型异常预测性能的影响，本章在所有数据集上均进行了主导周期掩码序列生成模块的内部消融实验。这些关键创新方法包括单变量时间序列划分、主导周期掩码序列生成，本次消融实验对象分别包括不进行单变量时间序列划分、使用随机周期掩码序列替代主导周期掩码序列后的MTAP-DM模型。实验结果如表4-5和4-6所示。实验结果表明不进行单变量时间序列划分后，MTAP-DM的异常预测性能出现下降，因为此时模型计算的是整个多维时间序列的主导周期，无法准确计算不同变量序列各自的主导周期，在面对每个变量序列上分布特点不同的异常前兆数据时无法针对性建模特征学习。使用随机周期掩码序列替代主导周期掩码序列后，模型的异常预测性能出现了显著下降，主要原因在于主导周期掩码序列屏蔽了大部分正常数据，有效减少正常数据对稀疏异常前兆数据特征学习的干扰，提高了模型对异常前兆数据的特征学习能力。

表4-5 在公开数据集上的主导周期掩码序列生成模块内部消融实验结果

消融方法	SWaT			MSL			WADI		
	P	R	F1	P	R	F1	P	R	F1
单变量序列划分	83.56	81.34	82.44	76.63	85.26	80.71	79.74	75.38	77.50
主导周期掩码序列	76.12	73.25	74.68	71.34	74.85	73.07	72.39	70.53	71.45
完整MTAP-DM	87.57	84.95	86.24	80.37	88.59	84.28	84.76	82.17	83.44

表4-6 在港口数据集上的主导周期掩码序列生成模块内部消融实验结果

消融方法	港口数据集1			港口数据集2		
	P	R	F1	P	R	F1
单变量序列划分	79.04	84.57	81.71	82.41	85.09	83.73
主导周期掩码序列	80.13	82.35	81.22	78.25	80.37	79.30
完整MTAP-DM	86.85	90.14	88.46	85.57	88.35	86.94

为了验证MTAP-DM的多尺度特征学习模块中部分关键创新方法对模型异常预测性能的影响,本章在所有数据集上均进行了多尺度特征学习模块内部消融实验。这些关键创新方法包括多尺度子序列分段、多尺度编码器,本次消融实验对象分别包括不进行多尺度子序列分段、使用单一尺度编码器替代多尺度编码器后的MTAP-DM模型,实验结果如表4-7和4-8所示。实验结果表明不进行多尺度子序列分段后,MTAP-DM的异常预测性能出现显著下降,因为此时模型无法学习和融合具有不同异常演变趋势和演变时间的异常前兆数据特征,降低了模型在处理具有不同异常类型数据时的异常预测性能和鲁棒性。使用单一尺度编码器替代多尺度编码器后,模型的异常预测性能有所下降,主要原因在于多尺度编码器能够有效处理具有不同异常类型的序列数据,而单一尺度编码器在处理此类数据时性能受限,无法充分捕捉不同异常类型的异常前兆数据特征,降低了模型的异常预测性能和泛化能力。

表4-7 在公开数据集上的多尺度特征学习模块内部消融实验结果

消融方法	SWaT			MSL			WADI		
	P	R	F1	P	R	F1	P	R	F1
多尺度子序列分段	82.16	80.28	81.21	76.49	74.28	75.37	79.32	74.28	76.72
多尺度编码器	85.24	81.71	83.44	78.03	82.57	80.24	80.84	76.37	78.54
完整MTAP-DM	87.57	84.95	86.24	80.37	88.59	84.28	84.76	82.17	83.44

表4-8 在港口数据集上的多尺度特征学习模块内部消融实验结果

消融方法	港口数据集1			港口数据集2		
	P	R	F1	P	R	F1
多尺度子序列分段	80.35	84.57	82.41	81.36	78.71	80.01
多尺度编码器	82.59	87.71	85.07	80.96	82.31	81.63
完整MTAP-DM	86.85	90.14	88.46	85.57	88.35	86.94

为了验证MTAP-DM的异常预测模块中部分关键创新方法对模型异常预测性能的影响,本章在所有数据集上均进行了异常预测模块内部消融实验。这些关键创新方法包括对融合后的多尺度特征进行序列重构、多尺度特征间相似度计算,本次消融实验对象分别包括移除序列重构、移除多尺度特征间相似度计算后的MTAP-DM模型。实验结果如表4-9和4-10所示。

表4-9 在公开数据集上的异常预测模块内部消融实验结果

消融方法	SWaT			MSL			WADI		
	P	R	F1	P	R	F1	P	R	F1
多尺度特征重构	73.29	77.51	75.33	70.82	79.18	74.71	68.31	74.64	71.34
多尺度特征相似度	81.27	75.42	78.23	78.56	83.21	80.81	80.18	76.07	78.11
完整MTAP-DM	87.57	84.95	86.24	80.37	88.59	84.28	84.76	82.17	83.44

表4-10 在港口数据集上的异常预测模块内部消融实验结果

消融方法	港口数据集1			港口数据集2		
	P	R	F1	P	R	F1
多尺度特征重构	81.57	84.04	82.79	80.05	82.41	81.21
多尺度特征相似度	83.54	82.56	83.05	81.74	85.65	83.65
完整MTAP-DM	86.85	90.14	88.46	85.57	88.35	86.94

实验结果表明不对融合后的多尺度特征进行重构误差计算后，MTAP-DM的异常预测性能出现了明显的下降，因为此时模型没有有效评估异常前兆数据的异常波动程度。移除多尺度特征间相似度计算后，MTAP-DM的异常预测性能也有所下降，因为此时模型没有对异常前兆数据的存在性进行有效检测，导致模型最终无法给出的准确、全面的异常预测结果。

4.5.4 参数敏感性实验

为了评估模型关键参数对MTAP-DM异常预测性能的影响，本章设计并进行了参数敏感性实验，重点分析历史时间窗口大小、预测时间窗口大小、分段序列尺度组合这三个参数的变化对模型性能的影响。本实验过程中仅调整上述三个参数，其余参数均保持与4.5.1.2节中的参数设置一致。

图4-4反映了在不同的历史时间窗口大小下，模型在不同的数据集上取得的不同性能，可以看到在MSL数据集、港口数据集1和港口数据集2上，模型在历史时间窗口大小为64时达到最优性能，而在SWaT和WADI数据集上模型分别在历史时间窗口大小为32和256时取得最优性能，这是因为SWaT数据集具有较短的异常演变时间，而WADI数据集则具有较长的异常演变时间，从而导致模型需要更多的历史数据来充分学习异常前兆数据特征。

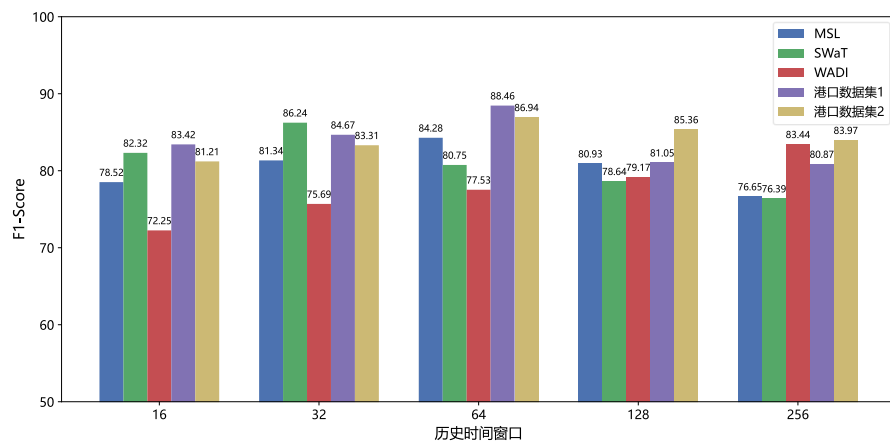


图4-4 不同数据集上历史时间窗口对模型性能的影响

图4-5反映了预测时间窗口大小对模型性能的影响，可以看到随着预测时间窗口增大，模型的预测性能会随之下降，这也符合异常预测的规律，即预测的时间窗口越长，模型预测结果出错的概率也就越大。

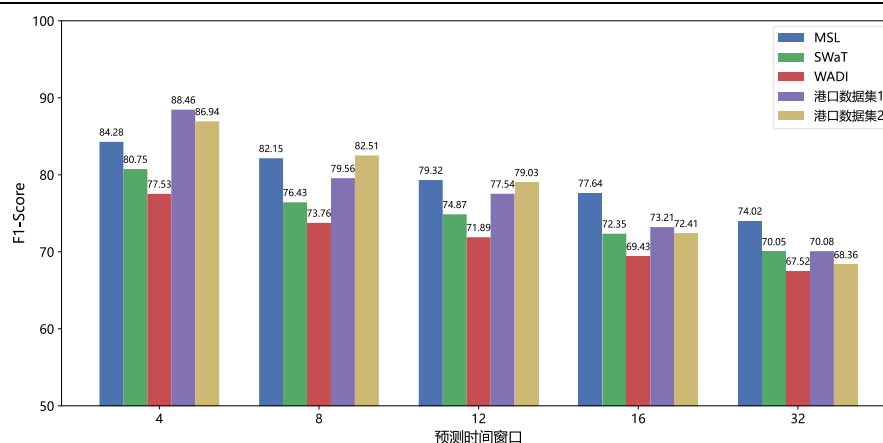


图4-5 不同数据集上预测时间窗口对模型性能的影响

图4-6反映了不同尺度组合对模型异常预测性能的影响，尺度组合会影响模型多尺度特征学习的效果，单一尺度的模型适用于具有特定范围演变时间的数据集，可能不能捕捉到所有演变时间模式，尽管多尺度方法通常能帮助模型捕捉更多的时间特征，但更多的尺度不一定带来更好的结果，如果最小尺度和最大尺度之间的差距过大，会导致模型难以有效学习到不同尺度之间的关系，导致模型预测能力受限。

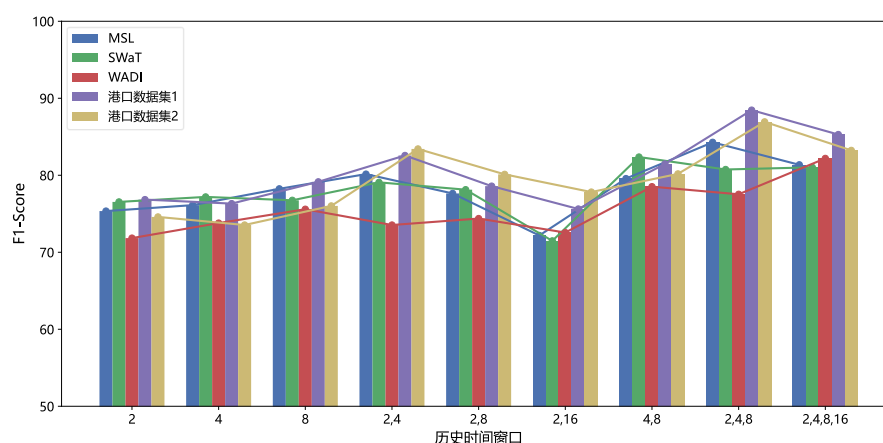


图4-6 不同尺度组合对模型性能的影响

4.5.5 MTAD-TSD和MTAP-DM联合实验

在实际的港口设备运行场景下，特定设备的异常检测任务和异常预测任务往往是需要同时进行的，这涉及到异常检测模型和异常预测模型之间的联合优化问题。为了验证本文提出的异常检测模型MTAD-TSD和异常预测模型MTAP-DM在集成到一个系统后针对同一个多维时间序列同时进行异常检测任务和异常预测任务时的性能，本章在多个数据集上进行了联合测试实验，结果如表4-11所示，表格中的前一项和后一项数值分别表示进行联合检测任务和单项检测任务时的F1-Score，两者中的较高值已用粗线标出。实验结果表明在进行联合测试任务时，MTAD-TSD和MTAP-DM在所有数据集上的F1-Score相比于进行单项异

常检测或预测任务时均有所下降，主要原因在于联合检测系统需要对输入数据进行统一的预处理操作，而这两种模型对时间窗口长度、滑动步长、维度归一化方式等要求不同，导致每个模型都没有得到最优数据输入，同时联合检测系统选择的模型参数如学习率、训练轮数等可能并不是各个模型最优的运行参数。但MTAD-TSD和MTAP-DM在进行联合检测任务时仍然取得了较高的F1-Score，表明联合检测系统具有较好的鲁棒性。

表4-11 MTAD-TSD和MTAP-DM联合实验

模型	SWaT	MSL	WADI	港口数据集1	港口数据集2
MTAD-TSD	83.29/ 87.52	86.37/ 90.02	89.18/ 94.58	84.39/ 89.97	87.47/ 90.88
MTAP-DM	84.31/ 86.24	82.43/ 84.28	81.64/ 83.44	86.21/ 88.46	85.06/ 86.94

4.6 本章小结

本章设计了一种基于主导周期掩码序列和多尺度特征学习相结合的多维时间序列异常预测模型MTAP-DM，以有效应对港口设备时序数据所具有的异常前兆数据稀疏、异常类型多样的特点。MTAP-DM聚焦于多维时间序列异常预测任务中的三个关键挑战：1. 正常数据对稀疏异常前兆数据特征学习的干扰；2. 不同类型异常前兆数据特征的准确学习；3. 异常前兆数据特征与高置信度异常预测结果之间关联关系的建立。针对第一个挑战，MTAP-DM通过频域视角分别对单变量时间序列进行周期性分析，利用快速傅里叶变换提取其主导周期成分，并构造主导周期掩码序列，以此减少正常数据对稀疏异常前兆数据特征学习的干扰；针对第二个挑战，MTAP-DM在主导周期掩码序列的基础上，采用多尺度子序列划分策略，构建不同时间窗口下的子序列视图，通过并行的多尺度特征学习模块对不同尺度的时序模式进行特征提取与融合，进而学习不同异常类型的异常前兆数据特征；针对第三个挑战，MTAP-DM根据异常前兆数据的多尺度特征提出了基于多尺度特征相似度的异常前兆数据检测方法和基于序列重构的异常前兆数据波动程度评估方法，并基于此方法建模异常前兆数据特征与异常预测结果之间的关联关系。

本章在三个公开数据集以及两个港口数据集上对MTAP-DM进行了大量的对比、消融实验和参数敏感性实验，对比实验结果表明MTAP-DM具有优异的异常预测性能，消融实验结果验证了MTAP-DM内部关键模块和机制设计的有效性，参数敏感性实验探讨了部分模型参数对MTAP-DM性能的影响。本章还在这些数据集上进行了MTAD-TSD和MTAP-DM的联合检测实验，实验结果表明了MTAD-TSD和MTAP-DM在进行联合检测任务时仍然能具有较好的性能表现，确保了本文所要设计的港口设备异常分析系统具有可靠的异常检测及预测性能。

第5章 港口设备异常分析系统

为了将本文提出的多维时间序列异常检测模型MTAD-TSD和多维时间序列异常预测模型MTAP-DM实际应用在港口设备异常管理场景下，本章基于这两个模型设计了一个港口设备异常分析系统，支持相关人员根据实际的港口设备异常管理需求选择异常检测或者异常预测任务，以提高港口设备的异常管理效率，减少设备异常带来的损失。同时本章为该系统设计了一个简洁的前端界面以方便相关人员操作。

5.1 系统设计需求

为了充分满足港口设备异常管理和相关操作人员的实际需求，本章设计的系统需要满足以下需求：

（1）模式选择

系统需要提供异常检测和异常预测两种模式来供用户选择，这分别针对港口设备实时监控和未来预警的需求，异常检测模式的设计目标是及时识别和标记已经发生的异常事件，异常预测模式的设计目标是通过历史数据进行建模与分析，预测未来可能出现的异常事件或趋势。通过为用户提供这两种模式的选择，系统能够在多种应用场景下提供精准的异常识别和预警能力，确保设备的健康和港口运营的高效性。

（2）数据集选择

系统需要允许用户加载和选择适用于异常检测与异常预测的多维度设备数据集，用户可以选择系统预先准备好的数据集，也可以通过系统支持的数据上传功能导入新的数据集，系统需要提供数据集管理功能，支持用户对上传的数据集进行查看、清洗、去噪和预处理等操作。

（3）系统参数设置

系统需要为用户提供对系统操作和异常检测模型的细粒度控制，用户可以自定义与模型相关的多个重要参数，如数据集历史时间窗口大小、模型的学习率、迭代次数、优化器等，系统应允许用户根据不同的数据集特性和业务需求灵活调整这些参数，以实现更精确的异常检测与预测结果。

（4）异常检测结果可视化

系统需要以直观和易于理解的方式展示异常检测与预测的结果，帮助用户快速识别潜在的设备故障或异常模式，包括对异常点进行统计、展示模型运行

的性能指标，如准确率、召回率、F1-Score等。

5.2 系统设计关键技术

为了满足系统设计的需求，本章在设计系统时主要使用了以下关键技术：

（1）PyTorch

PyTorch是一个开源的深度学习框架，其提供了强大的自动微分功能、灵活的神经网络构建能力以及高效的计算性能，是一个极具灵活性与扩展性的深度学习工具。在本系统中，PyTorch被用来搭建多维时间序列异常检测模型MTAD-TSD和多维时间序列异常预测模型MTAP-DM。

（2）Pandas

Pandas是一个基于Python的开源数据分析库，其提供了高效的数据读写功能，支持多种格式数据（如 CSV、Excel、SQL等）的读写操作，被广泛应用于数据处理、清洗、分析和可视化任务。在本系统中，Pandas被用来对模型运行数据进行加载、清洗、预处理等操作。

（3）MySQL

MySQL是一个开源的关系型数据库管理系统，其采用结构化查询语言来执行数据库操作，适用于各种规模的数据存储和管理。在本系统中，MySQL被用来构建数据库，存储和管理模型运行使用的数据集以及运行结果数据等。

（4）Node.js

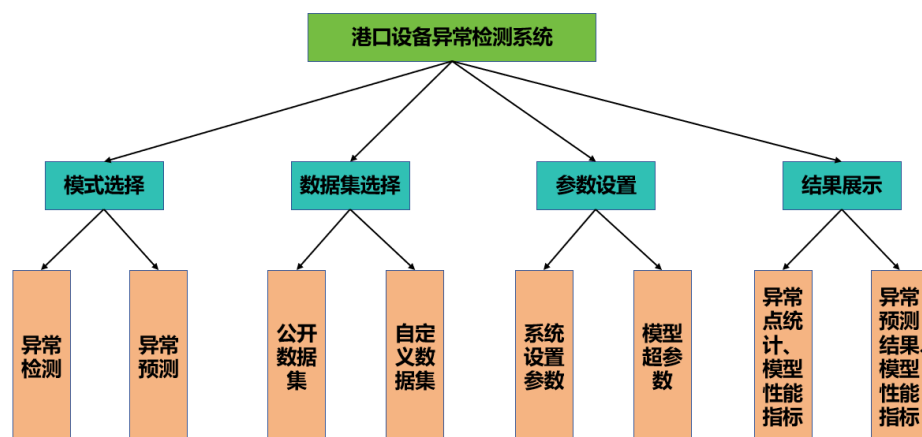
Node.js是一个开源、跨平台的JavaScript运行时环境，旨在帮助开发者利用JavaScript语言构建高性能、可扩展的网络应用。Node.js的核心特性包括其单线程模型和事件循环机制，这使得它能够在不创建多个线程的情况下，处理大量并发连接，其生态系统内提供了丰富的开源模块，可以通过npm（Node.js的包管理工具）轻松管理和使用各种功能扩展，Node.js的高性能和低延迟，使得它在构建现代Web应用、微服务架构和实时数据处理平台中得到了广泛应用。本系统使用Node.js搭建后端框架。

（5）React

React是一个用于构建用户界面的开源JavaScript库，其核心特点是组件化和虚拟DOM，它允许开发者通过创建可重用的组件来构建复杂的用户界面，其采用声明式编程风格，开发者通过描述界面的期望状态来告知React如何渲染UI，它还支持单向数据流，数据父组件传递到子组件，这有助于保持应用的可预测性和简洁性，从而提高开发效率和应用的可维护性。在本系统中，React被用来搭建系统前端界面。

5.3 系统结构与展示

本章设计的港口设备异常检测系统整体结构如图5-1所示，其主要有以下模块构成：



（1）模式选择模块

模式选择模块旨在为用户提供灵活的异常分析策略配置功能，允许用户根据具体应用场景和实际需求，自主选择适用的异常检测模式。系统当前支持两种主要模式：异常检测模式与异常预测模式。在异常检测模式中，系统集成了MTAD-TSD模型，该模型主要针对已发生的异常事件进行识别与定位。其核心思想是通过学习历史数据中的正常行为模式，将当前观测序列与之进行对比，从而判断是否出现异常。相较之下，异常预测模式内置的是MTAP-DM模型，其重点在于利用历史时间序列数据挖掘潜在的趋势变化，以此预测未来一段时间内设备的运行状态，该模式支持对潜在异常进行提前预警，从而为用户提供更具前瞻性的决策依据。通过该模块的模式选择机制，用户可根据任务目标和时效要求，灵活切换于面向当前状态的异常检测与面向未来趋势的异常预测之间，从而实现更精准、更高效的异常分析流程。



（2）数据集选择模块

数据集选择模块旨在为异常检测与预测任务提供灵活的数据输入与管理机制，支持用户根据具体应用需求选择或导入多维时间序列数据。该模块不仅支

持系统内置数据集的调用，还设计了自定义数据集的上传与管理功能，确保模型在多样化数据源下的可用性与可扩展性。在用户完成数据集选择或上传操作后，系统将自动对数据进行预处理检查，包括数据格式的统一、时间顺序的校验、缺失值的检测与填补等，以确保输入数据满足模型后续运行的结构性与完整性要求。目前系统提供了MSL、SWaT、WADI数据集作为默认数据集，用户也可以通过自定义数据集上传功能将新数据导入系统进行处理和分析。此模块不仅提升了系统在不同工业场景下的适应能力，也为用户进行异常检测与异常预测任务提供了高效、可控的数据基础保障。



图5-3 数据集选择模块

(3) 参数设置模块

参数设置模块为系统提供了高度可配置的模型运行控制接口，支持用户根据具体的异常检测或异常预测任务灵活设定相关参数。该模块涵盖了模型训练与推理过程中的关键超参数设置，包括历史数据窗口大小、模型训练轮次数、模型训练数据batch大小、学习率、模型嵌入层维度等，旨在提升系统在多种数据特性与应用场景下的适应性与性能表现。此外参数设置模块支持模式感知型的参数配置方案，即在不同任务模式下动态呈现相应的参数设置选项，增强了系统的交互性与智能化水平。

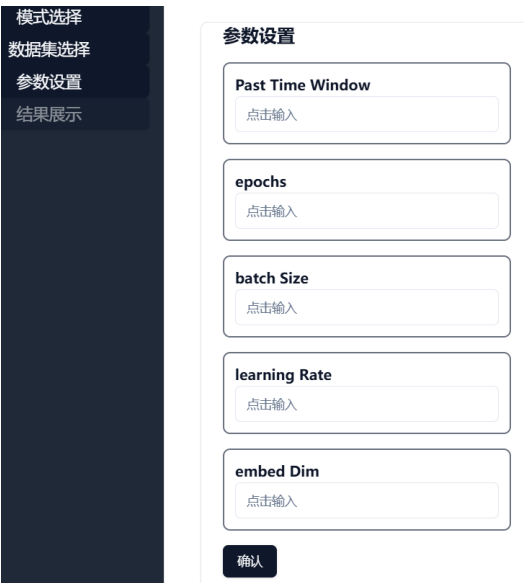


图5-4 参数设置模块

(4) 结果展示模块

结果展示模块旨在以直观、清晰的方式呈现系统在异常检测与异常预测任务中得到的分析结果，帮助用户全面理解设备运行状态，快速识别潜在故障或异常模式。该模块通过图形化手段对模型输出进行可视化展示，包括对异常点进行统计、给出异常预测结果、展示模型运行的性能指标，如准确率、召回率、F1-Score等，有助于增强系统的可解释性与用户操作的便捷性。通过结果展示模块，用户不仅能够直观掌握系统的异常分析结果，还可对模型的有效性进行综合评估，从而为后续的运维决策与风险管理提供可靠依据。

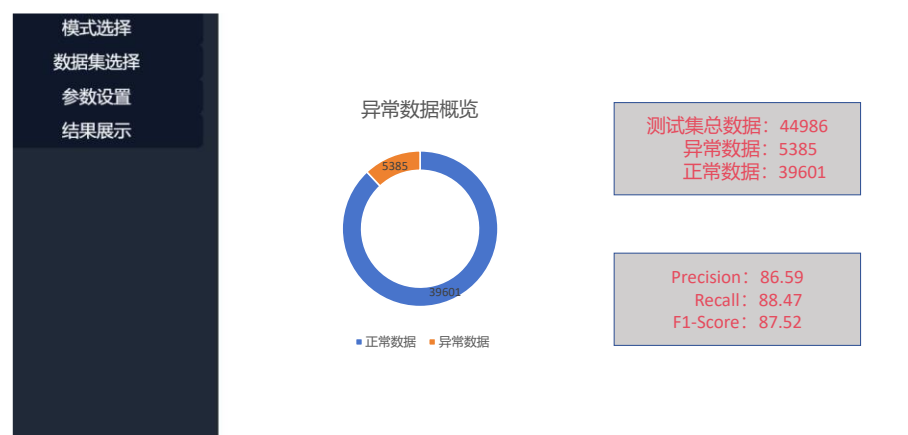


图5-5 结果展示模块

5.4 本章小结

本章介绍了基于第3章设计的多维时间序列异常检测模型MTAD-TSD和第4章设计的多维时间序列异常预测模型MTAP-DM设计的港口设备异常分析系统。本章首先介绍了该系统的设计需求，包括模式选择、数据集选择、系统参数设置、结果可视化展示等需求；接着介绍了该系统设计过程中所使用到的关键技术，包括PyTorch、Pandas、MySQL、React、Node.js等技术介绍；最后介绍了本系统的整体结构设计，主要包括模式选择模块、数据集选择模块、参数设置模块、结果展示模块的设计和使用方法。本系统的设计与开发将本文第3章设计的MTAD-TSD模型和第4章设计的MTAP-DM模型落地使用，将本文的理论创新应用于实际，具有很强的实际应用意义。

第6章 总结与展望

6.1 本文工作总结

随着港口物流规模的不断扩大与设备系统结构的日益复杂，港口设备的异常管理已成为保障港口运行安全、提升运行效率的重要环节。通过对港口设备生成的多维时间序列数据进行有效的异常分析，可以显著提高设备异常管理的效率。时间序列异常分析主要包括异常检测与异常预测两项任务，近年来，随着深度学习尤其是大模型技术的发展，Transformer和图神经网络等方法被广泛应用于相关研究中。然而港口设备因其复杂的运行工况、作业模式、操作行为使其所产生的时序数据具有高维异质、变量动态相关、局部变化模式复杂、异常前兆数据稀疏等特点，这给现有方法在港口设备异常分析中的应用带来了多方面挑战。本文聚焦于港口设备的异常检测与异常预测任务展开研究工作，具体做出了以下贡献和创新：

（1）在港口设备异常检测任务中，现有方法存在以下问题：1. 时空特征建模割裂，缺乏时间与空间特征的协同学习方法；2. 基于Transformer的方法对变量在时间维度的局部特征与异质性特征提取不足；3. 基于图神经网络的方法对变量在空间维度的动态相关性建模不充分。针对上述问题，本文设计了一种基于时空双维特征协同学习的多维时间序列异常检测模型MTAD-TSD，其核心设计包括：1. 对多维时间序列进行时间和空间特征的联合建模学习；2. 对Transformer的点积注意力进行改进并提出了单变量注意力机制UA，实现对变量异质性特征的有效学习，同时基于单变量注意力机制设计了时间特征学习网络Uaformer，实现对变量局部特征的充分建模；3. 设计了一种基于局部图结构自适应性构建的图结构动态学习方法，实现对变量间关联关系的动态构建。

（2）在港口设备异常预测任务中，现有方法存在以下问题：1. 难以从大量正常数据中准确学习稀疏异常前兆数据特征；2. 无法充分学习不同异常类型的异常前兆数据特征；3. 缺乏异常前兆数据特征与高置信度异常预测结果之间的关联机制。针对上述问题，本文设计了一种基于主导周期掩码与多尺度特征学习的多维时间序列异常预测模型MTAP-DM，其核心设计包括：1. 从频域视角对时间序列进行分析并生成主导周期掩码序列，减少正常数据对稀疏异常前兆数据特征学习的干扰；2. 对主导周期掩码序列进行多尺度特征学习及融合，充分建模多类型异常前兆数据特征；3. 基于多尺度特征相似度与序列重构方法，构建异常前兆数据特征与异常预测结果之间高效的关联机制。

(3) 本文在MSL、SWaT、WADI三个公开数据集以及两个港口设备运行数据集上对MTAD-TSD和MTAP-DM进行了大量的对比实验、消融实验、参数敏感性实验。对比实验结果表明本文设计的两个模型分别在多维时间序列异常检测和异常预测任务中具有优秀性能；消融实验结果验证了两个模型内部各个关键模块和方法设计的有效性；参数敏感性实验探讨了不同实验参数对模型性能的影响。

(4) 本文基于MTAD-TSD和MTAP-DM模型设计了港口设备异常分析系统，该系统能够实现模式选择、数据集选择、参数设置、结果展示等功能，可用于进行多维时间序列异常检测与异常预测任务，将本文的相关研究理论应用到实际工作中。

6.2 本文工作展望

尽管本文在面向港口设备的多维时间序列异常检测与异常预测研究方面取得了显著的进展，但仍然存在以下改进空间：

(1) 现有模型的计算复杂度较高，特别是在处理大规模数据集时，模型训练和推理的效率仍需进一步优化，未来的研究可以探索轻量化模型设计，或采用分布式计算和云计算等技术来提升模型对大规模数据的处理能力；

(2) 尽管模型在多个数据集上均取得了较好的异常检测与异常预测效果，但在实际应用场景中，港口设备的运行工况、作业模式、操作行为复杂，未来可以增加对模型适应性与泛化能力的研究工作，进一步提升其在港口复杂工况环境下的实际应用效果；

(3) 目前模型主要是基于港口设备时序数据进行异常检测及异常预测，未来工作中可以通过结合更为丰富的多源数据如视频监控数据、人工智能辅助决策数据等来为模型提供更多的上下文信息，从而进一步提高模型的准确性和鲁棒性。

参考文献

- [1] Wen Q, Zhang Z, Li Y, et al. Fast RobustSTL: Efficient and robust seasonal-trend decomposition for time series with complex patterns[C]//Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. 2020: 2203-2213.
- [2] Chen X, Ji J, Loparo K, et al. Real-time personalized cardiac arrhythmia detection and diagnosis: A cloud computing architecture[C]//2017 IEEE EMBS international conference on biomedical & health informatics (BHI). IEEE, 2017: 201-204.
- [3] 王婷, 王娜, 崔运鹏, 等. 基于半监督学习的无线网络攻击行为检测优化方法[J]. 计算机研究与发展, 2020, 57(4): 791-802.
- [4] 丁小欧, 于晟健, 王沐贤, 等. 基于相关性分析的工业时序数据异常检测[J]. 软件学报, 2020, 31(3): 726-747.
- [5] Liu P, Sun X, Han Y, et al. Arrhythmia classification of LSTM autoencoder based on time series anomaly detection[J]. Biomedical Signal Processing and Control, 2022, 71: 103228
- [6] Dornadula V N, Geetha S. Credit card fraud detection using machine learning algorithms[J]. Procedia computer science, 2019, 165: 631-641.
- [7] Zhou X, Hu Y, Liang W, et al. Variational LSTM enhanced anomaly detection for industrial big data[J]. IEEE Transactions on Industrial Informatics, 2020, 17(5): 3469-3477.
- [8] DOU S, ZHANG G, XIONG Z. Anomaly detection of process unit based on LSTM time series reconstruction[J]. CIESC Journal, 2019, 70(2): 481.
- [9] Tayeh T, Aburakhia S, Myers R, et al. An attention-based ConvLSTM autoencoder with dynamic thresholding for unsupervised anomaly detection in multivariate time series[J]. Machine Learning and Knowledge Extraction,

2022, 4(2): 350-370.

- [10] Pham V, Nguyen N, Li J, et al. Mtsad: Multivariate time series abnormality detection and visualization[C]//2019 IEEE international conference on big data (Big Data). IEEE, 2019: 3267-3276.
- [11] 胡珉, 白雪, 徐伟, 等. 多维时间序列异常检测算法综述[J]. 计算机应用, 2020, 40(6): 1553-1564.
- [12] 卓琳, 赵厚宇, 詹思延. 异常检测方法及其应用综述[J]. 计算机应用研究, 2020, 37(S1): 9-15.
- [13] 杨雅辉. 网络流量异常检测及分析的研究[J]. 计算机科学, 2008, 35(5): 108-112.
- [14] Qiangsheng H, Jiujun C, Qinma K. Network Anomaly Detection Based on Highorder AR Model[J]. Computer Engineering, 2010: 180-182.
- [15] Patel R, Saha G. Time Series Regression of Weather Parameters Over the Last Century for Cotton Crop[C]. Proceedings of the International Conference on Intelligent Systems and Signal Processing, 2018: 125-153.
- [16] Wang Y, Huang K, Tan T. Group activity recognition based on ARMA shape sequence modeling[C]//2007 IEEE International Conference on Image Processing. IEEE, 2007, 3: III-209-III-212.
- [17] Kadri F, Harrou F, Chaabane S, et al. Seasonal ARMA-based SPC charts for anomaly detection: Application to emergency department systems[J]. Neurocomputing, 2016, 173: 2102-2114.
- [18] Bayati A, Nguyen K K, Cheriet M. Multiple-step-ahead traffic prediction in high-speed networks[J]. IEEE Communications Letters, 2018, 22(12): 2447-2450.
- [19] Hardin J, Rocke D M. Outlier detection in the multiple cluster setting using the minimum covariance determinant estimator[J]. Computational Statistics & Data Analysis, 2004, 44(4): 625-638.
- [20] Yu Y, Zhu Y, Li S, et al. Time series outlier detection based on sliding window prediction[J]. Mathematical problems in Engineering, 2014, 2014(1): 879736.

- [21] 薛坤. 面向数据中心智能运维的KPI预测及异常检测模型研究[D]. 东华大学, 2023. DOI:10.27012/d.cnki.gdhuu.2023.000685.
- [22] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey[J]. *ACM computing surveys (CSUR)*, 2009, 41(3): 1-58.
- [23] Zhang X, Gu C, Lin J. Support vector machines for anomaly detection[C]//2006 6th World Congress on Intelligent Control and Automation. IEEE, 2006, 1: 2594-2598.
- [24] Schölkopf B, Platt J C, Shawe-Taylor J, et al. Estimating the support of a high-dimensional distribution[J]. *Neural computation*, 2001, 13(7): 1443-1471.
- [25] Primartha R, Tama B A. Anomaly detection using random forest: A performance revisited[C]//2017 International conference on data and software engineering (ICoDSE). IEEE, 2017: 1-6.
- [26] Xu D, Wang Y, Meng Y, et al. An improved data anomaly detection method based on isolation forest[C]//2017 10th international symposium on computational intelligence and design (ISCID). IEEE, 2017, 2: 287-291.
- [27] Chen J, Ouyang J Y, Feng A Q. DoS anomaly detection based on isolation forest algorithm under edge computing framework[J]. *Computer Science*, 2020, 47(2): 293-299.
- [28] Liu F T, Ting K M, Zhou Z H. Isolation-based anomaly detection[J]. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2012, 6(1): 1-39.
- [29] Mao W, Cao X, Yan T, et al. Anomaly detection for power consumption data based on isolated forest[C]//2018 international conference on power system technology (POWERCON). IEEE, 2018: 4169-4174.
- [30] Byers S, Raftery A E. Nearest-neighbor clutter removal for estimating features in spatial point processes[J]. *Journal of the American Statistical Association*, 1998, 93(442): 577-584.
- [31] Jain A K. Data clustering: 50 years beyond K-means[J]. *Pattern recognition letters*, 2010, 31(8): 651-666.
- [32] Ester M, Kriegel H P, Sander J, et al. A density-based algorithm for

-
- discovering clusters in large spatial databases with noise[C]//kdd. 1996, 96(34): 226-231.
- [33] Lee Y J, Yeh Y R, Wang Y C F. Anomaly detection via online oversampling principal component analysis[J]. IEEE transactions on knowledge and data engineering, 2012, 25(7): 1460-1470.
 - [34] Chalapathy R, Chawla S. Deep learning for anomaly detection: A survey[J]. arXiv preprint arXiv:1901.03407, 2019.
 - [35] Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks[C]//Proceedings of the 30th International Conference on International Conference on Machine Learning-Volume 28. 2013: III-1310-III-1318.
 - [36] Computation N. Long short-term memory[J]. Neural Comput, 2016, 9: 1735-1780.
 - [37] Chung J, Gulcehre C, Cho K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[C]//NIPS 2014 Workshop on Deep Learning, December 2014. 2014.
 - [38] Shen L, Li Z, Kwok J. Timeseries anomaly detection using temporal hierarchical one-class network[J]. Advances in neural information processing systems, 2020, 33: 13016-13026.
 - [39] Tax D M J, Duin R P W. Support vector data description[J]. Machine learning, 2004, 54: 45-66.
 - [40] Smyl S. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting[J]. International journal of forecasting, 2020, 36(1): 75-85.
 - [41] Hundman K, Constantinou V, Laporte C, et al. Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding[C]//Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. 2018: 387-395.
 - [42] Park D, Hoshi Y, Kemp C C. A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder[J]. IEEE Robotics and Automation Letters, 2018, 3(3): 1544-1551.

-
- [43] K Kingma D P, Welling M. Auto-encoding variational Bayes[C]//Proceedings of the 2nd International Conference on Learning Representations. 2014: 1-14.
- [44] Su Y, Zhao Y, Niu C, et al. Robust anomaly detection for multivariate time series through stochastic recurrent neural network[C]//Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019: 2828-2837.
- [45] Darban Z Z, Webb G I, Pan S, et al. CARLA: Self-supervised contrastive representation learning for time series anomaly detection[J]. Pattern Recognition, 2025, 157: 110874.
- [46] Darban Z Z, Yang Y, Webb G I, et al. DACAD: Domain adaptation contrastive learning for anomaly detection in multivariate time series[J]. arXiv preprint arXiv:2404.11269, 2024.
- [47] Bai S, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. arXiv preprint arXiv:1803.01271, 2018.
- [48] Pol A A, Azzolini V, Cerminara G, et al. Anomaly detection using Deep Autoencoders for the assessment of the quality of the data acquired by the CMS experiment[C]//EPJ Web of Conferences. EDP Sciences, 2019, 214: 06008.
- [49] Zhang Y, Wang J, Chen Y, et al. Adaptive memory networks with self-supervised learning for unsupervised anomaly detection[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(12): 12068-12080.
- [50] Lai K H, Wang L, Chen H, et al. Context-aware domain adaptation for time series anomaly detection[C]//Proceedings of the 2023 siam international conference on data mining (sdm). Society for Industrial and Applied Mathematics, 2023: 676-684.
- [51] Li Z, Zhao Y, Han J, et al. Multivariate time series anomaly detection and interpretation using hierarchical inter-metric and temporal embedding[C]//Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining. 2021: 3220-3230.
- [52] Li L, Yan J, Wen Q, et al. Learning robust deep state space for unsupervised anomaly detection in contaminated time-series[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(6): 6058-6072.

- [53] Hudson D A, Zitnick L. Generative adversarial transformers[C]//International conference on machine learning. PMLR, 2021: 4487-4499.
- [54] Schlegl T, Seeböck P, Waldstein S M, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery[C]//International conference on information processing in medical imaging. Cham: Springer International Publishing, 2017: 146-157.
- [55] Bowen Du, Xuanxuan Sun, Junchen Ye, Ke Cheng, Jingyuan Wang, and Leilei Sun. Gan-based anomaly detection for multivariate time series using polluted training set. IEEE Transactions on Knowledge and Data Engineering, 35(12):12208–12219, 2021.
- [56] Yang Y, Zhang C, Zhou T, et al. Dcdetector: Dual attention contrastive representation learning for time series anomaly detection[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023: 3033-3045.
- [57] Vaswani A. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017.
- [58] Tuli S, Casale G, Jennings N R. TranAD: deep transformer networks for anomaly detection in multivariate time series data[J]. Proceedings of the VLDB Endowment, 2022, 15(6): 1201-1214.
- [59] Xu J, Wu H, Wang J, et al. Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy[C]// Proceedings of the 10th International Conference on Learning Representations. 2022: 1-20.
- [60] Nam Y, Yoon S, Shin Y, et al. Breaking the time-frequency granularity discrepancy in time-series anomaly detection[C]//Proceedings of the ACM Web Conference 2024. 2024: 4204-4215.
- [61] Yue Z, Wang Y, Duan J, et al. Ts2vec: Towards universal representation of time series[C]//Proceedings of the AAAI conference on artificial intelligence. 2022, 36(8): 8980-8987.
- [62] Wang X, Pi D, Zhang X, et al. Variational transformer-based anomaly detection approach for multivariate time series[J]. Measurement, 2022, 191: 110791.

-
- [63] Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model[J]. IEEE transactions on neural networks, 2008, 20(1): 61-80.
- [64] Deng A, Hooi B. Graph neural network-based anomaly detection in multivariate time series[C]//Proceedings of the AAAI conference on artificial intelligence. 2021, 35(5): 4027-4035.
- [65] Han S, Woo S S. Learning sparse latent graph representations for anomaly detection in multivariate time series[C]//Proceedings of the 28th ACM SIGKDD Conference on knowledge discovery and data mining. 2022: 2977-2986.
- [66] Zhan J, Wang S, Ma X, et al. Stgat-mad: Spatial-temporal graph attention network for multivariate time series anomaly detection[C]//ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022: 3568-3572.
- [67] Boniol P, Palpanas T. Series2Graph: graph-based subsequence anomaly detection for time series[J]. Proceedings of the VLDB Endowment, 2020, 13(12): 1821-1834.
- [68] Dai E, Chen J. Graph-Augmented Normalizing Flows for Anomaly Detection of Multiple Time Series[C]// Proceedings of the 10th International Conference on Learning Representations. 2022: 1-16.
- [69] Doshi K, Yilmaz Y. Towards interpretable video anomaly detection[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2023: 2655-2664.
- [70] Ho T K K, Armanfard N. Self-supervised learning for anomalous channel detection in EEG graphs: Application to seizure analysis[C]//Proceedings of the AAAI conference on artificial intelligence. 2023, 37(7): 7866-7874.
- [71] Wang L, Zeng L, Li J. AEC-GAN: adversarial error correction GANs for autoregressive long time-series generation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2023, 37(8): 10140-10148.
- [72] Ferchichi A, Abbes A B, Barra V, et al. Multi-attention generative adversarial network for multi-step vegetation indices forecasting using multivariate time series[J]. Engineering Applications of Artificial Intelligence, 2024, 128: 107563.
- [73] Zhou H, Zhang S, Peng J, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting[C]//Proceedings of the AAAI

- hr/>
- conference on artificial intelligence. 2021, 35(12): 11106-11115.
- [74] Zhou T, Ma Z, Wen Q, et al. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting[C]//International conference on machine learning. PMLR, 2022: 27268-27286.
- [75] Liu Y, Hu T, Zhang H, et al. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting[C]// Proceedings of the 12th International Conference on Learning Representations. 2024: 1-25.
- [76] Guo N, Liu C, Li C, et al. Explainable and effective process remaining time prediction using feature-informed cascade prediction model[J]. IEEE Transactions on Services Computing, 2024, 17(3): 949-962.
- [77] Wu H, Hu T, Liu Y, et al. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis[C]// Proceedings of The 11th International Conference on Learning Representations. 2023: 1-23.
- [78] Zhu H, Yuan S, Liu X, et al. CasCIFF: A cross-domain information fusion framework tailored for cascade prediction in social networks[J]. Knowledge-Based Systems, 2024, 303: 112391.
- [79] Shi J, Shirali A, Narasimhan G. Boosting Time Series Prediction of Extreme Events by Reweighting and Fine-tuning[C]//2024 IEEE International Conference on Big Data (BigData). IEEE, 2024: 1450-1457.
- [80] Mecheri H, Benamirouche I, Fass F, et al. Prediction of rare events in the operation of household equipment using co-evolving time series[J]. Pattern Analysis and Applications, 2024, 27(3): 101.
- [81] Xiao T, Xu Z, He W, et al. XTSFormer: Cross-Temporal-Scale Transformer for Irregular-Time Event Prediction in Clinical Applications[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2025, 39(27): 28502-28510.
- [82] Wang J, Gao Y. Generalized mixture model for extreme events forecasting in time series data[C]//2023 China Automation Congress (CAC). IEEE, 2023: 7817-7822.
- [83] Chen S, Kalanat N, Topp S, et al. Meta-transfer-learning for time series data with extreme events: An application to water temperature prediction[C]//Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. 2023: 266-275.

-
- [84] Bhardwaj A, Singh S, Sam L, et al. A review on remotely sensed land surface temperature anomaly as an earthquake precursor[J]. International journal of applied earth observation and geoinformation, 2017, 63: 158-166.
- [85] Ghosh D, Deb A, Sengupta R. Anomalous radon emission as precursor of earthquake[J]. Journal of Applied Geophysics, 2009, 69(2): 67-81.
- [86] Lin Y, Chiang Y Y. A semi-supervised learning approach for abnormal event prediction on large network operation time-series data[C]//2022 IEEE International Conference on Big Data (Big Data). IEEE, 2022: 1024-1033.
- [87] Jhin S Y, Lee J, Park N. Precursor-of-anomaly detection for irregular time series[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023: 917-929.
- [88] You J, Cela A, Natowicz R, et al. Anomaly Prediction: A Novel Approach with Explicit Delay and Horizon[C]//2024 IEEE 20th International Conference on Intelligent Computer Communication and Processing (ICCP). IEEE, 2024: 1-8.
- [89] Zhao S, Wang W, Xu H, et al. Abnormality Forecasting: Time Series Anomaly Prediction via Future Context Modeling[J]. arXiv preprint arXiv:2410.12206, 2024.
- [90] Aminikhanghahi S, Cook D J. A survey of methods for time series change point detection[J]. Knowledge and information systems, 2017, 51(2): 339-367.
- [91] Hawkins D. Identification of outliers[J]. 1980.
- [92] Choi K, Yi J, Park C, et al. Deep learning for anomaly detection in time-series data: Review, analysis, and guidelines[J]. IEEE access, 2021, 9: 120043-120065.
- [93] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.
- [94] Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines[C]//Proceedings of the 27th international conference on machine learning (ICML-10). 2010: 807-814.
- [95] Cybenko G. Approximation by superpositions of a sigmoidal function[J].

Mathematics of control, signals and systems, 1989, 2(4): 303-314.

- [96] Veličković P, Cucurull G, Casanova A, et al. Graph Attention Networks[C]//Proceedings of the 6th International Conference on Learning Representations. 2018: 1-12.
- [97] Wang Y, Sun Y, Liu Z, et al. Dynamic graph cnn for learning on point clouds[J]. ACM Transactions on Graphics (tog), 2019, 38(5): 1-12.
- [98] Zhang H, Li P, Zhang R, et al. Embedding graph auto-encoder for graph clustering[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 34(11): 9352-9362.
- [99] Pan S, Hu R, Long G, et al. Adversarially regularized graph autoencoder for graph embedding[C]//International Joint Conference on Artificial Intelligence 2018. Association for the Advancement of Artificial Intelligence (AAAI), 2018: 2609-2615.
- [100] Kazemi S M, Goel R, Eghbali S, et al. Time2vec: Learning a vector representation of time[J]. arXiv preprint arXiv:1907.05321, 2019.
- [101] Xu H, Chen W, Zhao N, et al. Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications[C]//Proceedings of the 2018 world wide web conference. 2018: 187-196.
- [102] Hundman K, Constantinou V, Laporte C, et al. Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding[C]//Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. 2018: 387-395.
- [103] Mathur A P, Tippenhauer N O. SWaT: A water treatment testbed for research and training on ICS security[C]//2016 international workshop on cyber-physical systems for smart water networks (CySWater). IEEE, 2016: 31-36.
- [104] Ahmed C M, Palleti V R, Mathur A P. WADI: a water distribution testbed for research in the design of secure cyber physical systems[C]//Proceedings of the 3rd international workshop on cyber-physical systems for smart water networks. 2017: 25-28.
- [105] Huang X, Chen W, Hu B, et al. Graph mixture of experts and memory-augmented routers for multivariate time series anomaly detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence.

2025, 39(16): 17476-17484.

- [106] Luo D, Wang X. Moderntcn: A modern pure convolution structure for general time series analysis[C]// Proceedings of the 12th International Conference on Learning Representations. 2024: 1-43.
- [107] Wang C, Zhuang Z, Qi Q, et al. Drift doesn't matter: dynamic decomposition with diffusion reconstruction for unstable multivariate time series anomaly detection[J]. Advances in Neural Information Processing Systems, 2024, 36.
- [108] Campos D, Kieu T, Guo C, et al. Unsupervised Time Series Outlier Detection with Diversity-Driven Convolutional Ensembles[J]. Proceedings of the VLDB Endowment, 2021, 15(3): 611-623.