

DOI: 10.13382/j.jemi.B2508313

# 多视角对偶图注意力网络的软测量方法\*

郭中宇 岳玉麒 陈娟

(北京化工大学信息科学与技术学院 北京 100029)

**摘要:**在复杂的过程工业中,由于缺乏在线检测仪表或者由于恶劣的生产环境,导致一些生产过程中的关键变量难以测量,且无法实现在线测量,因此在过程工业中需要对这些变量进行软测量建模的研究。当前,基于深度学习的软测量建模大多关注的是单一视角下的特征建模,从而丢失了一些其他视角下有价值的信息,导致预测模型精度低。针对该问题,提出了一种多视角对偶图注意力网络(multi-view dual graph attention, Mv-DGAT)的工业软测量建模方法。该方法首先搭建多视角框架,构建基于最大信息系数的空间图注意力网络(spatial graph attention, SGAT),完成空间视角,并搭建基于多层次时序图结构和长短时记忆网络(long short-term memory, LSTM)的时间图注意力网络(temporal graph attention, TGAT),建立时间视角。其次,使用多头注意力机制进行时空特征融合预测。最后引入余弦相似度评估视角间的互补性,抑制冗余特征。所提出的方法在真实工业流程公开数据集上进行了实验,实验结果表明,所提出方法预测精度高,决定系数 $R^2$ 均达到0.85,均方根误差较对比模型降低10%以上。

**关键词:**软测量;多视角学习;图注意力网络;余弦相似度

**中图分类号:** TP274 **文献标识码:** A **国家标准学科分类代码:** 520.2060

## Soft sensing method based on multi view dual graph attention network

Guo Zhongyu Yue Yuqi Chen Juan

(College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China)

**Abstract:** In the complex process industry, due to the lack of online detection instruments or harsh production environments, some key variables are difficult to measure and cannot be measured online. Therefore, research on soft sensing modeling of these variables is needed in process industries. Currently, deep learning based soft sensing modeling mostly focuses on feature modeling from a single perspective, neglecting valuable information from other perspectives, resulting in low accuracy of the prediction model. To address this issue, this paper proposes an industrial soft sensing modeling method based on a multi-view dual graph attention (Mv-DGAT) network. This method first constructs a multi-view framework, builds a spatial graph attention (SGAT) network based on the maximal information coefficient to complete the spatial perspective, and constructs a temporal graph attention (TGAT) network based on a multi-level temporal graph structure and the long short-term memory (LSTM) network to establish the temporal perspective. Secondly, the multi-head attention mechanism is used for spatiotemporal feature fusion prediction. Finally, the cosine similarity is introduced to evaluate the complementarity between views and suppress redundant features. The proposed method was tested on a publicly available dataset of real industrial processes. The experimental results showed that the proposed method has high prediction accuracy, with determination coefficients  $R^2$  reaching 0.85 and root mean square error reduced by more than 10% compared to the comparison model.

**Keywords:** soft sensing; multi view learning; graph attention network; cosine similarity

## 0 引言

在工业生产过程中,对核心环节的关键变量实施在线测量与监控是确保产品质量以及生产流程安全推进的必要举措<sup>[1]</sup>。然而,一方面,一些被测变量没有在线检测仪表,另一方面,由于工业生产环境较为恶劣以及生产流程较为复杂,部分被测量无法直接测量。因此,针对这些变量采用软测量技术,借助过程中易于测量的辅助变量来间接计算难以测量的质量变量。

软测量建模主要分为两大方法,即基于过程机理和基于数据的方法。传统软测量技术多为基于过程机理建模<sup>[2]</sup>。该方法要求掌握领域理论知识,构建辅助变量与关键质量变量的数学表达式,实现对每时刻质量变量数值的精准计算。然而,在复杂过程工业中,对象的机理模型很难清晰完整地建立,难以实现基于过程机理的软测量建模。

随着工业逐渐向智能化和信息化的方向发展,在现代工业过程中分布式控制系统被广泛应用,因此可以获得大量的多批次间歇生产过程数据,为基于数据驱动的软测量建模提供了有效的数据支持<sup>[3]</sup>。在与被测关键变量关联较明确的数据上,多采用机器学习算法,包括支持向量回归<sup>[4]</sup>、高斯回归<sup>[5]</sup>等,用于捕捉辅助变量和关键质量变量间的非线性关系。Shi 等<sup>[6]</sup>构建基于变异贝叶斯推理的极限学习机(extreme learning machine, ELM),实验证明所提出的方法效果有所提升。全卫国等<sup>[7]</sup>提出改进麻雀搜索算法优化的最小二乘支持向量机(least square support vector machine, LSSVM)软测量模型,仿真实验显示该模型预测精度显著提升。徐雯等<sup>[8]</sup>提出半监督随机森林回归(random forest regression, RFR)的软测量模型,用采集的现场数据训练模型及预测,其效果稳定性良好。然而,工业过程中的数据往往时空特征复杂以及数据模态多元,机器学习难以准确表征过程的复杂特征,因此深度学习被广泛应用于软测量领域<sup>[9]</sup>。深度学习方法多围绕变量自身的时序特征和多元变量间的空间特征进行建模,以提高模型的推理能力。周建新等<sup>[10]</sup>提出基于改进鸚鵡优化算法与长短记忆神经网络(long short-term memory, LSTM)相结合的软预测模型,深度挖掘变量内部时序信息。Wu 等<sup>[11]</sup>提出了基于扩张卷积的卷积神经网络(convolutional neural networks, CNN)软测量建模方法,着重提取辅助变量间的空间特征。王通等<sup>[12]</sup>提出基于专家知识的生成对抗网络软测量数据扩充方法,提取数据深层特征以生成更优的虚拟样本。Li 等<sup>[13]</sup>提出 CNN 和双向长短记忆神经网络(bidirectional long short-term memory, BiLSTM)混合建模,从时空角度深度分析了过程数据。

随着深入的研究,学者们发现非欧图数据能更好地描述工业过程中多元变量间潜在的物理空间联系,因此图神经网络(graph neural network, GNN)被应用到软测量领域中,该方法将各个过程变量视为节点,并在有直接关联的变量之间建立边,从而将专家知识以图的形式融入模型<sup>[14]</sup>。Yan 等<sup>[15]</sup>提出自学习进化与节点感知图神经网络,结果优于现有软测量方法。Niresi 等<sup>[16]</sup>提出物理增强图神经网络,将物理原理融入图结构中,实验证明其优于仅考虑数据驱动的图神经网络。

目前也有学者提出了基于机理与数据的混合建模方法,分别建立多个基于机理以及基于数据的模型并融合。Ren 等<sup>[17]</sup>基于能量传递方程和流体动力学方程构建了直拉法生产单晶硅的机理软测量模型,引入 LSTM 网络对其进行修正,得到的混合软测量模型,有效提高预测精度。

上述软测量建模方法均在一定程度上提高了模型的准确性,然而,这些方法仅考虑了单一视角下的特征提取建模,忽略不同视角下的信息特征,导致预测精度低。因此,本文为进一步融合工业过程中时空两视角互补特征,提出了一种基于数据驱动的多视角对偶图注意力网络的工业软测量框架。

本文提出了结合时空特性的对偶图注意力网络。空间层面构建基于工艺流程和最大信息系数的 SGAT 模型,捕捉空间依赖关系;时间层面引入多层次时序图结构,结合长短时记忆网络与图注意力网络建立 TGAT 模型,融合长短期时序特征。提出了基于多视角对偶图注意力网络(multi-view dual graph attention, Mv-DGAT)的软测量框架,建立时空视角,全面提取数据特征。通过在损失函数中引入余弦相似度,度量视角间的互补关系,从而提升软测量模型的预测精度。通过两个工厂的公开现场数据集进行实验验证,证明模型具有准确性和泛化性。

## 1 多视角对偶图注意力网络软测量

针对单一视角建模难以全面提取系统的动态特征,本文提出了一种多视角学习框架的软测量模型,以图注意力网络(graph attention, GAT)<sup>[18]</sup>为主要技术,提取时空特征,其框架结构如图 1 所示。该模型包括了空间视角 SGAT 模块,时间视角 TGAT 模块以及特征融合回归模块,共同构成软测量框架。

SGAT 模块借助先验知识与最大信息系数(maximal information coefficient, MIC)协同度量关系,提取空间特征。时间视角搭建 TGAT 模块模型,采用 LSTM 和 GAT 混合建模,提取时间特征。多视角软测量建模以多头注意力机制(multi-head attention, MHA)为关键技术,进行特征融合及回归预测。最后在损失函数中引入余弦相似性函数,衡量视角间的互补性和特征冗余度,使模型更准

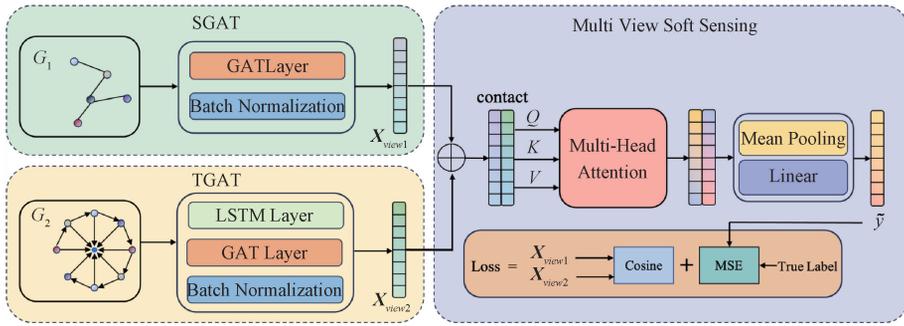


图 1 多视角软测量 Mv-DGAT 架构

Fig. 1 Multi-view soft sensing Mv-DGAT architecture

确和易收敛。

### 1.1 空间视角 SGAT

为了解决工业多元数据序列间存在空间耦合性的问

题,空间视角建立 SGAT 模块,以图的方式描述过程工业中各变量间的影响关系,获得更加全面清晰的空间特征表示。SGAT 模块的架构如图 2 所示。

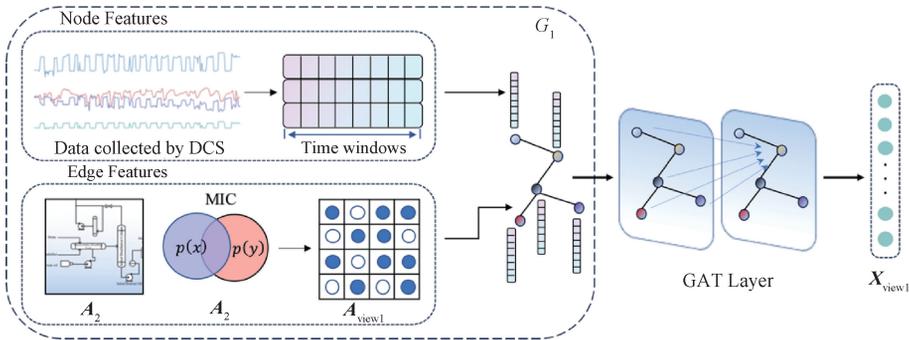


图 2 空间视角 SGAT 架构

Fig. 2 Space perspective SGAT architecture

首先,构建空间图  $G_1$ ,包含节点特征与邻接矩阵。构建空间图  $G_1$  的节点特征,针对每个变量采集到的  $n$  组数据,与这些数据相关的  $m$  个物理量作为节点。为了保持和时序视角预测的一致性并构建有效的子图结构,设定一个合理的时间窗  $t$ 。通过时间窗  $t$  对原始数据进行重构,从而将原始样本重构为  $n-t+1$  组新的样本数据。如此,总共得到  $n-t+1$  张子图,且每张子图均保持  $m$  个节点不变,而每个节点的特征维度为  $t$  维。子图的节点特征和标签如式(1)所示。

$$Node(i) = \begin{pmatrix} x_1((i-1) \times t + 1) & \cdots & x_m((i-1) \times t + 1) \\ \cdots & \cdots & \cdots \\ x_1(i \times t) & \cdots & x_m(i \times t) \end{pmatrix}^T$$

$$Label(i) = y(i \times t)$$

$$i = 1, 2, \dots, n - t + 1 \quad (1)$$

其次构建空间图  $G_1$  的邻接矩阵  $A_{view1}$ 。根据工艺流程,采用先验知识确定部分基础邻接矩阵  $A_1$ 。第 3 节实例分析中,通过实际工业过程硫回收装置的软测量问题

作为案例来具体介绍构建  $A_1$  的依据和方法。

随后再计算各个变量间的最大信息系数。MIC 是用于衡量两个变量之间非线性关系的统计方法,检测到线性和非线性的强关联,适合于发现数据中潜在的复杂关系。

首先将二维平面上的数据点进行网格划分,把变量  $X$  和  $Y$  的值域划分成若干等宽的区域。对于每种划分方式,可以计算  $X$  和  $Y$  在网格划分下的互信息,计算公式如式(2)所示。

$$MI(X_i, X_j) = \sum_{x \in X_i} \sum_{y \in X_j} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (2)$$

式中: $p(x, y)$  是  $X$  和  $Y$  同时出现的联合概率; $p(x)$  和  $p(y)$  是各自的边缘概率。

为了比较不同网格划分方法, MIC 对于每一种网格划分,按照式(3)对互信息归一化。

$$ave_M I = \frac{MI(X_i, X_j)}{\log_2(\min(|X_i|, |X_j|))} \quad (3)$$

式中: $|X|$  表示在该网格划分下的区域数量。

在所有的网格划分中,按式(4)选择归一化互信息的最大值,作为  $X$  和  $Y$  之间的最大信息系数。

$$MIC(X_i, X_j) = \max_{\text{grid partitions}} \text{ave}_M I(X_i, X_j) \quad (4)$$

计算出各变量间的 MIC 后,找到最大值和最小值,计算阈值  $\theta$ ,从而确定补充邻接矩阵  $A_2$  的值  $a_{ij}$ 。

$$\begin{cases} MIC_{\max} = \max(\{MIC(X_i, X_j) \mid i \neq j\}) \\ MIC_{\min} = \min(\{MIC(X_i, X_j) \mid i \neq j\}) \end{cases} \quad (5)$$

$$\theta = MIC_{\min} + 0.6 \times (MIC_{\max} - MIC_{\min}) \quad (6)$$

$$a_{ij} = \begin{cases} 1, MIC(x_i, x_j) \geq \theta \\ 0, MIC(x_i, x_j) < \theta \end{cases} \quad (7)$$

综合上述先验知识和 MIC 结果,得到空间视角的完整邻接矩阵  $A_{\text{view1}}$ 。

$$A_{\text{view1}} = A_1 + A_2 \quad (8)$$

将图数据  $G_1$  输入到 GAT Layer 中,其含有多个注意力头提取不同子空间的特征。每个注意力头单独计算其

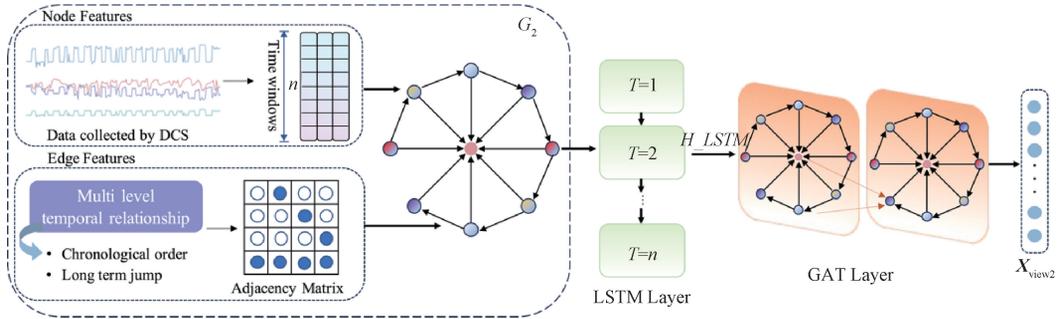


图 3 时间视角 TGAT 架构

Fig. 3 Time perspective TGAT architecture

首先构建时间图  $G_2$ 。此构建过程主要包含两个关键部分,即节点特征与邻接矩阵。首先构建时间图  $G_2$  的节点特征,将 DCS 采集到的时序数据使用类似空间视角方法进行重构。由于  $G_2$  的节点为时刻采样点,因此节点特征与  $G_1$  的节点特征呈转置关系。为了保持和空间视角预测的标签一致性,时间窗口  $t$  和子图标签与空间视角一致。节点特征和标签如式(12)所示。

$$\text{Node}(i) =$$

$$\begin{pmatrix} x_1((i-1) \times t + 1) & \cdots & x_m((i-1) \times t + 1) \\ \cdots & \cdots & \cdots \\ x_1(i \times t) & \cdots & x_m(i \times t) \end{pmatrix}$$

$$\text{Label}(i) = y(i \times t)$$

$$i = 1, 2, \dots, n - t + 1 \quad (12)$$

其次构建时间图  $G_2$  的邻接矩阵  $A_{\text{view2}}$ 。不同于空间图的边关系,每张时间子图均为静态边关系。每张子图按照采样的时间顺序连接作为基本图结构,并为了增加各节点对预测时刻的关注,引入所有节点到最后预测时刻的长跳跃连接。因此时空视角下的邻接矩阵  $A_{\text{view2}} \in \mathbf{R}^{n \times n}$  是一个静态矩阵,仅与时间窗口值,即节点个数有

注意力系数  $e_{ij}$ ,并进行归一化以满足概率分布<sup>[17]</sup>。

$$e_{ij} = \text{LeakyReLU}(\alpha^T \cdot [\mathbf{W} \cdot \mathbf{h}_i \parallel \mathbf{W} \cdot \mathbf{h}_j]) \quad (9)$$

式中: $\mathbf{h}_i$  为节点  $i$  对应特征向量; $\alpha$  表示可学习的权重向量。本文采用拼接的方式聚合  $\alpha_{ij}$ 。假设有  $K_1$  个注意力头,注意力分数的计算方法如式(10)所示。最后空间视角输出的完整表达如式(11)所示。

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik})} \quad (10)$$

式中: $\mathcal{N}(i)$  表示节点  $i$  的邻居节点集合。

$$X_{\text{view1}} = \frac{1}{K_1} \sum_{k_1=1}^{K_1} \text{Relu} \left( \sum_{j \in A_{\text{view1}}} \alpha_{ij}^{k_1} \cdot W^{k_1} \cdot x_j \right) \quad (11)$$

## 1.2 时间视角 TGAT

为了全面提取多元序列内的时序特征,时间视角建立 TGAT 模块,其架构如图 3 所示。

关,如式(13)所示。

$$A_{\text{view2}} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & & 0 & 0 \\ & & & \cdots & & \\ : & : & : & & : & \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{pmatrix} \quad (13)$$

在 TGAT 模块中,首先将  $G_2$  输入到 LSTM Layer 中,按时间序列输入内部时序子单元,将  $n$  维变量映射到隐藏空间中,提取数据短期时序特征<sup>[19]</sup>。取所有时刻的映射结果  $H\_LSTM$ ,代替  $G_2$  中的原始特征。

将更新特征后的  $G_2$  输入 GAT Layer 中,提取每个时刻对预测时刻的隐藏特征,利用与 SGAT 相同的聚合方式得到最终时序特征。假设有  $K_2$  个注意力头,时间视角输出的完整表达如式(14)所示。

$$X_{\text{view2}} = \frac{1}{K_2} \sum_{k_2=1}^{K_2} \text{ReLU} \left( \sum_{j \in A_{\text{view2}}} \alpha_{ij}^{k_2} \cdot W^{k_2} \cdot H\_LSTM_j \right) \quad (14)$$

## 1.3 软测量回归预测模块

在建立各个视角模型的基础上,搭建软测量框架。

其主要涵盖两个关键部分,构建软测量回归模型和基于时空特征优化损失函数,旨在提升模型在复杂系统中的预测精度。完整的多视角软测量模型框架如图4所示。

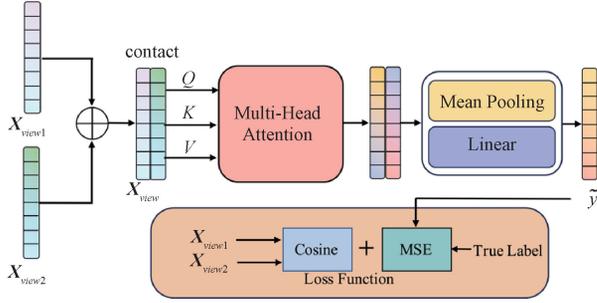


图4 软测量回归预测架构

Fig. 4 Soft sensing regression prediction framework

### 1) 搭建软测量回归模型

将上述时空视角所得结果  $X_{view1}$  和  $X_{view2}$  拼接为  $X_{view}$ , 将其输入到多头注意力机制中进行特征增强。本研究使用多头自注意力机制, 即  $Q, K, V$  的输入均为  $X_{view}$ , 每个注意力头的注意力分数计算如式(15)所示。

$$Attention_i = \text{softmax}\left(\frac{X_{view} W_i^Q (X_{view} W_i^K)^T}{\sqrt{d_k}}\right) X_{view} W_i^V \quad (15)$$

式中:  $W_i^Q, W_i^K, W_i^V$  为可学习参数,  $i$  为注意力头数。

最后将每个注意力头的输出进行拼接, 将其映射到模型的原始维度上。

$$MHA = \text{concat}(head_1, \dots, head_i) W^O \quad (16)$$

式中:  $W^O$  为可训练的投影矩阵。

在经过 MHA 后接经过平均池化层和线性层得到软测量关键变量预测值  $\tilde{y}$ 。

### 2) 改进损失函数

在单一视角下, 软测量模型的损失函数通常采用预测值与真实值的均方误差 (MSE)。然而, 在多视角学习框架中, 仅依赖均方误差作为损失函数不足以充分保证模型各视角之间的信息互补性及特征融合效果。两视角输出特征相似度过高会导致各视角的特征趋于一致, 从而弱化了多视角学习的核心优势, 即视角间的信息互补性。这种特征同质化现象可能限制模型对不同视角特征的差异化表达, 进而影响最终的特征融合效果和预测性能。

因此, 损失函数在均方误差的基础上引入了余弦相似度 (cosine), 以评估时空视角输出特征的相似程度。引入 cosine 和 MSE 后损失函数可以表示为:

$$Loss = \beta \cdot \text{cosine}(X_{view1}, X_{view2}) + (1 - \beta) \cdot \text{MSE}(\tilde{y}, y) =$$

$$\frac{\beta \cdot \sum_{i=1}^n \sum_{j=1}^d (X_{view1})_{ij} (X_{view2})_{ij}}{\sqrt{\sum_{i=1}^n \sum_{j=1}^d (X_{view1})_{ij}^2} \cdot \sqrt{\sum_{i=1}^n \sum_{j=1}^d (X_{view2})_{ij}^2}} + \frac{(1 - \beta)}{m} \sum_{i=1}^m (\tilde{y}_i - y_i)^2 \quad (17)$$

式中:  $m$  为样本数量;  $n$  和  $d$  为两视角输出特征的维度;  $\beta$  为可调整的权重参数。

## 2 实例分析

利用两个真实的过程工业数据集, 硫磺回收装置 (sulfur recovery unit, SRU) 工艺和联合循环发电设备 (combined cycle power plant, CCPP) 来验证 Mv-DGAT 的有效性。所有实验均在 NVIDIA A100 80 GB PCIe GPU (每块 VRAM 80 GB) 上进行。

本文建立 5 个基准模型比较: 梯度提升决策树 (eXtreme gradient boosting, XGBoost), 时间卷积网络 (temporal convolutional network, TCN), 质量驱动的堆叠自编码器 (quality-driven stacked auto-encoder, QD-SAE)<sup>[20]</sup> 以及基于 Transformer 的门控卷积神经网络 (Transformer Based on gated convolutional network, GCT)<sup>[21]</sup>。此外, 研究设计了消融实验, 系统地分析了 Mv-DGAT 中各模块与框架对预测性能的影响。

同时, 为检验模型的软测量预测效果, 选用决定系数  $R^2$ 、平均绝对误差 (MAE)、均方根误差 (RMSE) 作为预测效果的评价指标, 计算公式如式(18)~(20)所示。

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - \tilde{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (18)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \tilde{y}_i| \quad (19)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \tilde{y}_i)^2} \quad (20)$$

### 2.1 硫回收装置

在石油炼制过程中, SRU 的设计目的是去除环境污染物并从酸性气流中回收硫元素, 这对保护环境和提高生产效率具有重要意义。SRU 以含  $H_2S$  和  $NH_3$  的混合气体作为输入, 在反应炉内充分燃烧以去除气体中硫元素。由于尾气中常含有残留的  $SO_2$ , 容易损坏装置硬件设备, 所以要求对其浓度实时监测, 以防损坏设备带来的损失。因此, 使用软测量方法对尾气中的  $SO_2$  浓度进行实时测量是非常必要的。

本文选择了 5 种辅助变量对  $SO_2$  浓度进行软测量建模, 包含 10 081 组样本, 变量描述如表 1 所示。

表 1 SRU 过程中主要过程变量

Table 1 Main process variables in SRU process

辅助变量	工艺参数
$X_1$	主燃烧室 MEA 流量 (MEA_GAS)
$X_2$	主燃烧室空气流量 (AIR_MEA)
$X_3$	手动输入空气流量 (AIR_MEA_2)
$X_4$	副燃烧室 SWS 流量 (SWS_GAS)
$X_5$	副燃烧室空气流量 (SWS_AIR)

随后确定空间视角的邻接矩阵。首先空气流量是按照燃烧气体流量通过燃烧化学式进行配比,并在燃烧室空气不足时进行手动输入空气。因此,主副燃烧室中存在燃气量与空气量的耦合关系,即主燃烧室  $X_1-X_2$ 、 $X_1-X_3$ 、 $X_2-X_3$  以及副燃烧室  $X_4-X_5$ 。其次分析主副燃烧室间的耦合关系,主燃烧室中 MEA 燃气与空气混合燃烧后部分空气进入副燃烧室,用于 SWS 气体的燃烧反应。因此,主燃烧室空气总量会影响副燃烧室反应程度和燃气含量,建立  $X_2-X_4$  和  $X_3-X_4$  的边连接。由此,根据工艺知识获取相应辅助变量的层次依赖关系,得到基于工业流程的邻接矩阵。其次对辅助变量进行最大信息系数 MIC 分析,其结果如图 5 所示。



图 5 各个过程变量间 MIC 系数

Fig. 5 MIC coefficients between process variables

由式(6)、(7)可以确定基于 MIC 的邻接矩阵。由于两者为协调补充关系,因此将其进行矩阵叠加,得到完整的图邻接矩阵。为了更加直观地表示邻接矩阵  $A$ ,采用图结构的方式进行展示,如图 6 所示。其中蓝色线条对应工艺边关系,黑色线条对应 MIC 补充边关系。

数据输入批次为 64,训练样本和测试样本比例为 9 : 1,时间窗口大小设置为 10。为保证实验的公平性和有效性,采用经验试错法对每个模型超参数进行调优。

每个模型的主要超参数配置如下:各模型均训练 300 个 epoch,损失函数权重  $\beta$  设置为 0.01, Mv-DGAT 中 LSTM 结构为 [5-256-512-5], GAT 结构为 [5-512-512-5];MHA 使用 4 个注意力头。XGBoost 深度设置为 7,树数量为 150。QD-SAE 中 SAE 结构为 [5-512-512-5]。GCT 结构为[5-128-512-5]。

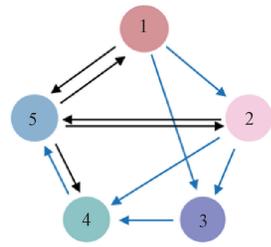


图 6 SRU 空间图结构

Fig. 6 SRU spatial graph structure

所有模型的预测效果和误差如表 2 所示,预测结果如图 7 所示,所提出的 Mv-DGAT 在所有评价指标上都优于其他对比模型。这一结果表明,本文提出的 Mv-DGAT 方法能够更有效地提取多元时序数据的时空深层表征特征,从而提高了软测量预测性能。

表 2 不同方法预测的 SO<sub>2</sub> 性能指标

Table 2 Prediction SO<sub>2</sub> indicators of different methods

方法	$R^2$	MAE	RMSE
XGBoost	0.622 5	0.026 1	0.035 7
TCN	0.649 5	0.026 9	0.034 5
GCT	0.805 1	0.019 0	0.025 7
QD-SAE	0.848 5	0.017 7	0.022 7
<b>Mv-DGAT</b>	<b>0.878 3</b>	<b>0.015 4</b>	<b>0.020 3</b>

图 7 进一步说明了传统简单模型 (XGBoost 和 TCN) 可以较为准确地跟踪质量变量,有效地拟合多元数据。然而,它们只关注单个时间点的主导变量和辅助变量,而忽略了工业过程数据的时间特性以及工业流程的空间特性。QD-SAE 通过堆叠多层编码器实现了单一视角下有效提取时间特征,充分利用时序数据内的时间信息从而提高了预测性能。GCT 对多元数据的短期模式进行编码,通过改进的门控卷积神经网络实现了单一视角下有效提取空间特征,描述数据的深层空间表征。然而,这限制了其时间特征提取能力,忽略了时序视角下模型的预测能力。

因此,本文提出的 Mv-DGAT 模型采用多视角学习的框架,兼顾过程数据中的时间和空间信息,同时,损失函数中的余弦相似性进一步对特征进行了筛选,从而进一步提高了软测量模型的预测精度。

## 2.2 联合循环发电设备

联合循环电厂运行系统利用燃气轮机响应快、带负载能力强等优势驱动蒸汽轮机发电,以实现高效的功率输出,提升能源利用效率。CCPP 以燃气为主要输入能源,燃烧发电。然而,在这个过程中难免发生副反应,产生大量的氮氧化物 NO<sub>x</sub>。由于该气体会引发光化学烟雾和全球变暖等环境问题,所以必须对其排放进行严格

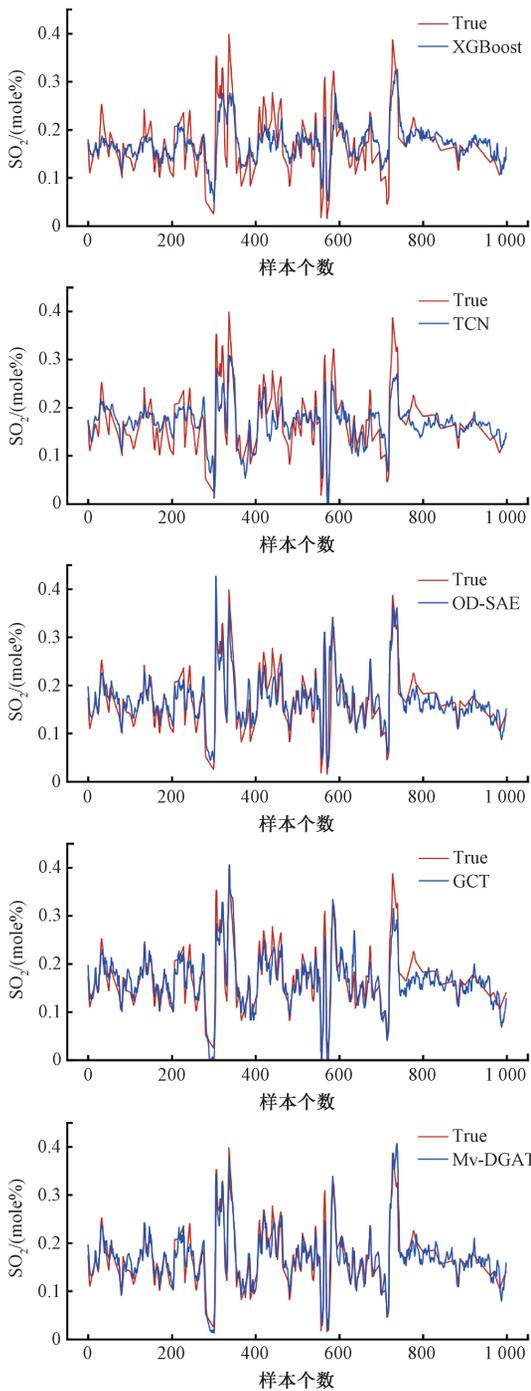


图 7 不同算法 SO<sub>2</sub> 浓度预测结果

Fig. 7 Prediction results of SO<sub>2</sub> using different algorithms

监控,以防污染环境。因此,对CCPP过程建立Mv-DGAT软测量模型,实现污染气体排放的连续监控和合理控制。

本文选择了9种辅助变量以及NO<sub>x</sub>浓度作为目标变量,使用2015年收集的7384组样本进行软测量建模,主要过程变量如表3所示。

CCPP和SRU采用同样的方式建立空间图,各辅助变量的MIC分析结果如图8所示。

表 3 CCPP 过程中主要过程变量

Table 3 Main process variables in CCPP process

辅助变量	工艺参数
X <sub>1</sub>	环境温度(AT)
X <sub>2</sub>	环境压力(AP)
X <sub>3</sub>	环境湿度(AH)
X <sub>4</sub>	空气过滤器差压(AFDP)
X <sub>5</sub>	涡轮机排气压力(GTEP)
X <sub>6</sub>	涡轮机入口温度(TIT)
X <sub>7</sub>	涡轮机出口温度(TAT)
X <sub>8</sub>	压缩机排放压力(CDP)
X <sub>9</sub>	涡轮机能量产量(TEY)

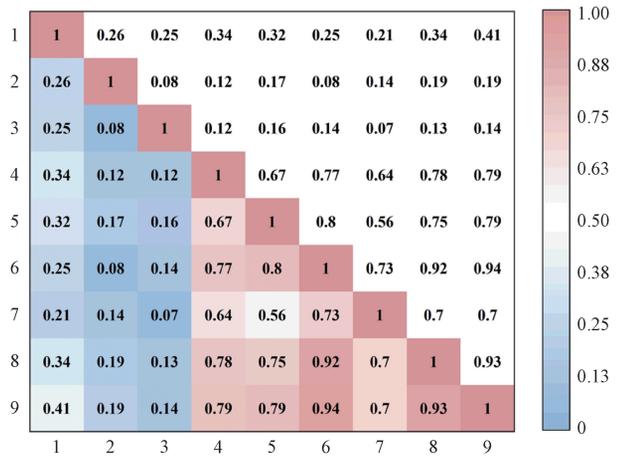


图 8 辅助变量 MIC 系数

Fig. 8 MIC coefficients between process variables

CCPP空间图结构如图9所示。其中蓝色线条对应工艺边关系,黑色线条对应MIC补充边关系。

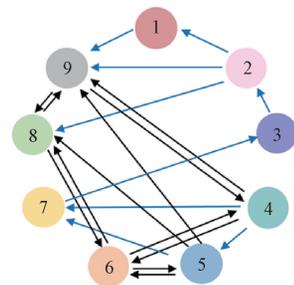


图 9 CCPP 空间图结构

Fig. 9 CCPP spatial graph structure

数据集预处理、划分以及训练轮次均与实例1保持相同,同样采用经验试错法对每个模型的超参数进行调优。每个模型的主要超参数配置如下:各模型均训练300个epcho,损失函数权重β设置为0.01,Mv-DGAT中LSTM结构为[9-128-512-9];GAT结构为[9-256-512-9];MHA使用4个注意力头。XGBoost深度设置为

9, 树数量为 200。QD-SAE 结构为 [9-256-512-9]。GCT 结构为 [9-128-512-128-9]。

预测结果如图 10 所示, 进一步说明仅将单时刻采样点作为输入的两种传统模型忽略了样本内时空模态信息, 导致预测准确性较差。虽然 QD-SAE 和 GCT 考虑了单一视角下数据的时间和空间特性, 使  $R^2$  和误差指标 (RMSE 及 MAE) 有明显提升, 但是它们仅考虑单一视角

数据特性, 忽略了其他视角下的互补信息, 导致预测精度较低。CCPP 实例的结果进一步表明, 本文提出的 Mv-DGAT 模型能够有效捕捉不同视角间的互补关系, 全面提取数据深度特征, 提高了软测量的精度。

所有模型的预测效果和误差如表 4 所示, 所提出的 Mv-DGAT 模型在所有评价指标上都优于其他模型。这一结果表明, 多视角学习框架下的对偶图注意力网络可以更有效地挖掘工业过程数据的深度特征, 利用子空间信息进行高效的软测量建模。

表 4 不同方法的预测 NO<sub>x</sub> 性能指标

Table 4 Prediction NO<sub>x</sub> indicators of different methods

方法	$R^2$	MAE	RMSE
XGBoost	0.694 9	3.883 3	5.340 4
TCN	0.748 2	3.530 9	4.852 1
QD-SAE	0.809 7	2.727 0	4.217 6
GCT	0.777 6	3.102 6	4.559 0
<b>Mv-DGAT</b>	<b>0.856 1</b>	<b>2.253 7</b>	<b>3.668 8</b>

### 2.3 消融实验

为了研究 Mv-DGAT 模型中各模块对软测量预测效果的影响, 设计了 3 个消融模型, 如表 5 所示。

表 5 消融实验模型介绍

Table 5 Introduction to ablation experimental model

方法	结构	验证对象
消融模型 1	空间视角 SGAT+MHA 回归预测	TGAT
消融模型 2	时间视角 TGAT+MHA 回归预测	SGAT
消融模型 3	多视角建模特征+MHA 回归预测	损失函数
Mv-DGAT	消融模型 3+改进损失函数	—

所有消融模型均使用与 Mv-DGAT 相同的超参数设置, 并在 SRU 和 CCPP 数据集上进行实验比较。

消融实验结果如表 6 所示。实验结果明显地说明多视角框架使模型的整体性能有显著提升。但是单纯的时空两视角特征进行融合存在信息冗余的问题, 并且对模型存在误导影响。因此, 消融模型 3 进一步表明, 在损失函数中引入 cosine 函数用于衡量不同通道间的互补性以及重复性是非常有必要的, 提高了模型的泛化性能和预测准确性。

此外, 考虑到模型中引入了多视角框架会只是增加了时间训练成本, 对比了该结构与单通道消融模型在训练时间与参数量方面的差异。结果表明, 多视角模型虽然训练时间提升约 31%, 但预测性能显著增强,  $R^2$  提高 6%, RMSE 下降 23%。但本研究所提出的软测量模型采用离线训练, 在线检测时需将模型嵌入硬件, 不影响在线检测时间, 离线训练时间增加对实时预测无影响; 而模型

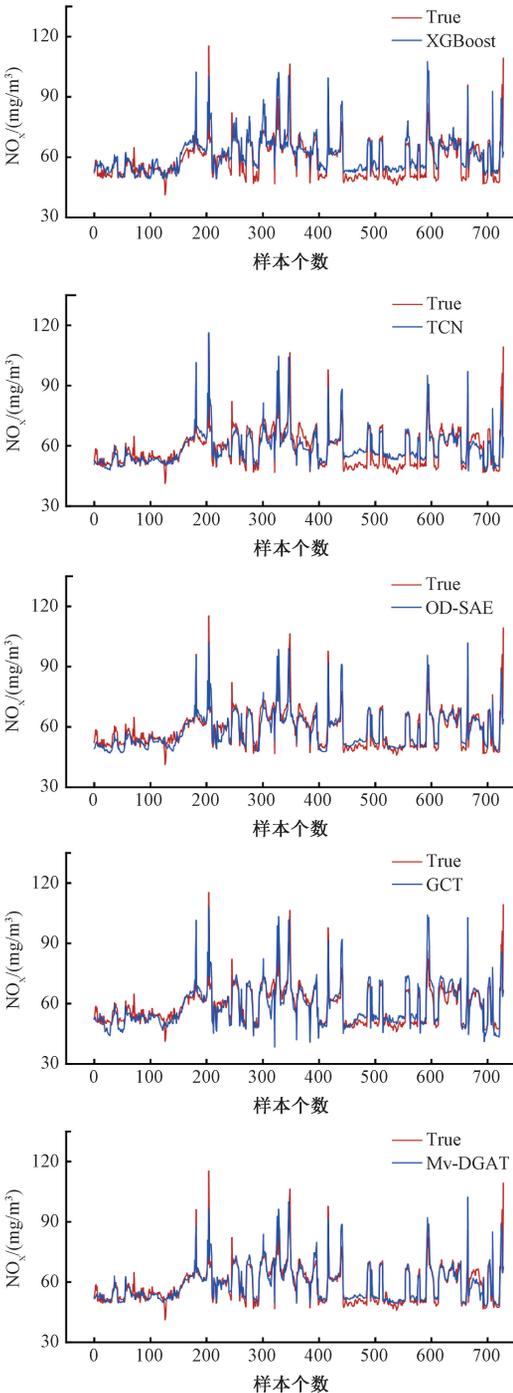


图 10 不同算法 NO<sub>x</sub> 浓度预测结果

Fig. 10 Prediction results of NO<sub>x</sub> using different algorithms

精度的提升对于工业实时预测任务更为重要。因此,此方法在计算成本可控的基础上实现了显著性能增益,具有良好的工程应用潜力。

表6 消融实验结果

Table 6 Results of ablation experiment

参数指标	消融模型1	消融模型2	消融模型3	Mv-DGAT
$R^2$	0.712 0	0.810 1	0.870 5	<b>0.878 3</b>
SRU MAE	0.024 0	0.018 7	0.016 5	<b>0.015 4</b>
RMSE	0.031 2	0.025 3	0.020 9	<b>0.020 3</b>
$R^2$	0.766 7	0.804 2	0.830 5	<b>0.856 1</b>
CCPP MAE	3.244 6	2.775 9	2.696 1	<b>2.253 7</b>
RMSE	4.670 4	4.279 0	3.980 4	<b>3.668 8</b>
离线训练时间/min	14.86	22.51	29.64	29.64
模型参数量/ ( $\times 10^6$ )	1.74	3.96	6.85	6.85

### 3 结 论

针对当前基于深度学习的软测量忽略了其他视角下有益信息的问题,提出了 Mv-DGAT 模型,设计了基于多视角学习框架的对偶图注意力软测量建模方法。Mv-DGAT 模型以图注意力网络为关键技术分别搭建了时间和空间视角模型,用于分别提取两视角深层特征。两视角特征拼接后使用 MHA 回归预测。其次,难点技术在于改进了适合于本模型的损失函数,将余弦相似度添加到损失函数中,提高两视角输出特征的互补性,提高预测性能。最后,将所提方法应用于两个实际工业过程数据(SRU 和 CCPP)软测量建模任务中,证明 Mv-DGAT 有效地提高了预测模型的性能。在未来,拟将采用即时学习以及大语言模型等技术提高模型的普适性,并将其拓展到半监督软测量领域。

### 参考文献

[ 1 ] YUAN X, QI S, SHARDT Y A W, et al. Soft sensor model for dynamic processes based on multichannel convolutional neural network [ J ]. Chemometrics and Intelligent Laboratory Systems, 2020, 203: 104050.

[ 2 ] WANG Y L, LIU D J, LIU C L et al. Dynamic historical information incorporated attention deep learning model for industrial soft sensor modeling [ J ]. Advanced Engineering Informatics, 2022, 52: 101590.

[ 3 ] LIU S, SUN W. Attention mechanism-aided data- and knowledge-driven soft sensors for predicting blast furnace gas generation [ J ]. Energy, 2022, 262(A): 125498.

[ 4 ] LI Z, JIN H, DONG S, et al. Semi-supervised ensemble support vector regression based soft sensor for key quality variable estimation of nonlinear industrial processes with limited labeled data [ J ]. Chemical Engineering Research and Design, 2022, 179: 510-526.

[ 5 ] XU Y J, CAI Z D. Gaussian mixture deep dynamic latent variable model with application to soft sensing for multimode industrial processes [ J ]. Applied Soft Computing, 2021, 114: 108092.

[ 6 ] SHI X, KANG Q, ZHOU M, et al. Novel L1 regularized extreme learning machine for soft-sensing of an industrial process [ J ]. IEEE Transactions on Industrial Informatics, 2021, 18(2): 1009-1017.

[ 7 ] 全卫国, 郭超宇, 赵如意. 基于改进麻雀算法优化 LSSVM 的再循环箱浆液密度预测模型 [ J ]. 电子测量技术, 2022, 45(1): 70-76.

TONG W G, GUO CH Y, ZHAO R Y. Prediction model of recirculation tank slurry density based on improved sparrow search algorithm optimized LSSVM [ J ]. Electronic Measurement Technology, 2022, 45(1): 70-76.

[ 8 ] 徐雯, 汤健, 夏恒, 等. 基于 Bagging 半监督深度森林回归的二噁英排放浓度软测量 [ J ]. 仪器仪表学报, 2022, 43(6): 251-9.

XU W, TANG J, XIA H, et al. Soft sensing of dioxin emission concentration based on bagging semi-supervised deep forest regression [ J ]. Chinese Journal of Scientific Instrument, 2022, 43(6): 251-259.

[ 9 ] SUN Q, GE Z. A survey on deep learning for data-driven soft sensors [ J ]. IEEE Transactions on Industrial Informatics, 2021, 17(9): 5853-5866.

[ 10 ] 周建新, 郑日成, 侯宏瑶. 改进鸚鵡算法优化 LSTM 的加热炉钢坯温度预测 [ J ]. 国外电子测量技术, 2023, 42(5): 174-179.

ZHOU J X, ZHENG R CH, HOU H Y. Prediction of the temperature of steel billets in the heating furnace using an LSTM optimized by an improved pelican algorithm [ J ]. Foreign Electronic Measurement Technology, 2023, 42(5): 174-179.

[ 11 ] WU H, HAN Y, JIN J, et al. Novel deep learning based on data fusion integrating correlation analysis for soft sensor modeling [ J ]. Industrial & Engineering Chemistry Research, 2021, 60(27): 10001-10010.

[ 12 ] 王通, 陈延彬. 基于改进生成对抗网络的动液面建模数据扩充 [ J ]. 电子测量与仪器学报, 2023, 37(2): 99-109.

WANG T, CHEN Y B. Data augmentation for dynamic liquid level modeling based on improved generative adversarial network [ J ]. Journal of Electronic Measurement and Instrumentation, 2023, 37(2): 99-109.

[ 13 ] LI Y, PENG T, SUN W, et al. A soft sensor model based on CNN-BiLSTM and IHHO algorithm for tennessee

eastman process [J]. Measurement, 2023, 218: 113195.

- [14] DONG G M, TANG M Y, WANG Z Y, et al. Graph Neural Networks in IoT: A Survey [J]. ACM Transactions on Sensor Networks, 2022, 19(2): 47.
- [15] YAN F, WANG C, WANG Z C, et al. SENGraph: A self-learning evolutionary and node-aware graph network for soft sensing in industrial processes [J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(7): 12284-12298.
- [16] NIRESI K F, BISSIG H, BAUMANN H, et al. Physics-enhanced graph neural networks for soft sensing in industrial internet of things [J]. IEEE Internet of Things Journal, 2024, 11(12): 34978-34990.
- [17] REN J C, LIU D, WAN Y. Modeling and application of Czochralski silicon single crystal growth process using hybrid model of data-driven and mechanism-based methodologies [J]. Process Control, 2021, 104:74-85.
- [18] VRAHATIS A G, LAZAROS K, KOTSIANTIS S. Graph attention networks: A comprehensive review of methods and applications [J]. Future Internet, 2024, 16(9): 318.
- [20] SONG X L, ZHANG N, SHI Y, et al. Quality-driven deep feature representation learning and its industrial application to soft sensors [J]. Journal of Process Control, 2024, 142: 103300.
- [21] GENG Z, CHEN Z, MENG Q, et al. Novel transformer based on gated convolutional neural network for dynamic soft sensor modeling of industrial processes [J]. IEEE Transactions on Industrial Informatics, 2021, 18(3):

1521-1529.

## 作者简介



**郭中宇**, 2023 年于北京化工大学获得学士学位, 现为北京化工大学硕士研究生, 主要研究方向为深度学习以及工业软测量建模。

E-mail: 2023200769@mail.buct.edu.cn

**Guo Zhongyu** received his B. Sc. degree from Beijing University of Chemical Technology in 2023. Now he is a M. Sc. candidate at Beijing University of Chemical Technology. His main research interests include deep learning and industrial soft sensing modeling.



**岳玉麒**, 2023 年于哈尔滨师范大学获得硕士学位, 现为北京化工大学博士研究生。主要研究方向为计算机视觉、数据驱动的工业建模和过程数据分析。

E-mail: yueyuqi980717@163.com

**Yue Yuqi** received her M. Sc. degree from Harbin Normal University in 2023. Now she is a Ph. D. candidate at Beijing University of Chemical Technology. Her main research interests include computer vision, data-driven industrial modeling, and process data analysis.



**陈娟** (通信作者), 博士、教授, 博士研究生导师, 主要研究方向为过程检测及自动化装置、智能检测技术与智能控制等。

E-mail: jchen@mail.buct.edu.cn

**Chen Juan** (Corresponding author) Ph. D., professor, Ph. D. supervisor. Her main research interests include process detection and automation devices, intelligent detection technology and intelligent control.