

## 基于知识蒸馏的钢铁高炉煤气系统建模方法

金 锋<sup>1,2†</sup>, 陈薇琳<sup>1,2</sup>, 赵博识<sup>3</sup>, 赵 珺<sup>1,2</sup>, 王 伟<sup>1,2</sup>

(1. 工业装备智能控制与优化教育部重点实验室, 辽宁 大连 116024; 2. 大连理工大学 控制科学与工程学院, 辽宁 大连 116024;

3. 马鞍山钢铁股份有限公司 能源环保部, 安徽 马鞍山 243003)

**摘要:** 钢铁企业高炉煤气系统具有波动性大、时变性强、不确定性高等特点, 对其未来产消趋势进行准确的建模预测有助于企业的高效决策与节能减排. 文章提出了一种基于知识蒸馏的高炉煤气系统建模方法, 为了提高训练过程中的拟合精度, 在教师网络中建立了基于长短期记忆网络的序列到序列模型来提取样本的中间特征. 进而, 提出了融入教师模型中间特征的知识蒸馏策略, 建立了考虑中间特征蒸馏损失与预测均方误差的损失函数, 对知识蒸馏过程及预测偏差进行评估. 通过国内大型钢铁企业高炉煤气系统实际运行数据的实验验证, 表明了本文所提建模方法的有效性, 可为后续的能源系统优化调度提供支撑.

**关键词:** 知识蒸馏; 时间序列; 高炉煤气系统; 钢铁企业

**引用格式:** 金锋, 陈薇琳, 赵博识, 等. 基于知识蒸馏的钢铁高炉煤气系统建模方法. *控制理论与应用*, 2024, 41(3): 428 – 435

DOI: 10.7641/CTA.2023.20864

## A knowledge distillation-based modeling method for blast furnace gas system in steel industry

JIN Feng<sup>1,2†</sup>, CHEN Wei-lin<sup>1,2</sup>, ZHAO Bo-shi<sup>3</sup>, ZHAO Jun<sup>1,2</sup>, WANG Wei<sup>1,2</sup>

(1. Key Laboratory of Intelligent Control and Optimization for Industrial Equipment (Ministry of Education), Dalian Liaoning 116024, China;

2. School of Control Science and Engineering, Dalian University of Technology, Dalian Liaoning 116024, China;

3. Ministry of Energy and Environmental Protection, Maanshan Iron & Steel Co., Ltd., Maanshan Anhui 243003, China)

**Abstract:** Blast furnace gas system in steel enterprises has the characteristics of high volatility, time-variability and great uncertainty, accurately modeling of its future generation and consumption flow plays a crucial role in efficiently decision-making, energy-saving and emissions reduction. In this study, a knowledge distillation-based modelling method for blast furnace gas system is proposed. Based on a long and short-term memory network, a sequence-to-sequence model is built in the teacher network to extract the intermediate features of the samples. And then, a knowledge distillation strategy is constructed which incorporates the intermediate features of the teacher model. Besides, in order to evaluate the capability of feature extraction, a new loss function is established by both considering that of the knowledge distillation process and the regression error of the actual energy data. Validation experiments are carried out by employing real-world data from the blast furnace gas system of a typical steel enterprise, and the results indicate the effectiveness of the proposed method when facing with the modeling problem, so as to provide powerful support for the optimal scheduling of the energy system.

**Key words:** knowledge distillation; time series; blast furnace gas system; steel industry

**Citation:** JIN Feng, CHEN Weilin, ZHAO Boshi, et al. A knowledge distillation-based modeling method for blast furnace gas system in steel industry. *Control Theory & Applications*, 2024, 41(3): 428 – 435

收稿日期: 2022–09–30; 录用日期: 2023–10–10.

†通信作者. E-mail: jin\_feng@dltu.edu.cn.

本文责任编辑: 周平.

国家重点研发计划项目(2017YFA0700300), 国家自然科学基金项目(62125302, 61833003, U1908218, 62103075), 辽宁省“兴辽英才计划”项目(XLYC2002087), 大连市科技人才创新支持计划项目(2022RG03)资助.

Supported by the National Key R&D Program of China (2017YFA0700300), the National Natural Sciences Foundation of China (62125302, 61833003, U1908218, 62103075), the Liaoning Revitalization Talents Program (XLYC2002087) and the Sci-Tech Talent Innovation Support Program of Dalian(2022RG03).

## 1 引言

钢铁工业是我国能源消耗大户,其能源消耗约占全国总量的14%<sup>[1]</sup>,而随着我国双碳目标的提出<sup>[2]</sup>,钢铁工业的节能降碳已经成为了行业关注的重点.其中,副产能源作为钢铁生产过程中重要的二次能源,约占钢铁企业总能耗的40%<sup>[3]</sup>,高炉煤气(blast furnace gas, BFG)作为炼铁过程中产生的副产能源的一种,具有产流量大、波动剧烈、时变性强、不确定性高等特点.因此,对其未来趋势进行准确、动态的建模与预测,可为能源系统的调度决策提供支撑,对减少能源浪费、降低碳排放具有重要意义<sup>[4]</sup>.

目前,国内外研究人员对于钢铁企业副产煤气的建模方法进行了大量的研究.文献[5]针对现有预测模型预测提前量较短的问题,提出了一种基于时间序列的反向传播神经网络(back-propagation, BP)的中期预测模型,在保证较高的预测精度的前提下,将预测时长延长至30 min.文献[6]提出了一种结合BP神经网络和最小二乘支持向量机(least square support vector machine, LSSVM)的高炉煤气预测模型,通过LSSVM进行趋势序列预测,BP神经网络进行波动序列预测,大幅度提升了预测精度.文献[7]则针对高炉煤气系统设备工况复杂、产销量波动大的问题,提出了一种结合小波分析、BP神经网络和LSSVM的复合预测模型,在文献[6]的基础上,将波动预测数据和趋势预测数据通过小波重构进行整合,将预测的平均绝对百分比误差(mean absolute percentage error, MAPE)控制在5%之内.此外,文献[8]提出了一种基于回声状态神经网络(echo state network, ESN)的在线预测方法,通过补偿信号作为特定输入消除噪声的影响,实现了比ESN更好的预测效果.在此基础上,文献[9]提出了一种基于变分推理回声状态网络集成模型的区间预测方法,该方法通过变分推理来近似推导出集成模型中所有不确定参数的联合后验概率分布作为模型参数值,实现了在预测精度、区间质量、模型稳定性,以及耗时方面有更好的表现.文献[10]建立了一种结合混合事件、机理和数据驱动的预测模型,根据每个高炉的运行状态、过程机理和历史数据,实现不同工况下不同预测模型的精准切换.文献[11]则提出一种基于双高斯核的在线区间预测方法,将区间构造过程中的雅克比矩阵计算转化为对核的计算,从而,在预测精度、区间效果和计算时效性3个方面都得到了明显的提升.然而,上述方法大都针对特定数据进行针对性建模,根据形态函数捕捉时序特征规律,未能充分挖掘数据序列内隐式的时序性周期性信息,即一段时间内特征之间的相互关系.

Hochreiter和Schmidhuber<sup>[12]</sup>于1997年首次提出了长短期记忆网络模型(long-short term memory, LS-

TM),该网络是一种递归神经网络,通过控制内部存储器单元输出的门控机制来学习输入序列数据中的复杂表示,实现了通过控制信息去留的方式取得更长的时序依赖<sup>[13-14]</sup>,同时,也被广泛应用于解决工业相关的建模问题.文献[15]针对高炉煤气系统在建模过程中存在的样本量大、工况繁多等情况,建立了基于LSTM以及季节性差分自回归模型的梯度驱动时序预测复合模型,该方法通过数据梯度变化判断高炉运行状态,实现了变工况下预测模型的切换,相比单一模型具有更高的预测精度.此外,文献[16]针对高炉煤气系统数据的波动特性,基于差分自回归模型(autoregressive integrated moving average, ARIMA), LSTM和小波变换(wavelet transform, WT)等多种方法,建立了ARIMA-WT-LSTM混合预测模型,从而,对样本进行分类并预测,提高了模型精度.然而,钢铁企业高炉煤气系统的实际生产过程受到多种内部和外部复杂因素的影响,具有高时变性和不确定性.上述研究主要利用LSTM模型学习时序特征之间的依赖关系的方式进行特征提取,但由于LSTM模型在LSTM单元内的同时进行特征提取与预测,因此,其特征提取的能力始终有限,很难学习到更加完整的时序特征,并将提取的特征信息全部应用于预测.

针对上述问题,本文提出了一种基于知识蒸馏的高炉煤气系统建模方法.建立了基于LSTM网络的序列到序列模型作为教师网络,同时,采用LSTM模型作为学生网络,从而,提取训练样本中的内部特征.进而,提出了融入教师模型中间特征的知识蒸馏策略,构建了考虑中间特征蒸馏损失与预测均方误差的损失函数,在每个训练轮次均对输入样本集的知识蒸馏过程及预测偏差进行评估,从而,实现将教师网络的中间层特征提取并迁移至学生网络.采用国内大型钢铁企业高炉煤气系统的发生量与消耗量数据进行实际验证,并分别将LSTM方法(单独学生网络)<sup>[17]</sup>、基于LSTM的seq2seq方法(单独教师网络)<sup>[18]</sup>和LSSVM方法<sup>[19]</sup>作为对比实验,结果表明,本文所提的基于知识蒸馏的建模方法具有更高的精度,可为后续的能源系统优化调度提供有力支撑.

## 2 问题描述

如图1所示为典型钢铁企业高炉煤气系统结构图,主要由发生单元、消耗单元、存储单元、管网以及压机等输送单元组成.高炉煤气是高炉炼铁过程中产生的副产能源,在优先供给自身配备的多座热风炉使用后,通过错综复杂的管网输送给热轧、冷轧、焦化、电厂等消耗单元.高炉煤气柜是系统的主要缓冲与存储设备,起到富余煤气临时存储与系统波动缓冲等作用.

由此可见,对于发生和消耗单元流量的准确预测,

将可为其优化调度过程提供有力支撑. 其中, 热风炉作为高炉炼铁过程的送风设备, 其在生产过程中需消耗大量的高炉煤气, 其燃烧、送风、焖炉等不同工况的切换具有周期性的剧烈波动的特点, 是导致高炉煤气系统流量的波动的重要因素. 高炉生产过程连续, 内部通过焦炭、铁矿石等发生还原反应冶炼铁水, 同时, 持续产生大量的高炉煤气输送到管网. 受热风炉以及原料成分、品质及炉顶压力变化、温度变化等复杂因素影响, 高炉煤气的发生流量存在较大幅度的波动和随机扰动. 因此, 无法直接通过热力学等机理方式对其进行准确建模与预测. 如图2所示为典型的高炉煤气发生流量与热风炉消耗流量的数据特征.

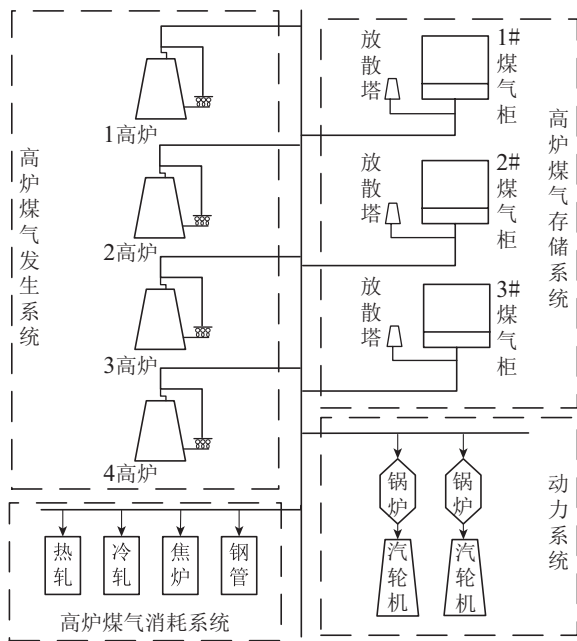


图1 高炉煤气系统结构图

Fig. 1 Structure of blast furnace gas system

在实际生产过程中, 受工艺机理和生产计划的双重影响, 煤气系统始终处于动态平衡, 一旦发生或消耗出现不平衡情况, 则需要调度人员及时给出相应的调整方案, 以达到新的平衡状态. 然而, 由于煤气系统设备单元多, 管网范围广, 生产工况复杂, 难以仅依靠人的经验来准确判断未来的能源供需趋势, 从而增加了调度方案的制定难度. 随着企业信息化的发展, 现场安装了大量的信息采集传感器, 在数据库中保存了大量的能源运行数据, 这也为基于数据的建模方法提供了良好的基础.

### 3 基于知识蒸馏的系统建模方法

#### 3.1 基于LSTM的学生网络

考虑到LSTM模型内部具有记忆机制的结构, 对输入信息具有较好的长时间记忆功能, 通过控制信息去留的方式, 可以有效地获取输入信息的时序依赖关系, 在时间序列建模上有一定优势. 因此, 采用LSTM

模型来建立本文的学生网络. LSTM网络在循环神经网络 (recurrent neural network, RNN) 的基础上, 引入了门控机制以选择性的保留存储在单元状态的信息, 从而, 更好地保留时序数据的长期依赖关系. LSTM的单元结构如图3所示. 更新门 $i_t$ 决定当前时刻的输入有多少保存到单元状态; 遗忘门 $f_t$ 选择上一时刻的单元状态保存到当前时刻状态; 输出门 $o_t$ 控制当前时刻的状态输出到当前输出.

$$i_t = \delta(w_i[h_{t-1} \ x_t] + b_i), \quad (1)$$

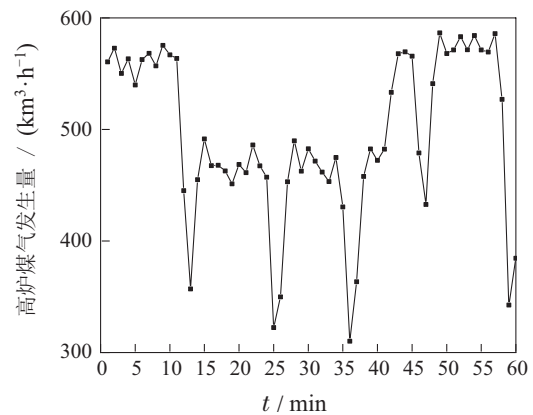
$$f_t = \delta(w_f[h_{t-1} \ x_t] + b_f), \quad (2)$$

$$o_t = \delta(w_o[h_{t-1} \ x_t] + b_o), \quad (3)$$

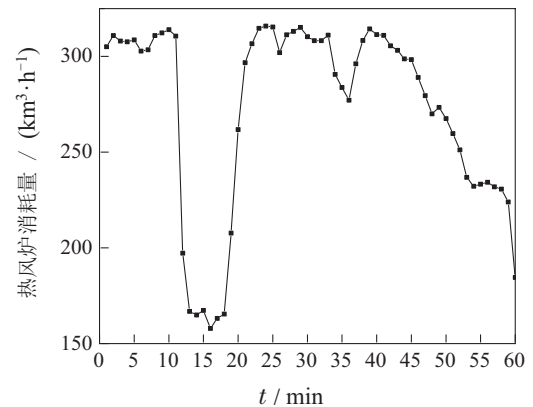
$$c_t = f_t c_{t-1} + i_t (\tanh(w_c[h_{t-1} \ x_t])), \quad (4)$$

$$h_t = o_t \tanh c_t, \quad (5)$$

其中:  $\delta$ 为激活函数;  $c_t, c_{t-1}$ 为 $t$ 时刻与 $t-1$ 时刻的单元状态;  $h_t, h_{t-1}$ 为 $t$ 时刻与 $t-1$ 时刻的隐藏层状态;  $x_t$ 为 $t$ 时刻的输入;  $w_f, w_i, w_c, w_o$ 均为自学习权重矩阵;  $b_f, b_i, b_c, b_o$ 均为偏置项.



(a) BFG发生量



(b) 热风炉消耗BFG流量

图2 典型高炉煤气产消数据特征

Fig. 2 Typical blast furnace gas production and consumption data characteristics

#### 3.2 基于seq2seq模型的教师网络

为了提高训练过程的拟合效果, 从而更好地学习

时间序列数据中的内部特征, 考虑到seq2seq模型在LSTM模型的基础上采用了编码器与解码器的结构, 其编码器与解码器分别具有输入特征提取和输出预测值的功能, 相对于单独的LSTM模型, 编码器可以学到更加完整时序信息, 并将提取的特征信息全部传送给输出端, 具有更强的特征提取能力. 因此, 采用基于LSTM的seq2seq模型来构建教师网络. seq2seq模型结构如图4所示.

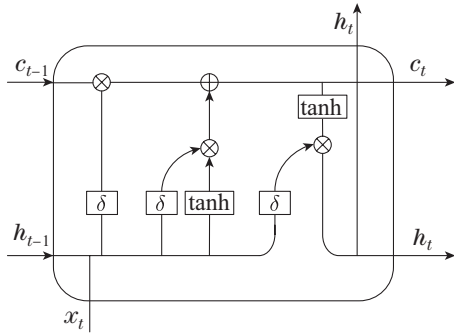


图3 LSTM单元结构图  
Fig. 3 LSTM cell structure

该模型由编码器和解码器两个部分组成, 编码器采用了LSTM单元对输入进行编码, 将每个时间步的值与上一个时间步的单元状态 $c$ 和隐藏层状态 $h$ 一起输入到下一个LSTM单元中, 直到生成最后的状态向量 $(h, c)$ 作为解码器的输入. 解码器由LSTM单元和全连接层组成, 每个时间步将状态向量 $(h, c)$ 与上一个时刻的预测值 $y_{t-1}$ 作为解码器的输入状态, 全连接层计算激活函数以确定预测值.

给定输入序列  $\{x_1, x_2, x_3, \dots, x_n\}$  和输出序列  $\{y_1, y_2, y_3, \dots, y_T\}$ , 模型通过输入序列求解输出序列的概率, 并选择概率最大的输出作为模型的预测结果. 在输入  $\{x_1, x_2, x_3, \dots, x_n\}$  情况下输出  $\{y_1, y_2, y_3, \dots, y_T\}$  的概率如式(6)所示:

$$p(y_1, y_2, y_3, \dots, y_T | x_1, x_2, x_3, \dots, x_n) = \prod_{t=1}^T p(y_t | y_1, y_2, \dots, y_{t-1}, c_t), \quad (6)$$

解码器 $t$ 时刻的概率分布可用式(7)表示, 即

$$p(y_t | y_1, y_2, \dots, y_{t-1}, c_t) = g(h_t, y_{t-1}, c_t), \quad (7)$$

其中:  $g$ 为LSTM单元的计算规则;  $c_t$ 为 $t$ 时刻的单元状态;  $h_t$ 为 $t$ 时刻的隐藏层状态.

### 3.3 融入教师模型中间特征的知识蒸馏策略

由于教师网络和学生网络之间存在的容量和结构差异, 导致二者的特征提取能力不同. 因此, 本文将教师网络提取的中间层特征作为学生模型中间层特征的提示(Hint), 可以使学生模型学习到教师模型的中间特征表达能力, 从而, 使得学生模型的预测性能可以接近, 甚至超越教师模型.

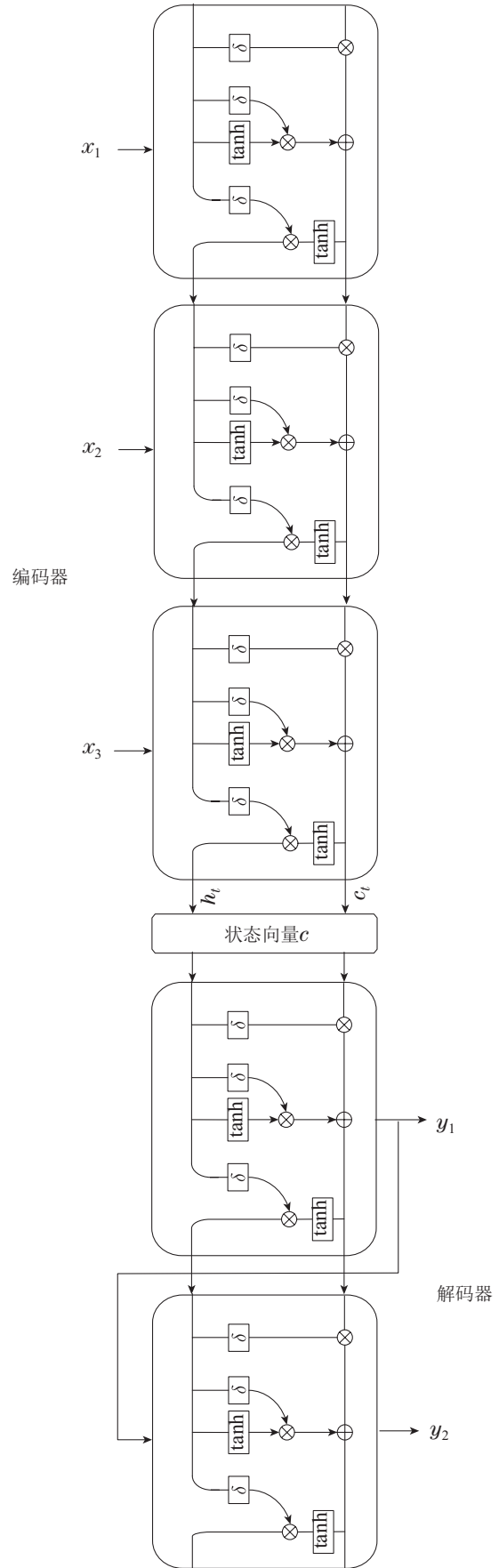


图4 Seq2seq模型结构

Fig. 4 Structure of the seq2seq model

选择教师网络编码器部分的单元状态 $c$ 和隐藏状

态 $h$ 作为教师模型的提示层,对学生网络的隐藏层进行引导.本文的中间层特征蒸馏方法如图5所示.设 $(X, y)$ 表示为一个训练样本集,  $X = (x_1, x_2, x_3, \dots, x_k)$ 为其输入,则中间层蒸馏损失 $L_{\text{Hint-KD}}$ 如式(11)所示:

$$H_X^T = \{H_{1,x}^T, H_{2,x}^T, \dots, H_{m,x}^T\}, \quad (8)$$

$$H_X^S = \{H_{1,x}^S, H_{2,x}^S, \dots, H_{m,x}^S\}, \quad (9)$$

$$\bar{h}_{i,X}^T = \frac{1}{m} \sum_{k=0}^m H_{i,x_k}^T; \quad \bar{h}_{i,X}^S = \frac{1}{m} \sum_{k=0}^m H_{i,x_k}^S, \quad (10)$$

$$L_{\text{Hint-KD}} = \sum_{x \in X} \left\| \frac{\bar{h}_{i,X}^T}{\|\bar{h}_{i,X}^T\|_2} - \frac{\bar{h}_{i,X}^S}{\|\bar{h}_{i,X}^S\|_2} \right\|_2^2, \quad (11)$$

其中: $m$ 为选择教师网络与学生网络的隐藏层数; $H_X^T, H_X^S$ 为输入 $X$ 对应的教师网络和学生网络的中间层; $\bar{h}_{i,X}^T, \bar{h}_{i,X}^S$ 为对教师网络和学生网络的中间层进行均值池化后的向量.

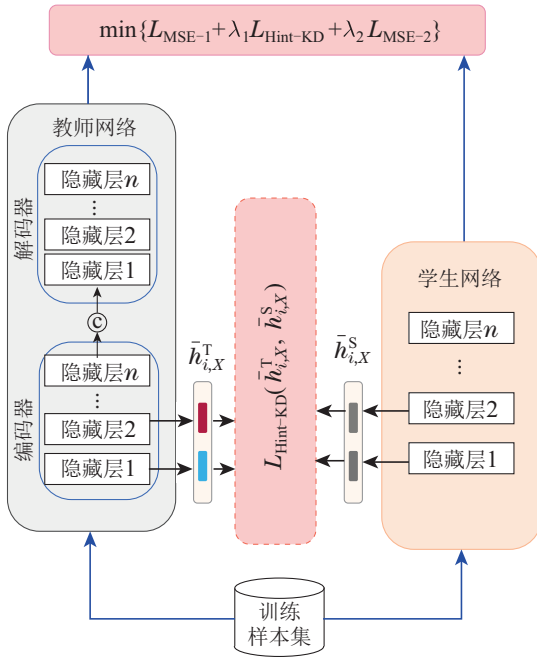


图5 中间层特征蒸馏方法示意图

Fig. 5 Diagram of the characteristic distillation method for the middle layer

传统的分类模型中间层特征蒸馏的损失函数一般表示为原始交叉熵损失、输出特征损失与中间层蒸馏损失相结合,即式(12)

$$\text{Loss} = \lambda_1 L_{\text{CE}} + \lambda_2 L_{\text{KD}} + \lambda_3 L_{\text{Hint-KD}}, \quad (12)$$

其中 $\lambda_1, \lambda_2, \lambda_3$ 为最小化损失函数的超参数.

输出特征蒸馏损失通过温度变量 $T$ 来引入负标签信息,从而软化输出概率分布来训练学生网络.然而,对于时间序列回归模型,其输出是连续无界变量而非离散变量,因此,教师模型无法向学生模型提供负标签概率信息.

针对该问题,本文建立了新的损失函数,即将数据

的原始均方损失与中间层蒸馏损失相结合,如式(13)所示:

$$\text{Loss} = L_{\text{MSE-1}} + \lambda_1 L_{\text{Hint-KD}} + \lambda_2 L_{\text{MSE-2}}, \quad (13)$$

其中: $L_{\text{MSE-1}}$ 为每个batch真实值与预测值的损失函数,并在每个batch进行梯度下降并更新权重; $L_{\text{Hint-KD}}$ 与 $L_{\text{MSE-2}}$ 为每个epoch计算得到的中间层损失与均方损失,并在每个epoch进行梯度下降并更新权重; $L_{\text{MSE-2}}$ 代替传统的输出特征蒸馏损失 $L_{\text{KD}}$ 对学生网络的输出进行惩罚; $\lambda_1, \lambda_2$ 为调节损失函数的超参数.

综上,本文所提的基于知识蒸馏的高炉煤气系统建模方法流程如图6所示,步骤如下:

**步骤1** 构建基于LSTM的seq2seq模型教师网络,根据样本数据集训练教师网络.

**步骤2** 构建基于LSTM模型的学生网络.

**步骤3** 将训练数据输入到学生模型中,学生模型的预测输出与真实值计算均方差损失,进行梯度下降并更新权重,并保存学生网络中间特征,将训练数据输入到训练好的教师模型中,得到并保存教师网络的中间特征.

**步骤4** 本次训练轮次结束后执行步骤5,否则,循环步骤3.

**步骤5** 将提取的教师网络的中间特征与学生模型的中间特征,计算中间层蒸馏损失,并计算学生模型的预测输出与真实值计算均方差损失和,构成总损失函数,进行梯度下降并更新权重.循环步骤3-4直至训练结束.

## 4 企业实际数据验证与分析

为了验证本文所提方法的有效性,选取国内某钢铁企业高炉煤气系统的实际运行数据进行实验验证,数据采集频率为1个点/分钟.采用Python3.7作为编程语言,初始学习率设为0.0001, batch-size为1,训练轮次(epoch)为80,优化器选择为Adam算法.采用基于LSTM的seq2seq方法、LSTM方法、LSSVM方法作为对比实验,并采用均方根误差(root mean square error, RMSE)和平均绝对百分比误差(MAPE)作为精度评估指标,其计算公式分别如式(14)-(15)所示:

$$\text{RMSE} = \frac{1}{n} \sqrt{\sum_{i=1}^n (Y_i - F_i)^2}, \quad (14)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{Y_i - F_i}{Y_i} \right|, \quad (15)$$

其中: $n$ 是预测数据点的个数; $Y_i$ 为实际的观测值; $F_i$ 为模型的预测值.

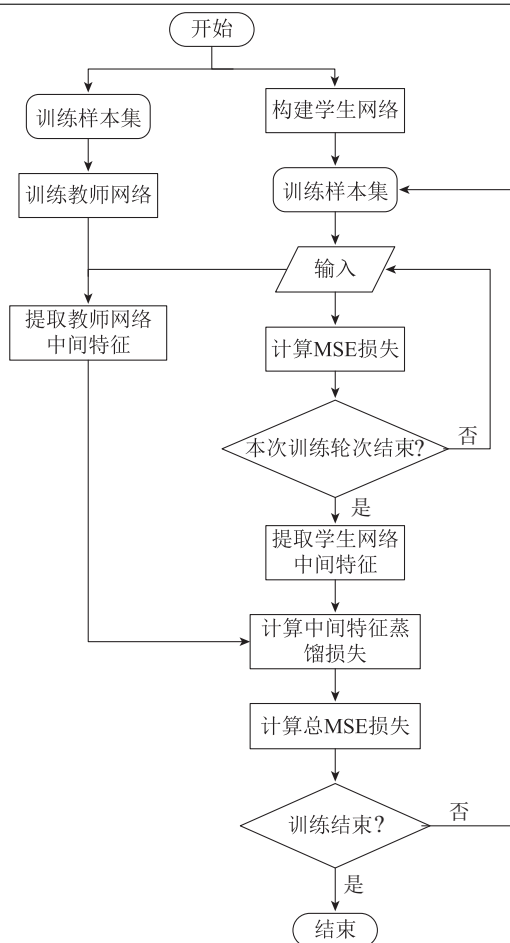


图6 建模方法流程框图

Fig. 6 Flowchart of the modelling method

### 4.1 高炉煤气发生流量建模

在高炉煤气发生流量建模的过程中,影响本文方法的预测效果的参数主要有LSTM模型的输入维度、隐藏层数、调节知识蒸馏损失函数的 $\lambda_1, \lambda_2$ . 调节参数 $\lambda_1, \lambda_2$ 可以调整中间特征蒸馏在整个训练中的所占的比例,  $\lambda_1$ 越大中间特征蒸馏在训练中所占的比重也越大, 由于 $L_{Hint-KD}$ 在每个epoch计算梯度更新权重, 调大 $\lambda_1$ 可以让中间特征蒸馏的损失函数更快地迭代到最小值. 通过交叉验证得到的高炉煤气发生量建模最佳参数如表1所示.

表1 参数优化结果

Table 1 Parameter optimization results

输入维度	隐藏层数	$\lambda_1$	$\lambda_2$
50	20	300	5

选取2000个连续时间序列作为训练样本, 预测时长为60 min. 如图7所示为高炉煤气发生量在各预测方法下的预测效果对比图, 表2则是各方法的预测精度统计结果.

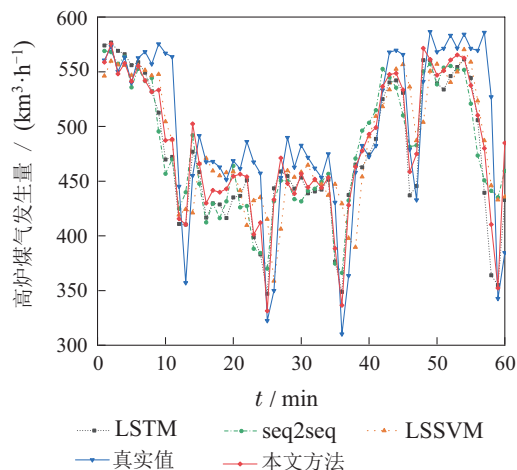


图7 各方法的预测效果比较

Fig. 7 Comparison of the prediction results of each method

表2 各方法的预测精度统计

Table 2 Prediction accuracy statistics by method

预测方法	RMSE	MAPE/%
LSTM	36.5734	7.6168
seq2seq	40.5323	8.6635
LSSVM	32.1050	7.0594
本文方法	30.7948	6.4896

由表2可知, 在高炉煤气发生量的预测建模过程中, LSTM方法虽然有较好的长时间记忆功能, 在时间序列建模上有一定优势, 但由于工业数据含有大量噪声, 该方法在高炉煤气发生量预测过程精度较低, 无法满足现场精度要求. 基于LSTM的seq2seq方法虽然有较好的特征提取能力, 但由于seq2seq模型的编码器将特征提取到的信息全部压缩再进行解码, 会出现信息损失, 因此, 其精度仍无法满足现场精度要求. LSSVM方法由于算法局部搜索能力较差, 搜索精度不高, 故其精度无法满足现场精度要求. 而本文方法综合考虑了融合教师模型的知识蒸馏策略, 通过结合LSTM模型处理时间序列的能力和基于LSTM的seq2seq模型的特征提取能力, 对于预测结果的精度有较大的提高, 同时, 预测稳定性也更高. 因此, 本文所提方法对于高炉煤气发生流量的建模问题具有较高的精度与稳定性.

### 4.2 高炉煤气消耗流量建模

在高炉煤气消耗流量建模的过程中, 影响本文方法的预测效果的参数与高炉煤气发生流量建模的过程相同, 即LSTM模型的输入维度、隐藏层数、调节知识蒸馏损失函数的 $\lambda_1, \lambda_2$ . 通过交叉验证得到的高炉煤气消耗量建模最佳参数如表3所示.

同样选取2000个连续时间序列作为训练样本, 预测时长为60 min. 如图8所示为热风炉高炉煤气消耗

量在各预测方法下的预测效果对比图,各方法的预测精度统计如表4所示。

表3 参数优化结果

Table 3 Parameter optimization results

输入维度	隐藏层数	$\lambda_1$	$\lambda_2$
50	25	500	5

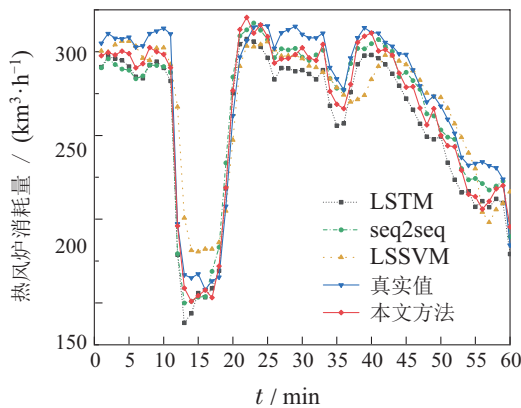


图8 各方法的预测效果比较

Fig. 8 Comparison of the prediction results of each method

由表4可知,在高炉煤气消耗量预测建模过程中,与LSTM方法、LSSVM方法和单纯的seq2seq模型相比,本文方法仍具有最优的精度与稳定性。因此,综上所述,本文方法的学生模型通过知识蒸馏融合教师模型,学习到了教师模型中良好的特征提取能力,从而达到或好于未经过知识蒸馏和教师模型的预测效果,对于高炉煤气发生量和消耗量的预测建模都具有实用价值。

表4 各方法的预测精度统计

Table 4 Prediction accuracy statistics by method

预测方法	RMSE	MAPE/%
LSTM	8.0286	6.7259
seq2seq	11.1385	4.2401
LSSVM	13.9640	5.7297
本文方法	11.1050	4.2146

## 5 结语

针对高炉煤气系统的产消过程存在高时变性和不确定性等特点,本文提出了一种基于知识蒸馏的高炉煤气系统建模方法。建立了基于LSTM的seq2seq模型作为教师网络,提高序列样本的内部特征提取能力,同时,构建了LSTM模型作为学生网络,提出了融入教师模型中间特征的知识蒸馏策略,将教师网络提取的中间层特征迁移至学生网络中,并建立了考虑中间特征蒸馏损失与预测均方损失的损失函数,在每个batch对输入样本预测计算均方损失,并在每个训练轮

次对整个输入样本集计算中间特征损失与预测均方损失,从而,对知识蒸馏过程及预测偏差进行评估。典型炼铁高炉煤气系统实际运行数据的实验验证表明,本文所提的基于知识蒸馏的建模方法具有更高的精度与稳定性,可为后续的能源系统优化调度提供有力支撑。

## 参考文献:

- [1] ZHANG Chunxia, QI Yuanhong, YAN Dingliu, et al. Energy-saving and environmental protection of ironmaking system in China. *Iron & Steel*, 2006, 41(11): 1-5.  
(张春霞, 齐渊洪, 严定鏊, 等. 中国炼铁系统的节能与环境保护. 钢铁, 2006, 41(11): 1-5.)
- [2] WANG Xindong, TIAN Jinglei, SONG Chengyuan, et al. Innovative practice technology and outlook in large iron and steel enterprise green manufacturing. *Iron & Steel*, 2018, 53(2): 6-14.  
(王新东, 田京雷, 宋程远, 等. 大型钢铁企业绿色制造创新实践与展望. 钢铁, 2018, 53(2): 6-14.)
- [3] LI Hongjuan, XIONG Wenzhen. Prediction and optimal operation on byproduct gas system in steel enterprises. *Iron & Steel*, 2016, 8(51): 90-98.  
(李红娟, 熊文真. 钢铁企业副产煤气预测及优化调度. 钢铁, 2016, 8(51): 90-98.)
- [4] HE F, ZHANG Q Z, LEI J S, et al. Energy efficiency and productivity change of China's iron and steel industry: Accounting for undesirable outputs. *Energy Policy*. 2013, 54: 204-213.
- [5] HAO Juxian, ZHAO Xiancong, HAN Yuzhao, et al. Medium-term prediction model for byproduct gas consumption in hot blast stove. *China Metallurgy*, 2018, 28(2): 17-22.  
(郝聚显, 赵贤聪, 韩玉召, 等. 热风炉煤气消耗量中期预测模型. 中国冶金, 2018, 28(2): 17-22.)
- [6] ZHANG Q, LI H, MA J L, et al. Dynamic forecasting and optimal scheduling of by-product gases in integrated iron and steel works. *Journal of Iron and Steel Research International*. 2019, 26: 538-546.
- [7] XU Huayan, MA Jialin. Composite prediction model of blast furnace gas based on data driven. *China Metallurgy*, 2019, 29(7): 56-60.  
(徐化岩, 马家琳. 基于数据驱动的高炉煤气复合预测模型. 中国冶金, 2019, 29(7): 56-60.)
- [8] ZHAO J, WANG W, LIU Y, et al. A two-stage online prediction method for a blast furnace gas system and its application. *IEEE Transactions on Control Systems Technology*, 2011, 19(3): 507-519.
- [9] LIU Ying, LIU Ruoyu, ZHAO Jun, et al. Variational inference-based echo state network ensemble model for intervals prediction of industrial data. *Control Theory & Applications*, 2018, 35(8): 1066-1073.  
(刘颖, 刘若愚, 赵珺, 等. 基于变分推理回声状态网络集成模型的工业数据区间预测. 控制理论与应用, 2018, 35(8): 1066-1073.)
- [10] SUN W Q, WANG Z H, WANG Q. Hybrid event-, mechanism- and data-driven prediction of blast furnace gas generation. *Energy*, 2020, 199: 117497.
- [11] ZHAO Jun, DU Yanan, SHENG Chunyang. Kernel-based method for predicting online gas flow interval in metallurgical enterprises. *Control Theory & Applications*, 2013, 30(10): 1274-1279.  
(赵珺, 杜雅楠, 盛春阳. 基于核的冶金煤气流量在线区间预测. 控制理论与应用, 2013, 30(10): 1274-1279.)
- [12] HOCHREITER S, SCHMIDHUBER J. Long short-term memory. *Neural Compute*, 1997, 9(8): 1735-1780.
- [13] YANG Li, WU Yuxi, WANG Junli, et al. Research on recurrent neural network. *Journal of Computer Applications*, 2018, 38(S2): 1-6, 26.  
(杨丽, 吴雨茜, 王俊丽, 等. 循环神经网络研究综述. 计算机应用, 2018, 38(S2): 1-6, 26.)

- [14] LIU J, SHAHROUDY A, XU D, et al. Skeleton-based action recognition using spatiotemporal LSTM network with trust gates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 40(12): 3007 – 3021.
- [15] BAO Xiangjun, WENG Sihao, CHEN Guang, et al. Comparison on multi-step prediction of blast furnace gas generation based on LSTM/SARIMA time series model. *Iron & Steel*, 2022, 57(9): 166 – 172. (包向军, 翁思浩, 陈光, 等. 基于时序模型的高炉煤气发生量多步预测对比. 钢铁, 2022: 57(9): 166 – 172.)
- [16] HE Dongfeng, GUAN Zhulin, HU Zhengbiao. Forecast of blast furnace gas production based on classification and ARIMA-WT-LSTM model. *Metallurgical Industry Automation*, 2022, 46(2): 103 – 109. (贺东风, 官竹林, 胡正彪. 基于分类和ARIMA-WT-LSTM模型的高炉煤气产生量预测. 冶金自动化, 2022, 46(2): 103 – 109.)
- [17] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 2014.
- [18] PALANGI H, WARD R, DENG L. Distributed compressive sensing: A deep learning approach. *IEEE Transactions on Signal Processing*, 2016, 64(17): 4504 – 4518.
- [19] HAN Z Y, LIU Y, ZHAO J, et al. Real time prediction for converter gas tank levels based on multi-output least square support vector regressor. *Control Engineering Practice*, 2012, 20(12): 1400 – 1409.

#### 作者简介:

**金 锋** 副教授, 研究方向为工业综合能源系统建模与优化调度,

E-mail: jin\_feng@dlut.edu.cn;

**陈薇琳** 硕士研究生, 研究方向为工业综合能源系统预测建模,

Email: chenweilin@mail.dlut.edu.cn;

**赵博识** 工程师, 研究方向为钢铁企业能源管理及动力运行管理,

Email: 172833391@qq.com;

**赵 珺** 教授, 博士生导师, 研究方向为工业人工智能、工业能源系统优化、能源云平台综合管控、大数据分析机器学习、复杂工业系统建模, E-mail: zhaoj@dlut.edu.cn;

**王 伟** 教授, 博士生导师, 研究方向为复杂工业过程的建模、控制与优化、流程工业生产计划与优化调度、知识自动化等, E-mail: wangwei@dlut.edu.cn.