

## 基于工业大数据的厚板板形预报系统研发

马宇飞<sup>1</sup>, 刘长鑫<sup>1</sup>, 孔伟<sup>2</sup>, 丁进良<sup>1</sup>

(1. 东北大学流程工业综合自动化国家重点实验室, 辽宁 沈阳 110819;

2. 宝山钢铁股份有限公司中央研究院, 上海 201900)

**摘要:** 厚板板形是衡量厚板产品质量的重要指标之一, 生产中最终板形的及时预报对于调整厚板生产操作与控制具有重要的意义。实际工业生产中, 厚板数据具有耦合信息多、冗余信息量大、数据呈现多源异构性等特点, 结合厚板板形预报的需求, 设计并开发了厚板板形预报系统。利用数据转存功能, 对工业大数据进行数据过滤和数据预处理, 去除数据中的耦合信息和冗余变量。利用 LSTM 神经网络、卷积神经网络以及 3D 卷积神经网络对不同维度的数据分别提取数据特征, 基于最大互信息系数将特征进行融合建立集成学习预报模型, 有效地解决了多源异构数据所带来的建模困难。采用国内某厚板生产线的实际工业数据进行验证, 结果证明了所开发系统的有效性。

**关键词:** 厚板板形; 预报模型; 多源异构数据; 系统开发

**中图分类号:** TP18

**文献标识码:** A

**doi:** 10.11959/j.issn.2096-3750.2021.00239

## Research and development of thick plate shape prediction system based on industrial big data

MA Yufei<sup>1</sup>, LIU Changxin<sup>1</sup>, KONG Wei<sup>2</sup>, DING Jinliang<sup>1</sup>

1. State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China

2. Central Research Institute, Baoshan Iron & Steel Co., Ltd., Shanghai 201900, China

**Abstract:** Thick plate shape is one of the important indicators to measure the quality of thick plate products. The timely prediction of the final plate shape in production is of great significance for adjusting the operation and control of thick plate production. In actual industrial production, thick plate data has many characteristics, such as multiple coupling information, large amount of redundant information, and multi-source heterogeneity of data. Combining the needs of thick plate shape prediction, a thick plate shape prediction system was designed and developed. The data dump function was used to filter and preprocess the industrial big data to remove the coupling information and redundant variables in the data. LSTM neural network, convolutional neural network and 3D convolutional neural network were used to extract data features from data of different dimensions, and the features were fused based on the maximum mutual information coefficient to establish an integrated learning prediction model, which effectively solved the modeling difficulties caused by multi-source heterogeneous data. The actual industrial data of a domestic thick plate production line was used for verification, and the results showed the effectiveness of the developed system.

**Key words:** thick plate shape, prediction model, multi-source heterogeneous data, system development

收稿日期: 2021-03-24; 修回日期: 2021-06-01

通信作者: 丁进良, jlding@mail.neu.edu.cn

基金项目: 国家重点研发计划 (No.2018YFB1701104); 辽宁省兴辽英才计划 (No.XLYC1808001); 辽宁省科技技术项目 (No.2020JH2/10500001)

**Foundation Items:** The National Key Research and Development Program of China (No.2018YFB1701104), The Xingliao Talent Plan of Liaoning Province (No.XLYC1808001), The Science and Technology Program of Liaoning Province (No.2020JH2/10500001)

## 1 引言

厚板是重要的钢材产品,被广泛应用于基础设施、船舶、工程机械、容器、能源和建筑等行业,在国民经济建设以及国防建设中占有重要的地位<sup>[1]</sup>。厚板产品品种繁多<sup>[2]</sup>、质量要求高、应用范围广,世界各国都把厚板的品种、质量作为衡量一个国家钢铁工业综合水平的尺度。

对于生产后的厚板产品,主要从尺寸、板形、性能、表面和内质 5 个方面对产品质量进行评价,而板形作为与生产过程关系最密切<sup>[3]</sup>、控制最复杂的一种性能指标,受到人们广泛的关注<sup>[4]</sup>。板形本质上是描述钢板内部残余应力的分布,直观上是指钢板的翘曲程度,只要钢板内部存在残余应力,即可称之为板形不良。因此,在厚板的生产过程中,及时地预报最终板形质量并识别最终板形故障类别,能够及时对生产过程中的控制参数进行调整,修正产品质量,提高成材率。

Laugwitz 等<sup>[5]</sup>通过分析矫直机中轧辊位置对板形的影响,建立有限元模型,并以此提供适当的调平策略,但该方法只能应用到特定的矫直机中,预测显性的板形缺陷。Schausberger 等<sup>[6]</sup>针对厚板的轧制过程,提出一种减少轧制过程中轮廓误差的反馈控制方法,并建立数学模型描述厚板的角速度与轮廓演变之间的关系,同时应用到实际的工厂中,验证算法能够明显地降低厚板的轮廓误差。Yu 等<sup>[7]</sup>通过轧制过程中钢板的温度变化,提出了一种梯度温度轧制的新颖工艺,通过建立厚板轧制过程的有限元模型,计算轧件的温度场及应力,实验表明该方法相比于传统的均匀温度轧制能取得更优异的性能。上述方法从机理的角度出发,分析了轧制过程以及矫直过程对厚板质量的影响,但只针对某一个单一工序进行分析,使结果过于局限,缺少对厚板生产过程的全流程工序进行分析。

工业大数据的迅速发展进一步促进了厚板板形质量预报技术的创新与发展。工业大数据及其应用将厚板生产全流程整个生命周期的工业数据进行综合分析<sup>[8]</sup>,利用机理与数据驱动相结合的方式建立预报模型,能够更加高效准确地对最终板形质量进行预报,使整个厚板生产流程向信息化、智能化的方向迈进<sup>[9]</sup>。

厚板生产过程中工序、设备复杂繁多,工序与工序之间的数据分布、数据类型、数据结构往往不

尽相同,时序数据、图像数据、视频数据相互混杂,并且由于各个工序中控制变量多,控制关系复杂<sup>[10]</sup>,因此厚板生产过程中的数据大多表现为高维异构数据。

针对高维异构数据,Guo 等<sup>[11]</sup>利用集成学习的方法,针对不同维度的异构数据基于深度学习方法分别建立分类模型,并利用对话状态追踪(DST, dialogue state tracking)将不同模型的输出结果进行统计得出分类结果,并利用医学图像数据进行验证,取得了良好的效果。Nguyen 等<sup>[12]</sup>提出了一种在异构集成系统中组合多个分类器的新方法,将基本分类器产生的预测多样性通过基于间隔的信息颗粒进行量化,然后通过考虑区间的边界和长度生成决策模型。上述方法需要针对不同高维数据建立多个预报模型,然而这会占用大量的计算资源,对在复杂的厚板生产现场的实际应用十分不友好。

为了满足目前厚板生产过程的实际需要,本文利用厚板全流程生产过程中的工业大数据,设计并开发了一套针对厚板最终板形质量的预报系统。基于深度学习算法对厚板生产过程中的高维异构数据进行特征提取,并将数据特征进行融合,建立对最终板形质量的预报模型,对指定的厚板产品进行预报,最终得到针对厚板板形的预报结果,为评估厚板最终板形质量、定位厚板生产过程中的数据异常进行服务,进而不断修正厚板产品质量,提高产品合格率。本文主要介绍了厚板板形质量预报系统的整体架构的设计与开发,并结合实际的厚板板形预报过程进行了验证分析。

## 2 厚板生产过程

由于厚板产品的特殊性,其生产工艺与普通热连轧过程有所区别<sup>[13]</sup>,生产过程更为复杂。厚板生产流程示意图如图 1 所示,显示了中国一家钢铁企业的厚板生产线。在生产过程中,板坯首先装入加热炉进行加热,经过 4 段加热后,板坯升温到所需温度,之后板坯进入轧制工序,经过粗轧机与精轧机的反复轧制后,生产出与要求尺寸和板形质量相符的钢板。之后钢板进入冷却工序,经过加速冷却装置,钢板的温度快速下降,保证钢板的韧性、硬度能够达到要求。之后钢板进入矫直工序,通过热矫直机的反复矫直,尽量消除钢板的不平直度。

在厚板的生产过程中,由于厚板产品由多个工序和设备的物理加工以及化学反应制造而成,不同

工序以及设备采集的指标变量是不同的，并且采集的数据的存储形式也是不同的，因此厚板的过程数据往往呈现多源异构性。数据主要包括一维时序数据、二维图像数据以及三维数据块。其中，一维时序数据包括厚板在加热炉内各个加热阶段的钢板平均温度、表面温度、中心温度、辊道温度、加热时间，轧制过程中轧制力、弯辊力、出口厚度、出口温度，冷却过程中水冷温度、环境温度、钢板测量温度、冷却时间等指标的统计量。二维图像数据包括厚板在加热炉内 6 个热电偶测量的二维炉温数据、冷却结束后厚板在宽度和长度方向上测量的二维温度数据。三维数据块包括在轧制过程中每个轧制道次下测量的厚板长度方向的轧制力、弯辊力、扭矩、辊缝、出口温度、出口平直度等指标数据。

在实际生产过程中，描述钢板最终板形质量的标签数据为板形仪测量的二维厚度矩阵，厚板板形质量示意图如图 2 所示，厚度矩阵越平整，表明最终板形质量越好。依据二维厚度矩阵数据的分布情况，将板形异常分为 4 种类别，分别是头尾翘曲，即在钢板的头尾两端，厚度值超限；厚度异常，即钢板中间部分的厚度值超限；中浪，即在钢板长度方向，厚度值的大小交替变换；左右边浪，即在钢板的左右两侧，厚度值的大小交替变换。

厚板的最终板形质量出现异常本质上是钢板内部应力分布不均导致的，而在厚板的生产过程中影响钢板应力分布的因素主要有钢板温度和轧制力。在加热过程中，加热炉内的炉温、钢板的表面温度和中心温度都会对钢板的整体温度分布产生

影响，进而影响钢板内部应力的分布情况；在轧制过程中，每个道次粗轧机和精轧机对钢板的轧制力、弯辊力、扭矩等变量都会对钢板内部应力产生影响；在冷却过程中，冷却前钢板的温度分布、冷却过程中的水流量、冷却后钢板的温度分布都会对钢板的内部应力分布产生影响；在热矫直过程中，每个矫直道次热矫直机对钢板的压力、矫直力等因素都会对钢板的内部应力产生影响。

因此，本文设计并开发了板形预报系统，针对厚板在加热、轧制、冷却、热矫直过程中的关键变量进行分析，对最终的板形质量进行预报。

### 3 厚板板形预报系统架构

#### 3.1 厚板板形预报系统功能

厚板板形预报系统主要具备以下几个功能：数据转存、数据预处理、预报模型训练、模型预报。厚板板形预报系统主要目的是针对厚板生产过程中产生的关键变量和数据，基于深度学习的智能建模算法，建立对厚板最终板形质量的预报模型，当用户输入钢板 ID 后，系统调取该钢板在数据库中的过程数据，使轧钢企业能够在钢板的实际生产过程中提早对最终板形质量进行预报，当预报值出现异常时，使企业能够及时调整变量参数，改善厚板的最终板形质量，提高钢板的合格率，减小轧钢成本，提高企业的经济效益。

1) 数据转存功能。数据转存功能主要是在钢板的实际生产过程中，当钢板数据更新后，对原始数据库中的钢板数据进行数据截取与数据筛选，去

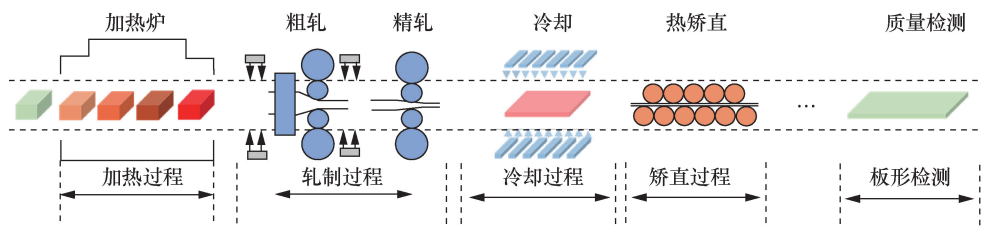


图 1 厚板生产流程示意图

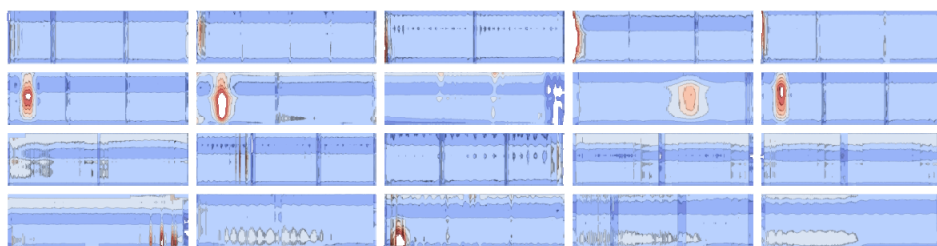


图 2 厚板板形质量示意图

除工业生产过程中的错误数据与冗余数据,将预报模型所需要的准确的钢板数据保存到数据库中的数据转存表中,以便后续数据预处理过程的数据调用。由于工业大数据维度高、冗余数据量大的特点,数据预处理过程往往需要较长的时间进行数据截取与数据筛选,因此,设计数据转存模块提取工业大数据中的准确数据信息,并保存在数据转存表中,能够有效地节省数据预处理过程的时间,使得预报系统运行更加高效且快速。数据转存过程主要实现两个功能:生产过程数据转存、标签数据转存。生产过程数据转存指的是对厚板加热过程、轧制过程、冷却过程、热矫直过程的数据截取与数据筛选,具体指的是针对加热炉中的炉温数据,依据板还在加热炉内的加热时间,提取对应时间内的炉温数据,组成二维变量保存到数据转存表中。针对冷却过程中测量的温度数据,结合钢板的位置信息与温度跳变区域,去除其中的冗余数据,截取描述钢板实际温度的数据,组成一维温度变量保存到数据转存表中。针对轧制过程中测量的轧制力、弯辊力等数据以及热矫直过程中的矫直力、温度等数据,由于存在空间尺度、时间尺度的多尺度数据,因此定义 3 种维度,即道次维度(描述随钢板生产时间的推进,钢板每个轧制道次所产生的数据)、指标维度(描述在空间尺度上,每个道次下同时产生的控制指标)、测量维度(描述每种指标下钢板的原始测量数据)组成一个三维数据块保存到数据转存表中。标签数据转存指的是针对钢板的二维厚度矩阵,判断其板形是否出现异常,以及板形的异常类别,并将结果保存在数据转存表中。

2) 数据预处理功能。在工业现场中存在许多未知的扰动,使得工业传感器的采样频率以及采样周期难以保证完全一致,并且不同钢板样本的尺寸不同,使得采集到的钢板样本之间的数据长度不同,因此需要设计数据预处理模块统一样本数据的长度。数据预处理模块从数据转存表提取待处理的钢板样本数据,经过数据预处理后统一数据长度,并将预处理后的数据用于模型训练与模型预报功能。

3) 预报模型训练功能。在轧钢工业现场中,每天生产的产品数量众多,并且厚板产品的型号规格和数据分布并不是一成不变的,会根据订单要求以及生产环境的变化而发生变化,因此系统需要根据不同的生产情况训练不同的预报模型,从而提升

对厚板最终板形质量的预报精度。首先,根据前端页面中输入的参数选取作为训练集和验证集的钢板数据,并对数据进行预处理;对预处理后的钢板样本数据进行归一化,使模型的训练过程更容易收敛;然后,按照数据维度的不同,将数据转换为不同的数据结构作为模型的输入,并利用设计好的模型结构对预报模型进行训练;待训练完成后,将训练好的预报模型保存在数据库的模型转存表中,方便板形预报时能够及时调用预报模型。

4) 模型预报功能。模型预报功能主要是在厚板的实际生产过程中,从数据转存表找到钢板对应的加热过程数据、轧制过程数据、冷却过程数据以及热矫直过程数据,将其作为预报模型的输入,得到该块钢板最终板形质量异常的概率,使用户在钢板的生产过程中提前了解钢板最终可能包含的板形质量异常,从而及时调整后续工序生产,改善钢板的最终板形质量,进而提高合格率。

### 3.2 厚板板形预报系统架构设计

本文设计的预报系统采用 3 层架构模式,将软件分为:人机交互层、业务逻辑层、数据存储层。系统架构设计如图 3 所示。

人机交互层位于最上层,是系统提供给用户的操作界面,与用户直接进行沟通。用户通过输入信息、单击按钮等操作访问业务逻辑层中封装的各项功能接口,并将系统返回的分析结果和数据以图表的方式展现给用户,从而使用户更方便地应用本系统所提供的数据转存、模型训练、模型预报等功能。人机交互层采用组件化的方式进行开发,将界面中的不同部分划分为一个又一个小组件,通过将不同的组件进行组合来实现一个完整界面的开发。组件与组件之间相互独立,利用协议和数据来进行连接和通信,组件内的数据不会发生耦合,充分降低了代码量,提高了工作效率。

业务逻辑层是连接人机交互层与数据存储层的桥梁,是整个系统中最复杂且应用逻辑最多的一层。业务逻辑层主要负责接收人机交互层中发送的请求,并根据不同的请求调用不同的服务,对数据存储层中的相关数据进行增、删、改、查等操作,并采用数据处理、数据分析、智能建模等算法实现预报系统所提供的各项功能,将数据分析结果返回给人机交互层。业务逻辑层基于微服务进行开发,每个微服务都能够完成单独的功能,根据系统所提供的不同功能将多个微服务进行拼装组合来满足

系统需求。这些微服务可以进行单独开发、编译以及测试，从而提高系统的可复用性以及可扩展性。

数据存储层位于系统的最底层，主要由原始数据库、转存数据库、模型数据库组成。负责存储厚板生产过程中采集的各项工业数据，数据存储层根据业务逻辑层中调用的不同服务为其提供相应的工业数据，并将业务逻辑层的部分数据处理结果保存到相应的转存表中，为后续服务提供数据，提高系统的运行效率。

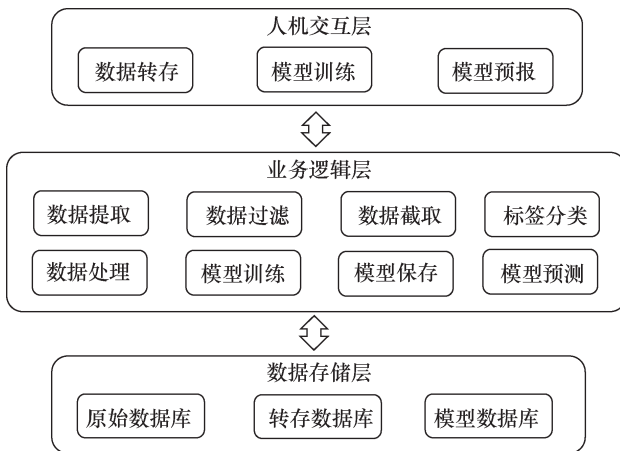


图3 系统架构设计

本系统基于 JavaScript 语言并结合 Vue 框架对人机交互层的功能进行开发，设计并开发出数据转存、模型训练以及模型预报的人机交互界面为用户服务。同时，本系统采用 Python 语言结合 Flask 框架完成对业务逻辑层的开发，并利用 PostgreSQL 数据库保存原始数据以及转存数据。

## 4 厚板板形预报模型设计

### 4.1 数据预处理

由于在实际现场中存在未知扰动，传感器的采样频率以及采样周期难以保证一致<sup>[14]</sup>，因此，不同样本之间的数据向量的长度不同，需要利用数据处理方法统一数据长度。传统的数据处理方法主要有数据插值<sup>[15]</sup>、数据放缩剪裁<sup>[16]</sup>等。

数据插值算法对数据变化过于敏感，工业现场中的数据点之间容易存在波动，导致数据异常点的产生，偏离原始数据信息。

数据放缩剪裁主要针对图像数据，对图像按照像素点进行有效的缩放，统一图片尺寸，或者对图像数据按照统一维度进行剪裁，只保留有效信息，但这种方法不适用于处理工业现场产生的序列数

据以及多维数据。

因此，本系统针对单维序列数据、二维图像数据以及三维数据的异构数据，采用补齐数据特征的处理方法，统一不同样本间的数据长度。针对加热炉温数据，在加热时间维度上填补零点数据，使炉温数据在长度方向上得到统一。针对轧制过程中的控制量和输出量数据，由于厚板产品规格不同导致轧制道次数不同，所以在数据的道次维度上填补零点，统一所有样本的轧制道次数。针对冷却前和冷却后钢板的温度数据，在钢板长度方向上填补零点，补齐样本数据长度。

### 4.2 预报模型设计

由于厚板生产过程会产生多源异构数据<sup>[17]</sup>，本文设计一种数据融合方法对异构数据进行融合建模，预报模型结构框图如图4所示。

将厚板生产过程产生的多源异构数据根据不同的维度划分为不同的数据块，一维数据包括冷却前的温度数据  $I^{(5)}$ ，二维数据包括加热炉的温度数据  $I^{(2)}$ 、冷却后的温度数据  $I^{(6)}$ ，三维数据包括轧制过程的控制变量数据  $I^{(3)}$  和输出变量数据  $I^{(4)}$ 。

为了不丢失隐含在不同数据块之间的关联信息以及时序信息，针对不同的异构数据块分别求取统计量（包含最大值、最小值、均值、方差等数据信息）用以近似表征异构数据块的特征。对于加热炉的温度数据  $I^{(2)}$ ，在加热时间维度上划分为 5 个部分，分别表示加热炉 5 个时间段的温度分布，并分别求取每个部分数据的最大值、最小值、平均值、方差组成一维特征向量  $\mathbf{S}_2$ 。对于轧制过程的控制变量数据  $I^{(3)}$  和输出变量数据  $I^{(4)}$ ，将粗轧阶段看作一个部分，精轧阶段分为前 3 个道次、后 3 个道次以及中间道次 3 个部分，分别求取每个部分数据块的统计量组成一维特征向量  $\mathbf{S}_3$  和  $\mathbf{S}_4$ 。对于冷却前的温度数据  $I^{(5)}$ ，在钢板的长度方向上划分为 5 个部分，分别求取每个部分数据块的统计量组成一维特征向量  $\mathbf{S}_5$ 。对于冷却后的温度数据  $I^{(6)}$ ，将钢板宽度方向分为 3 个部分、长度方向分为 5 个部分，分别求取每个部分数据块的统计量组成一维特征向量  $\mathbf{S}_6$ 。将所有的统计量合并，组成一维向量  $\mathbf{S}=\{\mathbf{S}_2, \mathbf{S}_3, \mathbf{S}_4, \mathbf{S}_5, \mathbf{S}_6\}$ ，将其作为预报模型的又一个输入  $I^{(1)}$ 。

利用统计向量  $\mathbf{S}$ ，求取不同原始数据块与最终板形质量  $\mathbf{Y}$  之间的最大互信息系数。最大互信息系数是度量两个连续型变量之间相关性的系数<sup>[18]</sup>，它通过对连续型变量实施不等间隔的离散化寻优来

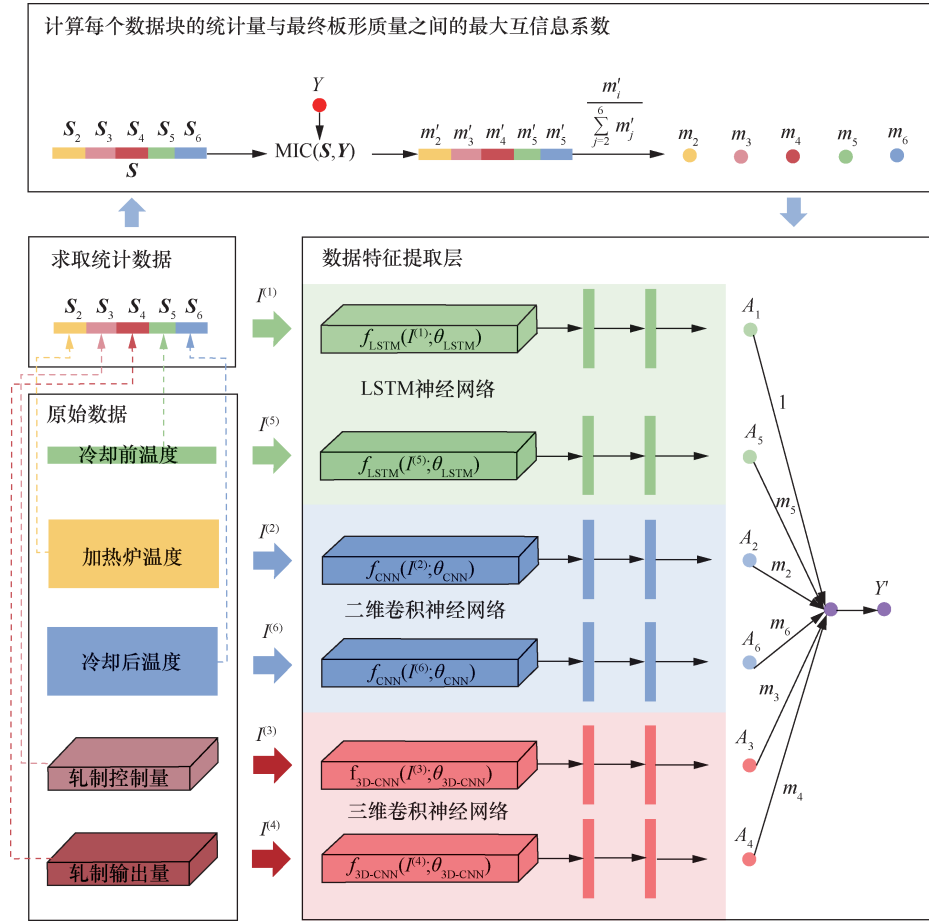


图 4 预报模型结构框图

挖掘非线性关联，并进一步通过标准化校正得到  $MIC(S, Y)$ 。之后根据求取的最大互信息系数，得到原始数据块占有的权重  $\{m_2, m_3, m_4, m_5, m_6\}$ ，即

$$m_i = \frac{m'_i}{\sum_{j=2}^6 m'_j} \quad (1)$$

之后，针对不同维度的原始数据，分别设计特定的神经网络结构提取数据中的特征。其中，针对一维时序数据  $I^{(1)}$  和  $I^{(5)}$ ，数据与数据之间存在时序相关性。因此，采用长短期记忆<sup>[19]</sup> (LSTM, long short-term memory) 网络提取数据特征。LSTM 网络是一种特殊形式的循环神经网络 (RNN, recurrent neural network)，相比于普通的 RNN 结构，LSTM 网络加入了输入门限、遗忘门限和输出门限，能够提取更多的有效信息，并且能够避免梯度消失的问题。通过 10 层的 LSTM 神经元组成提取一维时序特征的神经网络架构，将提取的 570 维统计指标  $I^{(1)}$  与冷却前的一维温度数据  $I^{(5)}$  作为输入，提取隐含在高维数据中的数据特征，即

$$A_j = f_{LSTM}(I^{(j)}; \theta_{LSTM}), j \in \{1, 5\} \quad (2)$$

其中， $A_j$  表示一维数据经过 LSTM 网络后提取到的特征， $f_{LSTM}(\cdot)$  表示 LSTM 网络结构， $\theta_{LSTM}$  表示 LSTM 网络参数的集合。

针对二维图像数据  $I^{(2)}$  和  $I^{(6)}$ ，由于其在横向方向和纵向方向的数据分布都有各自的意义，因此基于 Google 的 Inception V4 CNN 架构进行模型结构的搭建。Inception V4 是 Google 的开源 CNN 模型，它基于大型图像数据库 ImageNet 的数据进行训练。为了从多个尺度提取钢板数据在空间上的特征信息，本文利用多个不同大小的卷积核对输入数据进行识别，同时引入了残差模块使得信息前后向传播更加顺畅，解决了梯度弥散和网络退化的问题。将提取的各个钢板样本的炉温数据  $I^{(2)}$  以及冷却后的二维温度数据  $I^{(6)}$  作为输入，利用卷积神经网络提取二维数据中的数据特征，即

$$A_j = f_{CNN}(I^{(j)}; \theta_{CNN}), j \in \{2, 6\} \quad (3)$$

其中， $f_{CNN}(\cdot)$  表示卷积神经网络的结构， $\theta_{CNN}$  表示卷

积神经网络中参数的集合。

针对三维数据块  $I^{(3)}$  和  $I^{(4)}$ ，由于其在时间和空间维度上的数据分布都有各自的意义，因此采用 3D 卷积模块<sup>[20]</sup>提取数据特征。普通的二维卷积模块只能识别静态的图像数据，无法捕捉在时间维度上的信息。而 3D 卷积模块将二维卷积核进行扩展，形成一个立方体的三维卷积核，使得其能够在时间维度上捕捉数据的变化信息，从而使得最终提取的特征更加精确。本文利用 3D 卷积核替代传统卷积网络中的 2D 卷积核，通过多层的 3D 卷积组成提取三维时空特征的神经网络架构。将轧制过程中每个轧制道次的轧制力、弯辊力、扭矩、辊缝距离等数据  $I^{(3)}$  以及轧辊接触面积、出口厚度、出口平直度等数据  $I^{(4)}$  作为输入，利用三维卷积神经网络提取数据特征，即

$$A_j = f_{3D-CNN}(I^{(j)}; \theta_{3D-CNN}), j \in \{3,4\} \quad (4)$$

其中， $f_{3D-CNN}(\cdot)$  表示三维卷积神经网络结构， $\theta_{3D-CNN}$  表示三维卷积神经网络参数的集合。

之后，基于不同数据块与最终板形质量之间的最大互信息系数，为上述提取出的每个异构数据块的数据特征分配权重，同时将加权后的数据特征进行融合，表示为

$$A' = \{A_1, m_2 A_2, m_3 A_3, m_4 A_4, m_5 A_5, m_6 A_6\} \quad (5)$$

其中， $A'$  表示数据加权融合后的数据特征。

将加权融合后的数据特征输入一个单层神经

网络，从数据特征预测最终的板形质量标签信息。

$$Y' = f(A') \quad (6)$$

其中， $f(\cdot)$  表示一个单层的神经网络， $Y'$  表示从生产过程数据提取的产品质量标签信息。

利用预处理后的产品数据，对上述步骤建立起来的神经网络结构进行训练，最终得到针对板形最终质量的预报模型。

## 5 实验验证与分析

使用国内某厚板制造厂的实际工业数据对软件的数据实时转存功能、模型训练与预报功能进行实验验证。

### 5.1 数据实时转存功能验证

本文以 2020 年 1 月 1 日为起始点，对后续一年的数据进行转存。由于在实际的生产过程中，标签数据的更新比生产过程数据的更新晚 1~2 天，因此，为了保证数据更新的时效性，将生产过程数据的更新以及标签数据的更新分开。用户在时间选择模块选取要转存的时间段，并能够通过查询到的钢板 ID 进行精确选择，单击转存按钮，即可自动从生产过程数据以及标签数据提取厚板板形预报所需要的信息，并保存在转存数据库中，方便后续的模式训练以及预报工作。

数据转存界面如图 5 所示，用户能够从转存后钢板 ID 块的颜色状态分辨数据转存的结果。对于生产过程数据，将 ID 块分为 4 个部分，分别表示

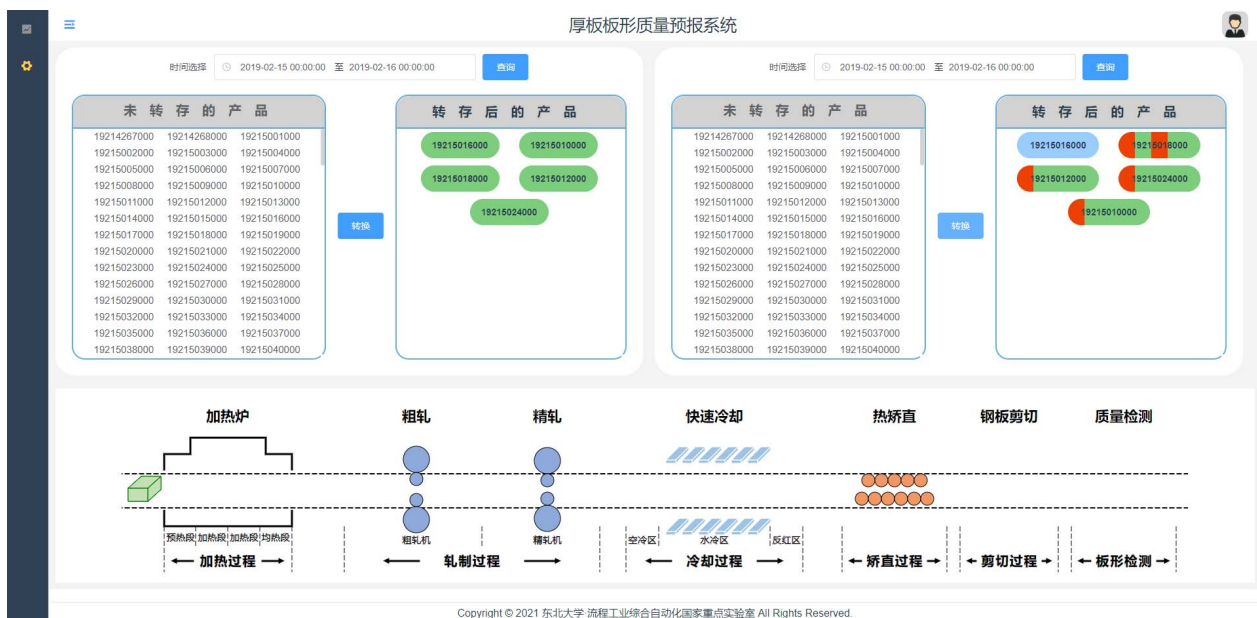


图 5 数据转存界面

加热过程、轧制过程、冷却过程以及热矫直过程数据的转存结果。其中，绿色表示该部分数据完整，转存成功；红色表示该部分由于生产工艺以及现场扰动的影响出现数据丢失，转存失败。对于标签数据的转存结果，将 ID 块分为 4 个部分，分别表示 4 种板形异常类别：头尾翘曲、厚度异常、中浪、左右边浪。其中，红色表示该块钢板存在对应的板形异常；绿色表示该块钢板不存在对应的板形异常；蓝色表示该块钢板的标签数据缺失，无法进行标签转存。

数据转存后，在转存数据表中能够查看转存后的数据信息，表明数据转存成功。

### 5.2 模型训练与预报功能验证

将转存后的生产过程数据作为预报模型测试的输入数据，对钢板的 4 种板形质量异常类别进行模型训练与预报。

本次实验选用的数据集一共包括 25 480 个样本，选定训练集数据后，系统会自动从数据转存表调取对应钢板的生产过程数据以及标签数据，对数据做预处理后，将生产过程数据作为预报模型的输入数据，将 4 种标签数据分别作为预报模型的输出数据，对预报模型中的参数进行训练，并将训练后的预报模型结构以及预报模型参数保存在数据库的模型转存表中，以使用户查看以及在后续的流程中调用。

板形质量预报界面如图 6 所示。用户在前端界面能够选择要使用的预报模型，并对其进行初始化，预报模型会自动加载到后台程序中。用户能够

在前端界面选择要预报的钢板 ID，单击预报按钮后，右侧会对生产过程中的关键指标数据进行展示，方便用户判定生产过程数据是否出现异常。下方会显示预报模型对 4 种板形质量异常的预报结果，对应数值越大，表明最终板形质量出现异常的概率越高，使用户能够及时了解钢板最终的板形质量，从而对生产过程进行调整，提高产品的合格率。

为了保证实验结果的准确性，从实验样本选取 5 组数据集进行测试。由于在实际的生产过程中，样本数据按照生产顺序依次产生，样本与样本之间存在时序关联，因此，不打乱样本顺序，选取 5 个时间点，将该时间点之前的样本数据作为训练集，将该时间点之后的 1 000 个样本数据作为测试集，5 组训练集与测试集样本个数如表 1 所示。

表 1 5 组训练集与测试集样本个数

NO.	数据集	训练集	验证集	测试集
1	Data1	5 601	622	1 000
2	Data2	12 759	1 417	1 000
3	Data3	16 838	1 870	1 000
4	Data4	18 354	2 039	1 000
5	Data5	20 477	2 275	1 000

对 5 组测试集钢板板形的预报结果的准确率进行统计，模型预报准确率如图 7 所示，显示了相同数据集下 4 种板形异常预报的准确率。算法对比如图 8 所示，显示了选择第 4 组数据集，针对 4 种板

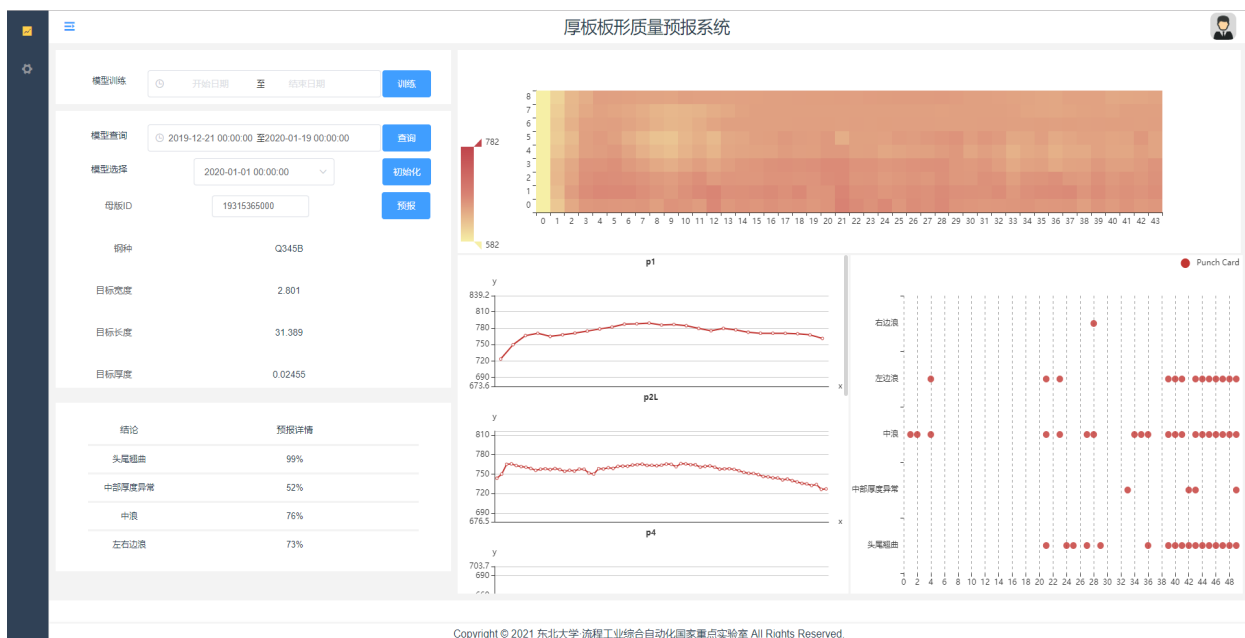


图 6 板形质量预报界面

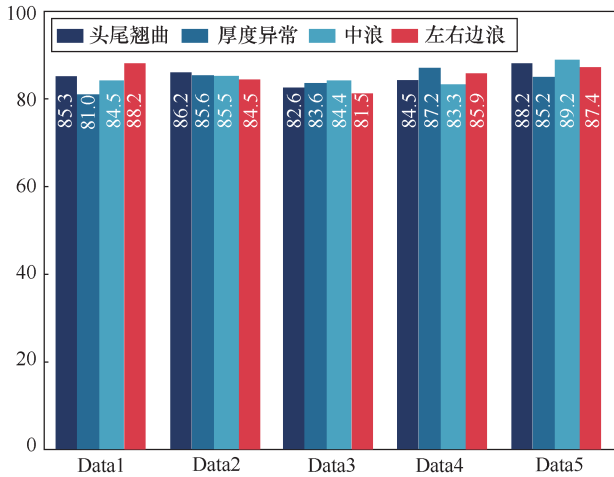


图 7 模型预报准确率

形异常类别，利用 LSTM 网络、支持向量机 (SVM, support vector machine)、决策树分别建立的预报模型的接收者操作特征 (ROC, receiver operating characteristic) 曲线，以及利用本文所设计的预报模型的 ROC 曲线。由图 8 可得，每个测试集对板形质量的预报准确率都达到了 80% 以上，并且本文所设计的预报模型明显优于利用传统方法所建立的预报模型。

## 6 结束语

本文设计的系统实现了厚板生产过程数据转存、板形质量预报模型训练、最终板形质量预报的功能，能够在厚板的生产过程中提前预报钢板最终的板形质量，使操作人员尽早了解厚板最终的板形状态，并依据预报的不同板形异常类别，对厚板的生产过程采取不同的应对措施，从而提高厚板的合格率。数据转存功能不仅有效地去除了实际工业现场中的冗余数据，更节省了数据处理时间，使系统运行更加高效。同时，预报模型有效地利用了工业现场中的多源异构数据，能够保证模型的预报准确率在 80% 以上，且模型的预报效果优于基于传统算法建立的预报模型。

### 参考文献：

- [1] SCHAUSBERGER F, STEINBOECK A, KUGI A. Optimization-based reduction of contour errors of heavy plates in hot rolling[J]. Journal of Process Control, 2016, 47: 150-160.
- [2] XIE B S, CAI Q W, YUN Y, et al. Development of high strength ultra-heavy plate processed with gradient temperature rolling, intercriti-

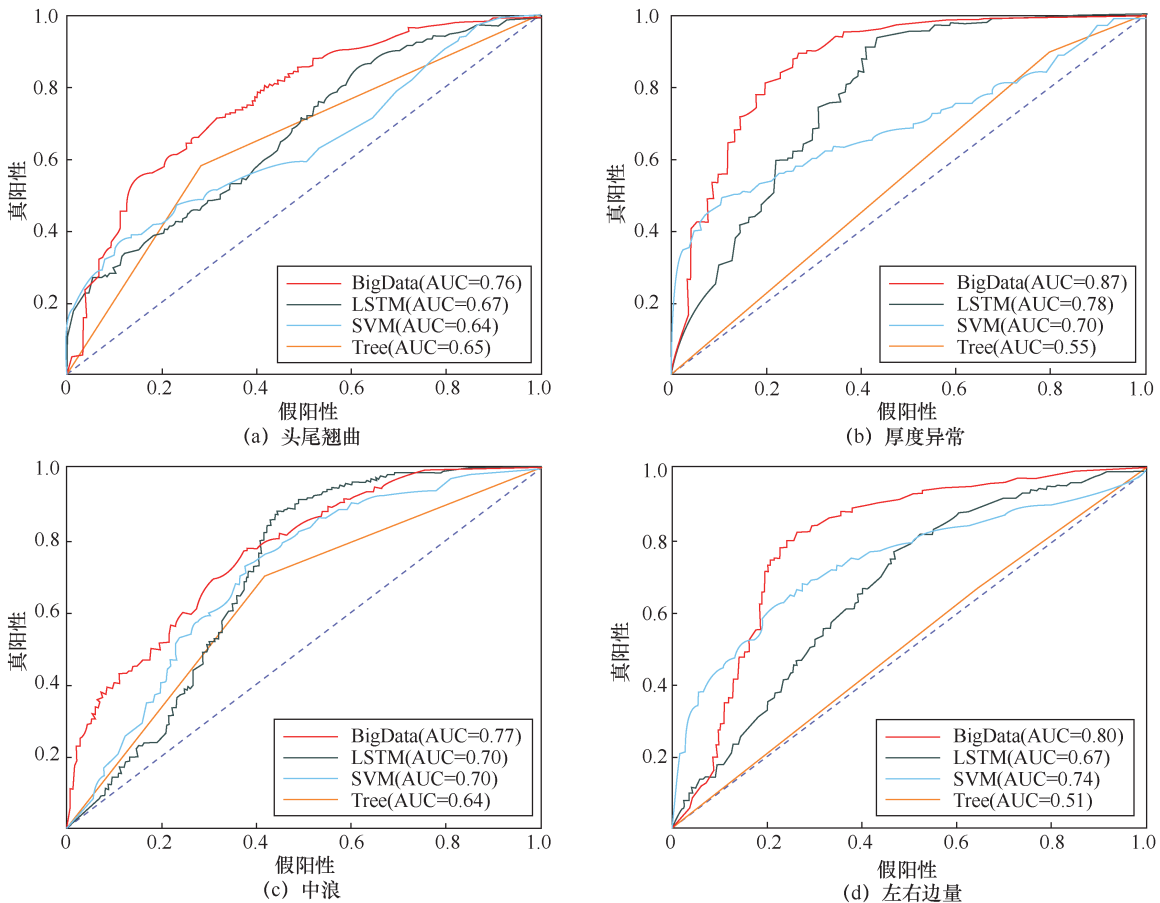


图 8 算法比对

- cal quenching and tempering[J]. *Materials Science and Engineering: A*, 2017, 680: 454-468.
- [3] MA X B, LIU H M, SUN J L, et al. Impact of main drive system of 5 m wide and heavy plate mill on screw-down load deviation[J]. *Engineering Failure Analysis*, 2017, 79: 913-927.
- [4] 王国栋. 高质量中厚板生产关键共性技术研发现状和前景[J]. *轧钢*, 2019, 36(1): 1-8, 30.  
WANG G D. Status and prospects of research and development of key common technologies for high-quality heavy and medium plate production[J]. *Steel Rolling*, 2019, 36(1): 1-8, 30.
- [5] LAUGWITZ M, SEUREN S, JOCHUM M, et al. Development of levelling strategies for heavy plates via controlled FE models[J]. *Procedia Engineering*, 2017, 207: 1349-1354.
- [6] SCHAUSBERGER F, STEINBOECK A, KUGI A. Feedback control of the contour shape in heavy-plate hot rolling[J]. *IEEE Transactions on Control Systems Technology*, 2018, 26(3):842-856.
- [7] YU W, LI G S, CAI Q W. Effect of a novel gradient temperature rolling process on deformation, microstructure and mechanical properties of ultra-heavy plate[J]. *Journal of Materials Processing Technology*, 2015, 217:317-326.
- [8] BASANTA-VAL P. An efficient industrial big-data engine[J]. *IEEE Transactions on Industrial Informatics*, 2018, 14(4): 1361-1369.
- [9] WANG T, KE H X, ZHENG X, et al. Big data cleaning based on mobile edge computing in industrial sensor-cloud[J]. *IEEE Transactions on Industrial Informatics*, 2020, 16(2): 1321-1329.
- [10] 焦四海, 丁建华, 闫博, 等. 厚板数据中心和智能制造的实践与探索[J]. *宝钢技术*, 2020(6): 8-15.  
JIAO S H, DING J H, YAN B, et al. Practice and exploration data center and intelligent manufacture of plate mill[J]. *Baosteel Technology*, 2020(6): 8-15.
- [11] GUO K H, XU T, KUI X Y, et al. iFusion: towards efficient intelligence fusion for deep learning from real-time and heterogeneous data[J]. *Information Fusion*, 2019, 51: 215-223.
- [12] NGUYEN T T, PHAM X C, LIEW A W C, et al. Aggregation of classifiers: a justifiable information granularity approach[J]. *IEEE Transactions on Cybernetics*, 2019, 49(6): 2168-2177.
- [13] 王国栋. 近年我国轧制技术的发展、现状和前景[J]. *轧钢*, 2017, 34(1): 1-8.  
WANG G D. Development, current situation and prospect of Chinese steel rolling technology in recent years[J]. *Steel Rolling*, 2017, 34(1): 1-8.
- [14] GENG D Q, ZHANG C Y, XIA C J, et al. Big data-based improved data acquisition and storage system for designing industrial data platform[J]. *IEEE Access*, 2019, 7: 44574-44582.
- [15] FABLET R, VIET P H, LGUENSAT R. Data-driven models for the spatio-temporal interpolation of satellite-derived SST fields[J]. *IEEE Transactions on Computational Imaging*, 2017, 3(4): 647-657.
- [16] JIANG X X, PAN S R, LONG G D, et al. Cost-sensitive parallel learning framework for insurance intelligence operation[J]. *IEEE Transactions on Industrial Electronics*, 2019, 66(12): 9713-9723.
- [17] LEE S Y, TAMA B A, CHOI C, et al. Spatial and sequential deep learning approach for predicting temperature distribution in a steel-making continuous casting process[J]. *IEEE Access*, 2020, 8: 21953-21965.
- [18] RESHEF D N, RESHEF Y A, FINUCANE H K, et al. Detecting novel associations in large data sets[J]. *Science*, 2011, 334(6062): 1518-1524.
- [19] ZHANG K, LIU N, YUAN X F, et al. Fine-grained age estimation in the wild with attention LSTM networks[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(9): 3140-3152.
- [20] CAO H C, LIU H, SONG E M, et al. Multi-branch ensemble learning architecture based on 3D CNN for false positive reduction in lung nodule detection[J]. *IEEE Access*, 2019, 7: 67380-67391.

## [作者简介]



马宇飞（1996-），男，东北大学硕士生，主要研究方向为工业大数据技术及其应用等。



刘长鑫（1983-），男，博士，东北大学讲师，主要研究方向为自动化技术、计算机软件及计算机应用等。



孔伟（1984-），男，宝山钢铁股份有限公司中央研究院高级工程师，主要研究方向为宽厚板轧制工艺及大数据应用。



丁进良（1976-），男，博士，东北大学教授，主要研究方向为复杂工业过程智能建模与智能优化与控制、生产全流程运行优化、工业大数据分析、机器学习、计算智能及其应用研究等。