

基于多任务学习的炼钢终点预测方法

程进*, 王坚

(同济大学 CIMS 研究中心, 上海 200092)

(* 通信作者电子邮箱 1410tankchj@tongji.edu.cn)

摘要: 钢水质量通常根据终点命中率来判断, 但炼钢过程影响因素众多, 机理分析难以准确预测终点温度和含碳量。鉴于此, 提出一种由数据驱动的多任务学习 (MTL) 炼钢终点预测方法。首先, 分析并提取炼钢过程的输入和输出要素, 结合炼钢两阶段吹炼特点选择多个子学习任务; 其次, 根据子任务与终点参数的相关性选择合适的子任务, 提升终点预测的准确度并构建多任务学习模型; 再对模型输出结果进行二次优化; 最后, 通过近端梯度算法对处理后的生产数据进行模型训练, 获取多任务学习模型的过程参数。以某钢厂为案例, 该方法相比神经网络在终点温度 12℃ 误差范围内和终点含碳量 0.01% 误差内的准确度提升了 10%, 误差范围 6℃ 和 0.005% 的预测准确度分别提升了 11% 和 7%。实验结果表明, 多任务学习在实际中能够提升终点预测的准确性。

关键词: 产品质量预测; 炼钢终点预测; 数据驱动; 多任务学习; 近端梯度算法

中图分类号: TP273+.5 **文献标志码:** A

Endpoint prediction method for steelmaking based on multi-task learning

CHENG Jin*, WANG Jian

(CIMS Research Center, Tongji University, Shanghai 200092, China)

Abstract: The quality of the molten steel is usually judged by the hit rate of the endpoint. However, there are many influencing factors in the steelmaking process, and it is difficult to accurately predict the endpoint temperature and carbon content. In view of this, a data-driven Multi-Task Learning (MTL) steelmaking endpoint prediction method was proposed. Firstly, the input and output factors of steelmaking process were analyzed and extracted, and a number of sub-learning tasks were selected to combine the two-stage blowing characteristics of steelmaking. Secondly, according to the relativity between the sub-tasks and the endpoint parameters, the appropriate subtasks were selected to improve the accuracy of the endpoint prediction, and the multi-task learning model was constructed, and the model output was optimized twice. Finally, the process parameters of the multitask learning model were obtained by model training of the processed production data through the proximal gradient algorithm. In the case of a steel plant, compared with neural network, the prediction accuracy of the method proposed increased 10% when endpoint temperature error was less than 12℃ and carbon content error was less than 0.01%. The prediction accuracy increased by 11% and 7% respectively with the error range less than 6℃ and 0.005%. The experimental results show that multi-task learning can improve the accuracy of endpoint prediction in practice.

Key words: product quality prediction; endpoint prediction for steelmaking; data driven; Multi-Task Learning (MTL); proximal gradient algorithm

0 引言

制造业全球化竞争日益加剧: 一方面, 传统制造企业面临本国和外国的双重压力, 提高产品质量刻不容缓; 另一方面, 绿色生产, 节约成本又要求制造企业能够有效提升成品率。产品质量预测是在生产前通过对原材料、设备和相关因素的分析, 提前对最终产品质量进行评估的一种方法, 通过此类预测企业可以根据结果相应地调整原材料配比或生产设备, 避免不必要的经济损失。尤其对钢铁企业来说, 因为单批次产量大, 质量瑕疵往往造成大量的经济损失, 产品质量的准确预测就显得至关重要。

炼钢是钢铁生产的必需环节之一, 尤其是转炉炼钢, 单批次的钢水产量达上百吨, 提升钢水质量对钢企降低成本、实现

节能减排具有关键意义^[1]。转炉炼钢是将铁水、废钢、铁合金以及一些辅料通过高炉加热, 通过复杂的化学和物理反应, 将铁水中的硫、磷等对钢铁性能产生负面影响的元素进行置换, 同时根据后续工艺要求获得符合产品性能条件的钢水的过程。其中, 氧气转炉由于单炉产量大, 吹炼时间短、热效率高、升温速度快等优点得到了广泛的应用^[2]。

通常, 钢水质量通过终点命中情况来判断其是否达标。终点参数中终点碳含量和终点温度是判断钢水符合产品质量的重要指标。整个炼钢过程也是要求出炉的钢水尽可能满足指标的规定范围。

但炼钢环节生产环境复杂、干扰因素多、生产多受企业实际情况影响, 传统的机理分析难以有效地预测钢水的质量, 随着数据科学的发展和成熟, 以及企业对数据资源的重视, 历史

收稿日期: 2016-08-24; 修回日期: 2016-10-24。 基金项目: 国家科技支撑计划资助项目(2015BAF22B00)。

作者简介: 程进(1989—), 男, 上海人, 博士研究生, 主要研究方向: 机器学习、计算机集成制造; 王坚(1961—), 男, 山东淄博人, 教授, 博士, 主要研究方向: 企业信息化、网络化制造与系统集成。

生产数据可以反映出生产中外部条件变化对产品性能的影响,利用有效的手段充分发挥数据的作用,为有效提升转炉终点参数预测提供了新角度。

1 数据驱动的转炉终点预测

为了最大概率命中终点参数,机理模型需要利用静态模型和动态模型对生产过程的各类输入参数进行预测(图1)。静态模型主要以铁水数据和终点指标为基准,根据热平衡和物料平衡构建数学模型,计算炼钢过程中各类原料的理论加入量。动态模型则根据测量参数及理论指标结合计算原料增加量,并在此进行吹炼和测量,使之能尽可能达到终点指标。

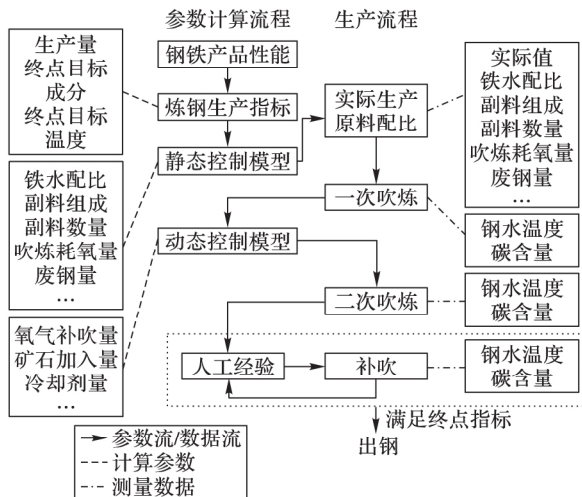


图1 机理模型驱动的转炉炼钢生产流程

Fig. 1 Basic oxygen furnace steelmaking production process driven by mechanism model

为了根据实际生产环境进行合理的预测,部分学者开始研究基于数据的预测方法。数据模型主要分为增量模型、统计模型和智能模型。增量模型和统计模型是对历史生产数据进行分析 and 修正,并根据当前生产任务从历史数据案例中发现相似和最好的生产工艺知识作为推荐^[3]。王心哲等^[4]利用因果关系分析对案例属性进行确认,提升案例推理对转炉静态控制模型预测的有效性。袁平等^[5]在案例推理基础上加入增量补偿,提高终点预报的精度,但案例推理的准确度受限于案例库规模,且难以对案例知识重现。随着人工智能的发展,大量学者开始将终点参数预测与之相结合,其中神经网络模型最为突出。黄金侠等^[6]和付佳等^[7]结合静态模型,计算部分输入参数的理论结果,并以实际终点参数作为输出,提升预测的精度。柴天佑等^[8]和张辉宜等^[9]则将径向基神经网络结构用于终点参数预测。Han等^[10]利用反向传播神经网络(Back Propagation Neural Network, BPNN)改进后的极端学习机用于连续时间的终点预测,并利用膜进化算法减小模型的过拟合问题。赵晓东等^[2]根据耦合多变系统原理,利用多输入输出神经网络进行终点参数预报。此外,王安娜等^[1]将粒子群算法用于BPNN权值计算,提高预测速度和精度,冯明霞^[11]通过回阻抗算法改进BPNN权值计算,提高学习收敛速度。

但以上数据驱动的终点预测面临同样的问题,炼钢企业虽然拥有大量的炼钢历史数据,同时也包含了大量的可变因素,如原料参数、设备、工作小组等,这些因素往往会影响工艺

过程,使终点参数造成较大的偏差。为了提高预测的精度和体现再现性原理,在数据处理阶段往往需要对离散属性进行筛选,保证条件和工艺操作的稳定,以获取相同的冶炼效果(图2)^[9]。最常用的筛选方法是根据相同钢种的生产数据来预测终点参数。钢企在生产计划以订单需求进行编排,造成了虽然历史生产数据总量很大,但训练模型可用的样本却不多的问题。而该问题也导致了神经网络的训练欠拟合、预测结果不稳定等现象。

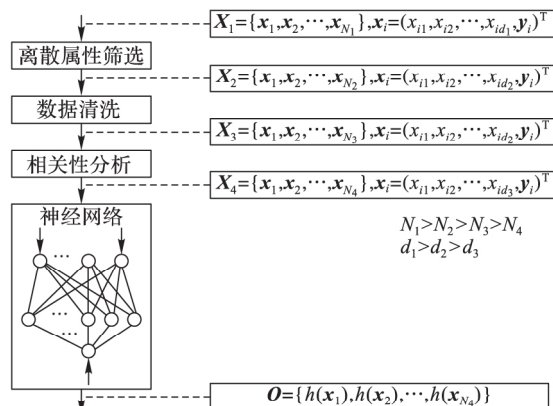


图2 终点参数预测神经网络训练流程

Fig. 2 Neural network training flow for predicting endpoint parameters

基于此,本文提出了一种基于多任务学习的炼钢终点参数预测方法。将终点温度和终点碳含量作为主任务,根据炼钢两阶段吹炼特点选择多个子学习任务并计算与主任务的相关度,选择相关度相对较大的预测任务同时与主任务作为多任务学习的输出,提升对主任务的预测能力,最后利用子任务预测结果二次优化主任务的预测精度,该方法可以有效改善在训练样本欠缺的情况下神经网络的训练精度,同时提升对终点参数的预测能力。

2 多任务学习

多任务学习(Multi-Task Learning, MTL)是利用归纳转移机理,对训练样本选取多个训练目标,根据目标间特定的相关信息来提升学习系统的泛化性能的方法^[12]。与传统机器学习中的单任务学习(Single-Task Learning, STL)不同,MTL充分利用了多个任务之间的相关信息来提升任务之间的可识别度或预测能力。传统的STL模型虽然可以对样本进行准确的分类,但由于训练样本的数量有限,导致部分数据特征被忽视而引起模型欠拟合,在实际的应用中不能获得较好的效果。MTL通过放松部分训练样本与主任务之间的关系,同时相关任务利用潜在数据特征,提升模型训练效果,从而实现模型的泛化能力。目前,MTL主要解决图像识别、文本识别等相关分类问题。

与单任务学习神经网络(Single-Task Learning Neural Network, STLNN)相比(图3)^[12],多任务学习神经网络(Multi-Task Learning Neural Network, MTLNN)采用多个任务同一隐层的方式,每个隐层的输出将作为多个任务输入。MTLNN通过多任务共享隐层的机制,可以有效地提取任务之间的共同信息,克服由于训练样本不足而导致的网络训练不足,但这种特性导致MTLNN的预测性能依赖于主任务和子任务之间的相关性,若主任务和子任务之间相互独立,网络性

能不仅不能提升,反而加大了训练难度,甚至可能影响预测能力。

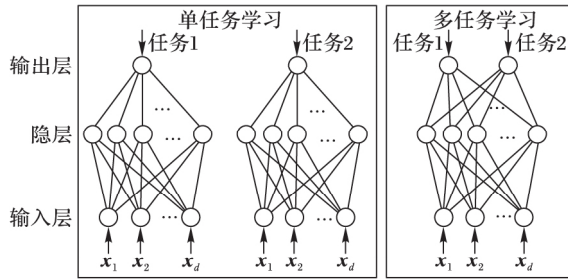


图3 基于神经网络的STL与MTL对比

Fig. 3 Comparison of STL and MTL based on neural network

3 转炉终点参数预测模型

转炉炼钢中静态模型通常确定了原料加入量的大致范围,钢企在配置后通常已经对铁水量(各类元素含量)、副原料量(白云石、石灰、镁球)、废钢量(废钢种类)、初始矿石量等参数进行了确定。一次吹炼结束后,根据钢水温度和碳含量和终点指标对比并利用动态模型预测原料补加量,这些量包括了补吹氧气量、矿石加入量、冷却剂等,该类变量可以作为子任务。二次吹炼的终点参数作为预测对象和MTL中的主任务。

3.1 任务相关性分析

多任务学习模型构建前,首先需要寻找主任务和与之相关的子任务。通常,多任务学习主任务与其相关任务要具有相同的输入变量,在确定了相关任务后,从中选择相关度较大的任务作为子任务为模型提供预测参数。

$y^{\text{main}} = (y_1^{\text{main}}, y_2^{\text{main}}, \dots, y_N^{\text{main}})$ 表示模型主任务的期望值向量,其中 N 表示样本数, y_i^{main} 为每个样本的期望值。

$Y^{\text{sub}} = \{y_1^{\text{sub}}, y_2^{\text{sub}}, \dots, y_L^{\text{sub}}\}$ 表示子任务期望值矩阵, L 表示子任务数, $y_i^{\text{sub}} = (y_{i1}^{\text{sub}}, y_{i2}^{\text{sub}}, \dots, y_{iN}^{\text{sub}})$ 表示样本实例第 i 个子任务的期望值。子任务之间的相关系数计算公式:

$$r_j = \frac{\sum_{i=1}^N (y_{ji}^{\text{sub}} - \bar{y}_j^{\text{sub}}) (y_i^{\text{main}} - \bar{y}^{\text{main}})}{\sqrt{\sum_{i=1}^N (y_{ji}^{\text{sub}} - \bar{y}_j^{\text{sub}})^2} \sqrt{\sum_{i=1}^N (y_i^{\text{main}} - \bar{y}^{\text{main}})^2}} \quad (1)$$

其中: \bar{y} 表示样本期望值的均值; $r_j \in [-1, 1]$ 表示第 j 个子任务与主任务的相关系数, $|r_j|$ 越大,子任务 j 与主任务之间的相关性越大。以钢水终点温度为例,根据实际生产情况和静态模型、动态模型将预测子任务分为总吹氧量、矿石补加量和二次吹氧量,分别计算所有样本中子任务期望值与终点温度期望值的相关系数(表1)。

表1 终点温度的主任务与子任务相关系数

Tab. 1 Correlation coefficients between main task and subtasks of endpoint temperature

子任务	终点温度	子任务	终点温度
总吹氧量	0.4742	二次吹氧量	-0.0214
矿石补加量	0.7534		

定义阈值 γ_r , 当 $|r_j| > \gamma_r$ 时,任务 j 作为主任务的一个子任务,否则将任务舍去。根据表1结果,二次吹氧量与主任务相关度较低随舍去,将总吹氧量和矿石补加量作为终点温度

预测子任务构建预测模型。

3.2 多任务神经网络模型

神经网络属于人工智能中一种复杂的网络模型,通过模拟人脑中神经元之间的信息传递过程实现对不同的信息反映并提供不同的行为方式。不同的神经网络结构可以用于分类、拟合或预测,其中BPNN由于对复杂非线性具有较强的映射能力,在对内在复杂问题的求解上具有优势。转炉炼钢过程虽然具有大量的机理模型对终点参数进行预测,但由于生产环境和设备等不稳定因素的影响其预测结果并不理想,由于神经网络具有较强的鲁棒性可以有效减小不良因素对结果的影响。根据BPNN结构构建转炉终点温度预测多任务神经网络模型。

神经网络权值矩阵:

$$W_1^{q \times (d+1)} = \begin{pmatrix} w_{11}^{(1)} & w_{12}^{(1)} & \cdots & w_{1d}^{(1)} & w_{1b_1}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & \cdots & w_{2d}^{(1)} & w_{2b_1}^{(1)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ w_{q1}^{(1)} & w_{q2}^{(1)} & \cdots & w_{qd}^{(1)} & w_{qb_1}^{(1)} \end{pmatrix}$$

$$W_2 \in \mathbf{R}^{3 \times (q+1)}$$

权值矩阵中除了输入参数权值外还加入一维常数权值向量。

图4中 d 表示输入参数的数量,输入参数为炼钢实际生产中实测的原料量、配比和相关记录数据, q 表示神经网络隐层数量,为简单说明问题本文采用单隐层的BPNN结构。输出参数则分为主任务和子任务,主任务为终点温度,子任务为总吹氧量和矿石补加量。

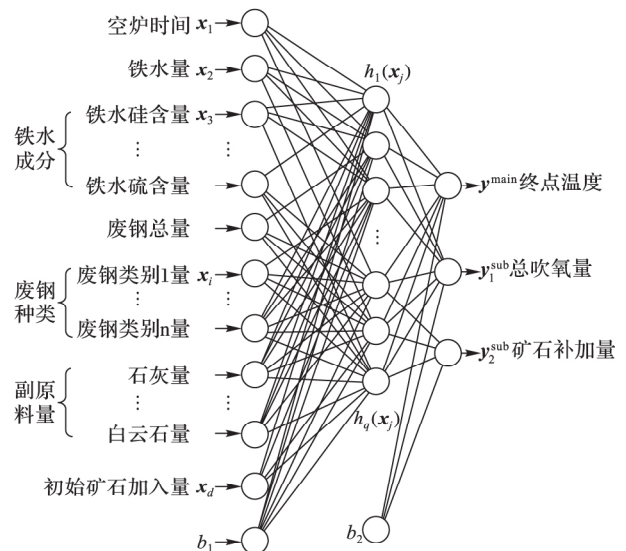


图4 基于MTL的终点预测神经网络模型

Fig. 4 Neural network model for endpoint prediction based on MTL

隐层网络结构:

$$H = W_1 X \quad (2)$$

$X = \{x_1, x_2, \dots, x_N\}$, $x_i = (x_{i1}, x_{i2}, \dots, x_{id}, b_1)^T$, b_1 为常数, $H \in \mathbf{R}^{q \times N}$ 。

隐层网络激活函数选择双曲正切函数:

$$g_1(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (3)$$

预测输出表示为:

$$\hat{Y} = W_2 G_1 \quad (4)$$

其中: $\hat{Y} \in \mathbf{R}^{3 \times N}$ $G_1 = [g_1(H) \quad D_2^{1 \times N}]^T$ D_2 中的元素值均为 b_2 b_2 为常数。

3.3 主任务预测值优化

为了提高预测精度,减小预测误差,充分利用多任务神经网络输出之间的相关性,通过多任务的改进联合概率预测主任务的误差变化正负和绝对值大小,从而达到对主任务的预测值优化的作用。

优化后的预测函数表示为:

$$\hat{o} = \hat{y}^{\text{main}} - \Delta \hat{y}^{\text{main}} \quad (5)$$

其中: \hat{o} 为最终主任务的预测值 $\Delta \hat{y}^{\text{main}}$ 表示误差修正量,具体表示为:

$$\Delta \hat{y}^{\text{main}} = (\varphi(\hat{y}^{\text{main}}, \hat{Y}^{\text{sub}} y^{\text{main}(L)}) - 0.5) \delta(\hat{Y}^{\text{sub}}) \quad (6)$$

函数 $\varphi(\hat{y}^{\text{main}}, \hat{Y}^{\text{sub}} y^{\text{main}(L)})$ 用于判断主任务预测值与期望值的误差正负和误差预测的准确度。 $y^{\text{main}(L)}$ 表示将主任务的取值范围等分成 $L-1$ 后组成的端点向量。当函数 $|\varphi|$ 趋向于 0.5 时,表示在统计上预测误差正负难以确定,此时不对结果进行优化,该函数表示为:

$$\varphi(\hat{y}^{\text{main}}, \hat{Y}^{\text{sub}} y^{\text{main}(L)}) = P(\Delta y_i^{\text{main}} \geq 0 | y^{(n-1)} \leq \hat{y}_i < y^{(n)}) + R_\Delta G_2 \quad (7)$$

其中: $R_\Delta \in \mathbf{R}^{1 \times 2}$ 表示主任务预测值误差变化(只保留误差变化正负,舍去具体值)与子任务(总吹氧量、矿石补加量)误差变化的相关系数 $G_2 \in \mathbf{R}^{2 \times 1}$ $G_2 = [g_2(\hat{y}_1^{\text{sub}} y_1^{\text{sub}(L_1)}) \quad g_2(\hat{y}_2^{\text{sub}} y_2^{\text{sub}(L_2)})]^T$ 。其中:

$$g_2(\hat{y}_i y^{(L)}) = P(\Delta \hat{y}_i^{\text{main}} \geq 0 | y^{(n-1)} \leq \hat{y}_i < y^{(n)}) - 0.5 \quad (8)$$

函数 $\delta(\hat{Y}^{\text{sub}})$ 求解预测值的误差优化量,根据子任务预测值、预测值误差分布和主任务误差相关性对预测值进行优化,在相关性一定的情况下,当子任务预测误差分布较大(方差较大)时,其预测值对主任务预测的优化准确性下降,应该减小其对结果的影响。 $\delta(\hat{Y}^{\text{sub}})$ 的表达式为:

$$\delta(\hat{Y}^{\text{sub}}) = r^{(2)} \hat{Y}^{\text{sub}} / \delta^{\text{sub}} \quad (9)$$

其中: δ^{sub} 表示每个子任务预测值与期望值的标准差 $r^{(2)} \in \mathbf{R}^{1 \times 2}$ 表示主任务误差与两个子任务的相关系数向量。相关性越大,子任务预测值对主任务误差优化的作用越大。

4 多任务神经网络训练

神经网络的预测精度取决于网络的结构和神经元的传递权值,对单任务神经网络而言,网络中权值训练仅需要考虑过拟合问题,最优权值表示为式(10):

$$W^* = \arg \min_W \Gamma(X, y, W) + \lambda P(W) \quad (10)$$

其中: $\Gamma(X, y, W)$ 表示为训练集的经验损失函数, $P(W)$ 为正则化项,用来避免网络过拟合, $\lambda > 0$ 为正则化参数。

多任务学习还需要考虑每个任务的经验损失函数在总经验损失函数中所占的比例问题,不同的比例决定了不同任务提供的共享信息的重要程度。本文中总经验损失函数函数定义为:

$$\Gamma(X, y, W) = \frac{1}{N} \left[\frac{1}{2} \sum_i (\hat{y}_i^{\text{main}} - \hat{y})^2 + \right.$$

$$\left. \frac{1}{2} \sum_{j=1}^3 \sum_{i=1}^N |r_j| [(\hat{y}_{ij}^{\text{sub}} - \hat{y}_j^{\text{sub}})^2] \right] \quad (11)$$

其中: \hat{y} 表示标准化在 $[-1, 1]$ 的数据。正则化项选择 W 的 l_2 范数 $\|W\|_2$ 。优化问题可以转换为:

$$\min_W \Gamma(X, y, W) + \lambda P(W) \quad (12)$$

其中: $\Gamma(X, y, W)$ 是可微函数, $P(W)$ 是一个闭的凸函数,可以利用近端梯度算法^[13]求解该问题,近端梯度算法表示为:

$$W^{k+1} = \text{Prox}_{\chi^k p}(W^k - \chi^k \nabla \Gamma(X, y, W^k)) \quad (13)$$

其中: χ^k 表示梯度下降步长,利用近端算子 $\text{Prox}_{\chi^k p}(\cdot)$ 将结果投影到 p 空间, χ^k 也表示投影参数。 χ^k 的选择决定了算法是否收敛,本文采用线搜索算法对 χ^k 求解^[14]。其中判决收敛的函数被定义为:

$$\bar{\Gamma}(X, y, W^k, W^{k+1}) = \Gamma(X, y, W^k) + \frac{1}{2\chi^k} \|W^{k+1} - W^k\|_2^2 + \nabla \Gamma(X, y, W^k)^T (W^{k+1} - W^k) \quad (14)$$

根据上述内容对多任务神经网络权值进行训练:

步骤 1 初始化各类参数 $\chi^k, \beta, \lambda, \eta$ 、迭代次数 S , 确定隐层结构和隐层数量。

步骤 2 对权值矩阵 W_1, W_2 初始化,记作 W^k 在 $[-1, 1]$ 随机取值。

步骤 3 对输入参数和目标变量标准化,参数范围约束在 $[-1, 1]$ 。

$$\bar{x}_i = \frac{2x - (x_{\max} + x_{\min})}{(x_{\max} - x_{\min})} \quad (15)$$

其中: x_{\max} 是每个输出参数额定最大值组成的向量, x_{\min} 是每个输出参数额定最小值组成的向量。

步骤 4 利用式(2)和式(3)计算标准化的预测值 \hat{y}^{main} 和 \hat{Y}^{sub} 。

步骤 5 将标准化后的输入参数代入目标函数:

$$f(W) = \Gamma(X, y, W) + \lambda \|W\|_2 \quad (16)$$

求得预测误差 E , 若 $E < \eta$ 或者训练代数 $k = S$, 则跳入步骤 9; 否则 $k = k + 1$, 进入步骤 6。

步骤 6 利用式(13)计算新的权值矩阵 W^{k+1} , 并将 W^{k+1} 代入到式(11)和式(14)。

步骤 7 若 $\bar{\Gamma}(X, y, W^{k+1}) \leq \bar{\Gamma}(X, y, W^k, W^{k+1})$, 则进入步骤 9; 否则进入步骤 8。

步骤 8 更新 χ^k , 让 $\chi^{k+1} = \beta \chi^k$, 返回步骤 5。

步骤 9 结束训练, 输出最优的权值矩阵 W^* 。

5 实验验证

5.1 数据处理

实验数据来自某钢厂 2012 年下半年 6 个月的转炉炼钢生产记录。生产记录总共 5865 炉次钢水生产过程数据, 其中包括了炉座信息、班组信息、钢水计划目标值、两次吹炼的原料加入量信息、补吹信息、过程测量值以及钢水最终的命中情况共 659 个属性类。一方面, 由于部分关键参数是通过人工记录的, 关键属性值会存在缺失或错误如“0”值或空值; 另一方面, 不同类型的炼钢工艺会有区别, 出于对企业知识产权的保护无法获得详细记录, 而这可能导致模型失真。鉴于以上情况, 将缺失数据剔除并基于出钢记号选择生产记录较多的钢作为研究对象, 并最终确定采用 AP655K1 类钢的 899 炉次生产记录。通过筛选得到输出参数 5 项(表 2) 22 个参数

作为输入属性(表 3)。

5.2 模型选择

首先计算终点参数与输出参数之间的相关系数(表 1 和表 4),令 $\gamma_r = 0.4$,则总吹氧量和矿石补加量同时作为终点温度和终点碳含量的子任务。以 699 炉次数据作为训练样本集,以 200 炉次作为测试集。分别构建以终点温度和终点碳含量为主任务,总吹氧量和矿石补加量为子任务的本文 MTL 模型、MTLNN 和 STLNN 模型。利用近端梯度算法训练以上三个模型的 BPNN,其中 $\chi^0 = 4$ $\beta = 0.5$ $\lambda = 0.3$ $\eta = 0.001$,迭代次数为 200 次。本文采用遍历的方式确定单隐层 STLNN 和 MTLNN 的最佳结构,隐层神经元数从 3 到 30 个,每种结构重复 100 次并选取在训练样本集上使式(16)最小的值(图

5)。其中:(a)为终点温度模型计算的总经验损失函数值,两个模型在隐层神经元数为 14 时获得最优;(b)为终点碳含量模型的值,两个模型在隐层神经元数为 17 时最优,min1 和 min2 分别对应 STLNN 和 MTLNN 最小总经验损失函数的解。

表 2 输出参数统计量

Tab. 2 Output parameter statistics					
比较项	终点温度/℃	终点碳含量/t	总吹氧量/m ³	矿石补加量/t	二次吹氧量/m ³
最小值	1 609	0.025	12 352	0	399
最大值	1 721	0.097	15 485	8.931	9 304
均值	1 679	0.053	13 598	2.995	1 403
标准差	14	0.009	431	2.062	455

表 3 输入数据统计量

Tab. 3 Input data statistics

比较项	空炉时间/min	废钢总重量/t	铁水总重量/t	铁水成分					铁水温度/℃	A 类废钢量/t	B 类废钢量/t
				碳含量/%	硅含量/%	锰含量/%	磷含量/%	硫含量/%			
最小值	66	10.208	249.609	2.325	0.098	0.205	0.102	0.009	999	0	0
最大值	13 862	38.124	281.510	8.856	1.404	0.493	0.147	0.105	1 600	16.041	16.927
均值	1 267	23.853	265.925	3.993	0.343	0.321	0.120	0.022	1341	0.788	1.542
标准差	1 536	5.211	5.863	0.477 0	0.119	0.038	0.007	0.012	75	2.478	3.342
比较项	C 类废钢量/t	D 类废钢量/t	E 类废钢量/t	F 类废钢量/t	石灰量/t	白云石量/t	镁球量/t	污泥球量/t	铁矿石量/t	发热剂量/t	抑制剂量/t
最小值	0	0	0	0	7.083	0.393	1.626	0	0	0	0
最大值	15.260	29.895	10.520	10.520	31.515	7.714	5.769	5.226	11.918	3.131	2.054
均值	5.256	12.847	0.043	3.359	13.336	3.550	3.416	2.025	3.905	1.484	0.258
标准差	3.378	6.288	0.656	2.329	2.360	0.801	0.435	0.608	2.188	0.790	0.178

表 4 终点碳含量的主任务与子任务相关系数

Tab. 4 Correlation coefficients between main task and subtasks of end-point carbon content

子任务	终点碳含量	子任务	终点碳含量
总吹氧量	-0.466 8	二次吹氧量	-0.110 8
矿石补加量	0.674 7		

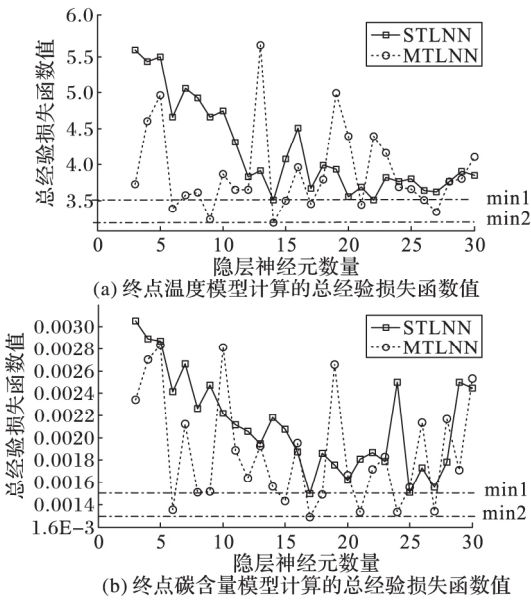


图 5 MTLNN 和 STLNN 模型训练结果
Fig. 5 Results of MTLNN and STLNN model training

5.3 MIV 分析

平均影响值(Mean Impact Value, MIV)分析用于研究神经网络中输入变量对输出变量的影响程度,本文利用该方法来识别 STLNN 和 MTLNN 模型中各输入属性对终点参数的影响。首先选取一个输入属性,将其取值分别增加 10% 和减少 10% 生成两个新样本集,其次,将新样本集分别代入训练好的 STLNN 和 MTLNN 模型中再次计算预测结果,最后,求两个新样本集预测结果平均差值绝对值。通过此步骤依次计算 22 个输入属性对终点温度的影响(图 6)。

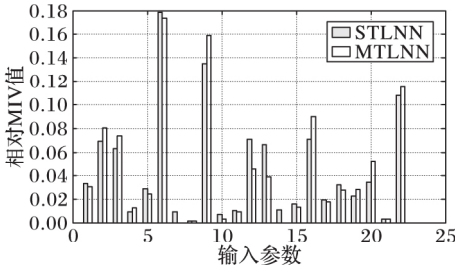


图 6 MTLNN 和 STLNN 模型的 MIV 分析

Fig. 6 MIV analysis of MTLNN and STLNN models

22 个输入参数依次对应表 3,MIV 值越大表示该属性相对其他输入参数对终点温度的预测影响程度越大,两个模型中锰含量、铁水温度和抑制剂量均为影响程度前 3 的属性,之后 MTLNN 模型中废铁总量、铁水量和石灰量均高于其他属性 0.025 以上,但 STLNN 模型 C 类废钢量和 D 类废钢量对模型的影响程度也依然很高,5 个属性的相对 MIV 值在 0.6 ~

0.7 表明了两个模型对终点温度预测时所参考的输入参数权重不同,而该特点将影响模型对新样本的适应度,即模型的泛化能力。

5.4 模型对比

将本文 MTL 模型、MTLNN 和 STLNN 模型应用到测试数据中,验证模型对未知输入数据的预测准确度。

图 7 中的实线表示预测值和实际值相等的标准线,其他线条向外依次表示图 7(a) 中 $|\Delta_{\text{误差}}| < 6^\circ\text{C}$ 和 $|\Delta_{\text{误差}}| < 12^\circ\text{C}$ 的标准线,图 7(b) 中 $|\Delta_{\text{误差}}| < 0.005\%$ 和 $|\Delta_{\text{误差}}| < 0.01\%$ 的标准线。通过统计分析得到表 5 的结果。总体上基于 MTL 的终点预测模型比单任务的预测准确度较高。本文方法比 STLNN 在终点温度 12°C 误差范围内和终点含碳量 0.01% 误差内的准确度提升了 10% ,误差范围 6°C 和 0.005% 的预测准确度分别提升了 11% 和 7% ,表明本文方法具有较好的泛化能力。与 MTLNN 方法相比 $|\Delta_{\text{误差}}| < 6^\circ\text{C}$ 和 $|\Delta_{\text{误差}}| < 0.005\%$ 预测准确度提升了 3% 左右,在 $|\Delta_{\text{误差}}| < 12^\circ\text{C}$ 和 $|\Delta_{\text{误差}}| < 0.01\%$ 时提升了 2% 左右,预测提升能力不大,但本文方法对预测值的优化可以让误差进一步地缩小。对比三种方法在测试样本集中的效果,由于训练样本集数量不足并且质量欠佳,STLNN 模型的泛化能力不足,对测试样本集的预测能力不足,相比而言多任务学习有效保证了在数据样本不足的情况下预测泛化能力,虽然在训练集上的优化能力不足,但对测试样本的预测能力较强,保证了模型的有效性。

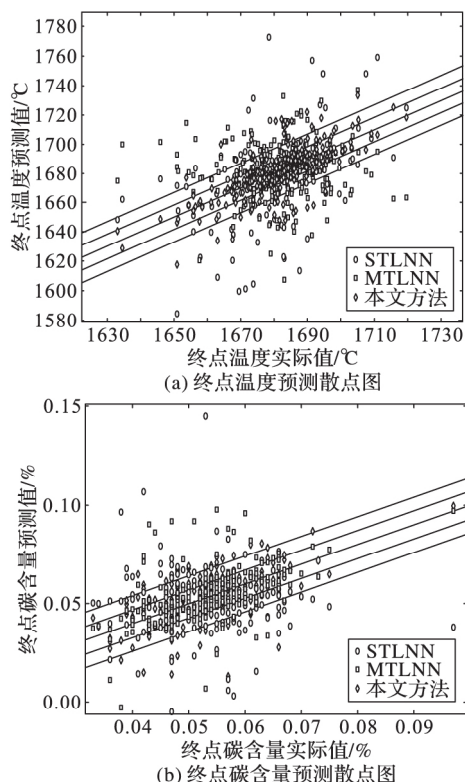


图 7 三种方法对终点温度和碳含量的预测结果比较

Fig. 7 Comparison of three methods for endpoint prediction of temperature and carbon content

图 8 表示三种方法的终点碳含量预测值与实际钢水碳含量的测量值的绝对误差波动范围。从总体上看本文 MTL 模型的预测误差波动幅度要小于 MTLNN 和 STLNN 模型。针对

训练数据集(Training): 本文 MTL 模型的预测误差波动幅度和 MTLNN 模型近似,在 $[-0.034, 0.023]$ 区间内; STLNN 模型误差较大,在 $[-0.033, 0.039]$ 区间内,负绝对误差比本文 MTL 模型低了 0.001 ,但正绝对误差高出 0.016 。在测试数据集(Testing)上: 本文 MTL 模型比 MTLNN 模型正绝对误差低 0.006 ,负绝对误差低 0.004 ; 同时本文 MTL 模型比 STLNN 模型正绝对误差低 0.014 ,负绝对误差低 0.007 。基于 MTL 模型的绝对误差波动区间小于 STL 模型,意味着 MTL 模型具有相对更准确的预测能力。本文 MTL 模型与 MTLNN 模型相比,对训练数据集的预测波动近似,主要由于对预测值进行优化中需要考虑误差正负问题,而部分点的误差正负概率通常不会极端表现为负误差或正误差,所以在优化中不仅会缩小误差,也有可能增大误差,但在测试数据集上误差范围缩小了 0.01 ,表明本文 MTL 模型具有更好的泛化能力,即实际的预测能力。

表 5 终点温度预测和终点碳含量预测对比
Tab. 5 Endpoint temperature prediction and endpoint carbon content prediction

对比项目	测试样本集	STLNN	MTLNN	本文方法
终点温度	$ \Delta_{\text{误差}} < 6^\circ\text{C}$	43.00	50.50	54.00
	$ \Delta_{\text{误差}} < 12^\circ\text{C}$	69.00	76.50	79.00
终点碳含量	$ \Delta_{\text{误差}} < 0.005\%$	49.00	53.50	56.00
	$ \Delta_{\text{误差}} < 0.01\%$	76.50	85.00	86.50

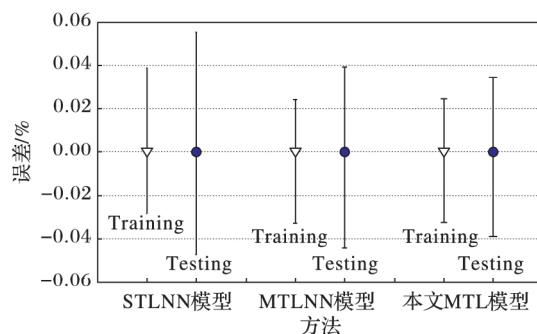


图 8 三种方法在终点碳含量预测中的误差线对比

Fig. 8 Error line comparison of three methods in prediction of endpoint carbon content

6 结语

针对钢水质量预测问题,本文提出了一种基于多任务学习的炼钢终点预测方法,该方法与传统数据驱动方法相比可以解决由于样本数量不足导致模型预测精度不足和泛化能力弱的问题。该方法主要特点在于利用多个子任务优化模型对主任务的学习能力,同时协助优化主任务。通过实际数据验证,该方法比传统神经网络在终点温度 12°C 误差范围内和终点含碳量 0.01% 误差内的预测准确度提升了 10% ,误差范围 6°C 和 0.005% 的预测准确度分别提升了 11% 和 7% ,同时表现出了较强的泛化能力。

本文只考虑了在单一钢种的情况下利用多任务学习方法进行预测,下一步工作将考虑更多结合离散属性,增加模型对多钢种的终点参数预测能力。

参考文献 (References)

[1] 王安娜,陶子玉,姜茂发,等. 基于 PSO 和 BP 网络的 LF 炉钢水

- 温度智能预测[J]. 控制与决策, 2006, 21(7): 814 – 816. (WANG A N, TAO Z Y, JIANG M F, et al. Intelligence temperature prediction of molten steel in LF based on PSO combined with BP neural network [J]. Control and Decision, 2006, 21(7): 814 – 816.)
- [2] 赵晓东, 徐生林, 杨成忠. 转炉炼钢多变量神经网络预报模型[J]. 化工学报, 2010, 61(8): 2111 – 2115. (ZHAO X D, XU S L, YANG C Z. Multivariable neural network predictive model for BOF steel making [J]. Journal of Chemical Industry and Engineering (China), 2010, 61(8): 2111 – 2115.)
- [3] 薄洪光, 刘晓冰, 马跃, 等. 基于粗糙集的钢铁行业工艺知识发现方法[J]. 计算机集成制造系统, 2009, 15(1): 135 – 141. (BO H G, LIU X B, MA Y, et al. Rough set based process knowledge discovery approach in iron and steel industry [J]. Computer Integrated Manufacturing Systems, 2009, 15(1): 135 – 141.)
- [4] 王心哲, 韩敏. 基于因果关系的 CBR 模型用于转炉炼钢静态控制[J]. 大连理工大学学报, 2011, 51(4): 593 – 598. (WANG X Z, HAN M. Causality-based CBR model for static control of converter steelmaking [J]. Journal of Dalian University of Technology, 2011, 51(4): 593 – 598.)
- [5] 袁平, 王福利, 毛志忠. 基于案例推理的电弧炉终点预报[J]. 东北大学学报(自然科学版), 2011, 32(12): 1673 – 1676. (YUAN P, WANG F L, MAO Z Z. CBR based endpoint prediction of EAF [J]. Journal of Northeastern University (Natural Science), 2011, 32(12): 1673 – 1676.)
- [6] 黄金侠, 金宁德. 转炉冶炼终点静态控制预测模型[J]. 炼钢, 2006, 22(1): 45 – 48. (HUANG J X, JIN N D. Static control predictive model for converter refining endpoint [J]. Steelmaking, 2006, 22(1): 45 – 48.)
- [7] 付佳, 陶百生, 陈春雨, 等. 基于 BP 神经网络的转炉供氧模型研究[J]. 冶金自动化, 2014, 38(4): 11 – 15. (FU J, TAO B S, CHEN C Y, et al. Research on oxygen model for BOF based on BP neural network [J]. Metallurgical Industry Automation, 2014, 38(4): 11 – 15.)
- [8] 柴天佑, 谢书明, 杜斌, 等. 基于 RBF 神经网络的转炉炼钢终点预报[J]. 中国有色金属学报, 1999, 9(4): 868 – 872. (CHAI T Y, XIE S M, DU B, et al. Endpoint prediction of basic oxygen furnace steelmaking based on RBF neural network [J]. The Chinese Journal of Nonferrous Metals, 1999, 9(4): 68 – 872.)
- [9] 张辉宜, 周奇龙, 袁志祥, 等. 样本自选择回归分析算法在转炉炼钢中的应用[J]. 钢铁研究学报, 2011, 23(12): 5 – 8. (ZHANG H Y, ZHOU Q L, YUAN Z X, et al. Application of the regression algorithm based on sample-self-selection in BOF steelmaking [J]. Journal of Iron and Steel Research, 2011, 23(12): 5 – 8.)
- [10] HAN M, LIU C. Endpoint prediction model for basic oxygen furnace steel-making based on membrane algorithm evolving extreme learning machine [J]. Applied Soft Computing, 2014, 19(1): 430 – 437.
- [11] 冯明霞, 邹宗树, 李强. 应用改进的神经网络模型预报转炉冶炼终点[J]. 炼钢, 2006, 22(1): 40 – 44. (FENG M X, ZOU Z S, LI Q. Heat endpoint prediction of combined blown converter by improved artificial neural network model [J]. Steelmaking, 2006, 22(1): 40 – 44.)
- [12] CARUANA R. Multitask learning [J]. Machine Learning, 1997, 28(1): 41 – 75.
- [13] ROCKAFELLAR R T. Monotone operators and the proximal point algorithm [J]. SIAM Journal on Control and Optimization, 1976, 14(5): 877 – 898.
- [14] LUO Z Q, YU W. An introduction to convex optimization for communications and signal processing [J]. IEEE Journal on Selected Areas in Communications, 2006, 24(8): 1426 – 1438.

This work is partially supported by the National Key Technology R&D Program of China (2015BAF22B00).

CHENG Jin, born in 1989, Ph. D. candidate. His research interests include machine learning, computer integrated manufacturing.

WANG Jian, born in 1961, Ph. D., professor. His research interests include enterprise informatization, networked manufacturing and system integration.

(上接第 859 页)

- [11] GUDMUNDSSON J, VALLADARES N. A GPU approach to sub-trajectory clustering using the Fréchet distance [J]. IEEE Transactions on Parallel and Distributed Systems, 2015, 26(4): 924 – 937.
- [12] 廖律超, 蒋新华, 邹复民, 等. 一种支持轨迹大数据潜在语义相关性挖掘的谱聚类方法[J]. 电子学报, 2015, 43(5): 956 – 964. (LIAO L C, JIANG X H, ZOU F M, et al. A spectral clustering method for big trajectory data mining with latent semantic correlation [J]. Acta Electronica Sinica, 2015, 43(5): 956 – 964.)
- [13] 龚玺, 裴韬, 孙嘉, 等. 时空轨迹聚类方法研究进展[J]. 地理科学进展, 2011, 30(5): 522 – 534. (GONG X, PEI T, SUN J, et al. Review of the research progresses in trajectory clustering methods [J]. Progress in Geography, 2011, 30(5): 522 – 534.)
- [14] ELNEKAVE S, LAST M, MAIMON O. Incremental clustering of mobile objects [C]// Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data. New York: ACM, 2007: 593 – 604.
- [15] YUAN G, SUN P, ZHAO J, et al. A review of moving object trajectory clustering algorithms [EB/OL]. [2015-01-09]. https://www.researchgate.net/publication/299434359_A_review_of_moving_object_trajectory_clustering_algorithms.

This work was partially supported by Natural Science Foundation of Hebei Province (F2016202144), the Science and Technology Research Project of Colleges and Universities in Hebei Province (ZD2014030), the Tianjin Research Project of Application Foundation and Advanced Technology (14JCZD JC31600).

SHI Lukui, born in 1974, Ph. D., professor. His research interests include machine learning, data mining.

ZHANG Yanru, born in 1990, M. S. candidate. Her research interests include machine learning, data mining.

ZHANG Xin, born in 1992, M. S. candidate. Her research interests include machine learning, data mining.