

DOI: 10.3901/JME.2021.02.239

# 板带力学性能预报的集成学习模型及其可靠性评价<sup>\*</sup>

李飞飞<sup>1,2</sup> 宋 勇<sup>1,2</sup> 刘 超<sup>1,2</sup> 李 博<sup>1,2</sup> 张世伟<sup>1,2</sup>

(1. 北京科技大学高效轧制国家工程研究中心 北京 100083;

2. 北京科技大学国家板带生产先进装备工程技术研究中心 北京 100083)

**摘要:** 热轧板带力学性能预报具有十分广阔的应用前景。但目前的力学性能预报模型精度不高、泛化能力较差,且无法评价预测结果的准确性,限制了其在实际应用中的效果。为了提高力学性能预报模型的精度和实现预测结果的可靠性评价,采用 BP 神经网络将目前直接预测力学性能结果的建模方法转换为对样本间的力学性能偏差分布预测的建模,并结合模型预测结果分布的离散程度设计可靠性评价指标,同时采用集成学习提高模型的泛化能力。通过试验验证,该集成学习模型具有较高的预测精度,进一步,根据可靠性评价指标分析,除了其中占比约 3.5% 样本的预测结果具有较大不确定性,剩余样本的屈服强度和抗拉强度在误差 $\pm 30$  MPa 的预测准确率达到 98.45% 和 98.97%,延伸率在误差 $\pm 5\%$  下的预测准确率达到 99.48%,有效地提高了模型预测准确率,在生产现场应用中具有一定的指导意义。

**关键词:** 偏差预测; 力学性能预测; 神经网络; 集成学习; 可靠性评价

**中图分类号:** TG156

## Ensemble Learning Model for Mechanical Performance Prediction of Strip and Its Reliability Evaluation

LI Feifei<sup>1,2</sup> SONG Yong<sup>1,2</sup> LIU Chao<sup>1,2</sup> LI Bo<sup>1,2</sup> ZHANG Shiwei<sup>1,2</sup>

(1. National Engineering Research Center for Advanced Rolling Technology,

University of Science and Technology Beijing, Beijing 100083;

2. National Engineering Technology Research Center of Flat Rolling Equipment,

University of Science and Technology Beijing, Beijing 100083)

**Abstract:** The prediction of mechanical properties of hot rolled strips has a very broad application prospect. However, the current mechanical performance prediction model is not accurate and the generalization ability is poor, and the accuracy of the prediction results cannot be evaluated, which limits its effectiveness in practical applications. In order to improve the accuracy of the mechanical performance prediction model and the reliability evaluation of the prediction results, BP neural network is used to transform the current modeling method of direct prediction of mechanical performance results into the modeling of mechanical property deviation distribution prediction between samples, and combined with the model. The degree of dispersion of the prediction results is designed to evaluate the reliability evaluation index, and ensemble learning is used to improve the generalization ability of the model. Through experimental verification, the ensemble learning model has higher prediction accuracy. Further, according to the reliability evaluation index analysis, except for the prediction results of about 3.5% of the samples, there is greater uncertainty, and the yield strength and tensile strength of the remaining samples. The prediction accuracy of the intensity  $\pm 30$  MPa is 98.45% and 98.97%, and the prediction accuracy of the elongation under the error of  $\pm 5\%$  is 99.48%, which effectively improves the accuracy of model prediction and has certain performance in production field applications. Guiding significance.

**Key words:** deviation prediction; mechanical performance prediction; neural network; ensemble learning; reliability evaluation

<sup>\*</sup> 国家自然科学基金(51674028)和广西创新驱动发展专项资金计划(GKAA17202008)资助项目。20200501 收到初稿, 20201030 收到修改稿

## 0 前言

热轧板带力学性能预测是指利用热轧产品化学成分和相关生产工艺参数建立能够准确预测板带最终力学性能的预测模型,对热轧板带力学性能预报模型的研究最早可以追溯至 20 世纪 50 年代, Irvine 和 Pickering 提出采用数学模型预测板带组织演变和最终力学性能<sup>[1]</sup>,在此之后,世界各国的冶金工作者在这方面开展了大量的模型研究工作<sup>[2-6]</sup>。这些机理模型大多是在实验室研究基础上建立的,由于实际现场生产数据存在较大的误差<sup>[7]</sup>,数据质量难以保证,导致其应用的效果不佳。

在数据量和算法逐渐丰富的今天,各种机器学习特别是深度学习模型相继出现,在一定程度上克服了冶金机理模型对数据质量要求高的问题。如马湧等<sup>[8]</sup>利用 BP 神经网络将带钢抗拉强度、屈服强度和延伸率在 10% 误差内的预测精度分别做到了 93.4%、92.1% 和 90.5%,贾涛等<sup>[9]</sup>基于改进的神经网络对集装箱热轧板力学性能进行预测,在  $\pm 6\%$  误差下的预测精度超过了 93%。但是,这种直接通过历史数据样本所建立的力学性能预报模型的泛化能力较差,模型预测结果的准确性和稳定性均无法得到有效保证<sup>[10-11]</sup>,而且过度追求高精度<sup>[12]</sup>,缺少预测结果的可靠性评价方法,阻碍了其在实际应用中的推广。

为了提高力学性能预报模型的泛化能力,并实现预测结果的可靠性评价,本文将目前直接预测力学性能结果的建模方法转换为对样本间力学性能偏差预测的建模,从训练样本集里随机均匀抽取若干子数据集,分别采用 BP 神经网络建立多个可预测任意两个样本间差异的基础子模型,并通过具有强泛化能力的 Bagging 集成学习算法<sup>[13-17]</sup>将这些子模型进行融合,以提高预测结果的准确率和泛化能力。在实际应用中,采用所建立的基础子模型,分别将建模时使用的训练样本作为比对样本,就可得到待测样本力学性能值的多个预测结果组成的集合,以集合中预测结果的均值作为最终力学性能预测值,而集合中预测结果的离散程度反映了该子模型此次预测的可靠性水平。在给定预测误差区间下,采用区间覆盖率(Prediction intervals coverage probability, PICP)作为评价模型预测结果可靠性的指标,根据各子模型预测结果的评价指标构建集成学习模型。另外,通过该评价指标可以识别该模型无法可靠预测的生产数据样本,设置较高的评价指标阈值能够提

升模型预测的准确性,但会导致模型给出不能可靠预测的样本数目增多。经试验验证,该集成学习模型具有较高的预测准确率,采用 PICP 指标能够比较精确地识别出测试集中存在的少量模型不能准确预测的样本,具有重要的现场指导意义。

## 1 构建集成学习模型

本文以某钢厂薄板坯连铸连轧 CSP(Compact strip production, CSP)生产线的相关生产数据为例,在不区分钢种的情况下,实现力学性能预测及可靠性评价,如图 1 所示,本文将主要从以下五个过程来构建集成 BP 网络模型。

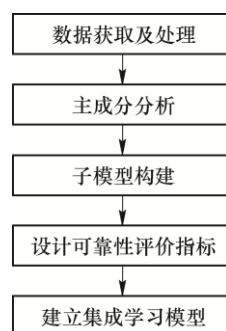


图 1 集成 BP 网络建模过程

### 1.1 数据获取及处理

获取到数据的力学性能采样点位于带钢尾部,因此对数据中的参数进行如下筛选:选取 C、Mn、P、S、Cr、Al、O、Nb、V、Ti、Mo 共 11 个化学成分<sup>[18]</sup>及各精轧机架入口带钢尾部温度、轧制力,以及原料厚度、出炉温度、终轧厚度、终轧速度、终轧温度、卷取温度共 6 个重要工艺参数<sup>[19]</sup>。共筛选出 31 个与热轧带钢力学性能相关的参数,其中在正式建模之前,需要对数据进行一些必要的处理,以避免异常数据及数据量纲不统一对后续建模的影响,主要涉及异常值处理<sup>[20-21]</sup>和规范化处理,为避免后续建模过拟合,采用 PCA(Principle component analysis, PCA)<sup>[22]</sup>进行降维处理。

本文采用 PCA 对初选的 31 个属性进行分析,图 2 展示了经 PCA 处理后新数据集的累积特征贡献率。

按照累积贡献率的大小选取 PCA 处理后的数据集,累积贡献率过大可能会将样本中的噪声数据选入,而贡献率过小又会丢失原数据中大量的重要信息,为了避免过多地丢失重要信息且尽可能少的选入噪声数据,本文选取累积贡献率约为 93%,最终得到 13 个特征共 3 300 条数据用于建模分析。

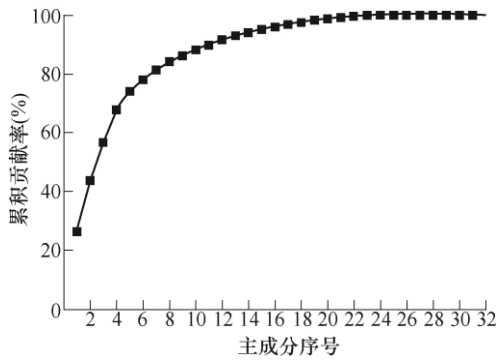


图2 PCA处理后数据累积特征贡献率

## 1.2 子模型构建

为了得到对模型预测结果的可靠性评价, 本文将力学性能预测问题转换为力学性能偏差估计, 通过获得待测样本与多个对照样本的力学性能偏差, 可确定待测样本力学性能分布的估计, 以该分布的均值为预测值, 而其离散程度来反映预测结果的可靠性。实际上, 力学性能偏差可以认为是样本间距离的映射, 可以采用神经网络进行逼近。误差逆传播(Back propagation, BP)是迄今最成功的神经网络学习算法, HORNIK 等<sup>[23]</sup>在 1989 年证明只需一个包含足够多神经元的隐层, 多层前馈网络就能以任意精度逼近任意复杂度的连续函数。如图 3 所示, 本文采用的 BP 神经网络隐层数目设置为 1 个, 由输入层、隐层和输出层三层构成。

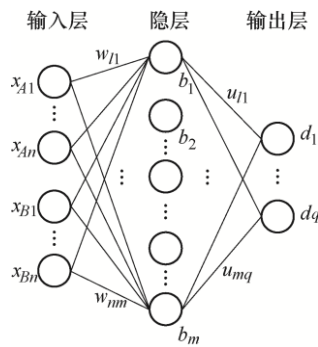


图3 用于偏差预测的BP网络设计

### 1.2.1 模型参数确定

为了能反映样本间的距离, 本文将任意 A、B 两个样本的所有特征值( $X_{A1} \sim X_{An}$ )、( $X_{B1} \sim X_{Bn}$ )作为 BP 神经网络的输入层, 而输出层为它们的力学性能偏差。选择具有良好非线性映射能力的 Sigmoid 函数作为神经元的激活函数, 该函数常用于二分类, 其值表示为类别概率, 通常可利用多个神经元实现多分类的效果。为此, 如果将不同性能偏差值看作不同类别, 可结合遗传算法中二进制编码的原理, 将样本间性能偏差值以二进制方式表示作为神经网络的输出层。CSP 热连轧产品大纲中任

意两块带钢力学性能极差一般在 500 MPa 以内, 因此可使用 10 位二进制编码表示力学性能偏差, 其中一位作为符号位, 其余 9 位表示偏差绝对值, 具体如图 4 所示。

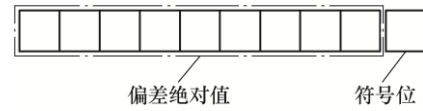


图4 偏差编码规则

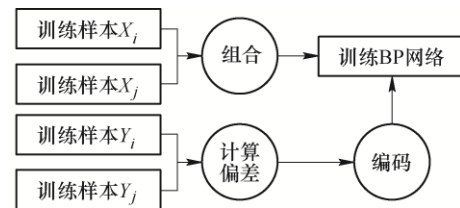
通过分析, 本文选取输入层神经元为 26 个, 输出层神经元为 10 个。另外, 为了达到比较好的拟合效果, 本文将隐层神经元数目设定为 27 个, 迭代次数为 80 000, 学习率为 0.05。

### 1.2.2 输入数据采样及处理

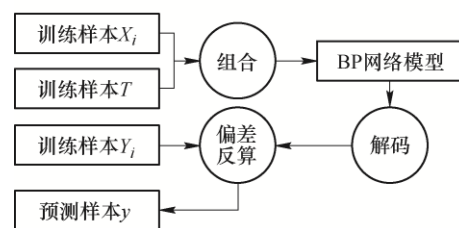
由于分类算法一般都假设数据分布是均匀的, 而实际的生产数据很大程度上是不均匀的, 因此需要在训练模型之前对数据进行均匀采样<sup>[24]</sup>。对此, 本文将经过 PCA 处理后的数据随机划分为训练集和测试集, 然后将训练集按照产品力学性能范围均分为 10 个性能区间, 从训练集的每一个性能区间中各随机抽取 60 条数据组成子模型训练集。同时, 将子模型训练集中任意两个样本进行组合作为 BP 网络输入层的训练数据, 从而每个子模型实际上得到了 360 000 个训练样本。

### 1.2.3 模型训练与预测

如图 5a、5b 分别为 BP 子模型的训练和预测过程。图中  $X_i$ 、 $X_j$  分别表示  $i$ 、 $j$  两个训练样本的特征值矢量作为 BP 网络的输入层参数,  $Y_i$ 、 $Y_j$  为对应训练样本力学性能实测值, 计算出差值并进行二进制编码后作为训练 BP 网络的期望输出。对于某



(a) BP子模型训练过程



(b) BP子模型预测过程

图5 BP子模型训练及预测过程

个测试样本  $T$ , 首先将训练样本  $X_i$  作为参考样本与  $T$  组合, 输入到训练好的 BP 模型进行预测, 得到  $X_i$  与  $T$  的力学性能偏差预测值, 然后根据  $Y_i$  可计算得到  $T$  的力学性能预测结果  $y_i^1$ , 接着交换  $X_i$  与  $T$  的输入位置, 预测得到  $T$  与  $X_i$  的力学性能偏差预测值, 再次根据  $Y_i$  计算得到  $T$  的力学性能预测结果  $y_i^2$ , 重复上述步骤, 最终得到 1 200 个力学性能预测结果  $\{y_1^1, y_1^2, \dots, y_{600}^1, y_{600}^2\}$ , 将该结果集的均值作为  $T$  的力学性能预测值。

### 1.3 设计可靠性评价指标

传统的力学性能预报系统大多是以预报一个力学性能数值为目标, 这种模型的预报精度受制于数据的质量, 且外延性差<sup>[12]</sup>。由于现场生产数据存在随机误差, 并且这些误差也会对模型训练造成一定干扰, 因此, 力学性能模型预报结果需要进行误差区间及可靠性评估。对于一个待测样本, 本文所建立的 BP 网络子模型能给出的是一个力学性能预测结果集, 该集合中预测结果分布的离散程度能够反映其对该样本预测的可靠性, 可实现模型在一定预测误差要求下的可靠性评价。

图 6 是 BP 神经网络对于某个待测样本的力学性能预测结果分布示意图。为了便于后文分析, 本文将预测结果集的均值记为  $\sigma_{\text{mean}}$ , 设置可靠区间为  $[\sigma_{\text{mean}} - \Delta\sigma, \sigma_{\text{mean}} + \Delta\sigma]$ , 其中  $\Delta\sigma$  表示预测值的允许误差, 可靠性评价指标可以设计为预测值在可靠区间内的数目占比(即区间覆盖率, PICP), 定义式<sup>[25]</sup>

$$\text{PICP} = \frac{1}{n} \sum_{i=1}^n c_i$$

式中, 当子模型预测结果  $y$  处于区间  $[\sigma_{\text{mean}} - \Delta\sigma, \sigma_{\text{mean}} + \Delta\sigma]$  时,  $c_i = 1$ , 否则  $c_i = 0$ 。

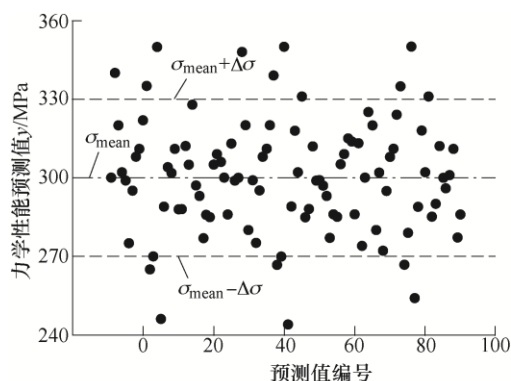


图 6 BP 网络预测值分布示意图

### 1.4 建立集成 BP 网络模型

如图 7 所示, 从训练样本集里随机抽取若干子训练集, 分别通过不同的子训练集训练得到多个 BP 模型(子模型), 可采用 Bagging 算法将多个子模型对

同一样本的预测结果进行集成处理输出。Bagging 是一种具有代表性的并行式集成学习算法, 该算法的原理是在同一数据集下每次有放回地选取一定量的样本用于模型训练, 基于每个采样集训练得到子模型, 再将这些子模型通过某种组合策略进行组合, 最终实现集成预测。

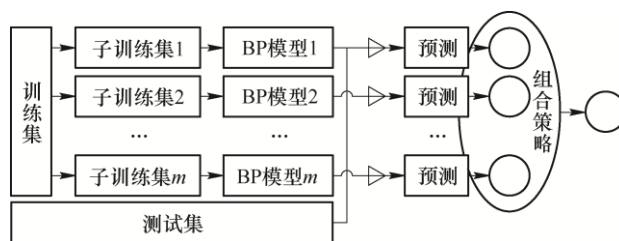


图 7 集成 BP 网络示意图

通过每个子模型可得到待测样本的力学性能预测结果  $\sigma_{\text{mean}}$  和 PICP 指标, 而 PICP 指标反映了  $\sigma_{\text{mean}}$  的可靠性。以待测样本  $k$  为例, 将第  $i$  个子模型预测的结果记为  $\sigma_i(k)$ , 对应的 PICP 指标记为  $p_i(k)$ , 集成学习后的待测样本  $k$  的力学性能预测结果  $\sigma_{\text{out}}(k)$  可设为所有  $m$  个子模型输出结果中  $p_i(k)$  取值最大时所对应的  $\sigma_i(k)$ , 而其最终 PICP 指标  $\sigma_{\text{out}}(k)$  则为  $\sigma_i(k)$  的最大取值。

## 2 试验结果及分析

### 2.1 子模型数量确定

图 8 是集成学习模型对带钢力学性能预测的准确率与子模型数量的关系, 集成学习能够显著提升子模型的预测准确率, 随着子模型数目的不断增加, 集成学习的准确率逐渐趋于稳定, 为保证模型的准确率及稳定性, 本文从学习曲线得到屈服强度、抗拉强度和延伸率的集成学习模型对应的子模型数目分别为 420、500 和 200。

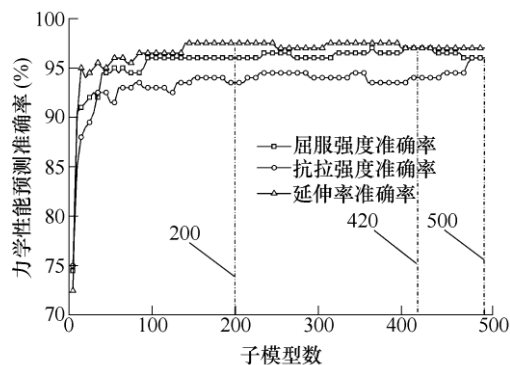
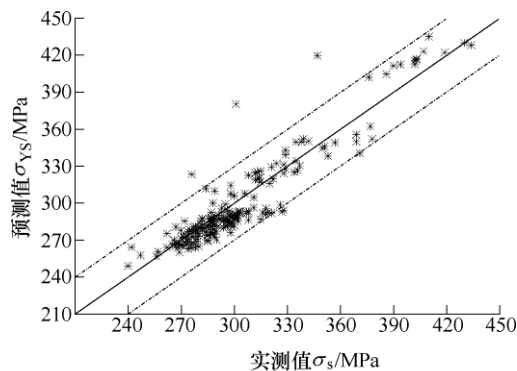


图 8 集成学习模型对带钢力学性能预测的准确率与子模型数量的关系

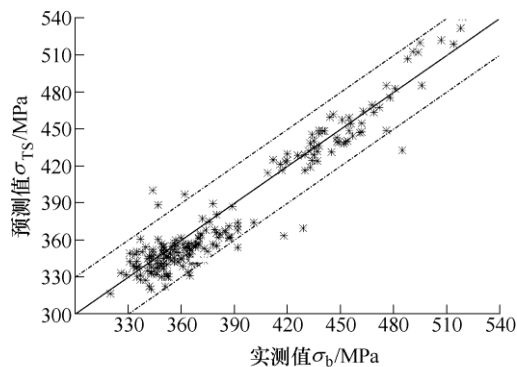


## 2.2 集成学习预测结果分析

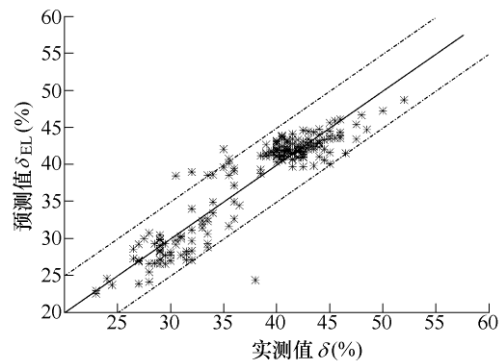
图 9 是采用集成学习模型对 200 个测试样本力学性能预测值和实测值的比较, 图中虚线表示误差界限( $\pm 30$  MPa 和  $\pm 5\%$ ), 预测结果的准确率超过了 95%。



(a) 屈服强度实测值与预测值对比



(b) 抗拉强度实测值与预测值对比



(c) 延伸率实测值与预测值对比

图 9 集成学习模型对带钢力学性能预测的结果

为了进一步验证集成学习模型对力学性能的预测精度和稳定性, 本文从原数据集随机抽取 200 个测试样本, 再利用其余样本来训练集成学习模型, 共重复以上过程 5 次, 得到 5 个集成 BP 网络模型的预测准确率如表 1 所示。从表 1 中可以看出, 集成学习模型对力学性能预报的准确率均可达到 95% 以上, 具有较高的预测精度和稳定性。

表 1 集成 BP 网络预测准确率 %

性能类型	测试集编号					综合
	1	2	3	4	5	
屈服强度	97.0	93.5	94.0	96.0	95.0	95.1
抗拉强度	96.0	94.0	95.5	94.5	96.0	95.2
延伸率	96.0	98.0	98.0	97.5	98.0	97.5

## 2.3 PICP 指标与力学性能预测误差之间的关系

图 10 是采用单个 BP 模型时各测试样本屈服强度预测误差与 PICP 指标与之间的关系。从总体趋势来看, 测试样本预测值的 PICP 越大, 模型预测结果的误差越小。当样本 PICP 达到 40% 以上时, 该部分测试数据完全能够在  $\pm 30$  MPa 误差区间内被准确预测。随着 PICP 的下降, 预测误差超过  $\pm 30$  MPa 的样本数目逐渐增多, 同时平均误差也在不断增大。

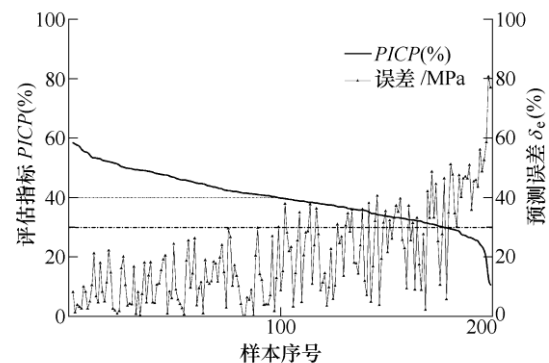


图 10 屈服强度预测误差与 PICP 的关系

图 11 是集成学习模型的预测性能, 通过分析依然可以得到和图 10 较为一致的结论, 相比图 10, 集成学习模型的预测结果准确率明显提高, 误差大幅度减小。而且, 集成学习模型对于样本预测值的 PICP 达到 55% 时, 对应样本的预测误差范围甚至能够缩小到  $\pm 20$  MPa 以内, 平均预测误差低至 6.62 MPa。

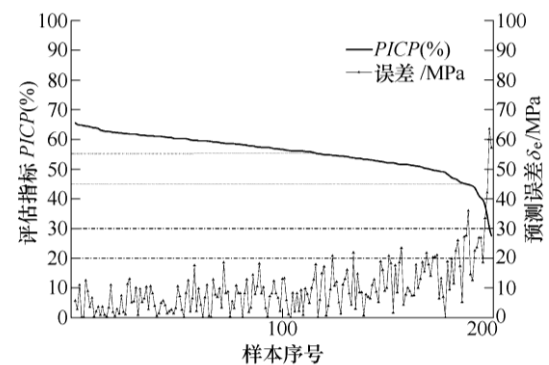


图 11 集成学习后屈服强度预测误差与 PICP 的关系

此外,虽然采用集成学习对 BP 网络模型的综合预测精度得到了很大提高,但依然会存在极少数样本预测值对应的 PICP 低于 40%,其对应样本的预测结果误差超过了 $\pm 40$  MPa。因此,PICP 可以在一定程度上反映模型预测值的可靠程度,PICP 越大,对应预测结果越可靠,误差越小。

从试验结果来看,当 PICP 大于阈值 45%时,表明集成学习模型的屈服强度预测值 100%能够落在 $\pm 30$  MPa 误差区间内。换句话说,当集成学习模型预测结果的 PICP 指标大于 45%时,可以说明其预测出来的力学性能结果在 $\pm 30$  MPa 误差区间具有非常高的可信度,即能够确定其可以准确预报这部分样本数据。设置较高的 PICP 阈值能够提升模型预测的准确性,但会导致模型给出不能可靠预测的样本数目增多,而不可靠样本一般需要人工取样对预测结果进行进一步检验,因此在生产应用中可以根据实际要求来优化设置 PICP 阈值。本文以 CSP 相关热轧数据为例,将屈服强度、抗拉强度和延伸率的 PICP 阈值分别设置为 44%、40%和 50%,即认为当 PICP 高于此阈值时的模型预测值是比较可靠的,按照该阈值对测试数据进行筛选分析,得到该模型能够可靠预测屈服强度的样本数占比为 97%,预测的准确率为 98.45%,抗拉强度可靠样本占比为 97%,准确率为 98.97%,延伸率可靠样本占比为 96.5%,准确率达到了 99.48%。

### 3 讨论与分析

采用样本间力学性能偏差预测的建模本质上是对样本间距离分布进行建模,使得属于同类样本靠近,异类样本远离,而所谓的同类和异类样本在数据上的直观表现恰好是实际性能差异的大小。简单来看,其与 KNN 有着一定的相似性,但是,KNN 的效果依赖距离度量的选取,一般采用的是一个比较随意的距离计算(如欧氏距离、马氏距离、夹角余弦等),其距离度量方式难以反映实际数据中的复杂关系。而本文的建模方式,其距离度量函数采用一定的网络结构来学习,能够依赖实际任务的复杂性进行调整,从而得到相对准确的一个反映参数差异到性能差异的映射关系。

### 4 结论

(1) 通过 BP 神经网络将直接预测力学性能结果的建模方法转换为对样本间力学性能偏差预测的

建模,并采样 bagging 集成学习对 BP 网络子模型进行融合,具有较高的预测性能,该模型对屈服强度和抗拉强度在误差 $\pm 30$  MPa、延伸率在误差 $\pm 5\%$ 下的平均预测准确率分别达到了 95.1%、95.2%和 97.5%。

(2) BP 子模型的预测值数据集在给定预测误差区间的覆盖率(PICP)可用于评价其预测结果的准确性。当 PICP 指标大于 45%时,其预测出来的力学性能结果在 $\pm 30$  MPa 误差区间具有非常高的可信度。

(3) 设置较高的 PICP 阈值能够提升模型预测的准确性,但会导致模型给出不能可靠预测的样本数目增多,在生产应用中可以根据实际要求来优化设置 PICP 阈值。本文以 CSP 热轧生产数据为例,将 PICP 分别设置为 44%、40%和 50%时,该集成学习模型可识别出大约 3.5%的少量样本无法给出确定性预测值,对于剩余样本其屈服强度和抗拉强度在误差 $\pm 30$  MPa 的预测准确率达到 98.45%和 98.97%,对延伸率在误差 $\pm 5\%$ 下的预测准确率达到 99.48%。

### 参 考 文 献

- [1] 干勇. 薄板坯连铸连轧(TSCR)热轧过程组织性能预报技术的开发[J]. 钢铁, 2003, 38(8): 10-15.  
GAN Yong. The R&D of process modeling in thin slab hot strip rolling[J]. Iron & Steel, 2003, 38(8): 10-15.
- [2] 余万华, 韩静涛, 余广夫, 等. HSMM 软件在攀枝花钢铁公司热轧板厂的应用[J]. 钢铁研究学报, 2006(11): 60-62.  
YU Wanhua, HAN Jingtao, SHE Guangfu, et al. Application of HSMM software in hot strip mill of panzhihua steel company[J]. Journal of Iron and Steel Research, 2006(11): 60-62.
- [3] GHRIBERNIG J, GHUBMER D. For the first time ever: Full metallurgical control of the mechanical properties of hot-rolled strip with VAI-Q strip[J]. Iron & Steel, 2001(01): 42-46.
- [4] 王蕾. 热轧带钢的相变和力学性能模型研究及应用[D]. 北京: 北京科技大学, 2017.  
WANG Lei. Research and application of phase transformation and mechanical properties models in hot strip rolling[D]. Beijing: University of Science and Technology Beijing, 2017.
- [5] 郑坚, 唐广波, 程杰锋, 等. 珠钢 CSP 热轧组织性能预报软件设计[J]. 轧钢, 2003(1): 15-17.

- ZHENG Jian, TANG Guangbo, CHENG Jiefeng, et al. Software design of microstructure and property prediction system for CSP line of Zhujiang Steel Co[J]. Steel Rolling, 2003(1): 15-17.
- [6] 刘正东, 董瀚, 干勇. 热连轧过程中组织性能预报系统的应用[J]. 钢铁, 2003(2): 68-71.
- LIU Zhengdong, DONG Han, GAN Yong. Application of process modeling to hot strip rolling[J]. Iron & Steel, 2003(2): 68-71.
- [7] 郭朝晖, 张群亮, 苏异才, 等. 关于热轧带钢力学性能预报技术的思考[J]. 冶金自动化, 2009, 33(2): 1-6.
- GUO Zhaohui, ZHANG Qunliang, SU Yicai, et al. Thoughts on mechanical property prediction of hot rolled strip[J]. Metallurgical Industry Automation, 2009, 33(2): 1-6.
- [8] 马湧, 王晓鹏, 马莎莎. 基于 Keras 深度学习框架下 BP 神经网络的热轧带钢力学性能预测[J]. 冶金自动化, 2019, 43(2): 6-10.
- MA Yong, WANG Xiaopeng, MA Shasha. Prediction of mechanical properties of hot rolled strip based on BP neural network under Keras deep learning framework[J]. Metallurgical Industry Automation, 2019, 43(2): 6-10.
- [9] 贾涛, 刘振宇, 胡恒法, 等. 基于贝叶斯神经网络的 SPA-H 热轧板力学性能预测[J]. 东北大学学报, 2008(4): 521-524.
- JIA Tao, LIU Zhenyu, HU Hengfa, et al. Mechanical property prediction for hot rolled SPA-H steel using bayesian neural network[J]. Journal of Northeastern University, 2008(4): 521-524.
- [10] 李维刚, 杨威, 赵云涛, 等. 融合大数据与冶金机理的热轧带钢力学性能预报模型[J]. 钢铁研究学报, 2018, 30(4): 302-308.
- LI Weigang, YANG Wei, ZHAO Yuntao, et al. Mechanical property prediction model of hot-rolled strip via big data and metallurgical mechanism analysis[J]. Journal of Iron and Steel Research, 2018, 30(4): 302-308.
- [11] 李维刚, 刘超, 杨威, 等. 热轧微合金钢力学性能预报模型研究[C]//第十一届中国钢铁年会论文集——S18: 冶金自动化与智能管控. 北京: 中国金属学会, 2017: 1-8.
- LI Weigang, LIU Chao, YANG Wei, et al. Prediction model of mechanical properties of hot-rolled micro-alloyed steel[C]//Proceedings of the 11th CSM Steel Congress: Metallurgical Automation and Intelligent Control. Beijing: The Chinese Society for Metals, 2017: 1-8.
- [12] 郭朝晖, 苏异才, 张群亮, 等. 热轧带钢性能预报技术研究中的几个误区[J]. 轧钢, 2013, 30(1): 29-32.
- GUO Zhaohui, SU Yicai, ZHANG Qunliang, et al. Wrong cognitions on research of hot-rolled strip property prediction[J]. Steel Rolling, 2013, 30(1): 29-32.
- [13] BREIMAN L. Bagging predictors[J]. Machine Learning, 1996, 24(2): 123-140.
- [14] 闫飞宇, 李伟卓, 杨卫卫, 等. 基于 Bagging 神经网络集成的燃料电池性能预测方法[J]. 中国科学: 技术科学, 2019, 49(4): 391-401.
- YAN Feiyu, LI Weizhuo, YANG Weiwei, et al. Prediction of fuel cell performance based on Bagging neural network ensemble model[J]. Scientia Sinica Technologica, 2019, 49(4): 391-401.
- [15] 谢琪, 程耕国, 徐旭. 基于神经网络集成学习股票预测模型的研究[J]. 计算机工程与应用, 2019, 55(8): 238-243.
- XIE Qi, CHENG Gengguo, XU Xu. Research based on stock predicting model of neural networks ensemble learning[J]. Computer Engineering and Applications, 2019, 55(8): 238-243.
- [16] 王征宇. 神经网络集成分类方法及其在并行计算环境中的应用研究[D]. 广州: 华南理工大学, 2015.
- WANG Zhengyu. Research on neural network ensemble classification methods and their applications in parallel computing environment[D]. Guangzhou: South China University of Technology, 2015.
- [17] 于彬, 李珊, 陈成, 等. 基于集成学习的人类 LncRNA 大数据基因预测[J]. 青岛科技大学学报, 2018, 39(1): 106-113.
- YU Bin, LI Shan, CHEN Cheng, et al. Prediction of human lncRNA big data genes based on ensemble learning[J]. Journal of Qingdao University of Science and Technology, 2018, 39(1): 106-113.
- [18] 张祝亭. 基于数据挖掘的马钢 CSP 热轧板卷的性能预测[D]. 马鞍山: 安徽工业大学, 2012.
- ZHANG Zhuting. The prediction of masteel CSP hot rolled strip based on data mining[D]. Maanshan: Anhui University of Technology, 2012.
- [19] 刘贵立, 张国英, 曾梅光. 用人工神经网络模型研究微量元素对钢力学性能的影响[J]. 钢铁研究, 2000(1): 48-50.
- LIU Guili, ZHANG Guoying, ZENG Meiguang. Studies

- on effects of trace elements on mechanical properties of steel by artificial nerve network model[J]. Research on Iron & Steel, 2000(1): 48-50.
- [20] LIU Feitong, TING Kaiming, ZHOU Zhihua. Isolation forest[C]//2008 Eighth IEEE International Conference on Data Mining, December 15-19, 2008. Pisa: IEEE, 2009: 413-422.
- [21] LIU Feitong, TING Kaiming, ZHOU Zhihua. Isolation-based anomaly detection[J]. ACM Transactions on Knowledge Discovery From Data, 2012, 6(1): 1-39.
- [22] 赵帅, 黄亦翔, 王浩任, 等. 基于随机森林与主成分分析的刀具磨损评估[J]. 机械工程学报, 2017, 53(21): 181-189.
- ZHAO Shuai, HUANG Yixiang, WANG Haoren, et al. Random forest and principle components analysis based on health assessment methodology for tool wear[J]. Journal of Mechanical Engineering, 2017, 53(21): 181-189.
- [23] HORNIK KM, STINCHCOMB M, WHITE H. Multilayer feedforward networks are universal approximators[J]. Neural Networks, 1989, 2(5): 359-366.
- [24] 从威. 非均衡分类的集成学习应用研究[D]. 南京: 南京信息工程大学, 2017.
- CONG Wei. Application research on ensemble learning of unbalanced classification[D]. Nanjing: Nanjing University of Information Science & Technology, 2017.
- [25] KHOSRAVI A, NAHAVANDI S, CREIGHTON D. Construction of optimal prediction intervals for load forecasting problems[J]. IEEE Transactions on Power Systems, 2010, 25(3): 1496-1503.

作者简介: 李飞飞, 男, 1993 年出生, 博士研究生。主要研究方向为复杂工业过程建模与优化控制。

E-mail: 15311449327@163.com

宋勇(通信作者), 男, 副研究员, 硕士研究生导师。主要研究方向为轧制过程控制模型及智能化。

E-mail: songyong@ustb.edu.cn