

深度学习在中厚板轧后超快速冷却系统中的研究与应用

张 田, 张子豪, 田 勇, 王昭东
(东北大学 轧制技术及连轧自动化国家重点实验室, 辽宁 沈阳 110819)

摘 要: 传热系数是冷却控制模型的核心参数. 传统学习模型对传热系数的修正存在不稳定、鲁棒性不强等问题. 为解决上述问题, 进一步提高控制精度, 基于深度学习技术, 建立了传热系数自学习的深度神经网络, 对神经网络框架的超参数进行了优化和算法选型, 增强控冷模型的稳定性. 通过在某钢厂 3 500 mm 中厚板生产线的应用验证, 采用深度学习的控冷模型对终冷温度预报精度有明显提高, 鲁棒性较强, 满足现场实际生产的需要.

关 键 词: 深度学习; 中厚板; 超快冷; 传热系数; 命中率

中图分类号: TP **文献标志码:** A **文章编号:** 1005-3026(2019)05-0635-06

Research and Application of Deep Learning Method for Plate Ultra Fast Cooling

ZHANG Tian, ZHANG Zi-hao, TIAN Yong, WANG Zhao-dong
(State Key Laboratory of Rolling and Automation, Northeastern University, Shenyang 110819, China.
Corresponding author: ZHANG Zi-hao, E-mail: 215070897@qq.com)

Abstract: The heat transfer coefficient (HTC) is a key parameter to determine control precision in a cooling-controlled model. However, the correction for the HTC isn't always stable and robustness by a traditional self-learning method. In order to solve such problems, the deep neural networks for the self-learning HTC were built based on the deep learning technology. The optimization of hyper parameter and the selection of algorithm in the network frame are studied. Therefore, the stability of the cooling control model can be greatly enhanced. The application of the 3 500 mm plate mill plant proves the better accuracy and robustness of the model, which can meet the requirements of actual on-site production.

Key words: deep learning; plate; ultra fast cooling; heat transfer coefficient; hit rate

随着 TMCP 工艺的广泛应用与不断创新, 作为 TMCP 工艺的主要组成部分控冷技术也取得快速发展. 其原理为通过改变轧后冷却条件来控制碳化物和相变的析出行为, 进而改善钢板的组织性能^[1], 以充分挖掘钢材潜力, 提高钢材强度, 改善其塑性和焊接性能. 在这一工艺中, 最重要的环节是冷却控制过程模型的建立, 这决定了最终温度控制的稳定与否. 目前国内大多数厚板线轧后控冷模型主要包括预计算模块、在线修正模块和自学习模块. 在很大程度上, 中厚板终冷温度的

控制精度由控冷模型中温度场模型的精准度所决定. 其中, 水冷传热系数是影响温度场模型的核心参数. 在实际生产中, 影响水冷传热系数的因素多且复杂, 如板材规格、化学成分、冷却水温及水流密度等, 层流冷却过程具有很强的动态随机性, 是复杂的非线性过程, 很难用精确的数学模型对水冷传热系数进行描述和控制^[2]. 只有当温控模型具备较强的自学习能力时, 才能适应复杂的工业生产条件. 对此, 只能通过建立自学习模型来优化控冷模型. 国内外很多学者都致力于温控模型的

优化研究,提出了很多具有一定适应性的工业温控模型.

一些学者基于控制规则和知识推理的方法提出了模糊控制模型^[3],其对于不明确规律机理的控制具有较强的适用性,但是模糊算法存在一些不足:过于依赖人工经验,并且由于静态误差的存在,会导致控制精度的大幅下降.另外还有一种基于层别划分的温控自学习模型^[4],具体做法是将冷却钢板的厚度、目标终冷温度及水温等条件参数,按照一定规则进行分类,相近属性的归为一类,新的待冷却钢板根据层别划分,参考学习已冷却钢板的历史实绩.但这种方法的局限在于受到划分的层别数量的限制:层别过细所需学习样本过多,层别过粗会出现层别跃迁现象,严重影响控制精度.除此以外,一些基于实例的智能算法也运用到统计预测领域. k 近邻法^[5]就是一种基于实例进行分类预测的经典算法.然而, k 近邻法也具有自身的局限亟待解决:一是缺乏考虑每个影响因素对冷却过程的真实影响程度;二是随着参考因素或样本数量的增加,会造成学习的数量成倍增长,大大降低了计算效率.

上述自学习模型算法都有其自身的特点,也同时存在局限性.针对目前主流自学习模型存在的问题以及现场对模型精度和稳定性提出的要求,提出一种新的方法,即深度学习.

1 深度神经网络研究

深度学习是机器学习中新兴的研究领域,其网络架构由多层非线性运算单元组成,可以从大数据中挖掘有效的特征表示,是一种从数据中挖掘有效特征的好方法,为解决复杂非线性问题以及机理未知过程的预测和控制提供了新的途径^[6].

1.1 深度神经网络关键参数的选择

首先确立输出层单元.控制模型中水冷传热系数与许多物理参数有关,如终轧温度、板材厚度、水流密度和水温等.水冷传热系数与这些物理参数之间存在着复杂的非线性关系,很难确定它们之间的函数关系.因此,水冷传热系数作为一个关键参数,它的精度直接影响到终冷温度的控制精度,所以,将水冷传热系数作为深度神经网络的输出层单元.

其次是确立输入层单元.输入层单元主要由影响水冷传热系数的相关参数所确定,由某中板厂数据库系统中初步采集出板材宽度、板材厚度、板材长度、化学成分等 20 多个影响因子组成.

采用主成分分析法对数据进行预处理,选出 11 个权重较大的影响因子(包括板材厚度、 $w(C)$ 、 $w(Cr)$ 、 $w(Cu)$ 、 $w(Mn)$ 、冷却水温度、空气温度、终轧温度、目标终冷温度、集管流量、辊速)作为神经网络的输入层参数导入神经网络模型.

选取隐含层层数、隐含层单元数、传播算法和正则化方法 4 个变量作为隐含层实验变量,并采用单一变量法训练深度神经网络对水冷传热系数因子进行预报,筛选出最优模型结构.

1.2 深度神经网络结构的建立

传统神经网络模型一般都是单隐含层,虽然单隐含层运算速度较快^[7],但运算精度不佳.而深度神经网络可分层逐级地挖掘输入数据的特征表示,有效降低了模型所需参数数目,并且增大了模型运算精度.因此,将各组实验的隐含层层数设置为 1~7 层,神经元数分别设置为 50~300 个,以 10 个为公差分别进行实验.选取几个特殊组合作比较,结果如表 1 所示.

表 1 不同隐含层层数、单元数对训练精度的影响				
Table 1 Effects of the number of hidden layers and hidden layer units on training precision				
单元数/个	1 层	3 层	5 层	7 层
50	83.422 7	95.599 9	97.369 7	97.370 3
100	86.560 7	97.788 6	98.197 7	98.517 9
200	89.871 5	97.559 1	98.183 9	97.731 8
300	90.919 2	96.713 5	95.962 8	95.973 3

精确度(Acc)是由均方根标准误差(RMS 误差)转换所得,即

$$Acc = 1 - \sqrt{\frac{\sum_{j=1}^m (y_j - c_j)^2}{m}}.$$

由表 1 可知,当设置隐含层为 1 时,精准度最高为 90.92%,而 3 层时,其精准度可以达到 97.79%,远远优于隐含层为 1 时.对于此模型,隐含层数或者隐含层单元数增多的话,其精确度的增加微乎其微,并且计算时间会显著增长,导致模型综合效率下降.故遵循训练时间、精确度相对最优化的条件,确定该神经网络模型隐含层为 3 层,隐含层单元数为 100 个.

1.3 激活函数的选择

传统神经网络中多应用 Sigmoid 函数^[8]与 Tanh 函数^[9],而现在神经网络用得最多的则是 Relu 函数或 Softplus 函数,如图 1 所示.

Sigmoid 函数表达式为

$$f(x) = \frac{1}{1 + e^{-x}}.$$

Tanh 函数表达式为

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

Relu 激活函数表达式为

$$f(x) = \max(0, x).$$

Softplus 激活函数表达式为

$$f(x) = \ln(1 + e^x).$$

Sigmoid 函数在历史上曾一度流行,可是由于其软饱和性,容易产生梯度消失,并且具有收敛缓慢的缺点,现一般不再采用它作为激活函数. Tanh 函数是 Sigmoid 函数的改进版,有异曲同工之处,但它也存在梯度饱和的问题.

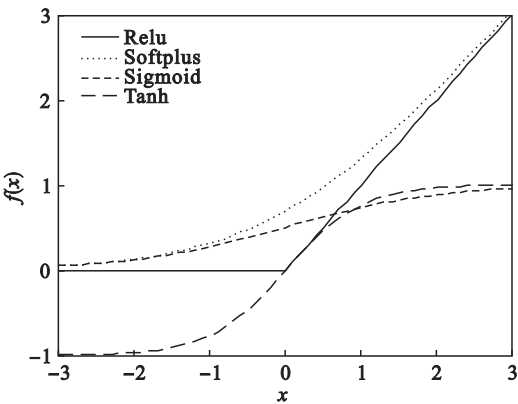


图 1 激活函数图
Fig. 1 Activation functions

近年来,Relu 激活函数变得越来越受欢迎. 如图 1 所示,Relu 函数与其平滑版本 Softplus 函数. 从信号方面来看,Relu 函数只对输入信号少部分选择性相应,这样可以提高学习的精度,更快更好地提取稀疏特征. Relu 激活函数模型相比于 Sigmoid 函数主要有以下 3 个优点:①单侧抑制;②相对宽阔的兴奋边界;③稀疏激活性. 由图 1 可以看出,Softplus 是 Relu 的平滑,其照顾到了 Relu 模型的前两个优点,却没有稀疏激活性.

将该 4 种激活函数应用于神经网络模型中, 对比其效果,如图 2 所示.

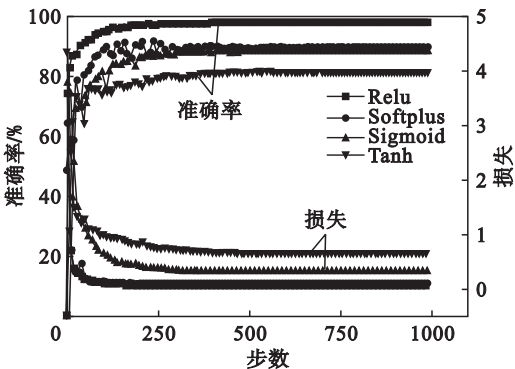


图 2 激活函数对比
Fig. 2 Comparison of activation functions

由图 2 可得,Relu 激活函数的模型损失达到 0.068,为最小值,其精确度也最高,为 98.33%. 从曲线可以看出激活函数为 Relu 时,收敛得更快,应用效果最佳,模型损失最小,故模型选择激活函数为 Relu 函数.

1.4 优化算法的选择

对于现阶段的深度神经网络来说,梯度下降法依然是其求最优解的基本方法. 但是单纯的梯度下降法已难以满足如今复杂的神经网络,于是在梯度下降法的基础上衍生出许多个不同的优化算法. 下面简单介绍一些常用的优化算法,并对其进行比较.

1) MBGD(批量梯度下降法)是将整个训练集分成若干个训练样本,对用 N 个训练样本,每次迭代更新采用 $t(1 < t < N)$ 个样本,是最常用的训练神经网络优化方法之一. 其训练时,参数更新的动荡相对较小,收敛过程更平稳,减小训练难度;但是,大部分神经网络所面临的都是复杂且非凸的目标函数,这也意味着优化很容易陷入局部最优的困境.

2) Momentum(动量梯度下降法)是计算梯度的指数加权平均数,并利用该值来更新参数值,具体过程为

$$\begin{aligned} v_{dw} &= \beta v_{dw} + (1 - \beta) dw, \\ v_{db} &= \beta v_{db} + (1 - \beta) db, \\ w &= w - \alpha v_{dw}, b = b - \alpha v_{db}. \end{aligned}$$

其中,动量衰减参数 β 一般取 0.9.

使用 Momentum 梯度下降时,通过累加减少了抵达最小值路径上的摆动,加快了收敛,当前后梯度方向一致时,Momentum 梯度下降能够加速学习;前后梯度方向不一致时,Momentum 梯度下降能够抑制震荡.

3) Adam 优化算法(自适应矩估计)适用于很多不同的深度学习网络结构,能计算模型各个参数的自适应学习率,使得整个训练过程更加稳定,具体过程为

$$\begin{aligned} v_{dw} &= \beta_1 v_{dw} + (1 - \beta_1) dw \\ v_{db} &= \beta_1 v_{db} + (1 - \beta_1) db, \\ S_{dw} &= \beta_2 S_{dw} + (1 - \beta_2) dw^2, \\ S_{db} &= \beta_2 S_{db} + (1 - \beta_2) db^2, \\ V_{dw}^{\text{corrected}} &= \frac{V_{dw}}{(1 - \beta_1^t)}, V_{db}^{\text{corrected}} = \frac{V_{db}}{(1 - \beta_1^t)}, \\ S_{dw}^{\text{corrected}} &= \frac{S_{dw}}{(1 - \beta_2^t)}, S_{db}^{\text{corrected}} = \frac{S_{db}}{(1 - \beta_2^t)}, \\ w &= w - \alpha \frac{V_{dw}^{\text{corrected}}}{\sqrt{S_{dw}^{\text{corrected}} + \epsilon}}, \end{aligned}$$

$$b;b-\alpha\frac{V_{db}^{corrected}}{\sqrt{S_{db}^{corrected}+\varepsilon}}.$$

其中:学习率 α 需要进行调参;超参数 β_1 被称为第一阶矩,一般取 0.9; β_2 被称为第二阶矩,一般取 0.999; ε 一般取 10^{-8} .

4) Adagrad 优化算法^[10] 主要思想:在进行参数更新时,不设置特定的学习率,在每次迭代时,对每个参数使用不同的学习率进行优化.其效果

是迭代至参数空间更平缓的方向时,会取得更大的进步.但是该方法会导致学习率大量减少,导致模型训练时间增长,效率偏低.

将以上 4 种优化算法进行训练对比,如图 3 所示,Adam 算法收敛速度最快,仅用 100 步迭代即可达到 0.027,且稳定损失最小,精确度最高,可达到 97.7%.经过对比后,综合各个方面的优势,模型选择 Adam 优化算法.

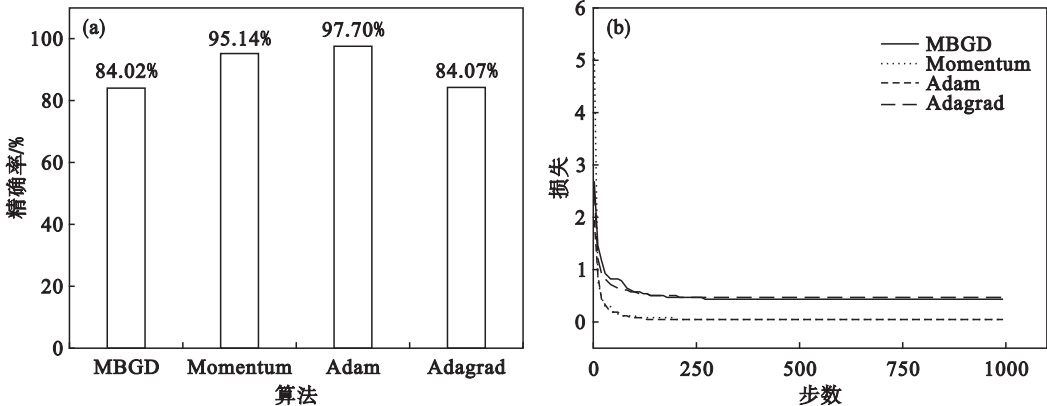


图 3 优化算法对比
Fig. 3 Comparison of optimization algorithms
(a)—精确率; (b)—损失.

1.5 正则化方法选择

由于深度神经网络中模型参数非常多,很容易产生过拟合现象.为了防止过拟合出现,研究者们提出了很多有效的技术.正则化是机器学习中非常重要并且非常有效的减少泛化误差的技术.本文比较常用的两个正则化技术:L2 正则化与 Dropout 正则化^[11].

L2 正则化主要思想:参数约束添加 L2 范数惩罚项.

$$\tilde{J}(w;X,y)=J(w;X,y)+\frac{1}{2}\alpha\|w\|^2.$$

通过梯度下降法推导出参数优化公式为

$$w=(1-\varepsilon\alpha)w-\varepsilon\nabla J(w).$$

其中, ε 为学习率.相对于正常的梯度优化公式,

对权重参数 w 乘上一个缩减因子.

Dropout 是一类通用并且计算简洁的正则化方法.简单来说,Dropout 是指在深度神经网络的训练过程中,在向前传播时,对于某层网络单元,按照一定的概率将其暂时从网络中丢弃,这样可以使一个神经元的出现不依赖其他的神经元^[12].

将 L2 正则化与 Dropout 正则化分别进行训练对比(图 4),由图 4 可以看出,无正则化时,存在过拟合问题,两损失差距过大.加入 L2 正则化后测试集损失为 0.07,训练集损失为 0.47,两损失差距缩小.并且相比于 L2 正则化,Dropout 正则化应用后训练集和测试集损失差距相比于应用 L2 正则化要小,说明 Dropout 正则化比 L2 正则

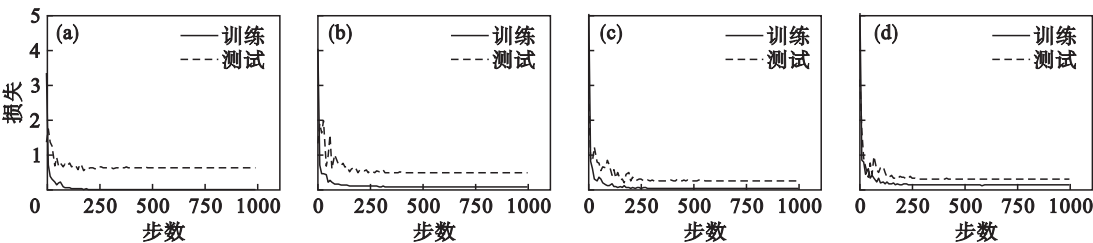


图 4 正则化对比
Fig. 4 Comparison of regularization
(a)—无正则化; (b)—L2 正则化; (c)—Dropout 正则化; (d)—L2 + Dropout 正则化.

化效果要好. 当 L2 正则化与 Dropout 正则化共同使用时,效果最好,训练集损失为 0.14,比单独使用 Dropout 正则化时要大一点,说明过拟合问题基本消失,测试集损失为 0.20,也更接近于训练集损失,且模型预测精确度较高. 故最终选择 L2 正则化与 Dropout 正则化共用.

2 仿真与应用

利用某钢厂 3 500 mm 中厚板线的冷却数据集,选取 5 000 组数据作为训练样本,进行模型训练,并且将训练好的模型应用于现场系统,进行传热系数的在线预报. 采用 2 000 组训练样本对深度神经网络进行训练,模型训练完成后,将该模型应用于现场实际生产中进行在线预报,将轧机二级传来的 PDI 数据作为输入层,用于预报传热系数.

连续采集 150 块预报的传热系数,与实测传热系数进行对比. 具体预报所得水冷传热系数如图 5 所示. 经统计,对传热系数的修正,准确率达到 96.7%. 由图 5 可看出,水冷传热系数因子的预报值和实测值拟合得很好,可看出深度神经网络模型可以很好地挖掘数据信息,并且能对一些复杂的参数进行准确的预报,说明深度神经网络

是可靠的.

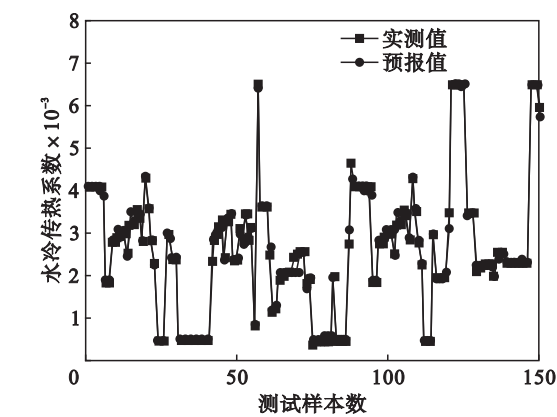


图 5 水冷传热系数因子预报值与实测值的比较
Fig. 5 Comparison between the predicted and measured values of water-cooling transfer coefficient factor

图 6 为冷后钢板实测温度. 由图 6 可以看出,在终轧温度波动严重的情况下,深度学习模型在预报终冷温度时,可过滤掉终轧温度的波动变化,并准确预报冷却时间与设定合适的冷却规程,以命中终冷温度,并且冷却后板材表面温度均匀. 通过两个对比可以总结出,应用深度学习后的冷却系统抗波动性强,鲁棒性更强.

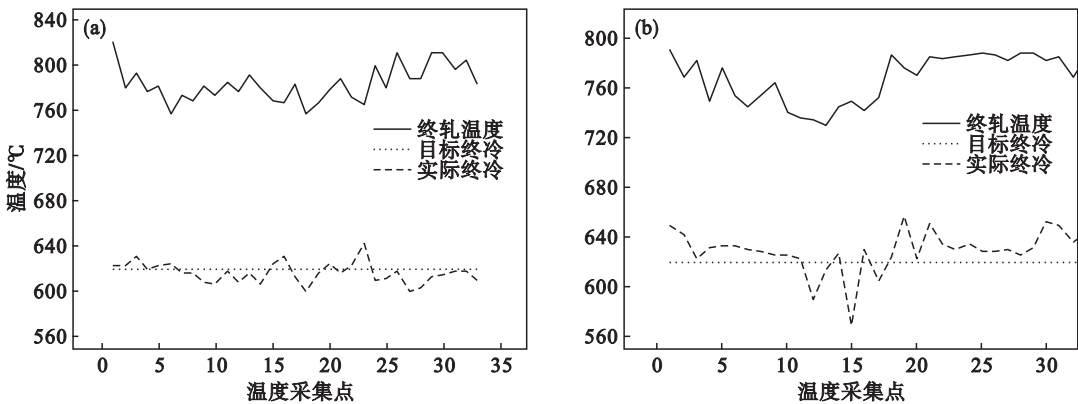


图 6 冷后钢板实测温度趋势
Fig. 6 Temperature tendency of plate after cooling
(a)—深度学习模型效果; (b)—原系统效果.

表 2、表 3 为命中率对比与部分在线应用结果. 由表 2 和表 3 可以得出,对于不同规格的板材,深度神经网络可以很好地预报出传热系数,而且对终冷温度的控制上满足工程的要求,且误差要小于原控冷系统,并更加稳定. 在控冷指标为 $\pm 20\text{ }^{\circ}\text{C}$ 条件下命中率可达 96%,相比于传统模型命中率提升了约 3 个百分点.

由上述对比可得出,经过深度神经网络模型学习的水冷传热系数相对传统的传热系数更加稳

定,模型鲁棒性更强,尤其当冷却板材规格改变或冷却环境受到外界干扰时,深度学习模型算法表现得更为稳定,对终冷温度控制精度更高.

表 2 命中率对比 Table 2 Comparison of hit rate			
数据对比	板材总 个数	命中板 材数/个	命中率 %
在线测试数据	300	288	96
原系统测试数据	300	279	93

表 3 部分在线应用结果
Table 3 Online results in application

板材厚度/mm	终轧温度/℃	水温/℃	气温/℃	w(C)/%	w(Cr)/%	w(Cu)/%	w(Mn)/%	传热系数 预报值	目标终冷 温度/℃	实际终冷 温度/℃	误差/℃
16.5	736	15.82	4.42	0.17	0.041	0.031	1.51	3791	630	622	8
10.1	787	15.87	4.47	0.12	0.018	0.015	0.84	2177	630	623	7
24.4	790	15.89	4.49	0.16	0.033	0.017	1.46	3500	650	641	9
20.3	760	15.76	3.59	0.16	0.033	0.017	1.46	4760	650	646	4
15.9	743	15.27	3.34	0.11	0.026	0.009	1.37	2177	610	613	3
46.7	767	14.92	3.73	0.17	0	0	1.41	802	580	570	10
16.5	748	15.23	4.85	0.16	0.026	0.014	1.47	3500	630	629	1
16.5	755	15.22	4.91	0.16	0.026	0.014	1.47	3653	630	625	5
32.5	772	15.19	5.02	0.17	0.034	0.018	1.44	6500	650	647	3
32.5	759	15.18	5.1	0.17	0.034	0.018	1.44	6423	650	652	2

3 结 论

1) 针对固定输入层、输出层的模型时,并不是层数越多、隐含层单元数越多越好. 恰到好处的层数与隐含层单元数既可以保证模型精确度,又可以减少模型训练时间.

2) 相比于传统激活函数,Relu 激活函数更接近于人类脑神经处理信息的方式,训练后的网络完全具备适度的稀疏性. 既可以提高模型的运算速度,又大大提升了模型预报准确度.

3) 相比于传统的梯度下降法,新开发出的各个优化算法对其学习率、梯度运行方向、梯度的更新程度等进行了优化. 其中 Adam 算法更加适用于该模型,预报精度最高,误差最小.

4) 该模型在经过训练后,在线应用时对于不同规格的板材,可以很好地预报出传热系数,而且对终冷温度的控制上满足工程的要求,且精度高于原控冷系统,并更加稳定. 在控冷指标为 ± 20℃ 条件下命中率可达 96%,比传统模型命中率提高约 3 个百分点. 并可推广至其他规格板材的轧后控冷模型中.

参考文献:

[1] 田勇,王丙兴,袁国,等. 基于超快冷技术的新一代中厚板轧后冷却工艺[J]. 中国冶金,2013,23(4):17-20.
(Tian Yong, Wang Bing-xing, Yuan Gong, et al. New generation TMCP for plate mill based on ultra-fast cooling technology [J]. *China Metallurgy*,2013,23(4):17-20.)

[2] 李栋. 基于神经网络的层流冷却控制模型研究[D]. 沈阳,东北大学,2010.

(Li Dong. Study of control model of laminar cooling based on neural networks[D]. Shenyang: Northeastern University, 2010.)

[3] Xie H B, Liu X H, Wang G D, et al. Optimization and model of laminar cooling control system for hot strip mills [J]. *Journal of Iron and Steel Research*,2006,13(1):18-22.

[4] Gong D Y, Xu J Z, Peng L G, et al. Self-learning and its application to laminar cooling model of hot rolled strip[J]. *Journal of Iron and Steel Research*,2007,14(4):11-14.

[5] Burba F, Ferraty F, Vieu P. *k*-nearest neighbour method in functional nonparametric regression [J]. *Journal of Nonparametric Statistics*, 2009, 21(4):453-469.

[6] 刘建伟,刘媛,罗雄麟. 深度学习研究进展[J]. 计算机应用研究,2014,31(7):1921-1930.
(Liu Jian-wei, Liu Yuan, Luo Xiong-lin. Research progress of deep learning[J]. *Application Research of Computers*,2014, 31(7):1921-1930.)

[7] 余滨,李绍滋,徐素霞,等. 深度学习:开启大数据时代的钥匙[J]. 工程研究,2014(3):233-243.
(Yu Bin, Li Shao-zi, Xu Su-xia, et al. Deep learning: the key of new world[J]. *Journal of Engineering Studies*,2014(3): 233-243.)

[8] Ebert T, Banfer O, Nelles O. Multilayer perception network with modified Sigmoid activation functions [C]// LNCS6319: Artificial Intelligence and Computational Intelligence. Berlin: Springer,2010:414-421.

[9] Kailik B, Olgac A V. Performance analysis of various activation functions in generalized MLP architectures of neural networks [J]. *International Journal of Artificial Intelligence and Expert Systems*,2010,1(4):111-122.

[10] Bartlett P L, Hazan E, Rakhlin A. Adaptive online gradient descent [C]//Advances in Neural Information Processing Systems 20. Vancouver,2007:1-15.

[11] Srivastava N, Hinton G, Krizhevsky a, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. *The Journal of Machine Learning Research*, 2014, 15(1): 1929-1958.

[12] Mendenhall M J. Improving quantitative structure-activity relationship models using artificial neural networks trained with Dropout [J]. *Journal of Computer-Aided Molecular Design*,2016,30(2):177-189.