2018 年 10 月 October 2018

DOI: 10.19476/j.ysxb.1004.0609.2018.10.14

基于深度网络训练的铝热轧轧制力预报

魏立新,魏新宇,孙浩,王恒

(燕山大学 工业计算机控制工程河北省重点实验室,秦皇岛 066004)

摘 要:在铝热轧过程中,轧制力预报精度直接影响着成品的产量和质量。为了提高铝热连轧轧制力预报精度, 提出一种基于深度学习方法的多层感知器(Multi-layer Perceptron, MLP)轧制力预报模型。模型利用 MLP 的函数 逼近能力来回归轧制力。模型以小批量训练为基础,利用 Batch Normalization 方法稳定网络前向传播的输出分布, 并使用 Adam 随机优化算法来完善梯度更新,以解决 MLP 模型难以训练的问题。仿真结果表明:模型使网络预 测与实测数据的相对误差降低到 3%以内,实现了轧制力的高精度预测。

关键词: 铝热轧; 轧制力预测; 深度学习; 多层神经网络; 优化算法

文章编号: 1004-0609(2018)-10-2070-07 中图分类号: TG339 文献标志码: A

在铝热轧轧制过程中,轧制力的预设定控制着轧 制规程的设置。并且影响着成品的厚度与其力学性能, 轧制力准确地预报可以减少带材头尾长度^[1-3]。对于传 统轧制力机理模型,由于复杂多变的轧制环境,以及 各个参量之间的相互耦合关系,在机理模型中引入了 许多假设,一旦测量产生误差或者假设不成立轧制力 的预报会产生很大波动^[4]。为了精确地预报轧制力, 智能模型被广泛应用到轧制领域。

MAHMOODKHANI 等^[5]将摩擦因数加入模型输 入,使用有限元分析与神经网络相结合对轧制力进行 了预报,实现了轧制力的实时预测。BYON 等^[6]建立 了 Shida 和 Misaka 两种流变应力方程并结合有限元分 析进行轧制力预测,得出了高温、中应变速率条件下 的轧制力模型。CHEN 等^[7]使用减法聚类将输入数据 划分为多个子空间,并对每个子空间建立支持向量机 SVM 模型,最后利用主成分分析法合成每个子模型输 出来预报轧制力,使得轧制力预报误差缩小到3.76%。 杨景明等^[8]改进遗传算法优化的 BP 神经网络对轧制 力进行了预测,结果表明该方法实现了轧制力的高精 度预测。赵文娇等^[9]建立基于信度分配的小脑模型神 经网络来预报轧制力,利用网络权值与样本激活地址 来更新权值信度,以完善梯度更新时网络的误差分配, 从而提高了了网络的收敛能力与预报精度。吕程等[10] 全面考虑神经网络输入项并按主要影响因素对训练样 本进行划分,使得网络预报的轧制力值更加准确。以

本文作者使用多层 BP 神经网络作为轧制力预报 的深度模型。并且为了增加网络的学习效率提升网络 性能,使用 Mini-batch 梯度下降作为模型的基础学习 方式来加快网络训练速度并且为网络的训练提供随机 性;使用 Batch Normalization 优化网络前向传播,稳 定每层网络的输入分布;使用 Adam 随机优化算法优 化网络反向传播,在梯度更新中为不同参数提供自适 应性学习率,完善梯度更新。

1 深度网络模型

1.1 多层 BP 神经网络

常规 BP 神经网络由输入层、一个或多个隐藏层 以及输出层组成,其结构如图 1 所示。由于神经网络 对复杂函数的表达能力随着网络层数的增加而增 强^[12],所以对于 *K* 层 BP 神经网络(*K*≥3),当其训练

收稿日期: 2018-03-27; 修订日期: 2018-07-23

上方法都优于传统的机理模型,提高了轧制力预报的 精度,但无论基于 BP、RBF 或 SVM 建立的轧制力模 型都属于浅层模型,其对复杂函数的表达能力与泛化 能力都受到一定制约。而随着深度学习网络的发展, 深度模型逐渐体现出其优点,深度模型具有更加优秀 的函数表达能力与更强的特征学习能力^[11]。因此,为 了进一步提高轧制力的预报精度,使用深度模型预报 轧制力并改进深度模型难以训练的问题。

基金项目:河北省自然科学基金资助项目(F2016203249)

通信作者:魏立新,教授,博士;电话: 0335-8387556; E-mail: wlx2000@ysu.edu.cn

得当时,其效果可超越常规 BP 网络与 SVM 等 *K* 为 2 的智能模型。但是随着网络层数的增加同时,也导致 了网络训练难度上升的问题,其中主要包括:随着网 络层数增加导致的训练时间过长;网络在训练过程中 发生的协变量转变;非凸优化导致网络收敛到局部最 小或鞍点以及梯度弥散导致的网络退化。本研究将通 过深度学习算法来解决上述问题。



图1 多层神经网络结构

Fig. 1 Multilayer neural network structure

1.2 Mini-batch 梯度下降

对于传统梯度下降算法每次迭代都需要在整个样本集上进行计算,当样本量较大时,会严重消耗计算资源,造成收敛速度过慢。同时由于在每次计算中都考虑所有的训练数据,丧失了训练的随机性,容易造成模型过拟合。随机梯度下降对传统梯度下降有了极大的提升,但是由于过度自由的训练(单个样本)导致损失函数在训练过程中产生较大波动。且训练中每次迭代只考虑单个样本,造成模型的整体性缺失以至于无法收敛。Mini-batch梯度下降可看做传统梯度下降与随机梯度下降的中和,该方法既保留了随机梯度下降

此梯度下降方法主要思想是将总训练集随机分成 batch_size 大小多个子训练集 ${X^i, Y^i}$, $i \in (1, 2, \dots \lceil m/batch_size \rceil)$,每次的迭代都在子训练 集上进行,但模型整体评价函数 epoch_cost 由所有子 训练集求出的损失函数平均值表示。具体流程如图 2 所示。

采用 Mini-batch 梯度下降作为网络的基础学习方式。MLP 模型训练时间随着网络层数与样本数量的增加而增加。为满足模型对多层网络与大样本的需求, 在模型训练中使用 Mini-batch 来减少训练时间,加快 模型收敛速度。同时也为 Batch Normalization 与 Adam 随机优化算法的使用提供基础。



图 2 单轮损失函数计算流程图

Fig. 2 Single-round loss function calculation flow chart

1.3 Batch Normalization

在训练单隐层网络时,对训练数据与测试数据进 行预处理,使两者属于相同分布,可以加快网络的训 练速度与提高模型精度。而对于多隐层网络,网络中 的每一层都可以看作一个单隐层的子网络。每一层的 输入都受到此层前所有层的影响。参数的微小变化会 随着网络的深度而增大,从而影响子层网络输入,造 成在训练过程中该层输入分布不断改变,导致 covariate shift的出现。后层网络将不得不去适应新分 布的产生,从而导致网络精度的下降与训练速度的降 低。

Batch Normalization 就是为了解决多隐层网络在 训练中产生的内部结点的分布变化^[13]。与预处理类似 Batch Normalization 在多层网络中对每层的输出进行 规范化处理,从而稳定所有子层网络的输入分布,并 且为了保持网络的表达能力,不固定每层的分布,而 是增加学习参数 $\gamma 与 \beta$ 分别作为规范化后数据的标准 差与均值。训练模型时单个 Mini-batch 操作如下(不考 虑激活):

对于 k 层网络 k=1, 2, …, K, 约定每层网络的神 经 元 个 数 为 n^k , 每 层 权 重 $W^k \in R^{n^k \times n^{(k-1)}}$, 输入 $X = Z^0 \in R^{n^0 \times batch_{size}}$, $U_{test}^k = 0$, $V_{test}^k = 0$.

1) 对于任意 k 层子网络,线性表达式为:

$$\boldsymbol{Z}^{k} = \boldsymbol{W}^{k} \boldsymbol{Z}^{k-1} \tag{1}$$

式中: Z^k 为k 第层网络的输出矩阵, $Z^k \in \mathbb{R}^{n^k \times batch_size}$ 。 2) 使用式(2)和(3)求出 k 层子网络每个神经元输

2) 使用式(2)和(3)水出 k 层于网络母个神经元输 出的均值与方差

$$\mu_j^k = \frac{1}{batch_size} \sum_{i=1}^{batch_size} z_{j,i}^k$$
(2)

$$v_j^K \leftarrow \frac{1}{batch_size} \sum_{i=1}^{batch_size} (z_{j,i}^k - \mu_j^k)^2$$
(3)

式中: $\mu_{j}^{k} = v_{j}^{K}$ 分别为 k 层子网络第 j 个神经元的输 出均值与输出方差。对 k 层子网络的每个神经元求解 可得均值向量: $U^{k} = (\mu_{1}^{K}, \mu_{2}^{K}, \dots, \mu_{n^{k}}^{K})'$, 与方差向量 $V^{k} = (v_{1}^{K}, v_{2}^{K}, \dots, v_{n^{k}}^{K})'$ 。

3) 利用指数移动平均法估计 k 层子网络总体均值 与方差,表达式如下:

$$\boldsymbol{U}_{\text{test}}^{k} = \beta_0 \boldsymbol{U}_{\text{test}}^{k} + (1 - \beta_0) \boldsymbol{U}^{k}$$
(4)

$$\boldsymbol{V}_{\text{test}}^{k} = \beta_0 \boldsymbol{V}_{\text{test}}^{k} + (1 - \beta_0) \boldsymbol{V}^{k}$$
(5)

式中: U_{test}^k 为第 k 层子网络的估计均值向量; V_{test}^k 为 第 k 层子网络的估计方差向量。

4) 规范化 k 层子网络,具体如下:

$$\hat{\boldsymbol{Z}}^{K} = \frac{\boldsymbol{Z}^{k} - \boldsymbol{U}_{\text{Broadcast}}^{k}}{\sqrt{\boldsymbol{V}_{\text{Broadcast}}^{k} + \varepsilon}}$$
(6)

式中: Z^{K} 为 k 层子网络的规范化输出矩阵; ε 为极 小数,该式为点运算。

5) 使用学习参数γ与β建立k层子网络新分布, 表达式为

$$\boldsymbol{Y}^{k} = \boldsymbol{Z}^{K} \,\boldsymbol{\gamma}_{\text{Broadcast}} + \boldsymbol{\beta}_{\text{Broadcast}} \tag{7}$$

式中: Y^k 为子网络新分布下的输出矩阵, 其中学习参数 $\gamma \in R^{n^k \times 1}$, $\beta \in R^{n^k \times 1}$, 式(7)为点运算。

参数γ与β的更新方式与网络权重 W相同,并且 因为训练中规范化了线性输出,可以省略偏差参数b。 模型在测试阶段,忽略步骤3),使用U^k_{test}与V^k_{test}完成 k 层子网络的规范化,其他参数不变。

Batch Normalization 在模型中的应用不仅可以极 大地加快网络的训练速度,同时减少前层网络参数改 变对后层网络的影响,使模型中每层网络的学习相对 独立,使网络的训练更有效率。并且与 Mini-batch 相 结合,利用子样本的均值与方差取代总体方差,为模 型引入噪音,使得模型具有更好的泛化能力。

1.4 Adam 随机优化算法

 \wedge

对于多隐层神经网络,随着网络层数的增加局部 极小点与鞍点的数目都成倍的增长。使得网络的训练 更容易陷入局部最小点甚至鞍点。传统的梯度下降算 法由于对初始点极为敏感与使用固定的学习率,不适 于在多隐层神经网络中使用。为获得更好的梯度更新 效果,可使用 BP 算法的改进方法,主要包括:动量 方法 Momentum、采用自适应学习率的 AdaGrad 方 法^[14]等。采用结合动量方法与自适应学习率的 Adam 随机优化算法来更新网络梯度。

Adam 算法^[15]是由 KINGMA 和多伦多大学的 BA 所提出的,该方法基于适应性矩估计,能够完成随机 目标函数的梯度优化。算法利用指数移动平均方法估 计网络梯度的一阶矩(均值)与二阶矩(方差),用来为不 同的参数提供独立的自适应学习率。并且使梯度的更 新具有动量效果。更适用于多隐层神经网络这种非凸 优化。

Adam 优化算法的参数主要包括梯度学习率 α , 一阶矩估计的指数衰减率 β_1 ,二阶矩估计的指数衰减 率 β_2 ,以及为防止出现除零错误的极小数 ε 。其主要 步骤如下。

Step1: 初始化参数向量 θ (θ 包括每层神经网络 权重 W与 BN 参数),梯度一阶矩向量 m_0 、梯度二阶 矩向量 u_0 ,并 *t*=0。

Step2: t = t+1, 计算损失函数f 对 $\boldsymbol{\theta}$ 的导数 \boldsymbol{g}_t 如 式(8)

$$\boldsymbol{g}_t \leftarrow \nabla_{\theta} f_t(\boldsymbol{\theta}_{t-1}) \tag{8}$$

Step3:使用指数移动平均估计梯度期望与平方梯 度期望,表达式如下:

$$\boldsymbol{m}_t \leftarrow \boldsymbol{\beta}_1 \cdot \boldsymbol{m}_{t-1} + (1 - \boldsymbol{\beta}_1) \bullet \boldsymbol{g}_t \tag{9}$$

$$\boldsymbol{u}_t \leftarrow \boldsymbol{\beta}_2 \cdot \boldsymbol{u}_{t-1} + (1 - \boldsymbol{\beta}_2) \bullet \boldsymbol{g}_t^2$$
(10)

式中: m_t , u_t 分别为梯度的一阶矩估计和二阶矩估 计,并且 $g_t^2 = g_t \times g_t$ 。

Step4: 引入偏差修正,使用(11)与(12)求出两个 期望的无偏估计。

$$\hat{\boldsymbol{m}}_t \leftarrow \boldsymbol{m}_t / (1 - \boldsymbol{\beta}_1^t) \tag{11}$$

$$\hat{\boldsymbol{u}}_t \leftarrow \boldsymbol{u}_t / (1 - \boldsymbol{\beta}_2^t) \tag{12}$$

式中: $\hat{m}_t = \hat{u}_t$ 分别为梯度期望与方差的无偏估计。 Step5: 按照式(13)完成梯度更新,返回 Step2。

$$\boldsymbol{\theta}_{t} \leftarrow \boldsymbol{\theta}_{t-1} - \alpha \cdot \hat{\boldsymbol{m}}_{t} / (\sqrt{\hat{\boldsymbol{u}}_{t}} + \varepsilon)$$
(13)

式中: α 为学习率; ε 为极小数。

同时使用 Adam 优化算法与 Mini-batch 梯度下降 来对模型进行训练。 Adam 算法目标是降低随机目标 函数的期望值,而目标函数的随机性主要来源于随机 的子样本, Mini-batch 在每个历元中都随机打乱样本 顺序为模型训练提供随机的子样本,满足 Adam 对随 机性的要求。在 Adam 算法中利用梯度估计均值与估 计方差为不同参数提供更好的自适应学习率,使梯度 得到规范化,完善梯度更新。并且 Mini-batch 单次迭 代使用大小为 Mini-batch 的子样本,迭代速度较快。 从而两者的结合提高了模型收敛精度与收敛性能。

2 轧制力预测

2.1 轧制力预测模型

该预测方法使用多隐层神经网络作为深度模型对 轧制力进行预报。并以 Mini-batch 梯度下降为基础, 分别使用 Batch Normalization 于 Adam 随机优化算法 优化网络的正向传播与反向传播。

对于神经网络预测模型,为了获得更好的函数逼 近能力,网络包含 5 个隐藏层。输入层单元数为 8, 输出单元数为 1,网络结构为[8,128,128,64,64, 32,1]。并且为了进一步克服非凸优化问题,使用 Xavier^[16]方法对参数进行初始化。网络前向传播中: Batch Normalization 在每层激活函数前使用, β_0 = 0.998, *batch_size* 取 64,隐藏层选用 Relu 为激活函数, 输出层激活函数选用 tanh。网络反向传播中:学习率 取值 0.001, Adam 算法中的参数按照原论文选取,一 阶 矩 估 计 衰 减 率 β_1 =0.9,二阶 矩 估 计 衰 减 率 β_2 =0.999,极小数 ε =1×10⁻⁸。

为防止模型训练可能会出现的梯度弥散,在隐藏 层中未使用 sigmoid 等饱和非线性函数,而是使用激 活函数 Relu,函数图像如图 3 所示。

Relu激活函数克服了许多传统激活函数的缺陷。 传统激活函数 sigmoid 函数结构复杂,使网络在梯度 计算时计算量变大,且由于其具有饱和非线性特性, 在饱和区处梯度大小接近于 0,造成了网络的梯度弥 散。而 Relu 函数的结构简单,不仅简化了网络的反向 传播,也抑制了梯度弥散的产生。更为重要的是由于



图 3 Relu 激活函数 Fig. 3 Relu activation function

Relu 激活函数的单侧抑制,使得网络自己引入了稀疏特性,从而缩小纯监督学习神经网络与包含无监督预训练神经网络的性能差距^[17]。使得网络在不经过无监督学习的情况下,仍能的到理想的预测结果。

网络结构简图如图4所示。

2.2 样本数据的选择

对于传统 SIMS 热连轧轧制力数学模型其计算公 式为

$$P = Bl'_{\rm c}Q_{\rm p}KK_{\rm T} \tag{14}$$

式中: P 为轧制力; B 为带钢宽度; l'_c 为考虑轧辊压 扁后的接触弧水平投影长度; Q_p 为摩擦力应力状态系 数; K 为金属的变形抗力; Q_p和 K 受到变形速度,变 形程度等影响; K_T 为前后张力对轧制力的影响系数。

综合考虑轧制力传统模型,由于摩擦因数无法用 理论公式精确表达,而且随着生产环境的变化而变化, 因此利用网络的自适应性,将其包含于网络内部,不 作为输入节点^[18]。且变形抗力主要与平均变形速度与 变形程度及温度有关。使用轧机进出入口厚度与轧辊 半径来表示累计的变形程度。最终选定轧制温度 *T*, 带钢宽度 *B*,轧机入口厚度 *h*₁,轧机出口厚度 *h*₂,前 张力*τ*_f,后张力*τ*_b,轧辊半径 *R*,以及轧制速度 *v* 最 为神经网络的输入。



图4 网络结构简图

Fig. 4 Network structure diagram

2.3 仿真与分析

利用建立起来的深度神经网络模型进行轧制力预 测仿真。预测模型的输入向量为(*T*, *B*, *h*₁, *h*₂, *R*, *τ*_f, *τ*_b, *ν*), 损失函数为

$$\min(\text{MSE}) = \min(\frac{1}{m} \sum_{i=1}^{m} (y_i - y'_i))$$
(15)

利用河南某厂铝热连轧生产线的在线数据,选取 其中的第2机架中的数据9000条对模型进行训练。在 训练前,对所有样本数据进行规范化处理。利用训练 好的网络模型,对不在训练集中的600多条轧制力数 据进行预测,其预测结果如图5所示。



图 5 预测值与实测值对比图

Fig. 5 Comparison between predicted and actual values

从图 5 中可以看出,该模型具有较高的预测精度, 其预测的相对误差可以控制在 3%以内,可以满足生 产要求。

表 1 所列为该模型与其他模型在限定时间 500 s 下,建模时间,平均相对误差,平均绝对误差与决定 系数之间的比较,其中遗传算法优化的 BP 网络学习

表1 模型预测结果对比

Table 1	Comparison	of model	prediction result	S
---------	------------	----------	-------------------	---

Model	Training time/s	Mean relative error	Mean absolute deviation/kN	Coefficient of determination
GA-BP	500	0.027	27.89	0.8836
MLP	500	Not converged	Not converged	Not converged
SVM	354.16	0.021	21.13	0.9145
Deep network	186.56	0.012	14.05	0.9604

率取 0.9, 隐藏层神经元数为 128。多层神经网络 MLP 未使用深度学习相关方法,但结构不变。为了消除随 机性对模型影响,使用 10 次预测结果的平均值来作 对比。

由表1可以看出,多层神经网络在无相关优化算法的情况下无法完成收敛。深度网络模型的建模时间为186.56s,其预测速度与预测精度上皆高于其他智能模型。

3 结论

1) 使用多隐层神经网络建立轧制力预测模型。并为了解决多隐层网络存在的固有缺陷,使用深度学习相关方法优化网络,其中主要使用了 Batch Normalization 加快网络的训练速度,解决因网络层数 增多引起的训练时间过长的问题,Adam 随机优化算 法的应用避免了网络在训练过程中陷入局部最优点以 及鞍点,提高了网络向最优解收敛的能力。

2) 该方法下轧制力预测的平均相对误差为
 1.2%。相对于优化算法优化的 BP 网络,其预报平均相对误差降低了 1.5%。并且使得网络的训练时间降低到 186.56 s,相对于 SVM 模型其收敛速度提升了接近 1/2。综上所述,该方法在实践中具有很大的应用潜能。

REFERENCES

- LIU X, LIU X H, SONG M, SUN X K, LIU L Z. Theoretical analysis of minimum metal foil thickness achievable by asymmetric rolling with fixed identical roll diameters[J]. Transactions of Nonferrous Metals Society of China, 2016, 26(2): 501–507.
- [2] ZUO Y B, FU X, CUI J Z, TANG X Y, MAO L, LI L, ZHU Qing-feng. Shear deformation and plate shape control of hot-rolled aluminium alloy thick plate prepared by asymmetric rolling process[J]. Transactions of Nonferrous Metals Society of China, 2014, 24(7): 2220–2225.

[3] 燕 猛,黄华贵,张彩云,杜凤山,张尚斌.板坯头/尾部平面 形状对铝合金厚板粗轧头尾切除量的影响[J].中国有色金属 学报,2017,27(6):1102-1108.

YAN Meng, HUANG Hua-gui, ZHANG Cai-yun,DU Feng-shan, ZHANG Shang-bin. Effect of the shape of the head/tail of the slab on the head and tail cutting of aluminum alloy thick plate[J]. Chinese Journal of Nonferrous Metals, 2017, 27(6): 1102–1108.

[4] OROWAN E. The calculation of roll pressure in hot and cold flat rolling[J]. Proceedings of the Institution of Mechanical 第28卷第10期

Engineers, 1943, 150(1): 140-167.

- [5] MAHMOODKHANI Y, WELLS M A, SONG G Prediction of roll force in skin pass rolling using numerical and artificial neural network methods[J]. Ironmaking & Steelmaking, 2016, 44(4): 281–286.
- [6] BYON S M, NA D H, LEE Y S. Flow stress equation in range of intermediate strain rates and high temperatures to predict roll force in four-pass continuous rod rolling[J]. Transactions of Nonferrous Metals Society of China, 2013, 23(3): 742–748.
- [7] CHEN Z M, LUO Z L. Rolling force prediction based on multiple support vector machines[C]// WATADA J. Proceedings of the 32nd Chinese Control Conference. New York: IEEE Press, 2013: 3306–3309.
- [8] 杨景明,顾佳琪,闫晓莹,车海军.基于改进遗传算法优化
 BP 网络的轧制力预测研究[J].矿冶工程,2015,35(1):111-115.

YANG Jing-ming, GU Jia-qi, YAN Xiao-ying, CHE Hai-jun. Prediction of rolling force based on improved genetic algorithm for optimizing BP neural network[J]. Mining and Metallurgical Engineering, 2015, 35(1): 111–115.

[9] 赵文姣, 闫洪伟, 杨 枕, 温玉莲, 孙祖乾. 基于 CA-CAMC
 网络的轧制力自学习预报模型[J]. 冶金自动化, 2016, 40(2):
 7-10.

ZHAO Wen-biao, YAN Hon-wei, YANG Zhen, WEN Yu-lian, SUN Zu-qian. Rolling force self-learning prediction model based on CA-CAMC network [J]. Metallurgical Industry Automation, 2016, 40(2): 7–10.

[10] 吕 程,王国栋,刘相华,姜正义,朱洪涛,袁建光,解 旗. 基于神经网络的热连轧精轧机组轧制力高精度预报[J].钢铁, 1998,33(3):33-35.

LÜ Cheng, WANG Guo-dong, LIU Xiang-hua, JIANG Zheng-yi, ZHU Hong-tao, YUAN Jian-guang, XIE Qi. High-precision prediction of rolling force in hot rolling finishing mill based on neural network[J]. Iron and Steel, 1998, 33(3): 33-35.

- [11] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527–1554
- [12] 毛勇华,桂小林,李 前,贺兴时. 深度学习应用技术研究[J]. 计算机应用研究, 2016, 33(11): 3201-3205.
 MAO Yong-hua, GUI Xiao-lin, LI Qian, HE Xing-shi. Research on deep learning application technology[J]. Journal of Computer Applications, 2016, 33(11): 3201-3205.
- [13] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]// BACH
 F. Proceedings of the 32nd International Conference on Machine Learning. Massachusetts: MIT Press, 2015: 448–456.
- [14] DUCHI J, HAZAN E, SINGER Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. Journal of Machine Learning Research, 2011, 12(7): 257–269.
- [15] KINGMA D P, BA J L. Adam: A method for stochastic optimization[C]// KINGSBURY B. Proceedings of the the 3rd International conference on learning representations. San Diego: arXiv, 2015: 1412–6980.
- [16] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[J]. Journal of Machine Learning Research, 2010, 9: 249–256.
- [17] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[C]// BACH F. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. Massachusetts: MIT Press, 2011: 315–323.
- [18] 杨景明,孙晓娜,车海军,刘 畅. 基于蚁群算法的神经网络 冷连轧机轧制力预报[J]. 钢铁, 2009, 44(3): 52-55.
 YANG Jing-ming, SUN Xia-ona, CHE Hai-jun, LIU Chang. Prediction of rolling force in neural network cold rolling mill based on ant colony algorithm[J]. Iron and Steel, 2009, 44(3): 52-55.

Prediction of aluminum hot rolling force based on deep network

WEI Li-xin, WEI Xin-yu, SUN Hao, WANG Heng

(Key Lab of Industrial Computer Control Engineering Department of Yanshan University, Qinhuangdao 066004, China)

Abstract: In the aluminum hot rolling, the prediction accuracy of the rolling force directly affects the output and quality of the finished product. In view of the inherent defects of traditional rolling force model, a MLP rolling force prediction model based on deep learning method was proposed. The model uses MLP's function approximation ability to regress the rolling force. Based on the Mini-batch training, the model uses Batch Normalization method to stabilize the output distribution of the network forward propagation, and uses the Adam stochastic optimization algorithm to improve the gradient updating so as to solve the difficult training problem of the MLP model. The simulation results show that the model can reduce the relative error between the network prediction and the measured data to less than 3%. Compared with the traditional mathematical model, this method realizes the high precision prediction of the rolling force, and realizes a high-precision prediction of rolling force.

Key words: aluminum hot rolling; rolling force prediction; deep learning; multilayer neural network; optimization algori

Foundation item: Project(F2016203249) supported by the Natural Science Foundation of Hebei Province, China Received date: 2018-03-27; Accepted date: 2018-07-23

Corresponding author: WEI Li-xin; Tel: +86-335-8387556; E-mail: wlx2000@ysu.edu.cn

(编辑 王 超)