

深層学習を用いた 12 段クラスター型圧延機の形状制御技術の開発

Development of Flatness Control in a 12Hi Cold Rolling Mill Using Deep Learning

北村 拓也	KITAMURA Takuya	JFE スチール スチール研究所 圧延・加工プロセス研究部 主任研究員
齋藤 晋	SAITO Susumu	JFE スチール 東日本製鉄所（京浜地区）冷延部 冷延技術室 主任部員
植野 雅康	UENO Masayasu	JFE スチール スチール研究所 サステナブルマテリアル研究部長【現 研究企画部長】
田内 佑樹	TAUCHI Yuki	株式会社日立製作所 デジタルシステム&サービス 社会ビジネスユニット
		インフラ制御システム事業部 社会・インダストリ制御システム本部 電気制御システム設計部

要旨

冷間圧延における平坦度制御は、圧延機出側に設置された平坦度計の実測結果と影響係数とを用いるフィードバック制御が一般的であり、特に 4 段や 6 段の冷間圧延機を中心に適用されている。平坦度制御において、影響係数は実機で実測する必要があるパラメータである。影響係数は、理想的にはサーマルクラウンやロール摩耗など平坦度に影響する数多くの因子を取り入れるべきであると考えられるが、因子が多いほどモデルの構築、管理に多大な労力を要する。本稿では、影響係数を必要としない新しい平坦度制御として深層学習を用いた平坦度制御を開発し、その有用性を検証するため 12 段クラスター型圧延機に導入した事例を報告する。

Abstract:

Flatness control in a cold rolling mill is generally performed by feedback control based on measurements from flatness meter installed on the exit side of the mill. This approach is usually applied to 4Hi and 6Hi rolling mills. In flatness control, influence coefficients are parameters that need to be measured on the actual mill. Ideally, numerous factors affecting flatness, such as thermal crown and wear of rolls should be considered in the influence coefficients. Constructing and managing a model that accounts for these factors is however unrealistic. In this paper, a new flatness control using deep learning that does not require measured influence coefficients was developed and installed to a 12Hi cluster cold rolling mill.

1. はじめに

冷間圧延における平坦度制御は、製品の品質を決定する重要な因子であると同時に、圧延ラインの生産性の観点でも重要な操業因子の一つである。4 段、6 段の冷間圧延ラインにおける平坦度制御には、圧延機出側に設置された平坦度計^{1,2)}で得られた実測平坦度を 4 次や 6 次関数に分解した後、それぞれの目標値との差分が最小となる平坦度制御アクチュエータの操作量を所定の影響係数をもとに最適化計算して算出する、フィードバック制御^{3,4)}が用いられる場合が多い。ここで影響係数とは平坦度制御アクチュエータを単独で操作した際の平坦度の変化であり、実機で実測するのが一般的である。

一方で 12 段クラスター型圧延機や 20 段ゼンジマー型圧延機における平坦度制御では、平坦度計が設置されている場合でも、上記平坦度の実測結果と影響係数とに基づく

フィードバック制御が用いられず、オペレータの手介入操作によって操業される場合がある⁵⁾。この理由としては、12 段クラスター型圧延機や 20 段ゼンジマー型圧延機では、一般的な 4 段、6 段型の圧延機と比較してワークロールが小径であることに起因してロールの弾性変形が複雑となり、実測平坦度が高次の項を含むため^{5,7)}、影響係数の妥当性が懸念されるためと考えられる。

この問題に対するアプローチ方法は大きく二つに大別できる。一つ目は精緻なオフラインモデルを構築した後、モデルの有するパラメータを合わせこむ方法³⁾、二つ目は実機で影響係数を実測する際にできるだけ多くの変数を導入する方法⁷⁾である。しかし、これらの方法には、オフライン計算でもロール発熱や摩耗の影響、他アクチュエータの現在位置などの多くの変数の影響に関しては考慮できない欠点がある。

他方、近年のインターネットインフラや GPGPU に代表されるコンピュータリソースの向上に伴い、生物の脳の基本構成要素であるニューロンを模したニューラルネットワークや

ディープラーニングが着目されている。ディープラーニングは、ニューラルネットワークの中間層を多層にした構造を有し、説明変数を用いて目的変数を最適に説明するための数理モデルを、ニューラルネットワークが有する重み係数をデータから学習することで得る方法である。なお、本稿ではニューラルネットワークを、ディープラーニングを説明するための上位概念の用語として使用する。ニューラルネットワークを平坦度制御に活用する報告は多く存在する^{8,9)}が、平坦度アクチュエータの操作量自体をニューラルネットワークの予測対象とした実機適用例は見られない。そこで本稿では 12 段クラスター型圧延機でディープラーニングを活用した平坦度制御を開発し、実機でその有用性を検証した。さらにニューラルネットワークが獲得する影響係数を実機で実測した影響係数と比較することで、その妥当性を検証した。

2. 深層学習を用いた平坦度制御モデル

2.1 12 段クラスター型圧延機の概要

本稿に示すシステムを導入した 12 段クラスター型圧延機の概要を図 1 に、設備仕様を表 1 に示す。本圧延機は入出側に平坦度計を有し、WR ベンダー、IMR ベンダー、レベルリング、上 BUR クラウン 3 種の計 6 つの平坦度制御アクチュエータを有する圧延機である。

2.2 深層学習の概要と学習条件

深層学習は、生物の脳の基本構造であるニューロンを模したニューラルネットワークの中間層を複数重ねた数理モデルであり、種々の意思決定問題に応用されている。深層学習における学習計算は、式 (1) に示すとおり、目的関数が最小となるように誤差逆伝播法を用いてネットワークの重み係数 w とバイアス係数 b を同定する最適化問題である。なお本稿では深層学習とディープラーニングを同義語として用いている。

$$w^*, b^* = \text{argmin}(L) \quad (1)$$

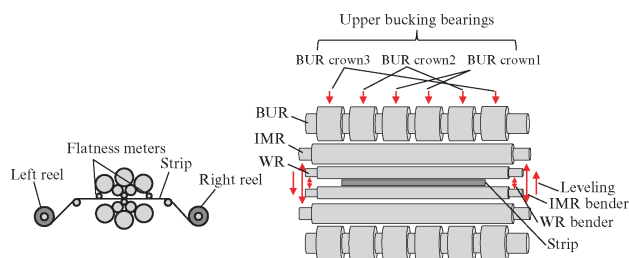


図 1 12 段クラスター型圧延機の概要図と平坦度制御アクチュエータ

Fig. 1 12Hi cold rolling mill and its flatness actuators

表 1 12 段クラスター型圧延機の設備仕様

Table 1 Specifications of the mill

Mill type	12Hi reverse
Max. rolling speed	600 mpm
Max. rolling force	6 000 kN
WR diameter	70-120 mm
Steel grade	High carbon steel
Thickness	0.3-2.3 mm
Width	400-1 070 mm

重み係数 w 、バイアス係数 b を最適化する理論は同等であるため本稿では重み係数で呼称する。添え字*は最適値であることを示す。 L は実測値と予測値の差から定義される損失関数であり、本稿では式 (2) で定義される平均平方二乗誤差を用いた。

$$L = \frac{1}{n} \sum_{i=1}^n (y_i^m - y_i^p)^2 \quad (2)$$

ここで y は目的変数である平坦度制御アクチュエータ操作量、添え字 m は実測値、 p は予測値、 i はデータバッチ、 n は学習計算におけるイタレーション数である。

各中間層において定義される活性化関数は式 (3) に示す ReLU 関数を用いた。

$$h(x) = \max(0, x) \quad (3)$$

本稿では表 2 に示すとおり、圧延速度、材料の寸法、および熱延板クラウンなどの基本的な情報に加え、当該ワークロールの圧延長や圧延ラインの加速度、圧延方向などを説明変数として用いた。圧延材の鋼種は、重複しない通し番号をすべての鋼種に付与して使用した。平坦度は関数近似すると特徴量を失う可能性があると考え、すべての幅方向チャンネルの出力を用いた。目的変数は表 3 に示すとおり、WR ベンダーやレベルリングなど計 6 種の平坦度制御アクチュ

表 2 ディープラーニングネットワークの説明変数

Table 2 Input information of the deep learning network

	Number of data	Unit
Measured flatness	44	I-unit
Target flatness	44	I-unit
Start/finish element	2	-
Pass No.	1	-
Position of flatness actuators	6	kN or mm
Width	1	mm
Entry/delivery thickness	2	mm
Entry/delivery speed	2	mpm
Entry/delivery acceleration	2	mpm/sec
Rolling force	1	kN
Lubricant flow	1	L/min
Entry/delivery tension	2	MPa
WR diameter	1	mm
Entry side strip crown	1	μm
Rolling direction	1	-
Rolling distance	1	m
Steel grade	1	-

表 3 ディープラーニングネットワークの目的変数と規格化に用いた定数

Table 3 Output information and constants for normalization

WR bender	9.2 (kN/sec)
IMR bender	9.2 (kN/sec)
Leveling	15 (μm/sec)
BUR crown1	10 (μm/sec)
BUR crown2	10 (μm/sec)
BUR crown3	10 (μm/sec)

エータの操作量とし、平坦度制御アクチュエータの操作量を時々刻々出力する教師あり学習として設計した。深層学習に用いる説明変数および目的変数は、適切に規格化することで学習効率および予測精度が向上することが知られている。本稿では説明変数に関しては過去実績の最大値もしくは最小値、目的変数に関しては表 3 に記載の操作速度で規格化した。図 2 および表 4 に本稿における深層学習モデルの詳細と学習条件を示す。学習データ数は 1 000 000、検証データとテストデータ数はともに 200 000 とした。中間層数は 7、中間層ノード数は 512、一度にネットワークに入力されるデータ数であるバッチサイズは 128 とした。学習の繰り返し数である epoch 数は 500 とした。学習計算における重み係数の最適化は式 (4)～(6) に示す adam 最適化¹⁰⁾を用いた。

$$v_t = \beta_1 v_{t-1} + (1 - \beta_1) G \dots \dots \dots (4)$$

$$s_t = \beta_2 s_{t-1} + (1 - \beta_2) G^2 \dots \dots \dots (5)$$

$$w_t = w_{t-1} - \alpha \frac{v_t}{\sqrt{s_t + \varepsilon}} \dots \dots \dots (6)$$

ここで、 β_1 、 β_2 はそれぞれ 1 次モーメント用係数、2 次モーメント用係数であり、Kingma ら¹⁰⁾ によって $\beta_1=0.9$ 、 $\beta_2=0.999$ が提案されている。 ε はゼロ除算を防止するための定数で 1×10^{-7} とした。 α は学習率であり、初期値を 1×10^{-3} とし、検証データに対する損失関数が連続して 10 epoch 改善しない場合に学習率を半減させて学習計算を継続した。さらに学習率に関係なく、検証データに対する損失関数が連続して 30 epoch 改善しない場合には学習計算を

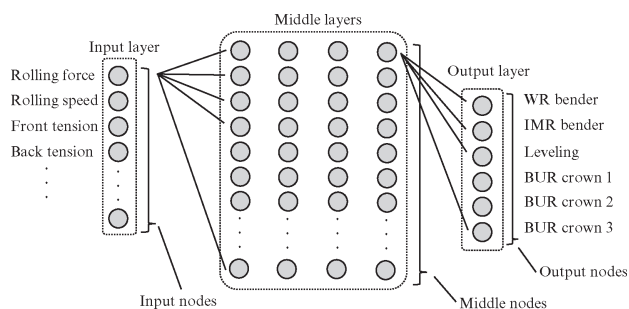


図 2 ディープラーニングネットワークの概要図

Fig. 2 Schematic diagram of deep learning network

表 4 ディープラーニングネットワークの学習条件

Table 4 Training conditions of deep learning network

Number of training data	1 000 000
Number of validation data	200 000
Number of test data	200 000
Number of input data	23
Number of output data	6
Number of middle layer	7
Number of node	512
Epoch	500
Loss function	mse
Optimizing method	adam

打ち切った。汎化性能向上のため、ニューラルネットワークのノードを一定の確率で消失させることで汎化性能を向上する Dropout¹¹⁾、およびデータバッチごとに分布を修正する Batch Normalization を導入した。Dropout 確率は 0.1 とした。

本システムの概要を図 3 に示す。説明変数である入力データはストレージに保存された後、変数ごとの単位換算や規格化、異常値の除去などが行われる。ここで平坦度制御アクチュエータ操作量は絶対値から 1 秒当たりの操作速度に規格化される。続いて後述するデータ重要度をすべてのデータに対して計算した後、再度ストレージに保存される。ネットワークの学習計算はこれらの処理とは独立して実施する。図中 dense layer とは式 (7) に示すとおり入力ベクトルと重み係数の内積計算である。

$$y_j = x_i w_{i,j} + b_j \dots \dots \dots (7)$$

ここで i は各層の入力データ数、 j は出力データ数である。学習計算が終了すると、既存モデルは新規モデルで上書きされる。その後、新規モデルに説明変数が入力され、ディープラーニングネットワークの順伝播計算により平坦度制御アクチュエータ操作量の予測値が出力される。最後に平坦度制御アクチュエータ操作量の上下制限約および時間当たりの操作量制約をチェックし、異常がなければ平坦度アクチュエータを操作する PLC へと出力される。プラント保護のため、モデル出力が一定の上限値を超えたら制御がオフとなる仕様とした。

本稿においてこれまで示した学習条件を用いた際、実機適用時に、平坦度制御アクチュエータが時間経過とともに徐々に操作される現象があった。これはモデルのテストデータに対する予測精度不足に起因すると考え、さらなるモデルの安定性向上策としてデータ重要度計算機能を導入した。データ重要度計算機能はディープラーニングネットワークの学習データを効率的に選定する機能であり、学習計算の効率化に寄与するのみならず、学習に不適なデータを取り除くことで、テストデータに対するモデル出力の安定性を向上させることができる。データ重要度計算機能の詳細を図 4 に示す。データの重要度は式 (8) で定義する。ここで ε は平

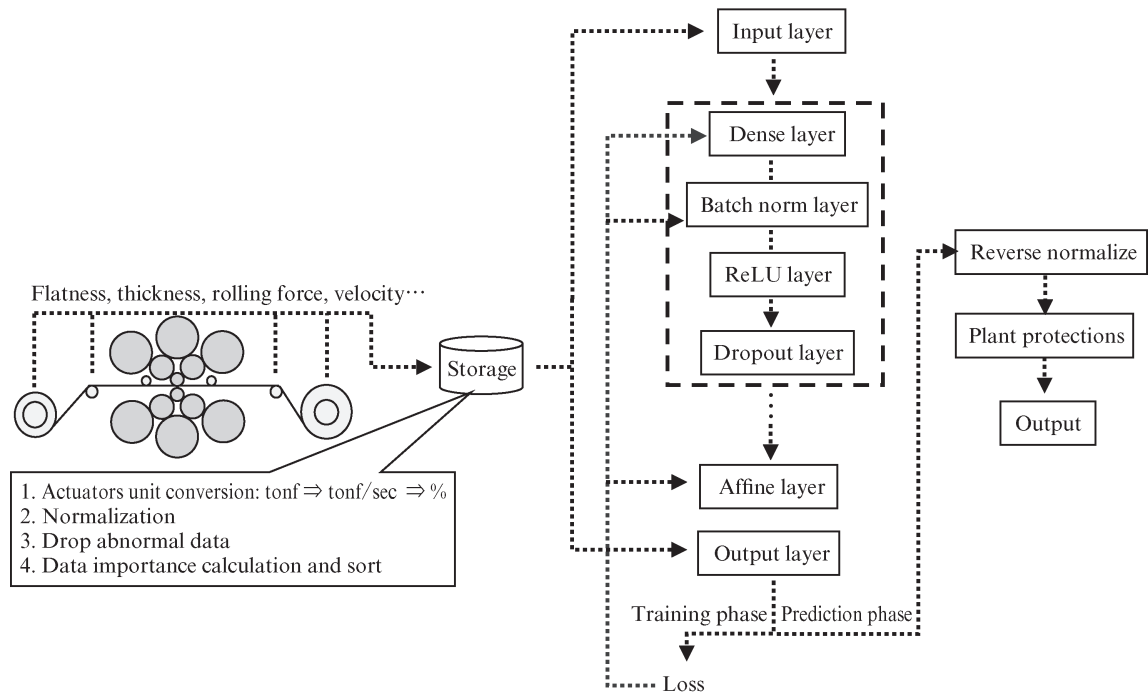


図 3 学習行程, 予測行程での計算フロー

Fig. 3 Details of calculation procedure in training and prediction phase

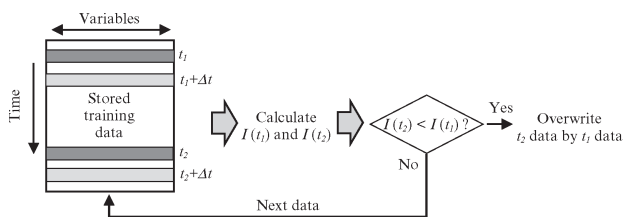


図 4 データ重要度計算の計算フロー

Fig. 4 Schematic diagram of the data importance calculation

平坦度で実測した各チャンネルの板の長手方向伸び, ch は幅方向チャンネル数, 添え字 $diff$ は目標値と実測値の差を示す。 Δt は重要度計算のための時間差である。式 (8) は, 特定のタイミングのデータが Δt 時間経過後にどれだけ目標平坦度に近づいたかを示す値である。この重要度をすべてのデータに対して計算した後, 重要度の高いデータから優先的にディープラーニングネットワークの学習計算に使用する。なお本稿に示す学習データ数および検証データ数はデータ重要度計算によって選定した後のデータ数である。なお, テストデータに対しては, 実運用におけるモデル出力を評価する必要があるため, データ重要度計算によるデータ選定は適用しない。

$$I(t) = \sum_{i=1}^{ch} [|\epsilon_i^{diff}(t)| - |\epsilon_i^{diff}(t + \Delta t)|] \quad \dots\dots\dots (8)$$

2.3 ハイパーパラメータ最適化とオフライン検証

ハイパーパラメータである中間層数, 中間層ノード数,

バッチサイズの 3 つのパラメータは, 交差検証とグリッドサーチを用いて最適化した。具体的には学習データを 5 分割した後, 1 つを検証データ, 残り 4 つを学習データとして定義する。学習計算の後, 検証データに対する損失関数を計算する。次にモデルの重みを初期化した後, 学習データと検証データを入れ替えて再度学習計算を行い, 検証データに対する損失関数を計算する。これを 5 回繰り返した後, 最後にテストデータに対する損失関数を計算する。以上の計算によって, 1 つのハイパーパラメータの組み合わせに対して, 5 回の検証データに対する損失関数の平均値とテストデータに対する損失関数が得られるため, それらをモデルの性能を示す代表値とした。さらにグリッドサーチによって, 検証データ平均値およびテストデータの損失関数が最小となるハイパーパラメータの組み合わせを決定した。

実機導入の前段階として, オフラインでモデルの予測時の挙動を調査した。ここではデータを時系列順に並べたうえで連続的に平坦度制御アクチュエータ操作量を予測させ, モデルの予測精度および時間方向の安定性を評価した。計算結果を図 5 に示す。図中央破線部が学習データとテストデータの切り替わりタイミング, 図 5 (a) がハイパーパラメータ最適化およびデータ重要度計算によるデータ選定適用前, 図 5 (b) が適用後である。学習データに対する予測精度はいずれの場合でも同程度であるが, ハイパーパラメータ最適化およびデータ重要度計算適用後は, 適用前と比較してテストデータに対してより良好に予測できていることが確認できる。このことから, 交差検証に基づくハイパーパラメータ最適化およびデータ重要度計算に基づく学習

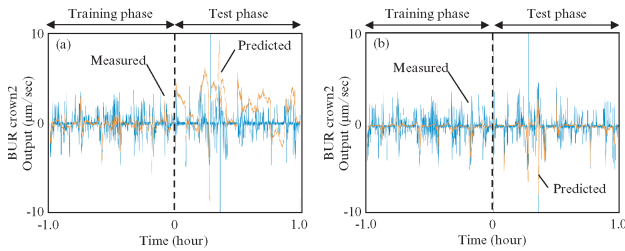


図5 ディープラーニングネットワークによる BUR クラウン 2 の計算値と実測値の比較
ハイパーパラメータ最適化およびデータ重要度計算
(a) 適用前 (b) 適用後

Fig. 5 Comparison of predicted BUR crown2 by deep learning network
(a) without and (b) with hyperparameter optimization and data importance calculation

データ選定は、モデルの汎化性能を向上させるのに有効な手段であることを確認した。

3. 実機導入

本モデルを実機へ導入し、平坦度制御アクチュエータ操作量と操作頻度をオペレータの手介入操作と比較した結果を図6に示す。図6は入側板厚 1.0 mm、板幅 978 mm のコイルの 4 パス目の圧延速度、平坦度制御アクチュエータ操作量、実測平坦度の 2 次成分と 4 次成分を比較した結果であり、図6 (a) がオペレータによる手介入操作、図6 (b) がディープラーニングを用いた制御である。平坦度制御アクチュエータは、主に操作される BUR クラウン 1 と BUR クラウン 2 を示した。ディープラーニングを用いた平坦度制御は、平坦度制御アクチュエータの操作量のみを出力するため、両条件とも圧延速度およびロールギャップなどはオペレータによる手介入操作である。圧延速度に着目すると、両条件で同等の圧延速度であることが確認できる。ディープラーニングを用いた平坦度制御では、オペレータの手介入操作と同等の平坦度制御アクチュエータ操作が出力されており、時系列的な変化も良好に一致している。圧延後平坦度もオペレータの手介入操作と同等の平坦度が実現できている。

続いて長期テストによって得られた結果を図7に示す。これは、1~10 パス、板幅 930~1 070 mm、計 1 959 パスの平均値と標準偏差を調査したものである。図7 (a) 縦軸は 1~4 次の平坦度成分ごとの各パスの目標平坦度と実測平坦度の差である。対称成分に着目するとディープラーニングを用いた本制御では手介入操作と同等の圧延後平坦度が得られていることが分かる。図7 (b) は圧延中の平坦度制御アクチュエータの手介入操作率である。ここで平坦度制御アクチュエータの手介入率は、全データに占めるいずれかの平坦度制御アクチュエータを手介入操作したデータの割合として定義した。ディープラーニングを用いた本制御では、手

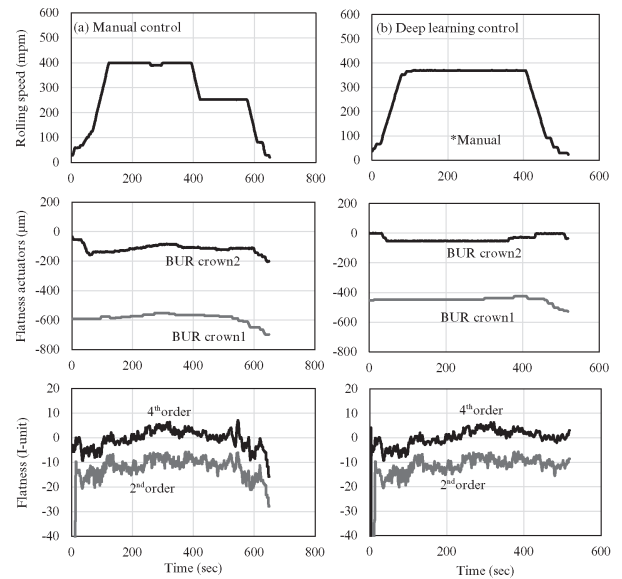


図6 手介入操作とディープラーニング制御の圧延実績の比較
Fig. 6 Comparison between manual control and deep learning control in the mill trial

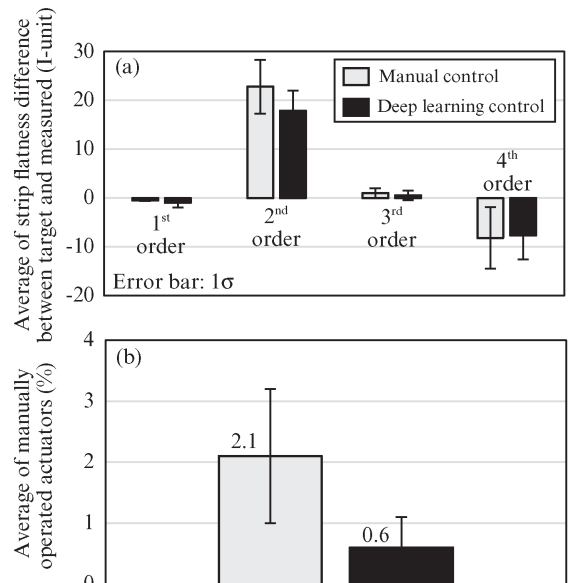


図7 実機テスト結果
(a) 平坦度実績 (b) 平坦度制御アクチュエータの手介入率
Fig. 7 Results obtained by mill trial
(a) strip flatness difference between target and measured (b) manually operated flatness actuators

介入操作と同等の平坦度を、少ない手介入回数で実現できていることが分かる。

図8に、本制御を用いて操作した場合と、同条件でオペレータの手介入によって操作した場合の、圧延開始からの平坦度と平坦度制御アクチュエータ手介入率の推移を示す。本制御を用いた場合は手介入操作と比較して迅速に目標平坦度に収斂しており、かつ対応する平坦度制御アクチュエータの手介入操作の割合が少ないことが確認できる。

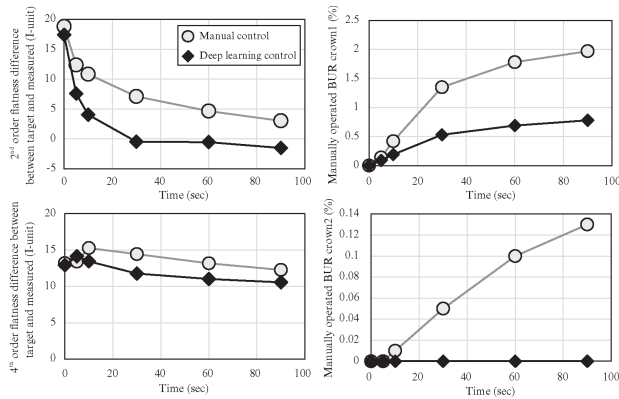


図 8 平坦度実績の 2 次成分と BUR クラウン 1 の手介入率，平坦度実績の 4 次成分と BUR クラウン 2 の手介入率の時間変化

Fig. 8 Comparisons of time-dependent behavior of strip flatness and flatness actuators under manual control and deep learning control

4. 考察

ディープラーニングを適用した平坦度制御が，オペレータが操作する平坦度制御アクチュエータと同等の平坦度制御アクチュエータ操作量を出力できる理由を考察するため，ディープラーニングネットワークが有する影響係数と実機で平坦度制御アクチュエータを独立に操作した際の実測値とを比較した。パス数は 8 パス目であり，板厚は 1.0 mm，板幅は 1 020 mm であった。

平坦度アクチュエータ操作量を目的変数としたニューラルネットワークでは，影響係数を算出するために逆解析が必要となるため，図 2 に示した条件から説明変数と目的変数のみを入れ替えて，目的変数に平坦度，説明変数に平坦度制御アクチュエータ操作量を用いて，損失関数が収束するまで十分に学習した。ニューラルネットワークにおいて，影響係数は入力と出力の関係性を示す値であることから，説明変数と目的変数を入れ替える影響は小さいと考えた。ディープラーニングが出力する影響係数は，平坦度制御アクチュエータの操作量自体に影響を受けるため，実測値と計算値の平坦度制御アクチュエータの操作量は同等とした。なおレベリングに関しては，トラブルを誘発する可能性があるため，影響係数の実測および比較は実施しなかった。

影響係数の比較結果を図 9 に示す。図 9 (a) が実測値，図 9 (b) がディープラーニングによる計算値である。各平坦度制御アクチュエータの操作量は図中に併記した。ディープラーニングによって計算した影響係数は，WR ベンダーでは実測値と比較して大きく，IMR ベンダーでは小さく計算されたが，いずれも操作すると腹伸びになるという定性的な一致が確認された。BUR クラウンはいずれも実測値と良好に一致した。

ディープラーニングが学習によって妥当な影響係数を獲

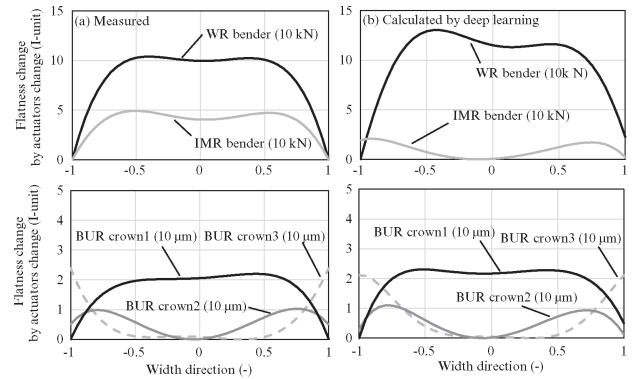


図 9 平坦度制御アクチュエータの影響係数
(a) 実測値 (b) ディープラーニング計算値

Fig. 9 Comparison between (a) measured and (b) calculated influence coefficients

得できるかに関しては，コンピュータビジョンの分野で報告が見られる。Hao ら¹²⁾ は同一のデータセットを複数の条件で学習した際の損失関数の可視化を試みており，学習によって決定される重み係数が損失関数の大小のみで一つに決まらない場合があること，さらには学習条件や良好な学習データ選定，パラメータ選定などによって妥当な結果を得やすくなることを報告している。本稿では前述した交差検証やハイパーパラメータ最適化，さらにはデータの重要度計算に基づくデータ選定ロジックによって良好な学習が行われていると考えられる。以上から，本稿におけるディープラーニングネットワークは，学習によって実測値と同等の影響係数を獲得する能力を有しており，その結果オペレータが操作する平坦度制御アクチュエータを正しく出力できたものと考えられる。

5. おわりに

本稿では，影響係数の実測を必要としない新しい平坦度制御技術として，ディープラーニングを用いた 12 段クラスター型圧延機の平坦度制御システムを構築し，実機でその有用性を検証した。

- (1) ディープラーニングを適用した平坦度制御を用いることで，より低い平坦度制御アクチュエータ手介入率で手介入操業と同等の圧延後平坦度を実現できた。さらに手介入操業と比較して，目標平坦度への収束が早いことが確認された。
- (2) ディープラーニングネットワークが学習によって獲得した影響係数を，同一の条件での実測値と比較した結果，ディープラーニングネットワークが獲得した影響係数は WR ベンダー，IMR ベンダーに対しては定性的に一致しており，BUR クラウンに関しては良好な一致を確認できた。

本技術は，JFE スチールのソリューションビジネスである

「JFE Resolus[®]」で提供中である。

参考文献

- 1) 原口成人, 梶原利幸, 秦和宜. 冷間圧延におけるストリップの形状. 日本機械学会誌. 1970, vol. 73, no. 616, p. 643-649.
- 2) 坪井邦夫, 川畑成夫, 山田健夫. 鋼板圧延プロセスにおける形状検出—渦電流平坦度計の開発—. 計測と制御. 1979, vol. 18, no. 6, p. 510-518.
- 3) 安部可治, 関口邦男. 冷間圧延の平坦度制御. 電学論 D. 2003, vol. 123, no. 10, p. 1213-1218.
- 4) Ogasahara, T.; Kitamura, T.; Aoe, S.; Tateno, J.; Asano, K. Dynamic Control of Flatness and Elongation of the Strip in a Skin Pass Mill. ISIJ Int., 2020, vol. 60, no. 12, p. 2886-2895.
- 5) 佐藤弘幸, 大島仁志, 前田兼一朗, 國弘祐司, 石井秀美. 周南製鉄所 No. 4ZM への自動形状制御システム導入. 日新製鋼技報. 1995, no. 72, p. 45-50.
- 6) 梶原利幸, 藤野伸弘, 杉山徳治. ストリップ圧延における自動制御. 日立評論. 1971, vol. 53, p. 842-846.
- 7) Zhou, G.; Li, H.; He, A.; Liu, C.; Sun, W.; Liu, Z.; Han, C. Simulation and control of high-order flatness defect in rolling wide titanium strip with 20-high mill. The International Journal of Advanced Manufacturing Technology. 2022, vol. 120, p. 5483-5496.
- 8) Peng, Y.; Liu, H.; Du, R. A neural network-based shape control system for cold rolling operations. Journal of materials processing technology. 2008, vol. 202, p. 54-60.
- 9) Raju, D. V.; Iqbal, A.; Trivedi, A. K.; Mukhopadhyay, A. Prediction of shape defects over length of cold rolled sheet using artificial neural networks. Ironmaking & Steelmaking. 2007, vol. 34, no. 2, p. 166-176.
- 10) Kingma, D. P.; Ba, J. L. ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION. 3rd International Conference for Learning Representations. 2015.
- 11) Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research. 2014, vol. 15, p. 1929-1958.
- 12) Li, H.; Xu, Z.; Taylor, G.; Studer, C.; Goldstein, T. Visualizing the Loss Landscape of Neural Nets. 32nd Conference on Neural Processing Systems. 2018.