

時系列センサーデータによる鑄造品質の要因分析に向けた 深層学習モデルの検討

吉井 碧海[†] 山本 泰生^{††} 西村 雅史^{††}

[†] 静岡大学情報学部情報科学科 〒432-8011 静岡県浜松市中区城北3丁目5-1

^{††} 静岡大学大学院総合科学技術研究科 〒432-8011 静岡県浜松市中区城北3丁目5-1

E-mail: [†]yoshii@yy-lab.info, ^{††}{yyamamoto,nisimura}@inf.shizuoka.ac.jp

あらまし 製造現場で利用される多くの部品は金型鑄造により成形されるが、鑄造工程は直接目視することができず、どのような条件下で不良が発生するかを把握することが困難な状況にある。本研究では、鑄造機に取り付けられたセンサーから得られる時系列データをもとに品質の予測・分析を行う。初めに鑄造機に取り付けられたセンサーデータを元に深層学習による鑄造製品の品質予測を行う。また、モデルの学習に用いられる教師データにはノイズラベルとなるデータが含まれている可能性もあることから、学習時に教師データのラベルの更新と信頼度の評価を行う。次に分類時の予測モデルの Attention 層を可視化することで予測を決定付ける要因となる製造の時間区間を分析する。実験結果から良品・不良品判定に影響を与える時間区間を部分的に絞り込むことができた。

キーワード 深層学習, 時系列センサーデータ, Attention, ノイズラベル

1 はじめに

自動車や家電などの工業製品のパーツを製造する手法の一つに金型鑄造・ダイカストがある。ダイカスト¹とはアルミニウムや亜鉛、マグネシウムなどの合金を高温で溶かし、金型に流し込む鑄造法の一つである。溶融金属を高速・高圧で注入するため、生産サイクルが短く、大量生産が可能である一方、空気やガスの混入による鑄巣と呼ばれる空洞が発生することがある。鑄巣が発生すると製品の耐久性や強度に影響を与えるため不良品として扱われるが、鑄造工程は直接人の目で見ることができず、どのような条件下で鑄巣が発生するのかを把握することが困難な状況にある。

他方、鑄造機械には複数のセンサーが取り付けられており製造時の機械の状態を時系列データとして鑄造製品毎に記録している。本研究では鑄巣の発生した不良品のセンサーデータの状態を用いて不良品が発生する際の重要特徴量を発見する課題に取り組んでいる。

これまで決定木ベースの機械学習モデルを用いて品質を決定づける要因を探索してきた。しかし、モデルの精度が十分ではなく、かつ人の手による特徴量設計には限界があるという問題があった。本研究では、はじめに再帰型ニューラルネットワーク (Recurrent Neural Network: RNN) モデルを使用することで元の時系列データを入力とする予測モデルを構築し、その精度を示す。また、精度向上と時間区間ごとの重要度を算出することを目的として Attention [1] を導入する。

本研究課題の難しさの一つとしてデータセットの中にノイズラベルが含まれている点が挙げられる。鑄造品は外観検査や圧力測定検査で不良品のスクリーニングを行っているが、外部に

現れない内側の鑄巣は見落とされており、良品とラベル付けされている可能性がある。ノイズラベルとなるデータが教師データに含まれている場合、モデルが適切にデータの特徴を学習することができず、精度に悪影響を与える。

この課題を克服すべく、本研究では Joint Optimization Framework for Learning with Noisy Labels [2] と呼ばれるラベル更新フレームワークを用いたスクリーニング法を提案する。[2] ではモデルの学習時にパラメータの更新と共にラベルの更新を行うが、本研究では更新過程のラベル値の変動をもとにサンプルの良品度を評価する。これにより、良品データの中から不良品の可能性の高いサンプルを排除することを狙う。

以上の手法を用いて鑄造製品の品質予測実験を行った結果、分類精度を大幅に向上することができた。また、Attention の重みを可視化することで製品の品質に影響を与える時間区間を一部特定することができた。

2 先行研究

鑄造工程における鑄巣発生についての研究は数多く行われてきた。品田ら [10] はアルミニウム合金ダイカストに観測される鑄巣の発生状況と、それに関係しているガスの組成と分布を質量分析計を用いて調べ、鑄巣に関係するガスの挙動を明らかにした。田中ら [11] は鑄物内部に空気やガスが含まれているかを検査するブリスト試験と X 線 CT を併用し、鑄巣の種類を判別するシステムを開発している。鑄造機械のセンサーデータを用いた研究例として鈴木ら [12] は自動生産における設備状態と運転条件の適切な調整方法の確立を目的として、設備の運転状態のモニタリングデータと鑄巣の分布状態を解析している。

昨今、製造業界ではデジタルデータの活用や品質・生産性の向上に向けて機械学習の導入も盛んに行われている。田中ら [13]

¹ : <http://www.diecasting.or.jp/index.php>

は画像による製品の外観検査を目的として半教師あり異常検知を用いた欠陥検出アルゴリズムを開発している。

著者らはこれまでに同鋳造製品を対象とした品質分析の研究を実施しており、ward 法による時系列データのクラスタリングや、時系列変化を可視化することによる良品・不良品の挙動比較、決定木ベースの XGBoost [9] により分類、ならびに重要特徴量の可視化を行ってきた。しかし、先行研究では良品・不良品を分ける明確な差異は確認できなかった。また決定木ベースの手法では、十分な分類精度のモデルを構築することができなかった。

先行研究の課題として以下の 2 点が挙げられる。1 点目は、入力に元の時系列データを使用していない点である。先行研究で使用していたデータは時系列データを元に専門家らによって設計された特徴量の定点値から構成される。この定点データでは、元の時系列情報が活かされていない。またドメイン知識に依存するため、未知の要因発見に繋がる可能性は低い。2 点目は学習データにノイズラベルが含まれている可能性がある点である。鋳造品は外観検査や圧力測定検査で不良品のスクリーニングを行っているが、外部に現れない内側の鑄巣は見落とされて、良品とラベル付けされている可能性がある。仮に鑄巣が発生するときの鋳造機械の状態をセンサーが観測していた場合、その鋳造製品に良品ラベルが付けられていればモデルの学習の妨げとなってしまう。

本研究では以上の 2 点の課題を克服する手法を提案する。

3 準備

本章では提案手法で使用する技術について述べる。

3.1 Attention

Bahdanau ら [1] は自然言語分野の機械翻訳タスクにおいて EncoderDecoder の改良版として Attention を提案した。EncoderDecoder では入力時系列データが Encoder から Decoder に渡される際に固定長ベクトルにまとめられてしまい、時系列情報を上手く利用できていないという欠点が存在する。Attention ではこれを解決するために、各隠れ層の出力ベクトルに対し加重平均を取る。それによって、出力に寄与する時間情報をより効果的に取捨選択することを可能にした。さらに Attention では加重平均をとる際の重みを利用することで、出力時にどのタイムスタンプをモデルが重要視しているかを可視化することができる。Attention は自然言語処理の分野に止まらず画像認識、音声認識、質問応答など様々な分野で応用されている。本研究では RNN をベースとするモデルに Yang ら [8] が提案した分類問題用の Attention を組み込むことで、良品と不良品を分類する際に重要となる時間区間の特定を試みる。

3.2 Joint Optimization Framework for Learning with Noisy Labels

機械学習において効率的に学習を進めるには正確なラベル付けが必要不可欠となる。一方で、データセットを作成するのは人間であり、タスクの難易度によってはドメイン知識を求め

られ、誤ってラベル付けされる可能性もある。データセット内にノイズラベルを含んでいる場合にも学習を進める方法について多くの先行研究がある [2] [3] [4] [5]。その中で、本研究ではパラメータの学習と同時に真のラベルの推定を行う、Joint Optimization Framework for Learning with Noisy Labels (以下、「JOF-NL」と略す)を用いる。JOF-NL では、内部の DNN モデルを自由に設定でき、かつ特定クラスのラベル更新が可能であるといった特徴を持つ。JOF-NL のフレームワークを図 1 に示す。

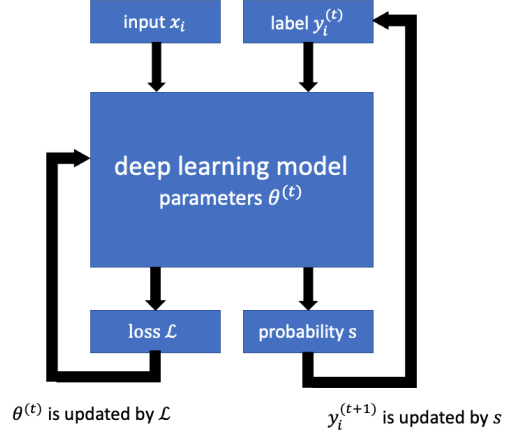


図 1 JOF-NL のフレームワーク

この手法は大きく 2 つのステップに分かれる。第一ステップではまずラベル更新を行う。ノイズラベルを含むデータによりモデルのパラメータの学習を進めたあと、あるエポックからパラメータの学習とラベルの更新を 1 回ずつ行う。パラメータを θ 、学習データを X 、学習データのサンプル数を n 、クラスラベルを Y 、ラベル数を c とするとコスト関数 L は式 (1) と定義される。

$$L(\theta, Y|X) = L_c(\theta, Y|X) + \alpha L_p(\theta|X) + \beta L_e(\theta|X) \quad (1)$$

$L_c(\theta, Y|X)$ が分類に関する損失関数であり、 $L_p(\theta|X)$ 、 $L_e(\theta|X)$ は正則化項、 α と β はハイパーパラメータである。 L_c はカルバック・ライブラー情報量を用いて式 (2) で表される。

$$L_c(\theta, Y|X) = \frac{1}{n} \sum_{i=1}^n D_{KL}(y_i || s(\theta, x_i)) \quad (2)$$

$$D_{KL}(y_i || s(\theta, x_i)) = \sum_{j=1}^c y_{ij} \log \left(\frac{y_{ij}}{s_j(\theta, x_i)} \right) \quad (3)$$

コスト関数が $L_c(\theta, Y|X)$ のみの場合、全ての x_i に対し、定数 y (ただし y はワンホット) を返すような学習を行う [2]。 L_p はこれを防ぐために設けられており、式 (4) で表される。ただし、 p はクラスの事前分布、 $\bar{s}_j(\theta, X)$ は式 (5) で与えられるミニバッチ B での出力確率の平均である。

$$L_p(\theta|X) = \sum_{j=1}^c p_j \log \frac{p_j}{\bar{s}_j(\theta, X)} \quad (4)$$

$$\bar{s}(\theta, X) = \frac{1}{n} \sum_{i=1}^n s(\theta, x_i) \approx \frac{1}{|\mathcal{B}|} \sum_{x \in \mathcal{B}} s(\theta, x) \quad (5)$$

$L_e(\theta|X)$ はクラスラベルをソフトラベルとして更新する場合、局所解に陥ることを防ぐために設けられたエントロピー損失項である。式 (6) に示す。

$$L_e(\theta|X) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^c s_j(\theta, x_i) \log s_j(\theta, x_i) \quad (6)$$

JOF-NL ではラベルの更新をハードラベルで行う場合とソフトラベルで行う場合があるが、本研究ではソフトラベルを使用している。なお、ラベルは現在のエポックから数エポック前までのモデルの出力結果で更新される。

第二ステップではモデルのパラメータを一旦初期化してから、ラベル更新後の訓練データを使って通常の教師あり学習を行う。

本研究では使用する鑄造データセットは不良品データについてはラベルが保証されているので、良品データについてのみラベルの更新を行う。

4 提案手法

本章では鑄造時系列センサーデータを用いて良品と不良品の分類予測を行う手法について述べる。

4.1 良品データのスクリーニング

JOF-NL により良品データのラベル更新ができるようになるが、更新されるラベルは最終エポックの更新値で定まり、エポック数によって最終的なラベル値が変化することがある。ここで、エポック毎のラベル値に着目すると値が不変なサンプルと変動するサンプルの2つに分けることができる。本研究では、ラベル値が常に良品であるサンプルを優先的に選別することを考える。

まず、製品番号 i における j 番エポックの不良品ラベルに対する尤度値を $l_{i,j}$ と書く。すなわち、 $l_{i,j}$ の値が1に近いほど不良と認識されていることになる。このとき、製品番号 i の評価値 e_i を式 (7) のように定める。ただし N はエポック数とする。

$$e_i = \max_{1 \leq j \leq N} l_{i,j} \quad (7)$$

e_i が低いものはラベル更新において安定して良品と判定されているものであり、不良品データとは異なる時系列パターンを有していると考えられる。よってこの評価値 e_i をもとに良品らしいサンプルを選定する。

4.2 使用モデル

本節では使用する深層学習モデルについて述べる。一層目に再帰型ニューラルネットワークの1つであるGRU [7] を用いる。GRU は時系列タスクでよく使用される LSTM(Long Short-Term Memory) [6] と類似した性能を持ちつつ、計算コストを減らしたアーキテクチャである。GRU ブロックの構造を図2に示す。

このGRU ブロックを時間軸の前から後ろの流れを読み取る

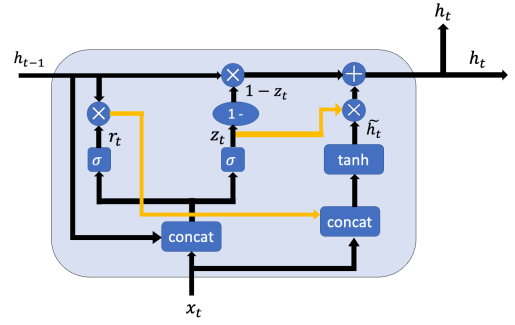


図2 GRU ブロックの構造

層と、後ろから前の流れを読み取る層の2つ用意する。

次に、GRU からの各出力を Attention 層の入力とする。Attention 層 [8] ではGRU からの出力をそれぞれに重みを乗算することで加重平均を行う。重み a 及び Attention の出力 v は以下の式 (8) から (10) により算出する。なお、 h はGRU の出力、 W_a, b_a はそれぞれ一層の多層パーセプトロンの重みとバイアス項、 u_a はコンテキストベクトル、 t はタイムスタンプである。

$$u_t = \tanh(W_a h_t + b_a) \quad (8)$$

$$a_t = \frac{\exp(u_t^T u_a)}{\sum_t \exp(u_t^T u_a)} \quad (9)$$

$$v = \sum_t a_t h_t \quad (10)$$

これにより分類に大きな影響を与えるタイムスタンプの情報を効果的に取捨選択することができ精度が向上するだけでなく、時系列の重要区間がどこかを可視化することができる。最後に出力 v に対して全結合層を用いることで良品・不良品の出力確率を計算する。

使用する深層学習モデルの概要図を図3に示す。 x が入力データ、 h はGRU ブロック、 a は Attention の重み、 t はタイムスタンプである。

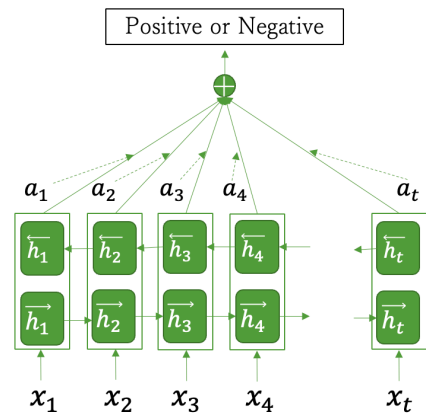


図3 使用モデル概要図

5 評価実験

本章では提案手法による分類精度の評価実験を行う。

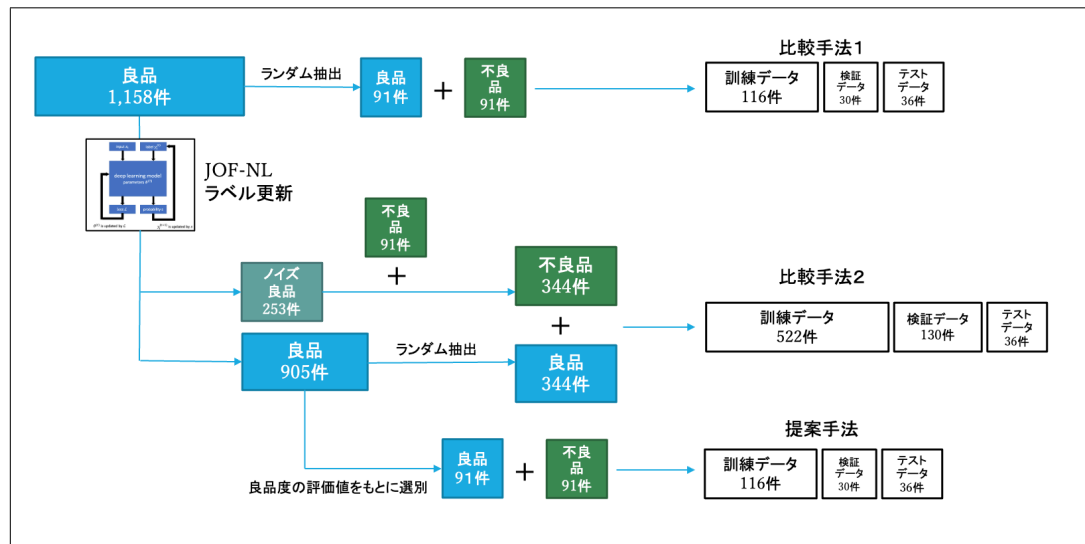


図 4 実験方法概要図

5.1 使用データ

使用データは金型鑄造機械に取り付けられた 52 種類のセンサーにより観測された製品ごとの時系列データである。サンプリングレートは 100Hz。タイムスタンプ数 15,000 で記録されており、良品サンプル数として 1,158 件、不良品サンプル数として 91 件をそれぞれ抽出し利用する。なお良品と不良品の抽出割合は不良率と関係ない点に注意されたい。前処理として製品ごとに平均 0、分散 1 の正規化処理を行い、ノイズ除去と過学習抑制のため区間幅 10 で平均を取り、一つのタイムスタンプとするビンニング処理を施している。

5.2 実験設定

5.1 節のデータを使用し、提案手法の分類実験を行う。提案手法は 2 つのステップに分かれる。使用モデルは両ステップとも 4.2 節のものである。第一ステップは 4.1 節で説明した通り良品のラベル更新とスクリーニングを行う。エポック数は 150 で 30 エポック目からラベルの更新を行う。更新するラベルは過去 10 エポックのモデルの出力確率の平均とする。最適化手法は SGD で学習率は 0.03、モーメントが 0.9、weight decay が 10^{-4} 、コスト関数の α が 6.0、 β が 0.4 である。良品度の評価値を算出する式 (7) をもとに不良品サンプル数と同じ 91 件を良品として選定する。

第二ステップでは第一ステップで得られたデータを用いて教師あり学習を行う。エポック数は 80、最適化手法に Adam、学習率は 0.025、GRU の隠れ層のノード数が 10 である。なお、バッチサイズは第一、第二ステップともに 100 とする。実験は Intel Xeon E5-165 の CPU と Nvidia GEFORCE GTX 1080 の GPU、131GB のメモリを搭載したマシン上で実施した。

5.3 実験方法

提案手法で良品データの選別を行うことにより、どの程度分類精度が向上するのかを確認するために比較手法を 2 つ用いて実験を行う。実験方法の概要図を図 4 に示す。

比較手法 1 はラベル更新やスクリーニングを行わず、良品サ

ンプル 1,158 件からランダムに 91 件を抽出し、不良品サンプル 91 件と合わせて 1 つのデータセットとする。比較手法 2 は JOF-NL によりラベル更新を行い、良品サンプルから新たに不良品ラベルに変更となった 253 件と元の不良品サンプル 91 件を合わせた 344 件を新たな不良品サンプルとする。良品・不良品の比率を合わせるため、良品サンプルも 344 件ランダムに抽出する。最後にラベル更新された 905 件から良品度の評価値をもとに選別した 91 件と不良品 91 件を合わせて作成されるデータセットが提案手法のものである。提案手法は 20% をテストデータ、64% を訓練データ、16% をハイパーパラメータチューニング用の検証データとする。比較手法 1、比較手法 2 も提案手法のテストデータと同サイズのテストデータをデータセットから作成し、残りを 8:2 で分割して訓練データ、検証データとする。これをテストデータの組み合わせを変えて 5 回実験を行い、その平均値を分類精度とする。テストデータは各手法それぞれで作成されるものを用いる場合と提案手法により作成されるものを共通で用いる場合の 2 ケースで実験する。

5.4 分類精度

検討手法の分類精度を検証する。評価指標は正解率と、不良品の分類に対する適合率、再現率、F1 とする。実験結果を表 1 に示す。

全ての評価指標において提案手法が最も精度が高く、次いで比較手法 2、比較手法 1 という結果となった。ラベル更新やスクリーニングを行わない比較手法 1 は正解率が 0.5 前後であり、有意に分類できているとは言えない。比較手法 2 は比較手法 1 よりも精度が向上しており、JOF-NL によりノイズラベルの更新ができていると考えられる。提案手法は比較手法 2 よりもさらに高い精度であり、精度自体もモデルの有用性を保証できる値であることから提案手法では第一ステップから不良品データとは異なるパターンを持つ良品サンプルを選定することができ、第二ステップでそのデータのパターンの違いをモデルが見つげ出すことができたと考えられる。

表 1 分類結果

	テストデータ (手法別)				テストデータ (共通)			
	正解率	適合率	再現率	F1	正解率	適合率	再現率	F1
提案手法	0.80	0.79	0.79	0.79	0.80	0.79	0.79	0.79
ランダム抽出 (比較 1)	0.49	0.48	0.34	0.39	0.51	0.35	0.35	0.34
JOF-LN (比較 2)	0.62	0.62	0.60	0.61	0.73	0.72	0.67	0.69

5.5 Attention の重みの可視化

5.4 節で提案手法では十分なモデルの分類精度が得られたことを確認できた。本節ではテストデータの分類予測時の Attention の重みを可視化することで良品と不良品の分類時に重要となる時間区間の特定を行う。

まず、有用な精度を実現できなかった比較手法 1 の、良品を正しく良品と予測できたときと不良品を正しく不良品と予測できたときの Attention の重みをそれぞれ図 5 と図 6 に示す。

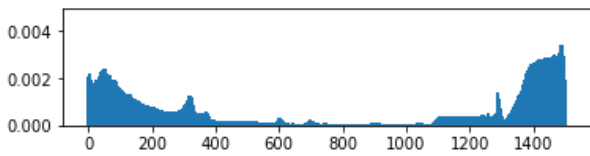


図 5 良品データを良品と予測した時の Attention の重み (比較 1)

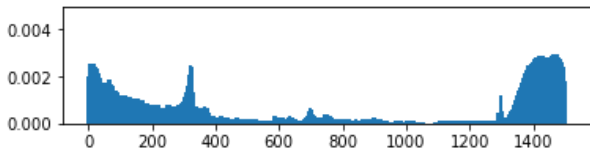


図 6 不良品データを不良と予測した時の Attention の重み (比較 2)

図 5 と図 6 とともにタイムスタンプ 0 から 400 の区間と 1300 から 1500 の区間で重要度が高くなる結果となり、良品判定と不良品判定で明確な違いは見られなかった。

次に図 7 と図 8 に提案手法で良品データを良品と予測したとき、不良品データを不良品と予測したときの Attention の重みを示す。

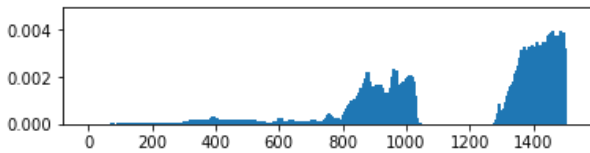


図 7 良品データを良品と予測した時の Attention の重み (提案手法)

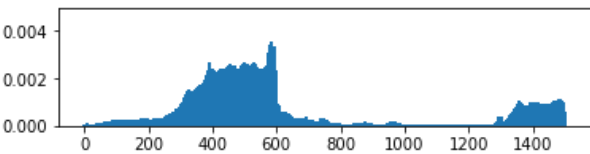


図 8 不良品データを不良と予測した時の Attention の重み (提案手法)

提案手法によって得られたデータセットで学習することによって、比較手法 1 とは異なる Attention の重みの分布となった。また、良品・不良品間でも重要とされる区間に違いが見られた。特に、図 8 ではタイムスタンプ 400 ± 200 区間が最も重要度が高くなっている。タイムスタンプ 400 前後は秒数に直すと鑄造開始から 40 秒付近で、鑄造工程に関わる専門家へのヒアリングの結果、金型へのアルミの装填が完了して型から取り出すあたりであることがわかった。工学的視点からも製品の品質を決定づける工程と考えられており、センサー情報にも品質に関わる特徴が含まれている可能性が高い。一方、図 7 ではこの区間は重要度が低いことから、特に不良品を決定付ける要因が潜在していると考えられる。

図 7 と図 8 の両方で重要度が高くなっている 1400 ± 100 区間や図 7 のみで重要となっている 800 から 1000 区間は既に鑄造品を型から取り出した後である。直接的に品質に影響を与える区間であるとは考えにくい、何らかの挙動から影響を受けた二次的な現象の可能性も含めて調査が必要であると思われる。

さらに今後は重要となる可能性の高い時間区間に絞った調査やセンサーごとの重要度などを調査し、より具体的な品質を決定づける要因を特定していきたい。

6 まとめと今後の課題

本論文では鑄造機に取り付けられたセンサーから得られる時系列データをもとに品質の分析を行うために深層学習を用いた分類実験を行った。結果として、ノイズラベルを含む可能性のある良品データから良品度を評価することでサンプルの選別を行い、分類精度を向上させることができた。また、Attention の重みを可視化することで分類予測時に重要となる時間区間を部分的に絞り込むことができた。

今後は具体的に品質に影響を与える要因を特定するため、センサーごとの重要度や、スクリーニングで選定されたサンプルとそうでないサンプルの間の差異などを調査する必要がある。

7 謝 辞

本研究においてデータ提供、ならびに貴重なアドバイスをいただいたヤマハ発動機株式会社吉岡一心様に厚く御礼申し上げます。本研究の一部は科研費 20K11935 の支援を受けている。

文 献

- [1] D. Bahdanau, K. Cho, and Y. Bengio. “Neural machine translation by jointly learning to align and translate” ICLR, 2016.
- [2] D. Tanaka, D. Ikami, T. Yamasaki, K. Aizawa. “Joint Op-

timization Framework for Learning With Noisy Labels” CVPR, 2018, pp. 5552-5560.

- [3] N. Natarajan, I. S. Dhillon, P. Ravikumar, A. Tewari. “Learning with Noisy Labels” NIPS, 2013, pp. 1196-1204.
- [4] S. Reed, H. Lee, D. Anguelov, C. Szegedy, D. Erhan, A. Rabinovich. “Training Deep Neural Networks on Noisy Labels with Bootstrapping” CVPR, 2014.
- [5] R. Tanno, A. Saeedi, S. Sankaranarayanan, D. C. Alexander, N. Silberman. “Learning From Noisy Labels By Regularized Estimation Of Annotator Confusion” CVPR, 2019.
- [6] S. Hochreiter, J. Schmidhuber. “Long Short-term Memory” In Proceedings of the Neural Computation, Vol. 9, No. 8, pp.1735–1780, 1997.
- [7] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555, 2014.
- [8] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, E. Hovy. “Hierarchical Attention Networks for Document Classification” Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.
- [9] T. Chen, C. Guestrin. “XGBoost: A Scalable Tree Boosting System” Proceedings of KDD’16.
- [10] Y. Shinada, Y. Ueda, K. Taki. “Porosity and Behavior of Gases in Aluminum Alloy Die Castings” 鋳物, 第 61 卷, 第 12 号, p. 920-925, 1989.
- [11] E. Tanaka, K. Kanazawa, K. Yano, R. Kawatani, J. Ogura. “Porosity Discrimination System for Aluminum Alloy Die Castings Using X-ray CT Imaging” 鋳造工学, 88 卷, 2 号, p. 85-91, 2016.
- [12] T. Suzuki, K. Fujishima, S. Takata, T. Osawa, N. Inoue, Y. Sakakibara. “Monitoring data analysis to identify process conditions of automated production equipment for producing non-defective products” 精密工学会春季大会学術講演会講演論文集, p. 893-894, 2017.
- [13] T. Tanaka, R. Kasahara. “Automated Visual Inspection Technology using Image Data” 日本画像学会誌, 第 55 卷, 第 3 号, p. 348-354, 2016.