

垂直 UT による欠陥検出を例にした 機械学習結果の解釈技術の調査

Study on an interpretation technique for machine learning results
with an example of defect detection by normal beam UT

発電技検

山本 敏弘

Toshihiro YAMAMOTO Member

A machine learning model is required to be interpretable for making its decision accepted by people. SHAP (Shapley Additive exPlanations) is a tool to show how much each of the features used for training a machine learning model contributes its decision. This paper shows an application of SHAP on a deep learning model for defect detection by normal beam UT as an example of nondestructive inspection problems.

Keywords: ultrasonic testing, UT, normal beam, AI, machine learning, deep learning, interpretation

1. はじめに

第三次人工知能 (AI) ブームの到来と言われて久しく、非破壊検査の分野でもこれまでに様々な AI の活用が試みられてきている。第二次 AI ブーム (1980 年代から 1990 年代前半) でも非破壊検査において AI を活用するための取り組みは行われており、超音波探傷試験 (UT) に関する例を挙げると、受信波形の解釈を自動化するためにエキスパートシステム、ニューラルネットワーク、パターン認識を利用する方法が考えられた[1-5]。これらのシステムが広く普及しなかった原因は、端的に言えば、実際の検査員の判断を模倣するには技術が不十分で、結果として多くの応用問題では満足な精度が得られなかったことである。第三次 AI ブームを代表する技術としてニューラルネットワークを高度化した深層学習 (ディープラーニング) があり、深層学習では大量の判断例を与えることでそこから自動的に判断の法則性を見つけ出し、その法則性が複雑な場合でもその判断を精度良く模倣するということが可能となった。このような AI 技術の進歩により AI の活用範囲が飛躍的に広がった一方で、深層学習に代表される AI 技術では判断過程がブラックボックス化され、判断の根拠が説明できないために社会的に受容され難いという問題が上がっている[6, 7]。経済産業省が設置した

スマート保安官民協議会の高圧ガス保安部会では「産業保安分野の AI は、高度な正確性と AI の判断の説明可能性を確保することが必要である」としている[8]。こういった社会的な要請を背景に、AI の説明 (可能) 性・解釈 (可能) 性に関する研究が進んでいる[9-11]。

本稿では、AI の一種である機械学習の解釈技術のうち、SHAP (Shapley Additive exPlanations) について取り上げ、垂直 UT の波形分析に利用した場合の結果とその解釈方法について調査した結果を紹介する。

2. 機械学習の解釈技術 SHAP

機械学習の解釈技術である SHAP はゲーム理論のシャープレイ (Shapley) 値の考えをもとにしたもので、シャープレイ値とは協力ゲームで得られた報酬をプレイヤーの貢献に応じて公正に分配するために考えられた指標である。SHAP では、これを流用して、与えられた多数の特徴量から機械学習システムがある判断を行うときに判断の決定に対する各特徴量の貢献度を数値化することで、どの特徴量が判断根拠になっているかを示すことができる[10-12]。

SHAP は、プログラミング言語 Python 用のパッケージとして実装されており、機械学習用のパッケージである scikit-learn や TensorFlow などと組み合わせて使用することができる[13]。本稿の例でもこれらを使用した。

3. 垂直 UT の波形分析

以下では、垂直 UT の波形分析に SHAP を利用した例を紹介する。

3.1 測定条件と測定結果

試験対象には JIS Z 2305 非破壊試験技術者資格試験の UT レベル 2 の実技試験対策のための平板試験体を用いた。材質は炭素鋼で、寸法は 120 mm×120 mm×25 mm であり、内部に大きさの異なる模擬欠陥が 2 個入っている。この試験体の UT 測定を行い、探触子には 5Z20N (5 MHz、径 20 mm の垂直探触子) を使用した。測定方法は、図 1 に示す 2 次元走査装置による自動探傷とした。探傷範囲は、試験体上面から外縁部を除いた 100 mm×100 mm の領域であり、0.5 mm ピッチで受信信号を記録した。

図 2 は、受信信号を記録したときの探触子の座標に対して各点の受信信号のある時刻の振幅の絶対値を色で表して描いた分布図である。色が白いほど高い値であることを意味する。これは、指定した時刻に対応する深さで反射した超音波の大きさを表す。つまり、白くなっている場所にはその深さに反射源 (欠陥) があるということである。これにより、図 2 で白く表れた場所のそれぞれ深さ 13 mm と深さ 17 mm に欠陥が入っていることが確認できた。

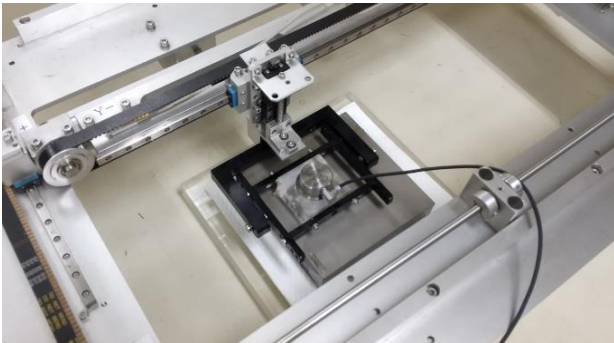


図 1 2次元走査装置

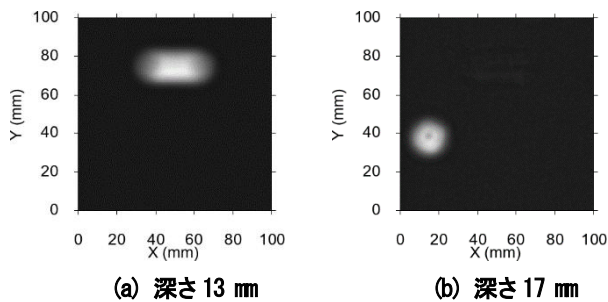


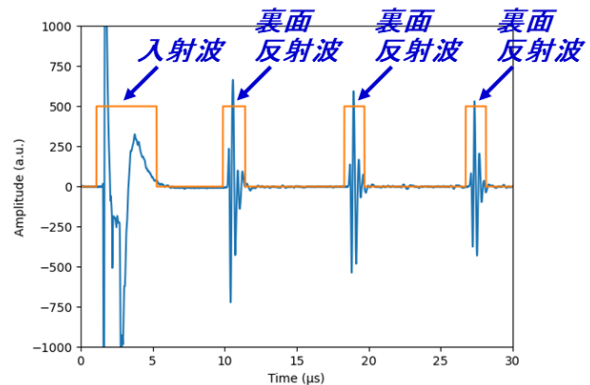
図 2 受信信号の振幅絶対値の分布

3.2 人間が与えたルールによる欠陥検出

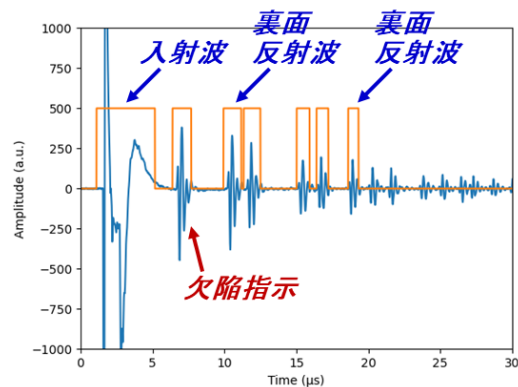
図 2 を見れば、この試験体のどこに欠陥が入っているかは明らかであり、この試験体の欠陥検出の自動化は、本来、機械学習/深層学習を使用しなくても然程難しくはない。本稿の趣旨は、あえて簡単な問題に対して深層学習を行い、その結果を分析することで深層学習の解釈技術の使用方法を示すことである。

深層学習について議論する前に、以下では、比較的単純な手順で示せるルールベースの自動化によりこの試験体の欠陥検出が自動化できることを確認する。

図 3 に欠陥がないときおよび欠陥があるときの受信波形の例を示す。欠陥がないときには、入射された超音波は試験体の裏面で反射し、その後、探触子で受信されるとともに探傷面でも反射して試験体の中で反射を繰り返す。これに対して、欠陥があるときには、超音波が裏面に到達する前に欠陥で反射し、それが探触子で受信される。このとき、欠陥で反射しなかった一部の超音波がそのまま裏面まで進んで反射すれば、図 3 (b) のように裏面反射波も同時に観測される。



(a) 欠陥がないとき



(b) 欠陥があるとき

図 3 受信波形の例

欠陥検出を自動化するために上述のような受信波形の差を判別する手順を考えた。まず、受信波形において各々の反射波は連なった波の塊である波束で表されるので、これらの波束を捉えることにする。ここでは、受信信号を絶対値で表し、一定区間ごとの移動平均を求めて、設定したしきい値を連続して超える部分を一つの波束として扱った。図3において振幅が500の四角の山がこの方法により波束として区分された範囲を表している。測定開始位置である座標(0, 0)には欠陥がないと仮定して、この位置で得られた受信信号を基準信号とし、基準信号の入射波以外の波束を裏面反射波のものとした。以後得られた受信信号において基準信号の波束の位置から指定した誤差以上離れた位置で波束が見つかり、その波束の最大値があるしきい値を超えていた場合、その点で欠陥があると判定した。図4は、以上のルールにより各測定座標での欠陥の有無を判定して2色に色分けしたものである。図2の振幅絶対値の分布と欠陥位置がよく一致しており、判定が妥当であると考えられる。

これによって得られた結果を深層学習に使用するデータとした。

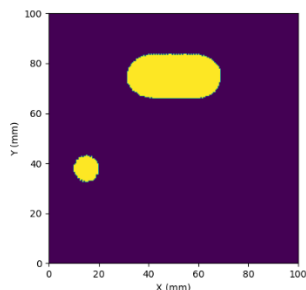


図4 ルールによる判定結果

3.3 深層学習による判定

ここでは、3.2で得られたデータを使って深層学習を行い、その推論結果を分析した例を紹介する。繰り返しになるが、この問題を自動判定するシステムは3.2で実現しており、本来は深層学習を使用する必要はないのであるが、簡単な例を使って深層学習の推論結果を分析する方法を確認する。

3.2で評価した箇所は40401点であり、このうち2689点で「欠陥あり」と判定された。今回は欠陥の有無の判定を目的としており、深層学習では欠陥の有無の割合が偏った訓練データを与えると、偏った推論が行われる傾向があるので、「欠陥なし」と判定された点から「欠陥あり」と同数の点を無作為に抽出して、「欠陥あり/なし」でそ

れぞれ2689点のデータを利用した。これらのデータから無作為に選んだ8割を訓練データとし、残りの2割は訓練済みモデルを評価するためのテストデータとした。

深層学習には、比較的一般的なものとして全結合型のニューラルネットワークを利用し、中間層を2層（各ノード数512）とし、中間層ごとにドロップアウトを20%とした。訓練は、ミニバッチ勾配降下法（バッチサイズ128）で行い、損失関数は交差エントロピー誤差とし、最適化手法はRMSPropとした。訓練データのうち2割は検証データとして訓練中の誤差評価に使用している。

プログラムを実行するたびにデータ選択が無作為に行われるので、訓練結果は毎回少し異なるのであるが、エポック数（訓練の繰り返し回数）を5とした場合（一般にはもっと大きな数にする）でも正解率が98%以上となった。

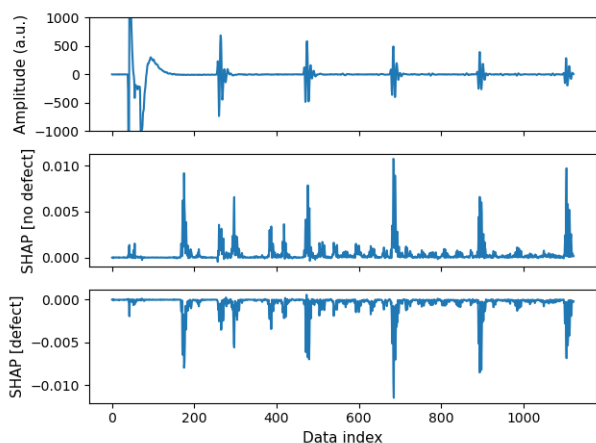
次に、機械学習の解釈技術であるSHAPの利用した結果について説明する。SHAPは、いくつかの機械学習方法に対応したものが開発されているが、ここでは深層学習用のものを使用した。機械学習による訓練済みモデルに入力データを与えて推論結果が得られるとき、SHAPは推論結果の決定に入力データの各特徴量がどれだけ寄与しているかを数値で示す。ここで扱っている問題で特徴量とは受信信号の時刻ごとの振幅値であるため、SHAPを使うことで訓練済みモデルがどの時刻の振幅に着目しているかを確認できる。

図5は、上記の深層学習による訓練済みモデルに入力データとして欠陥のないときの受信信号と欠陥のあるときの受信信号を与えたときのSHAP値をグラフにしたものである。一番上の段は入力データであり、横軸は入力データのデータインデックスであり、時間に対応する。中段は「欠陥なし」と推論する決定への寄与を表すSHAP値であり、下の段は「欠陥あり」と推論する決定への寄与を表すSHAP値である。負の値は、その推論への反対を意味する。これは2択の問題であるので、2種類のSHAP値はおおむね上下が逆になる。極自然な結果であるが、訓練済みモデルが受信波形の中の裏面反射波や欠陥反射波の波束に着目していることが分かる。ただし、早い時刻の波束ばかりでなく、何往復もして減衰した波束にも同じだけ着目していることが興味深い。

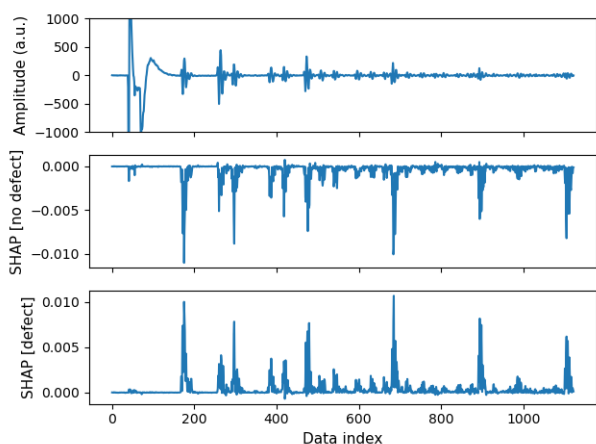
4. まとめ

本稿では、非破壊検査のための機械学習応用において

判断根拠の確認に SHAP を利用することを考え、垂直 UT の波形分析への適用例を紹介した。基本を押さえるために単純な問題を選んだので、得られた結果自体には意外性はないが、SHAP の機能が確かめられた。



(a) 欠陥がないとき



(b) 欠陥があるとき

図5 入力データに対する SHAP 値

参考文献

- [1] Ramesh Shankar, Ralph Williams and Michael Avioli Jr., “An expert system for power plant NDE,” Review of Progress in Quantitative Nondestructive Evaluation, Vol. 8A, 1989, pp. 665–672
- [2] 芳村敏夫, 森本和憲, 松浦洋司, 川人明美, “超音波探傷試験評価のためにファジィエキスパートシステム (欠陥エコーの波形判定におけるデータ処理の適用)”, 日本機械学会論文集 (C 編), Vol. 61, No. 582, 1995, pp. 724–729
- [3] 大石篤哉, 山田勝稔, 吉村忍, 矢川元基, “ニューラルネットワークと計算力学に基づく超音波欠陥同定 (複数欠陥の場合)”, Vol. 62, No. 598, 1996, pp. 1479–1487
- [4] 黒住保夫, 吉荒俊克, “超音波探傷波のパターン認識による欠陥識別”, INSS Journal, Vol. 1, 1994, pp. 191–205
- [5] 黒住保夫, 吉荒俊克, “超音波探傷波のパターン認識による炭素鋼試験片の欠陥識別”, INSS Journal, Vol. 2, 1995, pp. 127–136
- [6] 土木学会 SIP インフラ連携委員会, インフラ維持管理への AI 技術適用のための調査研究報告書, 土木学会 技術推進機構, 2019, p. 130
- [7] 橋本大樹, “AI による外観検査”, みずほ情報総研レポート, Vol. 21, 2021, pp. 65–74
- [8] スマート保安官民協議会 高圧ガス保安部会, 高圧ガス保安分野 スマート保安アクションプラン, スマート保安官民協議会 高圧ガス保安部会, 2020, p. 16
- [9] 原聡, “私のブックマーク : 機械学習における解釈性”, 人工知能, Vol. 33, No. 3, 2018, pp. 366–368
- [10] Neel Pradip Shah, Sheetal Jeshwani and Pavni Bhatt, “An introduction on interpretable machine learning,” International Journal of Innovative Technology and Exploring Engineering, Vol. 9, No.7S, 2020, pp. 107–111
- [11] Christoph Molnar, “Interpretable machine learning: A guide for making black box models explainable,” 2021/05/17 <https://christophm.github.io/interpretable-ml-book/> (和訳 : <https://hacarus.github.io/interpretable-ml-book-ja/>)
- [12] Scott M. Lundberg and Su-In Lee, “A unified approach to interpreting model predictions,” Advances in Neural Information Processing Systems 30 (NIPS 2017), 2017, pp. 4765–4774
- [13] Welcome to the SHAP documentation, 2018 (最終閲覧日 : 2021 年 5 月 18 日) <https://shap.readthedocs.io/en/latest/index.html>