

# インフラ維持管理業務での機械学習活用に向けたモデリング試行と結果の解釈に関する一考察

湧田 雄基<sup>1</sup>・山下 明美<sup>1</sup>・吉田 啓佑<sup>1</sup>・龍田 斉<sup>2</sup>・関 和彦<sup>3</sup>・  
有井 賢次<sup>4</sup>・熊谷兼太郎<sup>5</sup>・中畑 和之<sup>6</sup>・長沼 諭<sup>7</sup>

<sup>1</sup>正会員 北海道大学 数理・データサイエンス教育研究センター  
(〒060-0812 北海道札幌市北区北 12 条西 7) E-mail: wakuda@mdsc.hokudai.ac.jp (Corresponding Author)

<sup>2</sup>正会員 大日本コンサルタント株式会社 インフラ技術研究所  
(〒330-6011 埼玉県さいたま市中央区新都心 11-2L.A.タワー11 階) E-mail: tatsuta@ne-con.co.jp

<sup>3</sup>正会員 アイセイ株式会社 技術開発部  
(〒116-0013 東京都荒川区西日暮里 2-40-3) E-mail: seki-k@eyesay.co.jp

<sup>4</sup>正会員 株式会社長大 構造事業本部 技術統括部  
(〒730-0017 広島県広島市中区鉄砲町 7-18) E-mail: arii-k@chodai.co.jp

<sup>5</sup>正会員 湘南工科大学 工学部コンピュータ応用学科  
(〒251-8511 神奈川県藤沢市辻堂西海岸 1-1-25) E-mail: kumagai@sc.shonan-it.ac.jp

<sup>6</sup>正会員 愛媛大学 大学院理工学研究科生産環境工学専攻 環境建設工学コース  
(〒790-8577 愛媛県松山市文京町 3) E-mail: nakahata@cee.chime-u.ac.jp

<sup>7</sup>正会員 株式会社鴻池組 本社 土木事業総轄本部 技術本部 技術企画部 企画課  
(〒541-0057 大阪市中央区北久宝寺町 3-6-1) E-mail: naganuma\_ss@konoike.co.jp

本論文では、インフラマネジメント分野における AI (Artificial Intelligence) の活用を目的として、分析性能とモデルの解釈性に着目し、AI 活用の可能性についての考察を行う。特に、近年、機械学習のコンペティション等で好成績を上げているアンサンブル型学習手法を中心に、XGBoost, LightGBM, CatBoost, Random Forest, 決定木分析について、その数理的背景の概要を述べる。これらの手法により橋梁の劣化の推定を試行した結果について報告する。また、この結果について、個々の手法の特性をふまえ、AI のインフラマネジメント業務における活用の視点より考察を行った結果について報告する。

**Key Words:** インフラマネジメント, データサイエンス, AI, Machine Learning

## 1. はじめに

本稿では、近年、実用化が進んでいる機械学習について、得られるモデルの分析性能に加え、モデルの特性理解や考察を行うためのモデルの解釈性について考察を行う。特に、機械学習の数理的背景に基づき、手法毎の特徴について解説を行う。その上で、実際のデータ分析結果について、活用の視点より考察を行う。本研究は、土木学会 構造 AI 小委員会のワーキング・グループとして実施した結果について報告を行うものである。

この活動の一環として、インフラ維持管理分野において機械学習を活用した事例を中心に、直近3年間を対象に研究動向調査を実施した。近年の深層学習を用いた研究の興隆については多数報告されているとおりで<sup>1)</sup>。他方で、近年のデータ分析コンペティションで好成績を示すアンサンブル学習を用いた機械学習モデルの利用が

確認できた。例えば、Baruaら<sup>2)</sup>は、空港滑走路及び誘導路の舗装劣化予測を Gradient Boosting Machine Learning (GBM) を用いて行っているほか、Liら<sup>3)</sup>は、モンテカルロ法を取り入れた改良型勾配ブースティング決定木モデルによって、大型コンテナ船のラッシングブリッジの劣化予測を行っている。さらに、Chunら<sup>4)</sup>は、コンクリート構造物のクラック検出課題に対して、画像検出の後段に LightGBM を用いた2段階の推定アーキテクチャを提案している。このように、アンサンブル型手法を他の手法と組み合わせ、精度向上を狙った研究も発表されている<sup>5,6)</sup>。また、構築したモデルの解釈を行う事例についても取り組まれている<sup>7)</sup>。例えば、龍田ら<sup>8)</sup>は、橋梁管理データを対象として、XGBoost, LightGBM, CatBoostを用いた分析を行い、この結果について、SHAP (SHapley Additive exPlanations) を用いることで、変数毎の貢献度を算出し、この可視化について報告している。

以上の結果をふまえ、本研究では、分析手法としては、アンサンブル型機械学習手法（XGBoost<sup>9)</sup>、LightGBM<sup>10)</sup>、Catboost<sup>11)</sup>）に着目することとした。また、機械学習により構築したモデルの解釈手法としては、Feature Importance、Shapley 値に着目した。

## 2. データサイエンス試行の設計

### (1) データサイエンスのプロセス

本研究におけるデータサイエンスのプロセスは、東京大学「情報技術によるインフラ高度化」社会連携講座（第1期 2009～2013、第2期 2014～2018）における検討<sup>12)</sup>や著者らの既往研究で用いたインフラマネジメント分野におけるデータサイエンスプロセス<sup>13)14)</sup>に基づいた検討とした。同手法は、CRISP-DM<sup>15)</sup>を基に、インフラマネジメント分野のデータサイエンスに適合するようカスタマイズを施したものである。このプロセスのうち、業務理解、データ理解、データ整備、分析、評価に関するデータサイエンスの設計について、以降に述べる。

### (2) 対象データ

今回は、利用可能なデータとして、国土交通省が提供する「道路メンテナンス年報」<sup>16)</sup>を取り上げた。このデータでは、入手可能な表形式データにおいて、構造物健全性の判定区分Ⅰ～Ⅳの記載がある。構造物としては、橋梁、トンネル、シェッド、大型カルバート、横断歩道橋、門型標識等の6種類について提供されている。本検討では、橋梁を対象とした。

取得可能なデータには、12の変数（管理区分、点検年度、橋梁名、橋梁名(フリガナ)、路線名、架設年次(西暦)、橋長(m)、幅員(m)、管理者、行政区域:都道府県名、行政区域:市町村名、点検記録:判定区分)が含まれる。ファイルは、管理区分毎に、国交省管理、地公体管理、高速道路の3種類に分かれている。また、点検年次に応じて、2014年度～2019年度まで6カ年のデータに分かれている。点検1巡目のデータとして、2014年度～2018年度の5カ年があり、点検2巡目のデータとして2019年度がある。これらを総合し、18ファイル（約60.4Mbyte）を分析対象とした。

データ分析において対象とする業務については、点検計画の立案業務や補修計画立案を想定する。これらの計画において、5年後の次回点検時の判定区分を予測することができれば、よりリスクの低い補修計画立案が可能となるとし、仮説的課題設定を行った。今回のデータを用いて、1巡目の状態から2巡目の状態への判定区分

の進行を予測するモデルを構築することを検討した。

そこで、教師あり学習の目的変数としては、2巡目の点検において構造物の健全性判定区分がⅢ以上に進行したか否かを示す2値データを作成し、これを目的変数とした。具体的には、判定区分がⅢまたはⅣの場合を1（Class1）、ⅠまたはⅡの場合を0（Class0）とした。

2014年から2019年までの6年間のデータを統合すると、842,885件が利用可能であった。このうち、1巡目および2巡目のデータについて、結合に用いるマッチングキー（橋梁名、路線名、市町村名）と、結合の検証に用いる橋長・幅員、さらに、分析に用いる架設年次（西暦）のいずれかが空欄のデータレコードを除去すると、それぞれ、491,059件および86,231件となった。このデータ調製においては、架設年次が不明のレコードを除外した。また、1巡目のデータ(491,059件)と2巡目のデータ(86,231件)とを対応付けるデータ調製について、前述のマッチングキーが完全に一致するレコード同士を対応させる処理を行った。その結果、マッチングキーが完全に一致した49,339件のデータに対し、さらに、幅員または橋長が5%以上相違するデータレコードについては、今回の対象から除外する処理を適用した。ここで、5%という値は、統計における有効水準で多く用いられる値を試行的に設定したものである。この処理の結果、幅員と橋長について、比による誤差がいずれも5%未満となるデータを抽出すると、全データの約67%にあたる33,140件のデータセットを得た。なお、誤差が1%未満のデータは36.8%（18,137件）であり、誤差無く完全に一致するデータは、29.5%（14,536件）であった。

### (3) 試行設計

データサイエンスにおける試行設計については、表-1の条件とした。この設計においては、対象構造物や業務については、仮説的なものとなっている。

表-1 データ分析試行の設計<sup>12)</sup>

構成要素	設計内容
対象構造物	橋梁
対象業務	橋梁健全度判定業務
目標	<ul style="list-style-type: none"> <li>機械学習の性能評価により有用性を確認</li> <li>モデル解釈の有用性を確認</li> </ul>
目的変数	1巡目から2巡目への判定区分の進行有無
説明変数	9変数（データ調製結果による）
分析手法	決定木分析, Random Forest, XGBoost, LightGBM, Catboost
評価	ROC 曲線, P-R 曲線, 利用変数の影響度調査

#### (4) 分析と評価の流れ

本研究における分析試行は、事前学習と評価により構成する(図-1)。事前学習用のデータと評価用のデータを分割した交差検定(CV: Cross Validation)を行い、評価を実施する。交差検定においては、n-fold (n=5)を用いた。対象データ(33,140件)を分割することで、22,203件を学習に用い、10,937件を評価に用いた。

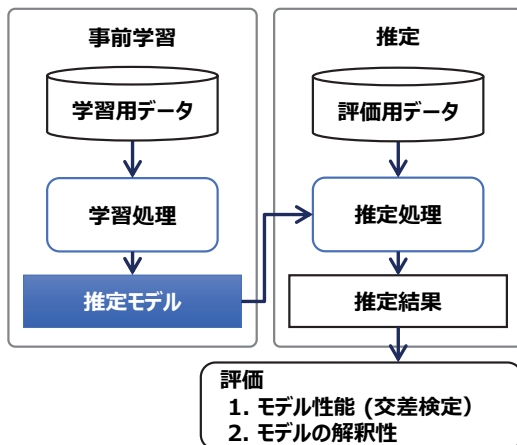


図-1 本研究における分析と評価の流れ

### 3. 手法

#### (1) 分析手法

本研究で扱うデータ分析手法について述べる。

本研究では、機械学習手法として、コンペティション等で多く用いられるアンサンブル型学習手法に着目する。具体的には、「XGBoost」、「LightGBM」、「CatBoost」を挙げる。また、これらの手法に対する対比のため、「決定木分析」および「ランダムフォレスト」も対象に含める。これらの手法の数理的理解に基づいた特徴付けを行う。本研究における目的変数は常に2値のカテゴリ変数(0,1と書く)である。このため、以下の特徴づけにおいても、目的変数が2値のカテゴリ変数である場合についてのみ述べる。

##### a) 決定木分析(分類)

単一の木構造を用いた教師あり機械学習による分析の総称である。本研究で用いた他の手法(「ランダムフォレスト」、「XGBoost」、「LightGBM」、「CatBoost」)と比較して、得られるモデルの精度では劣る傾向がある。しかしながら、モデルの木構造の可視化の点において他手法に勝る。

本研究では、scikit-learnに標準的に実装されているDecision Treeの一種であるCART(Classification and

Regression Trees)を用いた。CARTは、各分岐点でデータを二分割する二分木により分析を行う手法である。各分岐の最適化はgini係数により行った。

##### b) Random Forest

アンサンブル型学習手法の一種である。元の訓練データからランダムにいくつかのデータを重複を許して抽出し、新しい訓練データを構築する。その後、構築された新しい訓練データを元に決定木分析をおこなう。上記の操作を並列に行うことで複数のモデル(弱学習器)を得る。目的変数の推定の際には、複数の弱学習器から得られる出力値に基づいた多数決を行う。より多くの弱学習器が出力した推定結果を最終的な出力とする。このようにすることで、単一のモデルよりも、過学習の危険性を軽減することを目的とした手法である。

##### c) XGBoost

以降の手法もランダムフォレストと同様にアンサンブル型学習手法である。ただし、弱学習器の構成(決定法・集約法)は大きく異なる。ランダムフォレストにおいては、過学習の抑制が目的であったが、以降の手法の主な目的は予測精度の向上である。予測精度の向上を目指した結果、決定木分析やランダムフォレストよりも、過学習が生じやすい。

本手法では、目的変数が1である確率を勾配ブースティング法により算出する。XGBoostでは弱学習器にCARTに基づく回帰木を用いる。しかし、回帰木の構造において、葉の数や深さ、分岐構造の自由度の高さにより、最適な回帰木を決定することは、実装上、実現不可能である。回帰木の決定における葉の分割については、情報利得の最大化を基準とする。葉の分割を検討する順番はlevel-wise tree growthに則り、同じ深さの枝については、例えば左側から順に枝を成長させ、同じ深さの枝が成長しきった後に1つ深い枝の分割に移るといった手順となる。

##### d) LightGBM

LightGBMは、学習アルゴリズムにおいて、XGBoostと類似する手法である。目的変数が1である確率を、回帰木(CART)を用いた勾配ブースティング法で推定する点では、XGBoostと変わらない。XGBoostとの主な違いは各回帰木の構成方法にある。

一点目の違いは各特徴量をヒストグラム化している点である。ヒストグラム化により厳密な枝分かれを採らずに済み、計算コストを抑えることが可能である。

二点目の違いは「level-wise tree growth」ではなく「leaf-wise tree growth」により、回帰木を構成している点である。両者の違いは枝を伸ばす順番にある。なお、木を完全に

成長させた場合には同じ木が構成される。この leaf-wise tree growth は、情報利得の最も大きくなる枝から順に伸ばす構成方法である。leaf-wise tree growth を用いる場合は、「pre-pruning」と呼ばれる剪定アルゴリズムを適用することができる。pre-pruning とはさらなる分岐を行うことで予測誤差が下がるなら分岐を行い、そうではないのならそこで木の成長を止めるアルゴリズムである。pre-pruning は post-pruning に比較して精度が落ちる可能性があるものの、計算コストが小さい。なお、post-pruning とは決定木を完全に成長させてから、最も予測誤差の小さい状態を選択する剪定方法である。

ヒストグラム化、leaf-wise tree growth 及び pre-pruning により、lightGBM の計算コストは XGBoost よりも「軽くなる傾向がある。ただし、leaf-wise tree growth による決定木は複雑になりやすく、過学習には注意が必要である。

#### e) CatBoost

CatBoost も回帰木と勾配ブースティング法を組み合わせた手法である。「各回帰木が obvious decision tree である」という制約が課されている点が XGBoost, LightGBM との違いのひとつである。ここで、深さ  $d$  の分岐点あるいは葉が  $2^d$  個であり、同じ深さの分岐点の分岐条件が等しい決定木は obvious decision tree と呼ばれ、XGBoost や LightGBM において構成される決定木に比べ、単純な構造を持つため過学習に陥りにくい。また、各葉の重みはロジスティック関数により算出されている (CatBoost の弱学習器は各説明変数に対する回帰係数を推定する)。

また、CatBoost は、特徴量にカテゴリ変数が含まれる場合に有効とされる手法である。その理由は次のように、特徴量を差し替える点にある。訓練データ  $\{(x_i, y_i)\}_{i \in 1, \dots, n}$  ( $x_i \in R^m, y_i \in \{0, 1\}$ ) に対し、 $x_i = (x_{i,1}, \dots, x_{i,m})$  と書くことにする。ここで、特徴量にはカテゴリ変数と連続変数が混在していても構わない。各弱学習器の学習時に、その都度ランダムに選ぶ  $m$  次の置換を  $\sigma$  とし、新たな特徴量、 $\hat{x}_{\sigma(i)} = (\hat{x}_{\sigma(i),1}, \dots, \hat{x}_{\sigma(i),m})$  を

$$\hat{x}_{\sigma(i),k} = \frac{\sum_{j=1}^{i-1} [x_{\sigma(j),k} = x_{\sigma(i),k}] y_{\sigma(j)} + aP}{\sum_{j=1}^{i-1} [x_{\sigma(j),k} = x_{\sigma(i),k}] + a}$$

により定める。ここで、 $[\cdot]$  は  $[\cdot]$  内が True のときに 1 を返し、False のときに 0 を返す関数である。また、 $P$  は訓練データの目的変数の平均値 (本研究においては目的変数が 1 であるデータの割合) であり、 $a$  は  $P$  に重みをつけるパラメータである。 $x_{\sigma(i),k}$  のかわりに  $\hat{x}_{\sigma(i),k}$  を用いて弱学習器の学習を行う。このような特徴量の差し替えは、 $\sigma$  のランダム性によって過学習に陥るリスクを低減しつつ、カテゴリ変数と連続変数両方を区別することなく新

たに連続変数化 (数量化) している。

#### f) 手法毎の特徴についてのまとめ

XGBoost, LightGBM, CatBoost はどれも回帰木と勾配ブースティング法を組み合わせた手法であるという点で類似の手法である。しかしながら、一つ一つの回帰木の構成方法や勾配ブースティング法の進め方に違いがある。この違いによって、LightGBM の計算コストは XGBoost よりも軽くなる傾向がある。また、勾配ブースティング法の進め方の違いが CatBoost における汎化性能の向上を促す。ランダムフォレストは、勾配ブーストを用いた 3 手法に比べ、過学習の抑制に重きを置いた手法である。アンサンブル型学習手法という点では 3 手法と同様だが、弱学習器の構成法・集約法が大きく異なる。決定木分析は、単一のモデルを用いているため、過学習や予測精度の観点において、他の手法に比べ劣ることが多い。しかしながら、他の手法にはない解釈性能を有している。

表-2 に、分析手法毎の特徴をまとめて記載する。

表-2 今回試行した分析手法の数理的特徴付け

分析手法	特長	弱学習器の構成	過学習*
決定木分析	可視化による解釈性に期待	単体の回帰木	生じにくい
Random Forest	最適化が容易	バギング型	生じにくい
XGBoost	性能に期待できる		生じやすい
LightGBM	学習が比較的軽量	ブースティング型	生じやすい
CatBoost	特に、カテゴリデータを用いる分析において期待できる		生じやすい

\*ハイパーパラメータを適切に設定することが必要

### (2) モデルの解釈手法

本研究で扱うデータ分析手法について述べる。

機械学習により得られるモデルの解釈手法について、特に、Feature Importance と Shapley 値に着目し、ここでは、その仕組みについて確認する。

#### a) Feature Importance

Feature importance (以降、FI と表記) は特徴量ごとに選出される 0 以上の値である。値が大きいほど重要な特徴量であることが示唆される。FI は手法毎に算出方法が異なる場合がある。そのため、異なる手法において算出された FI の値を比較する際には注意が必要である。一方で、FI を値が大きいものから順に並べた場合に、その順番は考察の対象たりえる。従って、本研究では、モデルの解釈として、変数毎の FI を確認する手法について取り扱うこととした。以下では特徴量  $i$  に対して、 $i$  を分岐

条件に使った分岐点全体の集合を $E(i)$ とする。

XGBoost, Random Forest, 決定木分析の FI: デフォルトに設定されている gain を用いた。これは、 $E(i)$ に属する各分岐点ごとに情報利得を計算し、それらを足し合わせた値である。XGBoost, Random Forest では各弱学習器における gain の平均が算出されている。総和が 1 になるようにリスケールされている点には注意が必要である。

LightGBM の FI: デフォルトに設定されている, split を用いた。  $E(i)$ に属する分岐点の数を「split」と呼ぶ。

CatBoost の FI: デフォルトに設定されている Prediction Values Change を用いた。  $E(i)$ に属する各分岐点ごとに、分岐を行った場合と分岐を行わなかった場合の回帰係数の変化 (CatBoost の各葉の重みはロジスティック関数により推定されている) の二乗を計算し、各葉に属するサンプル数を考慮しつつそれらを足し合わせたものを  $PVG(i)$  と書く。  $PVG(i)$  の総和が 100 になるようにリスケールしたものを「Prediction Values Change」と呼ぶ。このリスケールのため、CatBoost によって構成されたモデル同士であっても feature importance の比較に意味はない。

#### b) SHAP (SHapley Additive exPlanations) <sup>17)</sup>

SHAP とはゲーム理論の Shapley 値を機械学習に応用したライブラリである。 Shapley 値によって、FI とは別のモデルの解釈が行える。 Shapley 値は次のように算出される。

訓練データ  $x$  における特徴量  $i$  の Shapley 値  $\phi_i(x)$  を算出する。  $F$  を特徴量全体の集合とし、  $S$  を  $i$  を含まない  $F$  の部分集合とする。  $S \cup \{i\}$  に含まれる特徴量のみから構成したモデルを  $f_{S,i}$  とし、  $S$  に含まれる特徴量のみから構成したモデルを  $f_S$  とする。

$$\phi_i(x) = \sum_{S \subset F \setminus \{i\}} \frac{|S|!(|F| - |S| - \{i\})!}{|F|!} (f_{S,i}(x) - f_S(x))$$

により、  $\phi_i(x)$  を定める。  $|F|$  が大きい場合は、  $i$  を含まない  $F$  の部分集合の数が非常に大きくなるため、近似計算で  $\phi_i(x)$  を求める (近似方法は手法毎に異なる)。 訓練データ  $x$  における特徴量  $i$  の Shapley 値は、  $x$  におけるモデルの 1 次近似の  $i$  に対する回帰係数に一致する <sup>17)</sup>。 このことから Shapley 値を求めることで、線形回帰モデル以外のモデルであっても近似的に回帰係数を求めることができ、モデルの解釈が可能になる。 本研究において Shapley 値が正であるとき、対応する特徴量 (変数の値) が増えるほど、出力が 1 になる確率が高まると考えられる。 逆に Shapley 値が負の場合は、対応する特徴量が大き

くなるほど、出力が 0 になる確率が高まると考えられる。 また、 Shapley 値の絶対値が大きい特徴量は重要な特徴量だと考えられる。 すなわち、 Shapley 値の絶対値の総和を変数の重要度として FI と同様の指標としてとらえることもできる。

## 4. 分析結果

### (1) 分析結果の確認手法

本研究では、分析結果の確認について、性能と解釈性の 2 つの視点により進める。 また、これらの視点を議論するにあたっては、AI を実務で活用することを想定した視座により、性能と解釈性についての考察を行った。

### (2) 性能に関する分析結果

分析結果の性能評価については、ROC 曲線 (Receiver Operating Characteristic curve) の描画と AUROC (Area Under the ROC curve) により実施した。 結果について、図-2 に示す。 ROC 曲線は、真陽性 (TP: True Positives) の率 TPR と偽陽性 (FP: False Negatives) の率 FPR の 2 指標により描画される曲線である。 関心のある事象 (正例) の推定結果の正誤に関し、構築したモデルの性能を評価可能である。

また、P-R 曲線 (Precision-Recall curve) を描画した結果について、図-3 に示す。 P-R 曲線は、Precision (精度) と Recall (検出率) により描画される曲線である。 応用時に、精度や検出率を重視してモデル性能を選択する場合には、P-R 曲線の利用が有効である。

### (3) モデルの解釈に関する算出結果

モデルの解釈については、FI と Shapley 値により行った。 FI については、表-3 に手法毎の比較を示す。 表では、5 種類の手法毎に、分析に用いた変数の FI 値を示した。 また、Shapley 値については、図-4 に手法毎の比較を示す。 図では、5 種類の手法毎に個別のグラフに分けて結果を示している。 個別のグラフでは、縦軸に分析に用いた変数、横軸に Shapley 値をとり、データレコード毎の Shapley 値を一つの点としてプロットしている。 また、点の色は、変数の値の高低を示している。

## 5. 分析結果の考察

### (1) 分析性能の視点による考察

#### a) AUROC による評価

分析性能については、AUROC 値で比較すると、

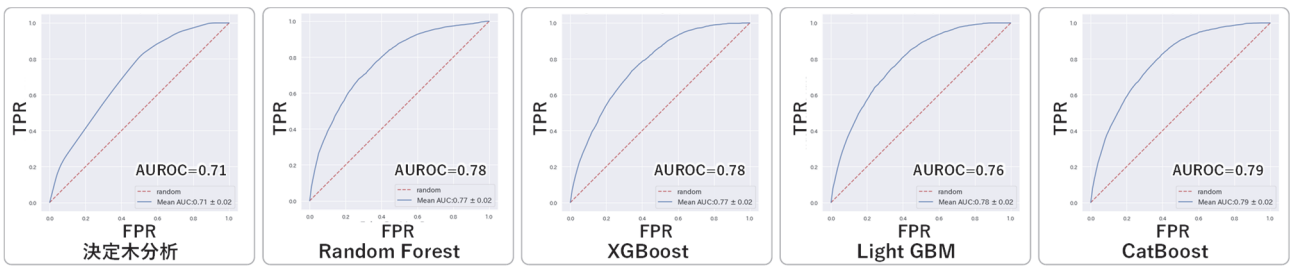


図-2 ROC 曲線を用いた手法毎の性能比較

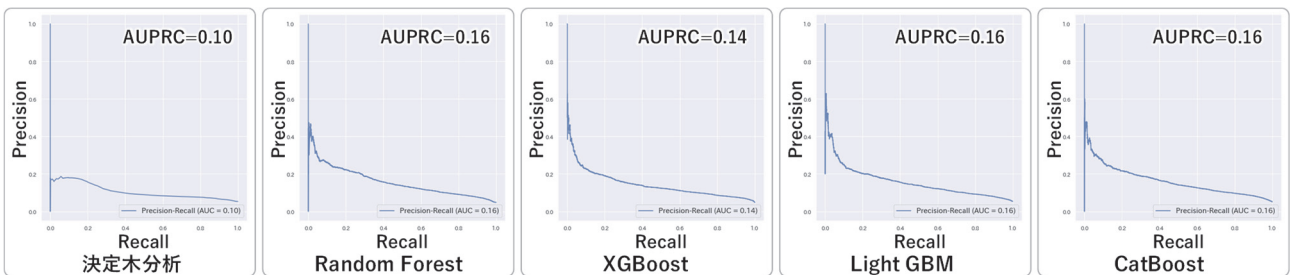


図-3 P-R 曲線を用いた手法毎の性能比較

表-3 Feature Importance (FI)算出結果表

変数	Feature Importance Value				
	決定木分析	Random Forest	XGBoost	Light GBM	CatBoost
橋長(m)	0.244	0.280	0.079	932	18,348
幅員(m)	0.000	0.113	0.073	883	14,283
管理区分:国土交通省	0.000	0.037	0.111	28	1,739
管理区分:高速道路	0.000	0.012	0.110	11	2,429
管理区分:地方公共団体	0.000	0.017	0.072	42	2,053
1巡目点検時点の判定区分値	0.577	0.273	0.402	118	33,424
1巡目点検時点の架設経過年	0.179	0.236	0.081	841	22,547
1巡目から2巡目点検への経過年	0.000	0.033	0.071	145	5,176

CatBoost (AUROC=0.79) が最良の結果となった。また、アンサンブル型手法の中では、LightGBM (AUROC=0.76) が最も低い値を示した。このことは、LightGBM が、学習時間の短縮を主眼とした手法であることから納得のいく結果である。また、今回試行した手法の中では、決定木分析 (AUROC=0.71) が最も低い結果となった。

アンサンブル型手法のうち、ブースティングを用いた XGBoost や LightGBM の性能が良好であるが、他方で、バギングによる Random Forest (AUROC=0.78) も好成績を上げている点が注目になる。Random Forest は、過学習に対して抑制的な動作をする手法であることに加え、ハイパーパラメータが少なく利用が簡単である点を考慮すると、Random Forest のメリットが目立つ結果となった。

以上の議論を通じて、AUROC は、モデルの全体的な

性能について評価が可能な指標であり、手法毎の性能比較には有用であると考察した。

一方で、今回構築したモデルについては、判別の閾値を調整することによって、業務課題に対して柔軟な対応が可能である。従って、具体的な業務課題として、例えば正答率 (Precision) や検出率 (Recall) の指定による要件が存在する場合において、これらに応じた評価を行うためには、AUROC のみでは、議論が不足することも認識しておく必要があると言える。

#### b) P-R 曲線による評価

前述のとおり、AUROC では、精度についての直接の評価ができない。従って、本研究では、Precision と Recall の関係により評価が可能な P-R 曲線を描画し、評価を試行した。

P-R 曲線の AUC である AUPRC (Area Under the P-R Curve) で比較すると、Random Forest、Light GBM、CatBoost (いずれも AUPRC=0.16) が最良の結果となった。これは、AUROC とは異なる結果となった。また、決定木分析が、今回試行した結果の中では、最も低い値 (AUPRC=0.10) となった。

以上で述べたとおり、P-R 曲線の確認によって、Precision に着目したモデル性能評価が可能であることが確認できる。なお、P-R 曲線の Recall 値 (本稿における図-3の横軸) は、ROC 曲線の TPR 値 (本稿における図-2の縦軸) と、定義上同一である。しかし、Precision に

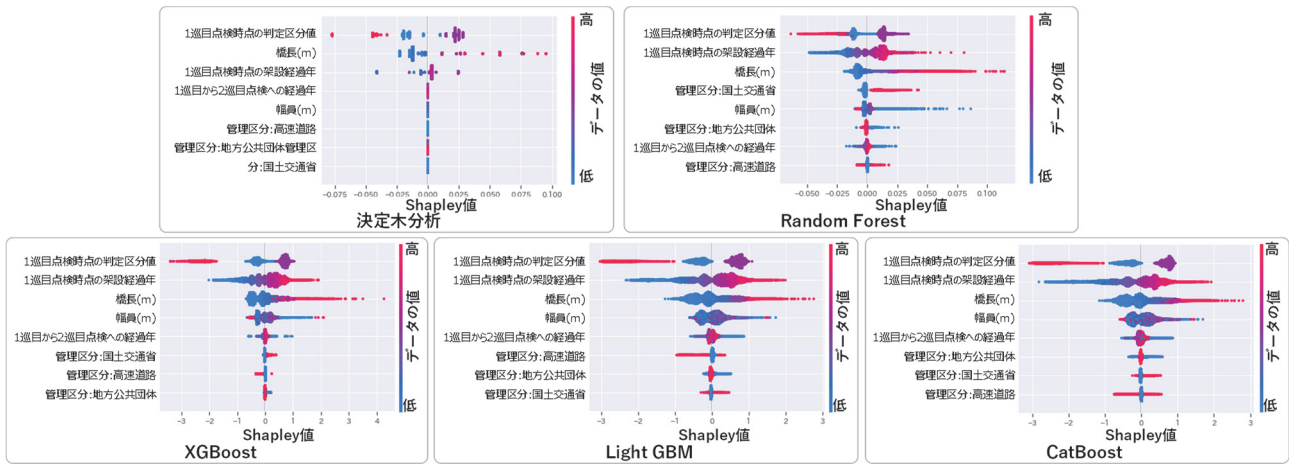


図4 手法毎の Shapley 値可視化結果の比較

については、ROC では直接に評価できない。従って、P-R 曲線で確認することも重要となる。今回の試行結果においては、例えば、Random Forest を用いた結果については、図-2 を参照した場合には、グラフは綺麗な弧を描いている。一般に、ROC 曲線では、曲線の膨らみが最大（即ち、対角線を示す破線への垂線の足の長さが最大）となる点を良好なモデルとして抽出する場合も多い。この考え方に基づけば、今回の例では  $TPR (Recall) = 0.68$ ,  $FPR = 0.27$  の点を選定することとなる。

他方で、図-3 の P-R 曲線を参照し、最良の条件を探すことを考える。P-R 曲線では、Precision と Recall 値の調和平均により求められる F 値 (F-measure) を用いて最良条件を議論する場合も多い。この方式に基づけば、Precision=0.16, Recall=0.39 の点を選定することとなる。この結果を図で示すと、図-5 のとおりとなる。

さらに、図-3 の P-R 曲線では、Recall=0.3 付近において、特に、Precision 方向の若干の凹凸が見られるが、図-2 の  $TPR (Recall) = 0.3$  付近を目視する限りでは、性能の凹凸は確認できない。今回の事例では、やや凹凸は少ないものの、一般に、P-R 曲線の Precision 方向の性能は、不安定な曲線を描く場合も多い。従って、ROC 曲線のみを確認していると、Precision の性能変化を見逃す可能性がある。

以上より、ROC 曲線と P-R 曲線は、利用の目的に応じて、適切に選択することが重要である。ROC 曲線は、AUC を参照することで、分析の全般的な性能を評価することが可能である。その一方で、とりわけ、インフラ維持管理分野においては、関心のある事象の正答率を示す「Precision」や、検出すべき事象の検出率を示す「Recall」が重要である場合も多いため、P-R 曲線を確認することが重要な場合も多いと考察する。

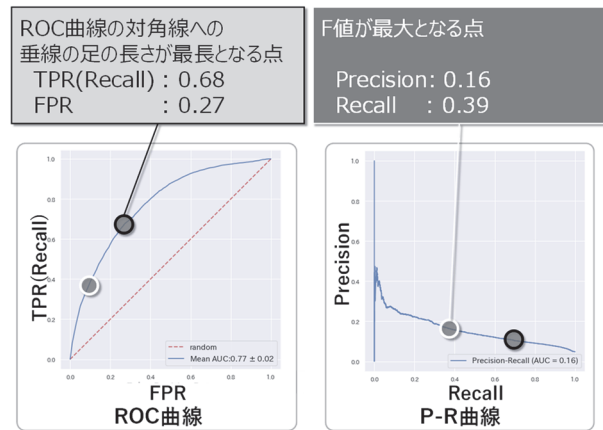


図5 ROC 曲線と P-R 曲線による評価の対比

## (2) 分析結果の解釈性の視点による考察

### a) Feature Importance による解釈

手法毎に FI 値が上位の変数を確認すると、第 1 位の変数については、決定木分析と XGBoost と CatBoost では「1 巡目点検時点の判定区分値」であり、Random Forest と LightGBM では「橋長(m)」となっている。

また、XGBoost については、「1 巡目点検時点の判定区分値」が特に高い値として目立つものであり、その他の変数の FI 値と大きな差があると考察できる。LightGBM では、「橋長(m)」、「幅員(m)」、「1 巡目点検時点の架設経過年」が特に高い値として目立つ変数であり、その他の変数の FI 値と大きな差が見られる。

以上のように、FI 値による変数の順序は、手法毎に大きく異なる結果を示している。この要因は、本稿で扱った 5 つの手法毎に FI の算出アルゴリズムが異なることが挙げられる。

従って、手法を跨いだ FI 値の差異についての議論は行わず、ここでは、Random Forest を例にとり考察を行う。

本稿で用いた FI の算出手法においては、gini 係数に基づいて、決定木を構成する際に、対象変数が不確実性や分散を低減させるためにどの程度貢献するかを評価することで計算している。この手法は、計算が高速である一方で、判別に対する重要性を正確に把握する結果になるには限らない。特に、値が連続的で Cardinality が高い（即ち、値のユニーク数が多い）変数について FI が高くなる傾向があることが知られている<sup>18)</sup>。今回の試行でも、Random Forest の FI については、「橋長」や「1 巡目点検時点の判定区分値」に次いで、「1 巡目点検時点の架設経過年」や「幅員」が高順位となっている。また、2 値で構成される管理区分に関する変数については、いずれも低順位となった。この結果は、前述の数理的背景を考慮した FI 値の算出傾向に当てはまるものであると考察できる。FI の利用においては、このような特性を考慮した上で土木や現場知見との照合が必要と考えられる。

#### b) SHAP による解釈

SHAP による結果について、影響変数の順序について確認すると、第 1 位の変数は、5 手法とも共通して「1 巡目点検時点の判定区分値」である。1 巡目点検時の構造物の状態に基づき、2 巡目点検時の状態が定まるということは、納得のいく結果であり、Shapley 値による解釈の有効性が示唆される。第 2 位の変数は、決定木分析を除き、「1 巡目点検時点の架設経過年」であった。第 1 位～第 3 位までの変数は、全ての手法で共有の変数について、順序が入れ替わっただけの結果であり、共通点の見いだせる結果となっている。

続いて、第 4 位の変数については、決定木分析は「1 巡目から 2 巡目点検への経過年」であり、Random Forest は、「管理区分:国土交通省」であった。その他の手法は、全て「幅員(m)」であった。また、第 5 位の変数については、ブースティング型手法は全て「1 巡目から 2 巡目点検への経過年」であるが、第 6 位から第 8 位までの変数については、全て異なる結果を示した。

Shapley 値によって、今回の試行のように少ない変数を用いた場合においても、構築されたモデルにおいて、判別に貢献している変数の順序が異なることが確認できた。特に、弱学習器の構成が異なる決定木分析や Random Forest について、変数の貢献順序が、他の手法に比べて特異的な結果となっている点について、納得のいく結果と言えよう。このように、Shapley 値を確認し、モデルに対する解釈を行い、外部知見と照らした考察を進めることが可能であると考察する。

#### c) Feature Importance と Shapley 値の対比

今回の結果より、FI と Shapley 値の結果を対比すれば、

「連続値を持つ量的変数の方が、2 値変数よりも高い値が算出される」といった大まかな共通点が見られた。しかし、変数の順序を議論する用途としては、FI による結果では手法毎に順序が大きく異なっており、Shapley 値の方が手法間のバラツキが少ない結果となっていた。特に、ブースティング系の 3 手法（図-4 下段）においては、Shapley 値では上位 5 変数まで一致していた。以上のことより、判別への貢献変数に関し、複数の手法において比較し、考察する用途としては、値の算出アルゴリズムの統一が容易な Shapley 値を用いた比較が望ましいと考えられる。

### (3) その他の考察

#### a) 対象データの分割

今回実施した性能評価では、手法毎の大きな差異は見られず、類似した傾向となった。一方で、性能評価について、手法毎の違いを考慮した上で性能順序については、納得感のある結果であり、ROC 曲線や P-R 曲線の AUC を用いた評価は、有効であることが確認できた。この手法について、今後は、例えば、管理者毎にデータを分割し、比較するなどの踏み込んだ分析も有効である可能性がある。

#### b) 説明変数の拡充

説明変数による判定への影響が可能であることが確認できたことより、今後、利用する説明変数については、判定区分に対する影響が考えられる変数を追加することも検討の余地があると考察した。例えば、地域性への影響が考えられる緯度・経度や、橋梁の疲労に影響が考えられる重交通、塩害や凍害との相関が考えられる気温データ等が挙げられる。

また、橋種を特定可能とする情報が追加で得られると、分析結果に基づく対策の検討に応用できると考えられる。さらに、交通量や、あるいはそれを特定可能な道路の級種区分情報などをあわせて利用できると分析や考察に活用できると考えられる。そのほか、管理区分について、地方公共団体は、県や市町村への区分により、管理可能な予算規模や体制による差異についても明確化できる可能性があるほか、海岸線からの距離や気候等の地理的特性も加味し得る。

これらのデータを追加することで、性能向上が期待されるとともに、結果の解釈についての議論を深められる可能性があると考えられる。

#### c) データの結合における課題

一方で、変数の追加については、レコードのマッチングを行う必要があるが、データソース毎にバラバラと考



えられる諸元情報等を元にしたマッチング作業は、労力  
 が大きいことが課題である。このマッチング作業は、多  
 くのデータサイエンスにおいて、データ調製・データク  
 レンジングに関する共通した課題であるとも言える。こ  
 の課題解決のためには、データセット間で共用可能な構  
 造物 ID 等の整備、ID だけでなく位置情報や諸元等から  
 類似性を評価可能となるデータの流通が期待される。令  
 和元年には、国土交通省より、定期点検における対象施  
 設の ID 付与に関するルールが示されているが、こうし  
 た動きが加速され、共通化された形式により取得可能な  
 データプラットフォーム等の早期実現が期待される。

## 6. おわりに

本研究では、アンサンブル学習による機械学習手法を  
 中心として、インフラ維持管理データを対象とした分析  
 を試行した。この結果について、性能評価とモデル解釈  
 性評価を行った。性能評価については、手法毎の数理的  
 特性を考慮した上で、納得のできる結果であった。この  
 ことより、評価指標の有効性が示唆された。

また、モデルの解釈性評価については、FI と Shapley  
 値いずれについても、利用における特性の把握は必要な  
 もの、モデル解釈は可能であることが示唆された。

謝辞：本研究開発の一部は、総務省 SCOPE #162103120,  
 #181503004 の委託および、JSPS 科研費 JP00377847 の助成  
 を受けたものである。

## 参考文献

- 1) 内閣府 戦略的イノベーション創造プログラム(SIP)  
 「インフラ維持管理・更新・マネジメント技術」、  
 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、公益社団法人土  
 木学会技術推進機構、2019.3.
- 2) Barua, L., Zou, B., Noruzoliaee, M. and Derrible, S.: A  
 gradient boosting approach to understanding airport  
 runway and taxiway pavement deterioration,  
*International Journal of Pavement Engineering*, pp. 1-  
 15, 2020.
- 3) Li, F. and Wang, D.: Reliability analysis of lashing bridge  
 of ultra-large container ship based on improved gradient  
 boosting decision tree-Monte Carlo method, *Chinese  
 Journal of Ship Research*, Vol. 15, No. 2, pp. 63-69,  
 2020.
- 4) Chun, P. j., Izumi, S. and Yamane, T.: Automatic  
 detection method of cracks from concrete surface imagery  
 using two-step light gradient boosting machine,  
*Computer-Aided Civil and Infrastructure Engineering*,  
 Vol. 36, No. 1, pp. 61-72, 2021.
- 5) Jun, Z., Jian, M., Jiemin, X. and Zuohong, M.:

Displacement prediction model for concrete dam based on  
 PSO-GBDT, *IOP Conference Series: Earth and  
 Environmental Science*, Vol. 358, No. 5, pp. 052043,  
 2019.

- 6) Wang, B., Yang, K., Wang, D., Chen, S. Z. and Shen, H.  
 J.: The applications of XGBoost in Fault Diagnosis of  
 Power Networks, *2019 IEEE Innovative Smart Grid  
 Technologies-Asia (ISGT Asia)*, pp. 3496-3500, 2019.
- 7) 安野貴人：予測メンテナンスとインフラ劣化学習モ  
 デル, 人工知能学会全国大会論文集 第 32 回全国大  
 会 (2018), pp. 3Z104-103Z104, 2018.
- 8) 龍田 齊, 横山 広, 永見 武司, 榊谷 浩, 近田 康夫, 山  
 田 宗明: 勾配ブースティング決定木を用いた橋梁損  
 傷原因および補修工法の推定と分析, AI・データサ  
 イエンス論文集, 第 1 巻, 第 J1 号, pp. 63-70,  
 2020.
- 9) Chen, T. and Guestrin, C.: Xgboost: A scalable tree  
 boosting system, *Proceedings of the 22nd acm sigkdd  
 international conference on knowledge discovery and  
 data mining*, pp. 785-794, 2016.
- 10) Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma,  
 W., Ye, Q. and Liu T. Y.: Lightgbm: A highly efficient  
 gradient boosting decision tree, *Advances in neural  
 information processing systems*, Vol. 30, pp. 3146-3154,  
 2017.
- 11) Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush,  
 A. V. and Gulin, A.: CatBoost: unbiased boosting with  
 categorical features, arXiv preprint arXiv:1706.09516,  
 2017.
- 12) 東京大学大学院情報学環「情報技術によるインフラ  
 高度化」社会連携講座 (第 2 期) : 社会連携講座の  
 活動成果の概要 2014-2018, <[http://advanced-  
 infra.org/archives/1650](http://advanced-infra.org/archives/1650)>, (accessed 2020.7.7).
- 13) 湧田雄基, 阿部真育: インフラ維持管理におけるデ  
 ータサイエンス活用による生産性向上, 建設機械施  
 工, Vol. 71, pp. 10-16, 2019.
- 14) Wakuda, Y., Isobe, M. and Ishikawa, Y.: Data science  
 support system for infrastructure maintenance on cloud,  
*17th International Conference on Computing in Civil and  
 Building Engineering (ICCCBE 2018)*, 2018.
- 15) Shearer, C.: The CRISP-DM model: the new blueprint for  
 data mining, *Journal of data warehousing*, Vol. 5, No. 4,  
 pp. 13-22, 2000.
- 16) 国土交通省 道路メンテナンス年報,  
 <[https://www.mlit.go.jp/road/sisaku/yobohozen/yobohoze  
 n\\_maint\\_index.html](https://www.mlit.go.jp/road/sisaku/yobohozen/yobohoze_n_maint_index.html)>, (accessed 2021.7.12).
- 17) Lundberg, S. M. and Lee, S. I.: A unified approach to  
 interpreting model Predictions, *NIPS'17: Proceedings of  
 the 31st International Conference on Neural Information  
 Processing Systems*, pp. 4768-4777, 2017 .
- 18) Strobl, C., Boulesteix, A. L., Zeileis, A. and Hothorn, T.:  
 Bias in random forest variable importance measures:  
 Illustrations, sources and a solution, *BMC bioinformatics*,  
 Vol. 8, No. 1, pp. 1-21, 2007.

(Received June 30, 2021)

(Accepted August 31, 2021)

# MODELING TRIAL AND INTERPRETATION OF RESULTS FOR MACHINE LEARNING APPLICATIONS IN INFRASTRUCTURE MAINTENANCE AND MANAGEMENT

Yuki WAKUDA, Akemi YAMASHITA, Keisuke YOSHIDA, Hitoshi TATSUTA, Kazuhiko SEKI, Kenji ARII, Kentaro KUMAGAI, Kazuyuki NAKAHATA, and Satoshi NAGANUMA

In this paper, we discuss the possibility of using Artificial Intelligence (AI) in infrastructure management, focusing on the analytical performance and interpretability of models. In particular, the paper outlines the mathematical background of ensemble learning methods, such as XGBoost, LightGBM, CatBoost, RandomForest, and decision tree analysis, which have recently achieved good results in machine learning applications. We report on the results of trial estimations of bridge deterioration determined using these methods. In addition, this paper discusses the analysis results from the viewpoint of AI application in infrastructure management, considering the characteristics of each method.