

論文 コンクリート橋梁のひび割れ損傷に対する回帰モデルに適した機械学習アルゴリズムの選択

岡崎 百合子*1・岡崎 慎一郎*1・浅本 晋吾*2・全 邦釘*3

要旨：コンクリート橋梁の劣化は、塩害や中性化等に起因する損傷と活荷重等の作用が互いに関連し、極めて複雑な様相を呈することから、その損傷予測に対し、膨大な情報の裏に潜むパターンを認識できる機械学習が注目されている。本研究は、コンクリート橋梁の劣化に関するデータのように、入力パラメータが多く、時間軸上に稠密でもなく等間隔でもないデータ群を対象として、機械学習の各アルゴリズムを用いた回帰モデルの性能比較を行い、最も適したアルゴリズムの選択を試みるものである。

キーワード：機械学習、橋梁劣化、重回帰、ガウス過程回帰

1. はじめに

高度経済成長期に建設された多くのインフラ構造物において劣化が顕在化している。この一方で、少子高齢化に伴い、これらを管理する財源や人材が必ずしも豊富ではなく、劣化に対する対応が今後不十分となる可能性が高い。このように、近年、維持管理に対する資本が制限される中で構造物のストック効果の最大化が望まれる社会情勢に加え、電子計算機の処理速度の飛躍的な向上という状況に伴い、こうしたインフラ構造物の損傷予測に対し、膨大な情報の裏に潜むパターンを認識できる機械学習が再び注目されている。特に、コンクリート橋梁の劣化は、塩害や中性化等に起因する損傷と、活荷重等の作用が互いに関連するところに特徴があり、極めて複雑な様相を呈するため、これまでの実橋梁の点検データ内に含まれる複雑な劣化機構を機械学習により抽出し、効率的な維持管理に反映することが期待されている。

機械学習は、複雑に相関するパラメータ群を用いて、合理的に回帰並びに識別モデルを構築できるという特徴があり、土木工学分野では、路面性状調査、河川や海面の潮位予測¹⁾、コンクリート分野では、コンクリート構造物におけるひび割れ識別²⁾、鉄筋の腐食予測³⁾等への利活用がなされている。

ここで、機械学習アルゴリズムは、ニューラルネットワーク、決定木など多く提案されており、各手法には得手不得手があるため、各手法の特性をよく理解したうえで、解くべき問題に対して用いるのが適切か吟味する必要がある。さらに、各アルゴリズムには、チューニングパラメータがあり、これらの操作により精度が大きく変動することも考慮しなくてはならない。しかしながら、解くべき問題に対してどのような機械学習アルゴリズムを用いるべきかについての比較検討について詳細報告し

ている例はほぼ見られない。

コンクリート工学において今後期待されているのは、先達により収集された、種々の入力情報に対するコンクリートの物性等の出力という貴重なデータの利活用である。これらのデータの特徴は、コンクリートの使用材料や曝露された環境等、入力パラメータは数多い一方で、強度や中性化深さ、ひび割れ等の出力は時間軸上には多く得られておらず、例えば強度であれば、7日、28日、91日というように、時間軸方向に等間隔でなく、情報がスパースとなるという特徴がある。したがって、コンクリート工学分野においては、入力パラメータの多さを勘案したうえで、時間軸上に稠密でもなく、等間隔でもないデータ群に適したアルゴリズムを選択する必要があると考えられる。

以上より、本研究は、コンクリート橋梁のひび割れ損傷の経時変化に関する回帰を例題として、機械学習の各アルゴリズムを用いた回帰モデルの性能比較を行い、最も適したアルゴリズムの選択を試みるものである。

2. 機械学習

2.1 概要

機械学習とは人間が自然に行っている学習能力と同様の機能を計算機で実現する技術・手法である。機械学習の多くは、既知の入力および出力データを教師としてモデルに学習させ、新しい入力に対する出力を予測できるようにする教師あり学習であり、予測のタイプは分類と回帰に大別される。本研究では、点検データから得られたコンクリート橋梁のひび割れ損傷度を教師データとし、次節に示した機械学習アルゴリズムを用いて構築した各回帰モデルに対し、性能比較を行う。なお、計算に使用したソフトウェアは Mathworks 社製 Matlab である。

*1 香川大学 創造工学部 創造工学科 建築・都市環境コース 博（工）(正会員)

*2 埼玉大学 工学部 建設工学科 博（工）(正会員)

*3 愛媛大学大学院 理工学研究科 生産環境工学専攻 環境建設工学専攻 博（工）(正会員)

2.2 機械学習アルゴリズム⁴⁾⁵⁾

(1) 重回帰

本モデルは、予測子（独立変数）が複数ある線形モデルにおいて、**図-1(a)**に示すように、回帰値と実測値の平均二乗誤差を最小とするように重みベクトルを学習させるものである。原理は非常に単純であり、解釈も容易であるが、柔軟性は非常に低い。

(2) サポートベクター回帰

サポートベクター回帰（以下、SVR という）は、分類問題において近年注目されているサポートベクターマシン（以下、SVM という）を回帰問題へ拡張したものである。SVM は、**図-1(b)**に示すとおり、マージンを最大化する最適な分離超平面を見つけるものであり、未学習データに対する高い識別性能である汎化性能を得る工夫が成されており、SVR の回帰問題への拡張である SVR も高い汎化能力が期待される。パラメータのチューニングが容易で、ロバスト性が高く、安定した結果が得られやすいという特徴がある。

なお、本研究では、基本的な線形 SVR と、非線形問題に対応するため、カーネルトリックを用いたガウシアン SVR（カーネル関数として汎用的なガウシアンカーネルを使用）により検証を行う。

(3) 決定木

決定木の学習は、元となるデータ群を属性値テストに基づいて部分集合に分割することにより行い、同処理は、すべての部分集合に対して再帰的に繰り返される。その結果生成された決定木は、フローチャート的な木構造を成しているため、学習の結果について可読性が高いという特長を持つ。一方で、決定木を大きく成長させ過ぎると過学習になり、汎化性能が損なわれる可能性が高くなるといった側面も併せ持つ。

こうした決定木の問題を克服するために開発されたアルゴリズムがランダムフォレストである。ランダムフォレストは、決定木を弱学習器とする集団学習アルゴリズムであり、相関の低い多数の決定木群を生成し、その平均をとることにより、十分な正確性を達成しながら汎化能力の確保を実現する。

(4) ニューラルネットワーク

ニューラルネットワーク（以下、NN という）とは、人間の脳神経系のニューロン（神経細胞）を数理モデル化したアルゴリズムである。NN のネットワーク構成法は様々であるが、代表的な NN は、**図-1(c)**で示す入力層、中間層、出力層から構成される階層型 NN である。入力層で入力された値が中間層の複数のユニットに伝播して計算を繰り返し、出力層では、中間層での各出力を入力として最終的な出力結果を得る。

中間層を複数にすることで、データ群に含まれる特徴

を段階的により深く学習することを可能とする深層学習が知られているが、非線形関数に対しては、1 層の中間層でも、ユニット数、すなわちニューロン数を適切に増やすことで十分に対応可能であることが知られている⁶⁾。したがって本研究では、中間層を 1 層とし、ニューロン数をパラメータにして複数の NN を構築した。なお、出力に用いられる活性化関数は汎用的なシグモイド関数とした。

(5) ガウス過程回帰

ガウス過程回帰モデルは、ベイズ統計学に基づいた設定においてカーネルを確率モデルに適用したノンパラメトリックモデルである。カーネルモデルは、モデル構築時に基底関数を予め設定するパラメトリックアプローチに対し、学習データ点を中心として定義されるカーネル関数 $k(x)$ の任意の 2 つのベクトルの類似度を示す) を導入することで目的応答 $y(x)$ を求める。**図-1(d)**に 1 次元のカーネルモデル概念図を示す。本モデルはノイズの多いデータに対して過学習となりにくく、また、多数の予測子に対応可能である一方、データ数 N に対し $O(N^3)$ の計算時間が必要となるという欠点を有する。

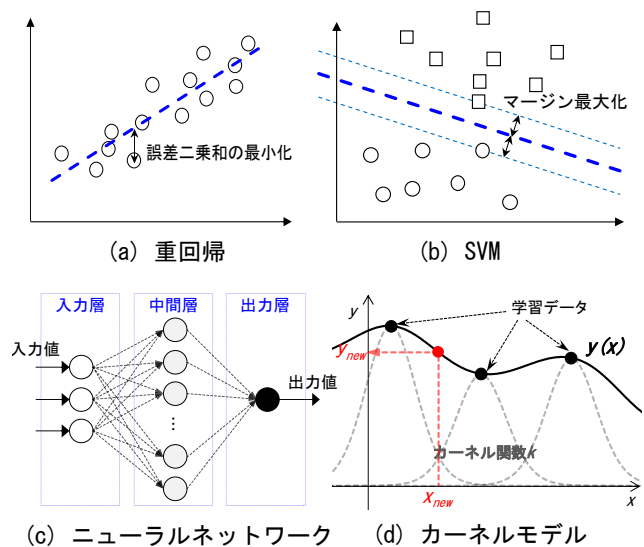


図-1 機械学習 各アルゴリズムの概念図

3. 検証例題

3.1 学習データ

(1) 概要

本研究では、香川県内で国が直轄する 347 橋の橋梁のうち、主桁にコンクリートを用いた 265 橋の橋梁のひび割れ損傷ランクの経時変化を回帰対象とした。**図-2**に対象橋梁の位置を示す。この橋梁群に対し、H17 年から H27 年に実施された定期点検のデータ等を学習データとして各アルゴリズムで回帰モデルを構築し、性能を比較検証する。



図-2 対象橋梁の位置

(2) データの内容

国道における定期点検では、対象橋梁は部材の径間ごとに要素が区切られ、損傷があれば、その種類とランクが要素に記載される。本研究では、部材としては主桁を対象とし、損傷種類として、最もデータの多いひび割れを用いた。なお、損傷ランクは低い方から a~e と記述されているが、以降の検討し易さを考慮して、これを 5 (健全) ~1 (劣化) に読み替えた。

ひび割れ損傷ランクに対する予測子の候補として、同じ定期点検表から得られる橋長等の橋梁諸元、交通量等の荷重条件の他、環境条件として、国土地理院のデータベースから年降水量等のデータを抽出した。さらに、同橋梁同径間内のばらつきを表すデータとして、「漏水の有無」と「空間位置」を加えた。「漏水の有無」は、定期点検結果で得られる損傷データのうち、「漏水・滞水」「漏水・遊離石灰」の損傷ランクが4以下となる要素で「有り」、それ以外で「なし」と分類し、「空間位置」は径間内での桁端部を「端部」とし、それ以外を「非端部」として分類した。表-1 にこれらの予測子候補一覧を示す。なお、後述するように、過学習防止のため、この中から各モデルで必要最低限の予測子セットを選択し、モデルを構築する。

表-1 予測子候補一覧

区分	項目				
橋梁諸元	橋長	支間長	全幅員	構造形式(橋種)	総径間数
設計条件	塩分対策区分	設計活荷重	示方書年代(上部構造)	材料	打設場所
荷重条件	交通量	大型台数			
環境条件	年降水量	年平均気温	海岸線からの最小距離	標高	
時間軸	供用年数	完成年度			
径間内ばらつき	漏水有無	空間位置			

(3) データの特徴

本研究で用いたデータの特徴として、まず、予測子の数が多いと想定されることが挙げられる。また、経時変

化を追跡する上で重要となる時間軸上の特徴に着目すると、図-3 のとおり、高度成長期である 1950~1970 年に多数の橋梁が建設されている一方、それ以前の橋梁は少ない。これは、供用年数が 40~50 年程度の橋梁が多く、50 年以上の橋梁は少ないことを意味する。

また、現在の点検要領で収集された点検データは平成 17 年からであり、その後は基本的には 5 年の間隔で点検が実施されていることから、図-4 の実測値が示すような、時間軸上にスパースなデータとなっている。

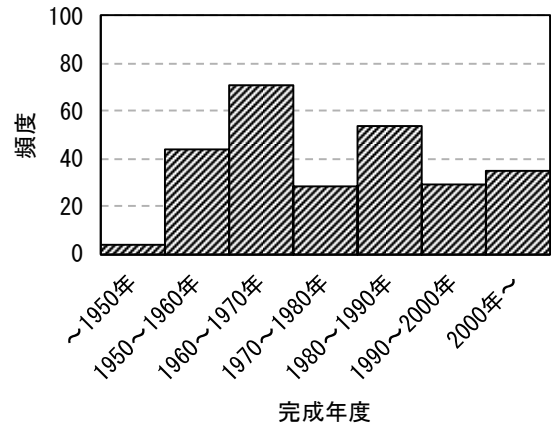


図-3 対象橋梁の完成年度 度数分布

3.2 データの前処理

本研究では、ひび割れ損傷に関する経時変化の回帰精度を検討対象とするが、前述のとおり時間軸上にスパースなデータとなっている。一般に、機械学習では、データの欠損がモデルの精度を低下させることが知られている。そこで、データ整理の段階で可能な限り時間軸上のデータの補間することとした。まず、供用開始直後は、初期欠陥等により生じるひび割れ発生の可能性があるので、本研究においてはひび割れ損傷ランクを 5 (健全) として学習データに追加した。さらに、補間するデータは 5 か年の倍数とすることで、対象橋梁群の時間軸を合わせて比較しやすくするとともに、対象データと直近の過去のデータとで同じ損傷ランクであった場合には、損傷ランクが期間内は変動しないと仮定し、データの補間を行った。一例を図-4 に示す。

以上の処理を行った結果、学習に用いた橋梁の数は 265 橋、要素数としては 6,664 要素となり、供用年数毎に 1 データとすると、計 40,695 データを学習に使用した。これらの学習データ群における供用年数の度数分布を図-5 に示す。供用年数が大きくなるほど、データ数が少なくなり、データのスパース性が增大する。また、図-6 に学習データ群におけるひび割れ損傷ランク 1~5 の頻度を示す。ランク 5 の健全なデータが大半であり、それ以外のものが 100 個のオーダーで存在している。

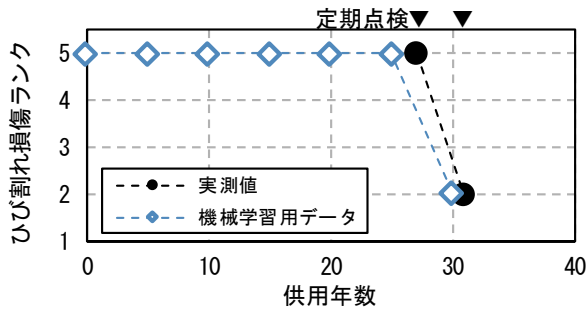


図-4 同要素の実測データおよび補間の一例

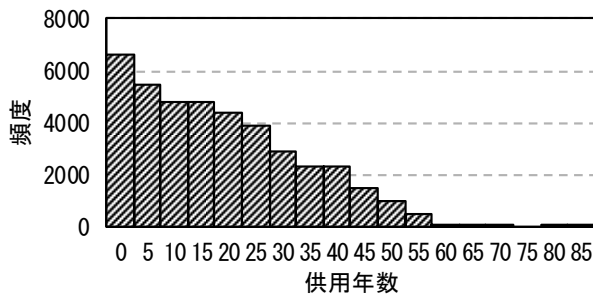


図-5 学習データにおける供用年数の度数分布

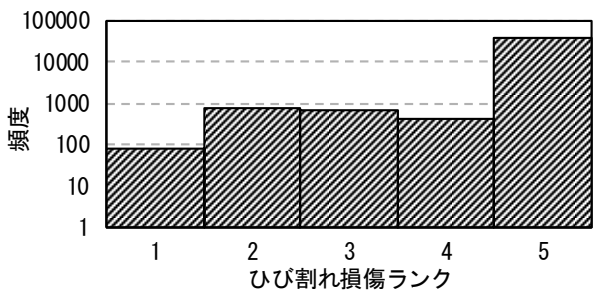


図-6 学習データにおけるひび割れ損傷ランクの度数分布

4. 機械学習結果

4.1 概要

コンクリート橋のひび割れ損傷ランクの経時変化に対し、2章で示した各アルゴリズムにより回帰モデルを構築した。

ここで、機械学習における過学習（過適合）について言及する。過学習とは、学習データに含まれる本来の規則性だけでなく、ノイズまでも学習することであり、学習データに対しては誤差が低くなる一方、未知の新規データでは誤差が増大し汎化性能が低い状態に陥ることを指す。過学習に対する対策は、変数選択法や収縮法、正則化法、交差検証等、多く提案されている。本研究では、各アルゴリズムに対し、変数選択法の一つである逐次特徴量選択を5分割交差検証時の平均二乗誤差（RMSE）を基準として実施した。なお、NNにおいては学習結果が不安定であったため、ベイズ正則化を行って100個の

ネットワークを構築し、それらの出力平均値を応答とすることで過学習を防止した。

4.2 各モデルにおける性能評価

(1) 5分割交差検証

図-7に、各アルゴリズムで構築したモデルでの5分割交差検証時のRMSEを示す。なお、SVRにおいては、線形SVRとガウシアンSVR、決定木においては、決定木とランダムフォレスト、NNにおいては、ニューロン数をパラメータにした交差検証においてニューロン数10でRMSEが収束したことから、ニューロン数を1, 5, 10としたものに関して検討した。

はじめに、線形回帰モデルと線形SVRの結果より、同じ線形モデルでの比較を行う。線形回帰モデルにおいては、RMSEが0.20である一方、線形SVMのRMSEは0.24となった。これは、回帰線の決定基準が、重回帰モデルでは全データに対する誤差の二乗和最小化であるのに対し、SVRでは、回帰線近傍の不感帯に属していない誤差の大きなデータの誤差絶対値の総和最小化であるためである。今回対象としたデータはランク5の健全データが多く存在し、SVRでは、供用年数が進んでもこれらの健全データが不感帯に属するように回帰線が調整されたと考えられる。

次に、非線形モデルであるガウシアンSVR、決定木、ランダムフォレスト、NN、ガウス過程回帰については、NNがニューロンの数の増加により多少の改善が見られるものの、いずれもRMSEは0.1程度であり、前述の線形モデルに比較し、性能が高くなっている。

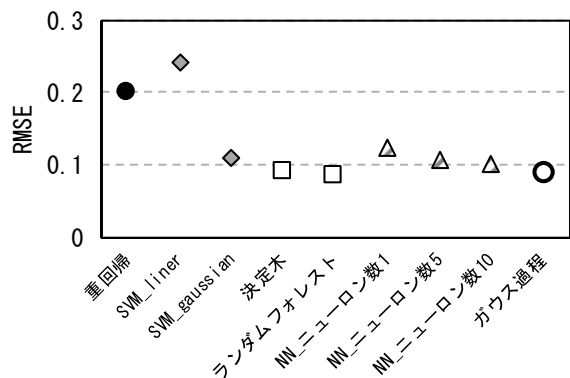


図-7 各手法による5分割交差検証時RMSE

(2) 再代入誤差

各モデルの再現性を詳細に確認するため、モデル構築に用いた実測値と回帰値の差を、実測のひび割れ損傷ランク別に検討する。

図-8にガウシアンSVR、ランダムフォレスト、ニューロン数10のNN、ガウス過程回帰モデルの実測値と回帰値の差のヒストグラムを、実測値のランクごとに示す。

いずれも、実測値のランク5から2まで比較的高精度に再現に至っているが、ランク1では実測との乖離が大きくなっている。これは、損傷ランク1となっている要素と同じ予測子ベクトルを有する別要素において、損傷ランクに大きなばらつきがあるためと考えられる。すなわち、定期点検時のひび割れ損傷ランクの判定誤差が大きい、もしくは表-1で想定した因子以外の影響要因が予測子として必要であると考えられる。

(3) 経時変化

(1)(2)でのモデルの性能評価では、学習に用いたデータ群に対して実施した。次に、回帰モデルの経時変化の傾向から、回帰モデルの性能を確認する。

図-9に、供用25年である橋梁の同径間内にある複数の要素での実測値と、各アルゴリズムによる回帰モデル値の経時変化を示す。なお、図-9左下図に示すとおり、同径間内には「漏水」が見られる要素と、そうではない要素が存在することから、各グラフ内に色別で示しており、漏水の有無によるひび割れ損傷への影響についても同時に確認している。

はじめに、実測値(点検データ値)の傾向を確認する。点検データによれば、漏水がない要素では、定期点検が実施された供用25年時までは、ひび割れは生じていな

い。この一方で、漏水が確認された要素では、供用20年時ではランク5で健全であったが、供用25年時の点検では損傷ランクが3となっており、本径間内において漏水がひび割れ損傷を促進させる影響を確認できる。

次に、各アルゴリズムによるモデルの回帰結果を確認する。まず、重回帰モデルについて確認すると、本モデルは線形モデルであるため、時間の経過に伴い、直線的にランクが低下している。この時間に経過に伴うランク低下の程度は、データ群においていずれの供用年数であってもランク5が一番多かったためか、僅かとなっている。また、漏水の有無による差異はほぼない。

ガウシアンSVRでは、漏水がある要素ではひび割れ損傷が低下する傾向を再現できているが、実測値のようにランク3まで低下しきれていない。これは、4.2(1)に前述したように、不感帯を設けているアルゴリズムであるため、ランク5を示す健全な要素のデータの多さに起因するものと考えられる。

ランダムフォレストの場合、実測値と同様に漏水が生じている場合だけ劣化が進行する傾向が再現できているが、回帰結果の形状は鉤状であり、滑らかではない。これは、データ群を階層状に2分割しながら説明しようとするアルゴリズムのためであり、本研究対象のように連

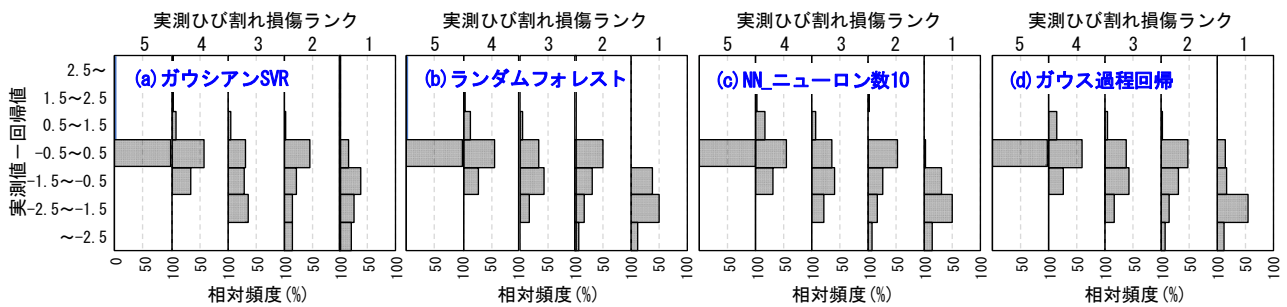


図-8 各アルゴリズムを用いた回帰モデルによる回帰値と実測値の差

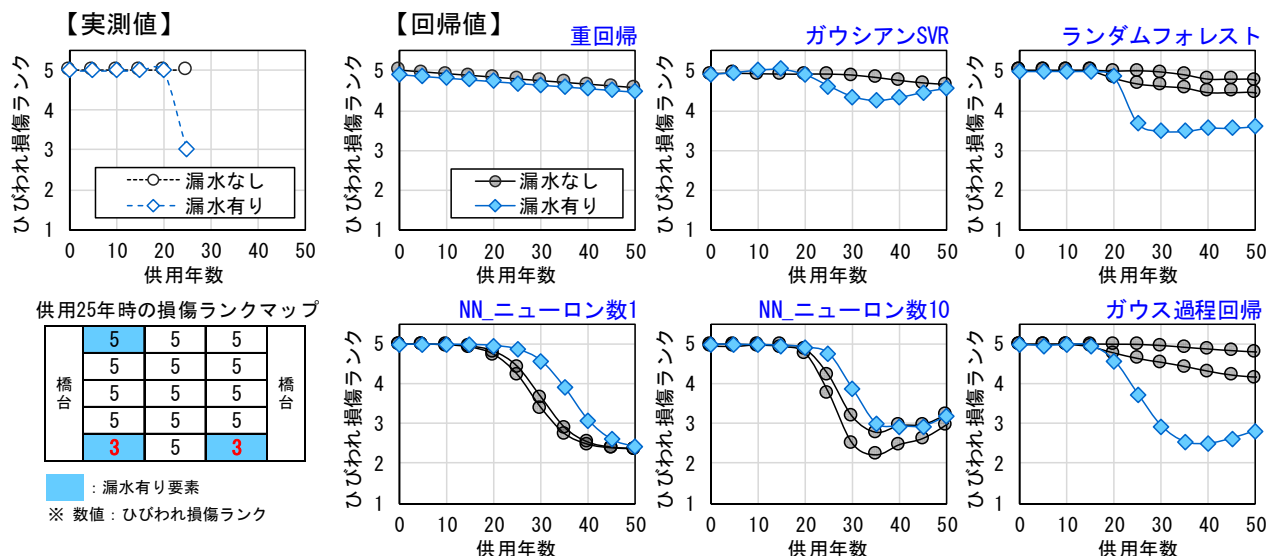


図-9 同径間内実測値と各アルゴリズムによる回帰結果の一例

続的な出力を求めるケースの場合、未知のデータに対する出力に誤差が大きくなる可能性がある。

NN では、ニューロン数が増加するほど、実測値のような、損傷ランクが急激に低下する傾向を表現できているものの、漏水の有無によるランク低下の影響が反映できていない。ニューロン数をさらに増加させることにより、漏水の有無によるランク低下の傾向の相違は表現できると思われるが、汎化性能の低下だけでなく、計算の収束性低下によるロバスト性の低下が懸念される。これに対して、NN では、多数のチューニングパラメータに対する感度分析を実施するなどの対策が取られるのが一般的であるが、これに対する経験とノウハウが必要となる。

ガウス過程回帰モデルでは、実測値と同様に漏水が生じている場合だけ劣化が進行する傾向が再現でき、かつ、回帰曲線も滑らかなものとなっている。これは、本モデルが、未知のデータを学習データ群との類似性を考慮して算定するため、漏水のような局所的で特徴的なデータに追従しやすく、データ数の多いランク 5 のデータのバイアスを受けにくいことに起因している。一方で、漏水有りの要素に対し、供用年数 40 年以降においてランクが若干回復しており、こうした傾向は、ガウシアン SVR、NN でも確認される。これは、本橋梁で得られているデータが供用 25 年までであり、それ以降の年数の回帰値はその他の橋梁の影響を大きく受けるためである。特に、3.1 に前述したとおり、供用 50 年以上のデータ数は少ない上、比較的健全な橋梁のみ残存した結果を反映したものと考えられる。この傾向は、データの追加もしくはパラメトリックモデルとの連成等により回避できると思われる。

以上より、今回対象としたスパースで、密度も局所的に異なるデータ群においては、ガウス過程回帰モデルが最も適していると考えられる。ただし、本モデルの構築時には、計算機のメモリを大量に消費し、計算量はデータ数 N に対して $O(N^3)$ オーダーとなるため、データ数によっては、近似手法を利用する等、計算の高速化を試みる必要がある。

5. 結論

コンクリート橋梁主桁のひび割れ損傷経時変化に対する回帰を事例として、機械学習の各アルゴリズムで構築したモデルの性能比較を実施した。コンクリート工学分野でよく見られるような、予測子が多く、出力がスパースで、密度も局所的に異なるデータ群においては、計算量が膨大となる問題はあるものの、ガウス過程回帰モデルが最も適していることを確認した。

謝辞

本研究で用いたデータは四国地方整備局より提供を受けた。また、内閣府総合科学技術・イノベーション会議「SIP インフラ維持管理・更新・マネジメント技術」(管理法人: JST)、および、JSPS 科研費若手研究 (A) (17H04932) により実施した。ここに謝意を記す。

参考文献

- 1) 高崎忠勝, 河村明, 天口英雄: ニューラルネットワークによる都市中小河川感潮域の水位推定, 土木学会論文集 B1, Vol.67, No.4, pp.1603-1608, 2011
- 2) 全邦釘, 嶋本ゆり, 大窪和明, 三輪知寛, 大賀水田生: ディープラーニングおよび Random Forest によるコンクリートのひび割れ自動検出手法, 土木学会論文集 F3, Vol.73, No.2, pp.297-307, 2017
- 3) 黒田裕伸, 松島学, 横田優, 佐々木孝彦, 飯島亨: ニューラルネットワークを利用したコンクリート中にある鉄筋の腐食速度の推定式の構築, 第 9 回設計工学に関するシンポジウム講演論文集, pp.133-140, 2005
- 4) Bishop,C.M: パターン認識と機械学習(上), 丸善, 2014
- 5) Bishop,C.M: パターン認識と機械学習(下), 丸善, 2014
- 6) 荒木雅弘: フリーソフトではじめる機械学習入門, 森北出版, 2014
- 7) Raschka, S.: Python 機械学習プログラミング 達人データサイエンティストによる理論と実践, インプレス, 2016