

地中埋設管の地震被害予測への機械学習の適用に関する 基礎的検討

岐阜大学工学部社会基盤工学科 能島 暢呂・大西 克茂

Fundamental Investigation on a Application of Machine Learning Techniques
for Estimation of Earthquake Damage to Buried Pipeline

Nobuoto Nojima and Katsushige Onishi
Department of Civil Engineering, Gifu University

要旨

地中埋設管の地震被害予測の高度化を目指して、現在用いられている管路被害予測式とは別のアプローチとして、機械学習の適用に関する基礎的検討を行った。2016年熊本地震における熊本市の給水管と配水管の被害データを用いて、4分の1地域メッシュ単位の給・配水管の復旧工事発生有無の予測と、工事率（各メッシュ内における給・配水管の復旧工事の管路延長 [km] あたり平均箇所数）の予測に対して、8種類の機械学習手法を適用した。その結果、各手法の特性と管路被害予測への機械学習の適用可能性が確認された。

1. はじめに

地震時には、地中に面的に広がる上水道の供給ネットワークには、多くの物理的被害が発生する¹⁾。今後発生が危惧される巨大地震に備えるために、施設の耐震化や・応急復旧計画の策定などの事前・事後対策を行う必要がある。そのような防災対策の判断材料として、被害・復旧に関する高精度な推定モデルが必要となる。

現在、行われている水道管の埋設管路被害率（導管の単位延長あたり平均被害件数 [件/km]）予測では、過去の被災事例に基づいて構築された標準被害関数に「形状や材質」、「口径」、「地盤条件」などに関する各種補正係数で補正を行う管路被害予測式が用いられる²⁾。しかしこうした予測式は事前に関数形を仮定する「パラメトリックモデル」であり、用いる関数型や補正項目の条件が合わない場合、予測精度が大きく低下する可能性がある。

一方で近年、様々な分野で注目を集めている機械学習³⁾は、事前に関数形を仮定することなく、データからモデルを構築する「ノンパラメトリックモデル」であり、柔軟なモデル表現が可能である。近年、防災分野においても、地すべりと土砂崩壊発生地点・液状化発生地点・水道や都市ガスの管路被害などの予測において機械学習の有用性が示されている。そこで本研究では水道の被害予測モデルの高度化と効率化を目指し、現行の予測手法とは別のアプローチとして、機械学習の適用を検討する。

本研究では、2016年熊本地震における熊本市の給水管と配水管の被害データを用いて、以下の2種類の検討を行う。一点目に、4分の1地域メッシュ（250mメッシュ）内における給・配水管の復旧工事発生の有無の予測を2クラスの分類問題として捉え、一般的な8種類の機械学習手法を適用する。二点目に、各メッシュ内における工事率（各メッシュ内における給・配水管の復旧工事の管路延長 [km] あたり平均箇所数）の予測を回帰問題として捉え、一般的な8種類の回帰の機械学習手法を適用する。

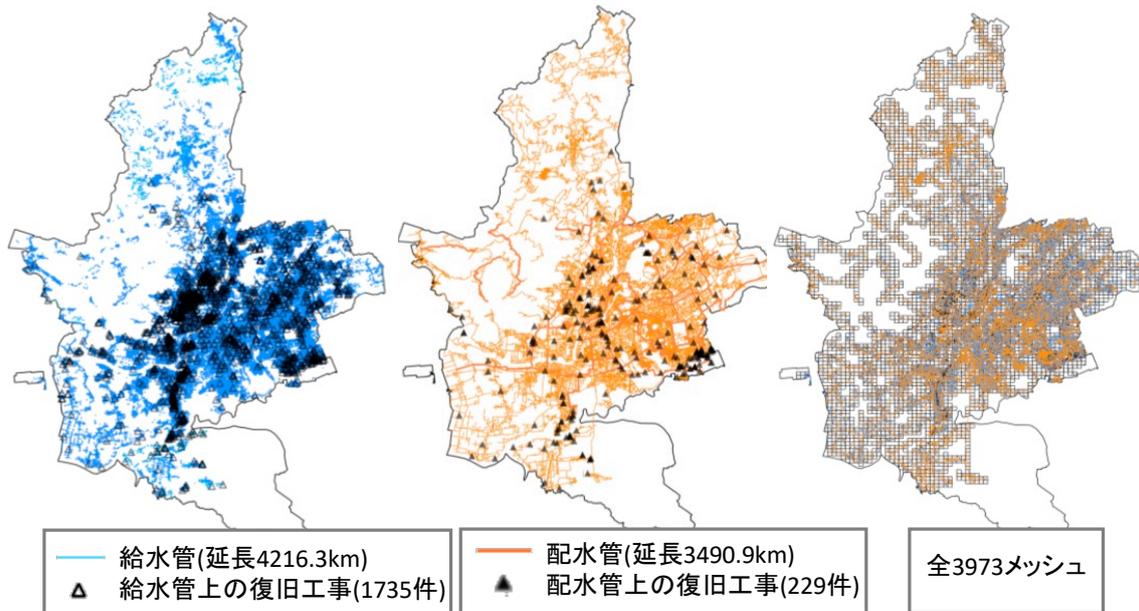


図1 熊本市の給配水管ネットワークと復旧工事発生箇所

2. 分析に使用したデータおよび分析手順と設定条件

(1) 熊本市の給・配水管ネットワーク

本研究では、2016年熊本地震により発生した、熊本市における給・配水管の被害修繕のために行われた復旧工事のデータを分析に用いる。図1に熊本市上下水道局⁴⁾より提供を受けた、給・配水管路のネットワーク図と、復旧工事が行われた地点を示す。給水管の管路延長は4216.3kmで、復旧工事件数は1735件である。また、配水管の管路延長は3490.9kmで、工事件数は229件である。いずれの管路についても市街地である東部で敷設密度が高く、工事件数が多い。4分の1地域メッシュを分析の単位とし、給水管または配水管が存在する3973メッシュを分析に用いる。なお本研究では、データ数の確保のため、給水管と配水管で発生した工事を分割せずに一まとめとして扱うこととする。

(2) 使用した変数およびデータセットの作成

モデルの説明変数として用いるのは以下の4種類である。各説明変数の空間分布および頻度分布（メッシュ数、工事箇所数、平均工事箇所数）を図2に示す。

- ①土地利用区分：土地の利用状態を示すデータである。国土地理院HP⁵⁾より入手した。平均工事箇所数（1メッシュあたりの平均工事件数）が最も多かったのは鉄道用地で0.93[箇所/メッシュ]である。
- ②AVS30：表層30mの平均S波速度である。地震ハザードステーション(J-SHIS)⁶⁾より入手した。平均工事箇所数が最も多いのはAVSが200～300[m/sec]のメッシュで0.814[箇所/メッシュ]である。
- ③計測震度：地震動マップ即時推定システム(QuiQuake)⁷⁾より入手した。2016年4月14日の前震と16日の本震について、メッシュ毎に計測震度が大きい方を計測震度データとして採用する。計測震度の増加に伴い平均工事箇所数も増加している。
- ④管路延長：各メッシュにおける給水管と配水管の延長距離の合計である。平均工事箇所数が最も多かったのは管路延長距離が8km以上のメッシュである。

本研究の被説明変数としては、「復旧工事の有無」および「単位延長距離あたり工事率」の2種類とする。説明変数と同様に4分の1地域メッシュ毎にデータの作成を行う。

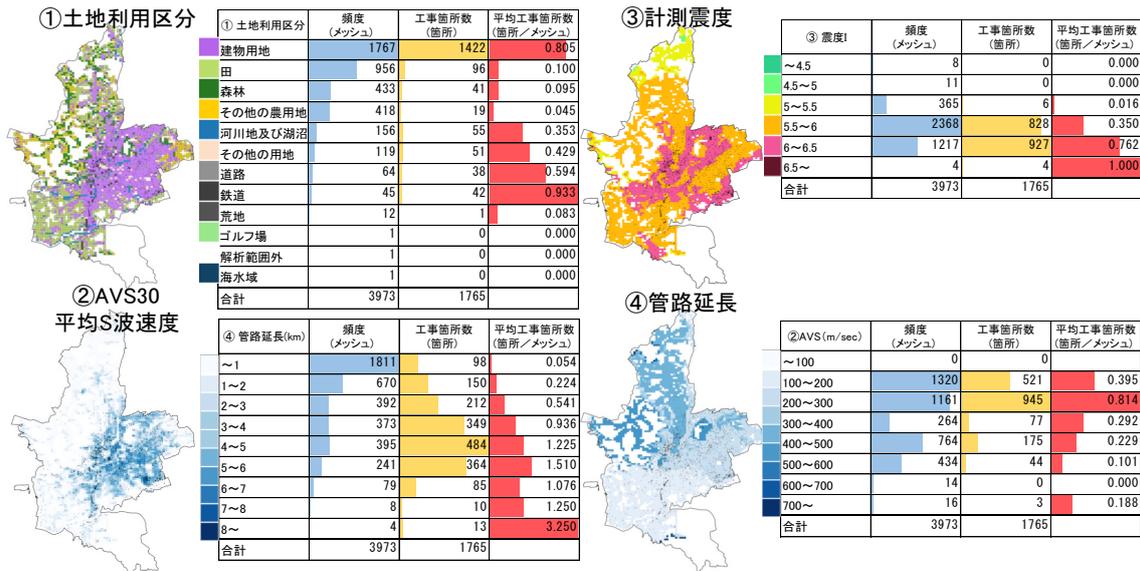


図2 各説明変数の空間分布と頻度分布 (メッシュ数・工事箇所数・平均工事箇所数)

- (a) 復旧工事の有無： 各メッシュ内で1つ以上復旧工事が行われていれば「工事有り」、工事がなければ「工事無し」とする。全 3973 メッシュのうち、「工事有り」が 883 メッシュ、「工事無し」が 3082 メッシュである (図 3 左)。
- (b) 工事率 (箇所/km)： メッシュ内にて復旧工事が行われた、883 メッシュについて工事率を算出する。各メッシュ内において、給水管および配水管にて発生した工事数の合計を、給水管および配水管の管路延長距離の合計 [km] で除して算出する。全 883 メッシュのうち、81,2%にあたる 717 メッシュが工事率 0~1[箇所/km] である (図 3 右)。

上記の「復旧工事の有無」を被説明変数として、上記の説明変数によるクラス分類を行うために機械学習を適用する。また、「工事率」を被説明変数として、上記の説明変数による回帰分析を行う機械学習により行う (図 4)。

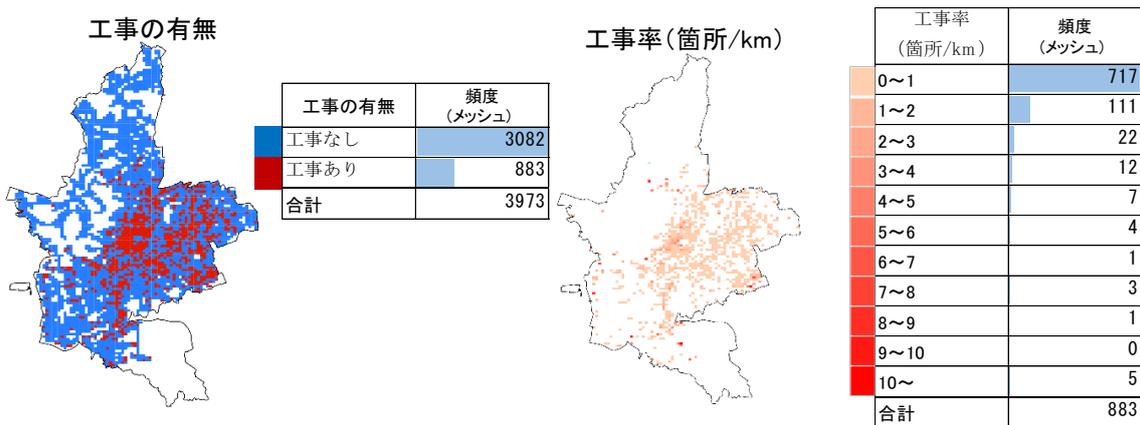


図3 工事の有無および工事率の空間分布とヒストグラム

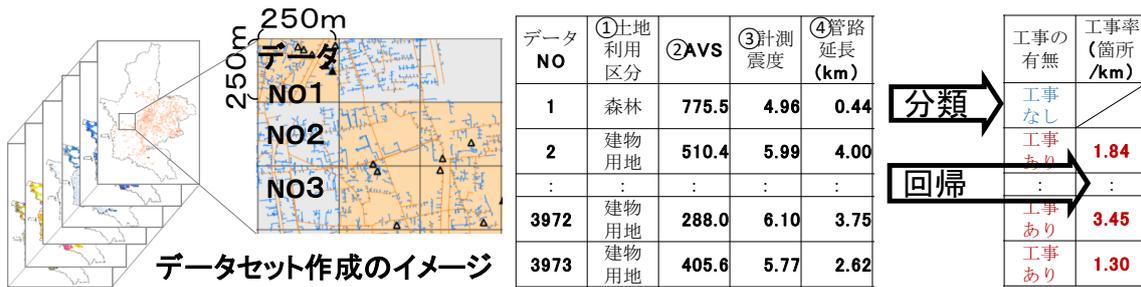


図4 データセット作成方法の概念図

(3) 分析手順と用いた手法および設定条件

上記のデータセットを用いてモデルを構築して、その検証を行う。図5にフローチャートを示し、以下に分析処理の流れを示す。

- 1) 質的変数である土地利用区分を0と1からなるダミー変数に変換する(図5の①)。
- 2) 説明変数と被説明変数からなる初期データを、モデルの学習に用いるトレーニングデータと、学習したモデルの検証を行うテストデータに分割する(図5の②)。分割割合は等分とする。

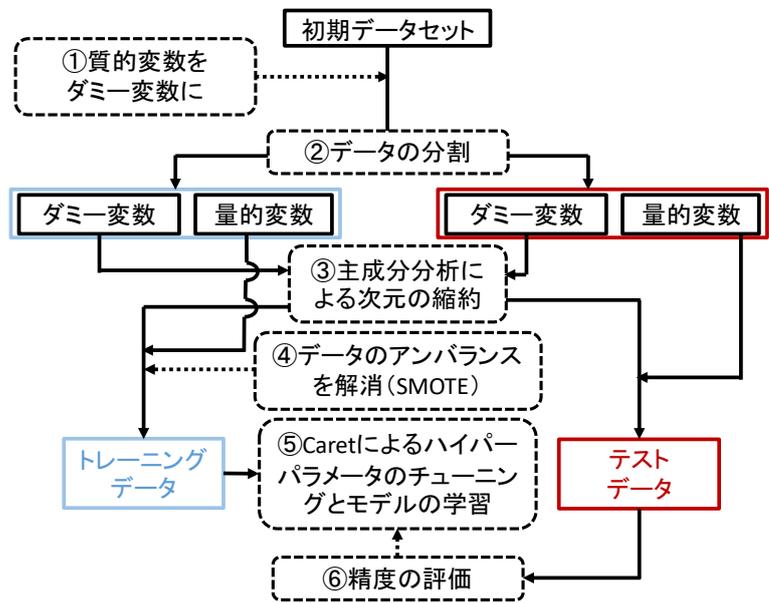


図5 分析の流れを示すフローチャート

- 3) 1) で土地利用区分をダミー変数に変換する際に、カテゴリの数だけ次元が増加する。そこでダミー変数に対して主成分分析⁸⁾を適用し、得られた主成分得点を高次から順に複数個選択することで次元の縮約を行う(図5の③)。検証の結果、1次のみ用いた場合に最も高い予測精度が得られたため、本研究ではその結果について示す。
- 4) クラス分類に適用する工事の有無のデータは、「工事無し」のデータが「工事有り」の約4倍とアンバランスである。そこでオーバーサンプリング法の1種である SMOTE⁹⁾を用いてデータのアンバランスを解消する(図5の④)。検証の結果「工事あり」を2倍とした場合に最も高い予測精度が得られたため、本論文ではその結果について示す。
- 5) トレーニングデータを用いてモデルの構築を行う(図5の⑤)。本研究では分類・回帰共に一般的に用いられている8種類の機械学習手法(表1)を適用する。分析には統計分析ソフト R のパッケージ Caret¹⁰⁾を用いる。なお各機械学習手法のハイパーパラメータについては、Caret のチューニング機能に依ってグリッドサーチを行い決定する。
- 6) 構築されたモデルにテストデータの説明変数を入力し、工事の有無および工事率の予測を行う。予測値とテストデータの実測値との誤差の評価指標を算出する(図5の⑥)。

3. 結果と考察

(1) 復旧工事の有無の予測への機械学習の適用

まず、復旧工事の有無の予測のために、8手法の機械学習によるクラス分類の予測結果と実測値を比較したマップを図6に示す。図中の表は全てのテストデータについて、予測値と実測値を比較した結果を示している。赤のTP(True Positive)とオレンジのFP(False Negative)が「工事有り」と予測された地点であり、TPが「正解」、FPが「空振り」である。青のFN(False Positive)と緑のTN(True Negative)が「工事無し」と予測された地点で、FNが「見逃し」、TNが「正解」である。いずれの手法においても、予測が誤りである地点は少なくないが、都市部にて被害が集中する傾向、郊外で被害が少ないという特徴が適切に予測できていることから、機械学習の適用可能性が確認できた。FN(見逃し)が最も少なかったのは単純ベイズの89地点である。FP(空振り)が最も少なかったのはSVM(線形カーネル)の271地点である。

次に各手法により得られた評価指標を図7に示す。それぞれの定義式は図中の表に示した通りである。全てのテストデータのうち正しく正解を予測できた割合を示す Accuracy は、手法による差があまり見られず0.74~0.80と8割程度であった。実際に工事が行われたメッシュを適切に工事有りと予測できた割合を示す TP rate は、空振りが最も少ない単純ベイズの0.80が最大であった。工事有りと予測したテストデータのうち、実際に工事が行われていた割合を示す PP value は、見逃しが最も少ないSVM(線形カーネル)の0.53が最大であった。

TP rateとPP valueは一般にトレードオフの関係にあるため、一般的には両者の調和平均をとったF値(F-Measure)がモデルの総合指標として用いられる。F値が最も高いのはSVM(線形カーネル)であった。線形カーネルはRBFカーネルとは異なり線形の識別面が構築される。今回対象としている管路被害のような不確実性の高い分類問題に対しては、フィッティング性能の高い分類手法よりも、ある程度大まかな分離を行う方が高い精度が得られるのではないかと考える。その他、単純ベイズおよびランダムフォレストによっても、SVM(線形カーネル)に準ずるF値が得られた。

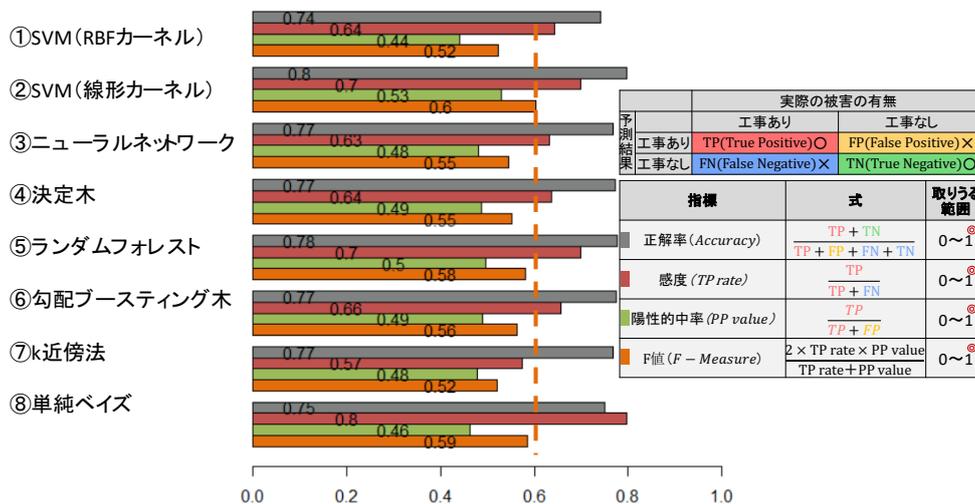


図7 全8種類の機械学習による分析の結果得られた精度評価指標

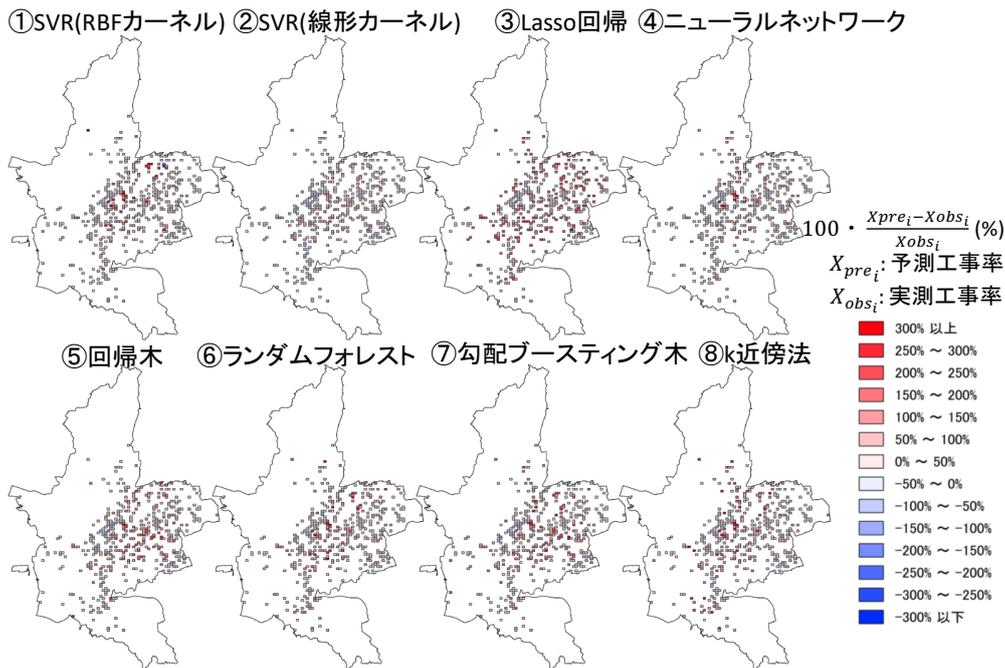


図 8 全 8 種類の機械学習手法による予測誤差の割合 (%)

(2) 工事率の予測への機械学習の適用

次に、工事率の予測のために、8種の機械学習手法による回帰分析を用いた予測値と、工事率の実測値に対する予測誤差の割合を示したマップを図8に示す。SVR（線形カーネル）を除いた7手法では、予測工事率が実測値よりも大きい地点（濃い赤）が散見され、安全側の予測結果となることがわかる。特に Lasso 回帰は予測が安全側となる傾向が強くみられる。

図9は、全8種類の機械学習による予測値と実測値の関係のプロットである。SVR（RBFカーネル）、ランダムフォレスト、勾配ブースティング木、k近傍法に関しては、予測誤差が見られるものの45度線周りにばらついている。また、上記4手法の相関係数は0.54から0.75の間にあり予測値と実測値との間に正の相関が確認できる。以上の結果は機械学習の適用可能性を示唆するものである。一方で、SVR（線形カーネル）、Lasso回帰、ニューラルネットワークについては、高被害率において過小評価となる傾向が高いことがわかる。これは、説明変数と被説明変数との間にある非線形的な関係が適切に学習できていないことを表す。また、回帰木については、変数に対して分岐条件を重ねて適用するため、回帰結果が離散的になるというメカニズムの特徴が顕著に現れている。

図10に、全8種類の機械学習手法による工事率の予測値と実測値の誤差の程度を示す精度評価指標として、MAPE(Mean Absolute Percentage Error) および RMSE(Root Mean Square Error) を示す。それぞれの定義式は図中の表に示した通りである。MAPE(%)は平均的な誤差割合を示す指標で、0に近づくほど好ましいモデルとなる。図8において大きな予測誤差を示す地点が、他の手法に比べ少ないことが確認されたSVR（線形カーネル）が最も高位である。一方で誤差の大きい地点が相対的に多くみられたLasso回帰は最も低位の評価である。RMSEは絶対値の大きい誤差を強調するため、図9において高工事率の地点の予測が相対的に適切に行われていることが確認された手法、すなわち、SVR（RBFカーネル）、ランダムフォレスト、勾配ブースティング木、k近傍法は相対的に高位の評価となっている。一方で高被害率において過小評価となる傾向がみられた手法、すなわち、SVR（線形カーネル）、Lasso回帰、ニューラルネットワークは相対的に低位の評価となっている。

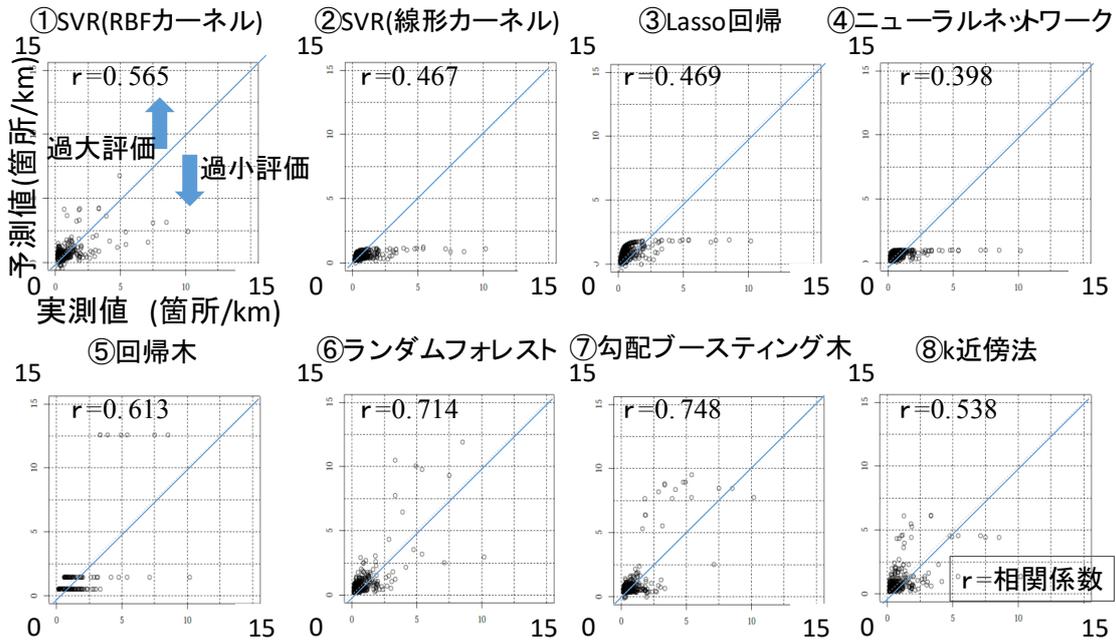
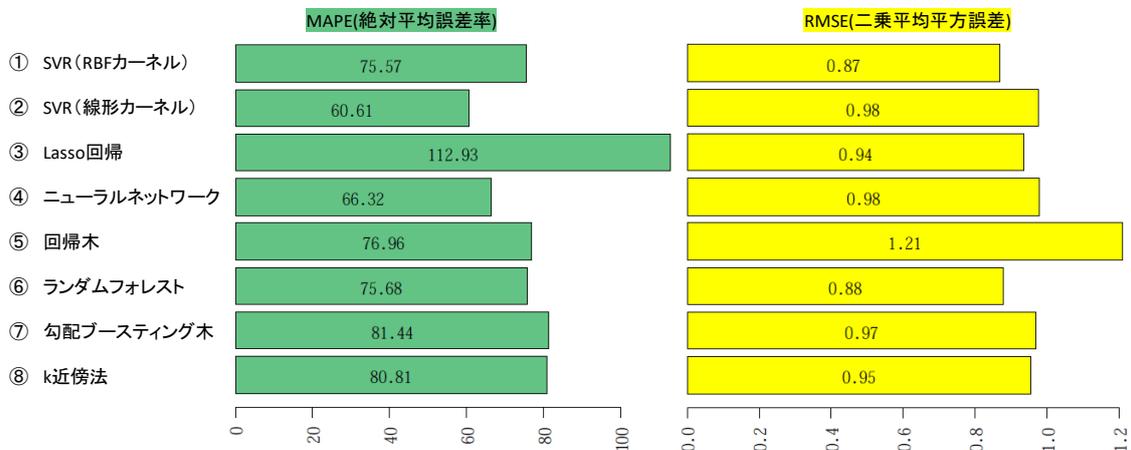


図9 全8種類の機械学習手法により得られた工事率の実測値と予測値の関係



$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{X_{pre_i} - X_{obs_i}}{X_{obs_i}} \right| \%$$

Mean Absolute Percentage Error

- ◆ 誤差の割合を算出
- ◆ 0に近いほど誤差が少ない

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_{pre_i} - X_{obs_i})^2}$$

Root Mean Squared Error

- ◆ 絶対値の大きい誤差を強調
- ◆ 0に近いほど誤差が少ない

X_{pre_i} : 予測工事率, X_{obs_i} : 実測工事率, i : データ番号, n : データ数

図10 全8種類の機械学習手法の精度評価指標

4. まとめ

2016 年熊本地震における 4 分の 1 地域メッシュの給・配水管復旧工事発生の有無の予測に一般的な 8 種類のクラス分類の機械学習を適用した。最も高い予測精度を示したのは SVM (線形カーネル) であった。その他、単純ベイズ、ランダムフォレストもこれに準ずる予測精度が得られた。いずれの手法も予測誤差が見られるが、都市部で被害が集中し、郊外では被害が少ないという傾向を掴めており、機械学習の適用可能性が確認できた。

同じく、4 分の 1 地域メッシュにおける工事率 (各メッシュ内における給・配水管の復旧工事の管路延長 [km] あたり平均箇所数) の予測にも一般的な 8 種類の回帰の機械学習を適用した。中でも SVR (RBF カーネル)、ランダムフォレスト、勾配ブースティング木、k 近傍法は比較的良好な予測結果が得られた。予測に誤差が見られるものの、予測工事率と実測値との間に一定の正の相関が見られ、機械学習の適用可能性を確認した。一方で、SVR (線形カーネル)、Lasso 回帰、ニューラルネットワークについては、高被害率において過小評価が目立ち、非線形的な関係を適切に学習できていない可能性があることがわかった。

本研究においては、管種や管径の情報を考慮していないため、こうした情報を組み込んだ分析で結果にどのような影響を及ぼすかを検討することが今後の課題である。

参考文献

1. 厚生労働省：熊本地震水道施設被害等現地調査団報告書，2016，
https://www.mhlw.go.jp/topics/bukyoku/kenkou/suido/houkoku/suidou/dl/140421-1_07.pdf
2. 水道技術研究センター：地震による管路被害予測の確立に向けた研究報告書 [概要版],2013.
3. 荒木雅弘：フリーソフトで始める機械学習入門，(株) 森北出版，pp27-149，2015.
4. 熊本市上下水道局：GIS データ，<http://www.kumamoto-waterworks.jp>，(参照 2019-01-12)
5. 国土地理院：国土数値情報 土地利用細分メッシュデータ，
<http://nlftp.mlit.go.jp/ksj/jpgis/datalist/KsjTmplt-L03-b.html>，(参照 2019-01-12)
6. (国研) 防災科学技術研究所：地震ハザードステーション，
<http://www.j-shis.bosai.go.jp/download>，(参照 2019-01-12)
7. (独) 産業技術総合研究所：地震動マップ即時推定システム (QuiQuake)，
<https://gbank.gsj.jp/QuiQuake/>，(参照 2019-01-12)
8. 中村永友：多次元データ解析法，R で学ぶデータサイエンス 2，共立出版，248p.，2009.
9. Rich data：How does SMOTE resolve the class imbalance problem?，
http://rikunert.com/SMOTE_explained，(参照 2019-01-12)
10. Max Kuhn: The caret Package, 2003,
www.kusastro.kyoto-u.ac.jp/~baba/wais/pagerank.html，(参照 2019-01-12)