

機械学習を用いた下水道管渠の劣化箇所予測

事業統括本部 下水道事業部

〃

中央研究所

○渡邊 眞明、福井 健吾、

谷裏 弘晃、東山 佳明

川口 智也

1. 背景と目的

下水道の管路施設の保全区分は状態監視保全に区分され、これら施設は地中に配置されていることから、施設の状態を把握するための調査には多大な費用を要する。現状、目視調査結果等により改築あるいは修繕が必要となる劣化が確認されるのは、概ね 20～40% であり効率的な調査とはなっていない。将来、老朽管の更なる増加を考えると、効率的に劣化を確認する調査が要請される。

本研究は、劣化箇所をスパン単位で予測する機械学習を用いたモデルを作成して二つの知見を得た。一つはモデルを性能評価した結果、凡そで F 値 60%～70% 程度を確保していること、もう一つは劣化を把握するための調査費用が削減できることであり、機械学習を用いた劣化予測モデルは有用であることを確認した。

2. 基礎調査

学習データは、国総研が有する管渠劣化データベース(以下、DB)を用いた。DB は全国の複数の自治体で実施した管路内調査(目視・TV カメラ調査)結果を収集したものである。

データ数は全部で 365,782 スパンあり、データの属性の欠損等を精査して、ここでは 97,916 スパンのデータを用いた。

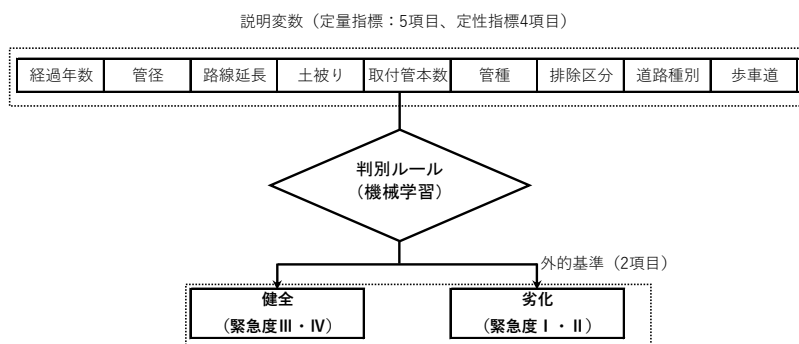


図-1 劣化箇所予測モデル

3. 学習データの作成方法

管渠の説明変数と外的基準を設定したうえで学習データを作成した(図-1 参照)。

(1) データ加工

1) 正規化 (0～1 の範囲に圧縮する)

説明変数のうち、経過年数や路線延長の定量指標は、指標間のスケールの影響を排除すべく正規化する。正規化を実施する際には、説明変数毎にそれぞれの頻度分布を確認したうえで最小値、最大値を設定する。

2) 類型化 (設定した条件に該当するものを 1、該当しないものを 0 とする)

説明変数のうち、管種や排水種別などの定性指標は類型化する。なお、定量指標も類型化した場合、モデルによる判別精度が向上することもあるため、全ての説明変数を類型化した場合のデータも作成する。

なお、定量指標について類型化をするときには、指標の特性を踏まえて区分を設定する。

(2) クラスタリング・ダウンサンプリング

学習データは、健全であるデータ数が劣化であるデータ数に比べ圧倒的に多く、これをそのまま学習データにすることは適切でなく、健全と劣化のデータ数を同数にするためのダウンサンプリングが必要である。

ダウンサンプリングは、クラスター分析(ここでは、k-means 法)を用いて、予めデータをクラスタリングしたうえで、クラスター毎に健全と劣化のデータを同数抽出した。

表-1 学習用データ作成の結果

項目	使用したデータ	ダウンサンプリング			学習用データ	
		健全	劣化	計	訓練データ	テストデータ
ケース①	97,916	17,719	17,719	35,438	24,800	10,638

4. 機械学習の性能評価

公共インフラに係る問題解決に AI(Artificial Intelligence)を用いる場合、学習結果の解釈や判別結果を説明できることに留意する必要がある。本研究では、機械学習のアルゴリズムに基づくモデルであり、それら留意事項に対応できる決定木、サポートベクターマシン(SVM)、ランダムフォレスト(RF)を採用した。

(1) 交差検証

性能評価は図-2 に示す通り、交差検証(クロスバリデーション)により実施した。交差検証は、作成した学習データのうち 70%を訓練データとし、残りの 30%をテストデータとする。訓練データは 5つのブロックに分割し、1つのブロックを検証データ、残りの 4つのブロックを学習データとする。検証データとするブロックを変えながら 5回にわたるイテレーションを通して、キャリブレーションによるパラメータ決定と性能評価の手続きを繰り返す。最適パラメータによるモデルを用いてテストデータを用いた性能評価を実施した。

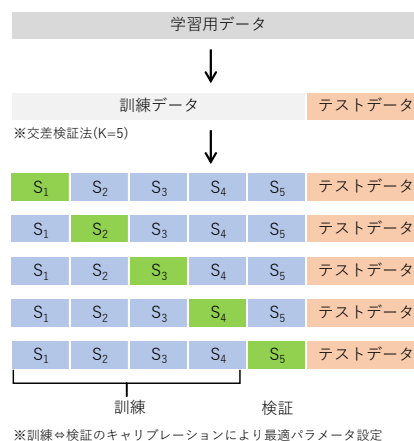


図-2 交差検証の概念図

(2) テストデータに対する性能評価

性能評価をした結果、説明変数を全て類型化したケースが、決定木、SVM、RF 共に良好であった。各学習に大きな差は見られないものの、決定木より SVM と RF が良好であった。表-2 の通り、経過年数別に評価すると、全データの F 値で約 67%であるものの、40 年以上を経過したデータでは 70%程度確保できており、今後、老朽化管が増加することを考慮すると本研究で構築した劣化予測モデルは十分な性能であると考えられる。

表-2 経過年数別の性能評価の結果(SVM、イテレーション1のテストデータ結果)

経過年数	データ数					Precision 適合率	Recall 再現率	Specificity 特異率	F-measure F値	accuracy 正解率
	TP	FP	FN	TN	計					
30 年未満	489	839	197	506	2,031	36.82%	71.28%	37.62%	48.56%	48.99%
30 ~40年未満	1,977	1,718	161	292	4,148	53.50%	92.47%	14.53%	67.79%	54.70%
40 ~50年未満	1,725	1,270	182	212	3,389	57.60%	90.46%	14.30%	70.38%	57.16%
50 ~60年未満	239	179	32	44	494	57.18%	88.19%	19.73%	69.38%	57.29%
60 年以上	370	156	24	26	576	70.34%	93.91%	14.29%	80.43%	68.75%
計	4,800	4,162	596	1,080	10,638	53.56%	88.95%	20.60%	66.86%	55.27%

この他、都市別に学習データを設定した場合、外的基準を腐食ならびにタルミ等の調査項目別に置き換えた場合の検討を加えたものの、全データの F 値は概ね 60~70%程度であった。これより、本研究では少ない説明変数で F 値の最も高いケースを劣化箇所予測モデルの学習データとして採用している。

5. 効果の検証

A 市における管路内調査が完了している地区(約 7,000 スパン)を対象に、劣化を把握するための調査に係る費用を試算した結果、本研究で構築した劣化予測モデル(SVM)を用いた場合、これまでの調査に比べ 10~25%の費用削減が確認された。