

観測データ間の関係を考慮した決定木による構造物の補強度合の判定

○増井 遼太 辻岡 竣祐 巽 啓司 (大阪大学) 楠木 祥文 (大阪府立大学) 尹 禮分 (関西大学)

Determination of Reinforcement Degree of Structures by Decision Trees Based on the Relationships between Observed Data

*R. Masui, S. Tsujioka, K. Tatsumi (Osaka University), Y. Yun (Kansai University) and Y. Kusunoki (Osaka Prefecture University),

Abstract— In recent years, there has been concern about the decrease in the labor force due to the declining birthrate and aging population, and the shortage of manpower has become a problem in the field of infrastructure development and maintenance as well. In this paper, we apply the decision tree to a problem of determining the reinforcement degree in constructing large-scale structures, which requires the experience and skills of experts. Thus, in this paper is formulated as a recognition problem to determine the degree of reinforcement of a structure by using potential relationships between data observed at a construction site. In addition, we also show how to resolve the inconsistency between the data. Moreover, we apply the decision tree method to real data of some large-scale structures in order to evaluate the formulation and the validity of the proposed method.

Key Words: Reinforcement of the structures, Decision tree, Recognition problem

1 はじめに

近年、少子高齢化に伴う労働人口の減少にともない、インフラ建設・整備や維持管理においても人手不足が深刻な問題となっている。加えて、年間総実労働時間が長く危険を伴う作業が多いなどの理由により、震災に伴う復興や東京五輪などの建設需要が高まっているにもかかわらず、20代、30代の若年建設業従事者数が半減している。また、今後20年で、50年以上経過した道路橋、トンネル、下水道などのインフラ施設の割合が急激に増加することが指摘されている一方¹⁾、インフラ設備の点検は近接目視により5年に1回、健全性評価が義務付けられている²⁾。インフラの建築・整備での危険度チェック等の判断を、AI (Artificial Intelligence) の導入により自動化・知能化する動きが見られる。大手ゼネコンでは、建設現場での専門家によるインフラの安全性評価を代替する形で、Deep Learning の手法を用いて、デジタル画像からのトンネルの掘削面の評価やコンクリート構造物のひび割れ検出を行うシステムを開発している⁴⁾。

本稿では、著者らが近年取り組んでいる、インフラ建設時の補強度合選択を機械学習を用いて置き換える研究について紹介する。著者らは、まず、補強度合の識別問題に、各々の構造物データごとに学習・テストを行い、高い識別率が得られることを確認した⁵⁾。また、観測ベクトルと補強度合の関係の解析から、観測ベクトルは同じで補強度合が異なる識別不可能データの存在を確認し、その矛盾を合理的に解消する機械学習の枠組みを提案している⁶⁾。本稿ではそれらの結果を紹介するとともに、多数の建造物データを用いた事前学習により作成した識別器が、未知の建造物の補強度合の識別にも有効であることも示す。

2 大型構築物建設時の補強度合決定

大型構築物における補強度とは、トンネル建設時であれば、上または横からの荷重を支える構造物である

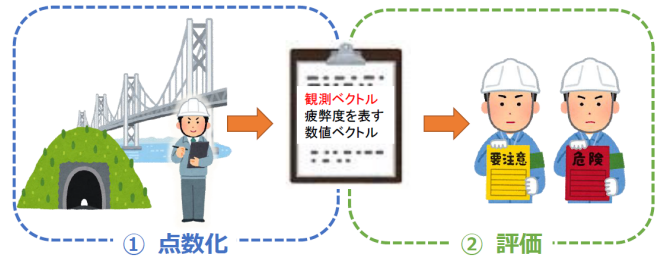


Fig. 1: 構造物建設時の2段階補強度合決定プロセス

支保工パターンの度合を表し⁹⁾、橋梁や道路であれば、補強・補修の必要性の程度を表すものとする⁸⁾。大型構築物の補強度決定は、図1にあるような2段階の判断過程により行われている。第1段階では、熟練者がインフラ建設状態に対して目視などの観察を行い、概ね30程度の項目ごとの危険度を数値ベクトルとして表す。この工程については、前述の通り、画像認識やDeep Learning などを用いたAI化が進められている⁴⁾。第2段階では、その観測ベクトルを基にして熟練者が最終的な補強度合を決定している。著者らは、比較的研究が行われていないこの後半部分の自動化を目指している。

大型構築物の建設時は、構造物全体を適当な区間に分割し、その区間ごとに逐次的に作業が行われるため、上記の補強度合も、各区間ごとに熟練者による観測により得られる観測ベクトルと、その区間の自然状況(立地環境や岩質など)を考慮して決められている¹⁾。

研究⁵⁾では、対象とする構造物の各区間で、観測ベクトルおよび自然状況に対しその区間の補強度合を予測する問題として定式化している。

¹⁾本来、各構築物ごとに、各々、適切な用語が用いられるが、本稿では、大型構築物一般の建設時の状況を一般的に述べるためこのような抽象的な表現を用いる。

区間 i での m 次元観測ベクトル

$$\mathbf{x}^i = (x_1^i, x_2^i, \dots, x_m^i)^\top \in \mathbb{R}^m$$

に対し、教師信号として、その区間 i での補強度合

$$y_i \in \{1, 2, \dots, l\}, i \in I := \{1, \dots, n\}$$

が与えられる。この値が大きいほど強い補強が必要になる。通常、各評価項目に対する観測ベクトル要素 x_j^i が大きいほど、補強度合のラベル値 y_i も大きくなると考えられる。

また、各区間での自然状況 r は、 \bar{r} 個の段階 $R := \{1, \dots, \bar{r}\}$ が存在し、構造物は、ほとんど同じ自然状況の段階のもとで建設される場合も、様々な異なる段階のもとで建設される場合もある。本稿では、構造物を建築する際の観測ベクトル、自然状況、補強度合を組み合わせて構造物データと呼ぶこととする。

評価マニュアルによると熟練者による補強度合の判断の際は、区間 i の自然状況 $r(i) \in R$ に応じて \mathbf{x}^i の各要素に対する重みベクトル $\mathbf{d}^r = (d_1^r, \dots, d_m^r)^\top$, $r \in R$ がこれまでの経験に基づいて調整されており、その重みを用いたスカラー評価点 $\sum_{j=1}^m d_j^{r(i)} x_j^i$ を参考にして補強度合が決定されている。例外として、構造物の端の区間に施される特殊な補強度合は、上記の値とは別の要因で決定される。本稿では、この補強度合の予測は行わない。

理想的な状況では、このスカラー値のみにより、補強度合ラベルが一意に定まることが期待できる。実際の構造物データでは、スカラー値が同じにもかかわらずラベルが異なるような状況が比較的多数存在し、そのような場合は、熟練者が経験に基づき、観測ベクトルの要素も参考にして決定する必要がある。Fig. 2 は、5 章で使用するデータのうち、構造物データ 1, 13, 31 の全区間 $i \in I$ のデータに対する上記スカラー評価点とラベルの分布を表している。この例からも、同じスカラー評価点でラベルが異なる区間が多数あることがわかる。

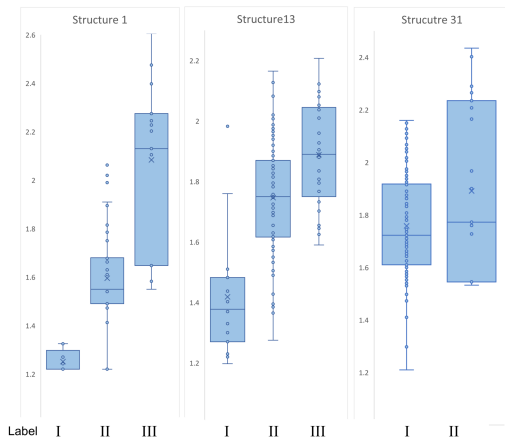


Fig. 2: 3つの構造物データでのスカラー評価点

3 補強度合識別問題の定式化

本研究の目的は、新しい構造物の観測ベクトル \mathbf{x} から適切な補強度合を出力するような識別関数

$$f: \mathbb{R}^m \rightarrow L := \{1, 2, \dots, l\}$$

を機械学習を用いて作成することである。このような m 次元観測ベクトルを入力として、補強度合を判別する問題を識別問題 (P1) と呼ぶことにする。また、(P1) での入力として、観測ベクトル \mathbf{x}^i のみを用いた定式化と、自然状況の重みも考慮したベクトル

$$\mathbf{d}^r \otimes \mathbf{x}^i = (d_1^r x_1^i, \dots, d_m^r x_m^i)$$

を用いるもの 2 通りを考える。前者を (P1a)、後者を (P1b) とする。

この識別問題に対して、著者らはパターン分類手法の一種であるサポートベクトルマシン (SVM) を適用した⁵⁾。データ構造を考慮した 2 種類の SVM を 3 つの異なる構造物データに適用し、汎化性検証を行っている。その際、多項式カーネルおよび RBF カーネルを用いる。検証の結果、おおむね 80 % 以上の比較的高い識別率が得られていると報告されている⁵⁾。しかし、カーネルを使用しているため、複雑で解析が困難になる問題がある。

そこで、本研究では、上記の識別問題 (P1) に対して、機械学習分野の代表的な教師あり学習法である決定木を適用することを考える⁷⁾。その際、決定木のアルゴリズムには CART 法を用いる。CART 法では、各節点で説明変数に対して閾値を選択し分類を行う。その際、分類条件に用いる説明変数は、分類前後での不純度の減少量で決定される。決定木で得られる識別関数は木構造で表現されるため、カーネルを用いた SVM よりも解釈が容易であることがあげられる。CART 法は、分類木と回帰木の両方に対応しており、目的変数に離散値と連続値を扱うことができる。また、2 分木を用いたアルゴリズムであるため、木の構造がシンプルになり結果が解釈しやすい利点がある。

4 識別不可能データを考慮した再定式化

本研究で使用する 28 個の異なる構造物データでの m 次元観測ベクトル・補強度合のデータを解析したところ、観測ベクトルは同じだが補強度合が異なる識別不可能データが 10 個の構造物データに多数含まれていることが分かった。このことは、本質的に識別率の上限となっており、識別問題 (P1) の定式化では不十分であると考えられる。各構造物データにおいて、識別不可能データの全データに対する割合を表 1 に示す。全データに対する割合が 30 % 近い構造物データもあり、無視できない割合であることがわかる。本研究では、この矛盾を合理的に解消するための機械学習の枠組みを提案する。

実際に、区間 i において、熟練者が補強度合を決定する際、区間 i の m 次元観測データだけでなく、それ以前の区間における建設済みの補強度合も考慮して補強度合を決定していると考えられる。この矛盾の解消のため、 m 次元観測ベクトルに加えて、対象とする区間の建設以前に建設済みの区間での情報を、以下のベクトルを追加入力とした再定式化を考えた。

1. 対象区間の 1 つ前の区間の補強度合と同じ補強度合ラベル連続数を要素とするベクトル

$$\mathbf{s}^i = (s_1^i, \dots, s_{\ell+1}^i)^\top, i \in I$$

Table 1: 識別不可能データの全データに対する割合 (%)

構造物番号	(P1)		(P2a)		(P2c)	
	識別不可能データの割合 [%]	誤識別データの下限割合 [%]	識別不可能データの割合 [%]	誤識別データの下限割合 [%]	識別不可能データの割合 [%]	誤識別データの下限割合 [%]
構造物 1	31.08	8.11	0.00	0.00	9.46	2.70
構造物 2*	27.78	1.11	0.00	0.00	0.00	0.00
構造物 3*	9.04	3.14	0.00	0.00	0.00	0.00
構造物 8*	5.51	1.20	0.34	0.17	0.00	0.00
構造物 9	1.20	0.40	0.00	0.00	0.00	0.00
構造物 10	3.19	1.06	0.00	0.00	0.00	0.00
構造物 14	17.24	3.98	0.00	0.00	1.33	0.53
構造物 16	6.76	2.01	0.00	0.00	0.00	0.00
構造物 20*	8.57	3.33	0.00	0.00	0.00	0.00
構造物 23*	8.33	4.17	0.00	0.00	0.00	0.00

2. 対象区間より前の3区間での評価点を要素とするベクトル

$$\mathbf{v}^i = (v_1^i, v_2^i, v_3^i), i \in I$$

上記の追加入力を、それぞれ連続ラベル数、過去評価点と呼ぶこととする。連続ラベル数は、特殊な補強度合についても考慮する。

さらに、その区間での自然状況 $r \in R$ も更なる追加入力として考える。このようにして得られた入力を用いて補強度合を判別する問題をここでは識別問題 (P2) と呼ぶこととし、連続ラベル数を用いた問題を (P2a)、過去評価点を用いた問題を (P2c) とする。

この識別問題 (P2) での識別不可能データの割合を表 1 に示す。すべての構造物のデータに関して、識別不可能データの数を大幅に減らすことができている。つまり、過去の区間での情報を考慮に入れることで、その区間だけでは区別できなかった補強度合が、適切に扱えるようになったと考えられる。

5 数値実験結果

ある大型構造物の 28 個の構造物データを用いた識別問題 (P1), (P2) に対し、決定木を適用し、モデル化の違いによる得られる識別器の性能を比較する。

本稿では、既知のデータにより学習した識別器を、新しい構造物データに適用した際の識別性能について検証するため、1つの構造物データをテストとして残りを学習として用いることで汎化性能を検証する。このような学習、テスト方法を統合識別と呼ぶこととする。統合識別を識別問題 (P1), (P2) として定式化を行い、決定木を適用し数値実験を行った。定式化する際、情報量過多による決定木の複雑化を避けるため、各構造物データにおいて隣接する補強度合クラスを併合させ、2クラス分類を行った。識別器の性能比較は、全区間での識別率をもとに行う。また、区間ごとに逐次的に作業する場合、補強度合が変化する区間を判断することが重要であると考えられるため、補強度合の変化区間での識別率にも着目する。補強度合の変化区間をラベル変化区間と呼ぶこととする。本実験で使用する計算機の環境は Table 2 に示す。テスト識別の際の 10 個の構造物データでの全区間での識別率とラベル変化区間での識別率を Table 3, 4 に示す。それぞれの構造物データで得られた識別率を比較し、最も高い識別率を青字で、2番目に高い識別率を赤字で示す。変化数は、補

Table 2: 計算環境

CPU	Core i7-9750H, 2.60GHz
メモリ	8.00GB
OS	Windows 10
数値解析ソフトウェア	Python (ver3.8)

Table 3: 構造物データの識別率 (%)

構造物番号	(P1a)	(P1b)	(P2a)	(P2c)
構造物 4	73.30	78.70	94.70	78.70
構造物 8	81.80	82.80	98.60	84.00
構造物 9	83.50	70.30	98.40	86.30
構造物 10	79.10	77.00	97.90	91.50
構造物 11	87.50	70.00	95.00	87.50
構造物 13	76.30	81.30	95.70	81.30
構造物 20	83.30	71.40	97.60	80.00
構造物 28	88.90	75.60	95.60	88.90
構造物 30	95.20	86.20	98.80	94.00
構造物 31	87.20	88.00	97.00	86.50
平均	83.61	78.13	96.93	85.87

強度合が変化する回数、追従数はそのうち正しく識別できた回数を示している。Table 3 から、決定木を適用した (P1), (P2) において 7 割から 9 割ほどの識別率が得られていることがわかる。特に、(P2a) では識別不可能データの有無にかかわらず、(P1) において 7 割から 8 割ほどしか識別できていなかった全ての構造物データにおいて、9 割以上の精度で識別できていることがわかる。しかし、Table 4 から、ラベル変化区間ではすべての構造物において識別出来ていないことが分かる。対して、(P2c) に関しては、全区間での識別率は (P1) と比較して、同程度の識別率が得られている。しかし、10%以上向上しているものがあることが確認できる。ラベル変化区間での識別率も比較的高い精度で識別出来ていることがわかる。これらの結果より、入力に過去情報を追加することは、有効な定式化であることがいえる。

(P1) のなかでは、全区間での識別率で、(P1b) が 5% ほど低い識別率が得られている。これは、構造物データごとに、その区間内に現れる自然状況や、その組み合わせが大きく異なっており、異なる特性を持ち、 \mathbf{x}^i の各要素に対する重みベクトル \mathbf{d}^r では、その特性を考慮できていないためだと考える。

