

局所外れ値因子法に基づく打音による鉄筋腐食判定に教師データへの誤ラベルデータの混入が与える影響

防衛大学校 学生会員 ○森藤 優一, 正会員 黒田 一郎

1. はじめに

近年、コンクリート構造物の非破壊検査手法の一つの打音法にAI技術の応用が試みられており、多くの研究が積み重ねられつつある。その中でも、局所外れ値因子法(以下、LOF)は、陽性・陰性の分類器として用いる場合には教師データに陰性データのみを要するという利点を持つ。陰性データとしては、構造物の新造時に収集したデータを活用するのが合理的であるが、既存の構造物を想定すれば、新造時の打音データが存在しない。その場合は、既存構造物の陰性と推定される部位で教師データを収集せざるを得ないが、それらには誤ラベルデータである陽性データの混入の可能性が無視できない。そこで、本研究では、LOFに基づく鉄筋腐食判定において、教師データへの陽性データの混入が判定結果に及ぼす影響について考察した。

2. LOFによる鉄筋腐食判定

LOFの概念図を図-1に示す。教師データ群全体の中から、テストデータA(図中の緑のマーカー)からの距離が近い順に近傍数 k 個だけの教師データ Q が選択され、テストデータAの局所密度 LD_A はA-Q間の到達可能性距離の平均値の逆数で表わされる。また、 k 個選択された教師データ Q も同様に、教師データ Q の近傍の k 個の教師データ R (図中の赤のマーカー)間との局所密度 LD_R を Q - R 間

の到達可能性距離の平均値の逆数で表わすことができ、 LD_R を LD_A で除することによりテストデータのLOF値が算出される。このLOF値が、設定した閾値 T_h よりも小さいならば、陰性とされ、大きいならば陽性とされる。分類器としてのLOFは、本来、教師データに陰性の特徴量のみを必要とするものであるが、陽性データが混入することが避け難いと考え、教師データの一部を陽性データで置き換えた場合の判定結果を考察した。

3. 入力データの収集と判定ケース

LOFで使用する入力データを収集するために表-1の示方配合条件で図-2の諸元のRC供試体を8体作成し、全供試体の内部鉄筋に対して目標腐食率(以下、腐食率)を1%に設定して電食に供し、表-2の手順で打音収録を実施した。打音収録では供試体A面(図-2)中央の $60\times 60\text{mm}$ の範囲を手動インパクトハンマで打撃した。A面直上300mmに設置したマイクロフォンにより収録したサンプル長さ40msの音圧時刻歴を、フーリエ変換によって周波数解像度25Hzのベクトル量であるスペクトル(帯域500Hz~10000Hz)に変換した。そのスペクトルをLOFの入力データとして、陰性である腐食前のデータ80個と、陽性である腐食率1%のデータ80個を用意した。

表-3に判定ケースを示す。4種の全ケースで、陰性のデータ80個から無作為に選抜された40個と、陽性のデータ40個の計80個のデータを判定対象のテストデータとし、教師データはテストデータと重複しないように選択した。教師データ数 N は全ての判定ケースで40個とし、教師データの状態 N_{st} を陰性のデータのみからなる「無混入」と、データ混入数 $N_m(=8)$ 個を陽性のデータで置換した「混入」の2段階に設定した。近傍数 k は2, 8, 20の

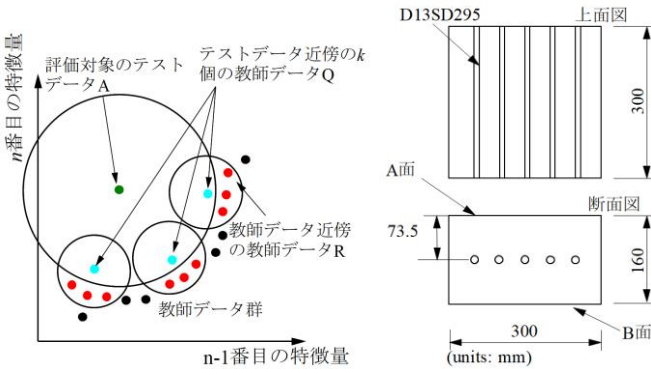


図-1 LOFの概要

図-2 供試体諸元

表-1 示方配合条件

粗骨材 最大寸法 G max (mm)	水セメ ント比 W/C (%)	空気量 (%)	単位量(kg/m ³)				
			水 W	セメント C	細骨材 S	粗骨材 G	AE剤 A
20	60	3	175	292	680	1060	3.5

表-3 判定ケース

判定ケース	状態	教師データ				近傍数 k	試行回数
		総数 N_{st}	陰性 データ数 N_N	陽性データ 混入数 N_m			
mix0-k2	無混入	40	40	0		2	5
mix8-k2	混入	40	32	8		2	5
mix8-k8	混入	40	32	8		8	5
mix8-k20	混入	40	32	8		20	5

表-2 データ収集手順

供試体番号	実験経過	
	腐食率 0% (電食前)	腐食率 1%
腐食供試体 No.1~8	打音収録	電食 → 打音収録

キーワード 非破壊検査, 打音法, 局所外れ値因子法, 鉄筋腐食, ROC 曲線, AUC 値
連絡先 〒239-8686 神奈川県横須賀市走水 1-10-20
防衛大学校 理工学研究科 TEL046-841-3810

3段階に設定し、近傍数 k が2での教師データの状態 N_{st} の違いの比較と、教師データに陽性データが混入した状態(N_{st} = 混入)での近傍数 k ($k=2, 8, 20$)の影響を調べるために、計4つの判定ケースを定めている。

なお、テストデータと教師データの選抜の影響を取り除くために無作為選抜をやり直して、各判定ケースで5回ずつの試行を行っている。

4. 判定結果

LOFによる判定結果の例として、真陽性率 TPR と偽陽性率 FPR の関係を表現した判定ケースmix0-k2のROC(受信者操作特性)曲線を図-3に示す。ROC曲線では真陽性率 TPR が1.0であると同時に偽陽性率 FPR が0.0である図中の左上の隅の位置(以下、左上隅)が一切の誤謬が生じていない理想的な判定結果であり、図-3の判定ケースmix0-k2は理想的な判定結果に近いものである。ROC曲線は以上のような定性的な評価しかできないため、定量的な評価を試みるため AUC 値を用いて各判定ケースの判定結果を評価する。 AUC 値は、ROC曲線と横軸に挟まれた部分の面積であり、この値が1.0であれば閾値を適切な値に設定すれば真陽性率 TPR を1.0にすると同時に偽陽性率 FPR を0.0にすることが可能であり、その際には正解率は1.0を実現可能である。

図-4に各判定ケースと AUC 値の関係をまとめる。図中の黒いドットは5回の各試行の AUC 値であり、白抜ききの着色された四角いマーカーはそれらの平均値である。 AUC 値は、判定ケースmix0-k2($k=2, N_{st}$ = 無混入)では1回の試行を例外として0.9以上であるが、判定ケースmix8-k2($k=2, N_{st}$ = 混入)では0.8を超す AUC 値は無く、平均値は0.65にとどまっております、誤ラベルデータの混入が学習器の性能を大きく損な

っている。そこで、近傍数 k を陽性データ混入数 N_m と同数にした判定ケースmix8-k2($k=8, N_{st}$ =混入)では、 AUC 値は、全ての試行で判定ケースmix0-k2($k=2, N_{st}$ = 無混入)と同値となるまで判定結果が改善しているが、近傍数 k を、さらに大きくした判定ケースmix8-k20($k=20, N_{st}$ =混入)では、教師データ数 N に対する近傍数 k の割合である k/N 比が大きくなる影響を受け、 AUC 値は低減した。

図-5は、各モデルの閾値 T_h と正解率 ACC の関係をまとめたものであり、図中の各モデルの正解率の推移を示す折れ線は、各5回の試行の正解率 ACC の平均値 ACC^* を示しており、各判定ケースの正解率平均値 ACC^* の最大値は AUC 値と同様に混入の影響を受ける。また、図-6は、図-5の縦軸の0.8以上を拡大表示した図である。正解率平均値 ACC^* の最大値は、mix8-k2($k=8, N_{st}$ =混入)が判定ケースmix0-k2($k=2, N_{st}$ = 無混入)に若干勝るものの、0.8以上の正解率を維持できる閾値 T_h の幅は、教師データの状態 N_{st} が無混入の方が明らかに大きい。教師データに陽性データが混入することで設定できる閾値 T_h の範囲は狭くなるものの、近傍数 k を陽性データ混入数 N_m と同数以上にすることで ACC^* の回復は可能となる。その際には、 k/N 比の増大に伴う悪影響に注意が必要である。

5. まとめ

本研究では、LOFに基づくRC供試体の鉄筋腐食判定において教師データへの誤ラベルデータの混入が判定結果に及ぼす影響について検討した。その結果、教師データに近傍数 k よりも大きい陽性データ数が混入することで学習器としての性能が低下するものの、近傍数 k を調整することで性能は回復することを明らかにした。

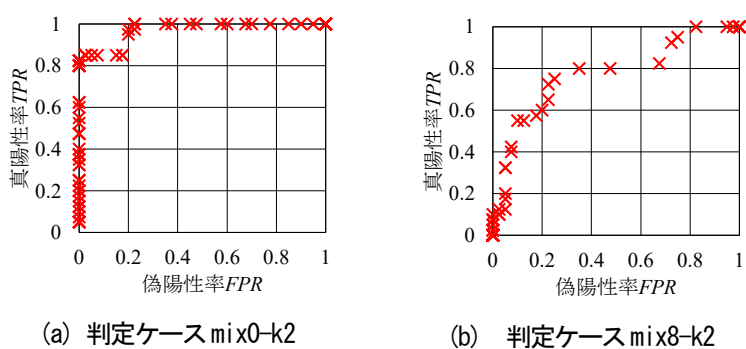


図-3 ROC 曲線の例

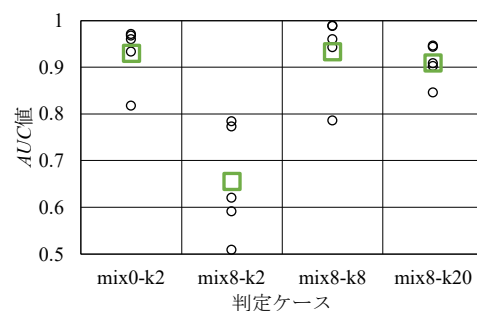


図-4 各判定ケースの AUC 値

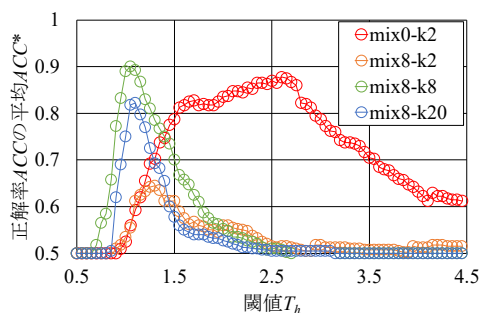


図-5 各判定ケースの ACC^*

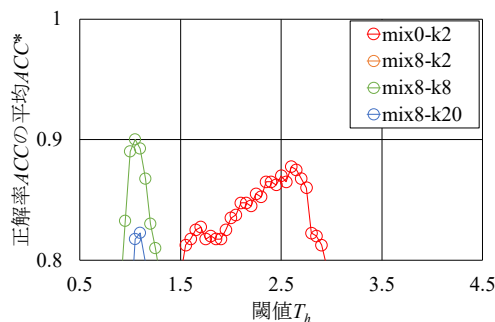


図-6 各判定ケースの ACC^* (拡大表示)