機械学習による降雨時の高速道路盛土のり面の変状発生予測

秋田大学　学生会員　○佐々木 翼　　正会員　荻野 俊寛
ネクスコ・エンジニアリング東北(株)　正会員　村山　暢

1. 背景

近年，降雨の量や強度の変化によって表層崩壊などの，のり面変状の頻度が増加している．このような変状の発生には各のり面が潜在的に有している変状に対する健全度(素因)と崩壊を誘発する降雨(誘因)の 2 つの要因が関係していると考えられる．本研究はこれら 2 つの要因を特徴量とした機械学習を用いてのり面の変状発生予測を行う．

2. 用いた機械学習アルゴリズム

機械学習によってのり面崩壊を予測するにあたり，本研究ではサポートベクターマシン(SVM)学習器および，SVM の 1 つである One Class SVM 学習器を用いた．SVM は代表的な教師あり学習アルゴリズムの 1 つであり，対象を区別し，線的あるいは面的に境界線を引く手段である．One Class SVM は SVM を外れ値検出(異常検知)に応用したものである．SVM では 2 つのクラスのデータを学習させ，分離超平面を引くことで崩壊・未崩壊を分類するが One Class SVM では 1 つのクラスの正常データを学習させ，分類境界を決定することで，その境界を基準に外れ値を検出する．本研究におけるのり面変状のように，異常がほとんど発生せず，異常クラスのデータが集まらないような場合に有効な異常値検知手法である．SVM ではハイパーパラメータと呼ばれるいくつかの変数を，分類精度が最大となるよう最適化する必要がある．SVM ではカーネルスケール s とボックス制約 C が，One Class SVM では s とサポートベクターの数に関係する係数 ν がハイパーパラメータとなる．

表-1　機械学習に用いた特徴量

属性	特徴量	次元数
素因	のり面潜在危険スコア	1
誘因	各タンク貯留量のピーク時刻 $(t_{p_1}, t_{p_2}, t_{p_3}, t_{p_4})$	4
	$t_{p_1}, t_{p_2}, t_{p_3}, t_{p_4}$における各タンク貯留量の値 (4×4=16)	16
	$t_{p_1}, t_{p_2}, t_{p_3}, t_{p_4}$における解析雨量の値 (4×1=4)	4
素因＋誘因	各タンク貯留量ピーク値の履歴順位	4
	各タンク貯留量ピーク値の偏差	4

表-2　パラメータの範囲

C	1, 1.7, 2.8, 4.6, 7.7, 12.9, 21,5, 39.9, 59.9, 100
ν	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0
Outlier Fraction	0.001, 0.002, 0.003, 0.004, 0.005, 0.01, 0.02, 0.05
Kernel Scale	10^{-1}, $10^{-0.75}$, $10^{-0.5}$, $10^{-0.25}$, 10^{-0}, $10^{0.25}$, $10^{0.5}$, $10^{0.75}$, 10^{1}

3. 機械学習に用いるデータの準備

対象としたのり面は NEXCO 東北支社管内の盛土のり面 16258 ヵ所のうち，2011 年までに崩壊したのり面 94 ヵ所である．機械学習に用いた特徴量を表-1 に示す．特徴量は最大で 33 次元であり，素因，誘因およびその両方に関する特徴量に分類できる．これらの特徴量の一部あるいは全部を学習に用いることで汎化性能の変化を検討した．素因には各のり面の潜在危険性の評点(素因スコア)[1]を用いた．

誘因の降雨履歴の抽出には土壌雨量指数を用いた．土壌雨量指数は 3 段直列タンクを用いた雨量指数で，各タンク貯留量の和が土壌雨量指数となる[2]．解析雨量から計算した土壌雨量指数および各タンク貯留量のピークを抽出し，降雨履歴と定義し，ピーク値を特徴量として学習に用いた．また，既存資料に記載された変状発生日に最も近い降雨履歴を崩壊時の降雨履歴とした．

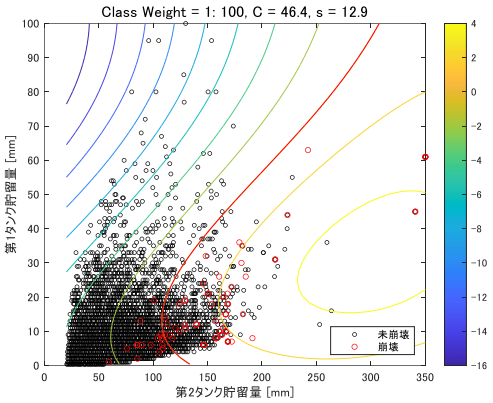

図-1　代表的な SVM の結果(図中の赤線が分離超平面．コンターが正の範囲は崩壊領域，負の範囲は未崩壊領域を表す．)

4. モデルの学習

極端に不均衡なデータセットの場合，SVM では少数派クラス(崩壊時)のデータは非常に重要である．本研究では全てののり面に対して 1 つの SVM モデルを作成することで，崩壊時のデータを全て使用した．一方，One Class SVM では崩壊時のデータは不要であるため，のり面ごとに未崩壊時のデータのみを用いてモデルを作成することができる．のり面ごとにモデルを作成することで，必要なのり面に対しより適切な分類境界を作成できると考えられる．本研究では代表的な 100 のり面についてモデルを作成した．いずれの SVM もカーネル関数にはガウス関数を用いた．ハイパーパラメータの最適化は過学習の防止のため 5 分割クロスバリデーションを採用し，探索範囲内で全ての組み合わせを総当たりで探索するグリッドサーチを用いた．設定したパラメータの探索範囲を表-2 に示す．

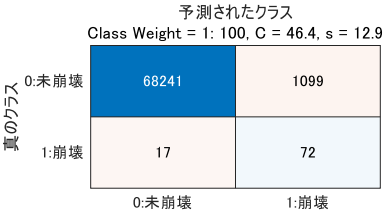

図-2　図-1 に対する混同行列

キーワード　盛土のり面　解析雨量　機械学習
連絡先　〒010-8502　秋田県秋田市手形学園町　1-1　TEL 018-889-2364

5. 汎化性能の比較

学習に第 1 および第 2 タンク貯留量のみを特徴量として用いた場合の代表的な SVM の結果を図-1 に示す．崩壊データの多くは崩壊領域に，未崩壊データの多くは未崩壊領域に存在し，分離超平面によっておおむね適切に分類されていることがわかる． しかしながら，未崩壊領域に存在する崩壊データ，崩壊領域に存在する未崩壊データもあり，一定数の誤分類が確認できる． 図-2 は分類結果の数を示した混同行列である．ここで，汎化性能の指標として，崩壊誤検知率 G を式(1)で定義する．

$$G = \frac{B}{A+B} \qquad (1)$$

ここに，A は真のクラス，予測されたクラスともに未崩壊のデータ数，B は真のクラスが未崩壊であり，予測されたクラスが崩壊であるデータの数である． 図-2 から G を計算すると，G=1099/(68241+1099)=1.6%となる． 特徴量の次元数を増加させて学習を行った場合の誤検知率の変化を図-3 に示す． 誤検知率は特徴量の次元が増えるにしたがって改善し，全ての特徴量を使用した 33 次元のとき，0.9%となった．

次に，One Class SVM で同様に特徴量の次元数を増やした時の誤検知率の変化を検証した． 始めに第 1 および第 3 タンク貯留量のみを特徴量として用いた場合の代表的な分離超平面を図-4 に示す． 図-4(a)は誤検知率が低いのり面の代表例である． 未崩壊データのほぼ全てが未崩壊領域に存在しており， 崩壊データは崩壊領域に分類することができている． 一方，図-4(b)は誤検知率が高いのり面の代表例である． 分離超平面が複数出現し， 明らかに過学習となっており，のり面によって，学習結果に差があることがわかる． 2 次元で学習させた場合の平均誤検知率は 18.6%で， SVM よりも汎化性能が悪かった． 図-5 は図-4(a)におけるパラメータの変化に伴う誤検知率の変化を可視化したものである． Kernel Scale およびνを小さくするほど誤検知率が上昇することが分かる． また， Outlier Fraction を小さくするとわずかながら誤検知率は低下する． SVM と同様に次元数を表-1 のように 4， 24， 32， 33 と増加させて学習を行った場合の誤検知率の変化を図-3 に示す． 誤検知率はどの次元数でも SVM の方が低く，さらに，次元数を増加させると， One Class SVM では平均誤検知率は上昇し，33 次元のとき，64.77%となった． 機械学習では特徴量の次元数を増やすほど， 学習に用いるデータ数も増加させることが望ましいが， のり面ごとにモデルを作成した One Class SVM では， 各のり面の学習データ数に限りがあるのにも関わらず特徴量を増やしたため誤検知率が悪化したと考えられる．

6. 結論

本報告では SVM および One Class SVM を用いてのり面変状の発生予測を行い，その汎化性能を比較した． 通常の SVM のモデルによる崩壊誤検知率は 1%以下となった． One Class SVM を用いるとのり面ごとにモデルを作ることができ，異なる分類境界を設定できるため，汎化性能の向上が予想されたが，各モデルの作成に用いるデータ数が減少するため，通常の SVM を用いてのり面全体で 1 つのモデルを作った場合の方が汎化性能は高い結果となった．

【参考文献】1) 長尾ら：被災のり面データに基づく東北地方の豪雨による高速道路のり面の崩壊素因の評価の試み，土木学会論文集 C，76(3)，pp. 235-253，2020. 2) 岡田ら;土壌雨量指数:日本気象学会．2001. 48 (5)， 349–356.

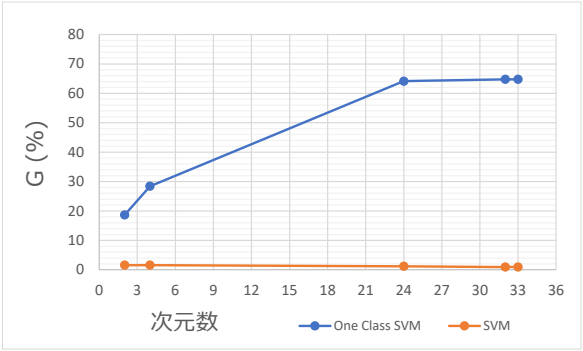

図-3 次元数の増加に伴う誤検知率の変化

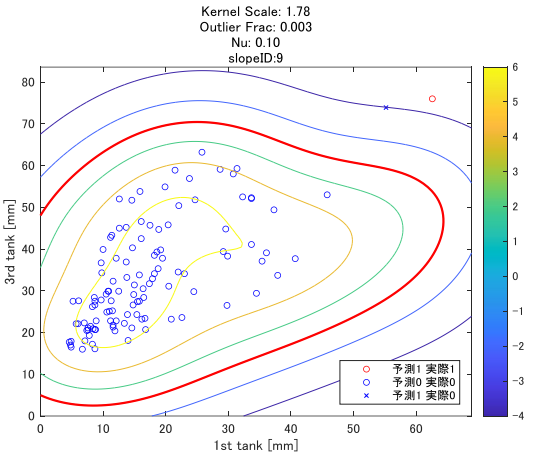

図-4(a) 代表的な One Class SVM の結果(図中の赤線が分離超平面．コンターが正の範囲は崩壊領域，負の範囲は未崩壊領域を表す．)

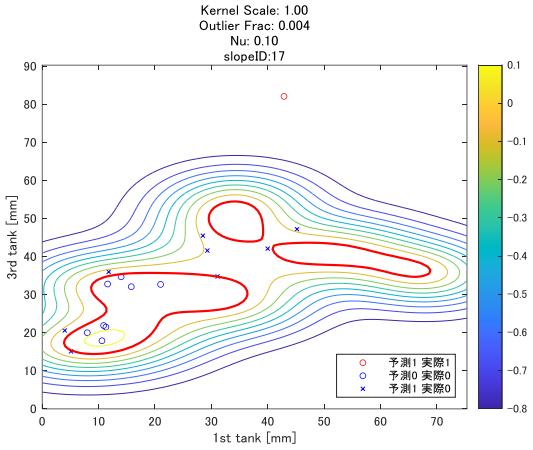

図-4(b) 代表的な One Class SVM の結果(図中の赤線が分離超平面．コンターが正の範囲は崩壊領域，負の範囲は未崩壊領域を表す．)

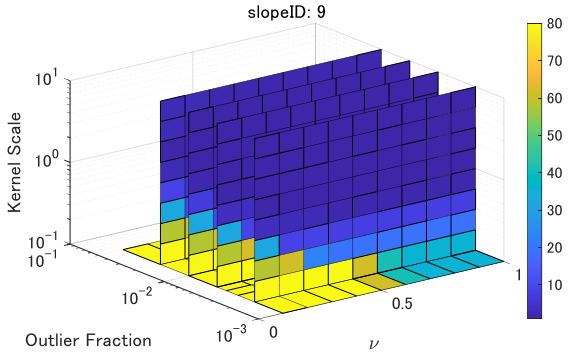

図-5 パラメータ最適化の代表例(2 次元)