

SVM を用いた土砂災害発生限界基準線の設定に関する研究

山口大学大学院 学生会員 ○江頭勲
山口大学工学部 正会員 大石博之
山口大学工学部 正会員 古川浩平

株式会社エイトコンサルタント 正会員 清家礼雄
香川大学工学部 正会員 尹 禮分

1. はじめに

現在、山口県では土砂災害の発生を予測し、警戒・避難を発令するための基準として、国土政策総合研究所指針に基づいた線形の土砂災害発生危険基準線 (Critical Line 以下 CL と表記) を設定している。しかし、現行の線形 CL は設定する上で客観性が確保できないことや、複雑な自然現象を線形で捉えようとするために生じる精度の低下が問題となっていた。このため、既往研究では RBF ネットワーク (Radial Basis Function Network : 以下 RBFN と表記) という、数学的手法を用いた非線形の CL が検討され、実用化へ向けての調整段階に入っている。しかし、RBFN ではパラメータの検討数が多いなどの問題がある。この問題を解決するため、本研究では、現在最も強力なパターン分類手法として注目されているサポートベクターマシン (Support Vector Machine : 以下 SVM と表記) を利用し、新たな非線形 CL の設定手法の確立を目指した。

2. 使用データの概要

本研究に使用する災害データは、全国的に土砂災害発生件数の多い山口県の下関市、秋吉台、山口市、防府市、岩国市を対象とし、2003 年度の調査データを引用して 1994 年から 2002 年までの災害データを使用した。また、非線形 CL の設定に用いる降雨データは、山口県の各気象台のアメダス観測所における 1994 年から 2002 年までの降雨データを使用した。

3. SVM 及び ν SVM の概要

SVM とは 1992 年に Vapnik らによって提案された手法である。SVM では、あるパターン分類問題が線形分離不可能な場合、ある非線形写像によって、線形分離可能な高次元空間にマッピングを行う。ここで、線形分離可能な状態とし、最適な分離超平面を求めることができる。これが基本的な考え方 (以下 2 群判別法と表記) である。

これに対し、1 群のデータ集団を、その分布状況の粗密から分類して 2 群に分離する SVM を ν SVM と呼ぶ。土砂災害の CL 設定への適用を考えた場合、上述の 2 群判別法は災害発生・非発生の両データが揃うことが必須であり、精度の高い CL を設定するためには、ある程度の数の災害発生データが存在することが条件となる。過去の被災履歴が少ない場合、1 群でもデータ分離が可能な ν SVM が有効な手法となってくる。本研究では、2 群判別法での SVM と ν SVM とによって CL の設定を行い、適用性の検証を行うこととした。

4. ν SVM のパラメータ検討

ν SVM で CL を設定する際のパラメータ ν は、全データのうち異常値データとして CL の外側に配されるデータ数の上限値を設定するものである。この ν と判別境界面の算出時に出力される $f(x)$ が本研究で設定するパラメータである。図-1 に $f(x)$ を 3 次元空間上に図示したものを、図-2 にその $f(x)$ を等高線で切った判別境界線を示す。これらのパラメータの検討を行い、より精度の高い CL を設定する。

パラメータを選定する基準として、災害の発生が少ない地域では数少ない発生を全て捉えて (散発発生を除く) かつその中で最も非発生の的中率が高い

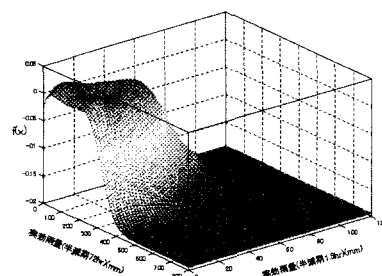


図-1 判別境界面の代表例

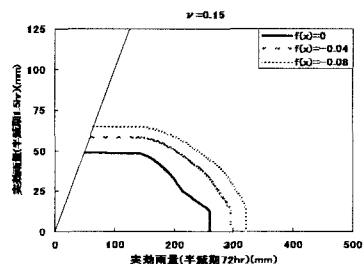


図-2 $f(x)$ 毎の判別境界線

パラメータを選択することとした。また、災害の発生が多い地域では非発生の的中率が90%以上でかつその中で最も発生の的中率が高いものを選択することとした。今回は発生の多い地域として下関市を採用した。表-1に各地域において選定した基準に当てはまる最適なパラメータ及び発生・非発生の的中率を算出した。その結果、各地域とも高い精度を得ることができた。

秋吉台ではパラメータ $\nu=0.1$, $f(x)=-0.02$ と $\nu=0.2$, $f(x)=-0.08$ が選定した基準に当てはまり、同等の的中率を示したが、CLの安全領域を考慮した上で $\nu=0.1$, $f(x)=-0.02$ を採用することとした。表-1に示した最適なパラメータによって各対象地域のCLを求めた。

5. 精度検証

各地域において選択したパラメータを用いて、線形CL、RBFNによる非線形CL（以下RBF-CLと表記）、2群判別法のSVM-CL（以下SVM-CLと表記）と比較検証を行なった。

図-3に下関市で設定したパラメータによる ν SVM-CL、RBF-CL、線形CLを比較した図を、表-2に発生・非発生の的中率を示す。その結果、 ν SVM-CLは従来の線形CLより高い精度を示し、現行のRBF-CLとも同等の結果を得ることができた。他の山口市、秋吉台、防府市においても下関市と同様の結果が得られた。

また、SVM-CLと ν SVM-CLとで比較検証を行った結果、下関市ではCLの形状に違いがあり、非発生の的中率は2群判別法が高く、発生の中率は ν SVMが高い精度を示した。しかし、図-5に示すように、2群判別法では発生の分布状況によっては大きく形状が異なってくるために常に高い精度のCLが設定できるとは言いがたい。これに対して ν SVMでは発生の分布状況に関わらず非発生のみでCLを設定できるので汎化性の高いCLが設定できると言える。以上のことから発生の多い地域においても ν SVMでCLを設定することの有効性が示されたものと判断した。

前述の最適なパラメータによるCLに対して2003年の秋吉台での災害データを用いて検証を行った。図-6にCLと災害発生スネークを示す。秋吉台では2003年に3件の災害が同時刻に発生しているが、CLは災害発生を的確に捉えており、選択したパラメータは妥当なものであったと判断した。

6. 結論

以上の結果より、 ν SVMでCLを設定する際にパラメータ ν と $f(x)$ の値を検討することで高い精度を有するCLが設定可能であることがわかった。しかし現状ではまだ山口県の5地域のための検証であるため、今後は多様な地域で検証する必要がある他、選択したパラメータで発生が捉えられない場合には、更新方法を検討する必要があると考えられる。

表-1 最適なパラメータ及び的中率

	秋吉台	山口市	防府市	下関市
最適なパラメータ ν	0.1	0.1	0.05	0.3
最適なパラメータ $f(x)$	-0.02	-0.02	0	-0.02
非発生の中率(%)	95.97	96.13	95.24	93.39
発生の中率(%)	100	100	100	89.65

表-2 的中率比較(下関)

		的中率(%)	的中降雨数(回)
発生	ν SVM($\nu=0.3$)	89.65	26/29
	RBF-CL	86.21	25/29
	線形CL	82.76	24/29
	2群判別SVM	75.86	23/29
非発生	ν SVM($\nu=0.3$)	93.39	113/121
	RBF-CL	93.39	113/121
	線形CL	91.74	111/121
	2群判別SVM	97.52	118/121

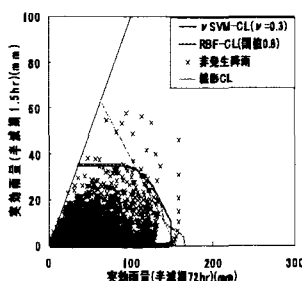


図-3 CL比較

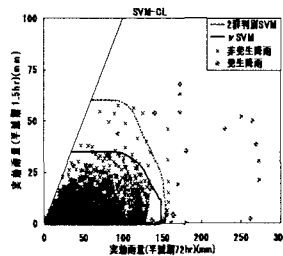


図-4 SVM・ ν SVM比較(下関)

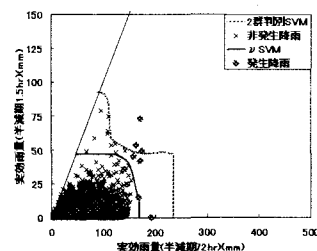


図-5 SVM・ ν SVM比較(岩国)

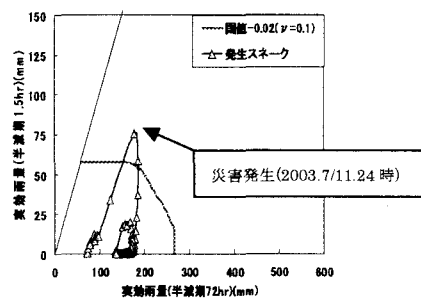


図-6 CLと発生スネーク