

SVM とラフ集合を用いた土石流の発生・非発生 Rule の作成方法に関する研究

大日本コンサルタント株式会社 篠崎嗣浩
西日本技術開発株式会社 大石博之

日本交通技術株式会社 森田真人
山口大学理工学研究科 古川浩平

1. はじめに

近年、局所的な集中豪雨が頻発し、最大時間雨量も増加傾向にあり、土石流災害は甚大化、凶悪化している。また、ハード対策においても重点整備が求められている社会情勢において、地域に広く散在する危険渓流の中から特に危険な箇所や比較的安全な箇所の分布を把握し、効率的に土砂災害の予防強化を行うことの重要性は高まっている。そのためには、多くの事例から土砂移動現象に影響する重要要因や発生・非発生の規則性（以下、Rule と呼ぶ）等の基礎知識を得ることが重要である。そのような中でラフ集合を用いた既往の研究では、地形・地質要因（以下、地形要因）と降雨要因を用いて、土石流の発生・非発生 Rule の作成を行い、災害発生の起因となった要因の追求に努められてきたが、実用的な Rule の作成にはいたっていない。

そこで、要因間の影響を加味して順位付けを行うことができるサポートベクターマシン（以下、SVM）と複数の要因から土砂災害の発生・非発生を説明するための Rule を抽出することに優れているラフ集合を用いて、できるだけ縮約された条件で多くの災害箇所を説明できる発生・非発生 Rule の作成方法を構築した。

2. SVM の概要

SVM は、非線形問題に対して線形分離可能な高次元特徴空間にマッピングを行うことで、最適な分離超平面を求めることができる手法である（図-1）。また、各データに対し分離超平面からの距離 $f(x)$ を算出することで危険度を評価し、順位付けを行うことができる。

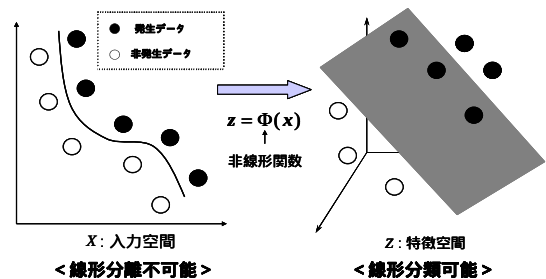


図-1 SVM の概念図

3. ラフ集合の概要

ラフ集合とは、要因間の関係から発生・非発生の規則（Rule）を求めることができる手法である。図-2(a)の着色部分内のデータを整合データといい、整合データが存在する領域を各 Rule として抽出する。なお要因の組合せを評価する指標としては、(1)式に定義される整合度を用いた。また抽出した Rule の精度と汎用性を表す尺度として(2)式、(3)式に示す確信度とサポートを用いた。

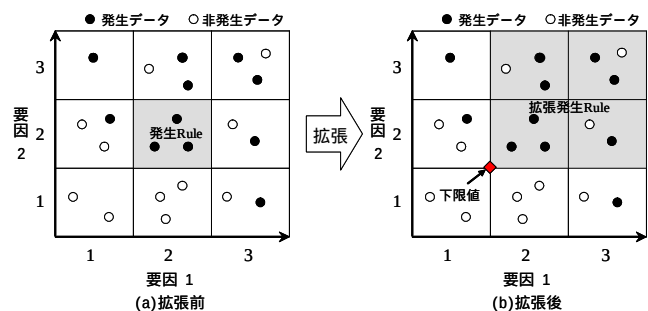


図-2 ラフ集合論

また、図-2(a)に示す着色領域が発生 Rule である場合、より危険と考えられる部分集合を含む領域を発生 Rule として拡張することで、発生 Rule に含まれる箇所数が増加し、より有効な Rule とすることができる拡張理論を取り入れる（図-2(b)）。

4. 使用データの概要

本研究の対象データは、土砂災害防止法制定の契機となるほどの大規模災害であった 1999 年 6 月 29 日に起こった広島災害の土石流のデータとした。その内訳は、発生流域は 213 箇所、非発生流域は 1,022 箇所となっている。なお、使用する要因として既往の研究にならい、地形要因 17 要因、降雨要因 2 要因を使用した。

$$\text{整合度} = \frac{\text{整合データ数}}{\text{全データ数}} \quad \dots (1)$$

$$\text{確信度} = \frac{\text{Rule領域内の整合データ数}}{\text{Rule領域内の全データ数}} \quad \dots (2)$$

$$\text{サポート} = \frac{\text{Rule領域内の全データ数}}{\text{全データ数}} \quad \dots (3)$$

5. 発生・非発生 Rule の抽出結果と考察

5.1 SVM による危険度の設定

本研究の解析手順を図-3に示す。SVM 解析において、発生・非発生を最も分離できる分離超平面を設定するため、危険箇所地域的中率が最大となる時のパラメータを用いることとした。ここでの中率とは的中データ数を全データ数で除した値と定義する。解析の結果、最も中率が高い $C=300$ 、 $r=3$ を最適なパラメータとすることとした。これより算出された $f(x)$ 値を用いて危険度を設定した。

5.2 SVM による代表データの抽出

危険度を基に対象データの順位付けを行い、その上位下位から特に危険なデータと安全なデータを抽出する。抽出方法については、元のデータベースの発生と非発生データの割合を乱さないようにするために、発生データ・非発生データからそれぞれ 14.1% ずつ抽出し計 174 箇所（発生：30、非発生 144）を抽出した。

5.3 ラフ集合による Rule 作成と結果

代表データ 174 箇所に対して整合度の要求水準 100% で要因数を絞り込み、その中でできるだけ少ない要因数で構成されるものを Rule として採用した。その結果、必要最小要因数は 4 要因（最急渓床勾配、主溪流長、溪流幅、時間雨量）となった。Rule の作成において汎用性の向上のため Rule 領域の拡張を行い、拡張した Rule のうち確信度が 90% 以上

（ある程度の誤判別を許容できる割合として設定）のもののみを発生・非発生 Rule として作成した。この代表データで作成された Rule を全体（1235 箇所）に適用した結果を表-1に示す。表-1 から 6 つの Rule での累計確信度 81.9%、累計サポート 84.8% と高い精度と汎用性をもった Rule の作成を行うことができた。

5.4 既往の研究との比較

既往の研究と本研究とを比較した結果を表-2に示す。なお、既往の研究では Rule 領域の拡張を行っていないため、1 つの Rule に対するサポートと適合個数について比較を行った。1 つの Rule に対するサポートの最大値を比較すると、既往の研究では非発生 4.0%、発生 0.5% であったものが、本研究では非発生 65.3%、発生 4.8% となっており、汎用性が大きく向上しているのが認められる。また、1 つの非発生 Rule で非発生データを、1 つの発生 Rule で発生データを適合できた数の最大値を比べると、既往の研究では、非発生 49 個、発生 6 個であったものが、本研究では非発生 717 個、発生 31 個となっており、Rule 適合の数が多いことが確認できた。

6. 結論

- ・代表データの抽出方法については、元のデータベースの発生と非発生データの割合を乱さないようにすることが最も効果的であることが分かった。
- ・代表データを抽出し、ラフ集合を用いて分析を行うことで、広島地区の土石流災害において全体の 84.8% を、わずか 6 つの Rule で説明できる実用性の高い Rule を得ることができた。

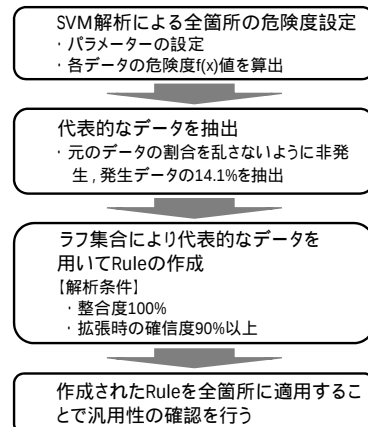


図-3 検討フロー図

表-1 Rule 抽出・適用結果

抽出されたルール	要因				判定	含まれるデータ数	整合データ数	矛盾データ数	残り箇所数 1235	累計	
	最急渓床勾配 (°)	主溪流長 (km)	溪流幅 (m)	時間雨量 (mm/hr)						確信度 = /	サポート = /1235
Rule12	*	*	*	60 ↓	非発生ルール	806	717	89	429	89.0%	65.3%
Rule1	*	*	2.0 ↓	*	非発生ルール	836	747	89	399	89.4%	67.7%
Rule11	20 ↓	0.20 ↓	*	*	非発生ルール	882	781	101	353	88.5%	71.4%
Rule26	*	*	*	70 ↑	発生ルール	941	812	129	294	86.3%	76.2%
Rule29	*	0.60 ↑	*	60 ↑	発生ルール	1014	844	170	221	83.2%	82.1%
Rule28	30 ↑	*	*	60 ↑	発生ルール	1047	857	190	188	81.9%	84.8%

表-2 既往の研究との比較一覧

項目	既往の研究 [ラフ集合のみ]	今回の検討 [SVM + ラフ集合]
使用データ数	[広島地区(3区)] 1235 溪流 (発生213 非発生1022)	
使用要因	地形・地質17要因 降雨 2要因	
整合度の要求水準	100%	100%
確信度の要求水準	発生、非発生: 0.900 以上	発生、非発生: 0.900 以上
累計確信度	—	81.9% (857/1047)
累計サポート	—	84.8% (1047/1235)
1Ruleでの最大サポート	非発生	4.0%
	発生	0.5%
1Ruleでの最大適合データ数	非発生	49
	発生	6