

SVM とラフ集合を用いた土砂災害の発生形態の差異に関する研究

山口大学大学院 学生会員 ○門井貴英 (株)四電技術コンサルタント 正会員 松本幸太郎
山口大学工学部 正会員 篠崎嗣浩 山口大学工学部 フェロー会員 古川浩平

1. はじめに

1999 年 6 月 29 日、梅雨前線による局地的な集中豪雨は、広島市及び呉市の各地で土砂災害を同時多発的に発生させ、死者・行方不明者 32 名もの多大な犠牲を出した。そこで土砂災害が集中した広島市周辺地区に着目すると、安佐北区、安佐南区、佐伯区の 3 区ではがけ崩れに比べ土石流が多く、呉市では逆の結果となっている（表-1）。3 区と呉市は同じ花崗岩風化土であり、同じ豪雨を経験しているにも関わらず、土砂災害の発生形態が異なっており、どの要因が影響して災害を引き起こしたのか詳細には分かっていないのが現状である。そこで、本研究では 3 区と呉市の発生差の大きい土石流に着目し、2 地区間の災害発生の起因となる要因を SVM (Support Vector Machine) とラフ集合により特定した。

2. 使用データ

使用データとして土石流危険渓流カルテと地形計測の 13 要因を用いた（表-2）。本研究で使用するラフ集合は離散値しか使用できないため、ラフ集合を使用するときのみカテゴリ分割を行い、要因データを作成した。

3. SVMの概要

SVMとは最も強力なパターン分類手法として注目されており、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。そうすることによって、線形分離可能な状態となり最適な分離超平面を求めることができる。これを利用して、土石流の発生・非発生を分離させ $f(x)$ 値による渓流毎の危険度の算出を行うことができる。（図-1）

4. ラフ集合の概要

4.1 評価指標

SVM により算出された発生・非発生の上位下位 5% のデータをラフ集合により解析を行い、簡潔で有効な知識 (Rule) を得ることとした。要因の組合せを評価する指標として次式に定義する整合度を用いた。また Rule の有効性を示す確信度とサポートを算出した。

$$\text{整合度} = \frac{\text{整合データ数}}{\text{全データ数}} \quad \text{サポート} = \frac{\text{Ruleに適合する箇所数}}{\text{全箇所数}} \quad \text{確信度} = \frac{\text{Ruleに適合するデータの箇所数}}{\text{Ruleに適合する箇所数}}$$

4.2 Ruleの拡張

汎用性の高い Rule 作成のため、適用領域の拡張を行った。図-2 (a) に示す網掛領域が発生 Rule である場合、より危険と考えられる領域を発生 Rule とすることで、Rule に含まれる箇所数が増加し、汎用性が高くより有効な Rule となる（図-2 (b)）。但し、非発生の矛盾データを含むことが多く、確信度は減少傾向になる。本研究では、有効な Rule を特定するため、この拡張理論を解析に取り入れることとした。

表-1 災害発生数

地域	土石流	がけ崩れ
呉市	1	81
広島市	安佐北区	77
	安佐南区	24
	佐伯区	50
	安芸区	0
	西区	0
	中区	0
	東区	0
	南区	0

表-2 地形要因

地形要因	水系模様
	流域平均勾配
	最急渓床勾配
	主溪流長
	流域面積
	流域長
	流域幅
	流域形状比
	谷深比
	0 次谷の数
	流域最大傾斜
	源頭部面積
	降雨集中度評価

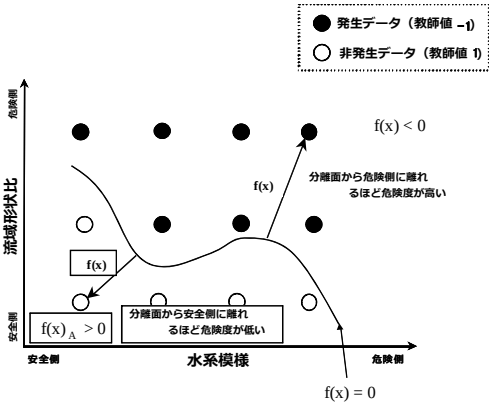


図-1 SVM2 次元イメージ図

5. 解析結果

5.1 重要要因の選定

Rule を構成する上で最適な要因の選定を行った。これは前章の Rule 領域の拡張を考えた場合、カテゴリ値の増加に伴い発生率が上昇する要因は、発生と非発生の分離性が良く、拡張後にサポートの向上に役立つだけでなく、確信度の低下を抑制できるからである。本研究ではラフ集合を用い、既往の研究より得られた整合度（90%）の要求水準を満たす組合せの中から6要因を最適な要因として選定した。

5.2 SVMによる危険度設定

前節の重要要因の選定で算出した要因を SVM で分析を行い、溪流の危険度を求めた。発生データの多かった3区を学習データとし、発生の少なかった呉市をテストデータとして解析を行った。SVM による危険度設定の結果、学習地域の3区は完全に発生と非発生に分離できており、テスト地域の呉はすべての溪流を非発生側で捉えていた。この結果から、呉市には土石流を引き起こす危険な溪流が少なかったと推測できる。

5.3 土石流の発生・非発生Ruleの抽出

SVM による解析の結果、発生と非発生の特徴的なデータである上位下位各5%をラフ集合にかけ Rule 抽出を行った。その結果、47個の Rule が得られた。その Rule を拡張し確信度90%以上でかつサポートの高い Rule を抽出した。表-3 に Rule の抽出結果を示す。表-3 が示すように、本研究で抽出した4つの発生・非発生 Rule により、全148溪流中99溪流（サポート約70%）を説明できる結果を得られた。

この結果より、3区の特徴的な Rule を特定できた。発生 Rule のうち最もサポートの高い Rule32 からは最急渓床勾配がきつく、山腹崩壊の可能性も高いという状況が推測される。この Rule を3区1479溪流に適用した結果（表-4）、確信度、サポート共に約70%と汎用性が高く多くの溪流を説明出来た。また、発生 Rule のサポートが約24%と3区の実際の発生率、約10%よりも高いことからして、3区には災害が潜在的に発生しやすい溪流が多いことが分かった。次に呉市283溪流にも4つの Rule を適用した結果（表-5）、非発生 Rule は確信度100%でサポート約50%と溪流の半分は完全に非発生であるという結果となった。また、発生 Rule のサポートが10%にも満たないことから、呉市では危険な溪流が少なかったことが分かった。4つの Rule で土石流全体の約70%が当てはまることから、結果として非常に汎用性の高い Rule が抽出できたといえる。

6. 結論

- 抽出した Rule を検証した結果、3区では最急渓床勾配がきつく、山腹崩壊の可能性も高い溪流であることに對し、呉市では全長が直線的で山腹崩壊の可能性が低く、なだらかな溪流だという Rule が抽出できた。
- 抽出された Rule は発生、非発生ともに2Rule で全体の7割を説明できる。
- SVM とラフ集合の組合せにより、少ない Rule で多くのデータを説明できる有用な Rule を抽出できた。今後は他地域にて本研究と同様の手法で Rule の選定が可能かを検証していく。

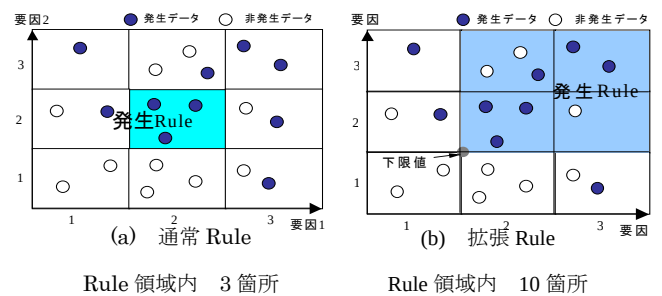


図-2 Rule 拡張の概念図

表-3 上位下位各5%Rule抽出

	要因						判定	①含まれるデータ数	②整合データ数	③矛盾データ数	④残溪流数 148	累計	
	水系模様	最急渓床勾配	流域形状比	0次谷	流域最大傾斜	源頭部面積						確信度	サポート
Rule9	3以下	*	*	1	4以下	*	非発生ルール	44	41	3	104	0.932	0.297
Rule12	*	2以下	4以下	*	*	1	非発生ルール	55	51	4	93	0.927	0.372
Rule32	*	3以上	*	3以上	*	*	発生ルール	87	81	6	61	0.931	0.588
Rule26	*	*	*	*	*	4以上	発生ルール	99	93	6	49	0.939	0.669

表-4 3区1479溪流Rule適用

	要因						判定	①含まれるデータ数	②整合データ数	③矛盾データ数	④残溪流数 1479	累計	
	水系模様	最急渓床勾配	流域形状比	0次谷	流域最大傾斜	源頭部面積						確信度	サポート
Rule9	3以下	*	*	1	4以下	*	非発生ルール	506	484	22	973	0.957	0.342
Rule12	*	2以下	4以下	*	*	1	非発生ルール	684	655	29	795	0.958	0.462
Rule32	*	3以上	*	3以上	*	*	発生ルール	921	701	220	558	0.761	0.623
Rule26	*	*	*	*	*	4以上	発生ルール	1033	718	315	446	0.695	0.698

表-5 呉市283溪流Rule適用

	要因						判定	①含まれるデータ数	②整合データ数	③矛盾データ数	④残溪流数 283	累計	
	水系模様	最急渓床勾配	流域形状比	0次谷	流域最大傾斜	源頭部面積						確信度	サポート
Rule9	3以下	*	*	1	4以下	*	非発生ルール	118	118	0	165	1.000	0.417
Rule12	*	2以下	4以下	*	*	1	非発生ルール	133	133	0	150	1.000	0.470
Rule32	*	3以上	*	3以上	*	*	発生ルール	139	133	6	144	0.957	0.491
Rule26	*	*	*	*	*	4以上	発生ルール	152	133	19	131	0.875	0.537