

イタラッキ地域における山地災害実態の解明

株式会社荒谷建設コンサルタント 正会員 古浦 孝雄 北海道立林業試験場 非会員 佐藤 創
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

近年 北海道では台風などによる森林被害が目立つようになってきた。豪雨被害としては2003年台風10号による日高、十勝地方を中心としたものがあり、大量の流木の発生が問題となった。厚別川流域全体(図-1)に堆積した流木は合計46,260m²であり、その内訳は、50%は台風による豪雨で山腹由来の樹木が新たに流木となったものであり、23%は河畔林が新たに流木になったもの、残り27%が元々流域内にあった古い倒木や流木であったことから、流木の発生する原因として斜面崩壊によるものが大半であり、今後の豪雨による流木の発生を抑制することが急務であることがわかる。



図-1 対象地域

そこで、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) という数学的な手法を用いることで、既往の研究では分離出来なかった危険箇所の明確な分離を行うことを目的とし、危険箇所の発生・非発生データの的中率の精度向上を図ることを試みた。

2. 使用データの概要

2.1 対象地域

本研究では、2003年台風10号によって被害を受けた厚別川の支流にあたる流域面積17.7km²、流路延長9kmのイタラッキ川流域(図-2)を解析の対象範囲とした。

2.2 解析データ

本研究で用いる要因は林相・地質・傾斜角・曲率の4要因となる。また、イタラッキ流域を10m×10mのメッシュに分割したため、データ数は合計65,124箇所となり、データ数が膨大となるうえに同一カテゴリ値となるデータが多くなるため、同カテゴリ値の箇所を集約したデータを用いることとした。

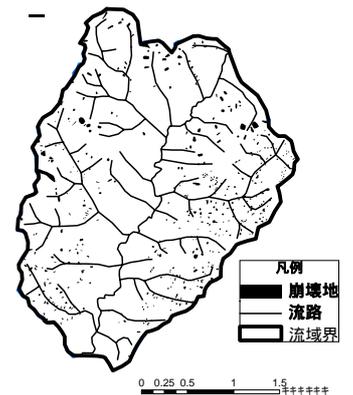


図-2 イタラッキ流域

3. SVMの概要

SVMとは1990年代にVapnikらによって提案された手法で、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。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特徴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。分離超平面から各危険箇所までの距離をf(x)値として算出し、どの程度危険なのかを相対的に判断することが出来る。SVMの基本的な考え方を図-3に示す。

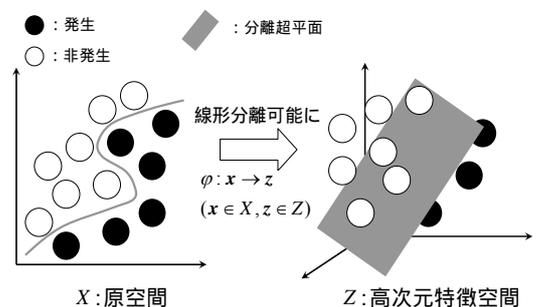


図-3 SVMの基本概念

4. 使用データの検討

4.1 データの集約

本研究で用いる65,124箇所のデータについて、同一カテゴリ値である箇所を集約したが、発生データと非発生データが混在しており、教師値をどちら側に扱うかという問題点が生じた。そこで本研究では集約デー

タについて、通常評価、安全側評価、崩壊率評価の3つのパターンのデータを作成し、どの集約パターンが最適であるかを検討した。ここで使用するデータは計算時間の短縮を考慮し、既往の研究で用いられた要因である林相・地質・SHALSTAB 値の3要因を使用した。

4.1.1 通常評価

集約した際に、同一カテゴリ値で発生と非発生を含むパターンが出てきた場合、一方は発生、もう一方は非発生という2つのデータを作成しSVMによる解析を行う(表-1)。

4.1.2 安全側評価

安全側評価では、集約した際に、発生と非発生を含むパターンについては、発生が1つでもあればそのパターンは発生として扱い、データを作成する(表-2)。

4.1.3 崩壊率評価

崩壊率評価ではカテゴリの組み合わせパターンに対し崩壊率を算出した(表-3)。この崩壊率は各パターンで集約された発生箇所を、そのパターンのTotalで割った値となる。ここでは崩壊率が5%以上であれば発生としてデータを作成した(表-4)。

4.2 検討結果

前述の3つのパターンをSVMにより解析した結果、発生の的中率が通常評価の検討では69%、安全側評価では84%、崩壊率評価では91%となった。よって、本研究では、3つのパターンの中で最も高い的中率を得ることができた崩壊率評価を採用することとする。崩壊率評価で得られたグラフを図-4に示す。

4.3 崩壊率の検討

4.1.3で、本研究では崩壊率評価によりデータを作成し解析を行う事としたが、崩壊率について詳細に検討することで、さらに的中率が向上する可能性がある。そこで、崩壊率を1%以上、5%以上、10%以上の3ケースについて、それぞれを発生として扱い、どの崩壊率が解析データに適しているかを検討した。

検討結果、いずれの崩壊率の評価でも、大半の非発生箇所は安全側で捉えることができているが、発生箇所に関しては、崩壊率が上昇するほど安全側に寄っていく傾向が見られた。よって本研究では、最も危険側で発生箇所を多く捉えることができた崩壊率1%以上を採用してデータを作成する。崩壊率1%以上を発生としたグラフを図-5に示す。

4.3 SVMの解析

4.3.1 最適パラメータの検討

SVMで分離超平面を構築するには、 C 、 r の2つのパラメータを設定する必要がある。この2つの値によって、どのような条件で分離超平面が形成されるかが決定される。 C はSVMの分離超平面が誤判別をどれだけ許容するかを示すパラメータで、値が大きいほど異常値を認めずデータを厳しく判別する。また、 r はデータを特徴空間に写像するための非線形関数として利用するガウス関数の半径を表している。

表-1 通常評価によるデータ

同カテゴリ値 パターンNo.	林相	地質	SHALSTA B値	教師値
1	1	1	1	1
2	1	2	1	1
3	1	3	1	-1
3	1	3	1	1
4	1	4	1	1
5	1	8	1	-1
5	1	8	1	1

表-2 安全側評価によるデータ

同カテゴリ値 パターンNo.	林相	地質	SHALSTA B値	教師値
1	1	1	1	1
2	1	2	1	1
3	1	3	1	-1
4	1	4	1	1
5	1	8	1	-1
6	1	10	1	-1
7	1	2	2	1

表-3 集約データ

同カテゴリ値 パターンNo.	林相	地質	SHALSTA B値	データ個数		崩壊率
				発生	非発生	
1	1	1	1	0	75	0
2	1	2	1	0	115	0
3	1	3	1	6	308	1.9%
4	1	4	1	0	11	0.0%
5	1	8	1	1	435	0.2%
6	1	10	1	1	255	0.4%
7	1	2	2	0	6	0.0%
8	1	3	2	3	57	5.0%

表-4 崩壊率評価によるデータ

同カテゴリ値 パターンNo.	林相	地質	SHALSTA B値	教師値
1	1	1	1	1
2	1	2	1	1
3	1	3	1	1
4	1	4	1	1
5	1	8	1	1
6	1	10	1	1
7	1	2	2	1
8	1	3	2	-1

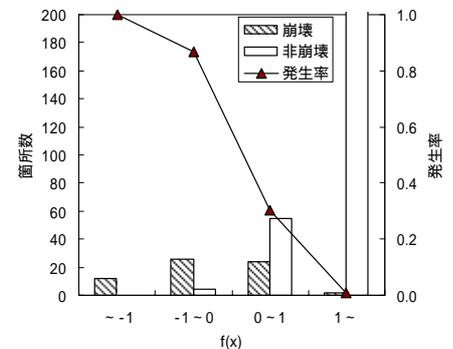


図-4 崩壊率評価での解析結果

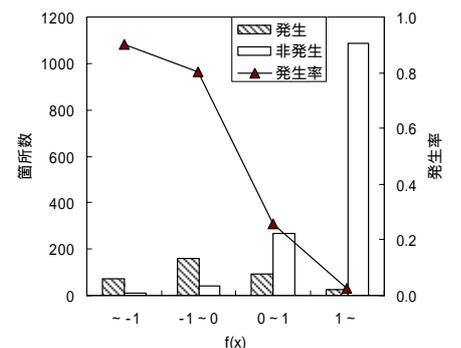


図-5 崩壊率1%以上の解析結果

本研究では、まず r について検討を行った。表-5 に解析結果を示す。結果を見てみると、 r が大きくなるにつれ、発生的中率は下がる傾向があり、 r が 4 以上になると発生的中率は 0% と、発生パターンを的中することができなくなった。そこで本研究では、パラメータ r は 0.1, 0.5, 1, 2, 3 に絞り、解析を行った。

表-5 的中率の変化

		r						
		0.1	0.5	1	2	3	4
C	発生的中率	100%	99.2%	66.3%	36.0%	22.8%	0%
	非発生的中率	100%	99.6%	96.6%	96.6%	97.6%	100%
	全データに対する的中率	100%	99.5%	90.4%	84.3%	82.4%	79.7%

4.3.2 SVM による解析

本研究では、崩壊率 1%以上を発生として 1,757 パターンの集約データを用いて SVM での検討を行った。検討を行う際のパラメータは r が 0.1, 0.5, 1, 2, 3 とし、 C は 5, 10, 50, 100, 200, 300 とする。以上のパラメータを用い、各パラメータの発生・非発生・全データのそれぞれに対する的中率を表-6 に示す。いずれのパラメータの組合せでも高い中率を得ることができたが、その中でも分離が出来ており、危険側ほど発生率が高く、安全側ほど発生率が低くなるパラメータを抽出した結果、 $C=200$ $r=0.5$ を採用することとした。図-6 に $C=200$ $r=0.5$ の解析結果のグラフを示す。 $f(x)=0$ を境に値が小さければ危険となり、大きければ安全となる。図-6 では危険側で発生パターンを、安全側で非発生パターンを多く捉えていることから非常に明確な分離を行うことができた。

表-6 各パラメータにおける的中率

		r					
		0.1	0.5	1	2	3	
C	5	発生的中率	100%	77.2%	43.3%	16.0%	0%
		非発生的中率	100%	97.6%	96.1%	98.2%	100%
		全的中率	100%	93.5%	85.4%	81.6%	79.7%
	10	発生的中率	100%	85.4%	49.4%	23.0%	0%
		非発生的中率	100%	98.5%	95.6%	97.8%	100%
		全的中率	100%	95.8%	86.2%	82.6%	79.7%
	50	発生的中率	100%	98.3%	60.7%	32.9%	14.3%
		非発生的中率	100%	99.5%	95.7%	96.9%	98.4%
		全的中率	100%	99.3%	88.6%	83.9%	81.3%
	100	発生的中率	100%	99.2%	66.3%	36.0%	22.8%
		非発生的中率	100%	99.6%	96.6%	96.6%	97.6%
		全的中率	100%	99.5%	90.4%	84.3%	82.4%
	200	発生的中率	100%	99.7%	69.1%	38.5%	24.4%
		非発生的中率	100%	99.8%	96.9%	96.1%	97.6%
		全的中率	100%	99.8%	91.2%	84.5%	82.8%
300	発生的中率	100%	99.4%	72.2%	39.6%	25.8%	
	非発生的中率	100%	99.9%	96.9%	96.0%	97.9%	
	全的中率	100%	99.8%	91.9%	84.6%	83.3%	

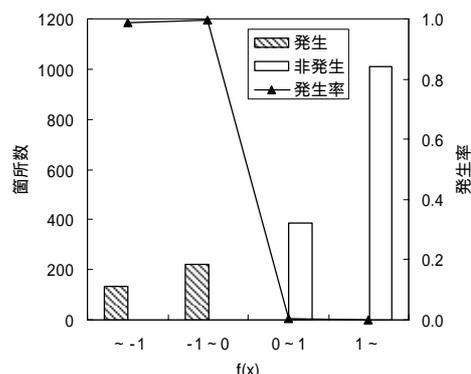


図-6 パラメータ $C=200$ $r=0.5$ グラフ

5. 分離データの特徴評価

本研究では、崩壊率評価を用い、集約したデータについて崩壊率 1%以上を発生とすることで、SVM を用いた明確な分離を行うことが出来た。しかし、崩壊率 1%以上のパターンを発生箇所として扱うという事は、集約してある 1 つのパターン内に最大で 99% の非発生を発生として扱い、分離を行った可能性があるということである。つまり、図-6 で表されたグラフの危険側にある発生パターンの中には多くの非発生箇所を含んでいるということになる。そこで、これらの危険側に分離された非発生箇所は発生データとして扱うのは妥当であるのか検証するため、ラフ集合を用いて、特徴的な違いを見るためのルール検討を行うこととした。

6. ラフ集合

6.1 ラフ集合の概念

ラフ集合論の基本概念は種別と近似である。本研究ではこの 2 つの概念を用いた発生・非発生データの縮約により、データベースから簡潔かつ有効な知識 (ルール) を得ることを目的とする。

図-7 にラフ集合の概念図を示す。発生箇所を●、非発生箇所を○とすると、発生箇所だけの集合は図-7 に示される部分集合となる。全体集合を 2 つの要因を用いて 9 個の領域に区分しても発生箇所だけの部分集合は完全に分離されない。このように完全に分離されない集合をラフ集合という。これらの要因によって集合が区分された時、同じ領域内のデータがすべて同種であるデータ領域を整合データ (図-7 の影の部分) とし、混

在するデータ領域を矛盾データと定義する。

6.2 ルール作成

本研究では集約した際に、矛盾を含まないデータを用いることで、信憑性の高いルールを抽出することを目的とする。集約した際に矛盾を含まないデータ、つまり同一カテゴリ値で発生箇所と非発生箇所が同時に存在しないデータのみを用いてラフ集合により非発生ルールを抽出し、そこで得られた非発生ルールを、危険側に含まれる非発生箇所に適用することで、本来ならば安全側に分離されるべきデータはどれ程あるのかを検討した。ここで抽出する非発生ルールの条件として、発生箇所を含まない、つまり確信度 100%のルールを選び、その中で最もサポートが高いルールを抽出することとする。

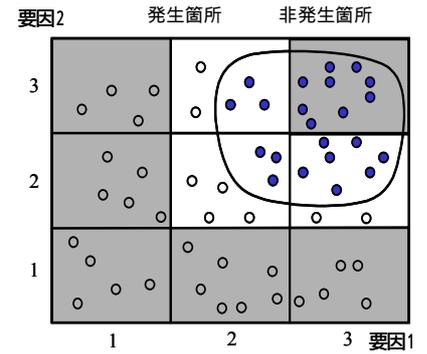


図-7 ラフ集合の概念

6.3 ルールの拡張

汎用性の高いルールの抽出を目的とするため、適用領域の拡張を行った。図-8(a)に示す網掛領域が発生ルールである場合、より危険と考えられる部分集合を含む領域を発生ルールとすることで、ルールに含まれる箇所数が増加し、より有効なルールとなる(図-8(b))。本研究では、有効なルールの特定のため、この理論を解析に取り入れることとした。また、ルール拡張を最大限に生かすため、離散データである林相と地質に関しては、各カテゴリの発生率を算出し、発生率の高い順に並び替えることとした。

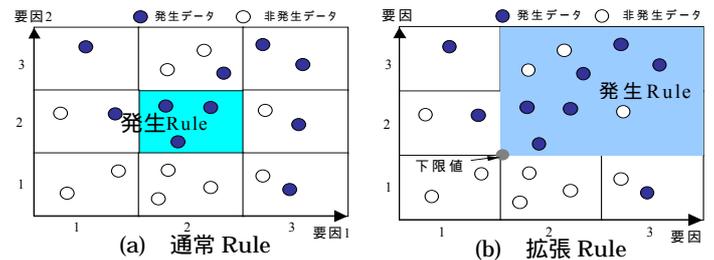


図-8 拡張の概念図

6.4 ルール抽出

ラフ集合によりルールを抽出した結果、68 個のルールを得ることが出来た。68 個のルールの中から、要求水準である確信度 100%であり、その中から最もサポートが高いルールを抽出した結果、2 つの非発生ルールを抽出することができた。抽出されたルールの詳細を表-7 に示す。

この2つのルールは、ルール36では「林相が7以下(カラマツ30~35年生以下)であり、かつ傾斜角が4以下(20°以下)」であれば非発生箇所となるルールであり、このルールで全データの約20%が説明できる。また、ルール1では他の要因に関係なく、「地質が1(段丘・沖積面)」であれば非発生箇所となるルールである。この抽出された2つの非発生ルールを、危険側にある非発生箇所に適用する。なお、危険側にある非発生箇所は合計33,456箇所あるが、ラフ集合によって得られた2つのルールを適用した結果、1,027箇所(3%)に該当した。今回抽出された非発生ルールに3%しか該当しないということは、残り97%は崩壊こそ起こっていないものの、SVMでは危険側に位置しているため、十分に崩壊の危険性を含んでいる箇所と言えるのではないかと考えられる。

表-7 抽出されたルール

	林相	地質	傾斜角	曲率	判定	ルール満足数	整合データ数	矛盾データ数	確信度	サポート
Rule36	7以下	*	4以下	*	非発生	287	287	0	1	0.208
Rule1	*	1以下	*	*	非発生	107	107	0	1	0.098

7. 結論

- データを離散化して集約することによって、解析時間の短縮を図ることができ、生値のデータを用いた場合に比的中率の向上を図ることが出来た
- 崩壊率1%以上を発生とすることで、多くの非発生箇所が危険側に分離されてしまうが、ラフ集合を用い非発生ルールを抽出することで97%のデータが崩壊の危険性が高い箇所であると推測できる。