

データ分類とニューラルネットワークによる斜面崩壊確率推定

防衛大学校 ○作田 健 香月 智 梶田幸秀
(財) 砂防・地すべり技術センター 安田勇次

1. 緒言

豪雨に起因する斜面崩壊予測に際し、対象地域の特性をつかむことが必要であり、現在までに蓄積されたデータ群を有効に利用した予測システムの構築が重要である。本研究では、図-1に示す地域におけるGIS、リモートセンシングデータ等のデータ群に対し、データマイニング¹⁾分野の基礎技術である類似性分析を活用し、さらに誘因である降雨データを含め、ニューラルネットワークによる斜面崩壊予測システムの構築を試みた。

2. データの類似性分析による崩壊確率の算定

2.1 観測データ

本研究で使用するデータは、愛媛県重信川上流域の砂防直轄区域内の崩壊記録である。この地域では、昭和39年、50年、52年および平成11年の斜面崩壊状況が航空写真として記録されており、植生・地質・勾配・標高等のデータとともに東西50m、南北50mごと(50,569グリッド)でデジタルデータ化されている。このうち、平成11年の崩壊状況写真は、平成11年9月14～15日の豪雨直後にとられたものである。本研究では、この時の降雨データを斜面崩壊の誘因データとして使用する。

2.2 崩壊確率の算定

本研究で取り扱う説明変数は、表-1に示すように、任意の数値範囲を有するデータ列となっている。例えば、標高が100～1300mと 10^2 のオーダーであるのに対して、時間最大雨量は40～120mmと 10^1 のオーダーである。これを同じ尺度で取り扱うために、0.0～1.0の規格化された変数にする。すなわち、

$$\bar{x}_i = \frac{(q_{i\max} - q_{i\min}) \cdot p_i + q_{i\min} \cdot p_{i\max} - q_{i\max} \cdot p_{i\min}}{p_{i\max} - p_{i\min}} \quad (1)$$

ここで、 \bar{x}_i ：正規入力データ、 $p_{i\max}$ 、 $p_{i\min}$ ：それぞれ入力項目の最大値および最小値、 $q_{i\max}$ 、 $q_{i\min}$ ：規格化された入力値の最大値（ここでは $q_{i\max}=1$ ）および最小値（ここでは $q_{i\min}=0$ ）、 p_i ：入力値。

データは図-2のように、多次元の超空間に点在する。そこで、データ間の類似度を、次式により求められるデータ j と k とのユークリッド距離により表すものとする。

$$R_{jk} = \sqrt{\sum_{i=1}^n (\bar{x}_{ij} - \bar{x}_{ik})^2} \quad (2)$$

ここで、 n ：次元数、 R ：類似度。

データの類似性分析を用い、斜面崩壊確率（崩壊ポテンシャル）を算定する。説明変数 x_j に対する崩壊確率は、適当な閾値 R_s を設定したうえで、データ x_j との類似度が R_s を超えない範囲のデータを抽出する。ここで、斜面崩壊確率を次のように定めた。

$$P_{sj} = \frac{n_{fs}}{n_s + 1} \quad (3)$$

$$P_{ej} = \frac{n_f}{n_s} \quad (4)$$

ここで、 P_{sj} ：統計崩壊確率、 P_{ej} ：推定崩壊確率、 n_s ：類似データの数、 n_f ：崩壊データ数、 n_{fs} ： x_j に対応する崩壊事象の有無を n_f に加えたもの。

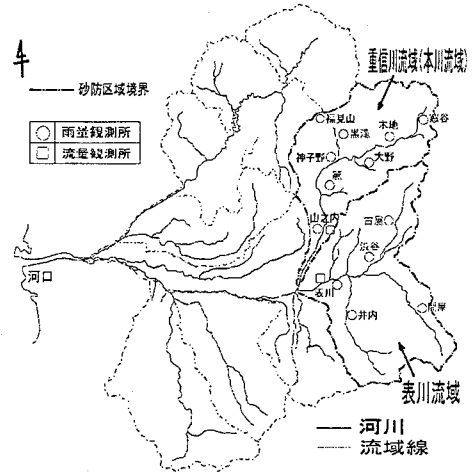


図-1 対象地域

表-1 規格化

(a) 元データ

No.	素因データ				誘因データ			崩壊
	x_1	x_2	...	x_j	...	x_{n-1}	x_n	
1	4	0	...	100	...	40	24	1
2	4	1	...	406	...	12	27	1
...								
i	x_{i1}	x_{i2}	...	x_{ij}	...	x_{in-1}	x_{in}	y_i
...								
m	8	0	...	1300	...	120	11	0

(b) 規格化データ

No.	素因データ				誘因データ			崩壊
	\bar{x}_1	\bar{x}_2	...	\bar{x}_j	...	\bar{x}_{n-1}	\bar{x}_n	
1	0.5	0	...	0.0	...	0.66	0.18	1.0
2	0.5	1	...	0.25	...	0.21	0.22	1.0
...								
i	\bar{x}_{i1}	\bar{x}_{i2}	...	\bar{x}_{ij}	...	\bar{x}_{in-1}	\bar{x}_{in}	\bar{y}_i
...								
m	1	0	...	1.0	...	0.33	0.07	0.0

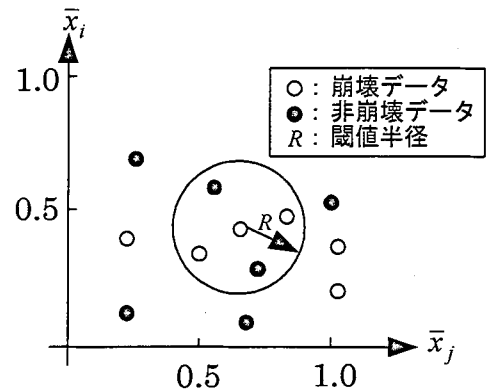


図-2 類似領域

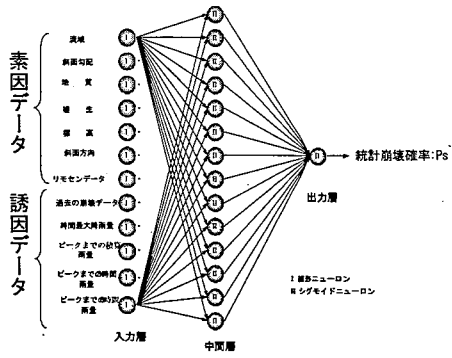
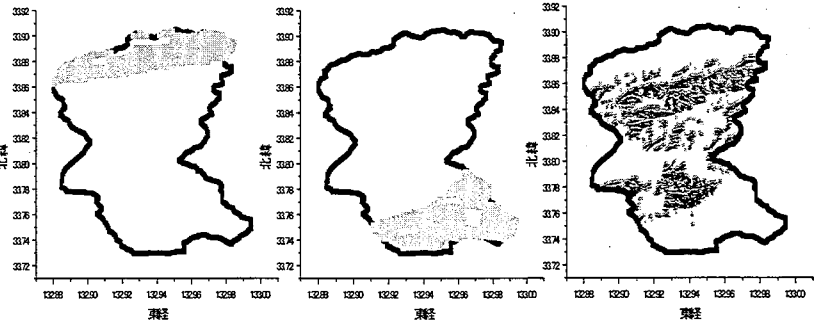


図-3 ニューラルネットワークシステム



(a) グループ1 (b) グループ2 (c) グループ3

図-4 グループ分け

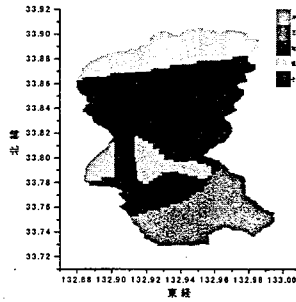


図-5 地質分布

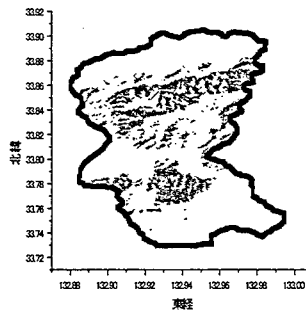


図-6 未学習データ分布

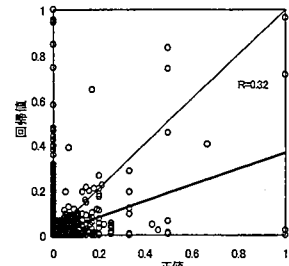


図-7 推定結果

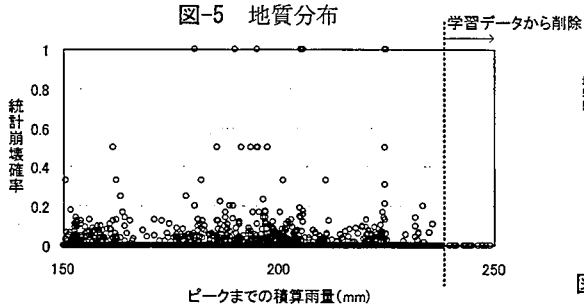


図-8 学習データと積算雨量

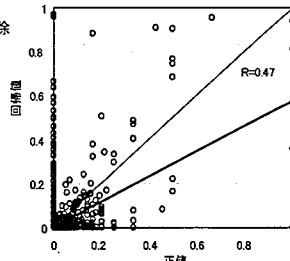


図-9 異常値を削除後の推定結果

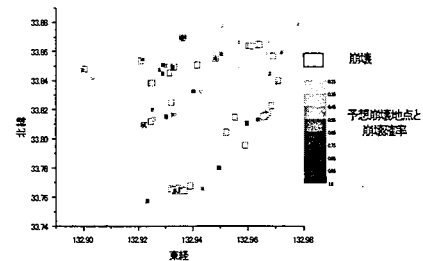


図-10 崩壊地点 (事実) と推定地点

3. ニューラルネットワークによる斜面崩壊確率推定

データの類似性分析を行うことにより、崩壊確率を算定する手法を提案した。しかし、この手法では、崩壊確率と降雨強度の関係を導きだすことができない。そこで、崩壊確率と降雨の関係を導きだすため、ニューラルネットワークシステムによるモデル構築を行う。ここで用いたニューラルネットワークシステムは、図-3に示す3層の階層型ネットワークである。

まず、 n 次元データ空間上において、降雨による誘因データを除いた素因データのみで、式(2)によりデータ間のユークリッド距離が近いデータのグループ分けを行った。ここで、各グループ内で、母集団が大きい3個グループの地域を図-4に示す。図-4の傾向から、図-5に示す地質の境界線がグループ分けに強く影響していることがわかる。ここでは、グループ3である本谷川流域付近の推定結果について述べる。

グループ3に含まれる全グリッド数は6,959であり、この内、崩壊グリッド数は90である。崩壊割合が流域全体では0.9% (444/50569) であるのに対し、グループ3では1.3%と高い値を示している。このデータを、類似性分析を用い、閾値半径 $R = 0.11$ として統計崩壊確率データに改めると、崩壊確率を有するグリッドは、1025グリッドとなる。グループ3のうち、3479グリッド (崩壊グリッド数=45, 非崩壊グリッド数=3434)

を学習データとして用い、ニューラルネットワークに知識獲得させ、残りの図-6に示す3480グリッド (崩壊グリッド数=45, 非崩壊グリッド数=3435) を未学習データとして推定させた。

学習データに対する収束結果は、相関係数が93%と高い。しかし、構築されたシステムを用いた推定結果は、図-7に示すように相関係数が32%と高くはない。そこで、学習データが有する降雨データの分析を行った。図-8に積算雨量と崩壊確率の関係を示す。通常、降雨の増大に伴い崩壊確率も高くなると考えられるが、用いた学習データは降雨が240mmを超えると崩壊確率が低下している。そこで、主観的判断により不自然と思われるデータを削除し、さらに、平成11年6月の非崩壊時の降雨データを追加データとして与えた。その結果、図-9に示すように、相関係数が47%へと向上した。また、推定結果と未学習データの崩壊地点をマッピングしたものを図-10に示す。これより、推定結果は、崩壊傾向をよく捉えている。

4. 結言

本研究では、平成11年9月の愛媛県重信川流域の豪雨を対象とし、斜面崩壊の発生を確率として評価した。また、斜面崩壊を推定する方法として、ニューラルネットワークを用い、素因データに誘因である降雨データを加えて、斜面崩壊確率を推定することを提案した。

参考文献

- 1) 山本ら：データマイニング、共立出版、1998.6