

69 データマイニングによる急傾斜地におけるがけ崩れ発生の特徴要因の抽出の試み

八千代エンジニアリング
山口県土木建築部砂防課
甲南大学理学部

○菊池 英明
鉄賀 博巳
中山 弘隆

山口大学工学部
山口大学工学部

榊原 弘之
古川 浩平

1. はじめに

斜面災害による人的被害の軽減のためには、災害発生の予測システムの確立が必要である。しかし、対象となる斜面が多数存在することから、すべての斜面についてその力学的挙動を詳細に反映した予測モデルを構築することは時間的、予算的制約上困難である。本研究では急傾斜地崩壊危険箇所データ（以下急傾斜地データ）に対してデータマイニング手法を適用し、地形要因から潜在的な危険性を評価する。

2. データマイニング

2.1 斜面防災計画のためのデータマイニング

がけ崩れや土石流等の斜面災害に対しては、それぞれ建設省による急傾斜地崩壊危険箇所調査、土石流危険渓流調査が実施され、大量のデータが蓄積されつつある。しかしデータの属性（調査項目）を相互に関連付けて、災害発生との因果関係を分析した例は存在しない。また、斜面災害を属性により分類し、現在の調査項目自体の妥当性について検討した事例も少ない。そこで本研究では大量のデータから有益な知識を抽出しようとするアプローチであるデータマイニングを急傾斜地でのがけ崩れ発生・非発生のデータに適用する。具体的手法としては、ラフ集合論を応用したアルゴリズムを用いる。

2.2 ラフ集合

ここでラフ集合の定義を簡略に示す¹⁾。複数の属性を有するデータの集合を考える。一つ一つのデータは各属性のいずれかのカテゴリーに属するとする。図1の場合、大きな四角形は斜面に関するすべてのデータの集合を意味する。個々のデータは地質区分(R_1)、傾斜度(R_2)、高さ(R_3)という三つの属性におけるいずれかのカテゴリー（地質区分の場合土砂・軟岩等、傾斜度の場合は角度）に属している。

斜面の崩壊現象を考えた場合、事前に収集したデータにおいて各属性のカテゴリーが全く同一であっても、崩壊する斜面と崩壊しない斜面が生じることは十分考えられる。これは、カテゴリーの組み合わせによって崩壊現象の必要十分条件を得ることができないことを意味する。このようにカテゴリーの組み合わせによって定義できないようなデータの部分集合をラフ集合と呼ぶ。ラフ集合は図1の楕円のように示される。

3. ラフ集合に基づいた特徴要因抽出手法

3.1 ルール抽出と属性削除

がけ崩れ発生・非発生データの集合がラフ集合であっても、そこから防災に有益な知識を抽出することは可能である。例えば図1において、地質区分が「土砂」の斜面において傾斜度が $40^{\circ}-60^{\circ}$ であった場合は必ず斜面崩壊が発生している。この場合(土砂、 $40^{\circ}-60^{\circ}$ →

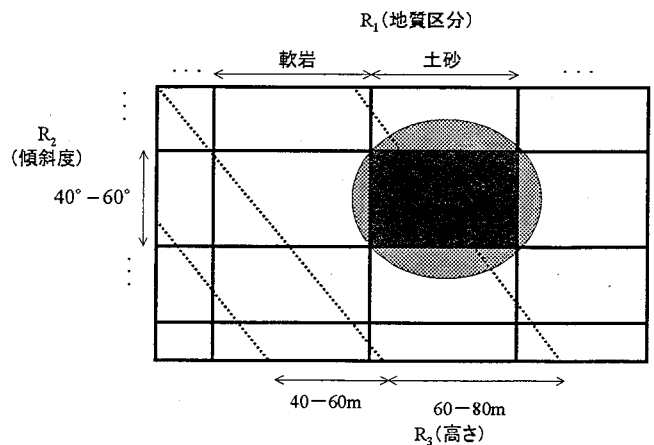


図1 ラフ集合

崩壊発生)というルールを抽出することが可能である。

さらに、ルールの正確性を低下させないのであれば、不要な属性をデータベースから削除することも可能である。図1においては、高さ属性が存在しなくとも、発生、非発生現象をある程度の正確さで分類することが可能であり、高さ属性の現象説明力は低いと考えられる。この場合、事前の調査項目から高さを削除することも可能であろう。

本研究では石井²⁾の開発した、ルールの正確性を維持しつつ属性を削除するアルゴリズムを急傾斜地データに適用した。

3.2 ルールの評価指標

一部属性を削除した結果として、縮約されたデータベースを得ることができる。その評価指標である整合度は以下のように定義される。

$$\text{整合度} = \frac{\text{縮約されたデータベースのもとで矛盾しないデータ数}}{\text{全データ数}} \quad (1)$$

ここで「矛盾しない」とは、各属性のカテゴリーが同じならば、結果（崩壊、未崩壊）が同一であることを意味する。図1の発生現象のデータにおいては、黒色の四角形の内部に属するデータは矛盾しないデータであり、灰色の部分のデータは矛盾するデータ（同一カテゴリーでも発生しないことがある）である。

4. 山口県急傾斜地崩壊データを用いた要因抽出結果

4.1 山口県急傾斜地崩壊データの概要

使用データは、山口県内の急傾斜地データと、災害報告書である。この2種類のデータベースを照合し、災害報告書に該当する危険箇所を「発生斜面」、該当しない箇所を「非発生斜面」とした。一部属性の欠けたデータを削除した結果、データ数は発生 430 箇所、非発生 3000 箇所、計 3430 箇所となった。

また各データの有する属性は以下の21通りである。傾斜度、高さ、斜面方位、斜面形状、横断形状、

遷急線、地表の状況、表土の厚さ、地盤の状況、岩盤の状況、斜面と不連続面傾斜関係、断層・破碎帯の有無、風化状況、植生の種類、樹木の樹齢、伐採根の状況、調査斜面・崩壊履歴、隣接斜面・崩壊履歴、湧水、対象斜面と活断層距離、対象斜面と活断層の方向

4.2 特徴要因の抽出結果

図2は整合度の制約の下で、必要最小限の属性数をプロットしたものである。整合度を高めようとするほど、必要な属性の数は増加する。しかし属性数を増やすほど、個々のカテゴリーの組み合わせに属するデータ数は減少し、ルールは個別の斜面に対応する傾向が強まる。その結果得られる知識の普遍性はむしろ低下してしまう。従って斜面防災計画の観点からの要請に基づいた整合度を選択する必要がある。

以下に整合度がそれぞれ70%、90%の場合の必要最小限の属性数と、属性の組み合わせを以下に示す。

a) 整合度 70% 属性数5, 属性の組み合わせ1通り
傾斜度, 斜面形状, 表土の厚さ, 樹齢, 活断層との距離

b) 整合度 90% 属性数7, 属性の組み合わせ5通り

1. 傾斜度, 斜面形状, 横断形状, 遷急線, 表土の厚さ, 地盤の状況, 樹木の種類
2. 傾斜度, 斜面形状, 横断形状, 遷急線, 表土の厚さ, 樹木の種類, 活断層との距離
3. 傾斜度, 斜面形状, 遷急線, 地表の状況, 表土の厚さ, 樹木の種類, 活断層との距離
4. 傾斜度, 斜面形状, 横断形状, 表土の厚さ, 地盤の状況, 樹木の種類, 活断層との距離
5. 傾斜度, 斜面形状, 遷急線, 表土の厚さ, 地盤の状況, 樹木の種類, 活断層との距離

また表1は整合度の制約90%の場合の5通りの組み合わせについて、各属性が削除されずに残留する組み合わせの数を示している。

表1 整合度90%における各属性の残留する組み合わせ数

傾斜度	斜面形状	横断形状	遷急線	地表の状況
5	5	3	4	1
表土の厚さ	地盤の状況	樹木の種類	活断層との距離	
5	3	5	4	

4.3 結果に対する考察

以上の結果より、次のような知見を得ることができた。

- ・整合度の制約を70%とした場合に残留する傾斜度, 斜面形状, 表土の厚さ等の属性は、表1(整合度90%)においても残留する組み合わせが多い。従って整合度の制約が変化しても説明力の高い属性にはかなりの程度での一貫性が認められる。すなわち一般的な認識に非常に合致した結果が得られた。
- ・がけ崩れの発生, 非発生を決定する要因として最も説明力の高い属性には、傾斜度, 表土の厚さ, 活断層との距離のように定量的な測定可能なものが多い。対照的に、定性的な評価しかできない属性(風化状況など)については、現象の説明力が相対的に低い。急傾斜地危険箇所の管理においては、地形属性の中に調査者の主観に左右されやすいものが多いことが問題とされてきた。しかし本研究の結果から、崩壊のポテンシャルを評価する際には定量的に測定可能な属性を優先的に調査すべきことが明らかとなった。
- ・現在の急傾斜地危険箇所調査の調査項目のうち、7属性によっても90%のデータについて発生, 非発生の判別が可能であることが明らかとなった。このことは比較的少数の地形要因によって急傾斜地の潜在的な危険要因を評価することが可能なことを示していると考えられる。

5. おわりに

本研究では山口県の急傾斜地崩壊危険箇所データからデータマイニングの手法により崩壊現象の特徴要因を抽出することを試みた。本研究で特定された地形要因は斜面崩壊の潜在的危険性を評価するのに有用であると考えられる。今後は崩壊を直接引き起こす誘因である降雨要因を含めた分析を実施する。

謝辞

甲南大学大学院の石井練一氏には、ラフ集合を利用したアルゴリズムによる計算について多大なご協力をいただきました。付して感謝いたします。

参考文献

- 1) 中村昭：ラフ集合—その基本概念と知識情報, 数理科学, No.373, pp.78-83, 1994.
- 2) 石井練一：ラフ集合論を用いたデータベースからのルール抽出, 甲南大学修士論文, 1999.

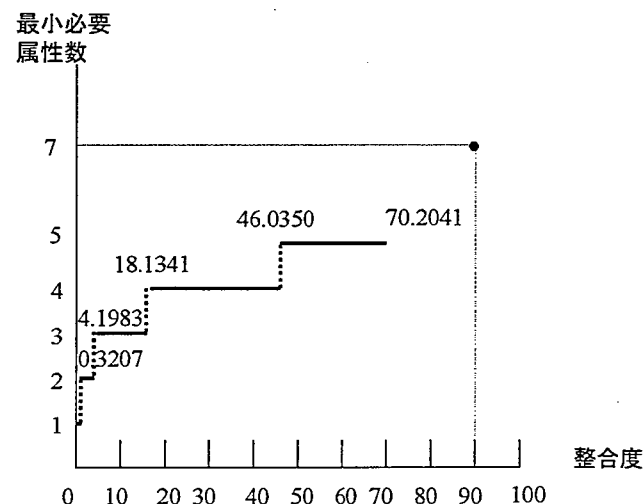


図2 斜面データにおける整合度の制約と最小必要属性数の相関