

ラフ集合を用いたデータマイニングによる がけ崩れ発生要因の抽出に関する研究

榎原 弘之¹・倉本 和正¹・菊池 英明²・中山 弘隆³・鉄賀 博己⁴・古川 浩平⁵

¹正会員 修 (工) 山口大学助手 工学部社会建設工学科 (〒755-8611 山口県宇部市常盤台 2-16-1)

²正会員 博 (工) 八千代エンジニアリング (株) 東京事業部水工部 (〒153-0061 東京都目黒区中目黒 1-10-23)

³工博 甲南大学教授 理学部応用数学科 (〒658-0072 兵庫県神戸市東灘区岡本 8-9-1)

⁴山口県土木建築部砂防課傾斜地保全係長 (〒753- 8501 山口県山口市滝町 1-1)

⁵正会員 工博 山口大学教授 工学部社会建設工学科 (〒755-8611 山口県宇部市常盤台 2-16-1)

本論文では、がけ崩れに寄与する重要要因の抽出を目的として、ラフ集合を用いたデータマイニングを山口県のがけ崩れ発生・非発生データに対して実施する。まず、地形要因データにおける重要要因を抽出し、比較的少数の要因によって、大部分のがけ崩れ発生・非発生を矛盾なく説明できることを示す。さらに、地形要因と降雨要因を組み合わせたデータへも同様の手法を適用し、降雨時に警戒対象とすべきがけの選別手法を示した。本論文により、ラフ集合によるデータマイニングを土砂災害の発生・非発生データの分析に用いることができることが明らかとなった。

Key Words : data mining, slope failure, rough sets, geological data, rainfall data

1. はじめに

土石流や地すべり、がけ崩れなどの土砂災害が毎年全国各地で発生し、尊い人命や貴重な財産が失われている。これら土砂災害は、いわゆる素因としての地形・地質要因と誘因としての降雨要因の双方が影響し引き起こされる。がけ崩れの場合、素因となる地山条件として土質・岩質などの物性、地質構造、地下水・表面水等の3種類が挙げられる。表層の小規模な崩壊は物性と地下水に依存し、大規模な崩壊は地質構造に依存するとされる¹⁾が、そのメカニズムは複雑であり、解明されていない点も多い。そこで、過去の災害データを分析し、危険性を判定する必要がある。特に、災害発生時及び非発生時の膨大な降雨データや危険箇所データの因果関係、特徴を抽出することは斜面の崩壊現象を解明する上で重要である。

地形・地質要因については、建設省によって土石流危険渓流、急傾斜地崩壊危険箇所等の調査が数年サイクルで実施され、危険度評価が行われている。一方、降雨要因については、気象庁のアメダスデータの膨大な蓄積が存在する。

しかし、このように、大量のデータが蓄積され、かつ近年著しく情報処理技術が発達しているにもかかわらず、地形・地質要因と降雨要因を組み合わせ、地

形ごとに災害の発生する降雨の条件を特定した例は少ない。その理由としては、各要因の崩壊現象に与える影響のウェイトの決定が困難であることが挙げられる。しかし、崩壊に関わる素因・誘因には何らかの因果関係があり、特徴的な要因があることは事実である。従って、これまでに集積された大量のデータを解析し、要因相互の複雑な因果関係を見出し、斜面崩壊の特徴的な要因を抽出可能にする手法が求められている。

古川らはこれまでに、土石流や高速道路の切土のり面崩壊の発生予測にニューラルネットワークを適用した^{2),3)}。しかし、ニューラルネットワークで構成されるネットワーク構造はブラックボックスとして取り扱われ、個別の要因の特徴抽出は難しい。加えて、学習データの精度が学習結果に大きく影響することから事前に学習データの選定が必要となるが、有効な選定方法が少ないなど問題も多い。

本論文では、がけ崩れの原因となる地形・地質要因、降雨要因の分析にデータマイニングの手法を適用し、がけ崩れ発生の特徴的な要因の抽出を試みる。具体的には、ラフ集合の概念を用いることによりノイズや矛盾を含む生データから有効な要素のみを取り出し、本質的に重要な要因のみを抽出する。特徴的な要因を抽出することで、従来行われてきた急傾斜地崩壊危険箇所点検の簡素化が可能であると考えられる。具体的な

ケーススタディ域としては、急傾斜地崩壊危険箇所数が全国5位であり⁴⁾、予測システムの重要性が高い山口県を対象とする。

2. ラフ集合によるデータマイニング

(1) ラフ集合

ラフ集合はPawlek⁵⁾によって提案された概念である。がけ崩れを例に、ラフ集合の概念を簡略に示す^{6),7),8),9),10),11)}。がけ崩れに関しては、個別のがけ(地形要因のみを考慮した場合)またはがけと個別の降雨を組み合わせたもの(地形・降雨両要因を考慮した場合)が一つ一つのデータとなる。個々のデータは、地形要因や降雨要因に関して複数の属性を有する。

図-1において、大きな四角形はがけに関するすべてのデータの集合を意味するとする。また四角形内部の点は個別のデータを表すものとし、黒点は発生(崩壊)、白点は非発生(未崩壊)を表すとする。内部の楕円は、発生データの集合を意味している。

個々のデータは各属性のいずれかのカテゴリーに属しているため、属性によってデータを分類することができる。図-1においては、属性 R_1 によってデータの集合は行方向に3つに分割され、属性 R_2 により列方向に4つに分割されている。従って、 R_1, R_2 の2つの属性によって、データの集合は12種類の領域に分割されることになる。

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、崩壊が発生するがけと発生しないがけ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発生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。図-1において、発生データの集合はラフ集合である。

属性の組み合わせが与えられたとき、各属性のカテゴリーが同一であっても、結果属性(上の場合においてはがけ崩れ発生、非発生)の異なるデータを「矛盾データ」と呼ぶこととする。図-1においては、灰色の領域に含まれるデータが矛盾データである。それに対して、同じ領域内のデータの結果属性がすべて同一である場合、それらのデータを「整合データ」と呼ぶこととする。図-1の白い領域に含まれるデータが整合データである。

(2) ラフ集合による重要要因の抽出

属性の組み合わせを評価する指標として、整合度を以下のように定義する。

属性 R_2 による分類(列方向)

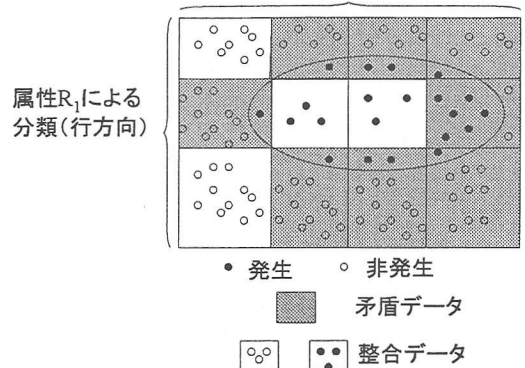


図-1 ラフ集合の概念図

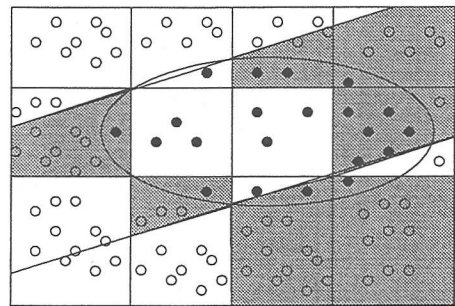


図-2 属性の増加による整合度の増加

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

属性数を増加させることにより、整合度は増加する。図-2はそれを概念的に示したものである。図-1と比較すると、2本の斜線によりデータの集合が20種類の領域に細分化されている。これは新たな属性が追加されたことを意味する。属性数が増加することにより、整合データ(同一領域内に白丸のみまたは黒丸のみが所属するデータ)が増加し、他方矛盾データが減少する。

広兼ら¹²⁾は、ラフ集合を斜面の崩壊危険度診断事例からのルール抽出に用いている。その際、矛盾データを許容しない(整合度を100%とする)ことを制約条件としている。専門家による診断は、ある種の法則性に従っていると考えられることから、矛盾データの数は少なく、比較的高い整合度を達成することができると考えられる。

一方、本論文が対象とする実際のがけ崩れにおいては、事前の調査では同様な要因を有するとされた斜面が、ある降雨によって一方は崩壊し、他方は崩壊しないことが起き得る。すなわち、矛盾データが生じる可能性が高い。さらには、すべての属性を加えても整合度が100%とならないこともあり得る。その場合、属性(以下要因と呼ぶ)数を増加させて整合度を高めることは、必ずしも有意な知識をもたらさない。むしろ、

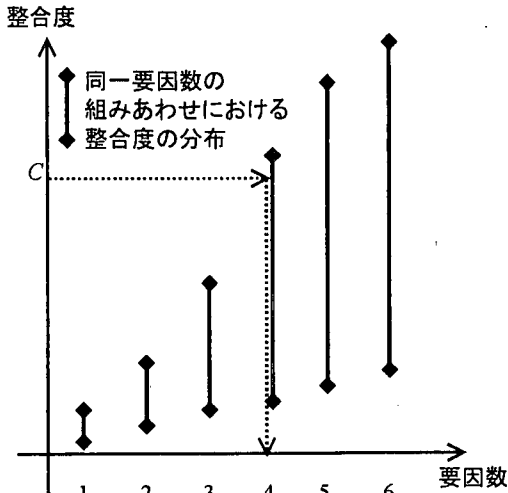


図-3 要因数と整合度の関係

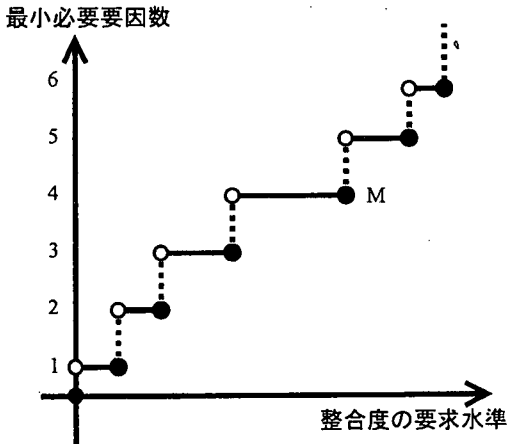


図-4 整合度の要求水準と必要最小要因数との関係

最低限必要な整合度をあらかじめ設定した上で、その整合度を最小の要因数で達成する組み合わせを求めることにより、がけ崩れに寄与する真に重要な要因を特定することができる。以下ではこのときの要因数を（整合度の要求水準に対する）最小必要要因数と呼ぶ。

最小必要要因数の決定方法を以下に示す。図-3は、同数の要因の組み合わせによって実現される整合度の分布を概念的に示したものである。同じ要因数であっても、組み合わせにより整合度は異なる。ここで、整合度の要求水準が図-3中のCであったとする。整合度Cを達成するためには、4つ以上の要因が必要になることがわかる。従って、整合度の要求水準がCの場合の最小必要要因数は4である。以上により、整合度の要求水準に対する最小必要要因数は、図-4に示すような非連続関数として与えられる。整合度の要求水準を高

表-1 急傾斜地危険箇所データに含まれる要因

| 要因 |
|---|
| 傾斜度(10°)、高さ(10m)、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(10cm)、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(20km) |

くするほど、必要な要因の数は増加する。

本論文では、中山ら^{13),14)}が提案している重要要因抽出手法を用いて、山口県のがけ崩れのデータに対するデータマイニングを実施する。以下に具体的な手順を示す。

- ① 整合度の要求水準を設定する。
- ② 要因数が1つの場合の整合度を求め、要求水準と比較する。少なくとも1つの要因において整合度が要求水準を上回った場合、最小必要要因数は1となる。
- ③ すべての要因について、整合度が要求水準を下回っていた場合、2つの要因の組み合わせに関する整合度を求め、要求水準と比較する。
- ④ 以下順次要因数を増加させ、整合度が要求水準を初めて上回った時点における要因数を最小必要要因数とする。

さらに、最小必要要因数の組み合わせのうち、整合度の上位数種類に含まれる要因を、重要要因として抽出する。図-4のMは、最小必要要因数が4の場合における整合度最大の組み合わせである。

3. 地形要因データに対する適用結果

(1) データの概要

平成8年度の山口県の急傾斜地崩壊危険箇所点検結果から得られる地形データと、昭和50年から平成10年までの24年間の災害報告書データを使用する。この2種類のデータベースを照合し、災害報告書のがけ崩れ発生箇所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。一部要因の欠けたデータを削除した結果、データ数は発生476箇所、非発生2957箇所、計3433箇所となった。また各データの有する要因は表-1に示す20種類である。

表-1の要因のうち、傾斜度、高さ、表土の厚さ、対

象斜面と活断層距離の4要因は元来連続値であるが、ラフ集合によるデータマイニングを実施するためには連続値を離散的なカテゴリーに変更(離散化)する必要がある。本論文では、各要因において、大半のデータが10個以内のカテゴリーに該当するように、表-1の括弧内に示す間隔で離散化した。

(2) 重要要因の抽出

表-1に示す20種類の要因すべてを用いた場合の整合度は99.65%であった。これは、すべての要因を用いても矛盾データが生じることを示している。次に、整合度の要求水準を80%とした場合、最小必要要因数は5となった。表-2は、5要因の組み合わせのうちで整合度が最も高い5通りを示したものである。最も整合度の高い組み合わせは「傾斜度、斜面方位、斜面形状、表土の厚さ、植生の種類」であり、そのときの整合度は82.70%である。一方、表-2の下部にいずれの組み合わせにも残存しない要因を示す。これらの要因のかけ崩れに与える影響は相対的に低いと考えられる。

表-2の結果は、わずか5種類の要因によって、全体の8割以上の箇所におけるかけ崩れ発生、非発生を矛盾なく説明できることを示している。それ以外の要因を含む組み合わせによって、82.70%を上回る整合度が達成されることはないため、その他の要因のかけ崩れに対する影響は相対的に小さいと考えられる。

しかし、要因の重要性を判断するに当たって、単に要因を2つの集合に分類するのみでは不十分な場合もあり得る。そこで本論文では、以下のような手順で要因抽出を繰り返すことによって要因を2つ以上のグループに階層分類する。

- ① 適当な最小必要要因数を設定し、その要因数で最も高い整合度を達成することができる要因の組み合わせを求める。
- ② ①で残留した要因を「階層1」に分類した上データから除去する。
- ③ 更新された(一部要因の削除された)データの下で再度最小必要要因数を設定し、残留する要因を「階層2」に分類した上、データから除去する。
- ④ 以下適宜繰り返す。

これによって、要因を説明力の強さにより2つ以上の階層に分類することができる。表-3は、要因数を5つに限定した場合に残存した要因を階層1、その他の要因のうち、さらに要因数を5つに限定した場合に残留する要因を階層2、階層1、2のいずれにも分類されなかった要因を階層3として分類したものである。各階層に属する要因を抽出したときの整合度はそれぞれ階層1(最も高い整合度を実現する5要因)において

表-2 地形データに対する適用結果

| 順位 | 傾斜度 | 高さ | 斜面方位 | 斜面形状 | 表土の厚さ | 植生の種類 | 樹木の樹齢 | 対象斜面と活断層距離 | 整合度(%) |
|----|-----|----|------|------|-------|-------|-------|------------|--------|
| 1 | ○ | — | ○ | ○ | ○ | ○ | — | — | 82.70 |
| 2 | ○ | — | ○ | ○ | ○ | — | — | ○ | 82.20 |
| 3 | ○ | — | ○ | ○ | ○ | — | ○ | — | 81.74 |
| 4 | ○ | ○ | ○ | ○ | ○ | — | — | — | 80.43 |
| 5 | — | — | ○ | ○ | ○ | ○ | — | ○ | 80.02 |

残存しない要因

横断形状、遷急線、地表の状況、地盤の状況、岩盤の状況、斜面と不連続面傾斜関係、断層・破砕帯の有無、風化状況、伐採根の状況、調査斜面・崩壊履歴、隣接斜面・崩壊履歴、湧水

表-3 要因の階層分類の結果

| | |
|-----|--|
| 階層1 | 傾斜度、斜面方位、斜面形状、表土の厚さ、植生の種類 (82.70%) |
| 階層2 | 高さ、横断形状、斜面と不連続面傾斜関係、樹齢、対象斜面と活断層距離 (51.50%) |
| 階層3 | 遷急線、地表の状況、地盤の状況、岩盤の状況、断層・破砕帯の有無、風化状況、伐採根の状況、調査斜面・崩壊履歴、隣接斜面・崩壊履歴、湧水 |

82.70%、階層2(階層1に属する要因を除いたうちで、最も高い整合度を実現する5要因)で51.50%である。階層分類の結果から、以下のような知見を得ることができた。

- ・階層1に属する要因のうち、傾斜度、表土の厚さの2要因は、定量的に測定可能な要因である。かけ崩れに対するこれらの要因の重要性はモデル分析からも指摘^{15),16)}されており、本論文の結果は実際のかけ崩れ発生・非発生データによりこれを実証するものといえる。
- ・一方、定性的な要因のうち、斜面方位、斜面形状、植生の種類などの要因の重要性も明らかとなった。
- ・風化状況、湧水等の要因は階層3に分類され、重要性が低い。このことから、少なくとも現在の調査方法によって得られた風化状況や湧水に関するデータ

によって、がけの崩壊に対するポテンシャルを評価することは困難であるといえよう。

5 要因の組み合わせのうちで整合度が最も高い組み合わせにおいて、整合データ、矛盾データにそれぞれ含まれる発生、非発生データの割合を示したのが表-4である。整合データにおいては、発生データの割合は10%以下なのに対して、矛盾データにおいては30%以上となる。すなわち、矛盾データには多くの崩壊データが含まれている。この結果から、がけ崩れに関するデータマイニングにおいては、整合データのみでなく、矛盾データも分析の対象とすべきことがわかる。これらの矛盾データでは、上の5要因が同一ながけが2つ以上存在し、その中がけ崩れが発生したがけと発生しなかったがけがともに存在する。図-5にその一例を示す。

このような矛盾データが生じる理由としては、①それぞれのがけが経験した降雨が異なる、②5つの要因以外の地形や地質に関する要因が影響している、という2種類が考えられる。そこで、さらに詳細な検討を実施する場合は、すべてのがけを対象とするのではなく、矛盾データに含まれるがけを重点的に調査すればよいと考えられる。以上のように、ラフ集合によるデータマイニングは、データの集合を、説明の容易なデータ(整合データ)と、詳細な調査が必要なデータ(矛盾データ)に分類することができる。

(3) 崩壊危険度判定における重要要因との比較

急傾斜地崩壊危険箇所点検においては、専門家の知識を基に、ファジィ積分による崩壊危険度判定を実施している¹⁷⁾。山口県では、全国的に実施されている3段階の評価ランクの他、最も危険な斜面(AAランク)を加えた4段階評価を行っている。

危険度判定において重視されている要因が、前節で示したがけ崩れ発生の重要要因と一致すれば、現行の危険度判定は有効であると考えられる。そこで本節では、がけ崩れの発生・非発生に代わり、危険度判定を結果属性とした場合の重要要因を抽出し、前節の結果と比較する。

地形要因データにおいて、危険度を結果属性とすれば、ラフ集合を用いたデータマイニングにより、危険度判定において重視されている要因を抽出することができる。表-5は危険度判定(対策施工箇所を除く)において5要因の組み合わせの中で整合度が最も高い5通りを示したものである。実際のがけ崩れの有無を結果属性とした場合(表-2、表-3)と、危険度を結果属性とした場合(表-5)のデータマイニングの結果を比較することにより、以下のような知見を得ることができた。

表-4 整合データ及び矛盾データ中の発生、非発生データ(箇所)の割合(括弧内はデータの実数)

| | 発生 | 非発生 |
|-------|-------------|--------------|
| 整合データ | 9.16%(260) | 90.84%(2579) |
| 矛盾データ | 36.36%(216) | 63.64%(378) |

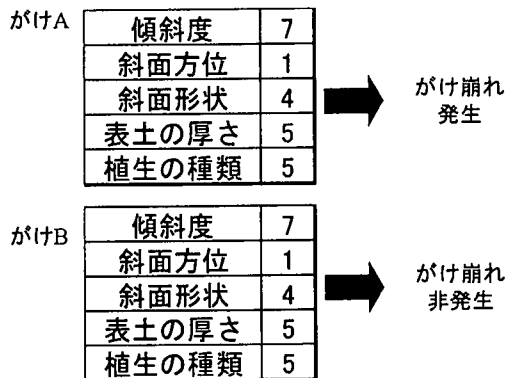


図-5 地形要因によるデータマイニングにおける矛盾データの例

表-5 危険度判定における重要要因

| 要因 | 傾斜度 | 高さ | 斜面方位 | 斜面形状 | 横断形状 | 表土の厚さ | 地盤の状況 | 樹齢 | 対象斜面と活断層距離 | 整合度 |
|----|-----|----|------|------|------|-------|-------|----|------------|-------|
| 順位 | | | | | | | | | | |
| 1 | - | ○ | ○ | ○ | - | ○ | - | ○ | - | 84.01 |
| 2 | - | - | ○ | ○ | ○ | ○ | - | ○ | - | 83.45 |
| 3 | - | - | ○ | ○ | - | ○ | - | ○ | ○ | 83.14 |
| 4 | - | - | ○ | ○ | - | ○ | ○ | ○ | - | 82.45 |
| 5 | ○ | - | ○ | ○ | - | ○ | - | ○ | ○ | 82.39 |

- 危険度判定においては、5種類の組み合わせの中で、傾斜度が含まれるものは1種類のみであった。このことから、実際のがけ崩れ(表-2、表-3)において傾斜度がほぼ常に重要な要因に含まれているのとは対照的に、傾斜度が危険度判定に及ぼす影響は相対的に低いと考えられる。この理由としては、危険度判定のためのメンバーシップ関数の合成に当たって、傾斜度のウェイトが比較的小さいことが考えられる¹⁷⁾。
- 斜面方位、斜面形状、表土の厚さ等の要因は、実際のがけ崩れ、危険度判定の双方において、重要な要因としてほぼ常に残存した。
- 危険度判定においては、横断形状、地盤の状況のよ

うに、実際のがけ崩れの有無を結果属性とした場合には重要要因とならないものが残存することがあった。このことから、危険度判定においては、調査項目全般が万遍なく反映されていると考えられる。このことは、実際のがけ崩れにおいて属性間で重要度に明確な差が生じることとは対照的である。

以上のように、実際のがけ崩れの有無を結果属性とした場合と、危険度を結果属性とした場合では、抽出された重要要因に一部差違が見られた。実際のがけ崩れのポテンシャルの高さが危険度判定に反映されるべきことを考慮すれば、現状の危険度判定手法に対しては、傾斜度をより重視することや、要因によって判定に影響を及ぼすウェイトに差を設定する必要があるといえよう。

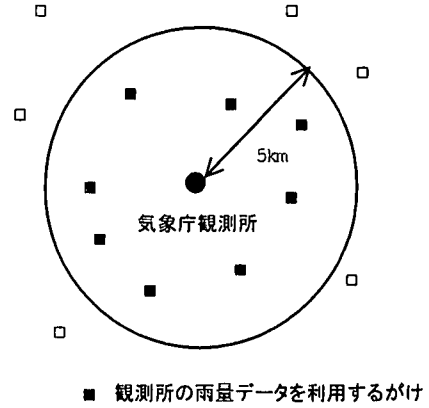


図-6 観測所と対象となるがけの位置関係

4. 地形・降雨要因データに対する適用結果

(1) 降雨要因を組み合わせたデータマイニング

過去の記録から、複数箇所でがけ崩れが発生したような豪雨時においても、急傾斜地崩壊危険箇所すべてにおいて崩壊が発生したわけではないことが明らかとなっている。そこで本章では、降雨要因と3.で抽出した重要な地形要因を組み合わせてデータマイニングを実施することにより、同一降雨におけるがけ崩れ発生・非発生データの分類を試みる。対象としては、山口県内でも特にがけ崩れの多発する地域の一つである下関市を取り上げ、この地域の代表的岩種である安山岩のがけのデータを用いる。

降雨要因を組み合わせるためには、信頼し得る降雨データが入手可能ながけに分析の対象を限定する必要がある。本論文では、土砂災害の予測に関する既往の研究²⁾に基づき、観測所からの直線距離が5km以内のがけにおける降雨が、観測所の実測値と一致していると仮定する(図-6参照)。降雨データは気象庁下関観測所の測定データを用いる。対象となるがけの総数は52箇所である。

地形要因については、表-2において整合度が最大であった傾斜度、斜面方位、斜面形状、表土の厚さ、植生の種類の5要因のみを用いることとする。

(2) 降雨データの設定

本章では、対象とする箇所のいずれかではがけ崩れが発生した昭和50年から平成10年までの24年間における計33回の降雨について、地形要因を組み合わせたデータマイニングにより、がけ崩れの発生した箇所、発生しなかった箇所を分類する。対象データは、上述した33回の降雨のそれぞれについて、その降雨でがけ崩

表-6 降雨データの設定

| 降雨 | 箇所 | 箇所1 | 箇所2 | 箇所3 | 箇所4 | 箇所5 | 箇所6 |
|--------------|----|-----|-----|-----|-----|-----|-----|
| 崩壊推定時 降雨1 | | ○ | ● | ○ | ○ | ○ | ○ |
| 崩壊推定時 降雨2 | | ○ | - | ● | ○ | ○ | ○ |
| 崩壊推定時 降雨3 | | ○ | - | - | ● | ○ | ○ |
| 崩壊推定時 降雨4 | | ○ | - | - | - | ○ | ● |

○... 非発生, ●... 発生

-... データマイニングに用いないデータ

れが発生した箇所を発生、その他の箇所を非発生としたものである。ただし、ひとたびがけ崩れが発生した斜面においては、対策工が施工されるなど、条件が変化することから、がけ崩れ発生以降のデータは用いないものとする。表-6はこれを模式的に示したものである。

次に、一連の降雨のどの時点の雨量を要因として採用するかが問題となる。倉本ら¹⁸⁾は、一連降雨のピーク時刻から12時間以上遅れて報告されているがけ崩れの多くが、早朝に報告されていることに着目し、これらの斜面崩壊の多くは実際には夜間に発生したものと推測している。そこで本論文では、以下のように定義する「崩壊推定時」の降雨要因が崩壊に直接寄与すると仮定し、分析を実施する。

定義

崩壊推定時

崩壊データのうち、崩壊報告時が午前6時台から午前

9 時台のデータに関して、前日午後 8 時から崩壊報告時までの時間で時間雨量が最大の時間を崩壊推定時とする。それ以外のデータに関しては、崩壊報告時を崩壊推定時とする。

降雨要因としては、時間雨量、累積雨量の 2 要因を用いる。また、時間雨量を 10mm ごと、累積雨量を 50mm ごとにカテゴリー分けした。

(3) 適用結果

まず、降雨要因と地形要因を組み合わせた場合の重要要因を抽出し、地形要因のみのデータの場合と比較する。

表-7 は、5 要因の組み合わせのうち、整合度が最大となる 2 種類を示したものである。降雨に関する 2 要因（時間雨量、累積雨量）が、ともに重要要因となっており、崩壊推定時における降雨要因ががけ崩れに影響を与えていることが明らかとなった。また計 45 回のがけ崩れのうち、20 回について整合データとして捕捉することができた。

表-7 の 1 番目の組み合わせ（傾斜度、斜面方位、斜面形状、時間雨量、累積雨量）における矛盾データに着目してさらに詳細な分析を行う。降雨データと地形データを組み合わせた場合、矛盾には次の 2 つのタイプが存在する（図-7 参照）。

タイプ 1：同時点・異地点間の矛盾

同一の降雨において、類似した地形要因を有するがけの中に、がけ崩れが発生したがけと、発生しなかったがけがともに存在する。

タイプ 2：同地点・異時点間の矛盾

同一のがけにおいて、過去の同規模の降雨では崩壊しなかったにもかかわらず、ある降雨によって崩壊する。

タイプ 1 の矛盾データについては、地形要因をさらに詳細に検討することにより、双方のがけの差異を見出すことができる可能性がある。一方、タイプ 2 の矛盾データについては、崩壊推定時以前の数時間にわたる降雨履歴の検討が必要であろう。ラフ集合によるデータマイニングは、このように大量のデータから詳細な検討が必要なデータを分類する役割を果たすことができる。

(4) 警戒対象の選別

がけ崩れが発生する可能性があるような降雨時においては、危険性の高いがけを重点的に警戒し、必要な場合は周辺住民を避難させるなどの措置が必要となる。そのためには、降雨との組み合わせでがけ崩れ発生の

表-7 降雨要因を組み合わせた場合の重要要因

| 傾斜度 | 斜面方位 | 斜面形状 | 表土の厚さ | 時間雨量 | 累積雨量 | 整合度 (%) |
|-----|------|------|-------|------|------|---------|
| ○ | ○ | ○ | — | ○ | ○ | 92.98 |
| — | ○ | ○ | ○ | ○ | ○ | 92.98 |

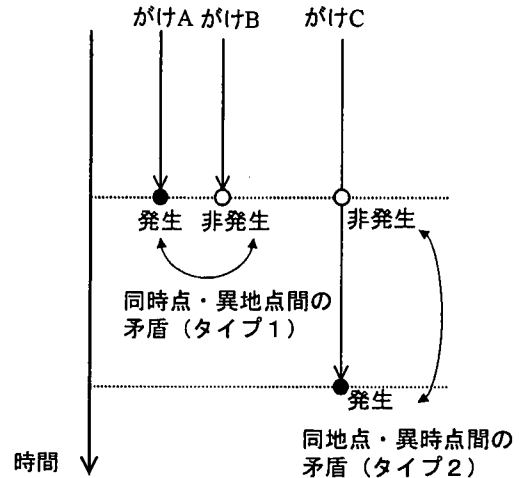


図-7 同時点・異地点間の矛盾 (タイプ 1) 及び同地点・異時点間の矛盾 (タイプ 2)

可能性が高いケースを選別する必要がある。第一に警戒対象に含めるべきケースは、整合データに含まれる発生データである。しかし、前節に示した 5 要因の組み合わせでは、45 回のがけ崩れのうち整合データに含まれるのは 20 回に限られている。

それに対し、矛盾データに該当する箇所全体を、潜在的に崩壊の危険性を有するとみなし、警戒対象とすることも考えられる。表-8 は、矛盾データに該当するがけのうち、非発生であった箇所について、その後の経過を分析したものである。左列は、0-10mm の降雨で崩壊しなかったがけ（類似した他のがけにおいてがけ崩れが発生したがけ）に関して、その後の経過を示したものである。15 箇所中 14 箇所はその後の降雨で崩壊しており、さらにそのうち 11 箇所は時間雨量 20mm 以下の比較的小さい降雨で崩壊している。調査期間中で未だ崩壊していないのは 1 箇所のみである。他方右列は 30-40mm の降雨で崩壊しなかったがけ 13 箇所の経過を示している。この場合も同様に、11 箇所において既にながけ崩れが発生している。以上より、矛盾データに該当し、現時点において未だ崩壊を経験していないがけにおいては、崩壊のポテンシャルは非常に高いと考えられる。

そこで、整合データとして捕捉されている発生デー

表-8 崩壊非発生箇所その後の経過

| 箇所数 | 推定時 時間雨量 | |
|------------------------|---------------|--------------|
| | 0-10mm | 30-40mm |
| 矛盾データ 該当箇所数 | 15 | 13 |
| 20mm以上の 時間雨量で 崩壊 | 3 (20%) | 7 (53.8%) |
| 20mm以下の 時間雨量で 崩壊 | 11 (73.3%) | 4 (30.8%) |
| 崩壊非発生 | 1 (6.7%) | 2 (15.4%) |

たと、非発生のデータがある程度含まれている矛盾データの全体をがけ崩れ発生の潜在的危険性があるとみなし、これら全体を「警戒対象データ」と呼ぶ。その際、矛盾データ中の非発生データの割合として、非発生エラーレベルを以下のように定義する。

定義

非発生エラーレベル

与えられた要因の組み合わせに関して、各要因のカテゴリがすべて同一のデータの集合を I とする。 I に含まれる非発生データの割合を、 I の非発生エラーレベルと呼ぶ。

I の非発生エラーレベルがあらかじめ設定した上限を下回っている場合、 I は警戒対象データに含まれる。この非発生エラーレベルの上限を、許容エラーレベル (EL) と呼ぶ。 EL=0.0 のときの警戒対象データの集合は、整合データ中の発生データに一致する。

図-8 は、前節の5要因 (傾斜度, 斜面方位, 斜面形状, 時間雨量, 累積雨量) に関して、 EL=0.0 及び EL=0.5 の場合に、全発生データのうちどれだけのデータが警戒対象データとして捕捉されたかを崩壊推定時の時間雨量別に示したものである。時間雨量 10-20mm 及び 20-30mm の場合、 EL=0.0 において既にすべてのデータが捕捉されていることがわかる。一方時間雨量が 10mm 以下及び 30mm 以上の場合は、許容エラーレベルを増加させることにより、捕捉される発生データが増加する。しかし、崩壊推定時の時間雨量が 10mm 以下の場合は、依然として捕捉されないデータも多い。

また、図-9 は、 EL=0.0 において捕捉されない発生データが存在する時間雨量が 10mm 以下及び 30mm 以上の場合について、発生データ、非発生データの割合を示したものである。時間雨量が 30mm 以上の場合は発生データが全体の4割近くを占めるのに対し、 10mm

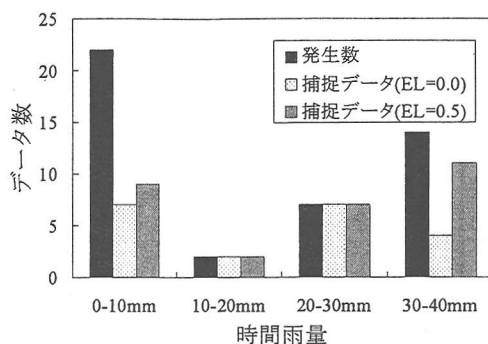


図-8 許容エラーレベルと捕捉データ数の関係

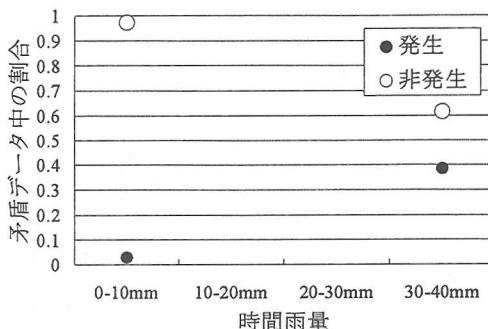


図-9 矛盾データにおける発生データ、非発生データの割合

以下の場合ほとんどが非発生データである。このことから、時間雨量の小さいときに発生するがけ崩れは、突発性が高く、予測が特に困難であることが分かる。従って、がけ崩れの予測にあたっては、時間雨量が小さいときのがけ崩れと大きいときのがけ崩れを分離し、それぞれに適合した予測モデルを構築する必要があると考えられる。

5. おわりに

以上、本論文では山口県の急傾斜地崩壊危険箇所点検データと、災害報告書、及び気象庁の降雨データを用いて、ラフ集合によるデータマイニングを実施し、重要要因を抽出した。地形要因データのみの重要要因抽出では、がけ崩れに関する経験的知識と合致した結果が得られた。また地形要因と降雨要因を組み合わせたデータを用いたデータマイニングでは、矛盾データを分析することにより潜在的ながけ崩れ発生の危険性の高い箇所を特定することができることを明らかにした。さらに予測困難な突発的ながけ崩れの存在を示すことができた。

本論文により、ラフ集合によるデータマイニングを

土砂災害の発生・非発生データの分析に用いることができることが明らかとなった。実際のがけ崩れの予測や管理への適用に関しては、以下のような可能性が考えられる。

①予測システムにおける入力要因の選別

重要要因として抽出された要因のみを用いることにより、ノイズの比較的少ない予測システムを構築することができると考えられる。

②重点的な警戒対象の絞り込み

より詳細な調査を実施する場合、対象を矛盾データに限定することにより、効率的な管理が可能となる。同様に、降雨時の重点的な警戒対象についても絞り込みが可能である。

今後の課題としては、がけ崩れの発生・非発生の条件となるルールの抽出が挙げられる。

参考文献

- 1) 奥園誠之：斜面防災100のポイント，鹿島出版会，1986。
- 2) 荒木義則，古川浩平，松井範明，大木戸孝也，石川芳治，水山高久：ニューラルネットワークを用いた土石流危険渓流における土砂崩壊のリアルタイム発生予測に関する研究，土木学会論文集，No.581/VI-37，pp.107-121，1997。
- 3) 菊池英明，古川浩平，大前信司，奥園誠之，西岡勲：ニューラルネットワークによる切土のり面の豪雨時のリアルタイム崩壊予測と予測に用いる降雨要因に関する一考察，土木学会論文集，No.616/VI-42，pp.63-76，1999。
- 4) 建設省河川局砂防部：砂防便覧（平成9年度版），1998。
- 5) Pawlek, Z.: Rough Sets, International Journal of Computer and Information Sciences, Vol.11, pp.341-356, 1982。
- 6) 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。
- 7) 中村昭：ラフ集合と論理・推論，数理科学，No.374，pp.86-91，1994。
- 8) 横森貴，小林聡：ラフ集合と意思決定，数理科学，No.375，pp.76-83，1994。
- 9) 谷田則幸，米村崇，横森貴：ラフ集合とデータ解析，数理科学，No.376，pp.78-85，1994。
- 10) 津本周作，田中博：ラフ集合と学習，数理科学，No.377，pp.78-83，1994。
- 11) ラフ集合とエキスパートシステム，数理科学，No.378，pp.76-83，1994。
- 12) 広兼道幸，古田均，中井真司，三雲是宏：斜面の崩壊危険度診断事例からのラフ集合を用いたルール型知識の抽出方法，土木学会論文集，No.582/III-41，pp.285-294，1997。
- 13) 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。
- 14) Nakayama, H., Hattori, Y. and Ishii, R.: Rule Extraction based on Rough Set Theory and its Application to Medical Data Analysis, Proceedings of 1999 IEEE International Conference on Systems, Man, and Cybernetics, Vol.5, pp.924-929, 1999。
- 15) 三森利昭，大倉陽一，堀江保夫：諸パラメーターが表層崩壊に与える影響の数値実験的検討，新砂防，Vol.46，No.1，pp.3-12，1993。
- 16) 平松晋也，水山高久，奥山清一：崩壊モデルを用いた各種崩壊要因の影響度評価，新砂防，Vol.43，No.5，pp.11-18，1991。
- 17) 全国地すべりがけ崩れ対策協議会：がけ崩れ対策の手引き—急傾斜地崩壊対策事業の実務—，1999。
- 18) 倉本和正，菊池英明，鉄賀博己，荒木義則，古川浩平：山口県の急傾斜地におけるがけ崩れ発生の現状，砂防学会研究発表会概要集，No.27，pp.131-132，1999。

(1999. 12. 8 受付)

A STUDY ON DETECTION OF CAUSES OF SLOPE FAILURES BY DATA MINING APPROACH USING ROUGH SETS

Hiroyuki SAKAKIBARA, Kazumasa KURAMOTO, Hideaki KIKUCHI, Hirotaka NAKAYAMA, Hiromi TETSUGA and Kohei FURUKAWA

In this paper, data mining is applied to the data on slope failures in Yamaguchi prefecture, for the purpose of the detection of critical causes of slope failures. First, the methodology based on rough set theory is applied to geological data, and it is shown that relatively small number of causes can explain most of the slope failures. Next, the same methodology is applied to the data with geological and rainfall factors, and the classification of slopes for preliminary warning and evacuation is proposed. This paper shows the effectiveness of data mining using rough sets in the analysis of slope failures.