Decision treeによる地すべり発生流域の推定とその検証 —ASTERデータを用いて—

Estimation and evaluation of drainage basins with landslides using decision tree technique and ASTER data

齋藤 仁^{a)*}・中山大地^{a)}・松山 洋^{a)}

Hitoshi SAITO, Daichi NAKAYAMA and Hiroshi MATSUYAMA

Abstract

In this study, drainage basins with landslides in the Akaishi Mountains are estimated using the decision tree technique which is a kind of data mining technique. Decision tree technique is characterized by the estimation processes which are explicitly and quantitatively represented by Tree-structures. The purpose of this study is to clarify the relationship between landslide and topography/lithology using Tree-structure. To estimate drainage basins with landslides, topographic characteristics (elevation, slope, profile curvature, plan curvature, dissection and undissection) and geological data were used as explanatory variables. Topographic characteristics were calculated from the DEMs (Digital Elevation Models) of EOS-Terra/ASTER. Objective variable is landslide area data.

Correctness of the estimation is 71.2%. The estimation processes quantitatively indicate that the lithology, plan curvature and the slope angle play important roles for estimating drainage basins with landslides. In addition, drainage basins where landslides occur frequently tend to have concave terrain and lower slope angle. While in the Setogawa group, distribution of landslides is not uniform, and landslides are concentrated in the drainage basins where slope angles are relatively small.

It is concluded that the decision tree technique can indicate the order of important explanatory variables clearly. This technique can also indicate the quantitative relationship between landslides and topography/lithology.

Key words : landslide, data mining technique, decision tree technique, EOS-Terra/ASTER, Akaishi Mountains

和文要旨

本研究では、データマイニングの一つであるDecision treeを用いて、赤石山脈における地すべり発生流域の推定を行う。また、Decision treeの特徴を活かして地すべり発生流域の推定過程をtree構造として定量的に示す。そして、tree構造から見える地すべり発生流域と地形・地質との関係を説明することが目的である。地すべり発生流域を推定するための説明変数は、数値標高モデル (DEM) の地形計測によって得られる地形量 (標高, 傾斜, Profile curvature, Plan curvature, 浸食量, 未浸食量), 及び地質データである。本研究ではEOS-Terra/ASTER衛星画像から作成された数値標高モデル (DEM) を使用した。また、学習/検証データとして既存の地すべり分布データを用いた。

その結果、71.2%の正解率で地すべり発生流域を推定できた。また、得られたtree構造より、地すべり発生流域を推定する上で重要な説明変数は地質、斜面形状、傾斜であることが定量的に示された。さらに地すべりが頻発する流域の特徴として、傾斜が緩く、直線斜面や谷型斜面が多いことがわかった。一方で、瀬戸川層における地すべりの分布は様ではなく、地すべりは斜面傾斜が相対的に小さい特定の場所に分布していることが示された。

このように、地すべり発生流域の推定にDecision treeを用いることにより、複数ある説明変数に明確な順位付けをして、地すべり地と地形・地質との関係を定量的に示すことができた。

キーワード: 地すべり, データマイニング, Decision tree, EOS-Terra/ASTER, 赤石山脈

1. はじめに

発生する地すべりの90%前後が既存の地すべりの再滑動または地すべり地の拡大によって起こると言われているように、地すべりは同じ斜面で繰り返し発生することが多い現象である (清水ほか, 2002)。このようなことから、地すべり地、地すべり地形に関する調査や研究が今日まで多く行われてきた。個々の地すべりの発生要因やメカニズムを研究対象とする場合には、現地調査や空中写真判読が有効である。しかし、発展途上国のような国土地理情報が整備されていない地域を対象とした地すべり地調査の場合、その全域で空中写真判読や現地調査を行うことは、時間的にも経済的にも困難である。

そこで、リモートセンシング・GISを用いて、広範囲

の地すべり地を予察的に抽出する研究が行われている。人工衛星データからは広域の情報を定期的に得ることができ、地すべり地の把握・記載に役立てることができる。例えば、瀬戸島ほか (1991) では、新潟県の新第三紀層地すべり地域を対象に、SPOTデータを用いた地すべり地の抽出が行われた。この研究では、土地被覆分類、傾斜区分、斜面縦横断形区分ごとに得点を設定し、3つのレイヤをオーバーレイすることで得点を加算し、地すべり地の抽出を行っている。しかし、このような従来の地すべり地抽出の研究は、はじめに設定した仮説に基づいたレイヤ演算手法であるために、仮説に依存した結論しか得ることができない。また、多種多様で大規模なデータが存在するときに、その一つ一つに仮説を立てて解析を行うことは非常に困難であり、その中に含まれる重要なシグナルを見落としてしまう危険性がある。

こうした問題を解決するのに、データマイニングとい

* 連絡著者 / corresponding author

a) 首都大学東京大学院都市環境科学研究科地理環境科学専攻
Department of Geography, Tokyo Metropolitan University
〒192-0397 東京都八王子市南大沢1-1
1-1, Minami-Osawa, Hachioji, Tokyo, 192-0397, Japan

う解析手法がある。データマイニングは、データの細かな一部分でしか成り立たないようなパターンにも注目することができ、「大規模なデータから有用なパターンを自動的に発見する手段」と定義される (Witten and Frank, 2005)。つまり、データマイニングを用いることで、仮説を必要とせず重要なパターンを発見することができる。この特徴を用いて、村岡ほか (2006) ではデータマイニングの一つであるDecision treeを用いてトルコ共和国南部を対象地域に地すべり危険地の推定を行っている。Decision treeは推定過程がtree構造として明示され、推定過程を解釈することが可能という特徴を持つ手法である。しかし村岡ほか (2006) では、地すべり危険地の推定が行われたのみで、得られたtree構造に基づく地すべり発生条件の解明は行われていない。

また、数値標高モデル (以下、DEMと表記) による地形計測を用いた、従来の地すべり地抽出の研究は、抽出が任意の大きさの方形区画単位あるいはPixel単位で行われてきた。しかし、実際の地形形成プロセスは方形区画や任意の点に対して働いているわけではなく、プロセスを受けた結果である現在の地形も必ずしも方形区画単位で把握できるものではない。このようなことから中山 (1998) は、日本のような湿潤気候下では地形形成営力の中で河川が占める割合は大きく、流域界を超えてその営力は他の流域に影響しないとし、地形量の計測を流域単位で行っている。また土砂災害対策を考えても、土砂移動を捉えるのは流域単位が一般的である (武居, 1983)。

現在日本国内において、日本全域をカバーし、かつ高空間解像度で一般的に入手可能なDEMは、主に「国土地理院数値地図50mメッシュ (標高)」(建設省国土地理院監修・財団法人日本地図センター編集、発行、1994) と「GISMAP Terrain (標高)」(北海道地図株式会社, 2005) がある。しかし、どちらも国土地理院発行の2万5千分の1地形図の等高線を基に作成したDEMであるため、2万5千分の1地形図以上の精度を持つことはできない。一方、より高精度・高解像度のDEMを得る手法として航空レーザ測量が挙げられる (八木ほか, 2003)。しかし、航空レーザ測量による地形計測は非常に高価であり、同時に広範囲のDEMを得ることは困難である。これに対し、近年ではステレオ視が可能な人工衛星画像からデジタル写真測量を用いることで高解像度のDEMを作成することが可能になってきている (野上ほか, 2001)。人工衛星画像を用いることは、グローバルなDEMの整備を可能とし、DEMが整備されていない国や地域においてもDEMを用いた広域の地形計測を行うことができる。なかでも、EOS-Terra/ASTER衛星画像からは空間解像度15mと比較的高い解像度のDEM (以下、ASTER-DEMと表記) を安価に得ることができる (財団法人資源・環境観測解析センター, 2005b)。また、空間解像度15mのDEMを用いて地形計測を行うことで、

山地における土砂移動を論じることが可能となる (田中・大森, 2005)。

以上のことより本研究では、データマイニングの一つであるDecision treeという手法を用いて、広域の地形量と地質データから地すべりが発生する流域の推定を行う。地形計測及び流域計算には空間解像度が15mのASTER-DEMを用いた。また、Decision treeの特徴を活かして、得られたtree構造を解釈し、地すべりが発生する流域と地形・地質との関係を解明することを本研究の目的とする。これらの目的を踏まえて対象地域は、多くの地すべり地が存在し、かつ多様な地形・地質条件であり、さらに地すべり地に関する多くの先行研究が行われている赤石山脈とした。

2. 対象地域概要

本研究の対象地域は赤石山脈南部山岳地帯であり、EOS-Terra/ASTERより得られる1シーンの範囲について解析を行う (図-1)。使用したASTER-DEMの陰影図を図-2に示す。対象地域となる赤石山脈南部山岳地帯は2,000~3,000m級の急峻な山々からなり、地すべり土塊の残存する多数の地すべり地が分布する (大森・須貝, 1994)。

対象地域は、主に糸魚川-静岡構造線と中央構造線に挟まれる地域である。西側では中央構造線を境に内帯と外帯に分かれ、外帯には領家帯が分布する。内帯には、西から順に三波川帯・秩父帯 (石炭紀~ジュラ紀)、四万十帯 (白亜紀~古第三紀)、新第三系が配列する (日本の地質『中部地方I』編集委員会編, 1988)。四万十帯は、北西から赤石層群、白根層群、寸又川層群、犬居層群からなる北帯と、三倉層群、瀬戸川層群に分かれ、年代はこの順に新しい (図-3)。また、特に古第三系の三倉層群において地すべりが頻発することが報告されている (Sugai *et al.*, 1994)。

図-1 対象地域 (太線内) と、地すべり地形分布図の使用範囲 (細線内)

Fig. 1 Study area (thick line) and the area where Landslide Maps Database is used (thin line)

図-2 使用したASTER-DEMの陰影図
Fig. 2 Relief map of ASTER-DEM

図-3 対象地域（黒枠内）及びその周辺の地質図（地質調査所 1995）

Fig. 3 Geological map around the study area (black rectangle), derived from Geological Survey Institute (1995)

3. 使用データと解析手法

3.1 数値標高モデル (DEM) を用いた流域分割

本研究で使用したのは、ASTERレベル 3 A01データ (DEMデータ添付オルソ画像, 財団法人資源・環境観測解析センター, 2005a) である。ASTER (Advanced Spaceborne Thermal Emission and Reflection radiometer) センサは、可視から熱赤外にわたる波長範囲を14バンドでカバーし、放射計部分は可視近赤外 (VNIR)、短波長赤外 (SWIR)、熱赤外 (TIR) の3つのサブシステムより構成される。可視近赤外 (VNIR) は、立体視できるバンドを有しており、空間解像度約15mのDEMを作成可能である。レベル 3 A01データは、ASTERデータより作成されたDEM付きの正射投影画像 (オルソ画像) である。本研究では、このレベル 3 A01データに含まれるDEM (ASTER-DEM) を用いた。また、ASTER-DEMの最大地表画素位置誤差と最大標高誤差は、それぞれ50mと15m以下である (財団法人資源・環境観測解析センター, 2005b)。ここで、対象地域における樹木の高さが20m程度であることを考えると、ASTER-DEMの標高値には少なからず植生の影響が現れていると言える。しかし、最大標高誤差と樹木の高さがほとんど変わらないため、ASTER-DEMから樹木の影響のみを取り出すことは非常に難しい。このため、本研究では3 A01データに含まれるASTER-DEMをそのまま用いて以下の解析を行った。なお、画像の取得日時は2004年11月10日であり、画像中の雲量は1%である。

地すべり地の推定は流域単位で行うため、まず対象地域における流域の分割を行う。特に、DEMを用いることで任意の流域面積を持つ流域を複数分割/抽出することが可能となる (中山, 1998)。本研究ではASTER-DEMを用いて、洪水追跡法 (野上, 1991, 1998など) により計算したDDM (流水線図) に基づいて流域の分割を行った。一方で、吉山 (1994) によると、流域は 2^n メッシュから構成される必要がある。そこで、対象地域において流域を密に分割するために、流域面積の閾値を 2^{11} メッシュ (約0.46km²) から 2^{15} メッシュ (約7.36km²) に設定し^{*1)}、約0.46km²~約7.36km²の面積をもつ流域を抽出した。本研究で使用した流域分割プログラムでは、設定した閾値の上限と下限の間でより流域面積が大きくなるように流域は分割される。結果、対象地域は1,343個の流域に分割された (図-4)。本研究では、得られた流域ごとに地質データと地すべり土塊データ、及び地

*1) 対象地域に存在する最大の地すべり土塊の面積が約 2^{15} メッシュであり、この地すべり土塊が複数の流域にまたがらないように、流域面積の閾値の上限を 2^{15} メッシュ (約7.36km²) と設定した。また閾値の下限については、 2^4 ~ 2^{14} メッシュまですべてを検討した。これは、流域面積は河川の合流により上流から下流に向かって指数関数的に増大するため、ある特定の面積を持つ流域のみを抽出することができないためである。このため、流域の分割条件となる流域面積の閾値はある程度の幅を持つ必要がある。幅が狭ければ (閾値の上限と下限との差が小さいと) はほぼ一定の面積を持つ流域のみが抽出されるが、流域に含まれない部分も多くなる。一方で、幅を広く (閾値の上限と下限との差を大きく) すると、流域に含まれない部分は少なくなる。しかし、閾値の下限が小さすぎると、流域面積が極度に小さい流域が分割される危険性がある。これらのことを考慮した結果、閾値を 2^{11} メッシュ (約0.46km²) から 2^{15} メッシュ (約7.36km²) とした時の流域分割が妥当であった。よって本研究では、流域面積の閾値を 2^{11} メッシュ (約0.46km²) から 2^{15} メッシュ (約7.36km²) とした。

図-4 解析対象とした1,343流域
Fig. 4 1,343 drainage basins analyzed

形量の統計解析を行う。

3.2 地質データ

地すべりと地質の関係は古くから研究され、特定の地質帯に地すべりが多発することが指摘されている（藤田, 1990）。よって本研究においても、地質データを用いて地すべりが発生する流域の推定を行う。対象地域の地質データとして、100万分の1日本地質図第3版CD-ROM版（地質調査所, 1995）を使用した。

100万分の1日本地質図第3版CD-ROM版にはArc/Info Export format (E00形式) のデータが収納され、地質はそれぞれ属性番号で与えられている。まず、Arc/Info (ArcMap9.0: ESRI社) を用いて、E00形式をshape fileに変換した。次にASTER-DEM (15mメッシュ) に合わせてUTM/WGS84 (Zone54) に投影変換し、15mメッシュのラスターデータに変換した。そして、3.1節で分割した流域ごとに集計を行い、面積で占める割合が一番高い地質をその流域の地質 (Geo) とした (表-1)。

3.3 地すべり地データ

対象地域における地すべり地データとして、独立行政法人防災科学技術研究所より公開されている地すべり地形分布図データベース (shape file版) を使用した (独立行政法人防災科学技術研究所総合防災研究部門, 2006)。このデータベースの特徴と、本研究でこのデータベースを使用した理由、及び使用範囲を以下に示す。

地すべり地形とは、地すべりの変動によって生じた、もしくは生じつつある地形の総称として用いられる (清水ほか, 2002)。地すべり地形分布図は、地すべり地形を空中写真判読等によって見出し、それを地形学的に図示した主題図である。本研究で利用した地すべり地形分

布図データベースの特徴として、(1)ごく初期段階の斜面変動も判読図示している点、(2)精度の高い判読水準を維持している点、(3)統一された基準で作成されている点、(4)滑落崖と地すべり土塊を区別して図示している点などが挙げられる (清水ほか, 2002; ハザードマップ編集委員会, 2005)。

本研究では、人工衛星データから広域的に地すべり地を推定することを目的としているため、広範囲の地すべり地に関する情報を同一基準で記載したデータが必要である。このような理由から、地すべり地形分布図データベースを対象地域の地すべり地データとして使用した。一方で、利用した地すべり地形分布図の縮尺は5万分の1であり、実際の現地調査や地すべり対策工事等で用いられる図面に比べると小縮尺である。しかし、本研究では個々の地すべり地の詳細な特徴を調べるのではなく、流域という単位で地すべり地の推定を行うため、5万分の1の縮尺で十分であると判断した。また地すべり災害において、実際に移動し被害をもたらすのは地すべり土塊である。よって本研究では、地すべり地形分布図に示される地すべり土塊の情報のみを用いて解析を行った。

地すべり地形分布図データベースはshape file形式でダウンロードすることができる (独立行政法人防災科学技術研究所総合防災研究部門, 2006)。本研究では、5万分の1地形図の図幅に相当する「富士山」「歟沢」「身

図-5 対象地域 (黒枠内), 及びその周辺における地すべり土塊の分布図 (地すべり地形分布図データベースより作成, 独立行政法人防災科学技術研究所総合防災研究部門, 2006)

Fig. 5 Distribution of landslide masses around the study area (black rectangle), derived from Landslide Maps Database (National Research Institute for Earth Science and Disaster Prevention, 2006)

表-1 各流域ごとに求めた流域特徴量
Table 1 Drainage basin parameters

		Definition	Variables
Landslide Mass Data		$\frac{LS\ area}{DR\ area} \times 100$	$LS\ ratio$
		$LS\ ratio \geq 5$	$LS - yes$
		$LS\ ratio < 5$	$LS - no$
	Geology mode		Geo
Topographic Characteristics			
Elevation	average	$\frac{1}{N} \sum_i Ele$	Ele_{av}
	standard deviation	$(\frac{1}{N} \sum_i (Ele - Ele_{av})^2)^{\frac{1}{2}}$	Ele_{sd}
	skewness	$\frac{1}{N * Ele_{sd}^3} \sum_i (Ele - Ele_{av})^3$	Ele_{sk}
	kurtosis	$\frac{1}{N * Ele_{sd}^4} \sum_i (Ele - Ele_{av})^4$	Ele_{ku}
Slope	average	$\frac{1}{N} \sum_i Slope$	$Slope_{av}$
	standard deviation	$(\frac{1}{N} \sum_i (Slope - Slope_{av})^2)^{\frac{1}{2}}$	$Slope_{sd}$
	skewness	$\frac{1}{N * Slope_{sd}^3} \sum_i (Slope - Slope_{av})^3$	$Slope_{sk}$
	kurtosis	$\frac{1}{N * Slope_{sd}^4} \sum_i (Slope - Slope_{av})^4$	$Slope_{ku}$
	mode		$Slope_{mo}$
Profile Curvature	concave	≥ 0.005	PrC_+
	flat	$-0.005 \sim 0.005$	PrC_0
	convex	≤ -0.005	PrC_-
Plan Curvature	concave	≥ 0.005	PlC_+
	flat	$-0.005 \sim 0.005$	PlC_0
	convex	≤ -0.005	PlC_-
Dissection	average	$\frac{1}{N} \sum_i dis$	dis_{av}
	standard deviation	$(\frac{1}{N} \sum_i (dis - dis_{av})^2)^{\frac{1}{2}}$	dis_{sd}
	skewness	$\frac{1}{N * dis_{sd}^3} \sum_i (dis - dis_{av})^3$	dis_{sk}
	kurtosis	$\frac{1}{N * dis_{sd}^4} \sum_i (dis - dis_{av})^4$	dis_{ku}
Undissection	average	$\frac{1}{N} \sum_i udis$	$udis_{av}$
	standard deviation	$(\frac{1}{N} \sum_i (udis - udis_{av})^2)^{\frac{1}{2}}$	$udis_{sd}$
	skewness	$\frac{1}{N * udis_{sd}^3} \sum_i (udis - udis_{av})^3$	$udis_{sk}$
	kurtosis	$\frac{1}{N * udis_{sd}^4} \sum_i (udis - udis_{av})^4$	$udis_{ku}$

延」「大河原」「赤石岳」「南部」「清水」「静岡」「井川」「千頭」「家山」「飯田」「時又」「満島」「佐久間」「天竜」の16図郭分を使用した(図-1)。まず、使用した図幅の地すべり地形分布図から地すべり土塊の情報のみを取り出した(図-5)。次にASTER-DEM(15mメッシュ)に合わせて、UTM/WGS84 (Zone54) に投影変換し、15 mメッシュのラスターデータに変換した。そして流域図(図-4)と重ね合わせ、流域ごとに地すべり土塊の占める面積率を算出した。

次に、1,343流域を「地すべり発生(地すべりが発生している流域, *LS=yes*)」と「地すべりなし(地すべりが発生していない流域, *LS=no*)」に分割する。ここで、地すべり土塊が面積で1%以上含まれる流域を地すべり発生の流域とすることを考えた。しかし、閾値1%ではほぼすべての流域が地すべり発生の流域となってしまう、推定モデルを構築する上で適当ではない。そこで閾値を2%, 3%と増加させ適切な閾値を検討した。

その結果、閾値を5%とした時に、地すべり発生の流域数と地すべりなしの流域数の比が617:726と、ほぼ1:1となった(図-6)。また、閾値を5%とすることで、617個の地すべり発生の流域に、全流域内の地すべり土塊面積の96%が含まれた。よって、本研究では地すべり発生の流域を定義する閾値は5%とした。

図-6 地すべり発生の流域(617流域)と地すべりなしの流域(726流域)の分布

Fig. 6 Distribution of drainage basins with landslides (*LS=yes*, $n=617$) and drainage basins without landslides (*LS=no*, $n=726$)

4. 流域地形量の算出

土砂移動現象である地すべりは、長い時間にわたる地形形成過程の一コマである。よって、地形計測により、地すべり地の形態的特徴とその成因にアプローチすることは重要である。本研究では、ASTER-DEMから最も基本的な地形量である標高と傾斜、斜面の曲率を表すProfile CurvatureとPlan Curvature (Burrough and McDonnell, 1998)、加えて浸食高と未浸食高(中山, 1998)の6個の地形量を算出した。算出した地形量は流域ごとに統計解析し、23個の地形特徴量を求めた(表-1)。以下にその詳細を述べる。

■標高(Elevation) 標高は、流域ごとに平均(Ele_{av})、標準偏差(Ele_{sd})、歪度(Ele_{sk})、尖度(Ele_{ku})を計算した。標高の平均(Ele_{av})は、流域の高度分布を表す基本的な統計量である(野上・杉浦, 1986)。また標高の標準偏差(Ele_{sd})は、野上(1995)によれば平均と共に流域の起伏に関する指標である。

歪度(skewness)は、平均値を中心として分布がどのように偏っているかを示す尺度である(石村, 1989)。右にスノが長い場合に歪度は正となり、左にスノが長い場合には歪度は負となる。左右対称の場合は0である。つまり、流域内の標高が相対的に低い方に集中していれば標高の歪度(Ele_{sk})は負となる。また、標高の歪度(Ele_{sk})は地形発達を程度を表す指標の一つである(野上・杉浦, 1986)。

尖度(kurtosis)は、分布の型の尖り具合を示す量であり、値が大きくなるほど分布が尖っていることを示している(石村, 1989)。標高の場合、特定の標高帯が多い、つまり平坦面が広がる流域ほど尖度(Ele_{ku})は大きくなる。

■傾斜(Slope) 表層物質の下方移動の際の営力に関係する量として重要なのは傾斜である。地すべり地と斜面傾斜の関係を研究した事例は多い(武居, 1983)。斜面傾斜が増大することは地すべり土塊を斜面下方に滑動させる力を増大させることであり、斜面傾斜は地すべりを形成する重要な地形条件である(藤田, 1990)。

本研究では、ERDAS IMAGINE 8.7 (Leica Geosystems社)を用いて傾斜の計算を行った。ここで傾斜は3×3メッシュの周囲8方向から算出される8近傍法である(Leica Geosystems, 2003)。傾斜は流域ごとに平均($Slope_{av}$)、標準偏差($Slope_{sd}$)、歪度($Slope_{sk}$)、尖度($Slope_{ku}$)、最頻値($Slope_{mo}$)を計算した。

平均($Slope_{av}$)は、その流域の斜面傾斜そのものである。Kawabata *et al.* (2001)は、地すべりが多発する地域では地すべりの作用によって斜面傾斜が減少としている。また平均($Slope_{av}$)や標準偏差($Slope_{sd}$)は標高と同様に、流域の地形の険しさや起伏に関する重要な指標である。

歪度($Slope_{sk}$)は平坦面と接するような山地流域において大きくなり、尖度($Slope_{ku}$)は同一傾斜の斜面が多

い流域で大きくなる。モード ($Slope_{mo}$) は、その流域における斜面傾斜の代表値である。モード ($Slope_{mo}$) が大きな流域は急峻で険しい流域である。

■Profile Curvature Profile Curvature (PrC) の計算は, Burrough and McDonnell (1998) に基づいて行った^{*2)}。PrCは、斜面の曲率を表す量である。PrC>0 なら凹斜面であり, PrC=0 なら直線斜面, PrC<0 なら凸斜面である。求めたPrCの度数分布では、-0.005と0.005に度数が集中した。よって、-0.005と0.005に閾値を設けて、PrC $\geq$ 0.005の場合を凹斜面, 0.005>PrC>-0.005の場合を直線斜面, PrC $\leq$ -0.005の場合を凸斜面と定義した。さらに流域ごとにそれぞれの斜面形の面積比を求め、凹斜面率 (PrC₊)、直線斜面率 (PrC₀)、凸斜面率 (PrC₋) を算出した。

■Plan Curvature Plan Curvature (PIC) の計算は, Burrough and McDonnell (1998) に基づいて行った^{*2)}。PICは、斜面の水平断面の曲率を表す量である。PIC>0 なら谷型斜面であり, PIC=0 なら直線斜面, PIC<0 なら尾根型斜面である。PrCと同様に、PICの度数分布にも-0.005と0.005に度数が集中した。よって、-0.005, 0.005と閾値を設けて、PIC $\geq$ 0.005の場合を谷型斜面, 0.005>PIC>-0.005の場合を直線斜面, PIC $\leq$ -0.005の場合を尾根型斜面と定義した。さらに流域ごとにそれぞれの斜面形の面積比を求め、谷型斜面率 (PIC₊)、直線斜面率 (PIC₀)、尾根型斜面率 (PIC₋) を算出した。

■浸食高 (Dissection) 接峰面高度と現地標高との差は、現在までの浸食の深さの近似値であると考えられる (中山, 1998)。本研究では、接峰面と現地標高との差を浸食高と定義した。接峰面の計算は中山・田中 (2002) に基づいて、9メッシュ (135m) $\times$ 9メッシュ (135m) の方眼法を反復計算することにより算出した。

浸食高は、深い谷が発達する起伏が大きな地域で大きくなり、平坦面が発達するような地域では小さくなる。求めた浸食高は流域ごとに平均 (dis_{av})、標準偏差 (dis_{sd})、歪度 (dis_{sk})、尖度 (dis_{ku}) を算出した。

■未浸食高 (Undissection) 本研究では、接谷面高度と現地標高との差を未浸食高と定義した (中山, 1998)。接谷面の計算は中山・田中 (2002) に基づいて、9メッシュ (135m) $\times$ 9メッシュ (135m) の方眼法を反復計

算することにより算出した。

未浸食高は、谷が深く急峻な地形のところほど大きくなる。逆に広い谷や平坦な地形の地域ほど未浸食高は小さくなり、その分散も小さくなる。未浸食高は流域ごとに平均 ($udis_{av}$)、標準偏差 ($udis_{sd}$)、歪度 ($udis_{sk}$)、尖度 ($udis_{ku}$) を算出した。

5. Decision treeを用いた地すべり発生流域推定モデルの構築

3章、及び4章で得られた流域毎の地形量、地質データ、及び地すべりデータ (表-1, 図-6) から、データマイニングの一つであるDecision treeを用いて、地すべり発生流域を推定するモデルの構築を行う。

本研究でデータマイニングを用いる理由は、「大規模なデータから有用なパターンを自動的に発見する」(Witten and Frank, 2005) というデータマイニングの特徴にある。この利点を活かして佐藤ほか (2005) は、データマイニングの一つであるニューラルネットワークを用いて斜面崩壊地の予測を行っている。しかしニューラルネットワークはブラックボックス的であり (森下・宮野, 2001)、斜面崩壊地の推定結果に対する理由付けや推定過程を明らかにすることは困難である。

Drainage basin characteristics database

ID	Elevation	Slope	Lithology	...	Landslides
001	300m	15°	A	..	○
002	600m	20°	A	..	○
003	500m	40°	B	..	×
004	400m	30°	A	..	○
..	..	..	..	..	..

図-7 「流域における地すべりの有無」に関するDecision treeの例

Fig. 7 Example of Decision tree technique as for "whether drainage basins contain landslides or not"

* 2) Profile Curvature (PrC) と Plan Curvature (PIC) の計算式 (Burrough and McDonnell, 1998) を以下に示す。

$$\begin{array}{|c|c|c|} \hline Z_1 & Z_2 & Z_3 \\ \hline Z_1 & Z_2 & Z_3 \\ \hline Z_1 & Z_2 & Z_3 \\ \hline \end{array} \quad d = 15m$$

$$A = \frac{(Z_1 + Z_2)/2 - Z_3}{d^2} \quad B = \frac{(Z_2 + Z_3)/2 - Z_1}{d^2}$$

$$C = \frac{-Z_1 + Z_2 + Z_3 - Z_3}{4d^2}$$

$$D = \frac{-Z_1 + Z_2}{2d} \quad E = \frac{Z_2 - Z_3}{2d}$$

$$PrC = \frac{2(AD^2 + BE^2 + CDE)}{D^2 + E^2}$$

$$PIC = \frac{-2(AE^2 + BD^2 - CDE)}{D^2 + E^2}$$

そこで本研究では、推定過程がtree構造として明示されるDecision tree (Witten and Frank, 2005) を用いた。Decision treeはデータベース中の注目する属性（目的変数）に関する重要な知識（説明変数）をtree構造によるルールの組み合わせで表現し、注目する属性を分類・推定するモデルである。つまり、注目する属性（目的変数）の推定過程が理解可能である。図-7にDecision treeの簡単な例（「流域における地すべりの有無」に関するDecision tree）を示す。図-7の上の表は、流域ごとの複数の説明変数（標高、傾斜、地質など）と、流域における地すべりの有無が記載されたデータベースである。このデータベースを基にして流域に地すべり地があるかどうかを推定する時に、Decision treeを適用することで、図-7の下に示したtree構造を得ることができる。tree構造には、流域を分類するための説明変数（標高、傾斜、地質など）と、流域における地すべりの有無の関係が示されている。まず、「地質がAであるかどうか」で2つのグループに分類される。次に地質がAの流域について、「傾斜が25度以下」の場合は「地すべり有り」の末端部に分類される。また「傾斜が25度より大きい場合」はさらに次の条件に分類される。図-7において、このtree構造を使って、データベース中のすべての流域はいずれかの末端部に分類される。また、treeは階層構造をなし、目的変数（「流域における地すべりの有無」）に対するより重要な説明変数ほど上部の階層に位置する。

本研究において、説明変数は流域ごとの地形量、地質データ（表-1）であり、目的変数ならびに学習／検証データが流域ごとの地すべり地の有無（図-6）である。福田ほか（2001）によれば、Decision treeを用いて目的変数の値を推定するモデルを構築する場合、推定精度が高くなるようにtree構造の大きさを調整する必要がある。つまり、tree構造を大きくしすぎるとデータベース中のノイズや外れ値にも適合してしまい、逆に小さすぎるとデータベースに含まれる有用な情報を表現できなくなってしまう。前者の状態は過学習（Overfitting）であり、必要以上の細かなtree構造のため、結果的に推定精度は低下する。さらに、tree構造が複雑になるため、tree構造の解釈が困難になってしまう。逆に、後者の場合は、目的変数に関する重要な知識がtree構造の中に含まれなくなり、推定精度が低下する。

このため、Cross validationという手法を用いて最適なtree構造の大きさを決定する（福田ほか, 2001）。Cross validationとは、教師付きのデータを分割して、一方をtree構築のための学習データ、もう一方を構築されたtree構造の推定精度を測る検証データとして、互いにtreeの構築と検証を行うことである。Cross validationは、モデルの構築データと検証データが独立しているため、適切な精度評価が行える。本研究では、最も一般的である10-fold cross validation (Witten and Frank, 2005) を用いて予測精度の評価を行った。ここで、本研究の対

象地域には1,343個の流域が存在している。10-fold cross validationでは、まず1,343個の流域を130個程度の大きさのほぼ等しい10個の集合に分割する。10個の集合への分割はランダムに行われる。そして1つの集合を除いた9個の集合でモデルを構築し、その後、除外した1個の集合でそのモデルの検証を行う。これを10回繰り返し、どの集合も1回検証データになるように、互いにtree構造の構築と検証を行うものである。そして、10-fold cross validationによる推定精度が一番高くなるtree構造が、最適な大きさのモデルとなる。

本研究では推定精度として一般的な「正解率」を採用し、10-fold cross validationによる正解率が最高になるモデルを採用した。正解率は次のように定義される。

$$Correct = \frac{YesYes + NoNo}{YesYes + NoNo + Over + Not} \times 100 \quad \dots\dots\dots(1)$$

Correct: 正解率 (%)

YesYes: 地すべり発生と正しく推定された流域数

NoNo: 地すべりなしと正しく推定された流域数

Over: 地すべり発生と誤って推定された流域数（過推定流域数）

Not: 地すべりなしと誤って推定された流域数（未推定流域数）

なお、Decision treeによるモデル構築のために、オープンソースのデータマイニングツールであるWEKA (Ver. 3.5.1) を用いた (The University of Waikato, 2006)。WEKAはニュージーランドのWaikato大学が中心となり開発されたソフトウェアである (Witten and Frank, 2005)。またtree構造の構築には、WEKAに実装されたJ48アルゴリズムを用いた。J48アルゴリズムはC4.5アルゴリズムをWEKAに実装したものである (Witten and Frank, 2005)。C4.5アルゴリズムでは、目的変数に関する相互情報量（ゲイン率）が最大になるように、目的変数の分割（分割のための説明変数の決定、閾値の決定）が行われる。ここで相互情報量とはエントロピー関数値の減少量である（福田ほか, 2001; Witten and Frank, 2005）。

また、tree構造の大きさを調整する手法として「枝刈り (Pruning)」が挙げられる（福田ほか, 2001; Witten and Frank, 2005）。とくに、事後枝刈り (post-pruning) は、tree構造を構築した後に過学習である部分木を取り除く手法である（福田ほか, 2001）。つまり、tree構造の主要な構造を変化させずに、かつ推定精度を大きく低下させることなく、簡単なtree構造を得ることが可能である。本研究ではtree構造の解釈を容易にするために、事後枝刈りを適用した（7.2節）。

6. 推定結果と精度

10-fold cross validationによる正解率が71.2%（表-2）で最高になった時のtree構造を、Decision treeによ

表-2 Decision treeによる地すべり発生流域推定結果の10-fold cross validationによるConfusion matrixと正解率

Table 2 Confusion matrix and correctness from 10-fold cross validation of the estimation by Decision tree technique

		Estimation	
		LS - yes	LS - no
Observation	LS - yes	398 (Yes-Yes)	219 (Not)
	LS - no	168 (Over)	558 (No-No)

LS - yes: 地すべり発生流域数

LS - no: 地すべりなし流域数

Yes-Yes: 地すべり発生と正しく推定された流域数

No-No: 地すべりなしと正しく推定された流域数

Over: 地すべり発生と誤って推定された流域数 (過推定の流域数)

Not: 地すべりなしと誤って推定された流域数 (未推定の流域数)

$$Correct = \frac{YesYes + NoNo}{YesYes + NoNo + Over + Not} \times 100$$

$$= 71.2 \quad (\%)$$

る地すべり発生流域推定モデル (図-8) として採用した。図-8は21個の要素・条件から構成され、末端部の数は22個である。例えば、モデルの上部について見ると、まず領家帯1 (Geo:133) の流域が地すべりなしと推定される。次に、領家帯1 (Geo:133) 以外の流域について、標高の平均 (Ele_{av}) が250m以下である場合には地すべりなしと推定される。そして次では、三倉層 (Geo:64) の流域が地すべり発生 (yes) と推定されている。ここでyesに分類された流域数は77個であり、うち9流域が誤り (実際は地すべりなし) である。このように、22個の末端部のうち、9個の末端部で地すべりが発生する流域 (yes) と推定され、13個の末端部で地すべりなしの流域 (no) と推定されている。

得られたモデルを用いて推定した地すべり発生流域の分布を図-9に示す。推定結果 (図-9a) では、主に中央構造線の東側や、大井川左岸の地域に地すべり発生

流域の分布が集中しているのがわかる。一方で、地すべり地形分布図データベースにより作成した地すべり発生流域の分布 (図-9b) も、主に中央構造線の東側や大井川左岸の地域に集中している。次に、このモデルによる地すべり発生流域の推定精度を図-10に示す。1,343流域のうち、398流域が地すべりが発生する流域として、また558流域が地すべりのない流域として正しく推定されている。一方で、過推定 (Over estimated) が168流域、未推定 (Not estimated) が219流域存在している。

7. 考察

7.1 推定精度の検証

Decision treeによる地すべり発生流域推定モデル (図-8) の10-fold cross validationによる正解率は、71.2%である (表-2)。本研究と同様にDecision treeを用いて、トルコ共和国北西部を対象地域に地すべり箇所を推定した研究 (村岡ほか, 2006) では、推定精度として正解率60~80%が得られている。その他にも、データマイニング手法を用いた研究として、佐藤ほか (2005) はニューラルネットワークと最尤法を用いて崩壊地点の予測を行い、約70%の正解率が得られている。これらの結果と比較しても、本研究で得られた71.2%という正解率は遜色ないものである。

また、本研究における正解率は(1)式によって求められ、全流域数に占める正しく推定された流域数の割合が正解率となる。ここで、地すべり地形分布図より作成した地すべり発生流域の分布 (図-6) を見ると、地すべり発生流域数と地すべりなしの流域数の比は617:726 (0.46:0.54) となっている。この場合、仮にすべての流域を地すべり発生流域と推定すると、正解率は(1)式より46%となる。逆に、すべての流域を地すべりなしと推定すると正解率は54%となる。このように地すべり発生流域か否かを一律に推定した場合、正解率は50%程度にしかない。しかし、本研究における正解率は71.2%である (表-2)。このことから、本研究の推定結果が十分な精度であることが示される。

一方で、過推定 (Over estimated) と未推定 (Not estimated) の流域もそれぞれ12.5%と16.3%ずつ存在している (表-2)。過推定 (Over estimated) は、地すべりなしの流域が地すべり発生流域として誤って推定された流域である。つまりこれらの流域は、地形・地質的に見れば地すべりが発生する流域と同様な特徴があり、将来的に地すべりが発生しやすい流域であるといえる。また防災の観点から見れば、過推定 (Over estimated) の流域は地すべりの発生を過大に評価していることとなるため、安全方向の過誤である。しかし、未推定 (Not estimated) は地すべり発生流域として推定できなかった流域であり、防災の観点から見ると重大な過誤である。未推定 (Not estimated) 流域は、主に四万十帯 (北帯) に対応する地域に多く分布していることがわかる (図-

図-10 Decision treeによる地すべり発生流域の推定結果の精度

Yes-Yes: 地すべり発生と正しく推定できた流域,
No-No: 地すべりなしと正しく推定できた流域,
Over estimated: 地すべり発生と誤って推定された流域(過推定), Not estimated: 地すべりなしと誤って推定された流域(未推定)

Fig. 10 Accuracy of the estimation by Decision tree technique

Yes-Yes: Correctly estimated as drainage basins with landslides, No-No: Correctly estimated as drainage basins without landslides, Over estimated: Incorrectly estimated as drainage basins with landslides, Not estimated: Incorrectly estimated as drainage basins without landslides

3, 10)。この未推定 (Not estimated) の流域を減らしていくことは、今後の課題の一つである。

7.2 Tree構造の解釈

本研究でDecision treeを用いた理由は、推定過程がtree構造として明示される点にある。よって、Decision treeによる地すべり発生流域推定モデル(図-8)のtree構造の解釈を行い、地すべり発生流域の推定過程を明らかにする。

得られたモデル(図-8)は末端部の数が22と多く、これではtree構造の解釈が複雑になりわかりにくい。一方で、tree構造の階層構造を考慮すると、下部階層の構造ほど重要度が小さい。そこでtree構造の解釈を容易にするために、tree構造の下部階層について「事後枝刈り(Pruning)」(福田ほか, 2001; Witten and Frank, 2005)を行った。その結果、末端部数が13個の簡単なtree構造が得られた(図-11)。このtree構造に注目すると、

枝刈りされたのは図-8における下部階層部分であり、tree構造の上部階層は変化していないことがわかる。よって、基本的なtree構造は枝刈り前後で変化していない。また、枝刈りされたtree構造における10-fold cross validationの推定精度は68.4%と、枝刈り前の推定精度71.2%から約3%の低下しか見られない。よって、枝刈りされたtree構造の解釈でも十分の精度を保っていると判断し、図-11を用いて地すべり発生流域の推定過程の解釈を行った。また図-11について、それぞれの末端部に分類された流域を色分けした(図-12)。

まず、tree構造の1番目と4番目の条件では、領家帯1 (Geo:133)と領家帯2 (Geo:134)の流域が地すべりなしの流域と推定されている(図-11, [1], [4])。本研究の対象地域においては、中央構造線以西の領家帯に地すべり土塊の分布は少ない(図-5)。

2番目の条件では、標高の平均 (Ele_{av}) が250m以下の流域が地すべりなしの流域と推定されている(図-11, [2])。標高の平均 (Ele_{av}) が250m以下の地域は、対象地域内の南東方向に分布する安倍川下流域の静岡市街地である(図-12)。この地域も地すべり土塊の分布は少ないため(図-5)、地すべりなしの流域と推定されている。

続く3番目の条件において、三倉層 (Geo:64)の流域が地すべり発生の流域と推定されている(図-11, [3])。三倉層は古第三系四万十帯に属し(日本の地質『中部地方I』編集委員会編, 1988)、対象地域中央南部(大井川左岸)に分布している(図-3)。第三紀層では地すべりが頻発することが知られ(藤田, 1990)、対象地域においても三倉層では地すべりが多発することが指摘されている(Sugai et al., 1994)。図-5より、三倉層には地すべり土塊が多く分布しているのがわかる。

5番目の条件(図-11, [5])では、尾根型斜面率 (PIC_{-}) が25%という閾値で大きく左側の枝と右側の枝に分かれている。左側の枝は尾根型斜面率 (PIC_{-}) が25%以下、つまり相対的に尾根型斜面の少ない流域である。このグループについてそれぞれの末端部の分類数に注目すると、地すべり発生の流域 (yes 2) が337流域であり、地すべりなしの流域 (no 4) の27流域と地すべりなしの流域 (no 5) の38流域に比べて圧倒的に多いことがわかる。つまり尾根型斜面率 (PIC_{-}) が25%以下の流域はほぼ地すべりが発生している流域であるといえる。地すべりが発生する流域では、地すべり地特有の地形から尾根型斜面が相対的に少ないと考えられる。

尾根型斜面率 (PIC_{-}) が25%以下のグループを詳しく見ると、傾斜の標準偏差 ($Slope_{sd}$) が11.9を超える38個の流域は地すべりなしの流域と推定されている(図-11, no 5)。図-12より、地すべりなしの流域 (no 5) の38流域は大井川、安倍川、富士川などの大河川の河床を含む流域であることがわかる。このような流域では平坦な河床と斜面が混在するために、尾根型斜面は少なく、

図-11 枝刈り後のDecision treeによる地すべり発生流域推定モデル (変数は表-1 に示す)

Fig.11 Pruned estimation model of drainage basins with landslides by Decision tree technique (Variables are listed in Table 1)

傾斜の標準偏差は大きくなると考えられる。この条件では、河川沿いの地すべりなしの流域が推定されているといえる。

さらに傾斜の歪度 ($Slope_{sk}$) により、地すべりなしの流域 (no 4) と地すべり発生の流域 (yes 2) が分類されている (図-11)。ここでは、傾斜の歪度が -0.9 を超える流域が地すべり発生の流域である。地すべりが多発する流域ではその地形的な特徴から平坦地が存在することが考えられる。よって傾斜の分布も小さい方に偏るため、傾斜の歪度も大きくなる。このような条件により地すべり発生の流域 (yes2, 337流域) が推定されている。また、図-12において地すべり発生の流域 (yes 2) の分布は、地すべり土塊の分布 (図-5) と良く対応している。

一方で、尾根型斜面率 ($PIC-$) が 25% を超えるグループは、標高の歪度 (Ele_{sk}) が -0.4 以下という条件で地すべり発生の流域 (yes 3) が推定されている (図-11, [6-b])。図-12より、地すべり発生の流域 (yes 3) は標高の高い地域の流域や、大きい河川沿いの流域面積が小さい流域であることがわかる。一般に標高の歪度は、同一高度の平坦面と谷が接しているような流域で小さくなる。よって、地すべり発生 (yes 3) に分類される流域では、河川の浸食により平坦面と谷の境界付近で地すべりが起きていると考えられる。

7番目の条件 ([7-b]) では、瀬戸川層 (Geo:66) の条件で2つのグループに分かれている。瀬戸川層は古第三系四十万帯である (図-3, 日本の地質『中部地方I』編集委員会編1988)。瀬戸川層は安倍川沿いに分布し、三倉層と共に地すべりが多数分布している (図-5)。

地質が瀬戸川層である流域は、さらに左側の枝で標高の Ele_{sd} (標準偏差) と $Slope_{av}$ (傾斜の平均) により地すべりなしの流域 (no6, no 7) と地すべり発生流域 (yes 4) とに分類されている。まず標高の標準偏差 (Ele_{sd}) が $121m$ 以下の流域が地すべりなしの流域 (no 6) である。地すべりなしの流域 (no 6) は図-5, 12より、地すべりなし (no 2) の流域と同様に安倍川下流部の地すべり土塊の存在が少ない地域の流域であることがわかる。

次に傾斜の平均 ($Slope_{av}$) が 32.3 度以下の、傾斜が緩い流域が、地すべりが発生している流域 (yes 4) と推定されている。このことはKawabata *et al.* (2001) で報告されている「地すべりが多発しやすい地域では相対的に斜面傾斜が小さい」という知見と整合する。また、緩斜面が存在するという地すべり地形の特徴 (武居, 1983) を考えても、地すべりが多発する流域では傾斜が小さくなることが考えられる。このように瀬戸川層の地域では地すべり地の分布は一樣ではなく、地すべり地は斜面傾斜が相対的に小さい特定の場所に分布していることがわかる。

瀬戸川層 (Geo:66) 以外の流域は、凹型斜面率 ($PrC+$) が 21.4% という条件で地すべり発生の流域 (yes 5) と地すべりなしの流域 (no 8) に分類されている (8-[b])。ここで分類数に注目すると、圧倒的に地すべりなしの流域 (no 8) に分類される流域 (477流域) が多い。よって、瀬戸川層 (Geo:66) 以外の流域はほぼ地すべりなしの流域であることがわかる。しかし、その中でも凹型斜面率 ($PrC+$) が 21.4% 以下という条件で、地すべり発生の流域 (yes 5) がわずかに推定されている。

図-12 Decision treeによる地すべり発生流域推定モデル (図-11) において、各末端部に分類された流域の分布

Fig. 12 Distribution of drainage basins classified into each leaf in pruned estimation model of drainage basins with landslides by Decision tree technique (Fig. 11)

以上のように、Decision treeによる地すべり発生流域推定モデル (図-11) のtree構造が明らかになった。次に、tree構造で用いられる説明変数とその階層構造に注目する。Decision treeでは、tree構造の上位にくる条件ほど重要である。まず、このモデルにおいて重要な説明変数として地質 (Geo) が挙げられる。地質と地すべりの関係は現在まで多く議論されており (藤田, 1990), 得られたtree構造においても地質区分と地すべり土塊の分布が強く関係していることが示された。次に重要な説明変数として、斜面形状 (PIC) が挙げられる。地すべりが発生する流域では、地すべり地特有の地形から尾根型斜面が相対的に少なくなることが示された。

Tree構造では傾斜 (Slope) も重要な役割を果たしている。地すべりを議論するうえで傾斜が重要であることも古くから指摘され (武居, 1983), 本研究におけるtree構造においてもそのことが示された。そして、地すべりが多発する流域において傾斜の平均が小さくなるということは, Kawabata *et al.* (2001) の見解と整合的であった。

このように、tree構造で用いられている説明変数は、従来から地すべりと関係があると指摘されているものや、地すべり地特有の地形を示すものである。しかしながら、

複数ある重要な説明変数に明確な順位付けをして、定量的に示したことはDecision treeの特徴であるといえる。またtree構造からは、同一の地質の地域であっても、地すべり地の分布は地形的な要因を受けていることが定量的な条件で示された。特に瀬戸川層の地域では、地すべり地は斜面傾斜が相対的に小さい特定の場所に分布していることが示された。

8. 結論

本研究では、赤石山脈南部山岳地帯を対象地域に、ASTER-DEMから得られる広域の地形量と、地質データから、地すべりが発生する流域の推定を試みた。推定には仮説に依存しない手法であるデータマイニングを用いた。具体的には、データマイニングの一つであるDecision treeを用いて、流域ごとに地すべり地の有無を推定するモデルを構築した。また、推定過程がtree構造として明示されるDecision treeの特徴を活かし、地すべり発生流域の推定過程の解釈を行った。その結果、以下の知見が得られた。

- ・ 本研究では、Decision treeを用いて、仮説に依存せずに地すべりが発生する流域を推定するモデルを構築した。それにもかかわらず、得られたtree構造は、従来から指摘される地すべり地と地形・地質との関係と整合的であった。tree構造より、地すべり発生流域を推定するために重要な説明変数は、主に地質、斜面形状、傾斜であることが定量的に示された。とくに、四万十帯に属する三倉層の流域で地すべりが頻発していた。また、地すべりが頻発する流域の斜面傾斜は緩く、斜面形状は直線あるいは谷型であることが示された。一方で、瀬戸川層においては地すべりの分布は様ではなく、地すべりが地形的な影響を受けていることがtree構造から明らかになった。このようにDecision treeを用いることで、複数ある重要な説明変数に明確な順位付けをして、定量的に示すことができた。
- ・ 地すべり発生流域の推定では正解率71.2%が得られた。この精度は、データマイニングを用いて地すべり地を抽出する、他の研究と比較しても遜色なく、推定に用いたtree構造が妥当なものであることを示している。
- ・ 今後の展開としては、対象地域を広げ、他の説明変数も加えて地すべり発生流域推定モデルの精度を向上させることが考えられる。また、現地調査も加えて、本研究で得られた結果の詳細な検証が必要である。

謝辞

本研究を進めるにあたり、ジオテクノス株式会社の村岡弘康さんからは多くの資料と助言をいただき、暖かく研究を見守っていただきました。心から感謝いたします。

参考文献

- Burrough, P. A. and McDonnell, R. A. (1998): Principles of Geographical Information Systems, Oxford: Oxford University Press, 333p.
- 地質調査所 (1995): 100万分の1 日本地質図第3版CD-ROM版, 地質調査所.
- 独立行政法人 防災科学技術研究所 総合防災研究部門 (2006): 地すべり地形分布図データベース, <http://www.bosai.go.jp/sougou/jindex.html>, 2006. 10. 10.
- 藤田崇 (1990): 地学ワンポイント 3 地すべり - 山地災害の地質学, 共立出版株式会社, 126p.
- 福田剛志・森本康彦・徳山剛 (2001): データサイエンス・シリーズ 3 データマイニング, 共立出版株式会社, 169p.
- ハザードマップ編集小委員会 (2005): ハザードマップ - その作成と利用 -, 社団法人日本測量協会, 234p.
- 北海道地図株式会社 (2005): GISMAP Terrain Ver. 2.00標準仕様書, 北海道地図株式会社, 9p.
- 石村貞夫 (1989): 統計解析のはなし, 東京図書株式会社, 340p.
- Kawabata, D., Oguchi, T. and Katsube, K. (2001): Effects of geology on slope angles in the Southern Japanese Alps - A GIS approach -, Transactions, Japanese Geomorphological Union, Vol. 22, pp. 827-836.
- 建設省国土地理院監修・財団法人日本地図センター編集, 発行 (1994): 数値地図ユーザーズガイド (改訂版), 財団法人日本地図センター, 471p.
- Leica Geosystems (2003): ERDAS field guide, seventh edition, Atlanta: Leica Geosystems, 672p.
- 森下真一・宮野悟 (2001): 発見科学とデータマイニング, 共立出版株式会社, 299p.
- 村岡弘康・Hakan, N.・Taner, S.・荒木強 (2006): ASTERデータを用いた地滑り危険地検知へのデータマイニング手法の適用 (その2), 日本リモートセンシング学会第40回学術講演会論文集, pp. 169-170.
- 中山大地 (1998): DEMを用いた地形計測による山地の流域分類の試み - 阿武隈山地を例として -, 地理学評論, Vol. 71A, pp. 169-186.
- 中山大地・田中靖 (2002): 数値標高モデル (DEM) 処理技術勉強会テキスト (2002年版), 日本地形学連合, 63p.
- 日本の地質『中部地方 I』編集委員会編 (1988): 日本の地質 4 中部地方 I, 共立出版株式会社, 348p.
- 野上道男 (1991): 地理情報とくにラスター型数値地図利用の高度化, 情報地質, Vol. 2, pp. 331-339.
- 野上道男 (1995): 細密DEMの紹介と流域地形計測, 地理学評論, Vol. 68A, pp. 465-474.
- 野上道男 (1998): DEM (数値標高モデル) からDDM (流水線図) を作成するアルゴリズムの改良とC言語プログラム, GIS - 理論と応用, Vol. 6, No. 1, pp. 95-102.
- 野上道男・杉浦芳夫 (1986): パソコンによる数値地理学演習, 古今書院, 275p.
- 野上道男・岡部篤行・貞広幸雄・隈元崇・西川治 (2001): 地理情報学入門, 東京大学出版会, 163p.
- 大森博雄・須貝俊彦 (1994): 地すべり土塊の形態的特徴とその地形学上の意義, 地形, Vol. 15, pp. 1-16.
- 佐藤浩・関口辰夫・神谷泉・本間信一 (2005): 斜面崩壊の危険度評価におけるニューラルネットと最尤法分類の比較, 日本地すべり学会誌, Vol. 42, pp. 293-302.
- 瀬戸島政博・白石一夫・森大 (1991): SPOTデータを用いた地すべり箇所の抽出と地すべり分布図の情報更新の試み, 写真測量とリモートセンシング, Vol. 30, No. 3, pp. 4-15.
- 清水文健・八木浩司・井口隆・大八木規夫 (2002): 地すべり地形分布図第14集「静岡」解説と読図の手引き, 防災科学技術研究所資料, 第221号, pp. 1-11.
- Sugai, T., Ohmori, H. and Hirano, M. (1994): Rock control on magnitude-frequency distribution of landslide, Transactions, Japanese Geomorphological Union, Vol. 15, pp. 233-251.
- 武居有恒監修 (1983): 地すべり・崩壊・土石流. 予測と対策, 鹿島出版会, 334p.
- 田中靖・大森博雄 (2005): 山地の地形計測のためのDEMの解像度に関する一検討, 写真測量とリモートセンシング, Vol. 44, No. 1, pp. 26-35.
- The University of Waikato (2006): WEKA, <http://www.cs.waikato.ac.nz/ml/weka/index.html>, 2006. 10. 10.
- Witten, I. H. and Frank, E. (2005): Data Mining - Practical Machine Learning Tools and Techniques, Second Edition, Oxford: Elsevier, 525p.
- 八木浩司・檜垣大助・吉松弘行・相楽渉・高木洋一・内山庄一郎 (2003): 空中レーザー高精度地形図の地すべり・微地形判読への応用, 日本地すべり学会誌, Vol. 39, No. 4, pp. 35-41.
- 吉山昭 (1994): 河道の勾配と流域の地形特性量との関係, 国土数値情報を用いた計測, 地理学評論, Vol. 67A, pp. 794-802.
- 財団法人資源・環境観測解析センター (2005a): ASTERユーザーズガイド第3編 3D Ortho Products (L3A01) Ver. 1.1, 財団法人資源・環境観測解析センター, 18p.
- 財団法人資源・環境観測解析センター (2005b): ASTERユーザーズガイド第3編 DEM Products (L4A01) Ver. 1.1, 財団法人資源・環境観測解析センター, 21p.
- (原稿受付2006年6月26日, 原稿受理2006年12月14日)