

機械学習法を用いた SPOT5/HRG データの 土地被覆分類とその精度比較

望月 翔太¹・村上 拓彦²

(受付 2015 年 7 月 8 日；改訂 2016 年 1 月 17 日；採択 3 月 1 日)

要 旨

近年、植生図作成においてオブジェクトベース画像分類を用いた研究事例が増えてきた。オブジェクトベース画像分類では、オブジェクトの基礎統計量や形状、テクスチャなど多くの特徴量を使用することができる。これにより、多次元の特徴空間を持つデータを分類する事が重要な課題となっている。これまで標準的に使用されてきた最尤法や近傍法では、多次元のデータから明らかになるパターンや関係性を特徴付けられないという指摘がある。この時、機械学習による分類が有効視されている。本論では、機械学習方法である CART, Bagging, Boosting, Random Forest, SVM を用いた画像分類を試みた。本研究の対象地は新潟県佐渡市である。2007 年 6 月に取得された SPOT5/HRG データ(パンシャープン画像)を使用した。分類クラスは広葉樹林, スギ林, アカマツ林, 竹林, 水田, 市街地, 道路, 裸地の 8 クラスとした。また、特徴量として、オブジェクトの基礎統計量や形状, テクスチャを含む 4 つのデータセットを用いた。得られた分類画像の精度を比較した結果, Boosting と Random Forest, SVM において高い精度が得られた。最も分類精度が低かった手法は CART だった。異なるデータセットでの分類結果から、使用する特徴量が多次元の場合は、Random Forest と SVM が有効な手法となる。一方、使用する特徴量が少ない場合、Boosting の精度が Random Forest と SVM に勝る。使用可能な特徴量の多さに応じて、どちらの機械学習法を採用するか決める必要がある。

キーワード：衛星リモートセンシング, 土地被覆, 画像分類, オブジェクトベース, 集団学習。

1. はじめに

地図化は衛星リモートセンシングの主要な活用方法の一つである。画像データから、土地被覆毎の分光反射特性に基づき、様々な画像分類手法を適用し、画像データのカテゴリ化を行う。衛星リモートセンシングデータを利用する目的は多様であり、生態系管理のための詳細スケールの植生図作成もあれば、全球スケールの純一次生産量推定のためのマッピングなどもある(村上・望月, 2014)。

現在、日本の土地被覆や植生分布を示した情報として、環境省の第 6, 7 回自然環境保全基礎調査植生調査の結果である 2 万 5 千分の 1 で作成された現存植生図を利用することができる。しかし、作成に要する作業量の膨大さから、作業完了までに長い年月を要し、自然環境の基礎

¹新潟大学大学院 自然科学研究科：〒950-2181 新潟県新潟市西区五十嵐 2 の町 8050

²新潟大学 農学部：〒950-2181 新潟県新潟市西区五十嵐 2 の町 8050

情報としては期待にそぐわないことが指摘されている(日置, 2007). また, マツ枯れのようなマツの大量枯損を引き起こす樹病や自然災害等による植生の変化をこの現存植生図から即時に捉えることは困難である(中川・村上, 2012). そのため, リモートセンシング分野において, 効率的に植生図や土地被覆図を整備する事は, きわめて重要な課題である(Xie et al., 2008). 特に判読が容易な市街地, 農地, 植林地などを中心として, 衛星画像の利用を前提とした植生図の整備手法を導入する事により, 整備の省力化や効率化, および敏速化が期待されている(鎌形他, 2006).

衛星画像から, 植生図や土地被覆図を作成する研究は数多く行われている(鎌形 他, 2006; 長澤, 2006; 伊藤 他, 2007; Xie et al., 2008; Adam et al., 2010). これまで, 最尤法やマルチレベルスライス法, ISODATA 法, k-近傍法など様々な分類手法が使用されてきた. また, 高分解能衛星を用いる事によって, 大区分の植生図作成の効率化が図れる事が報告されている(鎌形, 2006). さらに, 中分解能衛星(ALOS/AVNIR-2)においても, 比較的精度の高い植生図作成が可能である事が指摘されている(長澤, 2006). これらの研究では, 最尤法や再近接法による分類手法が取り入れられているが, 画像分類では, 目的とする分類クラスにより, 最適な分類手法や要求される空間分解能は変化する. つまり, 効率的に植生図や土地被覆図を整備するためには多くのケーススタディを元に情報を体系化する必要がある.

近年, リモートセンシングデータの画像分類において, 画像を構成する画素同士の空間的関係性を重視したオブジェクトベース画像分類を用いた研究事例が増えてきた(Blaschke, 2010). この分類方法では, 画素値の似た隣接し合うピクセル群をひとつの領域(オブジェクト)として区分し, その領域を最小単位として分類を行う. そのため, オブジェクトベース画像分類では, オブジェクトの基礎統計量(領域内における画素値の平均や標準偏差)や領域自体の形状, 画像のきめを表現するテクスチャなど多くの特徴量を使用する事ができる. これにより, 多次元の特徴空間を持つデータを分類する事が重要な課題となっている. 特に, これまで標準的に使用されてきた最尤法や近傍法では, 多次元のデータから明らかになるパターンや関係性を特徴付けられないという指摘がある(Melgani and Bruzzone, 2004). この時, 機械学習による分類が有効視されている(Gislason et al., 2006). 機械学習とは, ある程度の数のサンプルデータ集合を対象に解析を行い, そのデータから有用な規則や判断基準などを抽出する(Mitchell, 1997). 本論では, 複数の結果を統合・組み合わせることで精度を向上させる機械学習方法である集団学習(Opitz and Maclin, 1999)を用いたオブジェクトベース画像分類を試みた. また, 分類精度の比較対象として特徴量の分岐ルールが単純な樹木モデルを使用した. 複数の集団学習方法から画像分類を行い, その精度比較から最適な分類手法を明らかにする事を目的とした. なお, 本論で目的とする分類画像の空間解像度は1ピクセル 2.5m×2.5mという高分解能衛星画像の活用を想定している. 同時に, それぞれの分類手法において, 使用する特徴量(変量)の個数が与える画像分類への影響も評価した.

2. 研究対象地

本研究の対象地は新潟県佐渡市である. 佐渡地域は, 面積 855.26 km², 海岸線延長 280.6 km と日本海側最大級の離島であり, 北には標高 1,172 m の金北山を最高峰とする大佐渡山地, 南には標高 645 m の大地山を中心とした小佐渡丘陵を擁し, これらに挟まれる形で国仲平野が広がっている. 佐渡地域の 74%の面積が森林に覆われており, 主要樹種はコナラ, ミズナラをはじめとする広葉樹が全体の 70%を占め, 針葉樹はスギ, アカマツ, クロマツから成り, 竹林も分布している. 本研究では, 小佐渡地域の中央部に 9 km 四方のエリアを設け, 調査と解析を行った(図 1). このエリアには経塚山(標高:636 m)や女神山(標高:593 m)など小佐渡地域において比

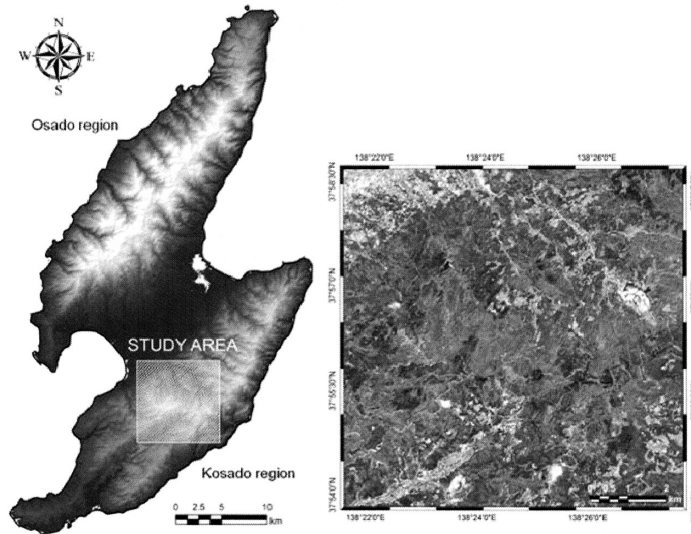


図 1. 研究対象地域。

較的標高の高い山地が含まれる。このエリアでは、コナラ、ミズナラ、ケヤキ等の落葉性広葉樹が大面積を占有している。また、スギやアカマツ、竹林等の針葉樹も分布している。2010年の8月と2011年の11月に36地点の現地調査を行い、周辺の樹種構成や土地被覆に関するサンプル(教師データ)を取得した。

3. 方法

3.1 使用データ

2007年6月に取得されたSPOT5/HRGデータを使用した。SPOT5データは、マルチスペクトル画像(空間分解能10m)と、パンクロマチック画像(空間分解能2.5m)で構成されている。マルチスペクトル画像は、可視緑色域(GREEN)、可視赤色域(RED)、近赤外域(NIR)、短波長赤外域(SWIR)の4バンド構成となっている。本論では、マルチスペクトル画像とパンクロマチック画像を掛け合わせるパンシャープン処理を施し、4バンド構成を有し、空間分解能が2.5mのパンシャープン画像を使用した。また分類精度を検証するため、水土里ネットより提供された簡易オルソフォト(2006年撮影, 空間分解能0.25m)を検証用のデータとして使用した。

3.2 オブジェクトベース画像分類

オブジェクトベース画像分類では、まず領域分割処理によって画素の集合体であるオブジェクトを生成する。本論では、各パラメータをScale Parameter: 20, Shape: 0.3, Compactness: 0.5と設定した。Scale Parameterとは、この値を基準として領域統合の判定を行う閾値であり、各領域内の画素値の標準偏差と領域の形状係数にそれぞれ重み付けを行った上で合計したものである。この値を大きくするほど領域統合の許容範囲が大きくなり、オブジェクトサイズが大きくなる。ShapeとCompactnessは領域統合のパラメータであり、線形から円形までのオブジェクトの形状を決定する。今回設定したScale ParameterとShape, Compactnessの値は、それぞれの土地被覆クラスを区分するために適したオブジェクトサイズになるように設定した。画像分類では、分類に使用する特徴量として、オブジェクトの基礎統計量や形状、テクスチャを含

表 1. 分類に使用した 4 つのデータセット.

データセット I :
オブジェクト内における輝度値の平均値
データセット II :
オブジェクト内における輝度値の平均値と標準偏差、バンド比
データセット III :
オブジェクト内における輝度値の平均値と標準偏差、バンド比
オブジェクト自体の形状
データセット IV :
オブジェクト内における輝度値の平均値と標準偏差、バンド比
オブジェクト自体の形状
画像のきめを表すテクスチャ特徴量

んだ 4 つのデータセットを用意した(表 1). データセット I では, 各オブジェクトの輝度値の平均のみ(特徴量: 4 個)を使用し, データセット II では標準偏差とバンド比を加えた(特徴量: 12 個). データセット III では, オブジェクトの基礎統計量(平均, 標準偏差, バンド比)に加え, オブジェクトの面積や長さ等, オブジェクトの形状を表す指標を採用した(特徴量: 19 個). データセット IV では, オブジェクトの基礎統計量と形状に加え, 画像のキメ細やかさを表すテクスチャ特徴量を採用した(特徴量: 25 個). テクスチャ特徴量には, 近赤外域の単バンド画像に対し, 同時生起行列(Gray Level Co-occurrence Matrix: GLCM, 高木・下田, 1991)を用いた. このように, データセット I から IV にかけて, 使用する特徴量を増加させた. 分類クラスは広葉樹林, スギ林, アカマツ林, 竹林, 水田, 市街地, 道路, 裸地の 8 クラスとした. 精度検証では, 分類結果と検証データから判定効率表を作成し, Kappa 係数(Stehman and Czaplewski, 2003)を求めた. 検証用のデータは, 2006 年に撮影された水土里ネット提供の空中写真と現地取得した樹木の位置情報である. それぞれのクラスに対し, 100 点ずつ検証用サンプルを取得した. 使用ソフトは, Definiens Developer 7.0(Definiens 社, ドイツ)である.

3.3 分類手法: CART

CART は, 決定木学習アルゴリズムの一つであり, 樹木モデルの基礎として広く用いられている. CART 法では, 説明変数を 2 進分岐させ, 2 進木を生成する. その際, 節の分岐の評価基準には尤離度とジニ係数が用いられる. ある基準を元に樹木を発展させ, 1 つの決定木を作成するシンプルなアルゴリズムである(Breiman et al., 1984). 本論では, データセットごとの特徴量の値を分岐条件とし, 各分岐点に対し, 分割する属性を樹木モデル全体のジニ係数の減少を最大にするという規準を採用した. CART による分類は, R2.10.0 の tree 関数を採用した.

3.4 分類手法: Bagging

Bagging は, 与えられたデータセットからブートストラップによって複数の学習データを作成し, そのデータを用いて作成した分類結果を統合・組み合わせる事で精度を向上させる集団学習の一つである(Breiman, 1994). ブートストラップサンプルはそれぞれ独立で, 学習は並列に行うことができる. ブートストラップサンプルは, 与えられたデータの経験分布とその推定量に基づいたリサンプリングにより得られたサンプルである. 複数の結果を組み合わせる事に

より、不安定な学習結果を安定させる事が可能となる。Bagging による分類は、R 2.10.0 のパッケージ ipred を採用した。

3.5 分類手法：Boosting

Boosting は、与えられたデータセットを用いて学習を行い、その学習結果を踏まえて逐次に重みの調整を繰り返すことで複数の学習結果を求め、その結果を統合・組み合わせ、精度を向上させる (Schapire, 1999)。Boosting の特徴として、単に複数の結果を組み合わせるだけでなく、分類精度の弱い学習結果に対し、重みを加える事で不安定な学習結果を安定させる事が挙げられる。Boosting には、重みの初期値の与え方や信頼度の計算と重みの更新方法に違いがある複数のアルゴリズムが提案されている。本論では、AdaBoost というアルゴリズムを使用した。Boosting による分類は、R 2.10.0 のパッケージ ada を採用した。

3.6 分類手法：Random Forest

Random Forest は、2001 年に提案された比較的新しいデータ解析方法である (Breiman, 2001)。基本的には Bagging と概念が似ており、与えられたデータセットからブートストラップサンプリングによって複数の学習データを作成する。両者の違いは、Bagging では全ての変数を用いるが、Random Forest では変数をランダムサンプリングし、そのサブセットを使用する。そのため変数間の相関が発生しやすい高次元のデータ解析に向いており、大量データに対して、効率的に動作する。Random Forest による分類は、R 2.10.0 のパッケージ randomForest を採用した。

3.7 分類手法：SVM

SVM (Vapnik, 1998) は、学習データに対して、2つのクラスを分割するような超平面の中でクラス間のマージンが最大となるような超平面を構成して判別する分類手法である。各クラスの汎化能力が低くならないようにマージンが決定されるため、高精度での分類が可能となる。この手法の利点として、データの特徴の次元が大きくなっても識別精度が良いことや、パラメータの算出が容易である事が挙げられる。SVM による分類は、R 2.10.0 のパッケージ e1071 を採用した。

4. 結果

各データセットと各分類手法における分類精度 (Kappa 係数) を表 2 に示す。まず、最も分類精度が良かった組み合わせは、Random Forest を用いたデータセット II の分類であった (Kappa=0.68)。次いで、データセット II における SVM とデータセット III における Boosting と Random Forest を使用した分類において精度が高かった (Kappa=0.67)。CART を用いた分

表 2. Kappa 係数を用いた分類精度の比較.

	CART	Bagging	Boosting	Random Forest	SVM
データセット I	0.51	0.60	0.63	0.60	0.60
データセット II	0.54	0.60	0.63	0.68	0.67
データセット III	0.56	0.64	0.67	0.67	0.67
データセット IV	0.56	0.61	0.66	0.66	0.67

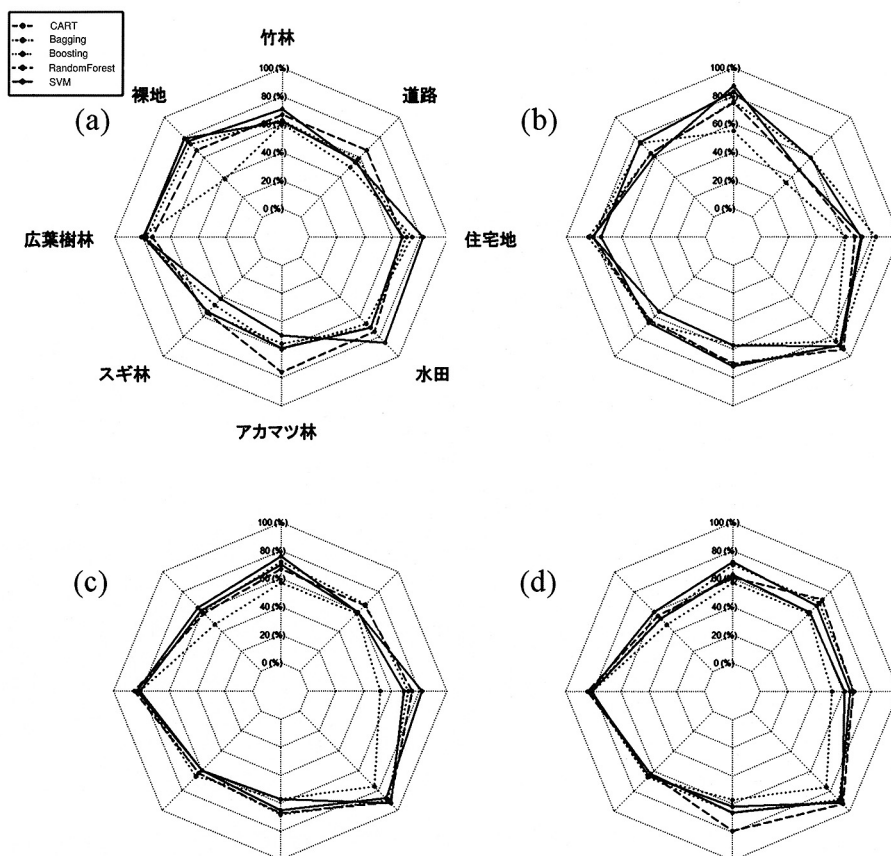


図 2. データセット毎の各分類クラスの信頼度 (User's Accuracy). (a) データセット I, (b) データセット II, (c) データセット III, (d) データセット IV.

類はどのデータセットでも Kappa 統計量が 0.6 以下であった。CART 法は、他の機械学習法に比べ、分類精度が低かった。Bagging は本論の手法間では中程度の精度を示した。

使用した特徴量 (データセット) 間での分類精度の違いは、データセット I とその他のデータセットとの間で顕著だった。分類に使用した特徴量の数が少ないデータセット I では Boosting の精度が良く、オブジェクト内の輝度値の標準偏差とバンド比を加えたデータセット II では Random Forest と SVM の精度が高かった。またオブジェクト形状とテクスチャ特徴量を加え、多次元の特徴量を使用したデータセット III と IV では、Boosting と Random Forest が同じ精度だった。多次元の特徴量を使用した場合、Boosting と Random Forest, SVM において若干の精度の低下が確認された。分類に使用する特徴量の増加は分類精度の向上には繋がらなかった。

次に、分類クラス毎の信頼度 (User's Accuracy) を図 2 に示す。個々の土地被覆の分類でも、Boosting と Random Forest, SVM の 3 つの手法による結果で精度が高かった。どのデータセットにおいても、広葉樹林と水田は精度良く分類できた。一方、アカマツ林や竹林、住宅域、道路は分類手法間で精度に違いが生じた。特に CART を用いた分類は、精度が悪かった。Boosting と Random Forest, SVM を使用した場合は全体的に精度が向上したが、スギ林やアカマツ林、

表 3. Random Forest を用いたデータセット II の判定効率表.

分類クラス	参照クラス								計	Use's Accuracy
	竹林	裸地	広葉樹林	スギ林	水田	アカマツ林	道路	住宅域		
竹林	67	1	4	3	0	12	0	1	88	0.76
裸地	12	64	3	0	2	0	11	8	100	0.64
広葉樹林	6	0	78	2	0	7	0	0	93	0.84
スギ林	0	0	0	79	0	38	1	6	124	0.64
水田	0	6	0	2	93	0	0	4	105	0.89
アカマツ林	5	0	2	10	0	43	0	1	61	0.70
道路	9	21	13	1	4	0	87	12	147	0.59
住宅域	1	8	0	3	1	0	1	68	82	0.83
計	100	100	100	100	100	100	100	100		
Producer's Accuracy	0.67	0.64	0.78	0.79	0.93	0.43	0.87	0.68		

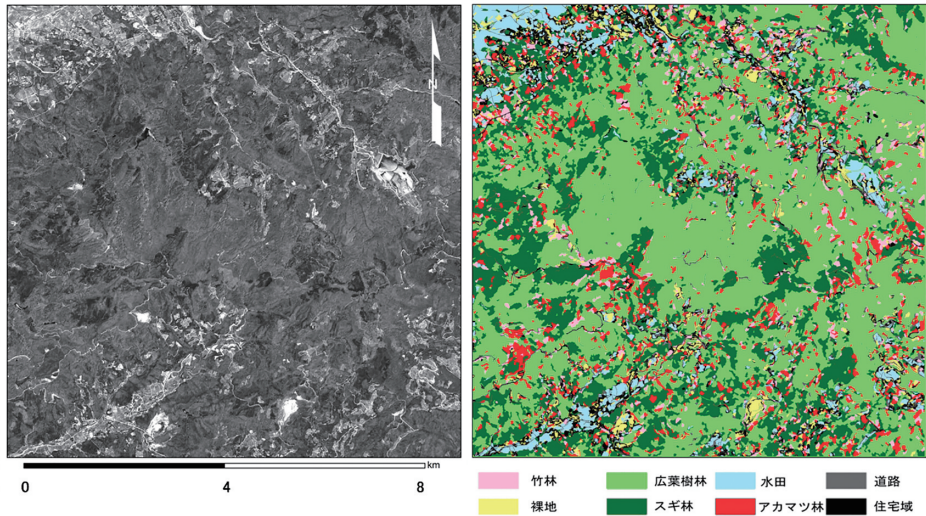


図 3. 各データセットと分類手法を用いた分類画像.

裸地で誤分類が生じた。最も精度の良かった、データセット II における Random Forest を用いた分類の判定効率表を表 3 に、土地被覆分類図を図 3 に示す。判定効率表からも、スギ林とアカマツ林の誤分類が多く確認された。また図 3 から、画像上で確認される樹種間における色調の違いを分類できた事が確認された。

5. 考察

本論は、SPOT5/HRG パンシャープン画像に対するオブジェクトベース画像分類のケースス

タディである。CARTとBagging, Boosting, Random Forest, SVMから得られた分類画像の精度を比較した結果、BoostingとRandom Forest, SVMにおいて高い精度が得られた。最も分類精度が低かった手法はCARTだった。これは、画像分類において機械学習が有効であることを示す。特に、Random Forestでは、アルゴリズムの似ているBaggingより高精度の分類が可能だった。Random ForestとBaggingの違いは、使用する特徴量の決定機構にある。Random Forestでは特徴量をランダムサンプリングするため、変量間に存在する相関の軽減に繋がったと考えられる(Breiman, 2001)。特に、これはデータセット間における精度の違いからも明らかである。使用する特徴量の少ないデータセットIでは、BaggingとRandom Forestの精度は同程度であるが、特徴量が多くなると両者の精度に差が生じた。このことから、多次元の特徴量を画像分類に使用する場合、Random Forestでのランダムサンプリングによる特徴量選択が有効である事が示唆される。

データセット別の分類精度から、データセットIIとIII, IVで精度の高い画像分類が実行できた事がわかる。データセットIIは、オブジェクトの基礎統計量であり、オブジェクト内の平均輝度値やその標準偏差から構成される。また、データセットIIIはオブジェクトの基礎統計量とオブジェクト形状から構成される。つまり、オブジェクトベース画像分類を行う時は、最低限オブジェクトの基礎統計量を使用する必要があると考えられる。オブジェクト内の平均輝度値のみのデータセットIは、4つのデータセット間で最も分類精度が低かった。また、テキストチャ特徴量を加えたデータセットIVでは、精度の向上は確認されなかった。これは、分類に使用する特徴量が増加したため、変数間の相関などにより適切に分類できなかった事が考えられる。特に、テキストチャ特徴量の異質性やコントラスト、標準偏差の間で高い相関が確認された。変数をランダムサンプリングするRandom Forestにおいても、使用する変数が多く、さらに相関し合う変数が多い場合、この影響は完全には除去できない事が推察される。さらに、テキストチャ特徴量の場合、空間分解能 $2.5\text{m} \times 2.5\text{m}$ のSPOTデータでは、空間分解能が粗く、画像のキメを表現するテキストチャを有効に使用できないことも考えられる。テキストチャ特徴量を使用する場合は、より高分解能データを用いた画像分類の実験が必要である(小出, 2008)。例えば、小阪 他(2007)では、空間分解能 1m の衛星画像を用いた林分タイプ分類において、テキストチャ特徴量が分類精度の向上に寄与したことを報告している。そのため、本論で用いた機械学習法を、高分解能衛星に適用することができれば、テキストチャ特徴量は分類精度の向上に貢献するかもしれない。

各分類クラスの信頼度を評価した結果、アカマツのクラスはスギや広葉樹のクラスと多く誤分類を生じていた。これは、スギの若齢林の示すスペクトル特徴量がアカマツと類似していたこと、アカマツと広葉樹の混交状態にある林分では分類が困難であったことが原因であった。スギの若齢林との誤分類箇所に関しては、SPOTデータよりも高分解能な衛星データを用いた上でのテキストチャ特徴量など、本研究では使用していない特徴量の検討が必要である(河邑 他, 2005)。広葉樹との混交状態にあるアカマツを抽出するには、適切なオブジェクトを作成することができなかったため、オブジェクトベース画像分類では困難であった。今後は、アカマツ広葉樹混交クラスの設定や分解能のより高い衛星画像での分類、もしくはピクセルベース画像分類との併用を検討する必要がある。さらに落葉期のデータ等も含めた多時期合成画像(小出, 2008)での分類も検討しなければならない。

総合的に判断すると、どのデータセットにおいてもBoostingとRandom Forest, SVMが高い分類精度を示している。このことから、機械学習を用いた画像分類では、これらの手法を用いる必要があるだろう。特に、使用する特徴量が多次元の場合は、Random Forestが有効な手法となる。SVMでは、使用する特徴量が多くなると、精度が低下したため、Random Forestの方が安定した結果を導き出すと考える。一方、使用する特徴量が少ない場合(例えば、データセッ

ト I), Boosting の精度が Random Forest に勝る。これは、両者のアルゴリズムの違いから説明できる。Boosting は、学習結果を踏まえて逐次に重みの調整を繰り返す手法である。そのため、使用する特徴量が少ない場合も、重みの調整により、ある程度 of 分類精度を得る事が可能となる (Schapire, 1999)。Random Forest では、分類精度の弱い学習に対し、Boosting で行われるような重み調整は実行されない。この事が、データセット I における分類精度の違いに繋がったと考えられる。ある程度の特徴量を使用する場合は、変数間の相関が問題になるため、Random Forest が有効になる。つまり、使用可能な特徴量の多さに応じて、どちらの機械学習法を採用するか決める必要がある。また、本論では領域ごとに独立に分類を実行したが、隣り合った領域に存在する空間自己相関の問題を考慮できていない。今後は、Markov Random Field (MRF: Tso and Mather, 1999; Poggi et al., 2005) などを用いて、全体を通じた統一的な問題として画像分類の処理を試みる必要がある。

謝 辞

本研究を行うにあたり、新潟大学農学部の相澤賢太郎氏、小板橋正充氏、中川恭兵氏、長谷川恭大氏には、佐渡の現地調査における地上基準点の取得にご助力いただきました。ここに、厚く謝意を申し上げます。

参 考 文 献

- Adam, E., Mutanga, O. and Rugege, D. (2010). Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review, *Wetlands Ecology and Management*, **18**, 281–296.
- Blaschke, T. (2010). Object based image analysis for remote sensing, *ISPRS Journal of Photogrammetry and Remote Sensing*, **65**, 2–16.
- Breiman, L. (1994). Bagging predictors, Technical Report, No.421, Department of Statistics, University of California, Berkeley.
- Breiman, L. (2001). Random forests, *Machine Learning*, **40**, 5–32.
- Breiman, L., Freedman, J. and Olshen, R. (1984). *Classification and Regression Trees*, Chapman & Hall, New York.
- Gislason, P. O., Benediktsson, J. A. and Sveinsson, J. R. (2006). Random forests for land cover classification, *Pattern Recognition Letters*, **27**, 294–300.
- 日置佳之 (2007). 日本全国を網羅する現存植生図の応用面から見た課題, 景観生態学, **11**, 107–112.
- 伊藤史彦, 喜多晃平, 長澤良太, 日野彰彦, 浅井樹, 中島尚子 (2007). 相観植生図作成における ALOS AVNIR-2 画像の利用可能性, 景観生態学, **11**, 125–132.
- 鎌形哲稔, 原慶太郎, 森大, 赤松幸生, 李雲慶, 星野義延 (2006). 高分解能衛星データのオブジェクト指向分類による植生図作成手法の提案, 写真測量とリモートセンシング, **45**, 43–49.
- 河邑眞, 辻野和彦, 辻子裕二 (2005). 森林の樹種に関する高分解能衛星画像特性の分析, 写真測量とリモートセンシング, **44**, 82–90.
- 小出馨 (2008). 植生フェノロジーの変化に着目した多時期衛星データによる林相区分の精度向上と森林の下層植生状態の推定, 写真測量とリモートセンシング, **47**, 4–12.
- 小阪尚子, 秋山侃, 蔡斌, 児島利治 (2007). 高分解能衛星画像のテクスチャ特徴量とスペクトル特徴量を用いたオブジェクト指向型林分タイプ分類, 写真測量とリモートセンシング, **46**, 27–36.
- Melgani, F. and Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines, *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 1778–1790.
- Mitchell, T. (1997). *Machine Learning*, McGraw-Hill, New York.

- 村上拓彦, 望月翔太 (2014). リモートセンシングによる植生マッピング, *日本生態学会誌*, **64**, 233–242.
- 長澤良太 (2006). デジタル植生図の技術的課題, *景観生態学*, **11**, 15–25.
- 中川恭兵, 村上拓彦 (2012). オブジェクトベース画像分類による森林タイプ分類における分類手法ならびに特徴量選択に関する検討, *新潟大学農学部研究報告*, **65**, 57–65.
- Opitz, D. and Maclin, R. (1999). Popular ensemble methods: An empirical study, *Journal of Artificial Intelligence Research*, **11**, 169–198.
- Poggi, G., Scarpa, G. and Zerubia, J. B. (2005). Supervised segmentation of remote sensing images based on a tree-structured MRF model, *Geoscience and Remote Sensing, IEEE Transactions*, **43**, 1901–1911.
- Schapire, R. E. (1999). A brief introduction to boosting, *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, 1401–1406.
- Stehman, S. V. and Czaplewski, R. L. (2003). Introduction to special issue on map accuracy, *Environmental and Ecological Statistics*, **10**, 301–308.
- 高木幹雄, 下田陽久 (1991). 変換, 『画像解析ハンドブック』, 475–637, 東京大学出版会, 東京.
- Tso, B. C. and Mather, P. M. (1999). Classification of multisource remote sensing imagery using a genetic algorithm and Markov random fields, *Geoscience and Remote Sensing, IEEE Transactions*, **37**, 1255–1260.
- Vapnik, V. (1998). *Statistical Learning Theory*, Wiley, New York.
- Xie, Y., Sha, Z. Y. and Yu, M. (2008). Remote sensing imagery in vegetation mapping: A review, *Journal of Plant Ecology*, **1**, 9–23.

Accuracy Comparison of Machine-learning-based Land-cover Classification Using SPOT5/HRG Data

Shota Mochizuki¹ and Takuhiko Murakami²

¹Graduate School of Science and Technology, Niigata University

²Faculty of Agriculture, Niigata University

Land cover mapping provides basic information for advanced science such as ecological management, biodiversity conservation, forest planning and so on. In remote sensing research, the process of creating an accurate land cover map is an important subject. Recently, there has been growing research interest in object-oriented image classification techniques. Object-oriented image classification consists of multi-dimensional features including object features, and thus requires multi-dimensional image classification approaches. For example, a linear model such as the maximum likelihood method of pixel-based classification cannot characterize the patterns or relations of multi-dimensional data. In multi-dimensional image classification, data mining and ensemble learning have been shown to increase accuracy and flexibility. This study examined the use of the object-oriented image classification by multiple machine learning algorithms for land-cover mapping. We applied four classifiers: Classification and regression tree (CART), Decision tree with Boosting, Decision tree with Bagging, Random Forest, and Support Vector Machine (SVM). The study area was Sado Island in Niigata Prefecture, Japan. Pan-sharpened SPOT/HRG imagery (June 2007) was used and classified into the following eight classes: broad-leaved deciduous forest, Japanese cedar, Japanese red pine, bamboo forest, paddy field, urban area, road, and bare land. We prepared four data sets with object and texture features. The number of features increases from data sets I through IV. As a result, CART was unsuitable for multi-dimensional classification. Random Forest, Decision tree with Boosting and SVM showed high classification accuracies. Furthermore, in the data set with the limited features, Decision tree with Boosting was an accurate classifier. Random Forest and SVM are effective for multi-dimensional image classification such as data sets II and III. Decision tree with Boosting is effective for image classification with limited features such as data set I.