

多時期の衛星データとサポートベクターマシンを用いた森林／非森林の分類精度の比較

田中真哉・高橋興明・家原敏郎（森林総合研究所）・齋藤英樹（森林総研九州支所）・栗屋善雄（岐阜大学）

要旨：多時期（複数の季節）のランドサット衛星 ETM+センサの観測データとサポートベクターマシンを利用して、北海道の東部地域を対象に森林／非森林の分類における分類精度と（1）チャンネルの組み合わせ、（2）ランドサットデータの取得時期と組み合わせ、の関係について検討した。データセットからトレーニングデータをランダムに選定して森林／非森林ラベルの予測を繰り返し、ランドサットデータとチャンネルの組み合わせごとに分類精度を算出して比較した。その結果、チャンネルの組み合わせとして23457を用いた場合に最も高い分類精度が得られ、本研究で利用した春、夏、秋のデータの中では夏のデータで最も高い分類精度が得られた。また、取得時期の異なる多時期のデータを組み合わせで分類に供することにより、さらに高い分類精度を得られることが分かった。近年、衛星データが比較的安価に利用できるようになっており、この現状を考慮すると多時期の衛星データを利用することは高い分類精度を達成するための有効な選択肢のひとつであると思われた。

キーワード：リモートセンシング、サポートベクターマシン、分類、季節変化、森林／非森林

I はじめに

広域の土地被覆の地図化を目的として、これまでにリモートセンシングデータの利用が検討されてきている。しかし、リモートセンシングデータの解析では地表面のスペクトルに基づいて分類が行われるため、特定の時期のデータのみに基づいて分類をした場合、地表面のスペクトルが類似して分離困難なカテゴリが生じることが多い。これらの誤分類を抑制するため、スペクトルの季節変化、すなわち多時期の衛星データの利用が検討されてきた（例えば、1,5）。ところが、著者らが北海道東部において、教師なし分類（ISODATA法）によって土地被覆分類を行った事例では、多時期の衛星データを利用して牧草地等の非森林植生と落葉広葉樹の分離が困難な事例が見られた（4）。このような森林と非森林の誤分類は、土地被覆図を炭素蓄積の推定や陸域生態系モデルへの入力として利用することを想定した場合に森林カテゴリの中での誤分類（例えば常緑針葉樹と落葉針葉樹など）に比べてより深刻な誤差を生じさせるおそれがある。

そこで、本研究では多時期の衛星データと近年注目されている機械学習法であるサポートベクターマシン（以下、SVMと記す）を利用した森林／非森林の分類について検討を行うことを目的とした。検討項目は、分類精度と（1）チャンネルの組み合わせ、（2）多時期の衛星データの取得時期と組み合わせ、とした。

II データと方法

1. 研究対象地 研究対象地は北海道東部、Worldwide

Reference System-2 のパス 106/ロウ 30 の範囲である（図-1）。対象地は常緑針葉樹、落葉針葉樹、落葉広葉樹、針広混交林、農地、牧草地、湿原、原野、市街地など多くの土地被覆カテゴリが混在している。比較的地形のなだらかな平野部にも森林が分布しており、人工林は比較的パッチサイズが大きい。冬季は積雪し、衛星データの判読からみると平野部でも3月後半まで雪で覆われることもある。また、高緯度に位置するため、標高の高い山地においては4月後半まで積雪が残るエリアが多い。対象域内に大雪山系を含み、衛星データの取得時期によっては植生のフェノロジーがシーン内で大きく異なる。対象域内東部、根釧地域には牧草地が広く分布し、格子状に防風林が作られている。

2. 衛星データ アメリカ地質調査所のウェブサイトからダウンロード可能なランドサット7号 ETM+（Enhanced Thematic Mapper Plus）のデータを解析に利用した。プロダクトの処理はレベル 1T で、Shuttle Radar Topography Mission によって作成された数値標高モデルによるオルソ幾何補正処理済みのデータである。打ち上げから衛星のスキャンラインコレクターが故障するまで、1999年から2003年までの期間のすべてのデータの中から、比較的雲の少ない良好なデータとして、着葉期（1999年8月11日、以後ID1）、黄葉期（1999年9月28日、ID2）、および落葉期（2002年4月29日、ID3）のデータを選定し、熱赤外（チャンネル6）を除いた計6チャンネルのデータを解析に利用した。陸域画素を対象とす

Shinya TANAKA, Tomoaki TAKAHASHI, Toshiro IEHARA (Forestry and Forest Products Research Institute, 1 Matsunosato, Tsukuba, Ibaraki 305-8687), Hideki SAITO (Kyushu Research Center, FFPRI), and Yoshio AWAYA (Gifu University), Comparisons of accuracy in forest/non-forest classification using by multi-temporal satellite data and support vector machine.

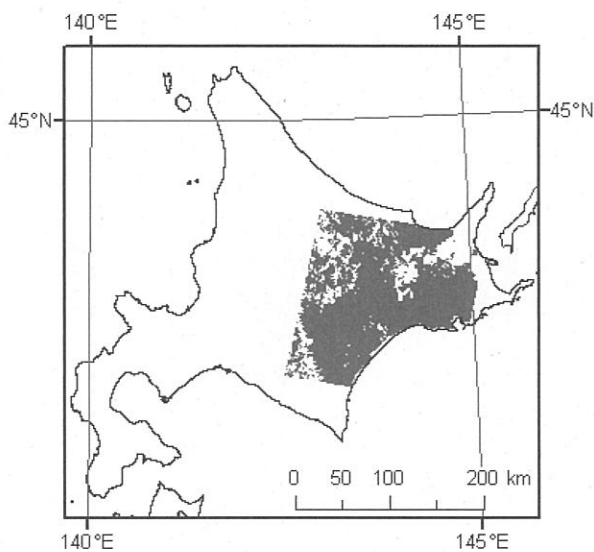


図1. 研究対象地の位置

灰色で塗りつぶされた領域が3時期分の雲・霧・霧およびそれらの影を除外して切り出した解析対象範囲を示す(図は Coverage A の範囲)。

るため、海域は海岸線の GIS データを利用して除外した。それぞれのデータ取得時期において雲、霧や霧、およびそれらの影を目視判読によって特定し、3 時期ともにこれらに覆われていない領域を解析対象エリアとして切り出した (Coverage A)。また、ID 3 ではシーン内の高標高域に積雪が残ることから、比較のためにこれらのエリアを除外して切り出した (Coverage B)。オルソ幾何補正済みの空中写真との比較から、ランドサットデータの位置誤差は 1 画素程度と判断されたため、解析者による精密幾何補正は行っていない。また、本研究では大気補正および地形効果補正 (地形による陰影の補正) は行わず、未補正のデジタル値を解析に用いた。

3. グラウンドトゥールズデータ 分類に利用するトレーニングデータおよびテストデータとして、森林/非森林のラベルを白黒空中写真とランドサットデータを目視判読により以下の手順によって作成した。(i) 対象領域にほぼ等間隔に点群を発生させた後、それぞれに半径 50m の円形バッファ (ランドサットデータの 3×3 画素を想定) を発生させた。(ii) 各円形領域に対してオルソ補正済みの空中写真と複数の季節のランドサットデータとを相互に判読し、森林/非森林のラベルを作成した。

(iii) データのそれぞれに位置誤差を含むこと、リサンプリングが三次畳み込み内挿によって行われていることを考慮し、半径 50m バッファ内に森林と非森林を含む場合に解析から除外した。また、複数の林分で構成されて

いる場合にその一部を同様に除外した。ランドサットデータの判読は、空中写真とランドサットデータの取得時期が異なることを考慮する目的で行い、ポイントの中心ピクセルに森林伐採や住宅地の造成等の土地被覆変化があったと判断される場合、さらに近傍 8 画素の範囲のおおよそ 20%以上の範囲で上記と同様の土地被覆変化があったと判断される場合にその地点を解析から除外した。

上記の手順によって最終的に Coverage A から 619、Coverage B から 551 点の森林/非森林ラベルが得られた (以後、それぞれをデータセット A および B と記す)。

4. 方法 本研究では SVM を分類器として利用した。SVM は性能の高い分類器として知られている。分類に用いる変数として、以下の組み合わせを検討した。(a) ID 1, ID 2, ID 3 のすべてのチャンネルの組み合わせ、計 18 変数。3 時期のデータのそれぞれからチャンネル 1 (青)を除いた 15 変数。同様にチャンネル 1, 2 (緑)を除いた 12 変数。以下同様に、(b) ID 1 と ID 2 の組み合わせ (12 変数, 10 変数, 8 変数)、(c) ID 1 と ID 3、(d) ID 2 と ID 3、(e) ID 1 のみ (6 変数, 5 変数, 4 変数)、(f) ID 2 のみ、そして (g) ID 3 のみ、である。上記のデータ取得時期の組み合わせ 7 通り、チャンネルの組み合わせ 3 通り、計 21 通りの解析をデータセット A, B についてそれぞれ行った。本研究においてチャンネル 1, 2 の有無を検討した理由は、当該チャンネルにおいて大気の影響が強いいため、また、その標高依存性が強いためである。

分類精度の変化とばらつきを調べるため、本研究では分類器の構築と予測を繰り返して分類精度を算出した。トレーニングデータは各データセットから重複を許さないようにそれぞれ 419 地点 (データセット A) および 351 地点 (データセット B) をランダムに選定し、残りをテストデータとした。この作業を 10000 回繰り返して予測誤差および Kappa 係数の平均、中央、最大、最小値を算出し、チャンネルおよびランドサットデータの組み合わせごとの精度評価値を比較検討した。

衛星による観測であるため、雲による被覆の影響を受け、1 年間の観測では必ずしも雲のない良好なデータを複数の時期で得ることができなかった。そのためランドサットデータの取得時期は複数の年次にまたがるが、今回の解析ではその間の被覆の変化は考慮していない。また、実際分類図作成においてはサブカテゴリ (各森林タイプ、水域、市街地、農地など) を考慮してトレーニングデータとテストデータを選定することが一般的であるが、今回の解析では考慮せずにランダムサンプリングを繰り返し、森林/非森林ラベルのみに基づいて解析を行った。

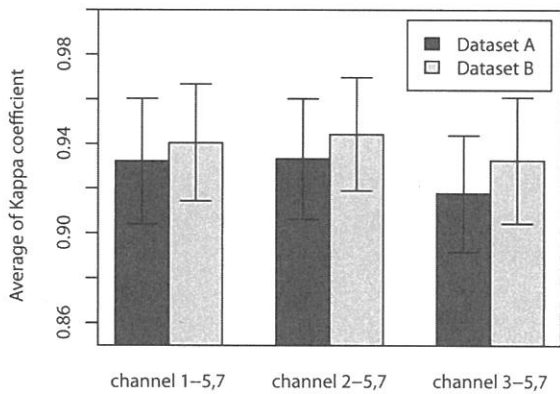


図-2. チャンネルの組み合わせと Kappa 係数の関係
10000 回の繰り返しによって算出した Kappa 係数の平均値を示す。エラーバーは標本標準偏差。チャンネルの組み合わせ以外のすべてのケースの結果を統合した。

III 結果

解析に利用するチャンネルと取得時期の組み合わせごとに Kappa 係数を比較したところ、組み合わせるチャンネルとしては 2 3 4 5 7 の場合に最も Kappa 係数の高い予測がなされた (図-2)。組み合わせるデータの時期は多いほど精度が高い傾向にあったが、3 時期を組み合わせた場合には雪に被覆されているエリアを含むか否かによって結果が異なった (図-3)。

対象範囲の広いデータセット A について、最も精度が高い結果が得られたチャンネル 2 3 4 5 7 を変数として解析した場合の精度評価値の詳細を表-1 に示す。単一時期のデータのみを比べた場合、本研究で用いた 3 時期の中で ID 1 (着葉期) と ID 2 (黄葉期) にはほとんど違いが見られなかった。予測誤差の最大値と Kappa 係数の最小値からは ID 1 の場合に最もよい分類結果が得られた。ID 3 (落葉期) の分類精度は ID 1 と ID 2 に比べて劣っており、7 通りの組み合わせの中で最も精度が悪かった。7 通りの組み合わせの中で精度の平均値と中央値から判断すると、最も精度が高いケースは ID 1 と ID 2 を組み合わせた場合で、Kappa 係数と予測誤差の平均値はそれぞれ、0.959, 0.018 であった。200 のテストデータ中の平均的な誤分類の数は 3 から 4 程度であり、予測されたラベルは現実の土地被覆とよく一致していた。一方、3 時期のデータを組み合わせた場合、平均値や中央値では ID 1 と ID 2 を用いた 2 時期の組み合わせに比べてわずかに低いものの、10000 回の繰り返しにおける最も精度が悪いケースにおいても 7 通りの組み合わせの中で最も精度が高かった。

今回の分類では 10000 回の分類を行ったため、全体的

な誤分類の傾向を調べることは難しい。そのため、作成した分類図のいくつかを目視によって比較した。その結果、森林と非森林の境界部分については分類結果が不安定で、また、若齢林が非森林として誤分類される地点もしばしば確認された。一方で農地の中心で森林と誤分類されるようなケースはまれであり、著者らの経験から ISODATA 法と比較して概ね良好な分類図が得られたと判断された。

IV 考察

チャンネルの組み合わせとして 2 3 4 5 7 で良い結果が得られた理由として、チャンネル 1 の大気の影響と標高依存性 (2) が影響していると思われた。チャンネル 2 についても同様に、より長い波長のチャンネルに比べて大気の及ぼす影響は大きいですが、本研究の場合、目視判読によって霞がかかったエリアについては事前に解析から除外しているためにノイズの影響が顕著でなく、情報量が多いことによる利得が大きかったものと推測される。

欧州の Carpathians 山地周辺を対象にランドサットデータを SVM によって森林/非森林に分類した事例において、Knorn らはトレーニングデータがクラスごとに 500、つまり計 1000 サンプル程度に達するとそれ以上精度の向上がみられないと報告しており (3)、本研究ではトレーニングデータがやや不足しており、さらにデータ数を増やすことによってより精度の高い分類結果が得られる可能性がある。ただし、誤分類数は 200 のテストデータ中から平均して 3、4 程度であり、すでに十分な精度を有していると思われた。また、実際の土地被覆図作成において安定して高い精度を達成するためには、本研究で考

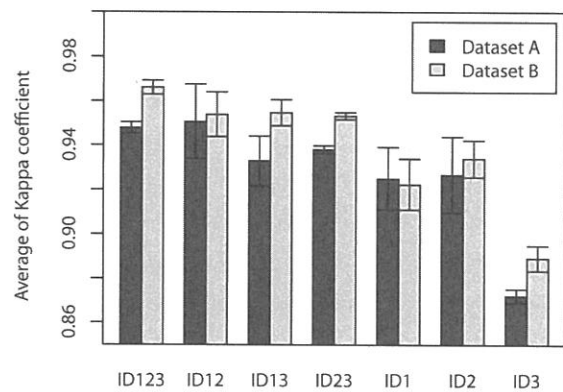


図-3. ランドサットデータの取得時期の組み合わせと Kappa 係数の関係

10000 回の繰り返しによって算出した Kappa 係数の平均値を示す。エラーバーは標本標準偏差。取得時期の組み合わせ以外のすべてのケースの結果を統合した。

表1. データセット A, チャンネル 2 3 4 5 7 を使って 10000 回の訓練と予測から計算された精度評価値

データの組み合わせ	予測誤差				Kappa 係数			
	Mean	Median	Max	Min	Mean	Median	Max	Min
ID 1+ID 2+ID 3	0.022	0.020	0.060	0.000	0.950	0.952	1.000	0.867
ID 1+ID 2	0.018	0.015	0.065	0.000	0.959	0.962	1.000	0.826
ID 1+ID 3	0.025	0.025	0.070	0.000	0.943	0.943	1.000	0.838
ID 2+ID 3	0.027	0.025	0.070	0.000	0.939	0.941	1.000	0.836
ID 1	0.030	0.030	0.070	0.000	0.931	0.932	1.000	0.841
ID 2	0.031	0.030	0.075	0.000	0.929	0.931	1.000	0.811
ID 3	0.055	0.055	0.115	0.010	0.873	0.874	0.977	0.722

ID 1: 1999 年 8 月 11 日取得, ID 2: 1999 年 9 月 28 日取得, ID 3: 2002 年 4 月 29 日取得

太字は 7 ケースのデータの組み合わせにおいて最も精度が高いものを示す

慮しなかったサブカテゴリ (各森林タイプ, 水域, 市街地, 農地など) のそれぞれから十分な数のトレーニングデータを得ることが必要であると思われる。

取得時期の組み合わせに関しては, 分類精度の点からは利用するデータの数が多の方がより良いと思われた。一方, より多くのシーンの組み合わせは雲の被覆による解析可能エリアの減少, それに伴うトレーニングデータ数の減少, そして解析コストの増加を生じさせるため, 最終的にはこれらのバランスを考慮して解析者が決定する必要がある。また, データと精度の関係はシーン内に含まれるカテゴリの種類や対象地の農事暦等に依存するため, これらにも注意しなければならない。Coverage A と B との比較からはシーン内で大きく地表面の状態が異なるデータの解析については注意を要すると思われた。本研究の結果からは少なくとも 2 時期のデータを用いることが妥当であると考えられた。

IV おわりに

本研究は, 多時期のランドサットデータと SVM を利用し, 起伏の激しい我が国においても精度よく森林/非森林を分類できうることを示した。大気および地形効果に対して未補正のデジタル値を利用した解析結果であるものの, その分類結果の精度は極めて高かった。現在, 多数の地球観測衛星が運用され, またこれらの衛星によって観測されたデータが比較的安価で入手可能になりつつある現状を踏まえると, 多時期の衛星データを利用することは常に雲の影響を受ける熱帯などの地域を除いては有効な選択肢のひとつであると考えられた。

SVM は様々な分野において 2 クラスの分類を行う標準的な手法として確立している。本研究では, 森林と非森林の分類に注目して研究を行ったが, 今後はさらに衛星データから森林の変化 (例えば, 病害) の発生のある/なしを特定することなどに利用できるかもしれない。

この点については興味深いところであり, 衛星データの有効な利用法として研究を行う必要があると思われる。

謝辞

本研究で使用したランドサットデータはアメリカ地質調査所から提供されたものである。また, 海岸線の GIS データは国土交通省国土計画局から提供された国土数値情報 (海岸線) を利用した。ここに記して関係各位に謝意を表す。本研究は農林水産省プロジェクト研究「農林水産分野における地球温暖化対策のための緩和及び適応技術の開発」の支援を受けて実施された。

引用文献

- (1) BROWN DE COLSTOUN, C.E., STORY M.H., THOMPSON, C., COMMISSO, K., SMITH, T.G., and IRONS, J.R. (2003) National Park vegetation mapping using multitemporal Landsat 7 data and decision tree classifier. *Remote Sensing of Environment* **85**: 316-327.
- (2) 飯倉善和・横山隆三 (1999) ランドサット TM 画像の大気および地形効果の補正. 日本リモートセンシング学会誌 **19**: 2-16.
- (3) KNORN, J., RABE, A., RADELOFF, V., KUEMMERLE, T., KOZAK, J., and HOSTERT, P. (2009) Land cover mapping of large areas using chain classification of neighboring Landsat satellite images. *Remote Sensing of Environment* **113**: 957-964.
- (4) 田中真哉・高橋與明・齋藤英樹・栗屋善雄・平田泰雅・家原敏郎・松本光朗・酒井徹 (2010) 多時期データの利用による北海道東部の土地被覆分類. 日本生態学会第 57 回全国大会講演要旨集: 517.
- (5) TOTTRUP, C. (2004) Improving tropical forest mapping using multi-date Landsat TM data and pre-classification image smoothing. *International Journal of Remote Sensing* **25**: 717-730.