

リモートセンシングデータからの森林管理情報マイニング

Remote Sensing Data Mining for Forest Management

吉田圭吾*1
Keigo Yoshida大木孝*1
Takashi Ohki寺邊正大*1
Masahiro Terabe関根秀真*1
Hozuma Sekine武田知己*2
Tomomi Takeda出口知敬*3
Tomonori Deguchi*1(株)三菱総合研究所
Mitsubishi Research Institute, Inc.*2(財)資源・環境観測解析センター
Earth Remote Sensing Data Analysis Center*3日鉄鉱コンサルタント(株)
Nittetsu Mining Consultants Co., Ltd.

We propose estimation methods of tree species, growth and density in a forest from earth observation data by sensors in artificial satellites and aircrafts, also known as remote sensing data. Such estimated information is essential for forest management. Recently-developed hyperspectral sensors and sensor fusion technology give us rich information about forestry, whereas they cause a high dimensional and small sample size problem because of the limitation of training data collected from field survey. We apply sparse regularization to solve such $n \ll p$ problems, where n and p mean sample size and number of dimensions. Experimental results show the validity of our approach.

1. はじめに

森林は、二酸化炭素の吸収や、再生産可能で炭素の貯蔵機能を有する木材の生産を通じ、地球温暖化の防止に重要な役割を担っている。健全な森林を維持するために森林の施業管理は重要であり、近年では、航空機・人工衛星に搭載されたセンサを用いて広域を観測する、リモートセンシングを森林施業管理に応用することに対して、期待が高まっている。

本論文では、森林施業管理を行う上で森林現況の把握に有用な情報である、樹種、樹木の密集度、および樹木の生育状況を、リモートセンシングデータから推定する問題を扱う。例えば、ある区域における樹種を正確に知ることは、植栽時の樹種と比較して侵入木の存在を確認することができ、また、樹木密集度の把握は、間伐実施の意思決定の基準として、樹木の生育状況に関しては、炭素吸収量の直接的な基礎情報として活用できる。これらの情報は、森林資源情報を整理した森林簿と呼ばれる台帳に記載されているが、実際には、森林簿の作成時と現況が大きく異なることが珍しくない。

従来のリモートセンシングのデータを用いた森林現況の推定では、センサの観測値が示す物理的な特性を基に、観測値を演算して生成される経験的に導かれた指標値*1を利用したり、最も推定精度が高くなるような観測結果の最適な組合せを探索的に選択することが行われていた [Franklin 01]。

しかし、最近ではハイパースペクトルセンサと呼ばれる、ある地点の光学特性に関して高次元のデータ(図1を参照)を取得するセンサが普及し始めたため、上記のような演繹的な方法では、森林現況の評価にセンシング情報を十分に活用することが難しくなった。また、複数種類のセンサ情報を組み合わせるセンサ融合の技術を利用して、より詳細に森林現況を把握しようとする動きがあるが、演繹的に推定モデルを構築する際に、多様な物理特性を持つセンサ情報を扱わなければならないことは、非常に大きな壁であると言える。

そこで、我々はリモートセンシングにより得られた観測データから帰納的に推定モデルを構築するアプローチを採用する。今回扱う問題の難しい点は、与えられた高次元のデータに対して、現地調査の実施地点数が限られるために少数の事例しか得

ることができない、ということである。これに対して、提案する森林管理情報の推定方法では、以下の理由からスパース正則化を利用する。

- パラメータ数が疎なモデルの構築を指向するという手法の特徴により、高次元データかつ少数事例という条件下でもモデルの複雑化を抑制して汎化能力を確保できる。
- 推定に有効なパラメータを絞り込めるため、推定モデルの可読性が高い。

すなわち、本論文では、高次元のリモートセンシングデータという入力に対して、スパース判別分析(SDA; Sparse Discriminant Analysis)による識別と、スパース正則化を用いた回帰(LASSO回帰)という、2種類のデータマイニングの問題を解くことで、樹種、樹木密集度、森林生育状況という3種類の森林管理情報の推定を試みる。また、異種センサのデータを同時に利用するセンサ融合によって、より詳細に森林管理情報を把握できることを実験により示す。

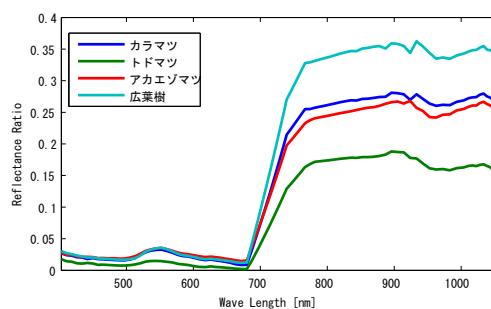


図1: ハイパースペクトルセンサによる樹木の反射スペクトル

2. 対象問題の設定

本節でまず、我々が解くべき問題について定義を行う。

リモートセンシング分野ではピクセル単位での分析が主流であるが、我々の目的は森林施業単位で森林管理上必要なデー

*1 例えば、赤波長と近赤外波長の値を用いて緑葉の特性を示す植生指数などがある。

連絡先: 吉田圭吾, (株)三菱総合研究所 科学・安全政策研究本部, 東京都千代田区大手町 2-3-6, Tel: 03(3277)0529, Fax: 03(3277)0567, e-mail: keigo@mri.co.jp

タを抽出することである。森林施業では通常、稜線などを境界として自治体ごとに設定された林班を、さらに樹種や林齢（樹木の年齢）によって細分化した、小班という単位で管理されている。また、各小班の森林現況の調査は悉皆的に行うことが難しいため、実際にはコドラートと呼ばれる調査区域を設けて、コドラート内の調査結果を小班の代表値として扱うことが一般的である。

つまり、我々が目的とすることは、森林施業単位である小班 S_i ($i = 1, 2, \dots, n$) と、それに対応する訓練データ $D = (\mathbf{x}_i, y_i)$ が与えられたときに、任意の入力 \mathbf{x} に対して目的変数 y を求めるような推定モデルを構築することである。ただし、 $\mathbf{x}_i \in \mathbb{R}^p$ は小班 S_i に対応するセンサデータを格納したベクトルであり、リモートセンシングデータからコドラート範囲に対応する各ピクセルにおける観測値を平均化して生成する。目的変数 y_i は、樹種の分類問題においては小班 S_i において支配的な樹種のラベルに、回帰問題においては現地調査で測定された、コドラート内の森林密集度と森林生育状況の測定結果に対応する。

3. スパース正則化による推定手法

本節では、樹種推定モデルの構築に使用するスパース判別分析、およびスパース正則化を用いた回帰分析の1つである LASSO 回帰について、簡単に述べる。なお、スパース判別分析は LASSO 回帰から導出されるので、まず LASSO 回帰について述べる。

3.1 LASSO 回帰

通常の回帰は、訓練データ集合 $D = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ が与えられたときに、経験誤差

$$\mathcal{L}(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^n (y_i - \mathbf{w}^T \mathbf{x}_i)^2, \quad (1)$$

を最小にするパラメータ \mathbf{w} を求める、という問題である。学習されたパラメータ \mathbf{w} を用いて、推定モデルは $\mathbf{y} = \mathbf{X}\mathbf{w}$ のように書くことができる。ただし、 $\mathbf{X} \in \mathbb{R}^{n \times p}$ は i 列目に \mathbf{x}_i を持つデータ行列を、 $\mathbf{y} \in \mathbb{R}^n$ は目的変数のベクトルを示す。

スパース正則化のモチベーションは、できるだけ少ない変数で説明する推定モデルを構築することであり、これはすなわち、 \mathbf{w} の非ゼロ要素の数をできるだけ少なくしながら経験誤差 (1) を最小化するパラメータ \mathbf{w} を求める問題であるので、次式のように表現できる。

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \mathcal{L}(\mathbf{w}) + \lambda \|\mathbf{w}\|_0, \quad (2)$$

ただし、 $\|\cdot\|_0$ は L_0 -ノルム、すなわち非ゼロ要素数を示し、 $\lambda \in \mathbb{R}^+$ はトレードオフパラメータである。式 (2) は組合せ解析により解く必要があるため NP 困難であり、現実的には凸最適化で解け、かつ L_0 -ノルムに最も近い L_1 -ノルムで代替する。

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \mathcal{L}(\mathbf{w}) + \lambda \|\mathbf{w}\|_1. \quad (3)$$

式 (3) で表される、 L_1 -ノルムの正則化項が付随した回帰が LASSO 回帰である [Tibshirani 96]。

3.2 スパース判別分析

判別分析では、クラス間の分散を最大にしつつ、クラス内分散を最小にする射影を求める。判別分析の場合、目的変数が離散値であるが、Optimal Scoring という方法により回帰問題として解くことが可能である [Hastie 94]。Optimal Scoring は、ラベルを示す質的変数を量的変数に変換することが基本的な考えである。 K クラスが存在するとき、クラス k ($k = 1, \dots, K$)

とパラメータ \mathbf{w}_j ($j = 1, \dots, J \leq K - 1$) へのスコア θ_{kj} を格納したスコア行列 $\Theta \in \mathbb{R}^{K \times J}$ を用いて、解くべき問題は次式のように定式化される。

$$(\theta_j^*, \mathbf{w}_j^*) = \underset{\theta_j, \mathbf{w}_j}{\operatorname{argmin}} \frac{1}{n} \|Y\theta_j - \mathbf{X}\mathbf{w}_j\|_2^2, \quad (4)$$

$$\text{s.t. } \frac{1}{n} \|Y\theta_j\|_2^2 = 1, \quad (5)$$

ただし、 $Y \in \mathbb{R}^{n \times K}$ はクラスへの所属を示すダミー変数を格納した行列を、 θ_j はスコア行列 Θ の第 j 列ベクトルを示す。スパースな判別分析の射影を求めるためには、式 (5) の制約条件のもとで次式の L_1 -正則化項が付加された問題を解けば良い。

$$(\theta_j^*, \mathbf{w}_j^*) = \underset{\theta_j, \mathbf{w}_j}{\operatorname{argmin}} \frac{1}{n} \|Y\theta_j - \mathbf{X}\mathbf{w}_j\|_2^2 + \lambda \|\mathbf{w}_j\|_1. \quad (6)$$

式 (6) は下記に示す手順により解く。

Step 1. まず、 Θ を固定して次式の L_1 -正則化項付き最適化問題を解く。

$$\mathbf{w}_j^* = \underset{\mathbf{w}_j}{\operatorname{argmin}} \frac{1}{n} \|Y\theta_j - \mathbf{X}\mathbf{w}_j\|_2^2 + \lambda \|\mathbf{w}_j\|_1. \quad (7)$$

Step 2. 続いて $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_J]$ を固定し、次式の最適化問題を解く。

$$\Theta = \underset{\Theta}{\operatorname{argmin}} \|Y\Theta - \mathbf{X}\mathbf{W}\|_2^2 \quad \text{s.t. (5)}. \quad (8)$$

式 (8) の解 Θ は、特異値分解 $Y^T \hat{Y} = Y^T (\mathbf{X}\mathbf{W}) = USV^T$ を用いることで、次式のように求めることができる。

$$\Theta = \left(\frac{1}{n} Y^T Y \right)^{-\frac{1}{2}} UV^T. \quad (9)$$

Step 3. Step 1, Step 2 を収束するまで繰り返し計算を行い、最終的に \mathbf{W} を更新する。

4. 実験設定

本節では実験の設定を述べる。実験は北海道下川町の町有林「敬和の森」と「21世紀の森」を対象に実施した。本実験では各小班に対して、樹種、森林生育状況の指標として樹高、森林密集度の指標として林冠率^{*2}と材積量^{*3}を求める。

4.1 実験データの準備

推定モデルの説明変数には、CASI-3 という航空機搭載のハイパースペクトルセンサと、陸域観測技術衛星「だいち」のフェーズドアレイ方式 L 波長帯合成開口レーダ (PALSAR) による、対象地域のリモートセンシングデータを使用する。

前者は観測対象の太陽光からの反射に関する光学的特性を取得するセンサであり、本実験では可視光から近赤外域にて連続的に設定された 84 波長帯の反射率を測定した。測定値に対しては輝度校正、大気補正、地形補正、および幾何補正などの前処理を施して使用する。後者はマイクロ波を用いて、対象物による散乱具合から対象の形状や物質に関する情報を取得するセンサである。実験では、水平照射水平受信 (HH)、水平照射垂直受信 (HV)、垂直照射垂直受信 (VV) の 3 偏波の後方

*2 単位面積当たりの、森林上層において葉が茂っている部分の表面積を示すが、本実験では空隙以外の部分の表面積率とした。

*3 木材の容積を示し、本実験では土地に生育する樹木の材積である立木材積を対象とする。

散乱のほか、リモートセンシング分野で一般的に用いられてきた、観測値から演算により算出される特徴量を合わせた、計9つの変数を使用する。測定値に対しては、地形陰影効果の補正を前処理として施す。ただし、モデル構築を行う前に、事前に平均が0、標準偏差が1になるように正規化を施す。表1に、使用するリモートセンシングデータの概要を示す。

表 1: 使用するリモートセンシングデータの概要

センサ名	観測データ		事例の数 ^{*4}
	データ内容	データの概要	
CASI-3	可視光-近赤外域 (400-1050[nm]) 間の 84 波長帯の反射率	光学的な特徴	84
PALSAR	後方散乱係数 (HH, HV, VV, 偏波) 表面散乱・二回反射・体積散乱成分 散乱エントロピ, α 角, Anisotropy	森林の植生量 や森林内部 に関する情報	9

航空機観測と同時期に、対象地域の現地調査を実施した。現地調査では、各小班に 25m 四方 (不可能な場合は 15m 四方) のコドラートを設置して毎木調査を行い、上層木の代表樹種、平均樹高などを測定した。材積量は樹高、樹木本数、胸高直径の測定値をもとに、樹種ごとに定められた北海道立木幹材積表に基づいて算出した値を使用する。林冠率については、各コドラートの中心点で全方位カメラを用いて全天写真を撮影し、画像における空隙のピクセル率と設定した。

なお、樹種分類においては、現地調査地点数の少なさを補うため、航空写真と森林簿情報を基に擬似コドラートを設置して、擬似コドラート内のピクセルに関する観測データの平均値をテストデータとして使用した。分類する樹種はカラマツ、トドマツ、アカエゾマツ、広葉樹の4つであり、内訳は、訓練データではカラマツ7点、トドマツ4点、アカエゾマツ4点、広葉樹4点、テストデータでは各樹種5点ずつである。

回帰問題では、現地調査を行った小班の75%相当を訓練データとしてモデル構築に使用し、残りの25%相当の予測を行う。その際に、訓練データの選び方による評価のバラつきに影響されないようにするため、訓練データをランダムに選び推定精度の評価を行う試行を1000回繰り返し、精度の平均と標準偏差を確認した。なお、事例の数が少数となるため樹種ごとの評価は行わなかった。

4.2 実験の評価項目

実験では2つの項目に関する評価を行う。

1) スパース正則化による森林管理情報抽出性能の評価

樹種の分類と林冠率を対象に、スパース判別分析、および LASSO 回帰を用いた推定を行い、精度に関する定量的な評価と選択変数に関する定性的な評価から、森林管理情報の抽出におけるスパース正則化の有用性を評価する。

分類モデル構築手法の比較の対象として、 k -近傍法 (k -NN) (ただし、入力形式として生データ (raw)、主成分 (PC)、カーネル主成分 (KPC) の3つを使用する)、 L_2 -ノルムを用いた正則化判別分析 (RDA)、カーネル判別分析 (KDA)、 ν -SVM を使用した。KDA、 ν -SVM では RBF カーネル $k(\mathbf{x}_1, \mathbf{x}_2) = \exp(-\|\mathbf{x}_1 - \mathbf{x}_2\|^2 / 2\sigma^2)$ を使用する。また、 ν -SVM の多クラス分類問題への適用には、誤り訂正出力符号 [Dietterich 95] を利用した。回帰モデル構築手法の精度比較の対象としては、リッジ回帰 (RR)、主成分回帰 (PCR)、カーネル回帰 (KR)、サポートベクター回帰 (SVR) を使用した。各手法のハイパーパラメータは LOOCV の誤差に基づいて決定した。

2) センサ融合による森林管理情報抽出性能の評価

材積量と樹高を対象に、CASI-3 と PALSAR という異種のセンサによる観測データを融合することで、森林管理情報の推定精度の向上について定量的な観点から評価する。回帰モデル

には LASSO 回帰を使用し、LOOCV の誤差に基づいてハイパーパラメータ λ を決定した。

以上述べた各実験の設定を表2に整理した。

表 2: 実験設定の整理

抽出情報	目的変数	説明変数		事例の数 ^{*4}
		CASI-3	PALSAR	
樹種	樹種ラベル	84 変数	-	39 (19)
樹木密集度	林冠率 [%]	84 変数	-	25 (18)
	材積量 [m ³ /ha]	84 変数	3 変数 (HH, HV, VV)	17 (12)
森林生育状況	平均樹高 [m]	84 変数	9 変数	18 (13)

5. 実験結果と考察

5.1 スパース判別分析による樹種分類

表3に、各手法によりテストデータの樹種分類を行った正答率を示す。正則化を用いる RDA, KDA, ν -SVM, SDA はいずれも80%の正答率を達成しており、SDAの正答率は90%を実現している。実際には、SDAが誤分類した小班では画像から異なる色彩の領域の混在が確認でき、複数の樹種が存在するために誤分類をした可能性が考えられる。

表 3: 樹種判別の正答率 [%]

推定器	k -NN (raw)	k -NN (PC)	k -NN (KPC)	RDA	KDA	ν -SVM	SDA
カラマツ	80	100	100	100	100	100	100
トドマツ	20	80	60	100	100	100	100
アカエゾマツ	80	60	80	80	60	80	60
広葉樹	80	80	80	100	80	60	100
全体	65	80	80	95	85	85	90

また、図2には SDA と RDA でそれぞれ求められた判別関数の一部を示す。図2より、SDA では非常にスパースな解が得られていることが分かる。選択された変数は、反射率の変化域や水分吸収帯など、樹種による差異が生じやすいとされる波長帯を示すものであった。図3には、森林簿情報、青・赤・緑 (450, 513, 656[nm]) を重ね合わせたトゥルーカラー画像、SDA で判別係数の大きかった3波長帯 (482, 521, 887[nm]) を重ね合わせたフォールスカラー画像を示した。フォールスカラー画像では各小班が樹種別に識別しやすく、SDA により樹種分類に適切な波長帯が選択されたことが確認できる。

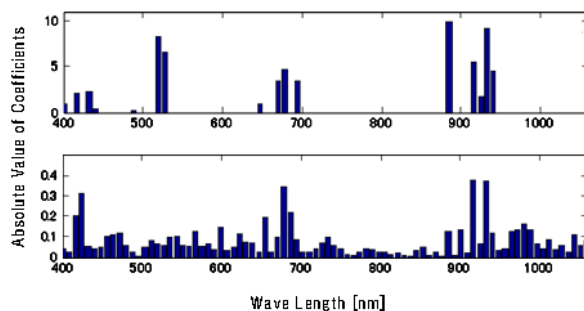


図 2: SDA (上) と RDA (下) の判別関数の比較

5.2 LASSO 回帰による樹木密集度 (林冠率) 推定

表4に、各手法により林冠率を推定した際の訓練誤差、予測誤差、および予測誤差の標準偏差を示す。なお、誤差は RMSE である。正則化を行う各手法とも約4%の誤差で予測が可能で

*4 () 内は訓練データの数を示す

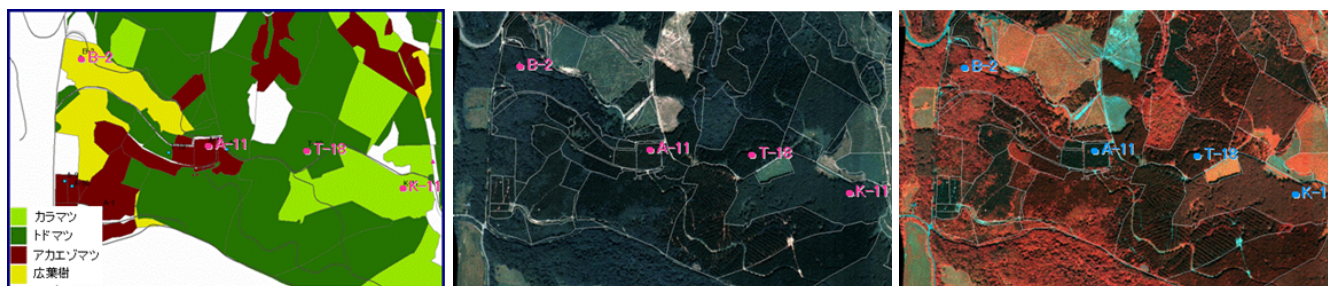


図 3: 森林簿 (左), ツールカラー合成画像 (中), フォールスカラー合成画像 (右) の比較

あり, 比較した手法の中では LASSO 回帰の予測誤差が最も小さかった. LASSO 回帰で獲得された推定モデルでは, 入力された 84 変数のうち 9 変数のみが使われる結果であったが, 他手法と遜色ない推定精度を達成している. 選択された 9 変数は, 植生の特徴を良く示す緑, 赤, 近赤外域の波長帯に対応しており, スパース正則化によって推定に必要な特徴のみが選択されていることが伺える.

表 4: 林冠率の推定誤差 (RMSE) [%]

推定器	RR	PCR	KR	SVR	LASSO
訓練誤差	1.99	2.24	1.62	0.75	0.45
予測誤差	4.73	4.36	4.03	4.36	3.93
σ (予測誤差)	1.43	0.97	0.92	1.23	1.02

5.3 センサ融合による樹木生育状況, 樹木密集度推定

表 5 に, 各センサ単独の場合と, センサ融合を行った場合のそれぞれについて, 材積量, および樹高推定時の訓練誤差, 予測誤差, 予測誤差の標準偏差を整理した. なお表 5 の中では参考までに, リモートセンシング分野で一般的な森林管理情報の推定方法である, 異なる 2 時期の後方散乱の複素相関係数である, コヒーレンスという値を用いて材積量を推定する方法と, PALSAR 偏波間の位相差を利用して樹高を推定する方法を演繹的分析と呼び, LASSO 回帰を用いた帰納的分析との比較を行っている. 材積量, 樹高ともに異種センサのデータを融合してモデル構築に利用することで, 推定精度が向上している様子が分かる. 特に材積量に関しては, 演繹的分析時と比べて誤差が約 48%減少するという結果が得られた. また, 材積量のセンサ融合時の推定モデルでは 93 変数のうち 4 変数 (CASI-3 1 変数, PALSAR 3 変数), 樹高の推定モデルでは 87 変数のうち 4 変数 (CASI-3 3 変数, PALSAR 1 変数) のみで使用されていた.

表 5: 材積量, 樹高の推定誤差 (RMSE)

	材積量 [m ³ /ha]			樹高 [m]		
	訓練誤差	予測誤差	σ	訓練誤差	予測誤差	σ
演繹的分析	148.6			5.03		
帰納的分析						
CASI-3	21.2	89.4	28.9	1.39	5.78	1.06
PALSAR	16.0	90.8	23.1	4.27	5.16	1.22
センサ融合時	14.4	76.6	24.4	2.19	3.18	0.84

6. まとめと課題

リモートセンシングデータから森林管理に有用な情報を抽出する, という問題を少数事例・高次元データからの回帰, 分類問題とみなし, スパース正則化を適用した. 現地調査結果と実際のリモートセンシングデータを用いた実験により, 適用手法が少数事例かつ高次元のデータに対しても十分な予測性能を

確保しつつ, 入力された多くの説明変数から推定に有用な少数の変数が抽出されることを確認した. また, この性質を活かして, 物理的特性が異なる複数のセンサの観測データを融合した際にも, モデルの複雑化を避けつつ, より高い精度で森林管理情報を推定できるモデルが構築されることを確認した.

より多くの検証地点を設定し, 推定モデルの汎用性を検証することが今後の課題である. また, 小班内に複数の樹種が混在している場合に対処するためにマルチレベル分類などを適用する必要がある. さらに, コドラート内は樹木の均一性が比較的保たれているが, 実際には侵入木, 風倒木, 部分的生育不良, 伐採などの影響により, 未調査地点とコドラート内で異なる分布を示すことが想定される. このような問題に対処するために, 例えば共変量シフト [Shimodaira 00] を用いることなどが考えられる.

なお, 経済産業省では衛星搭載型のハイパースペクトル/マルチスペクトルセンサを開発中である. 本研究の成果は, 開発中のハイパースペクトルセンサのデータ処理技術においても利用が期待される.

謝辞 本研究は, 財団法人資源・環境観測解析センターが実施する「平成 21 年度 次世代地球観測衛星利用基盤技術の研究開発」の一環として行われたものである. 本研究は下川町役場からの協力を受けて実施され, 森林関連データについては下川町役場から借用した. ここに記して感謝の意を表する.

参考文献

- [Clemmensen 08] Clemmensen, L., Hastie, T. and Ersboell, K., Sparse Discriminant Analysis, Technical report, IMM, Technical University of Denmark, 2008.
- [Dietterich 95] Dietterich, R. and Bakiri, G., Solving multiclass learning problems via error-correcting output codes, Journal of Machine Learning Research, 2:263–286, 1995.
- [Franklin 01] Franklin, S. E., Remote Sensing for Sustainable Forest Management, CRC Press, June, 2001.
- [Hastie 94] Hastie, T., Tibshirani, R. and Buja, A., Flexible Discriminant Analysis by Optimal Scoring, Journal of the American Statistical Association, 89(428):1255–1270, 1994.
- [Shimodaira 00] Shimodaira, H., Improving Predictive Inference under Covariate Shift by Weighting the Log-Likelihood Function. J. Statistical Planning and Inference, 90:227–244, 2000.
- [Tibshirani 96] Tibshirani, R., Regression shrinkage and selection via the lasso, J. Royal Statistical Society. Series B (Methodological), 58(1):267–288, 1996.