

IV-328 衛星マルチスペクトル画像分類におけるトレーニングデータ内の混在画素除去効果の検証

東京理科大学 正会員 大林成行

東京理科大学 正会員 小島尚人

東京理科大学 学正員 中島剛彦

1. はじめに：衛星マルチスペクトルデータを用いた土地被覆分類において広く一般に利用されている教師付き最尤法分類では、トレーニングデータの良否が分類精度に大きく影響を及ぼすことは言うまでもない。トレーニングデータは現地調査や航空写真等の情報を基にして土地被覆項目別に対応する領域を取捨選択するが、選定したトレーニングデータが必ずしも現地の土地被覆状況を代表しているとは限らず、トレーニングデータの選定については多くの問題が指摘されている。このような状況のもと、著者らはトレーニングデータ内に混在する画素を除去し、トレーニングデータそのものの精度（性能：Performance）を向上させるアプローチに焦点を絞って研究を展開した。その成果として詳細な現地調査を通じて選定したトレーニングデータに対して「教師付き分類（最尤法）」と「教師なし分類（ISO-DATA法）」をそれぞれ施し、両者の分類結果が一致した画素を抽出する方法を提案した。地上分解能の異なる衛星データ（MESSR、TM、H RV）に対する適用効果の確認を含めて、多くの検討を通じて本提案手法の実用性と普遍性を示した¹⁾。しかし、トレーニングデータ内の混在画素を除去することが目的とする分類図作成に意義のあることが否かについては、国内外を問わず今なお多くの議論が交わされている²⁾。そこで、本研究ではさらにトレーニングデータの再抽出条件の異なるケースとして、教師付き分類と教師なし分類をそれぞれ別々に利用して混在画素を除去する場合の分類精度と比較し、提案手法が分類精度を向上させる上で最も有効であることを示す。

2. 検討ケース：本研究ではH RVデータ（1988年10月14日：PATH332-ROW278）を使用した。評価領域は水田、畑地、市街地等の代表的な土地被覆項目が混在し、詳細な現地調査を入念に実施しやすい千葉県印旛沼周辺とした。設定した検討ケースを表-1に示す。ケース1は初期設定したトレーニングデータをそのまま使用するものである。ケース2とケース3が本研究で追加・検証したものである。ケース2は、初期設定したトレーニングクラスを教師としてトレーニング領域の全画素を最尤法によって分類し、自分自身のクラスに判別されなかった画素を混在画素として除去するものである。教師付き最尤法のみによる「トレーニングデータの洗練化」と言えるが、スペクトル的にクラス間の分離性が保障されるとは限らない。ケース3ではISO-DATA法を用いてトレーニング領域の全画素を分類した後にマハラノビスの分布間距離を指標として初期設定した教師クラスとの距離が最も小さいクラスに対応付け、一致した画素のみを抽出する。この方法ではスペクトル的にクラス間の分離性は保障されるが、必ずしも現地の土地被覆項目に対応しているとは限らない。ケース4は著者らが提案している手法であり、ケース2とケース3の手順を併用して両者の欠点を相互に補った画素を抽出するところに特徴がある¹⁾。著者らの知る限り現在これと同様なアルゴリズムは見当たらない。以上のケース1～ケース4によって再抽出される画素が属する集合を判り易くする表現すると図-1のようになる。

表-1 検討ケース

	混在画素の除去条件
ケース1	混在画素を除去しない
ケース2	最尤法（教師付き）による除去
ケース3	ISO-DATA法（教師なし）による除去
ケース4	ケース2とケース3を併用（提案手法）

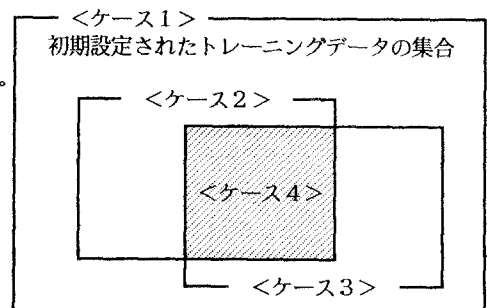


図-1 再抽出画素の属する集合

3. 分類精度の検証：本研究では、再抽出したトレーニ

ングデータを用いて教師付き最尤法による分類精度を検討する。トレーニングデータの精度は区分精度、誤分類率（OmissionおよびCommission Errorを考慮）によって検証し、画像全体にわたる分類精度の評価は、あらかじめ整備した評価用データを基に算出されるPCC（Probability of Correct Classification）を用いた。また、確率プロット法によって教師付き最尤法の前提条件である多次元正規性を検証し、クラス間の分離性についてはクラス間分散度を用いて評価した。

表-2 RMS誤差の比較（確率プロット図より算出）

	ケース1	ケース2	ケース3	ケース4
水田1	0.027	0.013	0.015	0.013
水田2	0.036	0.010	0.011	0.011
畑地	0.024	0.029	0.030	0.034
樹林	0.044	0.014	0.020	0.019
人工構造物	0.048	0.037	0.012	0.014
芝地	0.011	0.010	0.008	0.009
裸地	0.086	0.017	0.066	0.029
水域	0.072	0.049	0.050	0.049
平均値	0.044	0.022	0.027	0.022

表-3 分類精度の比較

	ケース1	ケース2	ケース3	ケース4
区分精度 (%)	水田1	93.3	100.0	100.0
	水田2	82.9	96.6	99.6
	畑地	67.7	99.1	94.8
	樹林	96.1	98.4	98.7
	人工構造物	93.3	100.0	99.3
	芝地	79.6	98.8	100.0
	裸地	94.0	99.1	96.4
	水域	97.4	99.1	99.4
	平均値	88.0	98.9	97.9
誤分類率 (%)	水田1	11.9	1.0	1.4
	水田2	42.1	4.4	6.4
	畑地	65.6	7.5	12.2
	樹林	11.1	1.6	4.1
	人工構造物	12.3	1.1	2.1
	芝地	34.9	2.1	1.6
	裸地	9.9	1.0	3.6
	水域	2.6	0.9	0.9
	平均値	23.8	2.4	4.0
クラス間分散度 (平均値)	1.83	2.17	2.12	2.25
PCC (%)	83.4	88.7	91.5	93.8

（1）多次元正規性の検証：確率プロット図から算出されるRMS誤差を表-2に整理した。ここで用いるRMS誤差の定義については割愛するが¹⁾、この値が小さい程多次元正規性が保障されることを意味する。ケース2とケース4の有意な違いは認められないが、ケース1に比べて多次元正規性が向上していることが伺える。このような多次元正規性の確認を踏まえて、分類精度の比較へと検討を進めた。

（2）分類精度の比較：分類精度の算出結果を表-3に整理した。区分精度を見ると全ての土地被覆項目についてケース4が最も高い値を示している。特に、畑地や芝地等のように他の土地被覆項目を多く含む傾向にあるトレーニングクラス内の混在画素の除去効果が高いことが判る。また、誤分類率でもケース4が最も低い値を示し、提案手法による適用効果が認められる。特に畑をはじめとし、稲の刈入れが済んでいる輝度値の高い「水田1」と、刈入れ前の輝度値が低い「水田2」の誤分類が少なくなり、分類図上でもその効果は確認している。このようにトレーニングデータそのものの精度が向上したとしても、必ずしも画像全体におよぶ分類精度が向上するとは限らない。そこで、PCCによって客観的に画像全体にわたる分類精度を評価する。このPCCを見てもケース4が最も高くなっている。クラス間分散度（平均値）もケース4が最も高い値を示している。本研究で用いたクラス間分散度は多次元正規性の前提の基に定義される統計指標であり、前述した多次元正規性の向上に加えて、教師付き最尤法分類による画像全体の分類精度の向上も裏付けることができる。この他にも多くの検証を通じて、提案手法の適用効果を確認している。

4. まとめ：トレーニングデータの選定方法をはじめ分類精度に関する研究は枚挙にいとまがない。混在画素を除去するアプローチ一つをとっても今なお多くの議論がある。このような状況の中にあって忘れてはならないことは、誰しもが手軽に利用でき、その適用効果が期待できる実用性のある処理アルゴリズムを提示することである。衛星データは益々高分解能化し、より高い画像処理／解析精度が要求される。本研究が今後我々が手にするより分解能の高い衛星データの土地被覆分類精度の向上にも寄与できれば幸いである。

【参考文献】1)大林成行、小島尚人、Tashpolat Tiyp、中島剛彦：トレーニングサンプルの抽出方法に関する考察、日本リモートセンシング学会第11回学術講演会論文集、pp.201～pp.204、1991年12月
2)Story, M.H., and J.B., Campbell: The Effect of Training Data Valiability on Classification Accuracy, ASPRS 52nd, Annual Meeting, Technical Papers, Vol.4, pp.370～pp.379, 1986.