

分散を考慮したGA・MLH分類法*

2U-5

小原 裕史 寺崎 浩 星 仰†

茨城大学‡

1 はじめに

リモートセンシング衛星から抽出される画像データの土地被覆分類は、数多くの画像単位分類手法が研究され、高速化、高精度化が図られている。衛星画像における分類では分光特性が特徴量として用いられている。画素ごとに与えられる特徴量を用いて、その画素をいくつかのカテゴリに分類する手法には各分類項目の特徴量をあらかじめ教える教師付き分類法とそれをしない教師なし分類法の2つに分けられる。教師なし分類法で代表的な高速処理ものにK-means法などが挙げられる。

従来、遺伝的アルゴリズムを用いた教師なし分類も行われてきたが、その手法では分散を考慮できないという問題点があった。そのため本研究ではある程度のデータのバラツキ(分散)を考慮にいれ、その値を用いて最尤法で分類を行うGA・MLH法を提案する。また、分散を用いた場合とそうでない場合の分類結果についても比較検討する。

2 遺伝的アルゴリズムによる分類法

2.1 変数の減量化

衛星画像データは複数のバンド、4から7バンドで構成される。そのため、遺伝的アルゴリズムにより極大点を探索するとき探索空間が小さいとそれだけ処理計算が短くなるので、探索空間を主成分分析により縮小する。この主成分分析に用いる関係式を以下に表す。

$$Y_j = \alpha_1 x_{1j} + \alpha_2 x_{2j} + \dots + \alpha_q x_{qj} \quad (j = 1, 2, \dots, n) \quad (1)$$

ここに、Y: 主成分、q: バンド数、x: 画像データ、n: データ数、 α : 係数、u: 減量化後のデータである。

$$\alpha_1^2 + \alpha_2^2 + \dots + \alpha_q^2 = 1 \quad (2)$$

$$V(Y) = \frac{1}{n} \sum_{j=1}^n \{ \alpha_1(x_{1j} - \bar{x}_1) + \dots + \alpha_q(x_{qj} - \bar{x}_q) \}^2 \quad (3)$$

上記の式(2),(3)から α_i を求め、以下の式(4),(5)で第二主成分まで算出する。

$$u_1 = \alpha_{11}x_1 + \dots + \alpha_{q1}x_q \quad (4)$$

$$u_2 = \alpha_{12}x_1 + \dots + \alpha_{q2}x_q \quad (5)$$

2.2 主成分領域での極大値の抽出

極大点の候補点の探索を行うため、初期個体を128点与え、ある探索を決め、この範囲内で最大となる極大点を遺伝的アルゴリズムを2回適用することによって決定する。各回の処理内容は下記の通りである。

1. 極大点の候補点を探索するため、ある世代間隔毎に、以前の世代より評価の高い個体を保存する。
2. 複数の極大候補点を結合するため、ある範囲内の複数の候補点を結合する。

2.3 遺伝的アルゴリズムの手順

1. 遺伝子型の決定(記号列への変換方法)

第1主成分をx軸、第2主成分をy軸とし、それぞれの値 $p(x,y)$ を9bit(0-511)で表し、適応度を $f(x,y)$ 、世代数を i 、適応度の増加率を p_i とする。

2. 初期遺伝子の決定

$p(x,y)$ に対し初期個体数 $M(=128)$ 点をランダムに配置する。適応度の評価

3. 以前の世代より評価が高い時、適応度を更新する。評価が低いときは、適応度は更新されない。選択

4. $p(x,y)$ について、エリート保存戦略と山登り法を使用し、 $p(x,y)$ を中心とする矩形(5×5)で、評価が最大のものを選択し、それを $p(x,y)$ とする。突然変異

5. 突然変異確率により、各遺伝子のあるbitのみの反転を行う。

6. 終了条件

GAにおける計算の終了条件は以下に示す。

- ・世代数がある閾値を越えた場合
- ・適応度の増加率の和がある閾値を越えた場合

3 GA・MLH分類法

3.1 分散値の決定

遺伝的アルゴリズムによって抽出されたクラスターはクラスターの平均値などを代用するが、そのクラスターのバラツキ状態を示す指標が見当たらない。そこでその分散となる値を以下のように求める。

各頻度分布が正規分布であると仮定する。式(4),(5)から構成される主成分分析後の (u_0, u_1) 空間においてGAにおける探索範囲の極大候補点の分散を計算する。

$$\sigma_{u0} = K_{u0,1} + K_{u0,2} + \dots + K_{u0,m} \quad (6)$$

$$\sigma_{u1} = K_{u1,1} + K_{u1,2} + \dots + K_{u1,m} \quad (7)$$

ここでKは極大候補点、mは極大候補点の数である。

*Classification method using Genetic Algorithm and Maximum Likelihood Method considering variance

†Yuji Obara, Terasaki Hiroshi and Takashi Hoshi

‡Ibaraki University

4-12-1 Naka-narusawa, Hitachi, Ibaraki 316, Japan

極大点の周りに多数の極大値候補点が密集していた場合、分散は小さくなり、極大値候補点が拡散してるほど分散は大きくなる。このように求めた分散値を用いて、最大尤度法により分類を行う。

3.2 最大尤度法

最大尤度法は Bayes の定理に基礎を置く統計的手法であり、教師付き分類法の中でも最も一般的に利用されている手法の一つである。

k クラスの平均を \bar{x} 、分散共分散行列を S_k とすると各画像データ $x = \{x_j | j = 1, n\}$ に対する第 k クラスの尤度 $f(k)$ は、次の式で表せる。

$$f_k(x) = \frac{1}{(2\pi)^{\frac{n}{2}} |S_k|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} d_k^2\right\} \quad (8)$$

ここで、 $|S(k)|$ は $S(k)$ の行列式であり、 $d(k)$ は次式となる。

$$d_k^2 = (x - \bar{x}_k)^t S_k^{-1} (x - \bar{x}_k) \quad (9)$$

そして、未知データは尤度 $f_1(x), f_2(x), \dots, f_i(x)$ のうち最大尤度 $f_k(x)$ を持つ第 k クラスへと分類される。

4 実験結果

主成分分析の結果画像を図 1、分散を考慮しない場合の分類結果の画像を図 2、分散を考慮した場合（最尤法）の分類結果の画像を図 3 に示す。使用した画像は JERS1.OPS の画像でサイズは 512×512 pixel である。図 2、図 3 を見比べると、分散の影響によりクラスの境界線にかなりの違いが見られる。

5 おわりに

以上から、分散を検出することによって GA・MLH 分類を可能にした。高速化のため、主成分分析によりバンド数の次元のデータを 2 次元のデータにして処理を行ったが、前述した方法ならば 3、4 次元データの分散も算出は可能である。今後、バンド数の次元で極大値を全探索した場合での分類結果の比較などを検討したい。

参考文献

- [1] 星 仰、寺崎浩：“GA による RS 画像データのクラスタリング”、計測自動制御学会第 23 回リモートセンシングシンポジウム講演論文集、pp.91-94,1997.10
- [2] 寺崎浩、星 仰：“遺伝的アルゴリズムを用いたトレーニングクラスの補充抽出法”、日本写真測量学会学術講演会論文集、G-2,pp.23-26,1997.5



図 1 主成分分析の結果

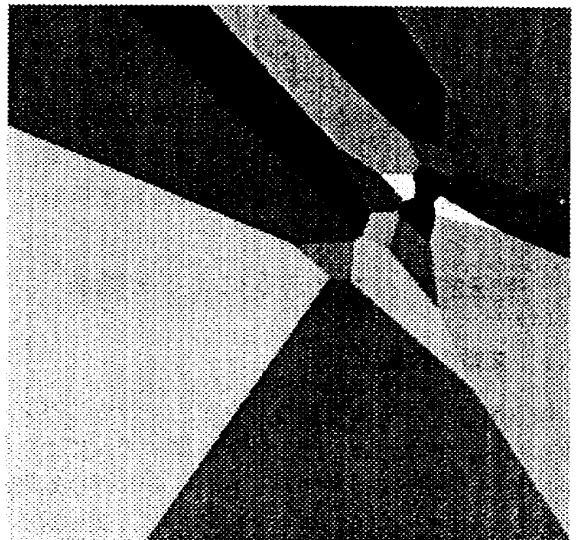


図 2 分散を考慮しない場合

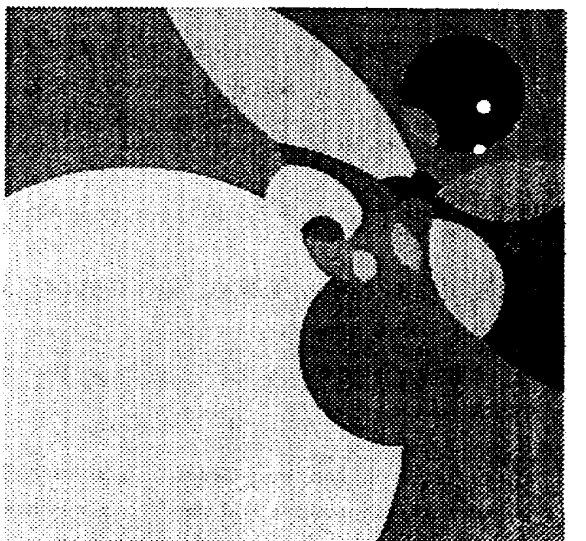


図 3 分散を考慮した場合