

IV-217

遺伝アルゴリズムによる
衛星画像データ分類への試み

茨城大学工学部 正員 星 仰
茨城大学大学院 山本真靖

1. はじめに

遺伝アルゴリズム (Genetic Algorithm : GA) は生物の進化のメカニズムをシミュレートすることを念頭におき、これを人間が人工的なモデル化するためのアルゴリズムであろう。このGAは探索システムの手法に応用されたり、学習システムにも応用することができる。この遺伝アルゴリズムはHollandが□進化的適応システムとしての遺伝アルゴリズム□を1975年に提唱してから注目され始め、1985年より隔年の国際会議が開催されるに至っている。本研究はリモートセンシング画像解析によく用いられているパターン認識の中で教師付分類と教師なし分類があるが、これらの画像データの分類への1手法として適応可能かどうかを手法面から考察し、適用性を模索してみる。

2. 遺伝アルゴリズムの概念

遺伝アルゴリズムには、個体(individual)と呼ばれる遺伝子によって特徴づけられるものとこれらを集めた集団(population)がある。遺伝子(gene)は遺伝情報を表現する基本的構成要素で、これの取りうる値を対立遺伝子(alles)と呼び、この遺伝子はストリングとして表現される。これを染色体(chromosome)といい、この存在場所を遺伝子座(locus)という。遺伝アルゴリズムの中では個体の環境に対する適応度(fitness)を調べたり、遺伝的操作(genetic operation)がなされる。この操作の中には選択(selection)、交叉(crossover)、突然変異(mutation)を含み、遺伝的操作の1サイクルを1世代(generation)という処理の必要性によっては並列化(parallelization)や移民(migration)なども行われる。一般的な遺伝アルゴリズムでは①初期化、②選択、③交叉、④突然変異の4段階が組み込まれ、2~4段階で1世代となる。項目②の選択の戦略(selection strategy)には各個体の適応度に比例した確率で選択をするルーレット選択、適応度の高い順に決定された確率で選択をするランキング選択および現世代で適応度の高い個体を次世代に残すエリート保存戦略とがある。項目③の交叉には染色体上の1ヶ所でのみ交叉する1点交叉、染色体上の複数ヶ所で交叉する多点交叉、マスクビットパターンに従い、どちらの親の遺伝子を継がせるかということを決定する一様交叉および遺伝子が連続値を取るとき両親の中間の値を子に継がせる混合交叉とがある。

3. リモートセンシング画像データへの試み

リモートセンシング画像データは巨視データのため、基本的には詳細なパターンから単純なパターンに置換することが多い。近年ニューラルネットワークやFuzzyによるクラスタリングも見受けられるが、基本的なクラスタリングは教師なし分類ではK-means, ISODATA, Fuzzy-Kmeans/ISODATAであろう。これらの手法は、基本的に画像データ総当たり法であり、効率的な探索法が要求されている。GAでは、この問題を直接解決できないかもしれないが、遺伝アルゴリズムの交叉、選択は少なくとも総当たり法ではないため何らかの糸口かつめる可能性を含んでいる。そこで、本研究ではリモートセンシング画像データに含まれる基本分類項目5種類(水域、植生域、都市域、裸地域、雲・雪域)を取りあげ、これらのクラスの極大値をどの程度の精度で抽出できるかということを実験的に試みることにした。表-1はLANDSAT, MSSのバンド4, 7を想定したCCTカウント値である。これをもとに頻度分布を約 $1,000 \times 1,000 \times 2$ の画素を想定して作成したものが図-1である。このモデルでは、点A(200)に最大値を持つため、GAを適用したとき点A近傍は得られても、他のB(170), C(150), D(150), E(100)の極大値は抽出困難と想定される。そこで本研究では、初期収束した個体付近にペナルティを与えるようGAを改良す

ることとした。使用パラメータは個体数：10、世代数：20、交叉確率：0.6、突然変異確率：0.2である。ペナルティ半径：Rと保存世代間隔：Gsaveについてはシミュレーションしてみることとした。図-2は Gsave の変化による最良個体の分布状態を示している図-3はペナルティ半径Rの変化による最良個体の分布状態を示している。この実験ではR=10、Gsave=3~4程度が良好な結果を得たといえよう。この結果、遺伝アルゴリズムはクラスタの中央値の近似値を算出するのに応用できる見通しができたと云えよう。

4. おわりに

遺伝アルゴリズムの問題点の1つは初期収束あるいは、局所解に陥りやすいことである。リモートセンシング画像データのクラスタリングではn個のクラスタに分類するにはいくつかの極大値が存在するため、これらを見い出す必要がある。本研究のペナルティゾーンの適用は最適解のみではなく、局所的最適解を求める可能性を見い出したということかでき、今後のクラスタリングへの可能性を示したといえよう。今後は、GAパラメータを更に検討し、リモートセンシング画像データのクラスタリングへ応用していく予定である。

各分類項目のCCTカウント値

	植生	人工構造物	裸地	表面水	雪・雲
Band 1	6.4	14.6	9.2	1.6	22.6
Band 2	8.4	13.2	9.8	1.4	23.4
Band 3	5.4	15.8	10.6	1.0	23.8
Band 4	20.0	13.8	11.4	4	24.4



図-1 頻度分布

