

クラスタ分類法による初期条件の設定

(株) 建設技術研究所 ○青山 定敬

日大生産工 工藤 勝輝 藤井 寿生 西川 肇

1 はじめに

衛星データを使って作成される土地被覆分類画像データは、洪水シミュレーションモデルや植生図の作成等、様々な分野で利用が期待されるデータである。

衛星データを使った土地被覆分類手法は、一般に最尤法に代表される教師付き分類法とK-means法やISODATA法に代表される教師なし分類法（クラスタ分類法）に大別される。

教師付き分類法は、分類者があらかじめ土地被覆の地物毎に画像の特徴量を表す統計データを抽出し、これに基づき該当する最適な地物に分類する方法である。この方法による分類は、分類精度があらかじめ抽出する教師データの統計量の精度に依存し、教師データ取得者の熟練度によって答えが左右されるという問題がある。

一方、教師なし分類法はコンピュータが自動的に土地被覆を分類するため、人の熟練度とは関係なく分類できるが、分類の初期条件設定によって答えが左右されるという問題がある。この場合、分類クラスタ数をあらかじめ設定するk-means法よりも、分類クラスタ数を画像の特徴によって自動的に設定するISODATA法の方がより人を介さずに分類画像を作成することができる。

本研究は、クラスタ分類法のISODATA法を土地被覆分類画像作成等の実利用に用いることを目的とし、ISODATA法における最適な初期条件設定方法について検討した結果を報告する。

2 使用ソフト及びデータ

ISODATA法による分類は、初期条件としてクラスタの最小サイズ（最小限の画素数）、クラスタ間の限界距離を設定することで、自動的に分類クラス数及びそれに応じた分類画

像を作成する方法であり、クラスタ間距離が限界距離内に入っているクラスタや最小サイズに満たないクラスタは、他の隣接するクラスタと統合し、最適なクラスタに分類される。

本研究で使用するISODATA法による分類画像作成ソフトウェアは、初期パラメータが分類結果に与える影響を把握するため、著者自作のものを用了。

また、分類画像を作成するための衛星リモートセンシングデータは、RESTECがインターネットで公開しているALOSのAVNIR-2オルソライトサンプル画像（山梨、2008年5月22日観測、処理レベル1 B2R）を使用し、研究用に甲府駅南側周辺を横200画素×縦200画素に切り出した画像を用了。分類画像作成に使用したバンドは、バンド1～4の4バンドである。

3 シミュレーション条件

(1) 初期クラスタ設定

ISODATA法では、あらかじめ各クラスタの輝度値を設定しなければならない。

分類結果は、この設定方法に依存する。

本研究では、各バンドにおいて輝度をあらかじめ設定した区分数に従い、等間隔で区分したクラスタを設定するという方法を採用した。この場合、区分範囲は、各バンドの最小値から最大値の範囲を用了。

2次元座標における初期クラスタ設定の概念図を図1に示す。

この初期クラスタの設定方法は、各バンド間をマトリックスで区分する全クラスタを対象とし、かつ最小値から最大値の範囲のみを対象とすることでよりきめ細かくクラスタを設定できるという特徴がある。この方法では例えば、4バンドをそれぞれ10分割した場合の初期クラスタ数は、10000（ $=10^4$ ）クラスタになる。

Initial Condition Setting in Cluster Classification Method

Sadayoshi AOYAMA, Katsuteru KUDOH, Hisao FUJII and Hajime NISHIKAWA

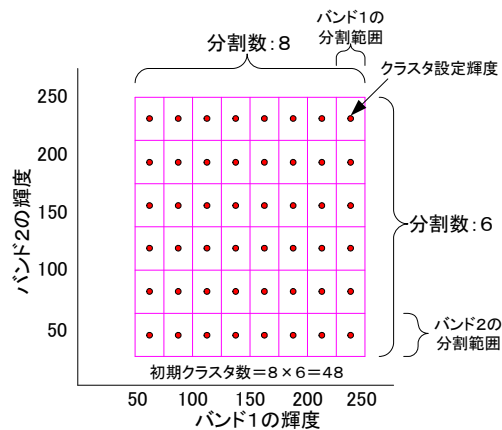


図2 初期クラスタ設定概念図
(2次元の場合)

各バンドの初期クラスタの設定輝度は、各バンドの輝度の最小値と最大値、そして分割数を基に、次式により設定した。

$$L = (\text{最大輝度} - \text{最小輝度} + 1) / N \quad (1)$$

$$Di = L \times i + L / 2 + \text{最小輝度} \quad (2)$$

ここで、 Di : クラス番号 i におけるクラスタ設定輝度、 L : 分割範囲、 N : 分割数、 i : 各バンドのクラス番号で $0 \sim (N-1)$ の範囲で設定。

本研究では初期クラスタをデータ全体に網羅的に設定することから、分類結果に対するクラスタ分離アルゴリズムの違いによる影響を排除するため、クラスタ分離処理は行わないことにした。

(2) 初期パラメータ設定

ISODATA法において初期条件を変化させた場合、分類クラス数及び分類精度がどのように変化するかを調べた。評価する分類クラスは実際の土地被覆状況との違いについて評価しやすいという理由で、水域を選定した。

調査は次の3パターンについて実施した。

ケース1 : 最小サイズを変化させた場合

ケース2 : 限界距離を変化させた場合

ケース3 : 初期クラスタ数を変化させた場合

ここで、限界距離はユークリッド距離を用いた。

4 検討結果

(1) 最小サイズによる影響

限界距離を10と13、各バンドの区分数を10とした場合の最小サイズによる分類結果(水域ピクセル数)への影響を図3に示す。

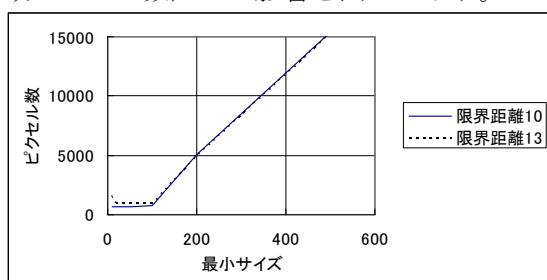


図3 最小サイズによる影響

(2) 限界距離による影響

最小サイズを10、各バンドの区分数を10とした場合、限界距離の違いによる分類結果(水域ピクセル数)への影響を図4に示す。

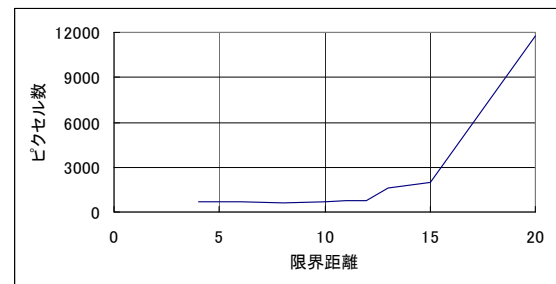


図4 限界距離による影響

(3) 初期クラスタ数による影響

初期クラスタ数をできるだけ細かく(数多く)設定する方が分類精度は高くなると考え、最小サイズを10、限界距離を10とした場合の初期クラスタ数による分類結果(水域ピクセル数)への影響を図5に示す。

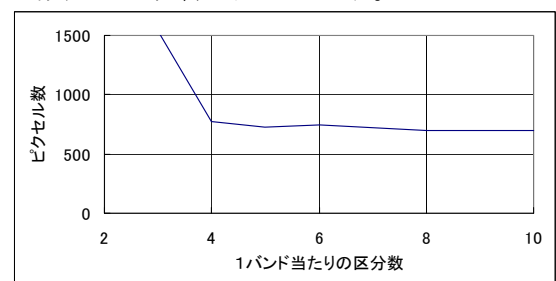


図5 初期クラスタ数による影響

5 まとめ

本研究で得られた結果は次のとおり。

(1) 最小サイズは、これが大きい場合は分類精度が悪くなるが、小さいほど分類精度が高くなり、ある一定の値以下では分類精度は限界距離等の他の影響により決まることが示された。このため、精度の良い分類を行うには、最小サイズはできるだけ小さい値を採用することが望ましい。

(2) 初期クラスタ数による分類精度への影響は、他の要因よりも強くないという結果を得た。このため、初期クラスタ数を数多く設定せずとも十分な分類精度を得ることが可能であることが分かった。

(3) 最小サイズを小さい値とした場合、分類精度に最も影響を及ぼすのは限界距離である。このため、ISODATA法の初期設定に当たっては、限界距離の設定に特に注意を払い、複数の限界距離による分類作業の実施と確認が望ましいと考える。

6 おわりに

本研究では初期クラスタの設定は各バンドの最小値と最大値の範囲内で等分としたが、密度分布等の他の設定方法についても検討が必要と考える。今後は実利用に向け、本研究を発展させるものとする。