スケーリングを目的とするミクセル推定における カテゴリ分解法と回帰推定法の比較

Comparison of Category Decomposition and Multiple Regression in the Mixture Analysis for the Scaling Techniques

竹内章司・稲永麻子

Shoji TAKEUCHI and Asako INANAGA

Abstract: The authors tested the applicability of category decomposition method based on the linear mixture model for the fusion of multiple-resolution satellite data such as Landsat-TM and NOAA-AVHRR. The goal of the application of this method is to estimate the mixing ratio of different categories within one pixel of the lower-resolution data using the classification result of the higher-resolution data, which is considered to be useful for the extrapolation of the information from the higher-resolution data over the wider coverage of the lower-resolution data. The authors tested the estimation accuracy by two kinds of decomposition methods, the maximum likelihood estimation and the minimum distance estimation and also by the multiple regression method. The experimental results showed that the most adequate estimation was obtained by the category decomposition based on the minimum distance estimation.

1. はじめに

現在の地球観測衛星の1つの動向として、ADEOSをはじめ、インドのIRSやSPOTの次のシリーズ(4号)など、分解能の異なる複数の光学センサーを同時に搭載する衛星が増えてくる傾向にある。このような状況から、分解能の異なるデータを組み合わせて利用することは、これからのリモートセンシングの実利用における1つの重要な方向であり、そのための有効な解析方法の開発が求められている。このような組み合わせ利用の1つの有効な方法として、スケーリングと呼ばれる方法がある。これは、高分解能センサーから得られる土地被覆の混合比の情報を、低分解能センサーデータの各画素に適用することにより、低分解能データがカバーするより広い地域での画素内の土地被覆混合比を求めようとするものである。このような解析方法においては、低分解能データの画素を高分解能データの画素によるミクセルとみなし、このミクセル

の土地被覆混合比を如何に精度良く推定するかが問題となる。このための推定方法としては、現在、カテゴリ分解法と回帰推定法が現実的な方法として考えられる。本報告は、この両者の方法の優劣について技術的な検討を加えた結果を報告するものである。

2. 土地被覆混合比の推定方法

2.1 カテゴリ分解法

近年カテゴリ分解法によるミクセル推定の研究が盛んになってきており、その中でも、最尤推定法の有効性が報告されている¹⁾。ここでは、この最尤推定法(ML)とそれを簡略化した最短距離法(MD)を、カテゴリ分解法として用いた。

カテゴリ分解法では、ミクセルの観測値は以下のような線形混合モデルにより得られると仮定する。

$$I_i = \sum_{k=1}^N B_k A_{ki}, \quad \sum_{k=1}^N B_k = 1 \quad (2.1)$$

ここで、 I_i はミクセルのバンド*i*の観測値、 B_k はカテゴリ*k*のミクセル内の占有比率、 A_{ki} はバンド*i*におけるカテゴリ*k*の代表値をあらわす。

最尤推定法(ML)では、 I_i を以下のような平均値 A_i と分散 σ_i^2 を持つ正規分布をなすものと仮定する。

(財)リモート・センシング技術センター

Remote Sensing Technology Center of Japan

〒106 港区六本木1-9-9六本木ファーストビル8F

「写真測量とリモートセンシング」VOL. 37, NO. 1, 1998

$$A_i = \sum_{k=1}^N B_k A_{ki}^*, \quad \sigma_i^2 = \sum_{k=1}^N B_k^2 \sigma_{ki}^2 \quad (2.2)$$

ここで、 A_{ki}^* と σ_{ki}^2 は、バンド i におけるカテゴリ k の平均値と分散をそれぞれあらわす。

占有比率 B_k の最適解は、以下に示す尤度 $Q(I)$ を最小化することにより求められる¹⁾。

$$Q(I) = - \sum_{i=1}^M \ln P(I_i) \quad (2.3)$$

$$P(I_i) = \frac{\exp\{-(I_i - A_i)^2 / 2\sigma_i^2\}}{\sigma_i(2\pi)^{1/2}} \quad (2.4)$$

なお、(2.2) の分散に関する式は、観測データのバンド間には相関がないことを前提としている。そのため、上記の方法を適用するためには、オリジナルデータを主成分分析等により無相関のデータに変換する必要がある。

一方、上記の方法を簡略化した方法として、分散を無視して平均値のみを考慮する方法が考えられる。この方法をここでは最短距離法 (MD) と呼ぶことにする。最短距離法においては、 B_k の最適解は、以下に示すユークリッド距離 D^2 を最小化することによって求められる。

$$D^2 = \sum_{i=1}^M (I_i - A_i)^2 \quad (2.5)$$

上記の2つのカテゴリ分解法で B_k の最適解を求めるために、各カテゴリの占有率を10%刻みで変化させながら起こりうるすべて占有率の組み合わせを発生させ、その中で、(2.3) あるいは (2.5) を最小にする占有率の組み合わせを最適解として求めた。

2.2 回帰推定法

ミクセルの推定方法として、物理モデルを使わずに統計的に直接土地被覆混合比を推定することもこれまで多く行われている^{2),3)}。この方法は、占有率とミクセルの観測値との回帰分析により、占有率を直接推定する方法であり、回帰推定法 (MR) と呼ぶことにする。 B_k の推定値は以下のような観測値の重回帰式で与えられる。

$$B_k = \sum_{i=1}^M a_{ki} I_i + C_k \quad (2.6)$$

ここで、 a_{ki} はバンド i におけるカテゴリ k のための回帰係数であり、 C_k はカテゴリ k の定数である。

3. ミクセル推定実験方法および実験結果

3.1 テスト用データセット

2種類のデータセットを作成して、ミクセル推定の実験を行なった。テストサイトは、東京を含む関東平野である。1つは、ランドサット TM データを高分解能データ (以下 HR データ) とし、TM データの平均化により擬似的な低分解能データ (以下 LR データ) を生成したシミュレーションデータセットである。HR データの分解能は、TM と同じ28.5m, LR データの分解能は、ADEOS の OCTS を想定して、684m とした。また、ADEOS の AVNIR と OCTS の組み合わせを想定して、バンドとしては TM データのバンド1から4のみを用いた。

2つ目のデータセットは、HR データとして TM データ、LR データとして TM データと同じ日に観測された NOAA/AVHRR データを用いた実データセットである。この場合の LR データの分解能は1.1 Km である。また、AVHRR データのバンドとしては、TM のバンド1～4が可視から近赤外までをカバーするため、同じ領域のバンド1 (可視) とバンド2 (近赤外) のみを用いた。したがって、実データセットの場合には、HR データと LR データとのバンドが互いに対応していないことが、シミュレーションデータセットとは異なる点である。また、LR データ (AVHRR データ) を HR データ (TM データ) に幾何学的に重ね合わせることも必要であり、ここでは、両画像の対応点を目視により選定した後、AVHRR データを TM の分解能に合わせてリサンプリングすることにより、両画像の重ね合わせを行なった。

3.2 TM データの土地被覆分類

土地被覆混合比を求めるための基礎データを得るため、HR データ (TM データ) に対して、教師有り最尤法分類により土地被覆分類を行なった。分類カテゴリとしては、都市 (Urban)、裸地 (Bare Soil/Baresl, 作付けされていない農地も含む)、草地 (Grass, 作付けされている農地も含む)、森林 (Forest)、および水域1、2 (Water1, Water2, 1は海水、2は河川水に対応) の計6種類を設定した。なお、分類に用いるデータとしては、オリジナルの TM データ (バンド1～4) に対して直交化のための主成分分析を行なって得られ

た第1, 第2主成分を用いた。得られた土地被覆分類画像に対して, LR データの分解能に合わせたメッシュ化を行い, LR データの画素毎に1画素内の各土地被覆カテゴリの占有率を算出した。

3.3 カテゴリ代表値の算出方法

カテゴリ分解法においては, カテゴリの代表値の決定が推定精度に大きく影響する。このため, シミュレーションデータセットを用いて, 2種類の代表値の決定方法を試み, その精度を比較した。1つの方法は, TM データの分類に用いたトレーニングデータの平均値と分散をそのままカテゴリの代表値として用いる方法であり, もう1つは, 分類画像の全域から, 各カテゴリに対応する TM 画素値を再抽出したデータから計算される平均値と分散を代表値として用いる方法である。

上記の2種類の代表値を用いて, ミクセル推定実験を行なった結果, ML および MD による推定方法ともに, 第2の方法による代表値の方が推定精度が高いことが確認された。これは, 第2の方法による代表値の方が, テストサイトのカテゴリのスペクトルをより良く代表するためと考えられる。以後, カテゴリ分解法の場合には, 第2の方法による代表値を用いた。

3.4 シミュレーションデータセットによる結果

図1には, シミュレーションデータセットによる, カテゴリ分解法 (ML, MD) と回帰推定法 (MR) による占有率推定誤差を示す。カテゴリ k の占有率推定誤差 (E_k) および全カテゴリの平均推定誤差 (E_t) は以下の計算式で求めた。

$$E_k = \left\{ \frac{1}{N_p} \sum_{n=1}^{N_p} (B_{k,n} - B_{k,n}^*)^2 \right\}^{1/2}$$

$$E_t = \frac{1}{N_p} \sum_{n=1}^{N_p} \left\{ \frac{1}{N_k} \sum_{k=1}^{N_k} (B_{k,n} - B_{k,n}^*)^2 \right\}^{1/2} \quad (3.1)$$

ここで, $B_{k,n}$ は画素 n のカテゴリ k の占有率の真値, $B_{k,n}^*$ はその推定値, N_p は対象地域の総画素数, N_k は, カテゴリの総数である。また, 水域と都市や森林との混同がかなり見られたため, 比較のため, 水域を除いた場合の結果も同時に示した。

図1の結果を見ると, 平均的には, 2つのケースともに最尤推定法 (ML) が誤差がもっとも大きな結果となっていることがわかる。MD と MR では, ケースにより誤差の大きさが異なり, どちらも優劣がつかない

図1 シミュレーションデータセットにおける土地被覆占有率の推定誤差。
ML: カテゴリ分解法 (最尤推定法)
MD: カテゴリ分解法 (最短距離法)
MR: 回帰推定法
上: 水域を含む場合, 下: 水域を除いた場合

い結果である。ML が最も悪い結果となったことは, これまでの研究とは矛盾する結果と言える。この原因としては, カテゴリ代表値の中で分散の推定が良くなかったことが考えられる。すなわち, 3.3に述べたように, 分類画像全域から画素値を集めるため, 1つのカテゴリに対するスペクトルのバリエーションが大きくなり, 正規分布仮定を満足しなくなったためと推測される。この問題を回避するには, より均一な画素値のみを集めることが考えられるが, そうすると, 一方ではカテゴリの数を増やす必要があり, このカテゴリ数の増大が推定誤差を大きくすることも考慮する必要がある。

3.5 実データセットによる結果

実データセットにおいては、カテゴリ分解法としてMDのみを用いた。この理由は、3.4でMLが最も悪い結果となったこと、および、HRデータとLRデータとの間でバンドが対応しないため、代表値の推定はLRデータのみから行なう必要があるためである。すなわち、LRデータにおいてはほとんどの画素が事実上のミクセル状態となり、その様な状態の画素値から分散を正しく推定することは非常に困難と考えられるからである。

MDにおいては、代表値としてバンド毎の平均値のみが使用される。実データセットにおいては、HRデータの分類画像とLRデータとを用いて、特定のカテゴ

リの占有率が90%を越えるLRデータの画素（擬似ピュアピクセル）をサーチし、それらの画素値の平均値を代表値として用いた。しかし、カテゴリによっては、上記の条件を満足するピュアピクセルが存在しない場合が起こり得る。本データセットでは、草地と水域2（河川水）がそうであった。

このような場合には、以下の式で求められる残差の回帰分析により代表値を求めた。

$$R_{i,n}=I_i-(\sum_{m=1}^l A_{mi}B_m) \tag{3.2}$$

ここで、 $R_{i,n}$ は、バンド*i*におけるカテゴリ*n*の残差を表し、 m はピュアピクセルによる推定が可能なカテゴリで、 A_{mi} 、 B_m はカテゴリ*m*のバンド*i*の代表値と占有率を示す。上記の $R_{i,n}$ と同じ画素内のカテゴリ*n*の占有率 B_n との間で回帰分析を行い、 B_n が100%になる $R_{i,n}$ をカテゴリ*n*のバンド*i*における代表値とした。この方法により、草地の代表値を求めることは出来たが、水域2については、LRデータの画素内の占有率が小さすぎて回帰分析が実行できず、最終的に水域2を無視して推定実験を行なった。

図2には、実データセットを用いたMDとMRによる推定誤差の結果を示す。図1と同様、水域を含むケースと除いたケースの両方の結果を示す。図2では、2つのケースともに、平均的にはMRの方が若干誤差が少ない結果となっている。

3.6 占有率分布パタンの比較

以上の実験結果を見る限り、推定誤差だけを考えると、回帰推定法（MR）の方がカテゴリ分解法（ML，MD）より若干優れているという結果になる。しかし、スケーリングにおいては、土地被覆カテゴリの占有率の分布パターンも重要な情報である。図3には、シミュレーションデータセットにおける都市と裸地の占有率の推定パターン、図4には、実データセットにおける都市、裸地および森林の占有率推定パターンをそれぞれ示す。いずれの図も、青から緑，黄，赤，ピンクと変化するにつれて占有率が大きくなることを示している。

両方の図から明白なことは、MRによる推定パターンは、ML，MDによるパターンと比べて極端にばやけていることである。この結果は、物理モデルを用いずに統計的観点のみから推定を行なう回帰推定法の限界を示していると考えられる。一方カテゴリ分解法では、いずれのケースでも真の分布とパターンとしては良く一致

図2 実データセットにおける土地被覆占有率の推定誤差。
MD：カテゴリ分解法（最短距離法）
MR：回帰推定法
上：水域を含む場合，下：水域を除いた場合

図3 シミュレーションデータセットにおける土地被覆占有率の推定パターン（上段：都市，下段：裸地）
（左端がTM データによる真値を表し，中央左はML，中央右はMD，右端はMR による推定パターンをそれぞれ表す）

図4 実データセットにおける土地被覆占有率の推定パターン（上段：都市，中段：裸地，下段：森林）
（左がTM データによる真値，中央がMD による推定パターン，右がMR による推定パターンを表す）

する推定パターンが得られていることがわかる。数値上の誤差だけを見る限り回帰推定法の方が良い結果ではあるが、このようなパタンの推定結果を見る限りは、やはりカテゴリ分解法の方が優れていると結論づけることが出来る。カテゴリ分解法の中のMLとMDとの比較では、MDの方がわずかにノイズになる傾向があるものの推定パターン自体には優劣は認められず、数値上の誤差が少ないMDの方が本データセットの結果を見る限り優れていると言うことが出来る。

3.7 占有率変化幅の推定誤差への影響

カテゴリ分解法においては、反復計算により占有率の最適解を求めており、その際、占有率の変化幅を10%とした。この値は一見かなり大きいように見えるので、この変化幅の影響も考慮する必要がある。まず、この変化幅の大きさを占有率の量子化誤差と見なすと、量子化誤差の標準偏差は $\sqrt{D^2/12}$ で与えられるので(D は量子化の幅)、10%の変化幅が与える量子化誤差は2.89%となる。現実の推定誤差値は10%~20%程度であり、仮に実際の推定誤差を15%とし、両者の誤差を独立と仮定すると、量子化誤差による推定誤差の増大分は0.3%となり、元の推定誤差と比較しても事実上無視出来る値である。実際、占有率の変化幅を5%に減らして実験して見たが、推定誤差の減少は平均でわずか0.4%にとどまった。これらの結果より、10%という占有率変化幅の大きさが推定誤差に与える影響は非常に小さいものと考えられる。もちろん、変化幅は小さい方が望ましいので、反復計算量との兼ね合いからさらに小さくすることも可能であろう。

4. むすび

スケーリングを目的としたミクセル推定において、カテゴリ分解法と回帰推定法によるカテゴリ占有率推

定誤差および占有率の推定パタンの比較を行い、両者の優劣を検討した。その結果、数値的な誤差では回帰推定法の方が若干良い結果を示したものの、占有率パタンの推定では、相対的にカテゴリ分解法の方がかなり良い結果が得られ、総合的にカテゴリ分解法の優位性を確認することが出来た。カテゴリ分解法の中では、従来提案されている最尤推定法とそれを簡略化した最短距離法を試みたが、本実験では最尤推定法の優位性は認められなかった。この原因としては、テストサイトがかなり広く、1つのカテゴリに属するスペクトルのバリエーションが大きくなりすぎたことが考えられるが、これについてはさらに検討が必要であろう。

また、本検討においては、高分解能データがカバーする地域のみで推定誤差等の検討を行なった。しかし、実際にスケーリングを適用する場合には、高分解能データの範囲の外での推定精度が大きな問題となる。これには、大気の影響等による低分解能データのスペクトルの場所的な変動が大きく影響する。スケーリングを実用的に用いるには、このような問題を含めて今後さらに研究を継続する必要があると言えよう。

(受付日1997.10.27, 受理日1998.1.12)

参考文献

- 1) 松本, 藤九, 土屋, 新井, 最尤推定法に基づくカテゴリ分解, 写真測量とりモートセンシング, Vol. 30, No. 2, pp.25-34, 1991年5月。
- 2) 竹内, 長, 吉村, 分解能の異なるデータの組み合わせによる土地被覆状況の解析, 日本写真測量学会平成2年度秋季学術講演会発表論文集, pp.105-108, 1990年11月。
- 3) Y. Suga, S. Takeuchi, H. Kimura and A. Inanaga, Environmental Monitoring of Land and Sea Surface Using Multisensors, Adv. Space Res., Vol. 17, No. 1 pp.(1)97-(1)106, 1996.