

解説

リモートセンシング・データのデジタル解析システム：LARSYS*

辺見隆三郎**

1. はじめに

資源衛星や気象衛星などに代表されるリモートセンシング技術とその観測された地球表面のデータ解析を行う画像処理技術は、資源探査・農林漁業・都市環境・地図作成・気象問題など多くの分野で幅広く注目を集めてきている。リモートセンシング技術への関心の急速な高まりにもかかわらず、反面、観測センサーやデータ収集のコストの問題、ハードウェア・ソフトウェア両方を含めたデータ解析システムの普及度の問題およびコンピュータによるデジタル画像解析に慣れていないことなど種々の問題が提起されている。

ここではリモートセンシング技術者に手軽に利用可能なデジタル画像解析システムとして米国のパデュー大学で開発された LARSYS についてその機能とデジタル解析の概要について紹介する。

2. システムの概要

2.1 LARSYS の沿革

LARSYS は米国航空宇宙局 (NASA) と農務省の援助のもとにインディアナ州のパデュー大学の研究機関：LARS***において 1964 年頃から研究が進められ開発されたリモートセンシング・データのデジタル解析の汎用システムである。LARS における初期の研究は農業の分野のリモートセンシング技術の開発であり 1967 年に一応その成果をみた。更にその後この技術は地質学・地理学・水理学などにも幅広く利用されるようになり 1972 年 4 月に LARSYS Version 3 が完成している。現在パデュー大学では IBM S/360 モデ

ル 67 をホスト・マシンとして LARSYS によるタイム・シェアリング・サービスを NASA を中心として全米各地に提供しており LANDSAT および航空機 MSS**** などのリモートセンシング・データの解析にあたっている。

日本アイ・ビー・エムのデータ・センター部門では 1976 年にこの LARSYS Version 3 を導入し更に機能向上を図りデータ・センターの全国的なネットワークを通じてリモートセンシング・データのデジタル解析処理サービスを行っている。

2.2 システムの概要

現在米国では NASA を中心にリモートセンシング・データの解析システムがいくつか稼働しているが LARSYS はそれらの中でもデジタル解析の入門者用として非常に使い易いシステムであり次のような特徴を持つ。

- 入力画像データ
- 統計的パターン認識
- 20 数個のプロセッシング・ファンクション
- 簡単なコントロール・データ
- 計算途中結果の記憶

2.2.1 入力画像データ

システムの主要入力データは、デジタル化されたマルチスペクトル画像である。すなわち、飛行機や人工衛星に搭載されたマルチスペクトル・スキャーやマルチバンド・カメラにより観測し得られる地表面の画像データである。図-1 (次頁参照) は、飛行機に搭載された MSS によるデータ収集の例である。このようにして得られた地表面の情報は通常アナログ量であるので A-D 変換された後で CCT*****として磁気テープに記憶され画像処理が行われる。LARSYS ではこの CCT のことを特にマルチスペクトル・イメージ・ストレージ・テープ***** (MIS Tape) と呼び特有のフォーマットを持った磁気テープとして処理される。

MIS テープ中の情報は地表面の矩形の一区画をラン・データ (run data) という単位で図-2 (次頁参照)

* Digital Analysis System of Remote Sensing Data: LARSYS by Ryuzaburo HENMI (DC Application Support, Data Center Service, IBM Japan).

** 日本アイ・ビー・エム(株)データ・センター・サービス DC アプリケーション・サポート

*** Laboratory for Application of Remote Sensing

**** Multi Spectral Scanner

***** Computer Compatible Tape

***** Multi-Spectral Image Storage Tape

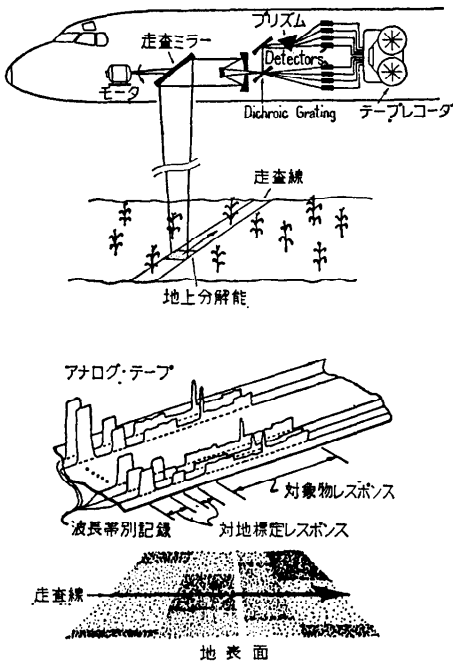


図-1 航空機 MSS によるリモートセンシング・データの収集

のようにデータ・収集の際のスキャン・ライン、スキャン・ライン上の画素に対応するカラムおよび観測センサーのスペクトル帯域に対応するチャンネルという3つの基準でマトリックスの形で整理されている。そして1つのスキャン・ラインとカラムにより定義されるサンプル(画素)が電磁波の反射率に応じて0から255のスペクトル値をスペクトル帯域の数だけ有する。

2.2.2 統計的パターン認識

リモートセンシングで観測したデータの解析の主な

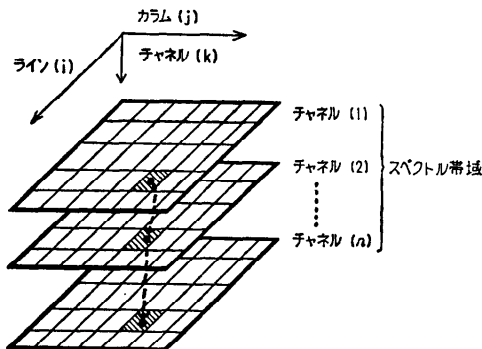
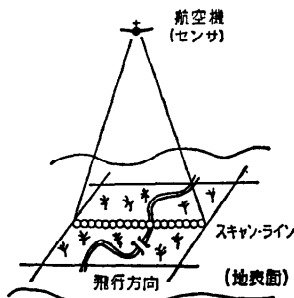


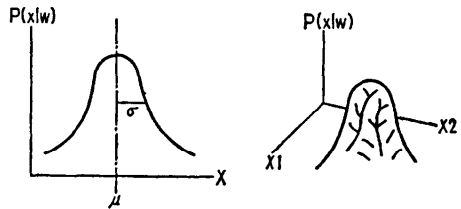
図-2 画像データの構成

目的は地表面の種々の状態の分類マップを作成することにある。例えば、土地利用・植性分布・農作物の生育状況・水質マップなどが挙げられる。LARSYSによるデジタル解析の主目的もこの種々の分類マップを作成することであるがシステムの中で採用している解析手法は、農業の分野におけるリモートセンシング・データの統計的パターン認識を中心に研究・開発された。

“リモートセンシングにより収集されるマルチスペクトル画像データは基本的にはガウス分布(正規分布)をなす。”という仮定のもとに統計的パターン認識のアルゴリズムを用いてデジタル解析を行う。ガウス分布は平均値(μ)と標準偏差(σ)により図-3のようにその確率密度関数: $p(x|\omega)$ が表現される。ガウス分布を仮定する大きな理由は、

- 地表面の現象は基本的にはガウス分布をなすであろう。
- コンピュータ処理の際にガウス分布による統計的パターン認識の手法を採用すれば処理コストに見合った結果が得られるであろう。

という2つの想定に基づいている。図-4(次頁参照)はパターン認識のブロック・ダイアグラムである。デ



$$P(x|\omega) = \frac{1}{(2\pi)^{1/2} \sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

図-3 ガウス分布

ータ収集の際のスペクトル帯域の数 (n) により n 次元のパターン空間が構成される。

2.2.3 プロセッシング・ファンクション

LARSYS は統計的パターン認識を中心にマルチスペクトル画像データの解析を行うが機能という観点から眺めると 20 数個のプロセッシング・ファンクションより構成されている。図-5 はこのプロセッシング・ファンクションと主なデータの流れを示している。各プロセッシング・ファンクションは大きく次の6つの機能に大別され、各ファンクションは単独でもあるいは連続してでもデータ処理が行える。

- ① ユーティリティ機能
- ② 画像データ評価機能
- ③ フィールド選択機能
- ④ 統計計算機能
- ⑤ クラス分類機能
- ⑥ 分析・評価機能

2.2.4 コントロール・データ

解析に用いるプロセッシング・ファンクションの選択、解析に用いる種々のパラメータ、必要な計算結果と出力フォーマットの指定などに使用されるコントロール・データの仕様の単純化にはかなりの考慮が払われている。データの記述は自由形式で、ユーザは希望する解析の手段、処理に用いる数値データなどをキーワ

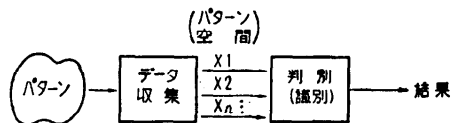


図-4 パターン認識のブロック・ダイアグラム

ド⊕コントロール・パラメータの形として簡単な英語的表現でコントロール・カードに記述したりあるいは直接ターミナルから LARSYS の中に入力可能である。この英語的表現も文法上の規則は非常に単純でしかもほとんどの場合キーワード、コントロール・パラメータには標準的な省略時解釈が準備されている。従って初心者でも簡単にデータ作成が可能でデータ・エラーも軽減されるように配慮されている。次にコントロール・データ作成の例を示す。

```

-COMMENT GRAY LEVEL MAP
*PICTUREPRINT
DISPLAY RUN (77080001), LINE (25, 400, 2), COL
(80, 550, 3)
PRINT HIST
CHANNELS 3, 4 7, 10
END
    
```

この例では、プロセッシング・ファンクション：PI

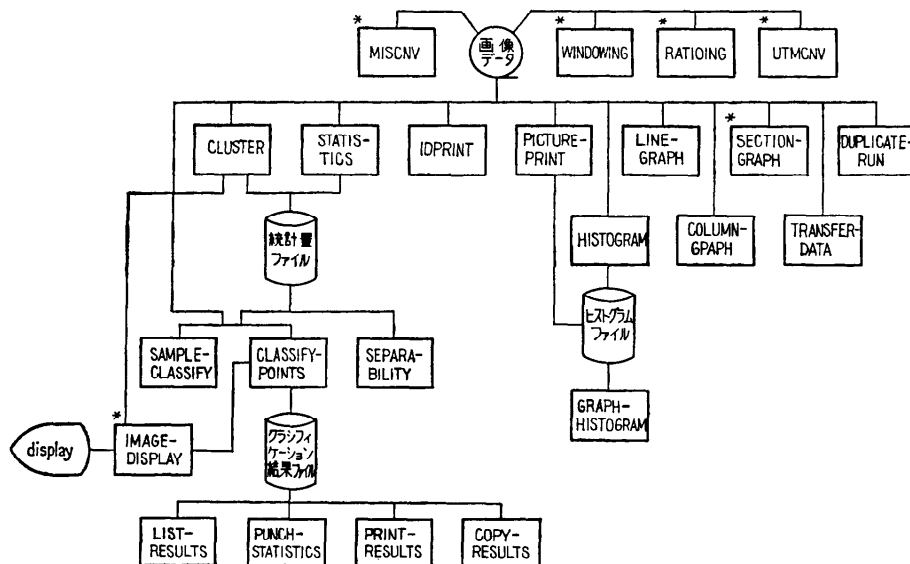


図-5 LARSYS のシステム構成図

(* 印は、日本 IBM データ・センターで追加した機能)

CTURE-PRINT によりラン番号: 77080001 のマルチスペクトル画像データについてライン: 25 から 400 まで 1 個おきにカラム: 80 から 550 まで 2 個おきにスペクトル帯域 3, 4, 7, 10 の 4 つのチャンネルについてデータの濃淡マップ*を作成する。同時に 'PRINT HIST' により各チャンネルごとのヒストグラムをグラフの形でプリントする。濃淡マップのレベル数は省略時解釈として 10 が適用され等頻度分割でマップが作成される。

2.2.5 計算結果の記憶

各プロセッシング・ファンクションの計算結果は通常磁気ディスク上に記憶され次の処理ステップの入力データとなる。計算結果を記憶する主要なファイルは、図-5 の中に示されている次の 3 つのファイルである。

① ヒストグラム・ファイル

マルチスペクトル画像データの各スペクトル帯域ごとのヒストグラム (頻度分布) を記憶する。

② 統計量ファイル

ユーザが選定したトレーニング・フィールド、トレーニング・クラスに関するガウス統計量 (平均値ベクトル・標準偏差ベクトル・共分散マトリックス) を記憶する。

③ クラシフィケーション結果ファイル

最尤法により各画素ごとに判別された結果のパターン・クラスの番号とその判別度を記憶する。

この 3 つのファイルの他にユーザが選定したトレーニング・フィールドやテスト・フィールドの領域を記憶するフィールド記述ファイルやメッセージ・ファイルを持つ。これらのファイルの内容は後日の処理のために必要に応じてカードにパンチ・アウトすることも可能である。

3. LARSYS を利用したデジタル解析の方法

3.1 解析の流れ

図-6 により LARSYS の一般的な解析の手順について説明する。観測された画像データは MIS テープの形で入力されるので初めにそのデータの質の検討・評価を行う。リモートセンシング・データにはセンサー・システムの特長や観測時の状況に起因した種々の

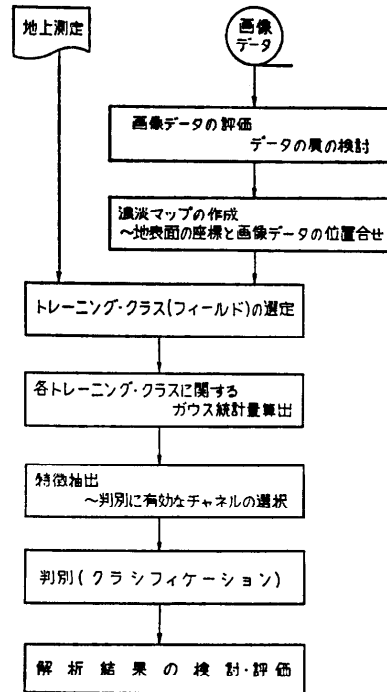


図-6 LARSYS によるデジタル解析の手順

雑音や歪みが含まれる。また遠隔から計測するために地表面とセンサーの間の空気層による光の散乱・吸収の変化も影響するので解析に入る前に十分にデータの質の検討を行う必要がある。場合によっては個々に補正プログラムを作成して CCT のデータを補正する必要も生じる。次に画像データから濃淡マップを作成する。濃淡マップは解析対象となる地表面と画像データの位置合せという重要な働きをなすので対象地域の特徴を示すと思われるスペクトル帯域について作成する。デジタル解析の正確度・信頼性を高めるために一般に対象地域の地上測定**を行うからこの地上測定の資料をもとに判別しようとするパターン・クラスを代表していると思われるトレーニング・フィールドを選定し各フィールド (領域) に関するガウス統計量を画像データの中から算出する。このガウス統計量から各パターン・クラスの判別関数***を設定し対象地域の画像データを 1 画素ごとにどのパターン・クラスに類似性が強いかを最尤法による判別****により判定する。最後にその判別結果の妥当性・効率などの評価・検討を行う。分類マップ*****を作成したりするのも結果の評価方法の 1 つである。MSS データは 4~24 のスペクトル帯域を持つので判別の際にすべてのスペク

* Gray Level Map

** Ground Truth

*** Discriminant Function

**** Maximum Likelihood Classification

***** Classification Map

トル帯域を利用してもかまわないが計算の効率などの観点から好ましくない。対象とするパターン・クラスの判別に有効なスペクトル帯域の組合せをあらかじめ自動的に抽出して解析に使用するスペクトル帯域の数を減らす方法として LARSYS では特徴抽出*という機能を持つ。図-7 は特徴抽出の機能を持つパターン認識のブロック・ダイアグラムである。図-4 と比較するとその違いがわかる。デジタル解析のコストは他の条件が同じ場合使用するスペクトル帯域の数にほぼ比例するので特徴抽出の機能は非常に有効である。

3.2 判別と特徴抽出

システムの大きな機能である統計的パターン認識の中から判別と特徴抽出について LARSYS のコントロール・データを例にしながら紹介する。

3.2.1 教師あり判別**と教師なし判別***

リモートセンシング・データの判別方法は

- ① 教師あり判別
- ② 教師なし判別

に大別される。一般に教師なし判別は、都市部のように対象とするパターン・クラスが明確に定義不可能な場合や湿原地帯など地上測定が困難な地域の解析に有効である。従ってリモートセンシングの分野では地上測定の情報の有無により両者を大別しているが LARSYS の解析の流れの中ではこの2つの解析方法は完全に別個のものではなくクラスタ解析****と地上測定の情報を組み合わせて最適なトレーニング・クラスすなわち可能な限りそのデータがガウス分布に近く設定できるようにプログラムが構成されている。クラスタ解析と地上測定の情報を組合せる理由としてリモートセンシングによるデータの収集と地上測定の時期のずれおよび地上測定誤差などが挙げられる。また地上測定にはかなりの期間と費用を要するので地上測定の情報が不十分でもデータ解析が可能な教師なし判別の手法は有効である。

LANDSAT MSSデータを用いて近畿地方の地表面分類マップ*****を作成する作業を考える。この時の解析対象となるパターン・クラスには、都市・農地・森林・草地・裸地・河川・海……が挙げられるであろう。

解析の第1の方法は対象地域に関する地上測定の情

* Feature Selection
 ** Supervised Classification
 *** Unsupervised Classification
 **** LARSYS では、教師なし分類の手法として“BALLとHALLの ISODATA”によるクラスタ解析を採用している。
 ***** Land Cover Map
 pool

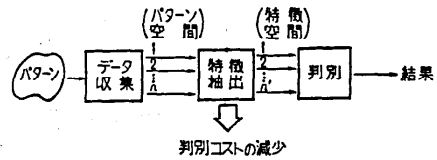


図-7 特徴抽出の機能を持つパターン認識のブロック・ダイアグラム

報が得られた場合には図-8 (次頁参照) のように地上測定の情報に基づいて対象となるパターン・クラスに該当する領域をトレーニング・フィールドとして選定しその領域について画像データの中から平均値ベクトル (M_i) を共分散マトリックス (K_i) を算出する。この M_i , K_i から各パターン・クラス毎の判別関数を設定し解析の対象となる全領域の未知データについて最尤法による判別を行い最終の分類マップを作成する (最尤法による判別については次節で述べる)。各トレーニング・フィールドのガウス統計量を計算する場合には次のようなコントロール・データを作成する。

```
-COMMENT CALCULATES GAUSSIAN
STATISTICS
*STATISTICS
PRINT HIST (F, C), CORRE (F, C), SPECTRAL
(F, C)
CHANNELS 4, 5, 6, 7
DATA
CLASS URBAN
RUN (72102401), LINE (60, 70, 1), COL (102, 131, 2)
CLASS AGRICAL
RUN (72102401), LINE (105, 147, 2), COL (341, 380,
2)
CLASS SEA
RUN (72102401), LINE (403, 472, 3), COL (181, 204,
2)
END
```

この例では、都市・農地・海のパターン・クラスについてその平均値ベクトル・共分散マトリックスを計算すると同時にヒストグラムとスペクトル・プロット図をプリントすることを指定している。また計算に用いるスペクトル帯域は MSS の 4, 5, 6, 7 チャンネルである。

解析の第2の方法は対象地域に関する地上測定の情報が無い場合に、図-9 (次頁参照) のように画像データの中からいくつかの領域を選定しこれらの領域に含まれるデータをひとまとめ*****にしてクラスタ解析

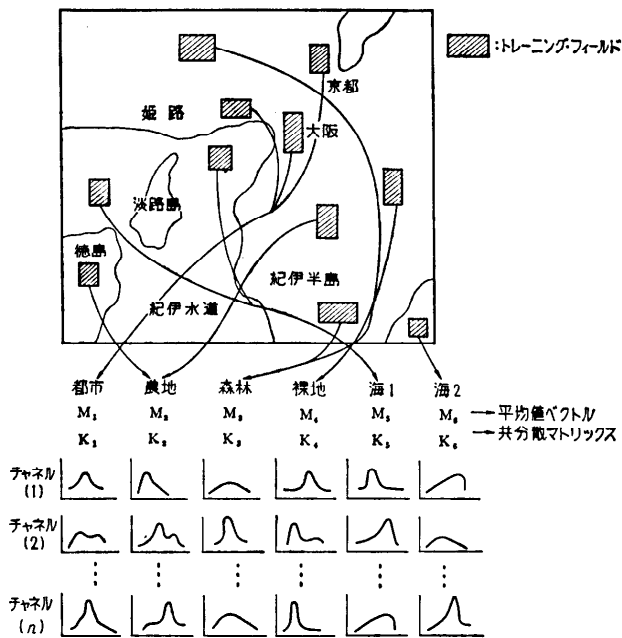


図-8 地上測定の数値に基づく教師あり分類の例

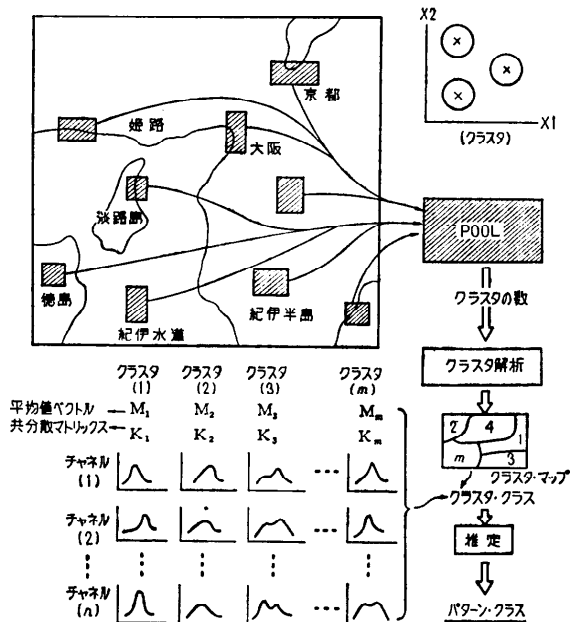


図-9 クラスタ解析による教師なし分類の例

を行い各クラス毎のガウス統計量を算出する。同時に選定した各領域毎のクラス・マップがプリントされるのでこのガウス統計量とクラス・マップから各クラス・クラスが対象とするどのパターン・クラス

に対応するかを推定する。その次に解析の対象となる全領域の未知データについて最尤法による判別を行い最終の分類マップを作成する。クラスタ解析のコントロール・データは次頁のようにして作成する。

```

-COMMENT UNSUPERVISED CLASSIFICATION
*CLUSTER
OPTIONS MAXCLS (13), CONV (99.8)
CHANNELS 4, 5, 6, 7
PRINT HIST
PUNCH STAT
DATA
RUN (72102401), LINE (504, 603, 2), COL (10, 48, 2)
RUN (72102401), LINE (101, 183, 2), COL (443, 497, 3)
RUN (72102401), LINE (349, 403, 2), COL (204, 276, 2)
END
    
```

この例では、クラスタの数を 13 個と定義し計算されたクラス・クラスの統計量をパンチ・アウトすることを指定している。解析に用いる領域は 3 個所で使用するスペクトル帯域は MSS の 4, 5, 6, 7 チャンネルである。

3.2.2 最尤法による判別

解析の対象となる m 個のパターン・クラスに対応するトレーニング・クラスの判別関数： $G_i(X), i=1, 2 \dots m$ を設定する。そして未知データ (X) を各判別関数に代入して図-10 のようにどの判別関数の値を最大にするかを計算してその未知データが属するパターン・クラスを判別する。各判別関数は次式で定義される。

$$G_i(X) = \log P(w_i) - \frac{n}{2} \log(2\pi) - \frac{1}{2} \log |K_i| - \frac{1}{2} (X - M_i)^T K_i^{-1} (X - M_i) \quad (3.1)$$

- 但し $P(w_i)$: パターン・クラス (i) の事前確率
- M_i : パターン・クラス (i) の平均値ベクトル
- K_i : パターン・クラス (i) の共分散マトリックス

最尤法による判別の中で“判別のあいまいさ”という問題が残る。未知データの中で実際にはどのパターン・クラスにも属さないものが現われても最尤法の原則により m 個のパターン・クラスのいずれかに属す

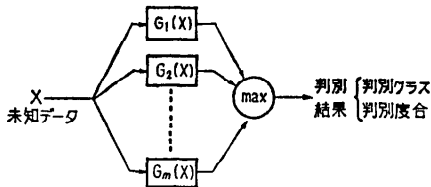


図-10 最尤法による判別

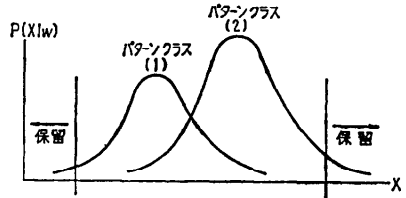


図-11 判別の保留

ると判別され最終の解析結果ではエラーと評価される。このような問題は例えば農地の作付けマップを作成する際のあぜ道・作業小屋・垣根などのデータとしてよく見うけられる。ガウス統計量を算出する際にこのようなパターン・クラスの属するトレーニング・フィールドを設定すれば良いわけであるがデータ点群が少なく設定しにくい場合も非常に多い。このような“判別のあいまいさ”を除去し判別の効率を高める手段として図-11 のようにいずれのクラスに属する確率も非常に低いデータの判別を保留することが考えられる。LARSYS では、スレッシュホールドという考え方を導入し未知データの判別の際に判定されたパターン・クラスの番号と同時にその判別の度合いを (3.2) 式で示されるような 1~234 の値で計算する。そして分類マップを作成したり判別効率を計算する段階でユーザがスレッシュホールド・レベルを 1~234 の範囲で設定し判別の保留を行う。

判別保留度:

$$C_{j-1} > \chi^2 = 2 \left[\log P(w_i) - \frac{n}{2} \log(2\pi) - \frac{1}{2} \log |K_i| - G_i(X) \right] \geq C_j \quad (3.2)$$

$j=1 \sim 234$: 保留度 (スレッシュホールド)

なお LARSYS では (3.1) 式, (3.2) 式に用いられる各パターン・クラスの前確率: $P(w_i)$ をユーザがコントロール・カードで指定できる。最尤法による判別とその結果を利用して分類マップを作成するコントロール・データの例を次に示す。

```

-COMMENT MAXIMUM LIKELIHOOD
CLASSIFICATION
*CLASSIFYPOINTS
RESULTS DISK
CLASSES URBAN (1/5/), AGRICUL (2/4/), GRASS
(3/1, 2/)
CLASSES FOREST (4/6, 7/), WATER (5/3/),
CLOUD (6/8/)
CLASSES SEA (7/9, 10/)
    
```

```

CHANNELS 4, 5, 6, 7
DATA
RUN (72102401), LINE (1, 2340, 1), COL (1, 3240, 1)
END
-COMMENT PRINTS LAND COVER MAP
*PRINTRESULTS
RESULTS DISK
PRINT TEST(P), OUTLINE (TEST)
SYMBOLS U, A, G, F, W, C, S
DATA
TEST 1
RUN (72102401) LINE (430, 502, 4), COL (130, 170, 3)
TEST 2
RUN (72102401), LINE (1021, 1100, 4), COL (381, 423, 4)
END
    
```

この例では既に磁気ディスク上に計算されている各パターン・クラスごとのガウス統計量を利用して最尤法による判別を行っている。

3.2.3 特徴抽出

リモートセンシング・データのデジタル解析において対象とするパターン・クラスを判別するのに冗長な情報を取り除き、パターン空間の次元数を少なくする手法が望まれる。クラスの判別に有効な情報の特徴と呼び元のパターン空間から特徴を抽出して次元数の低い空間すなわち特徴空間を構成することを特徴抽出と呼ぶ。

LARSYS では各トレーニング・クラスのガウス統計量から各クラス間の分離度*という尺度を計算してこの特徴抽出を行い、最尤法による判別に用いる最適なスペクトル帯域の組合せをユーザに提供する。図-12は、パターン・クラス (I) と (J) の分離度の良し悪しを平均値と標準偏差の違いから図示したものである。クラス (I) と (J) の分離度 (ダイバージェンス**) は各クラスの平均値ベクトル (M) と共分散マトリックス (K) から次式により算出される。

$$D(IJ) = \frac{1}{2} t_r [(K_I - K_J)(K_I^{-1} - K_J^{-1})] + \frac{1}{2} t_r [(K_I^{-1} + K_J^{-1})(M_I - M_J)(M_I - M_J)^T] \tag{3.3}$$

実際には LARSYS では $D(IJ)$ を次のような式で変換された分離度を採用している。

$$D_T = 2(1 - e^{-D(IJ)/8}) \tag{3.4}$$

* Separability
 ** Divergence

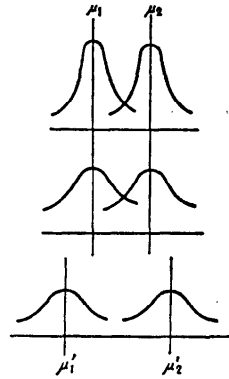


図-12 パターン・クラス間の分離

(3.3), (3.4) 式を用いて各トレーニング・クラス間の分離度をユーザが指定した数のスペクトル帯域のすべての組み合わせについて計算しそのクラス間分離度の平均値の最も大きいものを最適なスペクトル帯域の組合せとして定義する。

4. LARSYS を利用した解析例

図-13 (次頁参照) は近畿地方の LANDSAT MSS データ (1972 年 10 月 24 日観測) を用いて播磨灘・大阪・神戸・紀伊水道の水質調査を試みた例である。濁度に応じて 5 段階の水質分類マップを教師なし分類の手法を用いて作成した。

5. おわりに

特徴抽出および最尤法による判別を中心に LARSYS の統計的パターン認識について紹介した。LARSYS はその他にも領域ごとの判別、チャンネル間比演算、LANDSAT データの UTM 変換などの機能も持っているが紙面の都合上省略した。

リモートセンシング・データのデジタル解析は通常のコンピュータによる計算処理とは異なりユーザの解析目的・対象・手法の相違あるいは地上測定の情報など多種多様な要因が介入するのでその方法は一義的には決まらない。従って、我々はユーザの要求する機能・手法を取り入れてシステムのレベル・アップを図ると共に全国のリモートセンシング技術者にデジタル解析の有用性を理解していただくような活動を続けたい。

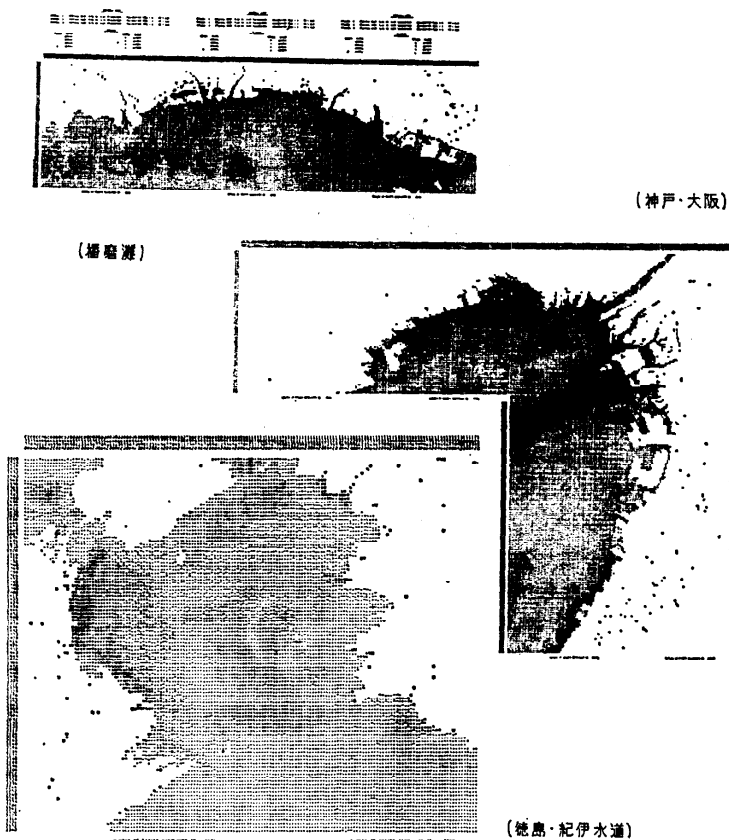


図-13 近畿地方の水質(濁度)分類マップ

参考文献

- 1) "LARSYS Users Manual", "LARSYS System Manual" Laboratory for Application of Remote Sensing. Purdue University.
- 2) "LARSYS 概説書" 日本 IBM, データ・センター・サービス.
- 3) 飯坂: "リモートセンシングにおけるデジタル画像解析" 画像技術 pp. 23~34 (1975. 6).
- 4) D. A. Landgreve: "Machine Processing for Remote Acquired Data" LARS Information Note 031573 Purdue University.
- 5) D. A. Landgreve: "An Evaluation of Machine Processing Techniques of ERTS-1 Data User Applications". LARS Information Note 121273 Purdue University.
- 6) D. A. Landgreve, F. C. Billingsley, J. D. Nichols "Machine Processing Methods for Earth Observational Data". LARS Information Note 100773 Purdue University.
- 7) 飯坂. 辺見: "リモートセンシング・データのデジタル解析例" 東京 Scientific Center Report, N: GE 18-1862, 日本 IBM.

(昭和53年4月19日受付)