

1. 分類の基礎となるデータ

複数の波長により観測される地表物体の分光輝度値を利用する。

図 1 によると、植生(Green vegetation)、土壌(Dark-toned soil)、濁水(Turbid water)を分光特性により分類できそうである。ここでは、植生と土壌の判別を例にする。

- ① 近赤外の波長帯域(バンド 4)によると、
植生：大きな値、
土壌：植生の 1/3 程度の値
- ② 赤の波長帯域(バンド 3)によると、
植生：小さな値、
土壌：植生より大きな値であるが、小さな値
- ③ 緑の波長帯域(バンド 2)によると、
植生：植生の赤よりも大きい小さな値、
土壌：植生と同じ値

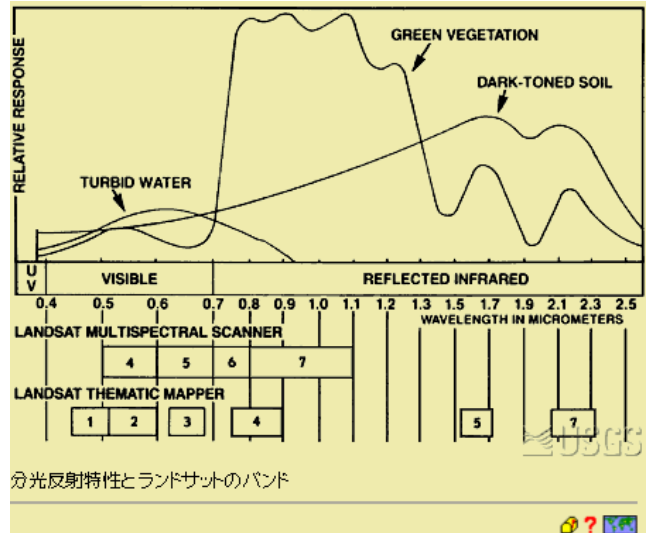


図 1 地表物体の分光反射特性とランドサットのバンド

2. 各地表物体の各波長におけるヒストグラム

対象物体の一つが、各波長における、強度の分布を示す。図 2 の例をとると、数値は、8 ビットのデジタル値である。0 から 255 まで表示可能である。

- ① 近赤外の波長帯域(バンド 7)によると、
植生：(赤の実線)大きな値 →75
土壌：(赤の破線)植生の 1/3 程度の値 →54
- ② 赤の波長帯域(バンド 5)によると、
植生：(緑の実線)小さな値 →18
土壌：(緑の破線)植生より大きな値であるが、小さな値 → 54
- ③ 緑の波長帯域(バンド 4)によると、
植生：(青の実線)植生の赤よりも大きい小さな値 →21
土壌：(青の破線)植生と同じ値 →39

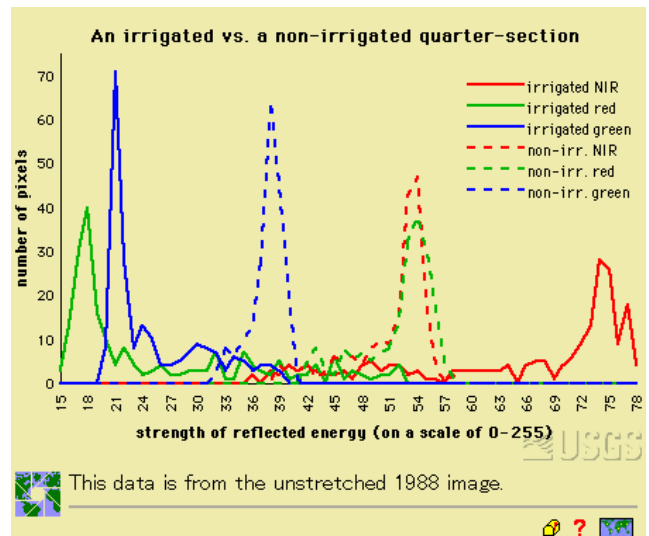


図 2 植生と土壌の各波長帯域におけるヒストグラム

3. 二つの波長空間における観測対象物のエネルギー分布

図 3 に示すように、2 次元空間において、観測対象物の分類の可能性が分かる。

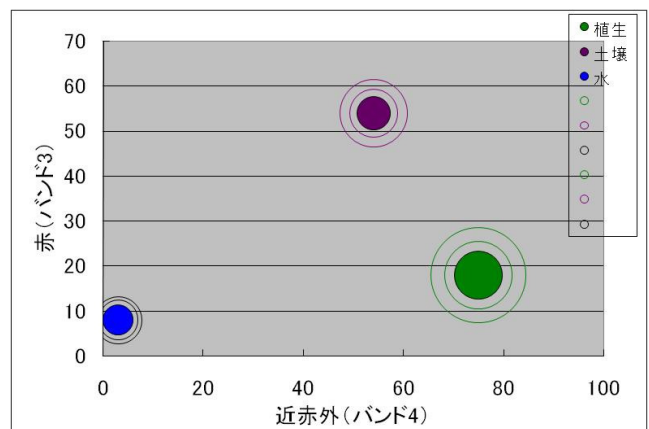


図 3 二つの波長帯域における観測対象物のエネルギー分布

4. 分類方法

(1) 教師なし分類 (Unsupervised classification)

各波長帯域を等分割し、分類対象地域の対象物を分類する。分類された対象物は何であるか分からない。

→ クラスタリング

(2) 教師あり分類 (Supervised classification)

既知の物体の各波長ごとのエネルギー分布を指定し、分類対象地域の対象物を分類する。

→ 各クラス間の距離関係を求め、判別する。

判別関数の求め方は、

- ① 尤度 (ゆうど) 判別関数 (もっともらしさ)
- ② マハラビノス判別関数
- ③ ユークリッド判別関数
- ④ 最短ユークリッド判別関数などがある。

5. 教師あり分類

(1) 最尤法 (Maximum Likelihood Method (MLH))

$$f(x_i, \alpha) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x_i - \alpha)^2}{2\sigma^2}\right\} \quad (1)$$

それぞれの現象 f が正規分布(ガウシアン分布)をなすものとする。ここで、 α は平均値、 σ は偏差である。このときの尤度関数 L は、正規分布の関数 f の積として定義される。

$$L = \prod_{i=1}^n f(x_i, \alpha) = \left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n \exp\left\{-\sum_{i=1}^n \frac{(x_i - \alpha)^2}{2\sigma^2}\right\} \quad (2)$$

この尤度関数が最も大きな値を示すカテゴリへ、 x を分類する。

この判別関数の対数を取り、さらにその変化が0を与えるところが最大の尤度となる。

$$\frac{\partial \ln L}{\partial \alpha} = \frac{\partial}{\partial \alpha} \left[n \ln \left(\frac{1}{\sigma\sqrt{2\pi}} \right) - \sum_{i=1}^n \frac{(x_i - \alpha)^2}{2\sigma^2} \right] = 0 \quad (3)$$

この判別関数が最小となるとき、尤度関数が最大値を与える。平均からの距離である。

各カテゴリが正規分布を示すとき、分類効率が最大を示す。現実には、正規分布に従わないことがあり、必ずしも分類効率が良いとは限らない。

(2) マハラノビス判別関数 (Maharanobis Discrimination Function)

カテゴリ間の相関構造を持つと仮定したとき、 σ が定数となり、(3)式は次のようになる。

$$g(x_i) = \frac{\partial}{\partial \alpha} \sum_{i=1}^n (x_i - \alpha)^2 \quad (4)$$

となり、 $g(x_i) = 0$ のとき、尤もらしくなる。この判別関数をマハラビノ関数と呼ぶ。

(3) 線形判別関数 (Linear Discrimination Function method (LDF))

(4)式のマハラビノス関数を展開すると、

$$g(x_i) = \frac{\partial}{\partial \alpha} \sum_{i=1}^n x_i^2 + \frac{\partial}{\partial \alpha} \sum_{i=1}^n \alpha^2 - 2 \frac{\partial}{\partial \alpha} \sum_{i=1}^n x_i \alpha \quad (5)$$

ここで、クラスに依存しない項 $\frac{\partial}{\partial \alpha} \sum_{i=1}^n x_i^2$ を除去すると、

$$g(x_i) = \frac{\partial}{\partial \alpha} \sum_{i=1}^n \alpha^2 - 2 \frac{\partial}{\partial \alpha} \sum_{i=1}^n x_i \alpha \quad (6)$$

のように、一次の判別関数(線形判別関数)が与えられ、分類の高速化が図られる。

(4) 最短ユークリッド判別関数

分散が各クラスとも同様であるとき、クラス間の積分が省略され、(4)式は、

$$g(x_i) = \frac{\partial}{\partial \alpha} (x_i - \alpha)^2 = \frac{\partial}{\partial \alpha} x_i^2 + \frac{\partial}{\partial \alpha} \alpha^2 - 2 \frac{\partial}{\partial \alpha} x_i \alpha \quad (7)$$

ここで、クラスに依存しない $\frac{\partial}{\partial \alpha} x_i^2$ を除去すると、

$$g(x_i) = \frac{\partial}{\partial \alpha} \alpha^2 - 2 \frac{\partial}{\partial \alpha} \alpha x_i \quad (8)$$

を最小とするクラスに分類される。

クラスの分散の情報を利用せずに、平均値の情報のみを利用するため、高速計算が可能。特に、データの分布が正規分布をなさないとき、有効である。

6. 教師あり分類の計算速度と判別精度

計算速度と判別精度は相反する要求であるが、計算機の世界速度上昇とともに、この概念に拘束されなくなってきた。

- ① 尤度 (ゆうど) 判別関数 (もっともらしさ)
- ② マハラビノス判別関数
- ③ ユークリッド判別関数
- ④ 最短ユークリッド判別関数

7. 最尤法の問題点

図4にあるバンドにおける3物体のヒストグラムと、最尤法による判別結果を示す。

水系、土壌+植生(少し植生の混じった裸地のイメージ)、植生がそれぞれの平均値と偏差の情報から判別される。問題は、植生よりデジタルカウントの大きいグループが判別されない点である。

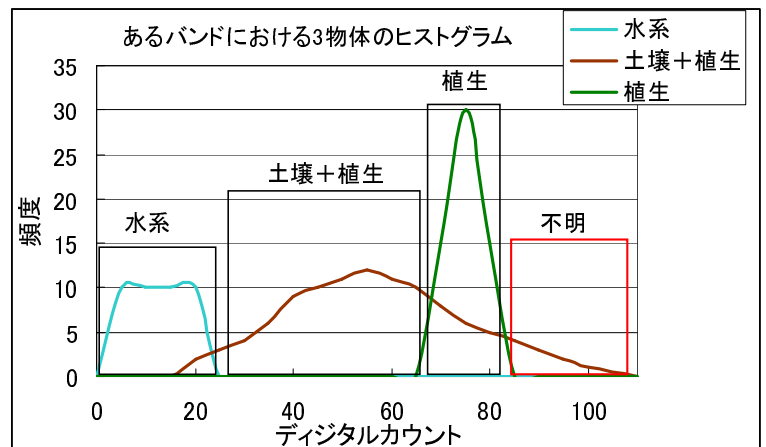


図4 あるバンドにおける3物体のデジタルカウントに対するヒストグラム