

IV-266

衛星リモートセンシングデータの最尤法分類におけるトレーニングデータの
正規性の評価に関する一考察

○東京理科大学 正会員 大林 成行

東京理科大学 正会員 小島 尚人

東京理科大学 正会員 Tashpolat Tiyyip

東京理科大学 学生員 大庭 将宣

1. はじめに 衛星データから土地被覆分類図を作成する場合、分類精度が高い手法として広く利用されているものに最尤法分類があげられる。しかし、得られる分類結果の精度は、あらかじめ教師データとして選定するトレーニングデータの良否に大きく左右される。トレーニングデータは理想的には現地の土地被覆項目をスペクトル情報によって忠実に代表できるものでなければならない。さらに最尤法分類を用いる場合にはトレーニングクラスの正規性、一様性が満たされていなければならないことは言うまでもないが、現実にはほとんどの場合このような前提条件が成り立つと仮定して最尤法分類を行っていることが多い。トレーニングデータの正規性の問題については、統計的仮説検定に代表される χ^2 検定によって正規分布との適合度を検証することもあるが、バンド別・クラス別に検定を行えるといった厳密性はあるものの、非常に手間を要することは誰しもが経験する。また、多次元正規性の検定についても多くの研究が行われているが、検定結果の表現方法の違いや処理の煩雑さ等、それぞれ一長一短がある。そこで、本研究では多次元正規性の検定を視覚的に評価できる「確率プロット法」に着目し¹⁾、トレーニングデータの「正規性の検定」や「取舍選択」といった繰り返し処理を支援できる検定手法として実用性があるか否かについて検討した。

2. 確率プロット法 確率プロット法とは、一つの母集団から N 個の標本($X_1, X_2, \dots, X_n$)を無作為に抽出するとき、その母集団の累積分布関数が $F(X)$ であるという仮説に対して標本の累積度数関数 $F_n(X)$ を作り、 $F(X)$ と $F_n(X)$ と比べてその類似性を検証するものである。 $F(X)$ と $F_n(X)$ を比較する場合にそのまま描いて比較するのではなく、仮説が正しいとき N 個の標本に対応する点がほぼ直線上に並ぶような視覚的に判断しやすいプロットの方法を工夫している。これにはP-PプロットとQ-Qプロットという2つの方法があり、いずれもプロットされた点が直線にどの程度近いかにによって仮説の正否の検討をつけようとするものである。 n 個のP変量標本値 X_{ij} ($i=1 \sim n, j=1 \sim P$) が与えられているとき、「仮説H: 標本値 X_{ij} はP変量の正規分布をもつ母集団からの無作為標本である。」についてP-Pプロット、Q-Qプロットを描いて検定する問題を考える。処理手順は次の3つのステップを踏む。

STEP 1: 各トレーニングクラスにおいて平均値 $\bar{X}$ から各個体 X_{ij} までのマハラノビスの距離を $d_1, d_2, \dots, d_n$ とし、これらを大きさの順に並べ替えた順序統計量を $d(1) \leq d(2) \leq \dots \leq d(n)$ とする。仮説Hのもとでは $d(i)$ の分布関数 $F(d(i))$ は近似的に自由度 P の χ^2 分布に従うことは理論的に証明されている。本文では紙面の都合上、 $F(d(i))$ の求め方については割愛する。

STEP 2: 式-1を解いて $Z_1, Z_2, \dots, Z_n$ を求める。

$$F(z_i) = \int_0^{z_i} \frac{1}{2^{\frac{P}{2}} \Gamma\left(\frac{P}{2}\right)} x^{P/2-1} e^{-x/2} dx = \frac{i - \frac{1}{2}}{n} \quad (i=1, 2, \dots, n) \quad \dots\dots\dots \text{式-1}$$

但し、 Γ : 自由度 P の χ^2 分布の確率密度関数

STEP 3: 次のように横軸、縦軸に対応する値を計算しP-Pプロット、Q-Qプロットを描く。

P-Pプロット: N 個の点に対して ($(i-1/2)/n, F(d(i))$)、($i=1, 2, \dots, n$) をプロット

Q-Qプロット: N 個の点に対して ($Z_i, d(i)$)、($i=1, 2, \dots, n$) をプロット

3. 確率プロットの適用 本研究では、著者らが検討しているトレーニングデータの選定方法²⁾を用いて混在ピクセルを除去し洗練化したトレーニングデータの正規性の向上を確認するために本方法を適用した。衛星データは1988年10月14日に観測されたHRVデータ(PATH332-ROW278)である。水田を対象と

してトレーニングデータの洗練化前後のP-Pプロットを図-1に示す。また、図-2にはQ-Qプロットを示す。これらはHRVの3つのバンドから構成される多次元データを用いた場合である。いずれの図を見てもトレーニングデータを洗練化した後のプロットの状態は原点を通る傾き1の直線に近づき、多次元正規性が向上していることが一目で確認できる。P-Pプロットでは、中央での直線との「ずれ」が大きく、Q-Qプロットでは直線の右上端部で「ずれ」が大きいことが判る。これは、分布の中央部で識別力が高いP-Pプロットと、分布の端の部分で識別力が高いQ-Qプロットの特徴をよく表しており、トレーニングデータの多次元正規性の検証においても十分に適用できると言える。また、このようなプロット図は、トレーニング

クラス別にコンピュータ処理によって即時に描画できることから、一般に利用されている χ^2 検定に比べて効率的であり、多次元変量で構成される正規性の高いトレーニングデータを取捨選択するといった繰り返し処理を支援する上でも極めて実用性があると言える。また、この確率プロット法はトレーニングデータの全個体を等しい重みで適合度の判定を行うことから χ^2 検定の思想と基本的に同じであり、この点から見てもトレーニングデータの多次元正規性の検証手法として推奨できる。

4. まとめ 本研究で適用した確率プロット法は、従来から用いられている確率紙と同様に確率分布の適合度を図式的に判定するものである。この方法は普通方眼目盛で視覚的にわかりやすくだしもが容易に多次元正規性を検証できる。さらに、トレーニングクラス別にデータをコンピュータに入力するだけで即時にプロット図を描くことができることから、多次元正規性の高いトレーニングデータの取捨選択を支援することにおいても効果があると言える。衛星データの処理/解析に関わる検討ではとく数学上の理論や手法論が先行し、実用性の高い要素的な手法の導入がなおざりにされていると言っても過言ではない。衛星データの利用をより一層促進する上で、このような要素的な手法の適用/提案が今後ますます重要になると考える。

【参考文献】1) E. LESAFFRE: Normality tests and transformations, Pattern Recognition Letters 1, PP. 187-PP. 199, 1983.

2) 大林成行、藤田圭一、小島尚人、中島剛彦: 衛星リモートセンシングデータの土地被覆分類におけるトレーニングデータの選定方法に関する考察、土木学会第46回年次学術講演会講演概要集、第4部門、1991年

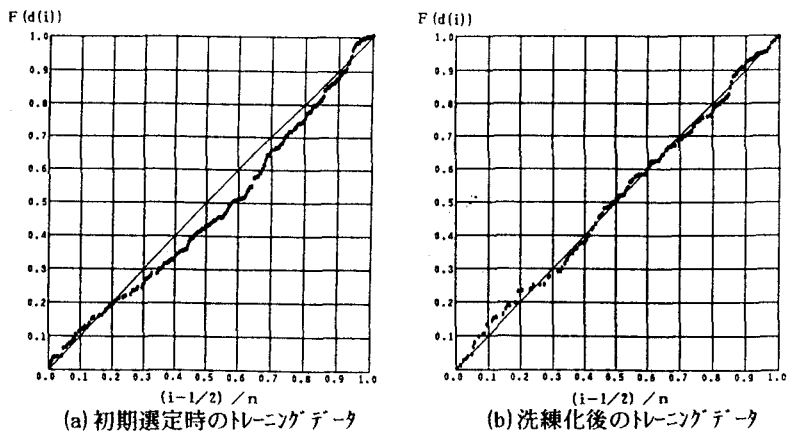


図-1 P-Pプロット(水田)

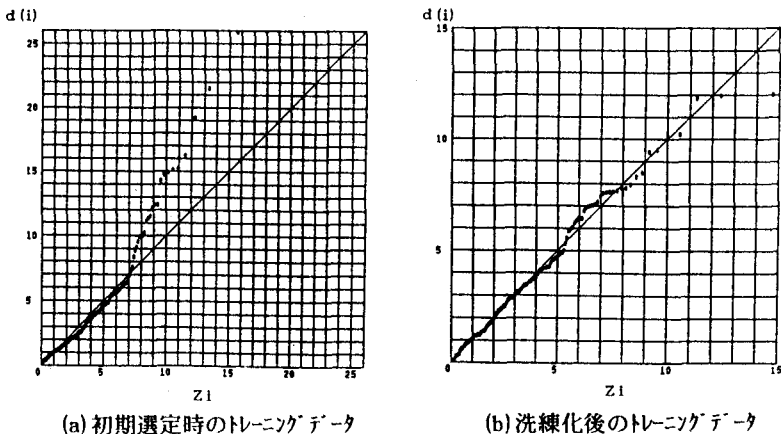


図-2 Q-Qプロット(水田)