《解　説》

分類結果の色彩表現

——探索的データ解析における カラー・グラフィックスの利用法——

大隅　昇*

1. はじめに

近年，統計情報を視覚的に伝達する手段として，データのグラフィカル表現法の研究や応用が盛んである．また，これらの手法を支援するハードウェアやソフトウェアの環境も，この十数年の間に，急速に改善されてきた．

ハードウェアの面では，マイクロコンピュータやグラフィクス・ワークステーションをはじめ，その周辺機器としてのグラフィクス・ディスプレイ，プロッタ，イメージ・リーダー，イメージ・レコーダ，グラフィクス・プリンタなどが廉価で手に入るようになってきた．

一方，ソフトウェアについては，統計グラフ用のプログラム・パッケージが多数登場し，ビジネスソフトウェアや統計システムの中でグラフィクス機能をかなり自由に利用できる環境にある．このようにコンピュータ・グラフィクスとその運用ソフトウェアは，もはやデータ解析にとって不可欠のものである．加えて最近は，操作性のよいラスター方式のカラー・グラフィクス・モニタが容易に入手できるようになり，色彩利用の技術的な面での進歩には著しいものがある．たとえば，CAD, CAM, コンピュータ・グラフィクス・アート，アニメーションなどの分野では，高価なハードウェアを利用し膨大な計算時間を費して，さまざまの色彩画像や精密色彩画を作成することが行われている．

データ解析の分野でも，統計グラフや統計マップに色彩を用いることをはじめとする種々の提案や実験がみられるようになってきたが，それらの多くは，まだ限られた環境の中での試みに留まっているように思われる．その理由の一つとして，色彩科学はいまだ厳密なものではなく，また色覚についての学説に定説がないことがあり，このため，色彩を用いることにいささか疑問が生ずるということがある．色彩の心理学的考察，色覚システムの生理学的あるいは生物学的機能の解明，色彩を構成する要素と色彩効果，色覚との関連，色彩の計量化など多くの検討事項があり，したがって体系的な色彩理論は未だ確立されていないのである．

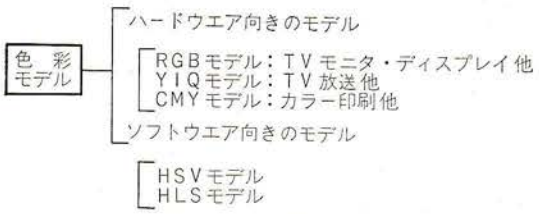
こうした事情を背景に，色彩を用いることにはかなりの勇気を必要とするが，それでもなおかつデータ解析にカラー・グラフィクスを利用することには魅力があり，しかも多くの利得が期待される．また何よりも大きな理由は，われわれが色彩の世界に住んでおり，したがって色彩でものを見ることをきわめて自然のことと感ずるからである．

この報告の目的は，データ解析，とくに自動分類において色彩を用いる場合の探索的な方法について若干の提案を行うことと，その簡単な適用例を示すことにある．とくにここで述べる方法は，マイクロコンピュータと廉価なカラー・モニタ・ディスプレイがあれば十分に利用できる，実用的でしかも操作性のよい環境を提供することを念頭に考えた．また，データ解析において，色彩操作とそのソフトウェアをどのように構築するかについての一つの提案でもある．

2. 色彩モデルとデータ解析のインターフェース

ここでまず，データ解析において色彩を利用するうえで考慮すべき事項について若干述べてみたい．すでに指摘のように，現在の色彩科学の知識では，色彩現象を表

* 統計数理研究所　調査実験解析研究系
パターン解析研究部門



図・1　色彩モデル

わすモデルを心理的あるいは知覚の観点から十分に説明することは難かしい．つまり，カラーグラフィクス機器の性能が技術的にどれほど向上しても，色彩科学には十分に解明されない面が残される．したがって知覚モデルとプログラム開発に適した記述的モデルとの差異を十分に認識したうえで色彩の利用法を検討せねばならない．実際，データ解析において色彩を用いるときに生ずる種々の拘束や危険性，錯視の問題，色の対比効果や調和の問題，ハードウェアの条件，制約などについてさまざまの提案がみられるようである[2),3),8),9),11),12)]．ここでの提案は，これらの報告で指摘の事項を十分配慮したうえで，データ解析における色彩利用の新たな展開を試みることにある．

簡単にいえば，コンピュータ・グラフィクスにおける色彩利用の依りどころとなる色彩モデルは，オストワルトやマンセルの提唱したカラー・システム（表色系）に近いものであり，色彩の調和の原理のうえに構築される記述的なモデルである[7),13)]，もちろん．前述のようにこれについての定説はいまのところあるとはいえない．しかし少くとも，こうした色彩モデルの考え方をとり入れた色彩操作を行うことを前提とした色彩利用を考えるべきであろう．また，カラー・モデルを扱うとき，ハードウェアの環境（機器の性能やアーキテクチュア）がおおいに関係する．とくに最近は，従来のストレージ・チューブ型のディスプレイにかわって，高精細度のラスター方式のカラー・モニタ・ディスプレイや，それを標準装備したマイクロコンピュータやグラフィクス・ワークステーションが登場し，色彩利用の計算機環境が大きく変化している[1),5)]．

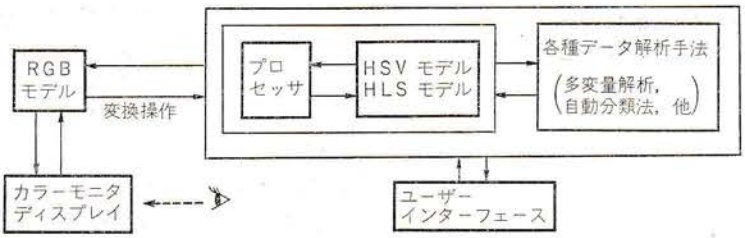
ところで，色彩操作のソフトウェアも，これらのハードウェア環境を十分意識して考えるべきである．一般に，ラスター方式のカラー・グラフィクス・モニタを，電気的あるいは工学的に操作するうえで適しているハードウェア向きの色彩モデルと，人間の感覚的要素を考慮しやすくユーザ・インターフェースの設計に適した色彩モデルとがある．カラーグラフィクスについて論じた多くの報告の中にみられる代表的な色彩モデルを要約すると図・1のようになる[7),13),14)]．

一方，人の色覚を記述的に表現する色彩モデルの代表として，HLSモデルやHSVモデルがある．これらの色彩モデルは，**色の3要素**，H：色相（hue），L：明度（lightness）またはV：V値（value），S：飽和度（saturation）により表わされる．色の3要素に基づく考え方は人の直観によく合い，しかも記述的，アルゴリズム的であるから，データ解析手法とのインターフェースのプログラム化に適しているといわれている．

さらに別の問題として，カラー・モニタの発色数の制限ということがある．多くの機器では，限られた数の色しか利用できない（たとえば，8～16色，あるいはせいぜい64色程度）．やや高級なカラー・モニタでは4096色，さらに最近のグラフィクス・ワークステーションでは1600万色程度の発色が可能であるが，実際に利用者が指定できる色の数は多くの場合制限される．人が．知覚できる色の数は30数万色とも100万色以上ともいわれているが，工学的な意味で発色数の問題は大きく改善の方向に進みつつある．

しかし，データ解析においては，発色数の豊富さもさることながら，われわれが色彩をどれくらい自由に操作できるか，がより重要である．たとえば，RGBモデルで色の3原色R（赤），G（緑），B（青）の混合比を指定して好みの色や希望する色を探すことはかなりやっかいである．これは，カラー印刷の基本3原色黄（Y），シアン（C），マゼンタ（M）の扱いについても同様である．一方，HLSモデルやHSVモデルに基づいて彩色を行うほうが，人の色感によくあいユーザ・インターフェースの機能が改善される．したがって色の表現用語である，色相，輝度，明るさ・暗さ，明度，色調，シェード，トーンなどの概念にもとづいて色彩を考えることが容易になる．また最近では，希望する色を用いて彩色や配色を行うために，自然言語に近い色彩表現用語を用いる色彩呼称システム（Color-Naming System；CNS）がよいという提案もある[6)]．

以上のことから，カラー・モニタの制御に適したRGBモデルと，データ解析手法とのインターフェース・ソフトウェアの開発に適したHLSモデルやHSVモデルとの間を結ぶ色彩モデル変換アルゴリズムとそのプログラムの準備が必要となる．いまこれらの関係を模式的に示すと図・2のように表わされるであろう．また，こうしたアーキテクチュアを実現するハードウェアとしては，既存のマイクロコンピュータやグラフィクス・ワークステーションがあれば十分である．むしろ，ユーザフレンドリーなインターフェースを提供する操作性のよいプログラムの開発が重要な課題となる．筆者は，こうした視点から，色彩モデルの利用に必要なプログラムの開発を



図・2　データ解析における色彩操作の概念図

進めてきた（それらの一部は参考文献15)にある）．これらのプログラムは，色彩を用いるデータ解析を行う際の基本的な道具として機能する．これから述べる考えはすべて，これらのプログラムの利用を前提としている．もちろん，繰返し指摘するが，こうしたアルゴリズム的色彩モデルは人の知覚カラーシステムを正確に表わすものではない．しかし，意味のない彩色や配色を行うグラフィカル表現や統計グラフを用いるよりはより現実的であり，また実用的であることはいうまでもなかろう．

3.　自動分類への色彩の利用

自動分類においては，計算結果の表示の手段として，グラフィカル表現が広く利用される．分類記号を用いた散布図，デンドログラム，あるいは木構造表現，グラフ的な表示法 (minimum Spanning Tree など)，クラスター楕円，クラスター濃淡図等々，多くのグラフィカル表現法がある．これらはいずれも分類の意味を視覚化し，情報を効果的に伝達したいということにほかならない．元来，分類操作は分析対象としたデータを区分し，標識を付与することである．このことは，とりもなおさず彩色操作に結びつく．つまり色彩により，似ているものは類似の色で，離れたグループはなるべく異なる色で彩色すると考えることはきわめて自然である．しかし，この操作の前提として，できるだけ客観的に彩色を行い，分類の意味を色彩として効果的に視覚化するための手順が必要となる．筆者が用意した色彩モデル間の変換プログラムと，それを装備したカラーグラフィクス・システムを用いて，自動分類における色彩利用の可能性と有効性を調べることが次の目標である．

3.1　多変量データの色彩パターン

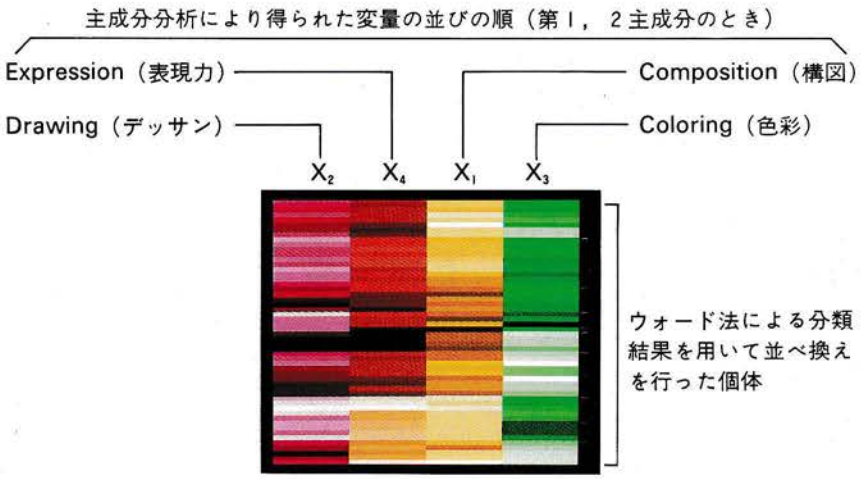
通常，与えられた多変量データ行列を用いて個体（あるいは変量）の分類を行い，その結果を適当な方法でグラフィカル表現すること（たとえば，デンドログラムを書く）が行われるが，このとき次のような疑問が生ずる．

1)　用いた変量のうちどの変量が個体の識別に役立っているのか．
2)　変量間の関連性（類似や差異）をどのように知るのか．
3)　分類手法の違いが分類結果におよぼす影響をどのように調べるか．
4)　とくに階層的分類手法を用いるとき，デンドログラムの書き方は一通りではない．デンドログラムの見栄えがグループ化の様子を正しく表わしているとはかぎらない．つまり，客観的にクラスター化の程度を知る目安とはなりにくいという問題にどう答えるか．
5)　類似度や非類似度の選択，データの標準化の有無などがクラスタリングの結果に及ぼす影響をどう知るか．

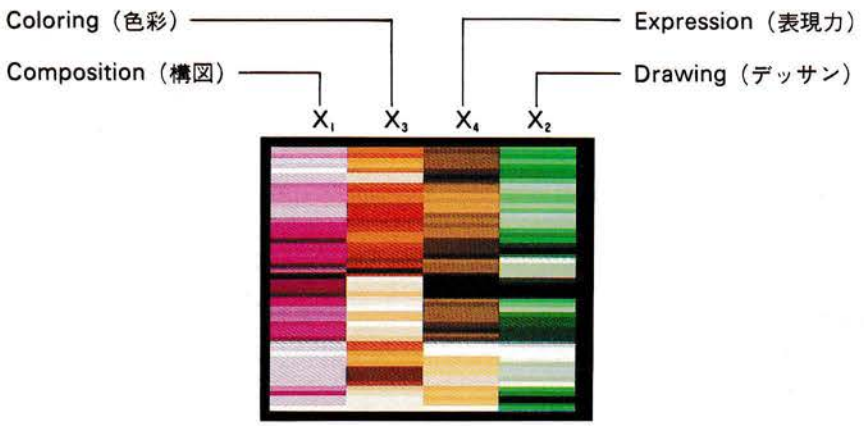
実は，これらはいずれも自動分類に共通の問題であり，いまだ明解な答があるわけではない．通常，分類結果を即座にデータ構造の解釈に用いることには若干の無理があって，分類後にその情報を用いて何らかの再分析を行ったり，クラスターの同定化を図る必要がある．こうした吟味の過程で，色彩利用を探索的手法として組み込むことで，効果的なデータ解析を進めることができるのではないか，というのがここでの主張である．しかも，その処理手順は，次に示すようにきわめて簡単であり，したがってプログラム化も容易である．

ステップ 1：まず，与えられた（n 個体）×（p 変量）の多変量データ行列，$\boldsymbol{X}=(x_{ij})$ $(i=1, 2, \cdots\cdots, n\,;\, j=1, 2, \cdots\cdots, p)$ の個体を分類する．分類手法は任意の方法でよい（たとえば，階層的手法，非階層的手法）．

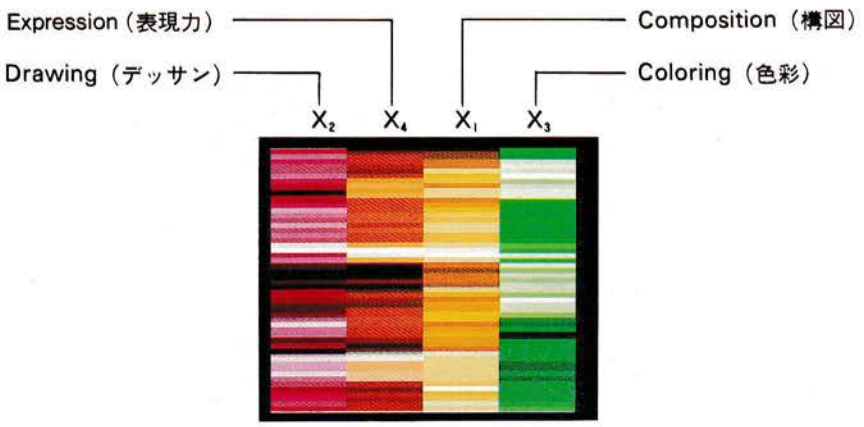
ステップ 2：分類結果に従って，そのデータ行列 $\boldsymbol{X}$ の個体側の並べかえを行う．たとえば，階層的分類法を用いた場合，デンドログラムの個体の並びの順を用いることができる．k-means 法のような非階層的手法を用いるならば，適当なクラスター数を指定して分類を行い，得られたメンバーシップリスト（各個体がどのクラスターに所属するかを示す識別コードの表）に基づいて個体をソートする．



（a） 第１，２主成分を用いた場合

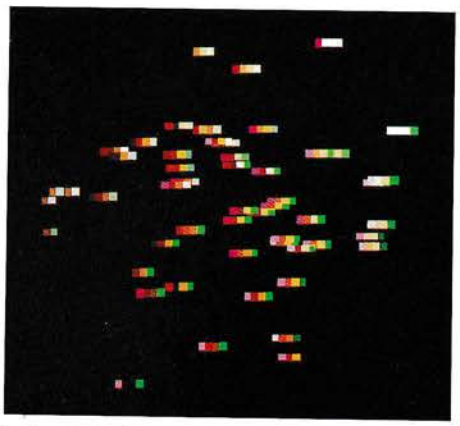


（b） 第２，３主成分を用いた場合

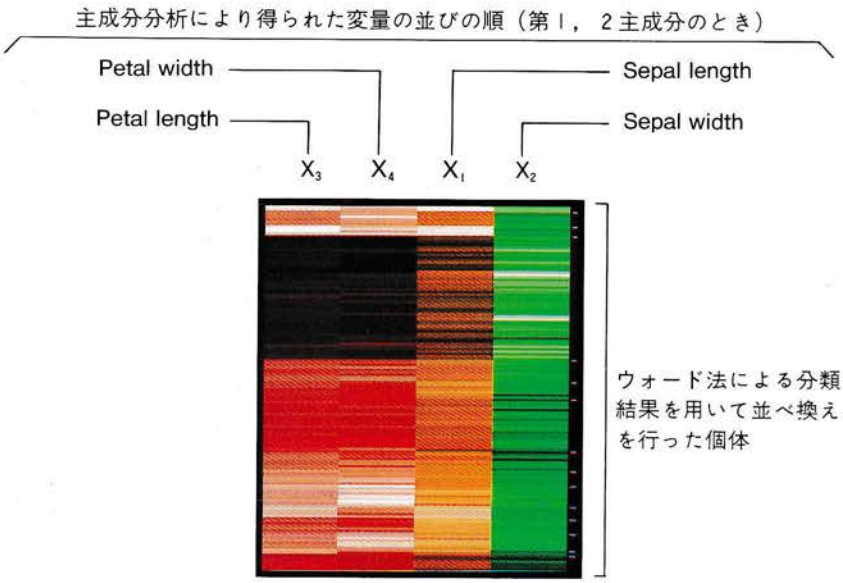


（c） 分類手法を k-means 法とした場合（第１、２主成分を用いた場合）

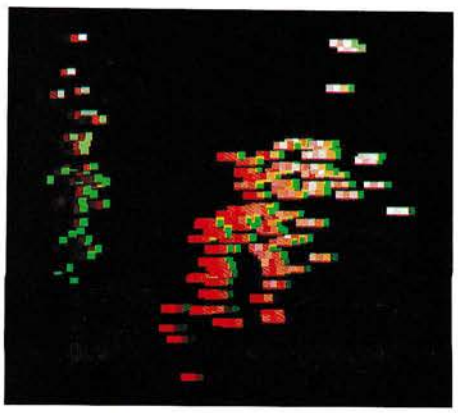
図・7 色彩パターン行列（画家の評価データ）



図・8　色彩プロット図の例（図・5（a）に対応する）



（a）　色彩パターン行例



（b）　色彩プロット図

図・9　Irisデータの分析例

ステップ 3：次に変量側の並べかえを行う．これは変量間の関連性を色彩の変化として表わすためである．変量の並べかえは，適当な分類手法を用いてもよいし，後述するように主成分分析を用いて決めてもよい．

ステップ 4：以上の手順により，個体と変量のそれぞれの並べかえを行って得られる行列を改めて $\boldsymbol{X}^*=(x_{ij})^*$ $(i=1, 2, \cdots\cdots, n\,;\, j=1, 2, \cdots\cdots, p)$ と表わす．

ステップ 5：次に，この $\boldsymbol{X}^*$ を適当な色彩モデルに基づく変換により色の3要素（H, L, S）で表わされる**色彩パターン行列** $\boldsymbol{Y}$ に変換する．さらに，この $\boldsymbol{Y}$ を RGB モデルにより色彩強度（intensity）に変換しカラー・モニタ・ディスプレイに送信すると色彩情報として表示される．

以上の方式による利点として，次のことがある．

1) 多変量データ行列に対して，分類手法を適用した結果を色彩により一目で眺められる．
2) しかも，そのデータ行列の個々の観測値の特徴が，直接カラーパターンとして表示される．
3) 用いた分類手法の差異，類似度や非類似度の分類に及ぼす影響などを色彩により比較できる．
4) 原理は単純であるから，プログラム化が容易で，しかも移植も簡単である（若干のグラフィクス機能を必要とするかもしれない）．

次に，上の各ステップについて，もう少し詳しい説明を加えておく．

変換データ行列 $\boldsymbol{X}^*$ の色彩パターン行列 $\boldsymbol{Y}$ への変換

行列 $\boldsymbol{X}^*$ を HLS 色彩モデルを用いて（HSV モデルでもよい），色彩パターン行列 $\boldsymbol{Y}$ に変換することを次のように表わす．

$$\boldsymbol{X}^*=(x_{ij}^*) \xrightarrow[\text{色彩モデル}]{\text{HLS}} \boldsymbol{Y}=\{y_{ij}\,;\, H_j, L_{ij}, S_j\} \qquad (1)$$

$$(i=1, 2, \cdots\cdots, n\,;\, j=1, 2, \cdots\cdots, p)$$

ここで，

H_j：色相　$0 \leqq H_j < 360$

L_{ij}：明度　$0 \leqq L_{ij} \leqq 1$

S_j：飽和度　（これは後述の方法で与える）

変量内のデータのばらつきの色彩化—明度の利用—

ある変量 j についての観測データ x_{ij}^* のばらつきが明度の変化として対応するように，また区間 $[0, 1]$ 内（あるいは0％～100％内）に入るように次の線形変換を行う．

$$L_{ij}^*=(x_{ij}^*-\min_i x_{ij}^*)/(\max_i x_{ij}^*-\min_i x_{ij}^*) \qquad (2)$$

$$(i=1, 2, \cdots\cdots, n\,;\, j=1, 2, \cdots\cdots, p)$$

また必要があれば，明度の大小とデータの示す数値の大小が対応するように，つまり数値が大きい（小さい）ほど明度が高く（低く）なるように，変量の特性が示す意味に合わせて逆変換を行うために，次のように**逆変数**を作る．

$$x_{ij}^*=\max_i x_{ij}^*+\min_i x_{ij}^*-x_{ij}^* \qquad (3)$$

（逆変換を必要とする変量 j について）

以上の変換により，各変量内のデータのばらつきが，明度差すなわち色の濃淡（明暗）として表わされる．また逆変数を用いることで，明暗が意味的にそろった向きに合わせることができる．

変量の並べかえ—飽和度と色相の利用—

変量側への配色は，分析者の利用目的に応じていくつかの場面が考えられる．

——性質が類似している，あるいはそう考えられる変量は近い位置において，色相の類似した色を割り当てる．

——逆に，性質が異なると思われる変量は，色相差を大きくとる．場合によっては，補色の関係におく．

——色相は固定して，明度差だけを用いて各変量内のデータの特徴を同一色の濃淡としてのみ表わす．

変量間の関連性を色彩の関連性として示すために，利用者が，自由に色を扱えることも大切であるが，これをある程度客観的に決める方式も必要であろう．つまり，利用者の多様な要求に応じて，いくつかのオプションを用意することが適当である．

オプション 1　飽和度は固定し，色相だけを各変量に対応させるとき

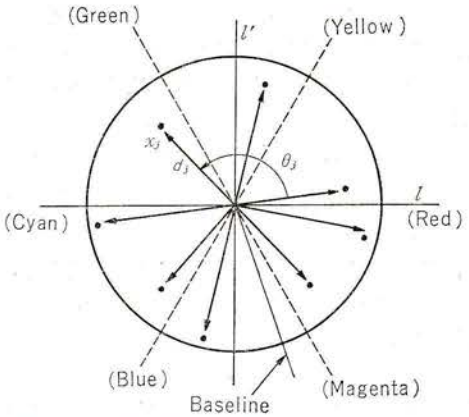
たとえば，次のように与える．

$$S_j=r \text{（通常は } r=1 \text{ とする）}$$
$$H_j=\theta_j, \quad \theta_j=360\times(j-1)/p \qquad (4)$$
$$(j=1, 2, \cdots\cdots, p)$$

飽和度を $r=1$ と与えると，HLS 色彩モデルの双円錘の表層上にある色だけ（したがって純色）を用いることに相当する．色相 H_j はある変量 j に対して，上のように等角度となるように与えてもよいが，変量の類似や差異を主観的に判断して，あるいは事前に何らかの情報があるときにはそれに従って，利用者が適当に与えてもよい．したがって，変量の並びの順は，原データ行列 $\boldsymbol{X}$ のそれと同じ順である必要はない．さらに，変量側の階層的分類を行い，得られたデンドログラムの変量の並びの順を用いることも考えられる．

オプション 2　変量間の類似性を考えて飽和度と色相を与えるとき

飽和度と色相を，オプション1のように利用者が任意に指定できる場合の利点はいろいろあろうが，同時に変量間の関連性の解釈に混乱をきたす危険がなくもない．



図・3　因子負荷量の布置と色彩モデルの関係

そこで次に，変量間の関連性を飽和度と色相の情報に自動的に変換する方式を考える．ここでは，主成分分析を用いることとし，次の条件を満たすような手順を用意する．なお，ここでの方法は，主成分分析に限らず，バイプロット法などでも用いることができるであろう．

——変量の関連性を配慮してその並べかえを行う．これを色相の変化に対応させる．

——次に，変量の寄与の程度を色調（tint）の変化として表わす．つまり，寄与の高い変量は鮮やかに，そうでない変量は飽和度を抑制して色をぼかす．

これらを考慮して，次の手順を用意する．

ステップ 1：与えられたデータ行列 $\boldsymbol{X}$ の主成分分析を行い，因子負荷量行列を作る．

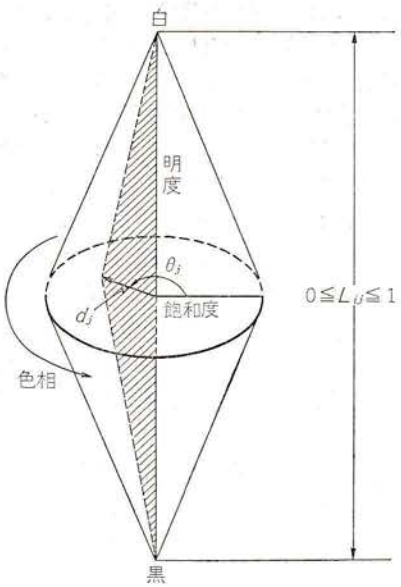
ステップ 2：任意の2成分 l, l' を指定して得られる因子負荷量の布置図を，HLS 色彩モデルの明度に垂直な面に対応させる．図・3，図・4は，この状態を模式的に表わしたものである．さらに，図・3のように，単位円の右側に赤を配置し，左回りに，基本色である黄（Y），緑（G），シアン（C），青（B），マゼンタ(M)を等間隔に順に置く．他の色相は，それぞれこの円に対して角度として与えられる．

ステップ 3：以上の準備のもとに，ある変量 j についての観測データ x_{ij} $(i=1, 2, \cdots\cdots, n)$ に対し，H, L, S を次のように与える．

$$H_j=\theta_j \qquad 0\leqq\theta_j<360$$
$$O_j=d_j \qquad 0\leqq d_j\leqq 1 \qquad (5)$$
$$L_{ij}=\{\text{これは，式(2)と同じ方式で与える}\}$$

ここで，d_j, θ_j は次のように与える．

まず，ある変量 j について，図・3の布置の原点からの距離 d_j を飽和度とする．次に，θ_j は，図・3のように時計の針と反対回りの方向に水平軸となす角を与える．



図・4　HLS 色彩モデルとデータの関係

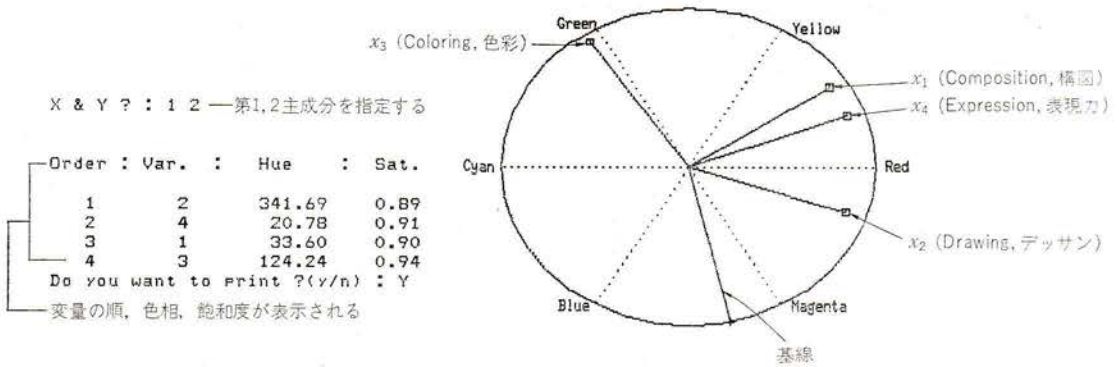
変量の順序は適当な**基線**を利用者が与えて，その位置から左回りに順に選んで得られる変量の並びを採用する．これは，ディスプレイ上に写された図・3に相当する図をみながら利用者がヘアーライン・カーソルを用いて指定する．こうして決められる H, L, S の関係は，ある変量 j について表わすと，図・4の太い点線に沿って分布すると考えればよい．

この方式に従うと，多変量データの視覚化が可能であるだけでなく，色彩が意味をもって機能する．しかも，多変量データを色の3要素（つまり3次元空間）を用いて表わしていることに注意しよう．さらに，この方式によると，色彩が次の意味をもつ．

1) 指定した因子負荷量の布置図の中で，近い位置にある変量は類似色相となる．一方，差異が大きく非類似関係にある変量，特に逆の関係にあるものは，互いに補色に近い色を示す．
2) 寄与の低い変量は飽和度が低減し無彩色（灰色）に近づく．反対に寄与の高い変量ほど鮮かになり，純色に近づく．つまり，変量の寄与の程度を色調の変化として観察できる．
3) 色の濃淡がデータのばらつきやひろがりを表わしている．とくに，データの分布が分離現象を示したり，はずれ値がある場合には明度差が顕著に現われる．

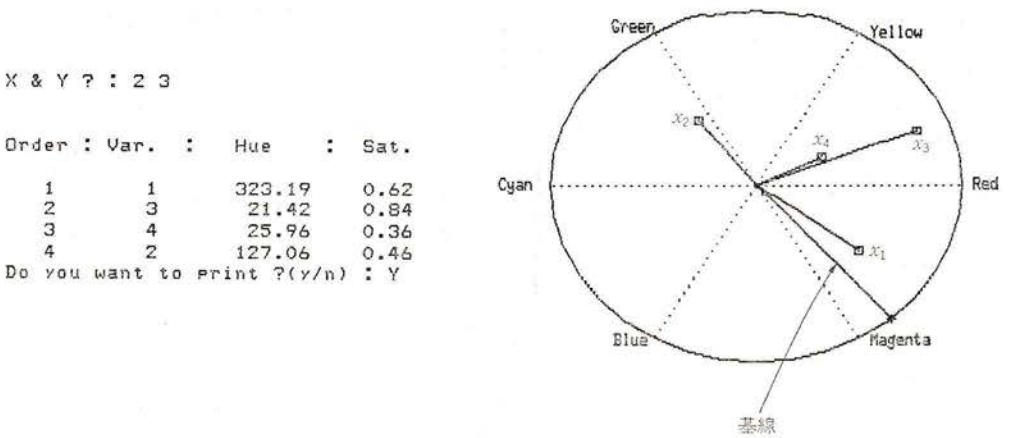
3.2 簡単な実験例

ここで，二つの例をみることにしよう．第1の例は，



Pilesの画家評価データの出力例（基線を図のように与えると，変量の順がx_2, x_4, x_1, x_3と定まる）

（a） 第1，2主成分を指定した場合



（b） 第2，3主成分を指定した場合

図・5 画家の評価データの分析

分類の分析例として知られる，画家の評価データを用いる．第2は，分類問題の実験データとしてよく用いられるIrisデータによる例である．

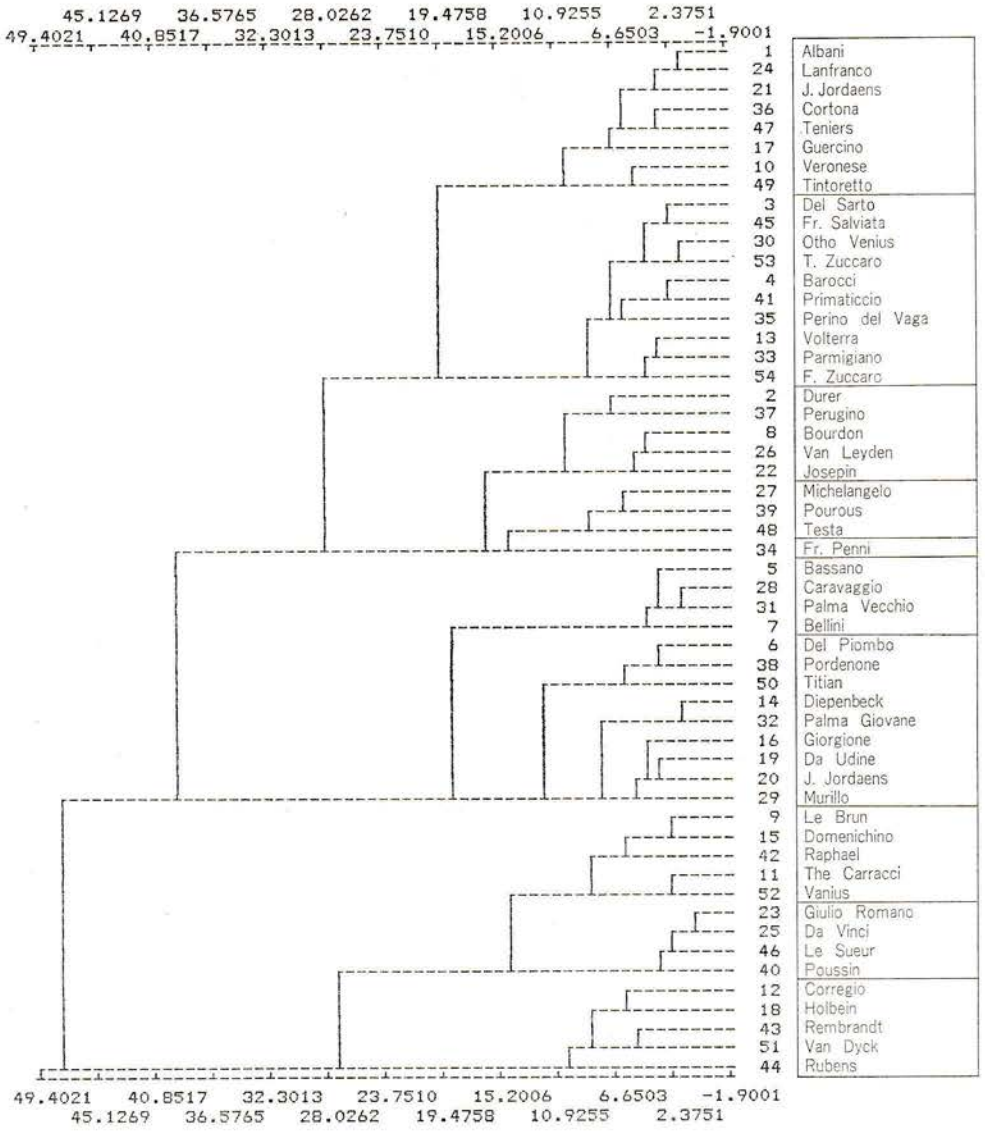
適用例 1

Davenportらが分類の例として引用したRoger de Pilesによる画家の評価データを用いて[4]，ここで述べた方式により色彩パターン行列を作成する．もとのデータは56人の画家の評価（aesthetic judgement）を行った評点データであるが，2人のデータに欠測があるので，これをのぞいた54人のデータを対象とする．なお，評価項目は，x_1（構図，composition），x_2（デッサン，drawing），x_3（色彩，color），x_4（表現力，expression）の四つである．

主成分分析による色相と飽和度の決定

与えられた大きさが54×4のデータ行列の主成分分析を行い，得られた結果（因子負荷量，主成分得点など）をファイルに格納する．この情報を用いてグラフィクス・ディスプレイ上に因子負荷量と色相，飽和度の関係が表示される．たとえば，第1，2主成分を指定すると図・5の(a)のように，因子負荷量の布置，変量の並びの順，色相，飽和度が表示される．ここでは，基線をマゼンタの左に置いたので，変量の順序がx_2, x_4, x_1, x_3となる．四つの項目は$\{x_2\}, \{x_1, x_4\}\{x_3\}$と分かれており，これが色相差となって現われる．各項目の色相は図からおよそどのような色を示すか読みとることができる．たとえば，x_3はほぼ緑色に近い．また，飽和度は第1，2軸の中ではいずれも値が大きい（寄与が高い）ので，どれも鮮やかな色相を示す．

次に，第2，3主成分を指定してみよう．このとき得られる結果が図・5の(b)である．1，2成分の場合に比べて飽和度が低くなり（とくにx_4）色調が低減する．x_3とx_4とは色相は類似しているが，x_3は明るい茶色に，x_4は



図・6　画家の評価データの分類結果（ウォード法）

色が混ざりこげ茶色となるはずである．これらを確認するため，色彩パターン行列をカラーモニタ上に出力するが，その前に個体の分類を行う．

自動分類による個体の並べかえと色彩パターン行列の生成

ここで，原データ行列を適当な分類手法でクラスター化し，個体の並び順を決める．たとえば，ウォード法や *k*-means 法などの標準的な手法でよい．いまウォード法を用いて得られるデンドログラムに基づいて，個体の並べ換えを行う．たとえば，図・6がその例である．一方，すでに変量の順序は図・5の(a)，(b)のように与えられている．これらの情報を用いて，原データ行列の行と列の並べかえを行い，変換行列 X^* を作り，これをさらに色彩パターン行列 Y に変換し，カラーモニタに出力する．こうして得られた画像が図・7の(a)，(b)（カラーページ参照）である．図・7の(a)，(b)（カラーページ参照）は，それぞれ図・5の(a)，(b)に対応する．また個体の並びは図・6のデンドログラムのそれに対応する．

さて，図・7の色彩パターンを眺めると，デンドログラムだけでは十分に見えなかったクラスター化の特徴がみられる．たとえば，次のようなことが観察される．なお，目安として，クラスター数を10群と指定して，図・7（カラーページ参照）の右端に印したように，クラスターの区分の目安を白い線で仕切っておく，こうして，

各テスターを観察すると次のような特徴が見られる．

——主として，構図(X_1)について優れているが，他の三つの変量についてはやや劣るという，グループとしてマニエリスイス派を中心とする次のクラスターがある．

{Albani, Lanfranco, Jordaens, Cortona, Teniers, Guercino, Veronese, Tintoretto}

——色彩(X_3)に対する評価がやや劣るが，デッサン(X_2)，表現力(X_4)，構図(X_1)の点で優れていて，しかも全体に評価が高いグループとして，次のクラスターがある．

{Le Brun, Domenihino, Raphael, The Carracci, Vanius}

——色彩(X_3)，構図(X_1)はよいが，表現力(X_4)，デッサン(X_2)の点でやや劣ると判断されたグループは，

{Corregio, Holbein, Rembrandt, Van Dyck, Rubens}

となる．

このように，デンドログラムの情報と色彩パターン行列，したがってもとの行列の情報との対比により分類の様子を視覚的に観察できる．このとき次の特徴は重要であろう．

——変量の並びの順は，主成分分析の計算を一度行えば固定される．

——一方，個体の並び順は，用いた分類手法により変化するので，クラスター化の程度や，用いた手法に対して，分類のキーとして強く働いた変量などの状況が視覚的にみられる．たとえば，図・7の(c)（カラーページ参照）は k-means 法を用いた例である．図・7の(a)（カラーページ参照）と比較すると，手法の違いが視察できる．

——行列 $\boldsymbol{X}^*$ の各要素への配色は，分類手法を変えても変化しないので，色の判断や評価に混乱を生じない．

——また，指定する主成分を変えると，その成分における変量の関連性や寄与の程度が，色相，飽和度と連動して変化し，これを意味のある色彩変化として観察できる．

——変量間の逆相関の関係や，場合によっては交絡化の様子などが明度差（明暗の分布）としてみられる．

主成分得点の色彩化

上にみたように，色彩パターン行列上の各個体の多変量観測データは色彩の帯として表示される．これを**個体の色彩ベクトル**と名づけることにし，この各個体の色彩情報を主成分得点の布置図の中の対応する得点座標の位置に置いて得られる図を**色彩プロット図**と名づける．画家のデータの分析を例にとれば，4変量の合成値である主成分得点の座標位置に4変量の色彩ベクトルの帯を置く．これにより縮約化された空間の中で，それぞれの個体が示す多変量観測値の特徴を色彩の変化として読みとることができる．

図・8（カラーページ参照）は，画家の評価データの主成分得点のうち，第1，2主成分を指定して作成した色彩プロット図である．図・5の(a)の色彩パターン行列の下の方のグループが，色彩プロット図の上方から右側にかけて分布している．また，色彩パターン行列の上方に位置するグループは，色彩プロット図の下方に集まっている．また，色相や明度が変量の関連性や観測データの数値の大小に対応するので，クラスターの意味や，主成分得点という合成値と原データの関係などを一目で眺められる．とくに，はずれ値や異常値，あるいはそれに近いデータは，色の変化も大きく表われる．図・8（カラーページ参照）では中央左下のデータ（画家名は Fr. Penni）が若干こうした傾向を示している．

なお，色彩ベクトルを，通常の2次元散布図の中に置くと，多変量のデータの色彩散布図として利用することができる．

適用例 2

次に，Edgar Anderson の Iris データをもちいた例を示そう．これは，3種類の Iris (Setosa, Versicolor, Virgi) について四つの項目 (Sepal length, Sepal width, petal length, Petal width) を測定したデータである．また，R. A. Fisher が判別分析の事例データとして用いたことで知られており，分類手法の検証や数値実験例で，広く用いられるデータとして有名である．

さて，得られた色彩パターン行列と，色彩プロット図（第1，2主成分について）が図・9の(a)，(b)（カラーページ参照）である．色彩パターン行列の作成時に用いた分類手法は k-means 法であり，クラスター数は15群とした．このクラスターの仕切りを図中の右側に白い線で示した．結果は，細かい説明をするまでもなく明らかである．変量間の関係，群の様子，プロット図の中での各変量の特徴，クラスター化の様子とクラスターを特徴づけている変量の分布，ややはずれ値と思われるデータの様子などを明瞭に把握できる．

4. むすび

コンピュータ・グラフィクスにおける色彩利用の考え方と，それに基づいて作成されたプログラムを自動分類法に応用するための簡単な方法について提案した．実際に，いつくかの例について適用し，データ解析を試みたが，その結果，利用に当たって次の事項に留意することが必要であろう．

1) データ数や変量数が多い場合には，やや無理がある．これは，用いるカラー・モニタの性能にも依存する（画素の数あるいは解像度）．ここで用いたモニタは，横256×縦240画素であるが，この場合，1画面に対して，個体数は100～180，変量数は10～20程度である．多量のデータでも画面を切り換えながら表示できるが，目で観察できる範囲は，やはり1画面内に限るほうがよい．もちろん，モニタの解像度がより高いものを使えば，個体数，変量数とも増やすことは可能である．

2) 個体数が非常に多い場合には，まず個体の分類を行い，クラスター数をやや多めに指定して，クラスター別の平均ベクトルを用いて，平均ベクトル行列の色彩パターン化を行うことが考えられる．筆者のプログラムでも，これをオプションとして用いることができる．クラスターごとの特徴を捉えるには，この方式のほうが適しているかもしれない．

3) ここでは，色彩映像を静止画として示したが，実際に表示を行うときには順を追って画面上に色彩ベクトルが表われるので，色彩の変化を動きとして眺めることができる．とくに，色彩プロット図の場合には，打点位置が夾雑になって見にくくなる傾向があるが，打点の順を目で追うことによりこれが避けられる．

4) 同時に利用できる色の数（カラーモニタの発色可能数ではなく，実際に利用できる色の数）が少ない場合には，ここで述べた方法はあまり意味がない．できれば256色（少なくとも64色）以上が必要であろう（筆者が用いたカラーモニタの場合は，3原色（R，G，B）の各色について16段階，したがって 163=4096色を用いることができる）．データに内在する変動や関連性を色の微妙な変化として表わすためには，画像数（解像度）の多少よりもむしろ利用可能な色の数が多いことを優先したほうがよい．

参考文献

1) Beatty, J. C. (1983) : "Raster graphics and color", The American Statician, 37, [1],60-75.

2) Becker, R. A. (1983) : "Integrating color into statistical software", Comp. Sciences and Statistics, The Interface, 14th Symposium.

3) Cleveland, W.S., McGill, R.(1983): "A color-caused optical illusion on a statistical graphs", The American Statistician, 37, [2], 101-105.

4) Davenport, M. et al. (1972): "The Statistical analysis of aesthetic judgment; An exploration", Applied Statistics, 21, 324-333.

5) Enderle, G., Kansy, K., Pfaff, G. (1984) : "Computer Graphics Programming-GKS The Graphical Standard", Springer Verlag.

6) Farhoosh, H., Schrack, G. (1986): "CNS-HLS Mapping using fuzzy sets", IEEE CG & A, June, 28-35.

7) Foley, J.D., Van Dam, A.(1982): "Fundamentals of Interactive Computer Graphics", Addison Wesley.

8) Hosmer, T., Edwards, R. (1984): "Imaging; A tool for data reduction", The American Statistician, 38, [4], 322-327.

9) Nichoeson, W.L., Littlefield, R. J. (1983); "Interactive color graphics for multivariate data", Comp. Science and Statistics, The Interface, 14th Symposium.

10) Ohsumi, N.(1983): "Practical Techniques for Areal Clustering", Data Analysis and Informatics III (eds., E. Diday and others), North-Holland.

11) Trumbo, B. E. (1981); "A theory for coloring bivariate statistical maps", The American Statistician, 35, [4], 220-226.

12) Wainer, H., Francolini, C. M. (1980) ; "An empirical inquiry concerning human understanding of two-variable color maps", 34, [2], 81-93.

13) リー・ボールドウィン (1985)：“コンピュータ・グラフィックスにおける色彩の考察”，「日経バイト」, March, 163-174.

14) 朝日新聞社編(1983)：「色の彩時記」.

15) 大隅昇 (1986)：“カラーグラフィクスのための色彩モデルの変換プログラム”，「統計数理」, 34, [1] (掲載予定).