

前処理付き階層的クラスタ分類法による色彩構成の分類

A Hybrid Hierarchical Clustering Method for Classification of Color Compositions

小林 光夫
根本優美子

電気通信大学情報工学科
電気通信大学大学院情報工学専攻

Mituo Kobayasi
Yumiko Nemoto

1 はじめに

配色の法則性を探るために筆者らは、絵画やデザイン画の色彩構成の自動分類を試みてきた[1, 2, 3, 4]. その方法は、まず、画像を構成する色の色空間における分布、すなわち色分布に着目し、各画像の色分布間の類似度あるいは距離を計算する. つぎに色分布の集合に対して、階層的クラスタ分類法を適用するというものである.

階層的クラスタ分類法では、クラスタ間の距離の定め方が重要であり、筆者らはクラスタのまとまりを良くするという最遠隣法 (Furthest Neighbor Method, 以下 FN 法と略す) を採用してきた. ところが、FN 法は、クラスタ間の分類を悪くするという欠点もあわせもつ.

本稿では、まず FN 法における欠点を分析する. つぎに、FN 法の欠点を補うべく考案した前処理付き FN 法を述べ、実例によりこの方法の有効性を示す.

2 階層的クラスタ分類法=FN 法の問題点

階層的クラスタ法とは、要素間の距離 (一般には類似度でよい) に基づいて、似ている要素を集めながら次第に大きなクラスタを作り、要素をいくつかのクラスタに分類する方法である. クラスタが作られていく様子は樹状図に表される.

クラスタ間の距離を測る一つの方法である FN 法は、2つのクラスタのそれぞれから取り出したすべての要素の組合せの中で、距離が最大のものをこれらのクラスタ間の距離とする方法である. FN 法の利点は、樹上図をクラスタ間距離 ϵ で切断した時点で、クラスタに含まれる要素間の最大距離、すなわちクラスタの“直径”が、必ず ϵ 以下になることである.

FN 法における問題点の分析を行うために、6つの要素 a, b, c, d, e, f だけからなるごく小さい集合のクラスタ分類を考える. 要素間の距離を表 1 に示すものとしよう. 表 1 をもとに FN 法を行うと、図 1 の樹状図を得る. この樹状図をクラスタ間距離 $\epsilon = 40$ で切断すると、二つのクラスタ G_1, G_2 が得られる.

図 2 は、最短距離の関係を示した図である. ①→②は、要素 a から見て最短距離にある要素は b であることを意味している.

図 2 および表 1 から、c は b からみて最短距離の関係にあり、その距離 17 も全要素間の距離の中で 2 番目に小さい値である. しかし、b と c はそれぞれ別のクラスタ G_1, G_2 に分類されており、いかにも不自然である.

もし仮に、図 2 の G'_1, G'_2 というクラスタ構成を考えるならば、 G'_1 と G'_2 の要素間の最短距離は 25 となり、クラスタの分離が良くなる. しかも、クラスタ G'_1, G'_2 の直径はいずれも ϵ 以下である.

このように FN 法は、クラスタの直径を一定値以下におさえることが出来るが、初期クラスタの要素の組合せによっては、距離の近い要素が別のクラスタに分類されてしまい、その結果クラスタの分離が非常に悪くなる場合がある.

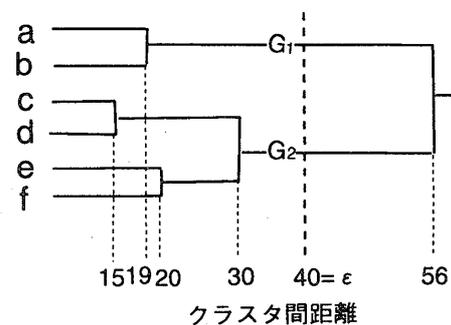


図 1: FN 法を用いたときの樹状図

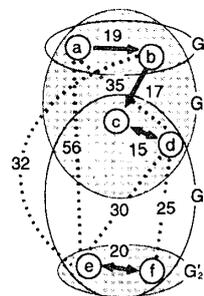


図 2: 要素間の最短距離の関係

表 1: 要素間の距離

	a	b	c	d	e	f
a	0	19	27	35	56	51
b	19	0	17	26	32	40
c	27	17	0	15	28	27
d	35	26	15	0	30	25
e	56	32	28	30	0	20
f	51	40	27	25	20	0

3 前処理付き FN 法=ND-FN 法

FN 法の利点を生かし、欠点であるクラスタの分離の悪さを改善するために、初期クラスタをうまく構成することを考える.

ある要素からみて最短距離にある要素が、別のクラ

スタに分類されることを防ぐために、以下のアルゴリズムにより初期クラスタを作る。

- (1) ある要素 a からみて最短距離にある要素 b を 1 つだけ選び、 $a \rightarrow b$ のように a と b を有向辺で結ぶ。このとき、最短距離が $\delta (> 0)$ 以下であることを要請する。もし、最短距離が δ より大きい場合は、辺は結ばない。(図3参照。) この操作をすべての要素について行うと森 (=木の集合) ができる。(証明略。) 各木を仮の初期クラスタとする。
- (2) 仮の初期クラスタのすべての直径を計算する。直径が $\epsilon (> \delta)$ を越える場合は、直径を構成する要素 a, b を結ぶ経路 (path) の中で、距離が最大である辺を削除し、そのクラスタ 2 分割する。(図4参照。) このようにして、すべてのクラスタの直径が ϵ 以下となるまで (2) の操作を繰り返す。

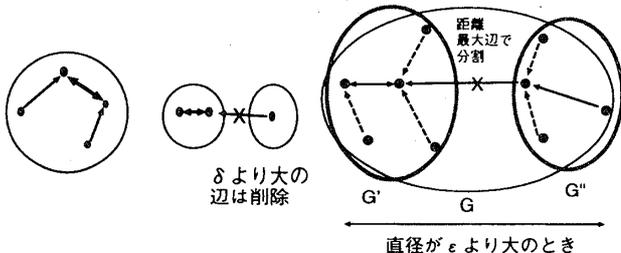


図3: 最短距離の関係よって作られたクラスタ

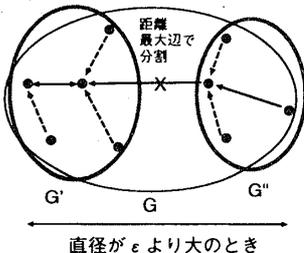


図4: 直径 $> \epsilon$ になるクラスタを 2 つに分ける方法

こうして作られたクラスタを初期クラスタとし、以後通常の FN 法を行う。この方法を ND(Nearest Distance)-FN 法と呼ぶことにする。

4 ND-FN 法を用いた分類結果の比較

単純で明快な色彩的特徴をもつ 3 種類のデザイン画群、

- (AF) アフリカン・エスニック・デザイン [5] 50 点,
- (FL) フローラル・パターン [6] 118 点,
- (SH) 日本の文様 [7] 213 点

について、FN 法と ND-FN 法でクラスタ分類を行い、結果を比較する。分類結果におけるクラスタ間の分離の良し悪は、クラスタ間の“間隔”，すなわち 2 つのクラスタのそれぞれの要素の組合せの中で最小になる距離、で評価することにする。

クラスタ分類を行う前に、まず、すべての画像をコンピュータに取り込み、逐次クラスタ化法 [8] を用いて代表色の抽出を行った。色空間は CIELUV を使い、代表色は、AF および SH については 10 色以内、FL については 10 色前後とした。この色数のとき、原画像と代表色画像はほとんど同じに見える。

つぎに、各デザイン画群ごとに、(代表色の) 色分布間の距離を文献 [4] に従い計算した。色分布間の距離の最小値、平均値、最大値は、表 2 の通りである。

表 2: 各デザイン画群における色分布間の距離

	AF	FL	SH
色分布間の距離	最小-平均-最大 9.1-37.7-103.4	最小-平均-最大 14.5-39.1-95.7	最小-平均-最大 6.8-38.7-119.0

各デザイン画群ごとに、色分布を要素とする集合に対し、FN 法と ND-FN 法を実行し、得られた樹状図をクラスタ間距離 40 で切断した。ND-FN 法におけるパラメタは、 $\delta = 20$, $\epsilon = 40$ とした。

分類結果のクラスタ直径およびクラスタ間隔の最小値、平均値、最大値を、表 3 および表 4 に示す。

表 3: FN 法による分類結果

	AF	FL	SH
クラスタ数	6	17	14
直径	最小-平均-最大 23.4-34.7-38.5	最小-平均-最大 23.8-32.7-39.4	最小-平均-最大 18.3-32.1-39.0
間隔	15.3-27.6-54.2	18.0-31.7-69.4	11.3-32.3-105.8

表 4: ND-FN 法による分類結果

	AF	FL	SH
クラスタ数	6	17	15
直径	最小-平均-最大 25.3-34.2-38.5	最小-平均-最大 26.6-33.5-39.4	最小-平均-最大 29.3-34.2-38.0
間隔	15.6-28.6-54.2	19.0-30.6-69.4	12.4-26.7-73.0

表 3 および表 4 の結果を比べると、どのデザイン画群についても、最小クラスタ間隔は、予想されたとおり、ND-FN 法のほうが FN 法より大きい。すなわち、クラスタ間の分離が良くなった。

5 おわりに

階層的クラスタ分類法の一つである FN 法の問題点を示し、それに対する解決策の一つとして ND-FN 法という前処理付き FN 法を考案した。3 つのデザイン画群の色構成に関する実験により、ND-FN 法の有効性を示した。

参考文献

- [1] KOBAYASI, Mituo., SUZUKI, Takuzi.: Quantitative Analysis of Color Features in Paintings — Classification based on Distance of Color Distribution in Color Space —, *Proceedings of AIC Color 97 Kyoto*, 1997.
- [2] 小林光夫, 鈴木卓治: 色分布の類似性に基づく絵画の分類, *日本色彩学会誌*, Vol. 21, supplement, pp. 84-85, 1997.
- [3] 小林光夫, 根本優美子, 鈴木卓治: 「アフリカン・エスニック・デザイン」における配色パターンのコンピュータによる分析, *日本色彩学会誌*, Vol. 21, supplement, pp. 86-87, 1997.
- [4] 小林光夫, 根本優美子, 鈴木卓治: 離散分布間の距離とその色彩分析への応用, *カラーフォーラム JAPAN'98 論文集*, pp. 137-140, 1998.
- [5] 福田明男: *アフリカ・エスニック・デザイン*, 柏書房, 1992.
- [6] E.A.Seguy: *Exotic Floral Patterns in Color*, Dover, 1974.
- [7] さくらほりきり: *日本の文様*, クレオ, 1995.
- [8] 小林光夫, 鈴木卓治: 逐次クラスタ化法を用いた画像の色彩分析, *日本色彩学会誌*, Vol. 19, supplement, pp. 72-73, 1995.