多変量データ系列における規則性を発見するための可視化手法

斉藤康彦†

大量のデータに基づく意思決定の過程では，生データを特定の視点から集約して可視化するために，各種のグラフが用いられる．しかし，同一の種類のグラフがいくつか存在する場合に，それらの間の関連を把握しながら，データの中に埋もれた規則性を発見することは，必ずしも容易でない．そこで，多変量データ系列における規則性の発見を支援するための可視化手法を提案する．多変量データ系列とは，同一の時間軸上で動く複数の折れ線を含む折れ線グラフや，時間的な順序にしたがって並べた帯グラフなどによって表現されるデータである．本手法では，多様な色相の多数の画素が織り成すテクスチャの系列として，多変量データ系列を表現する．これによって，視覚的印象に基づいて，類似するグラフのグループや例外的なグラフが検出できるようになる．さらに，どのようなグラフが系列中のどこに現れるかについての，大まかな傾向が把握できるようになる．本論文では，本手法によって可視化した日本の株式市場の変動から，日経平均株価の変動を予測するための規則性を導出する．また，ポートフォリオ管理に本手法が適用できることを示す．

A Visualization Technique for Extracting Rules from Series of Multivariate Data

YASUHIKO SAITO†

To visualize summaries of large amounts of data, decision-makers use various graphs. However, conventional graphs are not helpful, when one attempts to understand relations among a number of graphs to extract rules from data. This paper proposes a visualization technique for extracting rules from series of multivariate data. An example of a series of multivariate data is a series of pie charts arranged in chronological order. In the proposed technique, a series of multivariate data is represented as a series of textures composed of numerous colored points. The technique is useful in detecting groups of graphs similar to each other or groups of exceptional graphs. Furthermore, it allows one to perceive the macroscopic pattern of the arrangement of the graphs. A rule for predicting fluctuations of the Nikkei average is extracted from a series of textures for visualizing the Japanese stock market. In addition, an application of the technique to stock portfolio management is presented.

1. はじめに

人間の社会的活動の結果として発生する大量のデータに基づいて，迅速で的確な意思決定を行うための技術として，蓄積されたデータの中から（半）自動的に有用な規則性を発見するためのデータマイニングとともに，さまざまな視点から対話的にデータを分析するためのOLAP（On-Line Analytical Processing）が注目されている[2]．OLAPでは，棒グラフ，折れ線グラフ，帯グラフ，円グラフ，レーダーチャートなどを用いて，生データを特定の視点から集約した結果を，視覚的に分かりやすく表現する．

これらのグラフは，データの分布や変化を直観的に把握する上では，確かに有効であるが，同一の種類のグラフがいくつか存在する場合に，それらの間の関連を把握する上では，それほど有効でない．たとえば，地域別の売上構成比率の変動の規則性を発見する方法としては，各月の地域別の売上構成比率を表現する帯グラフを，月の順に並べることが考えられるが，これらの帯グラフの時系列から変動の規則性を発見することは，必ずしも容易でない．ある特定の地域のみを対象とするのであれば，その地域に対応する帯の部分を追跡していくことによって，何らかの規則性が発見できるかもしれない．しかし，全地域を対象とする場合には，たとえば，各地域に対応する帯の部分を個別に追跡した後に，その結果を総合して，全体としての規則性を導出する必要がある．

本論文では，多変量データ系列を可視化する手法である攪拌凝縮法[7]を提案する．本手法の目的は，多

† (株) アイネス システムリサーチセンター
Systems Research Center, INES Corp.

変量データ系列における規則性を発見しやすくすることである．多変量データ系列とは，同一の時間軸上で動く複数の折れ線を含む折れ線グラフや，時間的な順序にしたがって並べた帯グラフなどによって表現されるデータである．たとえば，各月の地域別の売上構成比率を月の順に並べたもの，あるいは，度数分布の時系列は，多変量データ系列である．

多変量データ系列における規則性を発見する場合には，多変量解析の手法を用いることができる．たとえば，地域別の売上構成比率の時系列から変動の規則性を発見する場合には，各地域の売上構成比率が変量になる．月 i における地域 j の売上構成比率を x_{ij} とすると，地域の数が n であるならば，月 i における地域別の売上構成比率の分布は，組 $(x_{i1}, x_{i2}, \cdots, x_{in})$ で表される．これは多変量データである．このような多変量データが月の数だけ存在する．したがって，たとえば，クラスタ分析[1]を用いて，これらの月をグルーピングすることによって，何らかの規則性が発見できるかもしれない．

しかし，クラスタ分析を用いても，現実の問題に適合した解が得られるとは限らない．なぜならば，クラスタ分析は，データ間の距離のみに基づくものであり，しかも，その距離をどのように定義したか，あるいは，多くのアルゴリズムのうちのどれを用いたかに応じて，可能な解のうちのひとつのみしか得られないからである．したがって，実際の分析の局面では，何通りかの可能な解を求めた後に，距離以外の何らかの情報に基づいて，最も適切であると考えられる解を選ぶ必要がある．

このような理由から，実際の分析の局面では，多変量解析のような統計解析の手法のみを用いるのではなく，分析者が多変量データそのものを直観的に理解することができるように，多変量データを可視化することが有効であると考えられる．そのような可視化手法は，これまでに数多く提案されている[10]．本論文で提案する攪拌凝縮法も，多変量データの可視化手法のひとつであるが，多数のグラフによって表現されるデータの系列を一覧し，視覚的印象に基づいて，類似するグラフのグループや例外的なグラフが検出できること，さらに，どのようなグラフが系列中のどこに現れるかについて，大まかな傾向が把握できることが，本手法の特徴になっている．

本論文の第2章では，攪拌凝縮法の基本的な概念と適用上の留意事項を示す．第3章では，本手法の具体的な使い方を，簡単な例題を用いて説明する．第4章では，株式市場の変動を可視化することによって，本手法の有効性を確認する．第5章では，これまでに提案された，画素やアイコンを規則的，かつ，稠密に配置した画像に基づいて，多変量データの傾向を把握する手法と，本手法との関係を明らかにする．

2. 攪拌凝縮法

本章では，攪拌凝縮法の基本的な概念と適用上の留意事項を示す．

2.1 概　　要

本手法は，分析の対象とする多変量データ系列に対して，攪拌凝縮，および，差分攪拌凝縮という操作を適用することによって，乱点図の系列である乱点図列を生成する．乱点図には，攪拌凝縮によって生成される全体乱点図と，差分攪拌凝縮によって生成される差分乱点図がある．また，並行乱点図，乱点地図，乱点織は，複数の乱点図を組み合わせたものである．

2.2 前　　提

本手法では，多変量データ系列と写像 h が前提として与えられる．

多変量データ系列とは，1変量データ分布の系列である．1変量データ分布とは，複数の1変量データを組にしたものである．1変量データとは，その分析で注目している特性が1種類のみのデータである．すなわち，変量 x によって特徴付けられるデータが n 個あるならば，組 $(x_1, x_2, \cdots, x_n)$ は，1変量データ分布である．

多変量データ系列を構成する1変量データ分布は，カテゴリの集合 C，値の集合 V，C から V への写像 v の組 $d = (C, V, v)$ として定義される．ただし，V の要素は，正の整数である．たとえば，ある月の地域別の売上構成比率の分布では，C の要素は地域であり，V の要素は（正の整数で表された）売上構成比率である．多変量データ系列は，i $(i = 1, 2, \cdots)$ によって順序付けられた，複数の1変量データ分布 d_i の系列として定義される．同一の系列を構成する d_i の間では，C と V が共通である．すなわち，i 番目の1変量データ分布は，組 $d_i = (C, V, v_i)$ として定義される．

h は，C から色相の集合 H への写像であり，マンセル系やPCCS（Practical Color Coordinates System）における色相環[3]を，C の要素の数，すなわち，カテゴリの数のセグメントに分割し，各カテゴリに各セグメントを代表する色相を対応させるものである．このとき，色相間の感覚距離がなるべく均等，かつ，最大になるようにする．

2.3 攪拌凝縮

攪拌凝縮とは，1変量データ分布の各カテゴリに対

```
INPUT : C, v, h, width, height, α
OUTPUT : P
for each c ∈ C {
    counter := 1.
    while counter ≤ v(c) × α {
        create p.
        p.hue := h(c).
        put p to W.
        counter := counter + 1.
    }
}
y := 1.
while y ≤ height {
    x := 1.
    while x ≤ width {
        get randomly p from W.
        p.x := x.
        p.y := y.
        put p to P.
        x := x + 1.
    }
    y := y + 1.
}
```

図 1 攪拌凝縮
Fig. 1 Mixing up colored points.

応する数値に比例する個数の点を，当該カテゴリに対応する色相で色付けして，ランダム，かつ，稠密に矩形領域に配置する操作である．なお，ここでの点とは，幾何学における点とは異なり，微小な面積を有するものとする．したがって，色付けすることができる．乱点集合とは，本操作によって得られた，色相と x 座標と y 座標の組の集合である．

本操作の手順を**図 1** に示す．*width* と *height* は，矩形領域の幅と高さである．α は，$\sum_{c \in C}(v(c) \times \alpha) \geq width \times height$ を満足する最小の自然数である．p は，乱点集合の要素となる組であり，$p.hue$，$p.x$，$p.y$ は，組 p の色相，x 座標，y 座標を示す．create *element* は，組 *element* を新しく生成する．put *element* to *set* は，要素 *element* を集合 *set* に追加する．get randomly *element* from *set* は，要素 *element* を集合 *set* からランダムに抽出する．抽出された *element* は，*set* から削除される．

2.4 全体乱点図

全体乱点図とは，乱点集合に属する組 (hue, x, y) のすべてについて，色相 *hue* の点を座標 (x, y) に描画したものである．全体乱点図列とは，i によって順序付けられた，複数の全体乱点図 g_i の系列である．g_i は，多変量データ系列を構成する 1 変量データ分布 d_i についての全体乱点図である．

図 2 に示すように，攪拌凝縮は，帯グラフから全体乱点図への変換である．帯グラフは，各カテゴリに対

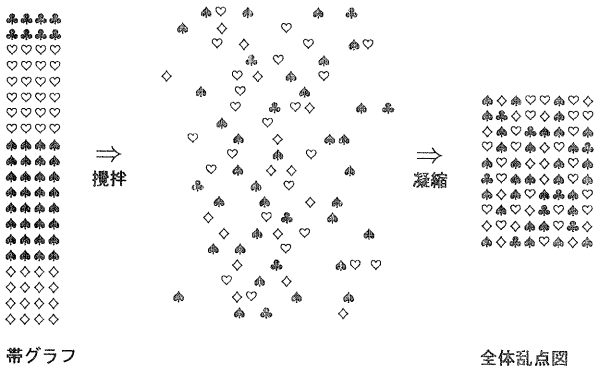


図 2 帯グラフから全体乱点図への変換
Fig. 2 Conversion of a band into a texture.

応する 4 色の画素（♣，♡，♠，◇）によって構成されている．全体乱点図は，これらの画素の集合体を，異なる色相の画素が混ざり合うように，十分に攪拌した後に，再び凝縮したものであるといえる．

全体乱点図は，多様な色相の多数の画素が織り成すテクスチャ（色合いと模様）である．帯グラフでは，カテゴリの数だけ存在する描画図形の大きさに本質的な意味があるのに対して，全体乱点図では，ただひとつの描画図形のテクスチャに本質的な意味がある．したがって，前者では，独立した複数の部分（描画図形の大きさ）から分布全体の様子を導出することになるので，「各部分を認識した上で，それらを総合する」という過程が介在するが，後者では，そのような過程を介在させることなく，単一の視覚的パターン（描画図形のテクスチャ）として，分布全体の様子が直観的に把握できる．さらに，テクスチャが識別できればよいので，表示に必要な面積が，帯グラフよりも小さくなる．したがって，より多くの分布についての全体乱点図を一度に表示することによって，一覧性が向上する．

これらの特徴のために，複数の分布が類似しているかどうか，あるいは，複数の分布がどのように違っているかを，分析者が全体乱点図列を一覧しただけで，それらのテクスチャから受ける視覚的印象に基づいて，判断することができる．すなわち，視覚的印象の類似の度合いが，分布間の差異の大きさを示し，そのときの色合いの違いが，分布間の差異の内容を示す．

2.5 差分攪拌凝縮

差分攪拌凝縮とは，まず，ふたつの乱点集合 P_a と P_b において，P_b に属する組の色相と同じ色相の組を，P_a から削除し，次に，これによって得られた集合に属する組を，その組の色相で色付けした点とみなし，これらの点をランダム，かつ，稠密に配置する操作であ

```
INPUT : Pa, Pb, width, height
OUTPUT : S(Pa, Pb)
for each pb ∈ Pb {
    if pa ∈ Pa and pa.hue = pb.hue, then
        delete pa from Pa.
}
y := 1.
while y ≤ height {
    x := 1.
    while x ≤ width {
        if Pa = φ, then stop.
        get randomly pa from Pa.
        pa.x := x.
        pa.y := y.
        put pa to S(Pa, Pb).
        x := x + 1.
    }
    y := y + 1.
}
```

図3 差分攪拌凝縮

Fig. 3 Mixing up the difference between two sets of colored points.

る．このとき，P_a を目的乱点集合といい，P_b を基準乱点集合という．差分乱点集合とは，本操作によって得られた，色相と x 座標と y 座標の組の集合である．

本操作の手順を**図3**に示す．$S(P_a, P_b)$ は，P_a を目的乱点集合とし，P_b を基準乱点集合とする差分乱点集合である．delete *element* from *set* は，要素 *element* を集合 *set* から削除する．

2.6 差分乱点図

差分乱点図とは，差分乱点集合に属する組 (hue, x, y) のすべてについて，色相 hue の点を座標 (x, y) に描画したものである．差分乱点図列とは，i によって順序付けられた，複数の差分乱点図 g_i の系列である．g_i は，多変量データ系列を構成する1変量データ分布 d_i の乱点集合を目的乱点集合とする差分乱点図である．また，このときの基準乱点集合は，すべての g_i について共通である．

差分乱点図も，全体乱点図と同じように，描画図形のテクスチャに本質的な意味がある．さらに，目的乱点集合の基準乱点集合に対する差異の大きさが，描画図形の大きさとしても示される．差分乱点図では，分布間の差異そのものが可視化されるので，より明確に差異の大きさと内容を把握することができる．

通常は，全体乱点図列の中から，適当な全体乱点図を選び，その乱点集合を基準乱点集合にするが，一様乱点集合を基準乱点集合にすることもできる．一様乱点集合とは，仮想的な一様分布についての乱点集合である．1変量データ分布 (C, V, v_i) の系列を分析する場合，仮想的な一様分布は，組 (C, UNI, uni) として定義される．ここで，uni は，C を定義域とし，singleton である UNI を値域とする写像である．一様乱点集合による差分乱点図では，構成比率が高いカテゴリのみについての分布の様子が示される．これによって，全体乱点図のテクスチャから受ける視覚的印象が強調される効果が生まれる．すなわち，この方法は，全体乱点図列において明確に認識できない分布間の差異を，明確に認識できるようにするためのものである．

2.7 並行乱点図

並行乱点図とは，2種類の多変量データ系列を構成する1変量データ分布 d_i と d'_i が，1対1に対応する場合に，d_i と d'_i についての全体乱点図 g_i と g'_i を，上下段に隣接して配置したものである．並行乱点図列とは，このような並行乱点図の系列である．

並行乱点図では，d_i と d'_i の分布の様子の論理積が，上下段のテクスチャの組み合わせによる視覚的パターンとして，一瞥しただけで把握できる．これによって，d_a と d_b が類似し，かつ，d'_a と d'_b が類似する a と b や，d_a と d_b は類似するが，d'_a と d'_b は類似しない a と b などが検出しやすくなる．また，多変量データ系列の間の相関を調べることができる．

2.8 乱点地図

乱点地図とは，多変量データ系列を構成する1変量データ分布が，地図上の領域（都道府県，市町村など）に対応付けられている場合に，各1変量データ分布についての全体乱点図のテクスチャを，地図上の対応する領域に貼り付けたものである．

地図と関連付けられたデータを扱う場合，従来は，棒グラフや円グラフなどを，地図上の対応する領域に貼り付けていた．しかし，これらのグラフは，地図に馴染みにくく，地図による表現の利点が損なわれることがあった．テクスチャは，これらのグラフよりも地図に馴染みやすく，特に，複数の地図における複数の領域の間の関係が把握しやすくなる．

2.9 乱点織

乱点織とは，乱点図列を構成する全体乱点図，または，並行乱点図を，上下左右の間隔を空けずに配置することによって得られる，抽象的なタペストリのような表現である．基本的には，矩形の乱点図を縦横に敷き詰めるように配置するが，それぞれの位置は，分析の目的に応じて決める．たとえば，多変量データ系列を構成する1変量データ分布が，日次のものであるならば，カレンダのように曜日を揃えて配置することもできるし，1月分を1行に配置することもできる．

乱点織は，多数の分布から構成される多変量データ系列を鳥瞰するためのものである．したがって，乱点織を構成する各乱点図を小さくして，一覧できる程度

の空間に多数の乱点図を表示する必要がある．その結果，乱点図のテクスチャの違いを認識することが難しくなる．しかし，乱点織では，テクスチャの違いを認識することよりも，各乱点図を1枚のタイルとみなして，タイルを敷き詰めて描いた1枚の絵の図柄を認識することが重要である．

乱点図と乱点織の間の関係は，より小さな粒度の表現の集積によって，より大きな粒度の表現が生成される点で，色付けした点と乱点図の間の関係に似ている．しかし，前者においては，乱点図をどのように配置するかによって，乱点織の図柄が変わるが，後者においては，色付けした点をどのように配置しても，ランダムに配置する限り，乱点図のテクスチャ（厳密には，テクスチャから受ける視覚的印象）は変わらない．

2.10 適用上の留意事項

2.10.1 数量の大きさと内訳

本手法では，ある数量の内訳，すなわち，構成比率が可視化され，数量の大きさは無視される．したがって，1変量データ分布の中でも，特に，円グラフやレーダーチャートによって表現されるデータの分析に適している．また，系列を構成する複数の帯グラフは，帯の長さが同じであるものとして扱われる．数量の大きさにも注目したい場合には，たとえば，テクスチャの矩形領域の大きさによって数量の大きさを表現する，あるいは，折れ線グラフを併用する，などの方法がある．

2.10.2 非時系列データ

本手法は，時系列データの分析に適している．しかし，時系列データへの適用に限られる訳ではない．複数の対象を何らかの順序にしたがって並べた系列があり，各対象に1変量データ分布が対応付けられている場合には，本手法を適用することができる．特に，乱点地図では，地図上の領域に1変量データ分布が対応付けられている．さらに，順序を考慮しないで適当に配置した乱点図の集まりから，類似する乱点図のグループや例外的な乱点図を検出することができる．

2.10.3 カテゴリの数

1変量データ分布 (C, V, v_i) における C の要素の数は，人間が識別できる色相の数が上限となる．これは，その色相で描画された領域の大きさに依存する．本手法では，原理上，1画素の大きさで識別できることが要求されるが，その場合には，経験的に，たかだか10色相程度である．

地域別，部門別，担当者別などのように，カテゴリ間に距離が定義できない場合には，異なるカテゴリが明確に識別されなければならない．したがって，カテゴリの数の上限は，10程度である．

これに対して，連続的な量を分割して設定したカテゴリのように，カテゴリ間に順序関係が存在するならば，順序関係に基づいて，カテゴリ間に距離が定義できる．このとき，カテゴリの数が多くなった場合には，距離の近い複数のカテゴリを識別する必要がないことがある．たとえば，0歳～99歳の年齢を1歳刻みに100カテゴリに分割する場合，分析の目的によっては，隣接する年齢の識別が重要でなくなる．したがって，カテゴリの数を100にすることができる．ただし，実際に識別されるのは，1歳刻みの年齢ではなく，識別可能な10色相に対応する年齢層である．

2.10.4 変量のデータ型

1変量データ分布 (C, V, v_i) における V の要素は，正の整数でなければならない．小数点以下の桁がある正の数を扱う場合には，整数にするために，全カテゴリについて 10^n 倍した後に，必要に応じて，小数点以下を切り捨てる．

3. 例　　題

本章では，攪拌凝縮法の具体的な使い方を，簡単な例題を用いて説明する．ただし，並行乱点図と乱点織については，次章で言及する．

3.1 客層分析

ある小売店における来店者数の客層別分布の時系列データに対して，本手法を適用する．このデータは，性別年齢別に設定した10客層のそれぞれについて，ある年の1月16日から2月8日までの24日間にわたって，来店者数を日次に集計したものである．

このときの乱点図列を**図4**に示す．(a) の全体乱点図列では，平日（月～金）と休日（土，日）で，テクスチャから受ける視覚的印象が異なっている．平日については，月水金のグループと火木のグループが，ぼんやりと区別できる．これらのグループの区別は，(b) の一様乱点集合を基準乱点集合とする差分乱点図列において，より明確になる．これによって，「休日の分布は，平日の分布と異なる」，および，「平日には，類似する分布が交互に出現する」という規則性が導出される．この規則性は，取り扱う商品の品揃えや発注量などを決定する上で参考になる．

本手法は，詳細な分析の前段階で，大まかな傾向を把握するために用いられる．テクスチャの違いが生じる原因を探るには，より詳細な分析が必要になる．たとえば，月水金に相対的に多い客層を調べるためには，客層別の構成比率を表現した帯グラフなどを参照しなければならない．

また，規則性に違反する例外を検出することも，そ

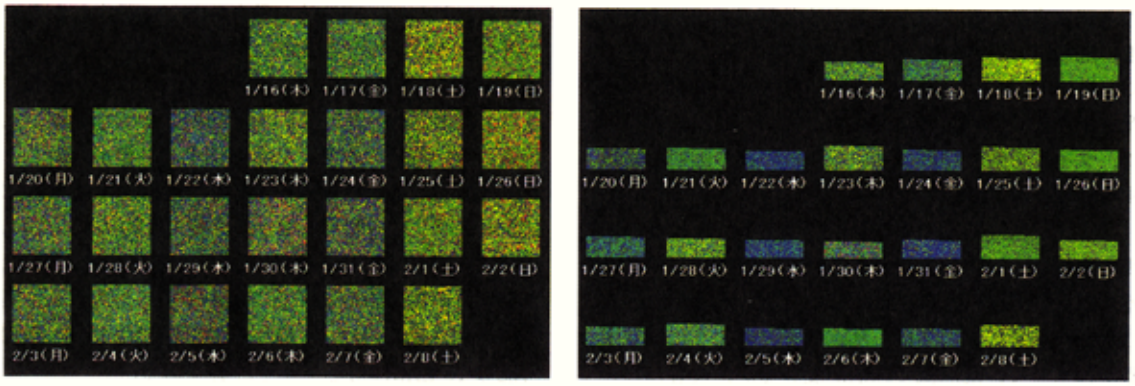


図 4 来店者数の客層別分布の乱点図列
Fig. 4 Series of textures for visualizing the distributions of customer types.

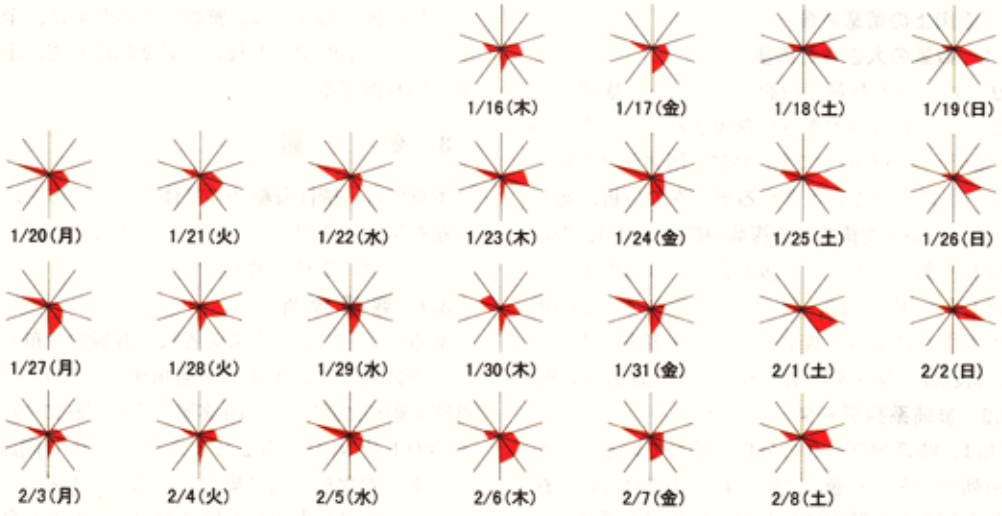


図 5 来店者数の客層別分布のレーダーチャート列
Fig. 5 A series of radar charts for visualizing the distributions of customer types.

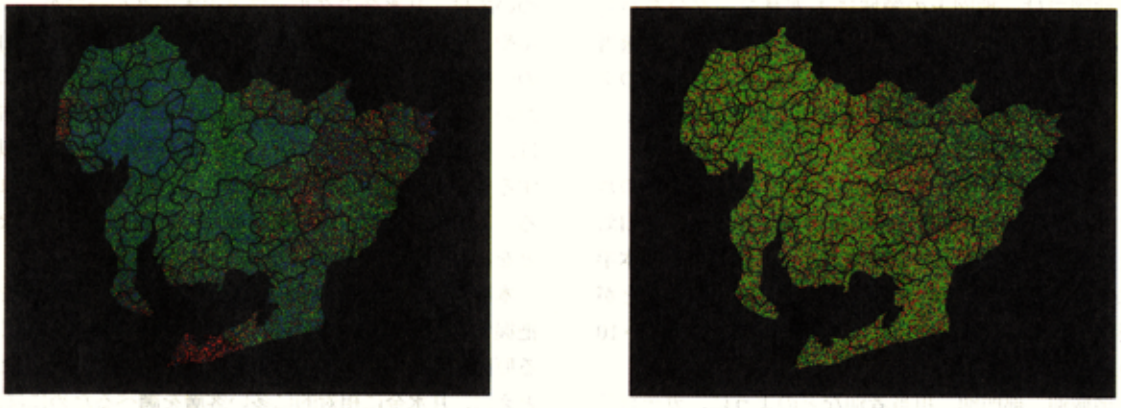


図 6 愛知県の人口構成の乱点織
Fig. 6 Maps of textures for visualizing the population compositions of Aichi Prefecture.

の後の詳細な分析における焦点を定める上で重要である．たとえば，1月17日（金）のテクスチャは，火木のグループに近い．したがって，例外ということになるが，このとき，規則性に違反する原因となる事象（「近所の学校が臨時休校であった」，「話題の新商品の発売日であった」，「雪が降った」など）を探ることは，次に続く分析の焦点になる．

ここで導出した規則性は，時間的な順序にしたがって並べた帯グラフやレーダーチャートによって，事後に確認することができる．しかし，これらのグラフの時系列から，この規則性を導出することは，必ずしも容易でない．**図5**は，同じデータをレーダーチャートで表現したものである．類似する多角形のグループを認識するには，まず，全多角形の特徴を把握し，記憶する必要がある．次に，いくつかのグループを暫定的に設けて，各多角形が属するグループを決めていくが，形が微妙に異なるものを同じグループとして扱うかどうかの判断は難しい．例外を含むことを認める場合には，さらに難しくなる．また，チャートの数が多くなった場合には，全多角形の特徴を把握することだけでも容易ではない．

3.2 地域人口分析

1995年の国勢調査に基づく各市町村の産業別就業人口と年齢別人口に対して，本手法を適用する．産業別就業人口は，第1次，第2次，第3次の構成比率である．年齢別人口は，年少人口（14歳以下），生産年齢人口（15歳以上64歳以下），老年人口（65歳以上）の構成比率である．

図6は，愛知県の人口構成の乱点地図である．(a)は，産業別就業人口で，第1次，第2次，第3次のそれぞれに，赤，緑，青を対応させている．(b)は，年齢別人口で，年少人口，生産年齢人口，老年人口のそれぞれに，赤，緑，青を対応させている．

(a)によると，渥美半島の先端は，赤が顕著であることから，第1次産業の就業人口が多い．それ以外では，東側の内陸も，かすかに赤を帯びている．この地域は，(b)によると，青を帯びていることから，老年人口が多い．したがって，第1次産業と老年人口の間の相関が認められる．これに対して，渥美半島の先端は，(b)によると，それほど青が強くないことから，第1次産業と老年人口の間には，弱い相関しか認められない．したがって，第1次産業の就業人口が多い地域の中でも，渥美半島の先端と東側の内陸では，地域性が異なると判断される．このように，複数の視点（産業別就業人口と年齢別人口）による複数の乱点地図から，地域性の違いを視覚的に把握することができる．

4. 株式市場の変動の可視化

本章では，株式市場の変動を可視化することによって，攪拌凝縮法の有効性を確認する[8]．

4.1 株式投資におけるリスク管理

将来の株価を的確に予測することは，株式投資におけるリスク管理の基本である．しかし，個別の企業の収益動向だけから，株価を予測することはできない．たとえば，金融政策や財政政策，為替相場の動き，貿易相手国の情勢などの社会的な事象に反応して，株価は変動するが，これらの事象は，個別の銘柄に対して直接的に影響を及ぼすばかりでなく，銘柄の集合としての市場に影響を及ぼした後に，その影響が個別の銘柄に及ぶ．したがって，株価を予測する上では，市場全体の状態を把握するための観測が不可欠である．そこで，本手法を用いた株式市場の観測を試みる．

本章では，日本の株式市場の変動を可視化したものから，日経平均株価の変動を予測するための規則性が導出できることを示す．また，ポートフォリオ管理[5]に本手法が適用できることを示す．ポートフォリオとは，複数の資産（ここでは，銘柄）の集合である．ポートフォリオ管理では，適切に選択した複数の資産に対して同時に投資することによって，リスクの分散を図る．

4.2 日足終値の変動による並行乱点図列

1988年～1997年の10年間に，東京証券取引所と大阪証券取引所で扱った全銘柄の日足終値に対して，本手法を適用する．日足終値とは，各立会日（証券取引所で取引が行われた日）の最後に取引された時点での株価である．データ件数は，約4,900,000件（銘柄の総数の平均と立会日の総数の積に相当する）である．

次のような並行乱点図列を用いて，株式市場の変動を可視化する．

上段は，次のような1変量データ分布 (C, V, v_i) についての全体乱点図である．C は，日足終値の変動率の絶対値の集合である．V は，銘柄の数の集合である．v_i は，変動率の絶対値 c に，立会日 i における変動率の絶対値 $= c$ である銘柄の数を対応させる写像である．これによって，変動率の絶対値に基づく銘柄の度数分布が得られる．各銘柄の変動率は，

(当日の終値 − 前日の終値)/前日の終値

として計算する．C から色相の集合 H への写像 h は，色相環を360等分し，赤を0として，順次，0～359の値を割り当て，この値を各カテゴリに対応させる．すなわち，

変動率の絶対値 $\times 2 \times 10^3$

を色相とする．ところで，色相環では，色相が360に近づくことは，0に近づくことになるので，その結果として，変動率の絶対値が大きいものと小さいものが，色相の上で識別しにくくなる．そこで，変動率の絶対値が0.13を超える銘柄を無視することによって，色相が260以下になるようにする．

下段は，次のような1変量データ分布 (C', V', v'_i) についての全体乱点図である．C' は，日足終値の変動の方向を示すzero，plus，minusを要素とする集合である．V' は，銘柄の数の集合である．v'_i は，zero，plus，minusのそれぞれに，立会日 i における変動率が，0に等しい（横ばい），0より大きい（上昇），0より小さい（下降）銘柄の数を対応させる写像である．これによって，変動の方向に基づく銘柄の度数分布が得られる．また，対象を上段の銘柄と一致させるために，上段と同じように，変動率の絶対値が0.13を超える銘柄を無視する．C' から色相の集合 H' への写像 h' は，zero，plus，minusのそれぞれに，赤，緑，青を対応させる．

4.3 日経平均株価の変動の予測

並行乱点図列と日経平均株価の間の相関を調べることによって，以下の規則性を導出した．

並行乱点図列における特異点は，それと同じ方向に日経平均株価が変動する立会日の系列の後半に出現する．

特異点とは，以下の条件を同時に満足する立会日である．

(1) 変動率の絶対値が相対的に大きい銘柄が多い．
(2) 変動の方向がplusの銘柄が，minusの銘柄に対して著しく多い，または，変動の方向がminusの銘柄が，plusの銘柄に対して著しく多い．

これらの条件を満足するかどうかは，並行乱点図のテクスチャに基づいて，分析者が直観的に判断する．特異点における上段のテクスチャでは，赤がやや弱くなって，黄や緑が強くなる．同時に，下段のテクスチャでは，緑と青のどちらか一方が支配的になる．緑が支配的な特異点の方向はplusであり，青が支配的な特異点の方向はminusである．すなわち，plus方向への特異点と，minus方向への特異点がある．これに対して，特異点でない立会日である平常点におけるテクスチャは，上段では，赤が強く，下段では，赤と緑と青が混在した中間的な色合いになる．特異点は，このような平常点の並びの中の例外として検出される．なお，今回の分析では，以下の要領で特異点を検出した．

(1) 古い時刻から新しい時刻に向かって，テクスチャをチェックしていき，チェックしている現在よりも約1月前からのテクスチャの並びに対する例外として，特異点を検出した．
(2) どのようなテクスチャの並びの後に出現するかによって，特異点であるかどうかの判断が変わることがある．テクスチャの変化が小さい並びの後では，特異点としての特徴がそれほど顕著でない場合にも，特異点として検出した．逆に，テクスチャの変化が大きい並びの後では，特異点としての特徴が顕著である場合にのみ，特異点として検出した．
(3) 特異点としての特徴が認められるテクスチャが連続する場合には，最初のものだけを特異点として検出した．

図7は，1991年の並行乱点図列と日経平均株価の折れ線グラフの一部である．この中で，7月1日，7月10日，8月22日をplus方向への特異点として検出し，6月19日，7月3日，7月8日，8月12日，8月19日をminus方向への特異点として検出した．

導出した規則性を次のように利用することによって，日経平均株価の変動を予測することができる．今日の取引が終了した時点で描いた今日の並行乱点図によって，今日が特異点であると判断されるならば，ここ数日の日経平均株価の上昇（下降）の動きが，近日中に止まる．止まるタイミングは，上昇（下降）の動きが何日間連続したかに依存する．昨日から上昇（下降）を始めた場合には，明日に止まり，一昨日から上昇（下降）を始めた場合には，明日か明後日に止まる．すなわち，n 日前から上昇（下降）を始めた場合には，n 日後までに止まる．

この規則性は，検出した74日の特異点のうち，約80%に相当する60日について成立する．しかし，分析の対象とした2,484日の立会日に対して，特異点の数が少なすぎるので，この規則性だけから，日経平均株価の変動を予測することは現実的でない．株価の予測では，現在，ファンダメンタル分析やテクニカル分析の各種の手法が用いられているが[5]，これらの手法と併用する必要があるといえる．なお，ファンダメンタル分析では，企業，産業，経済などの基礎的な諸条件に基づいて，株式の本来の価値を決定し，テクニカル分析では，株式の価格や数量などに基づいて，主に売買のタイミングを決定する．

4.4 バブル崩壊前後の日本の株式市場

図8は，1988年～1997年の並行乱点図による乱点織である．ここでは，上から下に月を順に並べ，左から右に各月の立会日を順に並べている．これによって，バブル崩壊前後の株式市場の長期的な変動を視覚

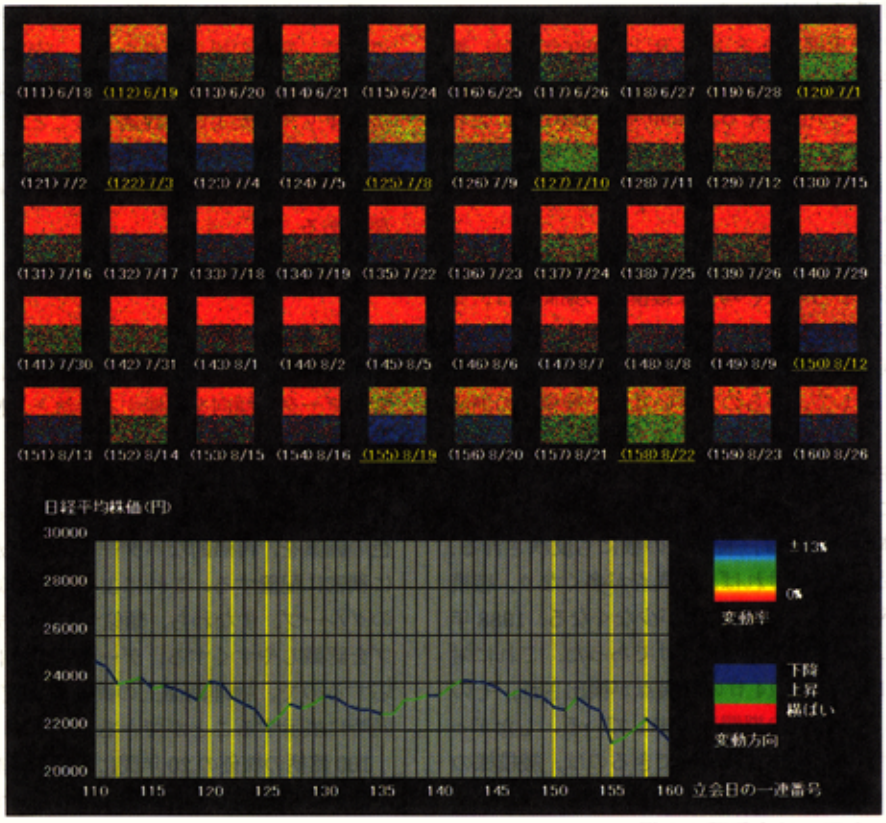


図 7 日本の株式市場の並行乱点図列と日経平均株価の折れ線グラフ
Fig. 7 A series of concurrent textures for visualizing the Japanese stock market and a line graph of the Nikkei average.

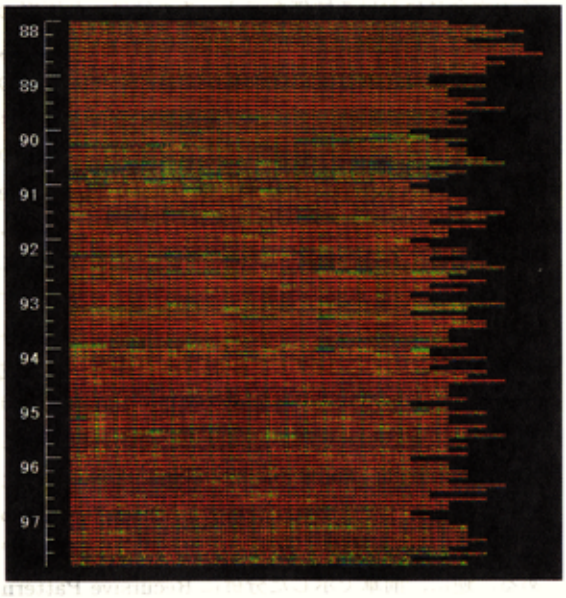


図 8 株式市場の乱点織
Fig. 8 A tapestry of concurrent textures for visualizing the Japanese stock market.

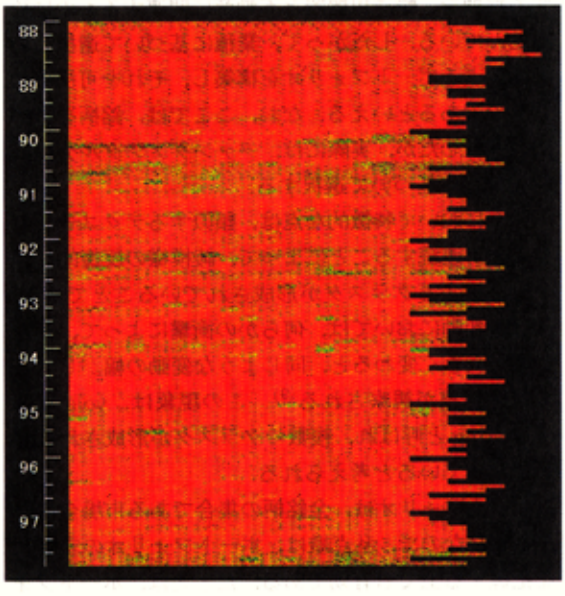


図 9 ポートフォリオの乱点織
Fig. 9 A tapestry of textures for visualizing a stock portfolio.

的に把握することができる．

バブルの絶頂期である1988年と1989年には，綾模様にむらがないが，バブルの崩壊が始まる1990年以降には，むらが出てくる．1990年と1991年に見られた大きなむらは，1996年頃までは，次第に縮小していく傾向にあるが，1997年末には，再び大きくなる兆しがある．

このような綾模様の変化は，市場全体の安定性を反映している．むらのない均一な綾模様の期間には，売りと買いの均衡が保たれている．したがって，市場の流動性が高く，リスクの小さい堅実な投資に適している．これに対して，むらの大きい乱れた綾模様の期間には，売りと買いのどちらかに偏ることによって，市場の流動性が阻害されることが多い．その結果，市場が不安定になり，リスクも大きくなる．

市場全体の安定性は，特異点の数にも反映される．綾模様の変化が大きければ，特異点が多くなり，綾模様の変化が小さければ，特異点が少なくなる．1988年に検出された特異点は，1日だけであった．これに対して，1991年には，14日の特異点が検出され，そのうちの8日が，図7の期間に集中している．したがって，前節で示した規則性が利用できるのは，綾模様の変化が大きい期間に限られる．

4.5 ポートフォリオ管理の支援

図9は，図8と同じ期間の並行乱点図の上段，すなわち，変動率の絶対値による乱点織であるが，対象とする銘柄を，輸送用機器と不動産に関連するものに絞り込んでいる．したがって，業種に基づいて選択した銘柄によるポートフォリオを構築し，それを可視化したものであるといえる．なお，ここでは，銘柄を機械的に選択したが，実際には，ファンダメンタル分析の結果などに基づいて選択する．

図9において特徴的な点は，類似するテクスチャの乱点図が連続することによって，綾模様の反復的なパターンによるクラスタが形成されていることである．株価の変動においては，何らかの衝撃によって，変動の幅が大きく変わると，同じような変動の幅がしばらく続く現象が観察される[9]．この現象は，volatility clusteringと呼ばれ，複数のクラスタが形成される原因になっていると考えられる．

ポートフォリオは，全銘柄の集合である市場全体の部分集合なので，乱点織は，ポートフォリオの安定性を把握する上でも有効である．たとえば，ポートフォリオを構成する銘柄を，水産・農林，食品，電力・ガスに関連するものに置き換えると，図9よりも，全体的に，黄や緑が後退して，さらに赤が強くなる．その結果，綾模様の現れ方が曖昧になることが確認された．これは，変動の幅がより小さく，安定性がより高いことを意味する．すなわち，綾模様の現れ方から，ポートフォリオの安定性を評価することができる．ポートフォリオの構築や運用に関する戦略を決定する場合には，このようなポートフォリオの安定性を考慮する必要がある．

5. 関連研究

本章では，これまでに提案された，画素やアイコンを規則的，かつ，稠密に配置した画像に基づいて，多変量データの傾向を把握する手法と，攪拌凝縮法との関係を明らかにする．

5.1 Recursive Pattern

Recursive Pattern[4]では，各データの値に対応する色の画素を，再帰的なパターンにしたがって，画面上に稠密に配置する．これによって，類似する色の画素のグループ，すなわち，類似する値のデータのグループを認識しやすくする．データと画素は，1対1に対応するので，物理画面上の画素の数だけのデータを，同時に表示することができる．

100銘柄の株価の20年間にわたる変動の様子を，本手法を用いて分析した事例が報告されている．この分析では，画素が再帰的なパターンにしたがって配置された矩形を，銘柄ごとに描く．すなわち，100個の矩形を描く．各画素は，立会日に対応し，画素の色は，その立会日における株価を示す．その結果，各矩形の内部には，時間軸に沿って色の縞が現れる．そこで，どの色の縞がどの位置に現れるかに基づいて，変動の様子が類似する銘柄のグループを認識することができる．前章で示した攪拌凝縮法による分析では，株価の変動率で銘柄を集計した結果を扱ったが，Recursive Patternによる分析では，株価をそのまま扱っている．攪拌凝縮法による分析で，株価をそのまま扱う場合には，たとえば，乱点図の矩形を，立会日ごとに描く．すなわち，20年間の立会日の総数だけの矩形を描く．各矩形の内部には，ランダムに配置された画素が織り成すテクスチャが現れる．各画素の色は，100銘柄の中のどれかに対応し，それが全体乱点図であれば，同じ色の画素の数は，対応する銘柄の株価に比例する．ただし，このような形で攪拌凝縮法を用いることによって，意味のある規則性が発見できるかどうかは疑問である．逆に，前章で示した分析にRecursive Patternを用いることは難しい．

5.2 Stick Figure Icon

Stick Figure Icon[6]では，折れ曲がった杖の形をし

た多数のアイコンを，画面上に稠密に配置する．杖は，複数の線分を繋ぎ合わせたものであり，繋ぎ合わせた箇所で，杖が折れ曲がる．データと杖は，1対1に対応するので，杖の数だけのデータを同時に表示することができる．杖を構成する各線分は，変量に対応し，線分の傾斜角度によって，その変量の値が示される．さらに，杖が配置される x 座標と y 座標によって，2変量の値が示される．分析者は，稠密に配置された多数の杖が織り成すテクスチャから，規則性や例外を表現するパターンを視覚的に検出することができる．

本手法は，杖が配置される座標によって，2変量の値が示されるので，2次元空間での位置情報を変量中に含むデータの分析に適している．このようなデータの分析には，乱点織も適している．乱点織では，杖の代わりに乱点図が用いられていると考えられる．

6. おわりに

本論文では，多変量データ系列の可視化手法を提案した．本手法では，多様な色相の多数の画素が織り成すテクスチャの系列として，多変量データ系列を表現する．本手法は，多変量データ系列における規則性を発見する上で有効である．株式投資におけるリスク管理に本手法を適用し，その有効性を確認した．

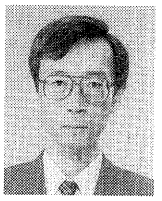
今後の課題は，本手法による分析を支援するシステムを使いやすくすることである．特に，カテゴリから色相への写像，および，乱点図の大きさや配置などのパラメータを，利用者が任意に指定できるインタフェースを提供するとともに，データベース管理システムやスプレッドシートとの連携を強化する必要がある．

参考文献

1) Anderberg,M.R.: *Cluster Analysis for Applications*, Academic Press (1973).
2) Berry,M.J.A. and Linoff,G.: *Data Mining Techniques: For Marketing, Sales, and Customer Support*, John Wiley & Sons (1997).
3) 金子隆芳: 色の科学, 朝倉書店 (1995).
4) Keim,D.A., Kriegel,H.-P. and Ankerst,M.: Recursive Pattern: A Technique for Visualizing Very Large Amounts of Data, *Proc. IEEE Visualization'95*, pp. 279-286 (1995).
5) 日本証券アナリスト協会 (編), 榊原茂樹, 青山護, 浅野幸弘 (著): 証券投資論［第3版］, 日本経済新聞社 (1998).
6) Pickett,R.M. and Grinstein,G.G.: Iconographic Displays for Visualizing Multidimensional Data, *Proc. IEEE Conf. Systems, Man, and Cybernetics*, pp. 514-519 (1988).
7) 斉藤康彦: 時系列データの束における変動の規則性や例外を発見するための可視化手法, 情報処理学会研究報告, 98-FI-52-2 (1998).
8) 斉藤康彦: 攪拌凝縮法による株式市場の変動の可視化, Computer Visualization Symposium'99 論文集, pp. 13-16, 日経サイエンス (1999).
9) 渡部敏明: 日本の株式市場におけるボラティリティの変動, 三菱経済研究所 (1997).
10) Wong,P.C. and Bergeron,R.D.: 30 Years of Multidimensional Multivariate Visualization, in Nielson,G.M., Hagen,H. and Müller,H.(eds.), *Scientific Visualization: Overviews, Methodologies, and Techniques*, pp.3-33, IEEE Computer Society Press (1997).

(平成1999年3月20日受付)
(平成1999年5月 6 日採録)

(担当編集委員 中谷多哉子)



斉藤 康彦（正会員）

1961年年生．1984年早稲田大学教育学部卒業．同年，(株) 協栄計算センター（現，(株) アイネス）入社．1991年より1995年まで，情報処理振興事業協会（IPA）にて，新ソフトウェア構造化モデルの研究開発に従事．現在は，(株) アイネスにて，情報可視化，および，ソフトウェア工学の研究に従事．博士（工学）．人工知能学会会員．