

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6602013号
(P6602013)

(45) 発行日 令和1年11月6日(2019.11.6)

(24) 登録日 令和1年10月18日(2019.10.18)

(51) Int.Cl. F 1
G 0 6 F 16/56 (2019.01)
G 0 6 F 16/54 (2019.01)

G O 6 F 16/56
G O 6 F 16/54

請求項の数 7 (全 16 頁)

(21) 出願番号 特願2015-6910 (P2015-6910)
(22) 出願日 平成27年1月16日(2015.1.16)
(65) 公開番号 特開2016-133899 (P2016-133899A)
(43) 公開日 平成28年7月25日(2016.7.25)
審査請求日 平成29年12月19日(2017.12.19)

(73) 特許権者 000001007
キヤノン株式会社
東京都大田区下丸子3丁目30番2号
(74) 代理人 100090273
弁理士 國分 孝悦
(72) 発明者 瀧本 将史
東京都大田区下丸子3丁目30番2号 キ
ヤノン株式会社内

審査官 鹿野 博嗣

最終頁に続く

(54) 【発明の名称】 情報処理装置、情報処理方法、及びプログラム

(57) 【特許請求の範囲】

【請求項 1】

画像中の所定個の特徴量間の類似度を算出する算出手段と、

前記所定個の特徴量の各々について、自身及び他の特徴量との前記類似度を要素とするベクトルを生成し、前記特徴量間の類似度に基づいて前記ベクトルに対応する点が原点から等距離に並ぶように当該ベクトルの次元数を削減する次元削減手段と、

前記特徴量間の類似度を基に、前記所定個の特徴量のそれぞれに対応する軸の配置を決定する形状決定手段と、

前記形状決定手段により決定した軸の配置に従って前記画像から抽出した特徴量を描画した図を出力する出力手段とを有することを特徴とする情報処理装置。

10

【請求項 2】

前記算出手段は、前記画像に含まれる各特徴量が示すデータの分布を基に前記特徴量間の類似度を算出することを特徴とする請求項 1 記載の情報処理装置。

【請求項 3】

前記形状決定手段は、前記所定個の特徴量のそれぞれに対応する軸を、前記特徴量間の類似度に応じた角度をなすように配置することを特徴とする請求項 1 又は 2 記載の情報処理装置。

【請求項 4】

前記所定個の特徴量のそれぞれに対応する軸の配置をユーザからの入力に応じて変更することを特徴とする請求項 1 ~ 3 の何れか 1 項に記載の情報処理装置。

20

【請求項 5】

前記特徴量間の類似度が所定の値よりも高い特徴量に対応する軸を 1 つにまとめて配置することを特徴とする請求項 1 ~ 4 の何れか 1 項に記載の情報処理装置。

【請求項 6】

画像中の所定個の特徴量間の類似度を算出する算出工程と、

前記所定個の特徴量の各々について、自身及び他の特徴量との前記類似度を要素とするベクトルを生成し、前記特徴量間の類似度に基づいて前記ベクトルに対応する点が原点から等距離に並ぶように当該ベクトルの次元数を削減する次元削減工程と、

前記特徴量間の類似度を基に、前記所定個の特徴量のそれぞれに対応する軸の配置を決定する形状決定工程と、

決定した軸の配置に従って前記画像から抽出した特徴量を描画した図を出力する出力工程とを有することを特徴とする情報処理方法。

10

【請求項 7】

画像中の所定個の特徴量間の類似度を算出する算出ステップと、

前記所定個の特徴量の各々について、自身及び他の特徴量との前記類似度を要素とするベクトルを生成し、前記特徴量間の類似度に基づいて前記ベクトルに対応する点が原点から等距離に並ぶように当該ベクトルの次元数を削減する次元削減ステップと、

前記特徴量間の類似度を基に、前記所定個の特徴量のそれぞれに対応する軸の配置を決定する形状決定ステップと、

決定した軸の配置に従って前記画像から抽出した特徴量を描画した図を出力する出力ステップとをコンピュータに実行させるためのプログラム。

20

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報処理装置、情報処理方法、及びプログラムに関する。

【背景技術】

【0002】

人がデータ解析を行う際、数字の羅列をそのまま眺めるのではなくデータを人が把握しやすい形態で表示するための視覚表現として多様なグラフ (graph) 又はチャート (chart) による表現方法を使うのが一般的である。対象となるデータや把握したい情報の性質によって適切に採用された可視化方法でデータを把握しやすくすることにより単なる数字の羅列では気付かないデータの傾向や情報を読み取ることが可能となる。ビッグデータの活用によって新しい問題解決方法や付加価値を提供することが昨今の技術領域のメインストリームになりつつある。例えば、多変量のデータを効果的に可視化して人の目による分析をサポートすることが重要となっている。

30

【0003】

一般的に、データを複数の項目で分析する場合、円グラフ (Pie Chart) や棒グラフ (Bar Chart、Bar Graph) 等が用いられる。また、レーダチャート (Radar Chart) や平行座標プロット (Parallel Coordinate Plot) 等が用いられる。これらのうち、棒グラフやレーダチャート、平行座標プロットは、多変量データの関係をそのまま知ることができる方法である。これら 3 種のグラフは、さまざまな評価に使用できるデータ可視化方法であり、一般的には多変量を比較するのに使用される。また、グラフ内の複数の項目同士を比較することもでき、同じ項目で評価値を出した他のグラフを比較することもできる。

40

【0004】

棒グラフや平行座標プロット、レーダチャートは、多変量解析に利用されるデータ可視化法である一方、変数の数の増加とともに軸数が増えてグラフ形状が複雑化していくため、人が一度に全体を把握することが困難となるということが知られている。認知神経科学で良く知られた事実として『マジック・ナンバー 7 ± 2 説』(非特許文献 1) というものがある。これは、人が情報を認識する際、7 ± 2 のチャンクに納まる情報であればうまく認識できるが、それ以上になると難しくなるという説である。この説によると、例え

50

ばグラフ形状の直感的な比較に優れているレーダチャートであっても、複数のグラフ形状比較をする際、隣り合う軸同士の意味的なまとまり（チャンク）のない一般的なレーダチャートの場合は各軸が1チャンクとなる。そのため、軸数が 7 ± 2 を超えたレーダチャートは人の形状認識能力を超えている可能性が高いことになる。

【0005】

下記特許文献1には、多変量データの可視化の際にデータを表現する特徴量の数の増加とともに横方向に伸びていく平行座標プロットにおいて、部分的な平行座標プロットを取り出して配置し、人の認識しやすいグラフを作成する技術が提案されている。また、下記非特許文献2には、レーダチャートの軸に割りつける変数の順番が重要であると説き、これらをPCAによって近くなったものをレーダチャート上での軸で近くなるように設定することが説明されている。

10

【先行技術文献】

【特許文献】

【0006】

【特許文献1】特開2013-161226号公報

【非特許文献】

【0007】

【非特許文献1】“The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information”, George A. Miller, The Psychological Review, 1956

20

【非特許文献2】『多変量解析事例集第1集』（吉澤・芳賀編，日科技連，1992）

【発明の概要】

【発明が解決しようとする課題】

【0008】

ここで、レーダチャート、平行座標プロット、棒グラフや棒グラフを折れ線表示したものすべてにおいて言える課題は、データ可視化を形状として把握しやすく可視化している傍ら、軸の相対的な位置や間隔（幅や角度）に情報価値がないことである。これらのグラフは、多変量の項目間を連結した線によって（棒グラフの場合は棒の並びによって）形状が構成されるにも関わらず、偶然隣り合った項目を結ぶことによって形状が確定する。そのため、可視化によって得られる情報において人の認識にとっては隣り合う軸同士の関係性を重点的に認識することになる。例えば、棒グラフでは遠くの棒同士の高さの差は近くの棒同士の差に比べて直感的な大小把握が難しくなり、平行座標プロットでは隣同士の変数間に相関があるかどうかしか判らない。よって、詳細にスコアの差を比較する必要のない項目を離したり、微小な差異を詳細に比較すべき項目を近くにしたりするといった、多変量データを効果的に可視化する際の工夫が必要となってくる。

30

【0009】

前記特許文献1に記載の平行座標プロットを部分的に取り出して配置する方法では、人が注目すべき部分が絞られるため、部分的な平行座標プロットに関して比較しやすくなる。しかし、グラフ全体を俯瞰してデータを把握するような場合には、多変量の部分的組合せを複数可視化して一つのグラフとするため、一瞥して情報を直感的に把握するのには不向きなグラフになったり、見落としが発生しやすくなったりする可能性がある。また、前記非特許文献2のレーダチャートの変数の順序決定方法によっても、近くにまとめるべき変数成分の見当がつかただけであり、それら変数を表す軸間の類似度等を反映したチャートを得ることができない。

40

【0010】

本発明は、このような事情に鑑みてなされたものであり、構成する軸同士の関連性や類似性が考慮された比較認識を容易に行うことが可能になるチャート（グラフ）を生成できるようにすることを目的とする。

【課題を解決するための手段】

【0011】

50

本発明に係る情報処理装置は、画像中の所定個の特徴量間の類似度を算出する算出手段と、前記所定個の特徴量の各々について、自身及び他の特徴量との前記類似度を要素とするベクトルを生成し、前記特徴量間の類似度に基づいて前記ベクトルに対応する点が原点から等距離に並ぶように当該ベクトルの次元数を削減する次元削減手段と、前記特徴量間の類似度を基に、前記所定個の特徴量のそれぞれに対応する軸の配置を決定する形状決定手段と、前記形状決定手段により決定した軸の配置に従って前記画像から抽出した特徴量を描画した図を出力する出力手段とを有することを特徴とする。

【発明の効果】

【0012】

本発明によれば、一瞥しただけでは把握しにくい多変量データであっても直感的に把握しやすい表現形態にて表示することができ、データ分析がしやすくなる効果が得られる。

【図面の簡単な説明】

【0013】

【図1】本実施形態における情報処理装置としての多変量データ可視化装置の構成例を示す図である。

【図2】本実施形態における処理動作の例を示すフローチャートである。

【図3】異常パターン毎にレーダチャートの形状が異なる例を示す図である。

【図4】異なる性質のデータの判別に不向きなレーダチャートの例を示す図である。

【図5】図4に示したレーダチャートを、異なる性質のデータを判別しやすく変換を施した例を示す図である。

【図6】第1の実施形態に係るレーダチャートの生成手順の例を示す図である。

【図7】第1の実施形態によるグラフの例を示す図である。

【図8】第1の実施形態に係るレーダチャートの生成手順の他の例を示す図である。

【図9】第1の実施形態に係るレーダチャートの生成手順の他の例を示す図である。

【図10】第1の実施形態に係る表示形式を選択可能なGUIの例を示す図である。

【図11】本実施例における情報処理装置を実現可能なコンピュータ機能を示す図である。

【発明を実施するための形態】

【0014】

以下、本発明の実施形態を図面に基づいて説明する。

【0015】

(第1の実施形態)

本発明の第1の実施形態について説明する。第1の実施形態は、複数の特徴量で表現される複数のデータが有り、各々のデータがそれらの中でどのような傾向にあるかを分析したい場合に、人が見て直感的に判り易い形でのレーダチャートやそれに類するチャートを生成する方法である。

【0016】

一例として、入力画像から複数の特徴量を抽出して自動で検査を行う外観検査でのユースケースを挙げて説明する。入力画像の正常・異常を自動で判定する外観検査では、入力画像から判定に必要な特徴量を複数抽出し、それらのスコア(評価値)を総合的に判断して正常画像か欠陥画像かを判定する。抽出される特徴量の示す各スコアと検出される異常の種類には相関があり、生産現場では判定結果である正常・異常のラベル以外にも判定スコアや元の抽出特徴軸におけるスコアにより測定の難しい欠陥の傾向を分析することができる。これによって、例えば生産ライン設計にフィードバックをかけることができる。

【0017】

特に、人の目による官能目視検査を自動化した外観検査装置では複雑なパターンかつ多様なレベルの異常データが存在するため、異常と判定した画像だけを見ても何を根拠に異常と判定したのか分析するのが難しい。そこで、抽出特徴を一覧性の高いチャート(グラフ)によって可視化することで、画像を見るだけではわかりにくかった傾向の把握を容易にすることができる。

10

20

30

40

50

【 0 0 1 8 】

図 3 に一例として、各々の入力画像から抽出した特徴量のスコアの傾向の違いを直感的に知ることができる例としてレーダチャートの形状の違いが異常パターン毎に異なる様子を示した。抽出特徴の数が 5 個である場合を示しており、各々の抽出特徴には I D が番号によって付与されている。図 3 (A) は、入力画像としての正常画像及びそれに対応するレーダチャートの例を示しており、図 3 (B) は、入力画像としてのムラ欠陥画像及びそれに対応するレーダチャートの例を示している。また、図 3 (C) は、入力画像としてのキズ欠陥画像及びそれに対応するレーダチャートの例を示しており、図 3 (D) は、入力画像としての異物欠陥画像及びそれに対応するレーダチャートの例を示している。図 3 (A) ~ 図 3 (D) に示すように、それぞれのレーダチャートの形状は異なっている。

10

【 0 0 1 9 】

このように、適切なレーダチャートが設定されると、ユーザがレーダチャートの形状を見るだけで欠陥の傾向を直感的に把握するのに役立つ。ここでは見易さのために各欠陥の事例は画像だけを見て判る入力画像を示したが、生産現場で発生する欠陥サンプルには人の目には正常サンプルとほとんど外観の変わらないようなものも存在する。そのような場合に、そのサンプルが何の異常の傾向を持つのか、といった確認や分析の際にもレーダチャートの形状を見るだけで直感的に理解することが可能となる。その他、こういった情報の可視化によって複雑な判定処理アルゴリズム自体をユーザが理解していなくとも、アルゴリズムがどのように画像を判定しているかを直感的にチェックすることができるため、検査装置の信頼度の向上や不具合の早期発見に役立つ。

20

【 0 0 2 0 】

しかし、図 3 に示したような抽出特徴の種類が 5 個程度と少なく、それぞれが各欠陥の傾向を把握するのに充分であるならば良いが、一般的に多様な異常パターンに対応することを前提とした識別器では抽出特徴の数は数十から数百まで増加することが多い。このような場合に何の工夫も無くレーダチャートを作ると、図 3 に示したように一瞥して異常の傾向を把握できるようなチャートとは異なるチャートが生成されることが多い。

【 0 0 2 1 】

例として、抽出特徴の数が 3 5 個である場合に、単純に抽出した特徴量を抽出特徴順に時計回りに軸を配しただけのレーダチャートを描画したことによって欠陥の傾向によるグラフの形状の違いを把握しにくくなる典型例を図 4 に示す。図 4 (A) ~ 図 4 (D) に示すレーダチャートは、それぞれ異なる理由で欠陥とされた検査対象の画像を入力としたときに生成されたグラフである。生産現場において 1 日に大量の検査対象がラインを流れていく中で人がこれら 4 つの形状の違いを一瞥して把握し、瞬時にそれぞれがどういった欠陥を意味しているのかを理解するのは困難である。

30

【 0 0 2 2 】

そこで、本実施形態では、グラフに対して以降で説明する処理を施すことによって直感的に形状の違いを把握しやすいグラフに変換する。図 5 (A) ~ 図 5 (D) に、図 4 (A) ~ 図 4 (D) を各軸が示すスコアはそのまま保存しながら、それぞれ認識しやすい形状になるように軸の順を並べ替えたレーダチャートを示す。これによると図 4 (A) ~ 図 4 (D) においては、不明瞭であった性質がよく判るようになる。例えば、図 5 (A)、図 5 (B)、図 5 (C) はそれぞれ異なる欠陥の種類を示す画像に対応するチャートであったこと、図 5 (D) は、図 5 (A) 及び図 5 (B) に示される 2 つの欠陥種が複合している可能性が高いこと等が即座に理解しやすい。

40

【 0 0 2 3 】

つまり、たとえ抽出特徴の数が増えたとしても、図 5 (A) ~ 図 5 (D) に示すように同じ欠陥信号に反応する特徴量を表す軸をそれぞれ近くに配置するだけで、形状を見るだけで欠陥の傾向を把握できるレーダチャートを作成することができるようになる。以下では、見やすいレーダチャート生成アルゴリズムの例を詳細に説明する。

【 0 0 2 4 】

前述したとおり、見やすいレーダチャートを生成するためには、関連性の高い軸を近く

50

に配置することが好ましい。決まった傾向の入力欠陥信号に対して決まった反応をする特徴量があれば、それらの軸が近くに配置されたレーダチャートでは該当の箇所がグラフの形状としていつも連動して変形する。そのため、ひとまとまりの領域に同じ意味付けをすることで数少ない部分領域を人は意識してグラフを読み取れば良いことになるためである。逆に、連動して反応する軸が散在したレーダチャートでは同じような意味を持つ特徴量の集合をまとめて認識することが人にとって困難となる。

【 0 0 2 5 】

ここで処理の流れに関して、図 1 の本実施形態における情報処理装置としての多変量データ可視化装置の構成例を示すブロック図、及び図 2 のフローチャートを用いて説明する。多変量データ入力部 1 0 1 にて、画像等の対象となるサンプルから n 個の特徴量が抽出されたデータが入力される (S 2 0 1)。そして、入力された n 個の特徴量を基に特徴量間距離算出部 1 0 2 にて各々の特徴量同士の類似度 (距離) を計算し (S 2 0 2)、得られた特徴量間の類似度を尺度とする特徴空間で表現された各特徴量を代表する n 個のベクトルが獲得される。

10

【 0 0 2 6 】

この n 個のベクトルを次元削減部 1 0 3 にて原点から等距離に点が並ぶように 2 次元 (又は 3 次元) に次元削減を行う (S 2 0 3)。このとき、データ可視化時に必要となる次元削減数やどの特徴量同士を近付けたい (又は遠ざけたい) 等のパラメータ類をユーザが指定したい場合、これらの変更を結果に反映させるため、データ可視化パラメータ入力部 1 0 4 から入力することも可能である。

20

【 0 0 2 7 】

そして、求まった原点から等距離に並んだ n 個の特徴量を代表するベクトルやユーザがデータ可視化パラメータ入力部 1 0 4 により入力したパラメータを基にしてチャート形状決定部 1 0 6 にてチャートの軸の配置が決定 (S 2 0 4) される (S 2 0 4)。この配置にならって多変量データ入力部 1 0 1 から入力されたデータのそれぞれが各軸毎に評価値をプロットされてチャートが完成し、ディスプレイ等の出力部 1 0 5 にて出力される (表示される)。

【 0 0 2 8 】

また、チャート形状決定部 1 0 6 にて決定したチャートの軸配置をチャート形状記憶部 1 0 8 で記憶しておく。そして、チャートの軸配置に使われなかった多変量データを多変量データ追加入力部 1 0 7 にて受付け、チャート形状記憶部 1 0 8 で保持されているチャートの軸の上にプロットすることで新たなチャートを作成して出力部 1 0 5 に出力する。

30

【 0 0 2 9 】

次に、図 2 に示したステップ S 2 0 2 での具体的な特徴量間の類似度 (距離) の算出例について説明する。特徴量同士の関連が高いかどうかは、例えばカルバック・ライブラー情報量で定義することができる。カルバック・ライブラー情報量によって特徴量同士の類似性を計算する場合、各特徴量が示すデータの分布をカーネル密度推定 (K D E) 等のアルゴリズムによって推定し、推定された分布間の距離によって定義する。特徴量 $I D$ が i と j の 2 つの特徴量 F_i と F_j のそれぞれが示すデータバラつきから推定した確率分布を P_i と P_j とすると、2 つの分布間の類似度は下記式 (1) で定義される。

40

【 0 0 3 0 】

【 数 1 】

$$D_{KL}(i, j) = (P_i \parallel P_j) = \int_{-\infty}^{\infty} P_i(x) \log \frac{P_i(x)}{P_j(x)} dx \quad \dots\dots\dots (1)$$

【 0 0 3 1 】

ただし、分布間の距離は式 (1) で示した以外にも、2 つの確率分布が多次元正規分布であると仮定して、B h a t t a c h a r y y a 距離を用いたり、相互情報量によって定

50

義する等、他の指標で算出しても良いことは言うまでもない。また、各確率分布推定と分布間の距離算出を同時に行う密度比推定と呼ばれる方法によっても同様の結果を得ることができる。その他、分布間の距離算出に使える指標としては、特徴量毎が示すデータのスコアランキングの差の2乗和の総和を使う方法がある。この場合にもスコアが大きい程、特徴量同士が似ていないものとして扱える。

【0032】

以上のようにして、全 n 個の入力特徴量 $F = \{F_i\}_{i=1}^n$ 同士の類似度を算出する。これにより、全 n 個の入力特徴量は同じく自身を含めて n 個の特徴量との類似度を要素とするベクトルが獲得される。このとき得られたベクトルを式(2)で表す。

【0033】

【数2】

$$G = \{G_i\}_{i=1}^n, G_i \in R^n, n \gg 1 \quad \text{..... (2)}$$

【0034】

次に、ステップS203に続くステップS203で行う処理について具体的な算出法の例について説明する。ステップS203では、2次元のレーダチャートの軸として n 個の特徴量を表す軸は2次元へ次元削減を行うため、次元削減後に対応する軸の描画のための点の座標を式(3)で表す。式(3)では m 次元に次元削減する場合を示しているが、通常のレーダチャートでは $m = 2$ とすれば良い。また、式(3)において B はレーダチャート生成のための埋め込み行列であり、定義は式(4)に示したとおりである。

【0035】

【数3】

$$g = \{g_i\}_{i=1}^n, g = BG_i \in R^m \quad \text{..... (3)}$$

【0036】

【数4】

$$B \in R^{m \times n}, 1 \leq m \ll n \quad \text{..... (4)}$$

【0037】

ここで、埋め込み行列 B を求めるためのいくつかの方法を以下に示す。求めるべき埋め込み行列 B を B^* と呼ぶことにし、また、近付けるべき特徴量同士の規則を記述した類似度行列 W によって式(5)で埋め込み行列を定義する。

【0038】

【数5】

$$B^* = \arg \min_{B \in R^{m \times n}} \left[\sum_{i,j=1}^n \|BG_i - BG_j\|^2 W_{i,j} \right] \quad \text{subject to } \forall G_i \in G; \|BG_i\|^2 = 1 \quad \text{..... (5)}$$

【0039】

ここで、類似度行列 W の要素である $W_{i,j}$ は、特徴量同士を近付ける規則を記述した行列であり、 i 番目の特徴と j 番目の特徴を近付ける場合には1に、遠ざける場合には0になるような関数として設定されれば何でも良い。つまり、類似度 $W_{i,j}$ が1の場合には式(5)で次元削減後の距離が最小化対象となり、0の場合には無視されるようにして B^*

10

20

30

40

50

を求めることになる。当然のことながら、近付ける優先度を対象によって変えたい場合には、 $W_{i,j}$ として0～1の間の値を設定しても良い。類似度 $W_{i,j}$ の例として使える式の例を式(6)、式(7)に示す。

【0040】

【数6】

$$W_{i,j} = \begin{cases} 1 & \text{if } G_i \in kNN(G_j) \text{ or } G_j \in kNN(G_i) \\ 0 & \text{otherwise} \end{cases} \quad k \in N \quad \text{..... (6)}$$

【0041】

10

【数7】

$$W_{i,j} = \exp\left(\frac{-\|G_i - G_j\|^2}{\gamma^2}\right) \quad \gamma > 0 \quad \text{..... (7)}$$

【0042】

式(6)は、n次元特徴空間での G_i が G_j のk近傍にあるかどうかで0か1かを決定する方式であり、式(7)は、定数 γ によって定義される距離ベースで算出される値をセットする方式である。その他の $W_{i,j}$ の例として、抽出特徴量に関する事前知識を反映させた値をセットしても良い。また、抽出特徴の性質上、ユーザの都合によりデータ可視化の際に近付けたくない特徴量が存在する場合には、ユーザによる該当特徴量ID指定の後、該特徴量同士の類似度 $W_{i,j}$ を0にセットすることでユーザ希望の結果を得ることができる。逆に、ユーザの希望で近付けたい特徴があれば同様の入力により、該当する特徴量間の $W_{i,j}$ を1にセットすれば良い。

20

【0043】

これらの指標に基き次元削減を行うが、式(5)にある次元削減後の原点からの距離が1になるという制約条件により、各特徴量IDを代表する点が原点からの距離1の円上に並ぶ結果を得る。このときの処理前と処理後のイメージ図、及びそのとき得られるチャートの例を図6に示す。

30

【0044】

図6(A)は、前述した全入力特徴量をカルバック・ライブラー情報量によって類似度ベースの距離により定義された10次元空間内に10個の各特徴量を代表するベクトル G_i が分布している様子を表している。図6(B)は、前述した方法にて算出された次元削減用の埋め込み行列によって2次元に次元削減された各特徴量を代表するベクトル g_i が原点周りの距離1の円上に分布していることを示している。これが、レーダチャートで各軸をどの位置関係に描画するかの指標になる。

【0045】

この円上の点 g_i が各特徴量を表す特徴量IDのi番目の軸を設定するための位置指標となるため、原点から全10個の点に向かって軸を設定し描画することによって、図6(C)に示したようなチャートを得ることができる。このチャートはレーダチャートのように軸を等間隔に配していないが、各軸の類似度が各軸の間の角度として表現されているため、意味的なまとまりを把握しやすくなっており、一般的なレーダチャートよりも理解しやすくなる場合がある。このように、軸の相対的な位置と角度に情報価値が無かったレーダチャートの欠点を補った、新しいチャートをClumsy Spider Chart(CSC)と称する。

40

【0046】

このチャート上に任意のデータの各軸スコアをプロットして描画したCSCの例を図6(D)に示す。図6(D)は説明を簡易にするために軸の数が少ない場合のCSCとして表現しているが、CSCが便利になるのはより高次元の多変量を扱うときである。高次元

50

になっても、連動してスコアが変わる軸が近く配置されているために、局所に注目する場合は近い軸同士のスコア変動をまとめて観察すればよく、全体のグラフ形状を見ると局所のまとまりを持ったスコアの動きのバランスを俯瞰して把握することができる。

【 0 0 4 7 】

以上説明したように、軸同士の関連性や類似性を考慮して軸の配置（並びや間隔）を設定してチャート（グラフ）を生成することによって、全体を俯瞰した際に複数の軸同士の意味的なまとまりを直感的に把握することが可能となる。これにより、巨視的な視点で見たときのグラフ全体の形状比較がし易くかつ微視的な視点では軸の近い関連性の高い物同士の差異に注目して比較することのできる、人の認識にやさしいチャート（グラフ）を生成することが可能となる。

10

【 0 0 4 8 】

図 7（A）、図 7（B）に高次元（108 次元）の多変量データの比較が直感的に優れていることがわかる C S C の例を示した。図 7（A）、図 7（B）によると、一般的なレーダチャートよりも、どの軸が相関が強く連動して動くことが直感的にわかるため、グラフの注目ポイントが暗に示されており、高次元になっても形状の認識がしやすいことがわかる。図 7（A）と図 7（B）に示したチャートを比較するときも全体的な形状の傾向は変わらないが、局所的な形状としての右上部分、左上部分の形状に違いがあるということが判り易い。軸の並びが適切な順になっていない一般的なレーダチャートであると、108 本もの軸の上のグラフを一瞥して傾向を把握するのは困難になるのは言うまでもない。

【 0 0 4 9 】

20

また、通常のレーダチャートのように軸の間の角度が等しい表示を好むユーザのために外観だけを一般的なレーダチャートに合わせるように軸の並びはそのままにしながら軸間角度を等間隔になるようにしても良い。その場合の変換過程の概念は、図 8（A）～図 8（D）に示しており、図 6（D）に対応するグラフを図 8（E）に示している。

【 0 0 5 0 】

詳細な分析ではなく、より短時間で直感的にデータの傾向を把握するためには、できるだけ冗長な情報になる可能性の高い部分は取り除いてユーザに提示した方が良い場合がある。そのため、さらに簡易チャート表示モード等を選択できるようにし、特徴量同士の類似度が所定の値より高い軸同士を 1 つの軸にまとめて表示しても良い。その際の処理手順を図 9（A）～図 9（D）に示す。例えば、次元削減後の軸の間の角度が設定した閾値以下であった場合にそれらの軸をまとめて 1 つの軸で表し、その軸で表されるスコアは元の複数の軸のスコアの平均値にする等により算出する方法がある。または次元削減前に近傍関係にある点の平均を取る等して 1 つの点にまとめ、その後次元削減をしても良い。図 9（E）は、図 6（D）や図 8（E）に示したデータを同様に図 9（D）のチャート上に表示した例であるが、図 9（E）をさらに一般的なレーダチャートの形式に見たいユーザのために全軸間の角度を等間隔になるように調整して表示しても良い。

30

【 0 0 5 1 】

なお、前述した各種チャートの表示切り替えは、例えば図 10 に示すように複数の画像に対応するチャートを表示させ、チャートを見ながらユーザが好みの表示形式をボタンをクリックするだけで簡単に選択できるようにしてもよい。これによりデータによって適した分析が容易になる。

40

【 0 0 5 2 】

また、前述した説明では、すべて一般的なレーダチャートに近い表示形態を想定したため、次元削減時の次元数を 2 次元とした例を用いて説明したが、より複雑な分析が必要な場合等は次元削減時の次元数を 3 次元としても良い。削減後の次元数が 3 である場合、前述したチャートの 3 次元表現となり、よりそれぞれの特徴量の関係が直感的に表現されたグラフを作成することができる。ただし、3 次元のチャートを確認するディスプレイが 2 次元表示の場合等は 2 次元のチャートよりも形状の把握が難しくなる場合がある。そのために 3 次元化したチャートの際は、緯度や経度によって R G B（表示色）の値を変化させながらカラーで表示するなどすれば良い。なお、本実施形態は外観検査を例として説明し

50

たが複数の特徴量を扱った場合のデータ可視化全般に利用できる方法であることは言うまでもない。

【 0 0 5 3 】

(第 2 の実施形態)

次に、本発明の第 2 の実施形態について説明する。第 2 の実施形態は、第 1 の実施形態で説明したチャート生成のための次元削減時において次元削減する前後において特徴量間の距離関係の保持度合いを基準にして次元削減を行い、人が見て直感的に判り易い形でのレーダチャート等を生成する方法である。以下では、第 2 の実施形態において、前述した第 1 の実施形態と異なる点について説明する。

【 0 0 5 4 】

次元削減する前後において特徴量間の距離関係の保持度合いを基準にして次元削減を行う場合、最小化すべきは次元削減する前後で距離関係が変わることによる誤差である。このときの誤差 2 乗和を例えば式 (8) に示した方法で定義すると良い。

【 0 0 5 5 】

【 数 8 】

$$J_{ee} = \frac{\sum_{i < j} (\delta_{ij} - \Delta_{ij})^2}{\sum_{i < j} \Delta_{ij}^2} \quad \text{..... (8)}$$

【 0 0 5 6 】

ただし、次元削減前の特徴量 G や g は、第 1 の実施形態で示したものと同一であり、 Δ_{ij} は式 (9) のように定義し、 G_i 、 G_j の間の距離を表す。

【 0 0 5 7 】

【 数 9 】

$$\Delta_{ij} = S(G_i, G_j) \quad \text{..... (9)}$$

【 0 0 5 8 】

また、 δ_{ij} は同様に次元削減後の g_i 、 g_j の距離を表す。ただし、 δ_{ij} には 2 次元平面に次元削減した後に原点からの距離が等しくなるように式 (10) に示す制約を加える。

【 0 0 5 9 】

【 数 10 】

$$\delta_{ij} = 2 \sin \frac{\theta_i - \theta_j}{2} \quad g_j = (\cos \theta_j, \sin \theta_j), \quad g_i = (\cos \theta_i, \sin \theta_i), \quad 0 \leq \theta_i < 2\pi \quad \text{..... (10)}$$

【 0 0 6 0 】

以上のようにした後、最急降下法により解を求めることができる。これによって求めた解は、2 次元平面の上で原点周辺の同心円上に配置された点として次元削減後の特徴量の代表点を得る。この点を基にして以降のチャートを獲得する手順は、第 1 の実施形態と同様である。なお、式 (8) 以外の誤差 2 乗和関数の例を式 (11)、式 (12) に示す。これらはどれも多次元尺度法等で利用されている関数である。

【 0 0 6 1 】

【 数 11 】

$$J_{ff} = \sum_{i < j} \left(\frac{\delta_{ij} - \Delta_{ij}}{\Delta_{ij}} \right)^2 \quad \text{..... (11)}$$

10

20

30

40

50

【 0 0 6 2 】

【 数 1 2 】

$$J_{ef} = \frac{1}{\sum_{i < j} \Delta_{ij}} \sum_{i < j} \frac{(\delta_{ij} - \Delta_{ij})^2}{\Delta_{ij}} \dots\dots\dots (12)$$

【 0 0 6 3 】

その他、式（ 8 ）、式（ 1 1 ）、式（ 1 2 ）に記した最小化すべき誤差 2 乗和関数は、次元削減前の特徴量 G のすべての関係性を保持するための誤差となっている。k 近傍までを近付けるといったパラメータを追加すると I S O M A P と同様な手順で次元削減を行うことができる。以上手順によって得られた次元削減後の各特徴量を表す代表点である g を基にしてチャートを生成する手順は、第 1 の実施形態と同様手順によって実現でき、適切なチャートを得ることができる。

10

【 0 0 6 4 】

（ 第 3 の実施形態 ）

次に、本発明の第 3 の実施形態について説明する。第 3 の実施形態は、高次元の特徴量で表現される複数のデータがあり、各々のデータがそれらの中でどのような傾向にあるかを分析したい場合に人が見て直感的に判り易い表示に平行座標プロットや棒グラフを表示する方法である。第 1 の実施形態又は第 2 の実施形態で説明した方法によって、レーダチャートの軸同士の間隔や順番を決定するアルゴリズムを用いることによって、同様に平行座標プロットや棒グラフの項目の順番と間隔を決定することができる。また、レーダチャートでは円形状になるように軸を配することができたため、平行座標プロットや棒グラフでも n 個の項目にてグラフを生成する際、末端である n 個目の項目の隣に 1 個目の項目からのグラフをさらに描画しても良い。

20

【 0 0 6 5 】

（ 本発明の他の実施形態 ）

また、本発明は、以下の処理を実行することによっても実現される。即ち、前述した実施形態の機能を実現するソフトウェア（プログラム）を、ネットワーク又は各種記憶媒体を介してシステム或いは装置に供給し、そのシステム或いは装置のコンピュータ（または CPU や MPU 等）がプログラムを読み出して実行する処理である。

30

【 0 0 6 6 】

例えば、前述した各実施形態における情報処理装置は、図 1 1 に示すようなコンピュータ機能 1 1 0 0 を有し、その CPU 1 1 0 1 により各実施形態での動作が実施される。コンピュータ機能 1 1 0 0 は、図 1 1 に示すように、CPU 1 1 0 1 と、ROM 1 1 0 2 と、RAM 1 1 0 3 とを備える。また、操作部（CONS）1 1 0 9 のコントローラ（CONSC）1 1 0 5 と、LCD 等の表示部としてのディスプレイ（DISP）1 1 1 0 のディスプレイコントローラ（DISPC）1 1 0 6 とを備える。さらに、ハードディスク（HD）1 1 1 1、及びフレキシブルディスク等の記憶デバイス（STD）1 1 1 2 のコントローラ（DCONT）1 1 0 7 と、ネットワークインタフェースカード（NIC）1 1 0 8 とを備える。それら機能部 1 1 0 1、1 1 0 2、1 1 0 3、1 1 0 5、1 1 0 6、1 1 0 7、1 1 0 8 は、システムバス 1 1 0 4 を介して互いに通信可能に接続された構成としている。

40

【 0 0 6 7 】

CPU 1 1 0 1 は、ROM 1 1 0 2 又は HD 1 1 1 1 に記憶されたソフトウェア、又は STD 1 1 1 2 より供給されるソフトウェアを実行することで、システムバス 1 1 0 4 に接続された各構成部を総括的に制御する。すなわち、CPU 1 1 0 1 は、前述したような動作を行うための処理プログラムを、ROM 1 1 0 2、HD 1 1 1 1、又は STD 1 1 1 2 から読み出して実行することで、各実施形態での動作を実現するための制御を行う。RAM 1 1 0 3 は、CPU 1 1 0 1 の主メモリ又はワークエリア等として機能する。

50

【 0 0 6 8 】

CONSC 1 1 0 5 は、CONS 1 1 0 9 からの指示入力を制御する。DISPC 1 1 0 6 は、DISP 1 1 1 0 の表示を制御する。DCONT 1 1 0 7 は、ブートプログラム、種々のアプリケーション、ユーザファイル、ネットワーク管理プログラム、及び各実施形態における動作を実現するための処理プログラム等を記憶するHD 1 1 1 1 及びSTD 1 1 1 2 とのアクセスを制御する。NIC 1 1 0 8 はネットワーク 1 1 1 3 上の他の装置と双方向にデータをやりとりする。

【 0 0 6 9 】

なお、前記実施形態は、何れも本発明を実施するにあたっての具体化のほんの一例を示したものに過ぎず、これらによって本発明の技術的範囲が限定的に解釈されてはならないものである。すなわち、本発明はその技術思想、又はその主要な特徴から逸脱することなく、様々な形で実施することができる。

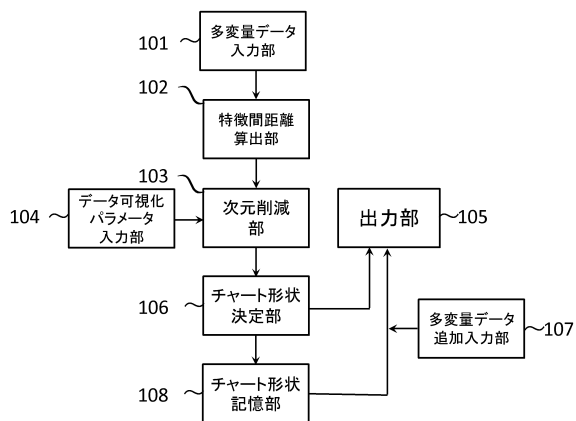
10

【 符号の説明 】

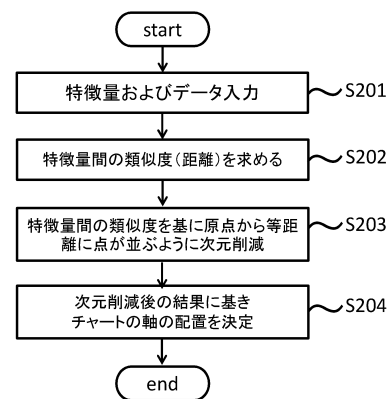
【 0 0 7 0 】

1 0 1 : 多変量データ入力部 1 0 2 : 特徴量間距離算出部 1 0 3 : 次元削減部 1 0 4 : データ可視化パラメータ入力部 1 0 5 : 出力部 1 0 6 : チャート形状決定部 1 0 7 : 多変量データ追加入力部 1 0 8 : チャート形状記憶部

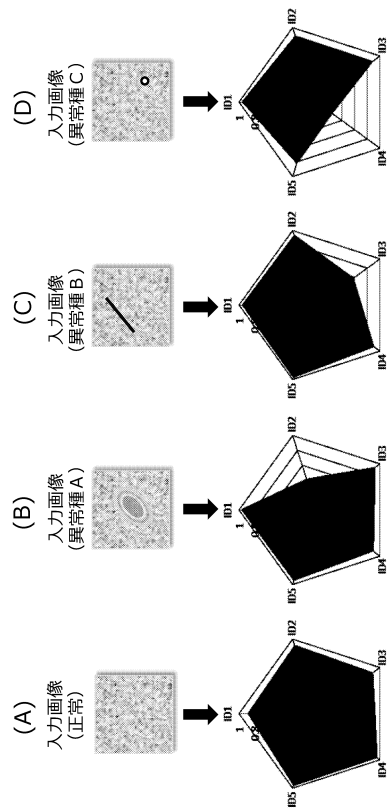
【 図 1 】



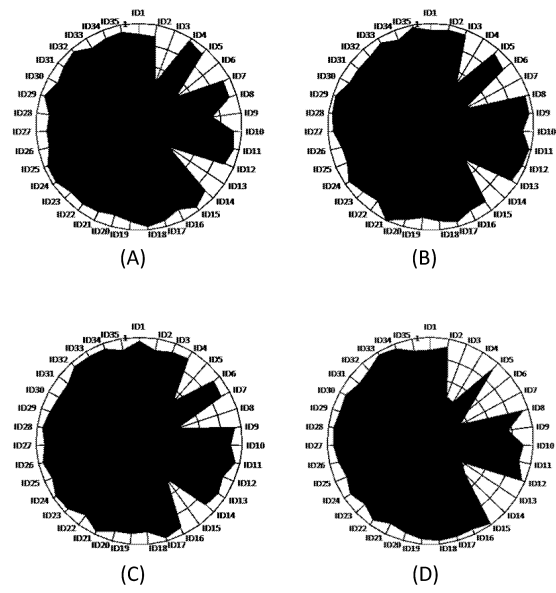
【 図 2 】



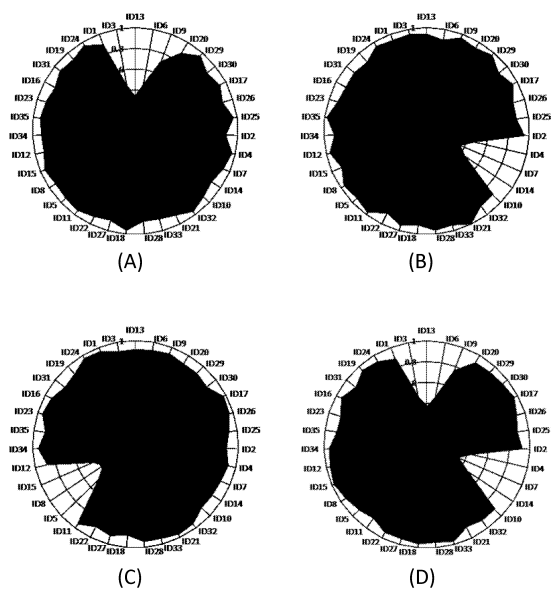
【図 3】



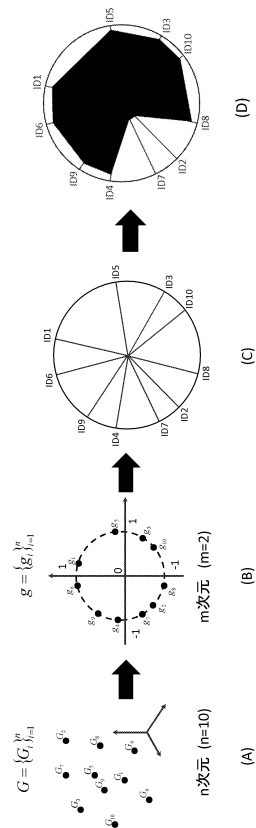
【図 4】



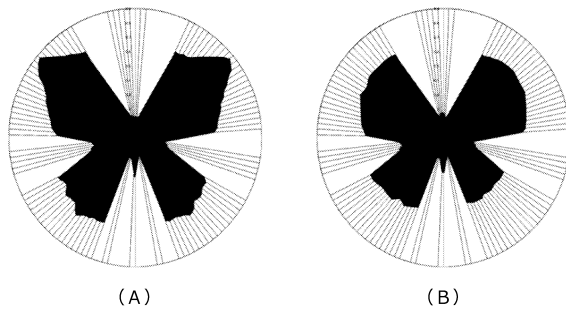
【図 5】



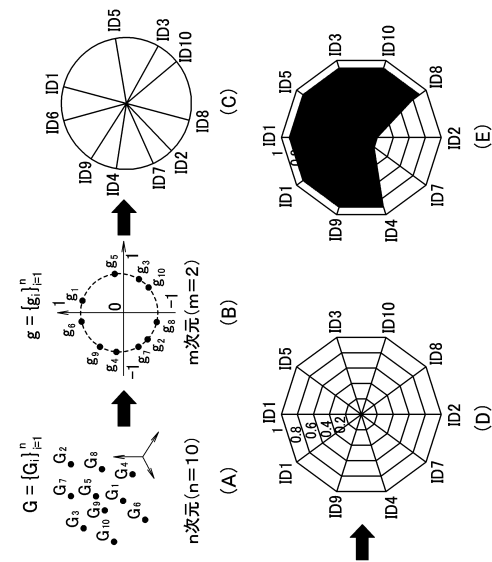
【図 6】



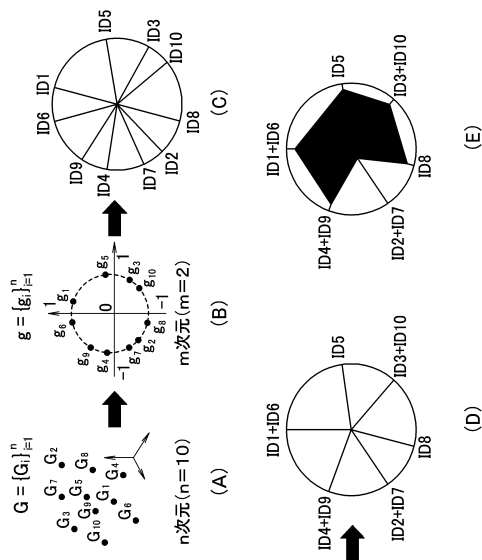
【図 7】



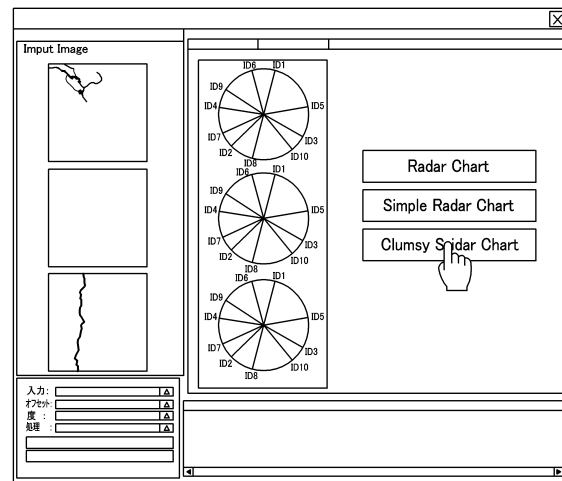
【図 8】



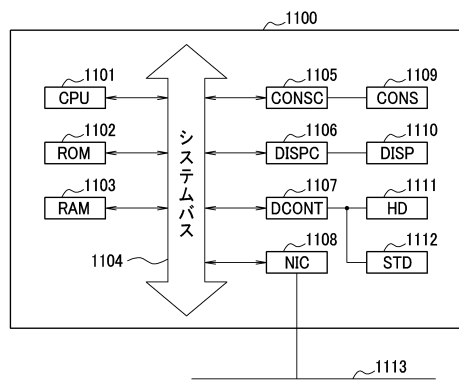
【図 9】



【図 10】



【図 11】



フロントページの続き

(56)参考文献 特開 2 0 0 8 - 0 0 3 8 6 9 (J P , A)
特開 2 0 1 0 - 2 5 0 0 2 2 (J P , A)
特開 2 0 1 0 - 1 6 5 2 3 4 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)
G 0 6 F 1 6 / 5 6
G 0 6 F 1 6 / 5 4