

## 明 細 書

発明の名称：

統計的距離行列の計算方法、統計的距離行列の可視化方法及び装置及びプログラム

### 技術分野

[0001] 本発明は、統計的距離行列の計算方法、統計的距離行列の可視化方法及び装置及びプログラムに係り、特に、統計的距離行列の精密化・表現・クラスタリングを用いた統計的距離行列の計算方法及、統計的距離行列の可視化方法及び装置及びプログラムに関する。

### 背景技術

[0002] 機械学習は、教師あり学習、教師なし学習、および強化学習という三種類に分けられる。具体的な目的として、分類、回帰、クラスタリング、次元削減、異常検知などの事例があり、解を求める方法として、サポートベクターマシン、K平均法、決定木など多くのアルゴリズムがある。しかし、どのアルゴリズムに対しても観測値の特徴に関する距離関数の設計および最適化をしなければならない。例えば、分類におけるサポートベクターマシンは、最も効率的なカーネル関数を使用し、二つの観測値カテゴリ間の特徴距離を測定する必要がある。クラスタリングにおけるK平均法は、ユークリッド空間またはマハラノビス空間において、観測値をクラスターに分割し、各観測値クラスター内の特徴の平方和距離を最小化することを目的としている。

統計的距離は、二つの確率分布に従って独立に生成された標本（観測値）集合を用いて、その二つの確率分布間の距離と定義される。一般的な距離が持つわけではない数学的特性を有するため、機械学習に対して、その測定がより効果的かつ適切になり、小さな外れ値に対して、よりロバストである（頑健性がある）。今まで使用されたマハラノビス距離（Mahalanobis Distance）[9]、バタチャリヤ距離（Mahalanobis Distance）[1]、ヘリンガー距離（Hellinger Dist

ance) [4]、カルバック・ライブラーダイバージェンス (Kullback-Leibler Divergence) [7]、チャノフ距離 (Chernoff Distance) [2]などのいくつかの重要な統計的距離は、画像セグメンテーション、テキストチャのセグメンテーション、色とテキストチャのマッチング、特徴抽出、音声認識、動作認識などを含む人工知能の分野に広く適用されている。

[0003] 従来技術としては、例えば、以下の文献が挙げられる。

特許文献1には、「畳み込み処理を含む学習器の判定制度を向上できる機械学習プログラム、機械学習方法および機械学習装置」(要約)が開示されている。

特許文献2には、「3つ以上の画像間の関連性を利用することで画像検索の精を向上させることができる」(要約)ようにした判定装置および判定方法が開示されている。

特許文献3には、「マハラノビス距離測定部を使用して、データベース内の複数の画像の中から問合せ画像を識別する」(アブストラクト)ようにした距離測定法を用いる画像検索について開示されている。

特許文献4には、「特徴の教師なし学習または半教師あり学習を実現するために、量子プロセッサが1つ以上のデータセットに階層型深層学習(HDLと呼ばれる)を実現するようにプログラムされる」(アブストラクト)ようにしたデータの量子プロセッシングのためのシステム及び方法が開示されている。

## 先行技術文献

### 特許文献

[0004] 特許文献1：特開2019-185483号公報

特許文献2：特開2019-185483号公報

特許文献3：米国特許第6,681,060号公報

特許文献4：米国特許第10,318,881号公報

### 発明の概要

### 発明が解決しようとする課題

[0005] しかしながら、今まで使用されたマハラノビス距離 (Mahalanobis Distance) [9]、バタチャリヤ距離 (Mahalanobis Distance) [1]、ヘリンガー距離 (Hellinger Distance) [4]、カルバック・ライブラーダイバージェンス (Kullback-Leibler Divergence) [7]、チャーノフ距離 (Chernoff Distance) [2]などの従来の統計的距離は、観測値の特徴要素のサイズに関係なく、二つの観測値集合間のグローバル距離を表す、スカラー値の出力しか提供しないため、観測値内においてすべての特徴要素間のローカル距離を詳しく表現できない場合が想定されるという明らかな課題がある。したがって、スカラー値の統計的距離を距離行列に変換し、距離の精密化を実現する方法が、非常に重要になる。

一方、距離行列の概念は既にグラフ理論[3]に導入されている。例えば、有向グラフの距離行列は、重み付き隣接行列によって定義されている。各辺に重みが割り当てられている場合、二つの頂点間の距離は、二つの頂点を接続する最短パスの重みの和として測定できる。パスが方向付けられているため、その距離行列は非対称であり、対称性を持つ数学的に厳密な意味での距離ではない。各頂点の標本が十分に多い場合、相関行列または相互相関行列を使用して、隣接行列の要素の重みを同定し、距離行列として定量化する。ただし、相関行列は、すべてのデータが確率空間に従うことを前提とすれば、情報が失われるため、機械学習のいくつかの高い要件を満たさない。

また、上述の特許文献1～4には、本件発明のような、統計的距離を行列に変換する工程及び／又はその要素クラスタリングに関する手法については開示も示唆もされていない。

[0006] 上述の課題を解決するために、本発明では、例えば、単純な脱トレース (detrace) 操作によって、従来の統計的距離を行列形式に変換し、今までにない高いパフォーマンスを持つ統計的距離行列を提案した。機械学習の分野に最も有名な画像データセットCIFAR-10 [6]に関する実験

を行い、複雑な場合でも統計的距離行列の効果があることを確認した。

本発明は、以上の点に鑑み、スカラー値の統計的距離を統計的距離行列に変換することで距離の精密化を実現することを目的とする。

### 課題を解決するための手段

[0007] 本発明の第1の解決手段によると、

統計的距離行列の計算方法であって、

処理部は、入力部又は他の装置又は記憶部から、比較対象となる第1のベクトルデータ  $a$   $a$  及び第2のベクトルデータ  $b$   $b$  を入力し、又は、前記処理部は、入力部又は他の装置又は記憶部から、比較対象となる第1の行列又は画像又は集合データ  $A$   $A$  及び第2の行列又は画像又は集合データ  $B$   $B$  を入力して前記第1のベクトルデータ  $a$   $a$  及び前記第2のベクトルデータ  $b$   $b$  にベクトル化し、

前記処理部は、前記第1のベクトルデータ  $a$   $a$  及び前記第2のベクトルデータ  $b$   $b$  から、それぞれ、第1の平均ベクトル  $\mu$   $\mu$  1 と第2の平均ベクトル  $\mu$   $\mu$  2、及び、第1の共分散行列  $\Sigma$   $\Sigma$  1 と第2の共分散行列  $\Sigma$   $\Sigma$  2 を計算し、

前記処理部は、第1の平均ベクトル  $\mu$   $\mu$  1、第2の平均ベクトル  $\mu$   $\mu$  2、第1の共分散行列  $\Sigma$   $\Sigma$  1、第2の共分散行列  $\Sigma$   $\Sigma$  2 により定義される統計的距離  $D$  の脱トレースである統計的距離行列  $D$   $D$  (ここで、 $D = \text{tr } D D$ ) を求め、

前記処理部は、統計的距離行列  $D$   $D$  を前記記憶部に記憶し又は表示部に表示させ又は出力部により出力させる、

統計的距離行列の計算方法が提供される。

[0008] 本発明の第2の解決手段によると、

統計的距離行列の可視化方法であって、

処理部は、入力部又は他の装置又は記憶部から、比較対象となる第1のベクトルデータ  $a$   $a$  及び第2のベクトルデータ  $b$   $b$  を入力し、又は、前記処理

部は、入力部又は他の装置又は記憶部から、比較対象となる第1の行列又は画像又は集合データA A及び第2の行列又は画像又は集合データB Bを入力して前記第1のベクトルデータa a及び前記第2のベクトルデータb bにベクトル化し、

前記処理部は、前記第1のベクトルデータa a及び前記第2のベクトルデータb bから、それぞれ、第1の平均ベクトル $\mu \mu 1$ と第2の平均ベクトル $\mu \mu 2$ 、及び、第1の共分散行列 $\Sigma \Sigma 1$ と第2の共分散行列 $\Sigma \Sigma 2$ を計算し、

前記処理部は、第1の平均ベクトル $\mu \mu 1$ 、第2の平均ベクトル $\mu \mu 2$ 、第1の共分散行列 $\Sigma \Sigma 1$ 、第2の共分散行列 $\Sigma \Sigma 2$ により定義される統計的距離Dの脱トレースである統計的距離行列D D（ここで、 $D = \text{tr} \ D D$ ）を求め、

前記処理部は、統計的距離行列D Dの各対角成分に対して対角成分を含む行及び列の距離成分を累積して距離累積ベクトル $\phi \phi$ を求め、

前記処理部は、距離累積ベクトル $\phi \phi$ を距離累積行列 $\Phi \Phi$ に行列化し、

前記処理部は、距離累積行列 $\Phi \Phi$ 又は距離累積行列 $\Phi \Phi$ による画像を前記記憶部に記憶し又は表示部に表示させ又は出力部により出力させる、統計的距離行列の可視化方法が提供される。

[0009] 本発明の第3の解決手段によると、

統計的距離行列の可視化方法であって、

処理部は、入力部又は他の装置又は記憶部から、比較対象となる第1のベクトルデータa a及び第2のベクトルデータb bを入力し、又は、前記処理部は、入力部又は他の装置又は記憶部から、比較対象となる第1の行列又は画像又は集合データA A及び第2の行列又は画像又は集合データB Bを入力して前記第1のベクトルデータa a及び前記第2のベクトルデータb bにベクトル化し、

前記処理部は、前記第1のベクトルデータa a及び前記第2のベクトルデータb bから、それぞれ、第1の平均ベクトル $\mu \mu 1$ と第2の平均ベクトル

$\mu \mu 2$ 、及び、第1の共分散行列 $\Sigma \Sigma 1$ と第2の共分散行列 $\Sigma \Sigma 2$ を計算し、

前記処理部は、第1の平均ベクトル $\mu \mu 1$ 、第2の平均ベクトル $\mu \mu 2$ 、第1の共分散行列 $\Sigma \Sigma 1$ 、第2の共分散行列 $\Sigma \Sigma 2$ により定義される統計的距離 $D$ の脱トレースである統計的距離行列 $D D$ （ここで、 $D = \text{tr } D D$ ）を求め、

前記処理部は、前記統計的距離行列 $D D$ に対してクラスタリング処理を実行して、各行に対応する番号若しくは $I D$ 、又は、各列に対応する番号若しくは $I D$ 、に対するクラスタラベルを要素とするクラスタラベルベクトル若しくは集合を求め、

前記処理部は、前記クラスタラベルベクトル若しくは集合をクラスタラベル行列に行列化し、

前記処理部は、前記クラスタラベル行列又はクラスタラベル行列による画像を前記記憶部に記憶し又は表示部に表示させ又は出力部により出力させる、

統計的距離行列の可視化方法が提供される。

[0010] 本発明の第4の解決手段によると、  
統計的距離行列の可視化装置であって、  
処理部  
を備え、

前記処理部は、入力部又は他の装置又は記憶部から、比較対象となる第1のベクトルデータ $a a$ 及び第2のベクトルデータ $b b$ を入力し、又は、前記処理部は、入力部又は他の装置又は記憶部から、比較対象となる第1の行列又は画像又は集合データ $A A$ 及び第2の行列又は画像又は集合データ $B B$ を入力して前記第1のベクトルデータ $a a$ 及び前記第2のベクトルデータ $b b$ にベクトル化し、

前記処理部は、前記第1のベクトルデータ $a a$ 及び前記第2のベクトルデータ $b b$ から、それぞれ、第1の平均ベクトル $\mu \mu 1$ と第2の平均ベクトル

$\mu \mu 2$ 、及び、第1の共分散行列 $\Sigma \Sigma 1$ と第2の共分散行列 $\Sigma \Sigma 2$ を計算し、

前記処理部は、第1の平均ベクトル $\mu \mu 1$ 、第2の平均ベクトル $\mu \mu 2$ 、第1の共分散行列 $\Sigma \Sigma 1$ 、第2の共分散行列 $\Sigma \Sigma 2$ により定義される統計的距離 $D$ の脱トレースである統計的距離行列 $D D$ （ここで、 $D = \text{tr } D D$ ）を求め、

前記処理部は、統計的距離行列 $D D$ の各対角成分に対して対角成分を含む行及び列の距離成分を累積して距離累積ベクトル $\phi \phi$ を求め、

前記処理部は、距離累積ベクトル $\phi \phi$ を距離累積行列 $\Phi \Phi$ に行列化し、

前記処理部は、距離累積行列 $\Phi \Phi$ 又は距離累積行列 $\Phi \Phi$ による画像を前記記憶部に記憶し又は表示部に表示させ又は出力部により出力させる、統計的距離行列の可視化装置が提供される。

[0011] 本発明の第5の解決手段によると、

統計的距離行列の可視化プログラムであって、

処理部が、入力部又は他の装置又は記憶部から、比較対象となる第1のベクトルデータ $a a$ 及び第2のベクトルデータ $b b$ を入力し、又は、前記処理部は、入力部又は他の装置又は記憶部から、比較対象となる第1の行列又は画像又は集合データ $A A$ 及び第2の行列又は画像又は集合データ $B B$ を入力して前記第1のベクトルデータ $a a$ 及び前記第2のベクトルデータ $b b$ にベクトル化するステップと、

前記処理部が、前記第1のベクトルデータ $a a$ 及び前記第2のベクトルデータ $b b$ から、それぞれ、第1の平均ベクトル $\mu \mu 1$ と第2の平均ベクトル $\mu \mu 2$ 、及び、第1の共分散行列 $\Sigma \Sigma 1$ と第2の共分散行列 $\Sigma \Sigma 2$ を計算するステップと、

前記処理部が、第1の平均ベクトル $\mu \mu 1$ 、第2の平均ベクトル $\mu \mu 2$ 、第1の共分散行列 $\Sigma \Sigma 1$ 、第2の共分散行列 $\Sigma \Sigma 2$ により定義される統計的距離 $D$ の脱トレースである統計的距離行列 $D D$ （ここで、 $D = \text{tr } D D$ ）を求めるステップと、

前記処理部が、統計的距離行列  $D$  の各対角成分に対して対角成分を含む行及び列の距離成分を累積して距離累積ベクトル  $\phi$  を求めるステップと、

前記処理部が、距離累積ベクトル  $\phi$  を距離累積行列  $\Phi$  に行列化するステップと、

前記処理部が、距離累積行列  $\Phi$  又は距離累積行列  $\Phi$  による画像を前記記憶部に記憶し又は表示部に表示させ又は出力部により出力させるステップ、  
をコンピュータに実行させるための統計的距離行列の可視化プログラムが提供される。

## 発明の効果

[0012] 本発明によると、スカラー値の統計的距離を統計的距離行列に変換することで距離の精密化を実現することができる。

## 図面の簡単な説明

[0013] [図1]「1. 行列に関する記号」についての説明図。

[図2]「2. 確率論に関する記号」及び「3. 階層的クラスタリングに関する記号」についての説明図。

[図3]本実施の形態に関するハードウェアの構成図。

[図4]第1の実施の形態の統計的距離行列の可視化方法に関する技術についての説明図。

[図5]第1の実施の形態の統計的距離行列の可視化方法に関するフローチャート。

[図6]CIFAR-10データセット[6]における飛行機、鳥、猫、および犬の画像の例を示す図。

[図7]飛行機と犬、鳥と犬、および猫と犬のケースに対する統計距離行列  $DDM$ 、 $DDKL$ 、 $ddb$ 、および  $DDC$  を示す図。

[図8]飛行機と犬、鳥と犬、および猫と犬のケースに対する距離累積画像  $\Phi M$ 、 $\Phi K L$ 、 $\Phi B$ 、および  $\Phi C$  を示す図。

[図9]階層的クラスタリングの処理についての説明図。

[図10]飛行機と犬のケースに対して、統計的距離行列 $DDB$ を使用した階層的クラスタリング結果を示す図。

[図11]鳥と犬のケースに対して、統計的距離行列 $DDB$ を使用した階層的クラスタリング結果を示す図。

[図12]猫と犬のケースに対して、統計的距離行列 $DDB$ を使用した階層的クラスタリング結果を示す図。

[図13]飛行機と犬のケースに対して、距離累積画像 $\Phi\Phi B$ から値の低い、普通、高い、合計3種類の要素領域を切り出した結果を示す図。

[図14]鳥と犬のケースに対して、距離累積画像 $\Phi\Phi B$ から値の低い、普通、高い、合計3種類の要素領域を切り出した結果を示す図。

[図15]猫と犬のケースに対して、距離累積画像 $\Phi\Phi B$ から値の低い、普通、高い、合計3種類の要素領域を切り出した結果を示す図。

[図16] $\rho$ の計算結果の例を示す図。

[図17]統計的距離に関する関連技術1についての説明図。

[図18]相互相関行列に関する関連技術2についての説明図。

[図19]階層的クラスタリングのフローチャート。

[図20]第2の実施の形態の統計的距離行列の可視化方法に関するフローチャート

[図21]第3の実施の形態の統計的距離行列の可視化方法に関するフローチャート

## 発明を実施するための形態

### [0014] 1. 記号の説明

まず、本明細書中で使用される記号について説明する。

図1に、「1. 行列に関する記号」についての説明図を示す。

図2に、「2. 確率論に関する記号」及び「3. 階層的クラスタリングに関する記号」についての説明図を示す。

なお、記号の上に付される“ $\wedge$ ”の記号は、電子出願ソフトで利用できる文字・記号の制約・都合上、文字の右上に記載するが、数式で示すように、文字の真上に記載されたものと同一である。また、数式等で示すように、太文字の  $\mathbf{a}$  等はベクトル、太文字の  $\mathbf{A}$ 、 $\mathbf{D}$ 、 $\boldsymbol{\mu}$ 、 $\boldsymbol{\Sigma}$  等は行列であるが、同制約・都合上、それぞれ、 $\mathbf{a} \mathbf{a}$  等と記載してベクトルを表し、 $\mathbf{A} \mathbf{A}$ 、 $\mathbf{D} \mathbf{D}$ 、 $\boldsymbol{\mu} \boldsymbol{\mu}$ 、 $\boldsymbol{\Sigma} \boldsymbol{\Sigma}$  等と記載して行列を表す。また、数式等で示すように、黒板文字の  $R$ 、 $E$  等は、同制約・都合上、 $R$ 、 $E$  等と記載する。また、[1]、[2]、 $\dots$  等は、参考文献の番号を表し、後述する（「1.1. 参考文献」参照）。

## [0015] 2. 統計的距離

定義 1 :

一般に、確率空間  $(\Omega, \mathcal{F}, P)$  が与えられた場合、可測空間  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  に分布  $P_X$  を持つ  $d$ -次元のランダムベクトル  $X_1$  に対して、その累積分布関数は  $F_1(x) := P(X_1 \leq x) = P_X((-\infty, x])$ 、 $x \in \mathbb{R}^d$  とする（「 $:=$ 」は、定義することを表す。）。 $F_1(x)$  が変数  $x$  に関して連続である場合、 $p_X(x) := dF_1(x)/dx$  は、 $X_1$  の確率密度関数と呼ばれる。同様に、別の確率空間  $(\Omega, \mathcal{F}, Q)$  が与えられた場合、可測空間  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  に分布  $Q_X$  を持つ  $d$ -次元のランダムベクトル  $X_2$  に対して、その累積分布関数は  $F_2(x) := Q(X_2 \leq x) = Q_X((-\infty, x])$  とする。 $F_2(x)$  が連続ならば、 $X_2$  の確率密度関数  $q_X(x) := dF_2(x)/dx$  が得られる。通常に、 $p_X(x)$  および  $q_X(x)$  は、それぞれ  $p(x)$  および  $q(x)$  と省略される。

上述の定義に基づいて、2つの確率密度関数  $p(x)$  と  $q(x)$  の間の非類似度を表すために、いくつかの測度が統計学に導入された。

[0016] バタチャリヤ距離  $D_B$  は、非類似度を定量化するために最初の計量として[1]に提案された（式（1））。そして、 $D_B$  を拡張したチャーノフ距離  $D_c$  は、[2]で導入された。ここで、平方根演算子は指数係数  $s$  に置き換えられた（式（2））。次式（3）のように定式化されたカルバックライブラダイ

バージェンスDKLは、[7]で提案された。計量の公理を満たしていないため、計量ではないことに注意を要する。また、[4]で紹介されているヘリンガー距離DHは、ヘリンガー積分によって次式（４）のように定義される、若しくは、次式（５）のように与えられる。

[数1]

$$D_B := -\ln \int_{\mathbb{R}^d} p^{\frac{1}{2}}(\mathbf{x}) q^{\frac{1}{2}}(\mathbf{x}) d\mathbf{x} \quad (1)$$

$$D_C := -\ln \int_{\mathbb{R}^d} p^s(\mathbf{x}) q^{1-s}(\mathbf{x}) d\mathbf{x} \quad (2)$$

$$D_{KL} := \int_{\mathbb{R}^d} [p(\mathbf{x}) - q(\mathbf{x})] \ln \left[ \frac{p(\mathbf{x})}{q(\mathbf{x})} \right] d\mathbf{x} \quad (3)$$

$$D_H^2 := \frac{1}{2} \int_{\mathbb{R}^d} \left[ p^{\frac{1}{2}}(\mathbf{x}) - q^{\frac{1}{2}}(\mathbf{x}) \right]^2 d\mathbf{x} \quad (4)$$

$$D_H := \frac{1}{\sqrt{2}} \int_{\mathbb{R}^d} \left\| p^{\frac{1}{2}}(\mathbf{x}) - q^{\frac{1}{2}}(\mathbf{x}) \right\|_2 d\mathbf{x} \quad (5)$$

これらの測度はすべて、対応するランダムベクトルの次元dに関係なく、2つの確率密度関数p（xx）とq（xx）の間にあるスカラーの非類似度のみを与える。

### [0017] 3. 統計的距離行列

最初に、従来の距離行列として広く使用されている、相互相関行列RRについて説明する。正規分布N（μμ1，ΣΣ1）とN（μμ2，ΣΣ2）にそれぞれ従い、d次元のランダムベクトルX1とX2を持つ2つの母集団（ベクトルの集合）{aak}と{bbk}、k=1，…，Nが与えられると、X1とX2の平均ベクトルμμ1，μμ2∈Rd、および共分散行列ΣΣ1，ΣΣ2∈Rd×dは、式（19）と（37）により得られる。

[数2]

$$\begin{aligned}\mu_1 &:= E[X_1] = \frac{1}{N} \sum_{k=1}^N \mathbf{a}_k, \\ \Sigma_1 &:= \text{cov}[X_1] = E[(X_1 - E[X_1])(X_1 - E[X_1])^T] = \frac{1}{N-1} \sum_{k=1}^N [(\mathbf{a}_k - \mu_1)(\mathbf{a}_k - \mu_1)^T] \quad (19)\end{aligned}$$

$$\begin{aligned}\mu_2 &:= E[X_2] = \frac{1}{N} \sum_{k=1}^N \mathbf{b}_k, \\ \Sigma_2 &:= \text{cov}[X_2] = E[(X_2 - E[X_2])(X_2 - E[X_2])^T] = \frac{1}{N-1} \sum_{k=1}^N [(\mathbf{b}_k - \mu_2)(\mathbf{b}_k - \mu_2)^T] \quad (37)\end{aligned}$$

[0018] 上式に基づいて、相互相関行列  $R$  は、式 (38) で計算される。ここで、 $\Sigma \Sigma 1$ 、 $\Sigma \Sigma 2$  は  $X 1$  と  $X 2$  の相互共分散行列、 $\text{diag}(\Sigma \Sigma)$  は行列  $\Sigma \Sigma$  の対角行列とする。

[数3]

$$\begin{aligned}R &:= \text{corr}[X_1, X_2] = (\text{diag}(\Sigma_1))^{-\frac{1}{2}} \Sigma_{1,2} (\text{diag}(\Sigma_2))^{-\frac{1}{2}}, \\ \Sigma_{1,2} &:= \text{cov}[X_1, X_2] = E[(X_1 - E[X_1])(X_2 - E[X_2])^T] = \frac{1}{N-1} \sum_{k=1}^N [(\mathbf{a}_k - \mu_1)(\mathbf{b}_k - \mu_2)^T] \quad (38)\end{aligned}$$

ここで、 $\text{diag}(\Sigma \Sigma 1)$  と  $\text{diag}(\Sigma \Sigma 2)$  に  $\Sigma \Sigma 1$  と  $\Sigma \Sigma 2$  の対角成分しかないため、情報が失われることに注意すべきである。

[0019] 一方、本発明及び／又は本実施の形態は、脱トレース (d e - t r a c e) 操作により、スカラー値の統計的距離を変換して得た、新たな統計的距離行列を提案する。脱トレース操作をわかりやすくするために、まず、バタチャリヤ距離  $D B$  の特定ケースと見なすマハラノビス距離  $D M$  [9] を紹介する。相互相関行列と同様に、平均ベクトル  $\mu \mu 1$ 、 $\mu \mu 2 \in \mathbb{R}^d$  と共分散行列  $\Sigma \Sigma 1$ 、 $\Sigma \Sigma 2 \in \mathbb{R}^d \times d$  をそれぞれ持つ2つの母集団  $\{\mathbf{a} \mathbf{a} k\}$  と  $\{\mathbf{b} \mathbf{b} k\}$  があれば、母集団の間のマハラノビス距離  $D M$  を次の二次形式で表す。

[数4]

$$\begin{aligned}D_M &= (\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2) \\ &= \text{tr}\{(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \Sigma^{-1}\} = \text{tr} D_M \quad (6)\end{aligned}$$

[0020] ここで、 $\Sigma \Sigma = \Sigma \Sigma 1 = \Sigma \Sigma 2$  とする。この二次形式は、式 (6) の2行

目で述べたように、トレース形式に変換できる。行列のトレースを脱ぐこと  
(removing the trace) により、マハラノビス距離行列  
DDMを式(14)で得られる。

[数5]

$$D_M = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \Sigma^{-1} \quad (14)$$

[0021] 対照的に、バタチャリヤ距離行列DDBに対応するバタチャリヤ距離DB  
は、定義1に従って連続的な可測空間で定義される必要である。本発明及び  
／又は本実施の形態では、2つの母集団(ベクトルの集合)  $\{a_k\}$  と  $\{b_k\}$   
からのd次元ランダムベクトル $X_1$ と $X_2$ は、それぞれ2つの正  
規分布 $N(\mu_1, \Sigma_1)$ と $N(\mu_2, \Sigma_2)$ に従うと仮定し、 $X_1$   
と $X_2$ の間のDBは次式のように定義される。なお、後述の「10. 付録  
バタチャリヤ距離DBの導出(式(7))」に詳細な証明を示した。

[数6]

$$\begin{aligned} D_B &= \frac{1}{4}(\mu_1 - \mu_2)^T(\Sigma_1 + \Sigma_2)^{-1}(\mu_1 - \mu_2) \\ &\quad + \frac{1}{2} \ln \left[ \det \Sigma_1^{-\frac{1}{2}} \det \left( \frac{\Sigma_1 + \Sigma_2}{2} \right) \det \Sigma_2^{-\frac{1}{2}} \right] \\ &= \text{tr} \left\{ \frac{1}{4}(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T(\Sigma_1 + \Sigma_2)^{-1} \right. \\ &\quad \left. + \frac{1}{2} \left[ \ln \left( \frac{\Sigma_1 + \Sigma_2}{2} \right) - \ln(\Sigma_1^{-\frac{1}{2}} \Sigma_2^{-\frac{1}{2}}) \right] \right\} = \text{tr} D_B \end{aligned} \quad (7)$$

[0022] 式(7)の第1項が式(6)に似ているので、式(6)と同様にトレース  
形式に変換できる。さらに、次式に基づいて、行列式の自然対数関数である  
式(7)の第2項もトレース形式に変更できる。

[数7]

$$\det A \det B = \det(AB) \quad (8)$$

$$\ln(\det A) = \operatorname{tr}(\ln A) \quad (9)$$

$$\operatorname{tr}[\ln(AB)] = \operatorname{tr}(\ln A) + \operatorname{tr}(\ln B) \quad (10)$$

[0023] ここで、 $A$  と  $B$  が  $R^d \times d$  における 2 つの正定行列である場合、上記のすべての数式が成り立つ。次に、トレースを脱ぐこと (dissolving the trace) により、バタチャリヤ距離行列  $DDB$  を式 (15) で表す。

[数8]

$$D_B = \frac{1}{4}(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T(\Sigma_1 + \Sigma_2)^{-1} + \frac{1}{2} \left[ \ln \left( \frac{\Sigma_1 + \Sigma_2}{2} \right) - \ln(\Sigma_1^{-\frac{1}{2}} \Sigma_2^{-\frac{1}{2}}) \right] \quad (15)$$

[0024] 2 つの正規分布  $N(\mu_1, \Sigma_1)$  と  $N(\mu_2, \Sigma_2)$  の間のチャノフ距離  $DC$  は、次式として定義される。これを後述の「10. 付録 バタチャリヤ距離  $DB$  の導出 (式 (7))」に示す方法で同様に導出できる。

[数9]

$$\begin{aligned} D_C &= \frac{1}{2} s(1-s)(\mu_1 - \mu_2)^T [(1-s)\Sigma_1 + s\Sigma_2]^{-1} (\mu_1 - \mu_2) \\ &\quad + \frac{1}{2} \ln \{ \det \Sigma_1^{-s} \det [(1-s)\Sigma_1 + s\Sigma_2] \det \Sigma_2^{-s} \} \\ &= \operatorname{tr} \left\{ \frac{1}{2} s(1-s)(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T [(1-s)\Sigma_1 + s\Sigma_2]^{-1} \right. \\ &\quad \left. + \frac{1}{2} \{ \ln [(1-s)\Sigma_1 + s\Sigma_2] - \ln(\Sigma_1^{-s} \Sigma_2^s) \} \right\} = \operatorname{tr} D_C, \quad s \in [0,1] \end{aligned} \quad (11)$$

[0025] トレース形式への変換後に、トレースを脱ぐことにより、対応する距離行列  $DDC$  は式 (16) として得られる。

[数10]

$$\begin{aligned} D_C = & \frac{1}{2}s(1-s)(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T[(1-s)\Sigma_1 + s\Sigma_2]^{-1} \\ & + \frac{1}{2}\{\ln[(1-s)\Sigma_1 + s\Sigma_2] - \ln(\Sigma_1^{1-s}\Sigma_2^s)\}, \quad s \in [0,1] \end{aligned} \quad (16)$$

チャーノフ距離  $D_C$  とその距離行列  $D D C$  は、 $D B$  と  $D D B$  から拡張されたものと考え、計算要求に応じて指数係数  $s$  を調整できるため、複雑なデータに対して柔軟に適応する能力を持つと考えられる。

[0026] 2つの正規分布  $N(\mu_1, \Sigma_1)$  と  $N(\mu_2, \Sigma_2)$  の間のカルバック・ライブラーダイバージェンス  $D_{KL}$  は次式のように定義される。

[数11]

$$\begin{aligned} D_{KL} = & \frac{1}{2}(\mu_1 - \mu_2)^T(\Sigma_1 + \Sigma_2)^{-1}(\mu_1 - \mu_2) \\ & + \frac{1}{2}\text{tr}(\Sigma_1^{-1}\Sigma_2 + \Sigma_2^{-1}\Sigma_1 + 2I_d) \\ = & \text{tr}\left\{\frac{1}{2}(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T(\Sigma_1 + \Sigma_2)^{-1} \right. \\ & \left. + \frac{1}{2}(\Sigma_1^{-1}\Sigma_2 + \Sigma_2^{-1}\Sigma_1 + 2I_d)\right\} = \text{tr}D_{KL} \end{aligned} \quad (12)$$

[0027] ここで、 $I_d$  は  $d$ -次元の単位行列である。式 (12) のようにトレース形式を記述した上で、式 (17) のようにトレースを脱ぐことにより、対応する距離行列  $D D K L$  を得る。式 (17) 第2項に対数演算はないことに注意すべきである。

[数12]

$$D_{KL} = \frac{1}{2}(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T(\Sigma_1 + \Sigma_2)^{-1} + \frac{1}{2}(\Sigma_1^{-1}\Sigma_2 + \Sigma_2^{-1}\Sigma_1 + 2I_d) \quad (17)$$

[0028] 2つの正規分布  $N(\mu_1, \Sigma_1)$  と  $N(\mu_2, \Sigma_2)$  の間のヘリンジャー距離行列  $D D H$  を次式のように求めてみる。

[数13]

$$\begin{aligned}
 D_H &= [1 - \exp(-D_E)]^{\frac{1}{2}} \\
 &= [1 - \exp(-\text{tr} \mathbf{D}_E)]^{\frac{1}{2}}
 \end{aligned}
 \tag{13}$$

しかし、式（13）は、 $\mathbf{D} \mathbf{D} \mathbf{B}$ に関する関数と見なすことができるが、完全なトレース形式に変換できないため、ヘリンジャー距離 $D_H$ に対して距離行列 $\mathbf{D} \mathbf{D} \mathbf{H}$ がないことが分かった。ここで、各統計的距離行列は、 $\mathbf{D} \mathbf{D} := [\delta_{uv}] \in \mathbb{U}^d \times d$ 、 $u, v = 1, \dots, d$ で定義される。

#### [0029] 4. ハードウェア及びプログラム

図3は、本実施の形態に関するハードウェアの構成図である。

このハードウェアは、中央処理装置（CPU）である処理部11、入力部12、出力部13、表示部14、記憶部15、インタフェース部（I/F）16を有する。また、処理部11、入力部12、出力部13、表示部14、記憶部15、インタフェース部（I/F）16は、スター又はバス等の適宜の接続手段で接続されている。記憶部15は、入力データファイル151、統計的距離行列ファイル152、距離行列ファイル153、クラスタ行列ファイル154、出力ファイル155等の各種ファイルを含む。

記憶部15の各ファイルに、「1. 記号の説明」で示したデータ及びその他の適宜のデータが必要に応じて記憶することができる。処理部11は、記憶部15に記憶されたデータを必要に応じて読み出し及び／又はそこに書き込むことができる。処理部11は、必要に応じて、入力部12、I/F16からデータを入力することができる。処理部101は、入出力データに基づいて、本実施の形態の各処理を実行し、例えば、各種の統計的距離行列の計算、階層的クラスタリング処理、類似性判定等を実行することができる。ま

た、処理部 11 は、必要に応じて、出力部 13 に出力し、I/F 16 を介して他の装置・部等へデータを出し、表示部 14 にデータを出し、出力することができる。

[0030] 本発明及び／又は実施の形態の統計的距離行列の計算方法又は装置・システムは、その各手順をコンピュータに実行させるため統計的距離行列の計算プログラム、計算プログラムを記録したコンピュータ読み取り可能な記録媒体、統計的距離行列の計算プログラムを含みコンピュータの内部メモリにロード可能なプログラム製品、そのプログラムを含むサーバ等のコンピュータ、等により提供されることができる。

また、本発明及び／又は実施の形態の統計的距離行列の可視化方法又は装置・システムは、その各手順をコンピュータに実行させるため統計的距離行列の可視化プログラム、可視化プログラムを記録したコンピュータ読み取り可能な記録媒体、統計的距離行列の可視化プログラムを含みコンピュータの内部メモリにロード可能なプログラム製品、そのプログラムを含むサーバ等のコンピュータ、等により提供されることができる。

[0031] 5. 統計的距離行列の計算・可視化及び応用例

#### 5. 1. 第 1 の実施の形態

図 4 に、第 1 の実施の形態の統計的距離行列の計算・可視化方法に関する技術についての説明図を示す。

図 5 に、第 1 の実施の形態の統計的距離行列の計算・可視化方法に関するフローチャートを示す。

[0032] 以下に、処理部 11 が実行する各ステップでの処理について説明する。

(ステップ S101)

処理部 11 は、入力部 12、記憶部 15 (入力データファイル 151)、又は、I/F 16 を介して他の装置から、比較対象となるデータを入力する。

入力 0 及び出力 0 は、d 個の要素で構成された特徴の 2 つの母集団 (デー

タ)である。

ここで、入力された母集団（データ）の特徴の形は、特徴ベクトル、特徴行列、あるいは特徴集合のうちいずれであってもよい。

(ステップS102)

処理部11は、入力データの特徴がベクトルである場合（条件1）、ステップS105に移行する。

ここで、

入力1：特徴ベクトルの2つの母集団  $\{a_k\}$  と  $\{b_k\}$

出力1：特徴ベクトルの2つの母集団  $\{a_k\}$  と  $\{b_k\}$

となる。

ここで、

$$a_t := \{a'_t\}, \quad t = 1, \dots, d$$

$$\{a_k\}, \quad k = 1, \dots, N$$

$$b_t := \{b'_t\}, \quad t = 1, \dots, d$$

$$\{b_k\}, \quad k = 1, \dots, N$$

処理部11は、入力データの特徴が行列である場合（条件2）、ステップS103に移行する。

処理部11は、入力データの特徴が集合である場合（条件3）、ステップS104に移行する。

[0033] (ステップS103)

処理部11は、入力データの特徴が行列である場合（条件2）、行列を  $vec(\cdot)$  によりベクトル化する。

ここで、

入力2：特徴行列の2つの母集団  $\{A_k\}$  と  $\{B_k\}$

出力1：特徴ベクトルの2つの母集団  $\{a_k\}$  と  $\{b_k\}$

となる。

ここで、

$$\begin{aligned} AA &:= \{a_{ij}\}, \quad i = 1, \dots, m, \quad j = 1, \dots, n, \\ &\{AA_k\}, \quad k = 1, \dots, N \\ BB &:= \{b_{ij}\}, \quad i = 1, \dots, m, \quad j = 1, \dots, n, \\ &\{BB_k\}, \quad k = 1, \dots, N \end{aligned}$$

処理部11は、関数  $\text{vec}(\cdot)$  より、入力された特徴行列  $\{AA_k\}$ 、 $\{BB_k\}$  をベクトル化して特徴ベクトル  $\{aa_k\}$ 、 $\{bb_k\}$  とする。すなわち、 $m$  行  $n$  列の行列を順次並べて、 $d$  個 ( $d = m \times n$ ) の要素のベクトルを形成する。例えば、1列目を1～ $m$ 番目、2列目を $m+1$ ～ $2m$ 番目、3列目を $2m+1$ ～ $3m$ 番目、 $\dots$ 、というように並べて行列化する。なお、行ごとに順番に並べるようにしても良い。

[0034] (ステップS104)

処理部11は、入力データの特徴が集合である場合(条件3)、集合の要素をベクトルで並べる。

ここで、

入力3：特徴集合の2つの母集団  $\{AA'_k\}$  と  $\{BB'_k\}$

出力1：特徴ベクトルの2つの母集団  $\{aa_k\}$  と  $\{bb_k\}$

となる。

ここで、 $\{AA'_k\}$ 、 $\{BB'_k\}$  は集合を表す。

$$\begin{aligned} AA' &:= \{a'_t\}, \quad t = 1, \dots, d \\ &\{AA'_k\}, \quad k = 1, \dots, N \\ BB' &:= \{b'_t\}, \quad t = 1, \dots, d \\ &\{BB'_k\}, \quad k = 1, \dots, N \end{aligned}$$

なお、例えば、後述の例のように、入力データが画像である場合、特徴行列が入力され、ステップS103が実行される。また、入力データの特徴を予め定めて限定することにより、定められた特徴によりステップS102、

S 1 0 3, S 1 0 4 のいずれかひとつ又は2つの処理を省略するようにしてもよい。

[0035] 一例として、C I F A R - 1 0 データセット[6]を使用し、統計的距離行列の効果テストを行った。このデータセットには、R G B 画像サイズは3 2 × 3 2 ピクセル；1 0 クラスの画像がそれぞれ6、0 0 0 枚、計6 0、0 0 0 枚の画像が含まれている。この例では、処理部1 1 は、ステップS 1 0 1 で画像A A とB B を入力し、ステップS 1 0 2 で特徴行列と判定し、ステップS 1 0 3 を実行する。

図6は、C I F A R - 1 0 データセット[6]における飛行機、鳥、猫、および犬の画像の例を示す図である。

図に示すように、計算を簡略化して区別可能な結果を得るために、2クラスの類似度は、弱いものから昇順で、飛行機と犬、鳥と犬、および猫と犬の間の距離行列を計算した。すべての画像 $AA := [a_{ij}] \in U^{m \times n}$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, n$ に対して、画像ピクセルの値域が $U = [0, 1]$ に指定され、画像の形が、ベクトル化する関数 $vec(\cdot)$ により、特徴要素のベクトル $aa := [a^t] = vec AA \in R^d \times 1$ ,  $t = 1, \dots, d$ のように再構成される。ここで、 $d = m \times n$ 。そして、2クラスの再構成された画像セット $\{aa_k\}$ と $\{bb_k\}$ ,  $k = 1, \dots, N$ は、それぞれ正規分布 $N(\mu\mu_1, \Sigma\Sigma_1)$ と $N(\mu\mu_2, \Sigma\Sigma_2)$ に従い、 $d$ -次元のランダムベクトル $X_1$ と $X_2$ を持つ2つの母集団と考えられる。ここで、 $X_1$ と $X_2$ の平均ベクトル $\mu\mu_1, \mu\mu_2 \in R^d$ , および共分散行列 $\Sigma\Sigma_1, \Sigma\Sigma_2 \in R^d \times d$ は、式(19)と(37)により得られる。

この例では、画像セット $\{aa_k\}$ と $\{bb_k\}$ が、飛行機と犬、鳥と犬、および猫と犬の順に設定された。各クラスにおいて、トレーニング画像のみが使用されるため、 $N$ は5 0 0 0とした。

[0036] (ステップS 1 0 5)

処理部1 1 は、式(19)、(37)より母集団の平均ベクトルと共分散

行列を算出する。

ここで、

入力：ステップS 1 0 2、S 1 0 3、S 1 0 4のいずれかの出力1

出力2：特徴ベクトルの2組の平均ベクトルと共分散行列 $\mu \mu 1$ 、 $\Sigma \Sigma 1$ と $\mu \mu 2$ 、 $\Sigma \Sigma 2$

となる。

本発明及び／又は本実施の形態では、統計的距離行列の計算に、一例として、2組の平均ベクトルと共分散行列 $\mu \mu 1$ 、 $\Sigma \Sigma 1$ と $\mu \mu 2$ 、 $\Sigma \Sigma 2$ を用いた。しかし、これらのパラメータは限定されるものではなく、データのサイズ等に合わせて適宜定めることができるものとする。

[0037] (ステップS 1 0 6)

処理部1 1は、式(1 4)、(1 5)、(1 6)、(1 7)の予め定められたいずれかひとつ又は複数より統計的距離行列を算出する。

ここで、

入力：ステップS 1 0 5の出力2

出力3：統計的距離行列D D

となる。

ここで、

$$D D := \{\delta u v\}, \quad u, v = 1, \dots, d$$

式(1 4)、(1 5)、(1 6)、(1 7)のうち統計的距離行列D Dとしてどの式を使用するかは、適宜の段階で入力部1 2等から設定しても良いし、デフォルトで設定されていても良い。また、複数の式を設定した場合、以下の処理は各式に対応してそれぞれ実行する。

また、統計的距離については、上述の式(1 4)、(1 5)、(1 6)、(1 7)等の例の他にも、二次形式、行列式の自然対数関数などの項で構成された、周知、公知又は任意のトレース形式で表現できる統計的距離を用い

ることができ、その周知、公知又は任意の統計的距離を脱トレース操作により統計的距離行列を算出することができる。

[0038] 図7は、飛行機と犬、鳥と犬、および猫と犬のケースに対する統計距離行列DDM、DDKL、ddb、およびDDCを示す図である。

各ケースについて、式(19)と(37)で計算された平均ベクトル $\mu_1$ と $\mu_2$ 、および共分散行列 $\Sigma_1$ と $\Sigma_2$ を、式(14)－(17)にそれぞれ代入し、3つのケースに対して、4つの統計距離行列DDM、DDKL、ddb、およびDDCを計算した。

図に示すように、この例では、3つのケースに対して、すべてのDDMとDDKLは無秩序かつ情報量が少ないように見えた。対照的に、ddbと指数係数 $s$ が0.3に設定されたDDCの中央にある、高い値を持つローカル距離 $\delta_{uv}$ は、グリッドのようなパターンで表された。画像 $a_k$ および $b_k$ のピクセル $a_t$ および $b_t$ が、それぞれ可測空間内のランダムベクトル $X_1$ および $X_2$ の対応する要素と見なされると、これらの高い値を持つローカル距離は、 $a_k$ と $b_k$ の区別に、効果的に使用できる。ddbは、 $s$ が $1/2$ に設定されているDDCの特定のケースだとしたら、DDCのような統計的距離行列が有効であると考えられる。

[0039] (ステップS107)

処理部11は、式(20)より距離累積ベクトルを算出する。

ここで、

入力：ステップS106の出力3

出力4：距離累積ベクトル $\phi$

となる。

[数14]

$$\phi_t = \sum_{u=1}^d \delta_{ut} + \sum_{v=1}^d \delta_{tv} \quad (20)$$

すなわち処理部11は、各画像ピクセル $a_t$ と $b_t$ に対して、式(2

0)のように、関連するすべての又は予め定められたローカル距離を累積し、得られた値を新たなピクセルに割り当てて、距離累積ベクトル  $\phi \phi := [\phi^{\wedge} t] \in \mathbb{R}^{d \times 1}$ ,  $t = 1, \dots, d$  を作る。

(ステップS108)

処理部11は、関数  $\text{vec}^{\wedge}(\cdot)$  より距離累積ベクトルを順次並べて行列化して統計的距離行列とする。すなわち、 $d$  個の要素のベクトルを順次並べて、 $m \times n = d$  を満たす  $m$  行  $n$  列の行列を形成する。

ここで、

入力：ステップS107の出力4

出力5：距離累積行列  $\Phi \Phi$

となる。

ここで、距離累積行列（距離累積画像） $\Phi \Phi := [\phi_{ij}] = \text{vec}^{\wedge}(\phi \phi)$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, n$  は、距離累積ベクトル  $\phi \phi$  により構成され、距離行列を表現する方法として使用する。ここで、 $\text{vec}^{\wedge}(\cdot)$  は、 $\text{vec}(\cdot)$  の逆のプロセスとした。

処理部11は、関数  $\text{vec}^{\wedge}(\cdot)$  より、入力された距離累積ベクトル  $\phi \phi := [\phi^{\wedge} t]$ ,  $t = 1, \dots, d$  の  $d$  個の要素のベクトルを順次並べて、 $m \times n = d$  を満たす  $m$  行  $n$  列の行列を形成することで行列化を行い距離累積行列  $\Phi \Phi$  とする。すなわち、 $\phi^{\wedge} 1 \sim \phi^{\wedge} m$  を1列目、 $\phi^{\wedge} m+1 \sim \phi^{\wedge} 2m$  を2列目、 $\phi^{\wedge} 2m+1 \sim \phi^{\wedge} 3m$  を3列目、 $\dots$ 、というように並べて行列化する。なお、行ごとに順番に並べるようにしても良い。

[0040] 図8は、飛行機と犬、鳥と犬、および猫と犬のケースに対する距離累積画像  $\Phi \Phi M$ 、 $\Phi \Phi KL$ 、 $\Phi \Phi B$ 、および  $\Phi \Phi C$  を示す図である。

統計的距離行列の効果に対して、それ自体よりも、距離累積画像は、もっと明瞭に反映できる。図に示すように、DDMとDDKLの距離累積画像  $\Phi M$  と  $\Phi \Phi KL$  は無秩序であり、ddbとddcの距離累積画像  $\Phi \Phi B$  と  $\Phi$

中Cは規則的なパターンを示した。主な表現は、高値のピクセルがすべて距離累積画像の中心に集中し、円または楕円に似た分布を示すことである。飛行機や犬などの類似性が低いクラスの場合、高値のピクセルの数が多くなり、それらの位置が画像の中心により集中している。対照的に、猫や犬などの類似性が高いクラス間では、高値のピクセルは少なく、中値のピクセルと一緒に画像の中心に広く集中している。したがって、距離行列イメージング法は、2クラスに対して画像ピクセル間の差異の大きさと位置を同時に定量化できる。

[0041] (ステップS109)

ステップS106の後、一方で、処理部11は、統計的距離行列DDに対して階層的クラスタリングにより特徴要素のクラスタを算出し、クラスタラベル集合を求める。すなわち、処理部11は、 $d$ 行 $\times$  $d$ 列の統計的距離行列DDに基づいて、 $d$ 個の要素IDに付けられたクラスタラベルの集合を求める。

ここで、

入力：ステップS106の出力3

出力6： $j$ 個の要素クラスタを表すクラスタラベル集合 $S_{d-j}$

となる。

処理部11は、関数 $vec^*(\cdot)$ より、入力されたクラスタラベル集合にある $d$ 個の要素のクラスタラベルを、ベクトルのように要素ID順に並べて、 $m \times n = d$ を満たす $m$ 行 $n$ 列の行列を形成することで行列化してクラスタ行列とする。すなわち、1番目 $\sim m$ 番目の要素を1列目、 $m+1$ 番目 $\sim 2m$ 番目の要素を2列目、 $2m+1$ 番目 $\sim 3m$ 番目の要素を3列目、 $\dots$ 、というように並べて行列化する。なお、行ごとに順番に並べるようにしても良い。

[0042] 図9に、階層的クラスタリングの処理についての説明図を示す。

この例では、図9(A)のように、処理部11は、統計的距離行列DDは9行 $\times$ 9列であり、行番号又は列番号に対応する9個の要素ID(C1, C

2, . . . , C 9) に付けられたクラスタラベルの集合を求める。

ここでは一例として、図 9 (B) のように樹形図が求められ、さらに、図 9 (C) のように、各要素 I D に対してクラスタラベルが付与されている。

C 1 : “1”、 C 2 : “3”、 C 3 : “1”、  
C 4 : “3”、 C 5 : “2”、 C 6 : “3”、  
C 7 : “1”、 C 8 : “3”、 C 9 : “1”、

[0043] 階層的クラスタリングの処理は、周知又は公知又は適宜の処理を用いることができる。

例えば、階層的クラスタリングについて、以下のサイトにわかりやすい例が示される。

<http://www.snap-tck.com/room04/c01/stat/stat20/stat2002.html>

以下に、これを参照して概要を説明する。

統計的距離行列 D D に対して、行番号又は列番号に対応する要素 I D を有する。最初に、各要素に違うクラスタラベルを付けるため、各クラスタは 1 つの要素のみを含む。処理部 1 1 は、入力した統計的距離行列 D D において距離  $\delta_{uv}$  が最も近い 2 つのクラスタを融合する。融合された 2 つのクラスタと他のクラスタとの距離を比較し、長いものは、融合後に生成した新たなクラスタと他のクラスタとの距離として、統計的距離行列 D D を更新する。さらに処理部 1 1 は、更新した統計的距離行列 D D にさらに距離  $\delta_{uv}$  が最も近い 2 つのクラスタを融合し、長い方の距離でクラスタを更新し統計的距離行列 D D を更新するという処理を、1 つのクラスタになるまで繰り返す。これにより、処理部 1 1 は、クラスタラベルの樹形図を形成し、樹形図に従い予め設定されクラスタ数に応じてグループ分けをして要素 I D に対するクラスタラベルを付する。この結果をクラスタラベル集合として形成する。処理部 1 1 は、さらにクラスタラベル集合を最初の要素 I D 順に並べ、行列化してクラスタ行列を形成することができる。処理部 1 1 は、クラスタ行列を記憶部 1 5 (クラスタ行列ファイル 1 5 4) に記憶することができる。

また、階層的クラスタリングに関する処理の一例として、後述の「9. 距離行列のクラスタリング」で説明する。

さらに、階層的クラスタリングに限らず、周知又は公知又は適宜のクラスタリングの処理を用いてもよい。

[0044] 図10は、飛行機と犬のケースに対して、統計的距離行列DDBを使用した階層的クラスタリング結果を示す図である（ラベル付き距離累積画像ΦΦB。(a)：3つのクラスタ；(b)：10つのクラスタ）。

図11は、鳥と犬のケースに対して、統計的距離行列DDBを使用した階層的クラスタリング結果を示す図である（ラベル付き距離累積画像ΦΦB。(a)：3つのクラスタ；(b)：10つのクラスタ）。

図12は、猫と犬のケースに対して、統計的距離行列DDBを使用した階層的クラスタリング結果を示す図である（ラベル付き距離累積画像ΦΦB。(a)：3つのクラスタ；(b)：10つのクラスタ）。

ここで、処理部11は、統計的距離行列DDBに基づいて、「9. 距離行列のクラスタリング」で後述するような階層的クラスタリングアルゴリズムを使用し、その距離累積画像ΦΦBのピクセル（特徴要素）をクラスタリングした。これらの図に示すように、飛行機と犬、鳥と犬、および猫と犬の3つのケースに対して、それぞれの中ΦBのピクセルを3つおよび10つのクラスタに分離し、ラベルを付けた。得られたクラスタパターンは、すべて円形または正方形（又は、楕円形、長方形等）で対称であり、中心点から放射状に広がる。このようなクラスタパターンはマンダラと見なすことができる。したがって、統計的距離行列のクラスタリング結果を直観的に説明するために、「情報マンダラ」という用語を確立した。（なお、マンダラという言葉は、「聖なる円」を意味するサンスクリット語の用語である。ヒンズー教、仏教、ジャイナ教、神道などのさまざまな宗教的伝統において、マンダラは、楽園、神々、または実際の神社を表すための図として使用されている。マンダラは円形または正方形であり、中心点から放射状に広がる色、形、およびパターンで設計されている。マンダラは正確に測定されると、幾何学的

な対称性を持つ。)

[0045] (ステップS 1 1 0)

処理部 1 1 は、: 行列要素にクラスタラベル (要素ラベル) を付ける。

ここで、

入力: ステップS 1 0 9の出力6、S 1 0 8の出力5

出力7: 類似性によりマンダラパターン

となる。

処理部 1 1 は、このステップで、関数  $\text{vec}^{\wedge}(\cdot)$  よりクラスタ集合を行列化してクラスタ行列としてもよい。すなわち、 $d$  個の要素のクラスタラベルを、ベクトルのように要素  $I$   $D$  順に並べて、 $m \times n = d$  を満たす  $m$  行  $n$  列の行列を形成することができる。あるいは、処理部 1 1 は、ステップS 1 0 9又はS 1 1 0で、実際にクラスタ行列を形成しなくても、クラスタ集合を距離累積行列  $\Phi$  と対応付けるようにしてもよい。

例えば、図9の例では、処理部 1 1 は、図9 (D) に示したように、クラスタラベル集合はベクトル化されると、9行であり、これを  $\text{vec}^{\wedge}(\cdot)$  の処理により3行×3列のクラスタ行列として、統計的距離行列  $DD$  は3行×3列に対応したクラスタラベルを付ける。

[0046] (ステップS 1 1 1)

処理部 1 1 は、クラスタラベル付きの距離累積行列  $\Phi$  を記憶部 1 5 (距離累積行列ファイル 1 5 3) に記憶することができる。また処理部 1 1 は、距離累積行列  $\Phi$  を、例えば画像として、表示部 1 4 に表示させ、及び/又は、出力部 1 3 若しくは  $I/F$  1 6 により結果を出力することができる。

ここで、

入力: S 1 1 0の出力7

出力: S 1 1 0の出力7

となる。

(ステップS 1 1 2)

処理部 11 は、距離累積行列中のパターンにより a a と b b との類似性を判定する。処理部 11 は、判定結果を、出力ファイルに 154 記憶することができる。類似性の判定の具体的処理については、「6. 距離累積画像におけるマンダラパターンの定量化」で後述する。

なお、処理部 11 は、各ステップ S 101 ～ S 112 等で計算・求めた各データ・出力結果等のいずれか又は複数を、適宜記憶部 15 又は記憶部 15 内の各ファイル（入力データファイル 151、統計的距離行列ファイル 152、距離行列ファイル 153、クラスタ行列ファイル 154、出力ファイル 155 等）に記憶してもよく、そこから必要に応じて読み出すことができる。また、処理部 11 は、各ステップ S 101 ～ S 112 等で計算・求めた各データ・出力結果等のいずれか又は複数を、表示部 14 に表示させ、及び／又は、出力部 13 若しくは I / F 16 により結果を出力してもよい。

[0047] 5. 2. 第 2 の実施の形態

図 20 に、第 2 の実施の形態の統計的距離行列の可視化方法に関するフローチャートを示す。

第 2 の実施の形態は、第 1 の実施の形態におけるステップ S 109 及び S 110 を省略したものである。

(ステップ S 101) ～ (ステップ S 108)

処理部 11 は、第 1 の実施の形態と同様の処理を実行する。

(ステップ S 111)

処理部 11 は、距離累積行列中のを、例えば画像として、表示部 14 に表示させ、及び／又は、出力部 13 若しくは I / F 16 により結果を出力することができる。（なお、クラスタラベルは付されていない。）

(ステップ S 112)

処理部 11 は、第 1 の実施の形態と同様の処理を実行する。

その他、処理の詳細は、第1の実施の形態と同様である。

[0048] 5. 3. 第3の実施の形態

図21に、第3の実施の形態の統計的距離行列の可視化方法に関するフローチャートを示す。

第3の実施の形態は、第1の実施の形態におけるステップS107及びS108を省略し、ステップS110をステップS110-2に置き換えたものである。

(ステップS101) ~ (ステップS106)、(ステップS109)

処理部11は、第1の実施の形態と同様の処理を実行する。

(ステップS110-2)

処理部11は、関数  $\text{vec}^{\wedge}(\cdot)$  よりクラスタラベル集合を行列化して統計的距離行列とする。すなわち、 $d$  個のクラスタラベル集合を順次並べて、 $m \times n = d$  を満たす  $m$  行  $n$  列の行列を形成する。

ここで、

入力：ステップS109の出力4

出力5：クラスタ行列

となる。

(ステップS111)

処理部11は、クラスタラベル行列を出力する。

(ステップS112)

処理部11は、第1の実施の形態と同様の処理を実行する。

その他、処理の詳細は、第1の実施の形態と同様である。

[0049] 5. 4. 第3の実施の形態の効果補足

距離累積行列ΦΦにおける各要素の値は、距離行列にある、その要素とすべての要素のローカル距離を累積した値である。したがって、距離累積行列ΦΦは、距離行列を圧縮して得たものと考えられ、距離行列より情報量は少ない。距離累積行列ΦΦの中に、ローカル距離の形跡も存在していないので、距離累積行列ΦΦから、距離行列を復元できない。

一方、クラスタリング結果としての樹形図は、距離行列にあるローカル距離の大きさによって、要素同士を階層的に並べたものなので、各要素の間の相対的な距離関係をすべて含んでいる。したがって、樹形図から、元の絶対的な距離行列でなく、相対的な距離行列を復元することができる。樹形図は、距離累積行列ΦΦとの関係は薄い。また、樹形図の計算に生成された $d$ 個の各 $j$ 個の要素クラスタを表すクラスタラベル集合  $\{S_{d-j}\}$  ,  $j = d - i$  ,  $i = 1, \dots, d$  は、マンダラ模様を表すのに使用することができる。

距離累積行列ΦΦとクラスタリング結果は両方ともマンダラ模様を表しているが、それらの模様は異なる。例えば、図12の(b)に示すように、クラスタ2と8のラベルを付けたΦの要素は、近い値を持っているので、簡単な閾値の設定によっては、区別ができないかもしれないが、10個の要素クラスタを表す、クラスタラベル集合(要素ラベル集合)  $S_{d-j}$  ( $d = 32 \times 32$  ,  $j = 10$ ) により区別することができる。

第1の実施の形態では、結果の見え方を一層わかりやすくするために、2種類のマンダラ模様を統合したが、第2の実施の形態又は第3の実施の形態のように、統合をしなくても構わない。

#### [0050] 6. 距離累積画像におけるマンダラパターンの定量化

(距離累積画像のセグメンテーションによる非類似度(マンダラパターンの強さ)の測定)

処理部11は、以下の処理を実行することで距離累積行列(距離累積画像)におけるマンダラパターンの定量化(距離累積画像のセグメンテーション

による非類似度（マンダラパターンの強さ）の測定）を実現することができる。

距離累積行列（距離累積画像） $\Phi\Phi := [\phi_{ij}] = \text{vec}^{\wedge}(\phi\phi)$ ,  $i = 1, \dots, m, j = 1, \dots, n$ のすべての要素に対しては、二つのしきい値で累積画像のセグメンテーションを行い、画像から値の低い、普通、および高い、合計3種類の要素領域を切り出す。値の高い要素領域と普通の要素領域を結合して得た領域は、多くの情報を持っているため、有効要素領域と呼ばれている。値の高い要素領域と有効要素領域の面積比（値の高い要素領域の要素数に対する有効要素領域の要素数の割合）を計算して求めた数値は、2クラスの非類似度、すなわち、マンダラパターンの強さと考えられる。

値の低い要素と普通の要素のしきい値 $\theta_1$ 、および普通の要素と高い要素のしきい値 $\theta_2$ は、それぞれ次式によって得られる。

$$\theta_1 := (1/3) \phi_{\max}, \quad \theta_2 := (2/3) \phi_{\max} \quad (42)$$

ここで、 $\phi_{\max}$ は、距離累積画像 $\Phi\Phi$ において、要素の最大値とする。それらのしきい値に基づいて、値の低い、普通、および高い、合計3種類の要素領域 $R_{\text{low}}$ 、 $R_{\text{medium}}$ 、および $R_{\text{high}}$ は、それぞれ次式で定義される。

$$R_{\text{low}} := \{ (i, j) \mid \phi(i, j) \leq \theta_1, \\ i = 1, \dots, m, j = 1, \dots, n \} \quad (43)$$

$$R_{\text{medium}} := \{ (i, j) \mid \phi(i, j) \leq \theta_2, \\ i = 1, \dots, m, j = 1, \dots, n \} \quad (44)$$

$$R_{\text{high}} := \{ (i, j) \mid \theta_2 \leq \phi(i, j), \\ i = 1, \dots, m, j = 1, \dots, n \} \quad (45)$$

そして、値の高い要素領域  $R_{high}$  と有効要素領域  $R_{medium} + R_{high}$  の面積比  $\rho$  は、 $R_{high}$  と  $R_{medium} + R_{high}$  の要素数とした  $N_{high}$  と  $N_{medium} + N_{high}$  の比率：

$$\rho := N_{high} / (N_{medium} + N_{high}) \quad (46)$$

で表される。

[0051] 図13は、飛行機と犬のケースに対して、距離累積画像ΦΦBから値の低い、普通、高い、合計3種類の要素領域を切り出した結果を示す図である（黒色：値の低い要素領域  $R_{low}$ ；灰色：値が普通の要素領域  $R_{medium}$ ；白色：値の高い要素領域  $R_{high}$ ）。

図14は、鳥と犬のケースに対して、距離累積画像ΦΦBから値の低い、普通、高い、合計3種類の要素領域を切り出した結果を示す図である（黒色：値の低い要素領域  $R_{low}$ ；灰色：値が普通の要素領域  $R_{medium}$ ；白色：値の高い要素領域  $R_{high}$ ）。

図15は、猫と犬のケースに対して、距離累積画像ΦΦBから値の低い、普通、高い、合計3種類の要素領域を切り出した結果を示す図である（黒色：値の低い要素領域  $R_{low}$ ；灰色：値が普通の要素領域  $R_{medium}$ ；白色：値の高い要素領域  $R_{high}$ ）。

これらの図に示すように、飛行機と犬、鳥と犬、および猫と犬のケースに対して、式(43)－(45)でそれぞれ  $N_{low}$ 、 $N_{medium}$ 、 $N_{high}$ 、および  $N_{medium} + N_{high}$  を求めてから、式(46)で面積を算出する。その結果は、以下の図##の通りである。

[0052] 図16は、 $\rho$  の計算結果の例を示す図である。

飛行機と犬、鳥と犬、および猫と犬のケース  $\rho$  は、値を高い方から低い順に並べており、人間の目と脳で2クラスの差異を認識する感覚と同じなので、2クラスの非類似度を表す尺度と考えられる。

(まとめ)

したがって、距離行列イメージング法で得られた距離累積画像中に対して、この「距離累積画像におけるマンダラパターンの定量化（距離累積画像のセグメンテーションによる非類似度（マンダラパターンの強さ）の測定）」の手法を使用すると、画像ピクセル間の差異の強さ（非類似度）や、位置、サイズおよび対称性（非類似性）等を同時に定量化することができた。その差異は、マンダラパターンのように、例えば、楕円形で対称であり、中心点から放射状に広がることを確認できた。

[0053] 7. 関連技術との比較

図17に、統計的距離に関する関連技術1についての説明図を示す。

以下に、本発明の実施の形態と関連技術1とを比較する。この関連技術1では、処理部11は、入力された特徴ベクトルの2つの母集団  $\{a_k\}$  と  $\{b_k\}$  に対して、式(19)、(37)より、特徴ベクトルの2組の平均ベクトルと共分散行列  $\mu_1, \Sigma_1$  と  $\mu_2, \Sigma_2$  とを求める。つぎに、処理部11は、2組の平均ベクトルと共分散行列  $\mu_1, \Sigma_1$  と  $\mu_2, \Sigma_2$  を用いて、式(6)、(7)、(11)又は(12)より統計的距離Dを計算する。

これに対して、例えば、本発明の実施の形態では、関連技術1のようなスカラー値である統計的距離Dではなく、画像等を表現可能な行列である統計的距離行列DDを計算する点で、両者は異なる。

[0054] 図18に、相互相関行列に関する関連技術2についての説明図を示す。

以下に、本発明の実施の形態と関連技術2とを比較する。この関連技術2では、処理部11は、入力された特徴ベクトルの2つの母集団  $\{a_k\}$  と  $\{b_k\}$  に対して、式(19)、(37)より、特徴ベクトルの2組の平均ベクトルと共分散行列  $\mu_1, \Sigma_1$  と  $\mu_2, \Sigma_2$  とを求める。つぎに、処理部11は、2組の平均ベクトルと共分散行列  $\mu_1, \Sigma_1$  と  $\mu_2, \Sigma_2$  を用いて、式(38)より相互相関行列RRを計算する。

これに対して、例えば、本発明の実施の形態では、関連技術2のような相

互相関行列  $R$  ではなく、画像等を表現可能な行列である統計的距離行列  $D$  を計算する点で、両者は異なる。

[0055] 8. 検討

まず、統計的距離行列  $DD$  および  $DDC$  が特徴距離測定に有効である理由について説明する。 $DD$  と  $DDC$  には、平均ベクトル  $\mu_1$  と  $\mu_2$ 、および共分散行列  $\Sigma_1$  と  $\Sigma_2$  に関する 2 次項がある。このような 2 次項は、 $DDM$  および  $DDKL$  にも存在する。一方、 $DD$  と  $DDC$  には、共分散比の対数のみを含み、平均ベクトルを含まない項もある。 $DDM$  には、共分散行列を含む項がなく、 $DDKL$  には、共分散比を含む項があるが、対数的に計算されていない。2 つの平均ベクトルが等しい、もしくは近似する場合、2 次項の値はゼロになる傾向がある。つまり、2 つの確率分布が大きく重なり合う場合、共分散比の項は 2 次項よりも重要な役割を果たす。これが  $DD$  と  $DDC$  が有効な理由と考えられる。

次に、クラスタリングが必要な理由を説明する。クラスタリングの前に、統計的距離行列  $DD$  は、ランダムベクトル  $X$  の要素間のすべてのローカル距離を表している。多数のローカル距離がゼロに非常に近い場合、 $DD$  はスパースになりやすい。一方、グラフ理論において、 $DD$  は、頂点が  $X$  の要素として定義され、辺がローカル距離  $\delta$  に割り当てられている、有向グラフに射影できる。ただし、このような疎グラフに対して、多くのアプリケーションでは、小さい値を持つ辺を処理すると、計算が複雑になることが想定される。したがって、例えば、グラフ内の重要でない辺の数を減らし、値を再割り当てした重要な辺に従って、頂点を階層的に再配置する必要がある。このような変形で得た木構造により、データのアクセス速度が大幅に向上し、コンピュータのメモリ領域を節約できるようになるだろう。

最後に、統計距離行列と、[5]で提案されたニューラルネットワークの新しい便利なモデルである、カプセルニューラルネットワーク (CapsNet) の重み行列との関係を検討する。本発明及び／又は本実施の形態で提案さ

れた統計距離行列は、視点不変性を持っている。したがって、ターゲットの傾きが画像内でどれだけ変化しても、ターゲットの区別に影響を与えないので、統計距離行列をそのままに使用できる。ただし、CapsNetの重み行列と比べると、統計的距離行列は、距離累積画像に基づいてより直感的であり、階層的クラスタリング手法を使用することでより解釈できるようになってきた。したがって、統計的距離行列と、「情報マンダラ」として表されるそのクラスタリング結果は、重み行列より優れていると考えられる。

[0056] (まとめ)

以上のように、まず、画像の比較実験を通じて、DDCのような統計的距離行列が、ピクセルの画素値が特徴要素として直接に計算され、他の距離と比べて、より効果的にターゲットを区別できることを明らかにした。次に、新たに提案された、統計的距離行列の表現方法としての距離蓄積画像は、価値の高いピクセルが画像の中央に集中していることを示した。さらに、統計的距離行列を階層的クラスタリングすると、すべて又はほとんどのピクセルクラスターは基本的に画像の中心を囲み、距離値に従って内側から外側に放射状のように配置されることを発見した。これらのパターンはマンダラに非常に似ているため、統計的距離行列とそのクラスタリング結果を「情報マンダラ」と呼ぶ。「情報マンダラ」は新たなエントロピーであり、畳み込みニューラルネットワークを理解するための重要な手段と考えられる。

[0057] 9. 距離行列のクラスタリング

距離行列の処理を加速するために、通常の階層的クラスタリング[10]を導入し、統計的距離行列に基づいてランダムベクトルの要素をクラスタリングする。

階層的クラスタリングアルゴリズムへの入力は、統計的距離行列 $DD$ 、すなわち、クラスタラベル集合 $SS$ および距離関数 $\delta$ と考えられる。クラスタラベル集合 $S := \{1, \dots, d\}$ は、ランダムベクトル $X$ （特徴ベクトル）の次元数 $d$ によって与えられる。さらに、 $S$ に含まれるクラスタラベルを、特

微ベクトルの要素に順番に付ける。距離行列  $D$  において、要素  $\delta_{uv}$  の位置を表す添字  $u, v \in S$  を変数として扱い、 $\delta_{uv}$  を関数化して得た距離関数  $\delta$  の写像は、 $\delta : S \times S \rightarrow R$  とする。 $\delta(u, v)$  には、 $\delta_{uv}$  の値が割り当てられる。ここで、 $\delta(u, u)$  は 0 に設定される。集合  $S$  に  $d$  個の要素があれば、組み合わせの数 ( $d^2$ ) 個の距離が存在する。

階層的クラスタリングアルゴリズムの出力は、樹形図  $L$  とクラスタラベル集合 (要素ラベル集合) によって定義される。樹形図は、データ構造と見なすことができ、数学的なグラフとして表現される。本発明及び／又は本実施の形態では、樹形図を使用している。基数  $d = |S_0|$  を持つクラスタラベル集合  $S_0$  が与えられた場合、すなわち、 $S_0 := \{1, \dots, d\}$ 、樹形図  $L$  は、クラスタラベル  $n_i$  に対応する順序組  $\langle u_i, v_i, \delta(u_i, v_i) \rangle$ ,  $i = 0, \dots, d-2$  のリストで構成される。ここで、 $u_i, v_i \in S_i$ 。クラスタラベル集合  $S_0$  は、初期クラスタラベル集合であり、ステップ  $i+1$  におけるクラスタラベル集合  $S_{i+1}$  は、 $(S_i \setminus \{u_i, v_i\}) \cup n_i$  として再帰的に定義される。各ステップで、 $n_i$  というラベルを付いた新しいクラスタは、距離  $\delta(u_i, v_i)$  を持つ、 $u_i$  と  $v_i$  というラベルを付いたクラスタを結合することによって構成される。手順には、 $d-1$  ステップが含まれているため、最終状態では、 $d$  個のすべての初期クラスタが一つのクラスタに含まれる。なお、要素クラスタの数を  $j = d-i$ ,  $j = 1, \dots, d$  とすれば、クラスタラベル集合  $S_{d-j}$  に、 $j$  個の要素クラスタに関するクラスタリング結果が含まれている。

[0058] 図 19 に、階層的クラスタリングのフローチャートを示す。

以下に、提案されている階層的クラスタリングアルゴリズムが示されている。

処理部 11 は、以下のように、フローチャートの沿ったアルゴリズムの各ステップ処理を実行することで、階層的クラスタリングを実現することができる。

(ステップS201) 入力

処理部11は、統計的距離行列DDを入力する。

すなわち、

クラスタラベル集合: S0

距離関数:  $\delta$

(ステップS202)

処理部11は、初期化をする。

クラスタラベルの数:  $d \leftarrow |S0|$

樹形図:  $L \leftarrow$  (空きの樹形図)

[0059] (ステップS203)

処理部11は、for  $i=0$  から  $d-2$  まで、以下のステップS203  
~S210までの処理を繰り返す。

(ステップS204)

処理部11は、 $(u_i, v_i) \leftarrow \argmin_{S_i \times S_i \setminus \Delta_i} \delta$ 、を  
実行する。

$\Delta_i$  は  $S_i \times S_i$  の対角成分

(ここで、「arg min」は、区間  $S_i \times S_i \setminus \Delta_i$  において、  
目的関数  $\delta$  を最小化する引数  $(u_i, v_i)$  の値を与えることを表す。  
また、区間  $S_i \times S_i \setminus \Delta_i$  は、 $\Delta_i$  を除く  $S_i \times S_i$  を表す。  
。)

(ステップS205)

処理部11は、Lヘトリプル(順序組)  $\langle u_i, v_i, \delta(u_i, v_i) \rangle$   
を追加する。

(ステップS206)

処理部11は、 $S_i \leftarrow S_i \setminus \{u_i, v_i\}$  を実行する。

(この式は、 $\{u_i, v_i\}$  を除く集合  $S_i$  を  $S_i$  に代入することを表す。  
)

(ステップS207)

処理部11は、新しいクラスタラベル $n_i$  ( $S_i$ に属さない)を作成する。

(ステップS208)

処理部11は、次式を使ってすべての $x \in S_i$ に対して $\delta$ を更新する。

$$\delta(n_i, x) = \delta(x, n_i) := f(u_i, x), \delta(v_i, x)$$

(ステップS209)

処理部11は、 $S_i \leftarrow S_i \cup \{n_i\}$  を実行する。

(ステップS210)

処理部11は、for  $i = 0$  から  $d - 2$  まで、以下のステップS203 ~ S210までの処理を繰り返す(end for)。

[0060] (ステップS211)

処理部11は、次の結果を、表示部14に表示させ、及び／又は、出力部13若しくはI/F16により結果を出力する。

樹形図: L

各ステップ $i$ におけるクラスタラベル集合:  $S_i, i = 0, \dots, d - 1$

すなわち、

各 $j$ 個の要素クラスタを表すクラスタラベル集合:  $S_{d-j}, j = d - i, i = 1, \dots, d$

ここで、 $\delta$ を更新するための凝集公式は、次式のように定義される。

$$f(\delta(u_i, x), \delta(v_i, x)) := \max(\delta(u_i, x), \delta(v_i, x)) \quad (18)$$

適当なカットオフ閾値が指定されているとき、このアルゴリズムは、ランダムベクトル(特徴ベクトル)の要素に対して安定したクラスタリング結果を提供できる。

[0061] 10. 付録 バタチャリヤ距離DBの導出(式(7))

$d$ 次元のランダムベクトル  $X_1$  と  $X_2$  は、2つの正規分布  $N(\mu_1, \Sigma_1)$  と  $N(\mu_2, \Sigma_2)$  にそれぞれ従い、それらに対応する確率密度関数  $p(x)$  および  $q(x)$  が、式(21)のように定義される。また、これらの平方根の積を式(22)で表す。

[数15]

$$\begin{aligned} p(x) &:= \det(2\pi\Sigma_1)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)\right] \\ q(x) &:= \det(2\pi\Sigma_2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\right] \end{aligned} \quad (21)$$

$$\begin{aligned} p^{\frac{1}{2}}(x)q^{\frac{1}{2}}(x) &= \det(2\pi\Sigma_1)^{-\frac{1}{4}} \det(2\pi\Sigma_2)^{-\frac{1}{4}} \\ &\quad \times \exp\left[-\frac{1}{4}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) - \frac{1}{4}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\right] \end{aligned} \quad (22)$$

[0062] 実数空間  $\mathbb{R}^d$  において、 $d$ 次元の変数ベクトル  $x$  に対して式(22)を積分すると、次式を得る。

[数16]

$$\int_{\mathbb{R}^d} p^{\frac{1}{2}}(x)q^{\frac{1}{2}}(x)dx = \det(2\pi\Sigma_1)^{-\frac{1}{4}} \det(2\pi\Sigma_2)^{-\frac{1}{4}} \quad (23)$$

$$\times \exp\left[-\frac{1}{4}(\mu_1^T\Sigma_1^{-1}\mu_1 + \mu_2^T\Sigma_2^{-1}\mu_2)\right] \quad (24)$$

$$\times \int_{\mathbb{R}^d} \exp\left[-\frac{1}{4}x^T(\Sigma_1^{-1} + \Sigma_2^{-1})x + \frac{1}{2}(\mu_1^T\Sigma_1^{-1} + \mu_2^T\Sigma_2^{-1})x\right] dx \quad (25)$$

[0063] この式に、因数式(23)、(24)、および(25)をまとめて簡略化するために、まず、式(26)を準備する。式(26)の導出に、式(27)と(28)を使用した。

[数17]

$$\begin{aligned}
& \exp \left[ \frac{1}{8} (\mu_1^T \Sigma_1^{-1} + \mu_2^T \Sigma_2^{-1}) (\Sigma_1^{-1} \Sigma \Sigma_2^{-1})^{-1} (\Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2) \right] \\
&= \exp \left[ \frac{1}{8} (\mu_1^T \Sigma_1^{-1} \Sigma_2 \Sigma^{-1} \mu_1 + 2 \mu_1^T \Sigma^{-1} \mu_2 + \mu_2^T \Sigma_2^{-1} \Sigma_1 \Sigma^{-1} \mu_2) \right] \\
&= \exp \left[ \frac{1}{8} (2 \mu_1^T \Sigma_1^{-1} (I - \Sigma_1 (\Sigma_1 + \Sigma_2)^{-1}) \mu_1 + 2 \mu_1^T \Sigma^{-1} \mu_2 + 2 \mu_2^T \Sigma_2^{-1} (I - \Sigma_2 (\Sigma_1 + \Sigma_2)^{-1}) \mu_2) \right] \\
&= \exp \left[ \frac{1}{8} (2 \mu_1^T \Sigma_1^{-1} \mu_1 - \mu_1^T \Sigma^{-1} \mu_1 + 2 \mu_1^T \Sigma^{-1} \mu_2 + 2 \mu_2^T \Sigma_2^{-1} \mu_2 - \mu_2^T \Sigma^{-1} \mu_2) \right]
\end{aligned} \tag{26}$$

$$\begin{aligned}
(\Sigma_1 + \Sigma_2)^{-1} &= \Sigma_1^{-1} - \Sigma_1^{-1} (\Sigma_2^{-1} + \Sigma_1^{-1})^{-1} \Sigma_1^{-1} \\
&= \Sigma_2^{-1} - \Sigma_2^{-1} (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1} \Sigma_2^{-1}
\end{aligned} \tag{27}$$

$$\begin{aligned}
(\Sigma_1^{-1} + \Sigma_2^{-1}) &= \Sigma_1^{-1} \Sigma_2 \Sigma_2^{-1} + \Sigma_2^{-1} \Sigma_1 \Sigma_1^{-1} = \Sigma_1^{-1} (\Sigma_1 + \Sigma_2) \Sigma_2^{-1} \\
&= \Sigma_2^{-1} (\Sigma_1 + \Sigma_2) \Sigma_1^{-1}
\end{aligned} \tag{28}$$

[0064] また、式(27)は、式(29)より成り立つことに注意しておく。ここで、 $A A$ 、 $B B$ 、 $C C$ はすべて正定行列とする。また、 $\Sigma \Sigma$ は式(30)より定義される。

[数18]

$$(A + C B C^T)^{-1} = A^{-1} - A^{-1} C (B^{-1} + C^T A^{-1} C)^{-1} C^T A^{-1} \tag{29}$$

$$\Sigma := \frac{\Sigma_1 + \Sigma_2}{2} \tag{30}$$

[0065] 次に、以下のように、式(25)を式(31)に変換する。ここで、 $d$ 次元の変数ベクトル  $y y$  は、式(32)より定義される。

[数19]

$$\begin{aligned}
& \int_{\mathbb{R}^d} \exp \left[ -\frac{1}{2} x^T \Sigma^{-1} x + \frac{1}{2} (\mu_1^T \Sigma_1^{-1} + \mu_2^T \Sigma_2^{-1}) x \right] dx \\
&= \int_{\mathbb{R}^d} \exp \left[ -\frac{1}{2} y^T y + \frac{1}{2} (\mu_1^T \Sigma_1^{-1} + \mu_2^T \Sigma_2^{-1}) (\Sigma_1^{-1} \Sigma \Sigma_2^{-1})^{-\frac{1}{2}} y \right] d \left[ (\Sigma_1^{-1} \Sigma \Sigma_2^{-1})^{-\frac{1}{2}} y \right] \\
&= \underbrace{\text{Eqn. (26)}}_{\text{first factor}} \times \underbrace{\text{Eqn. (25) / Eqn. (26)}}_{\text{second factor}}
\end{aligned} \tag{31}$$

$$y := (\Sigma_1^{-1} \Sigma \Sigma_2^{-1})^{\frac{1}{2}} x \tag{32}$$

[0066] そして、式(24)と式(31)の1番目の因数式を掛けて、式(33)を得る。

[数20]

$$\begin{aligned} \text{Eqn. (24)} \times \text{Eqn. (26)} &= \exp \left[ -\frac{1}{8} (\mu_1^T \Sigma^{-1} \mu_1 - 2\mu_1^T \Sigma^{-1} \mu_2 + \mu_2^T \Sigma^{-1} \mu_2) \right] \\ &= \exp \left[ -\frac{1}{8} (\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2) \right] \end{aligned} \quad (33)$$

[0067] 一方、変数変換法によって、式(31)の2番目の因数式を式(34)のように変換できる。さらに、式(23)に式(34)を掛けて、式(35)を得られる。

[数21]

$$\begin{aligned} &\det \left[ (\Sigma_1^{-1} \Sigma \Sigma_2^{-1})^{-\frac{1}{2}} \right] \int_{\mathbb{R}^2} \exp \left\{ -\frac{1}{2} \left[ y - \frac{1}{2} (\Sigma_1^{-1} \Sigma \Sigma_2^{-1})^{-\frac{1}{2}} (\Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2) \right]^T \right. \\ &\quad \left. \times \left[ y - \frac{1}{2} (\Sigma_1^{-1} \Sigma \Sigma_2^{-1})^{-\frac{1}{2}} (\Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2) \right] \right\} dy \\ &= (2\pi)^{\frac{1}{2}} \det(\Sigma_1^{-1} \Sigma \Sigma_2^{-1})^{-\frac{1}{2}} \end{aligned} \quad (34)$$

$$\begin{aligned} \text{Eqn. (24)} \times \text{Eqn. (34)} &= \det(2\pi \Sigma_1)^{-\frac{1}{2}} \det(2\pi \Sigma_2)^{-\frac{1}{2}} (2\pi)^{\frac{1}{2}} \det(\Sigma_1^{-1} \Sigma \Sigma_2^{-1})^{-\frac{1}{2}} \\ &= (2\pi)^{-\frac{1}{2}} \det \Sigma_1^{-\frac{1}{2}} (2\pi)^{-\frac{1}{2}} \det \Sigma_2^{-\frac{1}{2}} (2\pi)^{\frac{1}{2}} \det(\Sigma_1^{-1} \Sigma \Sigma_2^{-1})^{-\frac{1}{2}} \\ &= \det \Sigma_1^{\frac{1}{2}} \det \Sigma^{-\frac{1}{2}} \det \Sigma_2^{\frac{1}{2}} \end{aligned} \quad (35)$$

[0068] したがって、式(36)のように、式(33)と式(35)の掛け算に関する対数の負数は、バタチャリヤ距離DBとする。

[数22]

$$-\ln(\text{Eqn. (33)} \times \text{Eqn. (35)}) = \frac{1}{8} (\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2) + \frac{1}{2} \ln \left[ \det \Sigma_1^{-\frac{1}{2}} \det \Sigma \det \Sigma_2^{-\frac{1}{2}} \right] \quad (36)$$

[0069] 11. 参考文献

なお、ウムラウト等のドイツ語のアルファベットは、電子出願ソフトで利用できる文字・記号の制約・都合上、英語のアルファベットで記載する。

[1] A. Bhattacharyya, On a measure of divergence between two statist

ical populations defined by their probability distributions. Bulletin of the Calcutta Mathematical Society, 35:99-109, 1943.

[2] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sums of observations. Annals of Mathematical Statistics, 23:409-507, 1952.

[3] Frank Harary, Robert Z. Norman, and Dorwin Cartwright. Structural models: an introduction to the theory of directed graphs. Wiley, New York, 1965.

[4] E. Hellinger. Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. Journal für die reine und angewandte Mathematik, 1909(136):210-271, 1909.

[5] G. E Hinton, S. Sabour, and N. Frosst. Matrix capsules with EM routing. In International Conference on Learning Representations, 2018.

WO2019083553A1、US20200285934A1

[6] A. Krizhevsky, V. Nair, and G. E Hinton. CIFAR-10 (Canadian Institute for Advanced Research).

Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. 'Learning multiple layers of features from tiny images', Alex Krizhevsky, 2009. "The CIFAR-10 dataset", April 8, 2009

URL : <https://www.cs.toronto.edu/~kriz/cifar.html>

[7] S. Kullback and R. A. Leibler. On information and sufficiency. The Annals of Mathematical Statistics, 22(1):79-86, 1951.

[8] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. Le Cam and J. Neyman, editors, Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability, volume 1, pages 281-297, 1967.

[9] P. C. Mahalanobis. On the generalized distance in statistics. In

Proceedings of the National Institute of Sciences (Calcutta), 2:49-55, 1936.

[10] D. Mullner. Modern hierarchical, agglomerative clustering algorithms. arXiv e-prints, page arXiv:1109.2378, September 2011.

[11] A.Fitzgibbon, M.Pilu, R.B.Fisher, "Direct least square fitting of ellipse," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.21, no.5, pp.476-480, 1999.

[12] <https://mathworld.wolfram.com/Ellipse.html>

## 符号の説明

[0070] 1 0 0 P C

1 1 処理部

1 2 入力部

1 3 出力部

1 4 表示部

1 5 記憶部

1 6 インタフェース部 (I / F)

1 5 1 入力データファイル

1 5 2 統計値距離行列ファイル

1 5 3 距離行列ファイル

1 5 4 クラスタ行列ファイル

1 5 5 出力ファイル

## 請求の範囲

### [請求項1]

統計的距離行列の計算方法であって、

処理部は、入力部又は他の装置又は記憶部から、比較対象となる第1のベクトルデータ  $a$   $a$  及び第2のベクトルデータ  $b$   $b$  を入力し、又は、前記処理部は、入力部又は他の装置又は記憶部から、比較対象となる第1の行列又は画像又は集合データ  $A$   $A$  及び第2の行列又は画像又は集合データ  $B$   $B$  を入力して前記第1のベクトルデータ  $a$   $a$  及び前記第2のベクトルデータ  $b$   $b$  にベクトル化し、

前記処理部は、前記第1のベクトルデータ  $a$   $a$  及び前記第2のベクトルデータ  $b$   $b$  から、それぞれ、第1の平均ベクトル  $\mu$   $\mu$  1 と第2の平均ベクトル  $\mu$   $\mu$  2、及び、第1の共分散行列  $\Sigma$   $\Sigma$  1 と第2の共分散行列  $\Sigma$   $\Sigma$  2 を計算し、

前記処理部は、第1の平均ベクトル  $\mu$   $\mu$  1、第2の平均ベクトル  $\mu$   $\mu$  2、第1の共分散行列  $\Sigma$   $\Sigma$  1、第2の共分散行列  $\Sigma$   $\Sigma$  2 により定義される統計的距離  $D$  の脱トレースである統計的距離行列  $D$   $D$  (ここで、 $D = \text{tr } D D$ ) を求め、

前記処理部は、統計的距離行列  $D$   $D$  を前記記憶部に記憶し又は表示部に表示させ又は出力部により出力させる、

統計的距離行列の計算方法。

### [請求項2]

請求項1に記載の統計的距離行列の計算方法において、

統計的距離はバタチャリア距離  $D$   $B$  ( $= \text{tr } D D B$ ) であり、統計的距離行列はバタチャリア距離行列  $D$   $D B$  であること、

統計的距離はチャーノフ距離  $D$   $C$  ( $= \text{tr } D D C$ ) であり、統計的距離行列はチャーノフ距離行列  $D$   $D C$  であること、

統計的距離はカルバック・ライブラーダイバージェンス距離  $D$   $K L$  ( $= \text{tr } D D K L$ ) であり、統計的距離行列はカルバック・ライブラーダイバージェンス距離行列  $D$   $D K L$  であること、

統計的距離はマハラノビス距離  $D_M (= \text{tr } D D M)$  であり、統計的距離行列はマハラノビス距離行列  $D D M$  であること、

統計的距離は、トレース形式で表現できる統計的距離  $D (= \text{tr } D D)$  であり、統計的距離行列は  $D D$  であること、  
 のいずれかひとつ又は複数であることを特徴とする統計的距離行列の計算方法。

[請求項3]

統計的距離行列の可視化方法であって、

処理部は、入力部又は他の装置又は記憶部から、比較対象となる第1のベクトルデータ  $a$   $a$  及び第2のベクトルデータ  $b$   $b$  を入力し、又は、前記処理部は、入力部又は他の装置又は記憶部から、比較対象となる第1の行列又は画像又は集合データ  $A$   $A$  及び第2の行列又は画像又は集合データ  $B$   $B$  を入力して前記第1のベクトルデータ  $a$   $a$  及び前記第2のベクトルデータ  $b$   $b$  にベクトル化し、

前記処理部は、前記第1のベクトルデータ  $a$   $a$  及び前記第2のベクトルデータ  $b$   $b$  から、それぞれ、第1の平均ベクトル  $\mu$   $\mu$  1 と第2の平均ベクトル  $\mu$   $\mu$  2、及び、第1の共分散行列  $\Sigma$   $\Sigma$  1 と第2の共分散行列  $\Sigma$   $\Sigma$  2 を計算し、

前記処理部は、第1の平均ベクトル  $\mu$   $\mu$  1、第2の平均ベクトル  $\mu$   $\mu$  2、第1の共分散行列  $\Sigma$   $\Sigma$  1、第2の共分散行列  $\Sigma$   $\Sigma$  2 により定義される統計的距離  $D$  の脱トレースである統計的距離行列  $D D$  (ここで、 $D = \text{tr } D D$ ) を求め、

前記処理部は、統計的距離行列  $D D$  の各対角成分に対して対角成分を含む行及び列の距離成分を累積して距離累積ベクトル  $\phi$   $\phi$  を求め、

前記処理部は、距離累積ベクトル  $\phi$   $\phi$  を距離累積行列  $\Phi$   $\Phi$  に行列化し、

前記処理部は、距離累積行列  $\Phi$   $\Phi$  又は距離累積行列  $\Phi$   $\Phi$  による画像を前記記憶部に記憶し又は表示部に表示させ又は出力部により出力さ

せる、  
統計的距離行列の可視化方法。

[請求項4]           請求項3に記載の統計的距離行列の可視化方法において、  
前記処理部は、前記統計的距離行列 $D$ に対してクラスタリング処理を実行して、各行番号若しくは列番号に対するクラスタラベル、又は、各行番号若しくは列番号に対応する要素 $I$ に対するクラスタラベルを要素とするクラスタラベルベクトル若しくは集合又は要素ラベル集合を求め、  
前記処理部は、クラスタラベルベクトル若しくは集合又は要素ラベル集合を行列化したときの行及び列に対応する距離累積行列 $\Phi$ の行および列の各成分に、さらにクラスタラベルを付ける、  
ことを特徴とする統計的距離行列の可視化方法。

[請求項5]           請求項3又は4に記載の統計的距離行列の可視化方法において、  
前記処理部は、距離累積行列 $\Phi$ のパターンにより第1のベクトルデータ $a$ と第2のベクトルデータ $b$ との類似性を判定し、  
前記処理部は、類似性を表すデータを前記記憶部に記憶し又は前記表示部に表示させ又は前記出力部により出力させる、  
ことを特徴とする統計的距離行列の可視化方法。

[請求項6]           統計的距離行列の可視化方法であって、  
処理部は、入力部又は他の装置又は記憶部から、比較対象となる第1のベクトルデータ $a$ 及び第2のベクトルデータ $b$ を入力し、又は、前記処理部は、入力部又は他の装置又は記憶部から、比較対象となる第1の行列又は画像又は集合データ $A$ 及び第2の行列又は画像又は集合データ $B$ を入力して前記第1のベクトルデータ $a$ 及び前記第2のベクトルデータ $b$ にベクトル化し、

前記処理部は、前記第1のベクトルデータ  $a$   $a$  及び前記第2のベクトルデータ  $b$   $b$  から、それぞれ、第1の平均ベクトル  $\mu$   $\mu$  1と第2の平均ベクトル  $\mu$   $\mu$  2、及び、第1の共分散行列  $\Sigma$   $\Sigma$  1と第2の共分散行列  $\Sigma$   $\Sigma$  2を計算し、

前記処理部は、第1の平均ベクトル  $\mu$   $\mu$  1、第2の平均ベクトル  $\mu$   $\mu$  2、第1の共分散行列  $\Sigma$   $\Sigma$  1、第2の共分散行列  $\Sigma$   $\Sigma$  2により定義される統計的距離  $D$  の脱トレースである統計的距離行列  $D$   $D$ （ここで、 $D = \text{tr } D D$ ）を求め、

前記処理部は、前記統計的距離行列  $D$   $D$  に対してクラスタリング処理を実行して、各行に対応する番号若しくは  $I$   $D$ 、又は、各列に対応する番号若しくは  $I$   $D$ 、に対するクラスタラベルを要素とするクラスタラベルベクトル若しくは集合を求め、

前記処理部は、前記クラスタラベルベクトル若しくは集合をクラスタラベル行列に行列化し、

前記処理部は、前記クラスタラベル行列又はクラスタラベル行列による画像を前記記憶部に記憶し又は表示部に表示させ又は出力部により出力させる、

統計的距離行列の可視化方法。

[請求項7]

請求項6に記載の統計的距離行列の可視化方法において、

前記処理部は、クラスタ行列のパターンにより第1のベクトルデータ  $a$   $a$  と第2のベクトルデータ  $b$   $b$  との類似性を判定し、

前記処理部は、類似性を表すデータを前記記憶部に記憶し又は前記表示部に表示させ又は前記出力部により出力させる、  
ことを特徴とする統計的距離行列の可視化方法。

[請求項8]

請求項3乃至7のいずれかに記載の統計的距離行列の可視化方法において、

統計的距離はバタチャリア距離  $DB (= t_r DDB)$  であり、統計的距離行列はバタチャリア距離行列  $DDB$  であること、

統計的距離はチャーノフ距離  $DC (= t_r DDC)$  であり、統計的距離行列はチャーノフ距離行列  $DDC$  であること、

統計的距離はカルバック・ライブラーダイバージェンス距離  $DKL (= t_r DDKL)$  であり、統計的距離行列はカルバック・ライブラーダイバージェンス距離行列  $DDKL$  であること、

統計的距離はマハラノビス距離  $DM (= t_r DDM)$  であり、統計的距離行列はマハラノビス距離行列  $DDM$  であること、

統計的距離は、トレース形式で表現できる統計的距離  $D (= t_r DD)$  であり、統計的距離行列は  $DD$  であること、  
のいずれかひとつ又は複数であることを特徴とする統計的距離行列の可視化方法。

[請求項9]

請求項5又は7に記載の統計的距離行列の可視化方法において、  
前記処理部は、距離累積行列中の各要素に対して予め定められた複数の閾値でセグメンテーションを行い、複数種類の要素領域を切り出し、予め定められた各要素領域に基づく面積比  $\rho$  により類似性を判定することを特徴とする統計的距離行列の可視化方法。

[請求項10]

請求項5又は7に記載の統計的距離行列の可視化方法において、  
前記処理部は、前記パターンが、円形、楕円形、正方形又は長方形の形状で中心から放射状に広がるマンダラ状の画像であるマンダラパターンにより類似性を判定することを特徴とする統計的距離行列の可視化方法。

[請求項11]

統計的距離行列の可視化装置であって、  
処理部

を備え、

前記処理部は、入力部又は他の装置又は記憶部から、比較対象となる第1のベクトルデータ  $a$   $a$  及び第2のベクトルデータ  $b$   $b$  を入力し、又は、前記処理部は、入力部又は他の装置又は記憶部から、比較対象となる第1の行列又は画像又は集合データ  $A$   $A$  及び第2の行列又は画像又は集合データ  $B$   $B$  を入力して前記第1のベクトルデータ  $a$   $a$  及び前記第2のベクトルデータ  $b$   $b$  にベクトル化し、

前記処理部は、前記第1のベクトルデータ  $a$   $a$  及び前記第2のベクトルデータ  $b$   $b$  から、それぞれ、第1の平均ベクトル  $\mu$   $\mu$  1 と第2の平均ベクトル  $\mu$   $\mu$  2、及び、第1の共分散行列  $\Sigma$   $\Sigma$  1 と第2の共分散行列  $\Sigma$   $\Sigma$  2 を計算し、

前記処理部は、第1の平均ベクトル  $\mu$   $\mu$  1、第2の平均ベクトル  $\mu$   $\mu$  2、第1の共分散行列  $\Sigma$   $\Sigma$  1、第2の共分散行列  $\Sigma$   $\Sigma$  2 により定義される統計的距離  $D$  の脱トレースである統計的距離行列  $D$   $D$  (ここで、 $D = \text{tr } D D$ ) を求め、

前記処理部は、統計的距離行列  $D$   $D$  の各対角成分に対して対角成分を含む行及び列の距離成分を累積して距離累積ベクトル  $\phi$   $\phi$  を求め、

前記処理部は、距離累積ベクトル  $\phi$   $\phi$  を距離累積行列  $\Phi$   $\Phi$  に行列化し、

前記処理部は、距離累積行列  $\Phi$   $\Phi$  又は距離累積行列  $\Phi$   $\Phi$  による画像を前記記憶部に記憶し又は表示部に表示させ又は出力部により出力させる、

統計的距離行列の可視化装置。

[請求項12]

統計的距離行列の可視化プログラムであって、

処理部が、入力部又は他の装置又は記憶部から、比較対象となる第1のベクトルデータ  $a$   $a$  及び第2のベクトルデータ  $b$   $b$  を入力し、又は、前記処理部は、入力部又は他の装置又は記憶部から、比較対象と

なる第1の行列又は画像又は集合データ  $A$   $A$  及び第2の行列又は画像又は集合データ  $B$   $B$  を入力して前記第1のベクトルデータ  $a$   $a$  及び前記第2のベクトルデータ  $b$   $b$  にベクトル化するステップと、

前記処理部が、前記第1のベクトルデータ  $a$   $a$  及び前記第2のベクトルデータ  $b$   $b$  から、それぞれ、第1の平均ベクトル  $\mu$   $\mu$  1 と第2の平均ベクトル  $\mu$   $\mu$  2、及び、第1の共分散行列  $\Sigma$   $\Sigma$  1 と第2の共分散行列  $\Sigma$   $\Sigma$  2 を計算するステップと、

前記処理部が、第1の平均ベクトル  $\mu$   $\mu$  1、第2の平均ベクトル  $\mu$   $\mu$  2、第1の共分散行列  $\Sigma$   $\Sigma$  1、第2の共分散行列  $\Sigma$   $\Sigma$  2 により定義される統計的距離  $D$  の脱トレースである統計的距離行列  $D$   $D$  (ここで、 $D = \text{tr } D$ ) を求めるステップと、

前記処理部が、統計的距離行列  $D$   $D$  の各対角成分に対して対角成分を含む行及び列の距離成分を累積して距離累積ベクトル  $\phi$   $\phi$  を求めるステップと、

前記処理部が、距離累積ベクトル  $\phi$   $\phi$  を距離累積行列  $\Phi$   $\Phi$  に行列化するステップと、

前記処理部が、距離累積行列  $\Phi$   $\Phi$  又は距離累積行列  $\Phi$   $\Phi$  による画像を前記記憶部に記憶し又は表示部に表示させ又は出力部により出力させるステップ、

をコンピュータに実行させるための統計的距離行列の可視化プログラム。