

複層球面 GUI の設計 :

混合データのアイコン化と配置への応用

Multi-Layer Spherical GUI for Composite Data Iconification

堀江 輝樹[†] 前嶋 雅也^{†‡} 横手 良太[†] 松山泰男[†]

Teruki Horie Masaya Maejima Ryota Yokote Yasuo Matsuyama

1. はじめに

昨今のネットワーク社会の発展によって、誰もが容易にデータを生成できる時代になり、我々が扱うデータの量は急増している。大量のデータの背景に隠された情報を理解するためには、データを可視化するグラフィカルユーザーインターフェース (Graphical User Interface; GUI) がしばしば用いられる。GUI は、データを視覚的に表現する効果的な手段であり、エンドユーザーだけでなく、商品やサービスの研究開発を行う現場にも求められる必須の技術である。このとき、次のような性質を満たすことが大事となる。

- 多様な性質をもつ混合データ (composite data) を適切に配置できること。
- ユーザが扱いやすい GUI であること。

上記にある混合データの取り扱いについては、データの類似度 (similarity measure), すなわち距離 (distance, dissimilarity) を適切に定義することが主眼となる。従って、データ間の距離の定義についてはケーススタディの側面をもつ。一方、どのような GUI を設計するのかについては、距離を定義した次の段階の問題であり、汎用性を有する。本論文においては、まず距離が与えられたときにどのように新型 GUI, すなわち複層球面 GUI を設計するのかについてのアルゴリズムを与え、次いで離散型と連続型のそれぞれについての混合データに対する複層球面 GUI の設計を行うことにする。

データ間の関係性から視覚的情報を得る手法として、多次元尺度構成法 (Multidimensional Scaling; MDS) [1], [2] が挙げられる。MDS は、高次の特徴空間上にある 2 つのデータ間に定義される非類似度を保存したまま低次元空間に写像する次元削減法の一つ [1] である。この手法によって、データを 2 次元の可視空間上に展開することができるため、GUI として利用されることが多い。現在までに、様々な空間に対する MDS の派生形 [1], [3], [4], [5], [6] が提案されており、球面上にデータを配置する球面 MDS [3], [4], [6] もその一つである。球は、地球がそうであるように、我々にとって身近な空間的形状であり、その幾何的性質を直感的に理解することができる。したがって、2 次元平面と同様に GUI としての利用に適すると考えられる。

球面 MDS は、単層の球に対する MDS の適用であった。本論文では、より複雑なデータに対応し、その表現能力

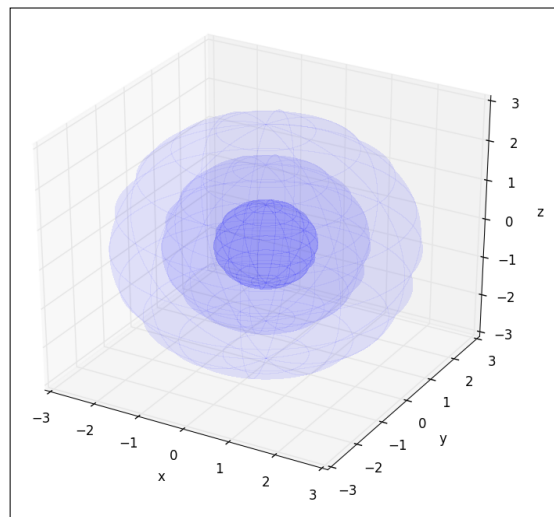


図1 入れ子になった球面インターフェース。

を向上させるための新しい手法を提案する。具体的には、図 1 に示すように複数の球を入れ子にし、複層構造を取るように球面 MDS を拡張することである。以後、これを複層球面 MDS (multi-layer spherical MDS) と呼ぶ。このとき球面上のデータ点は、次のような特性をもつように配置する。

- 同一球面上のデータ点は、与えられた非類似度に応じて適切な距離を取るよう配置する。
- 異なる球面上のデータ点の偏角についての距離が、データ間の相関関係を反映するように配置する。

上記の目標(a)は MDS の本質的な目的を達成するためのものである。一方、目標(b)は球面 MDS によって配置されるデータが、球の回転に対して不定性をもつことがないようにすることを目的としている。後述の節では、このことについてデータの性質を反映した詳細な議論を行う。

なお、本論文においては距離と非類似度を相互交換可能に用いている。従って、非負性と対称性は必要とされるが、三角不等式については必ずしも要求されない。

以上の目的を考慮して、本論文は次のように構成している。まず次節において、提案手法の基礎となる単層球面上での MDS について紹介する。3 節では、初めに複層化に当たってデータの性質に着目した検討を行う。その上で複層球面 MDS の手法を説明する。そして 4 節で GUI への応用結果を提示し、その表現能力について議論する。最後に、5 節で本論文の内容のまとめを述べる。

[†] 早稲田大学基幹理工学研究科情報理工学専攻
Department of Computer Science and Engineering, Waseda
University, Tokyo, Japan.

[‡] 現在、株式会社ハル研究所 HAL Laboratory, Inc.,
Tokyo, Japan.

2. 球面 MDS

この節では、単層の球面 MDS とそのアルゴリズムについて紹介する。特徴空間上、すなわち写像前の高次元空間上における n 個のデータ点は、単位球面上で次のように極座標系で表現される。

$$\Psi = \begin{bmatrix} \theta_1 & \cdots & \theta_n \\ \varphi_1 & \cdots & \varphi_n \end{bmatrix}^T \quad (1)$$

このとき、 i 番目のデータ点はデカルト座標系で、

$$\mathbf{x}_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \begin{bmatrix} \sin \theta_i \cos \varphi_i \\ \sin \theta_i \sin \varphi_i \\ \cos \theta_i \end{bmatrix} \quad (2)$$

と表現できる。データ点 \mathbf{x}_i と \mathbf{x}_j の距離は、次式に示すユークリッド距離の 2 乗を用いる。

$$\begin{aligned} d^2(\mathbf{x}_i, \mathbf{x}_j) &= \|\mathbf{x}_i - \mathbf{x}_j\|^2 \\ &= 1 - \left\{ \begin{aligned} &\cos(\theta_i + \theta_j) + \cos(\theta_i - \theta_j) \\ &+ \cos(\varphi_i - \varphi_j) [\cos(\theta_i - \theta_j) - \cos(\theta_i + \theta_j)] \end{aligned} \right\} \end{aligned} \quad (3)$$

データ点間の距離を測地線距離ではなくユークリッド距離で代替する理由は、次の 2 点に由来する。

- 球面 MDS では、与えられた非類似度の大小関係を保存することのみが重要であるから。
- 最適化における更新即の導出と実際の計算が、測地線距離を測る場合に比べて簡素になるから。

非計量 MDS の枠組みでは、任意の 2 つのデータ点間の非類似度 $\delta_{i,j} \in \mathfrak{R}_{\geq 0}$ ($i, j = 1, \dots, n$) (ただし $\mathfrak{R}_{\geq 0}$ は 0 以上の正の実数集合) と、データ点の配置 Ψ によって得られる距離の差の 2 乗和を評価する。すなわち、目的関数を

$$S(\Psi) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n [\delta_{i,j} - d(\mathbf{x}_i, \mathbf{x}_j)]^2 \quad (4)$$

と定義する。球面 MDS の目標は、この目的関数 $S(\Psi)$ を最小にするようなデータ点の配置 $\hat{\Psi}$ を得ることである。

$$\hat{\Psi} = \arg \min_{\Psi \in \mathfrak{R}^{2n}} S(\Psi) \quad (5)$$

これを求める方法として、しばしば最急降下法が用いられる。その手順は次の通りである。

- [1] データ点の初期位置 $\Psi^{(0)}$ をランダムに決定する。

$$\Psi^{(\text{old})} = \Psi^{(0)} \quad (6)$$

- [2] 各データ点の偏角に対する勾配を求める。

$$\nabla S(\Psi^{(\text{old})}) \quad (7)$$

ただし、

$$\nabla = \begin{bmatrix} \partial/\partial\theta_1 & \cdots & \partial/\partial\theta_n \\ \partial/\partial\varphi_1 & \cdots & \partial/\partial\varphi_n \end{bmatrix}^T \quad (8)$$

である。

- [3] データ点の位置を次式で更新する。

$$\Psi^{(\text{new})} = \Psi^{(\text{old})} - \varepsilon \nabla S(\Psi^{(\text{old})}) \quad (9)$$

上式の $\varepsilon \in \mathfrak{R}_{>0}$ (ただし $\mathfrak{R}_{>0}$ は 0 より大きい正の実数集合) は、最急降下法における学習係数である。

- [4] 解が収束していれば、

$$\hat{\Psi} = \Psi^{(\text{new})} \quad (10)$$

として終了する。そうでなければ、

$$\Psi^{(\text{old})} \leftarrow \Psi^{(\text{new})} \quad (11)$$

を実行して、手順[2]に戻る。

なお、データ点 \mathbf{x}_i の各偏角に対する勾配は次の通りである。

$$\frac{\partial [\delta_{i,j} - d(\mathbf{x}_i, \mathbf{x}_j)]^2}{\partial \theta_i} = \left\{ \begin{aligned} &\sin(\theta_i - \theta_j) + \sin(\theta_i + \theta_j) \\ &+ \cos(\varphi_i - \varphi_j) [\sin(\theta_i - \theta_j) - \sin(\theta_i + \theta_j)] \end{aligned} \right\} \times [1 - \delta_{i,j} d^{-1}(\mathbf{x}_i, \mathbf{x}_j)] \quad (12)$$

$$\frac{\partial [\delta_{i,j} - d(\mathbf{x}_i, \mathbf{x}_j)]^2}{\partial \varphi_i} = \left\{ \sin(\varphi_i - \varphi_j) [\cos(\theta_i - \theta_j) - \cos(\theta_i + \theta_j)] \right\} \times [1 - \delta_{i,j} d^{-1}(\mathbf{x}_i, \mathbf{x}_j)] \quad (13)$$

3. 複層球面 MDS

3.1 複層への拡張

前節で紹介した単層の球面 MDS は、すべてのデータ点を同一球面上に配置することを目標にしていた。本論文では、多数の球が入れ子になっている構造に対する球面 MDS を検討する。まず、各データ点に予め 1 つのラベルが付与されており、ラベルは複数ある球と一対一対応していることを仮定する。従って、入力是非類似度の集合に、各データ点に対するラベル集合を加えた組となる。

複層への拡張に当たっては、1 節で述べた 2 つの目標(a)と(b)を満たすように注意する。そのために、まず非類似度集合の性質に従って次のように分類する。

- すべてのデータ点間に非類似度が与えられている。
- 同一ラベルが付与されたデータ点のみに非類似度が与えられており、異なるラベルのデータ点間には非類似度が与えられない。

性質(i)を満たすような入力を以後、完全データ集合と呼ぶ。この場合、複層化する方法は自明である。すなわち、ラベルを無視した上で 2 節の球面 MDS を用いてすべての

データ点の配置を求めた後に、各データ点のラベルに応じて球面上に配置する。これにより目標(a)と(b)を同時に達成することができる。

一方、性質(ii)は非類似度の一部が欠損している。このような入力を、前述の完全データ集合に対して、不完全データ集合と呼ぶ。この場合、ラベルが異なるデータ点間には非類似度が与えられないため、前述した手順で球面上にデータ点を配置することはできない。同一ラベルのデータ点だけの集合に対して球面 MDS を適用することは可能であるが、異なるラベルをもつデータ点同士の配置の関係性がなくなるため、球の回転不定性が生じる。従って目標(a)は部分的に達成されるが、目標(b)は達成されない。このような性質をもつ実世界上のデータ集合の例として、次のようなものがある。

「一定期間ごとに追跡して行う統計調査で、前後の回で調査項目の個数や、項目の測定粒度が変化した。」

この状況では、データ点の次元や尺度が変化しているため、球面 MDS に必要な類似度、あるいは距離を厳密かつ正しく求めることができない。

そこで異なるラベルをもつデータ点間に、非類似度とは異なる情報を人為的に付与する。これをデータ点間の相関係数と呼ぶ。相関係数は 1, 0, -1 のいずれかの値を取り、それぞれ次のような意味をもつ値と定義する。

- (a1) 相関係数が 1 のとき、データ点間には高い等価性が認められることが分かっている。
- (a2) 相関係数が 0 のとき、データ点間には関係性がないことが分かっている。
- (a3) 相関係数が -1 のとき、データ点間には高い相違性が認められることが分かっている。

前述のような実世界上のデータ集合には、これらの性質が暗黙的あるいは明示的に見いだすことができる場合がある。例では、調査項目名が同じであれば高い相関をもつといえる。データ集合内の隠れた性質を利用するためには、このような相関係数を改めて定義することで、応用できると考えられる。本論文では、不完全データ集合に相関係数を含めたデータ全体を、相関付不完全データ集合と呼び、これに対する複層球面 MDS を提案する。

3.2 相関付不完全データ集合に対する複層球面 MDS

改めて、 p 層上に配置される n_p 個のデータ点の非類似度を $\delta_{p(h,i)} \in \mathcal{R}_{\geq 0}(h, i = 1, \dots, n_p)$ 、 p 層 i 番のデータ点の配置をデカルト座標系で、

$$\mathbf{x}_{p,i} = [x_{p,i}, y_{p,i}, z_{p,i}]^T \quad (14)$$

と表し、このデータ点の初期位置を、

$$\mathbf{x}_{p,i}^{(0)} = [x_{p,i}^{(0)}, y_{p,i}^{(0)}, z_{p,i}^{(0)}]^T \quad (15)$$

とする。また、 p 層 i 番のデータ点と q 層 j 番のデータ点の間の相関係数を $\rho_{q,j}^{p,i}$ ($p \neq q$) と表す。相関係数は予め付与する必要があるが、そのとき次の知見 1, 2 に留意する。

知見1 目標(b)を達成するためには、少なくとも 3 個の異なるデータ点に対して 0 以外の相関係数が与えられる必要がある。

知見2 (a1), (a3)のような相関係数を持つデータ点の集合

$$C = \{\mathbf{x}_{q,j} \mid \rho_{q,j}^{p,i} = \pm 1\} \quad (18)$$

の濃度 $|C|$ は、小さい方がよい。すなわち、1 つのデータ点に付与されている相関係数の個数は、疎であることを要求する。これは過去の研究[7]からもたらされる知見で、多変数関数の最適化においていくつもの局所最適解が存在する場合に有効な戦略である。後述する複層球面 MDS の目的関数では、相関係数がデータ点間に張られた“特別なゴム紐”のような役割を果たしているため、濃度 $|C|$ はこのゴム紐の個数と解釈することができる。

これらは回転不定性を解消しながら、相関係数によるデータ点の最適な配置を見つける重要な特性で、目標(b)の達成のための指針である。

相関係数が付与された不完全データ集合に対する複層球面 MDS のアルゴリズムについて、順を追って説明する。まず 1 層目について、初期位置をランダムに決定して通常球面 MDS を実行する。これによって得られた 1 層目のデータ点の配置を $\hat{\Psi}_1$ とおく。

次に 1 層目のデータ点の配置を固定して、2 層目の球面 MDS を行う。2 層目のデータ点の初期位置は、1 層目との相関係数 $\rho_{2,j}^{1,i}$ と $\hat{\Psi}_1$ を用いて、次のように定める。

(b1) $\rho_{2,j}^{1,i} = 1$ のとき、 $\mathbf{x}_{2,j}^{(0)}$ は $\hat{\mathbf{x}}_{1,i}$ の偏角についての近傍に配置する。すなわち、

$$\mathbf{x}_{2,j}^{(0)} = \hat{\mathbf{x}}_{1,i} + \sigma(2U[0,1]-1) \quad (16)$$

とする。

(b2) $\rho_{2,j}^{1,i} = 0$ のとき、 $\mathbf{x}_{2,j}^{(0)}$ は $\hat{\mathbf{x}}_{1,i}$ の配置とは無関係にランダムに配置する。

(b3) $\rho_{2,j}^{1,i} = -1$ のとき、 $\mathbf{x}_{2,j}^{(0)}$ は $\hat{\mathbf{x}}_{1,i}$ の偏角についての対蹠点近傍に配置する。すなわち、

$$\mathbf{x}_{2,j}^{(0)} = -\hat{\mathbf{x}}_{1,i} + \sigma(2U[0,1]-1) \quad (17)$$

とする。

$U[0,1]$ は、 $[0,1]$ の実数を取る連続一様分布を表す。また、 $\sigma \in \mathcal{R}_{\geq 0}$ は初期位置の近傍の範囲を制御する値である。2 層目のデータ点 $\mathbf{x}_{2,j}$ に対して、1 層目の複数のデータ点が ± 1 の相関係数をもつ場合は、式(16), (17)で得られた近傍点を偏角に関して平均をとった座標を初期位置とする。このような初期位置の決定方法は、次の知見 3 に基づく。

知見3 式(16), (17)中の一様分布による初期位置の揺らぎは、最急降下法による最適化において、データ点の配置が静止することを防ぐためのものである。

以上のように初期位置を決定した後に、2 層目の球面 MDS を実行する。このとき、2 層目のデータ点間の非類似度 $\delta_{2(j,k)}$ を用いた式(4)の最小化に加えて、1 層目のデー

タ点と2層目のデータ点間の相関係数に応じた次の評価値を最小化する。

$$S'(\Psi_2) = \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} \rho_{2,j}^{1,i} \|\mathbf{x}_{2,j} - \hat{\mathbf{x}}_{1,i}\|^2 \quad (19)$$

ただし、1層目と2層目のデータ点間の距離は、それらが同一の単位球面上に存在すると仮定したときのユークリッド距離の2乗である。

式(19)は、(a1)、(a3)の相関係数の意味から導かれる。すなわち、相関係数が1ならば、データ点が偏角について近傍に留まるよう、それらの間にある“ゴム紐”を縮める。一方、相関係数が-1ならば、データ点が偏角について可能な限り離れるよう、“ゴム紐”を伸ばす。この項を球面MDSの目的関数に追加することで、目標(b)を達成することができる。

以上の議論から、改めて式(4)と式(19)を用いて目的関数を

$$S(\Psi_2) = \sum_{j=1}^{n_2-1} \sum_{k=j+1}^{n_2} [\delta_{2(j,k)} - d(\mathbf{x}_{2,j}, \mathbf{x}_{2,k})]^2 + \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} \rho_{2,j}^{1,i} \|\mathbf{x}_{2,j} - \hat{\mathbf{x}}_{1,i}\|^2 \quad (20)$$

と定義する。ただし、式(19)の偏角に対する偏微分は次の通りである。

$$\frac{\partial S'}{\partial \theta_{2,j}} = \sum_{i=1}^{n_1} \rho_{2,j}^{1,i} \left\{ \sin(\theta_{2,j} - \theta_{1,i}) + \sin(\theta_{2,j} + \theta_{1,i}) + \cos(\varphi_{2,j} - \varphi_{1,i}) [\sin(\theta_{2,j} - \theta_{1,i}) - \sin(\theta_{2,j} + \theta_{1,i})] \right\} \quad (21)$$

$$\frac{\partial S'}{\partial \varphi_{2,j}} = \sum_{i=1}^{n_1} \rho_{2,j}^{1,i} \sin(\varphi_{2,j} - \varphi_{1,i}) \times [\cos(\theta_{2,j} - \theta_{1,i}) - \cos(\theta_{2,j} + \theta_{1,i})] \quad (22)$$

以上が2層目のデータ点の配置を得る過程である。3層目以降は、次の知見に基づいて2層目と同様の手法を繰り返し適用する。

知見4 層の個数が3以上ある場合、 $M(\geq 2)$ 層目以降の球面MDSを実行する際は、その直前に求めた $M-1$ 層目のデータ点の配置 $\hat{\Psi}_{M-1}$ と相関係数 $\rho_{M,j}^{M-1,i}$ のみを用いることを推奨する。これは、各層に対してアルゴリズムが単純な反復実行になるだけでなく、計算量の低減しながら目標(a)と(b)を同時に達成することができるからである。

すべての層についてデータ点の配置を求めたら、アルゴリズムは終了する。

4. 球面MDSのGUIへの応用

この節では、球面MDSをGUIに応用する実験と結果を提示する。実験では、次のような実世界データ集合を用いた。

- (1) テレビ番組のメタデータとユーザの番組視聴履歴。
- (2) 近赤外線分光器(Near-Infrared Spectroscopy; NIRS)によるタスク実行時の脳血流中の脱酸素化ヘモグロビン濃度変化量。

これは3.1節で述べた、データ集合の性質(i)と(ii)に対応している。すなわち(1)は性質(i)を満たしており、すべてのデータ点間で厳密かつ正しい非類似度が与えられている。一方(2)は、後述するように、データ集合内で互いに非類似度が与えられる2つの互いに素な集合に分割でき、それらの間には非類似度が定義できない。したがって、複層球面上にデータ点を表現するためには、提案手法の適用が必要である。

4.1 テレビ番組表の球面インターフェース

この実験は[6]を2層に拡張したものである。単層の場合の詳細を要約すると次のようになる。

データは、テレビ番組を録画する機器に蓄積された番組とそのユーザの視聴履歴である。非類似度は、番組のメタデータ(放送日時、放送時間、放送局、番組のジャンル、字幕)を元に算出する。また、各データのラベルはユーザの番組視聴履歴および番組間の非類似度を元に、「必ず観るべき(must-see)番組」と「観なくても良い(no-need)番組」の2種類を付与する。

結果を図2に示す。これらはタッチパネル上での操作に対応している。すなわち、図2は任意の場所に置かれた親指を中心にして、人差し指の操作で複層球面を回転させる様子を表している。外側の球には「必ず観るべき(must-see)番組」、内側の球には「観なくても良い(no-need)番組」のラベルが付与されたテレビ番組のアイコンが表示されている。3.1節の性質(i)についての議論の通り、このインターフェースはすべてのデータ点について球面MDSを実行して配置を求めた後に、各データ点のラベルによって描画する対象の球を区別している。したがって、各データ点の偏角についての距離は、ラベルとは無関係に非類似度を厳密に反映しており、目標(a)と(b)を同時に達成している。

4.2 脳血流の測定結果のマッピング

次に、提案手法が目標(a)と(b)を達成していることを確認するため、NIRSによる測定結果を可視化する実験を行った。被験者は、与えられた1つのタスクを10分間継続して実行した。タスクは次の2種類である。

[ノータスク] 何もせず、じっとしている。

[タイピング] コンピュータのキーボードを用いて、画面上に次々現れる単語を入力する。

実験中は、被験者のこめかみ部分を流れる血液中の脱酸素化ヘモグロビン濃度の変化量を、周波数6Hzで継続的に測定した[7]。10分間のタスクの実行によって、合計

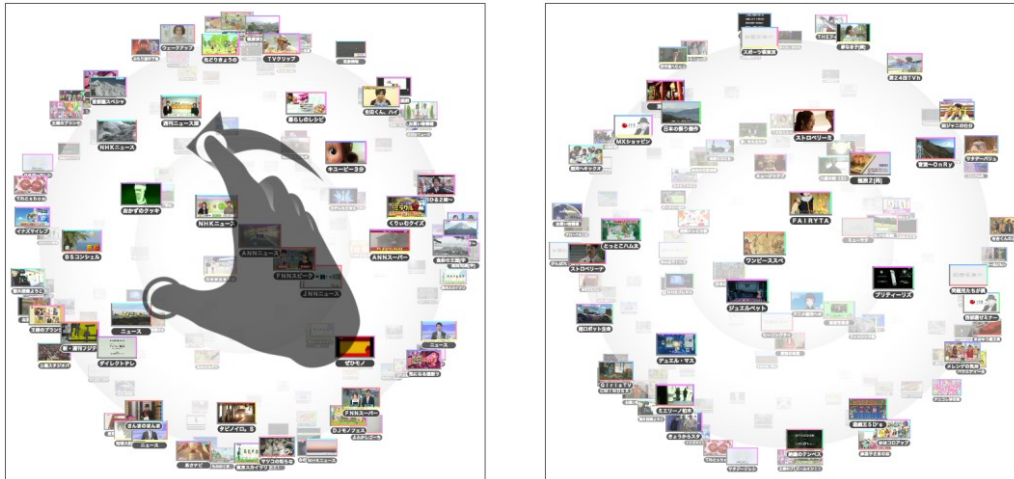


図2 2層に拡張された球面MDSによる実験(1)番組テレビ番組データのマッピング結果。左は表面を、右は裏面を示す。

3600個のサンプルが得られた。

次に得られたサンプルに対して前処理を実行した。まずサンプルの先頭から10秒ずつにチャンキングし、1チャンクごとにフーリエ変換を行い、各周波数に対するパワースペクトルを求めた。そして、測定結果全体で平均0、分散1となるよう正規化を行った。これにより、被験者1人につき、1本当たり60次元の測定結果が60本得られた。なお、この測定結果は継続認証についての研究[8]で用いられたものと同様の手順で得られたものである。

実験には延べ20人の被験者が協力した。被験者は、[ノータスク]を行ったA, B, C, D, E, F, G, H, I, Jの10人と、[タイピング]を行ったA, E, I, L, K, O, P, Q, R, Sの10人である。このうちA, E, Iは、両方のタスクを行っているが、それ以外の被験者はどちらか片方のタスクしか行っていない。また上記のタスクは、異なる日時に実施された。

これらの条件から、3.1節の(a1)から(a3)に示したデータ集合の意味を考慮して、A, E, Iのそれぞれのデータ点間に相関係数1を設定し、それ以外の被験者のデータ点間に相関係数0を設定した。そして、タスクごとに球を分け、外側の球に[ノータスク]、内側の球に[タイピング]のデータ集合が配置されるようにした。アルゴリズムの各パラメータは $\varepsilon=0.001$ 、 $\sigma=0.1$ で、最急降下法の収束条件は、式(4)、(20)に示した目的関数の勾配量の変化量が0.001を下回った場合とした。この実験では、通常の球面MDSと同様に目標(a)を達成し、なおかつA, E, Iのそれぞれのデータ点が、偏角に対して近傍に配置され、目標(b)を達成することが期待される。

複層球面MDSによる球面上への表示結果を、図3に示す。4つの図は球を異なる角度から見たもので、左上を正面にして、右上が裏側、左右側面がそれぞれ左下と右下の図である。外側の球に[ノータスク]、内側の球に[タイピング]のデータ点が配置されている。また、アルゴリズムによって得られた各データ点の配置について、層ごとに縦軸に θ [rad]、横軸に φ [rad]をとったグラフを図4に示す。上部が[ノータスク]すなわち外側の球、下部が[タイピング]すなわち内側の球に展開されているデータ点の偏角を表している。

この結果から、次のようなことがわかる。

- 脳血流の測定結果は、人間の脳から得られた情報であるため、一様な分布にはならず偏在する。したがって、球面上にデータ点が全く存在しない領域ができています。
- 相関係数を1に設定したA, E, Iは、それぞれ偏角についてほぼ同じ場所に位置しており、式(19)に示す相関係数の目的関数に対する効果が現れている。

特に後者の結果から、式(20)で導入した相関係数による“ゴム紐”効果によって、球の回転不定性が解消され目標(b)が達成されていることがわかる。

5. まとめ

本論文では、球を入れ子にした複層球面MDSを提案した。この手法は、一部のデータ点間の非類似度が与えられないような不完全データ集合に対して、新たに相関係数を導入して複層球面上にデータ点を配置する方法である。相関係数はデータ集合の明示的あるいは暗黙的な情報をもつ値であると定義し、その設定指針を示した。アルゴリズムは相関係数をもつ意味をデータ点の配置に反映して、不完全データ集合の最適な配置を得る。また最適化には最急降下法を用いたが、良質な解を求めるために2層目以降の球面上におけるデータ点の初期位置を相関係数によって補正する方法を導入した。そして、複層球面GUIとして不完全データ集合を可視化できることを、実験を通して示した。

MDSは、データ点間の非類似度のみを利用するため、多様なデータ集合に適用できるという利点がある。また、球面MDSは、球面上におけるデータ点の配置を直接得ることができるためGUIに応用される。提案手法は、球面GUIとしての表現能力を高め、一般化された複層球面に対するMDSを提供する。

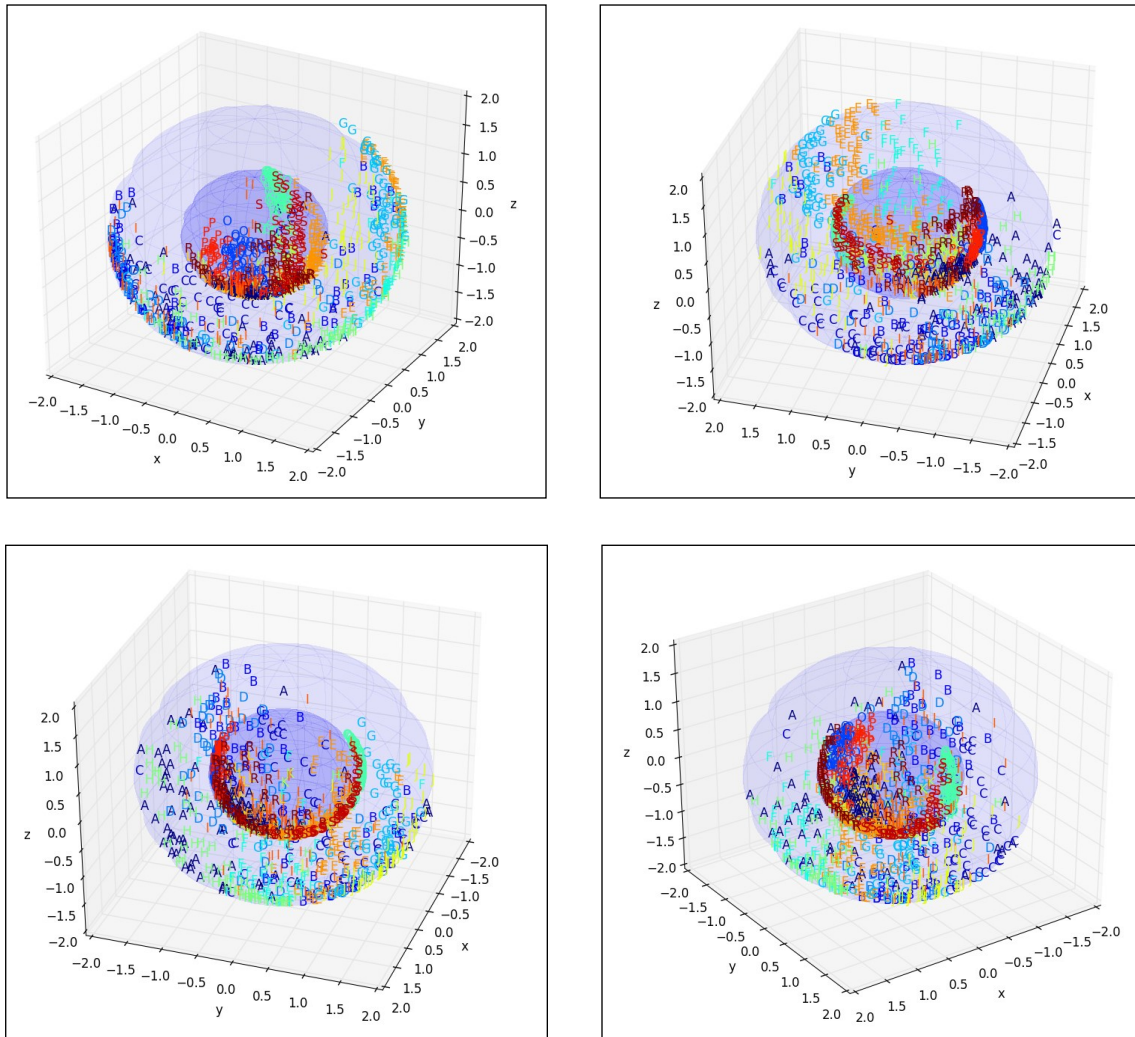


図3 複層球面 MDSによる実験(2)NIRS測定データの球面上へのマッピング結果。
 左上は正面、右上は裏面、左下は左側面、右下は右側面から見た球を示す。

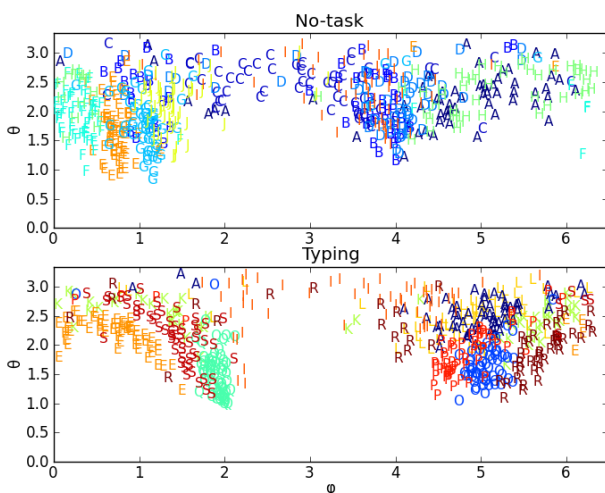


図4 複層球面 MDSによる各層におけるデータ点の偏角。

参考文献

- [1] T. F. Cox, M. A. A. Cox, *Multidimensional Scaling*, Second Ed., Chapman & Halls/CRC, Boca Raton, FL, (2001).
- [2] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, Vol. 29, pp. 1-17, (1964).
- [3] T. F. Cox, M. A. A. Cox, "Multidimensional scaling on a sphere," *Communications in Statistics - Theory and Methods*, Vol. 20, No. 9, pp. 2943-2953, (1991).
- [4] A. Elad, Y. Keller, R. Kimmel, "Texture mapping via spherical multi-dimensional scaling," *Scale Space and PDE Methods in Computer Vision*, Lecture Notes in Computer Science No. 3459, pp. 443-445, (2005).
- [5] Y. Matsuda, K. Yamaguchi, "An efficient MDS algorithm for the analysis of massive document collections," *Lecture Notes in Artificial Intelligence*, No. 3862, pp. 1015-1021, (2005).
- [6] M. Maejima, R. Yokote, and Y. Matsuyama, "Composite Data Mapping for Spherical GUI Design: Clustering of Must-Watch and No-Need TV Programs," in *Neural Information Processing*, Springer Berlin Heidelberg, Vol. 7667, pp. 267-274, (2012).
- [7] Hamamatsu Photonics, "NIRO 200 document," Shizuoka, Japan, (2003).
- [8] M. Shozawa, R. Yokote, S. Hidano, C-H. Wu, Y. Matsuyama, "Brain signal based continuous authentication: Functional NIRS approach," *Lecture Notes in Computer Science*, Vol. 7903, pp. 171-180, (2013).