

# 相関に基づくクラスタリングを用いた軸縮約可能な平行座標系

能野 琴<sup>1</sup> 吳 湘筠<sup>1</sup> 渡辺 一帆<sup>2</sup> 高橋 成雄<sup>1</sup> 藤代 一成<sup>3</sup>

東京大学大学院新領域創成科学研究科<sup>1</sup>

奈良先端科学技術大学院大学情報科学研究科<sup>2</sup>

慶應義塾大学理工学部<sup>3</sup>

## 1 はじめに

高次元データの可視化で広く用いられ、次元間の相関を把握できるグラフのひとつとして、図1に示されるような**平行座標系**と呼ばれるものがある。しかし、平行座標系においても、データの項目数や次元数の増加に伴い、データに対応する折れ線や次元軸が複雑に重なり合うため、相関の把握が困難になる。そこで本研究では、この問題を解決するために、相関の高い軸同士をまとめていくことで、描画対象となる**次元を選択**しながら、データに内在する次元間の相関を強調する平行座標系描画手法を提案する。

具体的には、次のように手法を構成する。まず、各軸同士の相関を類似度とみなして、軸をノードとするグラフを構築する。次に、グラフのスペクトル解析を通じて、各軸の位置を数直線上に写像する。数直線上の軸に合わせて、平行座標系の軸の順番を並び替えた後、適宜軸を併合することで平行座標系における軸の本数を減らし、最終的に少ない次元によるデータの可視化を実現する。

## 2 関連研究

平行座標系におけるデータの軸相関の視認性の向上のために、データのレンダリングを工夫するものが多く存在する。特に、グラフ描画でもよく用いられるエッジの束化(**edge bundling**)を折れ線描画に用いたもの[1]、また相関の高い折れ線を集合としてまとめ、半透明な領域として描くなどの手法[2]が提案されてきている。

これに対して、平行座標系のデータ表現そのものの表現に変更を施す手法も提案されてきている。文献[3]では、軸間のデータ相関が最大化

するように、軸の順番を入れ替える手法が提案された。最近では、データと次元の部分集合を適宜選択し、局所的に相関の高いデータを選択する対話的な機構が[4]において実現されている。本提案手法は、こちらのカテゴリに分類される。

## 3 提案手法

提案手法は以下の手順をとる。

- 1) データの各軸間の相関の計算
- 2) 軸間の相関に基づくグラフ構築
- 3) スペクトル解析による軸の順序付け
- 4) 系統樹表現を用いた軸の縮約

以下、それぞれの進め方を説明する。

### 3.1 データの各軸間の相関の計算

最初に、任意の軸ペア同士のデータ相関を計算する。具体的には、2つの軸に関する  $n$  個のデータ値をそれぞれ

$$x = \{x_i\}, y = \{y_i\} \quad (i = 1, 2, \dots, n)$$

とおいたとき、次の相関係数を計算する。

$$\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

ただし、 $\bar{x}, \bar{y}$  はそれぞれ  $x = \{x_i\}, y = \{y_i\}$  の相加平均である。上記の相関係数は、-1から1の値をとることに注意する。ここでは、算出した相関係数の絶対値を、各軸同士の類似度とする。

### 3.2 軸間の相関に基づくグラフ構築

次に軸をノードとみなした際に、高い相関をもつ軸間にエッジを挿入し、相関に関する隣接関係を表すグラフを構築する。このとき、グラフの隣接行列がスパースになるように、ある閾値以上の相関をもつ軸ペアの間のみエッジを挿入することとする。具体的には、軸間の相関係数の絶対値が0.15以上の場合のみエッジを挿入する。

Contracting the axes of parallel coordinates using correlation-based clustering

Koto Nohno<sup>1</sup>, Wu Hsiang-Yun<sup>1</sup>, Watanabe Kazuho<sup>2</sup>, Takahashi Shigeo<sup>1</sup>, Fujishiro Issei<sup>3</sup>

<sup>1</sup> Grad. School of Frontier Sciences, The Univ. of Tokyo

<sup>2</sup> Grad. School of Information Science, NAIST

<sup>3</sup> Faculty of Science and Technology, Keio Univ.

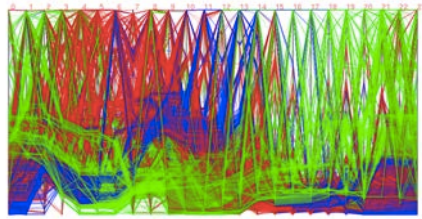
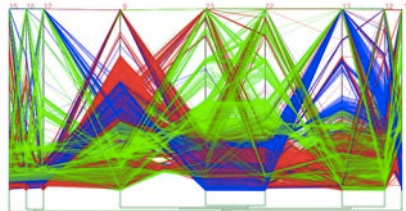


図1: 24次元データの平行座標系表示

図2: 24次元データを9つの軸にまで縮約した  
平行座標系表示

### 3.3 スペクトル解析による軸の順序付け

3.2項で構築したグラフのスペクトル解析[5]を通じて、相関の高い軸同士が並ぶような軸の順序付けを求める。具体的には、グラフのラプラシアン行列を、グラフに接続するエッジ数を対角成分にもつ行列とグラフの隣接行列の差として計算し、その固有値と固有ベクトルから、軸ノードをそのクラスタ分布が最も反映する軸へ写像することで、軸の順序を求める。

### 3.4 系統樹表現を用いた軸の縮約

3.3項で得られた軸の順序付けから、軸の縮約を行う。具体的には、最も相関の高い2つの軸を併合する操作を順次行い、系統樹を構成しながら軸の縮約を進めていく。

例えば、すべての軸の本数を  $x$  本としたとき、3.3項で得られた各隣同士の軸の距離を計算し、最も距離の小さい2軸を選択して縮約する。この順序を並べ替えた軸集合において、 $k$ 、 $k+1$ 番目の軸( $1 \leq k < k+1 \leq x$ )が最も近いとする。このとき  $k$ 、 $k+1$ 番目の軸の位置座標の平均を算出し、これを縮約された軸の座標値とする。この軸の縮約操作を、最も近い隣接軸ペアを順次選んで適用していく。

## 4 結果

提案手法を適用した例を、以下に示す。

まず本システムにおける、軸を縮約する際の平行座標系の描画においては、新しく生じる軸は既存の2つの軸の中間に位置するようにし、影響のない軸は位置を固定しておく。また、軸併

合操作にアニメーション効果を導入し、ユーザが自由に軸を縮約および展開できるようにしている。また、軸の併合の順番を、平行座標系の下に系統樹を描くことで、ユーザに軸の併合の履歴を見せるように実装されている。

図1は、24次元のデータを3つのクラスタに色分けし平行座標系に表示したものを示し、図2はそのデータにおいて軸を9本にまで縮約して描画した例を示す。図1では、項目数や次元数が多く描画されるデータの相関の把握が困難であるが、図2では次元間の相関を強調する表示が得られた。

## 5 おわりに

本稿では、高次元データに内在するデータ相関を強調できる、軸縮約可能な平行座標系表示の枠組みを提案した。今後の課題としては、部分空間を選択するためのインタフェースの構築があげられる。文献[4]では、2つの散布図を用いてデータと軸の両方を別々に散布図として描画していたが、平行座標系では1つの表現でデータサンプルと次元の両方の情報が描画できる。そのため、よりよい次元選択の操作ができれば、全体では視認性が低いデータ相関を、次元縮約を介して向上させることができる。

**謝辞** 本研究の一部は、科研費新学術領域計画研究25120014の支援により実施された。

### 参考文献

- [1] H. Zhou, X. Yuan, H. Qu, W. Cui, and B. Chen. Visual Clustering in Parallel Coordinates, *Computer Graphics Forum*, Vol. 27, No. 3, pp. 1047-1054, 2008.
- [2] M. T. McDonnell and K. Mueller, Illustrative Parallel Coordinates, *Computer Graphics Forum*, Vol. 27, No. 3, pp. 1031-1038, 2008.
- [3] W. Peng, M. O. Ward, and E. A. Rundensteiner. Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering, *IEEE Symposium on Information Visualization 2004 (InfoVis 2004)*, pp. 89-96, 2004.
- [4] X. Yuan, D. Ren, Z. Wang, and C. Guo. Dimension Projection Matrix/Tree: Interactive Subspace Visual Exploration and Analysis of High Dimensional Data. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 19, No. 12, pp. 2625-2633, 2013.
- [5] F. R. K. Chung. *Spectral Graph Theory*, American Mathematical Society, 1997.