3ZC-8

スパースモデリングを目的とした平行座標系表示の拡張

新部 祐輔 * 呉 湘筠 † 渡辺 一帆 ‡ 高橋 成雄 † 藤代 一成 *

慶應義塾大学 * 東京大学 † 奈良先端科学技術大学院大学 ‡

1 背景と目的

近年の計測技術の向上によって，解析者が獲得するデータは一層高次元化してきている．これにより解析時の計算量が次元数に対して指数的に爆発する状況が生じ，データ解析者による直感的な仮説の検証が困難となっている．

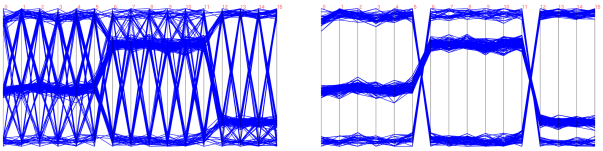
スパースモデリング [1] とは，そのような高次元観測データに普遍的に内在すると考えられるスパース (疎) 性を利用し，適切に説明変数を選択し，対象の本質を効率的かつ効果的に記述しようとするモデル化技術である．高次元データを十数〜数十次元に圧縮することが可能であり，先述の問題を解決するために有効な手法を提供するが，解析者が一般的に用いることのできるディスプレイ環境において，視覚的理解を促すには，さらに2〜4次元の表現へ効果的に圧縮する必要がある．

本研究ではスパースモデリングにおける効果的なビジュアルデータマイニングの環境を解析者へ提供することを目的とする．本稿では，多変量データを可視化する手法の一つである，平行座標系表示 [2] をその目的のために拡張する．スパースモデリングによって次元削減し，さらにバイクラスタリング [3] を適用して得られるデータを可視化する際に，その基底となることが望ましい，相関性の低い変数やデータサンプルの選択を視覚的に補助する．既存の手法でも平行座標系表示を拡張する様々な工夫が見られるが，バイクラスタリングを適用した表現は見られない．本研究では，このバイクラスタリングの寄与をよく表現しうる手法を新たに考案する．本稿ではそのなかから効果が高いと考えられる3種類の手法について報告する．

2 平行座標系表示とバイクラスタリング

平行座標系表示は，多変量データを可視化する際に多用される，代表的な情報可視化手法の一つである [2]．n 本の平行な軸で n 次元空間の各変量を，各軸間をつなぐ各折れ線で各データサンプルを表現する．ここでは，各変量・データサンプル間の相関が，折れ線同士の交わり方によって視覚的に表現されている (**図 1(a)**)．

一方，バイクラスタリングとは，データ行列に対して行方向と列方向の両方から同時にクラスタリングを行い，共通の意味をもつクラスタへブロック化する手法である [3]．本研究で扱うデータは，スパースモデリングによって次元削減されたデータに対し，この手法を変数方向およびデータサンプル方向へ適用したものを対象とする (**図 1(b)**)．意味をもつブロックとして解析者へ視覚的に提示することで，基底として最適な変数およびデータサンプルの選択を促進させられる．



(a) 適用前　(b) 適用後

図 1: 平行座標系表示とバイクラスタリング適用

3 提案手法

本稿では，解析者への視覚的フィードバックとして考えられる可視化表現のなかから，寄与が高いと考えられる手法を3種類提案する．

3.1 手法1：色相と彩度

色は，色相・彩度・明度に分解できる．これらのうち，視覚的に最もその大きな違いが現れるのは色相である．彩度と明度では，色相が表す色味の強弱を示す彩度の方が，色相への寄与が大きいと考える．提案手法では，色相を変数方向のクラスタへ，彩度をデータサンプル方向のクラスタへ，それぞれ割り当てる．これはデータサンプルの選択よりも，変数の選択の方がスパースモデリングにおいて，より重要であると考えるためである．

3.2 手法2：塗りつぶし

塗られている面積の大小が，視覚的な印象の強弱へ影響する点を考慮して，同一のデータサンプル

Extending parallel coordinate plots for sparse modeling
Yusuke Niibe*, Wu Hsiang-Yun†, Kazuho Watanabe‡,
Shigeo Takahashi†, Issei Fujishiro*
Keio University*, The University of Tokyo†,
Nara Institute of Science and Technology‡

クラスタに属するデータサンプルの隙間を，単色で塗りつぶす．バイクラスタリングによって得られた共通の意味をもつデータのブロックを，属する変数やデータサンプルの個数によらず，その値の幅によって強調して提示することができる．

3.3 手法3：3Dモデル化

各データサンプルについて，折れ線の代わりに直方体を連続して折れ線状に配置し，立体的に表現する．各データサンプルの法線が異なる2面を見せることで，陰影のコントラストを生じさせる．各バイクラスタ内における，コントラストの表現から，そのバイクラスタに属するデータサンプル数の大小を提示できる．各データサンプルクラスタは，要素数が多いものほど多くの面積を占めると仮定し，要素数の多い順に下から描き重ねる．

4 結果

本稿に掲載する画像は，PC（OS:Ubuntu13.10 64bit, CPU:Intel Core-i7 3.4GHz, RAM:3GB）上で，APIとしてOpenGL2.1 Mesa9.2.1を使用して生成した．提案手法を適用するデータセットとして，変数の個数が16個，データサンプルの個数が100個であるような人工データを使用する．

比較参照画像（**図2(a)**）は，各バイクラスタへ，無作為に選択した異なる色を割り当てた．塗りつぶしによる手法2と3Dモデル化による手法3は，色相と彩度による手法1（**図2(b)**）と組み合わせて用いることが想定されるため，それぞれの手法とともに適用した（**図2(c)**，**図2(d)**）．

4.1 手法1：色相と彩度

比較参照画像（**図2(a)**）では，異なる変数クラスタ内のバイクラスタにおいて，色相の近い色が2色選択されている．各々のバイクラスタが，実際には大きく性質が異なるにも関わらず，解析者へ与える視覚的な印象が近くなってしまっている．

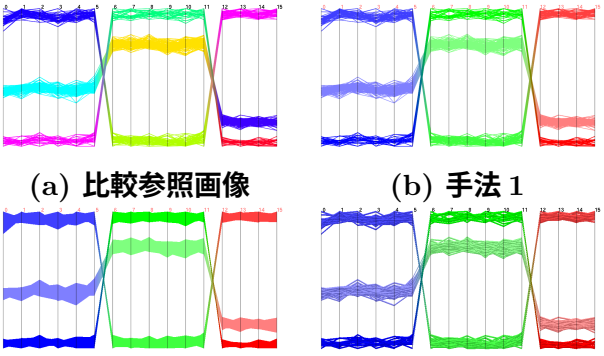
これに対し提案手法による画像（**図2(b)**）は，各変数クラスタ間の色相が大きく異なっているため，その差異が視覚的に顕著であることがわかる．さらにデータサンプルクラスタへ割り当てた彩度によって，各変数クラスタへの従属性を失わずに各データサンプルクラスタを表現している．

4.2 手法2：塗りつぶし

提案手法による画像（**図2(c)**）の方が手法1による画像（**図2(b)**）よりも，各バイクラスタが視覚的に強調されている．特に彩度の最も低いデータサンプルクラスタの色は背景色と近く，より彩度の高いクラスタよりも視覚的な刺激が弱い．このため，手法1による画像と提案手法による画像を比べたときに効果をより強く発現している．

4.3 手法3：3Dモデル化

手法1による画像（**図2(b)**）と提案手法による画像（**図2(d)**）を見ると，同程度の幅をもつバイクラスタでも，提案手法ではコントラストの表れる頻度が高いバイクラスタが，多くのデータサンプルを含んでいることが視認できる．



(a) 比較参照画像　(b) 手法1

(c) 手法1+手法2　(d) 手法1+手法3

図2: 提案手法による可視化の視覚効果の比較

5 結論と今後の課題

本稿では，スパースモデリングによって次元削減し，バイクラスタリングを適用したデータについて，その効果をよく表現する可視化手法をデータ解析者へ提供することを試みた．提案手法によって，相関性の低い変数やデータサンプルの選択が促進されれば，本研究によってスパースモデリングにおける効果的なビジュアルデータマイニングの環境を，データ解析者へ提供できると考える．そのため今後，ユーザへの評価実験の実施や，さらなる表現手法の考案が課題として挙げられる．

謝辞

本研究の一部は，平成25年度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] スパースモデリングの深化と高次元データ駆動科学の創成 (http://www.sparse-modeling.jp)

[2] Alfred Inselberg, *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. New York: Springer-Verlag, 2009.

[3] Sara C. Madeira and Arlindo L. Oliveira, “Biclustering Algorithms for Biological Data Analysis: A Survey,” *IEEE Transactions on Computational Biology and Bioinformatics*, Vol. 1, No. 1, pp. 24-45, January-March 2004.