

多次元情報の次元縮約過程における動的可視化手法の提案

森 健人[†] 大野 義夫[†] 藤代 一成[†]

[†] 慶應義塾大学 理工学部 情報工学科

1 背景と目的

情報可視化において次元縮約は非常に重要な操作の一つである。それは次元数の高さが情報の複雑さを生む大きな原因であり、それ故に適当な次元縮約がしばしばデータ解析の効率化や効果的な情報の提示を可能にするためである。

しかし対象データそれぞれに対して適当な縮約手法の選択するには専門的な知識や経験を要する場合が多く、元データと可視化結果だけから縮約操作の客観的な根拠を知ることは難しい。さらに多次元情報を次元縮約する場合は数値的処理が主となるため、その可視化結果に至る理由を理解することは一般の多くの人々にとってより難しいものとなる。

そこで本稿では次元縮約操作の前後関係に注目した可視化手法を提案し、多次元情報の縮約過程を動的に可視化することを目指す。そして次元縮約方法として主成分分析 (Principal Component Analysis, 以下 PCA と表記する) を用いた可視化システムを実装し、本手法の検証を行う。

2 アプローチ

情報可視化における数値計算プロセスは、与えられたデータのフィルタリング、それにより得られたデータの幾何空間上へのマッピングの2つに大きく分けられる。そして可視化とは一連のプロセスを経て得られた結果を参考にして、もし必要があればそれぞれの工程を再実行してより適当な可視化結果を探索することを指す。この場合次元縮約操作は可視化の前処理であるフィルタリングに相当する。

本稿では次元縮約過程の可視化における最適なマッピング手法は利用した次元縮約手法により決定できると仮定し、マッピング手法を固定した上での次元縮約操作過程の可視化を目指す。そこで縮約過程の可視化

に最適なマッピング手法を考察するとともにそれら次元縮約操作の前後関係を動的に可視化する手法を提案する。

そして以上を検証するため、最も基本的な次元縮約手法の一つである PCA を用いた次元縮約可視化システムを実験的に実装しその効果の検証を行った。

3 関連研究

次元縮約における数学的手法とその可視化については多くの研究が存在するが、その過程に注視したものは少ない。一方、解析の効率化のための対話的情報可視化システムとして P.I.A. Godinho ら [1] による PRISMA をはじめ幅広く研究されている。

また次元縮約操作に注目した可視化しているものとしては H.-C. Hege ら [2] による iPCA が挙げられる。この研究では複数のビューを用いて縮約操作を同時的かつ多角的に可視化することでインタラクティブに PCA を用いた解析を行えるシステムが提案されている。

4 PCA を用いた次元縮約過程の可視化

本稿では次元縮約過程の動的可視化手法として PCA を用いた多次元情報可視化システムを実装した。中心から座標軸が放射状に広がるスタープロットを基本的なマッピング手法として用いた上で、座標軸のモーフィングを用いて次元縮約過程の動的な可視化を行った。

4.1 スタープロットによる可視化

本システムでは多次元データの属性がスタープロットの軸に相当し、それら軸上に各々の値がプロットされる。PCA を行うことで抽出された主成分もまた新たな軸を形成し、値のプロットもそれに準じて移動を行う。このとき初期の軸ベクトルが主成分ベクトルの組み合わせで表現できる場合、軸をモーフィングさせてその変化 K の可視化を行う。

以上の一連の過程を軸の最小移動を条件に最適な主成分ベクトルの決定を行う。本システムでは統計学で一般的な軸回転の基本操作であるバリマックス回転に加え、最適なモーフィングのためのグラフ全体の回転を考慮した。

A Proposal of Dynamic Visualization Method for Reducing Dimensionality of Multidimensional Data

Kento Mori, Yoshio Ohno, Issei Fujishiro

Department of Information and Computer Science, Keio University

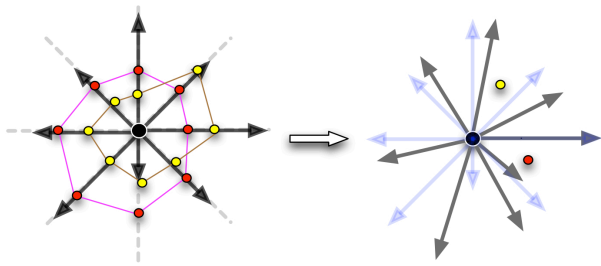


図1: スタープロットによるデータ可視化. PCAにより第2主成分まで抽出した際, 右図のように軸ベクトルが変化する

軸の移動量が最小となるような回転量は以下のように求めることができる. n 本の初期軸があるとき, 変化前の軸ベクトル v_n , 変化後の軸ベクトル v'_n から以下の式 (1) が得られる. このとき軸全体の回転は Q が最小になる回転行列 R として求める. また初期状態では軸の順序を同様に計算して配置する.

$$Q = \sum_{i=1}^n (v_i \times (v'_i \cdot R)) \quad (1)$$

PCA では抽出した主成分情報の累積寄与率が8割程度になることを目安に縮約する次元数が決められる. 本システムではこのようなスタープロットによる可視化により, ユーザが縮約する次元数を決定するもしくは抽出する主成分の数を選択することで, インタラクティブに段階的に次元縮約を行うことが可能である.

4.2 相関行列の可視化

PCAに限らず軸の相関関係を把握することは多次元情報解析において重要である. 相関関係の可視化は多くの情報可視化システムでも行われているが, 数値やグリフの表によるものがほとんどであった. そこで本システムではPCAによる操作に限定されていることからスタープロット上に同時に相関行列の可視化を行なった.

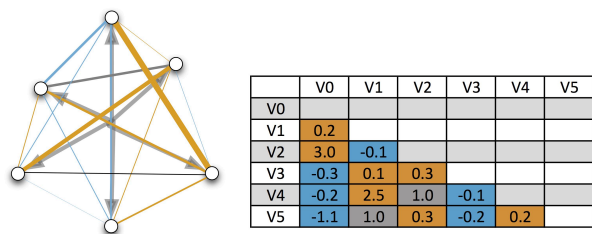


図2: 相関行列の可視化. 相関の高さとその正負を線の太さ正負を色で表現する

PCAの計算には指標のペアそれぞれの相関係数が必要であり, それは図2の右表のように与えられる. そしてこのとき各々の軸の相関は軸ベクトルの先端を結ぶ直線の色と太さによって表現させた.

5 結果

図2に実行結果例を示す. 提案システムは開発言語にObjective-C, アニメーションは主にQuartz Core-graphics(Core Animation)ライブラリを用いて実装した. 実装・実験はOS: Mac OS 10.6, CPU: Intel Core i7 2.66GHz, メモリ: 8GB, GPU: NVIDIA GeForce GT 330Mを用いた. 軸数100前後で十分インタラクティブな操作が可能であった.

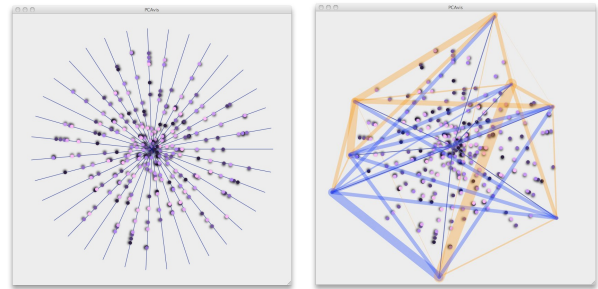


図3: システムの出力結果例

6 まとめと今後の課題

本稿では次元縮約過程を動的可視化手法を考察し解析の効率化と効果的な情報の提示について考察した. そしてPCAを用いた場合の具体的な可視化手法の提案を行い, 次元縮約の過程の把握を助けるシステムを実装した.

そしてこの提案手法は, 解析対象の副次的情報を全く考慮せず数値的処理のみに注目した可視化であることに注意する必要があるが, 他の次元縮約手法をでのマッピング手法の考察することで最終的に最適な次元縮約手法を選択するための静的な判断材料になりうると思う.

参考文献

- [1] P.I.A. Godinho, B.S. Meiguins, A.S. Goncalves Meiguins, R. M. Casseb do Carmo, M. de Brito Garcia, L. Hernandez Almeida, R.Loureno: IEEE information Visualization, 2007. IV '07. 11th International Conference 4-6 July 2007
- [2] H.-C. Hege, I. Hotz, and T. Muntzner: Eurographics/ IEEE-VGTC Symposium on Visualization 2009 Volume 28 (2009), Number 3