

多次元データを把握するための視覚化ツール

三治 信一郎 橋本 周司

早稲田大学 理工学部 応用物理学科
〒169-8555 東京都新宿区大久保 3-4-1
Tel. 03-5286-3233, Fax. 03-3202-7523
{sanji,shuji}@shalab.phys.waseda.ac.jp

概要 多次元データを視覚的に把握するための手法を提案する。多次元データを把握するための研究は広範囲な分野にまたがって行われているが、それらの多くはデータ全体を捉えようとするものであり、データの偏在性などがある場合、局所的なデータを解析することが難しい。

本稿では、具体的な企業データを用いて、同一ディスプレイ上でグローバルな情報とローカルな情報の解析を視覚的に行う方法について報告する。

A visualization tool for high-dimensional data mining

Shinichiro Sanji Shuji Hasimoto

Dept. of Applied Physics, Waseda University
3-4-1 Okubo, Shinjuku-ku, 169-8555 Tokyo
Tel. 03-5286-3233, Fax. 03-3202-7523
{sanji,shuji}@shalab.phys.waseda.ac.jp

Abstract We propose a new system to assist intuitive understanding of high dimensional data. It is difficult to describe the local characteristics of the data when the data are unevenly distributed. In this paper, the method of analytical visualization of the local and global information on the same display by using a statistical method and computer graphics. The system has been tested in the corporate data analysis.

1. はじめに

近年株価データなどデータの解析において多次元データを解析する必要性がでてきている。それに伴い、データの可視化の研究が行われている。可視化の目的は、高次元または複雑な構造をもつため理解しにくいデータを視覚的にわかりやすい形で提示することである。これより、そのデータの持つ意味を明らかにすることができる。

一方、コンピュータビジョンの分野では2次元の静止画像からの3次元形状の復元に関する研究がなされてきた[1]。また、2次元から3次元へと拡張したときと同様の手順を踏むことで4Dの情報を視覚化する研究が進められている[2]。これらの手法をデータの可視化に利用すれば、多次元のデータを提示する際問題となる情報量の増大に伴う人間の認知不足[3]の軽減が期待できる。

本研究では、最大4次元までのデータを大域的に提示するばかりでなく、ローカルな情報をインタラクティブに知ることのできる手法を提案する。

2. 研究目的

多次元のデータを提示し、直感的に捉える手法としては、S.Feinerらのn-Visionが知られている[4]。彼らは、多次元データを2つの3次元グラフを組み合わせで可視化し、データグループを用いてこれらのデータとのインタラクションを実現している。

また、3次元を超える多次元空間を把握するための手法としてハプティックインタフェースを用いて、高次元データとのインタラクションを実現した研究などがある[5]。

これらの研究では、科学技術データを主に扱っている。本研究ではより一般的な統計データを用い、分布の粗密等を効果的に提示することを目的とする。

多次元のデータを視覚化することで大域的な情報を把握することができるが、実際のデータ分布には偏りがあり、ローカルな特徴が全体とどのような関係を持っているかを知ることは難しい。

細部の情報解析を行うための手段として、主成分分析とクラスタ分析を行い、これらを同一画面上に提示する。さらに、詳細な情報にこれらの分析を再帰的に行うことで、ユーザの知りたい情報がどのような集団に属しているのか、さらに、その集団内での位置付けがどのようなものであるかということをはっきりとすることが可能になる。また、できる限り多くの情報を視覚的に把握することができるようにするために、本研究室において研究が進められている4次元情報提示システムを使用し、最大4次元までのデータ提示をおこなう。

3. 処理の流れ

図1は、本システムの処理の流れ図である。

まず、多次元のデータに対して主成分分析を行い、2から4次元の低次元のデータに変換し、ディスプレイ上に提示する。次に、主成分分析されたデータを用いてクラスタ分析を行い、クラスタに分割して表示する。その後でユーザがクラスタを指定すれば、再度そのクラスタ内で主成分分析を行い、結果をクラスタの重心を原点として同一画面上に提示することができる。この操作を繰り返すことで、ユーザは求めるデータの構造をインタラクティブに把握することができる。

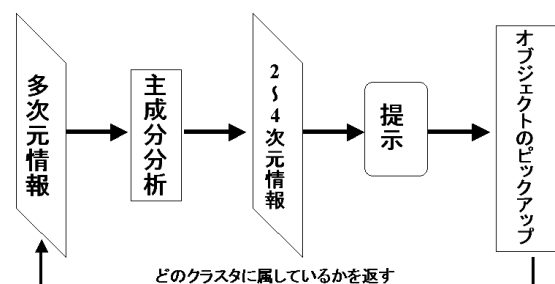


図1. 処理の流れ

4. 提示方法

4.1 統計手法

主成分分析とクラスタ分析が本システムで行う主な統計分析である。

主成分分析とは情報の損失をなるべく少なくして多次元の情報を低次元の情報で代表させる方法である。主成分分析の特徴は、多次元的情報を集約し、ほぼ 3、4 次元のデータで説明できる点と、集約された各変数の軸が幾何学的に直交しているという点である。

クラスタ分析とはデータを集団にわけ、分類付けする方法である。本システムでは、最短距離法・重心法など 5 種類の分類手法でクラスタ分析が行え、それぞれの手法の特徴の違いなども把握できるようになっている。なお、クラスタ分析で用いる非類似度は、ユークリッド距離を採用した。

4.2 提示方法

本システムでは、4 次元までのデータを提示できる。提示空間は 4 次元ユークリッド空間とし、3 次元から 2 次元上へ投影する方法を 4 次元から 3 次元へと拡張した表示方法を用いている（図 2、図 3）。つまり、4 次元データと 4 次元上の視点を入力し投影変換を行い 3 次元 CG として表示する。システムは、SGI Indigo2 上で開発し、3 次元表示するために液晶シャッター眼鏡（Stereo Graphics 社製 Crystal Eyes）を用いている。また、4 次元空間内ではマウス操作などにより、回転・縮小といった対話的な操作が可能である。

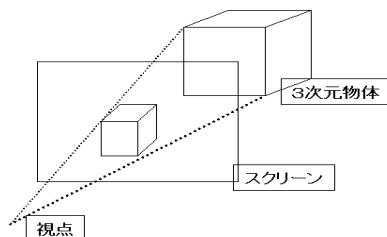


図 2 . 3次元から2次元への投影

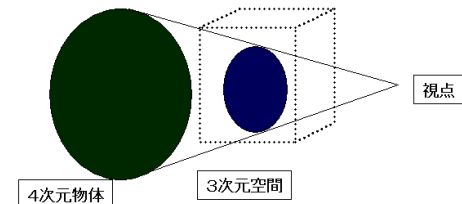


図 3 . 4次元から3次元への投影

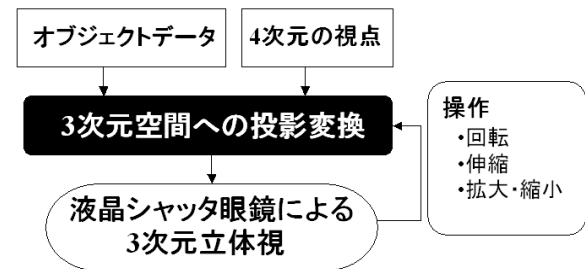


図 4 . 投影方法の概要

5. データ解析実験

東証一部上場の企業のうち、建築業・食料品などの業種の 350 社を無作為に選択し、従業員数、売上高など 24 項目のデータを用いて本システムによるデータ解析実験を行った。

図 5 は 350 社 24 項目のデータを主成分分析した結果を提示したものである。このうち、ユーザが指定したものが網状の球で囲まれている。さらに図 6 は、ピックアップしたデータ(黒点で表されている)が属するクラスタで再度主成分分析し、同一画面上に再描画したものである。ピックアップされたデータが属する個別の情報に関しては、データをピックアップした段階でテキスト情報として会社名、24 項目の情報と主成分得点が得られる。

実際のディスプレイ上で表示される画像では、主要なクラスタ 10 個に主だった色を割り当てるようにし、データがどのクラスタに属しているかを判りやすくしている。また、主成分分析において、主成分をいくつまで解析するかという基準は、「固有値 > 1」というカイザー基準を採用し、最大第 4 主成分まで求め提示した。

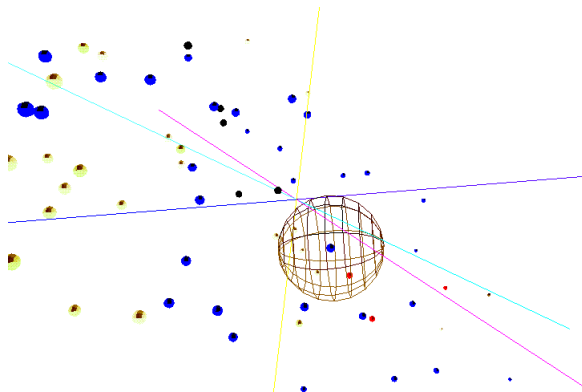


図5．オブジェクトのピックアップ

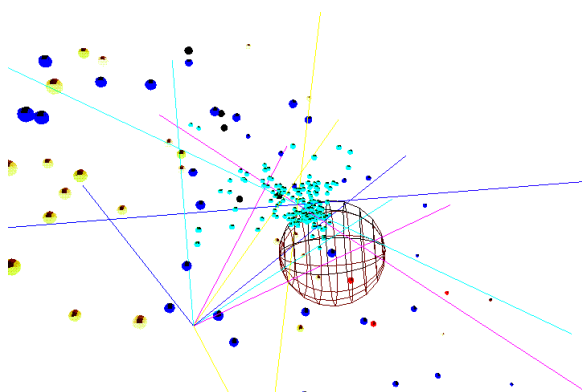


図6．ピックアップされたクラス内で再度主成分分析したもの

実際にデータの検証を行った結果、いくつかの興味深い事実が判った。例えば、一回目の主成分分析では、日本コカ・コーラの関連会社は同一クラスに属していたが、このクラス内で再度主成分分析した結果、地方特性等に左右され、3つほどのクラスに分類される。このような結果は、単純に一枚のグラフ上から得ることができないものである。

6. まとめ

本研究では多次元情報を把握するための手法として、最大4次元までの空間内で細部情報の解析が一度に行えるシステムを提案した。

これにより、従来では複数のグラフが必要な統計的なデータが同一ディスプレイ上で解析が行えるようになった。

しかし、本来人間の生活空間にない4次元空間をどのように把握するかという課題もある。

今後は、他の統計分析手法を実現するとともに、新しい入力デバイスを導入し、より使いやすいインタラクティブシステムを実現したいと考えている。

謝辞

本研究で使用したデータは(株)ゾイックスより提供されたものである。ご協力に感謝します。

参考文献

- [1] T. Kanade, "Recovery of the three dimensional shape of an object from a single view", *Artificial Intelligence*, Vol.17, No.1-3, pp.409-459, 1981
- [2] Makoto Murata, Shuji Hashimoto, "Interactive Environment for Intuitive Understanding of 4D Object and Space", in *Proceedings of the International Conference on Multimedia Modeling 2000 (MMM2000)*, pp.383-402, 2000
- [3] 小池英樹, "インタラクティブ3次元情報視覚化", *コンピュータソフトウェア*, Vol. 11, No. 6, pp. 20-31, 1994
- [4] S. Feiner and C. Beshers. "Worlds within world, metaphors for exploring n-dimensional virtual worlds", In *Proceedings of the ACM SIGGRAPH Symposium on User Interface Software and Technology (UIST'90)*, pp. 76-83, ACM Press, 1990
- [5] 橋本・岩田, "力覚を用いた多次元データブラウザ", *日本ヴァーチャルリアリティ学会論文誌*, Vol.2, NO.3, pp.9-16, 1997