

# 部分 PCP の空間充填による高次元データの可視化

末松はるか (指導教員：伊藤貴之)

## 1. 概要

高次元データの可視化には、Scatter Plot Matrix(SPM)や Parallel Coordinates Plot(PCP)がよく使用される。これらの可視化手法は 10 次元程度のデータには有効である。しかし、数十~数百におよぶ高次元データの可視化に適用させる場合、SPM であると各々の二次元散布図が小さくなり、PCP であると非常に横長な空間が必要となる。双方ともに、視認性の高い状態でディスプレイ上に表示するのが困難である、という問題が生じる。一方、高次元データの可視化の意義の一つに、次元間の相関関係の理解がある。高次元データではしばしば、任意の次元が多数の他の次元と同時に相関を持つことが想定される。そのため単一の PCP としての可視化は多彩な相関の発見を妨げ、相関関係を視覚的に発見するという可視化本来の重要な目的において十分ではない。

本研究では、PCP を発展させた数十次元以上の高次元データの相関性を理解する可視化手法を提案する。本手法では条件付き独立性の概念を適用して低次元 PCP (以後「部分 PCP」と称する) を生成し、この PCP 群を配置することでデータ全体の一覧表示を実現する。本手法によって、高次元データの次元間の相関を視覚的に把握することが容易になると考える。

## 2. 関連研究

### 2.1 高次元データにおける可視化手法

高次元データを可視化する代表的な手法として、Scatter Plot Matrix(SPM)と Parallel Coordinates Plot(PCP)[1]が挙げられる。

このうち SPM は、画面を格子状に分割し、二次元散布図(Scatter Plot)を配置する方法である。散布図は二次元間の相関を把握する時に有効である反面、三次元以上の相関を同時に見たい場合に適さない。またデータが数十次元を超えると、各格子のサイズが小さくなり視認性が低下する。以上の二点より本研究には採用しない。

一方で PCP は、次元一つ一つを縦軸として横に並列させて一覧表示する方法である。次元を横に整列させることで、二次元間の相関のみならず三次元以上で高い相関を示すものを同時にかつ直観的に把握することが可能となり、本研究に適していると考えられる。問題としては、次元数の増加に伴い非常に横長な画面空間を必要とする点が挙げられる。

### 2.2 FRUITNet

3 章で提案する手法では、FRUITNet (FRamework and User Interface for Tangled Segments Network) [2] というネットワーク可視化手法を用いて、多数の部分 PCP を画面

配置する。FRUITNet とは力学モデルと空間充填モデルを併用したアイテム集合付きネットワーク可視化手法である。この手法には、共通のカテゴリを有するノードを画面上の近くに配置するという特徴がある。それを活かして本研究では、次元の共有性の高い部分 PCP を画面上の近い位置に配置する。

## 3. 処理手順

### 3.1 条件付き独立性を用いた部分 PCP 生成

部分 PCP 生成には以下の条件を満たすことが必要となる。

- 1 各部分 PCP の次元同士は出来るだけ相関性を有する。
- 2 異なる部分 PCP 同士に属する次元同士は出来るだけ相関性を有さない。

そのための有力な方法として、条件付き独立性という概念を導入する。この手法では  $n$  個の変数の任意の組み合わせに対して条件付き独立性を判定し、任意の変数集合を与えた時、互いに独立にならない 2 変数は同じグループに入るように部分 PCP を生成する。条件付き独立性の判定は、 $n$  個の変数において互いに重ならない任意の 3 つの部分集合  $X_A, X_B, X_C$  を与えた時、次のように行われる。

1. それらを用いて計算される条件付き相互情報量(式(1))を計算する。
2. もしその値が極めて 0 に近い場合、変数集合 A と B は、C を与えた時に条件付き独立となると判断される。

例として 4 変数  $\{X_1, X_2, X_3, X_4\}$  の場合を考える。条件とする変数集合  $X_C = \{X_3\}$  として条件付き相互情報量を計算した時、 $X_A = \{X_1, X_2\}$ 、 $X_B = \{X_4\}$  となる条件付き相互情報量が 0 になった場合、4 変数は  $\{X_1, X_2, X_3\}$  と  $\{X_3, X_4\}$  に分割される。 $\{X_1, X_2, X_3\}$  と  $\{X_3, X_4\}$  のいずれも同様の手順でこれ以上分解されなくなった時、図 1 のようにこの 4 変数はこの 2 つのグループに分けられ、1, 2 の条件を満たす部分 PCP 生成を実現する。

$$I(X_A, X_B | X_C) \quad (1)$$

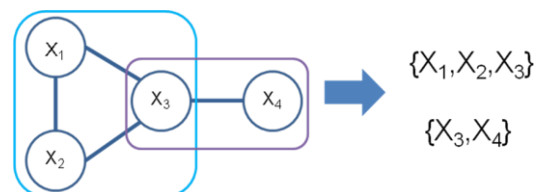


図 1 部分 PCP 生成

### 3.2 部分 PCP 群の配置

次に 3.1 で生成した部分 PCP 群に FRUITNet を適用させて図 2 に示すように配置する。

まず条件付き独立ではない部分 PCP ペアにエッジを生

成し、部分 PCP ペア間の次元の共有度の高さに応じて重みをつける(図 2 左上). この処理によって形成されたネットワークに対して、重みに応じたバネの力をエッジに与えた力学モデルを適用し、部分 PCP の適切な配置を得る(図 2 右上). この処理により、エッジの重みが大きい (= 次元の共有度が高い) 部分 PCP ペアは画面上で近くに配置される. また、重み付きエッジの長さや交差数の削減も同時に満たす. 続いて、この配置結果を座標値としてテンプレートに記述し、空間充填モデルにより配置を調整する(図 2 左下). 以上により、部分 PCP の画面上での重なりを回避し、かつ画面占有面積の増加を抑えながら再配置を実現する.

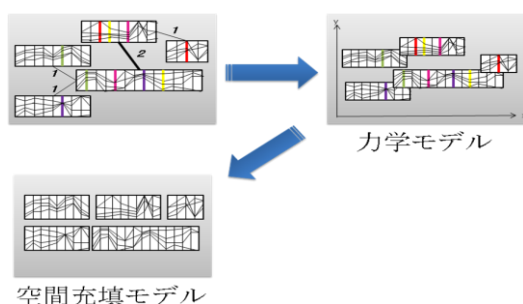


図 2 配置過程

#### 4. 実行結果

我々は UCI Machine Learning Repository で公開されている Segmentation Data[3]を使用して可視化を試みて図 3 のような実行結果を得た. このデータは折れ線数 210 本、全 19 次元の 7 枚の外部画像データで構成される. 各々の折れ線は 3×3 ピクセルの画像領域についてのデータであり、縦軸の次元は画像の領域重心、RGB 値、彩度等を表す. なお 7 枚の画像データはそれぞれ色分け表示されている. 各部分 PCP は 2~12 次元で構成されていて、エッジの総数は 20 本である. 我々の実装では、拡大縮小、平行移動、部分 PCP の ID や座標値の表示、などの各機能を搭載している.

数十次元以上の高次元データを可視化するにあたり、従来の PCP 手法では非常に横長なグラフが生成されてしまい、一般的に普及している汎用のディスプレイでの表示が困難な場合があった. それに対して本手法では、高次元データを低次元のサブグループに分割して各々を PCP として表示して、それをディスプレイサイズに合わせて配置する. そのため、視認性の高い状態でのデータ全体の可視化が容易となる.

また、次元の共有度が高い部分 PCP 同士が近くに配置されるため、部分 PCP 同士の比較も容易となる. 図 3 の左上部分を拡大した図 4 のうち次元 7, 9 は隣接する別々の部分 PCP から、次元 16, 3 と次元 12, 10, 5 のどちらも正の相関を有することがみてとれる. そのため従来の PCP 手法では視認できないような特定の次元に対する

他の三次元以上との相関性を視認することが容易になる.

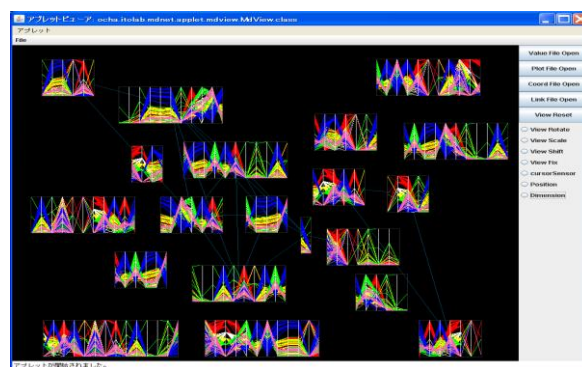


図 3 実行結果

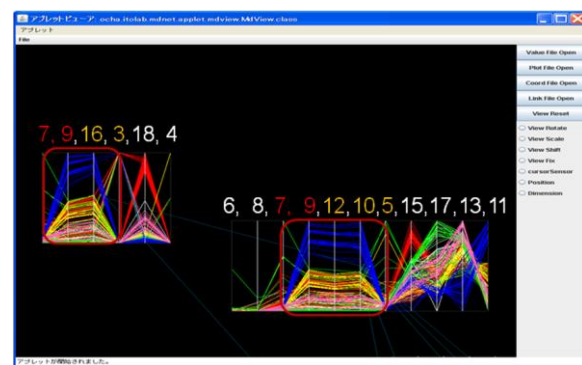


図 4 実行結果

#### 5. まとめと今後の課題

本研究では、高次元データのグループ化を行い、FRUITNet を利用して相関の高い部分 PCP が近くになるよう一画面に全データを配置した可視化手法を提案した. 以下に今後の課題を示す.

- 部分 PCP 生成の結果の検証
- 配置結果の検証
- 部分 PCP の見やすさの検証
- 折れ線のサンプリング
- 適応的な軸の反転

#### 謝辞

本研究を進めるにあたり多数のご助言をいただきました. NEC 藤巻遼平様、森永聡様、百瀬善文様、大阪大学河原吉伸助教に感謝の意を表します.

#### 参考文献

- [1] A. Inselberg, B. Dimsdale, "Parallel Coordinate: A Tool for Visualizing Multi-Dimensional Geometry," IEEE Visualization, pp. 361-370, 1990.
- [2] T. Itoh, C. Muelder, K.-L. Ma, J. Sese, "A Hybrid Space-Filling and Force-Directed Layout Method for Visualizing Multiple-Category Graphs," 2009 IEEE Pacific Visualization Symposium, pp. 121-128, 2009.
- [3] Image Segmentation Data Set  
<http://archive.ics.uci.edu/ml/datasets/Image+Segmentation>