

# 時系列データの三次元可視化の一手法

井元麻衣子 (指導教員：伊藤貴之)

## 1. 概要

近年の情報量の爆発的な増大に伴い、情報可視化の分野においても、さまざまな大規模データの可視化が試みられている。その中でも我々は、時系列データの可視化に着目している。我々は現在、気象の計測データやシミュレーション結果、遺伝子発現量の観察結果、原子力発電所の計測情報、などを例題として可視化の研究に取り組んでいる。

本研究では、大規模な時系列データを大局的にも局所的にも可視化するための、三次元可視化手法を提案する。本手法では大量の折れ線グラフ群を、類似度で奥行き方向に並べ替えて三次元空間に配置する。そして真上方向の視点から大局的に表示するとともに、その中から類似する有限本数の折れ線グラフを選択するユーザインタフェースを与える。同時に正面方向の視点から、選択された折れ線グラフを局所的に表示する。本手法によって、大規模時系列データの全体像を眺めながら、興味ある数値群を選択的に注視できる。

## 2. 関連研究

時系列データを大量の折れ線グラフ群で一画面に可視化する際に、線分同士の交差が多くなり、図が煩雑になるという問題点がある。Hochheiserら[1]は、複数の長方形を折れ線グラフ群に配置させるユーザインタフェースにより、類似する折れ線グラフを選択させる手法を提案している。しかし依然として、ユーザによる操作前の初期表示結果は煩雑であり、ユーザは興味深い折れ線を見つけるのが難しい場合がある。

それに対して内田ら[2]は、クラスタリングを利用した折れ線グラフの詳細度制御により、異なる動きをしているグラフを残して画面表示する手法を提案している。本研究の提案手法は、内田らの手法とは対照的に、類似するグラフを選びやすくするために三次元可視化を導入した手法である、と位置づけられる。

## 3. 提案手法

既存の時系列データ可視化手法の多くは、二次元(xy平面)上にグラフを描画する。それに対して本研究では、時系列データを表現する折れ線グラフを三次元空間(xyz空間)上に配置し可視化する一手法を提案する。これにより、大量の折れ線群を単一の二次元空間に描いた場合に起こりうる、折れ線同士の交差を解消する。

本手法では図1に示すように、x軸を時間軸、y軸を数値として、折れ線グラフをz軸上に等間隔に並べて可視化する。これにより、折れ線同士の複雑な絡み合いが解消され、折れ線1本の各時刻における数値を読み取れる。また、

三次元空間上に配置することで、複数の視点からグラフを観察できる。

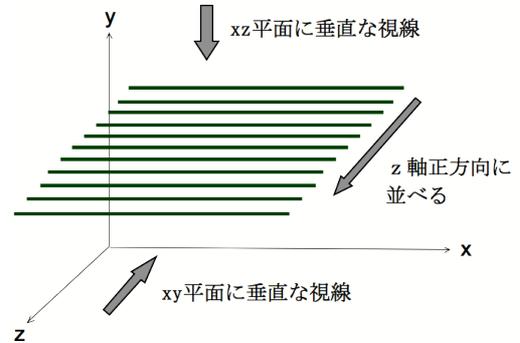


図1 xyz座標系に折れ線を並べたイメージ図

### 3.1 概要

提案手法は、おもに以下の3つの過程から成る。

1. 折れ線グラフをデンドログラムにより順列化
2. グラフ全体表示の表示と考察
3. 詳しく考察する特定の折れ線群の抽出と考察

### 3.2 折れ線グラフの並び替え

時刻  $i$  における  $j$  番目の数値を  $a_{ij}$  とし、 $j$  番目の数値群によって構成される1本の折れ線を  $A_j = \{a_{1j}, \dots, a_{nj}\}$  とする。このとき本手法では、任意の2本の折れ線間の距離を算出し、デンドログラムによって折れ線群を順列化する。ここで距離の算出には、いくつかの算出方式を予め用意する。ユーザが必要に応じて表示する算出方式を選択することで、目的に合った観点において類似する折れ線を比較観察できる。

ここでは、同一時刻に近い値を有する傾向が大局的に見られる折れ線が近くなるように折れ線を並べる場合について説明する。折れ線が  $N$  本あるとする。このとき、任意の2本の折れ線  $A_j, A_k$  間の距離の算出には最短距離法を用いる。まず、時刻  $i$  における2点のユークリッド平方距離  $\sqrt{(a_{ij} - a_{ik})^2}$  を時刻  $t_1$  から時刻  $t_n$  まで求め、以下の式

$$S_n(j,k) = \sum_{i=1}^n \sqrt{(a_{ij} - a_{ik})^2}$$

で和  $S_n(j,k)$  を算出する。

次に、 $S_n(j,k)$  の値が最も小さい2本の折れ線を併合し、1つのクラスタ  $K_1$  を生成する。ここで  $K_1$  は、含まれている2本の折れ線の各時刻における数値の重心をその時刻での数値として、それらにより生成される1本の折れ線とみなす。新しく折れ線を生成したことにより、折れ線の本数は  $N-1$  本となる。この操作を繰り返し折れ線の本数を減らしながらクラスタリングすることにより、デンドログラムを生成し、数値群を順列化する。このときデンドログラムの各要素を、 $S_n(j,k) \leq S_n(l,m) \leq \dots$  となるように、昇順に並べる。

### 3.3 グラフの全体表示と少数の折れ線群抽出

本手法では、xz 平面に垂直な視線と xy 平面に垂直な視線を用意し、両者を併用することで大域的な可視化と局所的な可視化の相互利用を容易にする。

まず図 2(左)に示す通り、xz 平面に垂直な視線により、折れ線グラフ群の全体を上から俯瞰する。このとき、折れ線グラフの各数値の変化を色で表現する。具体的には、y 座標値が大きければ暖色、小さければ寒色を与える。この配色により、各数値を概略的に表現する。

また、初期状態では視点をグラフ全体が一画面上に表示される位置に配置しているが、ユーザは視点を自由に操作でき、これによりユーザは興味のある少数の折れ線群を見つけることができる。

そして、特に着目したい折れ線群をその色分布から発見したら、その近くに視点の z 座標値を移動し、その部分にズームアップする。そして図 2(右)に示すように、クリックにより範囲を指定する。このとき、指定する折れ線群の本数は制限しない。この操作に伴って画面右側では図 3 に示すように、xy 平面に垂直な視線で、クリックされた範囲に含まれる折れ線群を可視化する。この場合も、視点の初期位置は折れ線群が全て一画面上に表示されるように配置するが、その後はユーザが自由に視点を操作できる。また、折れ線群のクリック指定を反復することも可能である。

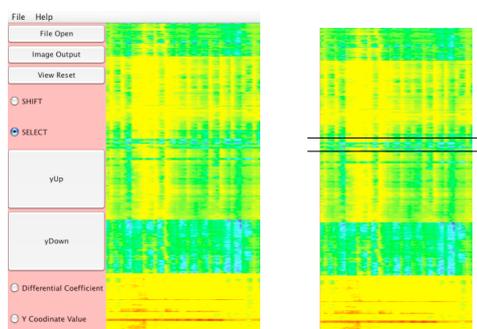


図 2 (左)グラフ全体を上から俯瞰する (右)2本の直線で示された箇所をクリックし、範囲を指定する

以上によって大量の時系列データの中から、注目する類似折れ線群を抽出し、その関連性や差異を観察できる。折れ線群の全体像を上から俯瞰することで、重要な意味をもつ折れ線グラフを見逃しやすいという問題点は解消される。

また xy 平面上での可視化において、折れ線グラフの y 座標値と色に別々の数値を割り当てることも効果的である。一例として微分係数から色を算出することを考えると、折れ線が右上がりの部分に暖色、右下がりの部分に寒色を与える。時系列データには数値の急激な変化を「異常」と捉える性質のものが多いため、このように微分係数に着目した可視化には異常発見などの意義があると考えられる。本手法では、折れ線の並び替えと色の割り当てに複数の選択肢を与えることにより、あるデータにおいて数パターンの描画を行うことがで

きる。

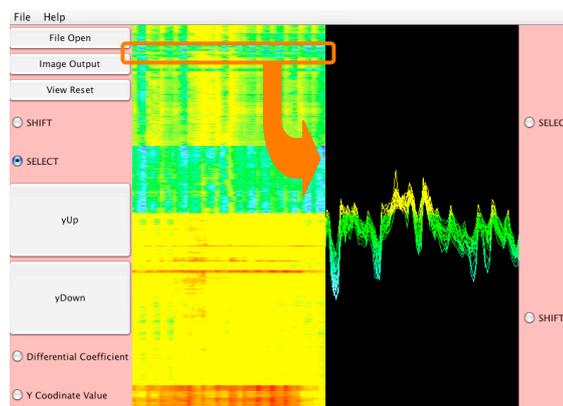


図 3 指定された範囲の中に含まれる折れ線群が画面右側に描画される。AMeDAS の気温データを用いた。

## 4. 適用事例

図 3 は AMeDAS が 2006 年 1 月に観測した全国 913 点の気温データを可視化した例である。この例において真上視点からクリックして抽出した気温は、時間変動が非常に類似しており、総じて朝と夜の気温差が小さかった。抽出された折れ線は主に秋田県・福島県・栃木県北部の気温に相当する。よって、これら一帯において、この時期に気温差の小さい天候が続いていたことがわかる。

## 5. まとめと今後の課題

本研究では、二次元(xy 平面)の折れ線グラフの時系列データを三次元空間(xyz 空間)上に、意味付けされた順番に配置して可視化する一手法を提案した。

今後の課題として、以下のような点を検討中である。

- ・ 色算出を厳密にして、数値の変化を鮮明に表示する。
- ・ 折れ線の並び方のアルゴリズムの一意性の確認と正当性の評価を行う。

## 参考文献

- [1] H. Hochheiser, B. Shneiderman, Dynamic Query Tools for Time Series Data Sets: Timebox Widgets for Interactive Exploration, *Information Visualization*, 3(1), 1-18, 2004.
- [2] T. Itoh, Y. Uchida, FRUITS Time: An Interactive Visualization Technique for Time-Varying Data, *Fifth International Conference on Flow Dynamics*, OS6-5, 2008.