

時系列データの三次元可視化の一手法

井元 麻衣子[†] 伊藤 貴之[†]

[†]お茶の水女子大学理学部情報科学科 〒112-8610 東京都文京区大塚 2-1-1

E-mail: [†] maiko_i@itolab.is.ocha.ac.jp, itot@is.ocha.ac.jp

あらまし 近年の情報量の急激な増大に伴い、情報可視化の重要性が向上している。その中で我々は時系列データの情報可視化に着目している。本報告では、時系列データの三次元可視化の一手法を提案する。本手法では、時系列データを表現する二次元の折れ線グラフを、三次元空間上に配置する。そして、真上方向の視点から時系列データ全体を俯瞰する可視化画面と、その中から興味深い折れ線群を選択させるユーザインタフェースを提供する。さらに、正面方向の視点から選択された少数の折れ線群を局所的に表示する可視化画面を提供する。本手法によって、ユーザ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できる。

キーワード 情報可視化, 時系列データ, 三次元可視化

A Three-dimensional Visualization Technique for Time-series Data

Maiko IMOTO[†] and Takayuki ITOH[†]

[†] Faculty of Information Science, Ochanomizu University 2-1-1 Otsuka, Bunkyo-ku, Tokyo, 112-8610 Japan

E-mail: [†] maiko_i@itolab.is.ocha.ac.jp, itot@is.ocha.ac.jp

Abstract Recent information explosion made information visualization techniques more important. We have been recently focusing on visualization of time-series data. This paper presents a 3D visualization technique for large-scale time-series data. The technique places line charts in a 3D space to visualize the time-series data. It then provides a view of the data from an upper viewpoint for the overview, and a user interface for the selection of interesting sets of time-series values. Moreover, the technique provides another view from an anterior viewpoint for the focus of the selected time-series values. Users can visualize large-scale time-series data both globally and locally by using the technique.

Keyword Information Visualization, Time-series data, 3D visualization.

1. 概要

情報可視化とは、文字情報として読むには大規模かつ複雑すぎる情報に、色や形を与えてわかりやすく表示する技術である。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。

本報告では、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。本報告の提案手法は、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. 提案内容

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。それに対して本報告では、

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。本手法では多数の時系列データに対して特徴量 (例えば微分係数の時間変化) を抽出し、関連性が強いと思われる折れ線が近くに表示されるように数値群を順列化し、各々の折れ線に z 座標値を与える。そして、xz 平面に垂直な視線と、xy 平面に垂直な視線を用意し、それらを併用することで大局的な可視化と局所的な可視化の相互利用を容易にする。その結果、時系列データ同士に対して、意外性のある新たな関連性を発見する可能性が広がると考えられる。

3.1. 概要

提案手法は、以下の 3 つの段階で構成される。

1. 折れ線グラフをクラスタリングにより順列化
2. グラフ全体表示の表示と観察
3. 詳しく観察する特定の折れ線群の抽出と観察

以下の節ではこれら 3 つの段階について説明する。

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。

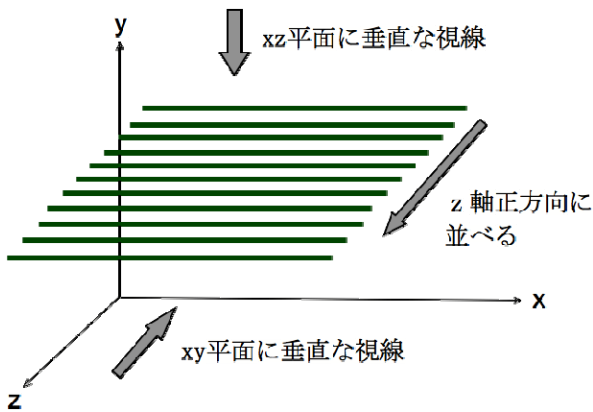


図 1. xyz 座標系への折れ線の並べ方。

3.2. 折れ線グラフの順列化

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 $A_j = \{a_{2j}, \dots, a_{nj}\}$ とする。このとき本手法では、以下のいずれかの判断基準により任意の 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、

- ・ 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、距離を算出する。

- ・ 同一時刻に近い値を有する傾向が局所的に見られる折れ線が近くなるように、距離を算出する。
- ・ 同一時刻でなくてもいいから、また値が大きく異なってもいいから、大局的または局所的に類似形状を有する折れ線が近くなるように、距離を算出する。

など、いくつかのパターンを予め用意しておく。ユーザが必要に応じて表示するパターンを選択することで、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 $\sqrt{(a_{ij} - a_{ik})^2}$ (図 2 参照) 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。

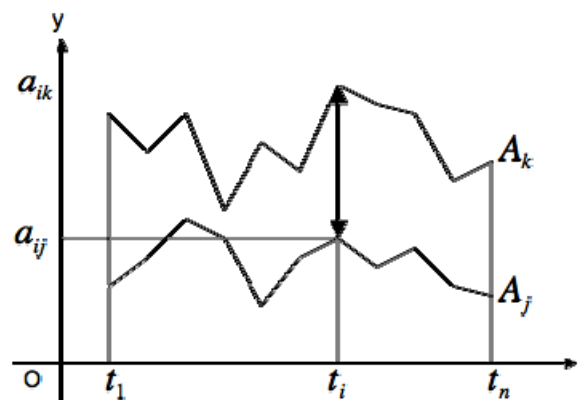


図 2. 2 本の折れ線の距離の算出。

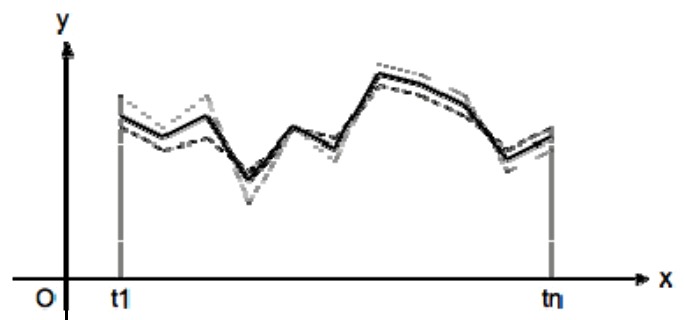


図 3. 新しく生成された折れ線。

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 (図 3 参照)。

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て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（図 4 参照）、数値群を順列化する。このとき、デンドログラムは、 $S_n(j,k) \leq S_n(l,m) \leq L$ となるように、昇順に並べる。

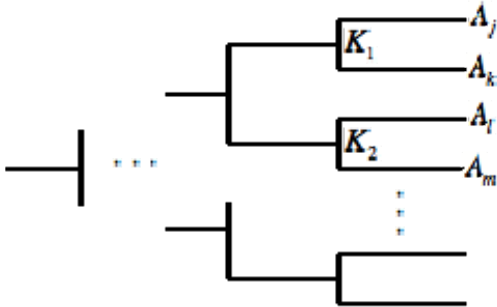


図 4. デンドログラムの生成.

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局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 5 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。これにより、具体的な数値として観測することのできない y 座標値を把握する。また折れ線をリボン状の面で表現し、シェーディングを施すことで、立体感を表現しやすくする。

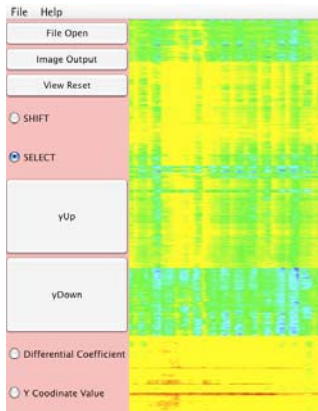


図 5. グラフ全体を上から俯瞰する.

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る。 y 軸方向の平行移動で一画面上に表示する折れ線群の本数を調整し、 x 軸・ z 軸方向の平行移動で時刻や表示する折れ線群を変える。この操作に

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 6 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 7 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、視点は指定した折れ線群が全て一画面上に表示されるように配置しているが、ユーザが自由に操作することができる。 z 軸方向の平行移動で一画面上に表示する折れ線群の本数を制御できる。また、 x 軸・ y 軸方向の平行移動で、時刻や表示する折れ線群を制御できる。さらに、折れ線群の指定は複数回行うことができ、本数もユーザが制御できる。

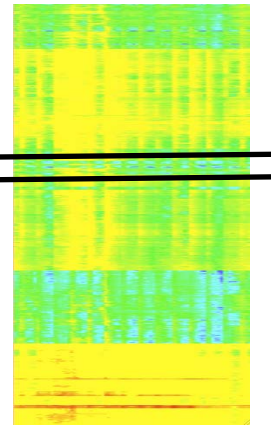


図 6. 範囲指定のために、2本の直線で示された箇所をクリック。

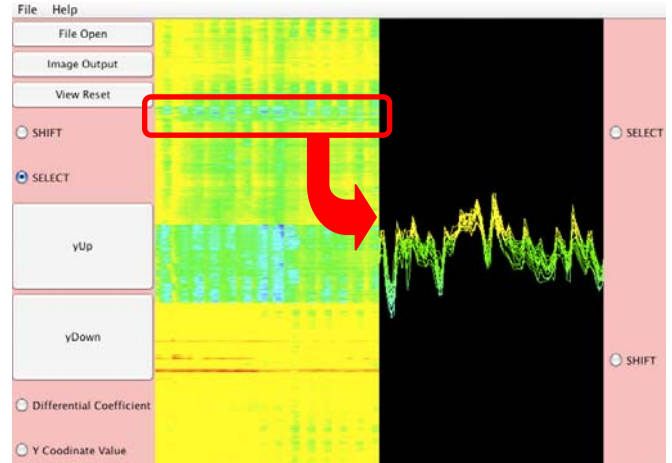


図 7. 指定された範囲の中に含まれる折れ線群が画面右側に描画される。

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落とすという問題点は解決されると考えられる。

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 (図 8 参照)。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が可能になる。

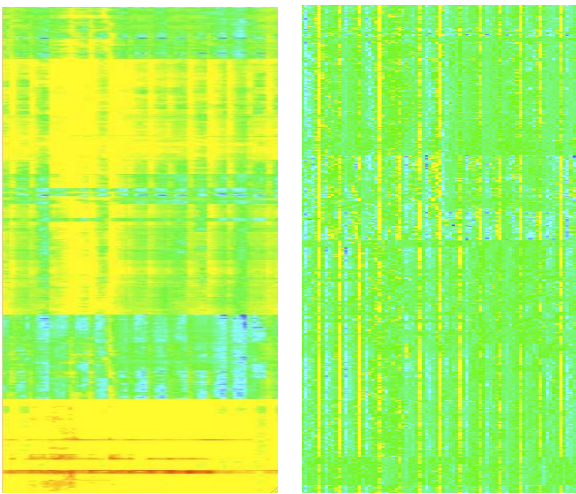


図 8. (左) y 座標値の値による色の割り当て。(y 座標値が大きければ暖色, 小さければ寒色.)
(右) 微分係数の値による色の割り当て。(折れ線の右上がりの部分に暖色, 右下がりの部分に寒色.)
※データ, 視点はともに同じである。

4. 適用事例

図 9 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い。抽出された折れ線は主に茨城県・石川県・奈良県の気温である。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。

ここで興味深い点として、円で囲まれた時間帯においては、気温が高く上がった地域と低い天候が続いていた地域に分かれている。これは真上視点での認識は難しいが、正面視点では明確に認識できる。このように本手法を用いることで、まず大局的に時系列データを眺めて興味深い類似データ群を抽出し、続いてそれらを局所的に比較することが容易になる。

5. まとめと今後の課題

本報告では、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可視化する一手法を提案した。

現在の本手法の問題点として、以下の 2 つを挙げる。

- ・ リボンの色算出が必ずしも適切ではない。
- ・ 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がなされていない

まずリボンの色算出については、グラフ全体を俯瞰し少数の折れ線群を抽出する際にユーザの判断材料として大きな比重を占めると考えられるため、ユーザの視覚的認知を高めるような色算出が必要である。現在は色算出に単純な一次関数を用いているが、今後は他の関数を併用することで、ユーザが興味深い時系列データをより直感的に見発見できることを目指す。また、折れ線の並び方のアルゴリズムについては、現在はユークリッド空間で距離を算出しているが、今後は他の空間での距離算出も検討する予定である。また、現在は順位付けが必ずしも一意には決定しないという問題点もある。この問題点については、今後さらに厳密なルールを加えることにより解決していく。

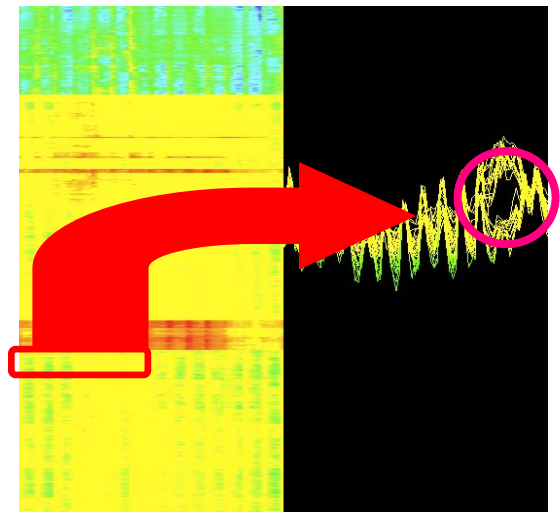


図 9. AMeDAS が観測した気温データを本手法に適用

文 献

- [1] H. Hochheiser, B. Shneiderman, “Dynamic Query Tools for Time Series Data Sets: Timebox Widgets for Interactive Exploration”, Information Visualization, Vol. 3, No. 1, pp. 1-18, 2004.
- [2] Y. Uchida, T. Itoh, FRUITS Time: An Interactive Visualization Technique for Time-Varying Data, Fifth International Conference on Flow Dynamics, OS6-5, 2008.