

FRUITS Time

大規模時系列データの詳細度制御を利用した可視化

内田悠美子[○], 伊藤貴之 (お茶の水女子大学大学院 人間文化創成科学研究科)

FRUITS Time: A visualization technique using level of detail control for large scale time series data

Yumiko UCHIDA and Takayuki ITOH

ABSTRACT

This paper presents a technique for visualizing large scale time series data. This technique represents variation over time in data as line charts.

We cannot figure out feature of time series data, when we draw the data which contain a large amount of elements as line charts in one display. To solve this problem, this technique uses a level-of-detail control technique. The technique applies a clustering technique based on shapes of charts, and selects representative line charts for each cluster. It displays only representative charts for level-of-detail control. Moreover, the technique can interactively control level-of-detail on the display. We suppose there are various applications of the technique since we can figure out which the chart has interesting behaviors from the visualization result applying level-of-detail control.

This paper shows the visualization result of volume data obtained from weather simulation and observed temperature data throughout Japan as examples of results of the proposed technique.

Keywords: Visualization, Time series, Line chart, Graph

1. 概要

時系列データの可視化手法の多くは、横軸に時間軸、縦軸に値を取り、各時刻の値をプロットした各点を線で結ぶ折れ線グラフ形式をとっている。この形式は、各要素の値の時間変化を直感的に捉えやすい利点ゆえに、広く用いられている。しかし、大量の折れ線群を単一の座標空間に描くと、折れ線同士の交差が多くなり、値の時間変化を適切に読み取るのは困難になる。

この問題を回避するために本報告では、特徴的な動きのある部分の強調や、類似する動きの要約的な表示などによって、大量の要素を含む時系列データを効果的に可視化する手法を提案する。本手法が対象とする時系列データの代表例として、気温変化、株価の変動、船舶の方向角速度、地震の揺れの記録などがある。

なお我々は、「絡まった線分群を有効に可視化するための仕組みとユーザインタフェース」に関するいくつかの研究に着手しており、そのコンセプトを FRUITS (=FRamework and User Interface for Tangled Segment-sequences) と名づけている。このコンセプトに基づく研究の中で我々は、時系列データを可視化する本手法を「FRUITS Time」と名づけている。

2. 関連研究

折れ線グラフ形式の時系列データ可視化手法に限らず、大量の線分群によって情報を可視化する手法には、線分同士の交差が多くなり、図が煩雑になるという問題がある。Holten ら[1]は、階層型グラフデータの可視化結果を構成する線分群を、スプライン関数を用いて束にすることによって、図が煩雑になるのを回避する手法を提案している。また Parallel Coordinates による多変数データの可視化において Ellis, Dix ら [2]は、線分群のサンプリング数を減らして表示することにより、可読性を高める手法を提案している。Hauser ら[3]は、多変数データの 1 変数に着目して線分の濃淡に変化を付けた表示や、線分の角度がある値に近いもののみを強調表示する手法を提案している。

一方で時系列データのグラフ表示における研究では、Wattenberg らによって、形状に基づくグラフ検索がおこなわれている[4]。ここでは、ユーザが求めるグラフ全体の形状を画面上で描画し、システムはこれに近い形状を持つグラフを複数提示している。

また、Shneiderman らによる[5]では、時系列データを表す大量の折れ線グラフ群から、グラフの傾きの角度・グラフの通過範囲等からのグラフ検索を提案して

いる。

3. 提案内容

3.1 可視化に用いるデータ

可視化に用いるデータの例として、気象シミュレーション結果のボリュームデータを可視化する。Fig.1は、代表的な格子点における100ステップのスカラ値の時系列変化を、折れ線グラフで可視化した例である。この可視化結果から、いくつかの格子点に対応する折れ線（以下、一連の折れ線を形成する単位を「要素」と称する）において、動きの類似性、連動性を確認できる。しかしこのような単純な可視化手法では、大規模なデータの場合に図が煩雑になり、こういった特徴は見落としやすくなるものと考えられる。

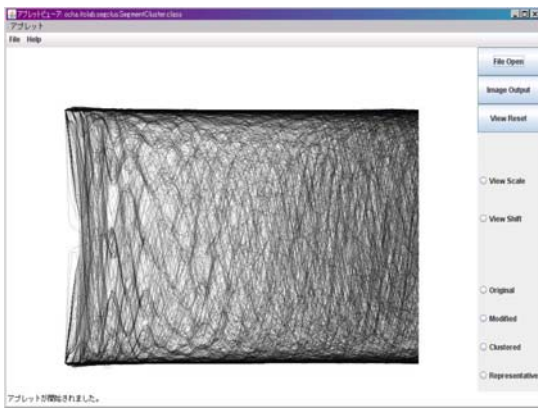


Fig.1: Time-varying scalar values at grid points in volume data.

3.2 本システムでの表示方法

本報告では、大規模な時系列データの変化の特徴を効果的に表現する手法を提案する。大規模なデータにおいては線分の数が莫大であるため、煩雑が生じやすい。しかしこの中には、類似した時間変化によって重なりあっている部分も多い。このような類似部分を要約表示することによって、全体の煩雑さを抑えることが出来ると考えられる。本手法では、データ全体での値の変動の特徴は保持したまま、グラフの本数を削減することで、時系列データの要約表示を実現する。

3.3 代表点の生成

本手法では、まず各要素の時系列変化を近似した代表点を生成する。最初に画面上にて、x軸を等間隔に区切って格子化する。そして各要素 C_k について、 i 本目の格子線と交差した点を P_{ki} として記録する (Fig.2左)。次に、y軸方向を等間隔に区切って格子化する。そして P_{ki} に隣接する格子点を選び、その点を新たに P_{ki}' とする (Fig.2右)。以下、 $P_{k1}' \sim P_{kn}'$ を要素 C_k の代表点と呼ぶ。

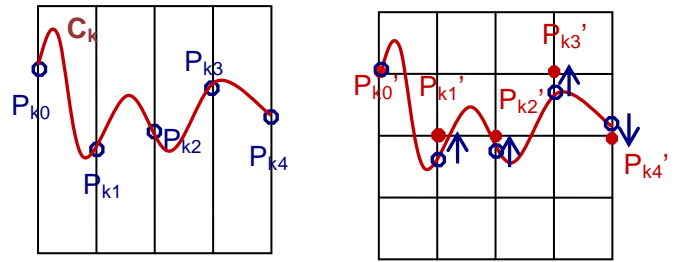


Fig.2: Calculation of representative points.

3.4 代表要素の選択

次にy軸方向の格子ごとに、要素をグループ化する。ここで m 列目の格子内において、 $P_{k(m-1)} = P_{i(m-1)}$ かつ $P_{km} = P_{im}$ を満たす要素 i があれば、 k は i と同一のグループに配置される。

つづいて、グループごとにグループ内の要素のクラスタリングを行う。ひとつの格子内に n 回のタイムステップにおける値がプロットされている場合を例にあげる。このとき提案手法では、グループ内の要素の各タイムステップにおける y 座標値を値にもった n 次元ベクトルを生成する。この n 次元ベクトルをグループ内の各要素について生成し、ここで得られたベクトルを用いてクラスタリングをおこなう。このベクトルは各グラフの時系列に伴う値の変動を表現している。よって、この処理によって、グラフの形状に基づくクラスタリングをおこなうことが可能である (Fig.3)。

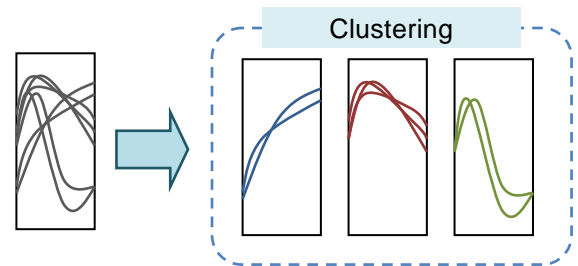


Fig.3: Clustering of graphs.

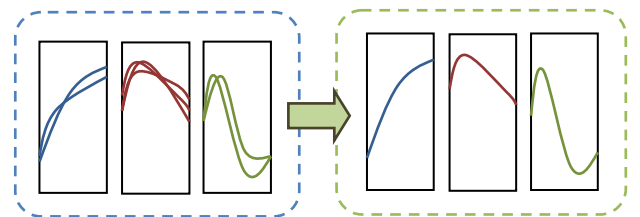


Fig.4: Representative element selection for each cluster

以上の処理によって得られた各クラスタから、クラ

スタの中心にあたる要素を、代表要素として選択する (Fig.4).

$m+1$ 列目以降の格子内においても、同様にグループ化・クラスタリングを行い、代表要素を選択する。このとき、以前に代表要素に選ばれたことのある要素がクラスタ内であれば、その要素を優先して選択する。無い場合にはそのグループは保留とし、次のクラスタに移る。

すべてのクラスタについて、代表要素の選択を試みた後、再度、保留としたクラスタの代表要素を選ぶ。このときも同様に、以前に代表要素に選ばれたことのある要素がクラスタ内であれば、それを優先して選択する。この処理を反復することにより、短い計算時間で、かつ適切に、少数の代表要素を選択できる。

3.5 表示

本手法では 3.4 節で選択された代表要素のみを表示することにより、詳細度制御を実現する (Fig.5).

なお、本手法では格子の幅を自由に設定できるものとする。格子の幅を狭くすることで、小さな特徴を捉えやすくなる。逆に格子の幅を広くすることで、大きな特徴のみが強調される。

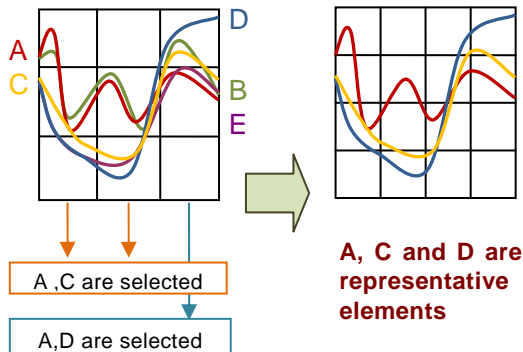


Fig.5: Selection of representative elements and display of them.

3.6 GUI

本手法では、詳細度制御後に表示したグラフのいずれかをマウスのクリック操作で指定すると、そのクリックした点を通過するグラフをハイライト表示する。また同時に、詳細度制御によって表示されなくなったグラフのうち、ハイライト表示されたグラフと同じクラスタに属するものも、同時にハイライト表示する。この操作により、詳細度制御によって画面上に表示されなくなった数値も可視化することができる。

4. 実行結果

Fig.6 は、Fig.1 で示した気象シミュレーションのボリュームデータを、提案手法を用いて詳細度制御を適用して可視化した例である。0~10 ステップ前後にお

いて、値が急激に下がっている様子が強調されている。また、20 ステップ前後での値の増減、80 ステップ前後での値の急降下と急上昇が確認できる。このように、本手法を用いることで、詳細度制御なしには捉えにくかった特徴的な時間変化を、視覚的に確認できる。

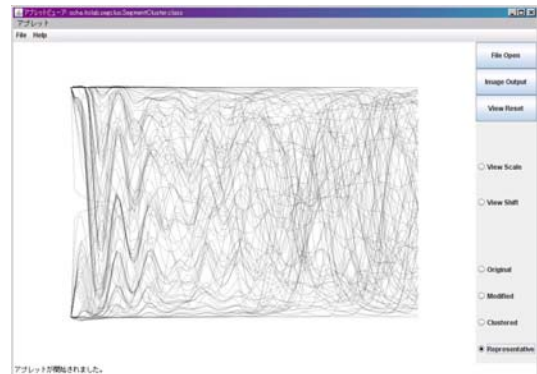


Fig.6: Applying the proposed technique for time-varying scalar values at grid points in volume data.

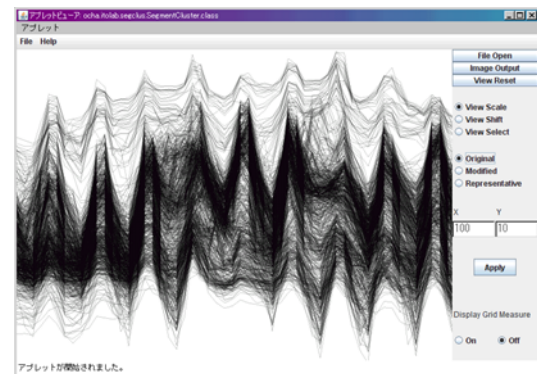


Fig.7: Visualization result of observed temperature data throughout Japan using line chart

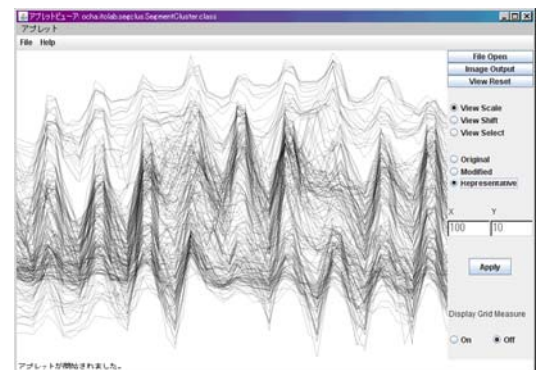


Fig.8: Applying level-of-detail control for Fig.7 using the proposed technique

Fig.7 はアメダスによって記録された日本の 1000 か所以上の計測点における、気温変化を折れ線グラフ化

したものである。このデータに提案手法を適用した可視化例が Fig.8 である。線分数はおよそ 6 分の 1 に減っているが、全体的な特徴を保持していることが確認できる。

このように提案手法を用いて詳細度制御をおこなうことにより、どのグラフが特徴的な動きをとっているかの判断が容易となり、重要な要素・時刻を知る手がかりとなる。

Fig.9 は、Fig.6 で示した可視化結果に対して、3.6 節で提案した GUI を用いて、クリックした点を通るグラフ、およびそれと同じクラスタにあるグラフをハイライト表示した例である。

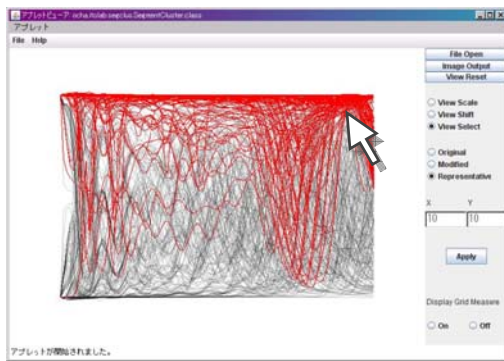


Fig.9: Display of not only selected graphs but also other graphs in the same cluster with the selected graphs

5. 今後の課題

現在の本システムの問題点として、詳細度制御後の表示結果からユーザの操作無しには、詳細度制御前に同一クラスタに含まれていた元のグラフの本数を知ることができないという点が挙げられる。この問題点を解決するために、同一クラスタに含まれるグラフの本数をグラフの線分の太さで表現する手法の実装を予定している。

今後の課題としては、要素の動きの特徴・ボリュームデータ上の位置などに基づき、折れ線の色分け機能を実装したい。また、ユーザがマウス操作により自在に要素を選択・抽出できるように、ユーザインタフェースを充実させたいと考えている。具体的には、ユーザがグラフの部分形状をマウスで入力すると、近い部分形状を持つグラフを検索する機能、および、グラフの通過範囲を指定することによる絞り込み機能等を検討している (Fig.10)。これにより、たとえばある時刻において値が大きく変動したグラフを探す、といった直観的なグラフ検索が可能となる。

また本研究では、このボリュームデータ・観測データのほか、新聞におけるキーワードの登場頻度の変

化・システムにおける計測データなど、さまざまな時系列データへの適用を検討している。

そして最終的には他の可視化システムと連動させ、時系列の変動の持つ意味を、多角的に解明するシステムの構築を目指したい。

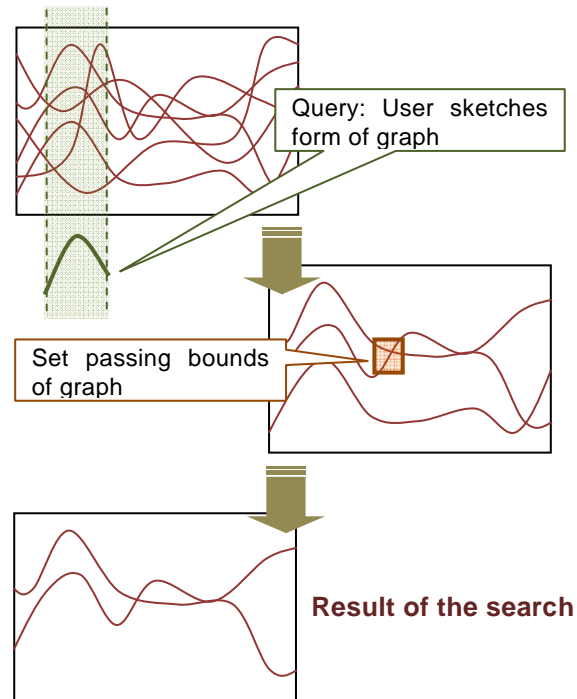


Fig.10: Search tool based on shapes of graphs and passing bounds of graphs

参考文献

- [1] D. Holten, Hierarchical Edge Bundles, IEEE Transaction on Visualization and Computer Graphics, Vol.12, No.5, pp.741-748, 2006.
- [2] G. Ellis, A. Dix, Enabling Automatic Clutter Reduction in Parallel Coordinate Plots, IEEE Transaction on Visualization and Computer Graphics, Vol.12, No.5, pp.717-723, 2006.
- [3] H. Hauser, F. Ledermann, H. Doleisch, Angular Brushing of Extended Parallel Coordinates, the Proceedings of the IEEE Symposium on Information Visualization 2002, Oct. 28-29, 2002.
- [4] M. Wattenberg, D. Jones, Sketching a Graph to Query a Time-Series Database, In Processing of the SIGCHI Conference on Human Factors in Computing Systems Extended Abstract (CHI2001), pp.381-382, 2001.
- [5] H. Hochheiser, B. Shneiderman, Dynamic query tools for time series data sets: Timebox widgets for interactive exploration, Information Visualization, Vol.3-1, pp.1-18, 2004.