

大規模時系列データの可視化と対話的な詳細度制御の一手法

内田 悠美子 (指導教員：伊藤 貴之)

1 はじめに

情報技術の普及に伴い、計算機システムのデータベースには非常に多種多様な、かつ膨大な情報が蓄積されている。その情報の中には、時系列データで表現できるものが数多く存在する。

時系列データの可視化に関する研究は数多く行われているが、既存手法の多くは、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び折れ線グラフ形式をとっている。この形式は、各要素(折れ線グラフ1本)の値の時間変化を直感的に捉えやすいという利点ゆえに、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。

この問題を解決するために本論文では、大規模な時系列データの詳細度制御を利用した可視化の一手法を提案する。本手法では時系列データの表示に折れ線グラフを用い、クラスタリングを利用してグラフの本数を減らすことで詳細度制御を実現する。

また本システムでは、詳細度制御後の可視化結果画面から、ユーザが着目したグラフをインタラクティブに抽出できるよう、グラフ検索のインターフェースを実装した。このインターフェースでは、ユーザは着目するグラフ形状を画面上でなぞる(スケッチする)ことにより、目的のグラフを抽出することができる。

2 関連研究

時系列データの可視化では、各要素を一本の折れ線グラフで表示する可視化手法が一般に広く普及している。近年は大規模なグラフ群から目的のグラフを抽出するための研究が進んでいる。一例として、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[1]。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[2]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。

またFuaらによる[3]では、多変数データのクラスタリングをおこない、クラスタごとの色分け、クラスタの中心を強調した表示をおこなっている。クラスタの中心を示す線の周りを、クラスタ内の要素数・広がりに応じた幅をもつ透明度の高いbandで囲っているため、データの全体の特徴をみることが出来る。クラスタの中心を強調し、他の要素を目立たなくするというアイデアは本論文に近いものである。

3 提案内容

3.1 詳細度制御

3.1.1 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標本化をおこなう(図1左)。さらに、 y 軸方

向を等間隔に区切って量子化をおこなう(図1右)。ここで、標本化・量子化に用いた間隔を格子と呼ぶ。

次に、 t 番目と $(t+1)$ 番目の標本化間隔ごとに、要素をグループ化する。ここで2つの要素の t 番目と $(t+1)$ 番目における量子化した値がそれぞれ等しい場合に、これら2つの要素は同一のグループとして扱う。

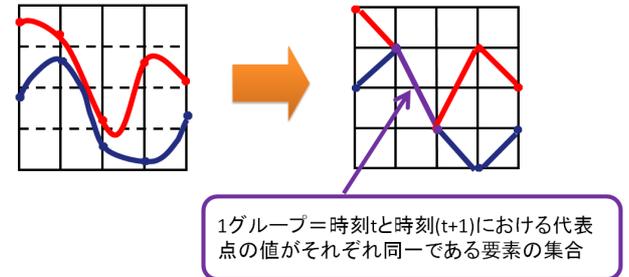


図 1: 代表点の生成とグループ化。

3.1.2 代表要素の選択

つづいて、量子化した値を元に戻し、それぞれのグループごとにクラスタリングを適用する。本手法では、 t 番目と $(t+1)$ 番目の標本化間に n 個のタイムステップが含まれている場合には、グループ内の各要素を n 次元ベクトルとして扱い、クラスタリングに用いる。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おり、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(図2)。

提案手法では、クラスタリング手法として非階層型クラスタリング手法であるk-means法を用いている。k-means法は、セントロイド c_i (クラスタ C_i の重心点)をクラスタの代表点とし、

$$\sum_{i=1}^k \sum_{x \in C_i} ((D(x, c_i))^2)$$

の評価関数を最小化するように k 個(予め指定した数)のクラスタを分割する。得られたクラスタ内部でセントロイドをとり再度クラスタに分割しなおすという方法を繰り返す。ここで x は各要素を示し、各データとクラスタのセントロイド c_i とのユークリッド距離 $(D(x, c_i))$ が最小となる分割を求めている。以上の処理によって得られた各クラスタから、クラスタのセントロイドに最も近い要素を、代表要素として選択する。

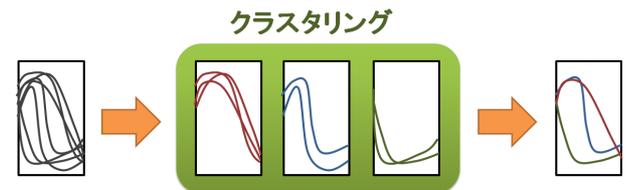


図 2: グループごとにクラスタリングをおこなう。

3.1.3 表示

3.1.2 節で選択された代表要素のみを表示することで、詳細度制御を実現する。ここで、代表要素の数は格子の大きさに依存する。本手法では格子の幅がユーザが自由に設定可能であり、詳細度制御の結果を見ながら代表要素の数を調節することが出来る。格子の大きさによる可視化結果の違いを図3に示す。また表示の際には、クラスタ内に含まれる要素数に応じて、代表要素の透明度を設定する。本手法ではクラスタ内の要素数が多いものほど、低い透明度で描画されている。この処理により、詳細度制御によって元のデータにおける本数の情報が失われるのを防いでいる。

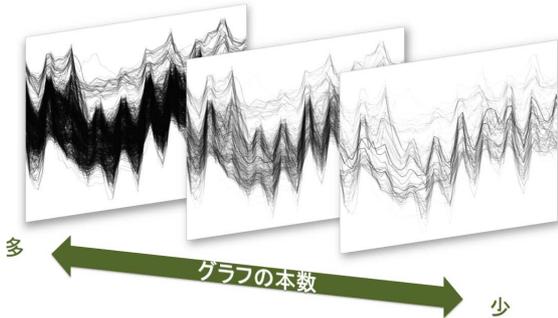


図 3: 格子の大きさによる可視化結果の違い。左: 元データ, 中央: 小さな格子を用いた詳細度制御, 右: 大きな格子を用いた詳細度制御。

3.2 GUI

3.2.1 スケッチ入力

本システムでは、詳細度制御適用後に表示した要素のいずれかの形状の一部をマウスでなぞる(スケッチする)と、なぞられた要素をハイライト表示することが可能である(図4)。

ここではスケッチ入力されたグラフ(入力グラフ)の各頂点のy座標値と、代表要素の頂点のy座標値を比較する。具体的には入力グラフの一つの頂点と、この頂点とx座標値が最も近い代表要素の頂点に関して比較をおこなう。入力グラフのすべての頂点に対して比較をおこない、すべての頂点との距離が閾値以下である要素を選択要素としてハイライトする。

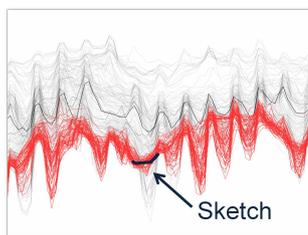


図 4: スケッチにより選択したグラフをハイライト表示。

また複数回のスケッチ入力に対し、論理演算をおこなうことが可能である。図4で示した1回目のスケッチ入力に基づく可視化結果に対して、2回目のスケッチ入力をおこなった可視化結果を図5に示す。ここでは、それぞれ2回の入力に対する論理和・論理積をと

り、演算によって得られた要素を選択要素としてハイライト表示している。

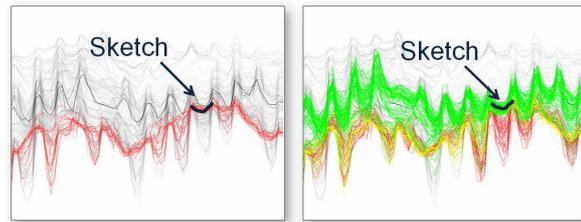


図 5: 左: 論理積演算をおこなった結果, 右: 論理和演算をおこなった結果

3.2.2 詳細度制御前の要素を復活表示

また本システムでは、詳細度制御によって表示されなくなった要素のうち、選択要素と同じクラスタに属する要素も、同時にハイライト表示することが可能である。この操作により、ユーザは全体を詳細度制御を適用した状態で確認することが出来る、さらに着目した部分のみ元の詳細度で見ることが出来る(図6)。

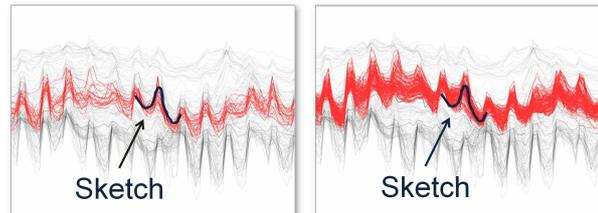


図 6: 詳細度前の要素を復活表示。左: 選択された代表要素のみをハイライト, 右: 選択要素と同じクラスタに属する要素も同時にハイライト。

4 まとめ・今後の課題

本論文では時系列データの対話的な詳細度制御、およびスケッチベースの要素選択機能を提案した。現在は気温観測データやボリュームデータを用いて実験をおこなっている。その他に本手法を適用可能なデータとしては、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が考えられる。今後は本システムで多様な時系列データの可視化をおこない、本手法の有用性を実証する予定である。またクラスタリングが適切であるかの検証、ユーザテストによる評価を検討している。

参考文献

- [1] M. Wattenberg, D. Jones, Sketching a Graph to Query a Time-Series Database, *Processing of the SIGCHI Conference on Human Factors in Computing Systems Extended Abstract*, pp. 381-382, 2001.
- [2] 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.
- [3] Y. Fua, M. Ward, E. Rundensteiner, Hierarchical Parallel Coordinates for Exploration of Large Datasets, *Proceedings of Visualization '99*, pp. 43-50, 1999.