

階層型データ可視化手法「平安京ビュー」への時系列情報の付加表示

橘春帆* 伊藤貴之**

(*) お茶の水女子大学大学院 人間文化研究科

(**) お茶の水女子大学 理学部 情報科学科

1. 概要

情報可視化とは、私たちが日常生活や日常業務で扱う情報を可視化する研究分野である。

情報可視化が対象とするデータは大きく分けて 7 種類に分類されるといわれる[1]。7 種類とは、1 次元データ、2 次元データ、3 次元データ、 n 次元データ($n > 3$)、時系列データ、階層型データ、リンクデータである。最近では、この 7 種類の複合的なデータ、例えば時系列かつ階層型、といったようなデータの可視化に関する研究事例が増えている。しかし、1 個のウィンドウには、その複合的な情報全体を表現するのは難しい。この問題を回避する一手段として、2 個のウィンドウにわたって情報を表現し、お互いを連携操作できるようにする、という手段が考えられる。

本研究では、階層型データ可視化手法「平安京ビュー」に時系列情報を付加表示する一手法を提案する。

本研究では、数百～数千個の数値が時系列に沿って変化するような情報を想定する。本研究では、この数値を何らかの規則に基づいて階層化し、時系列階層型データとして可視化する。この際に 2 個のウィンドウを用いて、一方では正方形ウィンドウ領域に階層型情報を表示し、他方では縦長のウィンドウ領域に時系列情報を表示する。この縦長のウィンドウ領域の縦方向に時間軸をおき、横方向に数値分布を表現することで、時系列にそった数値変化を表現する。

2. 関連研究

情報可視化手法「平安京ビュー」[2]は大規模な階層型データの可視化に向いている。平安京ビューは階層型データ全体を一画面に配置する情報可視化手法である。葉ノード、枝ノードの直交配置を意識したアルゴリズムで、平安京の地図に似ていることから命名された。

時系列データ中のある時刻における数値を、階層型データとして可視化した事例として、計算機ネットワークの不正アクセス検出結果の可視化[3]や、ウェブサイトのアクセス分布の可視化[4]がある。

本研究では、このように特定のシステムだけでなく、応用的にたくさんのシステムに使えるような手法の構築を目標としている。

3. 提案内容

本研究で扱うデータは、横軸が数値の種類、縦軸が時刻であり、表の各欄に実数値が格納された表形式のデータを前提とする。本研究で扱う数値は、例えば一定時刻ごとに記録された複数のログに記載された数値を指す。

本報告の提案手法では図 1 に示すように、一画面に 2 個のウィンドウを用いる。左側のウィンドウは、数値を何らかの規則に基づいて階層化した階層型データを表示するために用いる。右側のウィンドウは、時系列データを表示するために用いる。

階層型データを表示する方法には平安京ビュー[2]を使用する。平安京ビューは、長方形の入れ子構造で階層型データを表現するので、同じ階層に属する数値は同じ長方形の枠の中に表示される。

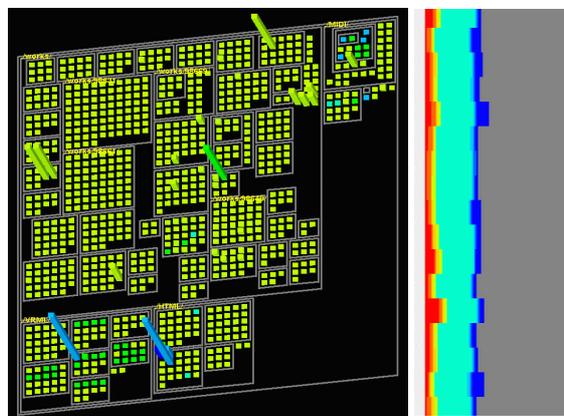


図 1 (左)「平安京ビュー」による階層型データの可視化。(右)提案手法による時系列データの可視化。

ここで図 1 のような可視化結果を、一般的に普及しているディスプレイ画面に表示することを考える。例えば横 1400 画素、縦 1050 画素の SXGA+規格のディスプレイを用いて、その左側に「平安京ビュー」を表示する正方形領域を確保したとする。このとき、右側の時系列データの可視化に利用可能な領域は 350×1050 画素という縦長の領域となる。提案手法では、この領域の縦方向を時間軸とし、横方向で特定時刻における数値分布を表現する。なお我々の実装では、時系列を表現する右側のウィンドウには、数値を記録した時刻

Visualization of hierarchical time series data using Heiankyo View

Haruho Tachibana, Takayuki Itoh

Ochanomizu University

{haruho,itot} @itolab.is.ocha.ac.jp

の数だけボタンがついている。ボタンをクリックすると、その時刻における各数値が、平安京ビューで表示されるようになっている。

なお、幅 350 画素といった細長い領域に、数百、数千といった大量の数値群の分布を表現することは容易ではない。そこで我々は、音声情報に用いられている「量子化テーブル」に似た概念を導入して、大量の数値群の分布を、少ない画素数の領域に表現する。音声情報の圧縮技術では、量子化テーブルに予め記録された量子化係数を乗じることで、人間の耳につきやすい重要な周波数領域では大きく量子化し、それ以外の周波数領域では小さく量子化する。この技術によって音声情報は、重要度に依存した情報量で合理的に記録される。それに倣って本手法では、注目すべき数値（例えば定常値を大きく外れている数値や、時間変化量の大きい数値など）には量子化を大きめに、それ以外の数値には量子化を小さめにすることで、重要度に依存した情報量で可視化を実現する。

4. 時系列表示のアルゴリズム

ある特定時刻における n 個の数値の分布を、横 m 画素の領域に表現することを考える。ここで、各数値を区間 $[0,1]$ に正規化した値を相対値 r とする。また、各数値の直前時刻における数値との差分を区間 $[-1,1]$ に正規化した値を差分相対値 d とする。この r および d を用いて、提案手法では以下のように重要度 p を求める。なお、 S_1 および S_2 は正の定数である。

$$p = S_1 |r - 0.5| + S_2 |d| \dots(1)$$

i 番目の数値の重要度を p_i とすると、 p_i の相対的重要度は以下ようになる。

$$\frac{p_i}{\sum_{i=1}^{512} p_i} \dots(2)$$

この相対的重要度を用いて、式(3)によって i 番目の数値のウィンドウ上の幅(画素数)を求める。なお式(3)の関数 f にどのような数式を用いるかについては、現在も検討中である。

$$f \left(\frac{p_i}{\sum_{i=1}^{512} p_i} \right) \times m \dots(3)$$

以上の処理によって各数値のウィンドウ上の幅を算出したら、これを左右に並べるようにして m 画素の領域を埋めることにより、ある特定時刻の数値分布を表現する。我々の実装では、各数値に対応する領域を、相対値 r でソートした順に左から並べる。さらに r から

色相を算出する伝達関数を導入し、これによって各数値に対応する領域を塗りつぶす。

5. 実行結果

図 2 に提案手法の実行結果を示す。楕円で囲んだ部分は、一部の数値が大幅に変化した時刻を表現している。これによって、数値変化を誇張し、時系列情報を監視するユーザに注意を促すような可視化が実現できていると考えられる。

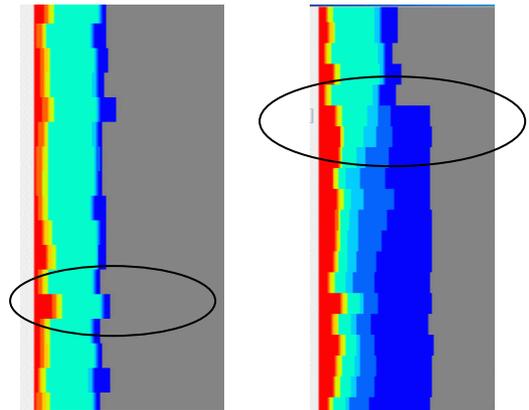


図 2. 実行結果。楕円で囲んだ部分に急激な数値変化が見られる。

6. まとめ

本報告では、階層型データ可視化手法「平安京ビュー」に時系列情報を付加表示する一手法を提案した。今後の課題として、さまざまな分野の階層型時系列データに提案手法を適用し、可視化結果を専門家に提示したり、被験者に評価させることで、提案手法の有効性を実証したい。また式(1)の係数 S_1 および S_2 の調節や、式(3)の関数 f に何を導入するか、といった点も今後の検討課題になるであろう。

参考文献

- [1] Shneiderman B., The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization, IEEE Symposium on Visual Language 96, pp. 336-343, 1996.
- [2]伊藤, 小山田, 平安京ビュー ~ 階層型データを基盤状に配置する視覚化手法, 可視化情報学会第 9 回ビジュアライゼーションカンファレンス, 2003.
- [3] 伊藤, 高倉, 沢田, 小山田, ネットワーク不正侵入監視のための視覚化の一手法, 情報処理学会第 9 回分散システムインターネット運用技術シンポジウム, pp. 63-68, 2004.
- [4] 山口, 伊藤, 池端, 梶永, 階層型データ視覚化手法「データ宝石箱」とウェブサイトの視覚化, 画像電子学会論文誌 Visual Computing 特集号, Vol. 32, No. 4, pp. 407-417, 2003.