

## 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 伊藤貴之

お茶の水女子大学大学院

## A Visualization and Level-of-Detail Control Technique for Large Scale Time Series Data

Yumiko Uchida Takayuki Itoh

Graduate School of Humanities and Sciences, Ochanomizu University

{yumi-ko, itot} @ itolab.is.ocha.ac.jp

### 概要

気象データや株価の変動など、私たちの身の回りには多数の時系列データが存在している。時系列データの解析のために可視化は重要な役割を担うことができる。そこで近年、時系列データの可視化に関する研究が多数報告されている。

本論文では、対話的な詳細度制御によって大規模時系列データの可視化を実現する一手法を提案する。本手法では時系列データを表す折れ線グラフをクラスタリングし、各クラスタの中心に近いグラフのみを表示する。これにより、グラフの表示本数を低減することで表示結果の煩雑さを抑え、かつデータ全体の特徴を逃さない表示を実現する。また本手法は、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選択するための、ユーザインタフェースを併せ持つ。これを用いることでユーザは、選択したグラフと同じクラスタ内にあるグラフを表示させることができる。よって、ユーザが着目した特徴をもつデータについて、対話的に詳細度を再制御できる。

著者らは、日本全国の気温データに提案手法を適用し可視化を試みた。本論文では、詳細度制御とグラフ選択のインタフェースの利用により気象現象を解析した例を紹介する。

### Abstract

We have various interesting time series data in our daily life, such as weather data and stock prices. Visualization is important means to analyze time series data, and there have been many works on visualization of time series data in recent days.

This paper presents a technique for visualization and interactive level-of-detail control of large number of time series data. The technique first generates clusters of time series values, then selects representative values for each cluster, and finally visualizes only representatives. The technique provides a simplified view without missing its interesting features since it reduces number of displaying polygonal lines by using a clustering algorithm. The technique also provides a user interface so that users can interactively select interesting representatives, and explore the time series values which belong to the clusters of the representatives.

We applied the proposed technique to the visualization of air temperature in Japan. This paper introduces that we analyzed weather phenomena by using the proposed level-of-detail control and user interface.

## 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。

時系列データの可視化に関する研究は、既に数多く報告されている。既存の可視化システムの多くは、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分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（折れ線グラフ一本）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、対話的な詳細度制御によって大規模時系列データの可視化を実現する一手法を提案する。提案手法では、時系列データの表示に折れ線グラフを用い、グラフの本数を減らすことで詳細度制御を実現する。詳細度制御の概要を図1に示す。詳細度制御処理では、一定間隔ごとにグラフ形状に基づくクラスタリングをおこない、各クラスタからひとつの要素を代表要素として選択する。クラスタリングによって選ばれた代表要素のみを表示することで、データ全体の特徴を保持した詳細度制御を実現することが可能である。

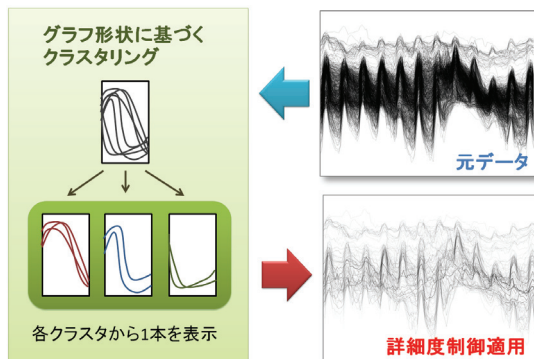


図 1: 詳細度制御の概要。

また提案手法は、詳細度制御後の可視化結果画面から、ユーザが着目した要素をインタラクティブに抽出できるように、要素選択のユーザインタフェースを提供する。このユーザインタフェースでは、ユーザは着目する要素の形状を画面上でなぞる（スケッチする）ことにより、目的の要素を抽出することができる。さらにユーザは、選択した要素と同じクラスタ内にある要素を表示させることができる。

提案手法を適用することで、ユーザは大量のグラフの中から特徴的な動きを発見することが可能である。さらに、要素選択のユーザインタフェースを利用することで、着目した部分のみ元の詳細度で見るといったことが可能となる。

さらに本論文では、提案手法である時系列データ可視化手法と、階層型データ可視化手法を連動するシステムの構築例、およびそれによる日本全国の気温データの可視化への適用事例を示す。著者らの実装では、観測所の属する地域・県ごとに階層を作成し、特定時刻における地域ごとの気温を階層型データとして可視化している。本論文では、本システムを用いて発見した要素の特徴的な変動を、階層型データ可視化手法を用いて分析した事例を示す。

なお、提案手法の基本的概念は、すでに著者自身による口頭講演にて発表されている [1, 2]。本論文はその実装等を詳細化するとともに、気温データへの適用事例により得られた興味深い結果を示すものである。

## 2 関連研究

### 2.1 時系列データの折れ線グラフによる可視化手法

時系列データの可視化では、横軸に時間軸、縦軸に値を取り、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おり、データの解析などの目的にも用いられている [3, 4]。近年は、大規模なグラフ群から目的の数値群を抽出するための研究がいくつかおこなわれている。一例として、Wattenberg らによって、形状に基づくグラフ検索 [5] が報告されている。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折れ線を複数提示している。また、Shneiderman ら [6, 7] は、時系列データを表す大量の折れ線グラフ群から、折れ線の傾きの角度・折れ線の通過範囲等に基づくグラフ検索を提案している。

Wijk ら [8] は、時系列データのグラフ表示とクラスタリングを利用した手法を提案している。この手法では、1日の各時刻における社内にいる従業員数の変化を日ごとに xy 平面でグラフ化し、z 軸に日付を割り当てた 3 次元表示をおこなっている。さらに、階層型クラスタリングを用いて従業員数の変化が似ている日ごとにクラスタを生成し、カレンダー型表示をクラスタに応じた色で塗り分けることで、時期や曜日による特徴を表現している。時系列データをクラスタリングする発想は本手法と同じであるが、ここでは一日のデータを一つの要素とみなしているのに対し、本論文では、ある時刻において複数の要素が存在するデータを扱っているという点が異なっている。

### 2.2 グラフ表示以外の時系列データ可視化手法

折れ線グラフ形式以外にも、時系列データの可視化の研究は数多くおこなわれている。

一例としては、一般にデータの分布状況を視覚的に認識するために用いられるヒストグラムにおいて、さ

らに時間軸を加えた3次元でのヒストグラム表示手法[9]などが挙げられる。

またWeberら[10]は、螺旋上に時間軸をとった表示手法を提案している。円周の長さをデータの周期に割り当てることで、似た特徴をもつ時刻が螺旋の中心を通る直線上に並ぶため、周期性のあるデータの表示に適している。

ThemeRiver[11]では、各要素の値の大きさを垂直方向の幅で表現し、複数の要素の時系列変化を積み重ねた表示をおこなっている。この手法は、値の大きさが塗り分けの幅に対応しているため、どの要素が大きな変化をとっているかを一目で知ることができる。しかし要素数が多い場合には各要素の幅が小さくなってしまったため、見分けるのが困難になるという問題点がある。

4章および5章で用いる日本全国の気温観測データの可視化については、Two-Tone Pseudo coloring[12]を用いた事例が報告されている。この手法は大規模な1次元データを、値に応じて色を塗り分けた横長の長方形領域で可視化する手法である。ひとつの値に対して、2つの色とそれぞれの色が占める領域の比率が割り当てられることが特徴である。この手法は時系列データの可視化において、ひとつの要素の時間変化の全体と細かな特徴の両方を観察する目的に適している。

## 2.3 時系列データ以外のグラフ表示可視化手法

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[13]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ら[14]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[15]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。

またFuaら[16]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可視化手法を提案している。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。

## 2.4 平安京ビュー：大規模階層型データ可視化の一手法

4章で提案するシステムでは、本論文が提案する時系列データ可視化手法と、大規模階層型データ可視化

手法「平安京ビュー」[17, 18]を連動している。そこで以下では、「平安京ビュー」について説明する。

図2は、「平安京ビュー」による階層型データの可視化結果の例である。この可視化結果において、階層型データ中の葉ノードは黒いアイコンで、枝ノードは長方形の枠で表示されている。「平安京ビュー」は、その可視化結果における葉ノードの格子状の配列が、まるで平安京の地図のように整然としていることから命名された手法である。

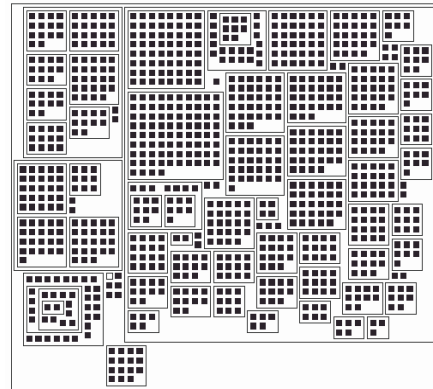


図 2: 「平安京ビュー」による階層型データの可視化の例。

図2からもわかるように「平安京ビュー」は、階層構造を二次元の長方形群の入れ子構造で表現し、その全体を一画面に表示することを目標とした手法である。この手法は、階層型データ中の葉ノードと枝ノードの親子関係よりも、階層型データ全体に分布する葉ノード群を全て一画面に表現することに主眼をおいた視覚化手法である。「平安京ビュー」で技術的に重要な点は、枝ノード群を表現する任意の大きさ・形状の長方形群を、限られた大きさの画面空間に有効に配置できるという点である。言い換えれば「平安京ビュー」は、計算機上の限られた大きさのウィンドウやディスプレイに、できるだけ大規模な情報を詰め込んで表現する技術、ということができる。

階層型データの最下位に属するデータ要素を限られた画面空間に一覧表示する手法としては、平安京ビューの他に Treemaps[19], Quantum Treemap[20]などが挙げられる。

これらの手法は、画面空間を再帰分割することにより階層構造を表現する手法である。最下位階層に属するデータ要素を表現する図形を、できるだけ隙間無く、しかも重なることなく画面に配置する」という点で、平安京ビューと類似していると言える。

しかし平安京ビューは文献[18]にて、長方形領域の縦横比、類似データ間における画面配置結果の類似度、などの数値評価結果において Quantum Treemapよりも良好な結果が得られたことが実証されている。そのため、各要素(アイコン)を一覧表示しながら、要素の持つ値を確認するという目的により適していると考えられる。

「平安京ビュー」では、以下の条件をできるだけ

満たすように、大量の情報を限られた画面空間に配置する。

- 葉ノードや枝ノードが画面上で互いに重ならないように配置する。
- データ全体の画面占有面積ができるだけ小さくなるように配置する。
- 全ての葉ノードが同じ大きさで表示されるように配置する。

なお、これらの条件を満たすような画面配置の処理手順については、文献 [17, 18] に詳しく解説されているので、参照していただきたい。

### 3 時系列データの詳細度制御と要素選択のための GUI

#### 3.1 概要

本論文で扱う時系列データを、以下の通り定義する。

$$V = (v_1, \dots, v_n)$$

ここで、 $n$  は要素の総数であり、 $v_i$  は

$$v_i = (v_{i1}, \dots, v_{im})$$

で定義される。ここで  $v_{ij}$  は、 $i$  番目の要素の  $j$  番目の時刻における値を示している。提案手法では、水平方向に 1 番目から  $m$  番目までの時刻をとり、垂直方向に値の大きさをとった線分で要素  $v_i$  を表現する。

本章で提案する時系列データ可視化手法のフローチャートを、図 3 に示す。提案手法では、一定間隔ごとに要素のグループ化をおこない、各グループ内にクラスタリングを適用する。ここで得られた各クラスタから、それぞれ 1 つの要素を「代表要素」として登録する。この代表要素のみを表示することで、詳細度制御を実現している。

詳細度制御適用後に、ユーザは要素選択機能を利用することが出来る。ユーザは可視化画面上で着目した要素の動きをなぞることで、システムに対しグラフ形状を入力する。システムは代表要素の中から、入力されたグラフ形状に近い形状を持つ要素を検索する。検索結果は「選択要素」として登録される。またユーザが選択部分をより詳細に確認したい場合には、選択要素と同じクラスタに属する要素を追加登録することが可能である。選択要素はハイライト表示され、ユーザは着目したグラフの詳細を得ることが出来る。

#### 3.2 詳細度制御

##### 3.2.1 代表点の生成

提案手法では、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 4(左) 参照)。さ

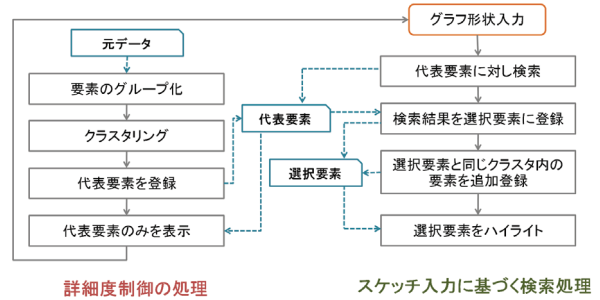


図 3: 時系列データ可視化手法のフローチャート。

らに、 $y$  軸方向を等間隔に区切って量子化をおこなう (図 4(右) 参照)。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  $(t+1)$  番目の標本時刻間隔ごとに、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。

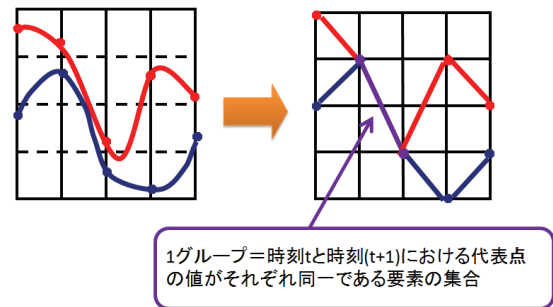


図 4: 標本化・量子化による代表点の生成。(左) 標本化。(右) 量子化。

##### 3.2.2 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。提案手法では、 $t$  番目と  $(t+1)$  番目の標本時刻間に  $n$  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う。 $n$  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。よって図 5 に示すように、このクラスタリングによってグラフの形状に基づくクラスタを生成できる。

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、

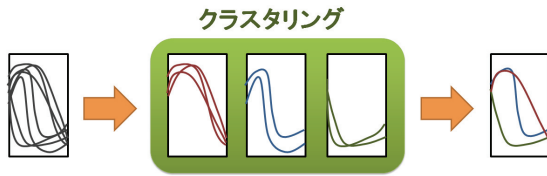


図 5: グラフの形状に基づくクラスタリング。

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

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。

このアルゴリズムを以下に示す (図 6 参照)。

1. データから  $k$  個無作為に要素を選択し、セントロイドとする。
2. 全ての要素を、 $k$  個あるセントロイドとの距離を計算し、最も近いセントロイドを含むクラスタに属させる。
3.  $k$  個のクラスタのセントロイドをそれぞれ求める。
4. 全要素と  $k$  個のセントロイドとの距離を計算し、最も距離の近いセントロイドを含むクラスタに属させる。

この 3, 4 を反復することにより、局所最適解に近づくことが出来る。本論文での適用事例に用いたデータにおいて、経験的に 10 回以内の反復処理によって解に収束することが多いことから、著者らの実装では本処理の反復回数を最大 10 回としている。

以上の処理によって生成された各クラスタについて、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。

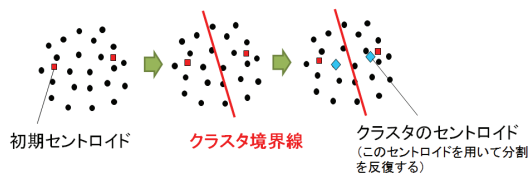


図 6: k-means 法によるクラスタリング。

### 3.2.3 表示

続いて提案手法は、3.2.2 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

で、代表要素の数は格子の大きさに依存する<sup>1</sup>。提案手法では格子の幅を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 7 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。本手法はクラスタ内の要素数に比例して、低い透明度で折れ線を描画する。この処理により、詳細度制御の適用後にも表示されている各々の折れ線が、何個の要素を代表しているか、という情報を提示出来る。

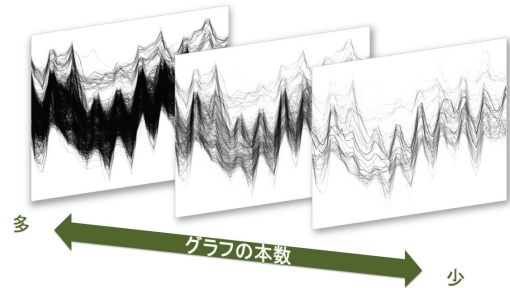


図 7: 格子の大きさによる可視化結果の違い。(左) 元データ。(中央) 小さな格子を用いた詳細度制御。(右) 大きな格子を用いた詳細度制御。

## 3.3 GUI

### 3.3.1 スケッチ入力

提案手法では、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 8 参照)。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。

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 8 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 9 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。

提案手法は、ハイライト色としてユーザに任意の色を選択させる機能を有する。また論理和演算において、1 回目と 2 回目のスケッチ入力に異なるハイライ

<sup>1</sup> 標本時刻の総数 (格子において鉛直方向に伸びる線分の総数) を  $N_t$ 、量子化値の総数 (格子において水平方向に伸びる線分の総数) を  $N_v$ 、k-means 法におけるクラスタ数を  $k$  とすると、代表要素は最大で  $k(N_t - 1)N_v^2$  となる。

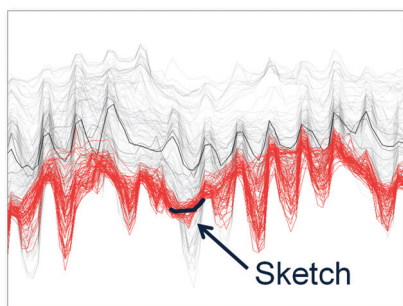


図 8: スケッチにより選択された要素のハイライト.

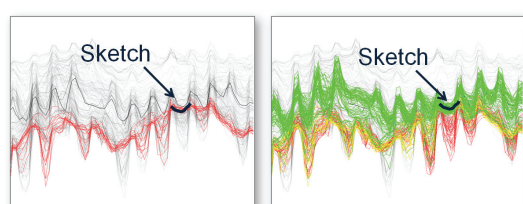


図 9: (左) 論理積演算をおこなった結果. (右) 論理和演算をおこなった結果.

ト色を選択した場合、1 回目の選択結果と 2 回目の選択結果の論理積となる要素は、2 つのハイライト色の中間色で表示される。

また入力グラフと異なる高さにあるが、形状が類似している要素をハイライトすることも可能である (図 10 参照)。ここでは入力グラフと代表要素の比較に、y 座標値の代わりに頂点間の傾きを用いている。

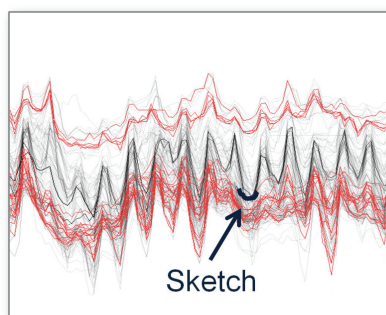


図 10: スケッチした位置と異なる高さに描かれた要素のハイライト.

### 3.3.2 詳細度制御前の要素を復活表示

また提案手法では、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 (図 11 参照)。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、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。また入力グラフとの比較は代表要素に対してのみおこなっているため、着目する部分の要素を効率よく抽出することが出来る。

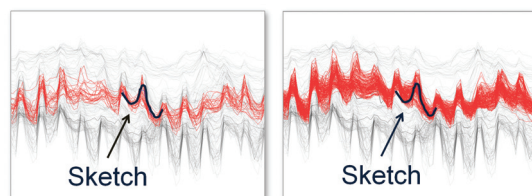


図 11: 詳細度制御前の要素の復活表示. (左) 選択された代表要素のみをハイライト. (右) 選択要素と同じクラスに属する要素も同時にハイライト.

## 4 時系列データ可視化手法と他の可視化手法との連動

身の回りにある時系列データの多くは、各々の要素が位置情報や階層構造を有している。これらを効果的に可視化するには、時系列データ可視化手法だけでなく、位置情報可視化手法や階層型データ可視化手法を連動させて、複合的な可視化システムを構築することが有効である。本論文では、位置情報可視化手法や階層型データ可視化手法と提案手法を連動させたシステムを構築し、日本全国の気温変化データを題材にして可視化を行った事例を報告する。

### 4.1 他の可視化手法との連動による気温変化データの可視化

著者らは日本全国の気温変化データに対し、提案手法を適用した。このデータは AMeDAS (Automated Meteorological Data Acquisition System) [21] によって観測されたものであり、全国約 1300 ヶ所の観測所で観測された気温・降水量・日照時間・積雪量・風向・風速が記録されている。著者らはこの中から、気温を観測している 913 地点の観測所についてのデータを作成した。本システムでは、各観測所を一つの要素として扱い、各観測所で観測された気温を 2 時間おきに記録したデータを用いる。

また解析をおこなうために他の可視化システムと連動して、時系列変化以外の情報に関して考察をおこなう。

図 12 は、AMeDAS の各観測所を示すアイコンを、日本地図上に配置した図である。ここではアイコンの色は各観測所で観測された気温を示し、アイコンの配置は各アイコンの示す観測所の所在位置に対応するものとなっている。このような表示手法は AMeDAS の

観測値を表示する際に、一般的に広く用いられているものである。しかし、この表示手法では一つ一つのアイコンが小さくなってしまいう上、アイコン同士の重なりも多く、全体を表示したまま各観測所の値を確認するのは困難である。

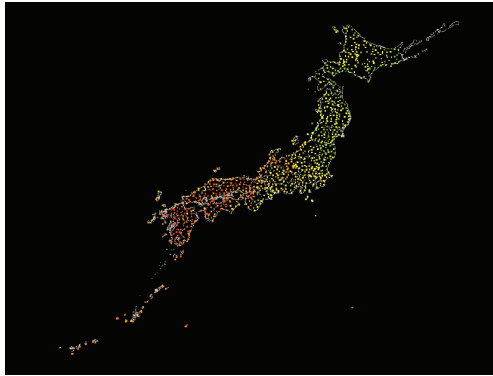


図 12: AMeDAS 観測所を日本地図上に配置した図。

この問題点を解決するため著者らは、階層型データ可視化手法「平安京ビュー」[17, 18]を用いて観測所データを可視化することを試みた。観測所データを階層型データとして取り扱うにあたり著者らは、図 13 に示すように、観測所を葉ノード、観測所の所在している地域を 2 段階の階層で表現した。

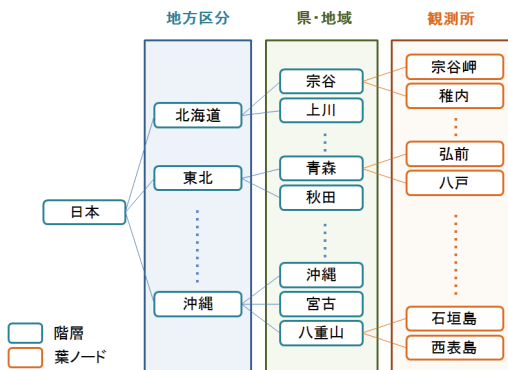


図 13: AMeDAS 観測所の階層構造化。

法においてユーザにより選択された折れ線グラフ群」に対応する「平安京ビュー上のアイコン群」をハイライト表示することで、地域的な気温変化の特性をより視覚的に把握しやすいものとした。

なお平安京ビューを用いた時系列情報表示の研究としては、文献 [22] がすでに発表されている。この文献が提案する手法では、平安京ビューの横に新たなウィンドウを設け、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色の割合で表現している。この手法は、突発的な数値変化を誇張するように表示することから、時系列情報の監視等に向いている手法であるといえる。また縦軸上に設置された時刻を示すボタンをクリックすることで、特定時刻における数値分布を平安京ビュー側で表示することが可能となっている。それに対して本システムでは、特定時刻における数値分布を平安京ビューで表示するのみでなく、特定要素のみのハイライト機能を設けている。このことから本システムは、突発的な数値変化に限らず任意の数値変化の分析に適した手法であると言える。

また文献 [23] では、平安京ビュー上で選択した要素 (アイコン) の時系列情報をグラフ表示する手法も提案されている。本システムは時系列変化をもとに抽出した要素をハイライト表示するものであり、データ探索操作の出発点をどこに置かうかという点で大きく異なるものである。

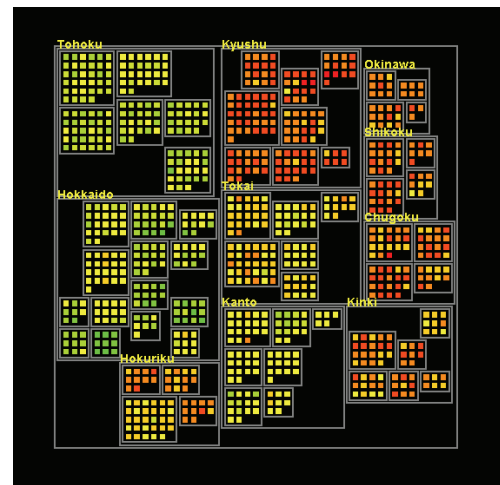


図 14: 「平安京ビュー」による AMeDAS 観測所の可視化。

## 4.2 時系列データ可視化手法と「平安京ビュー」との連動

図 14 は平安京ビューを用いて AMeDAS の観測所データを可視化した例である。アイコンが一つの観測所を示し、枠は観測所の所在している県・および地域、アイコンの色は観測された気温を示している。

一方、本論文で提案する時系列データ可視化手法では、折れ線グラフ一本が一つの観測所を表わし、各観測所の時系列変化を表示している。著者らは、「提案手

## 5 実行例

著者らは Java Development Kit 1.6 を用いて提案手法を実装し、Windows Vista (CPU 2.1GHz, RAM 2.0GB) 上で実行した。以下、提案手法である時系列データ可視化手法の処理時間と、AMeDAS データ可視化への適用事例について論じる。

## 5.1 時系列データ可視化手法の処理時間

我々は提案手法である時系列データ可視化手法の計算時間について、以下の通り解析および測定を行った。この解析に先立ち、本節では以下の通り変数を定義する。

- $n$  : 要素数
- $a$  : 標本化点の数
- $s$  : データ中のタイムステップの合計
- $r$  : 詳細度制御で得られた代表要素数
- $q$  : スケッチ入力したグラフの頂点数

上記のように定義した場合、3.2 節に示した詳細度制御にかかる計算時間は、標本化・量子化に  $O(na)$ 、クラスタリングに  $O(ns)$  であり、3.3 節に示したスケッチ入力に基づくグラフ選択に掛かる計算時間は  $O(rq)$  である。

上記の計算機環境にて、詳細度制御処理の計算時間を計測したところ、図 16(左) に示したデータ ( $n=913$ ,  $a=15$   $s=372$ ) において 250 ミリ秒。スケッチ入力に基づくグラフ選択の計算時間を計測したところ、図 17(左) に示した入力 ( $r=419$ ,  $q=30$ ) において 14 ミリ秒を要した。

このことから、提案手法による可視化処理が十分に高速であることがわかる。

## 5.2 AMeDAS データ可視化への適用事例

図 16(左) は 2006 年 3 月中旬における AMeDAS 気温データであり、図 16(右) は提案手法を用い詳細度制御を適用した可視化結果である。詳細度制御を適用したことにより、赤い枠で囲った部分 (3 月 11 日・12 日) において、値が上下している要素と、値があまり上下せず下がっている要素が存在していることが確認できる。用いたデータは 2 時間ごとの観測値を記録しているので、規則的な観測値の上下は日中と夜間の温度差を示している。

よって赤い枠で囲った部分において、値があまり上下していない要素が示す観測所では、この期間において日中も気温が上昇しなかったことを示している。このように日中も気温が上昇しない地域では、その期間において雨や雪が降るなど天候が悪かったということが考えられる。

ここで図 16 のデータに対し 3.3 節で示したスケッチ入力機能、詳細度制御前の要素を復活表示させる機能を用いた要素選択をおこなった。ここでは値があまり上下していない要素のうち、値が最も低く推移している要素群を抽出した (図 17(左) 参照)。またこれらの要素が示す観測所を、平安京ビューおよび日本地図上でハイライトした結果が図 17(中央)(右) である。ハイライトに用いた色は、図 15 のカラーマップにより算出している。この可視化結果から、これらの要素

は北海道のオホーツク海側を示しているということが分かる。

また同様に、値があまり上下していない要素のうち、値が中央付近で推移している要素群および最も高く推移している要素群をそれぞれ抽出した (図 18(左)、図 19(左) 参照)。これらの要素が示す観測所を平安京ビューおよび日本地図上でハイライトした結果がそれぞれ図 18(中央)(右)、図 19(中央)(右) である。この可視化結果から、値が中央付近で推移している要素群は東北・北陸地方日本海側、最も高く推移している要素群は北陸・中国・九州地方日本海側を示しているということが分かる。

図 17 から図 19 で抽出した地域では、日中も気温が上昇していないため、雨や雪が降っていた可能性が高いと考えられる。

またこれらの観測所において、特定の気温 (例: 5℃) を観測した時間をみると、北海道が最も早く、北陸・中国・九州地方が最も遅い (図 20 参照)。この結果は、寒冷前線が北から通過していった様子を示していると考えられる。

著者らは、この考察結果の検証をおこなうために、気象記録を公開したウェブサイト「気象人」[24] を参照しながら、2006 年 3 月中旬の気象情報の調査をおこなった。その結果、11 日には低気圧がオホーツク海にかかり、寒冷前線が夜遅くには北陸から東北日本海側にかけていたことが確認できた。また 12 日には -6℃ラインが 9 時に東北北部に、21 時には北陸から中国地方・九州北部にまで達していたことも分かった。さらに東北日本海側や北陸では、11 日から 12 日にかけて雨や雪が降っていたことも確認できた。以上のことから、可視化結果に基づく著者らの考察が正しかったことを実証できた。

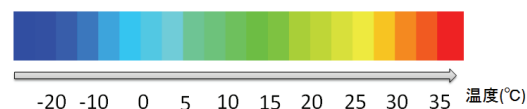


図 15: 観測値に対応するカラーマップ。

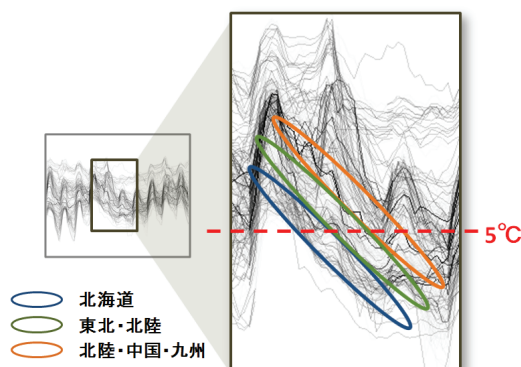


図 20: 地域による特定の気温を観測した時間の違い。

続いて、前線の影響が少なかった地域を調べるため

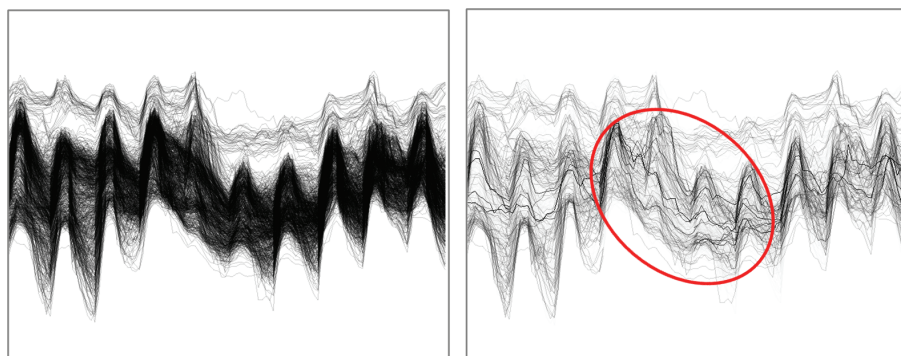


図 16: AMeDAS 気温データに詳細度制御を適用. (左) 適用前. (右) 適用後.

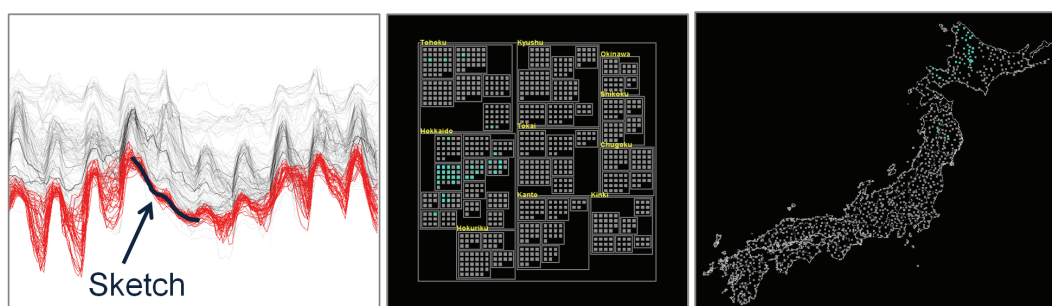


図 17: 気温があまり上下していない要素のうち、気温が低い要素を抽出. (左) スケッチ入力に基づき要素をハイライト. (中央) 平安京ビューで対応する観測所をハイライト. (右) 日本地図上で対応する観測所をハイライト.



図 18: 気温があまり上下していない要素のうち、気温が中央付近の要素を抽出. (左) スケッチ入力に基づき要素をハイライト. (中央) 平安京ビューで対応する観測所をハイライト. (右) 日本地図上で対応する観測所をハイライト.



図 19: 気温があまり上下していない要素のうち、気温が高い要素を抽出. (左) スケッチ入力に基づき要素をハイライト. (中央) 平安京ビューで対応する観測所をハイライト. (右) 日本地図上で対応する観測所をハイライト.



図 21: (左) スケッチ入力に基づき要素をハイライト. (中央) 平安京ビューで対応する観測所をハイライト. (右) 日本地図上で対応する観測所をハイライト.

に、新たな要素選択をおこなった。ここで、前線の影響が少なかった地域では、日中は気温が上がり、夜になると下がるため、観測値が上下すると考えられる。そのため、その観測値の上下を表す「上に凸」の入力グラフをスケッチした。

ここで図 16 のデータに対し、3.3 節で示したスケッチ入力機能を用いて「上に凸」の入力グラフをスケッチし、詳細度制御前の要素を復活表示させる機能を用いて要素選択をおこなった(図 21(左) 参照)。ここでは図 10 と同様に、異なる高さにある要素も同時にハイライトさせている。これらの要素が示す観測所を、平安京ビューおよび日本地図上でハイライトした結果が図 21(中央)(右) である。

この可視化結果から、太平洋側では日中は気温が上っており、雨や雪はほとんど降らなかったと考えられる。これは日本海を通過してきた寒気が日本海側で雪を降らせ、山を越えた太平洋側まで雪雲が流れてこないためであると考えられる。

しかし、これらの地域でも図 22 の青い枠で囲った部分においては、観測値が大幅に下がっており、夜にかけて急激に気温が下がったことが分かる。

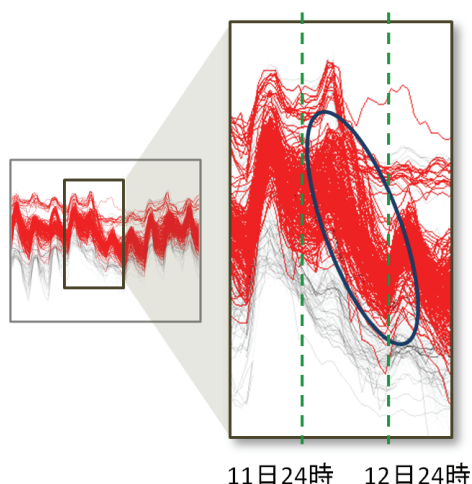


図 22: 観測値が大幅に下がった時間帯。

11 日 24 時と 12 日 24 時の全国の観測所の気温を

平安京ビューで可視化した図が図 23 である。図 15 と見比べればわかるように、12 日は前日の同じ時刻に比べて、10℃近く気温が下がっている地域が多い。この可視化結果から、12 日は全国的に寒気の影響が強くなり、日中は気温の上がった地域でも夜には冷え込んだことが視覚的に確認できた。

## 6 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、およびスケッチ入力に基づく要素選択を提案した。また AMeDAS データに提案手法を適用し、さらに平安京ビューおよび日本地図上で観測所を一覧可視化することで、その結果について考察した。

AMeDAS データの可視化における今後の課題として、平安京ビューで観測所の地理上の特徴を反映した表示をするという点が挙げられる。現在の平安京ビューの表示では、地域を示す灰色の枠の画面上での位置が、地理上での地域の位置と無関係なものとなっている。この問題を解決するために、階層型データを構成する各ノードに対して、理想的な画面位置を記録した理想座標値をあらかじめ用意する手法を取り入れることを検討している。平安京ビューは、理想座標値を反映しつつ、画面空間を節約した配置をおこなうことが出来る。この理想座標値を記録したデータは「テンプレート」と呼ばれ、テンプレートを用いた階層の配置手法 [25] がすでに発表されている。これを活用し、各地域の実際の地理上の位置を理想の座標値として記録しておくことで、各階層を地理上の位置と近い位置に配置することが可能になる。

また提案手法全般における今後の課題としては、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。

本論文では、AMeDAS データについて提案手法を適用した例を示したが、その他に気象シミュレーションデータ、新聞におけるキーワードの登場頻度の変化や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。

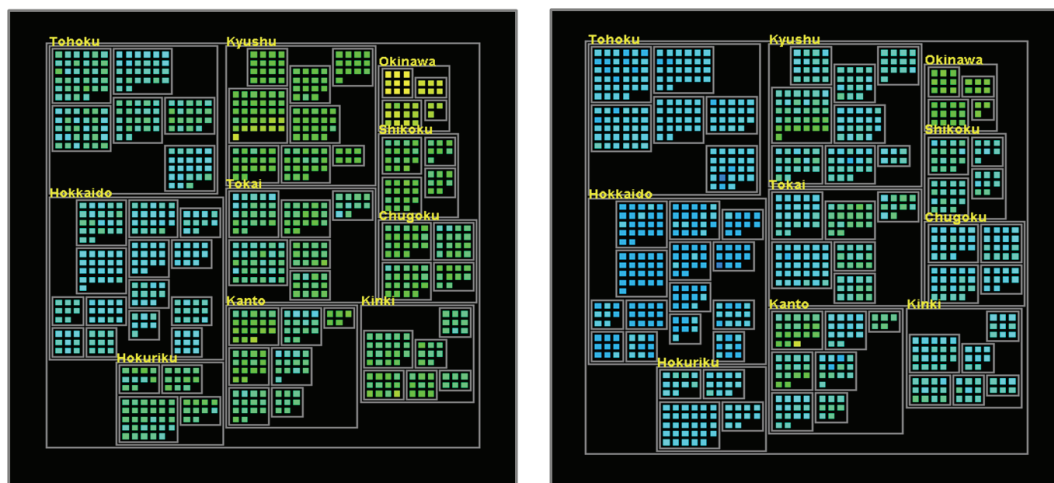


図 23: (左)2006 年 3 月 11 日 24 時における観測値. (右)2006 年 3 月 12 日 24 時における観測値.

## 参考文献

- [1] Y. Uchida, T. Itoh, A Visualization and Level-of-Detail Control Technique for Large Scale Time Series Data, *IEEE Information Visualization, Poster Session*, 2008.
- [2] T. Itoh, Y. Uchida, FRUITS Time: An Interactive Visualization Technique for Time-Varying Data, *Fifth International Conference on Flow Dynamics*, 2008.
- [3] J. Lin, E. Keogh, S. Lonardi, J. Lankford, D. Nystrom, Visually mining and monitoring massive time series, *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 460–469, 2004.
- [4] J. Lin, E. Keogh, S. Lonardi, P. Patel, Finding Motifs in Time Series, *Proceedings of the 2nd Workshop on Temporal Data Mining*, pp. 53–68, 2002.
- [5] 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.
- [6] H. Hochheiser, B. Shneiderman, Dynamic query tools for time series data sets: Time-box widgets for interactive exploration, *Information Visualization*, Vol. 3, No. 1, pp. 1–18, 2004.
- [7] H. Hochheiser, B. Shneiderman, Interactive Exploration of Time-Series Data, *Proceedings of the 4th International Conference on Discovery Science*, pp. 441–446, 2001.
- [8] J. Wijk, E. Selow, Cluster and Calendar Based Visualization of Time Series Data, *Proceedings of the IEEE Symposium on Information Visualization 1999*, pp. 4–9, 1999.
- [9] R. Kosara, F. Bendix, H. Hauser, Timehistograms for Large, Time-Dependent Data, *Proceedings of the 2004 Eurographics/IEEE TVCG Symposium on Visualization*, pp. 45–54, 2004.
- [10] M. Webera, M. Alexa, W. Muller, Visualizing Time-Series on Spirals, *IEEE Symposium on Information Visualization 2001*, pp. 7–14, 2001.
- [11] S. Havre, B. Hetzler, L. Nowell, ThemeRiver: visualizing theme changes over time, *Proceedings of the IEEE Symposium on Information Visualization 2000*, pp. 115–123, 2000.
- [12] T. Saito, H. Miyamura, M. Yamamoto, H. Saito, Y. Hoshiya, T. Kaseda, Two-Tone Pseudo Coloring: Compact Visualization for One-Dimensional Data, *Proceedings of the IEEE Symposium on Information Visualization 2005*, pp. 173–180, 2005.
- [13] D. Holten, Hierarchical Edge Bundles, *IEEE Transaction on Visualization and Computer Graphics*, Vol. 12, No. 5, pp. 741–748, 2006.
- [14] 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.
- [15] H. Hauser, F. Ledermann, H. Doleisch, Angular Brushing of Extended Parallel Coordinates, *Proceedings of the IEEE Symposium on Information Visualization 2002*, 2002.
- [16] Y. Fua, M. Ward, E. Rundensteiner, Hierarchical Parallel Coordinates for Exploration of Large Datasets, *Proceedings of Visualization '99*, pp. 43–50, 1999.

- [17] T. Itoh, H. Takakura, A. Sawada, K. Koyamada, Hierarchical Visualization of Network Intrusion Detection Data in the IP Address Space, *IEEE Computer Graphics and Applications*, Vol. 26, No. 2, pp. 40–47, 2006.
- [18] 伊藤, 山口, 小山田, 長方形の入れ子構造による階層型データ視覚化手法の計算時間および画面占有面積の改善, 可視化情報学会論文集, Vol. 26, No. 6, pp. 51–61, 2006.
- [19] B. Johnson, B. Shneiderman, Tree-Maps: A Space Filling Approach to the Visualization of Hierarchical Information Space, *Proceedings of the 2nd International IEEE Visualization Conference*, pp. 275–282, 1991.
- [20] B. Bederson, B. Shneiderman, Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies, *ACM Transactions on Graphics*, Vol. 21, No. 4, pp. 833–854, 2002.
- [21] 気象業務支援センター, アメダス 10 分値データ 2006 年版.
- [22] T. Itoh, S. Furuya, H. Ohshima, K. Okamoto, Hierarchical Data Visualization for Atomic Plant Data, *Journal of Fluid Science and Technology*, Vol. 3, No. 4, pp. 553–562, 2008.
- [23] M. Iwamaru, M. Asakura, S. Someya, T. Itoh, K. Okamoto, Collaborative Visualization System Using HeiankyoView, *IEEEJ Image Electronics and Visual Computing Workshop*, 2007.
- [24] 気象人, <http://www.weathermap.co.jp/kishojin/>.
- [25] 山口, 伊藤, 長方形の入れ子構造を用いた階層型データ視覚化手法の拡張, 情報処理学会論文誌, Vol. 44, No. 10, pp. 2469–2477, 2003.

内田 悠美子



2007 年お茶の水女子大学理学部情報科学科卒業. 2009 年お茶の水女子大学大学院人間文化創成科学研究科理学専攻博士前期課程修了. 情報処理学会会員.

伊藤 貴之



1990 年早稲田大学理工学部電子通信学科卒業. 1992 年早稲田大学大学院理工学研究科電気工学専攻修士課程修了. 同年日本アイ・ビー・エム (株) 入社. 1997 年博士 (工学). 2000 年米国カーネギーメロン大学客員研究員. 2003 年から 2005 年まで京都大学大学院情報学研究科 COE 研究員 (客員助教授相当). 2005 年日本アイ・ビー・エム (株) 退職, 2005 年よりお茶の水女子大学理学部情報科学科助教授. ACM, IEEE Computer Society, 情報処理学会, 芸術科学会, 画像電子学会, 可視化情報学会, 他会員.