

## Storyline を適用した実数値型時系列データ可視化の一手法

八木佐也香<sup>1)</sup>(学生会員) 伊藤貴之<sup>1)</sup>(正会員) 高塚正浩<sup>2)</sup>

1) お茶の水女子大学大学院人間文化創成科学研究科 2) シドニー大学情報理工学部

## A Storyline-based Visualization Technique for Consecutive Numerical Time-varying Data

Sayaka Yagi<sup>1)</sup> Takayuki Itoh<sup>1)</sup> Masahiro Takatsuka<sup>2)</sup>

1) The Graduate School of Humanities and Sciences, Ochanomizu University

2) Faculty of Engineering & Information Technologies, The University of Sydney

{sayaka, itot} @ itolab.is.ocha.ac.jp, masa.takatsuka @ sydney.edu.au

### 概要

我々の身の回りには数多くの時系列データが存在しており、その分析において、可視化は有効である。代表的な時系列データの可視化手法として、折れ線グラフとヒートマップがあるが、いずれの手法もデータの部分的な特徴と全体のクラスタの変遷の両方を同時に観察するのは難しい。本論文では実数値型の時系列データを対象に、要素間の交流や共起性の表現に用いられる storyline による可視化手法を提案する。本手法ではまず、実数値の類似度にもとづいて局所時間帯ごとに要素をクラスタリングする。続いて隣接する時間帯間にてクラスタ群のマッチングを適用し、その結果に沿ってクラスタの配置を決定する。この配置結果にしたがって各要素を描くことで、類似した要素同士が画面内で近接するような storyline を描く。また、各要素の実数値変化の重要度にもとづいて要素の不透明度を設定することで、重要度の高い部分を強調し、ユーザの知覚効果の高い可視化結果を得る。さらに本手法では、storyline 上で興味深い局所を対話的に指定することで、選択した要素群の実数値変化を別の画面領域に表示するというインタラクション機能を持つ。本手法を用いてクラスタの時間的変遷を観察することにより、長時間にわたって類似した値/外れ値を有する要素だけでなく、所属するクラスタが頻繁に切り替わるような要素に着目しやすくなる。また、インタラクション機能を用いて元の数値情報を参照することで、クラスタの切り替えの要因を分析することが可能である。

### Abstract

Information visualization is an effective approach to analyze time-varying data in our daily lives. We commonly represent time-varying values applying polyline charts or heatmaps; however, it is difficult to simultaneously observe short-term features of time-varying values and cluster transitions while applying either polyline charts or heatmaps. This paper proposes a storyline-based visualization technique for consecutive numerical time-varying data. Storyline is a visualization technique to show associative features among elements over time. Our technique first measures similarity of elements in each time-step, and divides the elements into clusters. The technique then defines the cluster layout by matching corresponding clusters between two adjacent time-steps, and draws similar elements as proximity storyline. Reflecting transparency on storyline as a visual variable, the technique also emphasizes the amount of line changes. Moreover, the technique provides a user interface so that users can interactively select interesting parts on storyline, and explore the numerical values by observing a polyline-based visualization. We believe it is important to focus on elements which switch their clusters frequently. We suppose that by making the appearances of numerical changes prominent based on the amount of changes, a user would be able to effortlessly pay his/her attention to where those changes occurred. This easy recognition of numerical changes would lead to further focused investigation on the causality through examination of the original numerical values and other associated information.

## 1 はじめに

ビッグデータ時代の到来に伴い、多種多様なデータベースに時系列性を有する情報が蓄積されるようになった。これらの時系列データの分析・予測はデータマイニングや知覚情報処理等の分野においては既に数多くの研究が進められている。一方で、時系列データの観察と理解、またそれを扱う業務での能動的な意思決定を支援する道具として、情報可視化が有用である。情報可視化には、目で見て全体像を把握し、対話操作によってデータ中の必要な部分を能動的に選択できる点に加え、特に時系列データにおいてはリアルタイムな観察のための瞬時的な理解においても有効である。一般的に情報可視化手法は、一次元、二次元、三次元、多次元、木構造、グラフ、時系列の7種類のデータ構造に大別される[1]。その中でも我々はこれまで、時系列データの可視化に関する新しい手法の開発に取り組んできた[2]。

多量の実数値で構成される時系列データの可視化では、データを構成する実数値の大小そのものだけでなく、実数値どうしの関連性や変化量などを観察することも重要である。例えば多数の商品の売上を一定時刻ごとに集計して作成される時系列データでは、商品単体の売上の推移だけでなく、どの商品と同時期に売れるか、季節や時間帯との相関はあるか、といった関連性を観察することによって、より深い分析が可能になり、売上予測や予兆発見などに貢献できると考えられる。

時系列データの可視化では折れ線グラフとヒートマップの2種類の視覚表現が特に多用されている[3]。しかしいずれの視覚表現においても、大量のデータ要素（以下「要素」と称する）の相互関係を全て表現するのは難しい。例えば折れ線グラフでは、要素に対応する折れ線を1個の画面空間に大量に描くことで、その相互関係に関する視認性は大きく低下する。そこでクラスタリングなどを適用して要素を構造化した上で可視化することが有効である。ここで時系列データにおいては要素間の相互関係も時間とともに変遷するため、クラスタ構造の時間的変遷を表現することも重要となる。一方でこれらの可視化手法において、データの部分的な特徴とクラスタ構造の時間的変遷を同時に表現するのは簡単ではない。

クラスタ構造の時間的変遷の表現に向けた可視化手法として、Sankey Diagram と storyline がよく利用され

ている。Sankey Diagram はクラスタ間の流量変化を観察するのに適している。一方、storyline は個々の要素について、クラスタから別のクラスタへの移動を観察するのに向いている。Storyline はもともと XKCD というウェブコミックに手書きイラスト“Movie Narrative Charts”として紹介されたものであり[5]、物語の登場人物の共起行動を時系列のダイアグラムで表現している。図1にあるように、x軸が時刻、各曲線が映画の各登場人物を表しており、複数の登場人物が同一のシーンに登場した場合に線が近接するような可視化結果となる。これらの可視化手法は任意の2要素間の流量や共起に関する時間変化を表現するのに用いられるのが一般的であり、各要素が独立に実数値を有するような時系列データに対してこれらの可視化手法を適用した事例はまだ少なく、議論の余地が残っている。

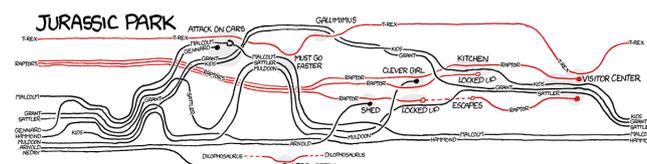


図1 手書きによる storyline の例 ([5] より転載)。

以上を踏まえて、我々は実数値で構成される時系列データを storyline として表現する可視化の一手法を提案する。本手法ではまず、時系列データを構成する各要素に対して局所時間帯ごとにクラスタリングを適用し、隣りあう局所時間帯間のクラスタ群にマッチングを適用する。その結果にもとづいて、クラスタに対応する長方形群を画面配置し、長方形群を通過する線分として各要素を描画する。描画時には次を工夫することで、より深い分析を可能にする。

- 各要素に割り当てられた数値やメタ情報から線分の色相を算出することで、要素のクラスタ移動の要因を理解しやすくする。
- 各要素の数値変化の重要度から線分の不透明度を算出することで、重要な数値変化を有する線分に視線を届きやすくする。

本手法を用いてクラスタの時間的変遷を観察することで、次の3種類の要素群を視覚的に分離して理解するのが容易になる。

- 長時間にわたって同一クラスタに属する要素群
- 外れ値を有するために小さなクラスタに属する要素群
- ある時には特定のクラスタに属し、またある時には別のクラスタに属する、というようなふらふらとした時間的遷移を有する要素群

例えば、夏には商品 A と B の売上の時間的推移が類似して、冬には商品 A と C の売上の時間的推移が類似している、という商品データがあるとすると、図 2(右)のように本手法でこれを可視化することで、商品 A が夏から冬にかけて別のクラスタに移動していることが storyline 上で明確に表現できる。これによって、商品 A に対する店舗での陳列には注意が必要である、というような知識の発見を期待できる。

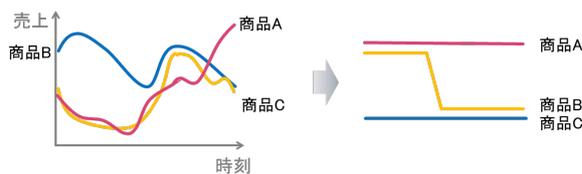


図 2 商品の売上データの可視化例の比較。

## 2 時系列データ可視化に関する関連研究

### 2.1 実数値型時系列データの可視化手法

実数値型時系列データの可視化には旧来から、折れ線グラフにもとづく可視化手法と、ヒートマップにもとづく可視化手法が特に多用されてきた [3]。

折れ線グラフにおいては、要素数の多いデータにおけるクラタリングがその可読性を妨げてきた。その解決策として折れ線のサンプリングが有効であり、有効なサンプリングを実現するために時系列データのクラスタリングがしばしば適用されてきた。我々は折れ線グラフにもとづく可視化手法において、局所時間帯ごとに要素をクラスタリングしてサンプリングすることにより、可読性が高く、かつデータの特徴を見逃さないような可視化を実現した [2]。しかしこのようなアプローチでは、どの要素が局所時間帯ごとにいつクラスタを変遷したか、といった情報を視認することができない。

またヒートマップにもとづく可視化手法において、時間帯全体にわたる類似度で要素を並べ替え、さらに局所時間帯ごとに見られる特徴的な数値パターンをマークス

ることで、データ大域にわたるクラスタ構成と、局所に見られる特徴的な現象を同時に表現する手法を提案した [4]。しかしこのようなアプローチでは、局所時間帯ごとにクラスタリングを適用できない。

### 2.2 Storyline を適用した可視化手法

時系列データを構成する各要素の共通性に関する変遷を可視化する、という問題設定は実数値型以外の時系列データにおいて活発に議論されている。人物や単語を要素とみなしたとき、その 2 要素間の距離（例えば共起性）の時間変化を可視化する、という問題がその最たる例である。この課題においてフレキシブルに各要素間の共通性を描く手法として、XKCD の手描きイラスト [5] に着想を得て storyline のメタファを用いた時系列データ可視化手法が近年数多く提案されている。Ogawa らはソフトウェア開発のリポジトリを題材として storyline のレイアウトを自動生成する手法を提案した [6]。この手法による可視化結果は、小～中規模のプロジェクトによるソフトウェア開発の構造変化の観察に有用である。一方でこの可視化結果には、線分の小刻みな蛇行や線分どうしの重なりが数多く含まれており、時間軸に沿って各線分の動きを追うことは困難である。Tanahashi らは Ogawa らの手法を拡張し、遺伝的アルゴリズムを用いて線分の配置を最適化する手法を提案した [7]。この手法では、線分のずれや重なり、画面上の空白を制御することで、Ogawa らの手法の問題点を解決し、線分間の共起性に着目しやすい結果を実現している。しかしこれらの手法にはスケーラビリティと計算時間に課題がある。数百以上の要素数を有するデータにおいて線分間の重なりが非常に多くなり、各要素を目で追うことが困難になる。この問題を解決するため、StoryFlow は各クラスタに詳細度制御を適用し、Storyline 全体の特徴を捉えやすくした [8]。StoryFlow では各要素の階層構造を考慮した配置に加え、インタラクティブな線分の並べ替え、バンドリングなどの機能も実現している。また、Reda らは storyline のメタファを用いつつも、数百規模の要素の描画が可能な手法を提案している [9]。Reda らの手法では、y 軸上にコミュニティを示すスロットを固定し、各スロット内に線分を配置している。このアプローチはユーザが特定のクラスタ間の変化に着目したい場合、有効であるといえる。しかしこれらの手法は、上述のように既知の関係性やキーワード・トピック等を入力情報と

して要素の共通性を表現しており、実数値で構成される時系列データを対象にしたものではない。また、要素間の距離などにもとづく強調表示などは行われていない。

実数値型の時系列データにおいても、局所時間帯ごとのクラスタリング結果に焦点をおいた可視化手法は最近になっていくつか報告されている。例として Lex ら [10] は、クラスタごと、局所時間帯ごとに断片化されたヒートマップを Sankey Diagram 風に接続する可視化手法を提案している。しかし Sankey Diagram と同様な表現を採用したこの手法では、時系列データを構成する個々の要素の変遷に着目することが難しい。また Turkay ら [11] は、局所時間帯ごとにクラスタを縦に並べ、各クラスタを通過するように要素を曲線で描く可視化手法を提案している。この手法は storyline の簡易な生成方法の一種と考えることもできる。本論文の提案手法における描画形式は Turkay らの描画形式に類似しているが、次節にて議論する点において提案手法は Turkay らの手法を改良しているといえる。

### 2.3 Storyline の適用に関する注意点

本節では、Storyline を実数値型時系列データに適用する際の注意点と、それらを改良するためのアプローチを議論する。

クラスタの位置関係: storyline による描画では、要素を表現する曲線の縦方向の移動を小さくすることが重要である。本手法では隣りあう局所時間帯において多くの要素を共有するクラスタが画面上で隣接するように配置することで、曲線の移動を抑える。上述の Turkay らの手法 [11] においても、要素の共有性を考慮した並び替えが行われている。一方で storyline による描画を見ただけでは、実数値型時系列データの数値自体の変遷を理解することは難しい。そこで、要素の共有性に加えて数値変化も加味してクラスタを配置することが有用である。

重要な要素の強調表示: storyline では多くの場合において、画面上で大きく移動する曲線が目立って見えるため、この曲線に対応する要素が重要な要素（あるいは大きな変化を有する要素）であると連想しがちである。よって実数値の時系列データを storyline で可視化すると、実数値の変化の小さい要素に対応する曲線が画面上で大きく移動しているために目立ってしまう、というような直感に合わない可視化結果を生むことがある。例えば図 3 において、左下部の赤の曲線は短時間で急激な変化を

有しているように見える。しかし多くの既存手法において、曲線の画面上での移動量は、個々のクラスタに含まれる曲線の本数やクラスタの配置などによって決定されるため、必ずしも実数値の変化量に比例するとは限らない。この問題を解決する手段として、曲線の形状以外の視覚変数によって重要な要素を強調表示することが有効である。

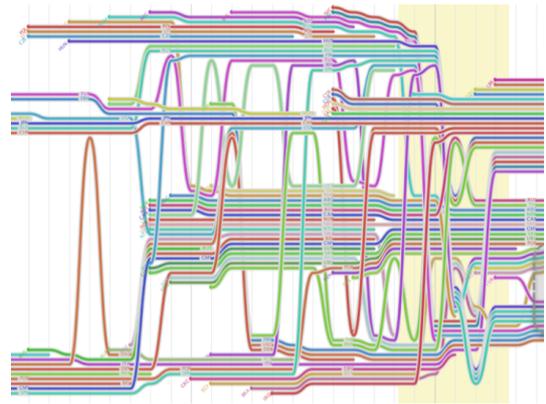


図 3 画面上での変化量が大きく見える storyline の例 (文献 [7] より転載)。

数値表示へのインタラクション: storyline による描画は数値自体の変遷を表現しない。そのため、既存手法ではヒートマップを併用する数値表現 [10] が用いられているが、厳密に数値を読み取ることは難しい。この問題を解決する一手段として、storyline 上で興味深い局所を対話的に指定し、別の画面領域で数値表示することが考えられる。

### 3 提案手法

本章では storyline を適用した実数値型時系列データ可視化手法を提案し、その各処理について論じる。本手法の処理の流れは下記の通りである。

1. 入力データが対象とする時間帯を「局所時間帯」と呼ばれる短い時間帯に分割し、各局所時間帯内の時系列実数値にクラスタリングを適用する。
2. 各局所時間帯を構成するクラスタ群を長方形とみなし、画面空間に配置する。
3. 入力データを構成する各要素を、クラスタに対応する長方形群を通過する storyline として描画する。その際に、各要素のクラスタ移動が強調されるように storyline の透明度を自動調節する。

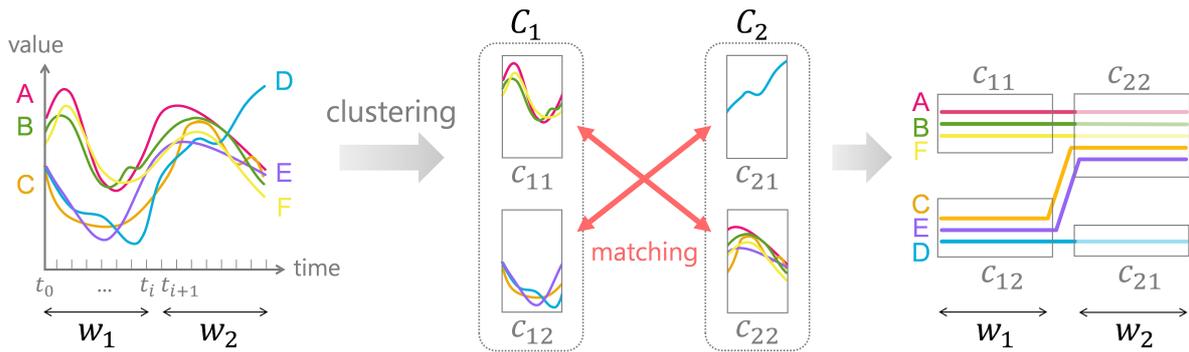


図4 本手法の処理の流れ。

図4に6個の要素集合(A~F)をもつ時系列データを用いた、処理の流れの概要を示す。

なお本章では、時系列データは  $m$  個の要素の集合  $D = \{a_1, a_2, \dots, a_m\}$  であり、要素  $a_i$  は  $n$  個の各時刻における実数値の集合  $a_i = \{a_{i1}, a_{i2}, \dots, a_{in}\}$  であるとする。また時系列データが対象とする時間帯は2時刻を区間とする  $N_w$  個の局所時間帯に分割されているとする。また、入力データの対象となる時間帯を  $N_w$  個に区切った局所時間帯のセットを  $W = \{w_1, \dots, w_{N_w}\}$  とし、 $i$  番目の局所時間帯  $w_i$  を次の通り記述する：

$w_i = \{t_{start}, t_{end}, C_i\}$  , ただし  $C_i = \{c_{i1}, c_{i2}, \dots, c_{ik}\}$  .  
 ここで、 $t_{start}$  ,  $t_{end}$  は  $w_i$  の開始時間および終了時間であり、 $k$  は局所時間帯  $w_i$  におけるクラスタ数、 $c_{ij}$  は  $j$  番目のクラスタを示す。

### 3.1 各局所時間帯におけるクラスタリング

まず時系列データを構成する各局所時間帯に対して、要素のクラスタリングを適用する。ある局所時間帯が  $j$  番目から  $k$  番目 ( $j < k$ ) までの  $n_w$  個の時刻で構成されている場合に、本処理では  $i$  番目の要素  $a_i$  から実数値  $[a_{ij}, \dots, a_{ik}]$  を抽出し、これを  $n_w$  次元ベクトルとして扱う。このベクトル群に対してクラスタリングを適用することで、各局所時間帯における要素を有限個のクラスタに分類する。現時点での我々の実装では非階層型クラスタリングを適用しているが、階層型クラスタリングを含め他のクラスタリング手法を適用することも可能である。

### 3.2 クラスタの画面配置

Storyline を適用した既存の可視化手法の処理手順は、次の2種類に大別される。

- XKCDのような表現を実現するために、曲線としての各要素の配置を最適化する方法 [6, 7]
- まず要素にクラスタリングを適用し、続いてクラスタの位置を決定する方法 [9, 10, 11]

本手法では数百・数千の要素を有する大規模な時系列データを対象とするため、前者の適用は非現実である。そこで本手法では後者を適用するが、本手法では時間的に隣接する局所時間帯間でクラスタのマッチングを取る必要がある。そこで本手法では、以下の手順でクラスタの共通度を求め、それに沿って隣接する局所時間帯間のクラスタをマッチングする。

#### 3.2.1 クラスタの共通度の算出

本手法では、 $i$  番目および  $(i+1)$  番目の局所時間帯を構成するクラスタ集合  $C_i, C_{i+1}$  に対して、以下の式(1)が最大となるような最大マッチング問題を適用する。ここで、 $c_{ij} \cap c_{(i+1)k}$  はマッチングされる2クラスタが共有する要素数、 $n_{cij}$  は  $c_{ij}$  に含まれる要素数、 $a_{cij}$  は  $c_{ij}$  を構成する実数値の平均、 $\alpha$  は  $0 \leq \alpha \leq 1$  を満たす定数である。

$$\sum \left( \alpha \frac{c_{ij} \cap c_{(i+1)k}}{n_{cij} + n_{c(i+1)k}} + (1 - \alpha) \frac{1}{|a_{cij} - a_{c(i+1)k}|} \right) \quad (1)$$

この式で第1項は両クラスタ間の要素の共有性を表すものであり、第2項は両クラスタ間の数値の不変性を表すものである。両者を考慮して最大マッチング問題を解くことで、要素を表現する曲線の移動量を抑え、かつクラスタ配置結果から数値変化を説明しやすい可視化結果を得る。

### 3.2.2 クラスタのマッチングと配置決定

本手法では各局所時間帯を構成する各クラスタを長方形で表現し, 3.2.1 節の処理によってマッチングされたクラスタが画面上で左右に隣接するように, 以下の処理によって各長方形の位置を決定する.

1. 画面左端の局所時間帯を構成するクラスタを上から順に配置する.
2. 左から  $i$  番目の局所時間帯について配置が終わったら, 左から  $(i + 1)$  番目の局所時間帯について, マッチングされたクラスタができるだけ画面上で左右に並ぶように各クラスタの位置を決定する.

### 3.3 Storyline の描画

続いて時系列データを構成する各要素を線分として描画する. 我々の実装では次の手順に沿って線分を描く.

- $i$  番目の局所時間帯にて当該要素が属するクラスタ  $c_{ij}$  の左右端および上下端の座標値を  $[x_{min_{ij}}, x_{max_{ij}}], [y_{min_{ij}}, y_{max_{ij}}]$  とする. このとき当該要素を表現するために 2 点  $(x_{min_{ij}}, y_{ij}), (x_{max_{ij}}, y_{ij})$  を結ぶ線分を描く. ただし  $y_{min_{ij}} \leq y_{ij} \leq y_{max_{ij}}$  とする.
- $i$  番目の局所時間帯と  $(i + 1)$  番目の局所時間帯を連結するために, 2 点  $(x_{max_{ij}}, y_{ij}), (x_{min_{(i+1)j}}, y_{(i+1)j})$  を結ぶ線分を描く.

このとき我々の実装では, storyline の色を以下のように定義する. 本手法では各要素に 1 次元の属性が付与されていると仮定し, その属性値に対応した色相を線分に与える. 1 次元の属性がない場合には例えば, 各要素の実数値  $a_{ij}$  で線分の各頂点に色相を与えることも可能である. また各要素の数値変化の重要度に応じて, 線分を構成する各頂点に不透明度を与える. 重要度が大きい部位において不透明度を高く設定することで, 観察者の注意を引くことができる. 現時点の我々の実装では実数値  $a_{ij}$  の微分値および各クラスタ内での要素の安定性にもとづいて重要度を算出している. 詳細は次節にて論じる.

また各要素のクラスタ移動の要因を分析するために, storyline 中の局所における実数値の変化を観察することが有用である. 我々の実装では, storyline を描画した画面領域でマウス操作で長方形を描くと, その長方形を

通過する要素群の実数値の時間変化を別の画面領域に表示するインタラクション機能を有する. 具体的には, 図 5 のように画面の上半分に storyline を描き, 上述の長方形描画操作によって要素群を選択すると, 折れ線グラフによってその要素群を画面の下半分に描く, という連携可視化を実装している.

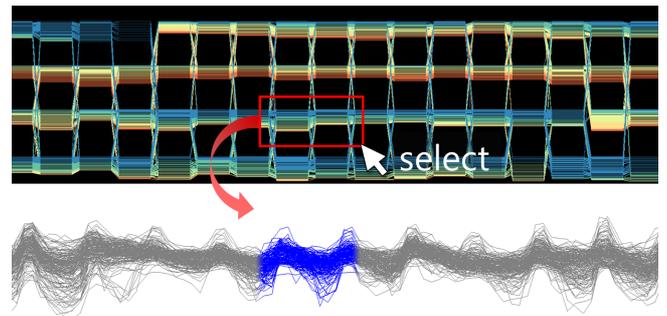


図 5 本手法におけるインタラクション機能.

### 3.4 Storyline の不透明度算出

前節で論じた通り, 現時点での我々の実装では, 次の 2 種類の基準を各時刻における重要度とみなし, storyline の不透明度算出に用いている.

隣接する局所時間帯間 においては, 要素の実数値  $a_{ij}$  の微分値を基準にする. 具体的には, 隣接する局所時間帯間でクラスタを移動する曲線について, 対応する要素の実数値の微分  $d_{ab}/|dt|$  を求める. ここで,  $d_{ab}$  はクラスタ AB 間の距離であり,  $|dt|$  は微分値を計算する時間幅を示す.

1 つの局所時間帯内 では, 要素の安定性を基準にする. 具体的には,  $i$  番目の局所時間帯について,  $(i - 1)$  番目の対応するクラスタと共通する要素は不透明度を高く, 別のクラスタから移動してきた要素は不透明度を低く設定する.

不透明度にもとづく視覚表現を採用した理由は以下の通りである. 人間の目は知覚的に変化の激しい部分に対して反応する [12] という Marr の理論を適用して, 我々は storyline 上での要素のクラスタ移動という重要な現象が発生している部位に視覚変化の大きな表現を適用したいと考えた. 可視化における視覚表現の手段として, Bertin [13] によって提案された視覚変数集合が知られている. それに対して MacEachren は, 画面内の同一座標上に複数の情報を載せる場合に不透明度を用いるのが有効であると述べている [14]. 本手法においても,

storyline を構成する線分において重要度の高い部位にのみ高い不透明度を付与することで、既に用いられている他の視覚変数（線分の形状や色相）に干渉することなく、ユーザの注意を重要度の高い部位に向けられると考える。また、不透明度から定量的な差異を読み取ることは困難であるものの、ユーザは不透明度を見て直感的にその部位における重要度を判断可能である。

## 4 実行結果

### 4.1 実行環境と処理時間

我々は JDK (Java Development Kit) 1.7.0 を用いて提案手法を実装し、Windows 8.1 (64bit) 搭載のラップトップ PC (CPU 1.80 GHz, RAM 4.0GB) 上で実行した。表 1 は上記の環境におけるクラスタリング、クラスタの画面配置およびインタラクションの平均処理時間を示す。なお、いずれの事例も元データ内の時間帯数は 240、また、クラスタ数を 3、局所時間帯サイズを 12 時間に設定した。

表 1 各処理における処理時間 (秒, 5 回の平均)。

要素数	クラスタリング	画面配置	インタラクション
86	0.40	0.08	0.03
376	1.83	0.19	0.18

上記の結果から、既存手法 [6], [7] に比べ、本手法は対話的に操作する上で十分高速であるといえる。

### 4.2 適用データ

本章では、全国 376 箇所の観測所について 3 時間ごとに観測された 2 種類のアメダス気温データを提案手法に適用した事例を紹介する。可視化結果において、x 軸は時刻、1 本の曲線は 1 観測所を示し、近接する曲線は同じクラスタに属することを表す。この事例ではクラスタ数を 3、局所時間帯サイズを 12 時間ごと、各クラスタ内の曲線の並びは観測所 ID 順とした。また観測所 ID から観測所を 8 つの地方に分類し、ColorBrewer 2.0[15] の配色方針に沿って各地方に色を割り当てた (図 6 参照)。

ここで、storyline の読み取りの凡例として、図 7 に 12 観測所の気温データを適用した実行結果の一例を示す。3 番目の局所時間帯において、各クラスタには上から順に、東北地方、東北から九州地方、北海道地方に属する観測所が含まれている。クラスタ 2 に着目すると、東北地方を示す青緑の線分と中部地方を示す黄緑の線分がそ



図 6 曲線の色と地方の対応。

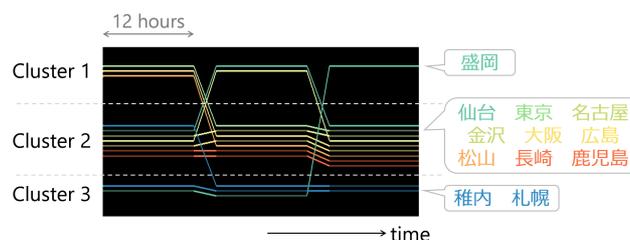


図 7 12 観測所の気温データを適用した実行例。

れぞれ 1 本ずつ不透明度が高く表示されている。これは 2 番目の局所時間帯ではクラスタ 1 に属する当該線分が、3 番目の局所時間帯でクラスタ 2 に切り替わっているためといえる。

また本章では、3.3 節で述べた要素選択機能を用いた折れ線表示についても示す。折れ線の表示においては、x 軸が時刻、y 軸が気温、各折れ線が各観測所を示す。

### 4.3 2009 年 4 月の事例

図 8 (上) は 2009 年 4 月の 1 カ月全体の気温変化を表す storyline である。上のクラスタから順に、青で示された北海道、青緑から黄緑で示される東北から中部、黄色から赤で示される近畿・中国から沖縄・南西諸島に属する観測所を多く含んでおり、ところどころ変動が見られるものの、全体的に似たような地方が安定して同じクラスタに配置されている。北海道の観測所を多く含むクラスタについて、前半 ( $w_a$ ) と月末 ( $w_b$ ) にそれぞれ破線で囲んだ 1 局所時間帯分のみ、青から橙で示させる北海道から九州にかけての観測所が含まれており、全国的に同じような気温の変化をしていたと推察できる。これらの局所時間帯について、長方形による要素選択機能を用いて、各クラスタに属する要素群を折れ線で描いた結果を図 8 (中)(下)に示す。 $w_a$ ,  $w_b$  いずれにおいても他の時間帯に比べ、全国的に似たような気温変化を示しており、気温差が小さくなっていることが確認できる。

局所時間帯  $w_b$  についてさらに観察するため、月末の

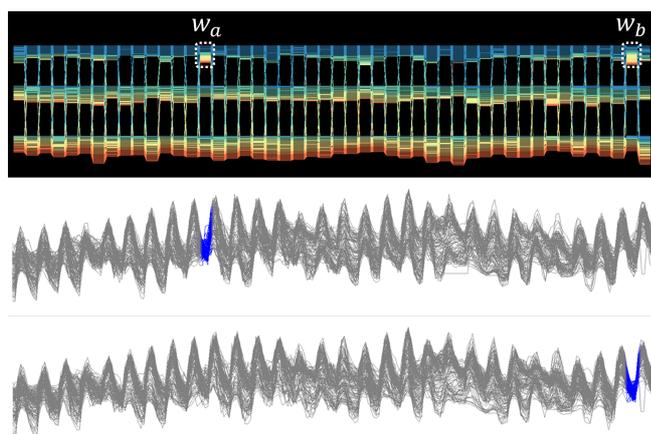


図 8 2009 年 4 月の気温変化 : (上) storyline による表示 . 局所時間帯  $w_a$  (中) /  $w_b$  (下) における , 選択したクラスタに属する要素の折れ線表示 .

1 週間分を拡大した結果を図 9 に示す . storyline 上の破線で囲まれた局所時間帯  $w_b$  について , 青から橙で示された要素が含まれるクラスタが 2 種類存在することが観察できる . クラスタ  $c_{b1}$  ,  $c_{b3}$  それぞれについて要素選択機能を用いて折れ線を表示した例を図 9 (中) (下) に示す . 各クラスタに含まれる折れ線群は似たような気温変化を示しているものの , 平均気温や気温差が異なっていることがわかる . これは日本海側と太平洋側で異なる気温変化を示したためと推察できる . 選択した折れ線群に着目すると , いずれのクラスタにおいても , 選択した局所時間帯以前は低温で推移していた折れ線群が , この時間帯だけ気温が高くなったため , 全国的に同じような気温変化となったと考えられる . 図 9 (下) に示されるクラスタ  $c_{b3}$  内の要素については特にその傾向が顕著であり , storyline 上でも青や青緑で表された北海道や東北地方の観測所の要素が不透明度が高く表示されている . このことから , 直接数値変化を示していない storyline による表示からでも , これらの地方で他の時間帯とは違う特徴的な気温変化があったと推察できる .

同じデータについて , 折れ線によって直接全観測所の気温を表示した例を図 10 に示す . 期間全体を通して似たような気温変化を示しているため , どの部分に着目すべきかが判断し難く , この図のみから上記のような知見を得ることは困難である . この結果から , storyline によって全体の傾向を俯瞰し , 細かな数値変化を見たい場合は折れ線グラフによる結果と比較するというアプローチが有効であるといえる .

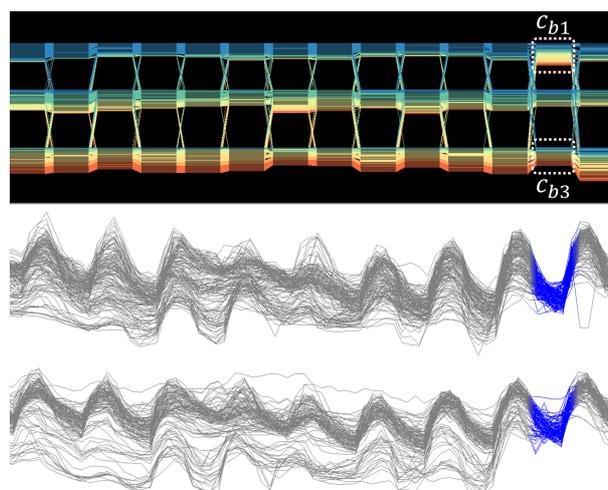


図 9 2009 年 4 月末 1 週間分の気温変化 : (上) storyline による表示 . (中) クラスタ  $c_{b1}$  / (下) クラスタ  $c_{b3}$  に属する要素の折れ線表示 .

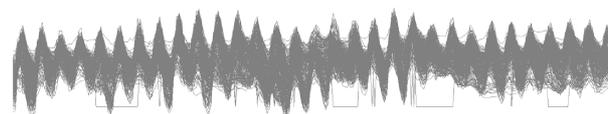


図 10 2009 年 4 月の気温変化を折れ線によって可視化した結果 .

#### 4.4 2009 年 11 月の事例

図 11 (上) は 2009 年 11 月の月上旬 2 週間の気温変化を表す storyline である . y 軸上中央に配置されたクラスタは比較的安定して推移しているが , 図 11 (上) の下部に配置されたクラスタに着目すると , 期間の前半は , 青緑から黄色で示される東北から中部地方の観測所を多く含んでいるが , 破線で囲まれた局所時間帯において , 青で示された北海道の観測所ばかりを含む表示に切り替わることが観察できる . この局所時間帯の前後で他のクラスタも観察してみると , 前半は北海道の観測所が複数のクラスタに分散しているが , 切り替え後は一番下のクラスタに集約されており , その後の時間も不透明度が低いことから , そのまま安定して推移していることがわかる . このことから , 11 月の 2 週目以降 , 北海道で他の地方とは極端に異なる気温変化を示していたと推察できる .

前述の局所時間帯の前後 2 箇所を選択し , 該当する折れ線を表示させた結果を図 11 (下) , 12 (下) に示す . 図 11 (下) は切り替えの前後 2 局所時間帯分 , 図 12 (下) は切り替え後 , 青の要素で安定して推移している 9 局所時間帯分を選択したものである . 図 11 (下) について ,

選択した局所時間帯の前数日間は折れ線群全体が似たような気温変化を示しているのに対し、選択部分以降は数日間にわたって気温の開きが顕著であることが観察できる。図 12 (下) の安定したクラスタを選択した結果と比較してみると、このクラスタに含まれる折れ線群（北海道の観測所群）が当該期間において、極端に低い気温変化を示していることがわかる。

これらの結果から、storyline 上のクラスタの安定性から、気温変化が他と異なる時間帯を推測できるといえる。

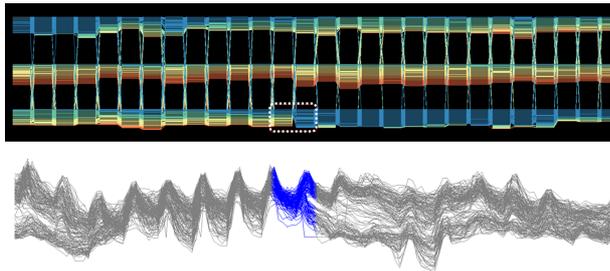


図 11 2009 年 11 月上旬 2 週間分の気温変化：(上) storyline による表示。(下) 選択した 2 局所時間帯に含まれる要素の折れ線表示。

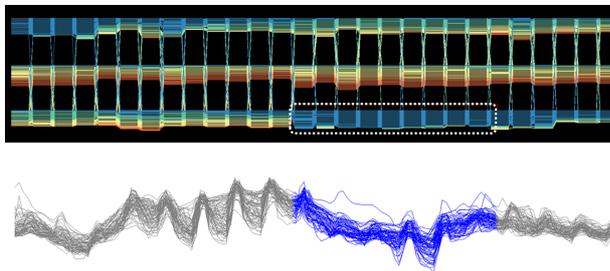


図 12 2009 年 11 月上旬 2 週間分の気温変化：(上) storyline による表示。(下) 選択した 9 局所時間帯に含まれる要素の折れ線表示。

## 5 まとめと今後の課題

本論文では、storyline を用いた実数値型時系列データの可視化手法を提案した。本手法では局所時間ごとに時系列実数値情報の類似度を算出し、隣接する局所時間帯間のクラスタ群にマッチングを適用することでクラスタの配置を決定し、類似した要素同士が画面内で近接するような storyline を描く。また、本手法では各要素の数値変化の重要度にもとづいて要素の不透明度を設定することで、重要度の高い部分を強調し、ユーザの知覚に近

い可視化結果を得る。

今後の課題として、気象以外のデータでの適用事例を示し、ユーザテスト等によって可視化結果の妥当性を検証することを計画中である。また手法部分の拡張として、各要素の実数値間の類似度を算出する単位となる局所時間帯を可変にすることで、よりデータの局所の特徴に応じたクラスタ切り替えの数や時間を検出したいと考えている。この点に関して、不等間隔離散化を適用したパターン抽出手法 [16] などを用いることが考えられる。さらに、3.1 節でも示したように、階層型クラスタリング手法やクラスタ数可変型のクラスタリング手法などの適用を検討したい。

## 参考文献

- [1] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. *Visual Languages, 1996. Proceedings., IEEE Symposium on*, 336–343, 1996.
- [2] S. Yagi, Y. Uchida, T. Itoh. A polyline-based visualization technique for tagged time-varying data. *16th International Conference on Information Visualisation*, 106–111, 2012.
- [3] 伊藤貴之. 視覚協創学 (5) マッピング技術: 特に時系列データ可視化技術の体系化に向けて. 日本計算工学会計算工学講演会, 17, 4, 2012.
- [4] M. Imoto, T. Itoh. A 3D visualization technique for large scale time-varying data. *14th International Conference on Information Visualisation*, 17–22, 2010.
- [5] R. Munroe. Movie narrative charts. <http://xkcd.com/657/>, accessed 4 June, 2015.
- [6] M. Ogawa, K.-L. Ma. Software evolution storylines. *ACM International Symposium on Software visualization*, 35–42, 2010.
- [7] Y. Tanahashi, K.-L. Ma. Design considerations for optimizing storyline visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2679–2688, 2012.
- [8] S. Liu, Y. Wu, E. Wei, M. Liu, Y. Liu. Storyflow: Tracking the evolution of stories. *IEEE Transactions on Visualization and Computer Graphics*,

19(12), 2436–2445, 2013.

- [9] K. Reda, C. Tantipathananandh, A. Johnson, J. Leigh, T. Berger-Wolf. Visualizing the evolution of community structures in dynamic social networks. *Computer Graphics Forum*, 30, 1061–1070, 2011.
- [10] A. Lex, C. Partl, D. Schmalstieg. Comparative analysis of multidimensional, quantitative data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 1027–1035, 2010.
- [11] C. Turkay, J. Parulek, N. Reuter, H. Hauser. Interactive visual analysis of temporal cluster structures. *Eurographics/IEEE Symposium on Visualization (EuroVis2011)*, 711–720, 2011.
- [12] D. Marr. *Vision: A computational investigation into the human representation and processing of visual information*. New York: Henry Holt and Co., Inc, 1982.
- [13] J. Bertin. *Semiology of graphics: diagrams, networks, maps*. University of Wisconsin press, 1983.
- [14] A. M. McEachren. *How maps work: representation, visualization, and design*. The Guilford Press, New York, 1995.
- [15] M. Harrower, C. A. Brewer. *Colorbrewer 2.0*. <http://colorbrewer2.org/>, accessed February, 2015.
- [16] M. Ohsaki, H. Abe, T. Yamaguchi. Numerical time-series pattern extraction based on irregular piecewise aggregate approximation and gradient specification. *New Generation Computing*, 25(3), 213–222, Springer, 2007.

八木 佐也香



2011 年お茶の水女子大学理学部情報科学科卒業 . 2013

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 . 現在 , 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後期課程に在学中 .

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 (客員助教授担当) . 2005 年日本アイ・ビー・エム (株) 退職 , お茶の水女子大学理学部情報科学科助教授 . 2011 年同大学教授 , 同大学シミュレーション科学教育研究センター長 , ACM, IEEE Computer Society, 芸術科学会 , 他会員 .

高塚 正浩



Masahiro Takatsuka received his Ph.D in Electrical and Computer Systems Engineering from Monash University in 1997. He is Director of the High-Performance Visualization Research Laboratory (ViSLAB) at the University of Sydney. He has been active researcher in the area of Scientific Visualization and Visual Analytics since 1998. He was a Senior Research Faculty at GeoVISTA Centre, the Pennsylvania State University between 1998 and 2002. He joined the University of Sydney in 2002 and was also a researcher at National ICT Australia between 2004 and 2007.