

動向情報の探索的データ解析を可能とする情報可視化システム

Information Visualization System for Exploratory Data Analysis of Trend Information

高間 康史 山田 隆志
Yasufumi Takama Takashi Yamada

首都大学東京
Tokyo Metropolitan University

This paper proposes the visualization cube for modeling an interaction in exploratory analysis of spatiotemporal trend information. The visualization cube consists of 4 axes; spatial and temporal axes, statistics-values axis, and type-of-views axis. Interactions for generating views are also defined as the operations on the visualization cube, which include drill down/up, comparison, spin, and transition. The interactive information visualization system for spatiotemporal trend information is developed based on the concept of visualization cube. The system was used in actual classes of an elementary school, of which the result shows the system has enough usability for 5th-grade elementary school children to perform exploratory data analysis.

1. はじめに

本稿では、動向情報に対する探索的データ解析のインタラクションモデルを構築するために、時空間動向情報を含む対象データを抽象的に構造化した可視化キューブを提案する。近年の計算機・ネットワーク環境の整備により、動向情報の集積・解析が可能となっており、意思決定などで有効に活用するために動向情報可視化システムなどの開発が行われている。しかし、探索的解析行動における自由度は非常に高く、効果的なシステムを実装するためには適切なインタラクションモデルの構築が不可欠である。提案する可視化キューブは、動向情報の可視化に関連する要素を4次元キューブとして整理し、これに対する操作としてインタラクションを定義する。可視化キューブに基づき動向情報可視化システムを実装し、小学校の授業で利用してもらった結果より、その有効性を示す。

2. 動向情報の可視化

「動向」とは、事態の変化していく方向や、社会や組織などの現状や今後のなりゆきなどを表す。動向を正しく把握することは、組織、個人における意思決定において重要であり、動向情報を収集・生成し、わかりやすく提示するシステムの開発は今後重要と考える。動向情報の要約と可視化に関するワークショップ (MuST) では、内閣支持率など20話題についてのタグ付き新聞記事をコーパスとして提供している。これらのコーパスを元に、テキストから動向情報の元となる数値データを抽出する研究や、動向情報に関する数値データからグラフを描画したり、動向を説明する自然言語文を自動生成するなどの研究が行われている [Kato 08]。

動向情報は、複数の時点・地点におけるデータを集約して初めて明らかになるものであるため、テキストによる要約や、情報可視化による集約が必要である。統計量の時系列変化 (時間的動向情報) は、棒グラフや折れ線グラフにより表現可能であり、Elucignage [Matsushita 07] などで採用されている。時間的動向情報に対し、台風や地震は、台風の進路や震源と各地における震度の関係などの空間的な変動も伴う点で異なる。これら、空間的な変動を伴う動向情報 (空間的動向情報) も可視化

連絡先: 高間 康史, 首都大学東京, 〒191-0065 東京都日野市 旭が丘 6-6, ytakama@sd.tmu.ac.jp

するためには、日本地図表現なども用いる必要があり、地震に関する動向情報を対象としたインタラクティブ情報可視化システム [高間 06] では、棒グラフと日本地図の連動により、時空間動向情報の可視化を行っている

3. 探索的動向情報解析のインタラクションモデル

2. 節で述べた時空間動向情報を対象とし、これに対する探索的データ解析に関するインタラクションモデルを定義するために、可視化キューブを提案する。動向情報を表す可視化表現を生成しながら探索的に解析を行う際に発生するインタラクションについて、本稿では以下の4種類に分類して考える。

1. ある統計量についての動向を示す可視化表現を生成する
2. 可視化の粒度を変更して可視化表現を生成する
3. 異なる統計量間などの比較を示す可視化表現を生成する
4. ある動向情報について、異なる可視化表現に切り替える

(1) は最も基本的なインタラクションであり、探索的分析の起点を与えるものである。時空間動向情報を対象とする場合、時間的あるいは空間的動向に関する可視化表現の生成に分けられ、それぞれに適した可視化表現は異なる。また、同一の統計量について、時間的動向から空間的動向に切り替える事により異なる観点から解析を行う事は、探索的解析において重要と考える。(2) は、年単位の気温変動を、月単位に変更したり、都道府県毎にビール販売量を集約していたものを、地方毎に集約し直して可視化する場合などに相当する。

(3) に関しては、ある年における気温とCO₂の空間的動向の比較や、今年と昨年における、1年間の気温変化の比較、といったように、比較対象を統計量、時間、空間のどれにするか、動向を表す軸 (動向軸) を時間軸、空間軸のどちらにするかでいくつか分類できる。統計量、時間については、時間的動向の比較が可能であるが、例えば関東地方と東北地方では1レベル下の都道府県が異なるため、空間について、空間動向の比較を行う事はできない。(4) については、ある可視化表現を用いて提示されている動向情報を、よりわかりやすい可視化表現

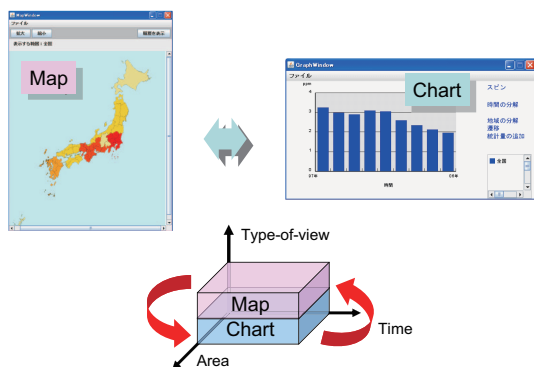


図 1: 遷移操作の適用例

に変更したり、異なる可視化表現で眺めることで新たな理解・解釈を試みる際のインタラクションである。

上記インタラクションをモデル化するために可視化キューブを提案する。可視化キューブは、時空間動向情報を含む対象データを抽象的に構造化したものであり、時間軸 (time)、空間軸 (area)、統計量軸 (statistics values)、可視化表現軸 (type-of-view) の 4 次元データキューブとして定義される。時間軸、空間軸は対象データの時間属性、空間属性にそれぞれ対応しており、それぞれ階層情報を持っている。統計量軸は、可視化対象となる統計量を扱い、可視化表現軸は、システムが利用可能な可視化表現を扱う。

可視化キューブにおいて、可視化表現の生成は、ある統計量について、時間軸・空間軸の一方を動向軸に、統計量・地点 (空間的属性)・時点 (時間的属性) に適宜記号・線種・色などを割り当てる操作としてモデル化される。探索的解析は、このようにして与えられた可視化表現に対し、前述のインタラクションを行い、新たな可視化表現を生成していくプロセスとしてとらえられる。これらのインタラクションをモデル化するために、可視化キューブに対するスピン、ドリルダウン/アップ、比較、遷移の 4 操作を定義する。

スピン操作は、前節の (1) における時間・空間動向情報の切り替えに相当する操作であり、横軸に割り当てる軸を時間軸、空間軸間で入れ替える事でモデル化される。ドリルダウン/アップは、OLAP における同名の操作と同様であり、(2) に対応する。比較は (3) に、遷移は (4) にそれぞれ対応する操作である。遷移操作は動向軸を固定したものでなく、遷移後の可視化表現に適したものに变更可能とする。

4. 評価実験

可視化キューブに基づく動向情報可視化システムを、Java により実装した。可視化表現として、現在のシステムでは時間的動向情報を可視化するグラフ、空間的動向情報を可視化する日本地図表現の二種類を実装しており、それぞれ独立したウィンドウとして表示する。システムの動作例として、遷移操作を適用した場合のスクリーンショットを図 1 に示す。

公立小学校の授業における、環境に関する調べ学習の中で、提案システムを調査ツールの一つとして使用してもらった。5 年生の 2 クラスに対し、提案システムをインストール済みのノート PC を各 2 台提供した。担任教諭と相談し、CO₂、SO₂、人口、河川品質、気温の 6 種類の統計量を用意した。

1 時限の間、システムを自由に利用してもらい、授業終了後に、操作に関する全体的感想 (Q1)、新しい発見があったか

表 1: 小学生へのアンケート結果

児童	Q1	Q2	Q3	Q4	Q5	操作時間 (分)
A	4	2	3	3	4	9
B	4	4	3	3	4	39
C	4	3	4	3	3	6
D	4	3	4	4	4	11
E	4	4	4	3	4	10
F	4	3	3	3	4	32
G	3	3	3	4	3	8
H	3	3	2	4	3	11
I	4	4	3	4	4	9
J	3	2	2	2	3	5
K	4	4	3	4	4	8
平均	3.7	3.2	3.1	3.4	3.6	13.5

(Q2)、十分に調べられたか (Q3)、思い通りの結果が得られたか (Q4)、今後も利用したいか (Q5) について 4 段階 (4 が高評価) で回答してもらった。システムを利用した児童は 11 名 (A~K) であり、パソコンの利用経験についてシステム利用前に行ったアンケートによると、マウスはある程度利用できるが、キーボードにはそれほど慣れていない児童がほとんどであった。

表 1 に示すアンケート結果より、全ての項目について肯定的な評価が得られていることがわかる。特に、Q1、Q5 の評価が高くなっており、好意的に利用してもらえたことがわかる。一方、Q2、Q3 が相対的に低い評価となっているが、これについては探索的データ分析のスキルにも関わることであり、短時間の利用で判断する事は難しいと考える。また、システム利用時間が短かったことも、これらの項目の評価が相対的に低くなった要因と考える。例えば、児童 J は Q2~Q4 に他の児童より低い評価 (2) をしているが、システム利用時間が 5 分であり、満足できるまでシステムが利用できなかったと考える。

5. まとめ

本稿では、動向情報に対する探索的データ解析におけるインタラクションについて考察し、これをモデル化するために可視化キューブを提案した。可視化キューブに基づき動向情報可視化システムを実装し、小学校の授業で利用してもらった結果、計算機スキルを持たない児童でも十分に利用可能であることを示した。提案するインタラクションモデルは、探索的行動の分析にも有効と考えており、システムデザインや知的作業支援の実現にも貢献することが期待できる。

参考文献

- [Kato 08] T. Kato and M. Matsushita: Overview of MuST at the NTCIR-7 Workshop –Challenges to Multimodal Summarization for Trend Information –, Proc. NTCIR-7 Workshop, pp. 475–488, 2008.
- [Matsushita 07] 松下, 加藤: 言語情報と数値情報の相補的利用を目指した可視化手法, 第 21 回人工知能学会全国大会論文集, 3H8-3, 2007.
- [高間 06] 高間, 山田: タグ付きコーパスを用いた地震記事からの動向情報抽出・可視化システム, 知能と情報, Vol. 18, No. 5, pp. 711–719, 2006.