〔特集論文〕「近未来チャレンジ 2010」

比較分析に着目した時空間的動向情報の探索的分析支援

Exploratory Analysis Support of Spatiotemporal Trend Information Focusing on Comparative Analysis

高間 康史
Yasufumi Takama

首都大学東京大学院システムデザイン研究科
Graduate School of System Design, Tokyo Metropolitan University
ytakama@sd.tmu.ac.jp, <http://www.krectmt3.sd.tmu.ac.jp/~takama/>

狩野 真次^{*1} (同 上)
Shinji Karino

keywords: information visualization, spatiotemporal trend information, exploratory data analysis, comparative analysis

Summary

This paper proposes a method for supporting an exploratory analysis of spatiotemporal trend information by focusing on comparative analysis. Recent growth of the Web has brought us various kinds of information, among which trend information is one of crucial information for our decision-making. In order to develop an exploratory analysis support system for spatiotemporal trend information, visualization cube has been proposed. It is an abstract data structure with 4 axes, on which 5 operations for generating views are defined. The prototype system has been developed based on the concept of visualization cube, of which usability has been evaluated through the experiments conducted at an elementary school. However, it was also pointed out that the data space that can be explored with a single interaction by the existing system is not so large. In order to solve this problem, the proposed method employs dual views by considering two visualization cubes. By operating two visualization cubes at the same time, six kinds of comparative analyses in a broad sense can be supported. The exploratory analysis support system is developed based on the proposed concept, of which the effectiveness is evaluated through experiments with test participants. Although no synchronization mechanism between views is implemented in the prototype system, effective synchronization functions are discussed based on the experimental results.

1. は じ め に

本論文では、広義比較分析の観点から、時空間的動向情報の探索的分析を支援する手法を提案する。大規模データの収集が容易になった現在、個々のデータを独立して扱うのではなく、集約・編纂して活用することが重要となっており、大規模テキスト情報を扱うためのテキスト要約技術、大規模・多次元データ空間を視覚的にわかりやすく提示する情報可視化技術などが研究されている。本論文で対象とする動向情報も、時間的・空間的に関連のあるデータ（統計データなど）を集約することにより、その時間的、空間的変動を扱うものであり、意思決定などへの活用が期待されている [加藤 04, Kato 08]。

動向情報など、集約されたデータを活用する際、分析目的や検証すべき内容が予め決まっているとは限らず、むしろ多角的な観点からデータを眺め、試行錯誤しながら気づいた点を掘り下げていくような、探索的分析が行われる場合が多い。高度情報化社会においては、多種・多

様なデータが社会にあふれており、これら进行分析、解釈し、問題解決や知識創造につなげる能力を統計教育などを通じて身につける事が重要であると指摘されている [統計委 05]。従って、探索的データ分析は現代の我々にとって必要不可欠な知的能力の一つであり、これを支援するシステムの実現が重要と考える。

計算機利用スキルを十分に持たないユーザでも特別な訓練の必要なく、動向情報の探索的分析が行えるシステムが提案されており、小学校での環境問題に関する調べ教育での使用を通じて有効性が示されている [高間 10]。このシステムでは、見通しのよいインタラクションを実現するために、時空間的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し、これに対する操作としてインタラクションを定義している。可視化キューブに基づき設計されたインタラクションモデルを用いることにより、特別な訓練をしなくても操作が理解しやすいシステムの実装が可能であることが示されているが、分析作業が特定の可視化表現に偏るなどの問題があり [Takama 10]、探索空間の広さの点で改良の余地が

^{†1} 現在は、株式会社バンダイナムコゲームスに勤務。

あることが示唆されている。

本論文では、情報の比較に基づく分析が、探索的データ分析において重要であるとの考えに基づき、広義の比較分析を支援するインタラクションを提供することにより、探索空間の拡大を実現する。具体的には、可視化キューブに対するプリミティブ操作の組合せとして広義比較操作を定義し、それらを統一的に実行可能なインタフェースとして、2画面表示インタフェースを提案する。プロトタイプシステムを開発し、実験協力者間で統計データに関する作問・解答タスクを行ってもらった結果より、2画面表示を活用した多様な比較分析が行われていることを示す。また、探索的分析において実際に用いられる操作、および2画面表示を用いた分析パターンに基づき、有効な2画面連動機能についても考察する。

2章では、動向情報、探索的データ分析、広義比較分析支援についての関連研究および可視化キューブについてまとめる。3章で、可視化キューブ上で可能な広義比較分析を整理し、プリミティブ操作の組み合わせにより各種比較分析を実行可能なインタフェースとして2画面表示によるインタフェースを示す。プロトタイプシステムを実装し、実験協力者に作問・解答タスクを行ってもらった結果を4章に示す。

2. 関連研究

2.1 動 向 情 報

「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。多様なセンシングや大量データ蓄積が容易になった現在、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となってきており、情報編纂技術の活用による動向情報の生成、分析支援が期待される。

動向に関する情報には、気温変化、株価変動、内閣支持率の変化など、様々なものが存在する。Webにおける検索キーワードのランキングや、ブログなどにおいて言及されているトピック（キーワード）の時間的変動なども、動向情報の一種であり、Amazonにおける書籍の売り上げランキングとブログ上での言及数の推移の相関 [Gruhl 05] や、ブログ中の評判情報を利用した商品の売り上げ予測 [Liu 07] など、実世界の動向との相関分析なども試みられている。

NTCIRでの評価型ワークショップとして実施された「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、動向情報を「幾つかの統計データの時系列データを基として、その変化を通時的にとらえつつ、それらを単に羅列するのではなく、総合的にまとめ上げることで得られるもの」と定義している。この定義に基づき、ガソリン（全国平均店頭価格など）、交通事故（全国の交通事故死者数）、台風、地震、政治動向（内閣支持率など）、日経平均株価など、計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 [加藤 04]。これらのコーパスを元に、自然言語処理、情報可視化の観点から、あるいはそれらを融合したアプローチにより、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]。

2.2 探索的データ分析

代表的なデータ分析の方法論には、検証的データ分析 CDA (confirmatory data analysis) と、探索的データ分析 EDA (exploratory data analysis) がある。CDAは、すでに持っている仮説の検証が目的であり、統計的検定などが用いられる [椿 07]。分析対象となるデータは、検証目的に合わせて実験計画を立て、収集することが多い。

一方、本論文で対象とする EDA は、すでにあるデータを出発点として、そこから仮説・知見などを得る事が目的である。分析対象となるデータは通常大規模、多属性であるため、全体的な傾向を把握したり、特異点や特徴的なパターンを発見するためには、データを多様な観点から眺めることが必要となり、その分析行為は探索的となる。

EDAを効果的に支援するためには、分析者に理解しやすい形式でデータを提示する事が重要であり、情報可視化技術が有効である [増井 04]。特に、単なる情報提示の手段ではなく、情報アクセスのための対話的インタフェースとして情報可視化技術を利用する、インタラクティブ情報可視化システムが研究されている [Chen 09, 松下 06, 高間 10]。InTREND [松下 06] は、分析におけるコンテキストを考慮することで、分析者とシステムの間のインタラクションを自然なやり取りとして実現している。コンテキストを保持することで、ユーザの自然言語質問を正しく理解することを可能としている。高間らは、時空間的動向情報を対象とした EDA を支援するインタラクティブ情報可視化システムを開発しており、見通しのよいインタラクションを実現するために、次節で紹介する可視化キューブを提案している [高間 10]。

また、データマイニングにおいても、大規模データから知見や興味深いパターンを発見するためには、分析者がマイニングプロセスに主体的に参加することが必要とされており、アクティブマイニング [Motoda 02] やビジュアルデータマイニング (ビジュアルマイニング) [Mozzafari 09, 白山 06] のコンセプトが提唱されている。

これらの EDA において、分析者が行う行為は多岐に渡るが、比較行為がその中心になることが指摘されている [Keogh 06, 白山 06]。Keogh らは、情報可視化やデータマイニングは “Compared to what?” に答える技術であるとしている [Keogh 06]。白山は、比較には、基準と比較しながら最も良い情報を探索する行為と、比較対象の差異に対して解釈を与えながら必要な情報を抽出していく

行為の 2 つの方法論があることを指摘しているが、EDA においては後者が主体になる。本論文では、後者の比較行為を「広義の比較分析」と見なし、これを支援する事により探索空間の拡大を実現する。

2.3 広義比較分析の支援

EDA を支援するシステムは、程度やアプローチの違いこそあれ、広義の比較分析行為を何らかの点で支援していると言える。本節では、利用するビュー (view) の観点から以下の 3 種類に分類してまとめる。ここで、ビューとは可視化表現を用いて描画されたものを指す。

- 単一ビュー内での比較を支援するアプローチ
- ビューの遷移を利用した比較を支援するアプローチ
- 複数ビューを用いた比較を支援するアプローチ

単一ビュー内での比較は、棒グラフや折れ線グラフなど一般的な統計グラフでも行われるものであり、同一の軸を用いるためデータ間の対応関係を容易に認知することができる。グラフ構造などを用いた可視化では、比較すべきデータ間の対応関係は必ずしも明確ではないため、共通部分や差異を強調するアプローチが比較支援としてよく用いられる。片上らは、学会誌・論文誌に掲載された論文の情報から、学会における研究動向の変遷を捉える事を目的として研究分野ネットワークと呼ぶ可視化手法を提案している [奥岡 09, 清水 07]。この手法では、共起度の高いキーワード同士を接続したネットワーク構造 (研究分野ネットワーク) を生成し、可視化を行う。会誌の発行年や、会議の開催年による研究分野の変化などを容易に把握できるように、差分ネットワークを提案している。差分ネットワークでは、キーワード数とエッジの重みについて、2 つのネットワーク間で差分をとり、差のプラス、マイナスそれぞれを赤色、青色で表現することにより、異なる時点間の差分を強調した描画を行う。

ビューの遷移を利用した比較は、同種の可視化表現内で遷移する場合、異種の可視化表現間で遷移する場合に大別できる。前者には紙芝居 Keygraph [Ohsawa 08] などがあり、時間的変動を見る場合などに利用されることが多い。後者の場合には、遷移間での対応関係を明示するために、アニメーションを利用する手法が提案されている [Heer 07, 松下 06]。

複数ビューを用いた比較の支援も、同種の可視化表現を並べる場合、異種の可視化表現を並べる場合に大別できる。前者には、Web サイトの比較閲覧を支援するために 2 画面表示のブラウザを利用する Comparative Web Browser (CWB) [Nadamoto 03] や、Web サイトのアクセス状況を時系列で可視化する手法 [Chi 98] などがある。CWB では、一方のブラウザ画面にスクロールや遷移などの操作を行うと、他方のビューでも対応する部分が自動的に表示される。Chi らのシステムでも、複数の Disk tree を並べることで Web アクセスの時系列変化を可視化するが、全ての Disk tree は同期して回転などを行う。後者で

図1 可視化キューブの構造

は、異なる可視化表現で同一、あるいは関連するオブジェクト集合を提示していることを明確にするために、クロスリンク [Sarma09] やクロスフィルタリング [Weaver10] といった手法が利用されることが多い。これらは、あるビューでオブジェクトの選択やフィルタリングを行うと、他のビューも同期して変更することで、ビュー間での対応箇所の同定を支援する手法である。

2.4 可視化キューブ

インタラクションモデルは、既存システムの理解、比較や、システム設計時における個々の機能およびその実現手段の検討などに利用可能であり、システム設計者にとって有益であるが、適切なインタラクションモデルに従い、首尾一貫した、わかりやすいインタフェースを設計することができれば、ユーザにとっても有益である。可視化キューブ [高間 10] は、時空間的動向情報を対象とした EDA に特化したインタラクションモデルを構築するために提案されたものであり、時空間的動向情報を含む対象データを抽象的に構造化している。

可視化キューブは、図1に示す様に、時間軸 (time)、空間軸 (area)、統計データ軸 (statistical data)、可視化表現軸 (type-of-view) から構成される 4 次元データキューブとして定義される^{*1}。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、「日本全体 → 地方 → 都道府県」や「年 → 月 → 週 → 日」などのように階層情報を持っている。統計データ軸は、可視化対象となる統計データを扱い、図1では降雨量 (precipitation)、気温 (temperature) が存在することを示している。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軸であり、図1では統計地図である日本地図表現 (map of Japan) と棒グラフ (bar chart) が利用可能であることを示している。

ビューの描画は、ある統計データについて、時間軸・空間軸の一方を動向軸とし、統計データ・地点 (空間属性)・時点 (時間属性) に対し、選択された可視化表現における視覚的要素 (記号・線種など) を割り当てる操作としてモデル化される。EDA は、このようにして与えられたビューに対してインタラクションを行い、新たなビューを生成していくプロセスとしてとらえられる。こ

*1 図1では、見やすさのために 3 次元キューブ二つとして示している。

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以下の4種類の操作が定義されている。
ドリルダウン/アップ 統計データの時間的・空間的粒度の調整

比較 統計データ軸，空間軸へのデータ追加
スピン 時間軸，空間軸間の動向軸の切り替え
遷移 可視化表現の切り替え

例えば，ドリルダウン操作を空間軸に適用することにより，地方単位での統計データ比較から，都道府県単位の比較に変更される。また，関東地方の気温変化を棒グラフにより表示している状況で，空間軸に東北地方を追加する比較操作により，両地方の気温変化を比較する事が可能となる。

3. 可視化キューブに基づく広義比較分析支援

3.1 広義比較分析の観点からの可視化キューブの拡張

可視化キューブは，見通しの良いインタラクションを実現する上で有用であることが示されているが，分析作業が特定の可視化表現に偏るなどの問題があり [Takama 10]，探索空間の広さの点で改良の余地があることが示唆されている。これは，既存システムで利用可能な可視化表現が棒グラフと統計地図の二種類に限定されていることに加え，一度のインタラクションで可能な操作が4軸の内一つの軸に変更を加えるか，あるいは動向軸の交換を行うことに限定されているため，実現可能な分析パターンのバリエーションが少ない事が主な要因であると考えられる。

可視化キューブを採用したシステムにおいて，利用可能な可視化表現を増やすことは容易であり，3.3節で後述するように，本論文で提案するシステムでは5種類の可視化表現を利用可能としている。

後者の問題に関して，見通しの良いインタラクションを実現可能という，可視化キューブの利点を減じることなく分析パターンのバリエーションを増やすために，2.2節に示した広義比較分析の観点を導入する。

2.3節に示した既存研究では，「比較すべき差異のタイプ」はシステムによって固定されており，その差異をいかに把握しやすくするかが支援の対象となっている。異種の可視化表現を利用する場合でも，どの可視化表現間で比較を行わせるかはシステム側で定められている。しかし，EDAにおける探索空間を拡大するためには，比較すべき統計データの組合せや可視化表現の組合せなど，比較すべき差異のタイプをユーザ自身が選択できる自由度を増やすことも重要と考える。これは，2.4節で述べたプリミティブ操作を組み合わせながら試行錯誤，探索的行動を行うという可視化キューブの特性を活かせるアプローチでもある。

可視化キューブでは，時間軸，空間軸，統計データ軸，可視化表現軸で構成される4次元キューブを想定し，こ

の上で可能なインタラクションを定義している。広義比較分析の観点から眺めれば，可能な差異は動向軸の違い（時間軸あるいは空間軸），統計データ軸上あるいは可視化表現軸上の違いの3種類，およびその組合せである。これらは，可視化キューブを二つ用意し，両者の差異を考える事で可能な組合せを列挙することが可能である。この中で，EDAに有効な差異のタイプを明らかにし，それを支援する機能を実現する事ができれば，動向情報分析におけるユーザの探索空間拡大に効果があると考えられる。

本論文では，差異のタイプに関して自由度の高いインタフェースを提供することにより，EDAにおいて実際に利用される比較分析行為について明らかにすることに焦点を当てる。このためには，可視化表現などの種類によらず，プリミティブ操作を組み合わせた広義比較分析を統一的に実行可能なインタフェースが必要であることから，本論文では可視化キューブを同時に二つ使用し，2画面表示を可能とするシステムを提案する。なお，2.3節で示したビュー間の同期や遷移アニメーションなどは，差異を明示するためのアプローチであり，差異のタイプに自由度を与える本論文のアプローチと組み合わせる事は可能である。例えば，一方の画面でビューを生成した際に，分析に有効な差異のタイプを推定し，他方のビューを自動生成すると共に，ビュー間の対応関係を明示するを行った2画面連動機能の実現が考えられる。このような連動機能については，本論文で示す評価実験の結果に基づき4.4節で考察する。

次節では，本論文で対象とする広義比較分析を可視化キューブの観点から整理し，可能な分析操作それぞれについて考察する。広義の比較と区別するために，2.4節で示したプリミティブ操作としての比較を以降では「狭義比較」と呼ぶ。これは，「異なる統計データ間，時点間，地点間などについて，同一軸上での差異を示す可視化表現を生成する操作」として定義されている。狭義比較では，異なる地点間での空間的動向比較は定義できないなどの制限があった。これは例えば，関東地方と九州地方の空間的動向を比較しようとした場合，1階層下の粒度に当たる都道府県は地方毎に異なることによる。

これに対し，本論文で対象とする広義比較は，対象間に存在する何らかの差異を，「動向軸（時間軸，空間軸）」「統計データ軸」「可視化表現軸」の3軸のうち少なくとも1軸以上が異なる状態として捉える。例えば，「昨年1年間の東京における日照時間と降水量の関係」は，統計データ軸上での差異の比較に相当する。また，「各地方のCO₂の違いを棒グラフ，円グラフの両方で可視化」した場合は，異なる可視化表現間での差異を比較する事により，差（棒グラフ）と比率（円グラフ）の両方を考慮した分析を行うことに相当する。

表 1 本論文で対象とする広義比較分析の分類

	動向軸	統計データ軸	可視化表現軸	対応する操作
変更なし	同一	同一	同一	-
1 軸変更	相違	同一	同一	スピン/DL
	同一	相違	同一	比較
	同一	同一	相違	遷移
2 軸変更	相違	相違	同一	スピン/DL +比較
	相違	同一	相違	スピン/DL +遷移
	同一	相違	相違	比較+遷移
3 軸変更	相違	相違	相違	-

3.2 比較分析操作の分類

前節で述べた通り、可視化キューブでは、探索的分析における操作を「動向軸（時間軸，空間軸）」「統計データ軸」「可視化表現軸」という 3 軸を用いて体系化しており、提案手法ではこれらの変更を組み合わせた比較分析が定義可能である。可能な組合せについて表 1 にまとめる。表において、「同一」は 2 画面間で同じ状態、「相違」は異なる状態にあることを示す。また、操作名において DL はドリルダウン / アップを意味する。可能な組合せは全部で 8 通り存在するが、3 軸全てが同一となる場合は差異が存在せず、全て相違となる場合は共通性がないため両方とも比較分析にはならないと考える。以下では、残る 6 種類の組合せについて、比較分析の観点からその意味について考察する。

§ 1 1 軸変更に対応した比較分析

3 軸の内 1 軸だけが異なる場合は、可視化キューブにおける 1 回のインタラクションに相当する。すなわち、動向軸のみが異なる場合は、スピン操作やドリルダウン / アップ操作の適用前後のビューを比較することに相当する。同様に、統計データ軸のみが異なる場合は比較、可視化表現軸のみが異なる場合は遷移にそれぞれ相当する。従来の 1 画面表示の場合、操作前のビューは失われてしまうが、2 画面表示により操作前後のビューを並べて比較する事が可能となるため、ユーザの認知的負荷軽減に貢献すると考える。

動向軸が異なる状態の比較は、空間的動向と時間的動向の両者を同時に確認したい場合や、異なる粒度による動向情報の比較などに有効と考える。

可視化表現軸が異なる状態の比較は、特性が異なる可視化表現で同一のデータを眺めることに相当する。これにより、大きさの違いを棒グラフで確認しつつ、割合は円グラフで確認するといった分析が可能となる。図 2 は、1998-2000 年における CO₂ の時間的動向を地方間で比較した例であるが、左画面は棒グラフ、右画面は円グラフをそれぞれ用いている。棒グラフにおいて、それぞれの棒が地方に対応する。棒グラフより、CO₂ の量は地方

図 2 可視化表現のみを変更した比較

毎に大きく異なるが、円グラフより、年毎の変動は各地方ともほとんどないことがわかる。

統計データ軸が異なる状態の比較は、統計地図など狭義比較操作の実装が困難な可視化表現においても、異なる統計データや時点に対するビューを並べることで比較が可能になる。また、棒グラフなど従来システムでも比較操作が可能であった可視化表現においても、多数のデータを比較したい場合に 2 画面に分割して描画することにより、ビューが煩雑になるのを防ぐ効果が期待できる。さらに、3.1 節で述べた、狭義の比較が定義不可能な空間軸上での空間的動向比較も、2 画面を並べて比較する事で広義の比較を行う事が可能となる。

§ 2 2 軸変更に対応した比較分析

表 1 に示すように、2 軸が異なる状態を比較する事は、従来システムでは二つの操作を実行した前後の状態比較に相当するが、提案システムでは、1 軸変更と同じくビュー間の比較として扱われる。以降では、2 軸変更に対応した比較分析を以下の様に命名する。

可視化表現固定比較 動向軸と統計データ軸が異なる比較

統計データ固定比較 動向軸と可視化表現軸が異なる比較

動向固定比較 統計データ軸と可視化表現軸が異なる比較

統計データ固定比較は、ある統計データの空間的動向情報と時間的動向情報を同時に見比べながら分析したい場合に用いる事で、各動向情報の可視化に適した特性を持つ可視化表現をそれぞれ利用可能になる。動向固定比較は、異なる統計データに関する同一の動向情報を比較したい場合に用いる事で、各統計データの可視化に適した特性を持つ可視化表現が利用可能になる。

一方、可視化表現固定比較は、同一の可視化表現を用いて、異なる統計データ、動向を比較する作業に相当する。EDA における役割は一見明確ではないため、あまり利用されないことが想定されるが、実際の利用状況については次章の評価実験にて考察する。

3.3 2画面表示による探索的分析支援システム

可視化キューブを2つ使い、2画面表示を採用した時空間的動向情報の探索的分析支援システムをJavaアプレットとして構築した。構築したシステムのスクリーンショットは図2に示した通りである。画面下部にある操作パネルより、操作対象とする画面(右・左)を指定し、2.4節に示した可視化キューブに対する操作をプルダウンメニューから選択し、対応するボタンを押すことで実行する。後述する実験において、実験協力者の多様な分析行為を観察可能とするために、現状では画面間の連動機能は実装していない。しかし、操作パネルは両画面に共通しているため、比較分析のために一部を変更したビューを生成することは容易に行えるようになっている。

分析対象とするデータは任意に追加可能だが、次節に示す評価実験には、以下に示す日本全国の統計データを用いている。収集期間はいずれも1998-2005年である。

- 人口: e-Stat (総務省統計局) *2
- CO₂, 窒素酸化物, メタン, 水質汚染度: 環境数値データベース (国立環境研究所) *3
- ガソリン価格: 石油情報センター (財団法人日本エネルギー経済研究所) *4
- 降水量, 積雪量, 日照時間, 平均気温, 風量: 過去の気象データ検索 (気象庁) *5
- ビール消費量: 酒のしおり (国税庁) *6

可視化表現は、統計地図、棒グラフ、折れ線グラフ、円グラフ、面グラフに対応しており、統計地図以外はグラフ描画用ライブラリ JFreeChart*7により実装している。統計地図は、既存システム [高間 10] と同じく SVG を用いて描画している。

4. 評価実験

4.1 実験の概要

提案手法により、探索空間が拡大されることを検証するために、実験協力者による評価実験を行った。評価実験の目的は、3.2節に示した各広義比較操作が、EDAにおいて実際にどの程度用いられるかを調査すること、および有効な2画面連動機能について考察するために、2画面を利用した分析パターンについて調査することにある。そのために、工学系大学生及び大学院生14名に協力を依頼し、2人1組となって作問・解答形式のタスクを行ってもらった。作問者には、3.3節の提案システムを用いて統計データを分析し、システムを用いて解答可能な問題を作成してもらった。解答者には、問題の解答となるビューを生成してもらい、その後出題者と解答者で

解答の正誤について確認してもらった。一組につき、作問・解答を交互に行いながら、計4回行ってもらい、その操作ログの分析を行った。

作問者は、事前に明確な目的を持たず、データを多様な観点で分析しながら、出題にふさわしいビューを探索する必要がある。一方、解答者も、解答となるビューは事前にわからないため、あいまいな目標状態を想定しながら試行錯誤を繰り返す必要がある。以上より、本実験の結果収集される操作ログを分析することにより、EDAにおける比較分析の利用状況が明らかになることが期待できる。

4.2 比較分析タイプの分析

表2は、7組(A~G)の実験協力者それぞれにおいて、表1にまとめた各分析操作の利用頻度をまとめたものである。表において「動向変」「統計変」「可視化変」は1軸変更に対応した比較分析であり、それぞれ該当する軸のみが異なる状態での比較分析を行った場合を示す。また「動向固」「統計固」「可視化固」は2軸変更に対応した比較分析を行った場合である。「0軸」は2画面が同じ状態の場合、「3軸」は全て異なっている状態を指す。

各組で操作回数(表中の合計)にばらつきがあるが、利用割合が20%以上となっている操作を太字で示している。E組において「0軸」の合計頻度が高くなっているが、操作がよく理解できていない段階で、操作ボタンを連打した実験協力者がいたためであり、それを除けばほとんど利用されていなかった。

表より、ほとんどの組において、統計データ軸のみ変更した比較操作が他よりも頻繁に利用されていることがわかる。また、2軸変更に対応した比較分析でも、統計データ固定による比較分析が他よりも少ないことから、統計データ軸に対する操作が最も利用されていると言える。この理由として、作問タスクでは異なる統計データ間の関係を問題にする場合が28問(4問×7組)中22問と多かった事が挙げられる。統計データ変更比較の用途として、3.2.1節では「統計地図など狭義比較操作の実装が困難な可視化表現での比較」「多数データの比較」を挙げていた。前者については、統計地図を用いた統計データ変更比較が実験協力者7組中6組で利用されており、想定通りの結果が得られた。多数データの比較については次節で考察する。

次に変更操作が多く適用された軸は動向軸であるが、これはドリルダウン・アップ操作が多用されたことによるものであり、3.2.1節で想定した用途の一つ、異なる粒度による動向情報の比較が実際に利用されていることを示している。同じ統計データであっても、粒度を変更することにより動向として異なる傾向を示すため、作問において重要であったと考える。

1軸変更に対応した比較分析では、異なる可視化表現を用いた比較が他よりも使用頻度が少なく、2軸変更

*2 <http://www.e-stat.go.jp/SG1/estat/eStatTopPortal.do>

*3 <http://www.nies.go.jp/igreen/index.html>

*4 <http://oil-info.ieej.or.jp/index.html>

*5 <http://www.data.jma.go.jp/obd/stats/etrn/index.php>

*6 <http://www.nta.go.jp/shiraberu/senmonjoho/sake/shiori-gaikyo/mokuji.htm>

*7 <http://www.jfree.org/jfreechart/>

表 2 比較分析タイプ毎の利用頻度の比較

組	0 軸	動向変	統計変	可視化変	動向固	統計固	可視化固	3 軸	合計
A	7	35	10	8	4	28	10	9	111
B	5	6	70	12	19	19	52	21	204
C	2	0	28	0	4	0	3	12	49
D	4	2	59	1	34	0	25	1	126
E	44	67	50	16	12	16	50	9	264
F	5	6	32	0	0	0	5	0	48
G	4	10	36	4	32	2	16	12	116
合計	71	126	285	41	105	65	161	64	918

対応した比較分析では、可視化表現を固定した比較の使用頻度が他よりも高くなっている。このことより、異なる可視化表現を用いた比較操作は他よりも少ない傾向にあることがわかる。しかし、動向固定による比較分析は比較的多く利用されているため、統計データの特性毎に適する可視化表現が異なる可能性が示唆される。また、3・2・2 節において、ほとんど利用されない想定していた可視化表現固定比較が比較的に利用されていたが、この理由については次節で考察する。

以上より、1 軸変更操作だけでなく、2 軸変更操作によるビューの比較も実際によく利用されていることがわかる。

4.3 解答ビューの分析

作問タスクにおいては、探索途中で多様な分析が行われ、その中には分析に有益であったものとそうでなかったものが混在していることが想定されるが、最終的に出題あるいは解答として選択されたビュー（解答ビュー）は、何らかの有益な知見を持ったものであると考えられる。また、解答ビューに利用された分析パターンは、探索的分析プロセス全体でも主流であったと考えられるため、これを手がかりとして分析行為を考察する。表 3 に作問者、解答者それぞれの解答ビューの比較結果を示す。

表において、作問、解答は作問者、解答者の解答ビューで用いられている比較分析タイプをそれぞれ表す。表記は表 2 と同様であるが、「1 画面」は、1 画面のみを利用して解答ビューを作成していることを表す。また、解答者、作問者の解答ビューが一致している場合は「一致」列に

を、利用している可視化表現のみが異なっている場合には を記している。

表 3 より、解答ビューでは統計データ変更が最もよく用いられていることがわかるが、このことは表 2 で統計データ変更が最もよく用いられていることと対応している。統計データ変更を用いた出題例を以下に示す。

- (1) 1998 年から 2002 年において日照時間と降水量に
関係があるか？(C 組第 2 問)
- (2) 関東地方の中で毎年、人口と CO₂ が最下位なの
はどこ？(F 組第 3 問)

- (3) 2003 年から 2005 年において地方毎の水質汚染度、
CO₂、メタンの関係は？(E 組第 3 問)

(1) の解答ビューを図 3 に示す。このタイプが最も多く見られた問題であり、解答ビューでは左右の画面でそれぞれ一つの統計データが描画されている。比較的簡単な出題パターンであるため多用されたと考える。C 組は全てこのタイプの出題であり、両者の解答ビューもほぼ一致している。また、(2) のタイプは一見 (1) よりも複雑な問題に感じるが、解答ビューとしては (1) と同様である。F 組も 4 題中 3 題が (2) のタイプであり、両者の解答ビューも一致している。表 2 において、C、F 組の操作回数は他よりも少ないことから、これらの組ではあまり深い探索が行われなかったと考えられる。

(3) のタイプは、3 種類の統計データについて描画することが必要なため、少なくとも一方の画面に複数の統計データを描画することになる。そのため、既存システムの比較操作に対応する操作を同一画面に対して適用する必要があり、(1)(2) よりも複雑な操作となる。3・2・1 節において、統計データ変更比較の用途の一つとして多数データの比較を想定していたが、ビューの視認性の観点のみではなく、ビュー生成の操作の容易さも利用される要因の一つと考えられる。

1 画面による解答ビューも比較的によく用いられているが、代表的な出題例を以下に示す。

- (1) 関東地方で 1 番人口が多い都道府県は？(A 組第 1 問)
- (2) 降水量が 2 番目に低い地方の人口が 2 番目に多い都道府県は？(A 組第 4 問)

(1) のタイプは、2 画面に分けて描画の必要がないタイプであり、最も単純と言えるためあまり見られなかった。

1 画面を用いた場合には、(2) のタイプに属するもののように、厳密には 1 画面で表現できないものの方が多かった。(2) の解答者の解答ビューを図 4 に示す。この様に、降水量が 2 番目に低い地方を示す画面（右画面）と、その地方（甲信地方）の人口動向について示す画面（左画面）が必要となるが、作問者はこれらの分析をシーケンシャルに行い、一方のみを解答ビューとしていた。これは厳密には正しい出題とは言えないが、今回の実験で

表 3 作問者・解答者による解答ビューの比較

組	作問順序	作問	解答	一致
A	1	1 画面	1 画面	
	2	1 画面	動向変	
	3	1 画面	1 画面	
	4	1 画面	可視化固	
B	1	統計変	統計変	
	2	動向固	統計変	
	3	統計固	統計固	
	4	1 画面	統計変	
C	1	統計変	統計変	
	2	統計変	統計変	
	3	統計変	統計変	
	4	統計変	統計変	
D	1	統計変	統計変	
	2	動向固	統計変	
	3	1 画面	動向固	
	4	統計変	統計変	
E	1	1 画面	1 画面	
	2	動向固	統計変	
	3	統計変	統計変	
	4	1 画面	1 画面	
F	1	統計変	統計変	
	2	動向変	1 画面	
	3	統計変	統計変	
	4	統計変	統計変	
G	1	1 画面	1 画面	
	2	統計変	統計変	
	3	動向固	統計変	
	4	動向固	統計変	

は比較的良好に見られた出題であった。両者の違いは、作問者がシーケンシャルに行った分析を、解答者は2画面に並べて行った事に相当し、提案手法による「重ねる比較」から「並べる比較」への拡張を反映したものと言える。これと同様な作問者と解答者の違いは、B, D, F組（ただし、F組の場合は解答者が1画面を利用）でも見られる一般的なものであった。

2軸変更が解答ビューで利用された頻度について表4にまとめる。表2では可視化表現固定比較が最も用いられていたが、解答ビューではほとんど用いられていないことがわかる。3.2.2節では、可視化表現固定比較はあまり利用されないと想定していたが、上述の様なシーケ

表 4 2軸変更による解答ビューの分類

比較分析	作問	解答	計
動向固	5	1	6
統計固	1	1	2
可視化固	0	1	1

図 3 C組第2問の解答ビュー

図 4 A組第4問の解答ビュー（解答者）

ンシャルな分析を行う際に利用されることが観測された。これらは分析の過程で利用されることが多いため、解答ビューではほとんど用いられなかったと考える。

また、最も利用されている動向固定比較は、表3よりB, D, E, G組で出題・解答に用いられているが、この結果は表2の利用頻度とも対応していることがわかる。

動向固定比較を用いた可視化ビューを図5に示す。これは、「2001年から2005年において都道府県別で見た関東地方の窒素酸化物とガソリン価格の関係は？」という出題（G組第3問）に対する作問者の解答ビューである。この例では、ガソリン価格は折れ線、窒素酸化物は円グ

図 5 G組第3問の解答ビュー（作問者）

ラフで描画している．多くの場合，可視化表現の選択は必然ではなく，分析者の好みで選択することができる．このため，このタイプの出題では，解答者が作問者と異なる可視化表現を選択することも多い．このことは，前節で述べたように，動向固定比較が比較的好く用いられていることからわかる．表 3 において，解答者が統計軸変更比較となっているのは，両統計データとも同じ可視化表現を用いていることを意味しており，作問者・解答者間の解答ビューにおける内容的な差はないと言える．

一方，統計データ固定比較は B 組第 3 問でのみ用いられており，その出題は「メタンの変化が一番大きい地方の変化の推移は？」というものであった．この問題では，メタンの変化が一番大きい地方を示すために，地方を粒度とした空間的動向を一方の画面で行い，該当する地方（東海地方）について，都道府県を粒度とした空間的動向を他方に表示している．この分析も，既存システムではシーケンシャルに行っていた分析を並べる比較で実行したものである．

以上より，多様なタイプの出題がなされていることから，提案システムは多様な分析を支援していることがわかる．さらに，同一の問題に対し作問者，解答者で異なるビューを作成していることから，画一的な分析手段を提供するのではなく，分析者毎の多様な分析行為を支援できていることがわかる．

4.4 2 画面連動機能に関する考察

実験結果より，頻繁に利用される比較操作がある一方，あまり利用されないものも存在することがわかる．また，多様な問題が作成される中でも，ある程度共通した分析パターンが存在することも明らかとなった．探索の自由度を高くすることで実験協力者による多様な分析行動を収集するため，現状のシステムでは画面間の連動機能は実装されていないが，実験により得られたこれらの知見に基づき適切な連動機能を実装することで，有効な探索方向にユーザを誘導することによる支援が期待できる．

表 2，表 3 より，統計データ変更比較が多用されていることから，異なる統計データ間の比較を支援する事が考えられる．これは「ある統計データと興味深い差異を持つ統計データを発見する」プロセスの支援と「二つの統計データ間の差異，共通点を発見する」プロセスの支援に大別できる．

実験協力者の操作ログにおいても，動向軸，可視化表現軸を固定して，統計データのみを連続的に変更していく分析過程が複数の実験協力者で観測されたが，これは前者のプロセスに相当すると言える．これについては，一方の画面で生成したビューに連動して，統計データのみを変更したビューを他画面に自動生成する機能を導入することによる支援が考えられる．

後者に関する連動機能は，統計データ変更，動向固定あるいは可視化表現固定比較のそれぞれについて検討す

ることが可能である．統計データ変更に基づく連動機能は，左右画面に異なる統計データを割当て，統計データ軸以外に対する操作を両画面同時に適用することが考えられる．一方，動向固定比較あるいは可視化表現固定比較の場合は，変更対象となる軸が一つとなる分変更の選択肢が限定されるため，他方のビューを自動生成する機能を実装することが可能と考える．例えば，ユーザが気温とビール消費量の関係について着目し，各地方の状況を知りたい場合，気温は差を把握するのに適する棒グラフ，ビール消費量は割合を把握するのに適する円グラフを用いてビューを自動生成するなどが考えられる．これを実現するためには，統計データ毎に適した可視化表現などについて，理論的および実際の利用状況の両面から明らかにしていくことが今後必要と考える．

4.3 節の最後に示した通り，統計データを固定した分析も，統計データ間の比較より頻度は少ないが行われており，これを支援する連動機能も可能である．該当する比較操作は動向変更，可視化表現変更，統計データ固定比較になるが，表 2 に示した使用頻度を考慮すると，動向変更比較について支援する事が重要と考える．すなわち，ある（複数）統計データについての空間的（時間的）動向が一方の画面で生成された場合に，他方の画面で異なる動向を示すビューを自動生成する．この時，使用されている可視化表現の特性により該当するビューを生成できない場合が発生することが想定されるため，他の可視化表現に自動的に切り替えるなどの機能が必要と考える．

5. ま と め

本論文では，可視化キューブに基づく時空間的動向情報の探索的分析支援システムの探索空間を拡大するために，可視化キューブ上に対するプリミティブ操作の組合せとして広義比較操作を定義し，それらを統一的に実行可能なインタフェースとして，2 画面表示インタフェースを提案した．提案手法では，可視化キューブを二つ同時に操作することにより，複数軸を変更したビューを並べての比較が可能になる．これにより，EDA において本質的な，広義の比較分析を支援する事が可能となる．

本論文では，提案手法により可能になる比較分析を 6 通りに分類し，分析における役割について考察した．また，プロトタイプシステムを実装し，作問・解答タスクに関する実験を行った結果に基づき，各比較分析の使用割合や分析における役割について考察した．実験結果より，1 軸変更操作だけでなく，2 軸変更操作によるビューの比較も実際によく利用されていることを示した．また，解答ビューを分析し，多様なタイプの出題がなされているだけでなく，同一の問題に対し作問者，解答者で異なるビューを作成するなど，多様な分析が行われていることを示した．

今後は，4.4 節で示した連動機能の実装などによる支

援機能の向上が挙げられる。連動機能は単なる操作効率の向上だけでなく、ナビゲーションなどの知的・能動的支援につながるものであり、今後操作ログの分析をさらに進め、有益な分析操作系列や、分析に行き詰まった状態などの検出手法についても検討する必要がある。また、現在は可視化キューブをインタラクションモデルの設計のみに用いているが、ユーザのメンタルモデルとしても有効である事が期待されることから、可視化キューブを明示的にしたインタフェースの実現なども今後の検討課題である。

謝 辞

本研究の一部は、国立情報学研究所との共同研究によるものである。

◇ 参 考 文 献 ◇

- [Chen 09] M. Chen, et al.: Data, Information, and Knowledge in Visualization, IEEE Computer Graphics and Applications, Vol. 29, No. 1, pp. 12–19, 2009.
- [Chi 98] E. H. Chi, J. Pitkow, J. Mackinlay, P. Pirolli, R. Gossweiler, S. K. Card: Visualizing the Evolution of Web Ecologies, CHI'98, pp. 400–407, 1998.
- [Gruhl 05] D. Gruhl, R. Kumar, J. Novak, A. Tomkins: Predictive Power of Online Chatter, SIGKDD, pp.78–87, 2005.
- [Heer 07] J. Heer, G. G. Robertson: Animated Transitions in Statistical Data Graphics, IEEE Transaction on Visualization and Computer Graphics, Vol. 13, No. 6, pp. 1240–1247, 2007.
- [加藤 04] 加藤, 松下, 平尾: 動向情報の要約と可視化に関するワークショップの提案, 情報処理学会自然言語処理研究会, 2004-NL-164(15), pp. 89–94, 2004.
- [Kato 08] T. Kato, M. Matsushita: Overview of MuST at the NTCIR-7 Workshop –Challenges to Multi-modal Summarization for Trend Information–, Proc. NTCIR-7 Workshop, pp. 475–488, 2008.
- [Keogh 06] E. Keogh, L. Wei, X. Xi, S. Lonardi, J. Shieh, S. Sirowy: Intelligent Icons: Integrating Lite-Weight Data Mining and Visualization into GUI Operating Systems, ICDM06, pp. 912–916, 2006.
- [Liu 07] Y. Liu, X. Huang, A. An, X. Yu: Arsa: A sentiment-aware model for predicting sales performance using blogs, Proc. SIGIR07, pp. 607–614, 2007.
- [増井 04] 増井: 情報視覚化の最近の研究動向, 情報の科学と技術, Vol. 54, No. 11, pp.558–567, 2004.
- [松下 06] 松下, 加藤: コンテキスト保持による探索的データ分析支援の枠組, 知能と情報, Vol. 18, No. 2, pp. 251–264, 2006.
- [Motoda 02] H. Motoda ed.: Active Mining, IOS Press, 2002.
- [Mozzafari09] E. Mozzafari, A. Seffah: From Visualization to Visual Mining: Application to Environmental Data, ACHI2008, pp. 143–148, 2008.
- [Nadamoto 03] A. Nadamoto, K. Tanaka: A Comparative Web Browser(CWB) for Browsing and Comparing Web Pages, WWW2003, pp. 727–735, 2003.
- [Ohsawa 08] Y. Ohsawa, T. Ito, M. Kamata: Kamishibai KeyGraph as Scenario Map Visualizer for Detecting Transient Causes from Sequential Data, Proc. of PAKDD 2008 Working Notes of Workshops on Data Mining for Decision Making and Risk Management, pp. 272–283, 2008.
- [奥岡 09] 奥岡, 片上, 新田, 研究分野ネットワークによる学術分野動向の解析と評価, 人工知能学会研究会資料, SIG-KABS-A803, pp. 21–26, 2009.
- [Sarma09] A. Sarma, L. Maccherone, P. Wagstrom, J. Herbsleb: Tesseract: Interactive Visual Exploration of Socio-Technical Relationships in Software Development, ICSE'09, pp. 23–33, 2009.
- [清水 07] 清水, 片上, 新田, 文献情報に基づく分野間のネットワーク分析, 電子情報通信学会知能ソフトウェア工学研究会,

Vol.106, No.473, pp.67–72, 2007.

- [白山 06] 白山: 知的可視化, 計算力学レクチャーシリーズ, 丸善, 2006.
- [高間 10] 高間, 山田: 時空間的動向情報の探索的分析を支援するインタラクティブな情報可視化システム, 人工知能学会論文誌, Vol. 25, No. 1, pp. 58–67, 2010.
- [Takama 10] Y. Takama, T. Yamada: Application of Visualization Cube to Analysis of Interaction Pattern in Exploratory Data Analysis of Spatiotemporal Trend Information, ISCIA2010, pp. 85–93, 2010.
- [統計委 05] 日本学術会議統計学研究連絡委員会, 知識創造社会に向けた統計教育の推進について, 日本学術会議報告, 2005.
- [椿 07] 椿: 発見・選択・検証の架橋, 人工知能学会誌, Vol. 22, No. 5, pp. 640–647, 2007.
- [Weaver10] C. Weaver: Cross-Filtered Views for Multidimensional Visual Analysis, IEEE Trans. on Visualization and Computr Graphics, Vol. 16, No. 2, pp.192–204, 2010.

〔担当委員: 阿部 明典〕

2011 年 1 月 17 日 受理

著 者 紹 介

高間 康史(正会員)

1994 年 東京大学工学部電子工学科卒業。1999 年同大学院博士課程修了。1999–2002 東京工業大学大学院総合理工学研究科助手, 2002–2005 東京都立科学技術大学助教授, 2005 年より首都大学東京システムデザイン学部准教授。博士(工学)。Web Intelligence や情報可視化, 知的インタフェースの研究に従事。主要著書は「インテリジェントネットワークシステム入門」(コロナ社)。IEEE, 日本知能情報ファジィ学会, 情報処理学会, 電子情報通信学会各会員。

狩野 真次

2009 年首都大学東京システムデザイン学部卒業。2011 年同大学院システムデザイン研究科博士前期課程修了。現在株式会社バンダイナムコゲームスに勤務。在学中は、情報可視化システムの研究に従事。