

可視化から何がわかるのか

What can we extract from the visualization?

白山 晋¹

Susumu Shirayama¹

¹東京大学人工物工学研究センター

¹RACE, the University of Tokyo

Abstract: Owing to the volume of data generated in recent computations and experiments, it is quite difficult to extract useful information from these data even if using scientific/information visualization techniques. Method or methodology to extract useful information from such data should be considered. Several concepts of very large scale visualization are proposed in this situation. Most of them are based on high-performance computing techniques or highly-efficient devices for computer graphics. In this paper, first, we insist on the necessity of adding cognitive models to the visualization process. And then, a flexible visualization methodology based on “*post visualization process*”, which includes a human recognition process and quantitative evaluations of visualized results is introduced. Finally, a possibility that a visualization agent designed from a process model helps to reduce the difficulty of handling huge data is described.

1. はじめに

データ、情報、知識を、見えるようにする、あるいは、わかりやすく表現するという要求は様々な状況で生じている。人間は、情報の多くを視覚から得るためにそのような要求は増し、様々なシステムにおいて視覚情報の果たす役割が注目されている。

可視化プロセスは、“みえないものを見るようにする”、“みえない関係性を顕在化する”プロセスとされるが、単に、データを視覚情報に変換することを、データの視覚化、あるいは可視化と呼ぶことも多い。なお、実験・観測・計算からのデータの可視化（サイエンティフィックビジュアリゼーション、SV: Scientific Visualization）に関しては、文献[1]での定義が定着している。

一方、（可視化する）対象によって目的も、背景も様々である。“見えるようにする”という要求自体をわかりやすく示すという目的もある。また、切実な要求のもとに行われる可視化や、何かしらの思想を持って行われる可視化もある。可視化は多様化し、技術論だけで一括りにすることはできなくなっている。例えば、取り調べの可視化[2]においては密室での出来事を白日の下に曝す必要性を述べている。その方法は、取り調べの全過程の録画・録音という直接的なものであり、近年の情報機器の発展をもってすれば実現は難しくない。この場合は、そのような対象に対して可視化を行うこと自体の意義が問われる。一方、トレーサー粒子の可視化を利用したPIVのような計測手法においては、3次元データの高精度な取得が課題であり、可視化技術の向上に意義がある。

これらの例は目的が明確であり、「可視化から何がわかるのか」という点においては、「その目的の

ために可視化を行えば、これこれがわかる」と言及できる。しかしながら、このような可視化の効果が明確に示される例は減少している。言い換えれば、視覚情報のもつ説得力のみに着目しているケース（可視化をすれば何か示せるだろう、という程度）が増えている。例えば、“見える化”、“診る化”、“魅せる化”などという造語とともに散見するものの多くは、今までに扱われなかった対象に対して何かしらのデータ化を試み、単に視覚情報化するというものである。確かに視覚情報には分かりやすさというメリットはある。しかし、対象を数値データや数理モデルなどに対応付ける部分に本質があり、敢えて可視化を強調する必要がないものが多い。ただ、正否はともかく、可視化自体の普及を促し、可視化事例が急増し、可視化の課題を明らかにしていることも事実である。

一般に、データの可視化は、情報、あるいは知識の抽出過程とされ、得られた可視化結果は、普遍的な情報や知識そのものを表すものと考えられてきた。しかし、可視化操作によって生データが視覚情報に変換される際にユーザーの経験、あるいは知識が利用され（例えば、ユーザーの判断にもとづくデータの選択、あるいは選別、情報集約）、可視化結果を読み解くというユーザーの解釈の中に暗黙知が存在する[3]。

可視化は探索（発見を含めて）と分析（解釈や利用を含めて）のプロセスに分けられるが、両プロセスにおいてユーザーに起因する曖昧性が含まれる。また、データの大規模化とともに、探索と分析に要する負荷が大きなものになっている。このため、ユーザーの解釈に依存する部分を明らかにするような方法・方法論や、可視化プロセスの半自動化が研究対象になっている[4]。これらの研究の中で最も一般

的な方法は、興味領域 (ROI: Region(s) Of Interest(s)) と詳細度 (LOD: Level(s) Of Detail(s)) にもとづく探索である。具体的には、特徴量にもとづきROIを定め(あるいは示唆し)、細部を調べるといったものである。また、クラスタリングなどを利用して情報に階層的な構造を持たせることが多い。階層構造によって階層毎の情報の把握は容易になる。

この方法では、詳細度にもとづいて予め階層化された情報 (LOD) に対して、指標を定めてROIのいくつかの候補を示し、人間の判断によって階層間を移動する[5]。階層間の移動において、詳細度の低い階層のある起点 (基点) から高い階層へ (ドリルダウン) とその逆 (ロールアップ) が、インタラクティブに行われる。このインタラクティブなプロセスに、人間が可視化結果を理解し、判断 (意思決定) とともにプロセスを進めるといったユーザーの解釈に依存する部分が含まれる。

本稿では、可視化という行為や可視化結果から何がわかるのかという問題意識を起点として、インタラクティブに進められる可視化プロセスにおいて、可視化結果 (画像) がどのように認知されるかという問題と、プロセス自体の情報化と構造化によってプロセスを半自動化できるかという問題について考察する。

2. 可視化の任意性

2.1 可視化の現状

可視化技術 (手法、システムなど) は一部を除き成熟し、可視化手法に関していえば、大部分は90年代前半に完成の域に達したものである。その後、計算環境の多様化や、可視化情報自体の肥大化のために、可視化研究の目的は、データの大規模化と分散化への対応、可視化結果の解釈の支援といったものに移行する。さらに、デジタルエンジニアリングなどの統合システム中での可視化結果の利用技術が議論されている。このような中で、テラバイト、ペタバイトといったデータの可視化がまだまだ難しいことは確かであり、研究としては直接的なもの (計算環境への対応) が多い。

一方、なぜ可視化するのかという動機、あるいは目的について深く考えることもまた難しい。

可視化の一番の目的が「みえないもの、みえない関係を見えるようにすること、分かり易く示すこと」であることは自明のようである。例えば、数値をグラフ化すれば分かり易い、色を付ければさらに分かり易くなることは直感的に納得できるだろう。80年代のはじめには、そのようにわかっている、グラフ化や色付けが難しかった。今では、それが簡単に (意識せずに) できるようになっている。データがあれば、それを視覚的にみるということはルーチンワーク化している。したがって、数値データの

羅列を眺めるということや、分析対象に応じて可視化の方法を工夫するということは少なくなっている。

しかし、積極的に (意識的に)、データの背後にあるものが何であるのか、それを明らかにしようとするれば、工夫が必要になる。例えば、流れの可視化において、「トレーサー粒子を利用する」と思いついても、どんな粒子でもいいのか、どの部分から流せばよいのか、さらには、粒子が流れることによって流れ自体を変えてしまわないか等の検討が必要である。そもそも、本当に流れを見ているのかも考えないといけない。

このようなことは、人間の技量 (知識) に依存するが、可視化自体のもつ任意性に起因するものでもある。この任意性が特に情報可視化の研究課題の一つになっている。

2.2 可視化の任意性と人間の理解

はじめに、情報可視化の特徴を述べておく。サイエンティフィックビジュアライゼーション (SV) との違いを端的にいうと、表示する際の座標軸の扱いが異なることである。SVでは、(数値) データが、2次元、あるいは3次元の物理座標に対応した表示空間において表現される。一方、情報可視化では、物理座標とは対応しないデータ、さらには4次元以上のデータを扱うことが少なくない。このため、データを選別・加工して、2次元、あるいは3次元表示空間に対応付ける必要がある。例えば、身長、体重、胸囲、座高というデータを扱う場合、それぞれの要素を独立変数として扱い、4次元空間を形成する。次に表示空間を踏まえたデータを選別・加工を行う。ここに任意性が生じ、意思決定が必要になる。

一方、一連の操作の中に、人間の理解と深く関連する部分が存在し、可視化プロセスの進め方に影響する。表1に01年度の一次エネルギー総供給量を示す[6]。これを可視化したものが図1である。図1(a)~(c)は、表で示される順番でグラフにしたものである。(c)は全体の割合を示すもので、(a)と(b)とは異なる観点からのグラフである。一見すると、(a)と(c)が表1で示す数値の適切な可視化例であるように考えられる。この例のような変数間で明確な関数関係が見出されないものに対しては、(b)の折れ線グラフによる表示は適当ではない (折れ線が線形近似と見なされるため)。次に、図1(d)と(e)を見てほしい。供給量の大きい順番で表示しただけで、数値自身は同じである。同じデータであっても、表示の方法によって可視化結果から読み取れるものが異なるという例である。このように順序を変えて、同じ素材を異なるように見せる方法は、情報可視化の中で、再配置 (Rearrangement) として知られているものである [7][8]。

別の例を図2に示す。図2は原油供給量の推移を棒グラフにしたものである。左図と右図は同じデータ

表1 一次エネルギー総供給量 (単位は $10^{15}J$) [6]

石炭	原油	石油製品	天然ガス
4347	9179	2071	2987
再生可能工	水力	原子力	
563	757	2879	

に認知し解釈しているか、という認知科学に属する
ようなものである。

可視化結果の解釈という点では、「対象のどこに注目しているのか」ということは重要な要素になる。初心者と熟練者の見方の違いなど、個人差が大きいことがわかっている。例えば、将棋やチェスなどでは、ある局面から次の一手を決める場合において、初心者よりも熟練者の視線の移動量は少ないことがわかっている[11]。一方、可視化を利用した探索においては、熟練者であっても特定領域間の頻繁な視線移動が行われることが明らかになっている[12]。

図1 一次エネルギー総供給量 (単位は $10^{15}J$)

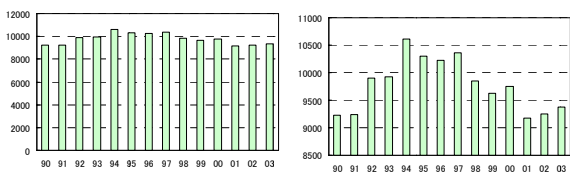
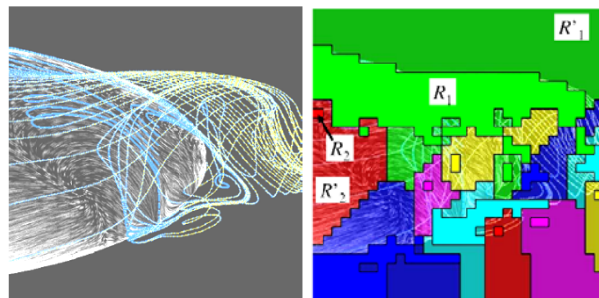


図2 原油供給量の推移 横軸は年度, 縦軸は供給量

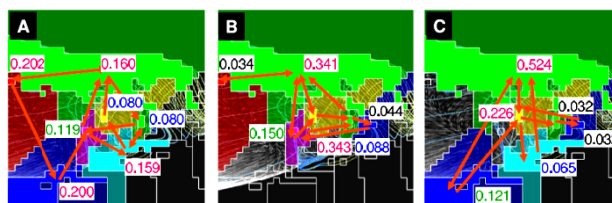
に対するものだが、縦軸のスケールを変えている。左図では、緩やかな変動と判断できるだろう。しかしながら、右図のように変化分を強調すれば、その印象は変わる。このような情報の見せ方は、*Focus + Context*と呼ばれる方法に属するものである[7][8]。

SVにおいて、見せ方によって異なる解釈を与えるようなことはあるのだろうか。例えば、2次元のカラー場の可視化手法に、等値線表示法と等値領域表示の一つであるカラーマッピングという方法がある[3]。空間分解能の低い計算結果や、数値誤差の大きい計算結果を可視化すると、等値線表示法では等値線にノコギリ状の特有なパターンが現れる場合がある。同じ計算結果をカラーマッピングで可視化すると、そのようなパターンは隠されてしまう。どちらの可視化法が良いかは、目的や、それを見るものの判断による。例えば、数値誤差を明らかにしたい場合は等値線表示法が、全体の傾向を見るためにはカラーマッピングが適しているという指摘がある。

このように、可視化には任意性が存在し、可視化結果から読み取れるものは、人間がどのように理解するかに関係する。このため情報可視化の分野を中心に、人間の理解をふまえた可視化研究が始められている[9][10]。それは人間が可視化結果をどのよう



(a) 可視化画像 (b) Saliency modelによる領域分割



(c) 視線移動のPageRankによる定量化

図3 可視化画像の見られ方

図3に、図(a)に示す可視化画像に対する3人の被験者の見方を示す。人間の視認性にもとづく領域分割法によって画像を分割し(図(b)), 注視点移動を定量的に評価したものが図(c)である(手法については文献[12]を参照されたい)。図(c)のAは流体力学の専門家の見方を示している(B,Cは初学者)。赤の数値で示す4つの領域に対してほぼ均等に視線が移動していることがわかる。全体を見ながら、いくつかの注目領域に対して比較を繰り返しながら情報を抽出していることが推測される。また、3者において見方が異なり、B,Cは専門家の注目している領域を見ていないこともわかる。我々が行っている他の視線計測でもこのような傾向は顕著である。このように、予想できることとはいえ、可視化画像から重要な情報が抽出されない可能性は常に考慮すべきであろう。

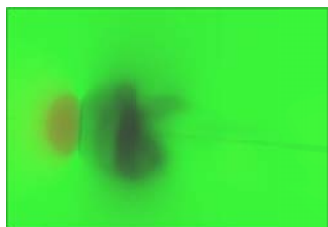
2.3 人間の理解を考慮した効率化

以上の結果は見方の重要性を示すものである。一方、可視化する人間の理解を考慮すれば、可視化自体を効率化できることもある。

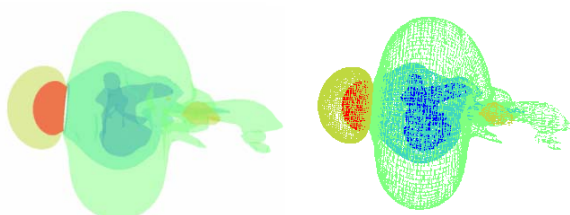
EbertとShawは、人間の視覚的認知過程に負荷をかけないようなユーザーインターフェースとステレオ

画像システムに基づいた可視化システムを構築し、そのシステムが可視化結果のより深い理解に貢献できることを示している[13]。非写実可視化と呼ばれる可視化は、可視化結果を、イラスト画のように加工し、本質的な部分を抜き出し示すものである[14]。

一般的な可視化システムを利用する際にも効率化ができる。図4は、流れに垂直に置かれた円板のまわりの圧力場を可視化したものである。



(a) ポリウムレンダリング



(b) 半透明等値面 (c) 等値面(ワイヤーフレーム)

図4 3種類の可視化法による同一のスカラー場の表現

図4(a)はポリウムレンダリング、(b)は半透明等値面、(c)はワイヤーフレームによる等値面であり、同じデータの可視化である。ポリウムレンダリングは計算負荷が大きい、スカラー場に対する最も効果的な可視化法とされる。しかし、図4に示す圧力場のようなものを可視化する場合、(b)や(c)と比べると、細かな構造の把握が難しいことがわかる。このように、可視化対象や目的によっては、ポリウムレンダリングは適切な可視化方法ではなくなる。では、(b)と(c)を比べるとどうだろう。(b)は半透明属性をもつポリゴン、(c)は線分であり、3次元操作までを考えると、データ量の面で(c)が有利である。一方、図に示すような静止画では、見え方という点で、(b)の方が優れているといえる((c)は線分の重なりがあり煩雑である)。しかし、3次元操作を行い、図を回転、縮小・拡大すれば、静止画だけでは線分の重なりの影響で識別が難しかった3次元の構造が、(b)と同等のレベルでわかるようになる。

3. 可視化エージェントへの展開

上述した可視化結果の解釈の方法や、人間の理解まで含めた可視化プロセスの検討の中で、可視化(あるいは可視化結果)の善し悪しが評価されるようになっていく。現時点では、被験者実験による統計的な評価や、利得の設定などであり[15][16][17]、我々が行っている視線移動からPageRankのような定量的指標を導出し評価に結び付ける[12]というような研

究は少ない。いずれにしろ、このような研究は端緒にすぎないといえるが、インタラクティブなプロセスの中で解釈に関連する定量的指標が導入され、評価基準の設定が行われれば、実用に向けて急速に発展する可能性がある。機械学習やエージェントによる作業の(あるいは人間自身の)支援は情報工学の展開を考えれば自然ななげだからである。

我々のグループでは、大規模計算データに対する可視化作業の負荷を軽減するために、可視化エージェントの設計と、人間と可視化エージェントの協働を検討している。

インタラクティブなプロセスという側面に注目すれば、可視化は、様々な可視化操作(手法)によってデータを視覚情報化し、膨大な量のデータの中から必要となる情報を抽出し、情報から知識に高め、さらに学習を行うものとなる。具体的には、図5に示すように可視化対象となるデータの選択の後、可視化操作(手法)の選択、可視化結果の解釈を、解析者が繰り返すことで進められる。可視化の意味づけや効果は、このサイクルの中で見いだされることが多く、このサイクルの中で学習が行われる。この一連のながれを、改めて可視化プロセスと呼ぶことにする。可視化プロセスの中で考えなければならないのは、「何のために」、「何を」、「どのように」、「どこで」見えるようにするか、「何がわかるのか」ということである。特に、「何のために」、「何を」見るか、そして「何がわかるのか」が重要である。

図5 可視化プロセス

可視化プロセスを一般的な意味での作業と考えると、目的と目的を達成するという観点が重要になる。「何のために」、「何を」が可視化の目的に相当し、「何がわかるのか」によって目的の達成度が決まる。しかしながら、可視化の多くは発見(探索)型であり、目的があいまいであることや、可視化の途中で目的が変更になることも多く、目的を同定することは難しい。

一方、可視化の目的が決められ、定型化されたものも少なくはない。定型化された可視化プロセスの場合、可視化の目的に対して、目的を達成するための方向性(方策、方針)を与える方法の提案はある。文献[18]では、主としてサイエンティフィックビジ

ュアリゼーション (SV) に対して, 2次元スカラー場の可視化には可視化操作の一つである等値線表示を選択するといった一般的な方法論が述べられている。これらは, 可視化の目的を達成するという点では不十分であることが多い。例えば, 「2次元スカラー場の可視化」には, 「物体に加わる局所的な力を知るために物体表面の圧力分布を調べる」や「数値振動の有無や生じている場所を調べる」という, 可視化プロセス全体における「何のために」に属する目的が存在し, 一般的な方法論ではそのような目的を達成することが難しいからである。

本研究では, 「何のために」に相当する目的を上位の目的とし, 可視化操作や見方 (表示) に直接関連する目的を下位の目的と呼ぶことにする。目的を, 「目的 = ねらい + 機能」とした場合に, 上位の目的が「ねらい」に, 下位の目的が「機能」に関連付けられる。発見 (探索) 型の場合は, 上位の可視化目的が一連の可視化プロセス後に同定されることが多い。例えば, 2次元スカラー場の可視化のために, 圧力場に対して等値線表示を行った結果, 「数値振動が生じている場所がわかった」というようなものである。

上位の目的と下位の目的を結びつけるためには可視化対象に依存する分野固有の知識が必要であることが多い。しかしながら, 物体表面の圧力分布を調べるために圧力場に対して等値線表示を行ったというように, 可視化プロセスを繰り返す中で関係性が明確になる場合もある (目的があいまいであっても行う作業自体は決まっている, という可視化の特徴が活かされた結果である)。

ここでエージェント化の可能性を考える。はじめに, 可視化の目的を同定することは難しいが,

- (1) 目的が階層的に表せる
- (2) 可視化プロセスの分析によって上位と下位の目的の対応付けが可能になる
- (3) 下位の目的に関しては可視化操作と目的が対応付けられる

ことを仮定する。可視化プロセス全体に対するエージェントを設計することは難しいが, どちらかの可視化目的が与えられた場合は一連の作業手順を決定できることが多い。本稿では, この点に着目し, 可視化作業を分析することによって, 可視化目的が与えられた場合の可視化エージェントに担わせることが可能な作業を抽出する。

以下では, 流れの可視化を例にするが, 可視化全般に対して同様の方法が利用できるものと考えている (講演では情報可視化の例も示す予定である)。

3. 1 情報化と構造化

情報工学におけるエージェントの設計は, 機械処理に適した形式にデータや情報を再構成し (構造化, さらに階層化), 再構成されたデータや情報に対して機械学習を含めたデータマイニングの方法を活用

して行われることが多い。本研究においてもこの方針を採る。

はじめに, データの情報化, データと情報の構造化が必要になる。第1のポイントはデータの記述法である。XML, RDF を利用してメタデータが追加できればよい。現実的には標準化の問題があり, そのような形で記述できるとは限らないが, 部分的であっても機械処理に適した形式で記述できればよい。

第2のポイントはパラメータ化と符号化である。可視化パラメータを分類すると,

- ・可視化対象の選択のためのパラメータ,
 - ・可視化操作に対するパラメータ,
 - ・表示方法に対するパラメータ
- となる。例えば, 圧力場に対する等値線の本数や3次元の視点である。

また, 可視化操作自体のパラメータ化も必要である。ここでは, 可視化操作を可視化の機能と考える。可視化プロセスの中で, どの機能が選択されたかを記号と数値で表すことによって, パラメータとして表現する。SV に関しては, 可視化操作の種類は多くはない。それらを,

- A: 格子線, B: 格子面, C: 等値線, D: 等値面,
- E: 等値領域, F: ボリュームレンダリング,
- G: ベクトル表示, H: 粒子追跡

のように記号化する。これらの操作に, 可視化パラメータが付随する。可視化パラメータを以下の3種類のカテゴリ変数群 (O_i, D_i, P_i) で分類する。

O : 操作に関わるもの (等値線 (次元, 分布, 範囲, 本数, アルゴリズム等), 粒子追跡 (状態, 時間刻み幅等) など)

D : 表示属性を示すもの (色, 透明属性, 矢印等)

P : 可視化精度に関わるもの (時間精度, 空間精度等)

また, 変数の意味, および取り得る値を予め定義しておく。等値線の範囲や本数, 粒子追跡の時間刻み幅など, 連続変数のものも多いが, 一旦カテゴリ変数で表し段階的に (階層構造として) 数値を与える。

例えば, 等値線に対するカテゴリ変数として,

- ・次元: $O_1 = \{2 \text{次元}, 3 \text{次元}\}$
- ・分布: $O_2 = \{\text{等間隔}, \text{不等間隔}\}$
- ・等値線の値の範囲: $O_3 = \{\text{自動}, \text{デフォルト}, \text{入力}\}$
- ・等値線の本数: $O_4 = \{\text{自動}, \text{デフォルト}, \text{入力}\}$
- ・アルゴリズム 1: $O_5 = \{\text{分割}, \text{直接}\}$
- ・アルゴリズム 2: $O_6 = \{\text{直接}, \text{フラグ利用}, \text{MCube}\}$
- ・表示属性: $D_1 = \{\text{色付き}, \text{モノクロ}\}$
- ・空間精度: $P_1 = \{\text{線形補間}, \text{双一次補間}\}$

というようなものを考える。ここで, “自動” は, 可視化対象のデータに依存してシステム側で自動的に設定するものである。“入力” は, 数式表現されたものの係数を入力するもの, テーブル表現されたものからの選択などを意味する。このような変数に対しては, カテゴリを選択後に具体的な値を決めるという形式にする。例えば, 等値線の値の範囲: $O_3 = \{\text{自$

動, デフォルト, 入力} において, “入力”の選択後, 最小値と最大値を要求する場合,

$$O_3 = \text{”入力”} \mid (\text{最小値, 最大値})$$

のように属性値が付随する. また, 表示属性や可視化精度に関しても同様に扱う. 例えば, 色の算出式やカラーテーブルといった更に詳細な情報が階層的に加えられることになる.

さらにカテゴリ変数を符号化する. 例えば, 等値線の場合,

$$C O_1 O_2 O_3 O_4 O_5 O_6$$

のように符号化する. 次元は”2次元”, 分布は”等間隔”, 値の範囲は”入力”, 本数は”デフォルト”, アルゴリズム1は”分割”, アルゴリズム2は”フラグ利用”であれば,

$$C002111$$

となる. このように可視化操作は, 形式的には操作を表す記号と可視化パラメータによって符号化できる.

3.2 構造化 (ワークフローの記述)

可視化対象となるデータの選択後, 可視化は目的に関わらず,

- (i) 可視化対象の場の決定 (対象量, 領域)
- (ii) 可視化操作の決定
- (iii) 可視化結果の表示
- (iv) 結果の解釈と判断

を繰り返しながら遂行される. 可視化操作と見方(表示)の繰り返しは, (i) (ii) (iii), あるいは(ii) (iii)の中で生じる. 表示を経て, 結果の解釈を行い, 継続か終了かの判断を行う.

そこで, 表示 (結果を見る) までの段階を一つの手続きとする. この手続きをパラメータの並びで表すことができれば, 作業に対するワークフローといった, 工程の繋がりや処理順序を記述し, 管理することができるようになる. これが第3のポイントである. このためには, 可視化操作のパラメータ化に加えて, (i)と(iii)についてもパラメータ化が必要になる.

(i)の可視化対象の場は,

- (i-1) 可視化対象となるデータの種類 (構造型, 非構造型, 散逸型, 複合型など)
- (i-2) 対象となる物理量 (圧力, 温度, 速度など)
- (i-3) 対象領域 {面(格子面, 切断面), ボリューム(格子領域, 切断領域)など}

等によって決まる. (iii)の表示は, (iii-1) 表示領域, (iii-2) 3次元の視点によってパラメータ化できる. 一つの手続きを S_k とし,

S_k {対象の場, 可視化操作(可視化パラメータ), 表示} によって表し, 可視化結果 V_k と対応付けて, 可視化結果がどのように生成されたかを示す. これは構造化された要素を, 順序をつけて関連づけることに相当する. 例えば, 可視化操作の部分だけ示すと, 等

値線, 格子線, ベクトルで表示するものは,

$$S_1 = \{ \dots, CAG, \dots \}$$

と表される. 実際には, C002111 のようにより詳細なパラメータがともなっている.

パラメータによるワークフローの記述後, {対象の場, 可視化操作(可視化パラメータ), 表示}と, 対象データ, および可視化結果を機械処理可能な形で結びつける. この際, セマンティックウェブにおける RDF と RDF スキーマ層を用いる. 例えば,

```
<rdf:RDF>
<rdf:Description about="http://S1.html">
  <data resource="field.dat" />
  <parameter resource="xyz.prm" />
</rdf:Description>
</rdf:RDF>
```

という記述は, 「S1.html が指定する結果は, field.dat と xyz.prm によって作られた」を意味する. このように, ある可視化結果を得るためのプロセスを機械処理可能な履歴情報として蓄えることが可能になる.

3.3 可視化エージェントの設計

エージェントの定義は様々であるが, 本稿では可視化エージェントを,

「手続き (操作や見方) の類型化から得られるルールにもとづいて, 可視化作業を行い, 可視化結果を蓄えていくもの」

と考える.

はじめに, ある可視化結果を得るための履歴情報から可視化エージェントに担わせることが可能な作業の抽出法を検討する. 履歴情報からの自動抽出が望ましいが, 背景として必要な知識が多く容易ではない. このため, 得られた履歴情報に対して, 目的, 可視化作業, 可視化結果の対応付けを行う. この際に, 可能ならば, Fekete ら[16]のように可視化に対する利得を与える.

より具体的には, X という対象に対して, 目的 Y_i と手続き S_k の意図をヒアリング等で明らかにし, 目的の達成度, 知見獲得の有無, 手続きの必要度などを4件法や5件法で調べ, 利得 F_k として与える. この繰り返しによって, 対象 X , 目的 Y_i に対して可視化結果 V_k が手続き S_k とともに蓄えられていく ($k=1, \dots, N$). また, 利得 F_k が与えられる場合もある. この履歴情報に対するデータマイニングによって手続き (操作や見方) の類型化を行い, 関連規則のようなルール抽出を試みる. 得られたルールをエージェントに与える. 例えば, 対象 X と目的 Y_i に対しては, エージェントはルールに従って可視化操作を行う. 単純であっても作業者の替わりに行うということが可視化作業の支援になる.

本稿の最後として, 一つの事例を示す. 普遍的なルールを導くことが難しいこと, また, 特に可視化が探索型となった場合, 利得の設定も煩わしいもの

になることは想像に難くない．そこで，簡単なルール抽出から始めた．例は，いくつかの2次元物体を過ぎる非圧縮性流れの計算データの可視化に対する履歴情報からのルール抽出である．

はじめに，手続き S_k の中で，精度を調べるためのものを抽出した．この場合の目的「計算精度を検討する」は流体計算では一般的なものなので，用意した選択肢や，自由形式の記述等により明らかにできる．ただし，今回は計算精度を検討するという目的を前もって与えた．

目的の達成度は，可視化結果 V_k から，(4:十分にできている, 3:どちらかといえばできている, 2:どちらかといえばできていない, 1:できていない) という4段階で主観的に判断した．

対象と目的が明らかとはいえず，すべてのパラメータに対するデータマイニングから有意なルールを抽出することは難しい．そこで得られた履歴情報と経験的な知見から以下のようにパラメータを簡素化した．可視化対象となるデータの種類の種類は構造型，物理量は， q_1 : 圧力, q_2 : 速度, q_3 : 渦度, q_4 : 流線とする．ただし，簡単のため，圧力と速度などを同時表示したものはルール抽出の対象から除外した．

可視化操作は，A: 格子線, C: 等値線, G: ベクトル表示とする．格子線のパラメータは，範囲(全体, 物体面, 領域指定)のみとする．それらを a_1 : 全体, a_2 : 物体面, a_3 : 領域指定, と符号化する．等値線に関しては，分布と値の範囲をまとめて符号化する．分布を(等間隔, 不等間隔), 値の範囲を(全体の最大最小, 指定領域の最大最小, 0 を中心とする入力値)とし, c_1 : 等間隔, 全体の最大最小, c_2 : 等間隔, 指定領域の最大最小, c_3 : 等間隔, 0 を中心とする入力値, のように符号化する．なお，利用されなかった操作は a_0 , c_0 , g_0 とした．

表示に関しては，平行移動, 拡大, 縮小に限定し, v_0 : デフォルトのまま, v_1 : 平行移動, v_2 : 拡大, v_3 : 縮小, v_4 : 平行移動, 拡大, v_5 : 平行移動, 縮小, v_6 : 拡大, 縮小, v_7 : 平行移動, 拡大, 縮小と符号化する．表2に手続き S_k の一部を示す．

表2 可視化手続きの例

手続き#	場	操作			表示	達成度
		A	C	G		
1	q1	a0	c1	g0	v0	2
2	q1	a1	c1	g0	v0	2
3	q1	a1	c2	g0	v2	3
4	q1	a3	c2	g0	v6	4

はじめに，達成度を除外し，表2のように多値のカテゴリ変数を 0,1 のフラグ型変数に展開してバスケット分析を行った．手続きの総数は30である．

抽出されたルールの例を示す．

rule1 {操作 A=a1, 操作 C=c2} => {場=q1}
rule2 {操作 A=a2, 操作 C=c6} => {場=q4}

ルールを読み解くと，rule1 は，圧力場の可視化を利用して計算精度を検討する場合，指定領域に対して等間隔で等値線を表示し，かつ格子線全体を表示する，というもので，rule2 は流線の場合は，0 を中心とした入力によって不等間隔で等値線を表示し，格子線は物体を表示する，というものである．ただし，得られるルールは多く，意味のあるものを見つけることは難しい．よく指摘されているようにデータマイニングの結果を精査する必要はある．

この他に，自己組織化マップによる手続きの分類，ベイジアンネットワークによる主要因子の推定などを行っている．方法，および結果の詳細については講演時に述べる．

4. おわりに

「可視化から何がわかるのか」と問われれば，何がわかるという以外答えようがない．可視化からわかることは対象依存であり，可視化を必要としないものも多いためである．また，解釈は個人に任せるべきものであると指摘されることもある．したがって一般論を述べることは難しい．

しかしながら，データ・情報の大規模化，多様化にともない，情報を一個人では消化できないという状況が新しい知見の創出を妨げていることが指摘され，視覚(可視化)情報のような人間が直感的に理解しやすいものへの変換は必須になっている．一方，同時に，可視化作業自体の難しさや非効率性が指摘されている．このため，作業を支援する，あるいは効率化する仕組みが必要になっている．

本稿で述べた人間の特性をふまえた可視化，および解釈まで考えた可視化，そして可視化エージェントがその仕組みの例である．

現在，日常的に行われている技術計算の規模は数十から百万点程度である．少ないながらも定型可視化プロセスが存在し，部分的な可視化結果からの全体推定も容易なため，格子生成から可視化まで含めて一両日で終了することが多いという(ただし，著者の経験であるが，探索型の場合は同規模の分析に数週間を要する)．しかし，数千万点を超える大規模な計算データの場合，インタラクティブな可視化は難しく分析の効率が悪い(数ヶ月を要するという報告もある)．この場合，可視化戦略というべきものが必要になる．可視化エージェントの利用はその1つである．例えば，計算点数以外を同じパラメータにした小規模な計算結果に対する可視化の履歴情報から直接的にルールを抽出する，あるいは手続き(操作や見方)の類型化の後にルールを抽出する．それらをエージェントに与え，エージェントを用いて大規模データに対する可視化を行い，効率的に有用な可視化結果を蓄えるということが考えられる．

なお，本稿で紹介したものは，研究の緒についた段階のものが多い．さらに多くの研究が望まれる．

謝辞

本研究の一部は科学研究費補助金(基盤研究(B))(課題番号 21300031)の援助を受けた。記して謝意を表する。

参考文献

- [1] McCormick,B.H., DeFanti,T.A. and Brown, M.D.: Visualization in Scientific Computing, Computer Graphics, Vol. 21, No. 6, pp.1-14 (1987).
- [2] 日本弁護士連合会編: *裁判員制度と取調べの可視化*, 明石書店 (2004).
- [3] 白山 晋: *知的可視化*, 丸善 (2006).
- [4] IEEE Symposium on Visual Analytics Science and Technology: <http://vis.computer.org/VisWeek2009/vast/>
- [5] Rauschenbach, U.: Progressive Image Transmission using Levels of Detail and Regions of Interest, Proc. of IASTED CGIM 98,pp.38-41 (1998).
- [6] 資源エネルギー庁編: *エネルギー2004*, エネルギーフォーラム (2004).
- [7] Spence, R. : *Information Visualization*, ACM Press Books, Addison-Wesley (2001).
- [8] many eyes:
<http://manyeyes.alphaworks.ibm.com/manyeyes/>
- [9] Kim,Y. and Varshney, A.: Saliency-guided Enhancement for Volume Visualization, IEEE Trans. on Visualization and Computer Graphics, **12.5**, pp. 925-932 (2006).
- [10] Chan,M-Y., Wu, Y., Mak, W-H., Chen, W. and Qu, H.: Perception-Based Transparency Optimization for Direct Volume Rendering, IEEE Trans. on Visualization and Computer Graphics, **15.6**, pp.1283-1290 (2009).
- [11] Blignaut,P.J., Beelders,T.R., So, C-Y.: The Visual Span of Chess Players, Proc. of the 2008 symposium on Eye Tracking Research & Applications, pp.165-171 (2008).
- [12] Egawa, A. and Shirayama, S.: Generation of an Importance Map for Visualized Images, Lecture Notes in Computer Science, **5875**, pp.135-146 (2009).
- [13] Ebert, D. and Shaw, C.: Minimally Immersive Flow Visualization, IEEE Trans. onVisualization and Computer Graphics, **7.4**, pp.343-350 (2001).
- [14] Strothotte, T. and Schlechtweg, S.: *Non-Photorealistic Computer Graphics: Modeling, Rendering and Animation*, Morgan Kaufmann Pub. (2002).
- [15] van Ham, F. and Rogowitz, B.E.: Perceptual Organization in User-Generated Graph Layouts, IEEE Trans. on Visualization and Computer Graphics, **14.6**, pp.1333-1339 (2008).
- [16] Fekete,J-D., van Wijk, J-J., Stasko, J.T. and North, C.: The Value of Information Visualization, Lecture Notes in Computer Science, **4950**, pp.1-18 (2008).
- [17] Plaisant, C.: The Challenge of Information Visualization Evaluation, Proc. of the working conference on Advanced Visual Interfaces 2004, pp.109-116 (2004).
- [18] Keller, P. R. and Keller, M. M.: *Visual Cues: Practical Data Visualization*, Institute of Electrical & Electronics Enginee, (1993).