

ことばからグラフへ

— 対話型可視化システム KEVIN —

KEVIN: An Interactive Visualization System for Data Exploration

松下 光範[†]

Mitsunori Matsushita

[†]NTT コミュニケーション科学基礎研究所

NTT Communication Science Laboratories

加藤 恒昭[‡]

Tsuneaki Kato

[‡] 東京大学 大学院

The University of Tokyo

Abstract: We propose an interactive visualization method suitable for exploratory data analysis. In such an analysis, statistical charts are used to help users understand the target data, and a displayed chart is redrawn interactively according to the user's input inspired by looking at the chart. We found that it was necessary to choose a suitable redrawn graph while considering two factors: the type of the displayed chart and the parts of the chart that change in accordance with the user's requirements. Our proposed method is developed based on this knowledge. As a first step of our proposed method, the changing parts and their features to meet the user's requirements for a displayed graph are interpreted. Then the semantic frame used to generate a statistical chart from data is updated by the results of the interpretation, and several candidates as appropriate chart-types are chosen based on frame and chart-type determination rules. Finally, the most preferable chart is chosen from among the candidates according to the type of the user's requirements, the displayed chart is transformed into the new chart with animation. By using this method, data can be visualized interactively according to changes in users' intentions without interrupting their thought processes.

1 はじめに

近年、計算機の処理能力の向上やネットワークの普及により、複数のユーザが多様かつ大量のデータを共有し、異なる意図の下で利用できるようになった。可視化はこのような膨大なデータを直観的に理解するために有効な手段の一つである。

ユーザがデータを可視化する際の基本的なプロセス [1] を図 1 に示す。このモデルでは、観測されたデータの中から必要なデータを取り出して加工し、データテーブルに変換する data transformation と、このデータテーブルを元にグラフ種や配置などの可視化構造を決定する visual mapping と、決定された可視化構造のパラメータを変化させ、焦点化や俯瞰化によってより効果的なグラフに変換する view transformation という 3 つのプロセスを経て、ユーザは目的のグラフを得る。

各プロセスに於けるユーザインタラクションの負荷を軽

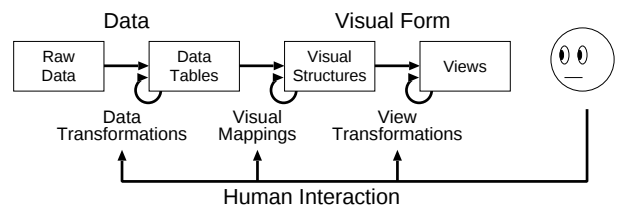


図 1: 可視化のモデル

減するために、多くの可視化システムが提案されている。

Data transformation については、必要なデータを選択を容易にする試みがなされている。例えば FilmFinder[2] では GUI のスライダーを操作することによって可視化するデータを選択する。また Visage[3] のように可視化対象のデータを GUI 上で直接選択し決定するものもある。

Visual mapping については、グラフ種の選択や、線種などのグラフオブジェクトを容易に選択する試みがなされている。例えば Microsoft Excel 等のスプレッドシートが有するグラフウィザード機能のように、グラフ種を視覚的に選択する方式がこれに相当する。Mackinlay[4] や

Fasciano[5] らはこのプロセスをデータ形式に基づいて自動的に行うことを試みている。

View transformation については、ユーザの視点や注目点などを明確にするための視覚効果を容易に変更する試みがなされている。例えば、Mittal は軸の範囲や原点の与え方などをデザインパラメータと捉え、これらをスライダーやチェックボタンで変更する事により、グラフ形状を操作する手法を提案している [6]。

このように、従来の可視化システムでは主として各プロセスに対する直接的なインタラクションが想定されている。特に GUI を利用して、これらの操作を簡単に処理する試みが多く見られる。言い替えれば従来研究の主眼は「どのように描画するか (how to draw)」という点に集中していると言える。

しかしながらこのようなプロセスを遂行するには、データの集約方法、データの網羅性や完全性の検証、データの特徴に適したグラフ種の選択方法など、様々な点を考慮しなくては適切なグラフが得られない。これは、本来ユーザが求めている「何を描画したいか (what to draw)」から各プロセスのパラメータへの変換がユーザの負担として残されていることを意味する。従って、得られるグラフの質はユーザの能力に依存し、場合によっては効果的でないグラフを描画してしまう事も考えられる。

またグラフを利用してデータを分析する場合、ある観点で描画されたグラフからユーザが新たな観点を想起してグラフを描き換えるのが一般的である。このようなグラフの利用を想定した場合、従来手法が採用しているモデルでは、ユーザがグラフを描き換えようとする度に各プロセスを一からやり直さなくてはならず、効率的ではない。

そこでこれらの問題点するために、日常的な言葉で表現された「何を描画したいか」という一連のユーザの要求から描画に必要なパラメータを抽出してグラフを描画し、対話的にそのグラフを描き換えていくシステムを提案する。

2 対話的なグラフ描画の特徴

本稿では二つの独立変数によって値が特定されるデータを対象に議論する。具体的な例として表 1 のような「時刻 × 場所 → 降水量」という観測データを取り上げる。

このデータに対して、以下に示す一連のユーザ要求が順に与えられる場合を考える。

- (1) 93 年と 94 年の四国地方の降水量を知りたい
- (2) 県毎に知りたい
- (3) 97 年まで知りたい

表 1: 観測データの例

県	観測日時	降水量
北海道	1990-01-01	5
北海道	1990-01-02	4
北海道	1990-01-03	0
...	...	...
沖縄	1995-12-30	0
沖縄	1995-12-31	3

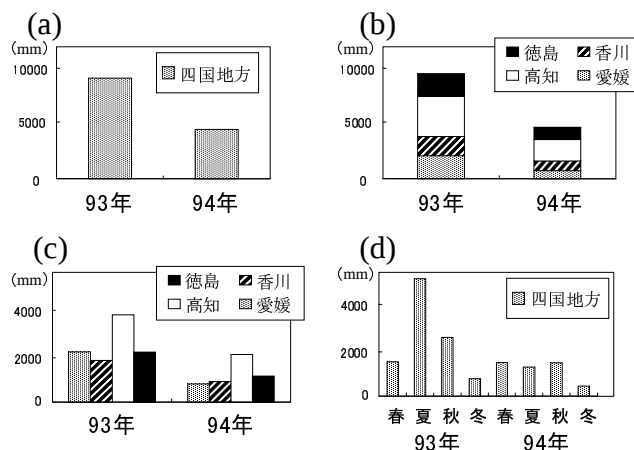


図 2: 追加ユーザ要求によるグラフ変化の違い

まずユーザ要求 (1) が与えられたとする。このユーザ要求 (これを開始ユーザ要求と呼ぶ) を満たすグラフは、例えば図 2 (a) である。このグラフに対し、ユーザが視点を変えてユーザ要求 (2) を加えた場合 (これを追加ユーザ要求と呼ぶ)、新しく描画されるグラフは図 2 (a) を県毎に詳細化したグラフであるべきで、図 2 (b) に示すような、棒グラフの各棒を県毎に細分化したグラフが適切であろう。

この連続したユーザ要求が全体として表しているのは

- (4) 93 年と 94 年の四国地方の県毎の降水量を知りたい

に他ならないが、このユーザ要求が開始ユーザ要求として単独に与えられた場合に適切なグラフは、図 2 (b) よりむしろ (c) である。これはユーザ要求 (1) と (2) を逐次的に与えた場合、四国地方全体の年間降水量に関するユーザの関心が (1) で示されているので、これを直感的に表現している (b) の方が適切だが、ユーザ要求 (4) ではそのような関心は示されていないので県毎の降水量が読みとり易い (c) の方が適切だからだ、と推測される。またグラフ (a) と (b) はその形状が類似しているのに対し、(a) と (c) は大きく異なるので、描画された要素同士の対応がすぐには判断しづらい。この点でも (a) に続けて描画する際には (c) よりも (b) が適当と思われる。

このことから、自然言語処理による対話システムが行う

ような断片的な発話の理解によってユーザ要求 (1) を前提として (2) を解釈し、それによって得た (4) に基づいてグラフを選択する方式では適切なグラフは選択できないことが分かる。そこで以下では、一連のユーザ要求に沿ったグラフを選択する際に考慮すべき事柄について検討する。

開始ユーザ要求 (1) に続けて

(5) 季節毎に知りたい

という追加ユーザ要求が与えられた場合、時間が詳細化されているので図 2 (b) のような積み上げ棒グラフよりも、(d) のように x 軸を詳細化した単純棒グラフのほうが適切であろう。

追加ユーザ要求 (2) と (5) ではどちらも同じ表現が用いられ、描画されている情報の詳細化を要求しているにもかかわらず適切なグラフが異なる。これは前者が描画される対象を詳細化 (i.e., 地方を県に詳細化) しているのに対して後者は軸を詳細化 (i.e., 年を季節に詳細化) しているためである。従って、追加ユーザ要求と適切なグラフとの関係を明らかにするには単に詳細化だと解釈するだけでは不十分で、開始ユーザ要求で示されたグラフ描画のための情報に対し、追加ユーザ要求によってグラフのどの部分にどのような変更が生じるかを考慮する必要がある。

また同一の追加ユーザ要求であっても、現在描画されているグラフ種によって、新たに描画されるべき適切なグラフが異なる。例えば描画される対象を追加するような追加ユーザ要求が与えられた場合、単純棒グラフが描画されている場合は複合棒グラフを描画するのが適切であるが、単純折線グラフが描画されている場合には複数折線グラフが適切であろう。従って、追加ユーザ要求によって新たに描画されるグラフの選択には、直前に描画されているグラフ種も考慮する必要がある。

以上の考察に基づき、以下では一連のユーザ要求に対して逐次適切なグラフを描画する方法について説明する。まず、開始ユーザ要求からグラフを描画する方法について述べ、その後に追加ユーザ要求が与えられたときのグラフ描画方法について述べる。

3 開始ユーザ要求からのグラフ描画

既に我々は自然言語で表現されたユーザ要求を意味フレームと名付けた形式で表現することでグラフ描画に必要な情報が獲得できることを示し、ユーザ要求からこの意味フレームを抽出する手法を提案している [7]。

例えば開始ユーザ要求 (1) は図 3 に示すフレーム形式に変換される。フレームはユーザ要求に依存して複数の行を持つ。各行は描画に要する一連の情報であり、これを「描

変数	ドメイン	粒度	型	制約条件
X	時刻	年	時間	$x_i \in \{93\text{年}, 94\text{年}\}$
Y	場所	地方	名義	$y_i = \text{四国地方}$
Z	降水量	mm	量	

図 3: 開始ユーザ要求 (1) のフレーム表現

画要素」と呼ぶ。各描画要素は 変数、ドメイン、粒度、型、制約条件の 5 つのスロットで表現される。以下に各スロットについて簡単に述べる。

変数スロットは描画要素を識別するための変数を値とする。

ドメインスロットはドメイン名を値とする。全ての描画要素はユーザ要求中に含まれる名詞連続によって特定される。例えば、“93 年”、“94 年”、“四国地方”、“降水量”といった名詞連続が開始ユーザ要求 (1) の中に現れる。これらの名詞連続は各々降水量ドメイン、日時ドメイン、場所ドメインに属する要素である。なお“93 年”と“94 年”は同じ日時ドメインに属している。これらのドメイン名がこのスロットの値として採用される。

粒度スロットはユーザが求めている描画要素の粒度を値とする。粒度とは、描画要素の記述単位のことである。

型スロットは描画要素のデータ型を示す。先行研究では、データ型は名義型、量型、順序型の 3 つに大きく分類されており、各データ型は更に幾つかの部分型に分類されている [1]。型を詳細に区別することはグラフ種を適切に判定する上で有効である半面、型の分類数に応じてグラフ種を判別するための知識が複雑になる。我々はこのトレードオフを考慮し、型スロットが取り得る値の種類を名義、時間、順序、量、区間、割合の各型に限定した。

制約条件スロットはユーザ要求中に記述されている描画要素の制約条件を示す。制約条件は、要素指定型 (e.g., $y_1 = \text{京都府}$)、上位概念指定型 (e.g., $y_1 \in \text{近畿地方}$)、範囲指定型 (e.g., $0 \leq y_i \leq 100$) のいずれかで記述することとする。

各描画要素はデータテーブル中のいずれかの属性と対応し、グラフの構成要素となる。図 3 では X と Y が独立変数で、グラフ上ではどちらか一方が x 軸になる (x 軸になる独立変数を x 軸変数、他方を対象変数と呼ぶ)。また Z が従属変数 (値) で y 軸になる。

このフレームに記述された情報に基づいて、観測データにデータ集約アルゴリズム [8] を適用することで、表 2 に示すテーブルが得られる。このテーブルと上述の意味フレームにグラフ種判別知識 [9] を適用することで、開始ユーザ要求に適したグラフが描画される。

表 2: 図 3 のフレームにより再構成されたテーブル

	93 年	94 年
四国地方	8840	4323

4 追加ユーザ要求からのグラフ種特定

2 章での考察から、開始ユーザ要求から図 3 のようなフレームが得られているという条件の下で、追加ユーザ要求から適切なグラフを描画するには

- (i). 追加ユーザ要求によって描画されているグラフのどの部分にどのような変更が生じるかを解釈する。
- (ii). (i) と現在の意味フレームから新しい意味フレームを合成し、得られた意味フレームにグラフ種判別知識を適用してグラフ種の候補を決定する。
- (iii). (i) と描画されているグラフの種類を元に、(ii) で得られたグラフ種候補の中から適切なグラフ種を選択し描画する。

という処理が必要である。

まず、追加ユーザ要求によって描画されているグラフのどの部分にどのような変更が生じるかを解釈する。

追加ユーザ要求は描画されているグラフを前提として行われるのでユーザは変更点にのみ言及すると考えてよい。例えば図 2 (a) のグラフが描画されている場合、ユーザは「93 年から 94 年の四国地方」という情報を省略し、単に「県毎に見たい」と要求するものと考えられる。そしてこの追加ユーザ要求で言及されていない情報については、現在描かれている情報から変化しないと見なせる。

追加ユーザ要求のタイプは様々考えられるが、グラフの変化を伴う要求として、追加 (～を追加したい)、削除 (～を削除したい)、粒度変更 (～毎に知りたい)、軸交換 (～を軸にしたい)、割合化 (～の割合が知りたい)、実量化 (～の量が知りたい)、推移化 (～の推移が知りたい)、の 7 種類を対象とした。

2 章で述べたように、同じタイプに属する要求であってもグラフのどの部分 (変数) に対する要求であるかによって適切なグラフは異なる。また、同じ「県毎に見たい」であっても、現在のグラフが市毎に描画されていれば粗大化であるし、地方毎であれば詳細化である。このように追加ユーザ要求の解釈にはその表現に加えて、現在のフレームを参照することが必要である。これを整理したものが表 3 である。文末を主とする表現パターンで“要求パターン”が、どの変数を対象としているかで“変化する変数”が、現在

表 3: 変化の種類と意味フレームの変更箇所の対応

要求パターン		変化する変数	変化内容	フレーム
追加	1	x 軸変数	追加	制約条件
	2	対象変数	追加	制約条件
削除	3	x 軸変数	削除	制約条件
	4	対象変数	削除	制約条件
粒度変更	5	x 軸変数	詳細化	粒度
	6	対象変数	詳細化	粒度
	7	x 軸変数	粗大化	粒度
	8	対象変数	粗大化	粒度
軸変更	9	独立変数	軸変数変更	変化なし
割合化	10	従属変数	割合化	型
実量化	11	従属変数	実量化	型
	12	なし	棒グラフ化	変化なし
推移化	13	なし	折線グラフ化	変化なし

の内容との関係で“変化内容”が決定する。これらは 13 パターンに細分化できることが分かった。なお、最右列はフレーム中で変更される箇所を示している。

次に、上記の処理で得られた“変化する変数”と“変化内容の詳細”から、現在の意味フレームに変更を加える。表 2 の最右列がフレーム中で変更される箇所である。例えば追加ユーザ要求 (2) は表 2 のパターン 6 に当てはまり、“変化する変数”は対象変数、“変化内容の詳細”は詳細化なので、図 3 中の描画要素 Y の粒度スロットが変更される。

これによって得られた意味フレームを元に、グラフ種判別知識を用いて描画可能なグラフ種の候補を決定する (なお本研究では一般的な統計グラフとして、単純棒グラフ、複合棒グラフ、単純積上げ棒グラフ、複数積上げ棒グラフ、構成比率棒グラフ、単純折れ線グラフ、複数折れ線グラフ、帯グラフ、円グラフ、多重円グラフを対象としている)。追加ユーザ要求 (2) によって得られる意味フレームの場合、複数折れ線グラフ、複合棒グラフ、複数積上げ棒グラフの 3 種類がグラフ種候補になる。

その後、上記の処理で得られたグラフ種候補の中から、描画されているグラフの種類と追加ユーザ要求パターンを考慮して適切なグラフ種を決定する。

グラフ種は上記で得られた“変化する変数”、“変化内容”と現在のグラフ種から決定される。そのため、グラフ種同士の形状の類似性や一般的な用いられ方に着目して、あるグラフから他のグラフへの変化の関係を整理した。

例えば単純棒グラフが描画されている場合、1、3、5、7 のいずれかの要求パターンに属する追加ユーザ要求が与えられた場合は、新たに描画されるグラフとして単純棒グラフが選択されるが、2 の要求パターンに属する追加ユーザ要求が与えられた場合は複数棒グラフが、6 の要求パタ

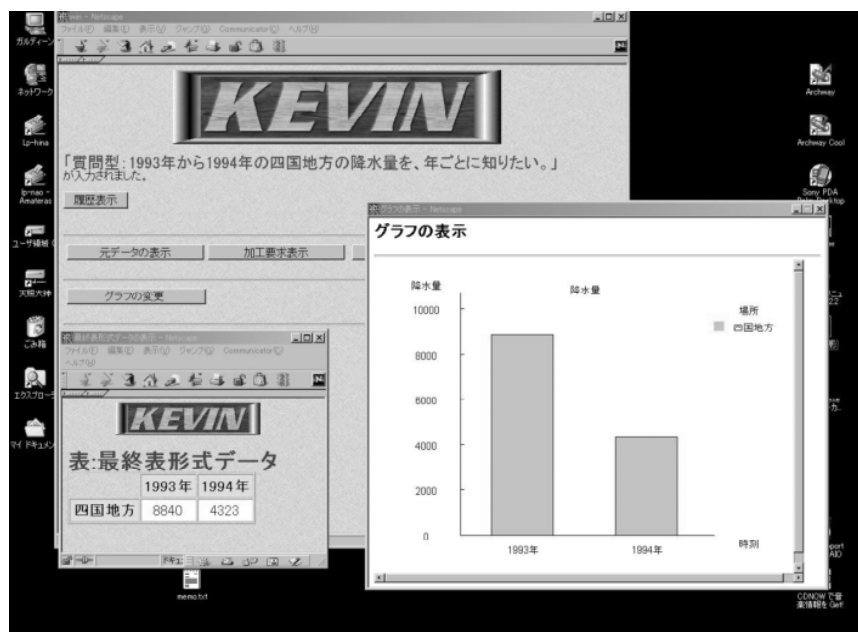


図 4: 対話型可視化システム KEVIN

ンに属する追加ユーザ要求が与えられた場合は積み上げ棒グラフが各々選択される。

以上の処理により、探索的なデータ分析の場面での適切なグラフ選択が実現できる。我々はこの仕組みを可視化システム KEVIN 上で実装した (図 4)。

2 章で適切なグラフ選択の基準のひとつとして、グラフに描画された要素同士の対応の容易性を挙げた。この点を重視して、KEVIN ではグラフ間の対応関係がより明確になるように各グラフ変化パターンに対応した変化アニメーションを実装し、変化の過程を連続して見せることでユーザにそれらの対応関係を直感的に理解させようと試みている。

5 おわりに

本稿では、日常的な言葉で表現された「何を描画したいか」という一連のユーザの要求から描画に必要なパラメータを抽出し、対話的にグラフを描画していく可視化システム KEVIN について述べた。

今後はグラフ変化に関する知識の精緻化や手法の有効性の定量的評価について検討したい。

参考文献

[1] Card, S. K., Mackinlay, J. D. and Shneiderman, B. (eds.): *Readings in Information Visualization - Using Vision To Think -*, Morgan Kaufmann Publishers (1999).

[2] Ahlberg, C. and Shneiderman, B.: Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays, *Proceedings of the Conference on Human Factors in Computing Systems (CHI'94)*, pp. 313-317 (1994).

[3] Derthick, M., Harrison, J., Moore, A. and Roth, S. F.: Efficient Mult-Object Dynamic Query Histograms, *Proceedings of the IEEE Information Visualization Conference (InfoVis '99)*, pp. 84-91 (1999).

[4] Mackinlay, J.: Automating the Design of Graphical Presentations of Relational Information, *ACM Transactions on Graphics*, Vol. 5, No. 2, pp. 110-141 (1986).

[5] Fasciano, M. and Lapalme, G.: PostGraphe: A System for the Generation of Statistical Graphics and Text, *INLG 96*, pp. 51-60 (1996).

[6] Mittal, V. O.: Visual Prompts and Graphical Design: A Framework for Exploring the Design Space of 2-D Charts and Graphs, *Proceedings of AAAI 97*, pp. 57-63 (1998).

[7] Matsushita, M., Yonezawa, H. and Kato, T.: A Frame Representation of User Requirements for Automated Data Visualization, *ECAI-2000*, pp. 631-635 (2000).

[8] 松下光範, 米澤勇人, 加藤恒昭: ユーザ要求に基づく数値データ可視化のための集約手法, 第 14 回人工知能学会全国大会, pp. 196-199 (2000).

[9] 米澤勇人, 松下光範, 加藤 恒昭: 数値データ可視化のためのグラフ判別知識の構築 — 新聞記事中のグラフに基づく分析 —, 第 60 回情報処理学会全国大会 Vol.2, pp. 199-120 (2000).