

表題に基づく統計データの自動可視化手法

松下光範[†] 米澤勇人^{††} 加藤恒昭^{†††}

本稿では日常的な言葉で表現されたユーザ要求に基づいて、多量の統計データを自動的に集約して可視化する手法を提案する。提案手法ではこのようなユーザ要求としてグラフの表題に着目し、それを解釈することで意味フレームと名付けた表現を得る。意味フレームは「何を描画するか」に関する情報を表現する話題項目フレームと「何を主張したいか」に関する情報を表現する比較フレームとからなる。意味フレームとして表現された情報を用いて、統計データからユーザが必要かつ適切と考える範囲と詳細度を持ったデータテーブルを作成する。この際、ドメイン階層知識とドメイン間関係知識によって構成された伝播ネットワークを用いる。領域に固有の知識はこの部分に集約されており、それ以外の部分は汎用的な構成となっている。そして、そのデータテーブルを適切に可視化するグラフ種を決定する。この際、機械学習により構築したグラフ種判別知識を用いている。最後に、比較フレームの情報に基づいて強調点と軸を設定する。これはユーザの注目点や値に対する意味付けに応じてグラフを適切に変化させる処理である。この手法を用いることで、ユーザはグラフ描画に必要なパラメータを意識せずに、要求を満たすグラフを効率的に得られる。

Automated Visualization Method of Numerical Data from a Caption

MITSUNORI MATSUSHITA,[†] HAYATO YONEZAWA^{††}
and TSUNEAKI KATO^{†††}

We propose a new visualization method that automatically creates a statistical chart from a large amount of numerical data corresponding to a graph's caption given by a user. In this method, an adequate chart is generated by the following steps: understand the graph's caption written in a natural language, select the necessary data from the observed data, convert and aggregate the selected data to a suitable granularity, and visualize the data with an appropriate statistical chart style. In this paper, we introduce a semantic frame in order to formally deal with the caption. The semantic frame consists of two subframes: one is mainly for selecting and restructuring data, and the other is mainly for choosing an appropriate chart-type and emphasizing focal characteristics. We also introduce an aggregation algorithm to create the necessary datatable. In this algorithm, a propagation network is used to manage all observed data; this network is constructed by focusing on the functional dependencies among data and knowledge. We classify the knowledge that is necessary for drawing a chart into two categories: hierarchical domain knowledge, and inter-domain knowledge. The former knowledge describes the subordinate relations between elements that belong to the same domain, whereas the latter knowledge describes the relation between two elements each of which belongs to a different domain. By using this method, a user can get an adequate chart from an enormous amount of data without considering the parameters necessary for drawing the chart.

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環境の充実により、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。たとえば、過去 10 年分の各市町村の降水量がデータとして蓄えられているとき、あるユーザは昨年の関東地方と近畿地方の年間降水量を比較するためにこれを使用するかもしれないし、他のユーザは京都市の降水量の推移を季節ごとに分析するためにこれを利用するかもしれない。可視化はこのような膨大なデータを直

[†] 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所

NTT Communication Science Laboratories, NTT Corporation

^{††} 西日本電信電話株式会社法人営業本部

Business Communications Headquarters, NTT West Corporation

^{†††} 東京大学大学院総合文化研究科言語情報科学専攻

Language and Information Sciences, Graduate School of Arts and Sciences, The University of Tokyo

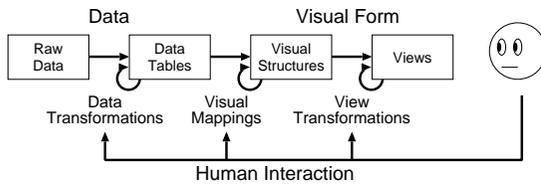


図 1 可視化のモデル

Fig. 1 Reference model for visualization.

観的に理解するうえで有効な手段の 1 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²⁾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可視化構造（グラフ種や配色、位置など）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。

各プロセスにおけるユーザインタラクションの負荷を軽減するために、多くの手法が提案されている。たとえば、data transformation におけるデータの選択を容易にする手法^{1),3)}、visual mapping においてグラフ種の決定や線種などのグラフオブジェクトの選択を自動化する手法^{6),11)}、view transformation において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手法¹⁴⁾などがその一例である。

これらの先行研究では主として「どのように描画するか (how to draw)」という点に興味集中しており、各プロセスに対する直接的な操作が考えられている。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、様々な知識が前提とされる。さらに、描画する意図が異なれば適切なグラフは異なるので、データのどの部分に着目し、それをどのような詳細度で見たいのか、データのどの側面を強調したいのか、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。したがって、得られるグラフの質はユーザの能力に依存するので、場合によっては効果的でないグラフを描画してしまうことも考えられる。

そこでこのユーザの負担を軽減するために、日常的な言葉で表現された「何を描画したいか」から可視化

の 3 つのプロセスでの処理内容を決定し、自動的に可視化する方式について提案する。

本稿では、まず 2 章で「何を描画したいか」を表すものとしてグラフの表題に着目し、それを理解し可視化のプロセスへと変換するという提案方式の概要を述べる。その後、3 章から 5 章で方式の詳細を説明する。最後に 6 章で考察する。

2. 提案方式の概要

前述のように本研究での課題は、「何を描画したいか」というユーザの意図を表現している表題から、適切かつ効果的な可視化、つまりグラフの描画を行うことである。具体的には (1) 可視化の 3 つのプロセスに与えるべきパラメータを決定するための情報を表題から適切に抽出する仕組みを用意すること (2) 可視化プロセスにおいて、人間が行っている処理やそこで用いている知識 (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) を明らかにし、それをシステムに代替させること、の 2 点が課題となる。

第 1 の点を詳しく見るために、グラフの表題について考えてみる。Zelazny によれば、グラフの表題は何を描画するかを示す topic である場合と、それに加えて何を主張したいかについても言及した message である場合とがある²¹⁾。グラフの表題の例を以下に示す。

- (1) 電器メーカー各社の 1997 年の月間出荷数
- (2) 1995 年の近畿地方における降水量は京都府が特に多かった

この例では (1) が topic 型の表題であり、(2) が message 型の表題である。グラフの描画という観点からこれらの表題を比較すると (1) と (2) が共通して指定しているのは、データのどの部分をどのような詳細度で取り上げるべきかという data transformation に関する指示と、そのデータのどの側面 (aspect) に着目して描画すべきかという visual mapping に関する指示である。これらは「何を描画するか」に関する情報と位置づけられる。そして (2) ではこれらに加えて、そこから何を読み取らせるべきか、読み手に何に注目してもらいたいかという view transformation に関する指示が含まれる。これらは「何を主張したいか」に関する情報となる。本研究では、自然言語で表現されたグラフの表題を解釈し、これらの情報を適切に抽出するために、意味フレームと名付けた表現を提案する。

なお、Zelazny は、グラフの表題にはできるだけ message 型の文を採用すべきであると述べているが、新聞記事や白書などで用いられているグラフを見ると、

現状では topic 型の表題が多く用いられている．この点を考慮し，本稿では両方の形式の表題を取り扱うことにする．

第 2 の点は，意味フレームとして抽出された情報から可視化プロセスへ渡すパラメータを決定する仕組みを構築することである．

Data transformation へ渡すべきパラメータは，観測データからユーザが必要かつ適切と考える範囲と詳細度を持ったデータテーブルを構成するための情報であり，これを得る処理をデータ集約処理と呼ぶ．本研究では，いくつかの変数によってその数値が関数的に定まるような，いわゆる統計データを対象とし，統計データの処理で重要な問題となる，それぞれの変数の値の間の階層性に関する集約処理とそれに関連するデータの網羅性の問題を扱った．より一般的な関係データは今回の対象とせず，集約処理においても，集約の際にどのような演算が必要となるか（たとえば，県の人口密度は市の人口密度だけからは集約できず，市の面積が必要になる）という集約演算の知識については扱っていない．

Visual mapping へのパラメータはそのデータテーブルを適切に可視化するグラフ種であるとした．ここでは，統計データの可視化に用いられる様々なグラフ種の中から適切なものを選択するという枠組みを考え，意味フレームの内容から適切なグラフ種へのマッピングを可能とするグラフ種判別知識を，機械学習により構築することを試みている．

最後の view transformation では，ユーザから与えられた message 型の表題において何に注目すべきか，その値から何を読み取らせたいのか（たとえば 1 円の変化は激しいものなのか，それともほとんど無変化であるのかという意味付けや評価）をより効果的に表示するための軸の設定と注視点の明示をパラメータとした．この 2 つは，グラフ作成における強調の代表的な手法である¹⁴⁾．

以上をまとめると提案方式は図 2 のようになる．

3. フレームによる表題の理解

ユーザから与えられたグラフの表題からグラフ描画の各プロセスにおける処理内容を決定するには，まずその表題からグラフ描画に必要な情報を抽出しなくてはならない．そこで，計算機がこの表題を取り扱う際の形式的表現として意味フレームと呼ぶ表現を提案する．このフレームは話題項目フレームと比較フレームという 2 つの部分フレームから構成される．話題項目フレームには「何を描画するか」についての情報が記

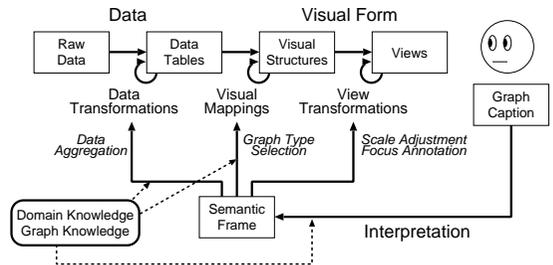


図 2 提案する可視化のモデル

Fig. 2 Proposed model for automated visualization from a graph caption.

述され，比較フレームには「何を主張したいか」についての情報が記述される．したがって，表題 (1) は話題項目フレームのみを用いて表現されるが，表題 (2) は両方のフレームを用いて表現される．

3.1 話題項目フレーム

ユーザから与えられた表題中で言及されている「何を描画するか」についての情報は話題項目フレームに記述される．話題項目フレームは表題の内容に応じて複数の行を持つ．各行が描画に必要な情報の断片を表現しており，その組合せによって描画するグラフが決定される．この各行を「描画要素」と呼ぶことにする．

表題の中で，何を描画するかに関する指定の部分は「90年の近畿地方の降水量」のようにキーワードの羅列に近く，それらの意味関係を推測する言語的な手掛かりは少ない．この点から，話題項目フレームでは表題中で言及されている対象について各々以下の情報を抽出し，それらの関係付けはデータ集約処理で行う．

蓄積されている観測データの中から必要な部分を取り出すには，表題中で言及されている対象とそのとりうる範囲が分かればよい．また，ユーザが要求する詳細度はデータテーブル中に蓄積されているデータの詳細度と必ずしも一致しないので，この違いを吸収するために蓄積データを適切な詳細度に変換し集約する必要がある．取り出したデータを適切な詳細度で集約するには，各描画要素について，ユーザの求めている詳細度が分かればよい．そこで，話題項目フレームにはこれらの情報を記述する．すなわち各描画要素を変数，ドメイン，粒度，型，制約条件の 5 つの-slot で表現する．各-slot の詳細を以下に示す．

- 変数

変数slotは描画要素を識別するための変数を値とする．

- ドメイン

ドメインslotはドメイン名を値とする．ここで，slot値となりうるドメイン名は後述する

変数	ドメイン	粒度	型	制約条件
X	産業	会社	名義	$x_i \in$ 電器メーカ
Y	日付	年月	時間	$y_j \in$ 1997
Z	出荷数	台	量	

図 3 表題 (1) の意味フレーム表現

Fig. 3 “What to draw” frame for caption (1).

ドメイン階層知識から得られる。

すべての描画要素は表題中に含まれる名詞連続によって特定される。たとえば，“降水量”，“1995年”，“近畿地方”，“京都府”といった名詞連続が表題 (2) 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“近畿地方”と“京都府”は同じ場所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

粒度スロットはユーザが求めている描画要素の粒度 (granularity⁴⁾) を値とする。粒度とは描画要素の記述単位のことである。たとえば，“京都府”は日本を都道府県の単位で記述した際のラベルであり，その属する粒度は県である。

このスロット値となりうる粒度は後述するドメイン階層知識から得られる。

● 型

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，各データ型はさらにいくつかの部分型に分類されている²⁾。型を詳細に区別することはグラフ種を適切に判定するうえで有効である半面，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，型スロットがとりうる値の種類を名義，時間，順序，量，区間，割合の各型に限定した。

● 制約条件

制約条件スロットは表題の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., $y_1 =$ 京都府)，上位概念指定型 (e.g., $y_i \in$ 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。

これらの定義から，たとえば表題 (1) (2) は各々図 3，図 4 に示すフレーム表現に変換される。

3.2 比較フレーム

message 型の表題に基づいてグラフを描画する場合，ユーザが主張したい内容によって適切なグラフは

変数	ドメイン	粒度	型	制約条件
X	日付	年	時間	$x_1 =$ 1995 年
Y	場所	県	名義	$y_i \in$ 近畿地方
Z	降水量	mm	量	

図 4 表題 (2) の話題項目フレーム

Fig. 4 “What to draw” frame for caption (2).

主題部位	$Y_k = \{ \text{京都府} \}$
比較対象	$Y - Y_k$
比較内容	多い

図 5 表題 (2) の比較フレーム

Fig. 5 “Focal point” frame for caption (2).

異なる¹⁴⁾ ので，ユーザ意図に沿ったグラフを得るには，ユーザが何を主張したいかという情報をグラフに反映する必要がある。そこで，この「何を主張したいか」に関する指定の部分を表現するために比較フレームを導入する。このフレームは主題部位，比較対象，比較内容という3つのスロットから構成される。各スロットの詳細を以下に示す。

● 主題部位

このスロットが値としてとるのは表題の注目点である。たとえば表題 (2) では“京都府”の降水量が注目点である。

● 比較対象

比較対象は主題部位が何と比較されているかを表す。たとえば，表題 (2) において，主題部位 (= 京都府) の比較対象は近畿地方の他の県である。

● 比較内容

比較内容は主題部位の比較対象に対する程度を示す程度表現を表す。

Message 型の表題の中には“急激に伸びた”のように主題部位の比較対象に対する様子を表す表現が含まれる。同じデータが対象であっても，この表現が“わずかに伸びた”の場合と“急激に伸びた”の場合では，ユーザの意図が異なる。この意図の違いをグラフに反映するために，このスロットの値に基づいて view transformation の処理を行う。その方法については後述する。

これらの定義から，たとえば表題 (2) の比較フレームは図 5 のようになる。

自然言語で表現されたグラフの表題から意味フレーム表現を得るためには，いわゆる意味文法と呼ばれる枠組み⁹⁾を用いた。これは表題の表現を，スロットを持つパターンとして記述し，これと分野知識を含んだ単語辞書を用いて，パターンマッチによりスロットを埋めていき，スロットの値を意味フレームの適当な部分の内容とするものである。たとえば，

(message 表題) → (topic 表題) は (条件値) が (比較表現)

(topic 表題) → ((条件値) { の | おける })⁺ (ドメイン名)

のパターンによって表題(2)が解釈される。なお、このパターン中の + 記号は Kleene plus である。(条件値)には“1995年”, “京都府”, “近畿地方”などの要素がマッチする。これがどのドメインのどの粒度に属する要素であるかは後述するドメイン階層知識を参照することで決定する。主に第1のパターンのスロットである(条件値)(比較表現)に埋まった値から比較フレームが, 第2のパターンの(条件値)(ドメイン名)から話題項目フレームが作られる。実装においては, 代表的な表題の表現について, 約40のパターンを用意した。

以降の章では, これら2つのフレームを用いて表現された表題の情報から決定される描画の各プロセスでの処理内容について述べる。

4. データ集約処理

データ集約処理では前章で提案した話題項目フレームに基づいて蓄積されている観測データの中から必要なデータを選択し, 適切な詳細度に変換して集約する。

この章ではまず集約に必要な知識について考察し, ドメイン階層知識とドメイン間関係知識の2種類の知識を導入する。そして, これらの知識からデータの集約可能性を判断するための伝播ネットワークを構築する方法と, 観測データの集約方法を説明する。

4.1 集約に必要な知識

本研究では, 各店舗の売上げや県ごとの降雨量などのように逐次追加されていく観測値を対象とする。これらはリレーションとして観測データテーブルに蓄積されるものとする。表1および表2に観測データテーブルの一例を示す。

議論を簡単にするため, 本稿では属性 A_1, \dots, A_{n-1} が互いに独立な候補キー属性で属性 A_n が非キー属性であり, A_n のみが量データであるようなデータテーブルを対象にして議論を進める。たとえば, 表1のデータテーブルでは“県”と“観測日時”が候補キー属性で“降水量”が非キー属性であり, 表2のデータテーブルでは“出荷品目”と“出荷月”が候補キー属性で“出荷数”が非キー属性である。本稿では d_{ij} を属性 A_i の要素とし, それらの関係を

$$\text{elem}(A_i) = \{d_{i1}, \dots, d_{im_i}\} \quad (1)$$

と表記する。ただし, $\text{elem}(A_i)$ の各要素は同一の粒度で記述されていると仮定する。

この観測データテーブルをユーザから与えられた表題に応じて集約する際, 観測データテーブル中には

表1 観測データテーブルの例(各県の降水量)

Table 1 An example of observed data (precipitation).

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

表2 観測データテーブルの例(電器製品の月間出荷数)

Table 2 An example of observed data (shipping amount of electrical appliances).

出荷品目	出荷月	出荷数
製品 A	1997-01	1,100
製品 B	1997-01	4,200
製品 A	1997-02	1,300
製品 B	1997-02	3,100
...
製品 C	1997-12	2,100
製品 D	1997-12	1,400

与えられた表題に適切な範囲と詳細度を持ったデータテーブルを構成するために必要な情報がすべて含まれているとは限らないため, 不足している情報を補う必要がある。たとえば表1において, 降水量を関東地方や関西地方などの地方ごとに集約するには, 各県がどの地方に分類されるかという情報が必要である。また表2において, 各製品の出荷数をその販売会社ごとに集約するには, 製品と会社の対応関係に関する情報が必要である。

これらの情報を補うには, 同一ドメインに属する要素間の包含関係についての知識と, 異なるドメインに属する要素間の対応関係についての知識を用意しなくてはならない。そこで, 前者をドメイン階層知識, 後者をドメイン間関係知識として以下のように定式化する。

ドメイン階層知識は, 様々な詳細度で記述されている要素を, その単位を粒度と見なして階層化したものである。ドメインの各要素は, その単位に応じた粒度に属する。図6にその例を示す。たとえば場所ドメインの要素の1つである“京都府”は“都道府県”の粒度に属する要素である。

粒度 G_i に M 個の要素 $d_{ix} (1 \leq x \leq M)$ が属するとき, それらの関係を

$$\text{gr}(G_i) = \{d_{i1}, \dots, d_{iM}\} \quad (2)$$

と表記する。任意の $G_i, G_j (i \neq j)$ について

本稿では同一カテゴリに分類される要素の集合を「ドメイン」と呼ぶ。たとえば“近畿地方”や“京都府”は場所ドメインの要素である。

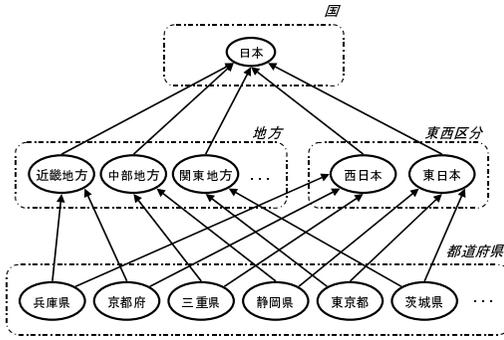


図 6 ドメイン階層知識の例
Fig. 6 Hierarchical domain knowledge.

$$gr(G_i) \cap gr(G_j) = \phi \tag{3}$$

であり、ドメイン D とそのドメインに属する粒度 $\{G_1, \dots, G_n\}$ の間には

$$D = gr(G_1) \cup \dots \cup gr(G_n) \tag{4}$$

という関係が成り立つものとする。

本稿では、粒度間の半順序関係を \sqsubseteq 、要素間の半順序関係を \preceq と表し、「 G_j は G_i より粗い」は $G_i \sqsubseteq G_j$ と表記する。このとき G_j を G_i の上位粒度と呼ぶ。また、各々の粒度に属する要素間の関係は $d_{ix} \preceq d_{jy}$ と表記する。

$G_i \sqsubseteq G_j$ のとき、すべての $d_{ix} \in gr(G_i)$ について $d_{ix} \preceq d_{jy}$ となる $d_{jy} \in gr(G_j)$ が存在するものとする(関数従属)。このとき上位粒度 G_j の要素 d_{jy} に対応する下位粒度 G_i の要素集合を

$$div_{G_i}(d_{jy}) \stackrel{\text{def}}{=} \{d_{ix} \in gr(G_i) | d_{ix} \preceq d_{jy}\} \tag{5}$$

と表す。 $G_i \sqsubseteq G_j$ かつ $G_i \sqsubseteq G_k \sqsubseteq G_j$ となる G_k が存在しないとき、そしてそのときに限り G_i から G_j へとリンクを張ることで、粒度の階層関係が得られる。得られた階層関係は後述する伝播ネットワークの構築に用いられる。

ドメイン間関係知識は各々異なるドメインに属する 2 つの要素集合の関係を示す。表 3 は会社と製品の関係を表したドメイン間関係知識の例である。本稿では、この知識の各カラムは単一粒度で記述されると仮定するので、この知識は異なるドメインに属する 2 つの粒度間の関係と見なせる。

この関係は関数従属を前提とする。すなわち、あるドメイン D_i 中の粒度 G_i に属するすべての要素 $d_i \in gr(G_i)$ に対して別のドメイン D_j 中の粒度 G_j に属する要素 $d_j \in gr(G_j)$ が関数従属関係にあるとき、これをドメイン間関係知識として扱い、 $G_i \rightarrow G_j$ と表す。

なお、本稿で実装したシステムにおいては、これら

表 3 ドメイン間関係知識の例
Table 3 Inter-domain relation knowledge.

会社	製品
会社 X	製品 A
会社 X	製品 B
会社 Y	製品 C
会社 Z	製品 D
会社 Z	製品 E

の 2 種類の知識は主として人手で作成したものを用いた。たとえば場所に関するドメイン階層知識は、郵便番号辞書をもとにして基本的な階層の知識(地方-県-市町村の 3 階層)を機械的に作成した後、人手で不要なノードの削除や異表記の追加などを行うことにより作成した。

これらの知識は 2 つの要素集合間の関数従属関係に基づいているため、既存のデータベース中に存在する要素集合間のリレーションを流用して自動的に作成することが考えられる。ただしその際には、網羅性や排他性の保証、粒度の均一性の確認などが必要となるため、具体的な自動作成方法については今後の課題である。

4.2 伝播ネットワークの構造

与えられた表題を満たすデータテーブルを作成するために、図 7 に示す伝播ネットワークを用いて観測データテーブルを拡張する。このネットワークは前述した 2 種類の知識と観測データテーブルから事前に作成する。

このネットワークには候補キー属性ノード、非キー属性ノード、粒度ノードの 3 種類のノードが存在する。候補キー属性ノードと非キー属性ノードは観測データテーブルの属性と対応し、粒度ノードはドメイン階層知識の粒度に対応する。これらのノードは関連リンク、粒度リンク、ドメイン間リンクの 3 種類のリンクで結合されている。

粒度リンク(図 7 中の実線)はドメイン階層知識の粒度関係を表すリンクであり、ドメイン間リンク(図 7 中の一点鎖線)はドメイン間関係知識に基づいて、異なるドメイン階層知識の粒度間に張られたリンクである。本稿では、この 2 つのリンクで結合されている部分ネットワークは、タスクや観測データの内容によって変化しない静的な知識であると仮定している。

関連リンク(図 7 中の点線)は、各候補キー属性ノード A_i から $elem(A_i) \subseteq gr(G)$ を満たす粒度 G

図 7 では各ドメインのノードの一部を示している。たとえば日付ドメインについては、Randall らが示したような知識¹⁶⁾を用いている。

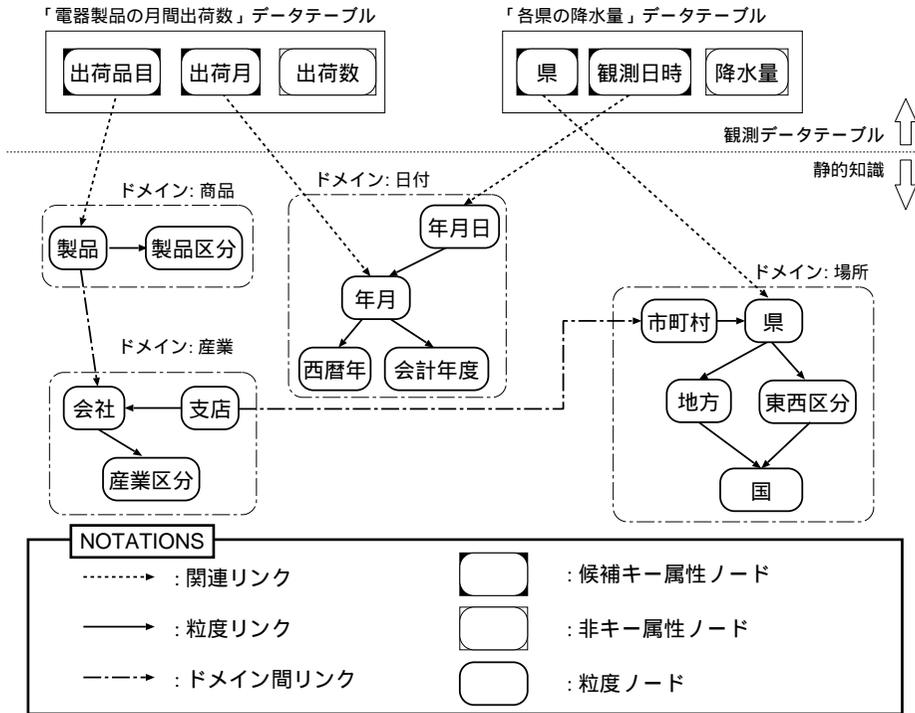


図 7 伝播ネットワーク

Fig. 7 Propagation network.

に張られたリンクである。 $elem(A_i)$ は単一粒度で記述されているので、ただ 1 つの粒度ノードと対応し、その関係は $A_i \rightarrow G$ と見なせる。もしこの条件を満たす粒度ノードが存在しなければ、知識が不足しているか、属性の要素が単一粒度で記述されていないことが原因と考えられる。図 7 から分かるように、表 1 および表 2 の観測データテーブルにはどちらも日付に関する属性が含まれているが、粒度が異なるためリンクが張られる先の粒度ノードは異なっている。

このネットワークのリンクはいずれも関数従属関係に基づいているため、ノード A, B, C の間に、 $A \rightarrow B$ と $B \rightarrow C$ が成り立つならば、推移律により $A \rightarrow C$ も成り立つ。すなわち A に属する要素に対応する C の要素が特定できる。

4.3 観測データテーブルの拡張

本研究では 2 次元の統計グラフによる可視化を対象としているので、その描画に必要なスプレッドシート形式の表を得ることが集約の目的である。この表を集約テーブルと呼ぶ。意味フレーム中の各描画要素を満たす要素を持つ集約テーブルを作成するために、まず伝播ネットワークを用いて観測データテーブルから求める集約テーブルを作成できるかを判断する。これは以下の手順による。

- (1) 意味フレーム中で量型の描画要素に注目し、その描画要素のドメインと一致する非キー属性を持つ観測データテーブルを選択する。ただし、このようなデータテーブルは複数存在するので、それらすべてを候補データテーブル集合として扱う。これらのデータテーブルの候補キー属性に対応するノードをトークンの伝播開始ノードとする。
- (2) 意味フレーム中で量型でない描画要素に注目し、それらのドメインと粒度からゴールノードを決定する。
- (3) トークンを伝播開始ノードから、リンクに沿ってネットワーク中を伝播させる。もしトークンが伝播開始ノードからゴールノードに到達可能であれば、伝播開始ノードの属性とゴールノードの属性が関数従属関係にあることが分かるので、その伝播開始ノードに対応するデータテーブル中の属性を、ゴールノードが示す粒度に変換することが可能だと分かる。そこで候補データテーブルのうち、すべてのゴールノードに到達可能な伝播開始ノードを有するものを、集約テーブルを構築するための基本データテーブルとして採用する。

- (4) この基本データテーブルに対し、その各候補キー属性ノードからゴールノードに至る最短パス上のリンク(リレーション)を自然結合演算²⁵⁾を適用して結合する。たとえばトークンが候補キー属性ノード A_i から G_k を經由してゴールノード G_j に到達した場合

$$R[A_i = G_k] S[G_k = G_j] T \quad (6)$$

という演算を実行する。ここで R は観測データテーブルであり、 S はリレーション $S(A_i, G_k)$ 、 T はリレーション $T(G_k, G_j)$ を意味する。

以下に図7の伝播ネットワークと図3に示した表題(1)の意味フレーム表現を用いて具体例を示す。図3では、 Z が量型の描画要素であり、そのドメインが出荷数なので「電器製品の月間出荷数」データテーブルが候補データテーブルとして選択される。

ゴールノードは、描画要素 X と Y が量型でない描画要素なので、それらの描画要素のドメインと粒度の値をもとに決定する。すなわち、産業ドメインの会社粒度ノードと日付ドメインの年月粒度ノードである。

トークンはデータテーブル中の候補キー属性ノードである出荷品目ノードおよび出荷月ノードから伝播を開始する。出荷品目ノードから伝播するトークンは商品ドメインの製品粒度ノードを經由して産業ドメインの会社粒度ノードに到達する。また、出荷月ノードから伝播するトークンは日付ドメインの年月粒度ノードに直接到達する。したがって、「電器製品の月間出荷数」データテーブルが基本データテーブルとして採用される。

次に製品、会社、年月の各属性を追加するために、このデータテーブルに対して以下の自然結合演算を適用する。

$$\begin{aligned} R[\text{出荷品目} = \text{製品}] S \\ [\text{製品} = \text{製品}] T \\ [\text{出荷月} = \text{年月}] U \end{aligned} \quad (7)$$

ここで R, S, T, U は各々「電器製品の月間出荷数」データテーブル、 S (出荷品目、製品) 関係、ドメイン間関係知識中の T (製品、会社) 関係、 U (出荷月、年月) 関係である。これにより、表4に示すデータテーブルが得られる。

表4 拡張されたデータテーブル
Table 4 Extended datatable.

会社	製品	年月	出荷数
会社 X	製品 A	1997-01	1,100
会社 X	製品 B	1997-01	4,200
会社 X	製品 A	1997-02	1,300
会社 X	製品 B	1997-02	3,100
...
会社 Y	製品 C	1997-12	2,100
会社 Z	製品 D	1997-12	1,400

4.4 データテーブルの集約

次に必要なのは、上記の処理により得たデータテーブルから、描画要素のすべての制約条件を満たすタプルを選択することである。たとえば図3のフレームに基づいて描画する場合、「 $x_i \in$ 電器メーカ」と「 $y_j \in$ 1997年」を満たすタプルが選択されなければならない。

基になるデータテーブルはドメインのすべての要素の網羅性を保証しないので、たとえばデータテーブルが場所ドメインについて京都府の情報しか有していなければ、近畿地方への集約は他県の情報不足のためできない。しかし、もしユーザが京都府についてのみ知りたいのであれば集約可能である。そこで、表題の制約条件に対する網羅性をチェックし、条件を満たすすべてのタプルを取り出すことが必要である。すなわち、制約条件の直積がデータテーブルに含まれているかを調べなければならない。実装では、上位概念指定型の制約条件はドメイン階層知識を参照してデータテーブルの最詳粒度の要素に展開される。たとえば、図3では、「 $x_i \in$ 電器メーカ」は「 $x_i = \{$ 会社 X, 会社 Y, ..., 会社 Z $\}$ 」に展開される。そのため、図3中の制約条件の直積は以下ようになる。

$$\begin{aligned} (X \text{ の制約条件}) \times (Y \text{ の制約条件}) \\ = (\text{会社} \in \text{電器メーカ}) \times (\text{年月} \in 1997) \\ = \text{div}_{\text{会社}}(\text{電器メーカ}) \times \text{div}_{\text{年月}}(1997) \\ = \{\text{会社 X, 会社 Y, \dots, 会社 Z}\} \\ \times \{1997-01, 1997-02, \dots, 1997-12\}. \end{aligned}$$

このチェックによってデータテーブルに必要なタプルの存在を確認したのち、余分な属性を削除する。もしタプルが不完全であれば、システムはユーザにデータの欠落があることを伝える。候補キーノードからゴールノードへ至るパス上にあるノード(上記例では製品粒度ノード)は余分なノードと見なされる。なお、これらの属性がこの処理の時点まで残っているのは、網羅性チェックが必要となる可能性があるからである。

上記の処理の後、すべての候補キー属性で同じ要素を持つタプルを1つに集約し、複数の要素を持つ属性の要素を表頭、表側とした2次元テーブルを作成する。

厳密には、自然結合演算は2つのリレーションが同一の属性名を有するときのみ適用可能であり、そうでない場合は、より一般的な演算である等結合演算を適用しなくてはならない。しかし、関連リンクで結合されている属性と粒度はその定義から同じ実体を示していることが明らかなので、等結合演算の代わりに自然結合演算を適用している。

表 5 集約テーブル
Table 5 Aggregated datatable.

	1997-01	1997-02	...	1997-12
会社 X	24,000	27,300	...	16,800
会社 Y	13,600	14,200	...	23,300
...
会社 Z	21,800	14,300	...	13,500

(e.g., 表 5) . この処理によって得られるのが、与えられた表題に最も適した集約テーブルである。なお複数の要素を持つ属性が 2 つある場合、集約テーブルの表頭と表側は、時間型ないし順序型の属性がある場合はそれを表頭にし、そうでないときは要素数が多い属性を表頭にするというヒューリスティクスを用いて決定している。

5. 統計グラフの生成

前章までの処理で、ユーザから与えられた表題を満たす集約テーブルが得られた。次はこの集約テーブルとフレーム表現された表題から適切なグラフを決定し、最後に軸や配色などのパラメータを決定する。

本章ではまずグラフ種を決定するのに必要な知識について述べ、提案した意味フレームに記述される情報がグラフ種を選択するのに十分であることを示す。その後、統計グラフを描画する際のパラメータの決定について述べる。

5.1 グラフ種決定に必要な知識

前章までの処理で得られた意味フレームおよび集約テーブルから描画するグラフ種を決定するには、これらから得られる情報に基づいてグラフ種を決定する知識が必要である。

一般にグラフ種は描画するデータの特徴や性質を考慮して決定される。Mackinlay や Fasciano らの先行研究では、グラフ種を決定する知識はヒューリスティクスによって構築されているが、このように構築された知識では網羅性や妥当性の保証が困難である。そのため、我々は新聞や白書などで実際に使われているグラフをもとに、機械学習によりグラフ種判別知識を構築した。このときグラフ種を特定するための決定要因として、ヒューリスティクスによって構築された知識を参考に、意味フレームおよび集約テーブルから獲得可能なものという観点から“データ型の組”、“単/複系列”、“集団性の有無”、“100% 割合”、“強調部位の

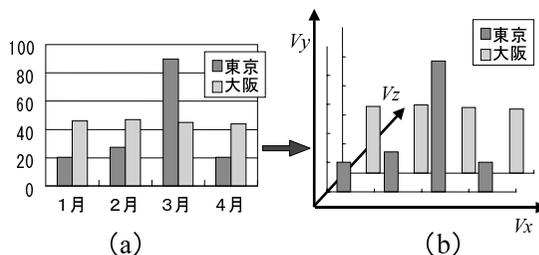


図 8 グラフの分解
Fig. 8 Chart decomposition.

特徴”の 5 つを選択した。

“データ型の組”とは各軸の持つ性質の組のことであり、量、割合、名義、順序、区間、時間、無のいずれかの値をとる。これら軸の持つ性質は描画する対象データの型を反映したものである。しかし複数棒グラフや積上げ棒グラフなどのように描画されるデータ系列が複数ある場合、そのデータ系列間の性質は軸として表現されない。そこでこれを表現するためにグラフを 3 次元に展開して扱った。たとえば図 8 中の (a) のグラフは (b) に示すように展開する。これにより、データ系列が複数ある場合にはデータ系列間の性質は V_z 軸の性質として表現される。

“単/複系列”とは描画されるデータ系列が単系列か複数系列かを区別するものである。たとえば図 8 の (b) では、東京と大阪という 2 つのデータ系列が描かれているため複数系列になるが、東京についてのみ描画した単純棒グラフでは、単系列となる。

“集団性の有無”とは、描画されるデータ系列が複数ある場合に、それらをいくつかの集団に分割して式 (5) を満たすような上位粒度の要素で置き換えられるかどうかを表す。たとえば県ごとの人口を描画する場合、東京都と大阪府を比較するグラフでは式 (5) を満たす上位粒度の要素は存在しないので集団性はないと判断するが、徳島県、愛媛県、香川県、高知県を描画する場合には四国地方という上位粒度の要素で置き換えることができるので、集団性があると判断する。

“100% 割合”とは軸属性が割合である場合に、その軸で表現されるデータの値の合計が 100 になるかどうかで判断する。

“強調部位の特徴”とはグラフ中の強調される部位がどの軸に存在するかを表すものであり、 V_x , V_z , 無のいずれかの値をとる。

これらのグラフ種決定要因を用いると、図 8 中の (a) のグラフは図 9 のようにコーディングできる。このようにコーディングしたグラフを新聞記事から収集

本稿では 2 次元の基本的な統計グラフを対象としているため、複数の要素を持つ属性が 3 以上ある場合には可視化構造を一意に決定できない。そのため実装では、このような場合にはユーザにさらなる制約条件の入力を促すようにしている。

- (ID0001
- (グラフ種 複合棒グラフ)
- (データ型の組 時間・名義・量)
- (単/複系列 複数系列)
- (集団性の有無 無)
- (100%割合 No)
- (強調部位の特徴 無))

図 9 サンプルデータ

Fig. 9 Feature expression of Fig. 8(a).

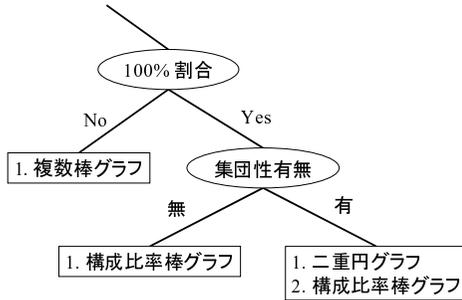


図 10 優先順位付き決定木

Fig. 10 Decision tree with priority levels.

し、各決定要因を分類属性、グラフの種類を分類クラスとして ID3 アルゴリズム¹⁵⁾を用いて決定木を作成する。これによって得られる決定木がグラフ種判別知識である。

このとき、作成された決定木に複数のグラフ種が上記の各決定要因で一意化できずに同一の葉ノードを構成する場合は、図 10 に示すようにそれらの候補に順位を付けることとした。この優先順位は、その葉ノードにたどりついたグラフの数により決定する。

5.2 グラフ種判別能力の検証

5.1 節の処理で得られたグラフ種判別知識によってどの程度正確に描画すべきグラフ種を特定できるかが本手法の精度を決定する。そこで、本節ではこの知識によるグラフ種判別能力について検証する。

意味フレームに記述される情報と、4 章の処理で得られた集約テーブルからグラフ種を決定するには、これらから上記の決定要因を取り出して、グラフ種判別知識を適用すればよい。

“データ型の組”は話題項目フレームの型スロットから求められる。たとえば図 3 の場合は(時間, 名義, 量)である。

“単/複系列”は集約テーブルから求められる。もし集約テーブルが $m \times 1$ (ただし $m > 1$) のスプレッドシートならば単系列であり、表 5 のように $m \times n$ (ただし $m, n > 1$) ならば複数系列である。

“集団性の有無”は、集約テーブルの表側と話題項目

フレームの制約条件から求められる。もし表側となった描画要素の制約条件が上位概念指定型で記述されているときは集団性があると判断し、それ以外であればないと判断する。

“100% 割合”は集約テーブルの値を計算することで判断できる。

“強調部位の特徴”は集約テーブルと比較フレームの主題部位から判断できる。すなわち、強調部位の要素が集約テーブルの表頭に属するのであれば V_x 、集約テーブルの表側に属するのであれば V_z 、topic 型の表題のように比較フレームがない場合は無と判断できる。

選択したグラフ種決定要因によるグラフ種判別能力を確認するため、以下の実験を行った。

まず、5.1 節で示したグラフ種決定要因に基づいて、1999 年 6 月から 7 月の 7 社分の新聞記事と 1997 年度通信白書で用いられているグラフから、265 個のサンプルデータ(折れ線グラフ 81, 棒グラフ 68, 円グラフ 44, 複合棒グラフ 30, 積上げ棒グラフ 18, 二重円グラフ 13, 構成比率棒グラフ 11)をコーディングした。

これらを 132 個の訓練パターンと 133 個のテストパターンにランダムに 2 分割し、訓練パターンによって作成した決定木に対して、テストパターンによる評価実験を行った。

この結果と、Visage などの従来手法で一般的に用いられている判断基準(x 軸の属性, y 軸の属性, 隠れ軸の属性, 依存関係, 100% 割合)を基に人手で作成したルールによるグラフ種判別能力とを比較したものが表 6 である。ここで、“一意に決定”とは優先順位が 1 番目の正しいグラフ種候補に決定したものであり、“候補が決定”とは優先順位に関係なく葉ノードに候補が存在していた場合である。“候補が決定”を正答率の評価に加えたのは、人間がデータの特徴や性質からグラフ種を判断する際に、最も適切なグラフ種以外にも、いくつかのグラフ種が候補となりうる場合が多いことを考慮したためである。

表 6 を見ると、機械学習によって作成した決定木の方が、ルール数(葉数)および 1 ルールあたりの平均グラフ候補数が少ないにもかかわらず、“一意に決定”では良い判別成績を示していることが分かる。この結果から、選択したグラフ種決定要因はグラフ種を判別するのに十分であると考えられる。

また“候補が決定”では人手で作成したルールの判別成績が優れているが、これは人手で作成したルールの方が、1 ルールあたりのグラフ種の数が多いため

表 6 機械学習で得られた決定木と人手ルールとのグラフ判別能力の比較
Table 6 Comparison between generated decision tree and heuristic rules.

	人手ルール	機械学習で得られた決定木 (200 回試行)		
		average	minimum	maximum
ルール数	32	24.12	19	29
平均グラフ数/1 ルール	3.25	1.341	1.208	1.526
一意に決定	58.9%	74.3%	62.7%	83.9%
候補が決定	96.2%	91.2%	81.4%	97.5%

ある。本研究で対象とする課題では、ユーザからのインタラクションを極力少なくしてシステムに代替させることを目指しているため、グラフ種を“一意に決定”できることが望ましい。その観点から見ると、人手で作成したルールに比べて、機械学習によって作成した決定木の方が適当だといえる。

5.3 グラフの描画

最後に、前章で説明したデータ集約処理によって得られた集約テーブルと意味フレームに基づいてグラフを描画する。

まず visual mapping に相当する処理として、5.1 節で示したグラフ種判別知識を利用してグラフ種の候補を意味フレームと集約データテーブルにより決定する。その後 view transformation に相当する処理として、比較フレームに基づいてグラフ中の重要な点を明らかにする強調方法と強調部位を決定する。

グラフを特徴だたせる効果は様々に存在する¹⁴⁾が、ここでは、色と線の太さ、位置、軸のとり方を利用して表現する。

集約テーブル中で、主題部位スロットの値から強調する対象を特定し、グラフ描画時にいくつかの強調操作を施す。たとえば、棒グラフがそのデータテーブルを描画するのに選択されているとき、主題部位スロットで特定される部位が濃い色と太い線で表示される。それに対して、比較フレーム中の比較対象スロットで特定される部位は、明るい色と細い線で表現される。

さらに、折れ線グラフの場合、適切に Y 軸のスケールを設定することで、グラフ上での変化の傾向が直観的に理解できる。たとえば“急激に伸びた”を表現するには、大きく変化するグラフを描画するために軸の表示幅を小さく設定し、“わずかに伸びた”を表現するには、小さく変化するグラフを描画するために軸の表示幅を大きく設定する²³⁾。この操作は比較フレーム中の比較内容スロットの値に応じて決定される。

最後に、システムはこれらのパラメータを用いてグラフを描画する。これによってユーザの注目点が明示されたグラフが得られる。

我々はプロトタイプシステム KEVIN¹²⁾ を実装した。本システムではグラフを描画する処理に Microsoft Excel を利用している。KEVIN の出力例として、次の 2 つの表題から得られるグラフを示す。

(3) 1995 年 1 月から 1995 年 6 月の神奈川県と東京都の降水量は、6 月の神奈川県が特に多かった。

(4) 1998 年の音声機器と映像機器の月間出荷数。

表題 (3) が入力としてシステムに与えられた際の出力結果を図 11 に示す。両県の降水量が複数棒グラフとして現れている。エッジの太さと配色が強調方法として採用されている。これらの効果により、「1995 年の神奈川県」の降水量が強調されたグラフが描画されている。

表題 (4) が入力としてシステムに与えられた際の出力結果を図 12 に示す。2 種類の機器の月別出荷数が複数折れ線グラフとして描画されている。表題 (4) は比較フレームに関する情報を有していないので強調操作は施されていない。

6. 関連研究との比較

情報可視化は大量のデータを効率的に取り扱うために効果的なアプローチであり、いくつかのシステムが提案されている^{13),22)}。グラフの自動生成の研究としては、統計データを効果的に可視化する研究の初期の成果として Mackinlay の APT システム¹¹⁾ があげられる。APT システムでは関係情報を可視化するのに必要なプリミティブとその合成演算を定義し、それらを利用合成する過程で、グラフ言語が必要な情報を記述できるかを判断する表現基準 (expressiveness criteria) と、グラフ言語が出力メディアと人間の視覚系を活用できるかを判断する効果基準 (effectiveness criteria) に照らし合わせて最適なものを合成するという visual mapping の自動化における基本的な枠組みを提案している。

Roth らは SAGE においてデータの様々な数学的

実装ではコーディングしたすべてのデータを用いて作成した決定木をグラフ種判別知識として採用した。

現行の実装では与えられた表題が正しいかどうかの判断はしていない。

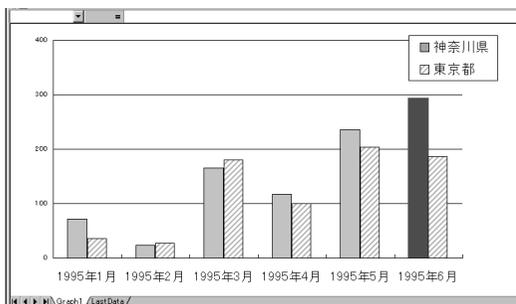


図 11 表題 (3) から得られたグラフ
Fig.11 Output example for caption (3).

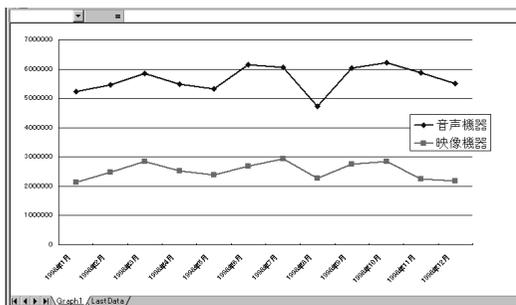


図 12 表題 (4) から得られたグラフ
Fig.12 Output example for caption (4).

な特徴がグラフの適切性に影響を与えることを論じ、グラフ生成の範囲を関係情報からそれ以外の様々なものに広げたことに加え、グラフの選択においてはユーザの情報探索ゴール (information seeking goal) が重要な役割を果たすことを指摘している^{(10),(17),(18)}。このユーザの情報探索ゴールという考え方を発展させ、Casnar は、ユーザがグラフを用いて解こうとしているタスクを反映してグラフが決定されるということを重視し、問題 (タスク) 解決のプリミティブと視覚的な認知のプリミティブとの関係からグラフの合成を試みている。

これらの流れを汲んだ最近の研究成果として、Post-Graphe ではユーザの情報探索ゴールと同一のものであるユーザの意図を前面に出し、それを細かく分類してグラフの設計を行うとともに、それにとまう自然言語による説明文の合成を試みている^{(5),(6)}。また、Green らは、この意図をメディア独立に表現することを試み、そのための内容表現言語として限定量化一階述語論理 (RQFOL) に基づくものを提案している⁽⁷⁾。

これらのシステムは、可視化における visual mapping の部分だけを対象とし、描画に用いるデータテーブル (data transformation の出力) とユーザの意図が曖昧さのない形で与えられることを前提としている。一方、我々は表題の中には visual mapping に関

する指定だけでなく、data transformation や view transformation に関する指示が自然な形で混じりあっていることに着目し、これらの過程全体を表題というものを入力として自動化することを試みている。これら一連の処理を行うことで、データの性質やその粒度がすべての過程で重要な役割を果たすことが明らかとなり、それらを適切な意味フレームで表現することの必要性が明らかとなった。

ユーザの要求を自然に表現するだけで大量のデータから必要なものを選択するという data transformation に関する仕組みは、可視化と離れて、自然言語インタフェースの研究として古くから行われている。たとえば、Woods の LUNER⁽²⁰⁾ や Martin らの TEAM⁽⁸⁾ がその典型であるように、これらの研究は自然言語の構成的意味合成、もしくはその解釈のテストベッドとして行われており、統語と意味の関連や限量子の扱いがその焦点であった。そのため、我々が扱ったような様々な粒度を持つデータ要素から構成され、その統計量が議論の対象となるような統計データベースについては、ユーザの要求中における統計的な手がかりがきわめて少ないこと、粒度について意味表現として採用された一階述語論理での扱いが困難であることから、あまり検討されていなかった。この問題は前述の Green らが提案する内容表現言語でも同様で、様々な粒度を持つデータ要素の扱いが統計データを様々な観点で可視化するという中で重要な役割を果たすと思われるが、彼らはこの点に関して十分な取扱いをしていない。

View transformation については、指導書的な記述は多いものの自動生成という点では Mittal の研究⁽¹⁴⁾ しか見あたらない。そこでは、指導書的な記述が自動生成の文脈で例示されるにとどまっている。我々は topic 型の表題だけでなく、message 型の表題をも取り扱うことで、可視化のプロセスの中での view transformation の意味付けを明らかにした。Roth や Fasciano が扱っている visual mapping におけるユーザの意図やゴールとは異なる種類の観点の指定が message 型の表題には含まれており、それが view transformation に対する指示となる。そのため、表題からこの部分を適切に抜き出し反映させることが重要である。

また、DBSENA⁽²⁴⁾ や OLAP (On-line Analytical Processing) ベースの可視化ツール⁽¹⁹⁾ がユーザの要求によって変化する詳細度の取扱いを試みている。しかしながら、これらのシステムでは自動的に可視化するために必要な知識や方法論が十分に記述されていない。

今回実装したシステムは、統計データを対象とし、visual mapping においても既存のグラフ種の選択とそのパラメータ（軸の割当てや目盛）の決定という枠組みをとっている、すなわち合成せずに選ぶだけなので、可視化の一部、特に visual mapping においては従来提案と比べて不十分な点もある。一方で統計データを扱う際に問題となるデータ要素の階層性と集約に関する扱いは従来の可視化ではまったくといっていいほど考慮されていなかった。また、visual mapping においても、ユーザの自然な要求とデータの特徴からどの程度までグラフ種が一意化できるのかという点について経験論的に取り組んでおり、この点で、そこで得られた結果は今後のより網羅的な可視化の第 1 歩となっている。最後に、ユーザから与えられる「何を描写したいのか」という要求から可視化の 3 つのプロセスすべてに関する指示を取り出し、それを自動化するという試みは見あたらず、このような可視化全体を対象とすることで表題に含まれる情報の位置付けを明らかにできたと考えている。

7. おわりに

本稿では、我々は表題に基づいて大量の数値データから自動的に適切なグラフを描画する可視化方式を提案した。これによって、ユーザはグラフ描画に必要なパラメータを意識することなく、効果的に要求を満たすグラフを得ることができる。

今後、提案手法の有効性を様々なドメインに適用して確認する。また既存のデータベースや WEB ページから必要な知識を自動獲得する方法についても検討する。

謝辞 提案手法の実装にご協力いただいた NTT アドバンステクノロジ株式会社の松島英之氏、松島英子氏、杉田理恵氏に感謝します。また、グラフ種判別知識の実装および評価にご協力いただいた大阪府立大学総合科学部の松本裕二氏（現在、富士通株式会社）に感謝します。最後に、貴重なコメントをくださった査読者の方々にも感謝します。

参 考 文 献

- 1) Ahlberg, C. and Shneiderman, B.: Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays, *Proc. Conference on Human Factors in Computing Systems*, pp.313–317 (1994).
- 2) Card, S.K., Mackinlay, J.D. and Shneiderman, B.(eds.): *Readings in Information Visualization — Using Vision To Think*, Morgan Kauf-

- mann Publishers (1999).
- 3) Derthick, M., Harrison, J., Moore, A. and Roth, S.F.: Efficient Multi-Object Dynamic Query Histograms, *Proc. IEEE Information Visualization Conference*, pp.84–91 (1999).
- 4) Euzenat, J.: An Algebraic Approach to Granularity in Qualitative Time and Space Representation, *Proc. 15th International Joint Conference on Artificial Intelligence*, pp.894–900 (1995).
- 5) Fasciano, M. and Lapalme, G.: Automatic generation of statistical graphics, *Proc. International Conference on Multimodal Communication*, pp.303–305 (1995).
- 6) Fasciano, M. and Lapalme, G.: PostGraphe: A System for the Generation of Statistical Graphics and Text, *Proc. 8th International Workshop on Natural Language Generation*, pp.51–60 (1996).
- 7) Green, N., Carenini, G., Kerpedjiev, S., Roth, S. and Moore, J.: A Media-Independent Content Language for Integrated Text and Graphics Generation, *Proc. Workshop on Content Visualization and Intermedia Representations*, pp.69–75 (1998).
- 8) Grosz, B.J., Appelt, D., Martin, P.A. and Pereira, F.C.N.: TEAM: An Experiment in the Design of Transportable Natural-Language Interfaces, *Artificial Intelligence*, Vol.32, pp.173–243 (1987).
- 9) Hendrix, G., Sacerdoti, E., Sagalowicz, D. and Slocum, J.: Developing a Natural Language Interface to Complex Data, *ACM Trans. Database*, Vol.3, No.2, pp.105–147 (1978).
- 10) Kerpedjiev, S. and Roth, S.F.: Mapping Communicative Goals into Conceptual Tasks to Generate Graphics in Discourse, *Proc. International Conference on Intelligent User Interfaces*, pp.60–67 (2000).
- 11) Mackinlay, J.: Automating the Design of Graphical Presentations of Relational Information, *ACM Trans. Graphics*, Vol.5, No.2, pp.110–141 (1986).
- 12) Matsushita, M., Yonezawa, H. and Kato, T.: A Frame Representation of User Requirements for Automated Data Visualization, *Proc. 14th European Conference on Artificial Intelligence*, pp.631–635 (2000).
- 13) Maybury, M.T. and Wahlster, W.(Eds.): *Readings in Intelligent User Interface*, Morgan Kaufmann Publishers (1998).
- 14) Mittal, V.O.: Visual Prompts and Graphical Design: A Framework for Exploring the Design Space of 2-D Charts and Graphs, *Proc. 14th*

National Conference on Artificial Intelligence, pp.57-63 (1997).

- 15) Quinlan, J.R.: Discovering Rules by Induction from Large Collections of Examples, *Expert Systems in the Micro Electronics Age*, Michie, D.(Ed.), Edinburgh University Press (1979).
- 16) Randall, D.J., Hamilton, H.J. and Hilderman, R.J.: Generalization for Calendar Attributes Using Domain Generalization Graphs, *Proc. 5th International Workshop on Temporal Representation and Reasoning*, pp.177-184 (1998).
- 17) Roth, S.F., Kolojejchick, J., Mattis, J. and Goldstein, J.: Interactive Graphics Design Using Automatic Presentation Knowledge, *Proc. Conference on Human Factors in Computing Systems*, pp.112-117 (1994).
- 18) Roth, S.F. and Mattis, J.: Data Characterization for Intelligent Graphic Presentation, *Proc. Conference on Human Factors in Computing Systems*, pp.193-200 (1990).
- 19) Stumme, G.: On-Line Analytical Processing with Conceptual Information Systems, *Proc. 5th International Conference on Foundations of Data Organization*, pp.117-126 (1998).
- 20) Woods, W.A.: Semantics and Quantification in Natural Language Question Answering, *Advances in Computers*, Yovits, M.(Ed.), Vol.17, Academic Press (1978).
- 21) Zelazny, G.: *Say It with Charts*, 3rd edition, McGraw-Hill (1996).
- 22) 角 康之: 情報可視化システムにおける適応的インタラクション, 人工知能学会誌, Vol.14, No.1, pp.33-40 (1999).
- 23) 米澤勇人, 飯田敏幸: 主張を反映した時系列データのグラフ表示, 第12回人工知能学会全国大会, pp.522-523 (1998).
- 24) 星野 隆, 綱川光明, 町原宏毅: DBSENA: マルチデータベース環境における情報資源管理と検索方式, 情報処理学会研究報告, Vol.98, No.2, pp.113-120 (1998).
- 25) 増永良文: リレーショナルデータベースの基礎

— データモデル編, オーム社 (1990).

(平成 13 年 3 月 19 日受付)

(平成 13 年 11 月 14 日採録)



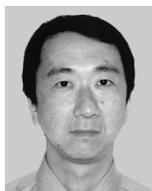
松下 光範 (正会員)

1969 年生まれ. 1993 年大阪大学工学部精密工学科卒業. 1995 年同大学大学院基礎工学研究科物理系専攻制御工学分野博士前期課程修了. 同年 4 月, 日本電信電話(株)入社, 現在に至る. 情報可視化, 自然言語理解に関する研究に従事. 1996 年度人工知能学会全国大会優秀論文賞受賞. 日本ファジィ学会, 人工知能学会各会員.



米澤 勇人

1971 年生まれ. 1997 年大阪府立大学大学院工学研究科電気・情報系専攻情報工学分野博士前期課程修了. 同年 4 月日本電信電話(株)入社. コミュニケーション科学研究所を経て 2000 年 4 月より西日本電信電話(株)に勤務. 現在に至る. 情報可視化に関する研究に従事.



加藤 恒昭 (正会員)

1959 年生まれ. 1981 年東京工業大学工学部電気電子工学科卒業. 1983 年同大学大学院総合理工学研究科電子システム専攻修士課程修了. 同年, 日本電信電話公社(現 NTT)に入社. 2000 年より, 東京大学大学院総合文化研究科言語情報科学専攻助教授, 現在に至る. 自然言語理解, 対話処理, マルチモーダルコミュニケーションに関する研究に従事. 工学博士. 電子情報通信学会, 人工知能学会, 言語処理学会, ACL 各会員.