ヒューマンインタフェース 65－10
音声言語情報処理 10－10
（1996. 3. 1）

情報可視化のためのデータビジュアル化モデル

黒川清、磯部成二、塩原寿子、鬼塚真
{kurokawa, isobe, shiohara, onizuka}@ciladb.dq.isl.ntt.jp

NTT情報通信研究所
〒238-03　横須賀市武1-2356

大量の文字数値情報の高度利用に向けて、情報の意味内容を直観的に理解できないといった課題を克服するために、多様かつアドホックな情報要求に迅速に応えられる情報視覚化手法が求められている。我々は、表現形式として単純で関係表現が容易なノードとラインという図形の組に着目し、これにより情報を表現するデータビジュアル化モデルを提案し、このモデルによる情報の意味内容理解の容易性について述べる。本研究のモデルはデータ視覚化のためのアプリケーション開発環境への応用が可能である。

A Study on Data Visualization Model for Information Representation

Kiyoshi KUROKAWA, Seiji ISOBE, Hisako SHIOHARA and Makoto ONIZUKA

NTT Information and Communication Systems Laboratories
1-2356, Take, Yokosuka-Shi, Kanagawa, 238-03 Japan

In many business areas, large amounts of character-based data is not so efficiently used. There are problems for realizing advanced applications. For example, the efficient graphical representation technique is for lack of character-based data representation capability. We propose the information visualization model including a graphical representation technique of character-based information for supporting powerful data analysis. This study can apply to an application development environment for data visualization.

1. はじめに

企業活動におけるサービスの計画、設計、運用、評価の各種業務は環境の変化に即応して柔軟に運営することが求められており、高度な情報分析に基づく経営の意思決定が重要な課題となっている。しかし、意思決定に必要な情報は様々な業務で発生し、大量の文字数値情報としてデータベースなどに蓄積されており、そのままの状態ではデータの持つ情報を直観的に把握することが困難で、迅速に意思決定することができないという問題がある。一方、コンピュータ技術の進展にともなって、大量情報を視覚化する技術が研究開発されている。しかし、サイエンティフィックビジュアライゼーションや地図情報システムなど数値、空間情報中心で、ビジネスの分野における多次元的な文字数値情報を一覧して分析することが困難であるという問題があった。[1]

我々はこれらの問題を解決するため、文字数値情報に含まれる分析対象のデータ集合を表示実体と呼び、個々の表示実体やその間の関係を表示実体の属性値を用いた図形表現によって視覚化するモデルを考案した。本モデルは、「ノードラインビューモデル」と呼び、表示実体をノード型かライン型に汎化されたオブジェクト情報に情報変換し結果を画面に表現するもので、表示実体の属性情報に基づいてオブジェクトの配置や形状を決定でき、ユーザの意図に合わせた図形表現が可能という特徴がある。

以下、2章で研究の目的、3章でノードラインビューモデルの定義と特徴、4章でモデルの評価実験と考察、5章でまとめを述べる。

2. 研究の目的

企業において分析対象となる情報は、文字数値情報としてデータベースに蓄積されていることが多く、データベース技術と可視化技術の融合が求められている。可視化のポイントは、対象情報を限定せず（実世界、概念など）、その情報特性（属性、値など）を利用して、情報変換手法を工夫することにある。[2]

本研究の目的は、意思決定支援のための視覚化手法の確立である。ビジュアル化への要求条件を整理すると、以下のようになる。

（1）データの汎用性

ビジネス分野の情報は、業務システム毎に蓄積されることが多く、データ構成や数値文字情報の組み合わせ方が業務に依存して多種多様である。どのような業務のデータも共通に扱えるようにするためには、業務システムのデータ構成とは独立性の高い汎用的データモデルを入力としたビジュアル化モデルが要求される。また、そのデータモデルは多種多様のデータ構成から変換可能な構成とする必要がある。

（2）表現の多次元性

ビジネス分野の業務帳票や設備管理表などのデータは多くの属性項目から構成されることが多い。従来、これらのデータは統計処理結果を複数の2／3次元チャートで表現するのが一般的で、インスタンス（属性項目値の組）との対応関係や多次元の属性項目を一覧することが困難であった。インスタンス個々の特徴と全体傾向を把握するためには、インスタンスレベルの表示が可能で、各インスタンスの多次元属性が図形表現として可視化できる必要がある。

（3）視点変更の多様性

ビジネス分野のアプリケーションは、あらかじめ想定したビジュアル化表現を実装している場合が多く、アドホックな要求に応えるための、迅速な表現パターンの変更が困難であった。業務変更やデータ分析の進展に応じたビジュアル化表現の変更要求に対して、迅速に対応するためには、多様な数値文字情報から図形情報への情報変換方法の提供によって、多くの表現パターンを選択できる必要がある。

次章では、これらの要求条件を満たすビジュアル化モデルについて述べる。

3. ノードラインビューモデル

3.1 モデル定義

ノードラインビューモデルは、図1に示すように、実世界や概念の実体を図形として表現する視覚化モデルであり、以下の3要素の組み合わせとして表現される。

Model = < Entity , View , Conversion >

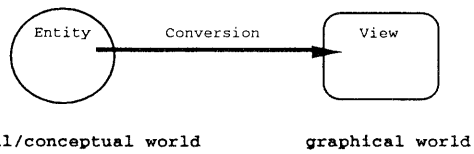

図1　ノードラインビューモデル

（1）Entity

Entityは実／概念データを実体として捉えたものでありその内包および外延を明示する。

$$Entity = \{ Class_i \mid i \geqq 1 \}$$

ただし、Classは1つ以上の属性attributeの順序づけられたn個組を持ち、

$Class_i = \{Attribute_i\}$

$$= \{ <attribute_{i1}, attribute_{i2}, \cdots, attribute_{in}> \mid n \geqq 1\}$$

である。このとき、この組はスキーマを表わす。スキーマが定義されるとき、その外延としての$Class_i$のインスタンス集合は、属性のドメインをdom(attribute)としたとき、

$Pow(instance_i) \subseteq dom(attribute_{i1}) \times dom(attribute_{i2}) \times \cdots \times dom(attribute_{in})$

である。また、各インスタンス（k番目）は属性値のn個組で表現することができ、

$instance_{ik} = \{< value_{ik1}, value_{ik2}, \cdots, value_{ikn}> \mid value_{ikj} \in dom(attribute_{ij})\}$

である。

（2）View

Viewはデータ視覚化の結果を表すものであり図形の集合として表現される。

$$View = \{ View_i \mid i \geqq 1 \}$$

ただし、Viewは以下のように表現され、

$$View_i = \{ Type_{ij} \mid j \geqq 1 \}$$

である。

(i)Type

Typeは表示オブジェクト型の集合である。ただし、

$Type_{ij} \in \{ Node\text{-}class_{ij}, Line\text{-}class_{ij} \}$

である。

このとき、それぞれのクラスは以下の2つ組で表現され、

$Node\text{-}class_{ij} = < Form_{ij}, Coordinate_{ij} >$

$Line\text{-}class_{ij} = < Form_{ij}, Coordinate_{ij} >$

である。

(ii)Form

FormはTypeの表示オブジェクト型属性を表わす集合である。

各$Form_{ij}$はラベル$label_{ij}$、カラー$color_{ij}$、サイズ$size_{ij}$、形$shape_{ij}$、のそれぞれの要素の組である。

$Form_{ij} = <label_{ij}, color_{ij}, size_{ij}, shape_{ij}>$

ただし、それぞれの値は画面表示されるオブジェクトの形状の値を表わし、

$label_{ij} \in$ Name (name of display object)

$color_{ij} \in$ Color (mixed RGB of display object)

$size_{ij} \in$ Size(scale of displayed object:big,small,middle,thick,thin)

$shape_{ij} \in$ Shape(shape of display object:square,circle,solid,break)

である。

(iii)Coordinate

Coordinateは表示オブジェクト配置座標の集合である。各$Coordinate_{ij}$は座標系の要素の組である。以下、表示空間が2次元の場合を示す。

$Type_{ij}$がNode-classの場合、座標はオブジェクトの中心座標を表し、

$Coordinate_{ij} = <x\text{-}coordinate_{ij}, y\text{-}coordinate_{ij}>$

$Type_{ij}$がLine-classの場合、座標はオブジェクトの始点／終点座標を表し、

$Coordinate_{ij} = < x\text{-}coordinate_{ijk}, y\text{-}coordinate_{ijk}, x\text{-}coordinate_{ijl}, y\text{-}coordinate_{ijl} >$

である。ただし、それぞれの値は表示空間内の値であり、

$x\text{-}coordinate_{ij}, x\text{-}coordinate_{ijk}, x\text{-}coordinate_{ijl} \in$ a vertical point of display space

$y\text{-}coordinate_{ij}, y\text{-}coordinate_{ijk}, y\text{-}coordinate_{ijl} \in$ a horizontal point of display space

である。

（3）Conversion

Conversionは情報変換（EntityからViewへの写像）を定義するものである。写像は、実体集合Entity、図形集合View、写像ルールの集合Ruleから構成され、以下のように表わすことができる。

$$Conversion : Entity \rightarrow View \equiv Rule \times Entity \rightarrow View$$

ただし、Ruleは$Rule_i$から構成され、$Rule_i$は属性マッピングのルール$Mapping_i$、属性値変換のルール $Method_i$の組として表現され、

$$Rule = \{ Rule_i \mid i \geqq 1\}$$

$$Rule_i = < Mapping_i, Method_i >$$

$Mapping_i \in \{$ View mapping $\}$

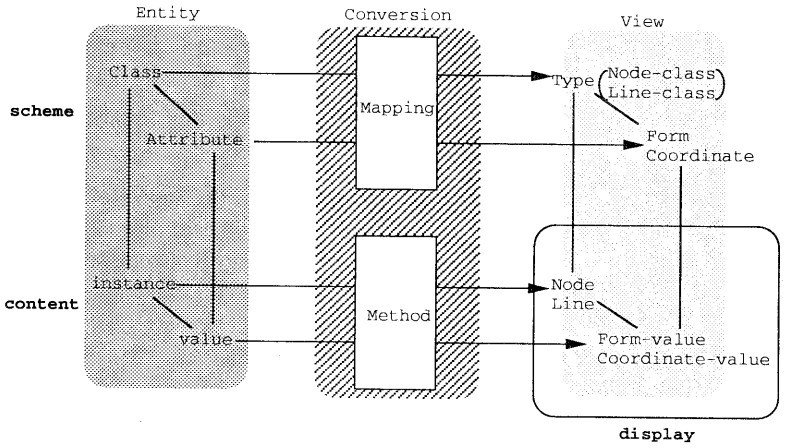

図2　ノードラインビュモデルの概念図

$Method_i \in \{$ translation method $\}$

である。このとき、次に示すEntityとViewの関係が成立する写像が定義される。

$$Entity = \{ Class_i \mid i \geqq 1 \}$$

$$View = \{ Type_k \mid k \geqq 1 \}$$

に対し、

$$Mapping_i(Class_i) = Type_k = \{Node\text{-}class_k, Line\text{-}class_k\}$$

$$Mapping_i(attribute_{ij}) = \{Form_k, Coordinate_k\}$$

$$Method_i(instance_i) = \{Node, Line\}$$

$$Method_i(value\ of\ attribute_{ij}) = \{Display\ value\ of\ (Form_k, Coordinate_k)\}$$

を満たす。

図2にノードラインビューモデルの概念図を示す。EntityのスキーマであるclassとAttributeを、Conversionの型、属性対応関係を表わす写像であるMappingにより、ViewのスキーマであるtypeとForm, Coordinateに変換される。Entityの内容（値）であるInstance, Valueは、値変換方法を表わす写像であるMethodにより、変換結果として表現される各値（Node, Line, Form-value, Coordinate-value）に変換される。

3.2 モデルの実装例

上記のモデルを実装したイメージを図3に示す。Conversionの手続きにより、Class1はNode-type、Class2はLine-typeにマッピングされ、また、それぞれの属性はForm, Coordinateにマッピングされている。Entityの値であるInstance, AttributeはMethodにより値変換されて、View上に表現される。

本モデルの表現能力はマッピング対象となる表示オブジェクトの属性数（ViewのForm,Coordinate属性数）と変換のメソッド（ConversionのMerhod）のバリエーションに左右される。

3.3 モデルによる表現の特徴

ノードラインビューモデルの表現法は、ノード型のみによる表現、ライン型のみによる表現、ノード型とライン型を組み合わせた表現の3種類に分類できる（イメージを図4に示す）。

（1）ノードのみによる表現

ノードのみの表現によるビジュアル化は、Entityの持つ複数の属性項目間の関係が把握し易い表現となる。一般の散布図との違いは、座標軸だけでなく、ラベル、サイズ、カラーなどのViewのForm属性を利用することにより、多次元の属性の一覧が可能となる点である。また、ノードの重なりを利用することにより、インスタンス間の依存／包含関係を表現することができる。

（2）ラインのみによる表現

ラインのみの表現によるビジュアル化は、値の推移や差分などの傾向が捉え易く、かつ、ViewのForm属性を利用することにより、Entityの持つ複数の属性項目間の関係を把握

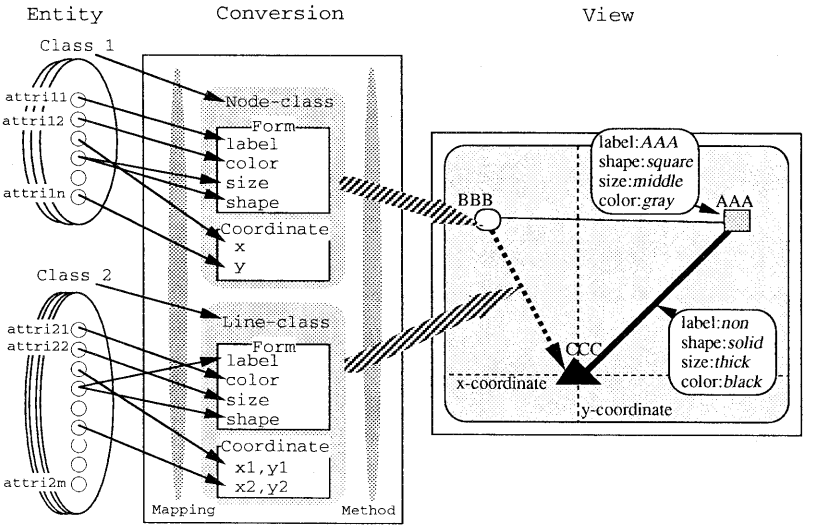

図3　ノードラインビューモデルの実装例

し易い表現となる。

（3）ノードとラインによる表現

ノードとラインの組み合わせ表現によるビジュアル化は、複数表示実体間のトポロジーが把握しやすい表現となる。一般のE-R図との違いは、EntityとViewとの対応関係がConversionで自由に変更でき、表現がクラスに限定されないことである。また、ノードとラインの重なりを利用することにより、個々の依存関係とノードライン間の依存関係を表現することができる。

ノードラインビューモデルにおける表現の最大の特徴は、ノード間、ライン間、ノードライン間の表現によって、インスタンス間の依存関係、相互関係の把握が容易にできることにある。

4. 評価

4.1 評価実験

ノードラインビューモデルの表現能力の有効性を評価するため実験を行った。

実験は、被験者に対して、実験データを表で表現したものとノードラインビュモデルで表現したものを順に提示し、その反応を記録した。この実験の観点は、ノードラインビューモデルを用いた表現により、予備知識を持たないユーザによる事実の抽出を支援することができるかということである。よって、事実の抽出件数と抽出時間を評価の指標として用いた。

データは2種類準備し、また、予備知識を持たない被験者を男女7名ランダムに選抜し実験を行った。実験結果を表に示す。

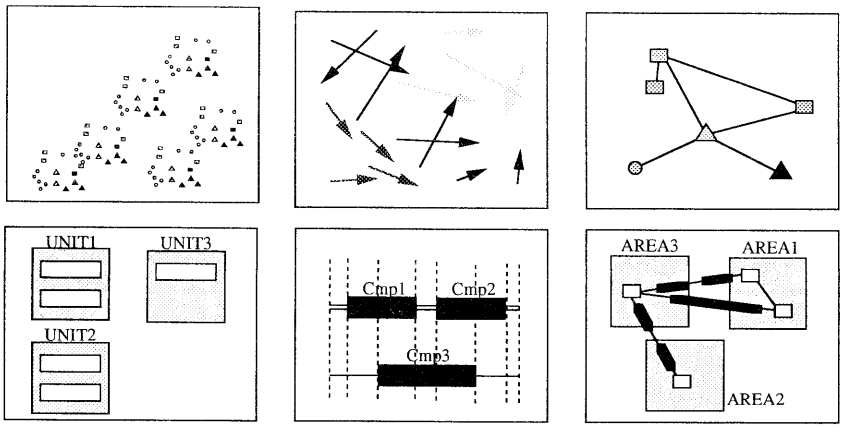

図4　表現パターン例

表　実験結果

	ﾉｰﾄﾞﾗｲﾝﾋﾞｭｰﾓﾃﾞﾙによる表現	表による表現
実験1	17件 111[s/件]	14件 132[s/件]
実験2	20件 98[s/件]	6件 110[s/件]

上段：事実抽出件数
下段：平均抽出時間

結果からわかるように、ノードラインビューモデルによる表現は、表による表現と比較して、事実の抽出、抽出時間とも優位な結果となった。

4.2 考察

実験結果より、事実抽出を支援するという観点からの、ノードラインビューモデルの有効性を確認できた。結果を詳細に見ると、全体の傾向を掴むにはノードラインビューモデルによる表現が良く、個々の具体的な情報を得るには表のほうが良いことがわかった。

また、実験後のアンケートにより、「ノードラインビューモデルは、表現したいと思う意図を直接反映することができる」との意見があった。これは、ノード型とライン型のオブジェクトが持つ属性を工夫することにより、様々な表現パターンが得られることを示唆する。

5. おわりに

本稿では、情報視覚化のためのビジュアル化モデルとしてノードラインビューモデルを提案し、その有効性について評価した。ノードラインビューモデルは、従来の表現法と比較して、多次元の属性を一覧できるといった特徴があり、大量データからの情報抽出に有効な表現方法であると言える。また、このモデルを情報ビジュアル化システムに実装することにより、エンドユーザによる簡易なアプリケーション開発とアドホックな情報要求へのタイムリーな応答を実現することが期待される。[3]

本研究の情報可視化は、日常業務で生成されるが見逃されがちな数値文字情報の価値を一層高めることができると考えられる。

参考文献

[1]石垣,他:"データベースのビジュアル化ツール",経営システムVol.5 No.3･4, 1995 .

[2]磯部,他:"ノードラインビューモデルに基づく数値文字情報のビジュアル化環境", 信学技報DE95-62, 1995.

[3]K.Kurokawa, et al.:"Information Visualization Environment for Character-based Datebase System", First International Conference on Visual Information Systems, 1996.