視覚化による多次元データ分析システム：INFOVISER

磯部成二[†] 黒川清[†]
塩原寿子[††] 飯塚哲也[††]

市場競争の激化にともなって，データに基づく戦略的意思決定が企業の重要課題となっている．これに対して，統計解析，多次元データ分析，データ可視化等の既存のデータ分析手法は，数値・文字が混在する複雑なデータを対象とする非定型な分析要求に柔軟に対応できない，データ準備からアクションまでのデータ分析プロセスの試行錯誤と業務背景知識の活用が難しい等の問題があった．筆者らは，複雑なデータを対象とするアドホックな分析要求に応えられるデータ分析システムの確立を目的に，情報視覚化モデルとデータ変換メカニズムを提案する．情報視覚化モデルでは，ユーザ定義に基づいて実体データをノード型かライン型の図形の形状や配置として多次元表現するという考え方を導入し，ユーザ定義可能なデータ変換メカニズムへの要求機能を明確化した．データ変換メカニズムは，多様な情報源から任意の実体データを抽出可能なソフトウェア部品と，実体データを多様な図形データに変換可能なソフトウェア部品から構成し，ユーザによる部品選択定義により豊富な表現が提供可能である．本モデルとメカニズムを適用した視覚的多次元データ分析システム（INFOVISER）を各種データ分析に適用し，本システムの有効性，情報源への適用性，表現の多様性を確認し，ビジネス分野のデータ分析への適用性を確認した．

A Visual and Multi-dimensional Data Analysis System: INFOVISER

Seiji Isobe,[†] Kiyoshi Kurokawa,[†] Hisako Shiohara[††]
and Tetsuya Iizuka[††]

Under the keen market competition, strategic decision making based on data analysis becomes the most important subject for enterprises. However, ad hoc analysis of complex data including numerical and string value, trial and error analysis through human interaction based on his back ground knowledge, were not entirely supported by existing data analysis methods such as statistics, multi-dimensional data analysis and visualization. Aiming to create a data analysis system that would support business data analysis as described above, we propose an information visualization model and its corresponding data conversion mechanism. In the model, a new concept which represent data as a node-type or line-type figure and its shape and layout attributes through user definition is proposed, and a precise definition of a user controllable data conversion function is given. The data conversion function has a powerful representation ability with two types of flexible software modules which enables to convert from source data to entity data, and from entity data to figure data. An experimental visual and multi-dimensional data analysis system named INFOVISER utilizing above model and mechanism was applied to actual data analysis and successful examples for wide business areas and variety of practical representation patterns were confirmed.

1. はじめに

企業の戦略的意思決定を支援する技術として，データウェアハウス，多次元データ分析（OLAP: On-line analytical processing），データマイニング等の研究が行われている．近年，企業ではデータウェアハウスの構築が進み，OLAP やデータマイニング等のデータ活用を支援する技術が注目されている．

厳しい環境下にあるビジネス分野のデータ活用を支援するためには，日々変動する外部環境に即応できる分析の柔軟性，分析者の意図によって試行錯誤できる人間介在インタフェース，多数の分析対象が複雑に関係し，数値・文字が混在するデータの分析機能を有するデータ活用支援技術が求められている．

従来，ビジネス分野では，統計解析や機械学習等の数値解析的手法，次元切替え等の人間操作による OLAP

† NTT ソフトウェア株式会社
NTT Software Corporation
†† NTT サイバーソリューション研究所
NTT Cyber Solution Laboratories

手法，大量の数値データの特徴を可視化して把握する視覚化手法等のデータ分析手法が用いられていた．しかし，これら手法のビジネス分野への適用には以下の問題があった．数値解析的手法は，対象が数値データに限られ，分析過程がブラックボックスのため分析者による理解や介在が難しく，試行錯誤的なデータ分析が進めにくい．OLAP手法は，次元切替えの操作性には優れているが，多次元キューブの限られたビューからデータを見るため全体の傾向把握は難しい．視覚化手法は，大量の連続値データの色表現による全体把握には優れているが，数値・文字混在の複雑なデータの関係は把握しにくい．また，3つの手法はともに，簡易にデータウェアハウス等の情報源から対象データを抽出し，編集する機能が不足しているため，外部ツールや手操作でデータを準備しなければならないという問題があった．以上の問題点は，多次元データの全体傾向把握と人間介在による試行錯誤的データ分析の両方を満足する手法がないことと，データ準備を支援する機能が不足していることに要約できる．

本論文では，多次元データを図形表現で一覧表示することにより，人間によるデータ分析の試行錯誤を支援するデータ分析方法を提案する．この視覚的データ分析手法は，多次元データを効果的に視覚化可能なデータ視覚化モデルと，簡易なGraphical User Interface（GUI）定義操作のみで視覚化可能な実体抽出と図形変換のメカニズムによって構成することを述べる．さらに，このモデルとメカニズムを適用した視覚的多次元データ分析システム（INFOVISER:Information Visualization Environment）[1),2)]の構成方法と，視覚化表現の有効性に関する評価結果について述べる．

以下，本論文では，2章で関連研究の動向，および問題点と解決策について述べる．3章で，解決策としてのデータ視覚化モデル，実体抽出と図形変換メカニズムについて述べる．さらに，4章で，これらを適用した視覚的多次元データ分析システム（INFOVISER）の構成法について述べる．最後に，5章で本システムの適用性評価の結果，および結論を述べる．

2. 関連研究

本章では，多次元データ分析と人間介在型データ分析の観点から，OLAPとデータ視覚化に関する研究動向を考察し，問題点と解決のアプローチについて述べる．

2.1 OLAP関連の研究動向

OLAPは，Coddによって提唱されたOLTP（基幹系処理）に対向する情報系処理のアーキテクチャで，データ前処理，データモデル，C/S環境，次元操作等の機能から定義される[3),4)]．一般に，簡易な次元切替えインタフェースを有する多次元データ分析ツールがOLAPツールと呼ばれている．

これまで，OLAPは，多次元データの効率的管理法や，多様なビューからの高速データ検索法が研究されてきた．最近は，分析者がテーブルの物理構造の意識やプログラミングをすることなく，簡易な次元切替え操作だけで多次元データ分析できるOLAP製品が市販されている．これらは，エンジンに多次元データベースかリレーショナルデータベースを，次元切替え・結果提示のユーザインタフェースにスプレッドシートを利用しているものが大半である．

このようなOLAPツールは，組織別・時期別・品目別に売上をチェックするといった，ビジネス分野では定型的な仮説検証には有効であるが，非定型の戦略的分析へ適用するには，(1) 変数が数値データしか扱えない，(2) 変数と次元の関係は事前登録を要し変更が難しい，(3) 事前登録した関係以外のビューから変数を見ることはできないといった問題がある．

2.2 データビジュアライゼーション関連の研究動向

データ分析に着目した視覚化手法を網羅的に調査したKeim[5)]の論文がある．この論文では，視覚化表現法には，幾何学的手法，アイコン手法，ピクセル手法，階層手法，グラフ手法があり，幾何学的手法はクラスタリングに，アイコン手法やピクセル手法は多次元の相関分析に，階層手法はカテゴリ化に適しているという主観的評価から，視覚化は発見的分析（Explorative Analysis），仮説検証的分析（Confirmative Analysis），事実提示（Presentation）の広範囲のデータ分析に有効な手法であることが報告されている．また，次元削減，データ集合のサブセット化，セグメンテーション，統計演算等のデータ前処理技術や表示イメージの変形や拡大・縮小，詳細表示等の動的な表示画面操作の技術も重要であることが述べられている．

特に，多次元データの一覧性を追求した視覚化技術としては，ピクセルマッピング法[6)~9)]，スティックフィギュア法[10)]，パラレルコーディネート法[11),12)]がある．ピクセルマッピング法は，データの値をピクセル（画素）の色に対応させ，比較対象となる次元間の関係の強さの順にデータをピクセル上に配置する方法で，百数十万のデータ値を一度に表示でき，多次元の相関を比較しやすいという特徴がある．スティックフィギュア法は，データの値によって木の枝の長さや角度を変更する方法で，レコード対応の木の領域的な特徴からデータの全体傾向がつかみやすいという特徴がある．

パラレルコーディネート法は，データ項目ごとに平行する座標軸を用意し，データ値を座標軸上に配置し，同一レコードの値間をラインで表示する方法で，データ項目ごとやレコードごとの値の分布傾向を容易に把握できるという特徴がある．しかし，ピクセルマッピング法やスティックフィギュア法では，ピクセル配置や枝形状変換の方法をユーザが簡単に変更することは難しく，視点を変えた試行錯誤的なデータ分析ができないという問題がある．一方，パラレルコーディネート法は，座標軸の入替え，表示対象レコードの選択等を対話的に行える技術も開発されており，視点変更の柔軟性は実現されているが，この手法だけでは，分類や階層等の分析が難しいという問題がある．

上記以外に，多次元散布図手法，階層配置手法，グラフ手法等の手法[13),14)]があるが，対象データや分析の視点が頻繁に変わるようなビジネス分野のデータ分析に，汎用的に適用可能な視覚化手法は見られない状況にある．

2.3 課題と解決のアプローチ

ビジネス分野の戦略的データ分析の特徴は，実体が多く関係が多岐というデータの複雑性と，環境条件の変化にともなって対象データや分析の視点が変わるといった非定型性と，業務経験や環境条件に基づく人間介在型の意思決定を必要とする試行錯誤性にある．本論文では，実体は顧客，商品，社員等の分析対象となるデータのまとまりを指し，データベースの概念設計で用いられる業務上の管理や処理の対象としての実体と同じ意味で使用している．

関連研究で述べたように，OLAP 手法はあらかじめ登録された変数と次元の関係の範囲内で視点を変更し，複数の次元が交差する範囲の変数値を把握するには有効であるが，登録以外の視点からのアドホックな分析や複数次元の全体の関係把握には適していない．一方，視覚化手法は，個々の表現手法ごとに適する分析内容は異なるが，人間のパターン認識能力や業務背景知識を活かした全体傾向の把握や分析過程への人間介在ができるという利点があり，複数表現手法の統合化ができれば有効な分析手法になる可能性を有している．

我々は，データの一覧性や人間介在に優れた視覚化手法に，OLAP の操作の容易性を取り入れた，視覚的多次元データ分析手法を提案する．以下，データの複雑性，分析の非定型性や試行錯誤性に対応するうえでの検討課題と解決のアプローチを示す．

(1) データの複雑性

ビジネス分野のデータには，顧客，商品，販売，物流等の多種多様な分析対象データが，業務処理に適する帳票，月次処理に適する一覧表，分析向きに再構築された生データ集約等の多様な形式で，相互に複雑な関係を持って蓄積されている．したがって，多様な構成の情報源から分析目的に応じて任意のデータ項目を含む実体を簡易に抽出できる環境の提供が課題となる．この課題を解決するために，ユーザ定義に基づいて，情報源から実体データへの変換を即時実行できる機能を提供することとした．

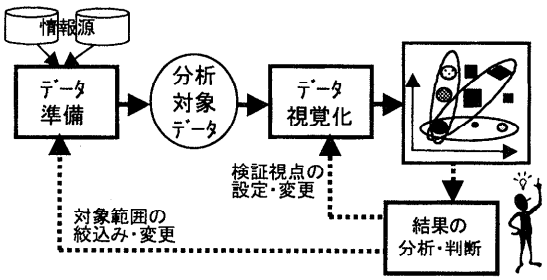


図 1 視覚化によるデータ分析プロセス
Fig. 1 Data analysis process by visualization.

(2) 分析の非定型性

ビジネス分野の戦略的なデータ分析では，環境条件の変化にともなって対象データや分析の視点が変わる．このような非定型の分析に対応するためには，データの意味・内容に依存しない視覚化機能，複雑な多次元データの全体傾向の一覧が容易な視覚化機能，分析目的や分析過程に応じた視覚化表現の変更機能の提供が課題となる．この課題を解決するためには，データの意味・内容に依存しないで複雑な多次元データの一覧ができ，簡易な指定で図形表現を自由に変更できる情報視覚化モデルを提供することとした．

(3) 分析の試行錯誤性

視覚化によるデータ分析手法のプロセスは，図 1 に示すように，分析対象のデータ準備，データ視覚化，視覚化結果の分析・判断の順に進められ，判断結果によって条件を変えて前のステップを再実行するというように，ビジネスに有益な知識を導出する試行錯誤の過程である．したがって，簡易に再実行の条件設定が行える試行錯誤インタフェースの実現が課題となる．この課題を解決するために，各プロセスの定義変更や保存が簡易に行える定義画面や結果画面のインタフェースを提供することとした．

3. 視覚的多次元データ分析手法と基本技術

本章では，視覚的多次元データ分析手法とその基本技術である情報視覚化モデルと視覚化のメカニズムについて述べる．

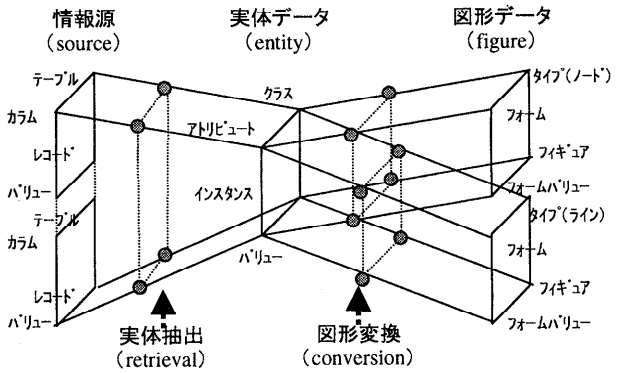


図 2 ノード・ラインビューモデル
Fig. 2 Node-Line View Model.

3.1 視覚的多次元データ分析手法

本論文の視覚的多次元データ分析手法は，簡易な定義で分析の視点に適した多様な図形表現が得られる情報視覚化モデルに基づき，ユーザ定義により選択的に分析対象データを抽出できる機能（以下，実体抽出メカニズムと呼ぶ）と，ユーザ定義により多次元データを任意の図形属性データに変換できる機能（以下，図形変換メカニズムと呼ぶ）を提供することにより，図1のフローに基づく，データ準備からデータ視覚化のプロセスと，視覚化結果の分析・判断に基づく人間介在型の試行錯誤的データ分析プロセスを支援する手法である．

3.2 情報視覚化モデル

既存の2次元や3次元のグラフ表現は，数値データやその集計結果を視覚化する機能が中心であり，文字データに対応できない，個々のデータを詳細に把握できないという問題があった．また，関連研究で述べた各種視覚化手法は，特定の分析目的には有効な表現であるが，入力データ形式，視覚化対象（ピクセル，図形，イメージ等），表現（配置，色等）は統一されていないため，ユーザが1つの環境から選択的に利用することは困難という問題があった．

本論文で提案する情報視覚化モデルは，上記問題の解決と，ユーザ定義による多様な視覚化手法を選択できることを目標に下記の方針で設計した．

(1) 情報源を画面上の図形として表現する過程に，実体データと図形データを設け，複数情報源からの1実体データの作成と，その実体データからの複数画面表現（以下，シーンと呼ぶ）の生成を可能とすることにより，各種情報源や各種 GUI ツールへの適用性のある，マルチユーザに共用可能なデータ分析支援環境を実現する．

(2) 実体データと図形データをオブジェクトモデルで統一的に管理し，実体データの1レコードを画面上の1図形オブジェクトに対応させ，ノード型（円形，矩形，三角形等）とライン型（実線，点線等）の図形種別の選択，複数の実体データの同一画面上への重畳表示，複数データ項目の図形属性による表示を可能とすることにより，豊富な表現パターンが選択可能で，複数データ項目間の関係と個々のデータ詳細把握が容易な視覚化機能を有するデータ分析支援環境を実現する．

(3) データ型対応のデータ変換処理と，データ項目名や文字・数値を用いた実体抽出定義と図形変換定義により，情報源の意味・内容には依存しない視覚化機能を有するデータ分析支援環境を実現する．

以下，この情報視覚化モデルは，ノード型とライン型の図形組合せ表現を基本とすることから，ノードラインビューモデルと呼ぶ．ノードラインビューモデルは，**図 2** に示すように情報源（Source），実体データ（Entity），図形データ（Figure）の3種類のデータと，情報源から実体データへの変換を行う実体抽出（Retrieval），実体から図形データへの変換を行う図形変換（Conversion）の2種類の変換処理から構成した．情報源はテーブル，レコード，カラム，バリューの関係型データ，実体データは，クラス，インスタンス，アトリビュート，バリューのオブジェクト型データ，図形データは，タイプ，フォーム，フィギュア，フォームバリューのオブジェクト型データから構成した．ここで，実体抽出は複数の関係型データからオブジェクト型データへのデータ変換処理を意味し，図形変換は実体の意味を表すオブジェクト型データから図形の表現を表すオブジェクト型データへのデータ変換処理を意味する．なお，図形データのアトリビュートは，情報の図形表現の設計に関する研究成果[15]を参考に，ラベル，サイズ，カラー，形等の形状と X 座標，Y 座標

等の配置のアトリビュートをノード型とライン型の2種類に分類して構成した．以下，インスタンス（フィギュア）とアトリビュート（フォーム）は，それぞれレコードとデータ項目という呼び方に統一する．

以上のように，本モデルは，オブジェクト形式による統一的データ管理と，テーブルから値までの各レベルごとのデータ変換処理により，情報源のデータ構成・データ内容およびGUIツールからの独立性が高く，簡易なユーザ定義に基づく図形型（ノード型とライン型）の組合せと豊富な図形の配置・形状表現により，多次元データ間の一覧比較が容易な視覚化表現を得られるという特徴がある．

3.3 実体抽出メカニズム

実体抽出には，複数の関係型データからなる情報源から任意構成の実体データを抽出し，オブジェクト型データとして登録できる機能が要求される．このため，多様な情報源からデータ抽出できるデータ検索部品，テーブルから値までの各レベルでデータ変換可能なデータ加工部品，オブジェクト形式で実体登録できるデータ登録部品の3種類の部品群から構成されるデータ操作部の部品実行方法および実行順序を自由に定義できる，簡易なユーザインタフェースを提供する方式とした．

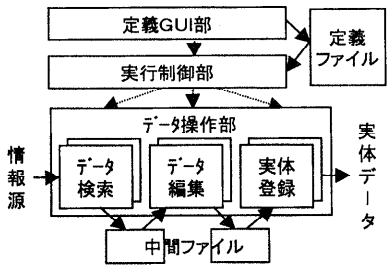


図3 実体抽出定義メカニズム
Fig. 3 Entity retrieval mechanism.

実体抽出メカニズムの構成を図3に示す．本メカニズムでは，定義GUI部から入力された抽出定義に基づいて，実行制御部がデータ操作部の各部品を順次起動することによりユーザ独自のデータ変換処理が実現される．本メカニズムの実現により，データ分析プロセスの試行錯誤過程において，従来はエディタやスプレッドシート等の外部ツールで行っていたデータ加工を，1つのデータ分析システム上の定義操作のみで対応可能とした．

3.4 図形変換メカニズム

データ視覚化には，複数実体間の関係と各実体内のデータ項目の関係を一覧したいという要求がある．図形変換部は，ユーザ設定可能なマッピング定義と図形変換定義に基づいて独立に動作するマッピング機能と値変換機能から構成し，マッピング定義と図形変換定義は，実体毎に独立した変換メソッド定義と複数実体を組み合わせた変換メソッド定義からの選択を可能とすることにより，ユーザの多様なデータ視覚化要求に対応できる方式とした．

図形変換メカニズムを図4に示す．本メカニズムでは，ユーザの定義に基づいて，値変換機能の各変換メソッドがマッピング機能によって引き渡されるデータ項目の値を図形属性の値に変換することにより，ユーザ独自の図形変換処理が実現される．また，ユーザの新しいシーン定義に応じて，データ管理機能が実体データと対応する新しい図形データを生成することに

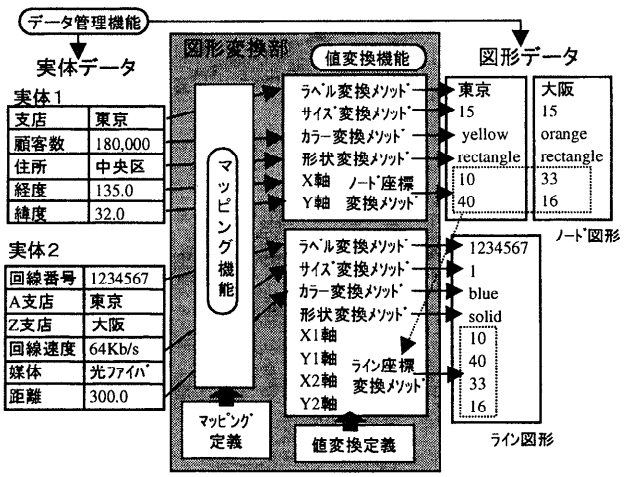


図4 図形変換メカニズム
Fig. 4 Figure conversion mechanism.

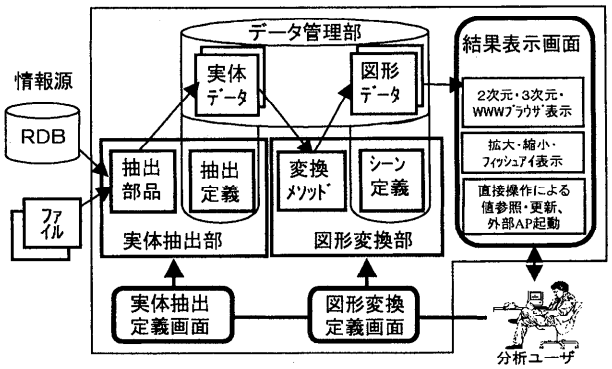


図 5 視覚的多次元データ分析システム
Fig. 5 Visual and Multi-dimensional Data Analysis System.

より，1 実体データから独立した定義を持つ N 個の図形データを生成できる情報共有環境が実現される．さらに，データ管理機能が実体間の関係をレコード間のリンクとして管理することにより，複数実体に関係する値変換処理が実現される．本メカニズムの実現により，ユーザ定義のみで多次元の図形表現が得られ，登録済の実体からは複数の図形変換を実行でき，1 画面に複数実体を関係付けて重畳表示できるため，分析要求に応じて表示対象の実体や図形表現パターンを変更するような試行錯誤のデータ分析を支援できる．

4. 視覚的多次元データ分析システムの構成

本章では，前章で述べたノードラインビューモデルと基本メカニズムを適用した視覚的多次元データ分析システムの実現法について述べる．本論文では，このシステムを INFOVISER（Information Visualization Environment）と呼ぶ．INFOVISER は，**図 5** に示すように，実体抽出部，実体抽出定義画面，図形変換部，図形変換定義画面，データ管理部，結果表示画面から構成した．

実体抽出部は，**表 1** に示す 22 種の実体抽出部品と，実体抽出定義画面を介してユーザにより設定された実体抽出定義から構成した．実体抽出定義画面は，**図 6** に示すように，データの検索・編集・登録の各種部品を作業領域に呼び出し，部品間を矢印付ラインで接続することにより部品の処理順序を指定できる構成とした．本機能の実装により，繰返しを含まないデータの検索・編集処理手続きについては，情報源のデータを変更することなく，実体を抽出することが実現できた．

図形変換部は，**表 2** および **表 3** に示す形状や配置に関する図形属性データへの変換メソッド（形状 20 種と配置 47 種）と，図形変換定義画面を介してユーザにより設定された図形変換定義から構成した．図形変換定義画面は，**図 7** に示すように，ノード型かライン型の選択を行う図形型選択，データ項目と図形属性の対応付けを行う属性マッピング，変換メソッドの選択とパラメータ設定を行う変換方法設定を，それぞれ宣言的に定義可能とした．また，検索条件設定による表示対象レコードの条件検索も可能とした．なお，ユーザによる実体抽出と図形変換の定義画面操作は，データ項目名と実体データの値参照による選択操作を基本とし，ユーザに親しみやすいインタフェースが実現で

表 1 データ検索部品一覧
Table 1 Data retrieval parts list.

部品分類	部品数	部品機能
データ検索部品	7種	RDB検索・RDB結合、ファイル入力、属性検索等
データ編集部品	13種	カラム選択、カラム結合、縦連結、コピー、重複排除、インデックス、数値演算、文字演算（結合・置換・分解）等
実体登録部品	2種	オブジェクトクラス登録、リンククラス登録

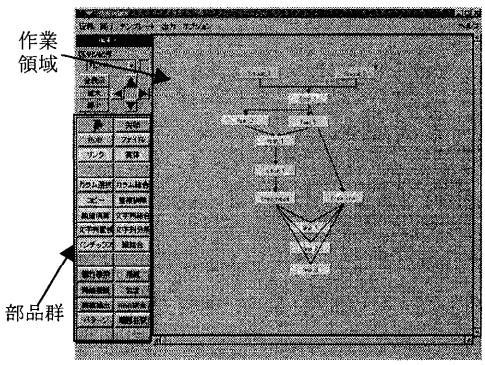


図 6 実体抽出定義画面
Fig. 6 GUI for entity retrieval definition.

表 2 形状変換メソッド一覧
Table 2 Shape conversion methods list.

種類	変換処理内容	図形属性	ノード図形	ライン図形
等分割	分割数、最大値、最小値と対応の図形サイズ、色、形を指定すると入力値を変換する。	カラー	○塗り潰し	○塗り潰し
		サイズ	○縦横	○線幅
		形状	△枠線幅	×
閾値	閾値と分割対応の図形サイズ、色、形を指定すると入力値を変換する。	カラー	○塗り潰し	○塗り潰し
		サイズ	○縦横	○線幅
		形状	○図形種	○線種
文字列	文字列対応の図形サイズ、色、形を指定すると入力値を変換する。	カラー	○塗り潰し	○塗り潰し
		サイズ	○縦横	○線幅
		形状	○図形種	○線種
自動	数値・文字データの値対応の色を、自動的に決められた規則で変換する。	カラー	○塗り潰し	○塗り潰し
		サイズ	×	×
		形状	○イメージ	×

○：実装、△：検討、×：未検討

表 3 配置変換メソッド一覧
Table 3 Layout conversion methods list.

ﾒｿｯﾄﾞ分類		ﾒｿｯﾄﾞ数	変換処理内容
基本	ノード	8	入力値を相対、絶対、順序的に散布配置
	ライン	8	入力値を相対、絶対、順序的に散布配置
ノード組合	格子図	2	指定された属性の値順で格子状に配置。
	包含図	7	依存関係のある図形を重ねあわせて配置。親図形と子図形の配置方法(相対、地理、自動)を独立に指定可能。
ノードライン組合	網状図	11	関係データをノードとノード間ラインで配置。ノード重複排除、複数ラインのオプション有。
	円状図	2	ノードを接続ライン数や指定の属性値に依り、同心円状に配置。
	樹系図	5	接続関係を辿ってツリー状に配置。
	接続図	4	ｴﾝﾄﾞｴﾝﾄﾞの構成区間、始点から順序配置。

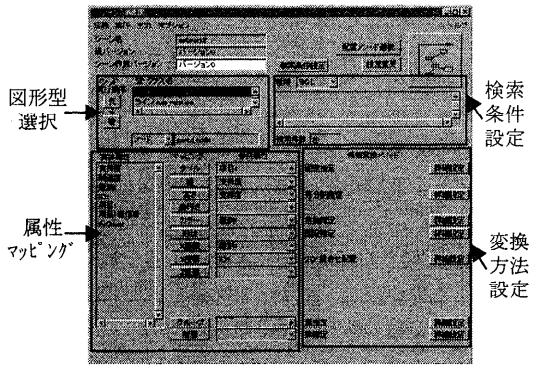


図 7 図形変換定義画面
Fig. 7 GUI for figure conversion definition.

きた．

データ管理部は，図 8 に示すように，図形変換部の処理フローから 11 種類のオブジェクトとその関連を管理する 3 つのブロックから構成した．実体定義部では実体抽出にかかわる部品と実行順序の関係を，実体/図形データ部では実体と図形間の関係を，シーン定義部ではシーン共通定義とシーン個別定義とシナリオ定義の関係を管理し，各ブロック間はシーン個別定義と図形，実体定義と実体の各オブジェクト間を関係付けることで，全体を統合的に管理した．なお，オブジェクトの参照は，参照・更新が必要な関係を OID による参照（図中の太線），その他を名前等による参照（図中の細線）で実現した．データ管理部の実装は，ユーザの要求に応じて動的オブジェクト生成要求が発生するため，オブジェクト・リレーショナルデータベース（ORDB）を適用した．

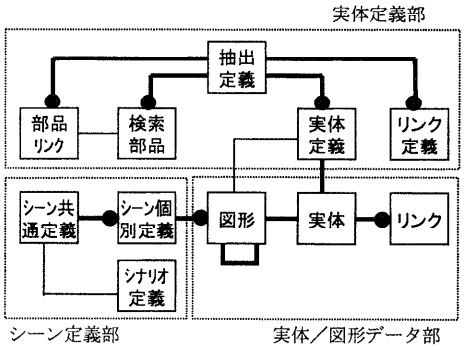


図 8 図形データのオブジェクトモデル
Fig. 8 Object model for figure information database.

結果表示画面は，2 次元表示用に C++，3 次元表示用に VRML，WWW ブラウザ表示用に Java の各形式へのデータ変換機能を持ち，2 次元は専用画面，3 次元は汎用 VRML ブラウザ，WWW は汎用 Web ブラウザへの出力を可能とした．また，表示領域の拡大・縮小，フィッシュアイ表示による座標変形，図形直接操作による実体データの値参照・更新や，外部アプリケーション起動等の付加的機能も実現した．

本システムの最も大きな特徴である表現パターン

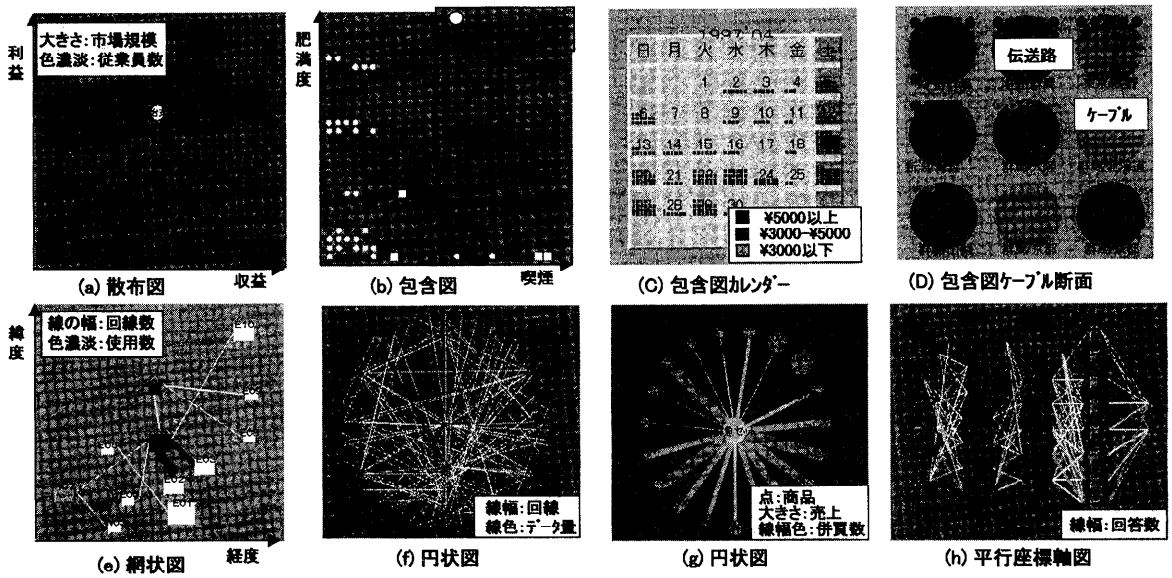


図 9 視覚化表現の適用性
Fig. 9 Adaptability of representation patterns.

の豊富さは，下式で表される．しかし，組合せ表現パターンの有効性は，データの特徴，分析目的に依存するため，平均的に有効なパターン数は多くの適用結果を待つ必要がある．

表現パターン数
$= \{$(形状表現数) + (配置表現数)$\}$
$= \{$(ノード実体数 × ノード形状表現数
+ ライン実体数 × ライン形状表現数)
+ (ノード実体数 × ノード配置表現数
+ ライン実体数 × ライン配置表現数)$\}$

ここで，形状表現数は表 2 に示す図形属性〔ラベル，サイズ（高さ・幅・奥行き），色，形〕対応の形状変換メソッド数，配置表現数は表 3 の配置変換メソッド数で決まる．

INFOVISER は，クライアント・サーバ構成で動作し，クライアントは Windows-95/NT，サーバは Windows-NT をプラットフォームとしている．情報源は，RDB（ORACLE, INFORMIX），ODBC，CSV 形式ファイルに対応している．

5. 適用性評価

本章では，前章で述べた視覚的多次元データ分析システム（INFOVISER）について，(1) 視覚化表現の有効性，(2) 情報源データ構成への適用性，(3) 視覚化表現の多様性，の 3 つの観点から適用性を評価する．

5.1 視覚化表現の有効性

ノード・ラインビューモデルの有効性は，実業務のデータ分析やパターン発見に効果のあった事例を主観的に考察して評価することとした．対象データは，経理，健康診断，通信網，商品購買，アンケート調査の 5 種類のデータとし，散布図，包含図，網状図，円状図，平行座標軸図の 5 種類の代表的配置方法に，図形形状の変化を重畳させた多次元表現の有効性について評価した．

(1) 散布図

本モデルの散布図表現（2 次元の場合）は，1 つの実体に対して最大 7 つのデータ項目を図形の属性〔X 座標，Y 座標，ラベル，サイズ（高・幅），色，形〕として表現し比較できるため，多次元データの相関関係分析に有効である．

たとえば，経理データには収益，費用，利益に関する詳細な内訳のデータ項目があり，支店の経営状態を多次元に比較したいという要求がある．図 9 (a) に，支店のある月の収益（X 軸），利益（Y 軸），市場規模（大きさ），従業員数（色）の関係を比較した例を示す．本例では，支店（円図形）の位置が 45 度右上がり線より上か下かで，収益に対する利益の割合（利益率）の良い支店と悪い支店にグループ化できる．さらに，悪い支店グループに着目すると，ほかに比較して市場規模が大きく（図形が大きく），従業員の多い（色の濃い）ことが分かり，従業員の稼動と人件費の詳細な要因分析の必要性が判断できる．本表現は，経理や健康診断データのように，多くのデータ項目からなる多次元データの分析に有効である．

(2) 包含図

包含図表現は，親子のように依存関係にある2つ以上の実体データを図形の重なりとして表現でき，それぞれの図形の形状や配置属性を独立のデータ項目によって視覚化できるため，セグメント化したデータの多次元相関関係分析に有効である．

たとえば，健康診断データには，喫煙や飲酒といった慣習と肥満度，血圧等の多くの測定データがあり，相互間の特徴的関係を知りたいという要求がある．図9(b)に，喫煙・肥満度グループ（親図形）と，飲酒（子図形の色）の関係を比較した例を示す．本例では，飲酒量の多い人（子図形の色が濃い）は，ほとんどの喫煙・肥満グループに属するが，飲酒量の少ない人（子図形の色が薄い）は，喫煙量が少ない（左側の親図形）と肥満度にバラツキは見られ，喫煙量が多い（右側の親図形）と肥満度が低くなるという傾向が見られる．また，包含図の拡張性を活かすことにより商品売上のような時系列データは，図9(c)のようなカレンダー表現することで，月・曜日・日の購買の特徴が一覧できる．通信網の設備管理データは，図9(d)のようなケーブル断面図表現することで，多種類あるリソースの在庫状況を一覧できる．本表現は，離散値を含むデータ項目にも図形を重複することなく，セグメンテーションと多次元相関を組み合わせた関係分析が可能である．さらに，順序配置，散布配置，地図配置等の配置方法を親図形と子図形に独立に適用することで多様な組合せ包含図表現が可能であり，適用範囲を拡張することができる．

(3) 網状図

網状図は，複数実体とそれら実体間の関係をノード図形とライン図形で表現できる．ノード図形は，データ項目に基づいて散布図的に多次元表現され，ライン図形はノード図形との関係を基に配置が，データ項目によって形状〔線幅，色，線種（実線，点線等）〕が表現されるため，ネットワーク的関係を含む多次元相関関係分析に有効である．

たとえば，通信網データにはビル・交換機・伝送装置といった点的データとケーブル・回線といった線的データがあり，地理的関係に加えて，点の特性（収容ユーザ数，接続回線数等）と線の特性（区間のケーブル種別，回線数等）を一覧したいという要求がある．図9(e)に，通信網の構成を地理的に表現し，ビルの収容ユーザ数（ノード図形の大きさ），接続回線数（ノード図形の色）と，ビル間の回線設備数（ライン図形の幅），使用回線数（ライン図形の色）の関係を比較した例を示す．ビルの規模と回線設備在庫の関係からアクションの必要な区間が一覧できる．また，本表現を散布図の例で示した経理データに適用すると，複数月分の同一支店の図形間をラインで結ぶことにより時系列の推移分析にも適用することができる．

さらに，網状図はノード図形の配置方法をカスタマイズすることにより，ノード図形を属性値やライン数に基づいて同心円状に配置する円状図や，Y軸（あるいはX軸）に平行する座標軸上にデータ項目ごとの値を示すノード図形を配置する平行座標軸図に容易に拡張することができる．たとえば，LAN内でサンプリング収集された通信相手マシン，通信回数，データ転送量等の通信量データから，LANの利用状態を把握したいという要求がある．図9(f)に，通信相手の多いマシン（同心円の中心近く）と，通信回数（ライン図形の幅），データ転送量（ライン図形の色）の関係を比較した例を示す．本例では，通信相手数にかかわらず通信データ量の多いマシンペアの存在が容易に把握できる．また，商品売上データには，バスケット分析のように併買商品の関係を知りたいという要求がある．図9(g)に，注目する商品（円の中心）と，併買された商品（外周図形）と，併買される確率（ラインの太さ）の関係を比較した例を示す．併買確率の高い商品の店頭配置の検討等に活用できる．さらに，アンケートデータには，多くの質問に対する回答があり，全体の質問と回答の関係を把握したいという要求がある．図9(h)に，各質問項目ごとの座標軸上に回答の値をプロットし，同一回答者の回答データ間をラインで接続することにより，回答の分布状況，回答者の傾向を比較した例を示す．本例では，どういった商品はどのような顧客層に好まれるかといった傾向を概観することができ，商品開発や宣伝方法の参考とすることができる．

以上のように，INFOVISERの視覚化表現は，多様な業務のデータ分析において，データ特徴と分析目的に合わせて視覚化表現を選択することにより，有益なパターンの発見を支援できることを示した．本評価で用いた図種別と発見されたパターンの関係を**表4**に示す．本表から，散布図はデータ項目間の関係，包含図はセグメント間の関係，網状図はデータ項目間の関係とレコード間の関係，円状図はレコード間の関係，平行座標軸図はレコードの値分布の特徴的パターン発見に有効であることが分かる．複雑な表現ほど，入力データへの制約はあるが，INFOVISERの実体抽出によりすべての場合に必要な実体を生成できることを確認できた．このように，視覚化表現と発見可能なパターンの関係が確立されてくると，データの特徴から

表 4 図種別と発見パターンの関係
Table 4 Relationship between visualization and discovery.

発見パターン＼図種別	散布図	包含図	網状図	円状図	平行座標軸図	業務例
データ項目間の関係	○		○			経理、健康診断
セグメント間の関係		○				健康診断、商品購買、通信網
レコード間の関係			○	○		通信網、経理、商品購買、トラヒック
レコードの値分布					○	アンケート
入力データへの制約	配置には連続値が必要	複数実体と関係が必要	実体と関係実体が必要	実体と関係実体が必要	——	

表 5 情報源構成への適用性評価
Table 5 Adaptability for source data format.

分類	データの例	分析の内容	データ抽出の定義量			
			部品数	マウス操作回数	定義時間(分)	規模(ステップ)
サマリ	例：97-98年度の支店別・月別営業成績	①支店間業績比較	2	19	6	290
		②支店別の比較	4	57	12	1880
		③期間別の比較	6	54	18	2300
トランザクション	例：97年度の商品売上レシート（POSデータ等）	①商品毎・期間毎の売上比較	5	55	15	1880
		②月日・顧客・商品間の関係比較	15	169	45	6210
		③併買商品の信頼度比較	10	105	30	4450
ネットワーク	例：ネットワークの構成(伝送路、交換機、ビル等)	①ネットワーク構成の全体把握	8	87	24	2630
		②特定伝送路中の回線構成把握	15	169	45	5520

自動的にパターンを選択し，有効な視覚化表現で表示するような自動的視覚化機能への適用の可能性が高まると考えられる．

5.2 情報源データへの適用性

情報源データへの適用性は，実体抽出機能が各種データ構成から特定の範囲や関係を持つ視覚化対象データを抽出できるか否かの機能的な満足性の観点から評価する．

情報源データ構成は，ビジネス分野において代表的な，(1) サマリデータ，(2) トランザクションデータ，(3) ネットワークデータの 3 種類とした．サマリデータには，経理，営業，企画等の伝票や成績が集約されたデータが，トランザクションデータには，カルテや商品購買のような伝票形式のデータが，ネットワークデータには，CAD，通信網構成のように入れ子関係を有するデータがある．

上記の例題データから，前述の 5 つの視覚化表現を作成するために必要な INFOVISER 定義操作量と展開されるソースステップ数の関係を**表 5** に示す．本表で示すように，3 種の例題データに対しては，情報源データの変更や外部ツールによるデータ加工をすることなく，INFOVISER の簡易な GUI 定義操作のみで分析対象データを抽出し，視覚化できることを確認した．本例における定義量は，平均で使用部品数約 8 個，マウスクリック約 90 回，定義時間約 24 分で，約 3 Kstep のプログラム開発に相当する作業を行うことができ，高い適用性を確認した．

以上のように，INFOVISER の実体抽出機能は，多様なデータ構成の情報源から少ない稼動で必要なデータを抽出できるという効果が確認できた．

5.3 視覚化表現の多様性

視覚化表現の多様性は，関連研究で述べた Keim の論文で引用されている視覚化手法の実現性の観点から客観的に評価することとした．

引用された視覚化手法と INFOVISER での表現可能性の検討結果を**表 6** に示す．本表で，○は完全に

表 6 既存視覚化手法の統合度評価
Table 6 Capability to existing visualization methods.

視覚化手法	INFOVISERの対応状況	判定
①Geometric	Scatterplots, Parallel Coordinates 共に可 色、形、大きさ、ﾗﾍﾞﾙ等の設定自由度大	○
②Icon-based	Chernoff Faces, Stick Figures, Shape-Coding は表情のｲﾒｰｼﾞ登録で代用可、Color Iconsは ４段階までの記号化は可能	○
③Pixel-oriented	Spiral Techniques, Circle Segmentの配置は不 可。図形をﾋﾟｸｾﾙと見做し少量ﾃﾞｰﾀには対応可。	△
④Hierarchical	Dimensional Stacking, Cone Treeは、配置法 の拡張で容易に対処可能。	▽
⑤Graph-Based	Orthogonal, Directed Acyclicは座標ﾃﾞｰﾀがあれ ば可。	○
⑥Distortion	Fisy-eyeはマルチ視点設定可。Hyperbolicは 円状図の拡張で対処可能。	○
⑦Interaction	Mapping, Zooming, Detail on Demandは可。 FilteringはSQL検索は可	○
	図形による検索, Linking & Brushingは不可。	×

○：対応可、△：擬似的に対応可、▽：簡易な追加で対応可、×：不可

実現できる場合，△は実現法は異なるが利用者からは擬似的に実現されたと見える場合，▽は比較的簡易な部品やメソッドの追加のみで対処できる場合，×は大きな追加を要する場合を示している．表中，△のPixel-orientedは，図形をピクセルと見なすことにより，数千件程度のデータまでは同じ機能を擬似的に提供できる．また，▽のHierarchicalは配置メソッドの拡張で対応できる．このように，大半の視覚化手法はINFOVISERで実現できることが確認できた．今後の課題は，INFOVISERの結果表示画面に，図形表現を用いたFilteringやLinking&Brushing等の高度な対話インタフェースを実現していくことである．

以上のように，INFOVISERは従来個別に研究されてきた各種データ視覚化手法を，ノードラインビューモデルという体系化されたデータ視覚化環境の中で統一的に実現できるところに特徴があり，試行錯誤的なデータ分析において分析対象のデータや視点が変更になった場合でも，簡易なGUI操作のみで，分析目的やデータ特徴に合った視覚化表現を選択しながら対話的にデータ分析を進めることができる．

6. まとめ

ビジネス分野における戦略的意思決定には，数値・文字が混在する複雑なデータを対象に，業務上の背景知識に基づいて人間が介在しながら，データ準備からアクションまでを非定型に試行錯誤できるようなデータ分析支援環境が求められている．

本論文では，これらの要求に応えられるデータ分析システムを確立することを目的として，情報視覚化モデルとデータ変換メカニズムを提案した．情報視覚化モデルでは，ユーザ定義に基づいて実体データをノード型かライン型の図形の形状や配置として多次元表現するという考え方を導入し，ユーザ定義可能なデータ変換メカニズムへの要求機能を明確化した．データ変換メカニズムは，簡易なユーザ定義により動作する情報源から任意の実体データを抽出可能なソフトウェア部品と，実体データを多様な図形データに変換可能なソフトウェア部品から構成することにより，ユーザ要求に応じた豊富な表現が提供可能であることを示した．また，試作した視覚的多次元データ分析システム（INFOVISER）を，実際のデータ分析に適用し，その有効性を確認した．本論文ではデータ分析への適用に絞って議論をしたが，INFOVISERは，周期の長い観測データの監視や情報検索ナビゲーション等の広範な分野への適用の可能性を有している．

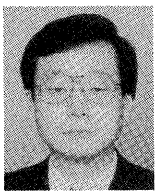
本研究では，分析者が検証したい仮説を有することを前提に，分析者が簡易なGUI定義画面操作を通じて仮説の軸を図形表現の軸に対応させて視覚化し，データを分析する方法を中心に述べた．今後，適用性を高めるためには，検証したいデータ項目が多い場合や仮説がない場合に，データの特徴から自動的に仮説の軸を発見してデータを視覚化する方式の研究や，データマイニングアルゴリズム等で抽出されたルールを視覚化して知識の精練を支援する研究等が有効であり，この分野での成果が期待される．

参考文献

1) 磯部，黒川，塩原：DB情報ビジュアル化技術，*NTT R&D*, Vol.45, No.1, pp.21–26 (1996).
2) Kurokawa, K., Isobe, S. and Shiohara, H.: Information Visualization Environment for Character-based Database Systems, *Proc. 1st*

International Conference on Visual Information Systems (VISUAL'96), pp.38–47 (1996).
3) Chaudhuri, S. and Dayal, U.: An Overview of Data Warehousing and OLAP Technology, *SIGMOD Record*, Vol.26, No.1, pp.65–74 (1997).
4) OLAP COUNCIL: *OLAP AND OLAP SERVER DEFINITION*, http://www.olapcouncil.org/research/glossaryly.htm
5) Keim, D.A.: Databases and Visualization, *Tutorial Notes of ACM SIGMOD International Conference on Management Data*, pp.1–81 (1996).
6) Keim, D.A.: Pixel-oriented Database Visualization, *SIGMOD Record*, Vol.25, No.4, pp.35–39 (1996).
7) LeBlanc, J., et al.: Exploring N-Dimensional Databases, *Proc. IEEE Visualization '90 Conference*, pp.230–237 (1990).
8) Keim, D.A. and Kriegel, H.P.: VisDB: Database Exploration Using Multidimensional Visualization, *IEEE Computer Graphics and Applications*, pp.40–49 (1994).
9) Beddow, J.: Shape coding of multidimensional data on a microcomputer display, *Proc. IEEE Visualization '90 Conference*, pp.238–246 (1990).
10) Pickett, R.M. and Grinstein, G.G.: Iconographic Displays for Visualizing Multidimensional Data, *IEEE Systems, Man and Cybernetics*, pp.514–519 (1988).
11) Lee, H.Y., et al.: Exploiting Visualization in Knowledge Discovery, *The 1st International Conference on Knowledge Discovery & Data Mining*, pp.198–203 (1995).
12) Inselberg, A. and Dimsdale, B.: Parallel Coordinates: A Tool for Visualizing Multi-Dimensional Geometry, *Proc. IEEE Visualization '90 Conference*, pp.361–375 (1990).
13) Becker, R.A., et al.: Visualizing Network Data, *IEEE Trans. Visualizing and Computer Graphics*, Vol.1, No.1, pp.16–28 (1995).
14) Mihalisin, T., et al.: Visualizing a Scalar Field on an N-dimensional Lattice, *Proc. IEEE Visualization '90 Conference*, pp.255–262 (1990).
15) Mackinlay, J.: Automating the Design of Graphical Presentation of Relational Information, *ACM Trans. Graphics*, Vol.5, No.2, pp.110–141 (1986).

(平成10年7月3日受付)
(平成11年2月8日採録)



磯部 成二（正会員）

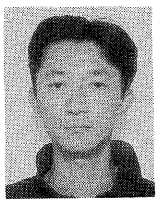
1949年生．1971年山形大学工学部通信工学科卒業．同年，日本電信電話公社（現NTT）電気通信研究所入所．以来，電子交換機ソフトウェア・ISDNオペレーションシステムの研究開発，通信網構成管理データモデルの研究開発，情報視覚化・データマイニングの研究開発等に従事．現在，NTTソフトウェア（株）担当部長（元NTTサイバースペース研究所）．IEEE-CS，電子情報通信学会各会員．

黒川 清（正会員）

1965年生．1988年九州工業大学工学部情報工学科卒業．1990年九州工業大学大学院工学研究科電気工学専攻博士前期課程修了．同年日本電信電話（株）入社．以来，データベース設計支援，情報資源管理，情報視覚化の研究開発に従事．現在，NTT東日本会社移行本部企画部グループ企業担当主査〔NTTソフトウェア（株）〕．主に情報システムの開発に従事．ACM，IEEE-CS各会員．

塩原 寿子（正会員）

1968年生．1990年大阪大学理学部物理学科卒業．1992年大阪大学大学院理学研究科物理学専攻博士前期課程修了．同年日本電信電話（株）入社．以来，データベース流通方式，情報視覚化・データマイニングの研究開発に従事．現在，NTTサイバーソリューション研究所研究主任．



飯塚 哲也（正会員）

1965年生．1989年群馬大学工学部情報工学科卒業．1991年群馬大学大学院工学研究科情報工学専攻博士前期課程修了．同年日本電信電話（株）入社．以来，データベース移行支援，データベース性能評価支援の研究開発，情報視覚化・データマイニングの研究開発等に従事．現在，NTTサイバーソリューション研究所研究主任．