

データ視覚化による情報分析手法の一検討

1 S-7

飯塚 裕一 黒川 清 磯部 成二
NTT情報通信研究所

1 はじめに

基幹系システムの中にある複数の業務データベース（DB）から取り出したデータの規則性や特徴を把握することにより、戦略的意思決定を支援するシステムの構築が望まれている¹⁾。データの規則性や特徴を把握するための分析は迅速かつ多角的に行わなければならないが、現状では分析プロセスおよび統計処理や知識処理結果の解釈に精通したユーザでなければ有効に行うことは不可能である。このことから意思決定者が簡易に操作することが可能で、結果を直感的に解釈できる情報分析環境が必要であると考えられる。

我々はデータベースに蓄積された文字／数値情報を図形情報へ変換する視覚化が、直感的な状況把握や情報分析の支援に有効であると考え、エンドユーザによるアプリケーション開発・実行環境を提供するDB情報ビジュアル化システム²⁾（INFOVISER）の研究開発を行っている。

本稿では、大量データの視覚化による情報分析をINFOVISERで行う際の課題と解決案について述べる。

2 INFOVISER

INFOVISERでは、文字や数値の情報集合をノード型やライン型のオブジェクトとして表示する。ノードラインビューモデル³⁾にもとづいてデータを視覚化する。この情報集合を表示実体と呼び、ユーザはGUI操作により任意に定義することができる。図形表現するための中心的な定義は情報変

換定義である。情報変換モデルを図1に示す。

文字／数値情報からの図形情報への情報変換は、実体オブジェクトが持つスキーマ（クラス、属性）および内容（インスタンス、属性値）などの情報から、画面オブジェクトが持つ図形の型や配置、ラベル、形状、大きさ、色などの情報へのマッピングおよびメソッドを定義することによって行う。

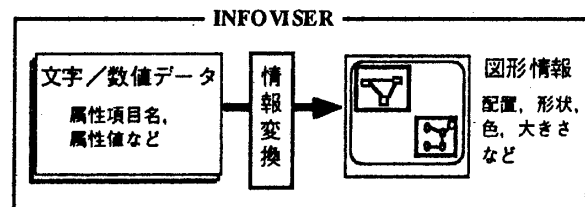


図1 情報変換

3 データ視覚化による情報分析

情報変換定義は、通常はユーザが背景知識をもとに、どのような視点で分析すれば有益な知見が得られるかについて仮説をたて、それに基づいて生成するか、あるいは試行錯誤を繰り返すことによって生成される。

しかし大量データを扱う場合には以下の課題がある。

- 1) 複数業務にまたがるデータを扱う場合にはユーザが十分な知識を持ち合わせないため定義が困難である
- 2) 多次元データの視覚化ではマッピングで参照する属性の定義が困難である
- 3) 仮説からは導き出せないような予想外の情報を見逃してしまう恐れがある

これらの課題を解決するためには、情報変換を有効に行うための支援が必要であると考えられる。

情報変換の支援を、大量情報から特徴を抽出

A Study on Information Analysis by Visualization

Yuichi IIZUKA, Kiyoshi KUROKAWA and
Seiji ISOBE

NTT Information & Communication Systems Labs.

し、得られた特徴に基づいて定義を生成する方針で検討する。

3.1 特徴の抽出

大量情報からの特徴抽出は、

- 1) 文字／数値データの解釈
 - 2) メタデータの意味的な解釈
 - 3) 視覚化により図形表示された結果の解釈
- によって得られると考えられる。

上述の2)については、相互運用性の高いIRD (Information Resource Dictionary) の構築が必要であること、また3)に関しては、図形表現の解釈はユーザの分析目的などに大きく影響されることなどで困難と考え、第一段階として1)の検討を行うこととした。特にデータ解析において有効である統計処理を用いる。

分析手法は、

- ・対象とするデータの種類 (質的／量的)
- ・分析目的 (相関、グループ分けなど)

によって分類され、適用可能と考えられる手法を用いて文字／数値データから属性間の相関や類似項目のグルーピングなどの特徴抽出を行う。

3.2 特徴に基づく定義生成

得られた特徴をもとに図形表現するための定義を生成する。このとき表示結果から読み取ることができる情報について考慮し、生データとの対応関係を明らかにすることが必要である。

図形表示結果から読み取ることができる主な事項としては、

配置：XYの相関、特異値の存在

形状：値の種別、特異値の存在

色：値の段階的变化、特異値の存在

大きさ：値の段階的变化、特異値の存在

および、これらの組合せによる複合的な情報にまとめられる。

そこで、生データでの項目間の相関や類似項目のグルーピングなどの特徴をもとにマッピング、メソッドの各定義の組合せ手法について検討し、有効と判断される結果を提案する。

上述の機能を実現させるシステム構成を図2に

示す。現状の定義機能には手を加えず、情報変換支援機能として選択的に利用できるようにする。

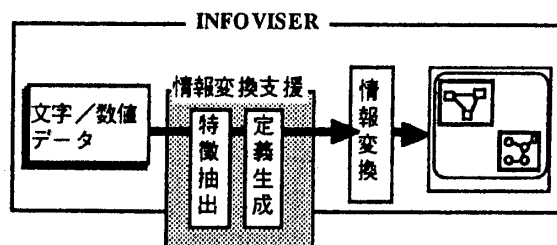


図2 システム構成

4 期待される効果

視覚化による情報分析では、図形表現することにより文字／数値データが持つ意味を解釈する。したがって、文字／数値データと図形の対応付けの手法が有益な分析のために重要な役割を担う。

本稿で述べたシステム側から分析に有効と思われる図形表現パターンの提案を行うことで、業務や分析等に関して専門的知識を持たないユーザが意思決定する際の支援が可能となる。また、専門家が分析を行う際の稼働の削減が期待される。

特徴抽出に関して機能向上をはかることで、より有効な図形表現の提案を行い、有益な情報の見逃しを防止することができると考えられる。

5 今後の方針

INFOVISERへ実装し、特徴に基づいた視覚化の実現を確認するとともに、特徴表示効果を評価する。また、視覚化に固有の情報分析について明らかにする。

参考文献

- [1] W.H.Inmon, 初めてのデータウェアハウス構築
トムソン・パブリッシング, 1996.
- [2] Isobe et al. "Object-Oriented Information Visualization Environment for Network Management"
ECOOP'94 Workshop7 Proceeding, 1994.
- [3] 磯部他 "ネットワーク情報ビジュアル化方式の検討" 電子情報通信学会IN研究会, SSE94-34, 1994.