

情報視覚化技術の多次元データ分析への適用

1S-6

黒川 清、磯部成二、塩原寿子

NTT情報通信研究所

1. はじめに

市場のグローバル化にともなう環境の劇的な変化により、高度な情報分析に基づく意思決定が重要な課題となっている。意思決定支援のためのデータベースの高度利用のためには、正確でタイムリーな情報の提供が必要である。特にデータベース中の大量の文字数値データを意思決定支援に活用することが求められており、このために有効な技術として情報視覚化技術が期待されている。

我々は、数値文字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ノードラインビューモデルを考案し、このモデルを適用したビジュアル化アプリケーションの開発実行環境の研究を進めている[1][2]。本稿では、我々が研究開発中の視覚化対象データへの視点設定方式が、多次元データ分析の実現方式に類似していることに着目し、情報視覚化技術の多次元データ分析への適用について述べる。

2. 視覚的な多次元データ分析の要求条件

多次元データ分析とは、データの多次元性を用いた動的なデータ分析手法である。この多次元性を持つデータ項目を自由に切り替え、さまざまな角度からデータを分析する。多次元データ分析ツールとして多くのものが存在するが、そのほとんどは、ユーザが直感的なオペレーションで操作できるようにするため、スプレッドシート型のインターフェースを持っており、多次元データを特定の視点から検索した結果を2~3次元のグラフとして表現しているものが多い。しかし、多次元データ分析の最終的な結果が、数字の羅列や一般的なビジネスグラフでは、情報の本質、すなわち、多次元のデータ相互の関係を見過ごしてしまう可能性がある。

この問題を解決するために、情報視覚化にもとづく多次元データ分析について検討を行った結果、以下の要求条件を満たす必要があることがわかった。
 関係表現の多次元性：分析対象データを抽象化し特徴を際立たせることでユーザの理解を支援すること。

視点変更の容易性：ユーザの意図に沿って多次元表現が自在に変更できること。

対象データの独立性：基幹業務DBに依存しないデータ構造であること。

3. アプローチ

前述の要求条件を満たすために、ユーザの直感的な思考を支援する機構である「ダイス&スライス」への、視覚化のベースとなるノードラインビューモデルの適用を検討した。

3.1 ノードラインビューモデル

ノードラインビューモデルは、実データをノードとラインの単純な図形情報に変換し視覚化するモデルである[3]。ノードラインビューモデルの概念を図1に示す。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Mappingにより、ViewのスキーマであるTypeとForm、Coordinateに変換する。Entityの内容であるInstance、Valueは、値変換方法を表わすMethodにより、変換結果として表現される各値(Node、Line、Form-value、Coordinate-value)に変換する。実データの属性情報は、表示されるオブジェクトの型に応じて、カラーなどの形状情報や座標などの配置情報に変換されて図形表現される。

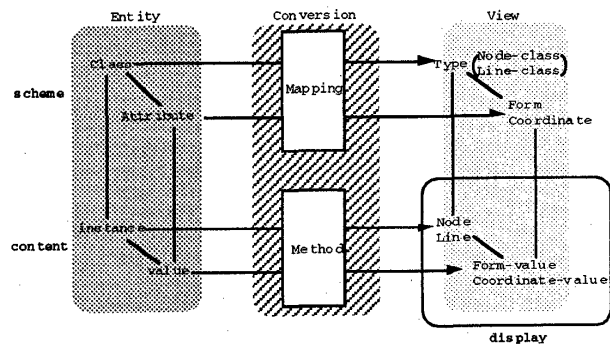


図1 ノードラインビューモデル

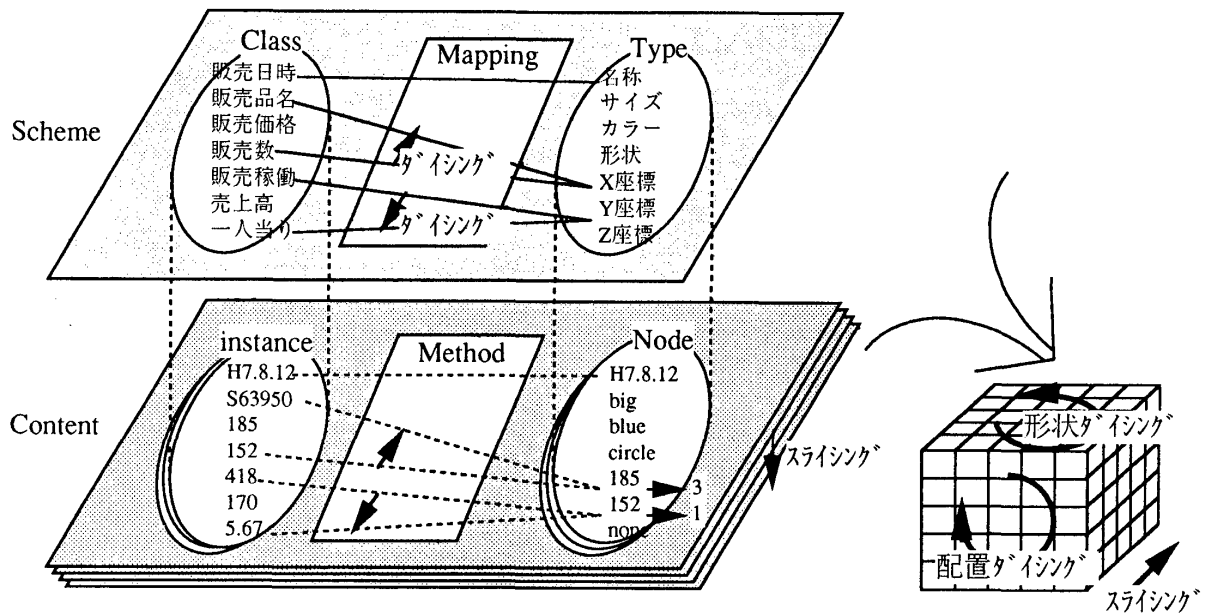


図2 ダイス&スライスの処理概念図

3.2 ダイス&スライス

ノードラインビューモデルに基づく情報変換方式について述べる。図2は、ダイス&スライスの処理概念を表わしたものである。

ダイシングに相当する機能は、情報変換元の実体オブジェクトの参照属性と情報変換先の表示オブジェクトの表示属性とのマッピングであり、ユーザが焦点を当てたい項目を必要に応じて選択することにより、動的に視点を変更することができる。形状(名称、カラー、サイズ、形)と配置(2次元の場合、X軸、Y軸、3次元の場合、+Z軸)をそれぞれマッピング対象とすることにより、形状ダイシング、配置ダイシングといった、視覚的な多次元データ分析に特徴的な機能を実現する。情報変換メソッドと併用することにより、特定属性値の強調表示も可能である。

スライシングに相当する機能はオブジェクトグループ化機能である。ユーザの指定したグループ化定義により、データ項目毎の属性値に応じて、実体オブジェクトのグルーピングを行う。その実体オブジェクトグループを頁に対応させて、その単位ごとに表示オブジェクトを生成することにより、複数頁で結果を連続表示することができる。

4. 考察

視覚的多次元データ分析における要求条件について考察を行う。

関係表現の多次元性：インスタンス個々の特徴と全体傾向を把握するための表現が可能で、各インスタンスの多次元属性が形状(色、形など)、配置により図形表現できる。

視点変更の容易性：マッピング、メソッドを変更することにより、視点変更を行うことができ、多くの表現パターンを選択することができる。

対象データの独立性：データの情報特性(属性、値)を利用して変換を行うことにより、業務DBのデータ構成に依存しない。

その他、表現結果からわかることは、視覚的データ分析の場合、分析者の意図に合わせてラベルや配置情報など表示情報をもとに、背景知識などを連想の道具として用いることにより、効率の良い分析を行うことができることである。しかし、これは逆に分析者の背景知識、分析スキルに依存するところが大きいといえる。

5. おわりに

本稿では、視覚化対象データへの視点設定方式が、多次元データ分析の中核技術実現方式に類似していることに着目し、情報視覚化技術の多次元データ分析への適用について述べた。今後は、本方式の有効性の評価を行うとともに、実装を進める予定である。

参考文献

- [1]黒川,他:"文字/数値情報の図形表現高度化の検討", 情報処理学会大51回全国大会 4-109, Sep. 1995.
- [2]K.Kurokawa, et al.:"Information Visualization Environment for Character-based Database Systems", Proceedings of The First International Conference on Visual Information Systems, pp.38-47, Feb. 1996.
- [3]黒川,他:"情報可視化のためのデータビジュアル化モデル", 情報処理学会研究報告 96-HI-65, Vol.96, No.21, pp.51-56, Mar. 1996.