

## データ分析支援ツール“DBSOLVER”（1）

7G-2

## -狙いと機能概要-

金枝上敦史 宮崎壽和† 萩原正敏 小林 博† 近藤了通†  
 三菱電機（株）情報システム研究所 †情報システム製作所

1. はじめに

近年のコンピュータのダウンサイジングと、RDBの普及により、システムへの拡張性と透過性の要求が増加し、益々データ中心アプローチによるシステム開発が重要となってきている。このデータ中心アプローチのシステム開発において、システム化対象の業務に現れるデータをいかに分析・整理し、拡張性があり、ユーザに扱い易いデータベースを構築するかがポイントとなる。

業務に現れるデータ項目名は不統一で大きなシステムでは数万もの名称が重複のあるままに出現する。

データ分析ではこれらデータ項目名を標準化と正規化を繰返すことにより、データ項目間の関連およびデータ項目グループ間の関連を明確化にしてゆく。そして、データ項目の重複をなくし、データベースの論理構造を決める。

しかし、従来データ分析は短期間の開発ではあまり実施されなかった。なぜならデータ分析には長い期間を必要とし、また分析のためのノウハウを必要とするため分析の専門家が必要であった。そこで短期間に分析でき、分析の専門家でなくても容易にデータ分析がおこなえることを目標にデータ分析支援ツール（DBSOLVER）を開発した。

今回の発表では、DBSOLVERの狙いと機能概要[本稿]、ツールを使ったデータ分析の流れ[1]、ER図統合機能[2]についておこなう。

2. 背景

我々はツール開発に当たってまずトップダウンとボトムアップの2つの分析手法を整理した。そして、ボトムアップ分析はデータ項目を洩れなく抽出するのに適しており、データベース設計に有効と考えた。しかしその反面ボトムアップ分析は分析者に多大な作業負荷をかけることも判明し、さらにこの分析を支援する有効なツールがないことから、ボトムアップ分析の各手順を一貫して支援するDBSOLVERを開発することとした。

3. ボトムアップ分析の問題

ボトムアップ分析とは伝票・帳票・画面定義や現行システムのファイル定義など（以下フォームと呼ぶ）に現れるデータ項目を元に標準化、正規化による分析をおこないデータモデルであるER図を作成することである。実際の分析は分析者がフォームのデータ項目に対して標準化と正規化、そしてER図の統合の3つの作業を試行錯誤を繰り返すことでおこなわれる。

従来このようなボトムアップ分析には次のような問題があった。

- ・分析結果を入力するツールはあるものの、分析作業を最初から支援する有効なツールがない。
- ・手による標準化、正規化の作業は膨大なデータ項目名の書き写しが必要で、そのために品質の低下が発生した。
- ・下流工程で利用可能なER図やデータ項目辞書を得るまでに長い期間と分析者の根気を必要とした。

我々は上記の問題を解決するためボトムアップ分析の手順をデータ項目入力、正規化、ER図統合、制約定義、ER図洗練の5つに分け、各手順をツール支援

Tools for Data Analysis: DBSOLVER Goals & Facilities  
 Atsushi KAWAGAMI, Toshikazu MIYAZAKI, Masatoshi IWAIWARA, Hiroshi KUBAYASHI, Norimichi KONDOH  
 MITSUBISHI ELECTRIC CORPORATION

すると共に、各手順でデータ項目名（属性名）の標準化ができるツールを実装した。

#### 4. ツールの特長

DBSOLVERは以下のような特長を持つ。

- (1) データ項目の書き写しなどの単純作業から分析者を開放し、分析者は本質的な分析に専念できる。
- (2) 書き写し作業をなくすことにより分析の品質を大幅に向上させる。
- (3) ER図の骨子を早期に作成することにより分析期間の短縮化ができる。
- (4) ER図の作成と並行してデータ項目辞書を作成できる。
- (5) 5つの分析手順の分析結果をリポジトリに保持しており、分析に使った伝票類のデータ項目とER図・データ項目辞書との対応関係を持つ。
- (6) リポジトリに格納された分析データを元に分析洩れ、分析誤りをチェックする。

#### 5. 機能概要

図1のとおり分析の各手順に対して以下の支援機能を持つ。

**データ項目入力**：フォームごとのデータ項目を管理し、フォーム上でのデータ項目名、データ特性や標準化後のデータ項目名を保持する。

**正規化**：データ項目の正規化結果をグラフィカルに表示し、分析者は正規化規則に従いデータ項目のグループ化をおこなうだけで、エンティティ、属性および関係はツールが管理し、ER図で表示する。

**ER図統合**：複数のフォームを正規化した結果得られるいくつものER図中には類似のエンティティが多く出現する。これらの類似するエンティティの検索を容易にし、同一と判断されたエンティティの属性、関係を自動で統合することができる。

**標準化**：標準化は分析者がデータ項目（属性）にドメインと標準データ項目名を設定することによりおこなわれる。これを支援するため分析者が設定したドメイ

ンから標準データ項目を抽出したり、標準データ項目を定義したフォームのデータ項目の情報を表示したりすることで、分析者が重複なく標準データ項目を定義できる。

**詳細特性定義**：分析者はER図の作成と共に、エンティティ、属性、関係の制約条件を定義する。それら定義を入力し分析データとして保持することができる。

**モデルチェック**：上記の標準化、正規化、統合の機能によりER図の骨格を早期に得ることができる。この骨格に対して分析洩れのデータや未標準のままのデータ項目、他のエンティティと関係が張られていない孤立エンティティの有無などを分析者に知らせ、分析洩れや分析間違いのチェックをおこなう。

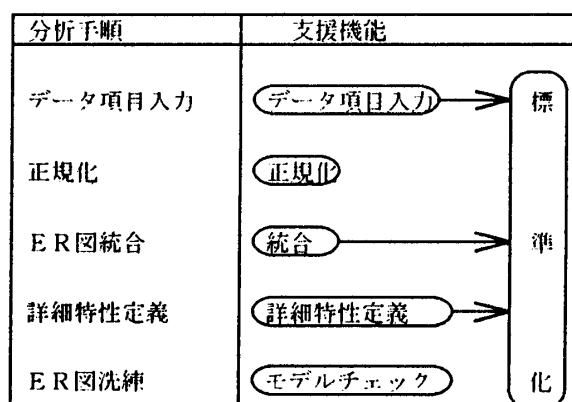


図1 支援機能と機能依存関係

#### 6. 今後の課題

今回、データ分析過程を支援するDBSOLVERを開発した。今後このツールの実使用を通して機能改良をおこなうと共に、データ分析以降の設計・製作・試験などの支援ツールとの連携を図る。さらに分析結果の実システムでの検証機能を構築し、ツールに蓄積された分析データの再利用の枠組みを明確にしてゆく。

#### 参考文献

- [1] 鈴木 他, データ分析支援ツール“DBSOLVER”(2) : 分析の流れ, 情報処理学会第50回全国大会7G-3
- [2] 山口 他, データ分析支援ツール“DBSOLVER”(3) : ERモデル統合機能, 情報処理学会第50回全国大会7G-4.