

データ分析支援ツール“DBSOLVER” (2)

7G-3

- 分析の流れ -

鈴木 由美子, 上野 浩一郎, 金枝上 敦史, 山口 陽一[†], 萩原 正敏

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 の普及の流れの中で、システム開発におけるデータ中心アプローチ (DOA) の重要性が高まっている。ここに報告するツール DBSOLVER[1] は、この DOA を実現するためのデータ分析作業を支援ものである。本ツールは、分析対象業務で扱われるデータ項目を標準化、正規化、統合して、システム全体のデータモデルを構築するデータのボトムアップ分析作業を支援する。本稿では、この DBSOLVER を用いたデータ分析の流れと特長を、標準化と正規化を中心に報告する。

2 データ分析の流れ

本ツールは、一連のボトムアップ分析作業を支援するべく、(1) 対象業務で使用される伝票、画面、帳票類 (以降フォームと呼ぶ) とデータ項目の定義、(2) データ項目の標準化、(3) 正規化、(4) 正規化により切り出されたエンティティ等の統合 [2]、(5) エンティティ、リレーションシップ (関連)、属性に対する制約等の詳細特性定義、(6) モデルチェックの6つの機能から構成されている。これらの機能を用いたデータ分析作業の流れは図1に示す通りであり、標準化の柔軟性、分析作業の効率化、分析情報の一貫性の保持の3点を特に留意して設定されている。

3 データ項目の標準化

3.1 標準化の位置付け

データ項目は各フォームで様々な使われ方をしている。標準化とはそうしたデータ項目を全体で一意に統一する作業で、システム全体で共通基盤となるデータモデルを構築するためには必須な作業である。

通常のボトムアップ分析手順では、この標準化またはそれに準ずるデータ項目名の統一は、正規化の前処理と

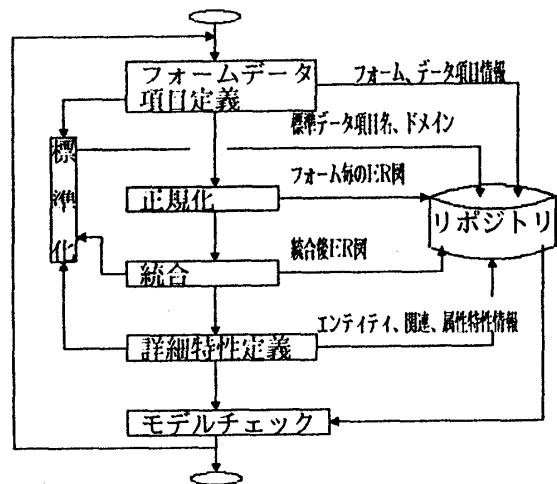


図1: データ分析の流れ

して位置付けられている。確かに、まずデータ項目の意味を良く理解して標準化しなければ正しく正規化、統合はできない。しかし、実際の分析では正規化、統合作業によりデータ項目の理解が深まり、同音異義語や異音同義語も判明していく。また、分析の初期に全データ項目を抽出し、標準化すること自体作業負荷が高く、データモデル構築を遅らせる原因になっていた。

そこで本ツールでは図1に示すように、分析を通じて順次標準データ項目を設定し、ER図が完成したときに、最終的なデータ項目の標準化が完了できるようにした。

3.2 標準データ項目の扱い方

標準データ項目と属性という粒度として同レベルの概念を比較した場合、全体をフラットに眺める標準データ項目よりは、業務全体のデータ構造を反映したER図の構成要素である属性の方が、その意味を把握し理解するのが容易である。そこで本ツールでは、データのタイプ、桁数、許容値、導出内容といった詳細特性は、属性に対して整理していく立場をとっている。ただ、同一または類似したデータ項目を検索するにはER図という構造だけでは不十分な場合があるので、データの取る値を示すドメインを設定し、ドメインによる類似データ項目の検索を可能としている。

従来のように、標準データ項目という視点を重視する

Tools for Data Analysis : DBSOLVER (2) - A Process of Data Analysis-

Yumiko SUZUKI, Kouichiro UENO, Atsushi KANAEGAMI, Youichi YAMAGUCHI, Masatoshi HAGIWARA
Mitsubishi Electric Corporation

場合は、属性を全体でフラットに捉え、その内容を標準データ項目名やドメインを用いて検索することも可能である(図3の標準データ項目情報を参照)。

4 正規化

正規化は、フォームから抽出されたデータ項目群をキーとそれに従属するデータ項目という視点から論理的なデータ項目のグループ(すなわちエンティティ)に分類する作業である。この作業はデータ項目名の書き写しや、キー設定の誤りなどによる分析のやり直しなどが多く、作業の効率化が望まれている。

4.1 正規化支援機能

本ツールは、キーとそれに従属するデータ項目の判定といった分析者の判断や名称入力以外は極力自動化すると共に、そうした判断や名称入力のガイダンスならびにチェックを行なっている。また、エンティティの切り出しで生じるキーと外部キーとの対応付けをグラフィカルに表示し、視覚的に正規化内容が理解できるように支援している。図2は正規化画面例である。データ項目の集まりがエンティティ、その中で下線が引かれているデータ項目(実画面では赤く表示されている)がキー、外部キーの参照先が矢線で表示されている。本ツールでは、フォーム(または統合後エンティティ)毎に正規化するが、その正規化結果はフォーム毎のER図として保存され、分析内容のレビューやビューの理解に利用することができる。

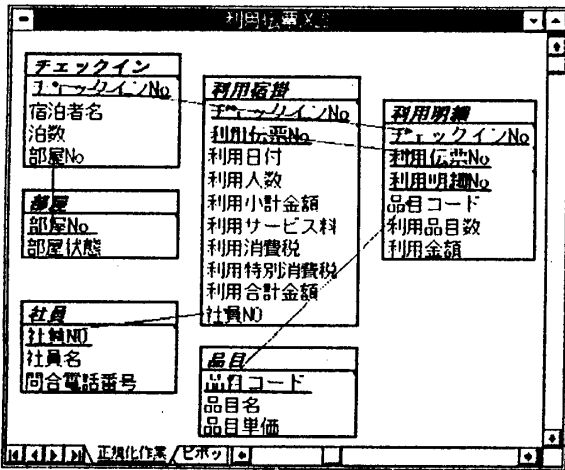


図2: 正規化画面例

4.2 作業の効率化

支援ツール開発に先立ち、11種類のワークシートを用いたボトムアップ分析技法を整備し、試行した。この技法において正規化には正規化支援シートが存在し、データ項目の転記の削減等の工夫がしてある。しかし、他作

業との分析結果の受渡しや正規化のやり直しなどで、作業の効率化には限りがあった。

本ツールでは、リポジトリによるデータの共有や作業のガイダンス等により、同じデータは一度入力すればよく、分析者は単純なルーチン作業から解放され本質的分析作業に集中することができる。データ項目の転記回数だけをみても1/3から1/6に減っており、作業の効率化が図られている。

5 メタモデル

DBSOLVERが管理する分析データとそれらの関連の概要を図3に示す。フォームとデータ項目情報、フォーム毎のER図、統合により生成されたER図(詳細特性情報を含む)、標準データ項目の情報が分析過程を反映して関連付けられている。この図3は、正しく分析した最終段階での構造を示している。実際は、分析誤りのデータや作業途中のデータを保存するために、図3とは矛盾する内容も保持する。例えば、本来標準データ項目と属性は1対1であるべきだが、正規化、統合の判断状況によっては1対多となる可能性がある。こうした矛盾はモデルチェック結果のレポートとして提供している。また、これら一貫して保持した情報より、ER図情報はもとより、フォームとエンティティとの対応表といった分析内容を解析するレポートも提供している。

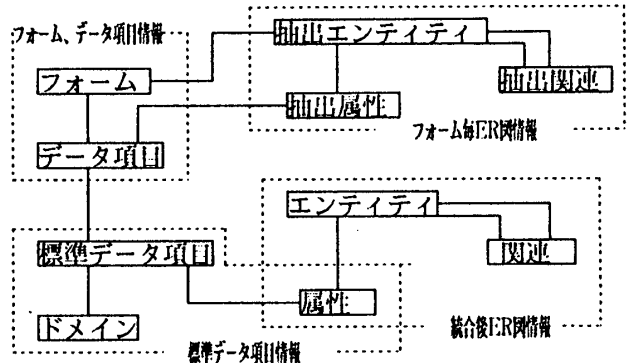


図3: DBSOLVERの概略メタモデル

6 おわりに

データ分析支援ツール DBSOLVERを用いたデータ分析の流れとその特長を説明した。現在、本ツールを実作業へ適用し、評価をしている。今後は、単なるルーチン作業の支援のみならず、そうした適用事例をベースに分析者の思考自体の支援を検討していくつもりである。

参考文献

[1] 金枝上, 他: データ分析支援ツール DBSOLVER(1)-狙いと機能概要-, 情報処理学会第50回全国大会, 7G-2.
 [2] 山口, 他: データ分析支援ツール DBSOLVER(3)-ER図統合支援機能-, 情報処理学会第50回全国大会, 7G-4.