

## データマイニング支援ツールの開発とその評価

三井田 浩<sup>†</sup> 和田雄次<sup>‡</sup>

東京電機大学大学院 情報環境研究科<sup>†</sup> 東京電機大学情報環境学部<sup>‡</sup>

### 1. はじめに

近年は、情報社会とも呼ばれ、様々な種類の情報が溢れている。それに伴い大量のデータが蓄積されるようになった。今日では、そういった大量のデータに対し、通常では分かりづらいデータに含まれる規則性やルールなどの有益な情報を発見することを目的としたデータマイニングを行う企業が増えている。データマイニング結果は分析者が確かめる必要があり、どの部分が有効なものかどうかを判断することが難しい場合がある[1]。

これらを踏まえ、本研究では被分析データとデータマイニング結果を合わせて可視化をすることにより、分析者がデータマイニング結果を理解し、より有効な情報を選択することを目的としたデータマイニング支援ツールの開発を行った。本稿では、データマイニング支援ツールの概要について述べ、トランザクションデータを本ツールに適用し、その有効性を報告する。

### 2. データマイニング可視化技術

本章では、従来のデータマイニング可視化技術とその問題に対し、本ツールによる問題解決のためのアプローチについて報告する。

#### 2.1 データマイニング可視化技術

データマイニングの可視化方法として代表的なものに決定木がある。この決定木には複数のルールが含まれていて、分類していくにつれ、枝やノードが増

加し、導き出されるルールも複雑なものとなる[2]

分析者は、このような決定木作成ツールにより作成された、たくさんのルールを含む決定木の中から有効なルールを発見し、それを活用するにはどうすればよいか問題となる[3]。

#### 2.2 本研究成果の特徴

2.1 節で述べた問題に対する解決方法の提案として、本研究ではデータマイニング結果を導くために用いた被分析データとデータマイニング結果で抽出された属性情報を合わせて可視化をし、比較する手法を提案する。これにより、被分析データに対してデータマイニング結果がどのようなルール（例えば、属性間での関連やアソシエーションルールなど）、または違いなどを視覚的に判断することができる。その結果、分析者にとってデータマイニング結果の有効活用度が高まることが期待できる。

### 3. データマイニング支援ツール

今回開発を行ったデータマイニング支援ツールでは、2.2 節で述べた特徴的な機能を実装し、分析者による操作により、被分析データとデータマイニング結果の比較を行う。

#### 3.1 データベース事例

現在では、至る所で様々な情報や履歴をデータとしてデータベースに保存をしている。今回、データマイニング支援ツールに用いたコンビニエンスストアのトランザクションデータには立地、取引された日時、購入商品など合計 20 の属性が含まれている。

#### 3.2 機能説明・全体の流れ

今回開発を行ったデータマイニング支援ツールのユ

The development of the data mining support tool and the evaluation  
<sup>†</sup>Hiroshi Miida: Graduate School of Information Environment, Tokyo Denki University  
<sup>‡</sup>Yuji Wada: Faculty of Information Environment Tokyo Denki University

一ケース図を図 1 に示す。主な機能は被分析データ可視化、データマイニング結果 (決定木ルール) の可視化、分析者の選択したルール情報の表示である。これらを GUI により分析者が操作をし、目的の情報について可視化を行う。

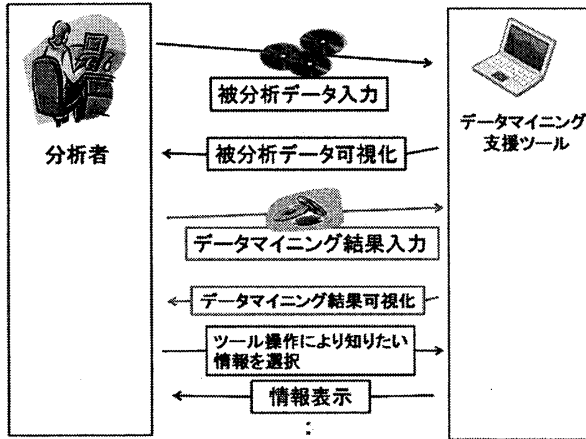


図 1 ツール使用のユースケース図

#### 4. 可視化事例

データマイニング支援ツールにより可視化した例を図 2 に示す。入力したデータからある選択属性値について三つの指標、サポート値、確信度、リフト値について、それぞれの値の大きさに応じて 4 段階にランク付けをし、可視化をする [4][5]。そして色分けされたラベルをクリックするとそのルールと指標が表示される。またデータマイニング結果を合わせて可視化をすることにより、購買商品に特化した

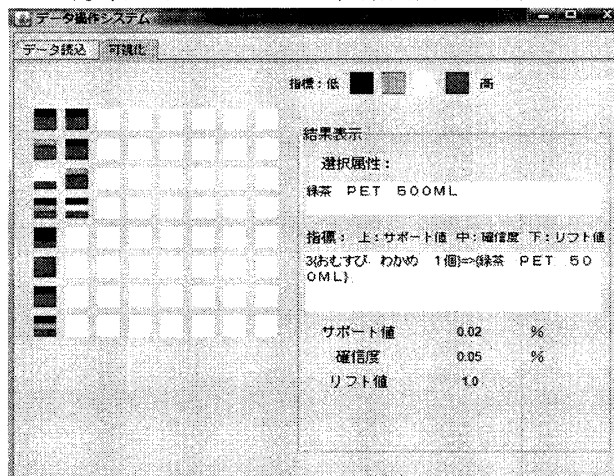


図 2 本ツールによる可視化例

ルールと全データを含むデータを合わせて比較することが可能となる [6][7]。その後、データ読み画面から別の知りたい情報を選択し直すことも可能である。

#### 5. 考察・まとめ

本ツールを用いる利点として、情報量が多く分析者がデータの内容を理解していない場合であっても、分析者のマウス操作により、知りたい情報を切り替えながら調べることができる点が挙げられる。今回は、開発を行ったデータマイニング支援ツールにより、知りたい属性値を選択することでアソシエーションルールの分析を行うことができた。今後は、分析者がより結果を理解しやすくする方法を考案し、実装させていく必要がある。

#### 謝辞

本研究は、科学研究費補助金 (基盤研究(C) 課題番号 20500095 「ユビキタスデータベース仮想化技術によるデータ理論の効率化に関する研究」) の支援による。

#### 参考文献

- [1] 福田剛志, 森本康彦, 徳山豪, データマイニング, 出版社(2001)
- [2] 伊藤貴之, 山口裕美: 「データ宝石箱」～ビジュアルデータマイニングの実現に向けて～, 情報処理学会研究報告, 2003-DBS-130(13), 2003-FI-71(13), (2003/5/23)
- [3] 日本アイ・ビー・エム(株)河村佳洋, 福田剛志: データマイニング活用ガイド
- [4] Ian H. DATA MINING, Practical Machine Learning Tools and Techniques, MORGAN KAUFMANN(2003).
- [5] 三井田浩, 和田雄次: データマイニング結果の可視化, 比較・分析支援ツールの開発, FIT2009
- [6] 金明哲: R によるデータサイエンス-データ解析の基礎から最新手法まで-
- [7] 間瀬茂: R プログラミングマニュアル