1ZA-4

データマイニングと OR の手法を用いた売上情報分析システムの構築

奥津 元靖* 松井 藤五郎† 大和田 勇人†

東京理科大学理工学研究科経営工学専攻* 同 理工学部 経営工学科†

1 序論

商品販売を行う企業ではより高い収益を上げるために様々な売上情報分析が行われている．そんな中で，近年商品の組合せによる相乗効果が注目を集めている．この商品の組合せを分析するには一般的に相関分析というデータマイニング手法が広く用いられている．

一般的に相関分析の手法を小売業に適用する事は有効である．しかし，近年ではオンラインショップの利用が急速に広がっており，取り扱う商品数の増大によるルール評価がより困難になっている．

そこで，本論文ではこの評価の困難さを軽減する為に，オペレーションズリサーチ（OR）の手法を用いてルールの利益効率性を評価し，より高い収益に繋がる商品の組合せを発見するシステムを構築する．

2 設計方針

本論文で構築するシステムには次の 3 点が必要になる．

- 相関分析機能
- 利益効率性の計算
- 導入・操作の容易性

相関ルールの導出には標準的な Apriori アルゴリズム [1] を用いる．ただし，あまりに低い確信度と信頼度を持つルールについては有用とは言えないので適切な閾値を指定する必要がある．

広く用いられている相関分析であるが，ルールの評価として用いられている確信度や信頼度の値は確率的な数値である．しかし，実際には発生する確率が高くても収益に繋がらないルールであれば意味が無い．例えばセール品のように薄利多売を目的としたものが高い評価を得ることもある．通常このようなルールは既知のものである．そこで，本論文では利益効率性という指標を導入しルールの評価を行う．利益効率性とは，ルールによって選ばれた商品の組合せがどれだけ利益に貢献しているかを示す指標である．効率と付いているのは費用や売上だけでなく，その商品を製造・輸送・販売するのにかかる時間や労力など，金銭面だけでは評価出来ない商品販売全体の効率も考慮する為である．

最後にシステムの導入についてであるが，本論文では全てのサーバや計算モジュールをオープンソース等のフリー環境で構築する．こうすることで，誰でも容易にこのシステムを導入する事が可能である．また，操作インターフェースを出来るだけ簡潔にする事で，利用面でも計算過程等を意識せずに操作出来るように実装を行う．

3 システム構成

本論文で構築するシステムではデータの一元管理と複数のシステム利用者がいる事を想定し，サーバ・クライアント形式とする．システム構成は次頁の図のようになっている．

本システムは web アプリケーションとして動作し，web サーバとデータベースサーバの 2 つのサーバからなっている．こうすることで利用者は web ブラウザから容易にアクセスする事が可能であり，見慣れた web ページ上の操作と同様に扱うことができる．web サーバにはシステムの処理をモジュール単位で実装する．

クライアントは web ブラウザから web サーバにリクエストを送ることで各モジュールにアクセスして処理を行う．各モジュールはリクエストを受け取ると必要な情報をデータベースサーバから取り出し，処理を行い，結果をクライアントに返す．モジュールの処理は順番に行われ，相関分析モジュールによって相関ルールを導きルールの利益効率性を計算する．

この利益効率性 p は費用に繋がる属性を x，収益に繋がる属性を y としたとき，次のような計算で求める．

$$p = \frac{u_1y_1 + u_2y_2 + \cdots + u_ny_n}{v_1x_1 + v_2x_2 + \cdots + v_mx_m} \tag{1}$$

ここで u，v は各属性に与えられる重みである．この利益効率性は属性の単位が単一でないため計算が困難である．そこで，本論文ではオペレーションズリサーチの分

Building sales analysis system by data mining and OR

Motoyasu OKUTSU*, Tohgoroh MATSUI†,and Hayato OHWADA†

Department of Industrial Administration, Graduate school of Science and Technology, Tokyo University of Science*, Department of Industrial Administration, Faculty of Science and Technology, Tokyo University of Science†

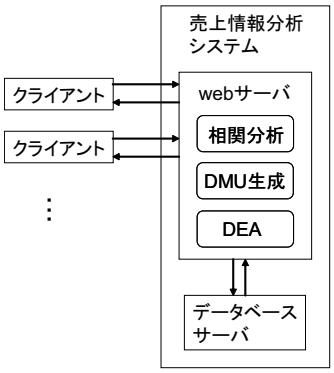

図 1　システム構成図

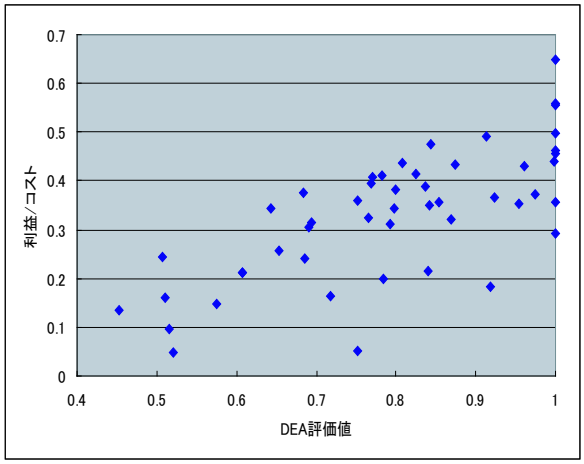

図 2　実験結果

野で用いられている Data Envelopment Analysis（以下，DEA と略す）[2] という手法を用いてこの計算を行う．

DEA では分析する対象を Decision Making Unit（以下，DMU と略す）と呼び，減少した方が良い属性集合（入力）と増加した方が良い属性集合（出力）のデータが含まれる．相関分析で得られたルールから DMU を生成し，DEA の計算を行いルールの評価を行う．

4　実装

本システムは導入の容易さと柔軟な拡張性を得るためにオープンソースの技術を用いて実装した．web サーバには Jakarta プロジェクトの Apach と Tomcat を用い，データベースサーバには MySQL を用いた．全てのモジュールは Servlet により実装されている．

相関分析については MUSASHI[3] を用いているため，Servlet とソケット通信を行うことによって MUSASHI を操作して相関分析を実行し，得られた結果を本システムで使用するプレーンテキストの形に整形した．DMU 生成モジュールはデータベースから必要な項目について検索・集計を行い結果をファイルに出力する．DEA 計算モジュールはシンプレックス法を用いた線形計画法ソルバーを実装して DEA の計算を行っている．各モジュールは結果を html 形式でブラウザに表示するようにした．

5　評価実験

本システムの検証を行うための実験を行った．MUSASHI プロジェクトから提供されているマーケットバスケット分析用のデータジェネレータを用いて商品・取引のデータを生成した．データジェネレータの設定は店舗数を 1 に制限する以外は初期設定のものを用いる．生成されたデータはトランザクション数 94864，商品数 646 となっている．生成された取引データを分析用と検証用に分け，分析用のデータに本手法を適用した．相関ルールの閾値は全てのルールを抽出するため 0 に指定し，相関分析によって得られたルールの上位 50 ルールについて DMU を生成し利益効率性の計算を行った．その結果を元に検証用のデータから得られる収益性について評価を行った．収益性は 利益/コスト の形で計算している．実験の結果は図 2 のようになった．

横軸がルールの利益効率性，縦軸が収益性である．利益効率性の高いルールが高い収益性を持っていれば良いことになるので，グラフでは右肩上がりの結果になる事が望ましい．

この結果から本システムで行った利益効率性の評価が高いものは高い収益性を持っていることが解る．

6　結論

本システムを構築することで売上情報を相関分析の手法を用いて分析し，得られた相関ルールを利益効率性により評価することが出来た．これにより企業にとってより効率的に高い収益をあげる商品の組合せを導くことが可能である．また，オープンソース環境を用いることで誰でも本システムを構築することが出来，柔軟な拡張を行うことが可能である．

参考文献

[1] R. Agrawal, R. Srikant. Fast Algorithms for Mining Association Rules, In Proc. of the 20th Int'l Conference on Very Large Databases (1994).

[2] A. Charnes, W. Cooper and E. Rhodes. Measuring efficiency using Data Envelopment Analysis, European Journal of Operational Research, 2:429-444(1978)

[3] MUSASHI Project, Mining Utilities and System Architecture for Scalable processing of HIstrical data, http://musashi.sourceforge.jp/