

データマイニングとORの手法を用いた売上分析システムの設計

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† 同 理工学部 経営工学科†

1 はじめに

今日、オンラインショッピングサイトでは、顧客が購入した商品の情報を分析することで、その購入動向がどのような特徴を持っているのかを見つける売上情報分析が盛んに行われている。その一つにアマゾン等に代表される相関ルールマイニングを用いた手法が挙げられる。これは、「商品 X を購入した顧客は商品 Y も購入する確率が高い」という相関ルールを調べ出し、顧客が購入した商品の関連性を分析することで、顧客が興味を持ていそうな分野の新商品を紹介したり、以前購入した商品の関連商品を宣伝するなどの用途で用いられている。しかし、この方法では商品同士の関連性を調べる事は可能であるが、実際の利益や投入と言った細かな売上分析まで掘り下げることは出来ない。

そこで、本研究では OR (Operations Research) の手法で事業体の相対評価を行う DEA (Data Envelopment Analysis) を用いて既存の相関ルールに基く特徴付けを元に、利益や投入まで考慮した売上分析の手法を提案する。DEA は元々、事業体の経営分析手法の一つで、最も優れた結果を出した事業体を元に効率的フロンティアを計測し、このフロンティアを基準として他の事業体の業績や効率を測定する方法論である。DEA についての詳細は参考文献の「DEA-経営効率分析法-」を参照して頂きたい。

2 DEA による売上分析

本研究では、相関ルールマイニングでは出来なかった関連する商品のより詳細な分析を DEA の手法と組み合わせることによって可能にする。DEA を使用する理由は以下の通りである。

- 項目数の多い商品の売上情報や商品情報でも扱える。
- 単位基準が異なり比較しにくいデータも DEA の手法では同時に扱うことが出来る。
- 顧客の購入動向を調べるにあたり、絶対評価ではなく相対評価を用いることで、売り方による差異を重視した分析結果を得ることが出来る。
- 顧客の特徴を加味した上で販売側が効率的に利益を上げる売り方を導くことが出来る。

DEA において、分析の対象となる各事業体を DMU (Decision Making Unit: 意思決定主体) と呼ぶ。今回 DMU として扱ったのは、顧客の商品購入履歴である。相関ルールと併用するにあたり、一度の購入で複数の商品を買う可能性があるデータを対象とする。ただし、トランザクション数が少なすぎる場合には、同時に購入した場合だけではなく一緒に購入したとみなす時間の範囲を設ける。

それでは次に入出力項目の抽出について具体的に説明する。入出力項目の決定には次のような事に注意する必要がある。

- 入力項目、出力項目とも数値データであり、原則としてすべての DMU において値は正。
- 比較するにあたり、入力対出力の効率性において、自分が得たい特徴を表している項目を選ぶ。
- ある出力を得る為の入力は値が小さい方が良く、逆にある入力から得られる出力は大きいほうが良いとする。もし逆になってしまう場合は逆数を取る事とする。
- 入出力項目の単位は任意に取ることが出来る。つまり、個数、金額、面積等である。

本研究ではこれらに留意して DEA 分析を行う。

DEA を用いるにあたり、分析対象の DMU にはある相関ルールで導かれた各レコートをを用いる。分析対象となる商品 X がある時、相関ルール $X \Rightarrow Y$ の商品 Y を関連商品とする。このルールに含まれる全ての取引が対象

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†

である．相関ルールについて分析対象と関連商品の二相関だけのルールを用いているのは，あまりに多次元の相関関係を用いると非常に小さな部分集合だけにマッチするルールになってしまうからである．これでは一般的な販売戦略に役立てることが出来ない可能性がある．また DMU が少なすぎると DEA の分析精度が落ちるという問題もあるので今回は二相関のみとした．

実際に DEA に用いる入出力項目は，入力項目として「分析対象の仕入値」，「関連商品の仕入値」，「分析対象の在庫保管費」，「関連商品の在庫保管費」の 4 項目．出力項目としては「分析対象の売値」，「分析対象と同時に売れた関連商品の売値」，「分析対象の販売個数」，「分析対象と同時に売れた関連商品の販売個数」の 4 項目である．項目に選んだ理由として，仕入値と売値，販売個数に関しては直接の取引に関する項目である為．在庫費用については，オンラインショッピングサイトでは通常販売店を持たないので，商品をストックする場所が必要になる．つまり，在庫費用も利益を効率的に生み出すためには考慮する必要があると考えたからである．これらの入出力項目を使って DEA の分析を行う．今回の DEA 処理に使用したプログラムは生田目崇氏による物である．

得られた結果から，参照集合として頻繁に現れる DMU はその商品の売り方に適していると言う事が出来る．これにより，販売戦略を決める足がかりとする事が出来る．さらに，非効率と判断された DMU に関しても，効率的フロンティアを参照することで，どうすればより効率的になるかという改善案を得る事も出来る．

3 分析システムの構成

本研究の分析システムの構成と流れを説明する．本研究で対象とするのはオンラインショッピングサイトの売上分析である．必要な情報（販売情報，顧客情報，商品情報，在庫情報）はネットワーク上に設置されたデータベースサーバに置かれているものとする．つまり，いつでも何処でも情報を取り出す事が可能である．また，売上分析をネットワーク上で行うことで，web サイトの管理と同時に使用でき，ネットワーク上の他のシステムとの連携も容易である．よって，分析システムも同じようにネットワーク上で動作する方が望ましいと考える．

実際の構成は図 3 のようになる．サーバクライアント形式で構成され，分析は 3 つの部分に分けて行う．一つ目は販売情報から相関ルールを導く部分．次に DEA ツールで使用する情報を集めて整形する部分．最後に前

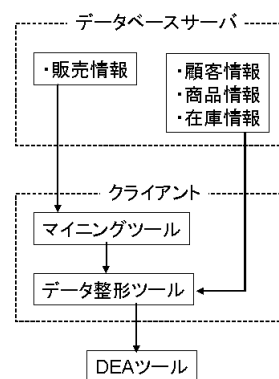


図 1 システム構成

述の DEA ツールによる分析である．

本手法で必要となる販売情報，商品情報，顧客情報，在庫情報はすべてデータベースサーバに置かれている．そして，相関ルールを出すためのツールは web ベースのクライアントとしてデータベースサーバにアクセスし，販売情報から分析対象のデータを検索し，相関ルールを導き出す．導かれた相関ルールを元に分析対象と関連商品についての情報を商品，顧客，在庫のデータベースから取り出し，DEA ツールにかけられるように整形する．今回使用する DEA ツールは Microsoft Excell で動作するので，所定のセルに数値を出力する．最後に Excell 上で計算を行い結果を出力する．

4 考察

DEA の手法を用いることで従来の相関ルールマイニングによる売上分析よりも詳細な分析が行える．この方法は，企業にとってより効率的な販売戦略を立てる助けになると考えられる．本手法で用いた DEA では DMU の数が増えれば増える程より精度の高い分析になるため，得られた分析結果を実際の販売に導入し，その販売情報を随時分析に取り入れる事が重要である．また，今回用いた DEA ツールでは入力，出力の項目数が 5 項目に制限されているため，より細かな項目を用いた分析を行うためには改善が必要である．さらに，Excell で動作しているため，全てのシステムを統合する必要もある．

参考文献

- [末吉 01] 末吉俊幸:DEA-経営効率分析法-(2001)
- [生田目他 03] 生田目崇 他:マーケティング・データ解析—Excel/Access による (2003)