

アクティブマイニングに基づく需要予測の戦略的販売促進への応用

Application of Demand Forecasting to Strategic Sales Promotion

矢田勝俊, 森本譲二

Katsutoshi Yada, George Morimoto

abstract

The purpose of this research is to present knowledge discovery process by applying some demand forecasting models to sales data in super market for constructing the strategic sales promotion. We shall provide a communication link between the marketing practioners and the active mining researchers to introduce the case of the consumer behavior analysis in the instant noodle market. It illustrates how the some models have been applied in practice of sales promotion planning and what advantages they have.

1. はじめに

近年の情報技術の目覚ましいイノベーションによって、多くのビジネスチャンスが生まれている。巨大なデータベースからの知識発見もそうしたビジネスチャンスの1つであり、チャンスを現実にするために様々な研究が行われている。アクティブマイニングは理論的にはもちろん、現実社会への貢献を同時に達成するために、データの収集やユーザーからの反応などを能動的にデータマイニングシステムに取り入れ、知識発見の効率化に寄与しようとするものである。

本稿では、需要予測問題に対するデータマイニング技術の適用の中で、ユーザーとのインタラクションを取り入れ、問題設定から実際に有用な知識発見、ビジネスアクションの創造までをトータルに説明することができるケースを紹介し、現実の知識発見プロセスの理解を深めることを目的としている。

本稿の構成は以下の通りである。まず従来の需要予測問題を扱っている研究をレビューする。次に本稿で扱うケースについて、分析の枠組み、プロセスを説明する。そしてルール抽出を行い、専門家による評価、解釈を検討する。最後に知識発

見プロセスに関する考察と今後の課題について述べる。

2. 従来の需要予測問題

需要予測に関する従来の研究は数多く存在する。近年では、実際のケース研究も報告されている。例えば、牛乳販売量の予測モデルについては、重回帰モデルを適用したもの(日経リサーチ 2000)やニューラルネットワークを用いたモデル構築を行った研究(日経リサーチ 2000, 豊田 2001)などが存在する。これらは比較的高い予測精度をもつモデル構築に成功しており、需要予測問題に対するデータマイニングの適用可能性を示している。

これら既存の需要予測に関する研究は、需要量の予測を目的とするものが多い。上記の研究では、あるスーパーにおける1日の牛乳の販売量を様々な属性から予測しようとするものである。したがって、モデルの評価基準は予測精度、つまり1日の販売数量をどれだけ正確に予測できたかである。

しかしこうした問題設定は、経営的観点から見ると必ずしも有用な示唆をもたらすとは限らない。基本的にビジネスでは、受動的なアクションよりもむしろ能動的なアクションが重要になる。近い将

来の需要量を正確に予測できたとしても、それをそのまま受け入れ、手をこまねく経営者はいないだろう。その需要量をさらに増加させるために様々な施策を実施し、利益の拡大を図るはずである。データマイニングの最終目的がビジネスに有用な知識を発見することならば、能動的なビジネスアクションを生み出すような知識を発見することが求められる。

従来の研究によく見られる問題点は、次の通りである。第一に需要の因果関係を明らかにするルールが明示化されていないために、能動的なビジネスアクションを創造することができない点である。通常の需要予測問題では、需要量の予測精度に焦点が置かれ、ルールの解釈可能性は比較的軽視されることが多い。ビジネスアクションの創造には現象の理解が不可欠であり、特に顕在化していない知識の理解が必要となる。

第二は、上述した詳細な現象の理解が不十分である点である。既存研究の多くは需要量に関する要因として、集約された POS データを用いることが多い。競合商品の販売数量や実売価格、特定地域の販売推移など、実際の POS 端末によって蓄積される元データを集約したものが用いられる。近年、多くの小売店では FSP カードといった顧客 ID を認識できるシステムの導入が進んでいる。既存研究の多くはそれらデータの特性を十分に反映したものにはなっていない（川田 2000）。過去の顧客の購買行為までモデルに反映させることができれば、精度の高い予測モデル構築が可能になるだけでなく、ビジネスアクションの創造を支援する知識発見が期待できる。

最後に、専門知識の導入ステップを明示的に扱っていない（矢田 2001）点である。既存研究では、属性の生成時点で何らかの専門知識の導入が行われたと見ることができるが、予測モデルの評価などを十分に専門家からヒアリングした形跡は

ない。多様な情報源の中からユーザーの目的や興味など価値観を反映した知識を発見し、新たな知識の発見の糸口を作り出そうとするアクティブマイニングの視点から考えれば、こうしたインタラクションがない知識発見には大きな限界があるといわざるを得ない。

本研究の立場は、ビジネスアクションの創造のために、予測精度よりもルールの明示化、解釈可能性を重視する。そして、そのルールがより現実の深い部分まで描写できるように、顧客個人レベルの行為も対象に入れた予測モデルの構築を目指す。そして、ビジネスアクションの創造に向けて、発見された知識を専門家に評価してもらい、インタラクティブな分析を行う。

3. 分析の枠組み

本研究で扱う分析ケース¹は、2001 年に実施された食品小売業チェーン A 社との共同研究の成果である。分析プロセスにそってケースの内容を紹介する。

3.1 分析目的の設定

分析対象を決定する前に、複数のメーカーに対して需要予測に関する分析ニーズについてヒアリングを実施したところ、新商品導入後の売上推移に関するニーズが高いことが明らかになった。そこで今回の分析対象としては、非常に多くの新商品が投入されるカップ麺市場を選択した。カップ麺とは、熱湯を注ぐだけで食べられる即席麺の中で、麺とかやくが容器に入っているものを指す。

通常の店舗では、約 100 種類のカップ麺が扱われているが、近年、その中に占める新商品の割合が 50% 近い店舗も珍しくなくなっている。ニーズの多様化と商品ライフサイクルの短期化がこ

¹ 本ケースは、2001 年 12 月、関西大学で行われたビジネスマイニングワークショップで発表された分析ケースをもとに作成した。

我々は分析目的を以下のように定めた。結果変数は新商品の需要量ではなく、その商品が発売後、長期間、売れつづけ定番商品になるのか、早期に棚から消えてしまう商品なのかに変更した。そして、それぞれの新商品の導入後1ヶ月間までのデータを用いて、継続的な販売に係る要因は何かを明らかにしようとした。今回の分析では、ルールが明示化され、実務家にも直感的に理解が容易である決定木を用いた。

3.2 対象の新商品

今回の対象となった新商品は 2000 年 8 月から

2001 年 3 月までに導入された新商品のうち、A 社の GMS 形態の店舗 R 店で取り扱いのある 86 商品を対象とした。上位 6 社で全体の 9 割を超えており、市場シェアは図 1 の通りである。カップ麺には、味噌、醤油、豚骨、塩などの味があり、通常、1 つのブランドにいくつかの味のバリ

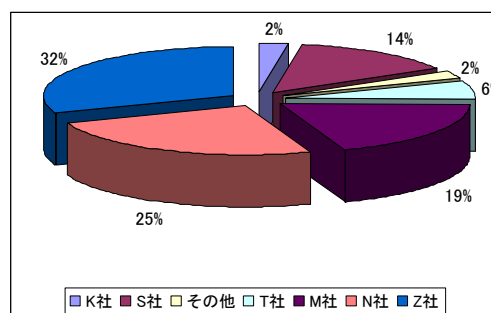


図1 カップ麺市場のメーカーシェア

エーションがある。対象となるデータのサイズは約 2.5GB、また、期間中に対象商品を購入した顧客数は 14945 人であった。

3.3 目的変数

対象商品の販売推移を図2で示している。横軸は対象商品それぞれの発売開始日からの期間、縦軸は週ごとの販売数量を表している。本研究においては、需要量を目的変数とせず、定番商品として生き残るかどうかを予測する。新商品の販売推移をクラスタリングすると、基本的に販売継続期間の長さによって4つの商品群に分類された。市場投入からの販売継続期間別の商品分布は、1ヶ

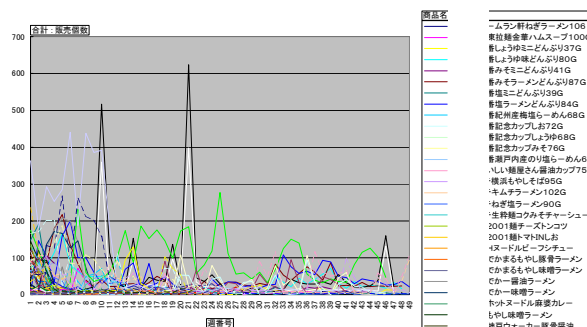


图2 商品別販売推移

月以内に消えていく商品が 19、2-4 ヶ月以内が 23、

5-8 ヶ月が 23、9 ヶ月以上継続して販売されている商品が 21 であった。今回の分析では、継続して販売される商品と短期間で売場から消える商品のルールを明らかにしたいため、5-8 ヶ月を除いた 3 つの商品群を目的変数と定義した。

3.4 説明変数

説明変数を作成する対象データは、発売前のデータと新商品発売後から 1 ヶ月以内までのデータを用いている。説明変数はその内容から 3 つに大別できる。

1)商品に関する定量レベルの属性

定量レベルの属性としては、発売後週ごとの販売数量や最大価格下げ幅、セール(通常売価よりも低い価格での販売)を行った週などを作成した。

2)商品に関する定性レベルの属性

定性レベルの属性としては、新商品のメーカー 6 社他を製造メーカーとして追加し、また新商品の味をもとに味噌、醤油、豚骨、塩、その他に分類した。

3)顧客レベルの属性

顧客レベルの属性としては、発売開始日から毎週のトライアル率、ヘビーユーザーの購入割合などを採用した。

3.5 トレーニングデータとテストデータ

上述のようにトレーニングデータの事例は、2000 年 8 月から 2001 年 3 月に発売された新商品 126 商品のうち、1 ヶ月以内に消えていく商品が 19、2-4 ヶ月以内が 23、9 ヶ月以上継続して販売されている商品が 21 の 63 商品をトレーニングデータとして採用した。

トレーニングデータから構築された予測モデルを検証するためのテストデータは、2001 年 4 月から 6 月までの期間に発売され上記の 3 つの分類に当てはまる 31 商品を採用した。内訳は 1 ヶ月以内に消えていく商品が 4、2-4 ヶ月以内が 16、9 ヶ月以上継続して販売されている商品が 11 であ

った。

表 1 は、これらの分析内容をまとめたものである。

分析目的		継続して売れ続ける新商品は、市場導入当初にどのような特徴をもつのか？
対象店舗		GMS形態の1店舗
対象データ	トレーニングデータ	2000年8月から2001年3月の期間に発売した新商品
	テストデータ	2001年4月から6月の期間に発売した新商品
目的変数		1ヶ月以内に消えていく商品、2-4ヶ月以内に消えていく商品、9ヶ月以上継続して販売されている商品
説明変数	定量レベルの属性	週別販売数量、最大価格下げ幅、セール実施週
	定性レベルの属性	メーカー名、味
	顧客レベルの属性	週別トライアル率、ヘビーユーザー購入割合

表1 分析の枠組み

4. 実験

ここでは、いくつかのデータマイニング技術を用い実際のルール発見を行う。抽出されたルールに対して専門家による評価、解釈を加え、考えられうるビジネスアクションを検討する。

4.1 複数のモデルの予測精度比較

上述の枠組みのもとで実験を行った結果が表 2 である。SPSS 社の Clementine6.0 を利用し、C5.0、ロジット回帰、ニューラルネットワーク、CART の 4 つの予測モデルを利用し、それぞれの精度をトレーニングデータ、テストデータで比較した。それぞれのモデルのパラメーターはデフォルト値を使用した。

(単位 %)		
	Training Data	Test Data
C5.0	58.73	41.94
Logit	92.06	45.16
NN	42.86	51.61
Cart	71.43	51.61

表2 予測モデルの比較

実験では、どのモデルもオーバーフィッティングが大きく、テストデータでは予測精度がかなり低下している。例えばロジット回帰は、トレーニングデータに対しては 92.06%と高い予測精度であったが、トレーニングデータに対しては 45.16%と低かった。ただし CART に関してはトレーニング

グデータ、テストデータともに比較的高い予測精度を維持しており、結果変数3分類に対してテストデータに対する予測精度が50%を超えていた。

4.2 専門家によるモデル、ルールの評価と再実験

実験結果を、メーカー、卸、小売店の担当者に見せ、それらの評価、検討を行った。事前のヒアリングで予測精度に関するニーズが低かったが、実際に上記のモデルのそれぞれの予測精度にはほとんど関心を示さなかった。

関心を示したのは、新商品の販売量と詳細な消費者行動との因果関係に関するもの、つまりモデルから抽出されたルールに大きな興味を示した。そこで、予測モデルの中で比較的ルールが明示的なC5.0、CARTのモデルのルールを詳細に検討し、意味解釈を行った。専門家との検討の結果、販売数量、メーカー名の属性がオーバーフィッティングを起こし、モデルの精度を落としていることが判明した。そこで、それらの属性を除き、再度、決定木でモデル構築を行った。図3は、C&RTによって再構築した需要予測モデルである。以下で抽出されたルールを考察してみよう。

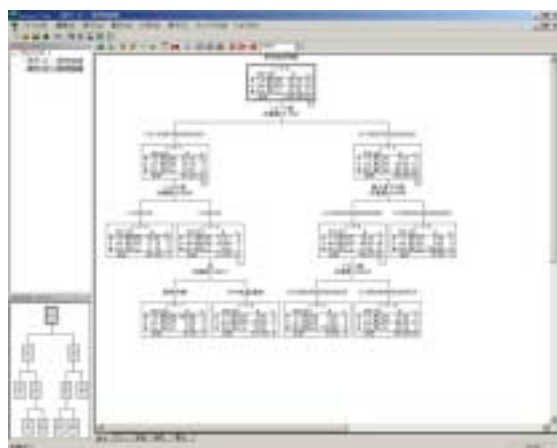


図3 C&RTによる予測モデル

全事例に対して最初の分岐ノードは、カップ麺のヘビーユーザーの全体に占める購入割合であり、ヘビーユーザーが発売開始後4週目の購入量全体

のうち、14%以下の場合、トライアル率に関わらず商品が店頭から消滅していく。そして新商品発売開始後4週目、ヘビーユーザーの購入割合が14%以上の場合、新商品が生き残る確率がかなり高くなり、次の分岐ノードでは最大値下げ幅が選ばれ通常売価の29%以上の値下げが一度でもあった場合、すべて9ヶ月以上継続して販売される商品になった。一方、最大値下げ幅が29%以下の場合でも、トライアル率が55%以下の場合、継続して販売される商品になっていることがわかる。トライアル率が低いということは、言い換えればリピート率が高いということであり、大幅な値下げがない場合でもリピート率が高い場合、継続して販売実績を作れる商品になる可能性が高いことを示している。

通常、新商品の市場導入問題において、最も重要な関心事の1つはプライシングの問題である。大幅な値下げは様々な顧客層のトライアルを引き起こすが、その後のブランド力にマイナスの影響をもたらしことが知られている。値引きされた価格が、購入顧客の参照価格としてその後の購買行為に影響を与える（上田 1999）ためである。しかし、実際の予測モデルではヘビーユーザーの購入割合が値下げの効果よりも将来の継続的な売れ行きに影響が大きいことが分かった。さらに大幅な値引きを実施しなくても、リピート率が高いときには、その新商品の生き残りに影響を与えることは専門家の関心を集めた。

4.3 ビジネスアクションの創造

グループディスカッションによって、これらのルールをもとにどのような販売促進戦略が考えられるかを検討した結果、いくつかのアイデアが提案された。まず新商品に関する販売促進の対象をヘビーユーザーに限定することが考えられた。多くの販売促進はすべての顧客を対象に実施されるため、コストパフォーマンスが非常に低い。将来

の売れ行きに最も影響を与えるヘビーユーザーに限定することによって、コストパフォーマンスの高い販売促進が可能になると考えられる。

具体的な販売促進の内容としては、ターゲット顧客（ヘビーユーザー）に対して3～4週目にポイント²サービス企画を行ったり、何度でも割引やポイントサービスが受けられる定期券サービスなどが考えられよう。これらの販売促進の目的は、ターゲット商品のリピート購入を促すことであり、上述のルールを念頭におくと、商品の定着に有効であると考えられる。

5. 考察と今後の課題

ケースから分かる通り、データの処理、データマイニングといった計算機を中心とした処理と、ルールの評価、解釈、ビジネスアクションの創造といった人間、組織が中心となる活動とが交互または並行に繰り返され、チャンス発見の研究で提起されている二重螺旋モデルのプロセスが観察される。このプロセスの中では分析者が中心になり専門家の評価や解釈とコンピュータのデータ処理を調整し、組織化している点で、二重螺旋モデル（大澤 2002）の変形した形になっているが、コンピュータと人間、組織が相互に関係し、様々な情報交換を行って有用な知識発見を達成するという点で本質的には同じプロセスと考えられる。

本研究では、これらのプロセスをモデル化するところまでは扱っていない。今後、二重螺旋モデルを購買履歴分析に適用した場合、一般モデルとどのような相違点が考えられるのか、実施上の留意点などを検討する必要がある。

需要予測問題としては、本稿の分析は需要量の予測問題を扱っていない。サプライチェーンの効

率化を進めるためには、本稿のような分析目的では十分ではないだろう。正確な需要量の予測を目的にした分析、モデル構築、その方法論が今後、重要になると考えられる。今後の課題としたい。

参考文献

川田英貴, 羽室行信, 加藤直樹, 矢田勝俊, 遺伝子解析技術を用いた購買履歴からのビジネスチャンスの発見, 2000 年度人工知能学会全国大会論文集, pp.155-156, 2000.

日経リサーチ, POS データに対するデータマイニング事例集, 日経リサーチ, 2000.

大澤幸生, 寺野隆雄, ヒューマン・システム・インタラクションに基づく知識評価と選択, 情報洪水時代におけるアクティブマイニングの実現研究成果報告書, pp.309-325, 2002.

豊田秀樹, 金鉱を掘り当てる統計学, 講談社, 2001.

上田隆穂, マーケティング価格戦略, 有斐閣, 1999.

矢田勝俊, 飯田洋, ビジネス知識発見の研究展望とアクティブマイニング, 第 46 回人工知能基礎論研究会資料, pp.213-217, 2001.

² ポイントとは、主に FSP プログラムなどで利用され、現金の代わりに商品券などと交換するための顧客に提供するサービス