

コンビニエンス・ストアにおける新商品発売時の データマイニングを用いた売れ行き予測

菅野 憲明*, 高山 毅*†, 池田哲夫*†

小売業においては、新商品の発売時には事前のマーケティング調査などにより売れ行き予測が行われている。しかし、事前調査の情報のみからの確に発売後の結果を予測するのは容易とはいえない。最近では売れ行き予測にデータマイニングを用いることが有力であるといわれている。スーパーマーケットでのデータマイニングを用いて新商品発売時の売れ行きを予測する手法の研究はあるが、コンビニエンスストアに関する学術的な分析は、現在までのところ充分になされていない。そこで本研究では、コンビニエンスストアにおいて、データマイニングを用いて新商品発売時の売れ行きを予測する手法を検討する。

Sales Forecast for New Items on Sale in Convenience Store Using Data Mining Techniques

Noriaki Kanno*, Tsuyoshi Takayama*†, and Tetsuo Ikeda*†

In retail trade, sales forecast is carried out using some marketing researches before putting a new item on sale. However, it is not always easy to forecast its sales precisely using only marketing researches. Recently, data mining techniques attract lot of attention in many commercial fields. There exist some researches concerning sales forecast for a new item in supermarket using data mining techniques. However, the research in the case of convenience store is insufficient. In this paper, we take an approach using data mining techniques to sales forecast for a new item in convenience store.

1. はじめに

小売業において売れ行き予測は重要である。小売業においては、新商品の発売時には事前のマーケティング調査などにより売れ行き予測が行われている。しかし、事前調査の情報のみからの確に

発売後の結果を予測するのは容易とはいえない。

売れ行きのデータが POS 等によってある程度電子化が進んでおり、売れ行きデータの集計が可能な場合、売れ行き予測にデータマイニングを用いることが有力であると最近では言われている。

コンビニエンス・ストア(以下「コンビニ」と略)は一般に売り場面積がスーパーマーケット(以降「スーパー」と略)よりも小さめであることが少なくなく、売れ行きが不調な商品を置く余裕がないことから、より精度の高い売れ行き予測が求められると予想される。そこで本研究では、コンビニ

* 岩手県立大学 大学院 ソフトウェア情報学研究科
Graduate School of Software and Information
Science, Iwate Prefectural University

† 岩手県立大学 ソフトウェア情報学部
Faculty of Software and Information Science,
Iwate Prefectural University

に着目し、その売れ行き予測の研究を行う。

コンビニでの新商品の売れ行き予測に関する学術的な分析は、現在までのところ充分になされていない[8]。比較的近い形態であるスーパーでは、データマイニングを用いて新商品発売時の売れ行きを予測する手法の研究[4]はあるが、「売り場面積が違う」、「ポイントカードによる顧客特定の普及度が違う」、「特価販売による販売促進の普及度が違う」などの要因により、スーパーでの手法、結果をそのまま適用することは出来ない。

以上より本研究では、コンビニにおいて、データマイニングを用いて新商品発売時の売れ行きを予測する手法を検討する。

本論文はこれ以降、次のように構成されている。まず 2 節では先行研究をまとめるとともに、コンビニ独自の手法の必要性について述べる。次に 3 節では本研究でのアプローチを述べる。そして 4 節では分析実験とその結果について述べる。最後に 5 節で結論と今後の展望を述べる。

2. 先行研究

2.1 売れ行き予測の現状

売れ行き予測の現状に関して、実践事例と研究事例を整理する。

まず実践事例であるが、文献[1]によれば、明治乳業とヴァル研究所の共同開発製品「客ごころ」では、気象データやイベント情報と過去 3 ヶ月分の来店数データを基に、重回帰分析を用いて来店数の予測をするが、何がいくつ売れるかまでは予測できない。文献[3]によれば、2002 年 7 月 11 日現在でも、来店数の予測に留まっている。

同じく文献[1]によれば、コンビニ大手のファミリーマートは ISDN を用いて全国 4000 の店舗に売れ筋分析データや天気予報等のデータを提供している。しかしながら、売れ行き予測については不完全であり、最終的な仕入れ量の判断は各店長に委ねている。

アイ・ビー・エム社の「INFOREM」は、可変

応答法と呼ばれる統計手法を用いて売れ行き予測を行い、在庫数量を考慮して最適な発注を自動的に行うシステムであるが[9]、旧来からの商品の売れ行き予測を行うもので、新商品の売れ行き予測には利用できない。

味の素株式会社では、トラッキング（追跡）モデルを用いて新商品の売れ行き予測を行っている[1]。しかしながらこれは、メーカー側から見た全国的な売れ行き予測であり、コンビニの店舗レベルでの商品単位の売れ行き予測を提供するものではない。

このように、現状では商品の売れ行き予測に特化した商用システムは、「コンビニの店舗レベルでの商品単位での新商品の売れ行き予測」という需要にこたえるものではない。

一方、学術サイドからの研究事例であるが、国立情報学研究所の論文検索システム NACSIS-IR や情報処理学会、電子情報通信学会、人工知能学会等の各ホームページ内の論文検索システムで検索しても、当該問題に関する研究事例は見当たらない。文献[8]での指摘の通り、コンピュータサイエンス領域からの学術的な検討が遅れている。

データマイニング以外の手法を用いた新商品の売れ行き予測に関しては、経済学での「トライアルリピートモデル」というモデルに基づく研究が代表的である[14]。ここでいう「トライアル」とは新規購入者、「リピート」は反復購入者のことである。どの程度の割合で新製品が試行的に購入されるか、その購入者のうちどの程度の割合が再度その新商品を購入するか、という分析から新商品の売れ行きを予測する。しかしながら、このモデルでは売れ行きの予測のために、6 週間程度の期間の売れ行きデータが必要になる[13]。

以上より、データマイニングを用いて約 6 週間より短い期間で売れ行きの予測ができれば、有意義といえる。

2.2 スーパーにおける予測事例

文献[4]ではスーパーにおける新商品の売れ行き予測に関して、C4.5[11]が採用しているものと同様のエントロピー最小化基準に基づくアルゴリズムで決定木を生成するという手法で分析を行っている。この事例では、データの抽出、事前の分析、決定木による分析という手順を踏んでいる。

データは「NEED-SCAN」POS データ[5]から抽出している。データの抽出期間は1997年10月から1999年11月までの104週間である。商品カテゴリとしては、スナック菓子を選択している。そして、1693商品のうち極めてローカルな商品、マイナーな商品、短期販売商品をふるい落とし328商品を抽出して使用している。得られた分析の結果、売れ行きのよい商品において約90%に共通する、

- 1) 発売後2週間目までに全国シェアを0.1%以上に拡大する
- 2) かつ、全国でのカバー率を10%以上に持つてくる

という法則を導き出すという成果を上げている。

しかし、これはメーカー側から見た全国規模の売れ行き予測であって、店舗単位での予測を提供するものではない。

2.3 コンビニ独自の手法の必要性

スーパーとコンビニには以下のような違いがある。

(1) 売り場面積

スーパーと比較してコンビニは一般に売り場面積が狭めのケースが少なくないので、売れないと判断された商品は早期に別の商品に置き換える必要がある。また、売り場面積が狭めということは、スーパーの場合よりもさらに、各商品の売れ行きを正確に予測して、適切な商品を適量仕入れる必要性が大きくなるといえる。

(2) ポイントカード

顧客を特定する方法にポイントカードがある。特定顧客の購入履歴は、売れ行きを予測する上で重要な情報である。スーパーではポイントカードの利用で顧客の特定が可能であるが、コンビニではポイントカードの利用がスーパーほど一般的ではないので顧客の特定が出来るとは限らない。

(3) 特価販売

スーパーでの売れ行きは、商品そのものの人気のほかに、特価販売の有無に多大な影響を受けるが、コンビニでは定価販売が主流であって、スーパーほど特価販売が一般的ではない。別の見方をすれば、スーパーの場合のように、特価販売によって売れ行きの意図的な制御を行うことが一般的にはなっていない。

以上のような相違点から、コンビニでの売れ行き予測はスーパーの場合よりも重要であるとともに、スーパーと異なったアプローチが必要である。

3. 本研究のアプローチ

3.1 分析方法

本論文では、発売直後の売れ行き動向とその後の売れ行き動向に関する相関関係の発見を目指す。

データマイニングの手法として代表的な決定木と記憶ベース推論を、以下の理由により採用する。決定木は、2.2節で説明したとおり、スーパーにおける売れ行き予測において、成果を上げている手法である。一方、記憶ベース推論は、新事例がデータベースに入ってきたのみで新しい分類を学習できるというメリットがあり[6]、次々と発売される新商品のデータに動的に対応しやすい。

文献[4]でも行われているように、データに前処理を施して、有用な結果が期待できそうな部分にデータを絞り、仮説を立てた上で、データマイニング手法を適用してそれを検証するという方法は、

組み合わせ爆発を避け、計算量を軽減するために有効である。そこで、本論文でも同様の手法をとる。

3.2 分析対象のデータ

3.2.1 データの概要

文献[8]中では「データ収集ボトルネック」として指摘されているが、データマイニングの研究を進めようとするときに、研究目的に合致する実データの入手は、必ずしも容易ではない。

コンビニにおける新商品の売れ行きを予測するという本研究の目的に合致し、入手可能なものとして、株式会社日本経済新聞社が有料で販売している「NEED-SCAN/CVS レシートデータ」を唯一、探すことができた。

これは、ある匿名のコンビニで販売されている商品の中から特定の商品を指定して、その商品が含まれているレシートの生データを期間3カ月以内で購入できるものである。たとえば、「チャーハンおむすび」、「山菜おむすび」といった商品指定が可能であり、おむすびでも具が違えば異なる商品として取り扱う。この購入価格であるが、大学等の研究・教育機関の場合で、

基本料金 10 万円、1 商品増やすごとに 1 万円追加 . . . (*)

であって、大学で購入する場合、その経済的負担は軽いとは言えない。一方で、現実と対応していない架空のデータでいくら分析をおこなっても、現実に利用可能な知見は得られない。また、実データを用いての分析も、それによってどの程度の知見が得られるかは、分析を実際に行ってみて初めてわかるところである。そこで、今回は予算額 30 万円以内で購入可能な範囲として、

おにぎり 10 商品、カップラーメン 10 商品 . . . (**)

についてレシートデータを購入し、どの程度の知見が得られるかに関する予備的な検討をおこなった。(*)より、この購入価格は税込みで 294,500 円

である。(**)を選んだ理由であるが、

- ・ 両者ともにコンビニにおける主力商品である [2]。
- ・ 賞味期限が比較的短めのものとは長めのもので挙動の差異を分析する。

の二点があげられる。

検討対象とした 2001 年 12 月 1 日から 2002 年 5 月 31 日の期間では、「NEED-SCAN/CVS レシートデータ」全体ではカップラーメンで 137 個、おにぎりで 46 個の新商品が投入されているが、(**)は売れ行きにおいてそれら全体の特徴に沿うように選んだ。

3.2.2 データの形式

生データはレシート単位のデータである。

データは CSV 形式で保存されており、データ項目は以下のとおりである。

店番号、レシート番号、YYYY, MM, DD, 曜日フラグ、休日フラグ、hh, mm, ss, 購入者性別フラグ、購入者年齢フラグ、購入商品名、JAN コード、単価、個数、値段、メーカー名、分類コード、分類名

3.2.3 データの加工

ここで以下の 2 ステップでレシートデータの加工を行う。

Step 1) 同時購入商品は考慮に入れないので、まずそれを削除する。

Step 2) 「発売からの日数」と「売上個数」へ加工する。

なお以降では、一日単位での販売個数を意味する言葉として「売上」を用い、時間的に幅を持った範囲での好不調を意味する「売れ行き」と区別して用いる。

3.3 事前の分析

パラメータの組み合わせの爆発を避け、有用な結果を期待できるパラメータの組み合わせの見通しを得ることを狙いとして事前の分析を行った。

ここでは決定木、MBRそれぞれの変数を決定するのに有用であった事前の分析結果を示す。

3.3.1 分析1：発売初日から60日までの売上個数の変化の分析

まず、発売初日から60日目までの売上個数の変化に着目して分析を行った。その結果、以下のような傾向を読み取ることが出来た。

- ・ 図1、図2のように、発売から30日後付近では発売初期の売上数と比較してほとんどの商品で売上数が半分以下になっている。
- ・ また、発売から60日後にはカップラーメンでは5商品、おにぎりでは7商品の売上数が0になっている。
- ・ カップラーメンで生き残っている5商品のうち4商品は、一日の売上個数が1~4個と低迷している。
- ・ 基本的に発売直後は売れ行きがよく、日数がたつにつれて売上個数が単調に減少していく傾向にある。

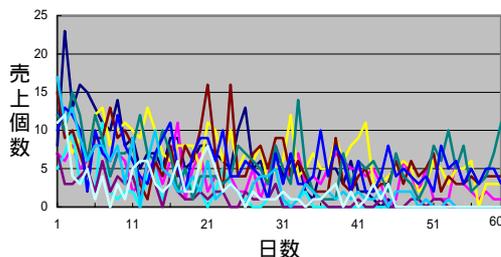


図1：カップラーメンの発売初日から60日目までの売上個数のグラフ。

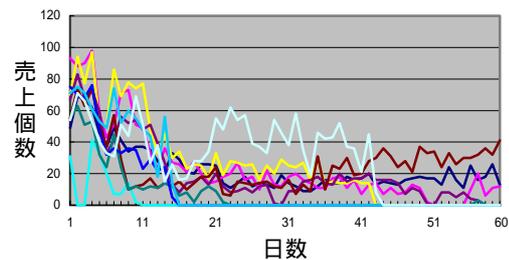


図2：おにぎりの発売初日から60日目までの売上個数のグラフ。

3.3.2 分析2：発売初日から10日目までの累計売上個数と発売27日目~33日目の累計売上個数の相関の分析

一日単位での売上を用いると数値の変動が大きいため、数値の変動を緩和するために、変数として累計売上個数を採用してみる。

発売初日から10日目までの累計売上個数と発売27日目~33日目の累計売上個数を比較してみる。図3、図4ではそれぞれ、「売れ行きが良いグループ」と「そうでないグループ」という2クラスがおおまかに見えてくる。

以上の分析から発売初日からの累計売上個数を中心に、発売30日目付近の売上個数が予測できるのではないかと推測される。

一般に連続値をカテゴリに分類する場合、その分類方法がデータマイニングをする上でベストか否かは断定できない。しかし、その分類方法をした上でデータマイニングをして有用な知見が得られ、その分類方法の周辺で試みたいいくつかの分類方法が、相対的に劣る精度や知見しか与えない場合には、その分類方法は少なくともベターな分類方法であったということはできる。ビジネスでは、ベストではなくてもベターなことがわかれば、何もわからないよりは良い、という場合が少なくなく、得られた知見は充分有益と言える[12]。

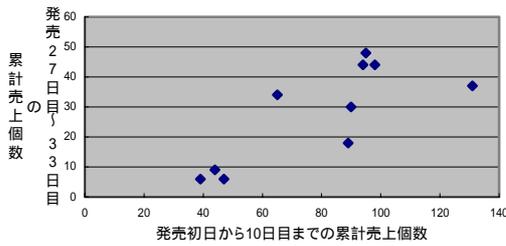


図3：カップラーメンの発売初日から10日目までの累計売上個数と発売27日目～33日目の累計売上個数の分布図。

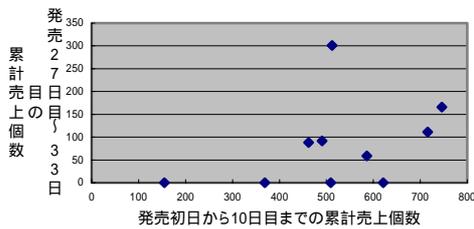


図4：おにぎりの発売初日から10日目までの累計売上個数と発売27日目～33日目の累計売上個数の分布図。

3.4 決定木による分析

決定木による分析では Ross Quinlan のホームページ[11]で公開されている C4.5 Release8 を使用する。

3.3.1 項, 3.3.2 項のケースにおいて, 意味のある結果が得られる可能性のあることがわかったパラメータを基に, 属性と属性値の決定をする。

3.5 記憶ベース推論による分析

3.5.1 分析の手順

データ数が少ないので以下の方法で分析を行う。

- 1) 10 商品のデータから 1 つを取り出し, 残りの 9 商品を記録セットとする。
- 2) 記録セットに対して取り出した 1 つのデータを与えて, その最近傍を求める。換言すると, 類似事例数を 1 とし, 売れ行きの好不調という二つのカテゴリで予測を行う。

3.5.2 推論の手順

説明変数にはカップラーメンでは「発売初日の売上個数」と「発売初日から 10 日目までの累計売上個数」の二つを, おにぎりでは「発売初日から 3 日目までの累計売上個数」と「発売初日から 10 日目までの累計売上個数」の二つを用いる。

目的変数は発売 27 日目～33 日目の 1 週間で売上が「好調」「不調」の 2 値と設定した。図 3, 4 を基に, カップラーメンでは 30 個以上を「好調」, 30 個未満を「不調」, おにぎりでは 1 個以上を「好調」, 1 個未満「不調」と分類した。

ItemA は取り出した 1 つ商品のデータ, ItemB は記録セットから取り出した商品のデータとする。ItemA と ItemB の距離 $d(A,B)$ は以下の式によって求める。

$$d_1(A,B) = | \text{ItemA の発売初日の売上個数} - \text{ItemB の発売初日の売上個数} |$$

$$d_{1-10}(A,B) = | \text{ItemA の発売初日～10日目の累計売上個数} - \text{ItemB の発売初日～10日目の累計売上個数} |$$

$$d(A,B) = \sqrt{d_1(A,B)^2 + d_{1-10}(A,B)^2}$$

$d(A,B)$ が最小の値をとったものを最近傍とする。最近傍が好調であったならば好調に, 不調であったら不調に分類する。

4. 実験

4.1 決定木による分析

4.1.1 カップラーメン

カップラーメンでは図 5 のような決定木が生成された。

図 3 を基に, 「不調」は発売 27 日目から 33 日目の累計売上個数が 10 個未満, 「やや好調」は 10 個以上, 40 個未満, 「好調」は 40 個以上と分類した。

それからは次のような傾向が読み取れる。

- (i) 発売初日から 10 日目までの累計売上個数が 50 個以上の製品は、発売 27 日目～33 日目の 1 週間に「やや好調」な売上個数を保っている。
- (ii) その中でも初日の売上が 10 個以上の商品は、さらに「好調」な売上個数を保っている。

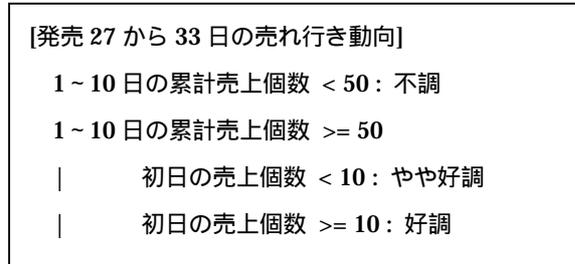


図 5：カップラーメンの決定木。

4.1.2 おにぎり

おにぎりでは図 6 のような決定木が生成された。

図 4 を基に、「不調」は発売 27 日目から 33 日目の累計売上個数が 10 個未満、「やや好調」は 10 個以上、100 個未満、「好調」は 100 個以上と分類した。

それからは、次のような傾向が読み取れる。

- (i) 発売初日から 3 日目までの累計売上個数が 180 個以上の商品は、発売 27 日目～33 日目の 1 週間も「やや好調」な売上個数を保っている。
- (ii) その中でも初日から 10 日目までの累計売上個数が 500 個以上の商品は、さらに「好調」な売上個数を保っている。

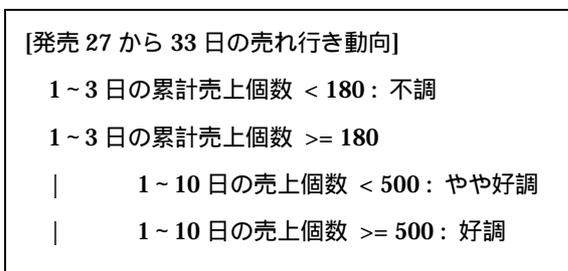


図 6：おにぎりの決定木。

4.2 記憶ベース推論による分析

記憶ベース推論による分析では、図 7、8 のようにカップラーメン、おにぎり共に発売 27 日目～

33 日目の 1 週間で売上が相対的に良い商品のグループと、そうでない商品グループの二つのクラスターに大分されることが解った。

しかし、おにぎりのうち 2 商品で明らかな不正解が出た。発売 27 日目～33 日目 1 週間で売上が 0 であるにもかかわらず、売上が良い商品のグループに分類されたのである。この 2 商品は 15 日目に急激に売上数を落とし、16 日目以降では売上数が 0 になっている。これは販売期間を 2 週間に絞った期間限定商品であったと考えられ、発売後 30 日近辺での売れ行きの動向を分析するには不適切なデータであると考えられる。

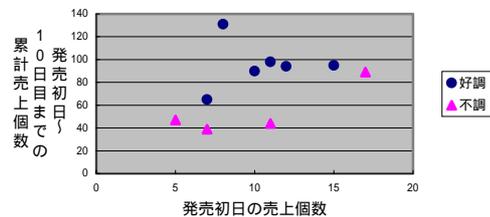


図 7：カップラーメンの分類

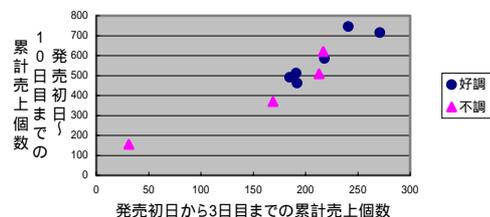


図 8：おにぎりの分類

4.3 分析結果の考察

決定木の利点として、データ量が少なくても決定木を生成できる[6][15]ことがあげられる。決定木でのカップラーメンのエラー率は 20%、おにぎりのエラー率は 30%であった。特筆すべきは、分岐の数が少ないにもかかわらず、比較的低いエラー率で新商品の売れ行き動向を分類できることである。ただし、前述の通り今回の評価はあくまで大量のデータを用いての本格的な分析を行う前の予備的な評価であって、エラー率として示してい

るのも学習データそのものの値であることを喚起しておく。通常の C4.5 の評価法に沿って、学習データとは別途の評価データを用いて決定木を評価した場合、今回記載しているエラー率を下限として、よりエラー率は高くなるものと推測される。

記憶ベース推論の方では、カップラーメンのエラー率は 20%、おにぎりのエラー率は 4.2 節で述べた不適切なデータを除けば 20%であった。

以上の結果より、データ量を増やして検証をするのに値する仮説が生成できたと考えられる。

5. 結論と今後の展望

コンビニにおける新商品の売れ行き予測を目的として、データマイニング手法として代表的な決定木と記憶ベース推論を用いて、新商品の発売直後の売れ行き動向とその後の売れ行き動向の相関の分析を行った。その結果、新商品発売時の売れ行きに関する仮説を生成できた。

今後は今回の分析結果を踏まえ、(i)データ量を増やして仮説を検証すること、(ii)「新商品投入後の日数」、「売上数」以外のパラメータも考慮していくことによって、より高精度な売れ行き予測を検討していくことが考えられる。

参考文献

- [1] 「売れ行き予測システム 天気で知る来店者数 在庫調整に威力発揮」、日経ビジネス, pp.66-68, 1996年9月9日号。
- [2] 「おにぎり」、「寿司」投入数と売れ行きの不均衡」、業界誌コンビニ, pp.29-31, 1999年6月。
- [3] 株式会社ヴァル研究所:「来店客数予測システム 客ごころ」, <http://www.val.co.jp/news/2002/0711.html>
- [4] 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」, 日経リサーチ販売有料資料(2000)。
- [5] 日経 QUICK 情報株式会社:「NEEDS POS 情報」, <http://www.nqi.co.jp/pos/scan.html>

[6] マイケル J・A・ベリー, ゴードン・リノフ著, SAS インスティテュートジャパン, 江原淳, 佐藤栄作共訳, :「データマイニング手法」, 海文堂(1999)。

[7] 松本和宏, 前田一穂, 柳沼義典, 新井洋一, 伊藤文信:「解説:金融ビジネスユースに適したデータマイニング手法: MBR ~融資申込み顧客の信用度の判定~」, 情報処理学会誌情報処理, Vol.42, No.7, pp.696-701, 2001年7月。

[8] 鷲尾隆:「ビジネスにおけるデータマイニングの現在・未来」, 情報処理学会誌情報処理, Vol.42, No.5, pp.467-471, 2001。

[9] 「IBM INFOREM のご紹介」: <http://www-6.ibm.com/jp/dist/s3-03.htm>, (2003年2月12日アクセス確認)。

[10] J.R.キンラン著, 古河訳:「AIによるデータ解析」, トッパン(1995)。

[11] Ross Quinlan - AI Group, CSE : <http://www.cse.unsw.edu.au/~quinlan/>

[12] Private Communications, エス・ピー・エス・エス株式会社プロフェッショナルサービスグループ・チーフスタティスシャン・荒 和志氏:情報処理学会第 65 回全国大会・チュートリアル(1) ビジネス分野でのデータマイニング「データマイニング, テキストマイニング入門」にて, '03-3-25, 於・東京工科大学。

[13] Private Communications, 明治学院大学経済学部教授・清水聰氏:情報処理学会第 65 回全国大会・チュートリアル(2) ビジネス分野でのデータマイニング「ビジネス分野での実際の事例と応用」にて, '03-3-25, 於・東京工科大学。

[14] 「マーケティングリサーチ用語 60」, <http://www.intage.co.jp/report/word60/121.html>

[15] 片岡浩巳, 小西修:「決定木の RDB 表現と知識探索支援システム」, 情報処理学会データベースシステム研究会研究報告, 125-79, pp.107-114, 2001。