

階層回帰モデルとベイズ推定による POS データの分析

An Analysis of Scanner Data using a Hierarchical-Bayes Regression Model

食品スーパーの戦略的マーチャダイジングを支援するために、戦略的な価格設定の基礎となる価格弾力性を分析した。交叉価格弾力性も考慮して分析するために、過学習を回避するのに有効なベイズ推定を用い、階層回帰モデルとベイズ推定によるプロトタイプシステムを開発した。日配品の週次データを分析したところ、有意な交叉価格弾力性は確認できなかった。日配品は価格変更が日単位で行われるため、週次データでは変動が顕著に現れなかったことが原因と考えられる。またサンプル外データに対する予測精度を検証したところ、最小二乗法と同程度であった。今回は分析精度の向上に有効な商圏情報が得られず、最小二乗法と同じ条件となったためと考えられる。今後は、今回のモデルを改良し、時系列構造解析の枠組みで日次データを分析することでこれらの課題を解決し、戦略的マーチャダイジングを支援するシステムを実現したい。

森川 浩司 Morikawa Koji
 捧 保浩 Sasage Yasuhiro

1. はじめに

食品スーパーにおける購買の約 6 割は非計画購買であると言われており¹⁾、店舗におけるマーチャダイジング(MD)は非常に重要である。商圏の需要を供給が上回るオーバーストアの下、買物客の購買単価は下落が続き、来店客数も伸び悩んでいる。このような経済状況においては、経験と勘に頼った従来の MD では効果は薄い。戦略から導きだされた MD をデータで検証し、戦略にフィードバックをかけ、MD の有効性を着実に高めていく仮説検証型の戦略的 MD が生き残りの鍵となっている。

しかしながら商品の価格設定をとってみてもそのような取組はまだ少ない²⁾。例えば価格弾力性に基づく価格設定を行う場合、価格弾力性は単品であればその商品の売上数量と販売価格のデータから容易に算出できる。しかし実際の売り場においては、ある商品の価格を下げると、その商品自身の売上数量は増えるが、同じ商品カテゴリーの他の商品の売上数量は減ることがある。したがって価格弾力性を分析する場合、交叉価格弾力性も含めた分析が必要となる。

交叉価格弾力性も考慮して商品カテゴリー全体について分析する場合、推定すべきパラメーターの数が単品の

場合と比べて大幅に増加する。また、商品の構成が分析対象期間内で変化しないことが前提となる。しかし日本においては商品の入れ替わりが激しいため、分析対象期間は、一般的に使用される最小二乗法が使えるほど十分長くないことがある。このような状況で最小二乗推定を行うとオーバーフィッティング(過学習)が起きることが知られている。この問題を回避できる推定法にベイズ推定がある³⁾。そこで今回はベイズ推定を用いて日配品の価格弾力性を分析した。また分析の精度を上げるため、階層回帰モデル⁴⁾を適用した。

2. 分析データの内容

25 店舗を展開する食品スーパーチェーンの POS データを用いて、普通牛乳 1L カテゴリーを分析した。分析対象期間は 2008 年度の 45 週間であり、分析単位は週である。このチェーン全体では約 20 種類の 1L 普通牛乳を扱っており、その中で売上上位 8 種類の商品を分析した。この 8 種類で売上数量、金額ともに普通牛乳 1L カテゴリー全体の 95%を占める。

25 店舗中のある 1 店舗ではこの 8 種類の牛乳を同時に揃えて売っている週が少なかったため、この店舗を分析

対象から除外した。またいくつかの店舗においては店舗改装により休業している期間があったためその期間を含む週は分析対象から除外した。したがって分析対象の週数は店舗によって異なる。ただし一番少ない店舗で 42 週である。

モデルの予測精度を検証するために、すべての店舗のデータについて後ろ 4 週間分のデータを予測精度検証用のホールドアウトサンプルとし、価格弾力性の推定には用いなかった。

各商品の販売価格帯と平均販売価格の関係を図 1 に示す。灰色の帯が販売価格帯、黒丸がその平均価格である。具体的な値は守秘義務に基づき明示していない。商品 2 の二商品は特売用商品の位置づけであり、交互に隔週で週末などの一日間に限って最安値で特売される。この二商品は隔週で特売されるため、特売されない週はほとんど店頭に並ばず、そのため売上数量は激減する。そこでこの特売用の二商品はまとめてひとつの特売品として扱うことにした。この特売品の価格と売上数量は各商品が特売される週のデータを採用した。この特売二商品は普通牛乳と成分調整牛乳であるが、普通牛乳のほうが成分調整牛乳よりも販売数量が多いという傾向がある。

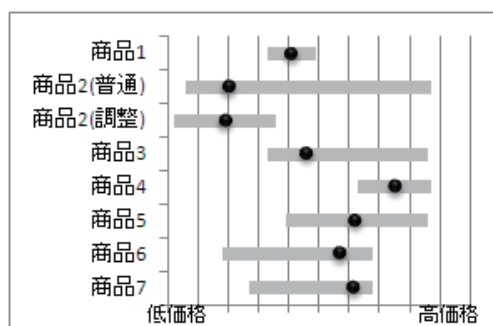


図 1 商品ごとの販売価格帯とその平均値

3. 分析モデルと分析手法

3.1 店舗内モデル

表 1 棚 MD の 5W2H

What	どのメーカーの商品か
Which	どんなフレーバーの商品か
Where	棚のどの位置に置くか
How many	どれくらいのフェース数で揃えるか
How much	いくらか
When	どれくらいの期間並べるか
Why	どんな「買う理由」を訴求するか

店舗の棚に着目したとき、MD の要素として表 1 の内容が考えられる。今回は利用可能なデータの観点から、商品の売上数量を価格とプロモーションで説明する回帰式を設定する。

店舗 s ($s = 1, \dots, S = 24$) における、商品 i ($i = 1, \dots, M = 7$) の、週 t ($t = 1, \dots, T_s$) における売上数量を q_{sit} とする。また店舗 s における、商品 i の、週 t における価格を p_{sit} 、週来店客数を N_{st} 、特売用商品のうち普通牛乳が特売になるときに 1 となり成分調整牛乳が特売になるときに 0 となるフラグを f_{st} とすると、回帰式は以下ようになる。

$$(1) \quad \ln(q_{sit}) = \alpha_{si} + \sum_{j=1}^M \eta_{sij} \ln(p_{sjt}) + \nu_{si} \ln(N_{st}) + \delta_{si} f_{st} + \varepsilon_{sit}; \quad \varepsilon_{sit} \sim N(0, \sigma_{si}^2)$$

ここで $N(\mu, \sigma^2)$ は平均 μ 、分散 σ^2 の 1 次元正規分布を表す。右辺第 1 項の切片はベースライン、すなわちその商品 (ブランド) の販売力を表す。右辺第 2 項の係数は価格弾力性を表す。 η において $i = j$ の項が自己価格弾力性であり、 $i \neq j$ の項が交叉価格弾力性である。右辺第 3 項の係数は来店客数弾力性を表し、これはチラシによる集客効果に相当する。右辺第 4 項は普通牛乳が特売される時の各商品への影響を表す。残差 ε_{sit} は買物客の購買行動の不確定要素を表す。残差 ε_{sit} は時間軸方向には独立であるが商品間では相関があるものとしている。これは、式 (1) でモデル化できていない要因による影響はこの商品の残差にも共通して現れることになり、商品の残差間には相関が生じると考えられるからである。

3.2 店舗間階層モデル

式 (1) の推定すべきパラメーターを $\beta_s^T = (\alpha_s^T, \eta_s^T, \nu_s^T, \delta_s^T)$ としてベクトル β_s の k 番目の成分を β_{sk} とするとき、 β_s の値は店舗平均値の周りに分布しており、その偏差は店舗の特徴 (店舗の商圈情報) の違いで説明できるとする回帰式を設定する。

$$(2) \quad \beta_{sk} = \bar{\beta}_k + c_s^T \gamma_k + u_{sk}; \quad u_s \sim N_{M(M+3)}(0, V_\beta)$$

ここで右辺第 1 項は $\bar{\beta}_k$ の店舗平均値である。右辺第 2 項のベクトル c_s は店舗の特徴を表すベクトルである。今回は商圈情報が入手できなかったため店舗で扱っている普通牛乳の種類数を代わりに使用した。種類数は 14 前後でありこの値は店舗間でほとんど差がない。ベクトル c_s はその店舗平均値が 0 になるように調整されてい

る。また係数 γ_k は店舗内モデルの各項に対応するように「切片 (ベースライン)」「自己価格弾力性」「交叉価格弾力性」「来店客数弾力性 (集客効果)」「普通牛乳特売の影響度」の 5 種類とする。 u_s は残差である。 $\theta^T = \{\beta^T, \gamma^T\}$ としたときの以上の関係を図 2 に示す。

以上のように、各店舗について推定すべき係数は店舗独立には推定せず、店舗平均値の周りに分布していて、その偏差は店舗の特徴の違いによって説明されるという二段構成 (階層構造) になっている。式(1)のベイズ推定における事前分布として式(2)を使用することから、このモデルは階層回帰モデルと呼ばれている。

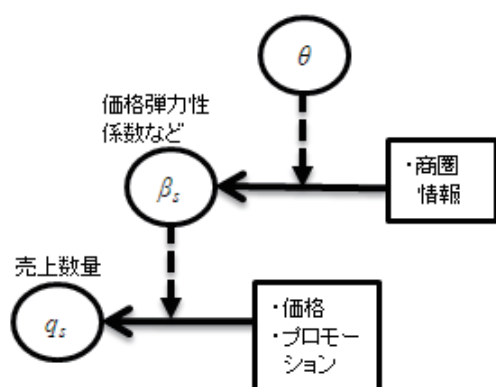


図 2 階層モデルにおける回帰の階層関係

3.3 分析手法

推定すべきパラメータは $\{\beta_s\}$, $\{\sigma_{si}^2\}$, θ , V_θ であり、これらに事前分布を設定してベイズ推定を行う。 β_s の事前分布は式(2)であり、 σ_{si}^2 , V_θ に逆ウィシャート分布、 θ に正規分布を仮定すると、完全条件付き事後分布も同じ分布族となる。完全条件付き事後分布は解析的に得られるのでギブスサンプラーによって事後分布からのサンプリングが可能となる。

サンプリング回数を 10000 回とし、最初の 2000 回は初期値依存期間として廃棄した。定常分布に収束していることや定常分布からランダムサンプリングできていることは標本経路のグラフとコレログラムを使って確認した。収束の判定に一般的に使用される Geweke の方法などを適用して数値的に一律に判定することはしなかった。ある店舗における商品 1 の自己価格弾力性係数の標本経路図を図 3 に示す。またそのコレログラムを図 4 に示す。

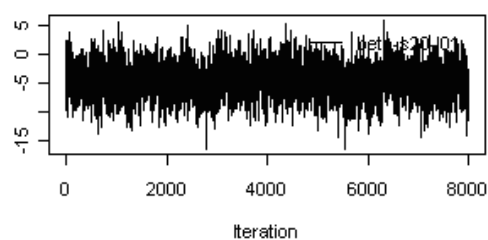


図 3 商品 1 の自己価格弾力性係数の標本経路

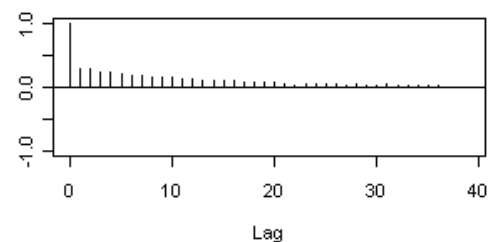


図 4 商品 1 の自己価格弾力性係数のコレログラム

4. 分析結果とモデル評価

4.1 分析結果

価格弾力性係数の、各店舗のサンプル平均値を店舗で平均した店舗平均値を表 2 に示す。一般に自己価格弾力性係数の符号は負、交叉価格弾力性係数の符号は正となる。

表 2 価格弾力性係数の店舗平均値

	商品1の	商品2の	商品3の	商品4の	商品5の	商品6の	商品7の
商品1に対する	-4.9	0.3	0.6	0.8	0.1	0.3	0.1
商品2に対する	1.9	-1.2	0.5	0.8	0.2	-0.2	-0.1
商品3に対する	1.6	0.4	-8.0	-7.8	3.2	-1.1	0.2
商品4に対する	1.0	-0.4	0.5	-1.9	0.5	0.1	-0.4
商品5に対する	2.9	1.2	0.7	5.3	-10.5	-0.1	-0.1
商品6に対する	2.5	2.4	0.5	-6.8	1.9	-4.5	0.1
商品7に対する	1.2	1.6	0.3	2.9	0.4	0.4	-3.7

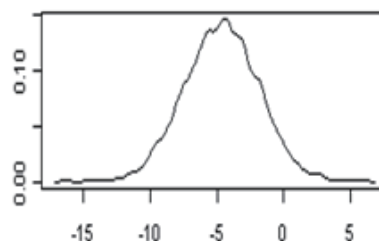


図 5 商品 1 の自己価格弾力性係数の確率密度分布

価格弾力性係数の確率密度分布の例として商品 1 の自己価格弾力性係数の確率密度分布を図 5 に示す。

自己価格弾力性係数については、商品 2 および商品 4 を除いて図 5 の商品 1 と同様に負である確率が高いとい

う結果が得られた。商品 2 および商品 4 の自己価格弾力性の確率密度分布は図 6 の分布と同様に 0 を挟んで正負の両側に大きく分布しており、負である確率が高いとは言えない結果となった。商品 2 の価格はほぼ特売時の価格であること、また商品 4 は値引きがほとんど行われないことから、ともに価格があまり変動していないことが原因と考えられる。

交叉価格弾力性については、商品 3 に対する商品 4 の係数が -7.8、商品 6 に対する商品 4 の係数が -6.8 と大きな負の値となった。この結果を検証するために、売上数量がもっとも多い店舗について交叉価格弾力性係数の確率密度分布を調べた。商品 3 に対する商品 4 の交叉価格弾力性係数の確率密度分布を図 6 に示す。図 6 にみられるように確率密度分布は 0 をまたいで正負の両側に大きく分布している。したがって平均値は大きな負の値となっているが、負である確率が高いとは言えない。商品 6 に対する商品 4 の交叉価格弾力性係数も同様であった。その他の交叉価格弾力性係数の確率密度分布も同様であり、正、あるいは逆に負となる確率が高いと言える係数はなかった。

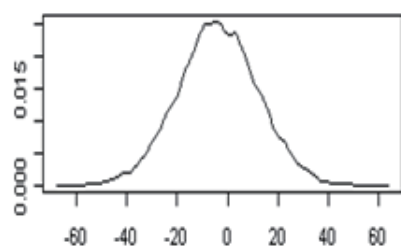


図 6 商品 3 に対する商品 4 の交叉価格弾力性係数の確率密度分布

4.2 交叉価格弾力性のないモデル

前節の結果を受けて、交叉価格弾力性の項の有無による差異を評価するため、式(1)において交叉価格弾力性の項のみを削除したモデルで同じ分析を行った。

係数の推定値は交叉価格弾力性ありのモデルとほぼ同じ結果となった。自己価格弾力性係数の、各店舗のサンプル平均値を店舗で平均した店舗平均値を表 3 に示す。

表 3 自己価格弾力性係数の店舗平均値

	商品1の	商品2の	商品3の	商品4の	商品5の	商品6の	商品7の
商品1に対する	-4.9						
商品2に対する		-0.7					
商品3に対する			-7.0				
商品4に対する				-2.2			
商品5に対する					-9.5		
商品6に対する						-5.4	
商品7に対する							-3.1

商品 1 の自己価格弾力性係数の確率密度分布を図 7 に示す。自己価格弾力性係数は交叉価格弾力性のあるモデルと同様に、商品 2 および商品 4 を除いて負である確率が高いという結果が得られた。

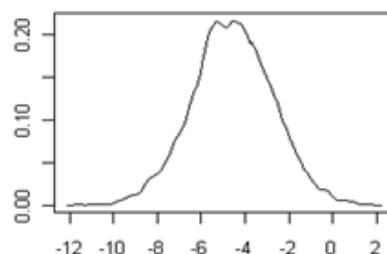


図 7 商品 1 の自己価格弾力性係数の確率密度分布

交叉価格弾力性ありのモデルと交叉価格弾力性なしのモデルについて、モデルの良さを評価する情報量基準 DIC (Deviance Information Criterion)の値と、ホールドアウトサンプルにおける売上数量の予測値の sMAPE (symmetric Mean Absolute Percentage Error, 一商品、一週あたりの調整平均絶対誤差率)の値を表 4 に示す。DIC, sMAPE とともに、値が小さいほうが良いモデルと言える。DIC の値からは交叉価格弾力性なしのモデルが支持される。一方、sMAPE の値からは交叉価格弾力性ありのモデルのほうが予測誤差は小さくなっている。

表 4 DIC 値と sMAPE 値の比較

	交叉弾力性あり	交叉弾力性なし
DIC	17029	11414
sMAPE	18%	20%

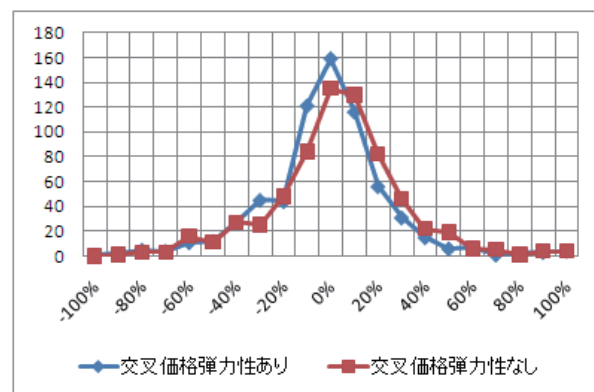


図 8 ホールドアウトサンプルにおける誤差率のヒストグラム

ホールドアウトサンプルにおける全商品の、売上数量の予測値の、誤差率のヒストグラムを図 8 に示す。青色

が交叉価格弾力性ありのモデル、赤色が交叉価格弾力性なしのモデルである。パラメーターの数が多い交叉価格弾力性ありのモデルのほうが sMAPE の値が小さくなっているのは、交叉価格弾力性係数が調整弁のような役割を果たしているためと考えられる。なお、一般的に使用される最小二乗推定による sMAPE の値は約 20% となった。したがって今回のデータについては、階層回帰モデルのベイズ推定による予測の精度は店舗独立の最小二乗推定と同程度であると言える。

4.3 分析のまとめ

今回の分析では、有意な交叉価格弾力性は確認できなかった。一般に牛乳の価格は日単位で変更される。今回は週単位で分析しているため、日単位の価格変動とそれに伴う売上数量の変動が平準化されて、価格変動に伴う売上数量の変化が見かけ上小さくなったことが原因と考えられる。もうひとつの原因としては、特売品と高級ブランド品のように購買層が異なるために交叉価格弾力性が 0 となっている可能性も考えられる。

自己価格弾力性については、商品 2 と商品 4 以外の商品については、有意な値を推定することができた。商品 2 と商品 4 については推定することができなかったが、価格があまり変動していないことが原因と考えられるため、妥当な結果であると言える。

予測精度については、階層回帰モデルのベイズ推定と店舗独立の最小二乗推定は同程度という結果となった。一般的には、階層化により店舗の特徴の違いを考慮することで予測精度が向上することが期待される。しかし今回は商圈情報が得られず、また代わりに用いた牛乳の種類数は店舗間であまり差がなく、結果として推定すべき係数が店舗平均値の周りにランダムに分布しているというモデルになったことが原因と考えられる。

5. モデリングの課題と今後の方針

分析結果に対する考察から、食品スーパーにおける商品の価格弾力性を分析するためには、日次データに対する分析が必須であると言える。日次データを分析するためには曜日効果などの考慮が不可欠である。その場合は、今回の重回帰分析のような静的な分析よりも、モデルの状態空間表現とカルマンフィルタなどを使用した時系列構造解析の枠組みによる動的な分析が有効であると考えられる。

今回は同一チェーンの複数店舗の POS データという

条件であったため、店舗間の共通性と商圈の違いによる差異を仮定して階層化した。時系列構造解析の枠組みで分析する場合、今回のような店舗をまたがる階層化の有効性は不明である。ベイズ推定の枠組みにおいては、階層化は事前分布としてどのような条件を設定するかの問題である。棚割も売上数量を左右する重要な MD 要素であることを考えると、店舗の共通性を設定せずに代わりに店舗独立に棚割情報を活用する方法も考えられる⁵⁾。

売上数量の価格弾力性を推定するためには価格の変動に伴う売上数量の変動があることが前提となる。今回の特売品や高級ブランド品のように価格の変動が少ない商品については価格弾力性を推定できない。高級ブランド品は分析対象から除外することも考えられるが、特売品の場合はその価格での存在自体が重要である。したがって、ある価格で特売品が存在するという情報をモデルの要素として表現できるようにする必要がある。

今回、商品の店舗内プロモーション情報は入手できたが、商品によっては常に何らかのセールを実施しているというデータとなっており、有効な情報とはならなかった。また、分析精度の観点からは、店舗外でのプロモーション情報であるチラシの情報も重要であるが、これは入手できなかった。分析精度の向上のためには、これらプロモーション情報の正確な把握とデータ化が不可欠である。

6. 戦略的 MD の支援に向けて

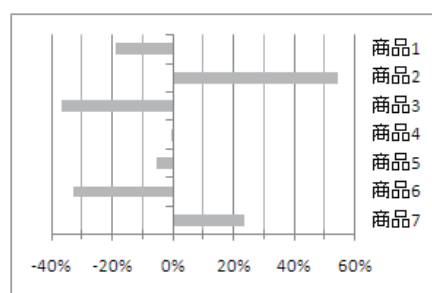


図 9 普通牛乳が特売されるとき他の普通牛乳の売上数量に対する影響度

交叉価格弾力性なしのモデルにおいて、普通牛乳が特売されるとき他の普通牛乳の売上数量に対する影響度を図 9 に示す。特売により商品 1 や商品 3、商品 6 の売上数量が 20% 以上減少することが分かる。商品 4 と商品 5 が影響をほとんど受けていないのは、これらの商品が高級ブランド品のため、購買層が異なっていることが原

因と考えられる。また特売により商品 7 の売上数量が増加しているのは、特売される商品 2 と同一のメーカーの商品であるため、値下げのタイミングが連動していることが原因と考えられる。

特売を実施する場合は、特売する商品による売上、利益だけでなく、他の商品に対する影響についても定量的に分析して販売価格や仕入れコストを決定することが必要不可欠である。

棚割すなわち棚のどの位置に何フェースで商品を揃えるかも重要な MD 要素である⁶⁾。今回は棚割に関する分析は行っていないが、店舗の MD 支援に向けて、表 1 の MD 要素を変更した場合に商品カテゴリ全体の売上数量、売上金額、利益がどのように変化するかをシミュレーションできるモデルを実現していきたい。

7. おわりに

定量的な分析に基づいて戦略的に販売価格を決定する際には交叉価格弾力性も考慮した価格弾力性の分析が必須となる。交叉価格弾力性も考慮して商品カテゴリ全体について分析する場合、推定すべきパラメーターの数が単品の場合と比べて大幅に増加する。また、商品の構成が分析対象期間内で変化しないことが前提となる。しかし日本においては商品の入れ替わりが激しいため、分析対象期間は、一般的に使用される最小二乗法が使えるほど十分長くないことがある。このような状況で最小二乗推定を行うとオーバーフィッティング（過学習）が起きることが知られている。この問題を回避できる推定法にベイズ推定がある。今回は階層回帰モデルとベイズ推定に基づいたプロトタイプシステムを開発して、日配品の価格弾力性を分析した。

分析の結果、情報量基準 DIC からは交叉価格弾力性のあるモデルよりも交叉価格弾力性のないモデルのほうが支持される結果となった。一般に牛乳の価格は日単位で変更されるが、今回の分析は週単位であるため、日単位の価格変動とそれに伴う売上数量の変動が平準化されて、価格変動に伴う売上数量の変化が見かけ上小さくなったことが原因と考えられる。

予測精度については、階層回帰モデルのベイズ推定と店舗独立の最小二乗推定は同程度という結果となった。一般的には、階層構造化により店舗の特徴の違いを考慮することで予測精度が向上することが期待される。しかし今回は商圈情報が得られず、また代わりに用いた牛乳の種類数は店舗間であまり差がなく、結果として推定す

べき係数が店舗平均値の周りにランダムに分布しているというモデルになったことが原因と考えられる。

今回はマーケティングの要素のうち価格に焦点を当てて分析を行った。今後は価格に加えて棚割、店舗内プロモーション、店舗外プロモーション（チラシなど）、在庫量、補充コストなどの情報も総合的に分析できるようにモデルの改良を進めて、戦略的マーチャンダイジングを支援するシステムを実現したい。

参考文献

- 1) 平久保：消費者行動論，ダイヤモンド社，(2005)
- 2) 株式会社 日経BP：全日食 儲かる価格教えます，日経ビジネス，2010.3.15号，pp.32-33，(2010)
- 3) 照井：ベイズモデリングによるマーケティング分析，東京電機大学出版局，(2008)
- 4) Alan L. Montgomery: Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data, Marketing Science, Vol. 16, Issue 4, (1997)
- 5) Erjen van Nierop et al.: Interaction between Shelf layout and Marketing effectiveness and its impact on optimizing shelf arrangements, Marketing Science, Vol. 27, No. 6, (2008)
- 6) Pierre Chandon et al.: Does In-Store Marketing Work? Effects of the Number and Position of Shelf Facings on Brand Attention and Evaluation at the Point of Purchase, Journal of Marketing, (2009)



森川 浩司 2001年入社
新規事業開発室
ビジネス・アナリティクス分野における
新規事業の創出
molikawa@hitachi-to.co.jp



捧 保浩 1988年入社
事業企画開発本部
ビジネス・アナリティクス分野のコン
サルテーション
sasage@hitachi-to.co.jp