

ブランドの価格属性を考慮したシークエンシャルパターンによるブランドスイッチ予測

西口 真央, 森田 裕之

1. はじめに

自社ブランドにロイヤルティを持つ顧客が、大量に他社ブランドにスイッチすることは、そのブランドを展開しているメーカーにとって大きな問題である。特に機能的側面の差別化が比較的困難な日常用品の商品分類を取り扱うブランドは、どのような要因が消費者のブランド選択に影響を及ぼすのか、また競合ブランドの価格などの変化に、自社ブランドにロイヤルティを持つ顧客がどの程度反応するのかを把握することは、マーケティング戦略を立案する上で重要である。ブランドスイッチの要因には様々なものが考えられるが、小売店舗での価格プロモーションによるブランドスイッチの分析は、日々のPOSデータからも限定的には可能である。

本研究では、あるドラッグストアのID付POSデータ¹を利用し、シャンプー類や柔軟剤類などの日常用品の特定のブランドにロイヤルティを持つ顧客のブランドスイッチ行動を予測するモデルを提案する。分類モデルのベースとなるのは、CAESP (Classification by Aggregating Emerging Sequential Patterns) [1]であるが、入力するデータに工夫を加えることで、購買時の商品価格および、競合ブランド商品との価格差を考慮することが可能な方法を提案する。計算実験の結果から、分析対象商品に対する顧客のブランドスイッチの予測、および出現したシークエンシャルパターンの考察を示す。

2. 分析対象データと基礎分析

使用するデータは、ドラッグストアの同一チェーン4店舗における口中衛生類、ヘアケア類、衣料用洗剤

類のID付POSデータである。データ期間は2008年1月1日から2009年12月31日までの2年間であり、データ的全レコード数は約87.2万件、購買実績のある顧客数は約6.8万人である。基礎分析の結果、ブランドスイッチの頻度やパターンは、商品分類内のブランド数に影響を受けていた。そこで本研究では、多くのブランドが競合しているシャンプー類、3ブランドが約80%の売上を占める柔軟剤類、そして2ブランドの複占状態であるライト系洗剤類の3つの商品分類を分析対象とした(図1)。

次に商品分類内における対象ブランドとその競合ブランドを決定するため、ブランド別売上金額と、各ブランド間の併買状況をJaccard係数によって確認する。競合ブランドは、本来複数のブランドが存在する可能性があるが、本稿では、全体の関係性を表現する最初の段階の研究として、対象ブランドと競合ブランドが1つずつ存在する場合の表現方法を提案する。

図2の左側はシャンプー類の上位5ブランド、右側は柔軟剤類の上位3ブランドの関係を表している。ノードの大きさは売上金額の大きさを相対的に、またエッジ上のラベルはJaccard係数の大きさを表している。シャンプー類ではラックス、TSUBAKIの順に売上が

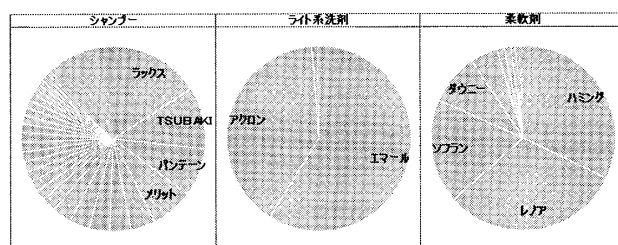


図1 売上金額構成比率

にしぐち まお, もりた ひろゆき
大阪府立大学 大学院経済学研究科
〒599-8531 堺市中区学園町1-1
受付 11.7.25 採択 11.11.5

¹ 平成22年度データ解析コンペティションにおいて、経営科学系研究部会連合協議会とカスタマー・コミュニケーションズ株式会社より提供されたデータ。

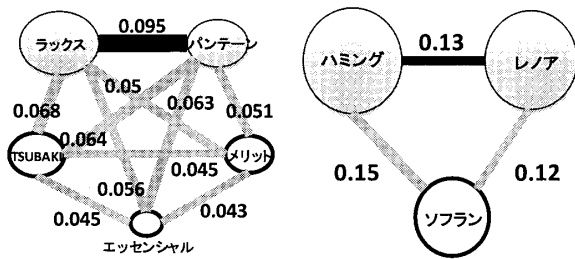


図2 上位ブランド関係図

高いが、Jaccard 係数をみるとラックスと TSUBAKI よりもラックスとパンテーンの方が若干大きい。柔軟剤類ではハミング、レノアが売上上位 2 ブランドであり、Jaccard 係数はどれも大差ない。ライト系洗剤類は前述のように、複占状態にあるため選択の必要はない。以上のような結果と考察から、各分類における対象ブランドと競合ブランドとして、シャンプー類ではラックスとパンテーン、柔軟剤類ではレノアとハミング、そしてライト系洗剤類ではエマールとアクロンを選択することにする。

3. 関連研究

文献[2]によると、ブランドの価値概念であるブランドロイヤルティとは当該ブランドを反復・継続して購入することを指し、現在および将来のキャッシュ・フローの安定的かつ確実な増加をもたらす要因の一つであるとしている。

ブランドスイッチに関する研究は、例えば Givon [3]の研究では、ブランドを消費することから得られる効用だけでなく、ブランドスイッチをすること自体からの効用も要素に取り込んだブランド選択モデルを構築している。また、文献[4]では消費者がブランドを確率的に選択することを前提としたモデルを提案している。文献[4]でも指摘されているように、従来のブランド選択モデルは、購買行動モデルの各プロセスにおける影響を可算的に考えるモデルと、各プロセスにおける満足閾値の連結条件を考えるモデルとに分類できる。これらのモデルは、選択のフレームワークとして妥当なものだと考えられるが、日用品に関しては、消費者があるブランドを選択して購買する際、そのブランドの商品の価格は影響を与えるものと思われる。また、それだけではなく、対象ブランドとそのブランドに競合するブランドの商品の値引き状態、そして 2 つのブランドの商品間の価格差も消費者の選択に影響を及ぼすのではないかと考えられる。例えば、あ

る消費者が、通常、ブランド X の商品を購入していたとする。あるとき店舗で、X と競合するブランド Y の商品が値引きされており、しかも X と比べると割安感を感じたため Y の購買を決定したとしよう。その場合、X の商品と Y の商品が単独で値引きされたかどうかだけではなく、それらの間の価格差も影響していると考えられる。本研究では、このような消費者の購買決定の状況を 1 つのアイテムとして表現することで、単にどのブランドが購買されたというだけではなく、どのような状況で購買が決定されているかということを表現する方法を提案する。

時系列データにおいてアイテムの発生順序を考慮したパタンのことを指すシークエンシャルパターン[6]を利用した既存研究としては、CAEP (Classification by Aggregating Emerging Patterns) [5]を応用した CAESP が存在する。これはトランザクションデータとして顧客の購買順序を記録したアイテムの購買履歴と、その顧客の所属するクラスを定義したクラスデータを与えれば、各クラスに顕著に出現するシークエンシャルパターンである顕在シークエンシャルパターンを利用して、顧客のクラスを予測する分類モデルである。本研究では分類モデルは CAESP を用いるが、アイテムとしてどのブランドを購入したかという情報だけでなく、購買時の対象ブランドおよびその競合ブランドの価格帯と値引きの状態、そしてブランド間の価格差を表現し、分類モデルの入力データとして利用することを提案する。結果として、どのようなシークエンシャルパターンが出現してブランドスイッチの要因になるのかを明らかにするとともに、出現するシークエンシャルパターンの解釈から各ブランドに対する相対的な価格プロモーションの方法について考察する。

4. 提案手法の説明

本節では、本研究でベースとなる CAESP の考え方を説明するとともに、これを利用した提案手法について説明する。

4.1 CAESP

CAEP は、頻出パターンをもとにした顕在パターン (Emerging Patterns) を利用する分類モデルであり、CAESP はシークエンシャルパターンをもとにした顕在シークエンシャルパターン (Emerging Sequential Patterns) を利用する分類モデルである。つまり、CAEP をシークエンシャルパターンに対応するように改良した手法が、CAESP であるといえる。

以下で説明する顕在パターンは、対象がシーケンシャルパターンであれば顕在シーケンシャルパターンとなる。

いま2つの異なるクラス¹, 2に属するレコードのデータベース D_1, D_2 が存在し、データベース D_1 に出現するパターン x のサポートは式(1)で表されるとする。ここで $\text{cnt}_{D_1}(x)$ は、データベース D_1 のパターン x を含むレコード件数を示し、 $|D_1|$ はデータベース D_1 の総レコード件数を示している。

$$\text{sup}_{D_1}(x) = \frac{\text{cnt}_{D_1}(x)}{|D_1|} \quad (1)$$

また、 x の D_2 に対する D_1 の増加率 (Growth Rate) は式(2)で示される。

$$\text{GR}_{D_1}(x) = \begin{cases} \frac{\text{sup}_{D_1}(x)}{\text{sup}_{D_2}(x)} & (\text{sup}_{D_2}(x) > 0), \\ \infty & (\text{sup}_{D_2}(x) = 0), \end{cases} \quad (2)$$

ここで最小サポート $\sigma (0 \leq \sigma \leq 1)$ と最小増加率 $\rho (\rho > 1)$ を与えたとき、 $\text{sup}_{D_1}(x) \geq \sigma$ かつ $\text{GR}_{D_1}(x) \geq \rho$ を満たすパターン x を、クラス1の顕在パターンと呼ぶ。

顕在パターンを利用した予測モデルである CAEP は、顕在パターンを計算した後、式(3)によって定義されるスコアを算出し、各メディアンで基準化したスコアを算出する。以下本稿では、簡単のためスコアと呼ぶことにする

$$\text{score}(d, c) = \sum_{x \in d, x \in E_c} \frac{\text{GR}_{D_c}(x)}{\text{GR}_{D_c}(x) + 1} \cdot \text{sup}_{D_c}(x) \quad (3)$$

ここで、 d はあるレコード、 c はクラス、 x は顕在パターン、 E_c はクラス c の顕在パターン集合、 D_c はクラス c に属するレコードのデータベースである。

$\text{GR}_{D_c}(x) / (\text{GR}_{D_c}(x) + 1)$ はパターン x のクラス c における事後確率 pb であり、この値は、クラス間のサポート比である増加率 $\text{GR}_{D_c}(x)$ の範囲を、 $1 \sim \infty$ から $0.5 \sim 1$ に変換したものである。この pb とサポートは、より大きな値であることが望ましいので、2つの積をスコアとして定めている。このスコアは、予測するレコードがどのクラスに分類されるかを決定する指標となる。CAEP は、予測するレコードに対して、各クラスのスコアを計算し、最大スコアを持つクラスを予測クラスとする方法である。Dong らの論文[5]では、上記の顕在パターンは頻出パターンを利用して分類モデル

を作成しており、それをシーケンシャルパターンに拡張した分類手法が CAESP である。

顕在シーケンシャルパターンの抽出においては、最小サポートと最小事後確率の他に、アイテムのギャップ長の上限とウィンドウサイズの制約を与えることが可能である。ギャップ長は、時系列データ中の2アイテム間に存在するアイテム数を指し、ウィンドウサイズは、シーケンシャルパターンを列挙する際に、考慮すべきトランザクションデータ中の連続したアイテム数の上限を表す。例えば、アイテム a, b, c が $\{c, a, a, b\}$ という順序で出現するトランザクションデータにおいて、ギャップ長とウィンドウサイズが無制限のとき、 $\{c, b\}$ というシーケンシャルパターンが存在する可能性があるが、レコード中の c と b の間には a が2つ存在するため、ギャップ長を1にすれば、 $\{c, b\}$ というシーケンシャルパターンは出現しない。また、このトランザクションデータでは、 c から b まで、のべ4つのアイテムが存在している。そのため、ウィンドウサイズが3であれば、 c からは、2回目出現する a までのトランザクションデータ中のシーケンシャルパターンを列挙するので、 $\{c, b\}$ というシーケンシャルパターンは、列挙されないことになる。このように、ギャップ長とウィンドウサイズを与えることで、アイテム間の購買間隔が、列挙されるシーケンシャルパターンおよび、それによる計算結果に影響を及ぼす。

また、頻出パターンとシーケンシャルパターンの違いは、大きく2点存在する。1つはアイテムの出現順序を考慮するかどうかという点であり、もう1つはアイテムの繰り返し出現を考慮するかどうかという点である。いま2つのアイテム a, b がトランザクションデータに存在するとする。このとき、トランザクション上の出現順序を考慮したシーケンシャルパターンでは、トランザクションデータに存在する a と b の順序によって、 $a \rightarrow b, b \rightarrow a$ の2つのシーケンシャルパターンが列挙される可能性がある。一方、頻出パターンでは発生順序は考慮されないため、トランザクションデータ上の出現順序によらず a, b のみが列挙される。このとき、出現順序を考慮することに意味があるならば、シーケンシャルパターンを列挙するほうがよいであろうし、出現順序を考慮することに意味がなければ、頻出パターンを列挙すれば十分である。またアイテム a がトランザクションデータ上、複数回出現しても、頻出パターンは、アイテム a が出現したかどうかだけがわかるが、シーケンシャルパターンでは、何度出現したか

¹ クラスは3以上でも適用可能であるが、ここでは簡単のため、2クラスの場合に限り説明する。

も表現することができる。しかし、これも繰り返し出現を考慮することに意味があるならば、シーケンシャルパターンを列挙する必要があるだろうし、出現の有無だけが重要であるならば、頻出パターンを列挙するだけで十分といえる。

後述の計算実験で CAESP と CAEP の 2 つの手法による実験を行い、その比較結果を示す。今回の計算実験は IGVcaesp[1] と IGVcaep[1] を使用した結果であり、前者は CAESP を、後者は CAEP を実装したプログラムである。

4.2 提案手法の概要

提案手法は、図 3 のような流れで行われる。図中の 1) の対象データの選択においては、a) 商品分類とその分類内の対象ブランドおよび競合ブランドの選択、b) 分析対象顧客の選択、c) モデル作成のために与える購買回数の設定、d) 各レコードの所属クラスの定義を決定する必要がある。a) は 2 節で選択したとおりであり、b) は特定のブランドにロイヤルティを持つ顧客を分析対象顧客として選択する。c) と d)、および今回選択された顧客数などについては 5.1 節で説明することにする。

次に、2) のアイテムの作成プロセスが、本研究での工夫点である。通常、購買履歴からパターンを抽出する際、商品名が利用されることが多い。しかしあるブランドの商品を購入する場合、日用品では、必ずしも商品の価格は一定ではない。また、競合ブランド商品の販売価格によっても、その商品を選択した意味は異なる。例えば、X ブランドの商品 a (100 円) と Y ブランドの商品 b (80 円) が存在し、a の購買を選択した場合と、a と b が値引きによってそれぞれ 60 円と 70 円のとときに a の購買を選択した場合では、必ずしも a の選択動機が同じとはいえない。このような状況をアイテムで表現できれば、競合ブランドとの相対的な状態も考慮した価格属性に基づく、商品購買選択を

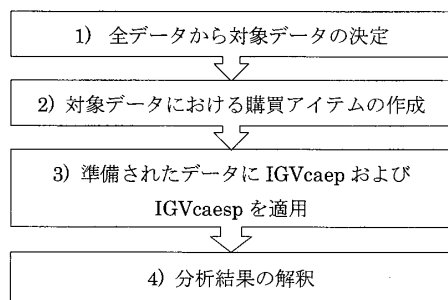


図 3 提案手法の流れ

明らかにすることができ、またそれを利用して列挙されたどのシーケンシャルパターンが、ブランドの継続やスイッチに影響を及ぼしているのかを示せるのではないかと考えた。

しかし、商品価格やそれらの間の価格差は通常整数値であるので、そのままの値を使ったのではアイテムが多様となり過ぎ、逆に有効なパターンの列挙を阻害する可能性がある。そこでこれらを 5 節で具体的に説明するように、適切に離散化して利用している。

4.3 販売価格と販売時点価格差の計算

対象ブランドや対象顧客などの必要な入力データが与えられたとき、1 つのブランドの中には数個~10 数個の異なる商品が提供データには存在していた。今回使用したデータには明確な商品名などは提供されていなかったため、対象データ中の各商品の販売価格の中の最大値を各定価と考えることにした。複数種類の商品コードの存在理由は、パッケージなどの商品の本質的な機能とはあまり関係ない要因が一因であった。そこでブランドを展開するメーカーのサイトからサイズなどによる種類数を確認し、それと同数になるように前述の定価を用いてクラスタリングを行い、商品を集約する。各ブランドには、数個の商品クラス c_i^b (b はあるブランド、 $i=1, \dots, k^b$ (k^b はブランド b のクラス数)) が存在する。また商品の値引き状態は、各商品で大きく異なるので、定価から販売価格を引いた額を昇順に並べて、件数が 4 つに均等分割されるように値引き区分を定める。

以上によって、対象ブランドのある商品とその販売額が与えられれば、その商品が所属する商品クラスと値引きの状態が特定される。例えば $b=1, i=1$ でそのクラスに所属している商品 3 を 125 円で販売したデータで、商品 3 の定価は 200 円、値引き区分は値引き 1: 5% 未満、値引き 2: 5% 以上 10% 未満、値引き 3: 10% 以上 15% 未満、値引き 4: 15% 以上の区分であったとする。この場合商品 3 は c_1^1 に所属し値引きは 4 となるので、「 c_1^1 : 値引き 4」 と表現される。

次に対象ブランドと競合ブランドとの商品の価格差を計算するために、競合ブランドの商品価格帯を特定する必要がある。選択方法はいくつか考えられるが、ここではデータから最も併買される可能性の高い商品を選択することにする。図 4 は一例として、ライト系洗剤類の価格帯クラスと参照関係のイメージである。エマールが $b=1$ 、アクロンが $b=2$ であり、 $k^1=k^2$

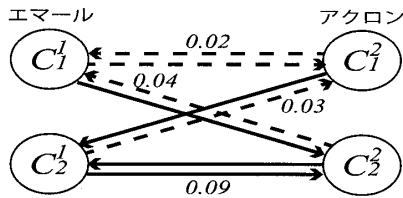


図4 価格帯と参照関係のイメージ

=2である。ノードは1つの価格帯クラスタを表し、有向枝のラベルは、トランザクションデータから計算した各価格帯クラスタ間の併買状況を Jaccard 係数で表したものである。各価格帯クラスタからは、競合ブランドの全ての価格帯クラスタに有向枝が存在しているが、その中で最大の Jaccard 係数を持つ有向枝を実線で、それ以外を破線で表し、実線で結ばれた価格帯が参照関係となる。このケースでは c_1^1 からは c_2^2 が参照関係となるが、逆に、 c_2^2 からは c_1^1 が参照関係にあるクラスタとなっている。参照関係にあるクラスタの中の競合価格の基準は、 c_1^1 の商品を購入したならば、 c_2^2 の中で、同日同店舗での最安値の販売価格とする。そして参照関係にある対象ブランドの価格から、競合ブランドの価格を引いた値を価格差とする。価格差の離散化の方法は、価格差が正の値（対象ブランドの価格が大きい）の値の集合を価格差の降順に並べて件数で2等分し、大きい方のカテゴリを“ \gg ”，小さい方のカテゴリを“ $>$ ”と標記する。価格差が負の値（対象ブランドの価格が小さい）も同様に2等分して、大きい方のカテゴリを“ $<$ ”，小さい方のカテゴリを“ \ll ”と標記する。また価格が等しい場合は、価格差なしを意味する“ $=$ ”の5つの状態で表現している。価格差の記号は、対象ブランドと、競合ブランドの間に配置して標記する。例えば、ある顧客が「 c_1^1 : 値引き4」の商品を125円で購買したとき、競合価格として選択された商品が「 c_2^2 : 値引き3」で、その日の価格は70円であったとする。このとき競合価格差は55円となり、50円以上の値引きは“ \gg ”で表現されたとすると、「 c_1^1 : 値引き4 \gg c_2^2 : 値引き3」と標記される。

以上のように、アイテムは最終的に、購買された商品のブランドとその価格帯、そして購買時の各ブランドの値引き状態および購買された商品のブランドと競合関係にあるブランドの類似商品との間の価格差を表現し、そのシークエンシャルパターンを列挙してモデルを作成する。

表1 対象データの定義

商品分類	分析期間	クラス決定期間	クラス定義	
			スイッチ	ロイヤル
シャンプー	2~11 回目	12,13 回目	対象ブランド 購買が1回以下	対象ブランド のみ購買
柔軟剤	2~11 回目	12~17 回目	対象ブランド 購買が3回以下	対象ブランド のみ購買
ライト系洗剤	2~6 回目	7~9 回目	対象ブランド 購買が2回以下	対象ブランド のみ購買

5. 計算実験

5.1 データ選択方法

ロイヤルティやブランドスイッチの定義は様々であるが、本研究では以下のようにそれぞれを定めることにする。ロイヤルティの定義は、ある1回以上の連続する購買機会において、 m 回以上（ただし $m \geq 1$ ）、対象ブランドを購買している状態と考える。すると、 m は大きいほうが、より安定したロイヤルティを持つ顧客集合が選択できると考えられる。逆に問題を難しく設定するには、多様なブランドロイヤルティを持つ顧客が混在する状態であると考え、そのためには、 m は小さく設定すべきであると考えた。そこで本実験では、初回の購買が対象ブランドを購買した顧客を分析対象とすることにした。対象データのその他の定義は表1のようにになっている。またブランドスイッチとは、クラス決定期間において、一定回数以下しか対象ブランドを購買しない状況と定義し、今回の実験では2購買機会あたり1回以下しか購買しない状態とすることにした。したがってクラス決定期間が2購買機会であれば1回以下、4購買機会であれば2回以下のようにすることとした。クラス決定期間において、ブランドスイッチをしたと判断された顧客はスイッチクラスに、また対象ブランドしか購買しない顧客は、ロイヤルクラスに属していると考えられるものとする。分析期間はトランザクションデータの顧客の購買回数の分布から判断した。購買回数を多く設定するほど、より多様なシークエンシャルパターンを得ることができるが、それだけ対象人数が減少してしまう。そこで両方の観点を総合的に考え、表1のように設定している。クラス決定期間も購買回数で判断し、クラス定義を保ちつつ2つのクラスの人数の偏りを抑えることができるよう期間を設定している。また、各クラスの対象顧客の

表2 各クラスの数

商品分類	所属クラス		合計人数
	ロイヤル	スイッチ	
シャンプー	159人	113人	272人
柔軟剤	141人	90人	231人
ライト系洗剤	93人	32人	125人

表3 計算時の各分類のパラメータ設定

商品分類	最小	最小	ギャップ	ウィンドウ
	サポート	事後確率	長の上限	サイズ
シャンプー	0.005	0.6	2	5
柔軟剤	0.005	0.7	2	5
ライト系洗剤	0.02	0.6	1	3

人数は表2のようになっている。

5.2 計算結果

前述の3つの商品分類に対してデータを作成後、IGVcaespによって計算を行った。計算時の各分類のパラメータ設定は表3に示す。

なおウィンドウサイズは、一年間の購買行動で予測をすることを目標にし、各商品分類での対象顧客の年間平均購買回数に設定している。ギャップ長の上限はその半分である半年以内の購買であれば、同じパターンに出現可能となるよう設定した。この値も平均購買回数に基づき設定している。また前述のブランド内の価格帯クラスタの区分や、参照関係、および値引き区分については、紙幅の関係から、ライト系洗剤類に限定して明示すると、それぞれ表4、表5のようになっている。

表4は、各ブランドの価格帯クラスタとその価格帯クラスタ間の参照関係を示している。例えばエマールの商品のうち、定価が[299, 313]の範囲に含まれる商品は価格帯クラスタ1に所属し、顧客がこの価格帯の商品を購入したなら、アクロンの定価が[217, 342]の範囲に含まれる商品のうち、同日同店舗での最安値の商品の販売価格を競合価格とすることを意味している。表5は、価格差と値引き率の離散化基準の一例を示している。価格差や値引き率の離散化の基準は商品ごとに異なり、全ての商品の離散化基準を掲載することは困難なため、例としてライト系洗剤類のエマールの価

表4 価格帯クラスタの定義と参照関係

ライト系洗剤			
エマール	参照関係	アクロン	クラスタ
定価(円)		定価(円)	番号
[299, 313]	←	[342, 427]	1
[265, 299]	←	[217, 342]	2

表5 価格差・値引き率離散化基準例

価格差(円)	表現記号	対象ブランド値引き率(%)	カテゴリ
38以上	»	45.5以上	4
0以上38未満	>	37以上45.5未満	3
-33以上0未満	<	29以上37未満	2
-33未満	«	29未満	1

格帯1に属する1つの商品の基準値を掲載している。価格差と値引き率は、全ての商品で4つの区分に離散化される。表5の例では、商品が購買されたとき、競合価格との差が38円以上であるなら価格差区分は4、そのときの対象ブランドの値引き率が29%未満なら値引き区分は1と表現される。競合ブランドの値引率も同様の手順で離散化を行う。以上のようにしてアイテム化し、設定したパラメータで計算した結果の評価値は、表6に示されている。表6では、提案手法の結果の有効性を検証するため、同じ入力データをCAEPによって予測した結果（以下の表ではCAEPと記述する）、購買ブランド情報のみをアイテム化してCAESPを適用した結果（以下の表では価格情報なしと記述する）、そして購買ブランドとその価格帯クラスタ情報のみをアイテム化してCAESPを適用し予測した結果（以下の表では価格帯情報ありと記述する）と比較している。また正答率は、全体のレコードに対するロイヤルクラスおよびスイッチクラスの予測された正解の割合だが、再現率、精度、F1値はスイッチクラスの予測をベースに計算した値である。そして、すべての結果は5回の交差検証によるものである。

全体的な結果として、“価格情報なし”と“価格帯情報あり”は、相対的にもっとも悪い結果を示していることがわかる。一部で“価格情報なし”の再現率が1.000になっている結果があるが、これはすべてをスイッチクラスと予測してしまったことによる結果であ

表6 計算結果の比較評価

	分析手法	正答率	再現率	精度	F1 値
シャンプー	提案手法	0.761	0.796	0.682	0.735
	価格帯情報あり	0.537	0.726	0.463	0.566
	価格情報なし	0.522	0.779	0.456	0.575
	CAEP	0.702	0.850	0.600	0.703
柔軟剤	提案手法	0.835	0.833	0.765	0.798
	価格帯情報あり	0.719	0.956	0.585	0.726
	価格情報なし	0.389	1.000	0.390	0.561
	CAEP	0.831	0.744	0.807	0.775
ライト系洗剤	提案手法	0.696	0.812	0.448	0.578
	価格帯情報あり	0.416	0.969	0.301	0.459
	価格情報なし	0.256	1.000	0.256	0.408
	CAEP	0.696	0.750	0.444	0.558

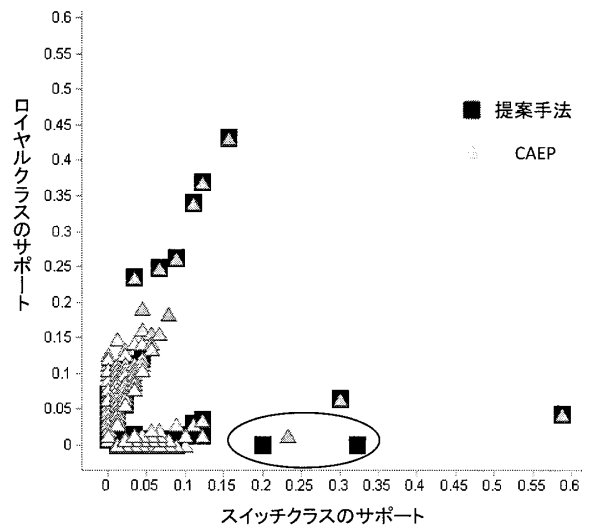


図5 柔軟剤類におけるパタン散布図

り、優れた結果であるとはいえない。この2つの比較では相対的に、“価格帯情報あり”の結果のほうがよいことがわかる。一方、提案手法をCAEPと比較すると、ライト系洗剤類の正答率では、評価値が同じであるが、他の正答率やスイッチクラスの予測に関するF1値では、提案手法の結果が勝っていることがわかる。これらの結果は、本研究で提案したアイテム化の方法と、シークエンシャルパターンを利用したCAESPを採用したことの複合的な効果であると考えられる。

結果の原因を詳しく見るために、提案手法とCAEPによって列挙された顕在シークエンシャルパターンと顕在パターンを、2つのクラスのサポートを軸に2次元空間に布置した1例が図5である。点が重なっている部分も多くあるが、例えば図中の中央下に楕円で囲ってある領域は、提案手法の点の方が多く出現していることがわかる。これらの点はスイッチクラスのpbがより1に近く、当該クラスより強い説明力になっていると考えられる。左下の点が重なっている領域においても、サポートはそれほど大きくないが、同様の傾向が確認され、それらの積み重ねによって提案手法の結果が改善していると考えられる。

次に、同じCAESPを適用しているが、アイテム化の方法が異なる“価格情報なし”、“価格帯情報あり”、そして提案手法について実際に出現した顕在シークエンシャルパターンのうち、比較的サポート値が大きなものを選択して顕在シークエンシャルパターンの図示化を行う。そのうち、比較的図がシンプルで見やすいライト系洗剤類に対する顕在シークエンシャルパターンを図

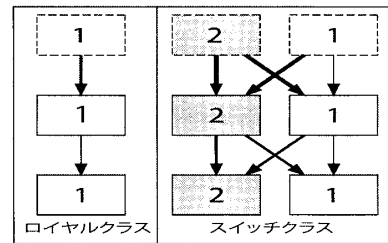


図6 “価格情報なし”の特微的なパタン

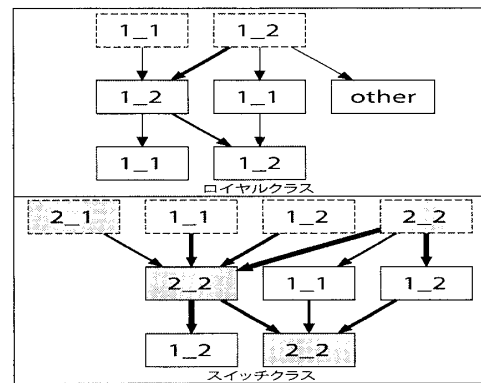


図7 “価格帯情報あり”の特微的なパタン

示したのが、図6、7、8である。3つの図は、それぞれ出現した顕在シークエンシャルパターンをパタン長2で区切り、その区切ったパタンの中で出現件数上位のパターンを組み合わせるグラフに表現している。図中の有向枝は購買順序を表し、有向枝の太さはパタンの出現数を相対的に表している。各ノード内の記号は、図6では“購買ブランド”を、図7では“購買ブランド_価格帯”を表現している。図8では、4節で説明し

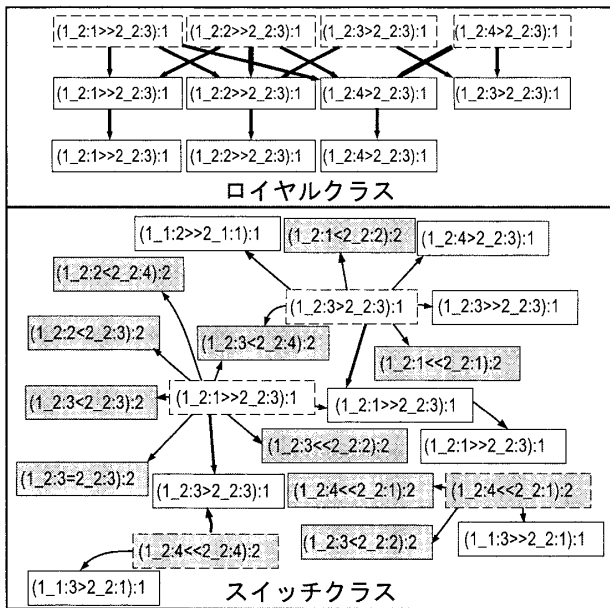


図8 提案手法での特徴的なパターン

たブランドの競合関係と選択された購買ブランドが、“(4節で説明したブランドの競合関係)：購買ブランド”のように表現されている。購買ブランドは、対象ブランドなら1、競合ブランドなら2、それ以外のブランドなら other と表記した。例えば、図8のスイッチクラスの“(1_2:4<<2_2:2):1”というアイテムは、対象ブランドであるエマールの価格帯2の値引き区分4の商品に対する競合ブランドのアクロンの価格帯と値引き区分が2で、価格はアクロンの方が非常に高い状況において、エマールが選択されたことを意味している。破線で描画されているノードは、シーケンシャルパターンとして最初に出現したアイテムを意味している。また、ノードがグレーになっているものは、競合ブランドを購入したノードを表している。

まず、図6からは、対象ブランドの商品を3回購買することで、ロイヤルティの高い顧客となる可能性が高いことがわかる。一方、競合ブランドを複数回購買していたり、対象ブランドと競合ブランドを交互に購買する顧客は、競合ブランドにスイッチする可能性が高いと考えられる。スイッチクラスの図にも、ロイヤルクラスと同じ対象ブランドを3つ連続して購買したように見える部分があるが、次の図7でも示されるように、これは最初対象ブランドを2回購買したが、次は競合ブランドを購入している場合と、その逆の場合が複合してそのように表示されていることに注意されたい。

次に、価格帯情報を加えた図7では、対象ブランド

の購買を続ける顧客の中でも、価格帯2に属する商品の購買を続ける顧客や、価格帯1の商品の購買の後に価格帯2の商品を購入する顧客がロイヤルクラスに特徴的な行動であることがわかる。一方、スイッチクラスでは、図6の傾向と同様に、競合ブランドの連続購買が顕著であるが、その中でも価格帯2の競合ブランドの連続購買が顕著であることがわかる。

最後にアイテム情報としてさらに値引き区分と価格差も加えた図8を見ると、ロイヤルクラスでは、“対象ブランドの価格帯2”の購買において、競合ブランドと比較して価格が相対的に高くても、購買が続いていることが確認できる。一方、スイッチクラスでも、対象ブランドを相対的に高くても購買しているノードは確認されるが、その次には、逆に価格が高くても、“競合ブランドの価格帯2”を購入していることがわかる。以上より、価格属性をアイテムで表現しなければ、単にブランドの購買の連続状態が表現されるだけであるが、価格帯や競合ブランドの価格差を表現することで、同じブランドでも、どの価格帯の商品が重要で、競合ブランドとのどのような価格差が重要であるのかという違いがわかった。これはブランドを展開する企業の立場から考えると、自社のブランドのどのような価格帯の商品の重要性があるのかを認識すること、そして競合ブランドがある値引き価格で販売しているという状況を所与としたときに、自社ブランドの商品に、価格プロモーションを実施するならどのような価格帯をどのような値引きで行うべきか、または価格プロモーションは行うべきではないのかのヒントになる。

例えば、上記の例でいえば、エマールのロイヤルクラスは、図8において、値段が安価であることを理由に購買を継続しているような顕著な傾向が見られなかったことから、価格を安くすればロイヤルクラスの顧客を維持できるわけではないことがわかる。そのため価格差が今回データから考慮されている範囲であるならば、競合他社との価格差については、相対的に高額であってもあまり敏感になる必要はない。逆にいえば、多少の値引きで競合ブランドより相対的に安い価格を提示しても、ロイヤルクラスの顧客を増やすことは難しいようである。一方、重要なのは、エマールを相対的に高い価格で購入してくれた顧客が、逆に相対的に高い価格でアクロンを購入しているタイミングが、ブランドスイッチをする危険なシグナルであることがわかる。上記の内容から総合的に考えると、これはエマールを購入している顧客の消費行動を価格によって誘

導することが困難であることが予想されるので、価格以外のプロモーションを行うか、または広告宣伝や新商品開発など別の手段によって、ロイヤルティを確保する手段を講じる必要があることがわかる。

6. おわりに

本研究では、購買されたブランドおよびその競合ブランドの商品価格、そして価格差の情報を表現に取り入れた CAESP を利用したブランドスイッチに関する分類モデルを提案した。アイテム化の違いや、CAEP との計算結果の比較から、今回適用したデータに対しては、提案手法が有効であることがわかった。また実際に出現した顕在シークエンシャルパターンを図示することによって、提案したアイテム化の影響を考察するとともに、実用例についても考察した。今後の課題としては、より複数の競合ブランド間の価格情報を表現に取り入れる工夫をすることによって、より一般的な手法となるように改善したいと考えている。

謝辞 本研究では経済産業省「情報大航海プロジェクト」の“パーソナル情報保護・解析基盤開発・改良と検証”において開発されたプログラムを一部使用して

いる。また、国立情報学研究所 宇野毅明先生が開発されたプログラムも使用させていただいている。ここに深謝の意を表する。

参考文献

- [1] 経済産業省, “パーソナル情報保護・解析基盤の開発・改良と検証—マイニング機能の改良詳細仕様書,” 平成 21 年度経済産業省情報大航海プロジェクト.
- [2] 経済産業省, “ブランド価値評価研究会報告書,” 平成 14 年度経済産業省企業法制委員会.
- [3] M. Givon, “Variety Seeking through Brand Switching,” *Marketing Science*, Vol. 3, No. 1, pp. 1-22, 1984.
- [4] 近藤文代, 黒田哲平, “CRM のための購買間隔を考慮したバラエティシーキングによる顧客セグメンテーション”, *マーケティングサイエンス* 16(1・2), pp. 95-114, 2009.
- [5] G. Dong, X. Zhang, L. Wong and J. Li: “CAEP: Classification by Aggregating Emerging Patterns,” *Proceeding of the second International Conference on Discovery Science*, 1999.
- [6] R. Agrawal and R. Srikant, “Mining Sequential Patterns,” *Proceeding of the 11th International Conference on Data engineering*, Taipei, Taiwan, 1995.