

ID付きPOSデータからの顧客行動パタンの抽出

会員番号 01206880 群馬大学工学部 関 庸一 SEKI yoichi

1 はじめに

小売業においては、ポイントカードやハウスカードの普及努力によって、POS (Point of Sales) データが顧客識別して収集されることが多くなっている。この場合、単なるPOSデータと異なり、顧客(カード)ごとに、購買時点を越えて名寄せしたデータ解析が可能となるため、時間軸上で顧客の行動パターンを把握することが可能となる。

しかし、このようなデータは数十メガ~数十ギガバイトになり、これを取り扱う方法は限定される。本研究では、これに対して一連の解析方法を与えることにより、CRM(Customer Relation Management)の基礎資料を得る一つの方法を示す。

ID付きPOSデータは一般に、個々の購買事例での購入数と単価などの値が、顧客×日付×商品の三相の反復をもって繰返されるデータとなる。また、付帯データとして、それぞれの相に顧客なら性別/年齢/住所など、日付なら曜日/天候/マーケティング活動など、商品なら商品分類/売場などが与えられる場合がある。顧客×日付の反復は来店の反復とも解釈できる。

本研究では顧客の特徴の把握を目的とするが、三相のままの解析は難しいので、顧客×日付と、来店×商品の二相データに縮約してから、パターン抽出を検討する。なお、以下での計算量の記述は、元データを顧客IDと来店日で辞書式に整序してあることを前提とする。

2 ある食品スーパーの解析事例

対象のID付きPOSデータは、福岡県のある食品スーパーで、2000年4月21日からの181日間に収集されたものであり、この間に購買のあった14,182人(男性2052人, 女性12130人)の3,031,812点、587,006,031円の購買分のデータ(約386Mbyte)である。

2.1 来店パターン分析

顧客×日付の二相データから顧客がどんなときに来店する生活パターンをしているかを分析するため、来店日付の曜日に注目し、顧客ごとの来店曜日分布を求め、こ

れから来店曜日パタンの指標を作成した。来店パタンの最も基礎的な指標は来店回数であり、これで各曜日の来店回数の個人間変動のうち80.5%が説明されていたが、残り2割の情報から主成分分析により来店回数とは独立な来店パターンを抽出し、顧客に主成分得点を与えた。具体的には顧客×曜日の $n \times 7$ 来店曜日分布行列 A から来店回数 $v = A1_7$ を差し引いた $\tilde{A} = A - v1_7^t/7$ に対して主成分分析をおこなった¹。この分析では、来店曜日分布と主成分得点を求める際にのみ、元レコードと顧客の逐次処理が必要である。

因子負荷量(表1)から、第1主成分は土日に負荷が高いので休日度、第2主成分は店頭割引が行われている火日金曜日に負荷が高いので、特売日度と解釈した。

表1: 因子負荷量

曜日	第1	第2	第3	第4
月	-0.2694	-0.3587	0.3908	-0.2977
火	-0.2477	0.8003	-0.1805	-0.0643
水	-0.1301	-0.3246	0.0963	0.3061
木	-0.1658	-0.0178	0.0272	-0.6140
金	-0.2421	0.0548	0.1514	0.6606
土	0.2064	-0.3119	-0.8188	0.0097
日	0.8488	0.1580	0.3337	-0.0003
寄与率	0.384	0.185	0.136	0.107
解釈	休日度	特売日度	-	-

2.2 購買商品パターン分析

来店×商品の二相データを用い、同時購買関係から顧客の食生活タイプの指標を作成した。同一日同一顧客IDのレコード群を一度の購買と解釈し(285,670来店)、各来店における購買商品の商品クラス(191クラス)上の購買個数分布を集計し、これに双対尺度法(数量化III類、Correspondence Analysis) [1]を適用し、顧客と来店の双方に共通の尺度上の得点を与えた。対象商品クラスは、この店舗で主要な売上げを占める食品に限定し、また、期間限定のギフト類などを除いた²。分析結果を図1に示す。第1軸には簡易食(弁当, おにぎり)↔伝統食材の順に並ぶので家庭調理度と、第2軸には食事↔菓子の順に並ぶので子育て度と解釈した。

¹ $1_k = (1, 1, \dots, 1)^t$ を k 次元ベクトルとする。

² 販売期間が限定された商品クラスがあると、そのクラスが特殊な次元として抽出されてしまう。

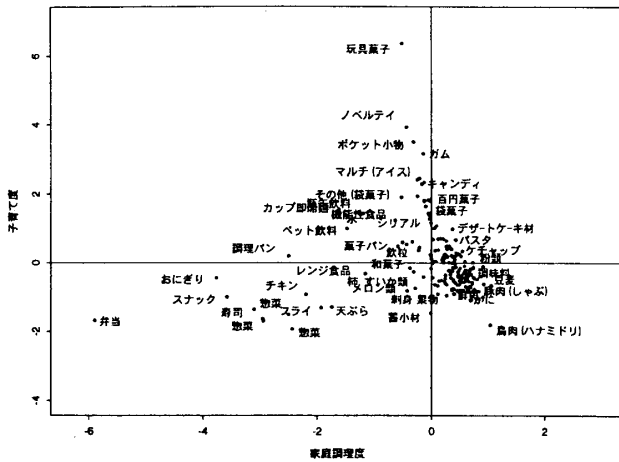


図1 双対尺度法による商品クラス布置

2.3 大規模データの双対尺度法

双対尺度法は、数学的には $n \times p$ 行列 F を $B = D_r^{-1/2} F D_c^{-1/2}$ と表側の合計に関して基準化³した行列を考え、直交行列 U, V と対角行列 Λ による B の特異値分解 $B = U \Lambda V^t$ を求めることで、行、列の得点を $X = D_c^{-1/2} V \sqrt{f_i}$, $Y = D_r^{-1/2} U \sqrt{f_i}$ と与えるものである⁴。データを $F = D_r Y \Lambda / f_i X^t D_c$ と分解することになる。

しかし、 n が大きい場合には特異値分解が実行できない。その場合でも、 p が適当な大きさであれば、 $p \times p$ 行列 $G = B^t B = D_c^{-1/2} F^t D_r^{-1} F D_c^{-1/2}$ を求め、これを特異値分解することにより、 $G = V \Lambda^2 V^t$ が得られ、これから、 $X = D_c^{-1/2} V \sqrt{f_i}$, $Y = D_r^{-1} F X \Lambda^{-1}$ と得点を計算することができる。この計算によれば、 G, Y の算出で、それぞれ n に関する逐次処理 1 回が必要だけで、 $O(n)$ のメモリは必要ない。

2.4 購買行動パターンと価格感度の分析

前節までで求めた指標を、各種の層別要因 (常連客⁵かどうかなど) ごとに平均し、その性格を検討した。図2は、当該店舗周辺の地図上に、住所の字ごとに来店曜日パターン指標値を示したものである。山がちとなり地形的に行き止まりになる南東側で、相対的に特売日度が高く、北西側では休日度が高いことがわかる。

³ F の行と列の総計ベクトルを $f_r = F \mathbf{1}_p$, $f_c = \mathbf{1}_n^t F$, $f_i = \mathbf{1}_n^t F \mathbf{1}_p$ として $D_r = \text{diag}(f_r)$, $D_c = \text{diag}(f_c)$ 。

⁴ ただし、 U, Λ, V の第1次元は無意味な解で利用しない。

⁵ 総顧客のうち最初と最後の一月の両方に来店し、かつ6か月間に6回より多くの来店があった客を常連客 (6489人, 45.8%) とした。残りの客を非常連客 (7963人) と呼ぶ。

さらに得られた指標と価格感度との関連を検討した。ここで、顧客の価格感度とは、次のように定義される購買ごとの (利益への) 寄与の個人平均として定義した。

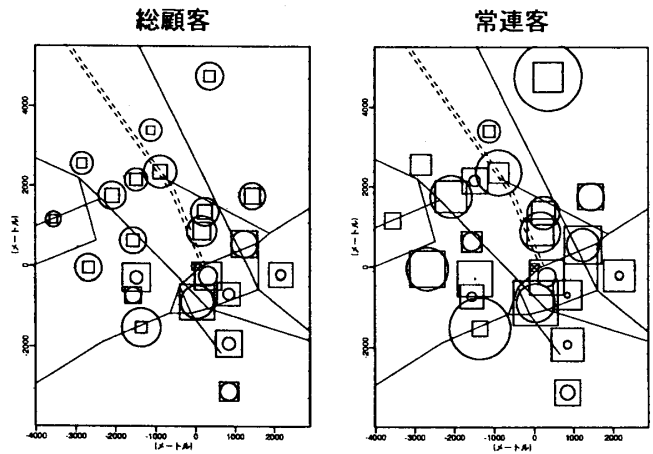
$$\text{寄与} = \text{購入価格} - \text{平均実売価格}$$

図3より、子育て度が高いと寄与が大きい。また、家庭調理度はほぼ個人平均 (-0.124) を谷として、そこから外れると高くなる傾向がわかる。育児に手間が掛ったり、自炊できない客層は時間コストが高く価格に敏感でないと予想される。また、伝統食中心の客層は高齢者に重心があることが別途わかっており、この層が価格に敏感でないと考えられる。

以上のように顧客行動パターンを数値化することで、客層の全体像を把握し、マーケティングのターゲットを絞ることが可能になると考える。

参考文献

- [1] 西里静彦, 質的データの数量化, 朝倉書店 (1982).



○：休日度/□：特売日度, 大きいほど得点が高い
図上の線は主要道路, 上が北で福岡市がある

図2 休日度/特売日度の地域分布

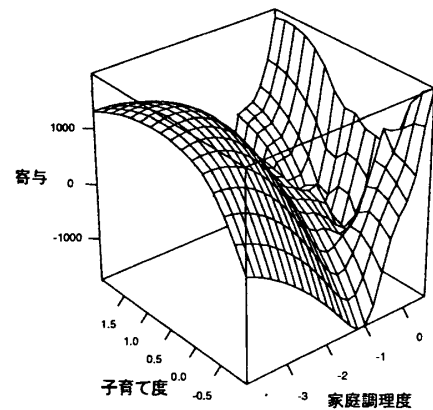


図3 購買商品パターンと寄与の関係 (常連客)