

ID 付き POS データからの顧客行動パターンの抽出

関 庸一

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メガ〜ギガバイト単位になり、これを取り扱う方法は限定される。本研究では、これに対してデータの持つ反復の構造に注目して、各反復、あるいは、反復の組み合わせの持つ情報を抽出するための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。

2. ID 付き POS データの構造と分析

ID 付き POS データは一般に、個々の商品が POS レジを通過した事実を記録したデータとなる。これに通過時刻 (日付) の情報と、カードから得られる購買顧客の ID 情報が合わせられて 1 レコードのデータとなる。同一商品の購買をまとめて、1 レコードにした場合を考えるとすると、各レコードは顧客、日付、商品の組で特定されることになり、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、付帯データセットとして、それぞれの相に、顧客なら入会時に収集された性別/年齢/住所など、日付なら曜日/天候/マーケティング活動状況など、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。

せき よういち
群馬大学 工学部
〒376-8515 桐生市天神町 1-5-1
受付 02.7.16 採扱 02.10.17

本研究では顧客の特徴の把握を目的とするが、以下の手順での分析を提案する。

1. まず、一つ一つの相の性格を明らかにする解析を行う。これにより、店舗における顧客の行動の傾向を把握して、対象データの絞り込み、層別などが必要であるかを判断する。
2. 三相データの中に現れる顧客パターンの抽出を行う。ただし、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、顧客の時間軸上での行動パターンや、購買商品の同時購買のパターンの抽出を検討する。
3. 以上で抽出されたパターンとの関係で、顧客の購買行動に関する傾向を見る。

次節以降で、解析事例の中で具体的な方法を示していく。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。また、分析では、UNIX 上のテキスト処理言語 perl で前処理を行った上で、Splus¹ を用いている。

3. ある食品スーパーの解析事例

3.1 データ概要

対象の ID 付 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 10 月 20 日までの 181 日間に収集されたものであり、この間にポイントカードで購買のあった 14,182 人の 3,031,812 点、587,006,031 円の購買分のデータ (約 386 Mbyte) である。

3.2 各相の概要の分析

3.2.1 商品概要

対象期間中に購買のあったのは 17,234 種の商品であったが、個々の商品 (ラララ特選コシヒカリ 10 kg 米, 酪農牛乳 1 L など) は、米, 鮮魚, 牛乳, 練り物などの 283 クラスに小分類されており、さらに、鮮魚, 野菜, 嗜好品, 乳製品などの 39 ラインに大分類され

¹ Version 3.4 J Release 1 for Sun SPARC: 1997 [1]

ていた。

このうちクラス分類には同一のクラス名があるなど、ユーザから見たクラス分類となっているといい難かったが、商品名からのクラスの再分類は困難であったので、そのまま用いて以下の分析を行っている。売上構成としては、食品が売上の主要な部分（92.7%）を成している（表1）。

各購買レコードには実売価格があるので、対象期間

表1 売上構成(%)

順位	ライン名	構成比	累積
1	鮮魚	8.48	8.48
2	野菜	6.53	15.01
3	嗜好品	6.41	21.42
4	乳製品	6.02	27.43
5	米	5.93	33.37
6	水物日配	5.61	38.98
7	調味料	5.37	44.35
8	パン	4.80	49.14
9	袋菓子	4.34	53.49
10	練物日配	4.27	57.76

中の平均をとることにより、商品ごとの平均実売価格が明らかとなる。個々の購買について、購買価格の平均実売価格との差額を（その購買の利益への）寄与、また、購入価格の平均実売価格との比を掛率と呼んで、その購買における、店舗と客との利得の按分の指標として用いる。

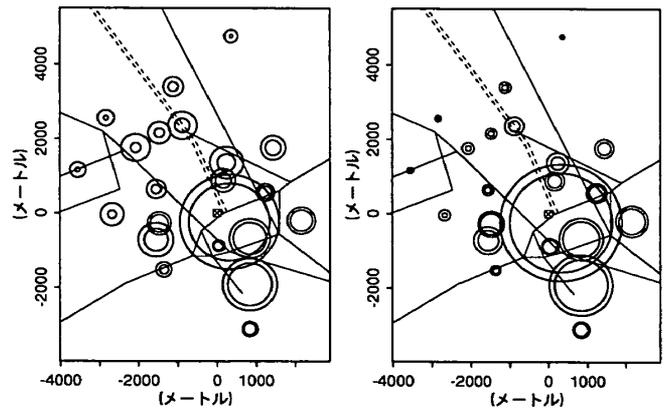
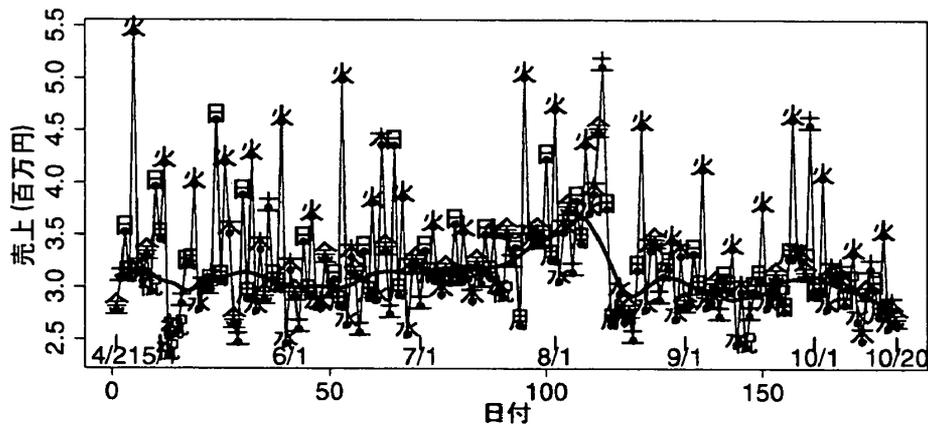
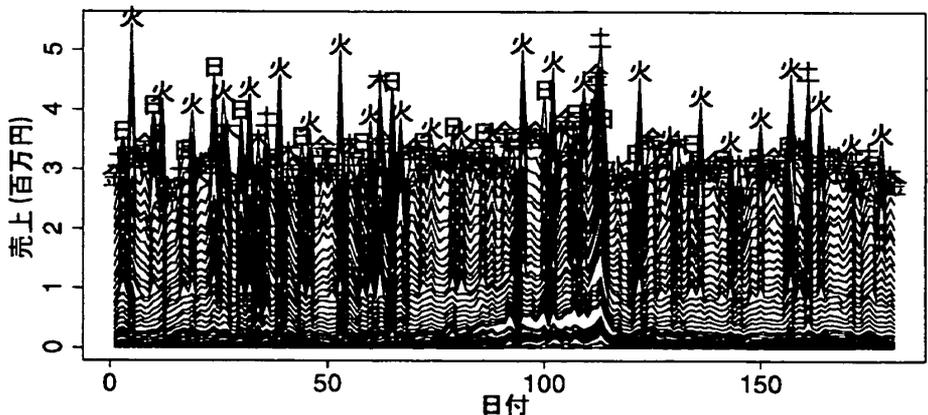


図1 顧客の地域分布（円の面積が量に対応、内円が常連客、外円が総顧客、図上の線は主要交通路、上が北で福岡市がある）



(a) 総売上高



(b) クラスごとの売上高の積み上げ図

図2 日々の売上高の時系列変化

掛率 = 購入価格 / 平均実売価格 (1)

寄与 = 購入価格 - 平均実売価格 (2)

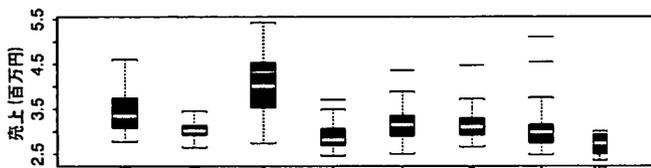
ただし、極めて大きな掛率の変動を示す“果物奉仕品”などという商品名もあり、商品内容が一貫して定まっていることが疑われる場合があった。この指標を用いる場合には、この点に注意する必要がある。

3.2.2 顧客概要

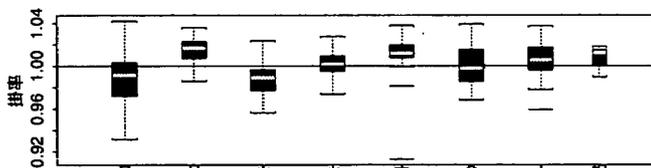
購買のあった顧客 14,182 人は、男性 2,052 人、女性 12,130 人、平均年齢は男性 52.5 歳、女性 49.6 歳、全体で 50.0 歳であった。6ヶ月の期間中、1人当たり 213.8 個、41,390 円の購買が行われている。

総顧客の中には、来店が1回だけの客が 2,536 名も

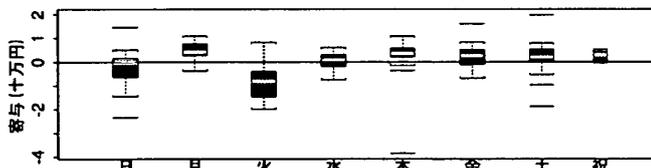
いるなど、来店の仕方には大きな差異がある。そこで、来店の継続性と頻度から、顧客を常連客と非常連客の二つの客層に分けた。具体的には、総顧客のうち最初と最後の1月の両方に来店した 7,172 人 (50.6%) と、6ヶ月間に6回より多くの来店があった 7,771 人 (54.8%) の共通部分を常連客 (6,489 人、45.8%) とした。残りの客を非常連客 (7,963 人) と呼ぶ。



(a) 売上高



(b) 平均掛率



(c) 寄与

図3 商品の曜日ごとの販売状況

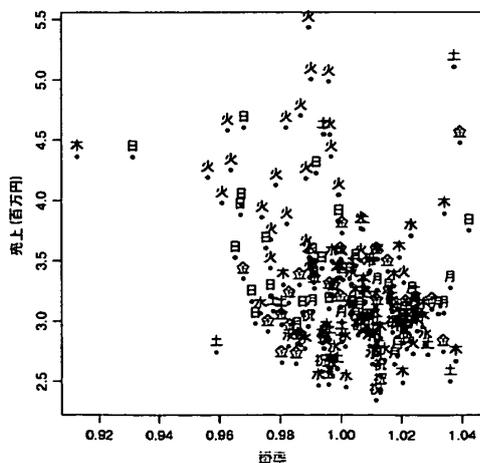
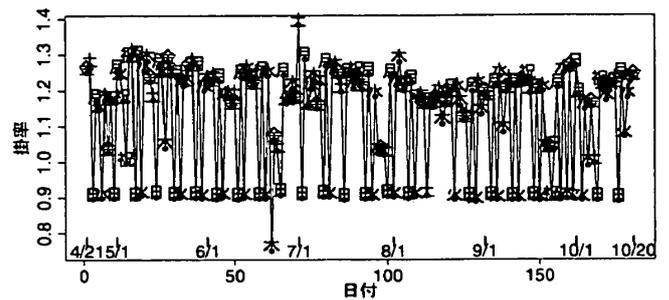
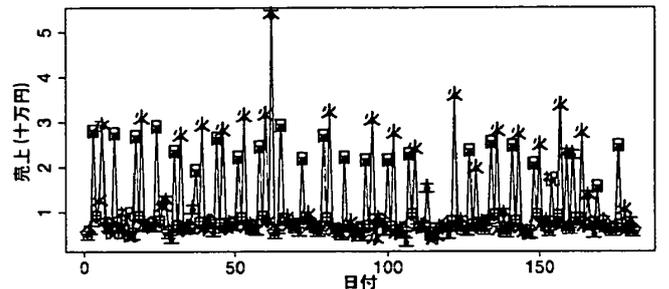


図4 日々の平均掛率と売上高

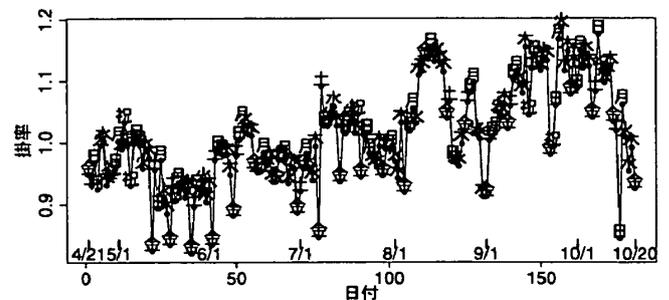


(a) 平均掛率

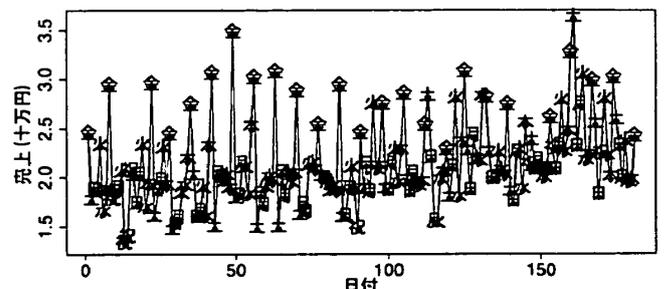


(b) 売上高

図5 冷凍食品の売上高分布



(a) 平均掛率



(b) 売上高

図6 野菜の売上高分布

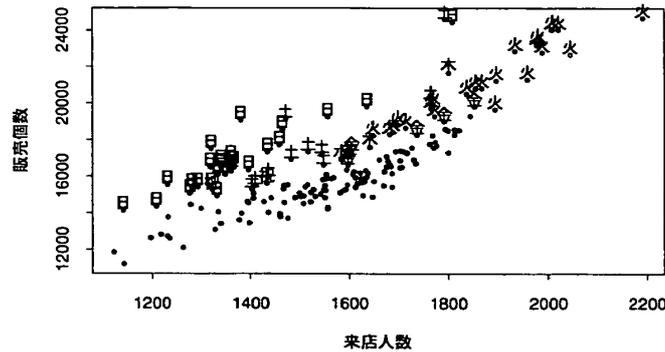
図1は当該店舗周辺の地図上に、住所の地区ごとの顧客のボリュームを表現したものである。顧客分布はほぼ半径数kmにあり、さらに購買額で見ると大部分が半径2km程の中にあることが読み取れる。

3.2.3 日付概要

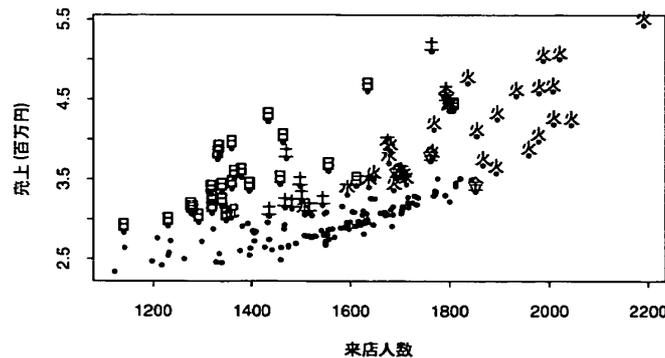
この店舗では、期間中1日平均16,750個、3,243,000円の売上があった。図2からその時系列変化を見ると、お盆の時期にギフト商品などのクラスに一時的売上があることが分かる。また、火、日曜日には、売上げが高い日が多いことも分かる。

曜日ごとに集計すると(図3(a))、火、日曜日に売上が多いことが確認できる。これは、これらの曜日で平均掛率が低い(図3(b))と考えられる。逆に、金曜日も掛率を低目にして販売している傾向が分かるが、それは、売上高に目立った影響を与えていない。ただし、寄与の面ではプラスに貢献している(図3(c))。

店舗全体として日々の平均掛率と売上高の関係(図4)を見ても明瞭な関係は見られないが、各ラインごとに売上と掛率の関係を把握してみると、特異な変動をしているものが分かる。例えば、冷凍食品(図5)は火曜日に日曜日に特売をしており、顧客もその日に購買することになっていることが分かる。その他、野菜



(a) 来店人数と販売個数



(b) 来店人数と売上高

図7 来店者数と販売状況

については(図6)、金曜日は安く価格設定され、それなりの顧客の反応があることが分かる。しかし、上述のように店舗全体の売上高には、目立った寄与をしていないことが分かる。

3.2.4 来店概要

同一個人の一日の購買を、1回の来店での購買とみなすと、期間中、285,670回の来店があったことに

表2 来店パターン因子負荷量

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	-	-

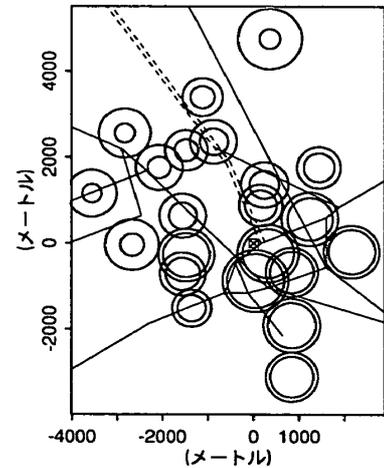
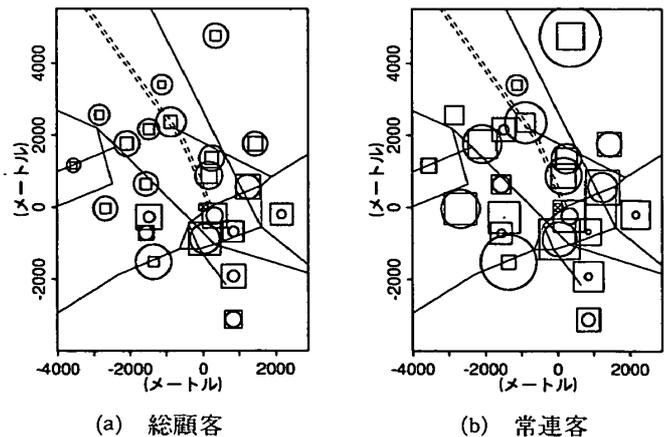


図8 来店回数の地域分布(外円:常連客/内円:全顧客)



(a) 総顧客

(b) 常連客

図9 休日度/特売日度の地域分布(○:休日度/□:特売日度, 大きいほど得点が高い)

しては、来店頻度が北西で低いことが分かる(図8)。しかし、常連客の来店頻度には大きな地域差はない。来店パターンとしては、総顧客では(図9(a))、中心ほど特売日度が高く、山がちで地形的に行き止まりになる南東側でも、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分かる。また、常連客に限ると(図9(b))、休日度、特売日度ともにハッキリし、全体に特売日度が高いことが分かる。これは、常連客が店舗でいつ特売が行われているかをよく知っており、それに合わせた来店行動をとっているためと思われる。また、ここでも北西の客は休日度が高い。ただし、数少ない遠北西の常連客は休日度が低いことが分かる。

また、図10より、年齢とともに休日度が低くなり、特売日度は70歳台までは高まっていくことが分かる。また、性別では、特に仕事を持つであろう50歳台までの男性は休日度が高く、女性は一般に男性に比べ特売日度が高いことが分かる。

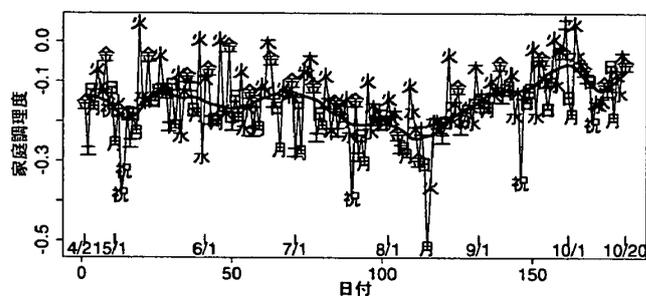
3.4 購買食品パターン分析

3.4.1 大規模データの双対尺度法

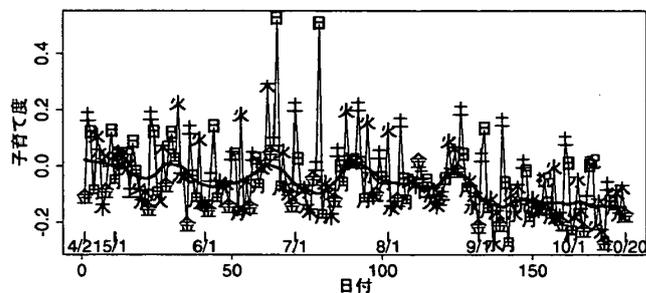
多くの対象についての同一カテゴリ上の頻度分布が与えられた場合、カテゴリと対象を同一尺度上の得点に数量化するには、双対尺度法(数量化III類, Correspondence Analysis) [2]を用いる。双対尺度法は、数学的には $n \times p$ 行列 F を $B = D_r^{-1/2} F D_c^{-1/2}$ と表側の合計に関して基準化²した行列を考え、直交行列 U , V と対角行列 Λ による B の特異値分解 $B = U \Lambda V^t$

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。ただし、 U , Λ , V の第1次元は無意味な解で利用しない。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 についての逐次処理1回が必要だけで、 $O(n)$ のメモリは必要ない。なお、 G を求めるには F の各行ベクトル F_i について、行列 $F_i^t F_i / f_{ri}$ を求めては i について累積していき、最後



(a) 家庭調理度



(b) 子育て度

図13 購入食品パターンの時系列変化

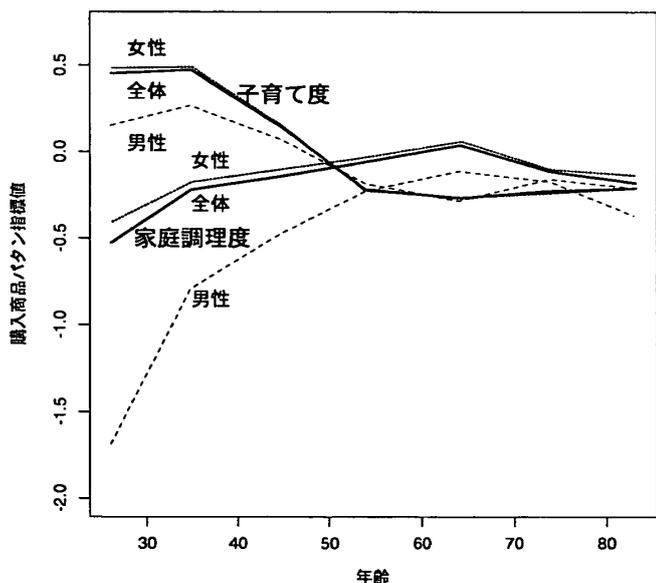
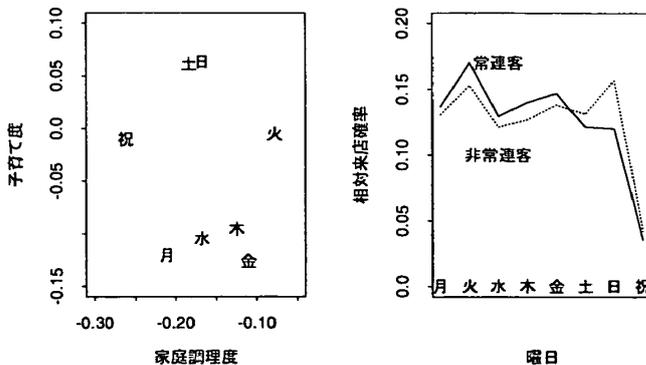


図12 常連客の性別年代ごとの購買食員パターン

² 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 $D_r = \text{diag}(f_r)$, $D_c = \text{diag}(f_c)$.



(a) 購入食品パターン

(b) 相対来店比率

図14 曜日の特徴づけ

に平行して求めておいた D_c で基準化すればよい。

3.4.2 購買食品パタンの抽出

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。1回の来店での購買商品（同一日同一顧客IDのレコード群）の商品クラス（191クラス）上での購買個数分布を集計し、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。これは、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ためである。得られた商品得点の2次元布置を図11に示す。第1軸は簡易食（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軸は和食材→洋食材→菓子の順に並ぶので子育て度と解釈した。

商品だけでなく、来店についても同一の尺度での得点を与えられているので、来店に与えられた得点を各顧客ごとに平均することで、顧客の購買食品パターンが得られる。これを客層ごとに平均したものを図12に示す。これより、50歳代以前の男性は子育て度、特

に家庭調理度が低く、女性も年齢とともに、子育て度が下ることが分かる。

また、来店に与えられた得点を日付ごとに平均することで、日付の購買食品パターンが得られる。この時系列変化を図13に示す。家庭調理度が祝祭日に下ること、子育て度が長期的に低下してきていることなどが

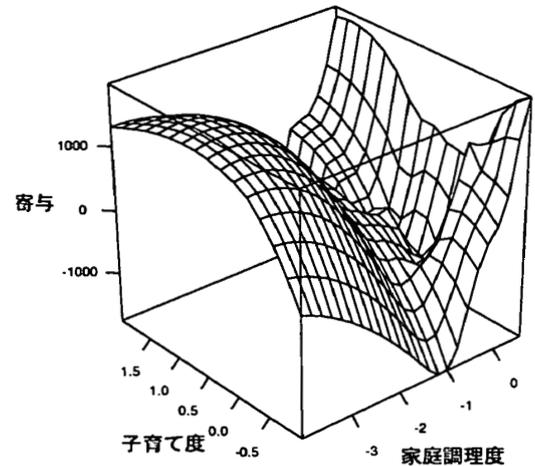
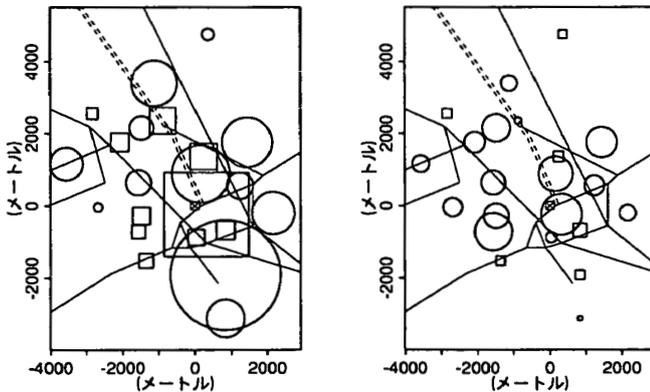


図16 購買食品パターンと寄与の関係（常連客）



(a) 常連客

(b) 非常連客

図15 寄与の地域分布（○：正，□：負，面積が絶対値）

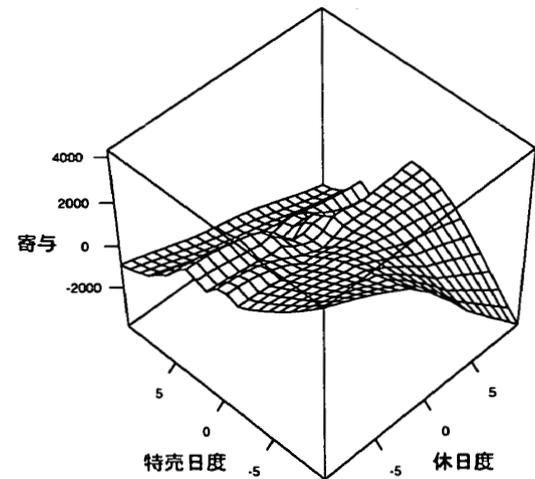


図17 来店パターンと寄与の関係（常連客）

表3 階層ごとの期間中の累計寄与

年齢層	女性 常連客		男性 常連客		非常連客	
	総寄与	人数 人当り	総寄与	人数 人当り	総寄与	人数 人当り
-29	-119849.2	279 -429.5	18734.9	29 646.0	-19273.3	613 -31.4
30-	-408697.7	875 -467.0	20502.9	69 297.1	-71363.2	1527 -46.7
40-	-417442.2	1582 -263.8	-67436.4	194 -347.6	24915.9	2092 11.9
50-	-620509.4	1349 -459.9	-85198.0	209 -407.6	10062.1	1841 5.4
60-	620628.7	899 690.3	122573.2	130 942.8	163146.1	988 165.1
70-	430850.5	534 806.8	103942.4	108 962.4	61490.9	493 124.7
80-	105999.0	165 642.4	90995.1	38 2394.6	26015.9	116 224.2
合計	-412992.1	5683 -72.3	211228.2	777 270.8	201763.9	7670 26.2

分かる。

また、それを曜日ごとに集計することで、曜日の特徴づけを見ることができ、図 14(a)からは、休日は家庭調理度が低く、休日や特売日は子育て度が高いという特徴があることが分かる。これは、図 14(b)に示されるように非常連客が常連客に比べ、相対的に日曜日の来店が多いことが一つの原因と思われる。

3.5 購買行動パターンと価格感度の分析

購買ごとの寄与を顧客ごとに合計したものは、顧客が平均価格を意識しながら購買を行っていたかの累積であり、顧客の価格感度の一つの指標と考えられる。また、寄与を店舗から見ると、標準的利幅へ加算される追加利益である。以下で、寄与の特徴を見た後、購買行動パターンの指標との関連を見る。

まず、寄与の地域特性(図 15)は常連客ではつきりした傾向を持つことが分かる。店舗最近の地区は大きく負であるが、購買量も大きいので、ボリューム利益の源泉であると考えられる。また、南東側地域では正であり、常連客も多かったので、安定顧客が多く高利益の源泉であると評価できる。一方、北西はロイヤリティの低い客が多く、寄与の絶対値も小さいことが分かる。この客層がこの店舗にとっては挑戦対象であると考えられる。

一方、年齢、性別で見ると、値引きをしなくとも購買してくれるのは、以下のような顧客層であることが分かる(表 3)。

1. 一般に高齢層(60歳代以降)
2. 人数は少ないが特に若い男性常連客
3. 40歳以降の非常連客

顧客行動パターンとの関連を見るため、顧客行動パターンの指標に対して、寄与がどの程度になるかを平滑化して表した結果を図 16, 17 に示す。購買食品パターンについては、子育て度が高いと寄与が大きく、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分かる。育児に手間がかかったり、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、60歳代に家庭調理度の高い伝統食中心の客層の重心があり、この層が価格に敏感でないと考えられる。

また、来店パターンとの関連を見ると、寄与が高い人は、休日度が低く、特売日避ける顧客であるか、休日度が高く特売日度が平均的な値の客層であることが分かる。

4. おわりに

購買データ一件一件の持つ情報は希薄であるが、それを集積して、興味のない反復を単純集計や平均で集約することで、多くの情報が引出せることを示した。その際、来店曜日や購買商品クラス上の分布のような多次元量については、それを低次元に数量化しておくことで、集約が容易となることを示した。個々の顧客は様々であるが、その一般傾向を把握することにより、マーケティングターゲットを絞ることなどが容易になると考える。

今回の解析では、確率モデルを構成しての検証的な分析は行わなかった。この点が今後の課題となる。

謝辞 本研究を進めるに当たって、解析に協力してくれた小茂田宏、シティーレズキア、金澤由佳、荒井学各位に感謝する。

参考文献

- [1] 伊藤幹夫他訳, Splus による統計解析, シュプリンガー・フェアラーク東京(2001).
- [2] 西里静彦, 質的データの数量化, 朝倉書店(1982).