

# 事象系列のストリング分析 —百貨店における買回り行動の分析—

関 庸一, 小茂田 宏, 石原 淳一郎

## 1. はじめに

小売業においては、ポイントカードなどの普及努力によって、POS (Point of Sales) データを顧客識別して収集することが可能となってきている。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。

本研究では特に、注目する事象の前後に顧客がどのような購買を行っていたかを分析する方法として、行動の事象系列を文字列に要約し、そこに表れる行動パターンの出現頻度を一般化線形モデル[1]などでモデル化する方法を提案する。この方法をある百貨店における POS データに適用し、店舗内における顧客の買回り行動を明らかにする。

## 2. 購買事象順列の構造と分析

ID 付き POS データは一般に、顧客、日付、商品の組で特定される個々の購買事例がレコードになり、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り返されるデータとなる。また、付帯データセットとして、それぞれの相の属性が利用できる場合がある。

本研究では顧客の来店時の行動パターンを把握する方法として、次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. 1回の来店での識別したい購買行動の単位 (部門または品番への訪問) とその分類カテゴリを確定する。また、来店行動の中で特に注目したい事象 (例えば分析対象とする商品の購買) を定める。
3. 前述の購買行動単位に文字を割り当て、1回の来店で発生した購買レコード群を対応文字の文字列に変換する。この際、同一単位に属し連続するレコードは1文字に集約する。
4. 作成した文字列のなかで注目事象が表れる位置で文字列を分割し、注目事象、事前購買行動文字列、事後購買行動文字列の三つにし、これに、来店に付随する情報 (顧客、日付など) を添付して一つのレコードに集約する。
5. 上記レコード群をサンプルとして、注目事象の生起頻度、および、それと他の行動との同時生起頻度を集計する。これをグラフ化したり、確率モデル化し、注目事象の生起と他の行動との関連を明らかにする。

このように来店を文字列にすることによって、文字列長から併買訪問回数が、また、文字種ごとの文字数から各部門への訪問数、文字の前後関係からどの部門の購買がどの部門への購買につながるかを、注目事象種類あるいは訪問の特徴ごとに集計することができる。これらの分析は文字列のパターンマッチング処理を利用することにより比較的容易に行うことができる。

次節以降の解析事例の中で具体的な方法を示す。なお、分析では、UNIX 上のテキスト処理言語 perl で前処理を行い、統計ソフト Splus[2, 3]を用いている。

## 3. ある百貨店の解析事例

### 3.1 データ概要

対象の ID 付き POS データは、平成 14 年度データ解析コンペティションで提供されたデータであり、あ

せき よういち  
群馬大学 工学部情報工学科  
〒376-8515 桐生市天神町 1-5-1  
こもだ ひろし, いしはら じゅんいちろう  
群馬大学 工学研究科情報工学専攻  
〒376-8515 桐生市天神町 1-5-1  
受理 03.7.25 採択 03.9.2



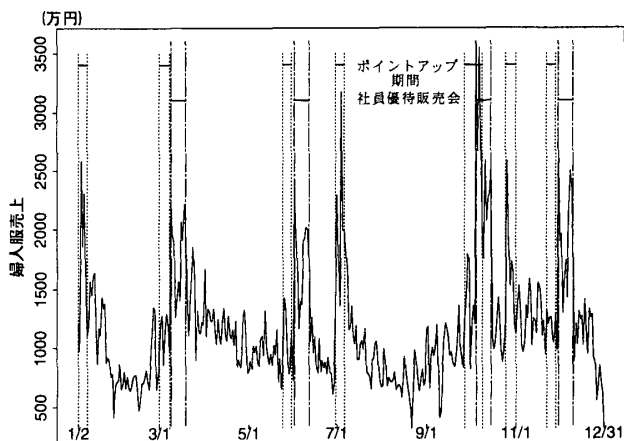


図5 婦人服の売上の時系列推移

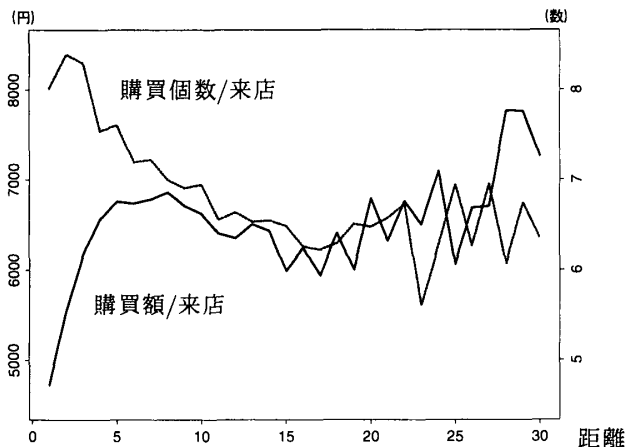


図7 店舗からの距離ごとの来店当たりの購買額・個数

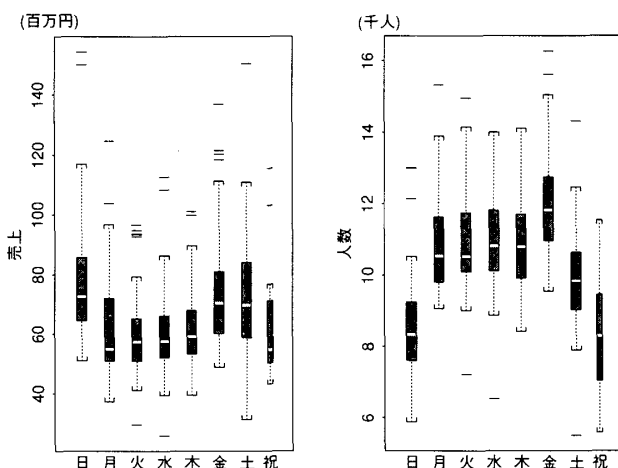


図6 曜日による来店人数や売上の違い

期に売上が伸びていた。これはお中元やお歳暮、正月用品などが含まれていると考えられる。また婦人服や婦人洋品はポイントアップなどのセールス期間によく売れていることがわかる(図5)。その他、曜日で特殊な売れ方をしている商品として寝装品、家庭用品があり、他の日に比べ日曜日によく売れていた。また、曜日で全体の来店人数、売上を見ると(図6)、金曜日に来店者数、売上共に高くなっていることがわかり、土曜、日曜は来店者数は少ないが、売上は高くなっていることがわかる。

### 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、3,792,215回の来店があったことになる。顧客は期間中に平均13.18回来店し、1回の来店で6.31個、平均10,484円の購買をしているが、来店が1回のみ顧客は26.79%であった。図7は顧客の来店当たりの購買個数と購買額を距離ごとに示している。店舗に近い顧客は小額の購買を高頻度でして

おり、平均購買単価は低い。

顧客の初回購入時刻を集計すると、午前中に小さなピークがあり、また、平日は17:00~20:00、土休日は16:00~17:00に最大のピークがあることがわかった。そこで、顧客の来店時間帯を午前10:00~12:59、午後13:00~16:59、夜17:00~20:00の三つに分類し、これと来店日の曜日と組合せた時間帯上で、顧客ごとの来店回数分布を求めた。この顧客×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、主成分分析により来店パターンの指標を作成した[4]。なお、この分析は当該店への来店習慣を想定しやすい顧客として、6回以上の来店があった顧客122,881人(42.7%)に限定して行った。因子負荷量(表1)から、第1主成分は平日夜に負荷が高いので**オフィス度**、第2主成分は平日午後に負荷が高いので**有閑度**、第3主成分は土日・祝日に負荷が高いので**休日度**と解釈した。

個人の主成分得点を求め、各顧客層の平均をとることにより、その特性を調べた。性別・年齢別(図8)では、20~40代の特に女性でオフィス度が高くなっており、男女共に年齢と共に減少し特に60代を過ぎると急激に低くなる。有閑度は女性の40~80歳くらいの、主婦層と考えられる顧客層で高くなっており、全体の傾向としては男女とも年齢と共に上昇していく。休日度は性別・年齢別であまり差異は見られないが、若干年齢と共に減少していく。また距離別(図9)では、距離12くらいまでは店舗からの距離に比例し、有閑度・休日度はだんだん減少し、オフィス度は増加していくことがわかる。距離12を過ぎたところから、それぞれほぼ一定の値になっている。

表1 来店パターン因子負荷量

時間帯	曜日	第一軸	第二軸	第三軸
午前	月	-0.3993	-0.5030	-0.3919
	火	-0.3930	-0.4977	-0.3995
	水	-0.3809	-0.4999	-0.4018
	木	-0.3944	-0.5314	-0.3557
	金	-0.3930	-0.4975	-0.4330
	土	-0.3240	-0.4964	0.0313
	日	-0.3165	-0.4795	0.1965
午後	祝	-0.3104	-0.5740	0.1882
	月	-0.4587	0.5502	-0.0946
	火	-0.4590	0.5625	-0.1009
	水	-0.4462	0.5445	-0.0831
	木	-0.4668	0.5167	-0.0436
	金	-0.4761	0.5771	-0.1335
	土	-0.2912	0.1484	0.4342
夜	日	-0.2783	0.0122	0.5694
	祝	-0.3500	-0.3541	0.3781
	月	0.8182	0.0985	-0.1062
	火	0.8301	0.1154	-0.1035
	水	0.8271	0.1006	-0.0859
	木	0.8193	0.0846	-0.0691
	金	0.8079	0.1325	-0.1216
土	0.2838	-0.1295	0.5603	
日	0.1635	-0.1882	0.6069	
祝	-0.0888	-0.4743	0.3955	
寄与率 (%)		32.93	14.48	8.16
解釈		オフィス度	有閑度	休日度

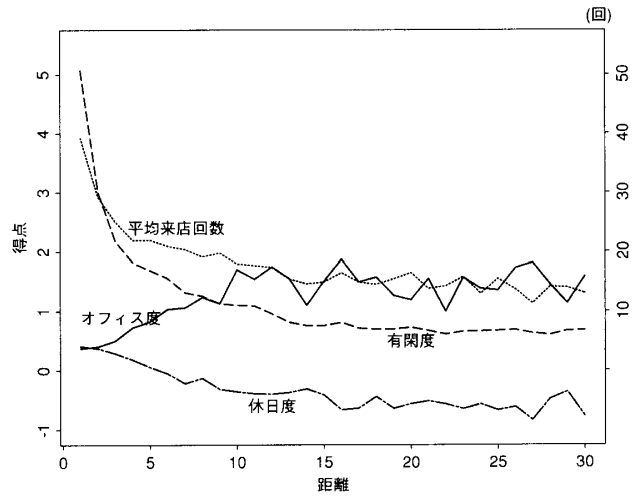


図9 距離別での来店傾向

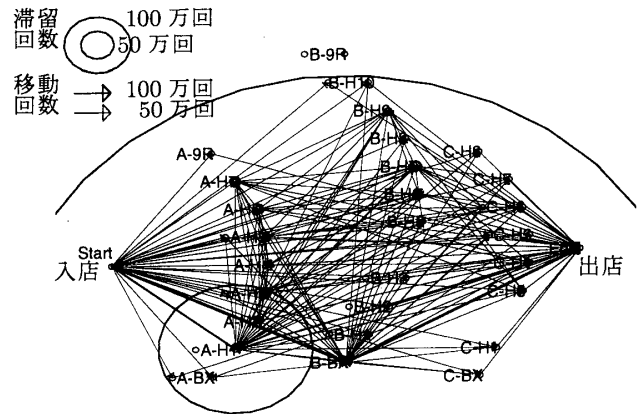


図10 総顧客のフロア移動傾向

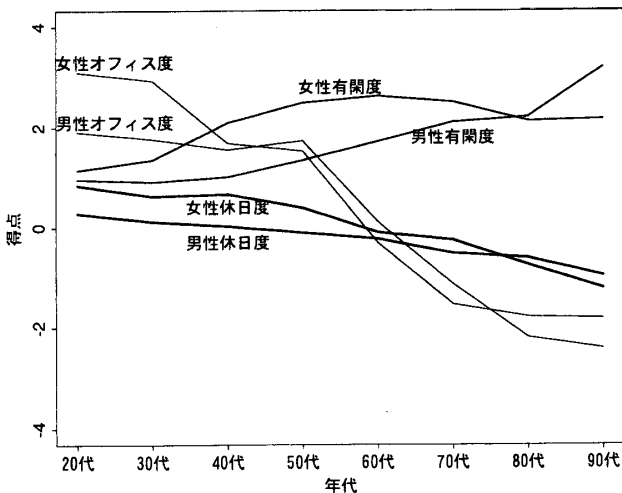


図8 年齢・性別での来店傾向

### 3.3 店内移動概況

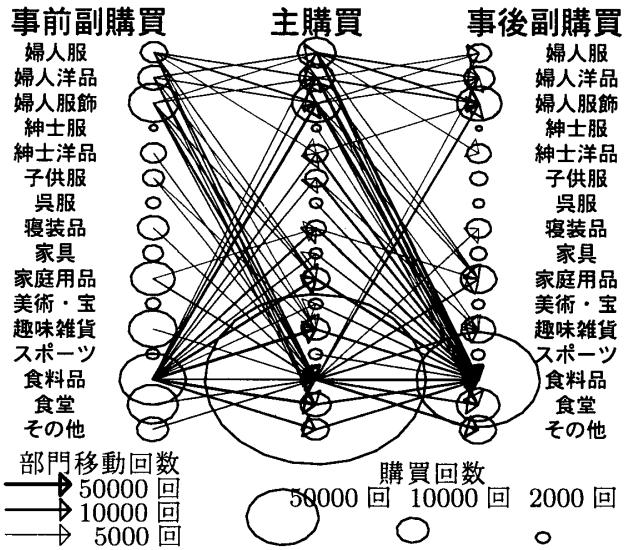
買回りについてフロア間の移動概況を検討するため、来店ごとに購買のあったフロアを調べ、あるフロアでの購買が次にどのフロアでの購買につながったかを集計した。ここでは、次のように定義する。

- フロア：館・階別（各階の食堂は別フロア扱い）
- 滞留：同一フロアでの反復購買
- 入店：ある来店の最初の購買の前
- 出店：ある来店の最後の購買の後

総顧客のフロア移動傾向（図10）から、A館1階とB館地下1階の食料品フロアへの移動やそのフロアでの滞留が多く、また食料品フロアからの移動もあり、下層階から上層階へ移動する噴水効果が見られていることがわかる。また、下層階での相互移動が少ないことが見て取れる。ただし、全体としてはあるフロアで購買が行われた後、特定のフロアへ移動するという顕著な傾向は発見できなかった。

### 4. スtring分析

まず、全来店を対象にしたString分析により部門買回りの概況を把握した上で、ここまでの顧客購買行動の分析から発見できた性別・年齢・距離での顧客の違いなどによる来店分類で、確率モデルに基づくスト



(30回以下の移動には矢線を引いていない)

図11 主購買と対応する副購買 (総顧客)

リング分析を行う。これにより、客層による購買行動の相違を明らかにする。なお、性別・年齢・距離といったデモグラフィックな要因で客層を規定すると、購買履歴をもたない新規顧客に対しても分類が容易にできるという利点がある。

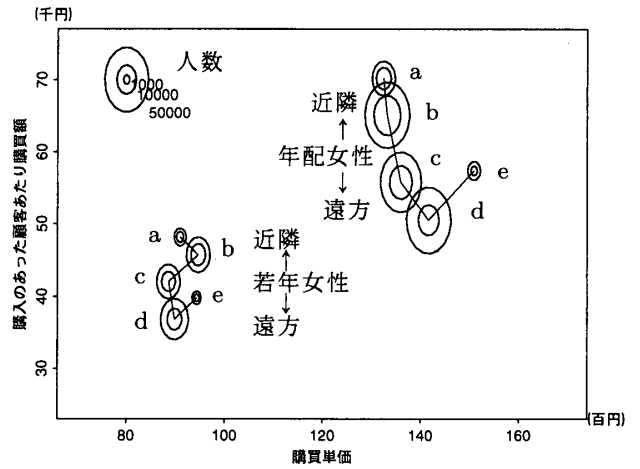
#### 4.1 部門単位の分析

各々の来店において、どのような部門の商品がどのような順序で購買されているかを、部門単位で買回り傾向を見ることによって分析する。注目事象としては、各々の来店で最大の購入金額のあった購買部門への訪問を考え、その購買部門を主購買部門と呼ぶことにする。それ以外の部門を副購買部門とした。

総顧客における主購買とそれに対応する副購買 (図11) から、食料品、次いで婦人関連の主購買が多いことがわかる。また、主購買と副購買との関連を見ると、食料品の副購買からはほとんどすべて各部門の主購買につながっており、地下食品売場から上層階への噴水効果はかなりあることがこの結果からもわかる。一方、婦人関連部門の購買では、相互に主購買と副購買がつながるほかは、食料品への買回りが多いこと、婦人関連部門の購買から家庭用品の購買につながる場合がかなりあることがわかる。

#### 4.2 婦人服購買の分析

前節のようにグラフで表現すると、部門間移動の概況がつかみやすいが多対多の関連の評価が難しいので、以下では婦人服部門内での購買に限定して、定量的な分析を行う。婦人服部門は、売上高で食品部門に次ぎ、利幅は食品部門に比べ多い商品群であり、この百貨店



(外円:顧客人数、内円:婦人服購買経験のある顧客人数)

図12 顧客層ごとの婦人服売上

における主力商品群と考えられる。

顧客としては、婦人服の主要顧客であり、来店者数も売上も大半を占める女性を対象を限定した。顧客の分類要因として、まず、年齢については、来店パタンの得点 (図8) などから40歳未満 (若年)・40歳以上 (年配) の二つに分類することとする。距離について、五つ (a: 0以上2未満, b: 2以上6未満, c: 6以上12未満, d: 12以上30未満, e: 30以上) に分類した女性の婦人服の売上状況を図12に示すが、これから若い女性は年配の女性に比べ、顧客当たり購買額も単価も低く、また、近隣は遠方に比べ、顧客当たり購入額が高いことがわかる。図中の円のまとまり方や売上や来店パタンの得点から、層の数を少なくするため、距離0以上12未満 (近隣) と12以上 (遠方) の二つに統合して分析を進める。また、来店の特徴として来店時間帯での違いを検討する。

購買行動単位としては、商品分類を1階層細かく見て、品番単位で購買行動を把握した。注目事象としては、婦人服品番の中で最大の購入金額のあった品番への訪問を考えた。節2の方法で品番単位で事象系列を作成し、そののち、評価を容易にするため、部門やフロアで頻度を集計して以下の分析を行った。

#### 4.2.1 婦人服品番の購買回数の分析

データを婦人服の購買のあった176,862来店に限定し、婦人服の主購買を、婦人服品番のあるフロア  $j = \{A: H1, B: H1, B: H2, A: H3, A: H5, B: H5, C: H5, A: H6, B: H6, C: H6, A: BX, \text{催事}\}$  (表2参照) に集約して検討した。

この婦人服の主購買回数について、来店特徴ごとに以下のようなポアソン分布に従うという仮定を置いて

表2 婦人服フロア概要

フロア	総来店数*	総売上(円)*	代表品番名
A:BX	16,344	150,801,341	MK クランプリコス, アンタイトル
B:H2	1,931	22,528,211	ストール ストール
B:H5	21,527	384,853,954	23ク, トランスワーク, エンスウィート
A:H1	9,942	82,793,941	フランドル シティ スペース
A:H3	2,948	25,963,375	サザビー
B:H6	45,721	974,752,509	プレツケンリッジ, シンプルライフ, バーバリー
C:H6	37,463	848,977,744	クイーンサイズ, ジェーンモア, レリアン
A:H5	64,084	799,392,867	ヤングキャラクター, スタイル・コム, ロペ
A:H6	16,952	490,448,402	イタリヤ, ジヤンニロジユデイチエ
B:H1	11,086	108,554,955	VIS ライフスタイル デイスへ, アナイ
C:H5	13,494	298,442,516	婦人フォーマル, マーガレットサイズ
催事	2,435	6,132,310	婦人催事服

\*: カードを利用した主購買分の総計

表3 婦人服フロアの購買回数とその他の要因との交互作用効果

婦人服 フロア	$\beta_j^{(1)}$	$\beta_{jk_1}^{(4)}$				$\beta_{jk_2}^{(5)}$		$\beta_{jm_1}^{(6)}$	$\beta_{jm_2}^{(7)}$
	フロア	金/平日	土/平日	日/平日	祝/平日	午後/午前	夜/午前	近/遠	年配/若年
A:BX	2.916*	0.680	0.478*	0.259*	0.937	0.880	0.397*	0.601*	0.832
B:H2	1.048	0.719	0.437*	0.367*	1.079	1.471*	2.129*	0.901	0.878
B:H5	2.429*	0.849	0.577*	0.752	1.130	0.375*	0.456*	0.707*	0.937
A:H1	1	1	1	1	1	1	1	1	1
A:H3	0.847	2.267*	0.839	0.655*	1.644	0.508*	1.009	1.580*	1.010
B:H6	1.295	1.110	1.010	0.593*	0.985	0.556*	0.318*	1.453*	1.076
C:H6	1.228	1.005	0.661*	0.802	1.789	0.362*	0.376*	1.675*	1.357*
A:H6	1.323	1.775*	0.944	0.537*	1.315	0.281*	0.434*	1.291	1.467*
A:H5	0.755	1.242	1.388	0.686*	1.435	0.705	1.140	0.972	1.469*
B:H1	0.534*	1.153	0.683*	0.814	2.099*	1.671*	0.888	1.138	1.667*
C:H5	1.504*	1.087	0.549*	0.451*	1.471	0.241*	0.442*	1.265	2.050*
催事	0.767	1.544*	0.451*	0.469*	0.711	1.356	0.771	0.968	2.198*

注: 数値は A:H1 (A館1階) を1とした相対値, 行順は若年の購買比率の大きい順

分析を行った。

$$x_{jk_1k_2m_1m_2} \sim Po(\lambda_{jk_1k_2m_1m_2} \times T_{k_1k_2})$$

ここで、 $T_{k_1k_2}$  は、時間帯  $k$  のこの1年間の累積営業時間であり、この項を導入することにより、購買回数期待値  $\lambda_{jk_1k_2m_1m_2}$  は時間当たりの量となる。また、来店特徴としては、時間帯  $k_1 = \{\text{平日, 金, 土, 日, 祝}\}$ ,  $k_2 = \{\text{午前, 午後, 夜}\}$ , 客層  $m_1 = \{\text{遠方, 近隣}\}$ ,  $m_2 = \{\text{年配, 若年}\}$  を考えた。ただし、時間帯とは来店の最初の購買時刻がいつであったかを意味する。以上、 $12 \times 5 \times 3 \times 2 \times 2 = 720$  ケースについて、婦人服主購買回数  $x_{jk_1k_2m_1m_2}$  を集計し、元データとした。

このポアソン分布の平均値  $\lambda_{jk_1k_2m_1m_2}$  を

$$\lambda_{jk_1k_2m_1m_2} = \beta^{(0)} \beta_j^{(1)} \beta_{k_1k_2}^{(2)} \beta_{m_1m_2}^{(3)} \beta_{jk_1}^{(4)} \beta_{jk_2}^{(5)} \beta_{m_1}^{(6)} \beta_{m_2}^{(7)} \times \beta_{k_1m_1}^{(8)} \beta_{k_1m_2}^{(9)} \beta_{k_2m_1}^{(10)} \beta_{k_2m_2}^{(11)} \quad (1)$$

と乗法的に分解する。各々のパラメータは、その条件

のもとで、購買回数の期待値が何倍になるかという意味をもつことになる。

ポアソン分布を仮定した一般化線形モデルで各パラメータを推定した結果を表3~5に示す<sup>1</sup>。なお、定数

<sup>1</sup> 実際には、対数リンク関数を用いた一般化線形モデルで加法的モデルとして、

$$\begin{aligned} \log(\lambda_{jk_1k_2m_1m_2}) &= \alpha^{(0)} + \alpha_j^{(1)} + \alpha_{k_1k_2}^{(2)} + \alpha_{m_1m_2}^{(3)} \\ &+ \alpha_{jk_1}^{(4)} + \alpha_{jk_2}^{(5)} + \alpha_{m_1}^{(6)} + \alpha_{m_2}^{(7)} \\ &+ \alpha_{k_1m_1}^{(8)} + \alpha_{k_1m_2}^{(9)} + \alpha_{k_2m_1}^{(10)} + \alpha_{k_2m_2}^{(11)} \end{aligned} \quad (2)$$

なるモデルを推定し、得られた各  $\alpha$  を指数変換して  $\beta$  としている。表3~5では、仮説  $\alpha = 0$  の下での  $t$  値が10を超えるものに \* を付している。要因のダミー変数化には treatment 対比を用いているので、表中の1は基準を示すのみで推定されたものではない。また、二元表形式の推定値での \* は第1列、第1行以外は交互作用の有意の有無を示す。

表4 曜日ごとの婦人服購買への効果

	$\beta_{k_1 k_2}^{(2)}$			$\beta_{k_1 m_1}^{(8)}$	$\beta_{k_1 m_2}^{(9)}$
	午前	午後	夜	近隣/遠方	年配/若年
平日	1	1.924*	2.749*	1	1
金曜	1.633*	2.358*	2.983*	0.856*	0.894
土曜	1.889*	4.397	3.873*	0.956	0.845*
日曜	2.303*	6.328*	4.414*	0.777*	0.864
祝日	1.200	2.430	1.875*	0.646*	0.943

表5 客層と時間の婦人服購買への効果

	$\beta_{m_1 m_2}^{(3)}$			$\beta_{k_2 m_1}^{(10)}$	$\beta_{k_2 m_2}^{(11)}$
	遠方	近隣		遠方/近隣	年配/若年
若年	1	0.909	午前	1	1
年配	0.883	0.933*	午後	1.352*	0.952
			夜	0.889	0.890

項  $\beta^{(0)}=0.416$  であり、これが平日午前中 A 館 H 1 フロアへの遠方若年層の時間当たり品番訪問回数となる。

表3に、婦人服フロアに関連したパラメータを、若年の購買が多い順に示してある。各フロアで婦人服に割り当てられた売場面積が不明なので、フロアの販売効率は評価できないが、A 館 BX, B 館 H 5 が時間当たりの訪問数が特に多いことがわかる。同表の他の効果を見ると、これらのフロアでは平日午前中の遠方若年者の来店が相対的に多いことがわかる。一方、表の下半分のフロアでは、どちらかというとな隣からの年配者が多いことがわかる。特に C 館 H 5 や催事は年配の購買が多いことがわかる。

表4の曜日ごとの効果を見ると、週末の午後以降に購買が多く、遠方の若年がそれに対応していると思われる。時間に依存しない購買回数全体の効果(表5)としても、遠方の若年が最大で、近隣の年配が次いでいる。婦人服購買はこれらの層で多いといえる。

#### 4.2.2 併買数の分析

前節で分析した720ケースの婦人服の主購買に対し、それぞれに併買訪問が何回(品番単位)あったかを検討した。併買の発生回数は、その発生部門  $i=\{\text{婦人服, 婦人用品, \dots, その他}\}$  と、その併買が主購買の前だったのか後だったのか( $l$ )を区別して集計し、頻度  $y_{ijk_1 k_2 m_1 m_2 l}$  を求めた。条件組合せの数としては  $720 \times 16 \times 2 = 23,040$  ケースが考えられるが、婦人服主購買回数なかった場合を除いて、22,943 ケースが元データとなる。

これが以下のポアソン分布に従うという仮定をおく。ここで、 $x_{ijk_1 k_2 m_1 m_2 l}$  をモデルに入れることにより、期

待値  $\mu_{ijk_1 k_2 m_1 m_2 l}$  は婦人服フロアへの品番訪問1回当たりの併買回数となる。

$$y_{ijk_1 k_2 m_1 m_2 l} \sim Po(\mu_{ijk_1 k_2 m_1 m_2 l} \times x_{ijk_1 k_2 m_1 m_2 l})$$

このモデルで、併買の期待値  $\mu_{ijk_1 k_2 m_1 m_2 l}$  を

$$\begin{aligned} \mu_{ijk_1 k_2 m_1 m_2 l} = & \gamma^{(0)} \gamma_i^{(1)} \gamma_j^{(2)} \gamma_{k_1 k_2}^{(3)} \gamma_{m_1 m_2}^{(4)} \gamma_l^{(5)} \\ & \times \gamma_{i m_1 m_2}^{(6)} \gamma_{j l}^{(7)} \gamma_{j l}^{(8)} \gamma_{m_1 m_2 l}^{(9)} \end{aligned} \quad (3)$$

と分解した。

ポアソン分布を仮定した一般化線形モデルで各パラメータを推定した結果を表6~8に示す<sup>2</sup>。なお、定数項  $\gamma^{(0)}=0.764$  であり、これは平日午前中 A 館 H 1 フロアでの近隣若年層の婦人服主購買1回当たりの併買訪問回数(品番数)となる。また、婦人服購買の前後での併買回数のパラメータは  $\gamma_2^{(5)}=1.381$  と推定され、上の条件では婦人服購買前より後の併買が4割近く多いことがわかった。

部門別の併買効果を表6に見ると、婦人服の主購買には、食料品の購買や婦人服飾の併買が多いことがわかる。これらは主購買の前に行われることが相対的に多い。一方、頻度は少ないが、紳士服や子供服、呉服などが婦人服に併買されるとすると、婦人服購買の後に購入されることが多いことがわかる。購買事象系列中で前に表れる部門は来店のきっかけである場合が多いと予想されるので、その意味で重要な部門と考えられる。

客層を見ると、近隣の年配で一般に買回りが多いことがわかる。特に、呉服、美術・宝などの副購買が目立つ。同様に、遠方の年配も全般に買回りが多いことがわかる。一方、近隣の若年は相対的に子供服、婦人服飾、趣味雑貨を中心とした買回りをし、遠方の若年は、婦人服飾、紳士用品などで他の客層と同程度である以外には、買回りが少ないことがわかる。客層ごとに併買の多い部門が異なることがわかるが、客層に対応した部門の魅力を向上させることができれば、その層からの婦人服の購買も増えることが期待される。

主購買の婦人服として何を購入したかでの併買効果の違いなどが表7からわかる。C: H 6, C: H 5, B: H 6 などの高層のフロアで婦人服を購入した客は併買を多くしていることがわかる。その場合、他のフロアに比べ、副購買をした後に婦人服の主購買をしている場合が多いことがわかる。客層単独の効果としても、

<sup>2</sup> 対数リンク関数を用いた加法モデルを脚注1と同様に考え、これにサンプル重み  $x_{ijk_1 k_2 m_1 m_2 l}$  を与えて解いている。表6~8では、仮説  $\alpha=0$  の下でのt値が500を超えるものに\*を付している。

表6 部門への併買効果とその交互作用

	$\gamma_i^{(1)}$	$\gamma_{im_1m_2}^{(6)}$			$\gamma_{il}^{(7)}$
	各部門/ 婦人服	近隣年配	遠方若年	遠方年配	後/前
		近隣若年	近隣若年	近隣若年	
婦人服	1	1	1	1	1
婦人洋品	0.948	1.091	0.902	1.132	0.670*
婦人服飾	1.742*	0.653*	0.958	0.673*	0.759*
紳士服	0.021*	1.674	0.443	1.655	1.401*
紳士洋品	0.132*	2.289*	0.983	2.327*	0.922
子供服	0.228*	0.623*	0.535*	0.523*	1.301*
呉服	0.015*	2.660*	0.634	2.415	1.309*
寝装品	0.086*	2.135*	0.903	2.017*	0.964
家具	0.035*	1.936	0.752	1.535	0.999
家庭用品	0.128*	2.493*	0.743	2.139*	1.059
美術・宝	0.009*	5.107*	1.217	3.846*	1.122
趣味雑貨	0.199*	0.836	0.610*	0.634*	1.160*
スポーツ	0.026*	3.466*	1.326	4.342*	0.758*
食料品	4.937*	2.174*	0.543*	1.327*	0.164*
食堂	0.206*	2.188*	0.650*	2.050*	0.905*
その他	0.063*	1.401	0.444*	0.946	0.859

年配の顧客が多く併買を行っている。

表8を見ると、祝日や金曜以外の夜など、比較的百貨店が空いている時間帯に併買が多いことがわかる。

### 5. おわりに

データ規模の問題から一般化線形モデルを個々の来店に当てはめることができなかつたので、個人属性や日々の環境変化の利用は性別年齢と曜日時間帯に限定したものとなった。例えば、販売促進活動の効果が図5のようにあることはわかっているが、これをモデルに取り込み個人のより詳細な分析を行えば、CRM (Customer Relationship Management) の基礎資料として有効な情報になると考えられる。さらなる分析法の開発が一つの課題となる。

来店データ1件1件のもつ情報は希薄であるが、それを部門や品番への訪問があったかどうかのみに集約することによって、多くの情報が引出せることを示した。個々の顧客はさまざまであるが、その一般傾向を層別して把握することにより、マーケティングターゲットに対する効果を予測することなどが可能になると考える。

**謝辞** データ解析コンペティションの過程で、貴重な御意見を下さった日本OR学会マーケティング・デー

表7 フロアや客層ごとの併買効果の相違

	$\gamma_j^{(2)}$	$\gamma_{jl}^{(8)}$
	各フロア	後/前
A:BX	0.900	1.051
B:H2	0.486	1.418
B:H5	1.567*	0.694*
A:H1	1	1
A:H3	1.237	0.912
B:H6	1.925*	0.732*
C:H6	2.046*	0.689*
A:H5	1.590*	0.795
A:H6	1.634*	0.838
B:H1	0.979	0.800
C:H5	1.817*	0.652*
催事	1.153	1.571

	$\gamma_{m_1m_2}^{(4)}$	$\gamma_{m_1m_2l}^{(9)}$
	各客層	後/前
近隣若年	1	1
近隣年配	1.766*	1.022
遠方若年	0.704*	1.039
遠方年配	1.649*	1.111

表8 曜日と時間帯での併買誘発効果の相違

	$\gamma_{k_1k_2}^{(3)}$		
	午前	午後	夜
平日	1	0.267*	1.117*
金曜	0.053*	0.471*	0.750*
土曜	0.944*	0.141*	1.141*
日曜	0.200*	0.315*	1.204*
祝日	3.985*	2.112*	5.453*

タ解析研究部会の方々に感謝する。また、本研究を進めるに当たって、解析に協力してくれた群馬大学の荒井学、村井崇浩、中西剛太郎、斎坂真吾、黒田和宏ら学生各位に感謝する。

### 参考文献

- [1] McCullagh, P. and Nelder, F. R. S.: *Generalized Linear Models*, Chapman & Hall, London, 1989.
- [2] 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.
- [3] Chambers, J. M. and Hastie, T. J. eds.: *Statistical Models in S*, Wadsworth, Inc., 1991.
- [4] 関庸一: ID付きPOSデータからの顧客行動パタンの抽出, オペレーションズ・リサーチ, 48(2), 75-82, 2003.