

時空間を基軸とした百貨店 POS データの分析

降旗 徹馬, 西野 和美, 寺崎 康博

1. はじめに

小売店舗の売上は顧客一人一人の購買の積み重ねで形成されている。これら顧客の個人レベルでの購買行動を理解することなしに売上の増大を目指した効果的なマーケティング戦略を立案することはできない。従来は POS システムから得られる集計レベルの売上が主な分析対象であったのに対し、近年では、情報技術の進展によって、POS システムに FSP (Frequent Shoppers Program) を組み合わせ、顧客別の購買実績を把握することが実務レベルで可能になってきている。これは同時に、顧客の購買履歴すべてをデータベースに蓄えることを意味し、データがますます大量化している。現在は、この非集計レベルの大量データをいかに分析し活用していくかが重要な課題となっている[3]。

本稿では、こうした問題意識のもと、ある百貨店の顧客別 POS データに対して、データマイニングとモデル分析を行った結果を報告する。具体的には、「顧客はどこから来てどのように購買するのか」という顧客の購買行動を理解するための基本的な問いに対し、

- ・顧客分布の時空間変動を把握する消費者密度推定手法を提示し、その変動を把握し知見を得る（空間的側面の分析）
 - ・顧客の店舗内滞留時間と購買額の間関係を明らかにする（時間的側面の分析）
 - ・購買パターン分析から有力な購買パターンを抽出し、売上との関係を調べる（顧客の店舗内行動分析）
 - ・これらが顧客の空間分布および滞留時間と密接に関連していることを示す（相互関連性の把握）
- といった分析結果を述べる。分析に用いたある百貨店

の POS データは平成 14 年度データ解析コンペティションより提供されたもので、顧客属性が付随する約 1200 万レコードの取引明細データ（期間は 2001 年 1 月～12 月）である。

2. 顧客分布の時空間変動解析

小売店舗にとって、顧客がどこから来店しているのかという商圈の把握は、基本的で重要な分析である。近年の情報技術の進展は、こうした分析に必要な顧客情報の入手を容易にしつつある。なかでも、FSP を通じて得られる顧客属性が付随する POS データは、顧客の購買履歴のみならず位置情報（住所）を含む顧客属性が付随しているため、顧客分布や顧客密度を推定するのに十分な情報を有している。このデータが与えられれば、メッシュやコロプレスマップ¹による視覚的表現や等値線図による商圈の表現、平滑化技法の一種である傾向面分析²およびノンパラメトリックな密度推定手法の一つであるカーネル密度推定[4]を用いて顧客分布や顧客密度を推定することができる。これら手法において、メッシュやコロプレスマップは個々の領域内の顧客分布の相違を、等値線図では等値線間での顧客分布の相違を把握できず、傾向面分析は地点データとして与えられる顧客分布を任意の領域に集約して処理する間接的な手法であるため、地点データから顧客密度を直接的に推定するカーネル密度推定が集計誤差を低く押さえる点で有用であるといわれている[1, 2]。

本稿では、このカーネル密度推定に基づいて顧客分布を直接的に推定する。さらに、地点ごとに異なる消費者の重み（購買額や購買頻度）を考慮するための拡張を行い、日単位、月単位、曜日単位といった時間ごとの変動も考慮できるように拡張を行う。

ふりはた てつま, にしの かずみ, てらさき やすひろ
東京理科大学 経営学部
〒346-8512 久喜市下清久 500
受付 03.7.25 採扱 03.10.29

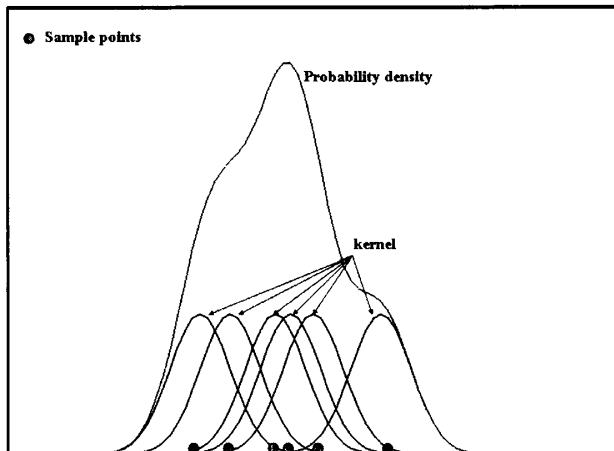
¹ 地域単位ごとに統計値を段階区分して彩色・陰影表示した地図。

² 種々の地域事象の全域的關係を規則正しい面的な傾向として見いだそうとする分析手法。

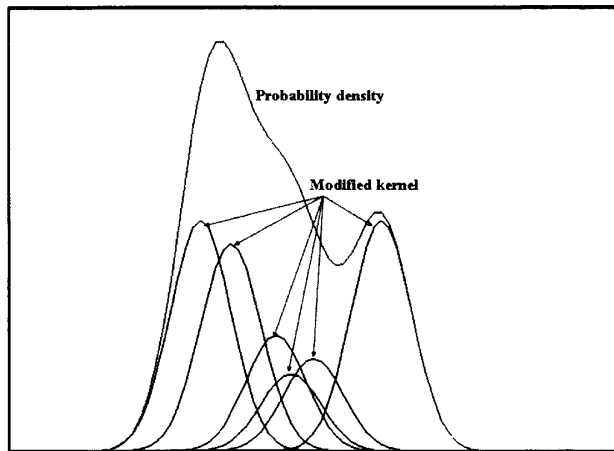
2.1 カーネル密度推定手法とその拡張

カーネル密度推定はサンプルデータを連続確率密度関数に平滑化する方法で、カーネル関数と平滑化のためのパラメータ $h(>0)$ を必要とする。その基本的な手続きは、図1(a)に1次元の例を示したとおり、サンプルとして与えられる個々の地点にカーネルと呼ばれる山をおき、それを累積させることによって密度分布を推定するものである。密度分布の形状は、パラメータ h に強い影響を受け、パラメータ h が大きくなると小さな変動ははっきりせずなだらかな分布になり、パラメータ h が小さいと凹凸の激しい密度面となる。また、密度分布の形状はカーネルとして選択される関数にはあまり反応せず、 n 個のサンプルに対するカーネル関数を $k(x)$ とすると、このカーネル関数は、

- (1) $0 < k(x) < C$, (2) $k(x) = k(-x)$, (3) $\int k(x) dx = 1$,
 (4) $\int k(x) x dx = 1$, (5) $n \rightarrow \infty$ のとき $h \rightarrow 0$ という性質を



(a) カーネル密度推定の例



(b) 拡張したカーネル密度推定の例

図1 カーネル密度推定とその拡張

有すればよく、正規分布や一様分布などの多くの確率分布がこの条件を満足する。顧客分布の密度推定においては2変量正規分布が一般的に用いられている。例えば、任意の地点を (x, y) とし、サンプル $i (=1, \dots, n)$ のカーネル関数を $k_i(x, y, h)$ 、サンプル i の座標を (x_i, y_i) とすると、

$$k_i(x, y, h) = \frac{1}{2\pi h^2} e^{-\frac{1}{2h^2}[(x-x_i)^2 + (y-y_i)^2]} \quad (1)$$

が用いられる。このカーネルを重ねることにより、

$$k(x, y, h) = \frac{1}{n} \sum_{i=1}^n k_i(x, y, h) \quad (2)$$

という顧客密度分布を得る。しかし、このカーネル関数の基本型のままでは顧客の購買力や買物頻度といった地点ごとに異なる重みを考慮できないため、本稿では、式(1)のカーネル関数に重みを加えて、

$$k_i(x, y, h) = \frac{w_i}{2\pi h^2} e^{-\frac{1}{2h^2}[(x-x_i)^2 + (y-y_i)^2]} \quad (3)$$

とする。ここで、 w_i は地点 i の重みを示し、図1(b)に示されるように、購買力や買物頻度といった地点固有の重みに応じてカーネルの山の高さを変えることを意味する。この拡張により、カーネル関数が満たすべき性質、特に確率密度としての性質を満足しなくなるが、 $k(x, y, h) / \iint k(x, y, h) dx dy$ と正規化するか、

$$k(x, y, h) = \frac{1}{n \sum_{i=1}^n w_i} \sum_{i=1}^n k_i(x, y, h) \quad (4)$$

とすることによって確率密度として利用可能である。この拡張により、個々の顧客の重みを考慮して密度推定を行うことができる。

さらに、顧客の空間的分布は、季節変動や曜日変動といった時間軸にそって変動する動的な分布と考えられる。そこで、カーネルは時点 t ごとに異なるものとし、カーネル関数を

$$k_i(x, y, h_t) = \frac{w_i}{2\pi h_t^2} e^{-\frac{1}{2h_t^2}[(x-x_i)^2 + (y-y_i)^2]} \quad (5)$$

と定義する。この時点 t は、便宜上、連続変数ではなく、日単位、月単位や曜日単位ごとに離散的にとらえることにする。これはパラメータ h をいくつかの時点に分割して推定するということだけであるが、顧客分布の時系列的な空間変動をとらえることができる点で意義がある。

この密度分布の形状はパラメータ h_t のみに依存するので、本稿では最尤法を用いたクロスバリデーションにより分布を特定する。すなわち、

$$\hat{k}(x, y, h_t) = \frac{1}{n-1} \sum_{i \neq t} k_i(x, y, h_t) \quad (6)$$

と定義し、対数尤度の平均

$$CV(h_t) = \frac{1}{n} \sum_{i=1}^n \ln \hat{k}(x_i, y_i, h_t) \quad (7)$$

を最大にする h_t を時点 t ごとに求める。

2.2 顧客分布の時空間変動

以上の分析手法を、ある百貨店の顧客別 POS データに適用した結果を示す。用いたデータには顧客の存在位置を示す座標値（需要点）が含まれているため、カーネル関数の重み $w_i (i=1, 2, \dots, 5030)$ として各需

要点 i ごとの購買額を用いた。全データを用いた全体の密度推定とともに、時点 t としては月ごと、曜日ごと、来店の最初の購買がいつであったかを意味する初回購買時間帯ごとの3区分について、それぞれ密度推定を行った。推定したパラメータ値 h_t は図2のとおりである。パラメータ h_t の値は、月別では冬場（2月）と夏場（8月）に大きくなり、曜日では週末に小さくなる傾向がある。初回購買の時間帯では、日中と夕食時に小さくなり、その間の17時台が最大になっている。これら推定値を基に顧客の購買額密度分布を視覚化した。その一つを図4に示す。図4はデータ全体を表現したものであり静止画であるが、月、曜日、初回購買時間帯ごとに推定した結果は動画として、市場における顧客分布の時空間的な変動を把握することができる³。

この視覚化した結果と、顧客の平均購買金額や平均距離（図3）と照合すると、冬場・夏場は中央部の密度が下り商圈が狭域になること、週末は意外にも平日に比べ商圈が狭域になるが中央部の密度は高くなること、1日の営業時間帯内でも商圈が変動していることなどの知見を得た。週末の売上は平日よりも高いので、常識的には週末の方が商圈が広いと考えられる。しかし、この事例では図3に示されるとおりに土日の平均距離は平日よりも小さくなっており、週末の売上は中央部の顧客の購買額の増大（客単価増）によって形成されている。この平日と週末の商圈の違いと夕方以降の商圈拡大には、図4の尾根状の部分の斜面が高くなることから、通勤行動の影響が寄与している可能性が高く、商圈よりも通勤圏の方が広い可能性がある。

式(1)に示されるカーネル関数を使用した従来からのカーネル密度推定の結果は、紙面の都合上省略するが、概して拡張したカーネル密度推定の方が中央部の密度が強調される形状を示している。これは節4にて明らかになるとおり、購買額が高く買物頻度も多い近隣居住者の影響を反映していることによる。

3. 顧客の店舗内滞留時間と購買額の関係

本節では、顧客の店舗内滞留時間に着目し、その購買額との関係を考察する。

顧客の店舗内での購買行動のタイプでは、ブランドレベルでの購入予定に従い予定ブランドを購入する計画

³ 顧客分布の月、曜日、入店時間帯ごとの時空間変動は <http://www.ms.kuki.tus.ac.jp/KMSLab/txf/corsj/> を参照。

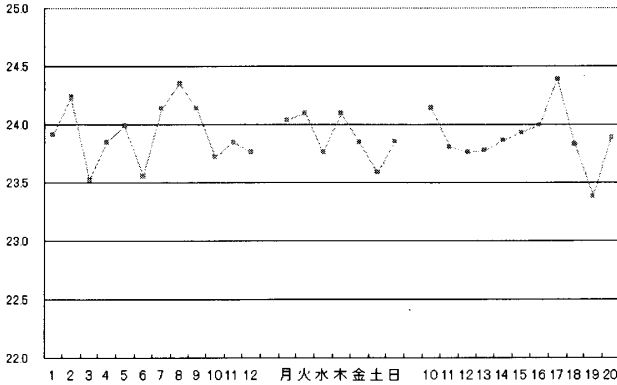


図2 パラメータ h の推定結果

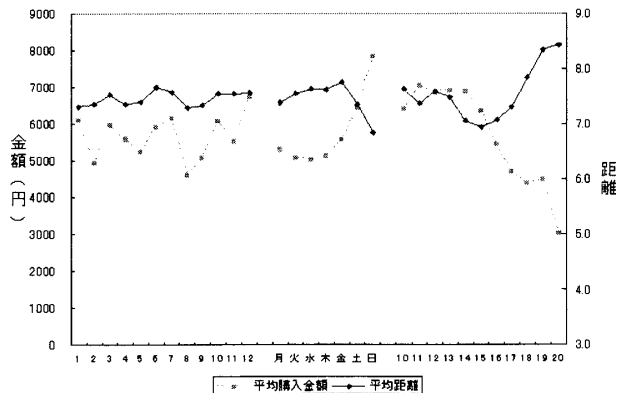


図3 平均購買額と平均距離の関係

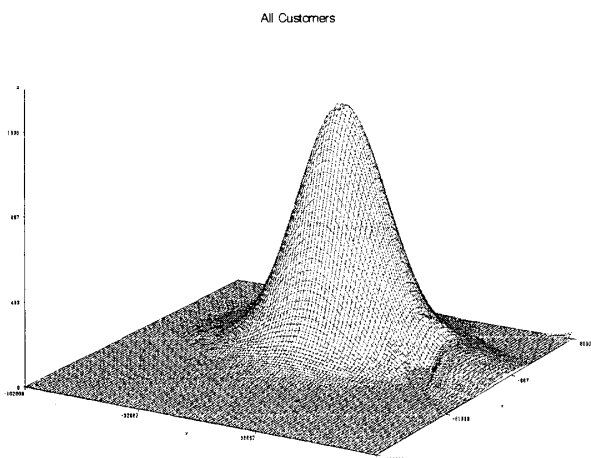


図4 購買額密度分布（全体）

購買の割合は1割程度といわれており、これ以外は想起購買、関連購買、衝動購買などの（広義の）非計画購買であるといわれている[5]。計画購買は顧客の商品購買に対する意思決定が入店前になされているのに対し、非計画購買は入店後になされる点に違いがある。このため店舗側は、売場内広告や関連陳列などによって非計画購買を高める努力を行っている。この非計画購買は顧客が店舗内で長く回遊しているほど、商品やサービスに接する機会が増え、これにより購買額が高くなると考えられる。顧客の店舗内での回遊性を高めることは店舗内での滞留時間を高めることにも関係しており、この滞留時間と購買額の関係について考察しておくことは、顧客単価を見る新しい視点になる可能性があり、店舗の売上増大戦略を立案する上でも十分な意義があると考えられる。

3.1 滞留時間と購買額の関係

まず、ある百貨店の顧客別 POS データから顧客の店舗内滞留時間と購買額の間を調べる。顧客が店舗に来店してから退店していくまでの正確な滞留時間を知ることはできないが、この取引データには、各顧客の商品購買の時刻が商品ごとに記録されており、最初と最後の購買時刻の差からおおよその店舗内滞留時間を測ることができる。もちろん、この店舗内滞留時間には来店から最初の購買までの時間と最後の購買から退店までの時間が含まれていないが、この過小評価分はデータの限界によって把握できないので、最初と最後の購買時刻の差を店舗内滞留時間と呼び、以降の議論で用いる。

図5に月別・曜日別の顧客の平均購買額および平均滞留時間の推移を示す。平均購買額の増える7月、10月、12月と週末に滞留時間は顕著に長くなっている。このように平均滞留時間の推移は、月別・曜日別とも

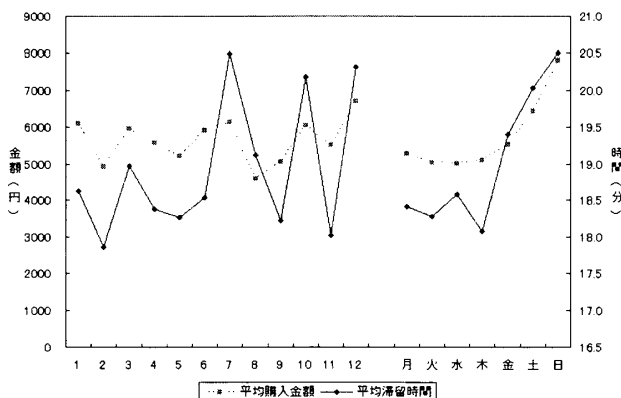


図5 月・曜日別の平均購買額・滞留時間の推移

に平均購買額の推移とおおむね連動しており、滞留時間が長くなると購買額が増えるという比例関係が成り立っているように見える。これを受け、平均滞留時間と平均購買額の間を関係を図示したところ、図6のような結果を得た。滞留時間と購買額は単純な比例関係ではないことがわかる。滞留時間が長くなるに従って、度数が少なくなるためばらつきがあるものの、滞留時間100分程度で平均購買額が最大になる性質を有している。顧客が購入する商品の品数・点数や年齢の平均を滞留時間ごとに求めると、これらは滞留時間に対してほぼ一定の値となることから、これらとは特別な関係はないと推察される。さらに、滞留時間と購買額の間を、性別、月別、曜日別に分けても、ピーク時の購買額に差があるものの、いずれの場合も100分程度で購買額が最大となり、曲線の形状も類似していることがわかった。

これは、「滞留時間100分の法則」とでも呼べるような、ある種の普遍的な関係である可能性があるため、モデル分析を行い考察する。モデルを用いるのは、他の小売店舗でも購買額や滞留時間の差こそあれ、この関係が観察される可能性があり、顧客単価を分析する新しい指標になる可能性があるからである。また、モデルで扱えれば POS データが与えられなくとも少ないサンプルでこの関係を推定することができる上、新たな店舗評価指標としての役割も期待できるからである。

3.2 モデル分析

図6に示されるとおりに購買額は滞留時間が100分程度のときにピークをむかえる。この時間までは滞留時間が長いほど購買額も多くなるという関係になっている。これは、多くの商品を購入するほど、その購買行動に時間を要するという常識と合致する。これには、

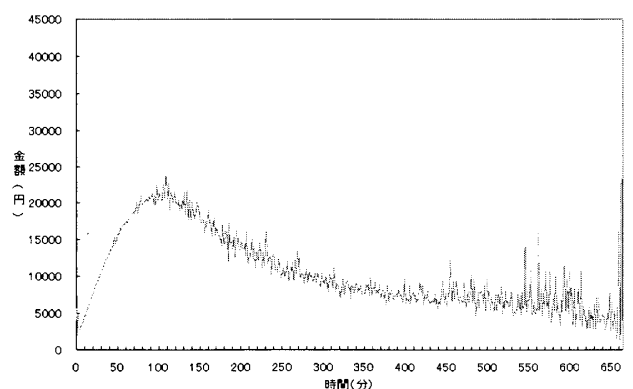


図6 平均滞留時間と平均購買額の間

計画購買で来店した顧客が店舗内に長く滞留しているほど多くの商品にふれることができ追加的に購入してしまうという効果も含まれていると想定できる。しかし、グラフは単峰型をしており、100分以降は変動しながら購買額が減少する傾向になっている。また、300分以上の常識的には考えられないような長い滞留時間が存在している。これをどのように解釈するかがモデル分析を行う上で問題になる。

度数ではピーク時をむかえる100分までで全体の約95.5%を占めており、残りは約4.5%ほどでしかない。このデータの少なさが100分以降でのばらつきの原因の一つであるが、他にも最初と最後の購買の間に他店を探索したり職場に戻ったりする店舗内に滞留していない退店時間の存在も無視できない。実際、滞留時間ごとの平均距離を調べると、600分以降では極端に短くなっている部分が存在した。これは百貨店の近隣に居住する顧客が日に2回（あるいはそれ以上）来店したと考えられ、大半は退店時間であると推察される。

このような退店時間が含まれているとすると、定義した滞留時間から退店時間を除くと、ピーク時より右側の分布は左に圧縮され、実際は正規分布になっていると考えることができる。したがって、ここでは時点 t における購買額 $n(t)$ は

$$n(t) = m \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \right] \quad (8)$$

と平均 μ 、標準偏差 σ の正規分布で与えられるものとする。ここで、 m は購買ポテンシャルである。

正規分布を仮定するもう一つの理由に、図6の変動は滞留時間70分程度から小さく変動しはじめているので、ここにも退店時間がわずかながら含まれ、実際のピーク時点は100分よりも短い可能性があり、このピーク時点を推定するためでもある。

この推定のためには、式(8)の m 、 μ 、 σ を求めればよく、3時点のデータさえ与えられれば推定することができる。ここでは時間間隔 j における購買額 $n(t_j)$ とし非線形最小二乗法で推定する。推定に当たっては、時間間隔を1分単位とし、比較的ばらつきが少ないピーク時の左側 ($j=0, 1, \dots, 100$) のデータを用いて、事前にグリッド検索により初期値を求めた上で推定を行った。ピーク時の右側のデータを用いないのは、度数も少なく、先述したとおりに退店時間の影響を受け、信憑性に欠けると判断したからである。

百貨店の顧客別 POS データに対し、性別、カード種別、月別、曜日別、時間帯別（午前12時まで、

表1 パラメータ推定結果

区分		m	μ	σ	$n(\mu)$
全体		2476243	87.5	47.7	20711.6
性別	男性	2456491	87.4	47.5	20648.7
	女性	2597094	88.2	49.1	21109.5
カード種別	ポイント	2271134	87.1	46.8	19378.8
	割引	2801177	84.1	48.1	23252.8
月	1	2296149	81.7	45.0	20360.3
	2	1984441	82.2	46.5	17018.3
	3	2818348	92.7	49.8	22581.1
	4	2396932	86.1	46.8	20452.1
	5	2341697	86.2	45.8	20398.8
	6	2530920	85.4	44.9	22481.4
	7	2496037	86.9	48.3	20613.0
	8	1946205	85.5	47.9	16192.7
	9	2279373	85.7	46.0	19767.4
	10	2655619	89.6	48.0	22079.3
	11	2879487	94.9	51.5	22309.6
	12	2807581	89.8	51.1	21933.3
曜日	月	2342012	86.9	47.1	19854.7
	火	2186338	84.0	45.8	19030.0
	水	2186764	84.5	45.6	19130.7
	木	2164251	82.8	45.0	19168.7
	金	2589827	92.5	50.7	20367.6
	土	2680659	88.0	47.6	22448.0
時間帯	日	3280652	94.3	54.4	24065.5
	午前	2400427	82.2	45.2	21168.1
	午後	2661303	90.6	50.4	21063.4
	晩	2180751	84.4	45.3	19190.5

午後は17時まで、晩は17時以降)にそれぞれパラメータを推定した結果を表1に示す。

図6に示した曲線の形状は、集計区分を変更しても類似しており、 μ と σ の間には大差がないが、ピーク時を表す μ はいずれも100分を下回っている。後述する購買パターン分析の図8に示す通りに、購買する売場が一つ増えると滞留時間は30分程度増える傾向があることから、最初の購買が行われるまでの過小評価分を加えた実際の滞留時間のピーク時が100分程度であると推察される。また、曜日の区分では日曜日に m と μ が大きくなっている。これは次節の購買パターン分析で明らかになるとおりに日曜日の購買パターンは平日に比べ変化するためである。購買ポテンシャル m は売上の変動と類似した推移を示している。興味深いのはカード種別で、ポイントカードを利用している顧客よりも割引カードを利用している顧客の方が購買ポテンシャルが高くなっている点である。これは、ポイントの蓄積よりも値引きの方が顧客の購買に与える影響が大きいことを示す一例である。

以上のように顧客の購買額は滞留時間100分程度で最大になる性質を有している。しかし、購買の度数で

見ると、レジを1回だけ通る滞留時間0分という購買が全体の約58%を占めている。関連購買をいかしたプロモーション、顧客を誘導するレイアウトや店舗内で休息できるスペースを確保するなど、店舗内により長く滞留してもらい、もう1品でも多く購入してもらうようにすれば売上が伸びるはずである。では、これから売上に貢献する購買パターンはどのようなものか、次節にて分析する。

4. 購買パターン分析

用いた顧客別 POS データには、売場号館 (A 館, B 館, C 館) と売場階の情報があり、これを用いて売場単位で購買パターンを調べた。取引明細ファイルを、顧客、日付、時刻ごとにソートした上で、一つの商品の購買を記録した各レコードの売場号館と売場階の文字列を結合し、それを顧客の1日の購買単位で1レコード (列を行に変換) にまとめるレコード圧縮操作により購買パターンを抽出した。抽出したパターンの格納されているフィールドにソートをかけて分析すれば、どの売場を選択したかの組み合わせ購買パターンになり、ソートをかけずに分析すれば購買順序を反映した順列購買パターンとなる。

4.1 組み合わせ購買パターンの分析

顧客が1回の買物で購買した売場の組み合わせを調べると14352パターンを抽出することができた。最大は11で、3売場選択から6売場選択のパターン数が多くを占めている。売上で見ると、「BBX」(16.0%)、「A1」(8.2%)、「A1 BBX」(6.9%)、「A2」(3.1%)、「A6」(3.0%)、「A5」(2.3%)という順に食料品を中心とした単一売場選択が上位を占めている (括弧内は売上に占める割合、「BX」は地下階を表す)。これらパターンの平均購買額と平均滞留時間の分布は図7のようになる。この分布は前節の滞留時間と購買額の関係と類似しており、購買額と滞

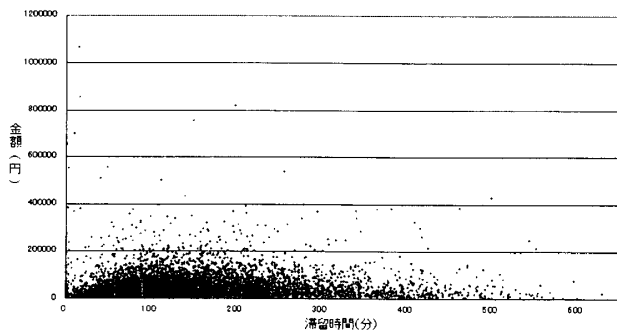


図7 購買パターンの平均滞留時間・購買額

留時間は密接に関係していることがわかる。

次に、顧客単価が高く買物頻度も多い有力な購買パターンを調べるために、購買パターンの売上高でデシル分析を行ったところ、表2の結果を得た (1デシルは1435パターン)。デシル1だけで全体の売上の94.6%を占めている。顧客の個人レベルでの購買行動を把握するためには、デシル1に関して、さらに個別に調べるべきであるが、1435パターンもあるため、度数が1000以上で平均購買額が1万円を超えるパターンに絞り込んだ。これは顧客単価が高く買物頻度も多い購買パターンを想定している。この条件で抽出すると、全部で88パターンであった。この88のパターンはパターン数全体では0.6%にすぎないが、全買物度数のなかでは16.3%を占め、売上では40.2%を占める有力な購買パターンである。これを滞留時間との関係で図示すると図8のようになる。図中に楕円で囲ったとおり、これら88の有力な購買パターンは、単一売場選択、近接2売場選択、遠接2売場選択、3売場選択と四つのグループに分かれて分布している。商品を購入する売場が一つ増えると滞留時間も30分ほ

表2 購買パターンのデシル分析

デシル	金額(万円)	割合(%)
1	2295804	94.6
2	56343	2.3
3	26524	1.1
4	16244	0.7
5	11030	0.5
6	7881	0.3
7	5705	0.2
8	4124	0.2
9	2820	0.1
10	1468	0.1
	2427944	100.0

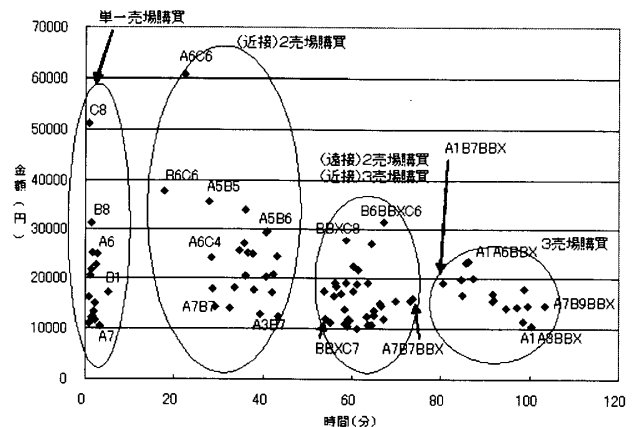


図8 有力な購買パターンの平均滞留時間・購買額

ど増加する傾向がある。購入する商品の点数も平均で2.7点、5.5点、10.1点、15.4点と増加する。しかし、節3で観察した滞留時間100分程度で購買額が最大になる性質は、この有力な購買パターンでは成り立たず、むしろ滞留時間が100分程度になる3売場選択の購買額はそれほど高くない。

さらに、これら購買パターンは月ごと曜日ごとにどのようなになっているかを調べた。2売場以上を選択する購買パターンを売上順にならべると表3および表4のとおりになる（紙面の都合上、上位3パターンとした）。売上の高い7月、10月、12月や週末のパターン数は多くっており、購買パターン数は売上と連動していることがわかる。また、月・曜日ともに「B7 BBX」（家具・寝装品と食料品）が1位であり、「B6 BBX」（婦人服と食料品）、「A2 BBX」（婦人服飾と食料品）が上位を占めている。しかし、週末には「A6 BBX」（紳士洋品と食料品）が上位になり、パターンが変化する。平均購買額を見ると「B7 BBX」は平日では1万円程度であるが、週末になると約5万円になる。日曜日に2位になる「A1 B7 BBX」も平日は1万円程度であるが日曜日には3万5千円になるなど顧客単価が週末は高くなっている。週末のみに上位になる「A6 BBX」の平均購買額は

表3 月ごとの売上順位

月	パターン	1位	2位	3位
1	3335	B7 BBX	B6 BBX	A6 BBX
2	2753	B7 BBX	A7 BBX	B6 BBX
3	3655	B7 BBX	A5 BBX	B6 BBX
4	3386	B7 BBX	A2 BBX	A6 BBX
5	3403	B7 BBX	A2 BBX	B6 BBX
6	3780	B7 BBX	A1 B7 BBX	A6 BBX
7	4649	B7 BBX	A1 B7 BBX	A6 BBX
8	3552	B7 BBX	A1 B7 BBX	A2 BBX
9	3245	B7 BBX	A2 BBX	A7 BBX
10	4641	B7 BBX	B6 BBX	A5 BBX
11	3343	B7 BBX	A2 BBX	A1 B7 BBX
12	5009	B7 BBX	A1 B7 BBX	A6 BBX

表4 曜日ごとの売上順位

曜日	パターン	1位	2位	3位
月	4714	B7 BBX	A2 BBX	B6 BBX
火	4519	B7 BBX	A2 BBX	B6 BBX
水	4339	B7 BBX	A2 BBX	B6 BBX
木	4359	B7 BBX	B6 BBX	A2 BBX
金	5668	B7 BBX	A2 BBX	B6 BBX
土	5991	B7 BBX	A6 BBX	A2 BBX
日	6082	B7 BBX	A1 B7 BBX	A6 BBX

曜日に関係なく1万7千円程度であり、週末の客数の増加によって上位にランクしていることがわかる。「A6」が含まれる購買パターンは客単価が高いため、平日にこの購買パターンを行う客数を増やすことができれば売上はさらに向上すると考えられる。

4.2 順列購買パターンの分析

以上のパターン分析は組み合わせを調べたため、顧客が購買した売場の順序を考慮していない。売場の選択の順序が購買額に影響を与えている可能性もあるため、順列購買パターンについても調べた。購買した順序も考慮すると35270パターンが抽出できた。

この順列パターンの概要を示すために、顧客はどの売場から購買を始めてどの売場で購買を終えるかをIn-Out分析により調べた。図9はIn-Outの流れを買物度数の割合で示した図である。図10はIn-Outを売上に占める割合で示した図である。買物度数で見ると食料品（「A1」および「BBX」）の購買が大多数を占めているため、「BBX」、「A1」、「BBX A1」、「A1

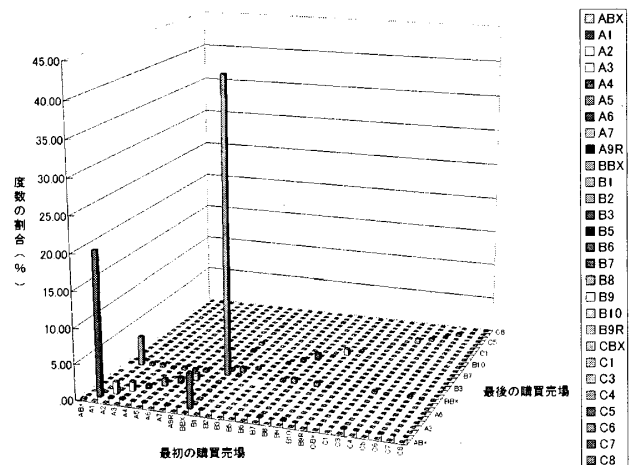


図9 In-Out分析（買物度数の割合）

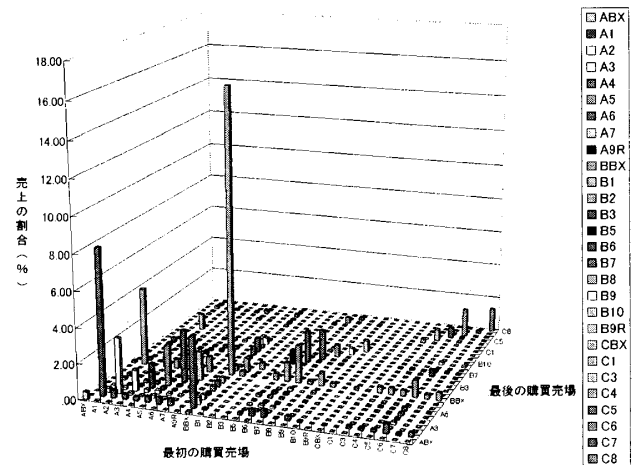


図10 In-Out分析（売上の割合）

BBX」の4パターンが突出している。これを売上に占める割合で見ると、これら4パターンは買物度数が多いため、売上に占める割合が高くなっているが、他の顧客単価の高いパターンも観察できるようになる。例えば、「A2」から「A7」, 「B6」, 「B7」, 「C6」, 「C8」といった単一売場選択や「A1」と「BBX」を選択する購買パターンである。購買順序の影響を調べるために、売上に占める割合を比べると、低層階の売場で買物を終えるパターンでは最初に買物をするパターンよりも売上に占める割合が高くなっている。例えば、「A1」で買物を始めるパターンは15.1%であるのに対し、「A1」で買物を終えるパターンでは17.1%となっている。これら低層階で買物を終えるパターンを除けば最初に選択されるパターンの方が売上に占める割合が高くなっている。

先に抽出した有力な88の購買パターン中で2売場以上選択したパターンで調べると、ほとんどのパターンでは売場の選択順序によって購買額が変わるといった現象は観察できなかった。どちらを先に選択しても平均購買額はほぼ同じであり、パターンごとに各購買の相関を見ても差異は確認できなかった。

以上をふまえると、全体としては低層階を最後に選択するように顧客を誘導する必要があると考えられるが、顧客単価の高い有力な購買パターンでは順序はその購買額にあまり影響していないと考えられるので、順序よりもこの購買パターンを選択してもらえるように務める必要があると考えられる。

4.3 顧客の空間分布との関係

先に有力な88の購買パターンを取り上げたが、この購買パターンを行う顧客はどこから来店しているのだろうか。節2で取り上げた顧客の空間分布との相互

関係を述べておかなければならない。この購買パターンを行う顧客を節2の手法で空間に展開すると図11のようになる。これは図4に示した百貨店の購買額密度のコアを形成している。これは有力な購買パターンを行った顧客を空間に展開したものであるが、顧客全体の購買額密度面と同じような形状をしている。この密度面を形成しているのは主に中央部周辺の顧客であり、これら顧客の特徴は、年間の買物頻度が多く、相対的に高齢である。

あわせて、年間の総買物頻度を調べたが、図11の分布よりもさらに狭域になっている。有力な購買の空間的な分布は、買物頻度の空間的な分布とも類似しており、頻度の少なくなる円環型地域の顧客がこの購買パターンを行ってくれるようになると、売上が向上すると考えられる。小売システムの最上位にある百貨店の商圈は他の業態よりも広域であるが、売上のコアを形成するのは近隣の顧客であり、重視すべき顧客である。以上のように、一見すると無関係と考えられがちな顧客の空間的分布と購買パターンは密接に関連していることも見いだせた。

5. おわりに

本稿では、ある百貨店の顧客別POSデータに対してデータマイニングとモデル分析を行った結果を報告した。

まず、顧客分布の時空間変動を把握するための密度推定手法の提案では、週末よりも平日の商圈の方が広く、週末の売上は平日の商圈を維持できれば向上させることができること、冬と夏には店舗近隣の居住者に対するアプローチが重要なことを明らかにできた。この手法を用いると、顧客属性(性別、年齢、など)や売場(紳士服売場など)ごとの顧客密度を時空間変動をまじえて比較・評価することや買物頻度やアクセシビリティとの関係を分析することによって、近隣効果が購買に与える影響を調べることもできると考えられる。

また、顧客の店舗内滞留時間と購買額の関係について調べ、滞留時間100分程度で購買額が最大になる性質を見いだした。これを受け、モデル分析を行い、このモデルが店舗内普及効果の予測モデルとしての役割を担える可能性や複数店舗間の新たな評価指標を提供できることを示唆した。その実証分析の結果からは、ポイントカードより割引カードの方が売上への貢献度が高いことがわかった。

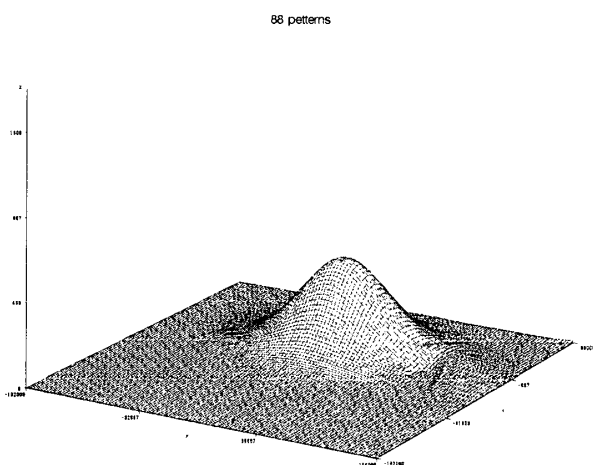


図11 有力な購買パターンの購買額密度分布

さらに、これらマクロ的なアプローチからの結果を詳細に調べるために、購買パターン分析を行い、有力な購買パターンの抽出や売上向上に結びつくと考えられる購買パターンについて言及した。

しかし、いくつかの課題が残された。その多くは時間的制約によるものであるが、特に、節3で述べた店舗内滞留時間と購買額の関係のモデル化が完成に至っていないことが課題である。当初は計画購買を中心とした顧客の影響とこの計画購買よりも非計画購買のウェイトが高い顧客の影響を明示的に説明できるモデルの構築を目指していたが、時間の過小評価分や退店時間の存在で滞留時間を正確に把握することができなかったからである。このモデル化を今後すすめていく必要がある。

謝辞 本研究を進めるに当たり、若狭宏嗣君と菅原崇君にはデータ解析作業の一部を手伝っていただいた。

また、匿名の査読者からは有益なるご意見を頂戴した。ここに記して謝意を表す。

参考文献

- [1] N. Donthu and T. R. Rust: "Estimating Geographic Customer Densities Using Kernel Density Estimation", *Marketing Science*, 8, 191-203, 1989.
- [2] N. Donthu: "Comparing Market Areas Using Kernel Density Estimation", *Journal of the Academy of Marketing Science*, 19, 323-332, 1991.
- [3] 守口剛: 「マーケティング・サイエンスにおける今後の研究の方向」, 『オペレーションズ・リサーチ』, 48, 507-515, 2003.
- [4] B. W. Silverman: *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, 1986.
- [5] 渡辺隆之, 守口剛: 『セールス・プロモーションの実際』, 日本経済新聞社, 1998.