

# スーパーマーケットの POS データに基づく 消費者の購買行動モデル構築

○下田稜 小山竜平 杉浦孝典 島田匠都 村井詩音 (静岡大学)

矢田勝俊 原田拓弥(関西大学) 李皓(静岡大学)

## Construct Consumer Purchasing Behavior Model by POS Data of Supermarkets

\* R. Shimoda, R. Koyama, T. Sugiura, T. Shimada, S. Murai (University of Shizuoka)

K. Yada, T. Harada (University of Kansai) and H. Lee (University of Shizuoka)

**概要**— 近年、小売店では POS データの活用が一般的となっている。活用方法によっては競争優位性が高まる為、POS データの扱いは重要である。本研究では、消費者の購買行動モデルを用いたシミュレーションを行うことを目的とし、POS データを基に世帯構成や購買傾向の推定を行い、エージェントベースの購買行動モデルを構築する。

**キーワード:** POS データ, 購買行動, モデル構築, シミュレーション

### 1 はじめに

線分析への一般的な有用性を明らかにすることは困難であった。

#### 1.1 背景

近年、商品についているバーコードをレジで読みとる行為は日常的なものとなっている。スキャナーを通じて取得した販売量や金額はスキャナーデータと呼ばれ、POSデータとスキャン・パネルデータに分けられる。POSシステムが存在する店舗では、消費者の単品購買動向が時系列に沿って記録されるPOSデータは、コンビニエンスストアから始まり、現在ではあらゆる小売店で活用されている。商品の売れ筋は、品目毎の販売データから容易に集計できる為、当初から売れ筋商品に焦点を当てて在庫管理や商品整理によく活かされているが、マーケティングの側面では更に効果的に利用することができる。マーケティングにおいて、優良顧客や顧客個人の購買記録に注目することは重要なことであり、そこで必要となるのがマーケティング指向のPOSデータの活用である。

あらゆる分野でデータの活用が期待されている社会において、製造業あるいは流通業内で競争優位に立つ為には、取得した膨大なPOSデータをいかに活用するかが鍵となる。集計したデータを基に仮説モデルを構築し、分析を繰り返しながらモデルの精緻化と知見を追求することで、他企業に対してポジショニングを行うことが出来る。

#### 1.2 先行研究

スーパーマーケットにおける消費者の購買行動モデルに関する既存研究として、矢田<sup>1)</sup>は既存の文字列解析手法を RFID によって得られた顧客動線に関するストーミングデータに適用させ、顧客の販売エリアへの訪問パターンを考察した。顧客動線データのうち売場への立ち寄り行為に焦点を当て、訪問パターンを文字列で表現することで、膨大なストリームデータを効率良く扱うことを提案したが、文字列解析技術の顧客動

また、矢田ら<sup>2)</sup>は顧客動線を用いた購買行動分析や状態空間モデルを用いた分析も行った。これらの分析により、顧客の回遊の複雑さが購買行動に与える影響を定量的に評価し、さらにはスーパーマーケットの業務改善に必要な洞察を得た。消費者行動にスケール階層の構造を見出す為、顧客動線データと ID 付 POS データを利用して実際の小売店内の消費者行動を考察した。

山田ら<sup>3)</sup>は、消費者の行動を考慮した店内レイアウト決定という問題に対してマルチエージェントシミュレーションを用いて、店舗内人流シミュレータを構築した。POP 広告の設置や商品配置・店内レイアウトの変更による顧客動線・店舗滞留時間の変化を分析すること、そして滞留時間を伸ばす為の商品配置を考案する上での意思決定支援ツールとして活かすことを目的とした。その結果、POP の設置や店内レイアウト変更による購買者の動線変化や、計画あるいは非計画購買者の滞留時間時間の延長が明らかになった。

POS データ分析に重点を置いた研究では、健康情報が消費行動に与える影響など、商品カテゴリーレベルでの商品需要分析が多く存在する。そのなかで川村<sup>4)</sup>は、加工食品を対象としたブランドレベルでの需要分析を行っている。ブランド毎に製品差別化が進んだ加工食品について分析を進める為、アメリカの小売店における POS データを扱った。具体的には、全米 65 地区の小売店での各商品の販売データのうち、マーガリンという品目に注目し、シェアが高い 5 ブランドを取り上げて需要体系推定を行った。その結果、ブランド間における価格の格差が消費者のブランド乗り換えに

与える影響が大きい事実を確認することができた。また、プライベート・ブランド品がブランドに対してこだわりのない消費者をターゲットにしていることを裏付けた。

### 1.3 本研究の目的

既存研究では、顧客動線やPOSデータを用いて店舗内での購買行動をモデル化する研究や、市場・需要予測を行う研究が多く存在している。しかし、POSデータに基づき購入店舗・商品の選択行動に至るまでの内部要因、さらには外部要因に着目し、エージェントベースで消費者の購買行動モデルを構築している研究は少ない。

そこで、本研究では消費者の購買行動モデルを用いたシミュレーションを行うことを目的として、あるスーパーマーケットの購買データ・顧客データ・店舗データ・商品データを用いた購買行動モデルを構築することを目的とする。消費者エージェントが食料品を購入する際、店舗までの距離や商品の価格、年齢や性別、そして家族構成などを考慮した上で店舗と商品を選択していることが考えられる。したがって、購買データを用いて世帯構成や購買傾向の推定、さらには消費者エージェントの商品購入シミュレーションを実施する。

## 2 データ分析

### 2.1 利用データ

本研究では、関西地方のあるスーパーマーケットのPOSデータを取り扱った。具体的な内容は以下の通りである。

#### [1] 顧客の購買データ

購入店舗、購入日時、顧客ID、レシートID、購入商品番号、商品名、数量、購入金額から構成されており、2015年10月から2016年9月(1年間)までの対象店舗の購買データを指す。

#### [2] 該当するスーパーマーケットの顧客情報

会員ID(顧客番号)、会員登録した店舗ID、会員の郵便番号、町丁目レベルの住所、性別、生年月日から構成される。

#### [3] 店舗情報

店舗ID、店名、駐車場の有無、所在地から構成される。

#### [4] 商品情報

取扱い商品について

商品ID、商品名、大分類、部門、中分類、小分類、商品をPOSへ登録した日が記載されている。

これらのデータを用いて、購買客に関する統計分析を行った。

### 2.2 分析手法

分析するにあたり、顧客の消費金額を3つの指標(総額、大分類、部門)で集計し、加えて3つの指標(年齢、性別、居住地と店舗間の直線距離)と顧客の購買行動(購入商品、消費金額)との相関を確認した。それらの集計結果を踏まえ、購買客各々がどういった顧客か(顧客の属性)を特定する。

分析内容の目的は、「優良顧客の抽出」と「顧客の属性特定」となっている。

まず、どの顧客が1年間のうちどれほどの金額を消費しているか、またどの商品をどの程度購入しているかといった具体的な値を把握する為、3つの指標(総額、大分類、部門)ごとに年間消費金額や平均消費金額を集計した。その際、性別や年齢という切り口も取り入れた。年齢に関しては、顧客情報の生年月日から年齢(2019年1月時点)を算出した。顧客年齢を5歳毎に区切り、階級分けを行った。また、それぞれの顧客が居住地からどの程度離れた店舗で購入したかを明らかにする為、購入店舗の住所(Googleマップ参照)と顧客情報に含まれる住所から距離を算出した。その後、顧客の消費金額と算出した距離の相関を確認した。

### 2.3 データの概要

2.1で述べたデータを用いて、性別という指標から顧客の来店頻度と使用金額の相関を確認した。赤は女性、青は男性を示す。

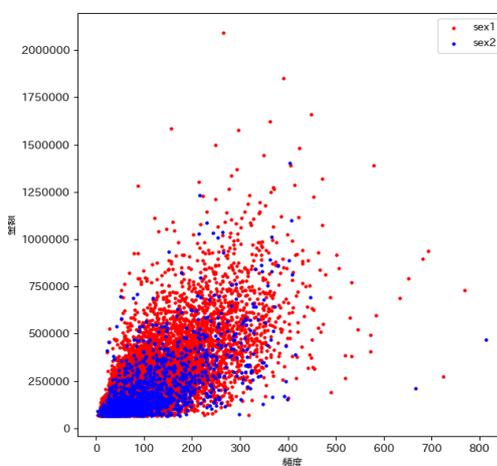


Fig.1 顧客の来店頻度と使用金額の相関

## 2.4 データ分析

集計データから優良顧客を抽出する為、デシル分析やRFM分析を行う。

そして、先述した5つの指標のうち3つの指標(総額, 大分類, 部門)を取り上げ, そこに性別・年齢を加え計5つの要素(総額, 大分類, 部門, 性別, 年齢)とした。これらの要素を用いて顧客の属性を特定する為に, 因子分析を行う。

はじめに, 「優良顧客の抽出」に関してデシル分析とRFM分析を行った結果を以下に示す。

購入金額に焦点を当てデシル分析を行った結果をFig. 2で示す。デシル分析では, 顧客を消費金額によって10段階にグループ分けした。Fig. 2では, 顧客全体の消費金額の約8割を, 購買客の上位10%が占めていることが分かった。(Fig. 2に示す) x軸がグループ, y軸が購入金額, 折れ線グラフが累積購入金額比率, そして棒グラフが顧客一人当たりの購入金額を表す。

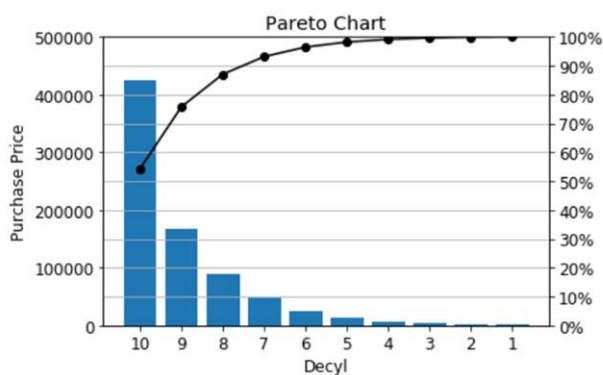


Fig. 2 購入金額で付けられたランク別の購入金額

RFM分析にあたり, 3つの評価要素(R:購買最近性, F:購買頻度, M:購買金額)それぞれでランク付けを行い, そこから総計ランクを算出する。ランク付けや, 算出する際の要素の重み付け(算出式)は, 企業の判断に委ねられている為, 一概には言えない。そこで本研究では, 評価要素の重要度分析に王ら<sup>5)</sup> が使用した算出式( $R*2+F+M$ )によって分析を行った。

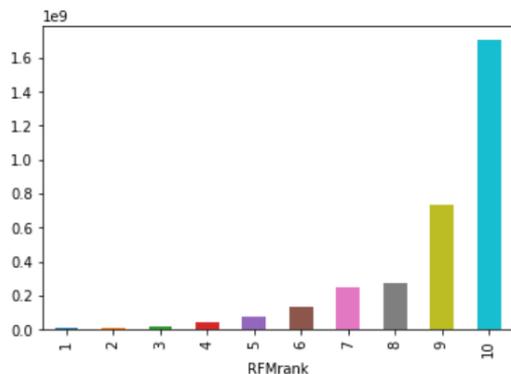


Fig. 3 ランク別の購買金額(M)割合

田中ら<sup>6)</sup> によれば, スーパーマーケット業界におい

て店舗売上の80%を占める, 顧客の20~30%という限られた層は優良顧客と分類される。本研究では顧客の上位30%にあたる約1万2千人を優良顧客とする。

また, RFM分析に際して抽出された3つの要素(最新購入日, 購入頻度, 総消費金額)は, 後述する属性特定に活用する。

次に, 顧客の属性のうち購買行動との関係性が深いものを特定する為に, 因子分析を行う, 今回は41476人の顧客のデータに対し, 統計ソフトRによって因子分析による顧客特定を試みた。使用した手法はプロマックス回転を用いた主因子法である。

商品カテゴリを表す部門(145種類)の中から, 特に購買実績が多かった28部門に加え, 先で述べたRFMrankを説明変数として設定した。

因子分析におけるの共通因子を3~7個で試行した結果, 5因子が5個の結果が最も解釈可能性が高かったため, 今回は5因子を前提とした分析を行った。

	Factor1	Factor2	Factor3	Factor4	Factor5
	生鮮食品	お惣菜	百円パン	アルコール	その他
精肉	0.877	-0.033	0.048	-0.024	0.072
鮮魚	0.829	0	-0.015	0.048	0.349
野菜	0.779	0.173	-0.029	-0.042	0.074
チキン	0.739	0.025	0.101	0.01	-0.109
塩干	0.642	0.12	-0.021	0.01	0.244
玉子	0.61	0.263	0	0.003	-0.077
シーズニング	0.591	0.416	-0.027	0.011	-0.122
金額	0.585	0.398	0.017	0.076	0.16
チルドフード	0.562	0.382	-0.015	-0.014	0.147
果物	0.54	0.234	-0.069	-0.076	0.265
ハム	0.539	0.284	0.047	0.003	-0.102
乳製品	0.461	0.305	-0.009	0.006	-0.169
ドライグッズ	0.445	0.53	-0.054	0.006	-0.075
菓子	0	0.693	-0.006	-0.031	0.028
パン	0.076	0.656	-0.006	-0.062	0.152
デリカ	-0.177	0.632	0.033	0.005	0.495
ソフトドリンク	-0.05	0.627	-0.043	0.059	0.135
頻度	0.2	0.606	0.059	0.076	0.174
和洋生菓子	0.084	0.53	0.046	-0.058	0.269
チルドドリンク	0.359	0.509	-0.013	-0.065	0.083
ティードリンク	0.253	0.497	-0.053	-0.017	-0.021
消耗雑貨	0.127	0.49	-0.106	0.057	0.089
寿司	-0.058	0.484	0.068	-0.002	0.416
アイスクリーム	0.077	0.476	0.043	-0.038	0.011
冷凍食品	0.166	0.457	0.068	-0.001	-0.07
RFMrank	0.197	0.451	0.11	0.04	0.055
百円パン売上仕入	0.027	0.009	0.877	-0.005	0.02
百円パン直営	-0.003	0.024	0.872	0.005	0.028
家庭雑貨	0.204	0.338	-0.088	0.043	0.051
リカー-SM	0.02	0.023	0	0.981	0.013

Table.1 各説明変数と共通因子の因子負荷

これらの結果を受けた, 我々は以下のように因子名を定めた。

Factor1の因子負荷量は精肉や鮮魚などの調理する部門が高かったため, 生鮮食品をよく買うような顧客が想定される。

Factor2はパンやデリカなどの調理不要の因子負荷量が高かったため, お惣菜類をよく買うような顧客が想定される。

Factor3は調理パンとの因子負荷量が高かったため, 調理パンをよく買うような顧客が想定される。

Factor4はリカー-SMとの因子負荷量が高かったため,

アルコール類をよく買うような顧客が想定される。

Factor5は突出した因子負荷量を示さなかったためその他とする。

	Factor1	Factor2	Factor3	Factor4	Factor5
因子寄与	5.822	5.219	1.61	1.009	0.994
因子寄与率	0.182	0.163	0.05	0.032	0.031
累積寄与率	0.182	0.345	0.395	0.427	0.458

Table.2 各 Factor の因子寄与率

### 3 考察

#### 3.1 「優良顧客の抽出」に関する分析について

RFM 分析を行う過程で、3つの評価要素それぞれのランクによって顧客を評価した際、購買最近性(R)のランク値が他の2要素に対し独立しているように思われた。そこで、「購買頻度は独立した評価要素である」という仮説を検証する為に、3つの評価要素の相関を確認した。

しかし、購買最近性 (recet\_rank) と購買頻度 (freq\_rank) の相関かつ購買最近性 (recet\_rank) と購買金額 (money\_rank) の相関は強いことが判明した。3つの評価要素それぞれの相関を Fig. 4 に示す。

	freq_rank	recet_rank	money_rank
freq_rank	1.00	0.75	0.92
recet_rank	0.75	1.00	0.72
money_rank	0.92	0.72	1.00

<Fig. 4> 評価要素間の相関

また、Fig. 4 にも示されたように購買頻度と購買金額の相関は非常に強い。したがって、本研究が利用した POS データでは、購買頻度と購買金額はほぼ同一視しても良い。すなわち、このスーパーマーケットでは、購買金額が高い顧客のほとんどが高い頻度で購買行動を行う。

#### 3.2 「顧客の属性特定」に関する分析について

因子分析の結果では、Factor1 は専業主婦がいる多人数世帯に直結すると考えられ、Factor2 は単身世帯や共働き世帯に関連する。また、Factor3 は洋風の朝食スタイルを好む因子であるため、比較的世帯メンバーが若いと思われ、Factor4 のアルコール購入量は、世帯にストレスが多いメンバーの存在を示唆できる。

これらの因子分析の結果をもとに、顧客の世帯成員推定ができると思われる。

## 4 おわりに

購入店舗・商品などの選択行動モデルを構築する。その為には、顧客の所在地と店舗間の距離、天気、曜日、季節、ポイントアップデーなどの外部要因に着目した上で、消費者エージェントベースの購買行動モデルを構築、さらには消費者エージェントの商品購入シミュレーションを実施する。その詳細については、研究会の際に説明する。

### 参考文献

- 1) 矢田勝俊：スーパーマーケットにおける顧客動線分析と文字列解析，統計数理，第56巻 第2号，pp199-213,統計数理研究所(2008)
- 2) 金子雄太, 矢田勝俊：スケールの階層性から探るスーパーマーケットの消費行動，オペレーションズ・リサーチ, Vol62, pp807-814(2017)
- 3) 山田健司, 阿部武彦, 木村春彦：計画・非計画購買者を考慮した店舗内人流シミュレーション，第19回人工知能学会全国大会論文集(2005)
- 4) 川村保：加工食品のブランドレベルでの需要分析—POS データ分析—，農業経済研究，第71巻 第1号，pp28-36(1999)
- 5) 王 于佳, 高 弘昇：RFM 分析における評価要素の重要度分析，NAIS Journal, Vol.10, pp51-56(2015)
- 6) 田中孝昌, 濱口智大, 西郷拓海, 津田和彦：スーパーマーケットの店舗別販売傾向と RFM 分析を利用した優良顧客分類，人工知能学会，第20回知識流通ネットワーク研究会(2017)
- 7) 阿部誠, 近藤文代：マーケティングの科学—POS データの解析—，朝倉書店(2005)
- 8) 総務省：社会の様々な分野で利用が始まったビッグデータ，平成25年版 情報通信白書(2013)  
<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h25/html/nc113320.html>
- 9) 石垣司, 竹中毅, 本村陽一：日常購買行動に関する大規模データの融合による顧客行動予測システム—実サービス支援のためのカテゴリマイニング技術—，人工知能学会論文誌, 26巻 6号 D, pp670-681(2011)