

商品重量を用いたスーパーマーケット再来店予測モデルの構築

Model building for prediction of Returning Supermarket Shoppers

田中孝昌¹ 津田和彦² 濱口智大¹ 西郷拓海¹ 西巻孝祥¹ 高橋理沙¹

Takamsa TANAKA, Kazuhiko TSUDA, Tomohiro HAMAGUCHI, Takumi SAIGO,
Takayoshi Nishimaki, Risa Takahashi

¹株式会社 NTT データ, NTT DATA CORPORATION

²筑波大学大学院ビジネス科学研究科

Graduate School of Business Science, University of Tsukuba

要旨：近年,我が国のスーパーマーケットチェーンの経営において顧客来店数の減少が課題となっている。その背景には,人口の減少および高齢化による食品市場の縮小,インターネット通販の普及による食品購買手段の多様化などがある。従来,スーパーマーケットチェーンは大規模な新規出店による経営拡大戦略を採ってきたが,このような外部環境の変化に直面したことにより戦略の転換を迫られている。具体的には,変化する既存顧客のニーズを捉え,継続的な来店に繋げる必要がある。そこで,本研究では複数のスーパーマーケットチェーンの会員を対象に再来店の予測モデルを構築する。モデル構築には,伝統的な顧客分析手法と商品重量を組み合わせた新たな手法を提案し,提案手法がチェーン個別の消費者ニーズの発見に有用であることを示す。

キーワード：スーパーマーケット,RFM 分析,ロジスティック回帰

1. はじめに

1-1.スーパーマーケットの課題

近年,我が国のスーパーマーケットチェーンの経営において来店数の減少が課題となっている。その背景には人口の減少および高齢化による食品市場の縮小,そしてインターネット通販の普及による食品購買手段の多様化がある [1][2]。スーパーマーケットチェーンの経営者は,食品市場が縮小していく中で従来の新規出店による経営拡大戦略を見なおさなければならない。そして,食品購買手段が多様化していく中でも自店舗の顧客に継続して来店してもらうために,既存顧客の再来店に対して何が寄与しているのかを知らなければならない。そのためには先ず食品購買手段の多様化とはどのような現象なのかを知る必要がある。

1-2.食品購買手段の多様化

食品購買手段の多様化はコンビニエンスストアの生鮮食品の取り扱いやドラッグストアの菓子類の取り扱いのような店舗販売によるものと食品の宅配を行う無店舗販売によるものに分けられる。そして,無店舗販売の代表例としてネットスーパーがある。一般にネットスーパーとは「生鮮食品,加工食品,日用品

などスーパーで扱っている商品をパソコンや携帯電話から注文を受けて,店舗から届けるサービス」と定義されている [3]。我が国では 2000 年 5 月に大手総合スーパーの西友が最初に進出している。その後 2006 年から 2009 年の 3 年間で 100 億円から 300 億円の市場に急成長を遂げており[4],ネットスーパーの登場は食品購買手段の多様化に大きな影響を与えていると考えることができる。

ネットスーパーの急成長の背景には経済産業省が社会問題として解決に取り組んでいる買い物弱者の問題がある。経済産業省の定義によれば,買い物弱者とは日常生活に必要な買い物に困難を感じている世帯の総称である。より具体的な消費者像としては,できるだけ外出を控えたい子育て世帯や重量のある商品を運ぶ負担の大きな高齢者世帯,生活圏内に店舗がない過疎地の世帯などが挙げられる。2011 年の経済産業省の調査では,そのような買い物弱者は 600 万人を超えており深刻な問題であること,対策のひとつとして食品の宅配サービスの強化が必要であることが報告されている [5]。

食品の宅配サービスには買い物弱者という社会問題への対応という背景があり,引き続きサービス規模は拡大していくと予想される。一方でスーパーマーケットチェーンは,自社で宅配サービスを運営するしないに関わらず,実店舗の経営努力によって宅

配に無い価値を提供し、既存の顧客に継続して来店してもらう必要がある。

1-3.商品重量への着目

本研究では宅配の価値と関係が深いと考えられる商品重量に焦点を当てる。顧客の再来店と購買商品の重量の関係を分析することでスーパーマーケットチェーンの経営に有用な知識の発見を試みる。具体的な手段としては顧客の再来店を予測するモデルを構築する。再来店予測モデルの構成要素に伝統的な顧客分析手法である RFM 分析[6]に用いられる Recency (直近の来店からの時間差) ,Frequency (来店の頻度) ,Monetary (購買総額) に加えて購買商品の重量を検討し、新たな知識の発見と再来店会員の判別の2点において提案手法の有用性を示す。スーパーマーケットの消費者行動を商品の重量を用いて分析することは著者が調査した限りでは過去の研究実績がなく、本研究の新規性を担保している。

2. 関連研究

スーパーマーケットの課題を扱った研究はこれまで多くなされてきた。飯塚ら[7]は購入金額による会員のランク分けを行い、ランクごとの来店回数や購買品目を分析することで高ランクの会員の特徴および高ランクに成長させるための施策を示した。店舗にとっての重要性を定義し、会員を判別を行い、施策の提案を行う点で本研究と関連がある。しかし、関連研究が購買品目を分析対象にしているのに対し、本研究では購買品目の重量に焦点を当てて分析を行う点で違いがある。高橋ら[8]は欠品を課題として捉え、品目の将来の売れ行きの予測に有用な目利き会員の判別を行った。岸本ら[9]は購買促進を実現する売り場のレイアウトをエージェント・ベース・シミュレーションを用いて検討し、施策を提案した。川田ら[10]は様々な音が無秩序に存在する売り場の音環境を課題として捉え、来店者の意識する音、不快に感じる音の調査を行った。これらの研究はスーパーマーケットの課題を扱った研究である点で本研究と関連があるが、本研究は会員の再来店と購買商品の重量に焦点を当てている点で違いがある。

表 1 関連研究と本研究の位置づけ

	本研究	飯塚(2003)	高橋(2010)	岸本(2009)	川田(2001)
把握内容	会員の再来店と購入商品の重量の関係	購入金額と購入品目の関係	新商品目の将来の売れ行き	販売促進に繋がる売り場レイアウト	来店者に快適な売り場の音環境
手法	ロジスティック回帰分析	共分散構造分析	協調フィルタリング	エージェント・ベース・シミュレーション	アンケート調査

3. ロイヤリティ計測モデルの構築

3-1.分析用データの説明

本研究には「経営科学系研究部会連合協議会主催、平成 27 年度データ解析コンペティションで提供された(株)アイディーズの i-code データ」を使用する。尚、本研究におけるチェーンとは POS データ上の ID を用いて類推したものである。又、チェーンの新規開店や閉店の影響を排除するため、データ期間の全てで毎日購買履歴が存在する店舗のデータへの絞り込みを実施した。

3-2.分析用環境の説明

本研究におけるデータの蓄積および整形はパブリッククラウド環境である Amazon Web Service の 2015 年 12 月時点で標準のサービスである Elastic MapReduce の Hadoop2.6.0,Hive1.0.0,Spark1.5.2 を利用した。またモデルの開発には R3.2.2 を利用した。

3-3.ロジスティック回帰モデルの説明

伝統的な統計手法であるロジスティック回帰モデルを用いて会員のロイヤリティスコア P_{ij} を式(1)として定式化する。偏回帰係数 β の添字にチェーン j を付与し、チェーンごとにモデルを構築することを表している。尚、パラメータの推定には R 言語の glm() 関数を使用する。

$$p_{ij} = \frac{1}{1 + e^{-(\alpha_j + \beta_{1,j}x_{1,ij} + \dots + \beta_{k,j}x_{k,ij})}} \quad (1)$$

p_{ij} … チェーン j の会員 i のロイヤリティスコア

$x_{k,ij}$ … チェーン j の会員 i の k 番目の説明変数

$\beta_{k,j}$ … チェーン j の k 番目の偏回帰係数

3-4.説明変数の設計

RFM 分析の 3 指標を用いたモデル (以降、RFM モデル) ,3 指標に重量を加えたモデル (以降、RFM+W モデル) ,2 つのモデルを構築するために下記 4 指標を会員の説明変数として利用する。

1. Recency スコア : 1 年以内の直近訪問を月単位でスコア化(直近が n ヶ月前ならスコアは $13-n$ 、但し当月の新規会員は 0)
2. Frequency スコア : 1 年以内の来店回数合計
3. Monetary スコア : 1 年以内の購入金額合計

4. Weight スコア : 1年以内の購入商品重量合計

下記の図 1 に本研究に利用する POS データに対応する商品マスタの登録商品のカテゴリ別平均重量を表す。飲料や穀物の重量が高くなっていることがわかる。

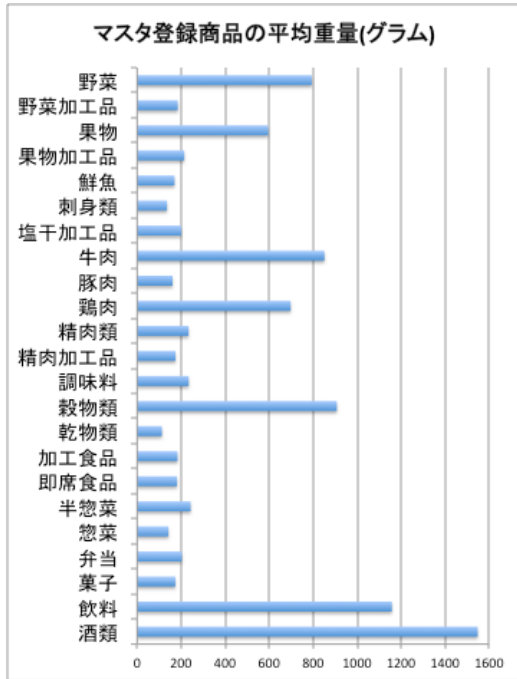


図 1 商品の重量

4. ロイヤリティ計測モデルの評価

RFMモデルとRFM+Wモデルを3つの観点で評価する。

4-1. 偏回帰係数の統計的有意性検定

を用いた説明変数の評価

表 2 にチェーンごとに構築した 2 つのモデルそれぞれの説明変数の偏回帰係数を示す。RFM 分析の 3 指標の偏回帰係数は 14 チェーンの両モデルで全て 5% 有意であり且つ符号は正となった。このことから本研究対象の全チェーンにおいて RFM 指標は将来の来店に正の効果があると解釈できる。一方、重量の偏回帰係数は 14 チェーン中 12 チェーンで 5% 有意となり (有意とならなかった偏回帰係数を表 2 中ではグレーアウトした), そのうち 3 チェーンの符号は正, 9 チェーンの符号は負となった。このことから本研究対象の約 9 割のチェーンで重量は将来の来店に効果があり, その効果はチェーンの特性に依存して正負が分かると解釈できる。

表 2 偏回帰係数の評価

チェーンID	エリア	偏回帰係数						
		RFM			RFM+W			
		R	F	M	R	F	M	W
1	北海道・東北	5.77E-01	2.86E-02	6.28E-06	5.77E-01	2.84E-02	5.72E-06	8.02E-07
2	関東	7.14E-01	1.93E-02	4.18E-06	7.14E-01	1.96E-02	4.63E-06	-7.93E-07
3	関東	5.79E-01	2.53E-02	6.49E-06	5.79E-01	2.56E-02	6.84E-06	-6.41E-07
4	関東	7.80E-01	1.55E-02	4.60E-06	7.80E-01	1.55E-02	4.59E-06	3.00E-08
5	中部	5.95E-01	1.98E-02	3.65E-06	5.95E-01	2.02E-02	4.23E-06	-1.50E-06
6	中部	6.78E-01	2.10E-02	4.91E-06	6.78E-01	2.05E-02	4.21E-06	1.57E-06
7	関西・中国・四国	6.36E-01	1.85E-02	4.12E-06	6.36E-01	1.82E-02	3.71E-06	7.80E-07
8	関西・中国・四国	7.05E-01	1.86E-02	4.76E-06	7.05E-01	1.83E-02	4.37E-06	7.11E-07
9	関西・中国・四国	2.93E-01	5.88E-02	6.90E-06	2.93E-01	5.91E-02	7.30E-06	-1.74E-06
10	関西・中国・四国	3.87E-01	3.71E-02	4.49E-06	3.86E-01	3.75E-02	6.79E-06	-5.48E-06
11	関西・中国・四国	1.47E+00	8.22E-03	4.44E-06	1.49E+00	8.41E-03	4.89E-06	-7.09E-07
12	九州・沖縄	3.71E-01	4.26E-02	6.02E-06	3.71E-01	4.28E-02	6.53E-06	-1.45E-06
13	九州・沖縄	2.99E-01	5.17E-02	6.13E-06	2.98E-01	5.31E-02	7.08E-06	-3.54E-06
14	九州・沖縄	5.89E-01	1.95E-02	4.46E-06	5.88E-01	2.00E-02	5.09E-06	-1.68E-06

4-2. 赤池情報量基準 (AIC) を用いたモデルの評価

AIC は世界的に用いられているモデル選択基準であり下記の式 2 で求められる。右辺第一項は学習データへの当てはまりの良さ, 第二項は罰則項としてモデルの複雑さを表しており AIC のより小さなモデルをバランスの良いモデルと判断できる。

$$AIC = -2 * (\text{最大化対数尤度}) + 2 * (\text{パラメータ数}) \quad (2)$$

表 3 はチェーン別にモデル間の AIC を比較している。本研究対象 14 チェーンのうち 12 チェーンにおいて RFM+W モデルのほうが優れていると評価できた。尚, ロジスティック回帰分析における AIC の評価として AIC の差が ±1 以内であれば同程度とみなすべきという主張があるが本書では大小のみで判断している [11]。

表 3 AIC によるモデルの評価

チェーンID	エリア	AIC	
		RFM	RFM+W
1	北海道・東北	35,030	35,031
2	関東	80,837	80,833
3	関東	109,577	109,575
4	関東	358,231	358,233
5	中部	225,455	225,410
6	中部	245,943	245,886
7	関西・中国・四国	184,197	184,184
8	関西・中国・四国	95,737	95,729
9	関西・中国・四国	75,667	75,652
10	関西・中国・四国	67,630	67,585
11	関西・中国・四国	30,043	30,041
12	九州・沖縄	46,463	46,457
13	九州・沖縄	79,587	79,565
14	九州・沖縄	272,475	272,438

4-3. 再訪会員判別の精度(Accuracy)を用いたモデルの評価

会員の過去1年間の購買履歴から当月の来店会員を判別する実験を行う。モデル間の判別性能の評価は Accuracy を用い以下の式(3)で算出する。

$$\text{accuracy} = \frac{\text{number of true positives} + \text{number of true negatives}}{\text{number of true positives} + \text{false positives} + \text{false negatives} + \text{true negatives}} \quad (3)$$

また、モデル間の Accuracy の差の評価は二項検定による5%有意差を確認する。尚、二項検定には R 言語 binom.test 関数を利用する。交差検証として学習データと検証データは長さが同一の別期間のデータを利用する。

評価結果を表4にまとめる。本研究対象の14チェーン中9チェーンで RFM+W モデルのほうが優れていると評価できた。但し、その差に関して5%有意であったものは1チェーンのみであり、判別の精度に関しては改善の余地があることがわかった。

表4 Accuracy によるモデルの評価

チェーンID	エリア	AIC	
		RFM	RFM+W
1	北海道・東北	81.10%	81.13%
2	関東	81.38%	81.39%
3	関東	80.00%	79.98%
4	関東	82.32%	82.32%
5	中部	81.57%	81.61%
6	中部	81.68%	81.71%
7	関西・中国・四国	81.04%	81.05%
8	関西・中国・四国	82.05%	82.06%
9	関西・中国・四国	82.48%	82.46%
10	関西・中国・四国	81.40%	81.35%
11	関西・中国・四国	88.05%	88.05%
12	九州・沖縄	82.17%	82.18%
13	九州・沖縄	80.15%	80.14%
14	九州・沖縄	81.59%	81.61%

5. まとめと今後の課題

本研究では会員の再来店予測モデルを構築した。モデルの説明変数として会員が購入した商品の重量を与えることで再来店との関係を確認した。結果として重量は来店に対してチェーンの特性によって異なる効果を与えることが確認できた。例えば負の効果となったチェーンに関しては、重量以外の条件が同一である会員であれば飲料などのより重い商品を購入していた会員のほうが再来店しない傾向にあることが確認できた。このことは当該チェーンが宅配

には無い実店舗独自の価値を十分提供できていないことを示唆しており、チェーンの経営者に経営努力を促す有用な知識の発見と言える。

また、今後の課題としてモデルの判別精度の改善に取り組む。商品の重量の意味について、例えば高齢者のように重たい商品を運ぶ負担が相対的に大きな層は重量に対する感度が高く、車を利用して複数店舗を買いまわるような層は重量に対する感度が低いと考察される。このような重量への感度の異質性をモデルに組み込むことで精度の改善が期待できる。

参考文献

- [1] 新日本スーパーマーケット協会: 2015 年度版スーパーマーケット白書, (2015)
- [2] 総務省: 平成 27 年度版情報通信白書, (2015)
- [3] 川辺信雄: ネットスーパーの生成と発展-バーチャル・ビジネスとリアル・ビジネスの統合-, 早稲田商学, Vol. 429, No. xx, pp. 23-78, (2011)
- [4] 経済産業省: 地域生活インフラを支える流通のあり方研究会~地域社会とともに生きる流通~, (2010)
- [5] 経済産業省: 買い物弱者を支えていくために~24 の事例と7つの工夫~, (2011)
- [6] 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)
- [7] 飯塚久哲, 米村大介, 豊田秀樹: 顧客ランクによる行動分析, オペレーションズ・リサーチ, 2003年2月号, pp. 94-99, (2003)
- [8] 高橋雅和, 山田隆志, 津田和彦, 寺野隆雄: 双方向リコメンデーションシステムによる小売流通マーケティングの改善手法, 電気学会論文誌. C, 電子・情報・システム部門誌, No. 130(2), pp. 317-323, (2010)
- [9] 岸本有之, 高橋徹, 高橋雅和, 山田隆志, 津田和彦, 寺野隆雄: エージェント・シミュレーションによる店舗内顧客行動と販売促進策の分析, 情報処理学会研究報告知能と複雑系, Vol. 2009, No. 16, pp. 87-92, (2009)
- [10] 川田一貴, 岩宮 眞一郎: スーパーマーケットの売場における音環境に関する意識調査, 情報処理学会研究報告音楽情報科学, Vol. 2001, No. 16, pp. 79-86, (2001)
- [11] 丹後俊郎, 山岡和枝, 高木春良: ロジスティック回帰分析-SAS を利用した統計解析の実際-, 朝倉書店, (1996)