

# ホームセンターの ID 付レシートデータの分析

M2013SS007 森口元気

指導教員：鈴木敦夫

## 1 はじめに

本研究で対象とするホームセンターは、オペレーションズ・リサーチ（以下、OR）を用いて売上増加や経費削減などに取り組んでいる。例えば、以下のような研究が挙げられる。

- 広告掲載商品の最適選定問題 [1]
- まとめ買いを考慮した商品の発注方式の提案 [2]
- ホームセンターの配送計画に関する問題 [4]

このように、これまで OR を用いた研究は様々あり、業務改善のために役立てられている。本研究では、このホームセンターにおける新しい取り組みとして、クレジットカードの ID 付レシートデータを用いた分析を行う。なお、本研究はこのホームセンターからの委託研究であり、大学 4 年生と役割分担をして行っている。

近年、POS システムのある小売業などでは、顧客 ID がついたレシートデータを収集し、店舗の運営に役立てている。顧客 ID は POS システムを利用するポイントカードやクレジットカードから収集しているものが多い。この顧客 ID があることで、顧客を様々な属性によってグループ化することができ、またそれぞれの属性にあった販売戦略をたてることが可能となっている。

これに関連して、昨今ではビッグデータが注目されている。ビッグデータとは、データのサイズ、種類、変化の頻度の 3 つの尺度をもって、分析が困難であるようなデータをさす。レシートデータも年単位のデータとなれば規模は大きくなり、顧客 ID がついていればデータの構造は非常に複雑になる。これを分析するには、ここ数年で発達してきたビッグデータの分析手法を適用する必要がある。

そこで ID 付レシートデータの分析について調査したところ、以下のような先行研究がなされている。

1. ドラッグストアの ID 付き POS データを用いた日用品購買パターンの分析 [3]
2. ID-POS の購買履歴情報に基づく購買人格の抽出と分析 [6]

1 つめの研究はドラッグストアの顧客 ID 付 POS データを用いて、2 年間の来店回数をもとに顧客を分類し、世代ごとに顧客の来店回数の違いによる購買パターンの違いを分析し、来店回数増加のための特徴を明らかにすることが目的となっている。様々な分析手法を用いたことで、年代ごとに中心となっている商品を明らかにするなどして、ビジネスプランの提案にまで至った。

2 つめの研究は、コンビニエンスストアが発行する ID-POS データを用いて、消費者のライフスタイルなどを明らかにし、購買人格というものを定義している。これによ

り消費者の様々な特性を抽出できるようになり、それを用いたシミュレーションを今後の課題としている。

さらに、現在では web マーケティングの研究も進んでおり、様々な情報を利用してシミュレーションモデルを作成した論文も発表されている [5]。これは、顧客をウェブサイトの閲覧履歴により分類し、いくつかのモデルを作成してシミュレーションを行ったり、様々なシナリオを発生させて分析を行うなどしてモデルの妥当性を検証し、シミュレーションモデルの提案を行っている。

以上の先行研究から、顧客の属性や購入される商品の把握などが、戦略的に扱えることを示している。ホームセンターでも、独自のクレジット機能付きのポイントカード（以下、会員カード）を発行している。この会員カードをホームセンターで精算するときに呈示することで、クレジット払い、現金払いに関わらず一定の割合でポイントがつく仕組みとなっている。またそれと同時に、レシートデータに会員カードの情報が紐付けされる。これにより、誰が、いつ、どの店舗で、どんな商品を、いくらで、いくつ買ったのか、などといったことがわかる ID 付レシートデータになる。この情報をうまく分析することができれば、販売促進活動などに役立たせることができると考えられる。

そこで本研究では、顧客の来店回数増加やそれに伴う売上増加を目的として ID 付レシートデータについて分析を行い、販売促進活動へつなげることを目指す。

## 2 使用するデータについて

本研究で使用するデータは、ホームセンターから提供された会員カードの利用実績である。このデータの期間は 1 年であり、全国にあるこのホームセンターのほとんどすべてが含まれている。このため、データサイズは合計 3.5GB 超の利用実績となり、大きなデータを取り扱うことになる。

なお、このデータの中には会員 ID、年齢、性別などの情報が含まれているが、個人を特定できるような、いわゆる個人情報には含まれていない。現在のところ利用しているデータの項目は、店コード、売上日付、売上時刻、レシート番号、部門、JAN コード、商品名、規格、数量、売上金額、会員 ID、支払方法、年齢、性別である。

ここで JAN コードとは、商品毎についている ID のことであり、規格が違えば JAN コードの値も異なる。また部門とは、商品が属する 1 つの集合であり、例えば植物や工具、日用消耗品といった大きな枠組みをさす。

## 3 分析

ホームセンターから提供されたデータは未加工であるため、少なからず欠損値が含まれている。この対策として、一般的には予測値で補完するか除外するといったことが行

われる。本研究では、必要な場合を除いて欠損値の処理をせずに分析を行う。これは以下のためである。

- 売上金額や個数が、集計した際に影響が出てしまうことを防ぐため。
- 会員の来店頻度や回数に影響を及ぼす可能性を排除するため。

また、異常値についても同様に扱うものとする。

### 3.1 使用するソフトウェア

集計や分析などで用いるソフトウェアは Excel, Access, さらに Visual Mining Studio (以下, VMS) である。VMS は NTT データ数理システムが開発した汎用データマイニングツールであり、各種集計、抽出、分析に使用する。全店舗のデータは 2 つの Access ファイルによって管理されているが、VMS を使うことで 1 つにまとめることができるため、分析が行いやすくなる。

### 3.2 分析事項

ホームセンターからの要請にしたがって、以下の項目について分析を行う。

- 優良会員の定義と抽出
- 会員の構造把握
- リピート性がある商品の把握
- 同時購入商品の把握
- 3 倍ポイントの効果の把握
- 会員の離反防止について

なお、これらの分析結果については、グラフ、表などで視覚化し、ホームセンターでの販売促進策に利用できるようにした。

#### 3.2.1 優良会員の定義と抽出

まずはじめに、対象期間中に来店した日数により会員を分類し、その中でも最上位のグループの会員を優良会員とする。優良会員を定義することで、優良会員と会員全体の購入する商品の違いを比較するなどが可能になる。表 1 に分類した来店日数とそれぞれの会員の構成比を示す。

表 1 来店日数とその割合

来店日数	会員の構成比
60 日以上	3.62%
48~59 日	3.18%
36~47 日	7.17%
24~35 日	15.70%
12~23 日	30.33%
1~11 日	40.00%
計	100.00%

来店日数で分類すると、年間 60 日以上来店している約 3.6% の会員が優良会員として定義される。以後、この優

良会員は分析対象となり会員全体との比較の対象となる。

#### 3.2.2 会員の構造把握

次に、会員全体と優良会員の構造を把握する。構造の把握とは、各年代の人数や男女比率などをとらえることである。会員の構造を調べることで、どのような年代、性別の会員が来店しているのかを把握することができ、今後の分析で対象を絞る時に役立つ。図 1 に会員全体の分布を示すヒストグラムを示す。横軸は年代、縦軸はその割合を示す。

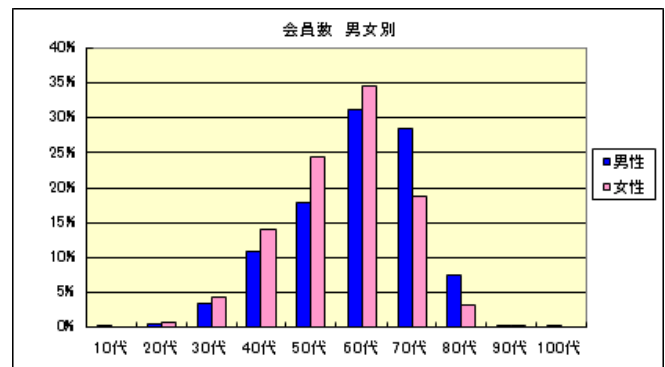


図 1 会員全体の年代の分布

図 1 を見ると、性別に関わらず 50 代~70 代の顧客が多くの割合を占めているのがわかる。一方、優良会員は、男性は似た傾向であるが女性は 50 代~60 代が多いことがわかり、客層が少し違うことが読み取れる。

#### 3.2.3 リピート性がある商品の把握

そこで次に、リピート性がある商品について調べる。これは「ある商品を年に  $n$  回 ( $n$  は 363 以下の自然数) 購入している人数」を集計する作業を行う。Access でデータを操作することで様々な場合におけるリストを得ることができた。以下にその例を示す。表 2 の商品は、年に 10 日

表 2 リピート性がある商品リストの一例

部門	JAN コード	商品名	規格	人数
001	0011111	○○○	△△	30

以上購入されている商品とする。このとき、部門 001 の商品○○○を 1 年のうち 10 日以上購入した会員は 30 人いる、ということになる。このようなリストを会員全体を対象に、 $n = 4, 6, 12, 24, 36, 52$  のそれぞれについて作成した。

$n = 6$  (平均 2 ヶ月に 1 日だけ買われる) のとき、主に日用消耗品やペット用品がリストにあがった。ところが  $n = 12$  以上になると、ペット用品に加えてペットの餌が多数リストにあがるようになり、日用消耗品はほとんどあがらなくなった。これは、日用消耗品の消耗速度、スーパーなどでも購入可能なことや、特売されやすい商品なので特売日だけの購入という場合も考えられるなど、様々な要因

が考えられる。一方ペット用品やペットの餌は  $n = 24$  以上でも多数リストにあがるため、他の商品に比べてリピート性がある商品であると考えられる。

単純に購入された日数だけで集計を行っただけであるので、購入間隔についてまだ考慮されていない。このため、特売商品や季節商品が一部リピート性のある商品のリストにあがってきてしまっているのを、これを考慮し、そういった商品をリストから除外することが必要となる。

### 3.2.4 同時購入商品の把握

続いて、優良会員と会員全体の同時購入の傾向について分析する。この同時購入を調べるために、アソシエーション分析を用いる。

アソシエーション分析とは、本研究で取り扱うようなレシートデータにしばしば用いられる分析手法である。無数にある購入商品の組合せを漏れなく抽出し、価値のあるルールを見つけ出すことを主な目的とする。このルールを支える信頼度とサポート値と呼ばれる値が、価値のあるかどうかの指標として用いられる。例えばレシートデータに「商品 A を買うと商品 B も買う」というルールがあるとすると、信頼度は、「着目するルールの現れた数」を「商品 A が現れたレコード数（出現回数）」で割ったもので定義される。つまり、商品 A を購入したとき、商品 B を買う確率であるとも言える。サポート値は、そのルール自体が発生する確率を示す。これを表 3 のようにまとめられたとする。すると、このルールは 0.5% の確率で発生し、商品 A を購入すると商品 B も 80% の確率で購入する、ということを示す。一般的に、どちらの値も高いほうが価値のあるルールであると言われている。

表 3 アソシエーション分析の結果の例

前提商品	結論商品	信頼度	サポート値
商品 A	商品 B	80 (%)	0.5 (%)

このアソシエーション分析を、優良会員、会員全体のそれぞれで性別でわけて行った。男性で比較すると、優良会員はペットに関連する商品のルールが多かったが、会員全体でみると食品や日用消耗品のルールが多いという結果となった。女性も男性と似たような結果となり、優良会員かどうかによって同時に購入する商品の傾向に違いをみることができた。

リピート性がある商品についてペットの餌を挙げたが、この結果とあわせるとより信頼できる結果になったと考えられる。これは、アソシエーション分析は商品がレシートに現れる回数に大きく関わっているためである。

### 3.2.5 3 倍ポイントの効果の把握

ホームセンターでは、毎月第 1 週の土曜日と日曜日に会員カードのポイント付与を 3 倍にする販売促進活動を行っている。また、キャンペーンや新店舗のオープンなどでも臨時で同様の販売促進活動を行っており、集客に力を入れ

ている。

このポイント 3 倍の日を含む週の売上は、他の通常の週の売上と比べて多いことが経験的に知られているが、数値的な根拠がなかったため調べる必要があった。そこで以下のような仮説をたて、t 検定を行う。

- 帰無仮説：ポイント 3 倍の日を含む週の売上の平均は、通常の週の売上の平均に等しい。
- 対立仮説：ポイント 3 倍の日を含む週の売上の平均は、通常の週の売上の平均に等しくない。

ここで、t 検定を行うとき等分散性の仮定が必要になる。これについて有意水準 5% で F 検定を行った結果、等分散性があると仮定することになった。これをもとに有意水準 5% で t 検定を行った結果、P 値は 0.05 以下となり、帰無仮説は棄却された。すなわち、ポイント 3 倍の日を含む週の売上は含まない週の売上と比べて多い、ということが明確となった。

また会員全体において、ポイント 3 倍の日を含む週と含まない週の「クレジット払い」の利用率について、折れ線グラフで視覚化して比較した。すると、ポイント 3 倍の日を含む週の利用率は上がっていることが確認できた。臨時で行ったクレジット払いのみポイント 3 倍の日は利用率が特に高く、会員はポイントに敏感であることがうかがえた。

### 3.2.6 会員の離反防止について

2013 年度において、表 1 からわかるとおり、カード会員であるにもかかわらず、ほとんど来店していない会員が 4 割を占めている。この 4 割のうち、新規以外の会員に何らかの施策を検討しなければ、やがて完全に離反してしまうことが予想される。

そこで取組みとして、全会員のそれぞれの厳密な来店間隔をはかり、来店間隔の平均と標準偏差を求めた。この来店間隔の平均と標準偏差の 2 倍を足した値が、各会員の離反可能性を示す参考値になると考える。実際に、季節性などの性質がある会員もいるので、詳細な分析が今後の課題となる。

## 4 モデルの検討

これまで、どのような会員が来店しているか、どんな商品が同時に購入されているかなどについてを調べてきた。これを踏まえて、どのような最適化モデルを作成し、販売促進活動に役立てていくかを検討していかなければならない。

販売促進活動にも様々な取り組みがある。例えば以下のような活動が考えられる。

- ダイレクトメール (DM) 送付、チラシの配布、メールマガジンの配信
- 店頭ディスプレイの配置、新規カード会員獲得
- キャンペーン (特別ポイント付与、値引き)、特売

このように広告、店舗での人的活動、安売りといった種類の仕方がある。本研究では来店回数の増加を主な目的と

して取り組んでいるため、影響が大きい販売促進活動は広告と安売りになる。ホームセンターの方と検討した結果、「特別なポイントの割り振り」についてモデル化する。

#### 4.1 モデル化

商品の種類が 15 万を超えるため、モデル化を行う準備として、まず商品の種類を絞る。ホームセンターの方と検討し、以下の条件をもとに商品数の絞り込みを行った。

- 通常の土日の年間売上数量が 50 以上である
  - 月初めの 3 倍日と通常の土日との年間売上数量の差が 50 以上である
  - 月初めの 3 倍日と通常の土日との年間売上金額の差が正である
  - 欠損値が含まれるデータや客注代は除外する
- 以上の条件を満たす商品は 318 点であった。

##### 4.1.1 記号の定義

定式化のため、以下の記号を定義する。

添字

$I$ : 商品の集合  $i \in I$

$J$ : 1 から 5 までの整数の集合  $j \in J$

定数

$s_i$ : 商品  $i \in I$  の土日の売上実績 (1 日あたり)

$q_{ij}$ : 各ポイントにおける商品  $i \in I$  の伸び率

$a_j$ : 付与するポイント

$C$ : ポイントの予算

$U_j$ : 付与するポイントの種類毎の商品数の上限

決定変数

$$x_{ij} = \begin{cases} 1 & \text{商品 } i \in I \text{ の } j \in J \text{ 番目の伸び率を選択,} \\ 0 & \text{それ以外} \end{cases}$$

ポイントの伸び率とは、3 倍ポイントの土日と普通の土日の売上実績を 1 日あたりに直し、その比を取ったものである。例えば、3 倍ポイントの土日の 1 日あたりの売上が 800 円、普通の土日の 1 日あたりの売上が 200 円の時、ポイント 3% 時の伸び率は 4 倍である。これをもとに 5%、7%、10% 時の伸び率を仮定する。また、ポイントの予算についてもパラメータとして様々な値を仮定する。

##### 4.1.2 定式化

定義した記号を用いて定式化をする。

目的関数

$$\max \sum_{i \in I} \sum_{j \in J} s_i q_{ij} x_{ij} - \sum_{i \in I} \sum_{j \in J} s_i q_{ij} a_j x_{ij} \quad (1)$$

制約条件

$$\sum_{i \in I} \sum_{j \in J} s_i q_{ij} a_j x_{ij} \leq C, \quad (2)$$

$$\sum_{i \in I} x_{ij} \leq U_j, \quad j \in J, \quad (3)$$

$$\sum_{j \in J} x_{ij} = 1, \quad i \in I, \quad (4)$$

$$x_{ij} \in \{0, 1\}, \quad i \in I, \quad j \in J \quad (5)$$

各式の意味は以下のとおりである。

式 (1): ポイントを考慮した売上の最大化

式 (2): ポイントのコストはポイントの予算以内

式 (3): 各ポイントに割り振ってよい商品の数の上限

式 (4): 各商品が属するポイントの種類は必ず 1 つ

式 (5): 決定変数はバイナリ変数

これにより、様々な状況のポイントの割り振りができるようになった。実行例はここでは割愛する。

なお、このモデルは Excel 上に最適化ソフトウェア What's Best!12.0 を利用して実装し、ホームセンターに提供した。

## 5 おわりに

ID 付レシートデータを分析することで、会員の基本構造やポイント 3 倍の日の効果などについて知ることができた。また特別なポイントを割り振るモデルを定式化したことで、様々な状況を想定してパラメータを調整することができ、どの商品に特別なポイントを付与すべきかを把握することができるようになった。

## 参考文献

- [1] 崔康幸, 岩瀬爽, 岡村彩音: 広告掲載商品の最適選定問題, 2012 年度南山大学情報理工学部情報システム数理学科卒業論文, 2013.
- [2] 早川陽介, 池田貴裕, 森嶋聡一: まとめ買いを考慮した商品の発注方式の研究, 2012 年度南山大学情報理工学部情報システム数理学科卒業論文, 2013.
- [3] 久松俊道, 朝日弓未, 山口俊和: ドラッグストアの ID 付き POS データを用いた日用品購買パターンの分析, オペレーションズ・リサーチ: 経営の科学, vol.57 no.2 pp.63-69 2012-02-01.
- [4] 市川雄太, 赤松真衣, 廣田健, 岡田直樹: ホームセンターの配送計画に関する研究, 2011 年度南山大学数理情報学部情報システム数理学科卒業論文, 2012.
- [5] 柴田一樹, 和泉潔, 磯崎直樹, 吉村忍: 閲覧行動タイプに基づいたウェブ広告配信シミュレーションモデル, 電気学会論文誌 C, vol.133 No.9 pp.1762-1769, 2013.
- [6] 山本仁志, 諏訪博彦, 岡田勇, 小川祐樹, 和泉潔, 磯崎直樹, 服部進: ID-POS の購買履歴情報に基づく購買人格の抽出と分析, 経営情報学会 2011 年春季全国研究発表大会要旨集, ([https://www.jstage.jst.go.jp/article/jasmin/2011s/0/2011s\\_0-170/\\_pdf](https://www.jstage.jst.go.jp/article/jasmin/2011s/0/2011s_0-170/_pdf)).