

広告掲載商品の最適選定問題

2006MI146 坂井寛治 2006MI148 佐々祐資

指導教員：鈴木敦夫

1 はじめに

現在、数多くの流通業者は顧客獲得や売上を増加させる1つの手段として、折込広告を使用している。本研究では、ホームセンターにおける折込広告の掲載商品を最適化することで、売上の増加につなげようと考えた。

ホームセンターで折込広告を作成するにあたり、どの商品を折込広告に掲載するかを、折込広告の制作部署が今までの経験を頼りに選定し、作成するというのが現在の一般的な方法である。また、現在のホームセンターで扱っている商品数は約5万種以上であり、各店舗で扱っている商品数は、そのうちの約2万種も存在する。この数多くの商品の中から、今までの経験のみで広告掲載商品を選定しているため、売上増加に直結しない商品に掲載してしまったり、売上に大きな効果がある商品を見逃してしまっていた可能性がある。

この状況を改善し、折込広告の作成方法を見直すことによって販売数を増加させることを目的とした研究が、過去に行われている。この研究はホームセンターからの委託を受け、昨年度、ある程度の成果を上げている。その内容は、過去のレシートデータをもとに、過去に折込広告に掲載されたことがある商品と、その商品が実際に購入されたときのレシートデータをもとに分析するものである。そして、その分析をもとに、広告掲載商品の最適選定を行うというものであった [4]。

この過去の研究では、同時購入という概念を中心に研究が進められている。それは、同時購入される確率が高い商品を折込広告に掲載すれば、その掲載された商品の販売数だけでなく、同時購入されやすい商品の販売数の増加にもつながると考えたからである。

また、過去の研究で対象としている商品は、過去に折込広告に掲載されたことがある商品であり、過去に掲載されたことがない商品は、対象とされていなかった。私たちは過去のレシートデータをもとに、購入履歴のある商品を対象に過去の販売傾向を分析する。

この他にも問題点がある。過去の研究では販売数の増加を目的としていた。しかし、販売数を増加させようとすると、価格が低い商品で過去の販売数が多いものに広告掲載商品が偏ってしまうという問題点が指摘できる。したがって、本研究では売上増加を目的とし、ある種類の商品に広告掲載商品の偏りが生じてしまわないように、商品選定を種類別に行うこととした。

これらのように、過去の研究を改善し同時購入という概念を発展させて、最終的には売上増加につながるような広告掲載商品の最適選定を行うことを本研究の研究目的とする。

2 データの詳細

ホームセンターにおける商品にはラインという7つのグループ分けが存在する。さらにこのラインを細かく分けると部門という小さい分類に分けることができる。商品は26個の部門に分割されており、折込広告では、1部門ごとに広告に掲載する数が決められている。

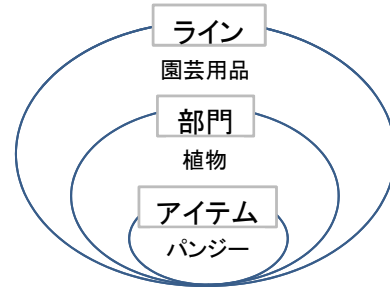


図1 商品分類

また、この研究で使用しているデータは2008年度の広告データと2008年度の年間のレシートデータである。本研究で注目している同時購入は、このレシートデータからアソシエーション分析を行って調べる。

3 同時購入の分析

3.1 アソシエーション分析

アソシエーション分析は、スーパーマーケットのデータなどに対して、例えば、商品Aを購入したときに商品Bを購入する確率などの購入された商品同士の関係性を分析するものである [3]。

アソシエーション分析を行うにあたって、「ルール」の長さというのが必要になる。「ルール」とは、「商品Aを購入したとき商品Bを購入する」といった前後関係を含む同時購入されたことがある商品の組合せのことをいい、「ルール」の長さとは、ひとつのルールの中に含まれる商品数を示す値となっている。つまり、「商品Aを購入したとき、商品Bを購入する」というときのルールの長さは2であり、「商品Aと商品Bを購入したとき商品Cを購入する」や「商品Aを購入したとき商品Bと商品Cを購入する」といったときはルールの長さは3ということになる。また、アソシエーション分析において、ルールの長さが2の場合の前後関係の“前”の部分にあたる商品Aを「前提」、 “後”の部分にあたる商品Bを「結論」という。

また、アソシエーション分析では前後関係が逆になった「商品Bを購入したとき商品Aを購入する」というときの分析結果も出る。それは、同時購入された商品の全

での組合せにおいて前後関係をみて分析しているからである。

ここで、アソシエーション分析の結果より、主に研究に使用するルール数と信頼度というデータの説明をする。

ルール数とは「商品 A を購入したとき商品 B を購入する」というルールがデータの中に現れた回数である。これは「商品 B を購入したとき商品 A を購入する」というルール数と同じになる。つまりルール数とは、商品 A と商品 B を同時購入した人の人数を示している。

また、これに関連した値として前提数と結論数というものがあり、前提数とは、前提の商品を購入した人数を示し、結論数とは結論の商品を購入した人数を示している。

信頼度とは、「商品 A を購入した人における商品 B を購入した人の割合」を示している。言い換えると「商品 A を購入した人が商品 B も購入する確率」ということになる。また、信頼度は次の計算式で求めることができる。

$$\text{信頼度} = \frac{\text{ルール数}}{\text{前提数}} \times 100$$

また、アソシエーション分析には Visual Mining Studio を使用した [3]。

3.2 同時購入金額の定義

ある商品 A に対して、その商品と同時購入される確率が高い別の商品 B があつたとする。すると、商品 A の価格には、本来の価格に加えて商品 B の価値が付加すると考えることができる。そこで私たちは、過去のレシートデータより、すべての商品にそれぞれ付加される価値を調べ、それを各商品の同時購入金額とし、次にその定義を示す。

例えば、100 円の商品 A と 150 円の商品 B があり、商品 A を購入した人が 100 人いたとし、そのうち 40 人が商品 B も購入したとする。また、商品 B を購入した人は 200 人いたとする。

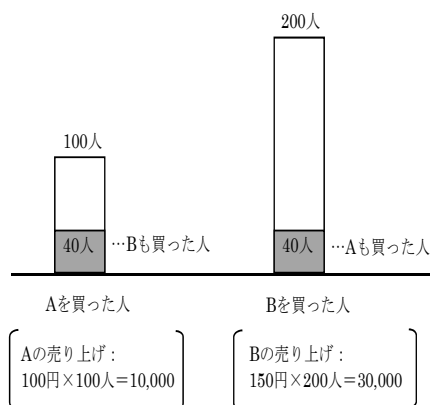


図 2 同時購入の例

ここで、商品 B を購入した 200 人のうち 40 人は商品 A も購入しているので、商品 B の売り上げ 30,000 円のうち

40/200 の 6,000 円は、商品 A を購入した 100 人全員の影響と考える。つまり、商品 A を購入した 1 人の商品 B に対する影響は、6,000 (円) ÷ 100 (人) = 60 (円) である。そして、この 60 円を商品 A の商品 B に対する同時購入金額と定義する。計算式をまとめると、次のようになる。

$$\begin{aligned} & 150(\text{円}) \times 200(\text{人}) \times \frac{40(\text{人})}{200(\text{人})} \times \frac{1}{100(\text{人})} \\ &= 150(\text{円}) \times \frac{40(\text{人})}{100(\text{人})} \\ &= \left[\text{アソシエーション分析の結論の価格} \right] \\ & \quad \times \\ & \quad \left[\text{商品Aの商品Bに対する信頼度} \right] \\ &= \left[\text{商品Aの商品Bに対する同時購入金額} \right] \end{aligned}$$

このように商品 A の商品 B に対する同時購入金額を求め、同様に商品 A につく商品 B 以外の同時購入金額をすべて求め、合計したものを商品 A の同時購入金額と定義する。

4 折込広告商品の最適選定

4.1 定式化の概要

まず、目的関数の概要を次にまとめる。

目的関数では、商品 i の同時購入金額を商品 i の価格に足して商品 i の価値付き価格とし、この価値付き価格に商品 i の販売数予測値を掛け、最後に変数を掛けて和をとることで売上を最大にする。ここでいう変数は商品 i 、つまりアソシエーション分析の結果における前提に対応している。したがって、広告掲載商品の対象は前提の商品となる。

しかし、このままでは、例えば商品 A と商品 B が採用されたとき商品 A と商品 B がともに購入されていた場合、商品 A の商品 B に対する同時購入金額と、商品 B の商品 A に対する同時購入金額が重複してしまっていることになる。この重複部分を解消するための引き算の式を目的関数に組み込む必要がある。

次に、制約条件については、折込広告に掲載する商品数に関する制約と、変数がバイナリー変数であるという変数制約がある。

そして、最後に非線形になっている部分を線形化する。定式化までの具体的な手順を次に示す。

4.2 ネットワークの作成

アソシエーション分析の結果には、前章で述べたとおり「前提」、「結論」という前後関係が存在する。アソシエーション分析の結果を利用する際、この前後関係をうまく表現し定式化するために、ここでその分析結果をネットワーク化する。

まず、レシートデータに載っている商品、つまりその期間に購入されたことがある商品をノードとする。そし

て、同時購入された商品同士を枝で結び、有向グラフを作成する。次に有向グラフの定義を示す [2]。

$N(V, E)$: 有向グラフ
 V : ノード (購入履歴のある商品) の集合
 E : 枝の集合 ($E = (i, j) | i, j \in V$)

つまりここでは、商品 i と商品 j が同時購入されたことを示す。さらに、枝の始点である商品 i をアソシエーション分析の結果における前提とし、枝の終点である商品 j を結論とする。また、枝の長さを P_{ij} とおき、その値をアソシエーション分析の結果の信頼度とする。

次に、先ほど述べたように、重複している部分を引くための式を作るにあたって、もうひとつの有向グラフを作る必要がある。引き算の部分は、広告掲載の対象商品同士が同時購入されたときの同時購入金額の重複を解消するための式なので、広告掲載の商品対象のみでネットワークを作成する。つまり、アソシエーション分析の結果における前提の商品だけで有向グラフを作成する。次にこの有向グラフの定義を示す。

$N'(V', E')$: 有向グラフ
 V' : ノード (アソシエーション分析の結果における前提の商品) の集合
 E' : 枝の集合 ($E' = (s, t) | s, t \in V'$)

また、この有向グラフは、広告掲載の対象商品同士での同時購入についてのものであり、先ほどと同様に前提と結論という前後関係が存在する。つまり、 N' は N の部分集合なのである。

これらの有向グラフをふまえて定式化をしていく。

4.3 記号の定義

まず、添字と定数の定義を示す。

L : 部門 l の集合 ($l \in L$)
 I_l : アソシエーション分析における前提の商品 i の添え字集合 (前提の商品は部門 l に限定) ($i \in I_l, l \in L$)
 J : アソシエーション分析における結論の商品 j の添え字集合 ($j \in J$)
 A_l : 部門 l における広告掲載商品数
 N_i : アソシエーション分析における前提の商品 i に対する結論の商品 j の添え字集合 ($j \in N_i$)
 n_i : 商品 i の販売数予測値 (レシートデータからの商品 i の販売数の和)
 w_i : 商品 i の価格
 P_{ij} : アソシエーション分析における前提の商品 i の、結論の商品 j に対する信頼度
 N'_i : アソシエーション分析における前提の商品 i に対する結論の商品 k の添え字集合であり、 $k \in I_l$ に限定 ($k \in N'_i$)
 r_{ik} : 商品 i が前提、商品 k が結論のときのルール数 (組み合わせ数)

次に変数の定義を行う。

$$x_i = \begin{cases} 1 : \text{商品 } i \text{ を掲載する} \\ 0 : \text{商品 } i \text{ を掲載しない} \end{cases}$$

4.4 定式化

目的関数は次のような式になる。

$$\max \sum_{i \in I_l} n_i \left\{ w_i + \sum_{j \in N_i} P_{ij} w_j \right\} x_i - \sum_{i \in I_l} \sum_{k \in N'_i} P_{ik} w_k r_{ik} x_i x_k \quad (1)$$

この目的関数における $\sum_{j \in N_i} P_{ij} w_j$ の部分は商品 i の同時購入金額であり、 $P_{ik} w_k$ の部分は商品 i の商品 k に対する同時購入金額である。

制約条件は次のようになる。

広告掲載商品数に関する制約

$$\sum_{i \in I_l} x_i \leq A_l \quad (l \in L) \quad (2)$$

変数制約

$$x_i \in \{0, 1\} \quad (i \in I_l) \quad (3)$$

最後に、(1) は非線形式で表されているため、 y_{ik} という変数を導入して線形化する [1]。

y_{ik} という変数の定義は次のようになる。

$$y_{ik} = \begin{cases} 1 : \text{商品 } i \text{ と商品 } k \text{ を両方とも掲載する} \\ 0 : \text{その他の場合} \end{cases}$$

目的関数は次のようになる。

$$\max \sum_{i \in I_l} n_i \left\{ w_i + \sum_{j \in N_i} P_{ij} w_j \right\} x_i - \sum_{i \in I_l} \sum_{k \in N'_i} P_{ik} w_k r_{ik} y_{ik} \quad (4)$$

そして、制約条件に線形化のための次の条件を加える。

$$\begin{aligned} y_{ik} &\leq x_i \quad (i \in I_l, k \in N'_i) \\ y_{ik} &\leq x_k \quad (i \in I_l, k \in N'_i) \\ x_i + x_k - 1 &\leq y_{ik} \quad (i \in I_l, k \in N'_i) \\ y_{ik} &\in \{0, 1\} \quad (i \in I_l, k \in N'_i) \end{aligned}$$

4.5 実行するにあたって

実行するにあたって LINDO SYSTEMS INC. 社の LINGO 11.0 を使用し実行した [5] .

LINGO とは線形・非線形最適化問題を解くツールである .

5 実行結果

最適解を求めるにあたって, 2008 年度 12 月 11 日から 15 日までの 5 日間のレシートデータを使用した . このレシートデータをアソシエーション分析し, 前提の商品を部門別に分け, それぞれに広告掲載商品数の制約をかけ部門ごとに最適広告掲載商品を求めていった . また, 制約の広告掲載商品数というのは, レシートデータと同じ期間に実際に折込広告に掲載されていた部門ごとの商品数を利用した .

そして, 実行結果と実際の広告掲載商品とを比較してみた . 実際の掲載商品にはなく, 新たに最適広告掲載商品として選定された商品の数を, 部門ごとに次にまとめた . また, 売上予想金額についても比較した . ここで私たちは比較対象として, 実際の広告掲載商品を私たちが作った目的関数に当てはめ, 売上金額を算出し, これを利用した . 私たちが求めた売上予想金額というのは, 同時購入金額を求めるなどして算出したものであり, 架空の金額である . そのため, 実際の売上金額と比較するのではなく, 目的関数に当てはめて, 比較した .

表 1 広告掲載商品数, 売上予想金額との比較 (部門別)
2008 年度 12 月 11 日 ~ 15 日までレシートデータより
① 広告掲載商品数, ② 新たに選定された商品数,
③ 売上予想金額, ④ 比較対象の売上金額 (1),
⑤ 倍率 (2)

| 部門 | ① | ② | ③ | ④ | ⑤ |
|----|-----|-----|-------------|-------------|------|
| 1 | 46 | 42 | ¥ 882,886 | ¥ 442,516 | 2.00 |
| 2 | 26 | 24 | ¥ 1,342,564 | ¥ 437,743 | 3.07 |
| 3 | 23 | 22 | ¥ 383,206 | ¥ 154,419 | 2.48 |
| 4 | 18 | 18 | ¥ 227,385 | ¥ 170,179 | 1.34 |
| 5 | 28 | 25 | ¥ 412,036 | ¥ 219,382 | 1.88 |
| 6 | 19 | 19 | ¥ 359,725 | ¥ 69,025 | 5.21 |
| 7 | 19 | 13 | ¥ 992,781 | ¥ 529,645 | 1.87 |
| 8 | 48 | 33 | ¥ 4,656,053 | ¥ 3,892,972 | 1.20 |
| 10 | 27 | 23 | ¥ 975,660 | ¥ 228,252 | 4.27 |
| 11 | 25 | 20 | ¥ 5,005,939 | ¥ 529,879 | 9.45 |
| 12 | 126 | 108 | ¥ 9,082,566 | ¥ 3,968,741 | 2.29 |
| 14 | 18 | 14 | ¥ 532,552 | ¥ 510,323 | 1.04 |
| 15 | 22 | 21 | ¥ 925,863 | ¥ 111,060 | 8.34 |
| 17 | 52 | 42 | ¥ 9,252,477 | ¥ 931,989 | 9.33 |
| 18 | 34 | 31 | ¥ 4,817,524 | ¥ 993,335 | 4.85 |
| 19 | 4 | 4 | ¥ 234,033 | ¥ 37,542 | 6.23 |
| 20 | 6 | 6 | ¥ 91,764 | ¥ 44,956 | 2.04 |
| 22 | 46 | 31 | ¥ 147,768 | ¥ 69,951 | 2.11 |

| 部門 | ① | ② | ③ | ④ | ⑤ |
|----|----|----|-------------|-------------|-------|
| 23 | 30 | 28 | ¥ 1,760,813 | ¥ 141,927 | 12.41 |
| 25 | 13 | 11 | ¥ 298,639 | ¥ 153,127 | 1.95 |
| 27 | 2 | 2 | ¥ 37,262 | ¥ 14,365 | 2.59 |
| 28 | 38 | 24 | ¥ 2,839,097 | ¥ 1,966,856 | 1.44 |
| 30 | 4 | 4 | ¥ 989,119 | ¥ 583,707 | 1.69 |

- 1: 広告掲載商品を目的関数に代入した時の売上金額
2: (倍率) = (売上予想金額) ÷ (比較対象の売上金額)

6 考察

比較した結果, 最適広告掲載商品は, すべての部門においてほとんどの商品が実際に折込広告に掲載されていた商品とは異なる新たに選定された商品となった . これに伴い, 売上金額も全ての部門において予想金額の方が上回るという結果が得られた . しかし, 売上予想金額は同時購入にのみ注目し, 算出したものであり, 単品購入には注目しなかった . そのため, 売上金額が多少, 少なく見積もられていると思われる .

7 今後の課題

私たちは, 同時購入という概念を軸に研究を進めたので, 単品購入にまで注目して最適解を求めることが必要である .

それにより, より最適な選定商品, 売上予想金額を求めることができると考える . また, 私たちが定式化したものを解く際に使用した LINGO には, スプレッドシートから表を読み込む際の表の大きさに限界があり, 全商品を対象として問題を解くことができず, 部門ごとに問題を解くことになってしまった . そのため, 全商品の売上金額を比較することができなかつたので, 表の作りなどを工夫し全商品を対象として問題を解き, 広告掲載商品を選定する必要がある . このような問題点を解決し, より良い商品選定をしていくことが今後の課題として挙げられる .

参考文献

- [1] F.S.Hillier, G.J.Lieberman: Introduction to Operations Research Eighth Edition, Mc Graw-Hill, 2005 .
- [2] 福島雅夫: 『数理計画入門』 . 朝倉書店, 東京, 1996 .
- [3] 加藤直樹, 羽室行信, 矢田勝俊: 『データマイニングとその応用』, 朝倉書店, 東京, 2008 .
- [4] 野々垣壇: 『折込広告の最適選定問題について』, 2008 年度南山大学数理情報学部数理学科学科卒業論文, 2008 .
- [5] 横山雅夫: 『LINGO を用いた線形問題ならびに整数計画問題の解き方』
<http://www.ipc.fukushima-u.ac.jp/kouhou/01/P24-32.pdf>