

顧客データを用いた効果的クーポン配布手法: 価格設定とブランド選択についての可読性の高いルールの発見をめざして

大阪産業大学経営学部 羽室行信 HAMURO Yukinobu
01104684 京都大学大学院工学研究科 *加藤直樹 Naoki KATOH
関西大学商学部 矢田勝俊 YADA Katsutoshi

1 序

インターネットをはじめとした近年の情報技術の発展により、顧客の購買履歴など、非常に詳細な顧客データを低コストで獲得できるようになってきており、ビジネスにおける購買履歴の重要性は次第に認識されるようになってきている [3]。このような背景の中、データマイニングもしくはデータベースにおける知識発見が注目され、大規模なビジネスデータから興味深い知識を半自動的に引き出す技術や手法が模索されている [2]。本研究では、詳細な顧客購買データを用い、ディスカウントクーポンの最適配布方法、すなわち、どの顧客にいくらのディスカウントクーポンを配布すれば効果的かについて、ブランド選択の観点から論じる [1, 4]。クーポンに関する研究においては、従来より、その割引額や期間、対象商品、顧客特性などがどのように顧客獲得につながっているかに対して、多くの関心が払われてきた。また、アメリカでは最近、顧客毎に異なる値引額のクーポンを配布する customized coupon がインターネットの発達とあいまって注目を集めている。

上述の背景の中で、我々は、ディスカウントクーポンの最適な配布について、以下の二種類の問題の定式化する。さらに、これらの問題を解くためのヒューリスティックな手法を提案する。

提案する手法の有効性を測るために、薬局チェーン店における巨大な購買データを用いて実験を行った。今回の実験では、洗濯洗剤のみを扱っているが、本手法は、容易に他の商品カテゴリーにも適用することができる。

2 問題の定式化

2.1 価格設定

価格設定は、そのブランド選択に影響を与える主要な要因の一つと言ってよいであろう。我々は、価格設定に関して、以下に説明する二種類の価格設定を考えた。

1. 時系列相対価格: ある店におけるある特定のブランドに関して、ある価格以下であれば購入するが、その価格より高ければその商品を購入しないと決めている顧客は多いであろう。そこで、そのブランドが通常価格よりどのくらいディスカウントされたかを第一の価格設定として考える。あるブランドがある一定期間において通常価格より値段が下がった時、ディスカウントされていると認識するものとする。このような価格設定を時系列相対価格と呼ぶ。
2. ブランド間相対価格: ある顧客は、他のブランドと値段を比較することによって購入ブランドを決定するかもしれない。例えば、ブランド A にロイヤルな顧客は、ブランド B がブランド A より 100 円以上安い時に限って、ブランド B を購入するかもしれない。そこで、我々は、全ての 2 ブランド間の価格差を第二の価格設定として用いた。

2.2 目的関数

本研究では次の二つの目的関数を考える。

(1) 顧客ゲイン最大化: 最初の問題は、ある予算制約のもと、対象となるブランドをクーポンの配布により購入し始める顧客 (顧客ゲイン: customer gain と呼ぶ) を最大化する問題である。予算には、クーポンの配布に必要なとなる郵送費などのコスト (配布コスト: fixed cost と呼ぶ) に加えて、実際にクーポンが利用されたことにより発生するコスト (換金コスト: redemption cost と呼ぶ) が含まれる。これらのコストは、実際のクーポン配布において、メーカーが小売店に対して負担するコストである。

(2) 効果的コスト最大化: 二番目の問題では、ディスカウントクーポンの限界効用を考慮する。ディスカウント価格が p の時、対象商品を購入するであろう顧客が N_p 人いるとする。これら N_p 人の顧客の中には、ディスカウント価格が p よりも低い p' であったとしても、すすんで購入する顧客もいるであろう。そのような顧客の人数を $N_{p'}$ 人とする、 $(p' - p)N_{p'}$ のコストは浪費されたといえることができる。実際にかかったコストから浪費コストを引いたコストを効果的コスト (effective cost) と呼ぶ。そこで、二番目の問題は、ある予算制約のもとで、この効果的コストを最大化する問題である。

2.3 導出ルール

ここで、これらの問題を解く際に得られるルールについて、それらのルールの可読性を高めることを我々は重視している。なぜならば、得られたルールがたとえ優れたものであったとしても、現場の人間にとって全く解釈できなければ、実際のアクションプランに結びつくような洞察を得ることが困難となるからである。例えば、See5 [5] のような決定木によって生成されるルールは、実際のビジネスにおいて適用した場合、その解釈が困難であるケースが多い。そこで、ルールは単純なものとなるように制限した。

顧客データには n 個の属性 A_1, A_2, \dots, A_n があるとすると、各属性 A_i のドメイン $dom(A_i)$ は適当に離散化されていると仮定する。つまり、 $dom(A_i) = \{a_{i1}, a_{i2}, \dots, a_{i, s_i}\}$ とする。本稿で扱うルールは次の二通りである。また、ディスカウント価格についても、離散値をとるものと仮定する。

1. 1次元ルール: 属性 A_i に関するルール R_i を (S_i, d_i) と表す。ここで $S_i \subset dom(A_i)$ で d_i は d_i はディスカウント価格を表す。すなわち、 R_i は $S_i \subset dom(A_i)$ なる顧客にディスカウント価格 d_i のクーポンを配布するというルールである。このような顧客グループを $C(R_i)$ と記す。 $R = \bigcup_{1 \leq i \leq n} R_i$ とおく。
2. 2次元ルール: 属性 A_i と A_j に関するルール R_{ij} を (S_i, S_j, d_{ij}) と表す。すなわち、 R_{ij} は $S_i \subset dom(A_i)$ かつ $S_j \subset dom(A_j)$ なる顧客にディスカウント価格 d_{ij} のクーポンを配布するというルールである。また、 $R' = \bigcup_{1 \leq i, j \leq n} R_{ij}$ とおく。

ルール集合 R に対する顧客ゲインを $gain(R)$ 、効果

的コストを $effective(R)$, 配布コストと換金コストの総和を $cost(R)$ とする。また総予算を B とする。

そのとき問題は次のように定式化される。

$$P_1 : \{ \text{maximize } gain(R) \mid cost(R) \leq B \}$$

$$P_2 : \{ \text{maximize } effective(R) \mid cost(R) \leq B \}$$

二つの次元ルール R_i, R_j で定まる顧客集合 $C(R_i)$ と $C(R_j)$ は互いに素ではないことに注意されたい。互いに素であってもナップザック問題となるので、難しい問題と言える。したがって以下で述べるような貪欲解法を用いて解を得ることとした。

3 アルゴリズム

問題 P_1 に対して 1 次元ルールを対象として説明する。 $\{d_0, d_1, \dots, d_m\}$ をディスカウント価格の集合とする。 d_0 はディスカウントなしを意味する。属性 A_i が値 a_i を取る顧客集合を G , その人数を g とする。 G に属する顧客がディスカウント価格 d_j で対象商品を購入する確率を p_j とする。このとき、顧客ゲインの期待値は $(p_j - p_0)g$ であり、その時の費用は $(d_j p_j + c)g$ となる。ここで c は単位クーポン当たりの配布コストである。そして、

$$(p_j - p_0) / (d_j p_j + c)$$

の比を考える。この比が最大となるような顧客グループ G とディスカウント価格 d_j の組に対して d_j のディスカウントクーポンを配布していく。この手続きを予算を使い切るまでおこなう。最後の半端な部分は予算制約を満たすように顧客グループからランダムサンプリングをおこなう。また、同じ顧客に二種類のクーポンが配布されないようにするために、一度クーポンを配布した顧客は除くという処理が必要である。

4 実験結果

前節の貪欲アルゴリズムを、薬局チェーンの実販売データに適用して分析をおこなった。このデータは 1996 年から 1998 年の 3 年間における、30,358 人の顧客についての購買履歴である。

今回の実験では、衣料用粉末洗剤の商品カテゴリにおいて、「ブランド A」を製造するメーカーがクーポンを配布することを仮定した。衣料用粉末洗剤のカテゴリでは、無名のブランドも含めると非常に多くの種類のブランドが存在するが、売上トップ 3 ブランドで全シェアの 80% 以上を占める寡占状態にある。そこで今回はこれらの 3 つのブランド (ブランド A, B, C) に限定した。

用いた顧客属性は、(1) 対象ブランド A を購買したかどうか、(2) ブランド B に対する相対価格、(3) ブランド C に対する相対価格、(4) 時系列相対価格、(5) 年齢、(6) 特殊洗剤購買の有無、(7) 洗剤の総購入数量、(8) ブランド A の購入シェア、(9) ブランド B の購入シェア、(10) ブランド C の購入シェア、(11) その他のブランド購入シェア、(12) 医薬品購入シェア、(13) 医療用品購入シェア、(14) 健康食品購入シェア、(15) 日用雑貨品購入シェア、(16) 化粧品購入シェア、(17) ベビー用品購入シェア、(18) 来店あたり購入金額、(19) 来店あたり購入数量、(20) 来店あたり粗利、(21) 洗剤のブランドスイッチ回数、(22) 来店回数である。

問題 1 において導出されたルールは以下の通りである。

1. 1 次元ルール

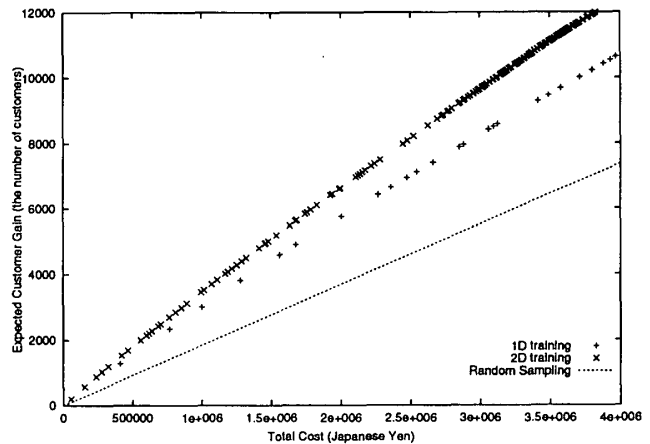


Figure 1: 1D, 2D ルールの効果

- ブランド B シェア = 50% ~ 0.58 でブランド B 相対価格 -200 円なら購入確率 9.0% / 73.0%
 - 日用雑貨シェア = 20% ~ < 40% で通常より 99 円安くすれば購入確率 33.2% / 73.1%
 - 年齢 = 18 ~ 32 でブランド C 相対価格 -200 円なら、購入確率 25.4% / 82.9%
 - ブランド B シェア 8 ~ 25% でブランド B 相対価格 -100 円なら、購入確率 17.8% / 62.0%
2. 2 次元ルール
- ブランド B シェア = 50% ~ 58% & 日曜雑貨品シェア = 30% ~ 41.4% で時系列相対価格 = 100 円なら 購入確率 17.8% / 87.9%
 - 対象ブランド A のシェア = 0% & 粗利 -3198 円 ~ 23 円でブランド B 相対価格 -200 円なら、購入確率 11.0% / 76.1%
 - ブランドスイッチ回数 = 2 & 粗利 -3198 円 ~ 23 円で時系列相対価格 = 87 円なら 購入確率 24.5% / 82.4%

References

- [1] Bawa, K., and Shoemaker, R. W., "The Effect of a Direct Mail Coupon on Brand Choice Behavior," *Journal of Marketing Research*, Vol. 24, 1987, pp.370-367.
- [2] Fayyad, U. M., Shapiro, G. P., and Smyth, P., Uthurusamy, R. Ed., *Advances in Knowledge Discovery and Data Mining*, AAAI Press, 1996.
- [3] Y. Hamuro, N. Katoh, Y. Matsuda and K. Yada, Mining Pharmacy Data Helps to Make Profits, *Data Mining and Knowledge Discovery*, Vol.2, No.4, (1998), pp.391-398.
- [4] Mulhern, F. J., and Padgett, D. T., The Relationship Between Retail Price Promotions and Regular Price Purchases, *Journal of Marketing* Vol. 59, 1995, pp.83-90.
- [5] Quinlan, J.R., See5/C5.0, <http://www.rulequest.com/>, Rulequest Research, 1999.