

人気感度と先行ポイントを利用した 顧客セグメント化とその応用

白井 康之, 森田 裕之, 中元 政一, 小山 聡

1. はじめに

インターネットを活用した商取引は、日々進化を遂げている。商品やサービスをクーポンで提供する際に、大幅な割引価格や特典を顧客に提供する一方で、販売期間、利用条件、販売成立条件などを厳しく設定するフラッシュマーケティングも近年の新たな方法の1つである。この方法では、出品店舗は数日程度のクーポン購買可能期間を設定し、その取引が成立する最小購買数量を設定する。その際、クーポンは通常より有利な購買条件（例えばより大きな値引き）を設定することで魅力を高める。一方、顧客は、そのクーポンを購買したければ、最小購買数量を超えて、購買可能期間内に取引を成立させる必要があるため、自ら他人に購買を促す宣伝をする。この相互依存関係を利用することで、例えば大手サイトの1つであるグルーポンは業績を拡大してきた。現状の国内市場は、[1]によれば、グルーポンとポンパレが共に月間15億円前後の販売額でトップシェアを二分し、それ以外の8サイト（2013年2月時点）の販売実績を合計しても、グルーポンの半分程度のシェアである。しかしこの業界の市場占有率は未だ流動的であり、どの企業にもチャンスが存在している。そのため顧客の動向を正確に捉え、適切なセグメンテーションを用いてCRMを展開することは重要な課題である。

本稿での分析対象データは、上述の共同購入クーポンサイトで、2011年7月1日から1年間にサンプリングされたクーポン購入履歴データである¹。個人属性

データとして年齢・性別が利用可能であるが、属性のバランスやプライバシーを考慮したサンプリングが行われている。一方、クーポンに関しては、名称、定価、販売額、割引率、ジャンル、そして内容説明文などが利用可能である。ただし上述のような取引成立の条件となる最小購買数量のデータはなく²、完売もあらかじめ最大購買数量が設定されていて、全部購買された場合のみ完売記録が利用可能である。本データの分析に関し、著者らが留意した点は以下のとおりである。

1. 本ビジネスに関係するプレイヤーは、サイト運営企業、出品業者、購入者である。
2. 各出品業者は、一時点に1クーポンのみ出品可能である。また同じ商品を再出品しても別クーポンとして識別されている。
3. データはサンプリングされているため、顧客の購買や閲覧の総計を算出しても、顧客行動全体を表現しているわけではない。
4. クーポンの割引率分布は、50%以上が大半で、全体的に高く設定されているが、値引きの根拠や妥当性が不明であり、これ自体を定量的に解析することは困難である。
5. クーポンの販売額は96円～10万円、平均が4338.63円で、標準偏差は5459.36である。一方、顧客の年間平均購入数は10.75、標準偏差が15.83で偏差は大きい。
6. サイト運営企業の収益は、販売額の一定割合と考えられるが、100%割引クーポンの販売があることを考えれば、別の収益構造も予想される。

以上のように、購入額ならびに割引率は商品による変動が極端に大きく、その妥当性も不明であるため、購入額を前提とした分析は困難である。また、期間が指定された商品出品であるため、商品の継続的な販売を

しらい やすゆき, なかもと まさかず
科学技術振興機構 ERATO 湊離散構造処理系プロジェクト
〒060-0814 札幌市北区北14条西9丁目 北海道大学情報科学研究科工学系 C306
もりた ひろゆき
大阪府立大学現代システム科学域知識情報システム学類
こやま さとし
北海道大学大学院情報科学研究科

¹ 平成24年度データ解析コンペティションにおいて、経営科学系研究部会連合協議会より提供されたデータ。

² 当該サイトでは最小購買数量が1である場合も多い。

前提とした分析も困難である。そこで分析の方針としては、サイト運営企業の観点から、単発販売クーポンを前提とした顧客とクーポンの購買の関係性に焦点を当てるものとする。この場合、顧客を評価する主な観点は購買回数や購買量となるが、購買を掘り下げて考えると、同じ購買でも多くの顧客に人気のあるクーポンを買っているケースもあれば、そうではない個性的なクーポンを購入しているケースもある。人気クーポンに素早く反応する顧客は当然重要だが、必ずしも人気のあるクーポンのみが出品されるわけではないので、個性的なクーポンに反応する顧客の動向を知ることが重要となる。本分析で対象とするクーポンサイトの特性を考えれば、将来どのような内容のクーポンが出品されるかを事前に予測することは難しいが、顧客が人気クーポンあるいは個性的なクーポンに対して、どのような購入態度を持っているかを分析することは可能である。すなわち、将来出品されるクーポンの種別や価格といった属性ではなく、クーポンの人気度合い、ならびにクーポンの人気度合いへの顧客の反応を分析することにより、顧客行動の将来予測ならびにこれに基づくマーケティング方法を検討することとした。

そこで本稿では、顧客の人気クーポンへの反応度合いとして、検索エンジンなどでサイトの重要性を計算する方法である Hypertext Induced Topic Selection アルゴリズム [4] (以下 HITS と呼ぶ) の応用を提案する。HITS では、通常、ユーザのサイト巡回履歴から、サイト間の参照度 (hub score) と被参照度 (authority score) を計算して、それらのスコアに基づき、「重要度」の高いページを抽出する。その際、サイト間の接続関係は一般グラフで表現できるが、本データでは、各顧客と各クーポンをノードで表現すると、購買関係は顧客ノードとクーポンノードの二部グラフで表現することができる。この二部グラフに対して HITS を適用し、顧客の人気クーポンに対する反応度合い (以下「人気感度」と呼ぶ) とクーポンの人気度合い (以下「クーポン人気度」と呼ぶ) を算出することが、今回新規に提案する 1 つのポイントである。

次に、発売直後にほかの顧客に先んじて購買する顧客と、一定の時間が経過し、ほかの顧客の動向を見てから購買する顧客では、クーポンサイトへの参加態度や購買の動機が大きく異なっている点に着目する。図 1 は、発売開始の正午からの時間経過とクーポンの購買数量を集計した結果である³。発売開始後 2 時間まで

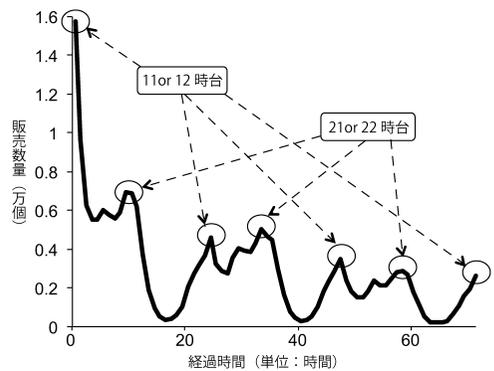


図 1 発売開始後の購買数量の変化

が最大ピークで、その後別のピークが存在し、24 時間周期で繰り返す。本稿で対象とするフラッシュマーケティングでは、一般に購買可能期間が短く、かつ完売数の設定もあるため、購買のタイミング、すなわち、早期に購買するか、あるいは他者の購買反応を見て購買するかといった相違は、セグメント化の重要な 1 つの次元となる。そこで本稿では、購買順序の相対的位置関係を「先行ポイント」として定義する。

本分析では、以上のような人気感度と先行ポイントの 2 つの次元で顧客を 4 つにセグメント化し、いくつかの分析結果から、当該フレームワークの重要性を考察する。

2. 顧客とクーポンに関する指標

本研究では、以上述べたように、顧客とクーポンに関する特徴量として、HITS アルゴリズムを応用した顧客の人気感度ならびにクーポン人気度を導入する。また、顧客の購入手動を特徴づけるものとして、先行ポイントを定義する。はじめに、分析対象とするデータの選別方法について触れ、次に各指標の計算方法と基礎的な分析結果を記す。

2.1 分析データの概要

本分析で使用したデータは、全データの一部がサンプリングされたものであり、その結果として販売数が 1 つのみであるクーポンや、購入回数が 1 回のみである顧客がデータ中には多数存在している。以降に示す分析では、これらがノイズとなる可能性があるため、顧客とクーポンによって二部グラフを構成し、その際各ノードのエッジ数が 2 以上となる部分グラフを抽出した。抽出された部分グラフに含まれる顧客 7,573 人、ならびにクーポン 6,646 件を分析対象とする。なお、提供データに含まれる商品ジャンル区分は、主にサイトでの実運用を目的としたもので、分析上ノイズを含

³ 本サイトは大部分のクーポンを正午に発売している。

表 1 顧客・クーポン別の購入・販売数分布

購入頻度	顧客数	販売数	クーポン数
2 回以上 5 回未満	2,663	2 個以上 5 個未満	3,131
5 回以上 10 回未満	2,439	5 個以上 10 個未満	1,719
10 回以上 15 回未満	1,272	10 個以上 15 個未満	731
15 回以上 20 回未満	573	15 個以上 20 個未満	344
20 回以上 25 回未満	283	20 個以上 25 個未満	209
25 回以上 30 回未満	160	25 個以上 30 個未満	109
30 回以上	183	30 個以上	403
合計	7,573	合計	6,646

むことが懸念されたため、「グルメ」「リラクゼーション」「医療・ビューティ」「宅配食」「旅行」「美容品」の 6 カテゴリに再分類を行った。

顧客別の購入数の分布、ならびにクーポン別の販売数の分布は表 1 のとおりである。

2.2 人気感度とクーポン人気度

Web ページのランキング手法としてよく知られた方法として、PageRank[3] や HITS アルゴリズム [4, 5] がある。このうち、HITS アルゴリズムは、すべての Web ページを authority と hub の 2 つから構成されるものと考え、「良い hub からリンクされているページは良い authority である」、「良い authority にリンクしているページは良い hub である」といった「重要度」に関する再帰的な定義をもとに、Web ページ間のリンク構造から各ページのランキングを行う。

本稿では、この HITS アルゴリズムを顧客とクーポンの関係に応用する。ただし、消費行動においては、Web ページのように、「良い」という概念が明確ではない。このため、「人気クーポンに敏感な顧客により購入されているクーポンは人気クーポンである」、「人気クーポンを購入している顧客は、人気クーポンに敏感な顧客である」と考え、顧客に関しては「人気クーポンへの敏感度合い（人気感度）」、またクーポンに関しては「クーポン人気度」といった観点からランキングを行う。

具体的な方法は以下のとおりである。顧客 i の「人気感度」を u_i 、クーポン j の「クーポン人気度」を m_j とする。このとき、人気感度ベクトル $\mathbf{u} = (u_1, u_2, \dots, u_{n_u})^T$ 、ならびにクーポン人気度ベクトル $\mathbf{m} = (m_1, m_2, \dots, m_{n_m})^T$ を求めることを考える（ここで、 n_u は顧客数を、 n_m はクーポン数を表す）。

\mathbf{u} ならびに \mathbf{m} は、初期値 $\mathbf{m}^{(0)}$ を与えた上で、以下の式を反復計算した結果として得ることができる。ここで N は繰り返し回数を表し、 $\mathbf{u}^{(N)} = (u_1^{(N)}, u_2^{(N)}, \dots, u_{n_u}^{(N)})^T$ 、 $\mathbf{m}^{(N)} = (m_1^{(N)}, m_2^{(N)}, \dots, m_{n_m}^{(N)})^T$ であるとする。また、以下で、 $i \rightarrow j$ は顧客 i がクーポン j を購入したことを

表す。

$$\begin{aligned} \bar{u}_i^{(N+1)} &= \sum_{j:i \rightarrow j} m_j^{(N)} \quad (i = 1, \dots, n_u) \\ \bar{m}_j^{(N+1)} &= \sum_{i:i \rightarrow j} u_i^{(N+1)} \quad (j = 1, \dots, n_m) \\ \mathbf{u}^{(N+1)} &= \bar{\mathbf{u}}^{(N+1)} / \|\bar{\mathbf{u}}^{(N+1)}\|_2 \\ \mathbf{m}^{(N+1)} &= \bar{\mathbf{m}}^{(N+1)} / \|\bar{\mathbf{m}}^{(N+1)}\|_2 \end{aligned}$$

ここで $\|\cdot\|_2$ は L_2 ノルムを表す。すなわち、上式の各 $\mathbf{u}^{(N+1)}$ 、 $\mathbf{m}^{(N+1)}$ は各要素の二乗和が 1 になるように正規化されている。

さて、ここで顧客 i のクーポン j の購入有無（0 または 1）を a_{ij} ($1 \leq i \leq n_u, 1 \leq j \leq n_m$) とし、

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n_m} \\ a_{21} & a_{22} & \dots & a_{2n_m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_u 1} & a_{n_u 2} & \dots & a_{n_u n_m} \end{pmatrix}$$

とすると、

$$\mathbf{u}^{(N+1)} = \mathbf{A} \mathbf{m}^{(N)}, \quad \mathbf{m}^{(N+1)} = \mathbf{A}^T \mathbf{u}^{(N+1)}$$

と表すことができ、

$$\mathbf{u}^{(N+1)} = \mathbf{A} \mathbf{A}^T \mathbf{u}^{(N)}, \quad \mathbf{m}^{(N+1)} = \mathbf{A}^T \mathbf{A} \mathbf{m}^{(N)}$$

となる。この際、 $\mathbf{u}^{(N)}$ ならびに $\mathbf{m}^{(N)}$ は、 $\mathbf{A} \mathbf{A}^T$ 、 $\mathbf{A}^T \mathbf{A}$ の主固有ベクトルにそれぞれ収束することが知られており、本稿では、これらの（正規化された）主固有ベクトルを \mathbf{u} 、 \mathbf{m} とした。

顧客 7,573 人、ならびにクーポン 6,646 件に対して、上記に基づき、 \mathbf{u} 、 \mathbf{m} を算出した。また、6 カ月時点での状況を見るために、顧客に対する前期人気感度を算出した。具体的には、2011 年 7 月、8 月に購買実績がある顧客（3,936 人）に対して、2011 年 12 月末時点の購入実績を基に人気感度として算出した。

なお、顧客母集団での人気感度を降順にソートして順位をつけたものを、以下、人気感度ランクと呼ぶ。人気感度ランク 1 位の顧客は最も人気に敏感な顧客であるといえる。

2.3 先行ポイント

次に、顧客の購入手動における指標として、顧客 i の先行ポイント P_i を定義する。顧客 i の購入クーポンの集合を I_i とし、クーポン $k \in I_i$ における顧客 i の時間順の購入順位を $r_k(i)$ とする。また、クーポン k の全顧客による購入数を M_k とする。例えば、あるクーポン k について 100 名の購入者がいて、顧客 i の購入が時間順にみて 10 番目であったとすると、 $M_k = 100$ 、

$r_k(i) = 10$ となる。

先行ポイント P_i は顧客 i より後の購入数と前の購入数の差分の和を購入クーポン数で正規化したものとして、以下のように定義する。

$$P_i = \frac{1}{|I_i|} \sum_{k \in I_i} \{(M_k - r_k(i)) - (r_k(i) - 1)\}$$

先行ポイントは 0 を平均として、負になる顧客は、顧客自身の購入が全体平均から見て遅れていることを表し、正になる顧客は、ほかの顧客に先んじて購入を行っているといえる。

また、人気感度同様、6カ月時点での状況を見るために、2011年7月、8月に購買実績がある顧客(3,936人)に対して、2011年12月末時点の購入実績を基に前期先行ポイントを算出した。

なお、顧客母集団での先行ポイントを降順にソートして順位をつけたものを、以下、先行ポイントランクと呼ぶ。先行ポイントランク1位は、最も他ユーザに先行してクーポンを購入する顧客であるといえる。

3. 分析結果

3.1 基礎指標の分布

以下、人気感度、クーポン人気度、ならびに先行ポイントに関して、基礎的な分析結果を示す。

人気感度

表2に、人気感度の高い顧客トップ10を示す。上位は、人気クーポンを好んで購入するタイプで占められる。例えば、トップの顧客は、クーポン人気度で1, 8, 20, 47, 98, 618位の商品を購入している。また、第2位以降の上位顧客も、ジャンルはそれぞれ異なるものの、クーポン人気度の高い商品を複数回にわたって購入を行っている。

表3は、人気感度の性別、年代別比較である。ここで見るように、全般的には女性のほうが男性よりも人気感度が高く、特に、中年層以降の女性は、結果として人気の高いクーポンを購入する頻度が高いことがわかる。

また、半年時点での人気感度とその後のアクティビティの間にどのような関係があるかを見たのが表4である。ここで見られるように、6カ月経過時点での人気感度が高ければ高いほど、7カ月以降のアクティビティ(購入回数)も高くなっていることがわかる⁴。

⁴ 有意水準1%でこの差が有意であることを確認している。なお、「大」、「小」はそれぞれ上位半数、下位半数を表している(ただし、同順位の場合は上位に含めている)。

表2 人気感度の高い顧客トップ10

性別	年齢	購入回数	ユーザの 人気感度	ジャンル別商品購入回数						購入 総額 (定価)	購入 総額 (購入額)
				グルメ	リファケイション	医療・ビューティ	宅配食	旅行	美容品		
m	36	11	0.11348	2	0	0	0	1	8	81710	23025
f	36	33	0.09160	0	8	6	3	0	16	298125	76353
m	36	28	0.08995	19	0	0	1	1	7	124274	38772
f	55	50	0.08380	22	5	0	14	5	4	301307	131878
f	38	11	0.07688	1	1	0	2	1	6	115335	30422
f	52	30	0.07675	23	0	0	3	0	4	173451	54692
f	50	32	0.07412	25	1	0	0	2	4	246892	111273
f	34	13	0.07326	4	0	0	1	0	8	65947	22447
f	28	40	0.07277	22	5	2	5	3	3	412171	152693
f	42	15	0.07259	3	1	0	2	3	6	170490	45115

表3 人気感度の平均ランクの比較

性別	20代以下	30代	40代	50代	60代以上
女性	4,440	4,071	3,450	3,195	3,032
男性	4,495	4,066	3,931	3,695	3,834

表4 6カ月経過時点の人気感度と7カ月目以降の平均購入回数

分類	7カ月目以降の平均購入回数
人気感度大	5.306
人気感度小	3.201

クーポン人気度

表5に、クーポン人気度のトップ10を示す。表中、割引率や定価も示しているが、クーポン人気度と割引率や定価との間には有意な関係は見られなかった(したがって必ずしも価格面で人気が決まっているわけではない)。人気商品には、美容品やグルメジャンルのものが多くなっているが、どのような特徴が認められるのかは今後の分析課題である。

また、詳細なデータは省略するが、クーポン人気度ランクをジャンル別に平均をとると、美容品やグルメ、宅配食では全体的に人気商品が多いのに対して、医療・ビューティや旅行は平均して人気商品が少なくなっている。これらのジャンルでは、特定の期間・用途・顧客を対象とした商品が多く、このような商品特性がクーポン人気度に大きく関与していることも考えられる。クーポン人気度は必ずしも高ければ良いというわけではないが、出品者がこうした商品の販売状況や位置づけを把握しておくことは重要であると考えられる。

先行ポイント

図2に先行ポイントの分布(顧客数)を示す。先行ポイントは、0を中心としてほぼ左右対称に分布している。また、半年時点での先行ポイント(前期先行ポ

表 5 クーボン人気度のトップ 10

ジャンル	商品名	販売数	クーポン人気度	割引率 (%)	定価
美容品	ウルオイト美容マスク	463	0.58400	78	7,200
グルメ	バームクーヘン 1260 円	353	0.17600	50	1,260
グルメ	ワッフルセット 54%OFF	221	0.16700	54	1,100
旅行	朝食付きペアチケット	290	0.16000	55	17,000
美容品	ホホワイトモイスチャークリーム	251	0.15700	81	7,980
グルメ	プリン 6 個 1860 円→ 500 円	198	0.13700	73	1,890
美容品	エクスパイアージュ	180	0.13700	92	40,000
美容品	プラチナコラーゲンゲル	162	0.13500	64	5,565
グルメ	つけ麺プラストッピング 2 品 450 円	192	0.13400	51	920
宅配食	手包みジャンボ餃子 50 個	215	0.13400	50	5,750

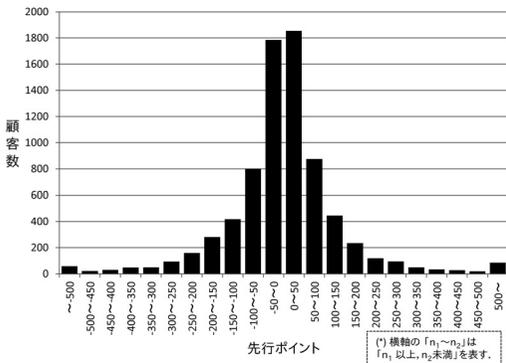


図 2 先行ポイントの分布

イント) とその後のアクティビティの間にどのような関係があるかを見たのが表 6 である。本結果から明らかかなように、前期先行ポイントが高ければ高いほど、7 カ月目以降のアクティビティも高くなっている⁵。先行ポイントの高い顧客は、常にさまざまな商品をチェックしている顧客であり、このような熱心な層が長い期間での優良顧客になる可能性を持っているといえる。また、逆に、先行ポイントの低い顧客で、かつ総販売数の多い商品を購入している場合には、他者追随型の消費行動となっていることも想定されるため、自発的に商品購入を決断する優良顧客へ移行させるためには、何らかのモチベーションを付与することが必要であると推測できる。

人気感度と先行ポイント

すでに示したように、前期人気感度、前期先行ポイントと 7 カ月目以降の平均購入回数との間にはそれぞれ有意な関係がある。さらに、これらの情報を組み合わせることにより、より精緻な購入回数予測が可能であ

⁵ 有意水準 1% でこの差が有意であることを確認している。なお、「大」「小」の定義は、「人気感度」と同様である。

表 6 前期先行ポイントと 7 カ月目以降の平均購入回数

分類	7 カ月目以降の平均購入回数
先行ポイント大	4.444
先行ポイント小	4.061

表 7 前期先行ポイント・人気感度と 7 カ月目以降の平均購入回数

	先行ポイント大	先行ポイント小
人気感度大	5.531	5.069
人気感度小	3.304	3.103

る。表 7 に、前期先行ポイント、前期人気感度の組合せと、7 カ月目以降の平均購入回数の関係を示した。表から見られるように、これら 2 つの指標を組み合わせることにより、例えば、人気感度が大きくかつ先行ポイントが大きいグループは、ともに小さいグループに比べて、ほぼ倍近い購入回数を期待することができる。

また、図 3 は、人気感度ランクと先行ポイントランクの 12 カ月経過時点での分布を示したものである。この図より、人気感度ランクと先行ポイントランクはほぼ独立な変数として扱うことができ、人気感度の高い(順位の小さい)グループ、すなわち人気商品に敏感な顧客の中でも、先行ポイントの高いグループ(人気商品を先行して購入するグループ)と先行ポイントの低いグループ(他者追随型のグループ)が存在していることがわかる。

前期購入回数との比較

上記で示した変数とは別に、6 カ月経過時点(前期)の購入回数は、後期の購入回数の有力な説明変数となることが想定される。

実際、前期の購入回数で上位、下位に分割すると、前者の 7 カ月目以降の平均購入回数は 5.817 回、また、

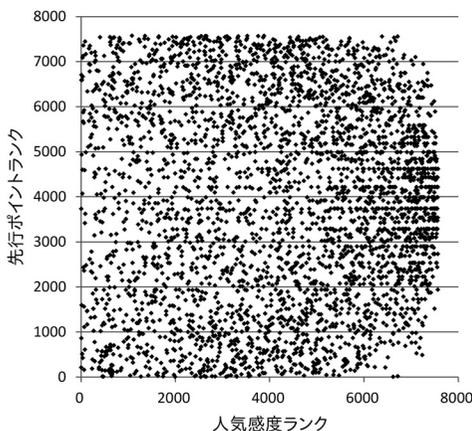


図3 人気感度ランクと先行ポイントランクの分布

表8 先行ポイント・人気感度と前期購入回数の組合せによる7カ月目以降の平均購入回数

	先行ポイント大	先行ポイント小
前期購入回数大	6.046	5.570
前期購入回数小	2.466	2.466

	人気感度大	人気感度小
前期購入回数大	6.136	4.997
前期購入回数小	2.560	2.435

後者の7カ月目以降の平均購入回数は2.466回となっている。

ここで、上記で提案した人気感度や先行ポイントと組み合わせた結果が表8である。これらの変数を組み合わせることによってさらにその後の購入回数を精度良く予測できることがわかる。さらに、前期購入回数大、先行ポイント大、人気感度大で組み合わせると7カ月目以降の平均購入回数は6.32となる。以上のことから、人気感度や先行ポイントは、前期購入回数とは独立した指標として、特に将来の優良層の識別において、非常に意味のある指標となっていることがわかる。

3.2 応用分析

本項では、顧客の人気感度、クーポン人気度、ならびに先行ポイントを用いたマーケティング展開への応用に関して考察する。

ここでは、人気感度ランクと先行ポイントランクを利用してセグメント化を行う。まず、継続的に取引をしている顧客に限定するため、2011年7～8月に購買履歴の存在する顧客を(3,936人)を対象とする。この中に、後期(2012年1月以降)に購買履歴がない顧客が373人存在している。そこでまず、この373人とそれ以外の3,563人を、それぞれ休止顧客、継続顧客

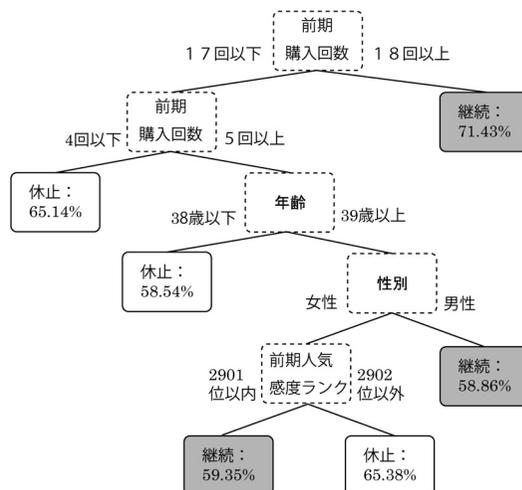


図4 休止と継続を分類する決定木モデル

と定義し、分類要因を決定木分析によって考察する。

次に、前述した前期人気感度ランク、前期先行ポイントランクが定義されている2011年7～8月に購買履歴のある継続顧客に対して、1年経過時点での人気感度ランク、先行ポイントランクを比較して顧客の変化について分析を行う。

まず決定木分析では、目的変数は、休止顧客と継続顧客を記述したクラスであるが、クラス間の人数の偏りが大きいため、継続顧客から373人をランダムサンプリングし、両クラスを同数の状態として入力する。説明変数は、6カ月以内の行動データから性別、年齢、前期人気感度ランク、前期先行ポイントランク、前期の月ごとの利用日数の近似直線の傾き(以下、前期トレンドと呼ぶ)、そして前期購入回数を与える。決定木の計算は、Weka⁶のJ48アルゴリズムを用い、検証方法は5回の交差検証法を採用した。枝狩りのパラメータについては、入力サイズに合わせて調整し、それ以外の各種パラメータはデフォルト値を採用した。検証による結果の正答率は57.239%で、休止顧客と継続顧客のF-Measureはそれぞれ、0.551と0.592であった⁷。出力されたモデルは図4のようにになっている。図中では、分岐ノードを破線で、葉ノードは実線で表し、継続クラスはノードをグレーにしている。モデルの内容を見ると、前期購入回数が重要な要素の1つとなって

⁶ <http://www.cs.waikato.ac.nz/ml/weka/>

⁷ ここでの分析結果は、継続顧客のサンプリングにも依存する。本分析においては、ランダムサンプリングを10回行って分析結果の差異を比較したが、正答率の差は最大で3%程度であり、ツリーに出現するノードはほとんど差がなかった。

おり、それが6カ月で18回以上存在すると継続の可能性が高く、4回以下であると休止の可能性が高いことがわかる。実際、目的変数を7カ月目以降の継続と休止にした場合、6カ月時点での購入回数が非常に多い場合(18回以上)や少ない場合(4回以下)では、それぞれ継続と休止の可能性が高いことは直感的にも明らかであろう。一方、その中間にある場合には、39歳以上では、男性が継続の可能性が高く、女性はさらに前期人気感度ランクが高いと継続になる可能性が高くなっている。本分析ではサンプル数が限られているため、精度の高い決定木を構築することは困難であるが、本論文で導入した人気感度ランクは、前期購入回数から後期のアクティビティがトリビアルではない顧客に対して、後期の購入有無に関する説明力の高い変数となっていることが想像できる。

次に人気感度ランクと先行ポイントランクを、それぞれ x 軸と y 軸として、各軸を2分すると、4つのセグメントができる。これらのセグメントを、ロジャースのイノベーションの普及における過程⁸を参考に以下のように定義する [7]。

アーリーアダプタ (以下, EA) 先行して人気の商品を購入する顧客。購買リーダー的な存在である。

アーリーマニア (以下, EM) 個性的なクーポンを早期に購入する顧客。一見EAの行動と類似するが、フォロワーが少ない。

レイトマジョリティ (以下, LM) ほかの顧客の反応を観察してから人気クーポンを購入する顧客。購買フォロワー的な存在である。

ラガード (以下, LD) 個性的な商品をゆっくり購入する顧客。タイミングや嗜好するクーポンから保守的態度である。

これらのセグメントを利用し、顧客の6カ月時点と12カ月時点での、人気感度ランクと先行ポイントランクによるセグメントのポジション変化を分析する。その際のセグメント区分は、各軸のメディアンを採用する。

表9は、両時点での各セグメントの性別人数を比較したものである。全体としては、性別を問わず、各セグメントの総数は安定しているが、若干の傾向としては、男性はLMが減少して、LDが増加しているようである。これは比較的男性のほうが、人気クーポンであるよりは、個性的なクーポンを好む傾向を表してい

表9 顧客セグメントと性別人数

	EA		EM		LM		LD	
	女性	男性	女性	男性	女性	男性	女性	男性
半年時点	532	373	505	369	489	387	492	416
1年時点	558	388	481	354	495	340	484	463

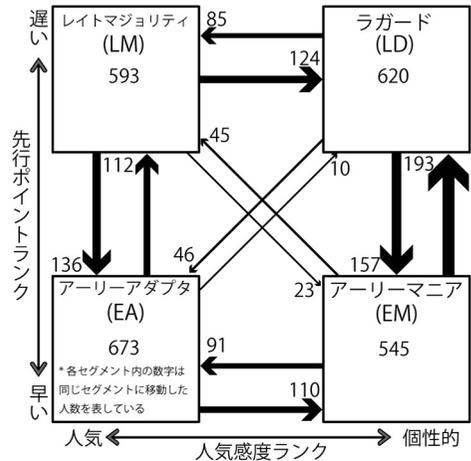


図5 人気感度ランクと先行ポイントランクによるセグメント

るのかもしれない。

次に各顧客のセグメント移動パターンと各該当人数を表しているのが、図5である。図中の四角のノードが、各セグメントを表し、中の数字は両時点で同じセグメントに属した人数を表している。また、セグメント間の移動人数は有向枝に付随した数値で表されており、有向枝の太さは、移動人数の大きさを相対的に表現している。セグメント間の移動について見てみると、先行ポイントランクのみ、または人気感度ランクのみを変化させる縦と横の移動は、ある程度存在するが、両方を同時に変化させる斜め方向の移動は圧倒的に少ない。前期から後期に至る一度の移動しか確認できていないため、確実なことはいえないが、セグメント人数が安定していて、斜めの移動が少ないことを考慮すると、人気感度ランクおよび先行ポイントランクが、顧客の購買態度を表現する次元になっており、短期間においては、いずれかの態度が変化しても、両方を同時に変更する急激な変化は少ないことがわかる。今後さらに継続的に、数期間の顧客行動変化を確認することができれば、より結果を明らかにすることが可能であろう。

これらの移動顧客が、どのように全体的なパフォーマンスに影響するかを分析するための1つの例として、

⁸ 本稿では、クーポン購買に対する態度とそのタイミングで顧客を区分する際に、[7]よりアーリーアダプタ、レイトマジョリティ、そしてラガードを顧客分類として引用し、新たにアーリーマニアというグループを定義した。

表 10 各移動グループと全体の販売数の回帰モデル

変数	係数	標準誤差	t 値	P 値
切片	191.687	15.655	12.245	5.21E-29
EA⇒EA	2.826	0.370	7.645	1.97E-13
EA⇒LM	6.781	1.839	3.687	2.62E-04
EM⇒EA	14.094	1.862	7.567	3.29E-13
EM⇒EM	3.122	0.909	3.435	6.62E-04
EM⇒LM	10.719	2.833	3.784	1.81E-04
LD⇒EA	12.614	2.611	4.831	2.03E-06
LD⇒EM	6.340	1.801	3.520	4.87E-04
LD⇒LM	9.555	2.136	4.473	1.04E-05
LM⇒LD	-4.374	1.716	-2.548	1.12E-02

移動グループに顧客を分け、その取引人数と全体の販売量の関係を回帰分析で確認する。目的変数は日々の販売量、説明変数は、各日の各移動グループの取引人数である。モデルは変数減少法を用いて作成し、最終的に有意な説明変数として残った変数が表 10 に示されている。重回帰モデルの重決定 R^2 は 0.767 で 1% 有意である。

出現した変数の回帰係数は、LM⇒LD を除いてすべて正の値である。特に、EM⇒EA や LD⇒EA は回帰係数が比較的大きく、日々の販売量に対して正の影響力が大きいことがわかる。これらの移動グループは、前期における購買態度から、後期にはより人気に敏感になり、先行して購買行動を行うようになった顧客グループを表しており、このような顧客行動の変化が全体的な販売数を増加させる要因になっていると考えられる。したがって、販売数を増大させるためには、EM あるいは LD の顧客を EA に誘導するようなプロモーションは 1 つの有効な手段となりうる可能性がある。例えば、クーポンの中でも比較的 EM や LD のセグメントの顧客に人気があるようなクーポンを選択して販売を強化することが可能であれば、それは日々の取引量の増大に正の影響をもたらすものと予想される。このように、移動グループから顧客の購買態度や特性を明確化し、全体的な顧客種類のバランス、およびプロモートする顧客のターゲティングに本提案方法は有用だと考えられる。具体的なプロモート方法については、各顧客グループが好むクーポンを確認しながら、それらの出品組合せやタイミングをコントロールすることで可能になるだろう。

4. まとめ

本稿では、某共同購入クーポンサイトのデータを対象として、HITS アルゴリズムを応用して、各顧客の人気感度と各クーポンの人気度を算出する方法を提案し、また、購買行動のタイミングを特徴づけるものとして先行ポイントを定義した。これら人気感度と先行ポイントにより、顧客の購買行動を 4 つのセグメントにポジショニングするとともに、時間経過による変化についても分析を行った。人気感度や先行ポイントは、顧客の行動を特徴づけるうえで、有意な変数であり、将来の購買予測のほか、さまざまなプロモーションへの応用が可能であると考えられる。

今後の課題として、今回の人気感度や先行ポイントだけでなく、クーポンの購入種類、あるいは購買の多様性を新たな次元として追加した分析を検討している。

謝辞 本分析は、経営科学系研究部会連合協議会主催「平成 24 年度データ解析コンペティション」で提供されたデータを使用して行ったものである。関係各位に謝意を表する。

参考文献

- [1] 共同購入クーポンの検索サービス「クーポン JP」, <http://coupon-jp.com/>
- [2] 野村早恵子, 小山聡, 早水哲雄, 石田亨: WEB コミュニティ発見のための HITS アルゴリズムの分析と改善. 電子情報通信学会論文誌 D-I, J85-D-I(8), 2002.
- [3] S. Brin and L. Page: The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Seventh International World-Wide Web Conference*, 1998.
- [4] J. Kleinberg: Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM (JACM)*, **46**(5), 1999.
- [5] 宇野裕之: ウェブページのランキング技術. オペレーションズ・リサーチ, **57**(6), 2012.
- [6] D. Gibson, J. Kleinberg and P. Raghavan: Inferring Web Communities from Link Topology. *Proc. of the 9th ACM conference on Hypertext and Hypermedia*, 1998.
- [7] E. M. Rogers (三藤利雄訳): イノベーションの普及, 翔泳社, 2007.