

支払い方法選択行動分析による “高価値”顧客の発掘

山口 景子, 中島 上智, 岡 賢一

1. はじめに

クレジットカード会社が得られる収益は主に二種類存在する。ひとつは加盟店から得られる収益、もうひとつはカード会員である顧客から得られる収益であり、後者は更に年会費といった固定収益と、利子・手数料といった変動収益に分類される。しかし「年会費無料」を掲げた顧客獲得競争が激化する今日においては、顧客からの固定収益はそれほど多くを期待することは出来ない。このような状況下において、クレジットカード会社が収益を確保し続けるためには、より多くの変動収益、特に利子収益をもたらしてくれる顧客を“高価値”顧客として選別し囲い込む必要がある。

ここで、どの支払い方法を選択するかという問題は、利用者の利用履歴や属性等と密接に関連する。したがって、カード利用者の支払い行動を利用履歴データを用いて個別に分析し、顧客の“高価値”性を示す指標を推定し、その結果から顧客を選別する必要がある。この既存顧客に関する分析を第2節で行った。

ところで、支払い行動を個別に分析するためには、ある程度のデータ量が必要である。クレジットカード会社はデータ量が豊富に存在する既存顧客だけでなく、カードを作成したばかりという新規顧客やカードを所有しているだけという未利用顧客等、データ量が圧倒的に少ないあるいは全く存在しないという顧客も抱えている。これらの顧客は、データ量の不足ゆえに既存顧客のように利用履歴データを用いて個別に分析することは出来ない。しかし、何らかの手法で“高価値”顧客かどうかを判別することで、プロモーションの効率化が図れるのではないかと考えた。そこで、利用履歴データ以外のデータと顧客の“高価値”性を示す指

標とを関連付けることで、データ量が不足する顧客の“高価値”性の判別となる指標の抽出を試みた。この分析は第3節において展開する。

本研究では「カード利用時に利子付きの支払い方法を選択するまたは選択する可能性のある顧客＝“高価値”顧客」と定義して分析を進めていく。

2. 既存顧客の分析

2.1 データの概要

データは平成16年度データ解析コンペティション提供データを使用した。データ期間は2002年1月から2003年12月までの24ヶ月間である。本分析では支払い方法選択行動を問題としているため、カード利用日時、各カード利用時点における支払い金額、支払い方法（一括・リボルビング・複数回払い等）、ショッピング利用かキャッシング利用か等が含まれる日別利用明細データを使用した。

本研究における分析を行う上での前提として、ショッピング利用とキャッシング利用は同じカード利用でも別種の行動と見なすことにする。その理由としては一回の平均利用額の大きさ、利用発生頻度、利率等がショッピング利用とキャッシング利用とではかなり異なっているという点が挙げられる。一回の平均利用額の大きさ、利用発生頻度、利率等の高低はそのまま各カード利用を特徴付けられる重要な要素であると考えられる。したがって、これらの要素がかなり異なっていることから、同じクレジットカード利用であっても、ショッピング利用とキャッシング利用は別種の利用であると考えるのが適当ではないかと考えた。したがって、以下ではショッピング利用とキャッシング利用に関してデータを分割し分析を行った。

2.1.1 キャッシング利用分析のデータ

2002年の利用履歴データを推定用に、2003年のデータを検証用に使用した。各年において10回以上キャッシングを利用している300名の顧客をサンプル

やまぐち けいこ, なかじま じょうち, おか けんいち
東京大学 大学院経済学研究科
〒113-0033 文京区本郷 7-3-1
受付 05.7.25 採択 05.11.14

として抽出し分析した。

2.1.2 ショッピング利用分析のデータ

キャッシング利用行動には、「手持ちの現金がないが、クレジットカードでは支払うことができない」「急に現金が必要になった」というような緊急性が影響を与える場合が多いと考えられる。しかし、ショッピング利用行動には購買行動の季節性等の特殊な状況が存在する可能性があり、通年でデータ処理してしまっただけではその季節特有の状況を平準化してしまうおそれがある。

そこで、データの期間は出来る限り購買行動の季節性を参考に細かく取った方が望ましいこと、ショッピング行動に影響を及ぼすと考えられる季節的な行事（お中元・お歳暮の購買、クリスマス、卒業・新入学行事に関する購買など）は複数月に渡って長期に行われるものではないということとを考慮し、一ヶ月というデータ期間を設定した²。また、推定用データと検証用データは同じ季節であることが望ましいと考えられるため、2002年のある一ヶ月を推定用に、2003年の同じ月を検証用に使用した。なお、ある一ヶ月のデータを用いた推定のみでは、当該月だけがモデルの当てはまりが良く、他の月では当てはまりが悪いという可能性が存在する。したがって、そうではないことを確認するためにランダムに三ヶ月分抽出し分析を行った。ただし、これらの三ヶ月は連続する月とはならないように抽出した。

この結果3月、6月および12月をサンプル月として抽出し、2002年および2003年の各月においてショッピング利用履歴があり、かつ2002年の各月において10回以上利用履歴のある顧客300名をサンプルとして抽出³した。

2.2 選択行動モデルの設定

本研究では、支払い方法の意思決定に対して選択行

動モデルを当てはめる。通常、カード利用時に顧客が直面する選択肢は多種多様であると考えられるが、本研究では顧客からの利子収入を念頭においているため、単純に「利子付きの支払い方法（高利子収入の支払い方法）」か「利子の付かない支払い方法（低利子収入の支払い方法）」かという二つの選択肢を考える。したがって、選択行動モデルとして二項プロビットモデルを当てはめることが出来る。

ショッピング利用データに関しては、ID番号 $i(i=1, \dots, n)$ の顧客が $t(t=1, \dots, T_i)$ 回目のカード利用時に利子付きの支払い方法を選択した場合には $y_{it}=1$ 、そうでない場合は $y_{it}=0$ というフラグを、キャッシング利用データに関しては、リボルビング払いを選択した場合には $y_{it}=1$ 、一括払いを選択した場合は $y_{it}=0$ というフラグを立てて区別する⁴と、二項プロビットモデルは

$$y_{it} = \begin{cases} 1 & z_{it} \geq 0 \\ 0 & z_{it} < 0 \end{cases} \quad (1)$$

$$z_{it} = \alpha_i + x_{it}'\beta_i + \varepsilon_{it} \quad (2)$$

と書き表せる。 β_i および x_{it} は $k \times 1$ のベクトルであり、 k は説明変数の数である。ここで、 z_{it} は支払い方法の選択に関する潜在変数（例えば、カード利用時における顧客の効用）であり、 x_{it} は潜在変数 z_{it} に影響を与える変数となっている。本研究では、説明変数 x_{it} として各利用時点における支払い金額、累積利用回数、限度額に対する充足率等を考えたが、モデル選択の結果、各利用時点における支払い金額のみを使用することとした。したがって式(2)は

$$z_{it} = \alpha_i + \beta_i \text{PRICE}_{it} + \varepsilon_{it} \quad (3)$$

と表現出来る。ここでは β_i はスカラーである。また ε_{it} は攪乱項であり、互いに独立で同一な標準正規分布に従うと仮定する。

さらに、 α_i および β_i に関して確率変数モデルを仮定する。すなわち、

$$\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} \sim N \left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \begin{pmatrix} \sigma_\alpha^2 & 0 \\ 0 & \sigma_\beta^2 \end{pmatrix} \right) \quad (4)$$
$$\sim MVN(\mu, \Sigma_\mu)$$

と考え、式(4)のパラメータ $\alpha, \beta, \sigma_\alpha^2, \sigma_\beta^2$ を推定した。 $y = \{y_i\}_{i=1}^n, y_i = \{y_{it}\}_{t=1}^{T_i}, z = \{z_i\}_{i=1}^n, z_i = \{z_{it}\}_{t=1}^{T_i}, \theta = (\mu, \Sigma_\mu), \delta_i = (\alpha, \beta_i), v_{it} = \alpha_i + \beta_i \text{PRICE}_{it}$ とすると、モデルの尤度は、

¹ ベイズ的アプローチによる顧客毎のパラメータの推定精度を維持するため、文献[8]で用いられているサンプル数を参考とした。

² 一顧客当たりのショッピング利用回数は、キャッシング利用回数よりもはるかに多い。ゆえに期間を一ヶ月としても推定精度を維持するために必要なサンプル数を十分確保出来る。一方、キャッシング利用分析では、一ヶ月当たりのキャッシング利用回数が少ないという問題から、月ごとに分析することは難しい。

³ ただし、このサンプル抽出ルールでは検証用データにおける利用回数に制約を置いていないため、2003年の各月におけるショッピング利用回数は10回未満となる場合が存在する。

⁴ 推定用および検証用データにおける $y_{it}=1$ の総選択回数と $y_{it}=0$ の総選択回数の比率は、キャッシングデータがおおよそ1:1、ショッピングデータが1:4となっている。

$$f(y|\theta) = \prod_{i=1}^n \prod_{t=1}^{T_i} [1 - F(v_{it})]^{y_{it}} [F(v_{it})]^{1-y_{it}} \quad (5)$$

の形で与えられる。ただし、

$$F(v_{it}) = \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(z_{it}-v_{it})^2}{2}\right\} dz_{it} \quad (6)$$

である。式(4)から $\pi(\delta_i|\mu, \Sigma_\mu) \sim MVN(\mu, \Sigma_\mu)$ であるから、ハイパーパラメータの事前分布をそれぞれ $\pi(\mu)$, $\pi(\Sigma_\mu)$ と表記すると、事後分布の同時密度関数式(7)が得られる。

$$\pi(\theta|y) \propto f(y|\theta) \pi(\mu) \pi(\Sigma_\mu) \prod_{i=1}^n \pi(\delta_i|\mu, \Sigma_\mu) \quad (7)$$

ここで、 $\pi(\mu)$, $\pi(\delta_i|\mu, \Sigma_\mu)$ が正規分布に、 $\pi(\Sigma_\mu)$ の各対角成分がインバースガンマ分布に従うと仮定し、ギブスサンプリングによって式(7)を求める。以下の過程に基づきサンプリングを行う。

- (1) $z, \{\delta_i\}_{i=1}^n$ および θ の初期値を発生させる。
- (2) $z_i|y_i, \delta_i, \theta$ for $i=1, \dots, n$ のサンプリング。
- (3) $\delta_i|y_i, z_i, \theta$ for $i=1, \dots, n$ のサンプリング。
- (4) $\theta|y, z, \{\delta_i\}_{i=1}^n$ のサンプリング。ただし、
 - (a) $\mu|\{\delta_i\}_{i=1}^n, \Sigma_\mu$
 - (b) $\Sigma_\mu|\{\delta_i\}_{i=1}^n, \mu$
 の二種類のサンプリングを行う。

- (5) (2)へ戻りサンプリングを繰り返す。

ここで、過程(2)の $z_i|y_i, \delta_i, \theta$ for $i=1, \dots, n$ のサンプリングは、 z_{it} が切断正規分布

$$z_i|y_i, \beta_i \sim TN_{I(z_{it})(v_{it})} \propto \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(z_{it}-v_{it})^2}{2}\right\} \times \tilde{I}(z_{it}) \quad (9)$$

に従う⁵としてサンプリングを行う。ただし

$$\tilde{I}(z_{it}) = \begin{cases} I[z_{it} \geq 0] & (\text{if } y_{it}=1) \\ I[z_{it} < 0] & (\text{if } y_{it}=0) \end{cases} \quad (10)$$

である。

本分析において興味があるのは、各顧客固有のパラメータ α_i および β_i である。式(3)より、 α_i の推定値が高いほど利子付き支払いを選択しやすいことになるため、定数項 α_i は顧客に固有の「利子に対する許容度」と解釈することが出来る。つまり α_i の値が正である顧客は利子に対して寛容であり、 α_i の値が負である顧客は利子付きの支払い方法を嫌う傾向があると考えることが出来る。一方、各利用時点における支払い金額 β_i は、利子に対する許容度への支払い金額の影響度と解釈出来る。つまり、 β_i の値が高い顧客は、高額の商品を購入する際には利子付きの支払い方法を選

択する傾向にあり、 β_i の値が低い顧客は支払い方法の選択に商品の金額の高低があまり影響しないと考えることが出来る。

2.3 分析結果

2.3.1 推定結果

表1および表2はそれぞれ、キャッシング利用データおよびショッピング利用データを用いて推定されたモデルの統計量である。キャッシング利用データに関しては、推定モデルの U^2 統計量が0.66、平均正答確率が85.36%、正答率が90.68%であり、推定の段階では説明変数 x_{it} が1変数であるのにも関わらずそれなりに良い当てはまりを示している。ショッピング利用データに関しては、 U^2 統計量が0.67、平均正答確率が86.01%、正答率が90.18%であり、キャッシング利用データ同様それなりに良い当てはまりを示していると言えよう。この統計量は2002年12月のデータを用いて算出したものであるが、他の二ヶ月も同様の当てはまりを示した⁶。

表3および表4は推定値 α および β の各種統計量を示している。キャッシング利用データおよびショッピング利用データにおいて、推定値 α および β はともに有意水準95%で有意であり、顧客に固有の利子に対する許容度と利用金額の大きさが利子付き支払い選択行動に影響を与えていると言える。また、 α および β の分散の推定量である σ_α^2 および σ_β^2 ともに有意

表1 モデル全体の統計量〈キャッシング〉

統計量	
Log-Likelihood	-1301.6
Null Log-Likelihood	-3845.58
U^2 statistics	0.66155
Mean Prob. of Correct Choice	85.36
Percentage of Correct Choice	90.68

表2 モデル全体の統計量〈ショッピング〉

統計量	
Log-Likelihood	-1318.1
Null Log-Likelihood	-4052.83
U^2 statistics	0.67478
Mean Prob. of Correct Choice	86.01
Percentage of Correct Choice	90.18

⁶ 3月は U^2 統計量が0.70、平均正答確率が87.12%、正答率が91.52%であり、6月は U^2 統計量が0.70、平均正答確率が86.96%、正答率が91.11%であった。

⁵ 文献[2]の Data Augmentation の手法を参照した。

水準95%で有意であることから、顧客によって α_i および β_i の値が異なっていることが伺える。そこで、横軸に定数項、縦軸に各利用時点における支払い金額の係数の値を取り、推定されたパラメータをプロット

表3 パラメータ推定量<キャッシング>

	推定値	標準偏差	2.5%点	97.5%点
α	0.3695	0.1566	0.0714	0.6873
β	0.0812	0.0176	0.0507	0.1197
σ_α^2	2.4737	0.1626	2.1453	2.7852
σ_β^2	0.1017	0.0231	0.0616	0.1511

表4 パラメータ推定量<ショッピング>

	推定値	標準偏差	2.5%点	97.5%点
α	-2.2452	0.1307	-2.5202	-2.0169
β	1.2501	0.1028	1.0716	1.4752
σ_α^2	1.5965	0.1252	1.3856	1.8662
σ_β^2	0.9565	0.1247	0.7581	1.2354

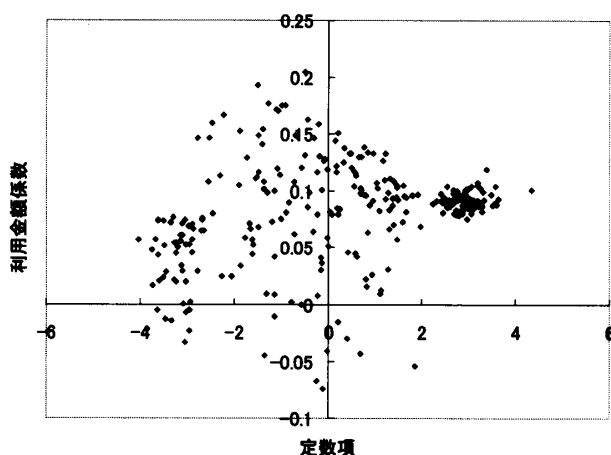


図1 推定値の散布図<キャッシング>

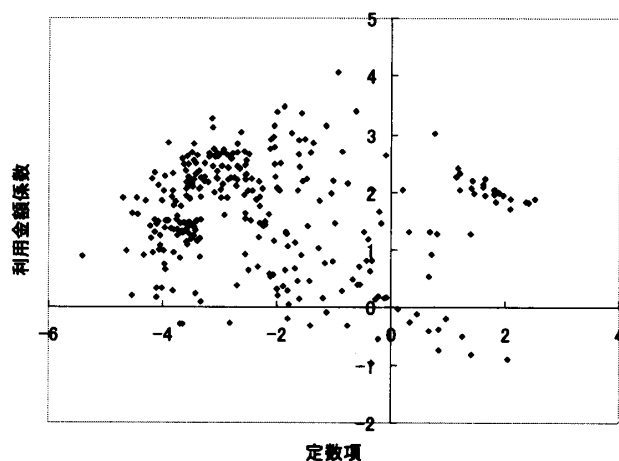


図2 推定値の散布図<ショッピング>

したものが図1および図2である。二つの図から、選択行動の特徴にはっきりした違いがあることが見て取れる。

さらに、支払い方法の選択行動と推定値 α_i および β_i の関係を数値的に明確化するために正準判別分析を行った。判別得点 z について、 $z \geq 0$ ならば支払い方法を一度でも選択したことのあるサンプルに、 $z < 0$ ならば利子付きの支払い方法を選択したことがないサンプルに判別すると仮定すると、求められた判別関数はそれぞれ、

$$z = -1.07 + 0.56\alpha_i + 10.63\beta_i \quad (11)$$

$$z = -0.12 + 0.57\alpha_i + 0.82\beta_i \quad (12)$$

となった。式(11)はキャッシング利用データの推定値、式(12)はショッピング利用データの推定値から算出された判別関数である。式(11)および式(12)から α_i および β_i の係数が共に正であることが分かるため、 α_i および β_i の値が高い顧客ほど利子付きの支払い方法を選択する顧客、すなわち利子を許容する顧客と見なせる。これは2.2節で述べた α_i および β_i の解釈とも一致する。

2.3.2 検証結果

ここでは、2.3.1節で求められた推定結果の有効性に関する検証を行った。表5がキャッシング利用データを用いたモデルの検証結果、表6がショッピング利用データの検証結果である。ショッピング利用データの正答率がおよそ85%とそれなりの予測力を持っており、本モデルが有効であることを示している一方で、キャッシング利用データの正答率は52%であり、本モデルによってキャッシング利用行動を予測することは難しいことを示している。キャッシング利用行動は、前述のように「手持ちの現金がないが、クレジット

表5 キャッシング利用データの検証

統計量	
Mean Prob. of Correct Choice	49.00
Percentage of Correct Choice	51.94

表6 ショッピング利用データの検証

統計量	
Mean Prob. of Correct Choice	82.74
Percentage of Correct Choice	85.69

7 3月の検証結果は平均正答率が79.11%、正答率が82.63%であり、6月の検証結果は平均正答率が76.23%、正答率が80.12%であった。

トカードでは支払うことができない」「急に現金が必要になった」というような緊急性が影響を与える場合が多いと考えられるため、ある時点における利用金額を用いた静的な分析ではなく、どのような頻度でどのくらいのキャッシング利用が発生しているかというような動的な分析が望ましいのではないかと考えられる。したがってこれ以降は本モデルの有効性が示されたショッピング利用データのみを焦点を当て、分析を進めていく。

ところで、本モデルによる推定および検証結果は、利用金額というたった一つの説明変数しか用いていないのにもかかわらず、上述のようにかなりの精度を持ったものである。この「利用金額が高額であればクレジットカードによる支払いがされやすく、さらに利子付きの支払いが選択されやすい」という結論は、直感的にも非常にわかりやすい結論である。そこで、本モデルを支払い方法選択行動分析に適用する妥当性および必要性を示すため、顧客を区別せずに単純な線形判別分析を行い、本モデルの推定および検証精度との比較を行った。データには2002年および2003年の12月のショッピング利用データを用い、本モデルと同様に2002年のデータで推定を、2003年のデータで検証を行った。その結果、線形判別分析による判別正答率は79%、予測率は74%であり、本モデルによる正答率および予測率は表2および表6からそれぞれ90%、86%である。線形判別関数を用いて約80%も判別出来るということから、利用金額が支払い方法の選択に及ぼす影響はかなり大きく、利用金額のみでも支払い選択行動を大部分説明できると考えられる。しかし上述の結果から、本モデルは利用金額が線形的に影響を及ぼす部分だけでなく、さらにそれ以外の部分をも説明しているのだと考えられるため、本モデルの適用は妥当でありまた必要であるといえよう。

本モデルの他の利点として、モデルのパラメータを顧客ごとに推定することが可能なため、従来であれば「全体の傾向として」しか捉えられなかった顧客のカード利用行動を顧客毎に管理することが出来、よりきめ細やかなマーケティング戦略の策定が可能な点が挙げられる。これは、マス・マーケティングからOne-to-Oneマーケティングへと発展してきた潮流[1,3,6,9]に則ったものであり、かつ顧客価値の“高価値”化を図り収益を確保し続けるためにも必要な視点であると考えられる。さらに、本モデルはベイズ的アプローチによるパラメータの推定を行っているため、

モデルの拡張や事前情報の追加が比較的容易に出来る点も強みであるといえる[4]。本研究ではこのモデルの特性を生かし、第3節において属性データを用いたモデルの拡張を行っている。

2.3.3 結果の考察

ここで再び散布図を別の視点から眺めてみると、(a)点線で囲まれた層、(b)破線で囲まれた層、(c)実線で囲まれた層という三つの特徴的な顧客層が観察される(図3)。(a)のサンプルは α_i の値がかなり小さく β_i もさほど大きくないため、常に利子付きの支払い方法を嫌う「顕在的“低価値”顧客」であり、(b)のサンプルは α_i および β_i が共に正であるため、常に利子を許容する「顕在的“高価値”顧客」であると考えられる。一方、(c)のサンプルは α_i の値は負であるが値が小さすぎることとはなく、 β_i の値が大きいことから、購入するものの金額次第で利子を許容する「潜在的“高価値”顧客」であると考えられる。

クレジットカード会社としては、利子付きの支払い方法を利用してくれる“高価値”顧客に対して、より積極的にプロモーション等を実施することが効率的であると考えられる。しかし同じ“高価値”顧客であっても、顕在的“高価値”顧客は常に利子を許容し、自発的に“高価値”で居続けてくれる顧客であるが、潜在的“高価値”顧客は利子を許容するかどうかは購入するものの金額次第で決定するため、何らかのアプローチによって潜在的“高価値”顧客からより顕在的な“高価値”顧客へ導くことが可能であると考えられる。そこで潜在的“高価値”顧客へのアプローチを時期・内容・通知方法の三点から考案した。時期であるが、購入するものの金額次第で“高価値”顧客となるため、クリスマスや誕生日等のイベント時期や引っ越し後

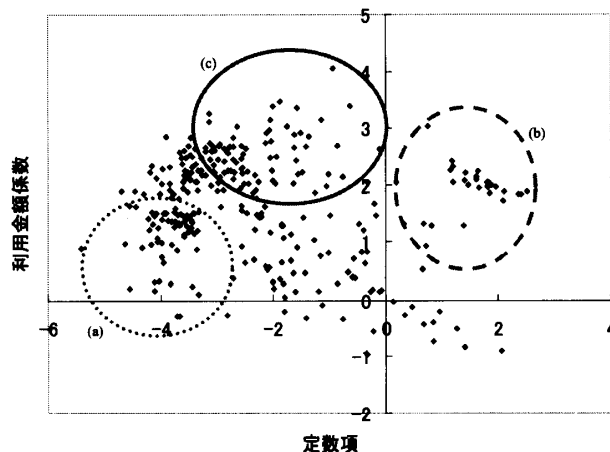


図3 三種類の特徴的な顧客層

(住所変更届受理直後)等,高額出費が予想される時期にアプローチを行う。プロモーション内容は,利子率の引き下げやポイント制度やマイレージサービスの優待といった,クレジットカードによる高額出費に対する優待を行い,顧客に対してはDM, eメール等によってプロモーションの通知を行う。

3. 新規・未利用顧客への応用

これまでは利用履歴データが充実している既存顧客を対象に分析を行ってきた。しかし,クレジットカード会社が抱える顧客には,利用履歴データが少ない新規顧客や,そもそもクレジットカードを保持しているだけで利用履歴データが存在しない未利用顧客が存在する。これらの顧客はデータの不足から, α_i および β_i といった顧客の特徴を示す数値を推定することが出来ない。この状況の下ではこれらの顧客が“高価値”かどうか分からないため,プロモーションが無駄うちとなる可能性が存在する。しかし,それらの顧客に関しても属性データは存在する。したがって,もし属性データと顧客の“価値”について何らかの関連性を見出すことが出来たなら,その関連性をもとに利用履歴データが少ないあるいは存在しない顧客でも,顧客属性から“高価値”顧客かどうかを判別することが出来る。そこで既存顧客の利用履歴データを用いて,顕在的“高価値”顧客の代表的特徴と考えられる α_i および潜在的“高価値”顧客の特徴 β_i とその顧客の属性との関連付けを試みた。

3.1 属性データによる関連付け

本分析の目的は,既存顧客の利用履歴データの分析によって得られる“高価値”顧客の特徴 α_i および β_i とその“高価値”顧客の属性データがどのように関連しているかを分析することである。そこで本分析においては,2節で用いたショッピング利用データ(2002年および2003年の各月においてショッピング利用履歴があり,かつ2002年の各月において10回以上利用履歴のある顧客300名のショッピング利用履歴データ)を再び用いることにした⁸。属性データには,性別・年齢・年収といった基本的属性データに加え,新たに限度額比率なる変数を作成し使用した。限度額比率は

$$\text{限度額比率} = \frac{\text{キャッシング利用の限度額}}{\text{総限度額}} \quad (13)$$

⁸ 論文には12月のデータの分析結果を掲載した。紙面の関係上,3月および6月の分析結果は割愛した。

のように算出した。一般的に,キャッシングを利用するあるいは利用する意思のある顧客は利子に対して寛容あるいは無意識な傾向があると考えられるため,限度額比率は利子に対する寛容度の指標になると考えられる。

属性データとの関連付けを行うにあたり,前節で用いた二項プロビットモデルのパラメータである α_i および β_i に属性データ組み込むモデルの階層化を行い,データを用いて分析した。すなわち式(14)のように,パラメータベクトル $\delta_i = (\alpha_i \beta_i)$ が属性データ γ_i の線形関数となったモデルを考える。

$$\delta_i = Q\gamma_i + \eta_i \quad \text{where } \eta_i \sim N(0, \Gamma_\delta) \quad (14)$$

ここでは γ_i は $(d+1) \times 1$, Q は $2 \times (d+1)$, η_i は 2×1 である。

属性データを組み込むにあたり,予備的に α_i および β_i に全属性データを回帰させる分析を行った。その結果, β_i に関しては全てのデータセットにおいて影響力があると考えられる属性がいくつか発見されたが, α_i に関しては,全てのデータセットにおいて影響力があると考えられる属性が存在しなかった。したがって,以下では β_i のみに属性データを組み込んだ二項プロビット分析を行う。式(14)から

$$\beta_i = \rho\gamma_i + \zeta_i \quad \text{where } \zeta_i \sim N(0, V_\beta) \quad (15)$$

とする。ここでは Q は $1 \times (d+1)$, ζ_i はスカラーとなる。

第2節で得られた考察から, β_i の値は潜在的“高価値”顧客かどうかの指標であるといえる。したがって, β_i に関連付けられる属性を発見することによって,潜在的に“高価値”である可能性の高い顧客を属性から判別出来ることになる。

3.2 分析結果

表7は本節で用いたモデルの推定によって得られたモデル全体の統計量であり,表8は推定されたモデルのパラメータ推定量である。今回のモデル推定の結果,有意水準95%で有意だった属性データは限度額比率のみであることが分かった。したがって,表8の C_p

表7 モデル全体の統計量<推定>

統計量	
Log-Likelihood	-1327.3
Null Log-Likelihood	-4052.83
U^2 statistics	0.67249
Mean Prob. of Correct Choice	85.87
Percentage of Correct Choice	90.25

表8 パラメータ推定量

	推定値	標準偏差	2.5%点	97.5%点
C_ρ	-2.1368	0.1240	-2.3864	-1.9136
ρ_L	1.8955	0.1686	1.5709	2.2358
$\sigma_{C_\rho}^2$	1.5270	0.1116	1.3152	1.7484
$\sigma_{\rho_L}^2$	0.9154	0.1101	0.6888	1.1232

表9 モデル全体の統計量〈検証〉

統計量	
Mean Prob. of Correct Choice	83.07
Percentage of Correct Choice	85.96

は定数項、 ρ_L は限度額比率の係数を表している。 ρ_L の値が正であることから、「限度額比率の高い顧客は β_i の値が高い傾向にあり、ゆえに利用金額が高い場合には利子付きの支払いを選択する傾向にある顧客である」ということが出来る。

表9は推定されたパラメータを用いて行った検証結果である。モデルの正答確率は約83%、正答率は約86%であることから、このモデルおよびパラメータの解釈は有効であると考えられる。

したがって、新規顧客あるいは未利用顧客に対して2.3.3節における潜在的“高価値”顧客に対するプロモーションを行う際には、限度額比率が高い顧客を中心にアプローチすることが効率的であるといえる。

4. おわりに

本研究は、「カード利用時に利子付きの支払い方法を選択するまたは選択する可能性のある顧客＝“高価値”顧客」と定義し、(1)既存顧客に関して、利用履歴データを用いた支払い方法選択行動分析と“高価値”顧客の類型化、(2)新規顧客および未利用顧客への応用を踏まえ、“高価値”顧客の特性と顧客の属性データの関連付け、の二分析を行った。特に(2)の分析で得られた限度額比率と潜在的“高価値”顧客との関連性は、クレジットカード会社の限度額設定ルールがはからずとも潜在的“高価値”顧客の判別に活用できる可能性があることを示唆している。

今後の課題として、以下の点が挙げられる。(1)の支払い方法選択行動分析において、本研究では顧客の特徴を表す α_i および β_i に確率変数モデルを仮定したが、 α と β の共分散 $\sigma_{\alpha\beta}$ を0と設定する非常にシンプルなモデルとなっている。今後解析を進める際には $\sigma_{\alpha\beta}$ をモデルに組み込んで推定することが望ましいと考えら

れる。また、クレジットカードを利用する際に、どの支払い方法を選択するかという意思決定には、どのような商品を購入したか、どのような店舗(環境)で利用したかといった要因も影響を与えると考えられる。したがって、支払い方法の選択行動モデルを精緻化する一つの方法として、利用金額に関する情報だけでなく何をどこで購入したのかといった情報も反映出来るモデルへの拡張が考えられる。

最後に、本研究では“高価値”顧客の特徴の分析および顧客の属性と“高価値”顧客の特徴の関連性を調査する分析を行ってきたが、この分析には大きな限界が存在する。本研究で前提としている“高価値”顧客は、あくまで収益のみの“高価値”顧客であり、顧客の破綻リスクが全く考慮されていない。したがって、顧客の破綻リスクを考慮出来る分析の併用による、真に高価値な顧客の探索が望ましい。顧客の破綻リスクを別途推定し、モデルに組み込むあるいはモデルを併用して分析を進めることも、クレジットカード会社にとって真に高価値な顧客を判別し、その顧客価値を最大化する上で必要であろう。

謝辞 本研究を進めるにあたり、ご指導頂きました阿部誠先生(東京大学)に深く感謝申し上げます。また、貴重なデータをご提供頂いたデータ解析コンペティション関係者の皆様に併せて感謝申し上げます。

参考文献

- [1] G. M. Allenby and P. E. Rossi, “Marketing Models of Consumer Heterogeneity,” *Journal of Econometrics*, Vol. 89, No. 1/2, pp. 57-78 (1999).
- [2] J. H. Albert and S. Chib, “Bayesian Analysis of Binary and Polychotomous Response Data,” *Journal of the American Statistical Association*, Vol. 88, Issue 422, pp. 669-679 (1993).
- [3] P. E. Rossi and G. M. Allenby, “A Bayesian Approach to Estimating Household Parameters,” *Journal of Marketing Research*, Vol. 30, No. 2, pp. 171-182 (1993).
- [4] P. E. Rossi, R. E. McCulloch and G. M. Allenby, “The Value of Purchase History Data in Target Marketing,” *Marketing Science*, Vol. 15, No. 4, pp. 321-340 (1996).
- [5] P. E. Rossi and G. M. Allenby, “Bayesian Statistics and Marketing,” *Marketing Science*, Vol. 22, No. 3, pp. 304-328 (2003).

- [6] P. Guadagni and J. D. C. Little, "A Logit Model of Brand Choice Calibrated on Scanner Data", *Marketing Science*, Vol. 2, No. 3, pp. 203-238 (1983).
- [7] P. H. Franses and R. Paap, *Quantitative Models in Marketing Research*, Cambridge University Press, Cambridge (2001).
- [8] R. L. Andrews, A. Ainslie and I. S. Currim, "An Empirical Comparison of Logit Choice Models with Discrete Versus Continuous Representations of Heterogeneity", *Journal of Marketing Research*, Vol. 39, No. 4, pp. 497-487 (2002).
- [9] W. A. Kamakura and G. J. Russell, "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, Vol. 26, No. 4, pp. 379-390 (1989).
- [10] W. H. Greene, *Econometric Analysis, 4th edition*, Prentice Hall, New Jersey (2000).