

クレジット・カード利用顧客の デフォルト予兆発見分析

矢野 順子, 佐治 美歩, 中川慶一郎, 高橋 彰子, 山中 啓之, 生田目 崇

1. はじめに

近年、クレジット・カード市場は、様々な提携カードの発行などによる競争激化を背景に、カードの新規発行数が増加し続けている。実際に、1995年からの10年間で、日本国内のクレジット・カード枚数は20%増加し、また信用供与総額は80%という急速な伸びを示した。カード発行会社は、多くの顧客が高額利用することにより、より多くの収入を得ることになるが、その一方で、無計画なクレジット・カード利用などに起因する支払遅延や自己破産（以下、デフォルトとよぶ）の件数も増加の一途を辿っているという現状がある。この一因としては、クレジット・カードを利用することに対する心理的な障壁が低くなったことにより、利用金額・頻度が増えたことが挙げられる。例えば、通信販売やインターネットを通じたオンライン・ショッピングではクレジット・カード決済の選択が購入者の信用力を示す一つの手段となっており、クレジット・カードが幅広く利用されている。また、従来は現金や口座振替などであった公共料金の支払においても、クレジット・カード決済が可能となってきた。さらに、支払方法についても請求後にリボ払いへの切替えができるなど、以前より柔軟に支払方法を選択できるようになった。こういったことから、クレジット・カード決済を利用する顧客が増加している[5]。

クレジット・カードの利用形態は、大きくショッピングとキャッシングに分けられる。クレジット・カード発行会社は、加盟店からの利用手数料ならびにキャ

やの じゅんこ, さじ みほ, なかがわ けいいちろう,
たかはし しょうこ, やまなか ひろゆき

(株)NTT データ技術開発本部

〒104-0033 中央区新川 1-21-2

なまため たかし

専修大学 商学部

〒214-8580 川崎市多摩区東三田 2-1-1

受付 05.7.25 採択 05.11.14

ッシングによる金利を主な収入源としているが、その一方で、加盟店への代金支払いやキャッシングの元金についての責任を持つため、デフォルトによる損失（以下、不良債権とよぶ）を引き受けることになる。デフォルトの件数増加に伴い、クレジット・カード会社が抱えるデフォルト・リスクに対する負担は、さらに大きくなっていくと考えられる[1,3]。

デフォルトは、顧客が自身の支払能力を超えてカードを利用することで起こる。デフォルトを未然に防ぐためにクレジット・カード発行会社では、従来より自社の判断基準に加え、信用機関からの情報など利用し、カード申し込みに対する与信（利用限度額の設定）を行っている。しかし、申し込み時のデータからは、カード利用中に顧客がデフォルトをおこすかどうかを動的に判断することはできない。もし、顧客のクレジットカード利用履歴情報から、将来デフォルトをする危険性が高いであろう顧客を早期に発見し、その顧客に対して適切なアプローチを行うことができれば、顧客のデフォルトを未然に防止し、デフォルト・リスクに対する負担を軽減することができる。また同時に、それぞれの顧客にあった支払方法や契約プランを提示するといった、きめ細やかなサービスを提供するための情報を得ることができ、最終的には顧客の満足度向上にも繋がると考えられる。

そこで本論文では、不良債権を未然に抑制することを目的として、将来デフォルトする可能性が高いと思われる顧客（以下、デフォルト予備群とよぶ）を早期に発見するためのフレームワークを提案する。本提案では、デフォルト・リスクの算出モデルを構築し、顧客別に検証を行う。さらに、デフォルト予備群のうち、実際にデフォルトした顧客（以下、デフォルト顧客とよぶ）について、カード利用の特徴を抽出し、考察する。

2. データの概要

、本論文で使用したデータは、平成16年度データ解

析コンペティションで提供されたデータである。分析対象とするデータは、2002年1月～2003年12月の2年間に一度も利用がない顧客102人を除く、54,862人の顧客別属性データおよびクレジット・カードのトランザクション・データである。

データ項目の詳細を以下に示す。

顧客属性データ

- 顧客属性
 - 加入日, 生年, 年齢, 性別, 住所区分, カード区分, 支払コース, 配偶者, 子人数, 居住区分, 勤務区分, 年収, 職種
- 限度額
 - 総限度額, キャッシング限度額, ショッピング限度額

トランザクション・データ

- 月別トランザクション・データ
 - 月別利用金額・回数, 月別総利用限度額, 月別利用残高, 支払結果
- 日別トランザクション・データ
 - 利用日, 利用金額, 支払方法区分, 利用区分, 利用店舗名, 端末番号

ただし、利用店舗については、一部の端末番号のみしか把握できず、またトランザクション処理日が利用日となっているため、すべてのトランザクション・データについて実際の利用日を正確に捕捉することは事実上不可能である。また、キャッシングや分割払いにおける金利、利用停止、デフォルトといった内部管理情報は提供されていない。

3. 分析フロー

本論文では、不良債権発生を未然に抑制することを目的として、「デフォルト予備群」を早期に発見するためのフレームワークを図1の分析フローのように構築する。

Step 1では、提供データだけでは把握できない月別トランザクション・データのショッピングおよびキャッシングの利用履歴情報から、「デフォルト顧客」を定義する。Step 2では、Step 1で定義したデフォルトに基づき、ロジット・モデルを用いて顧客別デフォルト・リスクを算出するモデルを示す。Step 3では、構築したモデルの妥当性を検証する。なお、検証にあたっては、検証用データにより顧客別のデフォルト・リスクをモニタリングし、結果的にデフォルトした「デフォルト予備群」をどの程度抽出できたかを評

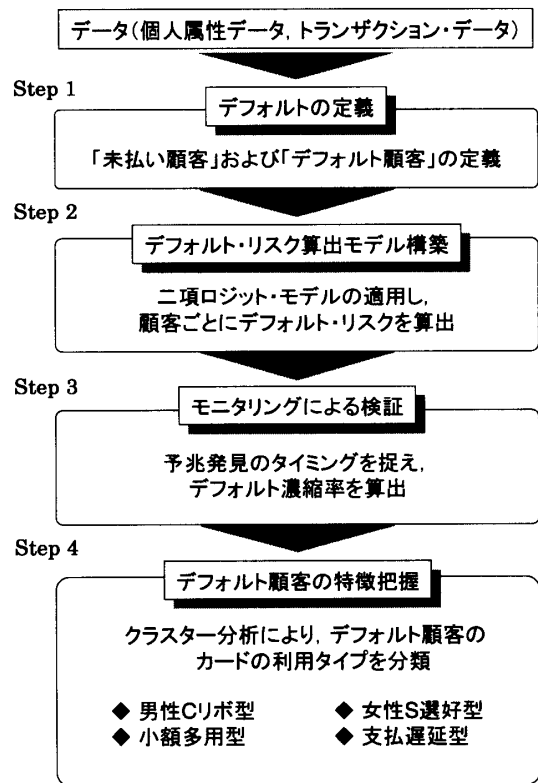


図1 分析フロー

価する。さらに、デフォルト顧客のデフォルトの予兆をどれだけ事前に捉えることができたのか、「予兆発見のタイミング」について分析する。最後に Step 4では、Step 3の検証において、予兆発見可能であったデフォルト顧客の属性情報および利用履歴情報から、クラスター分析を通して、典型的なデフォルト顧客の利用の特徴を把握する。以下で各 Step を詳述する。

4. デフォルトの定義

Step 1では、クレジットカードのトランザクション・データである支払遅延情報、ショッピングおよびキャッシングの月別利用金額、残高から、「未払い顧客」を定義し、さらに支払不能顧客である「デフォルト顧客」を定義する。

4.1 未払い顧客の定義

支払が遅っている状態とは、利用をしていないにも関わらず、支払をしないために毎月の残高が増加している状態と考えられる。このような状態を示す条件を以下のように設定し、本論文ではこれらすべてを満たす顧客を「未払い顧客」と定義することにする。

- 前月に比べて当月の残高が増加している。(残高の増加は金利分と考える)
- 前月および当月の利用がない。(利用停止になっ

ている可能性がある)

- ・ 前月の支払が遅延している。(結果的に支払を行っている顧客も含まれる)

したがって、顧客 i の j 月目における支払が「未払い」であるかどうかの判定 f_{ij} は、

$$f_{ij} = \begin{cases} 1, & \text{if } b_{ij} - b_{i,j-1} \geq 0 \\ & u_{i,j-1} = 0 \\ & p_{i,j-1} = 1 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

と記述できる。なお、 p_{ij} は、支払結果のダミー変数で1のとき「未払い」、0のとき「支払済み」と定義する。また、顧客 i の j 月目の利用金額を u_{ij} 、支払残高を b_{ij} とする。

なお、利用金額および残高は、対象月のキャッシングおよびショッピングの合計金額とする。

4.2 デフォルト顧客の定義

本節では、前節で定義した未払いに基づき、次の条件を満たす顧客を「デフォルト顧客」と定義する。

- ・ 未払い状態が3ヶ月連続観測される。

ここで、未払いが3ヶ月続いた場合の3ヶ月目の月をデフォルト月とする。このように定義したのは、店舗や利用カード区分によって請求月が異なり、カード利用から請求まで最大2ヶ月のずれがあると考えられ、その結果、利用停止になるまでに最終利用日より最大3ヶ月かかる可能性があるという判断による。実際、一般的な銀行発行クレジット・カードにおける顧客評価モデルでは、取引実績に90日超の延滞や破産などの信用情報記録がある場合に、「デフォルト・リスクの高い顧客」と定義されている[2]。

したがって、顧客 i が j 月目にデフォルトを引き起こすかどうかの判定 d_{ij} は、

$$d_{ij} = \begin{cases} 1, & \text{if } f_{i,j-2} + f_{i,j-1} + f_{ij} = 3 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

となる。

4.3 デフォルト顧客の分布

本節では、前節で定義したデフォルトについて、まず、顧客別属性データとの関係を明らかにする。表1、2にデフォルトとの関係が強かった属性に関するクロス集計結果を示す。表1を見ると、若年層ほどデフォルトする確率が高いことが分かる。また、表2より、低限度額の顧客ほどデフォルトを起こす確率が高いことが分かる。さらに、 χ^2 検定を行った結果、 χ^2 値はそれぞれ、82.7、454.8となり、有意水準5%で年代および総限度額の違いについてデフォルトに差がある

表1 年代とデフォルトの関係

	20代	30代	40代
デフォルト (人)	107	150	115
非デフォルト (人)	5564	13993	12505
合計 (人)	5671	14143	12620
デフォルト率 (%)	1.89	1.06	0.91
	50代	60代	70代
デフォルト (人)	81	43	17
非デフォルト (人)	11620	7387	3382
合計 (人)	11701	7430	3399
デフォルト率 (%)	0.69	0.58	0.50

表2 総限度額とデフォルトの関係

	総10万	総20万	総30万
デフォルト (人)	95	88	122
非デフォルト (人)	2654	3176	6539
合計 (人)	2749	3264	6661
デフォルト率 (%)	3.46	2.70	1.83
	総50万	総80万	総100万～
デフォルト (人)	170	37	1
非デフォルト (人)	31506	7161	3415
合計 (人)	31676	7198	3416
デフォルト率 (%)	0.54	0.51	0.03

と判定される。表中で、「総10万」は、総限度額が10万円以下を意味する。このような属性情報は、ほぼ固定的な要素であるため、初期与信等には有効であると考えられるが、時間とともに変化する顧客ごとのデフォルト・リスクを把握することはできない。したがって、この結果からデフォルトを起こす可能性の高い顧客を事前に抽出することは困難である。

そこで、次節では顧客のカード利用履歴を考慮したデフォルト・リスク算出モデルを構築する。

5. デフォルト・リスク算出モデル

Step 2では、顧客別にデフォルト・リスクを算出するモデルを構築する。5.1小節では、モデル構築に用いる変数および対象期間について示し、5.2小節では、モデル構築結果について述べる。

5.1 モデルの構築

前節で定義したデフォルト情報に基づき、顧客別デフォルト・リスクを算出するために、二項ロジット・モデル[4]を適用する。二項ロジット・モデルは、式(3)に示すように、デフォルト状態を $S=1$ とした場合に、顧客 i のデフォルト確率 p_i は、デフォルトに関する m 次元の共変量ベクトル x_i とパラメータ $\beta_0, \beta_1, \dots, \beta_k$ に対して、

$$p_i = \Pr\{S=1|x_i\}$$

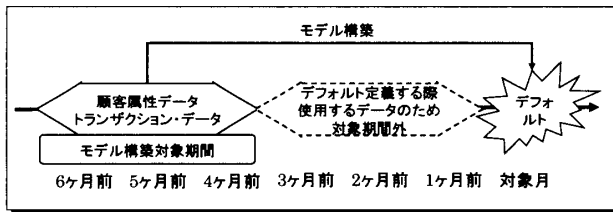


図2 説明変数の対象期間

$$= \frac{1}{1 + \exp\{-(\beta_0 + \sum_{k=1}^m \beta_k x_{ik})\}} \quad (3)$$

となる。

モデル構築に用いた学習データは、提供データのうち、2003年1月～2003年12月の1年間のデータである。分析では、顧客属性データの他に3ヶ月間のトランザクション・データを加工した変数である「キャッシング利用率」や「ショッピングリボ残高増加率」などを含めた23変数を説明変数の候補として、ステップワイズ減少法により変数選択を行った。

ただし、支払遅延を1回も起こしていない顧客は、全体の約85% (45,924人) を占めており、本論文では、支払遅延のある顧客のうち、デフォルトに近づくと思われるデフォルト予備群を抽出したいという目的から、モデル構築にあたっては、対象期間中に支払の遅延が1回以上ある顧客を対象とした。

デフォルト顧客の説明変数の対象期間を図2に示す。なお、非デフォルト顧客の対象期間は、ランダムに選択した。

図2に示すとおり、対象月においてデフォルトと判定された場合、モデル構築に用いる説明変数の対象期間は、1ヶ月前から3ヶ月前までの情報を使用している。これは、デフォルトの判定条件として使用した情報を排除する必要があるためである。

5.2 選択された変数

学習データを用いてモデルを構築した結果を表3に示す。表3のt値の結果より、デフォルトに最も関与する要因として、「ショッピング回数」、「ショッピングリボ残高」、「キャッシング回数」、「キャッシングリボ残高」、「キャッシング利用率」が挙げられる。「ショッピングリボ残高」、「キャッシングリボ残高」のように「リボ払い」がデフォルトに高く寄与している理由には、リボ払いの支払方法を利用することにより、毎月の支払額を一定させることができるため、顧客は、本来の支払能力を超えた高額商品の購入や借入れが手軽にできてしまい、結果として支払できなくなってしまいうケースが多いと思われる。なお、表3中

表3 二項ロジット・モデルの係数の推定値

属性カテゴリ	選択された説明変数	係数	t値
総限度額	20万	6.14×10^{-1}	2.77***
	30万	7.93×10^{-1}	3.99***
キャッシング限度額	0万	6.60×10^0	2.08**
	5万	5.91×10^0	1.88*
	10万	5.09×10^0	1.62
	20万	5.00×10^0	1.59
	30万	4.97×10^0	1.59
	50万	5.16×10^0	1.65*
	60万	4.39×10^0	1.39
年齢	20歳代	-3.58×10^{-1}	-2.06**
	30歳代	-4.30×10^{-1}	-2.86***
性別	女性	2.55×10^{-1}	1.94*
居住区分	賃貸公営住宅	3.29×10^{-1}	1.71*
	賃貸マンション	2.84×10^{-1}	1.74*
職業	会社員	2.71×10^{-1}	1.96**
年収	400万～500万	3.26×10^{-1}	1.95*
残高	SR	1.02×10^{-6}	4.20***
	CR	1.40×10^{-6}	4.02***
支払区分	一括回数	-1.10×10^{-1}	-8.20***
	リボ回数	-7.98×10^{-2}	-4.29***
支払回数	ショッピング回数	8.65×10^{-2}	7.66***
	キャッシング回数	2.64×10^{-6}	4.19***
利用率	キャッシング利用率	8.85×10^{-1}	3.09***
	SR残高増加率	2.21×10^{-1}	2.01**
支払遅延	回数	6.45×10^{-1}	10.45***
定数項	定数	-1.05×10^1	-3.35***

***, **, *はt検定でそれぞれ10%, 5%, 1%有意になった値を示す。

の「SR」は「ショッピングリボ」を、「CR」は「キャッシングリボ」を意味する。

6. モニタリングによる検証

Step 3では、前節で構築したモデルを検証用データに適用し、顧客別にデフォルト・リスクをモニタリングすることにより、その妥当性を検証する。なお、検証は次に述べる2つのパターンについて行う。

検証パターン1では、モデル構築顧客を除く全顧客50,598人を対象として、検証対象月の4ヶ月前、5ヶ月前、6ヶ月前の3ヶ月間のデータを用いて、前節で構築したモデルに適用して検証を行う。また検証パターン2では、モデル構築の条件に揃え、分析対象期間中に支払を1回以上延滞している顧客のうち、モデル構築顧客を除く4,674人を対象とする。なお、適用期間は検証パターン1と同様である。モニタリングによる検証に用いたデータは、2002年1月～2003年8月の20ヶ月間のデータである。

6.1 濃縮率

本論文で提案するモデルを実際のビジネスへ適用することを考えた場合、デフォルト未然防止のためにすべてのクレジット・カード契約者の利用履歴を常に監視することは現実的ではなく、一定のレベル以上にデフォルトを起こす可能性が高い顧客と判断されるグル

ープに対してアプローチをすることが効率的である。そこで、構築したモデルの妥当性を評価するため、デフォルトを起こす可能性の高い顧客の抽出度合いについて「濃縮率」を定義し、定義された濃縮率をもとにモデルの検証を行う。濃縮率 c は、モデルを適用しない場合のデフォルト確率に対するモデルを適用した場合のデフォルト確率の比と定義する。モデルを適用しない場合と適用した場合の顧客数をそれぞれ n_1 , n_2 デフォルト顧客数 m_1 , m_2 とする。デフォルト確率は、それぞれ $\frac{m_1}{n_1}$, $\frac{m_2}{n_2}$ となるので、濃縮率は、

$$c = \frac{m_1/n_1}{m_2/n_2} = \frac{m_2 n_1}{m_1 n_2} \quad (4)$$

で計算される。

検証結果を表4, 5に示す。また、濃縮率を表6にまとめる。表6より、全体のデフォルト率は1%を切っているが、構築したモデルを適用することにより、デフォルト率が10.4%となるようなリスクの高い顧客グループを抽出することができており、本分析モデルが有効であることを示している。同様に、支払遅延を1回以上行っている顧客に限定した検証パターン2においても、本分析で示したデフォルト予兆を発見する方法を導入することにより、濃縮率が3.27倍とより高いリスクの集団を抽出することができていることがわかり、モデルの有効性を検証できた。

表4 検証パターン1の適用結果

		モデルから		合計 (人)
		デフォルト	非デフォルト	
検証	デフォルト	106	65	171
	非デフォルト	915	49512	50427
合計		1021	49577	50598

表5 検証パターン2の適用結果

		モデルから		合計 (人)
		デフォルト	非デフォルト	
検証	デフォルト	106	65	171
	非デフォルト	781	3722	4503
合計		887	3787	4674

表6 濃縮率の比較

		デフォルト確率 (%)		濃縮率 (倍)
		モデル非適用	モデル適用	
検証	パターン1	0.34	10.38	30.72
	パターン2	3.65	11.95	3.27

6.2 予兆発見のタイミング

前節で述べた予兆発見されたデフォルト状態が、実際にデフォルトを起こす何ヶ月前に判定されていたか、そのタイミングをモニタリングすることにより検証する。なお、検証の対象となる顧客は、検証パターン1, 検証パターン2の両者において、デフォルトと判定された171人である。図3に示した結果から、デフォルトを起こす月の半年以上前の時期に、将来デフォルトを起こす顧客の約60%を抽出できていることが分かる。これは、モニタリングによる効果を示しており、モデルにより抽出された顧客に対して、事前にデフォルト防止のためのアプローチが可能になると考えられる。

7. デフォルト顧客の特徴把握

Step 4では、前節において、予兆発見可能であった顧客に関して、デフォルト・リスクのみならず、属性や利用の特徴も考慮し、デフォルトを引き起こす可能性が高い顧客像を考察することを目的とする。ここでは、利用の特徴を分類して把握するために、予兆発見可能であった顧客106人に対して、K-means法によるクラスター分析を行う。なお、分析では、予兆発見月の前3ヶ月間のデータを基に作成された、表3に示す変数を説明変数とし、クラスター数は4とした。

各クラスターの代表的な変数別平均値を表7に示す。なお、表中の「C 限度額~10万」は、キャッシング限度額が10万円以下を意味する。この結果から、クラスターの特徴を表8にまとめる。クラスター1およびクラスター2に含まれる顧客は、キャッシングのリボ払いの残高が非常に多い。また、クラスター4の顧客は、キャッシング、ショッピングともにリボ払いはそれほど行っていないことが分かる。図4をみると、クラスター

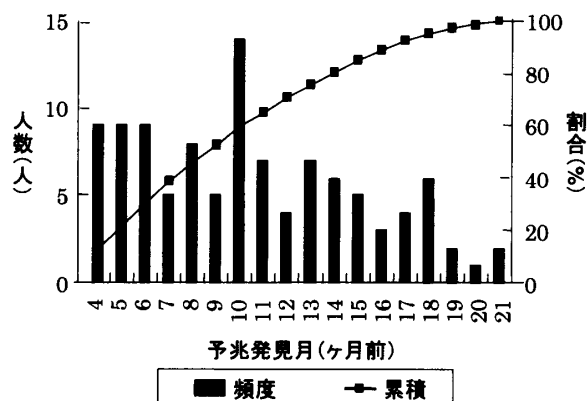


図3 デフォルト顧客の予兆発見タイミング

1は圧倒的にリボ払い回数が多く、クラスタ3はリボ払い回数、一括払い回数ともに多く、さらに他のクラスタと比較すると支払遅延回数も多いことが分かる。クラスタ3に属する顧客は、支払遅延回数が多いため優良顧客とはいえないが、メイン・カードとして利用している可能性が高いと思われる。また、図5より、クラスタ1は圧倒的に男性の割合が高く、クラスタ2は女性の割合が高いことが分かる。これらの結果より、クラスタ1は、男性でキャッシング・リボ払いが多い

ので「男性キャッシングリボ選好型」、クラスタ2は、女性の割合が多く、ショッピング・キャッシング共に多いので「女性ショッピング選好型」、クラスタ3は、支払残高が少ないが利用回数が多いので「小額多用型」、クラスタ4は、「支払遅延型」と名づけた。このように、利用形態の似た顧客グループに分類し、典型的なデフォルト・タイプを定義することにより、例えば、クラスタ1の「キャッシングリボ選好型」の顧客に対しては、リボ払いの最小単位の変更サービスや、「支払遅延型」の顧客に対しては、貸出利率の引き上げや早期に未払い残高の催促を行うなど、事前に特定したデフォルト・リスクの高い顧客に対して、デフォルトを防止するために、より適した商品やサービスを提供できると考えている。

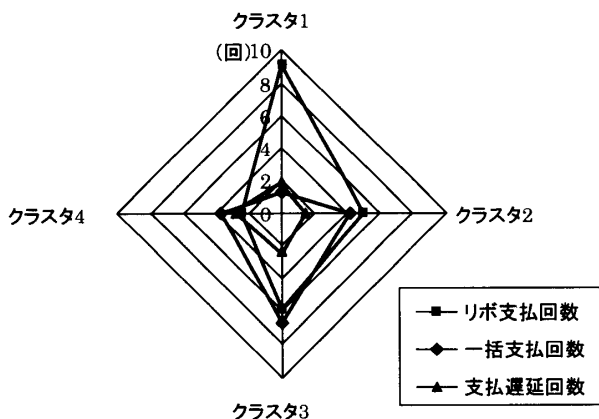


図4 クラスタ別支払区分別回数および支払遅延回数

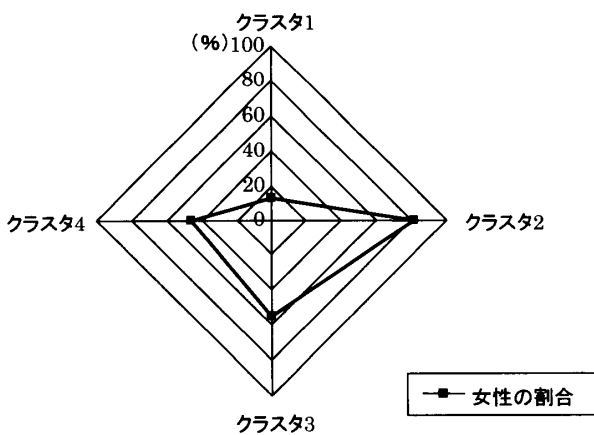


図5 クラスタ別男女比率

8. おわりに

本論文では、顧客のデフォルトの抑制を目的とし、クレジット・カードの利用履歴情報から、顧客別デフォルト・リスクの算出モデルを構築し、デフォルトする可能性の高い顧客の予兆発見を行うモデルを提案した。また、モニタリングによるモデル検証を行った結

表7 各クラスタの変数の平均値と構成比（残高と回数以外は平均値）

	クラスタ1	クラスタ2	クラスタ3	クラスタ4
C限度額~10万(%)	0.00	7.41	0.00	100.00
C限度額10万~30万(%)	4.17	14.81	96.55	0.00
C限度額30万~(%)	95.83	77.78	3.45	0.00
女性(%)	12.50	81.48	55.17	46.15
男性(%)	87.50	18.52	44.83	53.85
CR残高(千円)	393.90	250.69	91.67	42.73
SR残高(千円)	62.58	88.69	82.96	43.41
SR残高増加率(%)	9.03	16.68	17.33	45.50
C利用率(%)	91.36	79.53	83.47	84.74
支払遅延回数(回)	1.75	1.44	2.55	2.81
一括支払回数(回)	1.25	4.11	6.66	3.65
リボ払い回数(回)	9.13	4.96	5.90	2.46

表8 クラスタごとの特徴

クラスタ1: 男性キャッシングリボ選好型	キャッシング限度額が非常に高い キャッシングリボ払い残高、回数ともに非常に多い
クラスタ2: 女性ショッピング選好型	キャッシング限度額ぎりぎりまで利用している ショッピングリボ払い残高が多い 女性の比率が高い 支払遅延回数が多い
クラスタ3: 小額多用型	利用回数が多い 一括払い回数 リボ払い残高はそれほど多くない
クラスタ4: 支払遅延型	支払遅延回数が多い ショッピングリボ残高、キャッシングリボ残高が多い キャッシング限度額が低い

果、デフォルト前の利用状況から、デフォルトの可能性が高い顧客をデフォルト予備群として、検出することができた。さらに、典型的なデフォルト顧客の利用の特徴をタイプ別に分析した。これらの分析のフレームワークを利用することで、デフォルトを起こす前に、その可能性の高い顧客を見つけ出し、モニタリングすることにより、デフォルトを未然に防止するためのより適したサービスを提供できると考えられる。

また、今後の発展としては、カードの利用を促進するために、デフォルト・リスクの算出モデルに加え、クレジット・カードの利用の変化から個々の顧客のライフスタイルの変化を捉え、その変化に応じた支払方法や契約プランを提案するシステム構築を行うことなどが挙げられる。

謝辞：本稿をまとめるにあたり、日本オペレーション

ズリサーチ学会マーケティング・データ解析研究部会の皆様から、多くの有益なコメントをいただきました。ここに感謝の意を表します。

参考文献

- [1] 青島幹郎：「クレジットカードの実務と業務知識」，シーメディア（2002）。
- [2] E. Mays and C. Hudson: *Credit Risk Modeling*, (スコアリング研究会訳：「クレジットスコアリング」)，シグマベイスキャピタル（2001）。
- [3] 伊藤隆敏，川本卓司，谷口文一：“クレジットカードと電子マネー，”日本銀行金融研究所ディスカッション・ペーパー・シリーズ，No. 99-j-16（1999）。
- [4] 木島正明，小守林克哉：「信用リスク評価の数理モデル」，朝倉書店（1999）。
- [5] 谷村賢治：「生活リスクと環境知」，朝倉書店（2004）。