

# 多機能クレジットカード利用履歴データに対する潜在クラスモデル分析

情報数理応用研究

5218C019-1 世古裕都  
指導教員 後藤正幸

## An Analysis based on Latent Class Models to Activate Non-royal Users of Multifunctional Credit Card

SEKO Yuto

### 1. はじめに

近年、多くの国でクレジットカードシステムが普及し、消費者に広く浸透・利用されるようになった。最近では鉄道会社や EC 事業会社などが、VISA などの国際ブランドと提携して発行する多機能クレジットカードが広く利用されている。多機能クレジットカードは、通常のクレジットカードとしての機能に加え、ポイントカードとしての機能も有するものを指す。ポイントはクレジットの決済額に応じて、またはグループ加盟店での現金決済時に提示することで得られる。このポイントはグループ加盟店で支払金額の割引機能として利用することができる。このことから、ユーザの多機能クレジットカード利用行動は、クレジット機能とポイント機能の利用行動の組み合わせにより多様化の傾向を有している。

このような状況下、多機能クレジットカード発行会社では、多機能クレジットカードの運営を通じた自社へのロイヤルティ向上とグループ加盟店への利益貢献の双方が望まれている。そのため、クレジット機能の利用が多く、かつグループ加盟店でのポイント利用頻度も高いユーザを、発行会社では優良ユーザと捉え、クレジット機能とポイント機能のそれぞれの利用を促すような施策が求められている。

Shimizu et al.[1] は上記のような問題に対し、多機能クレジットカードのユーザ数とユーザ満足度の向上を目的に、潜在クラスモデル [2] を用いたユーザクラスタリング手法を提案している。具体的には、ユーザごとの多機能クレジットカードの利用頻度を表現する潜在クラスモデルと、1 ユーザがどの業種のどの店舗で利用したかを表現する潜在クラスモデルを構築し、これら 2 つの潜在クラスモデルを駆使したクロス分析の結果を示している。そして、利用店舗は比較的類似しているが利用頻度が異なる 2 つのユーザ群に対して、高頻度利用ユーザ群の利用店舗を低頻度利用ユーザ群に勧めることで高頻度利用ユーザ群への移行を促せると主張している。一方で、どのクラスの利用店舗とも類似していない低頻度利用ユーザ群や、潜在クラスで表現されていない利用頻度と利用店舗を持つユーザに対しては利用増加策を検討するための情報が得られないなどの問題がある。

そこで本研究では、クレジットとポイントの双方の機能を利用可能な業種が多岐にわたることに着目し、ユーザが利用した業種に基づいて、クレジット利用額とポイント利用率の変化に寄与する業種を分析可能な潜在クラスモデルを提案する。また、提案モデルを小田急電鉄株式会社発行の多機能クレジットカードである小田急ポイントカード (以下、OP カー

ド) 利用履歴データに適用し、実データ分析の結果から、分析対象ユーザのカード利用を促す施策を検証する。

### 2. 準備

#### 2.1. 問題設定

本研究で対象とする事例は、小田急電鉄株式会社から提供された OP カード利用履歴データである。クレジット機能は小田急グループ加盟店の内外の様々な業種で利用できる一方、蓄積されたポイントは、百貨店やスーパーマーケットなど様々な小田急グループ加盟店でのみ利用可能である。

#### 2.2. 従来研究

従来のクレジットカードにおける分析では、不正利用検知に関する研究 [3] や、利用履歴データから顧客の特徴を抽出する研究 [4]、利用履歴データとパネル調査のデータを用いた顧客クラスタリングの研究 [5] などが主に行われている。

また Shimizu et al.[1] は、多機能クレジットカードのユーザ数とユーザ満足度の向上を目的に、カードの所持目的と、購買傾向の多様性を表現する 2 つの潜在クラスモデルを提案している。前者は、クレジット機能とポイント機能のそれぞれについて、各店舗での利用金額と利用回数などを共起させ、ユーザごとの各店舗での利用傾向から、カード所持目的を表現するモデルを提案している。後者は、ユーザが利用可能な全店舗に対して、各ユーザがその店舗を利用したか否かを共起させ、ユーザごとの店舗利用傾向から購買傾向の多様性を表現するモデルを提案している。そしてこれら 2 つのモデルの結果をクロス分析し、利用店舗の傾向は比較的類似しているが利用頻度が異なる 2 つのユーザ群に対して、高頻度利用ユーザ群の利用店舗を低頻度利用ユーザ群に勧めることで高頻度利用ユーザ群への移行を促せると主張している。

一方で、利用店舗の傾向がどの潜在クラスとも類似していない低頻度利用ユーザ群や、潜在クラスで表現されていない利用頻度と利用店舗を持つユーザ群に対して施策の検討ができないなど、施策検討の対象となるユーザ群が限定的であるという問題がある。そのため、得られた結果から考えられる施策も限定的とならざるを得ない。

### 3. 提案モデル

#### 3.1. 概要

上述の通り、多機能クレジットカード発行会社は、自社の利益向上に繋がるクレジット機能と、グループ加盟店への利益貢献に繋がるポイント機能の双方の利用促進が求められる。また、より多くのユーザに対して施策の検討が可能となるような分析モデルが求められている。そこで本研究では、これら

2つの観点から、各機能の分析を業種の軸で分析可能とする、クレジット分析モデルとポイント分析モデルの2つのモデルを提案する。まず、クレジット分析モデルでは、クレジット決済額の増加がマージン増加に繋がる点に着目し、年間利用金額の増減に寄与する業種を分析する潜在クラスモデルを提案する。すなわち、1人のユーザのクレジット年間利用額の変化と、利用した業種での年間の利用額変化を共起させることで、どの業種での利用額変化が年間利用金額の増減に寄与するかを分析可能なモデルを示す。一方、ポイント分析モデルでは、蓄積されたポイント利用に伴う来店頻度の向上が利益増加に繋がること、また発行ポイントの利用率が高いユーザはグループ加盟店での来店頻度が高い優良ユーザであるという点に着目し、優良ユーザと非優良ユーザそれぞれの利用傾向を表現する潜在クラスモデルを提案する。具体的には、1人のユーザの年間のポイント利用度合いと、そのポイント利用先を共起させることで、ポイント利用度合いの高い優良ユーザの利用傾向と、利用度合いの低い非優良ユーザの利用傾向を比較可能なモデルを示す。

また、本研究の目的は、クレジット機能の利用額の増減に寄与する業種、およびポイント利用率の高低と利用先業種の傾向について分析し、ユーザの利用傾向をモデル化することである。そのため、各モデルで共起させる各変数について、変化率や利用率の値についてではなく「どの程度増加・減少したか」に関心がある。そこで、各指標をそれぞれ複数の閾値でカテゴリ化する。

### 3.2. 準備

あるユーザのある年度と前年度のクレジット機能の年間利用金額を  $\alpha_{\text{all}}$ ,  $\alpha'_{\text{all}}$  とすると、クレジット年間利用額の増減  $r_{\text{all}}$  は以下の式 (1) で定義される。

$$r_{\text{all}} = \frac{\alpha_{\text{all}} - \alpha'_{\text{all}}}{\alpha'_{\text{all}}} \quad (1)$$

同様に、 $A$  個の業種集合のうち、 $a$  番目の業種でのある年度と前年度のクレジット機能年間利用金額を  $\alpha_a$ ,  $\alpha'_a$  とすると、業種別年間利用額増減  $r_a$  は以下の式 (2) で定義される。

$$r_a = \frac{\alpha_a - \alpha'_a}{\alpha'_a} \quad (2)$$

ポイント利用について、あるユーザのある年度の年間付与ポイント数を  $\beta$ 、利用ポイント数を  $\gamma$  とすると、ポイント利用率  $\epsilon$  は以下の式 (3) で定義される。

$$\epsilon = \frac{\gamma}{\beta} \quad (3)$$

また、ポイントを利用可能な  $B$  個の業種集合のうち、 $b$  番目の業種での利用ポイント数を  $\mu_b$  とすると、その業種での利用割合  $\theta_b$  は以下の式 (4) で定義される。

$$\theta_b = \frac{\mu_b}{\sum_{b=1}^B \mu_b} \quad (4)$$

ただし、 $\sum_{b=1}^B \mu_b = \gamma$  である。

以下、クレジット機能の各指標について、 $r_{\text{all}}$  は  $I$  段階に離散化し、その要素集合を  $\mathcal{E} = \{e_i : 1 \leq i \leq I\}$  とする。 $r_a$  は  $J$

段階に離散化し、その要素集合を  $\mathcal{F}^a = \{f_{j_a}^a : 1 \leq j_a \leq J\}$  とする。ポイント機能の各指標について、 $\epsilon$  は  $H$  段階に離散化し、その要素集合を  $\mathcal{X} = \{x_h : 1 \leq h \leq H\}$  とする。 $\theta_b$  は  $G$  段階に離散化し、その要素集合を  $\mathcal{Y}^b = \{y_{g_b}^b : 1 \leq g_b \leq G\}$  とする。

### 3.3. 定式化

クレジット分析モデルについて、 $K$  個の潜在クラス集合を  $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$  とし、 $\mathbf{f} = (f_{j_1}^1, \dots, f_{j_a}^a, \dots, f_{j_A}^A)$  とすると、確率モデルは式 (5) で定義される。

$$P(e_i, \mathbf{f}, z_k) = P(z_k)P(e_i|z_k) \prod_{a=1}^A P(f_{j_a}^a|z_k) \quad (5)$$

ポイント分析モデルについて、 $L$  個の潜在クラス集合を  $\mathcal{W} = \{w_l : 1 \leq l \leq L\}$  とし、 $\mathbf{y} = (y_{g_1}^1, \dots, y_{g_b}^b, \dots, y_{g_B}^B)$  とすると、確率モデルは式 (6) で定義される。

$$P(x_h, \mathbf{y}, w_l) = P(w_l)P(x_h|w_l) \prod_{b=1}^B P(y_{g_b}^b|w_l) \quad (6)$$

### 3.4. パラメータ推定

提案モデルのパラメータは EM アルゴリズム [2] を用いて推定する。EM アルゴリズムは、対数尤度を局所最大化するパラメータを E-step と M-step の反復計算により求めるパラメータ推定手法である。

クレジット分析モデルについて、 $N$  人のユーザのうち  $n$  番目のユーザの年間利用額の増減度合いを  $c_n \in \mathcal{E}$ 、 $A$  個の要素からなる業種別の増減度合いベクトルを  $\mathbf{s}_n = (s_{n1}, \dots, s_{na}, \dots, s_{nA})$ ,  $s_{na} \in \mathcal{F}^a$  とすると、 $n$  番目のユーザの利用傾向は  $(c_n, \mathbf{s}_n)$  で与えられる。このとき、対数尤度  $LL_{\text{credit}}$  は以下の式 (7) で示される。

$$LL_{\text{credit}} = \sum_{n=1}^N \log \sum_{k=1}^K P(c_n, \mathbf{s}_n, z_k) \quad (7)$$

ポイント分析モデルについて、 $M$  人のユーザのうち  $m$  番目のユーザのポイント利用度合いを  $u_m \in \mathcal{X}$ 、 $B$  個の要素からなる業種別の利用度合いベクトルを  $\mathbf{t}_m = (t_{m1}, \dots, t_{mb}, \dots, t_{mB})$ ,  $t_{mb} \in \mathcal{Y}^b$  とすると、 $m$  番目のユーザの利用傾向は  $(u_m, \mathbf{t}_m)$  で与えられる。このとき、対数尤度  $LL_{\text{point}}$  は以下の式 (8) で示される。

$$LL_{\text{point}} = \sum_{m=1}^M \log \sum_{l=1}^L P(u_m, \mathbf{t}_m, w_l) \quad (8)$$

## 4. 実データ分析

### 4.1. 分析条件

分析データの対象期間は、2015年4月1日から2017年3月31日である。対象となるユーザは、クレジット分析モデルでは2015年度と2016年度の双方でクレジット機能の利用があった421,609人のユーザのうち、クレジット機能の利用額増減が  $r_{\text{all}} < 100$  のユーザ  $N = 421,398$  人とした。 $r_{\text{all}} \geq 100$  のユーザは前年度の途中から利用を開始したユーザである可能性が高く、必然的に翌年の利用額が大幅に上昇していると考えられるため、今回の分析では除外した。ポイント分析モデルでは、2016年度にポイント機能の利用 (蓄

表 1: 業種ごとの利用額変化 (抜粋)

$z_k$	$P(z_k)$	利用額の変化	百貨店		スーパー	スーパー	ショッピング	ショッピング	通信	レストラン・
			(G)	百貨店	マーケット (G)	マーケット	センター (G)	センター	販売	飲食
$z_1$	0.0816	増加	0	0	0	+	0	0	0	0
$z_2$	0.0285	増加	+	0	+	+	+	0	0	0
$z_3$	0.0620	増加	+	0	0	0	0	0	0	0
$z_4$	0.0319	やや増加	+	+	0	+	+	+	0	-
$z_5$	0.1929	増加・減少	0	0	0	0	0	0	0	0
$z_6$	0.1350	増加・減少	-	-	0	-	0	-	-	+
$z_7$	0.0795	やや減少	0	0	0	0	0	0	-	0
$z_8$	0.0776	減少	-	0	0	-	-	0	0	0
$z_9$	0.0250	減少	-	-	-	-	-	-	0	-
$z_{10}$	0.2859	減少	-	0	0	0	0	0	0	0

表 2: クラス別の主な利用先業種

$w_l$	$P(w_l)$	利用度	主な利用先
$w_1$	0.0624		ショッピングセンター (G)
$w_2$	0.0234	低	趣味・娯楽 (G)
$w_3$	0.1306		スーパーマーケット (G)
$w_4$	0.1256		百貨店 (G)
$w_5$	0.0982	中	ショッピングセンター (G)
$w_6$	0.3406		百貨店 (G)
$w_7$	0.0394		百貨店 (G), スーパーマーケット (G), ショッピングセンター (G)
$w_8$	0.0500		百貨店 (G), スーパーマーケット (G), ショッピングセンター (G), レストラン・飲食 (G)
$w_9$	0.0499	高	百貨店 (G), スーパーマーケット (G), ショッピングセンター (G)
$w_{10}$	0.0799		百貨店 (G), スーパーマーケット (G), ショッピングセンター (G)

積のみも含む)があった 377,142 人のユーザのうち、 $\epsilon \neq 0$  のユーザ  $M = 185,670$  人とした。対象となる業種は、クレジット分析モデルでは、全 42 業種中、2015 年度と 2016 年度の 2 年間において対象ユーザの 15% 以上が利用した業種である  $A = 14$  業種とした。ポイント分析モデルでは、全 18 業種中、ポイント利用先の上位の  $B = 6$  業種とした。

離散化については、 $r_{all}$ ,  $\epsilon$ ,  $\theta_b$  については各カテゴリに該当するデータ数が均等となるように  $I = 29$ ,  $H = 9$ ,  $G_b = 11$  とした。 $r_a$  については「 $r_a < 0$ 」, 「 $r_a = 0$ 」, 「 $r_a > 0$ 」の 3 段階に離散化し、 $J_a = 3$  とした。潜在クラス数は解釈性の高いクラス数を探索的に求め、 $K = L = 10$  とした。

#### 4.2. クレジット分析モデルの分析結果

表 1 に、各クラスの生起確率、利用額の変化、各業種での利用額変化の結果を示す。ここで利用額の変化は、利用額の増減に関するパラメータ  $P(e_i|z_k)$  に基づいて解釈を与えたものである。また、業種別の利用額変化について増加したもののについては +, 減少したもののについては -, 変化しなかったもののについては 0 と表記する。

表 1 より、例えば百貨店 (G) やスーパーマーケットでの利用では、その業種での前年度からの利用額が増加 (減少) するクラスで、クレジット機能の年間利用額も増加 (減少) していることが分かる。すなわち、前年度からのクレジット機能の利用額が増加する要因として、これらの業種での利用が増加したことが考えられる。一方で、通信販売やレストラ

ン・飲食での利用では、その業種での前年度からの利用額変化が、クレジット機能の年間利用額の変化と対応していない。また、同じ百貨店でも、グループ加盟店の百貨店 (G) の利用額増減がクレジット年間利用額の変化と対応づくのに対し、グループ加盟店外の百貨店の利用額は前年度から変化がない場合が多く、クレジット年間利用額の変化と対応がつかない。このことから、業種名だけでなく、加盟店か否かの情報を加味した分析が有用であることが示唆される。

以上のことから、クレジット機能の年間利用額の変化に寄与する業種は百貨店 (G) やスーパーマーケット、ショッピングセンター (G) などであることが分かる。つまり、クレジット機能の年間利用額が変化しているユーザの多くはこれらの業種での利用額を変化させているユーザである可能性が高く、これらの業種での利用を促すことがクレジット機能の年間利用額増加に繋がりやすいことを示唆している。

#### 4.3. ポイント分析モデルの分析結果

表 2 に各クラスの生起確率、利用先業種、利用度合いについての結果を示す。ただし、ポイント利用度合いについては、利用度合いのパラメータ  $P(x_h|w_l)$ 、利用先についてはパラメータ  $P(y_g^b|w_l)$  に基づき解釈を与えたものである。

クラス  $w_1-w_3$  ではショッピングセンター (G) や趣味・娯楽 (G) など特定の業種での利用が主であり、ポイントの利用度合いが低いクラスである。このことから、これらのクラスではポイント機能に対する意識が低く、低頻度でのポイン

ト利用であり、また利用する業種は限定的であるクラスと解釈できる。クラス  $w_4-w_6$  では、クラス  $w_1-w_3$  と同様に、百貨店 (G) やショッピングセンター (G) など特定の業種での利用が主である一方、ポイントの利用度合いがそれらのクラスと比較して高いクラスである。このことから、ポイントに対する意識は低くはないが、利用先が限定的であり、特定の業種を利用した際にポイントを用いるクラスと解釈できる。クラス  $w_7-w_{10}$  ではいずれも百貨店 (G)、ショッピングセンター (G)、スーパーマーケット (G) の3業種が主な利用先であるが、 $w_7$  と  $w_8-w_{10}$  で利度合いに違いがある。このことから、クラス  $w_8-w_{10}$  ではポイントに対する意識が高く、その利用先は百貨店 (G) やスーパーマーケット (G) などに多く見られることから、日常の購買活動で積極的にポイントを利用しているクラスであると解釈できる。また、クラス  $w_7$  は、クラス  $w_8-w_{10}$  と利用傾向が潜在的に近いが、利度合いに違いがあるクラスと解釈できる。

このことから、ポイント機能の優良ユーザ、つまりポイントに対する意識が高くグループ加盟店を繰り返し利用するユーザの利用傾向は、日常的に買い物を行う業種での利用が多く、様々な業種で横断的に利用しているといえる。故に、ポイント利度合いの低いユーザのうち、グループ加盟店の百貨店 (G) やスーパーマーケット (G) で日常的に購買活動を行うユーザに対するポイント利用の呼びかけやポイント機能に対する意識付けを重点的に行うことで、グループ加盟店の利用増加に繋がると考えられる。

#### 4.4. 2つのモデルのクロス分析結果と考察

双方の分析モデルの条件を満たすユーザ  $V = 147,177$  人についてクロス分析 [1] を行う。以下の表 3 に、 $(z_k \times w_l)$  の所属人数をセグメントごとに平均した結果を示す。

表 3: クロス分析結果

		$w_1-w_3$	$w_4-w_7$	$w_8-w_{10}$
		低	中	高
$z_1-z_4$	増加	2352.0	1362.9	728.8
$z_5-z_6$	増加・減少	2525.9	1496.8	844.7
$z_7-z_{10}$	減少	2188.6	1283.4	729.9

表 3 より、クレジット利用率の変化に依らずポイントの利用度が低いユーザが多いことが分かる。また、ポイント利用度が低く、クレジット利用額が増加したユーザ群は、クレジット決済額に応じてポイントが付与されることから、利用可能なポイントが増加しているといえる。そのため、施策を優先的に打つべき対象として、これらのユーザ群に対してポイント利用を促す施策を講じることは、他のセグメントのユーザ群に比べポイント利用をした際の効果が大きいと考えられる。

具体的な施策を打つべき業種としては、先述の結果より、百貨店 (G) とショッピングセンター (G) は、主たるクレジット機能の利用とポイントの利用先において共通であった。つまり、これら 2 業種でのクレジット決済を促すことがクレジット利用額増加に繋がり、結果的に蓄積ポイントの増加に直結する。そのポイントを次回来店時に利用してもらうよう促すことで、クレジット機能とポイント機能の双方の利用を向上させる相乗効果が生まれると考えられる。故に、クレジット

機能とポイント機能の利用促進をこれら 2 業種で行うことが効果的であると考えられる。

## 5. 考察

本研究では提案したクレジット分析モデルとポイント分析モデルの 2 つのモデルを独立に構築した。これら双方のモデルを用いることで、ユーザの付与ポイント増加から、カード利用機会の増加までの一貫した利用を促す施策の検討が可能となる。また、これらのクロス分析の結果から、施策効果が大きいユーザ群を特定可能であり、双方のモデルを同時に適用することは有用であると考えられる。

一方で、本研究では独立に構築した 2 つのモデルの組み合わせで利用傾向を表現したが、2 つの特性を同時に表現した単一のモデリングも考えられる。しかしながら、複雑な構造のモデリングとなるため、パラメータ推定時に過学習を防ぐ策を講じる必要があると考えられる。また、本研究のように一方の機能のみを重点的に利用するユーザの利用傾向も考慮する場合、利用のない機能のデータに関しては欠損データとなり、パラメータの学習時に工夫も必要となる。これらの点を考慮すると、本研究で提案した 2 つのモデルはよりシンプルで精度と取り扱いの容易性の面から合理的であると考えられる。

## 6. まとめと今後の課題

本研究では、多機能クレジットカードの有するクレジット機能とポイント機能の 2 つの機能に着目し、これらの機能の利用傾向を表現する潜在クラスモデルを提案した。また、実際の多機能クレジットカードの利用履歴データに提案モデルを適用して分析を行い、クラスごとに利用傾向を表現できることを確認した。さらに、得られた結果のクラス間を比較することで、実際にクレジット機能とポイント機能の利用促進の一助となる施策に繋がる可能性を示した。

本研究ではユーザの利用した業種に着目した分析手法の提案を行ったが、ユーザの属性情報など、今回は考慮していない様々な着眼点がある。また、得られた結果に対する施策の定量的な評価や潜在クラス数の客観的な決定方法なども検討の余地があり、これらについては今後の課題とする。

## 参考文献

- [1] Ryotaro Shimizu, et al., "Consumer Purchasing Behavior Analysis Model for Purchase History Data Stored in Credit Card," *The 19th Asia Pacific Industrial Engineering and Management Systems*, ID 195, Full Paper, 2018.
- [2] C. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, No.2, 2010.
- [3] Chan, Philip K., et al. "Distributed Data Mining in Credit Card Fraud Detection." *IEEE Intelligent systems* 6, pp.67-74, 1999.
- [4] 中原孝信, 森田裕之, "百貨店のクレジット購買データを用いた関連購買による顧客特徴分析," *オペレーションズ・リサーチ*, Vol.50, No.7, pp.488-494, 2005.
- [5] Nie, Guangli, et al., "Credit Card Customer Analysis based on Panel Data Clustering," *Procedia Computer Science* 1.1, pp.2489-2497, 2010.