

顧客成長のための施策立案を導く特徴転移型クラスタリングモデルに関する研究

経営情報学研究

5219F016-6

平野洋介

指導教員

後藤正幸

A Study on Feature Transfer Based Clustering for Designing Customers Growth Measures

HIRANO Yosuke

1. はじめに

近年、顧客に様々な特典を提供し、企業と顧客が長期的に良好な関係を築くためのツールとして、ポイントカードやクレジットカードが広く普及している。本研究で対象事例とする小田急電鉄株式会社が発行する小田急ポイントカード（以下、OP カード）もその一つである。OP カードには、OP ポイント専用カードに加え、クレジット機能付きポイントカードである OP クレジットがある。ポイント専用カードは、加盟店での購買時にカードを提示した際に、加盟店利用による OP ポイントが付与される。蓄積された OP ポイントは、加盟店での支払い手段の 1 つとして利用可能である。一方、クレジット機能付きポイントカードは、クレジットカードとして加盟店内外で使用することで、クレジット決済額に応じて OP ポイントが加算される機能を持つ。クレジット機能付きポイントカードの所持ユーザ（以下、クレジットユーザ）は、ポイント専用カードの所持ユーザ（以下、ポイントユーザ）に比べて、ポイントが蓄積されやすくその OP ポイントをきっかけとした加盟店の利用が見込まれる。また、ポイント蓄積に向けた加盟店内外での継続的なクレジット機能の利用も期待できる。そのため、ポイントユーザをクレジットユーザに成長させることで、クレジット機能とポイント機能の相乗効果により優良ユーザとなる可能性を高め、ユーザとの長期的に良好な関係の構築が期待できる。すなわち、企業側の観点からは、クレジットユーザを増加させることは 1 つの望ましい戦略である。しかし、クレジットカードの作成は、ユーザにとって心理的ハードルが高いため、ポイント専用カードを顧客に配布することでポイントユーザとしての入会を促している。その後、ポイント専用カードからクレジット機能付きポイントカードへの切り替えを促進する様々な施策を講じることで、優良顧客への成長が期待できる。施策を講じる際、多様な利用傾向を持つポイントユーザ全体に対して一律の施策を講じたとしても効果的であるとは限らない。逆に、ユーザ個別に施策を検討することも施策コストの側面から現実的ではない。ゆえに、施策対象とするべきユーザセグメントを定め、そのセグメントの特性に合わせた施策を検討する必要がある。そこで、本研究ではポイントユーザを加盟店の利用傾向に基づいてクラスタリングし、セグメントごとに施策を検討するセグメントマーケティングを考える。

ユーザを統計的特徴によってセグメントに分類するにあたり、潜在クラスモデル [1] に基づくソフトクラスタリングの適用を考える。ここで、本研究の対象は切り替え促進施策で

あることから、対象となるポイントユーザがステップアップする先となるであろうクレジットユーザのセグメントを明確にすることで、施策対象群の特定とともに切り替え促進に向けたより効果的な施策の検討が可能になると考えられる。すなわち、利用傾向を基にポイントユーザのクラスタリングを行う際、成長目標となるクレジットユーザのグループと紐づける形でクラスタ化できれば、このようなセグメントごとに利用傾向を反映した施策に結び付けることができる。

そこで本研究では、切り替え施策立案に向けてユーザの利用傾向を表現した潜在クラスモデルにより、クレジットユーザの各クラスの特徴を用いて、ポイントユーザをクラスタリングする手法を提案する。具体的には、転移学習 [2] の考え方を援用する。クレジットユーザのデータから潜在クラスモデルを学習させ、推定したクレジットユーザの潜在クラスのパラメータをポイントユーザの潜在クラスモデルに転移し固定したうえで、ポイントユーザのデータで追加学習してクラスタリングを行う。これにより、クレジットユーザの所属する潜在クラスを事前に定め、その潜在クラスの特徴に類似した利用傾向のポイントユーザを対応させることで、クラス別の施策立案を強力に支援することができる。最後に、提案手法を実際の利用履歴データに適用し有用性を示す。

2. 準備

2.1. 問題設定

本研究では、小田急電鉄株式会社が発行する OP カードを対象事例とする。OP カードにおける優良ユーザは、クレジット機能、ポイント機能双方において利用の多いユーザと定義される。戦略的に優良ユーザを増やすことを考えると、いくつかのユーザ成長モデルを想定することが有効である [3]。図 1 に示すように、優良ユーザへの成長ルートとしては、「グループ加盟店の利用頻度⇒クレジット利用頻度」の順に向上させるルートと、「クレジット利用頻度⇒グループ加盟店の利用頻度」の順に向上させるルートが考えられる。このもとの Shimizu et al. [3] においては、クレジット利用履歴データ、ポイント利用履歴データに対して、カードの所持目的、購買行動の多様性を表現する潜在クラスモデルを用いて、これらを統合して分析する手法が提案されている。この研究では、図 1 に示す通り、如何にしてクレジット機能、ポイント機能の双方において利用の促進を図るかを検討している。

これに対して、本研究では現状でクレジット機能を全く利用していないポイントユーザに着目する。ポイントユーザは図 1 における横軸方向にポジショニングされる。これらの

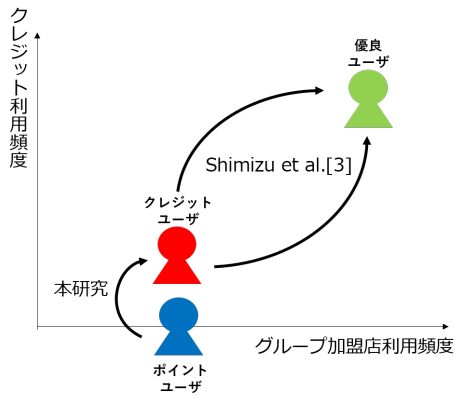


図 1: 優良ユーザへの成長ルートモデル

ユーザの中には、適切な施策を講じることでクレジットユーザへ成長する可能性を持ったユーザが含まれている。ポイントユーザをクレジットユーザに成長させることで、OP クレジットがユーザの日々の購買行動における決済手段の選択肢の1つとなり、OP クレジットにより蓄積された OP ポイントをきっかけとした加盟店での再利用も見込まれる。加えて、加盟店での利用ポイント蓄積に向けたクレジット機能の利用も期待することができる。このように、クレジットユーザに成長させることでクレジット機能とポイント機能の相乗効果が期待できるため、ポイントユーザへの OP クレジットへの切り替え促進は重要な課題である。以上より、本研究では OP ポイント専用カードから OP クレジットへの切り替え施策の検討を対象とする。

2.2. 事前分析

本研究では、ある一定期間内に入会したユーザを対象とした施策を想定し、2017年1月1日～2018年12月31日において OP ポイント専用カードに入会したユーザを分析対象とする。対象ユーザの中で 2018年12月31日時点で OP クレジットに切り替えを行ったユーザと OP ポイント専用カードを継続利用するユーザをそれぞれクレジットユーザ、ポイントユーザとし、ユーザ数の合計を表 1 に示す。表 1 より、ユーザ数に大きな偏りがあることがわかる。

表 1: 各カード所持ユーザ数

利用カード種類	ユーザ数
OP クレジット	22,836
OP ポイント専用カード	78,518

3. 提案手法

3.1. 提案への着想

ポイント専用カードからクレジット機能付きポイントカードへ切り替えを促す施策を考える際、これらの施策の多くは切り替え時のポイント付与などのインセンティブコストが必要となるため、費用対効果の高い施策を検討する必要がある。その際、OP カードの特性として 2 つの考慮すべき点が挙げられる。まず、1 点目は利用店舗の業種や地域の多様性である。OP カード加盟店には幅広い地域に様々な業種が存在し

ており、各ユーザの店舗の利用傾向は多種多様である。そのため、利用金額や利用頻度、ポイントによる支払い回数など各加盟店をどの程度利用しているかを表す利用状況に加えて、利用店舗の属性を考慮したモデル化が必要である。この時、利用状況、利用店舗の属性についてそれぞれモデル化することも考えられる。しかし、利用店舗は利用状況に強く影響を与えることが考えられるため、単一のモデル化が望ましい。

2 点目としては、同時期に入会した会員においてポイントユーザがクレジットユーザに比べて極端に多いことである。本研究で対象とするのは切り替え促進の施策であることから、クレジットユーザの利用パターンも同時に考慮することで、どのようなクレジットユーザへステップアップするのか明確になる。しかし、同時期に入会したユーザでは、クレジットユーザの割合はポイントユーザよりも極端に少なく、これらのユーザを同時にクラスタリングした場合、ポイントユーザの特徴に引っ張られてしまい、潜在クラスの推定にクレジットユーザの特徴が反映されない。そこで本研究では、クレジットユーザのクラスを事前に定め、そのクラスの特徴に基づいてポイントユーザをクラスタリングする 2 段階の学習を考える。その際、クレジットユーザとは異なる特徴を有したポイントユーザも存在することが考えられるため、これらのユーザに対しては新たなクラスを設定する必要がある。

3.2. クラスタリングモデルの定式化

本研究では潜在クラスモデルを用いて、購買行動によってユーザのクラスタリングを行う。ここで、ユーザが加盟店をどの程度利用しているかを表す利用状況ベクトル、ユーザがどの地域に所在する店舗をどの程度利用しているかを表す地域利用頻度ベクトル、どの業種の店舗をどの程度利用しているかを表す業種利用頻度ベクトルの 3 つのベクトルを用いてユーザの特徴を表現する。そして、利用状況ベクトル、地域の利用頻度ベクトル、業種の利用頻度ベクトルの間に潜在変数を仮定し、ユーザの利用傾向を表現する潜在クラスモデルを構築する。この時、各ユーザが持つ利用状況ベクトルについては多次元正規分布を仮定する。しかし、ユーザの各地域の利用頻度、各業種の利用頻度についてはそれぞれに対してどの程度利用しているかを把握する必要がある。そこで、各ユーザが持つ地域の利用頻度ベクトル、業種の利用頻度ベクトルにはそれぞれ独立した多項分布を仮定する。ここで、潜在クラス集合を $Z = \{z_k : 1 \leq k \leq K\}$ とし、各ユーザはいずれかの潜在クラスに所属すると仮定する。また、各ユーザに対し利用状況ベクトル $\mathbf{x} = (x_1, \dots, x_H)$ 、地域利用頻度ベクトル $\mathbf{s} = (s_1, \dots, s_I)$ 、業種利用頻度ベクトル $\mathbf{t} = (t_1, \dots, t_J)$ が定義されているものとする。このとき、これらの共起事象の確率モデルを式 (1) のように定義する。

$$p(\mathbf{x}, \mathbf{s}, \mathbf{t}) = \sum_{k=1}^K p(\mathbf{x}|z_k)p(\mathbf{s}|z_k)p(\mathbf{t}|z_k)p(z_k) \quad (1)$$

3.3. 提案モデル

式 (1) の潜在クラスモデルを用いてクレジットユーザの利用傾向をもとに、ポイントユーザをクラスタリングする手法

を提案する。

各ユーザはいずれかの潜在クラスに所属すると仮定する。ここで、クレジットユーザの潜在クラス集合を $Z^c = \{z_k^c : 1 \leq k \leq K\}$ 、ポイントユーザにおける施策対象外ユーザの潜在クラス集合を $Z^p = \{z_l^p : 1 \leq l \leq L\}$ と定義する。この時、クレジットユーザにおける $\mathbf{x}, \mathbf{s}, \mathbf{t}$ の生成モデルを式 (2) のように定義する。

$$p(\mathbf{x}, \mathbf{s}, \mathbf{t}) = \sum_{z \in Z^c} p(\mathbf{x}|z)p(\mathbf{s}|z)p(\mathbf{t}|z)p(z) \quad (2)$$

これに対し、ポイントユーザにおける $\mathbf{x}, \mathbf{s}, \mathbf{t}$ の生成モデルとしては式 (3) を仮定する。ただし、 $z \in Z^c$ に対しては、 $\tilde{p}(\mathbf{x}|z) = p(\mathbf{x}|z), \tilde{p}(\mathbf{s}|z) = p(\mathbf{s}|z), \tilde{p}(\mathbf{t}|z) = p(\mathbf{t}|z)$ である。

$$\begin{aligned} \tilde{p}(\mathbf{x}, \mathbf{s}, \mathbf{t}) &= \sum_{z \in Z^c \cup Z^p} \tilde{p}(\mathbf{x}|z)\tilde{p}(\mathbf{s}|z)\tilde{p}(\mathbf{t}|z)\tilde{p}(z) \\ &= \sum_{z \in Z^c} p(\mathbf{x}|z)p(\mathbf{s}|z)p(\mathbf{t}|z)p(z) \\ &\quad + \sum_{z \in Z^p} \tilde{p}(\mathbf{x}|z)\tilde{p}(\mathbf{s}|z)\tilde{p}(\mathbf{t}|z)\tilde{p}(z) \end{aligned} \quad (3)$$

上のモデル式は、クレジットユーザよりもポイントユーザの利用パターンの多様性が高いことを仮定している。

3.4. 学習アルゴリズム

これらのモデルの学習では、クレジットユーザに式 (2) の潜在クラスモデルを適用して推定した潜在クラスのパラメータの値を、ポイントユーザに式 (3) の潜在クラスモデルを適用する際の学習に転移する。具体的には、第 1 段階のクレジットユーザからパラメータの意味を固定し、次にポイントユーザのデータから新たに設定したクラスのパラメータのみを推定する 2 段階の学習を行う。ここで、クレジットユーザのクラスの特徴を明確にとらえるため、転移したパラメータの値を固定する必要がある。ただし、パラメータ推定にはデータに対する過学習を防ぐことに加え、クラス数の選択のため変分推論 [4] を適用する。以下に提案手法の学習アルゴリズムの具体的なステップを示す。

[Step1]

クレジットユーザのデータ集合に潜在クラスモデルを適用し、式 (2) のモデルのパラメータ $p(\mathbf{x}|z), p(\mathbf{s}|z), p(\mathbf{t}|z), p(z)$ を推定する。

[Step2]

ポイントユーザのデータ集合に対して、潜在クラスモデルを適用する。この時、[Step1] で推定したパラメータ $p(\mathbf{x}|z), p(\mathbf{s}|z), p(\mathbf{t}|z)$ を転移し固定して、式 (3) のモデルの $z \in Z^p$ に対するパラメータ $\tilde{p}(\mathbf{x}|z), \tilde{p}(\mathbf{s}|z), \tilde{p}(\mathbf{t}|z), \tilde{p}(z)$ と $z \in Z^c$ に対しては $\tilde{p}(z)$ のみを推定する。

4. 実データ分析

4.1. 分析条件

本分析の対象データは、OP カードの利用履歴データである。2017 年 1 月 1 日～2018 年 12 月 31 日までに OP カードに入会したユーザを分析対象とし、クレジットユーザは 2018 年 12 月 31 日までに OP クレジットに切り替えた

ユーザ、ポイントユーザは 2019 年 12 月 31 日まで OP ポイント専用カードを継続利用しているユーザとした。ユーザ数はそれぞれ 22,836 件、78,518 件である。ユーザの利用履歴データ期間は 2019 年 1 月 1 日から 12 月 31 日までとした。利用状況ベクトルの次元数は $H = 6$ であり、地域利用頻度ベクトル、業種利用頻度ベクトルの次元数はそれぞれ $I = 20, J = 18$ である。分析に用いる変数の用語の定義を表 2 に示す。各潜在クラス数は変分推論を適用することからクラス数を多めに取り、それぞれ $K = 15, L = 15$ と設定した。

表 2: 各変数における用語の定義

変数名	定義
x_1 : 利用金額平均	期間内での 1 回あたりの利用金額
x_2 : 利用頻度	期間内での加盟店の利用回数
x_3 : 利用業種数	期間内で利用した業種数
x_4 : 付与ポイント合計	期間内に付与されたポイント数
x_5 : ポイント支払い回数	期間内にポイントで支払いを行った回数
x_6 : ポイント支払い金額	期間内での 1 回あたりのポイントでの支払い金額
\mathbf{s} : 業種利用頻度	期間内での業種別店舗利用頻度ベクトル
\mathbf{t} : 地域利用頻度	期間内での地域別店舗利用頻度ベクトル

4.2. 分析結果

[Step1], [Step2] の推定において潜在クラス比率が 0 より大きくなった潜在クラスについて、各潜在クラスの所属割合と利用状況を表す変数における平均値を表 3 に示す。クラス数に着目すると、提案手法の学習アルゴリズム [Step1], [Step2] の推定において、それぞれ 2 クラス、8 クラスで潜在クラス比率が 0 となった。これはパラメータの推定に変分推論を適用したことで、必要最小限のクラス数が選択された結果と考えられる。また、各潜在クラスにおける代表的な利用地域と利用業種を一部抜粋して表 4 に示す。

結果の分析の具体事例として、まず、潜在クラス z_9^c, z_{10}^c に着目すると、利用状況については同様の傾向を示すことがわかる。表 4 から z_9^c に所属するユーザは世田谷・狛江において保育を、 z_{10}^c に所属するユーザは五月台～唐木田においてその他の業種を利用しており、利用地域と利用業種については異なる傾向がみられる。すなわち、利用状況が同程度でありながら利用する店舗が異なるユーザ群の存在が窺える。このように、利用店舗の属性を考慮したことで、詳細なユーザ特性の把握が可能となった。

次に、潜在クラス z_9^p, z_{11}^p に着目する。表 4 から z_{11}^p に所属するユーザは世田谷・狛江といった地域でハウスクリーニングや保育といった業種を利用している。すなわち、 z_{11}^p と z_9^c は利用業種、利用地域については似通った傾向を持つ。ここで、表 3 において、 z_{11}^p は z_9^c と比較すると利用頻度や付与ポイント合計といった値が小さい。すなわち、世田谷・狛江エリアで保育を利用するポイントユーザの中で利用状況に差があることがわかる。潜在クラス z_{11}^p に所属するユーザは、加盟店においては限定的な業種を利用している可能性が考えられる。そのため、OP カードをメインのクレジットカードとすることに魅力を感じにくいことが想定でき、施策優先度の低いユーザ群であると解釈できる。このようにクレジットユーザのクラスを考慮し、ポイントユーザをクラスタリング

表 3: 各クラスの所属割合と利用状況の平均値

z	z_1^c	z_2^c	z_3^c	z_4^c	z_5^c	z_6^c	z_7^c	z_8^c	z_9^c	z_{10}^c	z_{11}^c	z_{12}^c	z_{13}^c	z_1^p	z_2^p	z_3^p	z_4^p	z_5^p	z_6^p	z_7^p
$p(z)$	0.215	0.100	0.044	0.096	0.075	0.053	0.023	0.021	0.069	0.014	0.009	0.265	0.016	0.108	0.082	0.043	0.025	0.015	0.015	0.004
$\tilde{p}(z)$	0.232	0.105	0.077	0.074	0.057	0.040	0.039	0.035	0.019	0.011	0.010	0.006	0.002	0.108	0.082	0.043	0.025	0.015	0.015	0.004
利用金額平均	5731.2	1910.6	2953.6	3297.9	1849.0	1717.4	1860.6	1603.7	1926.0	1751.3	23637.6	2298.2	1347.8	1729.8	1518.7	3761.0	2140.9	2484.1	2373.1	28289.5
利用頻度	36.7	74.3	47.3	47.8	82.1	103.6	133.1	98.8	88.2	85.8	89.8	19.2	151.8	73.9	63.7	10.6	69.9	44.1	83.0	173.5
利用業種数	2.0	3.1	2.4	2.6	3.7	3.9	4.6	3.9	4.0	4.3	3.3	2.1	4.1	2.9	1.6	1.4	3.3	2.1	2.8	2.6
付与ポイント合計	7017.3	3245.7	3245.5	5061.8	3693.5	4717.8	7849.5	4317.3	4952.3	4912.3	13777.1	2767.8	7675.0	1446.8	1473.0	471.1	1464.7	1420.8	2123.3	34787.3
ポイント支払い回数	3.8	3.9	3.5	3.8	4.5	6.5	8.1	5.7	6.0	6.0	6.0	1.7	6.4	4.5	5.5	1.6	4.8	3.5	4.4	24.2
ポイント支払い金額	1682.6	814.6	826.0	1257.0	907.3	1066.6	1125.5	935.2	1103.5	1132.5	4512.8	983.9	1823.1	348.5	362.0	84.2	314.9	281.4	453.4	988.6

表 4: 各クラスにおける代表的な利用地域と利用業種 (一部抜粋)

z	z_1^c	...	z_9^c	z_{10}^c	...	z_1^p	...
代表的な利用地域	新宿 オンライン	...	世田谷・狛江	五月台～唐木田	...	世田谷・狛江	...
代表的な利用業種	百貨店	...	保育	その他	...	ハウスクリーニング 保育	...

したことで、施策対象群を特定できたと考えられる。

最後に、クラスの特徴から施策案について検討する。例えば、施策対象ユーザにおいてボリュームゾーンになると考えられる潜在クラス z_1^c について考察する。潜在クラス z_1^c に所属するユーザは、表 3 より購買単価が高く、貯まったポイントはある程度利用しているユーザ群であると解釈できる。また、表 4 から地域としては新宿やオンライン、業種としては百貨店を利用している。OP クレジットによりどのくらいポイントが貯まりやすくなるのかについて、新宿の小田急百貨店やオンラインショッピングを例として示すことが効果的なのではないかと考えられる。また、OP カードは年会費がかかるカードであるが、クレジットによる支払いを年間で1度でも行えば翌年は年会費がかからない。一般に、オンラインショッピングでは支払いの際にクレジットカードを用いて決済を行うことが考えられるため、オンラインショッピングにおいて OP クレジットを利用すれば年会費がかからなくなることを強調した施策により、切り替えへのハードルが下がり、施策効果は高まると期待できる。

5. 考察

本分析結果における施策の検討には、 $p(z), \tilde{p}(z)$ を考慮することが重要である。本研究で対象としたポイントユーザからクレジットユーザへの切り替え施策は、施策への反応度が高もともと非常に小さいことが知られており、最終的なゴールである優良ユーザの増加を考えたときに、小規模なユーザ群に対する施策ではその先にある優良ユーザの増加に対する寄与はあまり見込めない。すなわち、 $\tilde{p}(z)$ がある程度大きなクラスに対して施策を講じる方が効果的と考えられる。しかし、 $\tilde{p}(z)$ がある程度大きな場合でも $p(z)$ が小さいクラス、すなわち、クレジットユーザとしては稀なユーザ群の潜在クラスに所属するポイントユーザは、成長プロセスとして切り替え施策が効果的であるとは限らない。そのため、 $p(z)$ の値も考慮する必要がある。同様に潜在クラス集合 Z^p のクラスに所属するユーザは、ポイントユーザ固有のユーザ群であり、施策対象としては考えづらい。しかし、本分析における z_{11}^c や z_7^p に所属するユーザは利用状況の値から加盟店をよく利用するユーザであり、優良ユーザへと成長する可能性が高い

ユーザ群であると考えられ、ユーザの成長を促すという観点からは施策対象とすべきユーザ群であるとも考えられる。したがって、本分析結果に対して多角的な視点から、施策対象群を定め、施策案を検討していくことが望ましい。

6. まとめと今後の課題

本研究では、ステップアップする方向性を定めて、クレジット機能付きポイントカードへの切り替えを促す施策の検討を可能とする特徴転移型のクラスタリング手法を提案した。具体的には、ユーザの利用傾向を表現した潜在クラスモデルにより、クレジットユーザの各クラスの特徴を用いて、ポイントユーザをクラスタリングする手法を示した。提案手法を実際の利用履歴データに適用して分析を行い、ユーザの利用店舗の属性を考慮することによって詳細なユーザの利用傾向の把握が可能であることを確認した。また、クレジットユーザのクラスの特徴を用いてポイントユーザをクラスタリングすることで施策対象群を特定可能であることを示した。最後に、クラスの特徴から具体的な施策案について検討を行い、提案モデルを用いることで施策対象群を定め、そのセグメントの特性に合わせて有用な施策の検討が可能であることを示した。

本研究では、ユーザ属性を考慮していない。しかし、ユーザの属性情報は施策立案を検討するうえで有用な示唆の獲得に寄与する可能性がある。また、分析結果から導かれる施策については、実応用上での実証的な評価が望ましく、併せて今後の課題とする。

参考文献

- [1] J. Hagenars and A. McCutcheon, *Applied Latent Class Analysis*, Cambridge University Press, 2002.
- [2] 神島敏弘, “転移学習,” 人工知能, Vol.25, No.4, pp.572-580, 2010.
- [3] Ryotaro Shimizu, et al., “Latent Variable Models for Integrated Analysis of Credit and Point Usage History Data on Rewards Credit Card System,” *International Business Research*, Vol.13, No.3, 2020.
- [4] C. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, No.2, 2010.