

顧客成長を促す施策立案のための 特徴転移型クラスタリングモデル

平野 洋介^{1,a)} 楊 添翔¹ 雲居 玄道¹ 阿部 永² 立花 徹也³ 後藤 正幸¹

受付日 2021年1月8日, 採録日 2021年6月7日

概要: 近年, 顧客に様々な特典を提供し, 企業と顧客の良好な関係を築くためのツールとして, 購買金額に応じて顧客にポイントが還元されるポイントカードシステムが広く普及している. 現在では, 従来のポイント専用カードにクレジット機能が追加されたクレジット機能付きポイントカードも利用されるようになった. クレジット機能付きポイントカードの所持ユーザ (クレジットユーザ) は, ポイント専用カードの所持ユーザ (ポイントユーザ) に比べて, ポイントが蓄積されやすく加盟店の利用増加が見込まれる. そのため, 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替えを促す様々な施策検討の際に施策対象のポイントユーザを, どのクレジットユーザのセグメントにステップアップしてもらうのかを明確にすることができれば, 施策対象群の特定とともに切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. そこで, 本研究では, 切替え施策立案に向けて, クレジットユーザをカード利用傾向からクラスタリングし, 得られたクレジットユーザの各クラスの特徴を転移させることで, ポイントユーザをクラスタリングする手法を提案する. 提案手法では, クレジットカードの各ユーザクラスに, 特徴が類似している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可能となる. 最後に, 提案手法を実際の購買履歴データに適用し有用性を示す.

キーワード: 購買行動分析, 潜在クラスモデル, トピックモデル, クラスタリング, 転移学習

Feature Transfer Based Clustering for Designing Customers Growth Measures

YOSUKE HIRANO^{1,a)} TIANXIANG YANG¹ GENDO KUMOI¹ HARUKA ABE² TETSUYA TACHIBANA³
MASAYUKI GOTO¹

Received: January 8, 2021, Accepted: June 7, 2021

Abstract: In recent years, the point card system has been widespread as a tool for building good relationships between companies and customers by the special favor. Some of companies provide point cards with credit card function for their customers. Compared to users of point card, users of point card with credit card function are more likely to become good customers. Therefore, when we consider various measures to encourage users to switch from point-only cards to point cards with credit functions, if we can clarify that what kind of the credit user segment the point-only card user will step up, it will be possible to identify groups and consider more effective measures to promote switching. In this study, we propose a method for planning of user switching measures. First, we cluster the card usage tendency of credit users. Then, we transfer the obtained credit users' characteristics of each class to the point users for clustering. In our proposal, it is possible to strongly support policy planning for each class by associating point users with similar characteristics with each class of credit users. Finally, we clarify the effectiveness of our model by using actual data.

Keywords: purchasing behavior analysis, latent class model, topic model, clustering, transfer learning

¹ 早稲田大学
Waseda University, Shinjuku, Tokyo 169–8555, Japan

² 小田急電鉄
Odakyu Electric Railway Co., Ltd., Shinjuku, Tokyo 160–8309, Japan

³ 小田急エージェンシー
Odakyu Agency Inc., Shinjuku, Tokyo 160–0023, Japan

a) yosuke199688@akane.waseda.jp

1. はじめに

近年、顧客に様々な特典を提供し、企業と顧客の良好な関係を築くためのツールとして、ポイント専用カードやクレジット機能付きポイントカードが広く普及している。

このとき、多くの顧客を有する一般ユーザであるポイント専用カードの所持ユーザ（以下、ポイントユーザ）を上位ステージユーザである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）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が高く、ユーザとの長期的に良好な関係を構築できると考えられる。しかし、最初から上位ステージユーザとして入会させることは、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一般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替えを促進する様々な施策を講じることで、優良ユーザとなりうる可能性の高い上位ステージへの成長が期待できる。しかし、施策を講じる際、それぞれ異なる利用傾向を持つ一般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、1人1人のユーザに個別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集合を定めてそのユーザの特性に合わせた施策を検討する必要がある。そこで、本研究では一般ユーザ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 [1] を考える。

ユーザを統計的特徴によってセグメントに仕分けする手法としては、潜在クラスモデル [2] に基づくソフトクラスタリングがよく知られている。潜在クラスモデルをベースとした様々なモデルは、近年ではトピックモデルとも呼ばれ [3]、顧客の購買行動分析に様々な形で適用されており [4], [5], [6]、ユーザが持つ利用傾向の類似性による確率的なクラスタリングによってセグメントを抽出する [7], [8]。これに対して、本研究では切替え促進施策の立案のため、一般ユーザに対する適切なクラスタリングを検討したいという目的がある。もし、一般ユーザのみをクラスタリングした場合、個々の潜在クラスの特徴は、上位ステージユーザの利用傾向をいっさい考慮せずに与えられる。しかし、クラスタリングされた一般ユーザのセグメントが、利用傾向の類似性によって、上位ステージユーザのセグメントに紐付けられていれば、「どのような上位ステージユーザに成長してもらうような施策を検討するか」という観点での検討が可能であり、より具体的な施策立案につながる。すなわち、どのような上位ステージユーザのセグメントにステップアップするのかを明確にすることで施策対象群の特定とともに切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。

そこで本研究では、切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、上位ステージユーザの各クラスの特徴を用いて、一般ユーザをクラスタリングする手法を提案する。しかし、単純に一般ユーザと上位ステージユーザを合わせてクラスタリング手法を適用しても、一般ユーザの数が圧倒的多数であるため、上位ステージユーザの特徴が加味された潜在クラスは生成されない。また、アンダーサンプリングで一般ユーザを減らそうとすると、分析対象である一般ユーザのデータ数を減らしてしまうことになる。そこで本研究では、転移学習 [9], [10] の考え方を導入することで、上位ステージユーザの利用傾向をとらえつつ、一般ユーザのクラスタリングをそれに合わせる形で実行する転移学習型クラスタリングを提案する。まず、上位ステージユーザのデータから潜在クラスモデルを学習させ、その後推定した上位ステージユーザの潜在クラスのパラメータを転移させて、一般ユーザのデータを潜在クラスモデルで学習させて、クラスタリングを行う。これにより、上位ステージユーザと一般ユーザの双方に利用傾向の特徴が同じ潜在クラスが生成される。上位ステージカードの各ユーザクラスに、特徴が類似している一般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可能となる。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. 関連研究

2.1 潜在クラスモデルと購買行動分析

近年、蓄積された購買履歴データからユーザの購買行動を分析し、マーケティング施策に応用する研究はさかに行われている。たとえば、ID付きクレジット購買履歴データを用いた顧客の特徴によるセグメント化に関する研究 [11]、クレジット決済による購買履歴データと店舗チェックイン履歴情報を用いた優良ユーザ推定手法 [12]、顧客の購買パターンを活用した顧客生涯価値予測 [13] など様々な観点から研究が行われている。

購買行動分析の事例研究においては、潜在クラスモデルが用いられた分析が多く報告されている [14]。潜在クラスモデル [2], [15], [16] とは、観測データの背後に潜在変数が存在することを仮定し、データの潜在的な構造を確率的にモデル化する手法である。購買履歴データに潜在クラスモデルを適用することで、確率的なクラスタリングによりセグメントを抽出することが可能であり、購買行動分析アンケートデータと ID-POS データを用いた顧客と商品の同時カテゴリ生成とそのカテゴリを活用した顧客行動の計算モデル化による顧客行動予測システムに関する研究 [7] や、潜在クラスモデルによりユーザを分類し、その所属確率を考慮した購買予兆発見モデル [8] などが提案されている。最近では、購買行動の背後に単純に潜在クラスを仮定して EM アルゴリズムなどの最尤推定を目指した方法だ

けでなく、過学習を抑え、生成モデルを構築する目的で階層ベイズモデルをベースとした様々な拡張がなされ、変分ベイズ法やMCMC法などの手法を用いた学習も一般的となっている [3]。潜在クラスモデルは、購買履歴データの分析 [4], [5], [6] 以外にも画像分類 [17], 推薦システム [18] などに幅広く利用されている。

また、クレジットユーザを対象とした研究として、Shimizuら [19] においても潜在クラスモデルが用いられている。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を提案している。この手法を用いた分析結果からクレジット利用の促進、顧客満足度向上に向けた施策案について検討している。

2.2 クラスタリング手法と転移学習

本研究では、クレジットユーザに対するクラスタリング結果として得られる潜在クラスモデルの推定確率を固定し、それにフィッティングさせる形でポイントユーザをクラスタリングする手法を、特徴転移型クラスタリングとして提案する。

ここで、転移学習 [9], [10] とは、ある問題に対して、別の関連した問題のデータや学習結果を再利用することである。転移学習において、転移する知識の送り側（元ドメイン）と転移する知識の受けて側（目標ドメイン）の双方で教師なし学習である場合は、教師なし転移学習（unsupervised transfer learning）と呼ばれるが [9], [10], 本研究で提案する特徴転移型クラスタリングはこの領域に属する手法の1つである。転移学習の手法としては、圧倒的に教師あり学習に関するモデルや手法が多いものの、教師なし転移学習のモデルについてもいくつかの研究がなされている [10]。たとえば、Wangら [40] は転移学習の枠組みで、転移学習型次元削減手法を提案している。Yangら [20] は、heterogeneous transfer learning という枠組みを提案し、ソーシャル Web から取得した補助的な異種データから知識を転送することで画像クラスタリングの結果が改善できることを示している。また近年では、深層学習モデルを用いて、LIGO データや画像データのクラスタリングを行う転移学習型のモデルがいくつか提案されている [21], [22], [23], [24]。

一方、あるモデルで学習されたデータの結果の一部を他のモデルの初期値として転移し、クラスタリングを行う手法としては、これまでもいくつかの提案がなされている。転移学習を用いた潜在クラスモデルに関する研究としては、文書分類問題において、ソースドメインとターゲットドメインの共通トピックの活用 [25], ドメイン間の差異、クラスの共通性を活用したアプローチ [26] が提案されている。このとき、学習したパラメータを初期値として与えるという観点から特徴転移型クラスタリングも転移学習の

一部であると考えられる。パラメータの初期値を定めたクラスタリング手法として文書クラスタリングにおいて、代表元を定めその補正值を初期値とする文書クラスタリング手法が提案されている [27], [28]。また、活用したいデータを Nonnegative Matrix Factorization [29] により学習し、得られた局所表現を適用先のデータをクラスタリングする際に初期値として転移させる手法 [30] も示されている。購買履歴データへの適用事例として清水ら [31] はアンケートデータと購買履歴データを統合的に分析する際にアンケートデータから得られた因子得点を Gaussian Mixture Models [32] に適用し、得られたパラメータを購買履歴データとアンケートデータの両方を用いた潜在クラスモデルの学習時に初期値として転移させる方法を示している。これらの手法では、元ドメインで学習した知識を初期値とし、目標ドメインの学習データを用いて再学習をするという方法となっている。

以上の従来手法では、分析対象とする目標ドメインに対して、十分な学習データや情報が得られていないというセッティングで、元ドメインで学習した知識を援用しようとするものがほとんどである。これに対し、本研究で扱う対象は、ドメインが異なるユーザ（ポイントユーザとクレジットユーザ）の購買履歴データにおいて、データ数では目標ドメインであるポイントユーザの方が、元ドメインとなるクレジットユーザよりも圧倒的に多い。しかし、具体的な施策立案のためにクレジットカードへのステップアップ後の利用傾向が重要であるという観点から、ポイントユーザの利用傾向についてクレジットユーザのクラスと関連付けてクラスタリングを行う手法を提案している。その際、クレジットユーザの利用傾向が重要であるため、ここで得られた学習結果は固定し、新たにポイントユーザの購買履歴データから得られる知識を付加的に学習しようとしている点で、従来手法と差異がある。

3. 準備

3.1 事前分析

本研究で対象とするのは、小田急電鉄株式会社が発行するポイントカード（以下、OPカード）[33]である。OPカードには、クレジット機能付きポイントカードであるOPクレジットとポイント専用カードであるOPポイント専用カードが存在する。一般ユーザであるポイントユーザに対して上位ステージであるクレジットユーザへステップアップしてもらうための施策を考えるため、クレジットユーザ、ポイントユーザの利用傾向の差異を把握する必要がある。そこで、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について分析を行う。対象ユーザの中で2018年12月31日時点でOPクレジットに切替えを行ったユーザとOPポイント専

表 1 各カード所持ユーザ数
Table 1 Number of users for each card.

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

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分かる。そのほか、対象事例の詳細については付録 A.1 に示す。

3.2 問題設定

本研究で、OP カードの優良ユーザは、図 1 に示すとおり、クレジット機能、ポイント機能双方において利用が多いユーザととらえられる。戦略的に優良ユーザを増やすことを考えると、いくつかの顧客成長モデルを想定することが有効である。図 1 に示すように、優良ユーザへの成長のルートとしては、「グループ加盟店利用頻度が向上後、クレジット利用頻度が向上する」ルートと「クレジット利用頻度の向上後、グループ加盟店利用頻度が向上する」というルートが考えられる。このもとで、Shimizu ら [19] においては、クレジットユーザの現在のカード利用方法や所持目的により、ユーザをポジショニングし、現状のポジションに合わせてクレジット機能、ポイント機能の双方における利用の促進を図る施策の検討に向けた分析を行っている。

これに対して、ポイントユーザの成長に向けては異なる 2 方向から顧客成長のためのマーケティング施策が検討できる。ポイントユーザのままグループ加盟店での利用頻度を向上させる施策として、新規利用店舗の割引クーポンの発行、利用していない加盟店サービスの情報提供、ポイントユーザからクレジットユーザへの切替え促進施策としてはクレジットカードへの切替えキャンペーン、クレジット入会案内のダイレクトメールなどが考えられる。ポイントユーザであるユーザの利用頻度や購買額を増やすことを目的とした分析や施策立案は、一般的な販売促進施策の考え方であり、多くの取り組みがなされているが、本研究においては、多様な施策方法を増やすため、「ポイントユーザからのクレジットユーザへのステップアップ」という観点から現状ではクレジット機能をまった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 つとなることが期待できる。また、OP クレジットを利用すれば、OP ポイントが蓄積され、その 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。加盟店での利

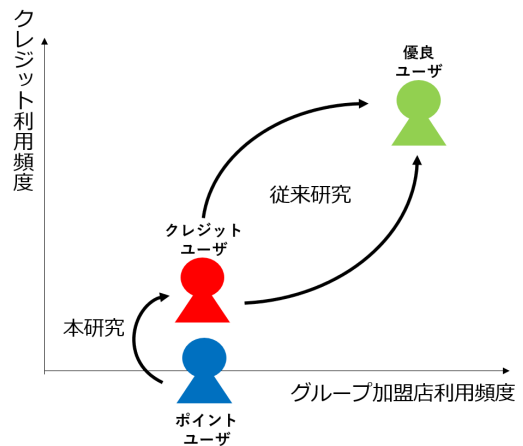


図 1 優良ユーザの位置づけ
Fig. 1 Positioning of good customers.

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 クレジットへの切替え施策の検討を対象とする。

4. 提案手法

4.1 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替えを促す施策を考える際、これらの施策の多くは切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、効果的な施策を検討する必要がある。つまり、ポイントユーザのような一般ユーザ、クレジットユーザのような上位ステージユーザが存在するユーザ群に対して、一般ユーザから上位ステージユーザへの切替え促進の効果的な施策が必要となる。このとき、必要となるのは、一般ユーザと類似する利用傾向を持つ上位ユーザ群の特定である。これにより、より効果的な施策が検討可能となる。

ここで、一般的に上位ステージユーザはユーザ全体の中で少数であることが多い。これは、表 1 に示すように OP カードにおいても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。そのため、ユーザ全体をクラスタリングすると大多数である一般ユーザの特徴によってクラスタリングがなされてしまい、上位ステージユーザの特徴が反映されない。そこで、上位ステージのクラスを事前に定め、そのクラスに基づいて、一般ユーザをクラスタリングする 2 段階の学習を行うことを考える。その際、上位ステージユーザとは異なる傾向を有した一般ユーザも存在すると考えられる。そのため、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。

また、本研究で対象とする OP カードの特性として利用

店舗の業種や地域の多様性がある。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、加盟店全体における利用金額や利用頻度といった利用状況に加えて、利用店舗の各属性における利用頻度を考慮したモデル化が必要である。このとき、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、利用店舗は購買単価や利用頻度などの利用状況に強く影響を与えることが考えられるため、単一のモデル化を行うことが望ましい。

4.2 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、ユーザの加盟店内をどの程度利用しているかを表す利用状況ベクトル、ユーザの利用店舗においてどの地域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行動を表現する。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。このとき、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、表 A.1、表 A.2 より、ユーザの各業種の利用頻度、各地域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 $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$ 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。

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

ユーザの総数を N 人としたとき、 n 番目のユーザの利用状況ベクトル、地域利用頻度ベクトル、業種利用頻度ベクトルをそれぞれ $\mathbf{x}_n, \mathbf{s}_n, \mathbf{t}_n$ 、所属する潜在クラスを $\mathbf{w}_n \in \mathcal{Z}$ と記述する。また、 $\mathbf{z}_n = (z_{n1}, \dots, z_{nK})$ を $\mathbf{w}_n = z_k$ に対応する k に対してのみ $z_{nk} = 1$ 、それ以外の k に対しては $z_{nk} = 0$ という要素からなる 1-hot ベクトルとする ($z_{nk} \in \{0, 1\}$)。このとき、すべての確率変数の同時分布を書き下すと式 (2) のようになる。ただし、潜在クラス比率を $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)$ 、 k 番目の潜在クラスにおける、多次元正規分布の平均ベクトル、精度行列は $\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k$ 、各地域の出現確率を $\boldsymbol{\theta}_k = (\theta_{k1}, \dots, \theta_{kI})$ 、各業種の出現確率を

$\boldsymbol{\phi}_k = (\phi_{k1}, \dots, \phi_{kJ})$ と表す。

$$\begin{aligned} & p(\mathbf{x}_n, \mathbf{s}_n, \mathbf{t}_n, \mathbf{w}_n, \boldsymbol{\pi}_k, \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k, \boldsymbol{\theta}_k, \boldsymbol{\phi}_k) \\ &= p(\mathbf{x}_n|\mathbf{w}_n, \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k)p(\mathbf{s}_n|\mathbf{w}_n, \boldsymbol{\theta}_k)p(\mathbf{t}_n|\mathbf{w}_n, \boldsymbol{\phi}_k) \\ & \quad \times p(\mathbf{w}_n|\boldsymbol{\pi}_k)p(\boldsymbol{\pi}_k)p(\boldsymbol{\mu}_k|\boldsymbol{\Lambda}_k)p(\boldsymbol{\Lambda}_k)p(\boldsymbol{\theta}_k)p(\boldsymbol{\phi}_k). \quad (2) \end{aligned}$$

ここで、潜在変数とその他の確率変数に関する分解を仮定すると潜在変数とパラメータに分解した変分近似は、式 (3) のように表される。

$$\begin{aligned} & q(\mathbf{w}_n, \boldsymbol{\pi}_k, \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k, \boldsymbol{\theta}_k, \boldsymbol{\phi}_k) \\ &= q(\mathbf{w}_n|\boldsymbol{\pi}_k)q(\boldsymbol{\pi}_k)q(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k)q(\boldsymbol{\theta}_k)q(\boldsymbol{\phi}_k). \quad (3) \end{aligned}$$

潜在クラスモデルの多くはパラメータ推定に EM アルゴリズム [34], [35] が用いられている。EM アルゴリズムを用いた場合、学習データに対して過学習する可能性が考えられる。本研究で提案した手法では、あるユーザ群のデータを用いて推定したパラメータを未知のユーザ群をクラスタリングするために用いる。そのため、あるデータのみ fitting することは避ける必要があることから、式 (2) のパラメータ推定には変分推論 [36], [37] を適用する。以下の変分 E-step (式 (4)), 変分 M-Step (式 (5)–式 (15)) を繰り返すことで更新される。ただし、 $\text{Dir}(\cdot)$ はディリクレ分布を表し、 $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$ 、 $\boldsymbol{\gamma}_k, \boldsymbol{\delta}_k$ はそれぞれディリクレ分布のパラメータであり、 $\alpha_0, \gamma_0, \delta_0$ はそれぞれハイパーパラメータである。 $\mathcal{N}(\cdot)$ 、 $\mathcal{W}(\cdot)$ はそれぞれガウス分布、ウィシャート分布を表し、 $\mathbf{m}_k, \boldsymbol{\beta}_k$ はガウス分布のパラメータ、 $\boldsymbol{\nu}_k, \mathbf{W}_k$ はウィシャート分布のパラメータ、 $\mathbf{m}_0, \boldsymbol{\beta}_0, \boldsymbol{\nu}_0, \mathbf{W}_0$ はハイパーパラメータである。このとき、グラフィカルモデルは図 2 のようになる。ただし、 $\boldsymbol{\mu} = \{\boldsymbol{\mu}_k\}$ 、 $\boldsymbol{\Lambda} = \{\boldsymbol{\Lambda}_k\}$ 、 $\boldsymbol{\theta} = \{\boldsymbol{\theta}_k\}$ 、 $\boldsymbol{\phi} = \{\boldsymbol{\phi}_k\}$ である。

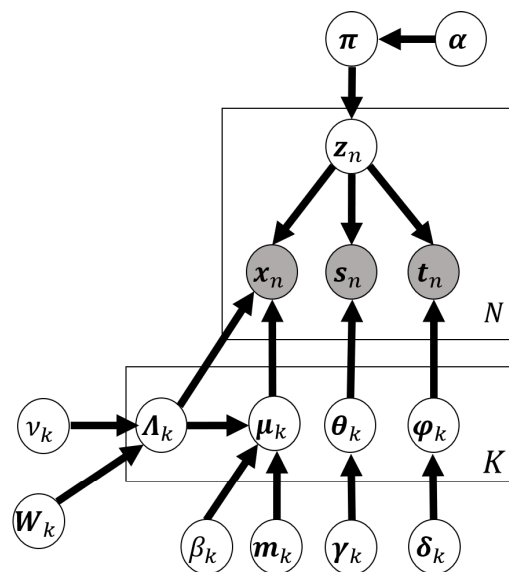


図 2 グラフィカルモデル
Fig. 2 Graphical model.

[変分 E-step]

$$q(z_{nk}) = r_{nk} z_{nk} = \left(\frac{\rho_{nk}}{\sum_{k=1}^K \rho_{nk}} \right)^{z_{nk}}. \quad (4)$$

[変分 M-step]

$$q(\pi_k) = \text{Dir}(\pi_k | \alpha_k). \quad (5)$$

$$\alpha_k = \alpha_0 + \sum_{n=1}^N E[z_{nk}]. \quad (6)$$

$$q(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k) = \mathcal{N}(\boldsymbol{\mu}_k | \mathbf{m}_k, \beta_k \boldsymbol{\Lambda}_k) \mathcal{W}(\boldsymbol{\Lambda}_k | \mathbf{W}_k, \nu_k). \quad (7)$$

$$\beta_k = \beta_0 + \sum_{n=1}^N z_{nk}. \quad (8)$$

$$\mathbf{m}_k = \frac{1}{\beta_k} + \beta_0 \mathbf{m}_0 + \sum_{n=1}^N r_{nk} \mathbf{x}_n. \quad (9)$$

$$\begin{aligned} \mathbf{W}_k^{-1} &= \mathbf{W}_0^{-1} + \sum_{n=1}^N r_{nk} (\mathbf{x}_n - \bar{\mathbf{x}}_k)(\mathbf{x}_n - \bar{\mathbf{x}}_k)^T \\ &\quad + \frac{\beta_0 \sum_{n=1}^N z_{nk}}{\beta_0 + \sum_{n=1}^N z_{nk}} (\bar{\mathbf{x}}_k - \mathbf{m}_0)(\bar{\mathbf{x}}_k - \mathbf{m}_0)^T. \end{aligned} \quad (10)$$

$$\nu_k = \nu_0 + \sum_{n=1}^N z_{nk}. \quad (11)$$

$$q(\boldsymbol{\theta}_k) = \text{Dir}(\boldsymbol{\theta}_k | \boldsymbol{\gamma}_k). \quad (12)$$

$$\boldsymbol{\gamma}_k = \boldsymbol{\gamma}_0 + \sum_{n=1}^N r_{nk} \mathbf{s}_n. \quad (13)$$

$$q(\boldsymbol{\phi}_k) = \text{Dir}(\boldsymbol{\phi}_k | \boldsymbol{\delta}_k). \quad (14)$$

$$\boldsymbol{\delta}_k = \boldsymbol{\delta}_0 + \sum_{n=1}^N r_{nk} \mathbf{t}_n. \quad (15)$$

ただし, $\ln \rho_{nk}$, $\bar{\mathbf{x}}_k$ は式 (16), (17) のように定義する.

$$\begin{aligned} \ln \rho_{nk} &= E[\ln \pi_k] + \frac{1}{2} E[\ln |\boldsymbol{\Lambda}_k|] - \frac{D}{2} \ln 2\pi \\ &\quad - \frac{1}{2} E_{\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k} [(\mathbf{x}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Lambda}_k (\mathbf{x}_n - \boldsymbol{\mu}_k)] \end{aligned} \quad (16)$$

$$+ \sum_{i=1}^I s_{ni} E[\ln \theta_{ki}] + \sum_{j=1}^J t_{nj} E[\ln \phi_{kj}].$$

$$\bar{\mathbf{x}}_k = \frac{1}{r_{nk}} \sum_{n=1}^N r_{nk} \mathbf{x}_n. \quad (17)$$

4.3 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.

各ユーザは潜在クラス z 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 $\mathcal{Z}^c = \{z_k^c : 1 \leq k \leq K\}$, ポイントユーザにおける追加されるユーザの潜在クラス集合を $\mathcal{Z}^p = \{z_l^p : 1 \leq l \leq L\}$ と定義する. この

とき, 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 (18) のように定義する.

$$p(\mathbf{x}, \mathbf{s}, \mathbf{t}) = \sum_{z \in \mathcal{Z}^c} p(\mathbf{x}|z) p(\mathbf{s}|z) p(\mathbf{t}|z) p(z) \quad (18)$$

ただし, $\sum_{z \in \mathcal{Z}^c} p(z) = 1$ である.

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 (19) を仮定する.

$$\begin{aligned} \tilde{p}(\mathbf{x}, \mathbf{s}, \mathbf{t}) &= \sum_{z \in \mathcal{Z}^c} p(\mathbf{x}|z) p(\mathbf{s}|z) p(\mathbf{t}|z) \tilde{p}(z) \\ &\quad + \sum_{z \in \mathcal{Z}^p} \tilde{p}(\mathbf{x}|z) \tilde{p}(\mathbf{s}|z) \tilde{p}(\mathbf{t}|z) \tilde{p}(z) \end{aligned} \quad (19)$$

ただし, $\sum_{z \in \mathcal{Z}^c \cup \mathcal{Z}^p} \tilde{p}(z) = 1$ である.

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たモデルとなっている.

4.4 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 (1) 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 (1) の潜在クラスモデルを適用する際の学習に転移させる. 具体的には, 第 1 段階のクレジットユーザから学習した潜在クラスモデルの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. 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, 潜在クラスモデルのパラメータ推定において EM アルゴリズムなどの最尤推定法による推定手法を用いた場合, 学習データに対する尤度関数を極大するため, 過学習が起りやすい. 本モデルでは推定したパラメータを新たな学習を行う際に転移させることから学習データに対する過学習を防ぐ必要がある. そこで, パラメータに事前分布を仮定することで, 事前分布がある種の正則化のような役割を演じるため, 学習データにフィッティングしすぎることを防ぐ効果が知られる変分推論法を適用する. 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対して潜在クラスモデルを適用し, 式 (18) 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. このとき, Step1 で推定したパラメータ $p(\mathbf{x}|z)$, $p(\mathbf{s}|z)$, $p(\mathbf{t}|z)$ を転移させ固定して, 式 (19) のモデルのパラメータ $\tilde{p}(\mathbf{x}|z)$, $\tilde{p}(\mathbf{s}|z)$, $\tilde{p}(\mathbf{t}|z)$, $\tilde{p}(z)$ のみを推定する.

表 3 各クラスの所属割合と利用状況の平均値
Table 3 Affiliation ratio and Average usage in each class.

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
$\bar{p}(z)$	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	16.6	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）
Table 4 Representative area of use and industry of use in each class.

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	...

表 2 各変数における用語の定義
Table 2 Definition of terms in each variable.

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
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
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

5. 実データ分析

5.1 分析条件

本分析の対象データは、小田急電鉄株式会社に蓄積された 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。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とり^{*1}、それぞれ $K = 15$, $L = 15$ と設定した ($|Z^c \cup Z^p| = 30$)。また、ハイパーパラメータは経験的に決定し、 $\alpha_0 = 0.1$, $\gamma_0 = 0.1$, $\delta_0 = 0.1$, $\beta_0 = 1.0$, $\nu_0 = 6.0$, W_0 は 6×6 の

*1 変分推論では潜在クラスを多めにとること、不必要な潜在クラスの推定確率が 0 となり、潜在クラスが自動的に削減される。多めにとるとは、この減少が生じる数まで大きく設定することを意味する。

単位行列である。ただし、 m_0 については各要素が平均 0、分散 1 (標準偏差 1) の正規分布 (標準正規分布) に従う乱数である 6 次元のベクトルである。イテレーション数については、200 と設定した。

5.2 分析結果

[Step1], [Step2] の推定においてそれぞれ潜在クラスの推定確率 π_k , π_l 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の推定確率が 0 となった。これにより、事前に設定したクラス数 K , L は、 $K' = 13$, $L' = 7$ 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、必要最小限な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分かる。表 4 から z_9^c に所属する世田谷・狛江において保育、 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、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うかがえる。ゆえに、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になったと考えられる。

次に、潜在クラス z_1^p に着目する。表 4 から z_1^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している。このことから z_1^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_1^p は z_9^c と比較すると利用頻度や付与された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分かる。潜在クラス z_1^p に所属するユーザは、加盟店の中では限定的な業種の利用の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と考えられる。以上のことから、施策優先度の低いユーザ群であると解釈できる。このようにポイントユーザに対してクレジットカードユーザのクラスを考慮してクラスタリングを行ったことで、施策対象群を特定することが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。たとえば、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、た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 から利用地域、利用業種については新宿やオンラインにおいて百貨店を利用している。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を用いて決済を行うことが考えられる。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を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。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なくなることを強調することで、切替えへのハードルが下がり、施策効果は高まると考えられる。

6. 考察

6.1 モデル化について

本研究においては、利用状況、利用店舗の業種、地域について1つの潜在クラスモデルを用いてモデル化を行った。Shimizu ら [19] のように利用状況についてのモデルと利用店舗の業種、地域についてのモデルを構築し、それぞれの結果に対してクロス分析を行うということも考えられる。しかし、ユーザの利用状況と利用店舗との間には関連性があり、利用状況、利用店舗の属性についてそれぞれ個別にモデル化を行った場合、結果に広がり生まれ、クラスの解釈が困難になる。実際それぞれ個別にモデル化した場合のクラス数と提案モデルで推定されたクラス数（推定確率が0ではない正の確率を持つ潜在クラスの数）は表 5 に示すような結果になる。このとき、利用状況分析モデルには Gaussian Mixture Models [32] を用いており、モデル式は式 (20), (21) のようになる。また、利用店舗分析モデルには Probabilistic Latent Semantic Analysis [38] を用いており、確率モデルは式 (22), (23) のようになる。

$$p(\mathbf{x}) = \sum_{z \in Z^c} p(\mathbf{x}|z)p(z) \tag{20}$$

表 5 利用状況と地域、業種の利用頻度を個別にモデル化したときのクラス数

Table 5 Number of classes in individual models for usage volume and store attributes.

モデル	利用状況分析モデル	利用店舗分析モデル	提案モデル
クラス数	70	55	20

$$\tilde{p}(\mathbf{x}) = \sum_{z \in Z^c} p(\mathbf{x}|z)\tilde{p}(z) + \sum_{z \in Z^p} \tilde{p}(\mathbf{x}|z)\tilde{p}(z) \tag{21}$$

$$p(\mathbf{s}, \mathbf{t}) = \sum_{z \in Z^c} p(\mathbf{s}|z)p(\mathbf{t}|z)p(z) \tag{22}$$

$$\tilde{p}(\mathbf{s}, \mathbf{t}) = \sum_{z \in Z^c} p(\mathbf{s}|z)p(\mathbf{t}|z)\tilde{p}(z) + \sum_{z \in Z^p} \tilde{p}(\mathbf{t}|z)\tilde{p}(\mathbf{s}|z)\tilde{p}(z) \tag{23}$$

また、事前に設定する潜在クラス数は、提案手法と同様に変分推論を適用することから多めにとり、利用状況分析モデルでは、 $K = L = 50$ 、利用店舗分析モデルでは、 $K = L = 30$ とした。

表 5 から分かるとおり、利用状況分析モデル、利用店舗分析モデルともに提案モデルに比べてクラス数が大きくなっており、結果に広がり生まれつつあることが分かる。これらの結果を用いてクロス分析を行った場合、本研究で対象とする切替え施策の検討においては、2つのモデルの結果から関係性の強いクラスの組み合わせを見つけ出す必要があり、不向きであると考えられる。以上より、提案モデルを用いることで、ユーザの購買行動をうまくとらえ、施策の検討を行ううえで効果的なモデル化ができていると考えられ、本研究の提案モデルは有用であるといえる*2。

6.2 特徴転移型クラスタリングの有効性

4.1 節において述べた、クレジットカードユーザとポイントユーザを同時にクラスタリングすることとの差異について考察する。実際に式 (1) の潜在クラスモデルに対してクレジットカードユーザとポイントユーザのデータを同時に適用したときと提案手法の各クラスの所属ユーザ数におけるポイントユーザの割合を図 3 に示す。図 3 を見るとほとんどのクラスにおいて、所属ユーザの大部分がポイントユーザで構成されるクラスになっていることが分かる。すなわち、各クラスの特徴はポイントユーザの影響を強く受けており、クレジットカードユーザの特徴を反映しきれていない。また、1つのクラスにクレジットカードユーザが固まってしまう。ゆえに、成長先であるクレジットカードユーザの特徴を明確にす

*2 一般的にモデルの評価に尤度や AIC が用いられる。しかし、本研究で提案した転移型のクラスタリング手法の場合、尤度や AIC による公平な比較を行うことを考えた場合に、単純な比較が難しい。加えて、過学習を避けるために導入した変分ベイズ推定は、尤度関数を極大化する最尤推定を行っていない。そのため、尤度原理によって推定をしていない統計モデルに対して、対数尤度や AIC を用いてデータへのあてはまりを評価することでは、モデルの良し悪しを測る評価関数と観点が異なるため、推定されたクラス数による比較を行った。

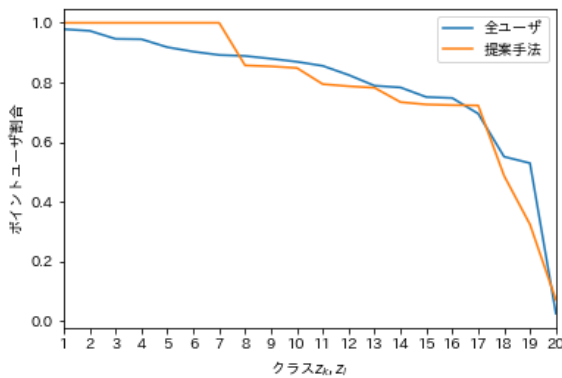


図 3 各クラスにおけるポイントユーザの割合

Fig. 3 The clustering results for point card user by all users.

るという本研究の目的が達成できていない。しかし、提案モデルでは、クレジットカードユーザのデータのみで事前に学習を行ったパラメータを用いている。このため、ポイントユーザのみのクラスが明確に分かれているうえ、クレジットカードユーザの特徴を明確にとらえられている。

一方で、クレジットカードユーザのクラスの特徴のみを用いて、その特徴にあるユーザが当てはまるかを判断するリジエクトルを用いたポイントユーザの分類方法も考えられる。しかし、リジエクトルによる分類では、ルール基準の定め方によって本研究において Z^p のいずれかに所属するようなポイントユーザが Z^c のいずれかのクラスに所属してしまう可能性がある。このように、リジエクトルでは、そのユーザの所属クラスを正しく判断することは困難である。ゆえに、ユーザの利用傾向に対して適切な施策を検討するには、提案モデルが有用であると考えられる。

6.3 特徴転移型クラスタリングの安定性

本研究では、特徴転移型クラスタリングを提案した。この際、クレジットカードユーザのみで学習したパラメータを転移させ、ポイントユーザのクラスタリングを行う。そのため、クレジットカードユーザでの学習結果により、ポイントユーザのクラスタリング結果に差異が生じる可能性がある。

そこで、クレジットカードユーザから半数をランダムサンプリングし、提案手法に適用する。このランダムサンプリングを繰り返し、施策対象群であるポイントユーザのクラスタリング結果をランド指数 [39] 用いて比較する。ランド指数は、クラスタリング結果の類似度を示す指標であるため、この値によりクラスタリングの安定性の評価が可能となる。

比較した結果を表 6 に示す。この結果より、分散が小さく平均が大きいことから安定したクラスタリングが可能であると考えられる。

また、ポイントユーザのみの特徴を含むクラス数はすべての実験でクレジットカードユーザ全体を用いた結果と同様に 7 となった。この結果から、ポイントユーザの特徴について、安定した分析が可能となると考えられる。

表 6 ランド指数

Table 6 Rand Index.

平均	0.809
分散	0.00193

一方で、[Step1] クレジットユーザのクラス数は、クレジットカード全体を用いた場合よりも減少するケースもみられた。これは、データ数が半数となっているためと考えられる。この結果から、たとえばクレジットカードユーザから事前に、企業側にとってより望ましい優良ユーザなどを抽出することによって、ポイントユーザの施策対象を絞り込むことも可能であると考えられる。

7. 分析結果からの施策案検討について

提案モデルにより、クレジットカードユーザのクラスの特徴を考慮したクラスタリングを行うことが可能になった。これにより、本研究で目的としていたポイントユーザのステップアップする方向性を定めたクレジットカード機能付きポイントカードへの切替え施策の検討が可能になった。

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カードユーザへの切替え施策は、施策への反応度が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、なるべく $\tilde{p}(z)$ の大きいクラスに対して施策を講じる方が効果的であると考えられる。また、 $p(z)$ が小さいクラスはクレジットカードユーザとしては稀なユーザ群であり、OP クレジットへ切り替えにくいユーザ群であるとも考えられる。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にくい。しかし、本分析における z_{11}^c や z_7^c 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、優良ユーザにな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。

8. 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、クレジットカード機能付きポイントカードへの切替えを促す施策の検討を目的とし、特徴転移型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カード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、クラスの特徴から施策案について考察を行った。

本研究では、ユーザの属性情報を考慮すること、地域の分け方を検討することで施策立案に有用な示唆を得られる可能性がある。また、分析結果から導かれる施策を講じることによる実応用上での定量的な評価が望ましい。以上については、今後の課題とする。

参考文献

- [1] 佐々木土師二：変革するマーケティング，繊維製品消費科学，Vol.30，No.2，pp.46–54 (1989).
- [2] Hagenaaers, J.A. and McCutcheon, A.L.: *Applied Latent Class Analysis*, Cambridge University Press (2002).
- [3] Blei, D.M., Ng, A.Y. and Jordan, M.I.: Latent dirichlet allocation, *Journal of Machine Learning Research*, Vol.3, No.1, pp.993–1022 (2003).
- [4] 阿部 誠：消費者行動のモデル化：消費者の異質性，オペレーションズ・リサーチ，Vol.48，No.2，pp.121–129 (2003).
- [5] 佐藤 圭：マーケティング研究におけるトピックモデルの適用に関する一考察，経営研究，Vol.68，No.3，pp.125–148 (2017).
- [6] 里村卓也：トピックモデルによる顧客データの統合的分析，オペレーションズ・リサーチ：経営の科学，Vol.63，No.2，pp.67–74 (2018).
- [7] 石垣 司，竹中 毅，本村陽一：日常購買行動に関する大規模データの融合による顧客行動予測システム，人工知能学会論文誌，Vol.26，No.6，pp.670–681 (2011).
- [8] 久松俊道，外川隆司，朝日弓未，生田目崇：ECサイトにおける購買予兆発見モデルの提案，オペレーションズ・リサーチ：経営の科学，Vol.58，No.2，pp.93–100 (2013).
- [9] Pan, S.J. and Yang, Q.: A Survey on Transfer Learning, *IEEE Trans. Knowledge and Data Engineering*, Vol.22, No.10, pp.1345–1359 (2009).
- [10] 神嶋敏弘：転移学習，人工知能，Vol.25，No.4，pp.572–580 (2010).
- [11] 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。
- [12] 土井千章，石井 暁，荒木尊士，稲村 浩，太田 賢，重野寛，片桐雅二：店舗チェックイン履歴情報を用いた優良顧客推定手法と来店促進への応用，情報処理学会論文誌 コンシューマ・デバイス&システム (CDS)，Vol.7，No.2，pp.115–124 (2017).
- [13] 蓮本恭輔，雲居玄道，後藤正幸：非負値行列因子分解を用いたプラットフォームビジネスにおける顧客生涯価値予測，情報処理学会論文誌，Vol.60，No.7，pp.1283–1293 (2019).
- [14] Zhang, Q., Yamashita, H., Mikawa, K. and Goto, M.: Analysis of Purchase History Data Based on a New Latent Class Model for RFM Analysis, *Industrial Engineering & Management Systems*, Vol.19, No.2, pp.476–483 (2020).
- [15] 後藤正幸，小林 学：入門パターン認識と機械学習，コロナ社 (2014).
- [16] 平井有三：はじめてのパターン認識，森北出版株式会社 (2012).
- [17] Bosch, A., Zisserman, A. and Muñoz, X.: Scene classification via pLSA, *Proc. European Conference on Computer Vision*, pp.517–530, Springer (2006).
- [18] Wasilewski, J. and Hurley, N.: Intent-aware diversification using a constrained PLSA, *Proc. 10th ACM Conference on Recommender Systems*, pp.39–42 (2016).
- [19] Shimizu, R., Yamashita, H., Ueda, M., Tanaka, R., Tachibana, T., Goto, M., 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, pp.106–117 (2020).
- [20] Yang, Q., Chen, Y., Xue, G.-R., Dai, W. and Yu, Y.: Heterogeneous Transfer Learning for Image Clustering via the Socialweb, *Proc. Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pp.1–9 (2009).
- [21] 中山英樹：深層畳み込みニューラルネットワークによる画像特徴抽出と転移学習，電子情報通信学会技術研究報告，SP2015-45，Vol.115，No.146，pp.55–59 (2015).
- [22] George, D., Shen, H. and Huerta, E.: Classification and Unsupervised Clustering of LIGO Data with Deep Transfer Learning, *Physical Review D*, Vol.97, No.10, p.101501 (2018).
- [23] Han, K., Vedaldi, A. and Zisserman, A.: Learning to Discover Novel Visual Categories via Deep Transfer Clustering, *Proc. IEEE International Conference on Computer Vision*, pp.8401–8409 (2019).
- [24] 高橋伸弥，李 玉潔，鶴田直之，藍 浩之：深層自己符号化器と転移学習による画像中からのミツバチ個体検出手法の検討，人工知能学会全国大会論文集 第 34 回全国大会 (2020)，Vol.204GS1305，人工知能学会 (2020).
- [25] Xue, G.-R., Dai, W., Yang, Q. and Yu, Y.: Topic-bridged PLSA for cross-domain text classification, *Proc. 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.627–634 (2008).
- [26] Krithara, A. and Paliouras, G.: TL-PLSA: Transfer learning between domains with different classes, *Proc. IEEE 13th International Conference on Data Mining*, pp.419–427, IEEE (2013).
- [27] 雲居玄道，石田 崇，後藤正幸，平澤茂一：PLSIを用いた文書分類手法に関する一考察，電子情報通信学会技術研究報告，AI2010-33，Vol.110，No.301，pp.13–18 (2010).
- [28] 平澤茂一，石田 崇，足立鉦史，後藤正幸，酒井哲也：文書分類技法とそのアンケート分析への応用，経営情報学会 全国研究発表大会要旨集 2005 年度春季全国研究発表大会，p.60，一般社団法人 経営情報学会 (2005).
- [29] Lee, D.D. and Seung, H.S.: Learning the parts of objects by non-negative matrix factorization, *Nature*, Vol.401, No.6755, pp.788–791 (1999).
- [30] 荻野広樹，吉田哲也：トピックグラフに基づく NMF を用いた転移学習，情報処理学会論文誌数理モデル化と応用 (TOM)，Vol.4，No.3，pp.73–83 (2011).
- [31] 清水良太郎，坂元哲平，山下 遥，後藤正幸：アンケートデータを考慮した EC サイトの購買履歴分析モデルの提案，経営システム，Vol.27，No.2，pp.70–76 (2017).
- [32] Reynolds, D.A.: Gaussian Mixture Models., *Encyclopedia of Biometrics*, Vol.741 (2009).
- [33] 小田急電鉄：小田急ポイントカード，小田急電鉄 (オンライン)，入手先 (<https://www.odakyu-card.jp/>) (参照 2020-11-23).
- [34] Dempster, A.P., Laird, N.M. and Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol.39, No.1, pp.1–22 (1977).
- [35] McLachlan, G.J. and Krishnan, T.: *The EM Algorithm*

and Extensions, Vol.382, John Wiley & Sons (2007).

[36] Bishop, C.M.: *Pattern Recognition and Machine Learning*, Springer (2006).
 [37] 須山敦志, 杉山 将: 機械学習スタートアップシリーズ ベイズ推論による機械学習入門, 講談社 (2017).
 [38] Hofmann, T.: Probabilistic Latent Semantic Analysis, *arXiv preprint arXiv:1301.6705* (2013).
 [39] Rand, W.M.: Objective criteria for the evaluation of clustering methods, *Journal of the American Statistical Association*, Vol.66, No.336, pp.846–850 (1971).
 [40] Wang, Z., Song, Y. and Zhang, C.: Transferred dimensionality reduction, *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp.550–565 (2008).

付 録

A.1 対象事例の概要

本研究で対象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）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。

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、加盟店利用によるポイントが付与される。一方、クレジット機能付きポイントカードは、ポイント専用カードの機能に加えてクレジット機能を持つ。このクレジット機能は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ポイントが加算され、蓄積されたポイントは加盟店内での支払い手段の 1 つとして利用可能である。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クレジット機能の長期的な利用も期待できる。このとき、ポイント専用カードの所持ユーザ（以下、ポイントユーザ）を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）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が高く、ユーザとの長期的に良好な関係を構築できると考えられる。

本研究においては、対象となる店舗が所在する地域を 20 種、対象となる店舗の業種を 18 種とする。表 1 に示した 101,354 人のユーザの中で 2019 年 1 月 1 日～12 月 31 日の 1 年間で地域ごとに利用したユーザ数を表 A-1 に、業種ごとに利用したユーザ数を表 A-2 に示す。表 A-1, 表 A-2 より、OP カードの加盟店は、小田急線沿いを中心に様々な地域、業種にわたって存在していることが分かる。ユーザは様々な業種、地域を利用しており、ユーザが利用する

表 A-1 各地域における利用ユーザ数

Table A-1 Number of users at each area.

地域名	利用ユーザ数
新宿	50,142
町田・相模原	34,192
座間～愛甲石田	23,337
川崎	19,534
世田谷・狛江	16,649
東林間～片瀬江ノ島	11,197
伊勢原～足柄	7,596
小田原～箱根湯本	6,810
五月台～唐木田	3,430
渋谷	3,104
横浜	887
埼玉	33
東京東部	105
その他*1	1,927
オンライン*2	779
特別ポイント付与*2	19,534
特別ポイント利用*2	43
ポイント失効*2	1309

表 A-2 各業種における利用ユーザ数

Table A-2 Number of users for each industry.

業種名	利用ユーザ数
百貨店	61,508
ショッピングセンター	60,101
スーパーマーケット	49,232
レストラン・飲食	35,293
特別ポイント	20,455
鉄道	12,852
趣味・娯楽	9,908
花	5,974
タクシー	4,618
ホテル・旅館	1,431
旅行業	758
美容・エステ	97
ゴルフ場	63
公共料金・保険	42
通信販売	40
航空・その他運輸	33
保育	27
不動産	23
ハウスクリーニング	10
その他	465

店舗の属性によってユーザの特性は大きく異なると考えられる。

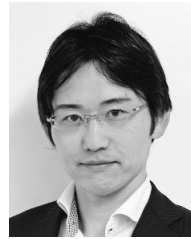
*1 公共料金などの地域を特定不能やゴルフ場などの静岡地域。

*2 特殊な履歴データのため、1 地域と見なす。



平野 洋介

1996年生。2019年神奈川大学工学部経営工学科卒業。2021年早稲田大学大学院創造理工学研究科経営デザイン専攻修士課程修了。機械学習を用いたデータ分析に関する研究に興味を持つ。



立花 徹也

2003年法政大学経済学部卒業。マーケティング会社を経て、現在、小田急エージェンシーに在籍。データサイエンティスト。食品、アパレル、流通、クレジットカード等幅広い業種のCRM、データ分析、リサーチの経験

を持つ。



楊 添翔 (正会員)

1988年に中国の温州市に生まれる。2015年早稲田大学大学院国際通信研究科修士課程修了。現在、同大学院創造理工学研究科経営システム専攻博士課程在籍。2019年より同大学創造理工学部経営システム工学科助手。機械学

習に基づく購買履歴データに関する研究に従事。



後藤 正幸 (正会員)

1994年武蔵工業大学大学院修士課程修了。2000年早稲田大学大学院理工学博士課程修了。博士(工学)。1997年早稲田大学理工学部助手。2000年東京大学大学院工学研究科助手。2002年武蔵工業大学環境情報学部助教授。

2008年早稲田大学創造理工学部経営システム工学科准教授。2011年同大学教授。情報数理応用とデータサイエンスの研究に従事。著書に、『入門パターン認識と機械学習』コロナ社(2014)、『ビジネス統計—統計基礎とエクセル分析』オデッセイコミュニケーションズ(2015)等。IEEE, INFORMS, 電子情報通信学会, 人工知能学会, 日本経営工学会等各会員。



雲居 玄道 (正会員)

2008年早稲田大学理工学部経営システム工学科卒業。2008年同大学理工学術院総合研究所嘱託研究員。2015年浄土真宗本願寺派総合研究所研究助手。2017年早稲田大学大学院創造理工学研究科博士後期課程。2019年よ

り同大学創造理工学部経営システム工学科助手。現在に至る。情報数理応用・テキストマイニングの研究に従事。IEEE, 経営情報学会, 日本気象学会各会員。



阿部 永

2014年小田急電鉄株式会社入社。小田原駅、小田急百貨店での研修を経て、カード戦略部(現カード推進部)に配属。お客さま相談担当、会員サービス担当、会員獲得担当、営業担当を経て、2019年より企画・戦略担当。主

に小田急ポイントカード会員の分析や会員向けのコンテンツ制作、新規サービス検討を行う。