

会員ステージランク制を対象とした顧客購買行動分析モデル
に関する研究

A Study on Customer Purchase Behavior Analysis Model
for Membership Stage Rank System

2024年2月

楊 添翔
Tianxiang YANG

会員ステージランク制を対象とした顧客購買行動分析モデル
に関する研究

A Study on Customer Purchase Behavior Analysis Model
for Membership Stage Rank System

2024年2月

早稲田大学大学院 創造理工学研究科
経営システム工学専攻 情報数理応用研究

楊 添翔
Tianxiang YANG

目次

第1章	序論	1
1.1	研究背景	1
1.2	研究対象	1
1.2.1	Frequent Shoppers Program	1
1.2.2	会員ステージランク制	2
1.2.3	会員ステージランク制の特徴	2
1.2.4	消費者側のメリット	3
1.2.5	企業側のメリット	4
1.3	研究目的	5
1.4	論文の構成	7
第2章	顧客購買行動分析と周辺研究	8
2.1	顧客購買行動に関する分析手法	8
2.1.1	顧客セグメンテーション分析	8
2.1.2	バスケット分析	10
2.1.3	顧客満足度分析	10
2.1.4	顧客生涯価値分析	11
2.2	会員ステージランク制を対象とした顧客購買行動に関する分析手法	11
2.3	本章の結論	13
第3章	確率的潜在意味解析の転移学習に基づく会員ステージを考慮した購入 行動分析モデル	14
3.1	本章の目的	14
3.2	Probabilistic Latent Semantic Analysis	14
3.3	EM アルゴリズム	15

3.4	転移学習	17
3.5	課題と本研究への展開	18
3.6	提案手法	19
3.6.1	概要	19
3.6.2	定式化	19
3.6.3	提案手法による顧客購買行動分析アルゴリズム	19
3.7	人工データを用いたシミュレーション実験による精度評価	21
3.7.1	グループの定義	22
3.7.2	人工データ生成手順	22
3.7.3	シミュレーション条件	24
3.7.4	評価指標	24
3.7.5	結果	25
3.8	実データ分析	26
3.8.1	分析対象データ	26
3.8.2	実験結果	27
3.9	考察	34
3.9.1	定量的な検証	34
3.9.2	定性的な解釈	35
3.9.3	潜在クラス数の決定	35
3.9.4	提案の有効性	36
3.10	本章の結論	37
第4章	会員ステージ間の潜在クラス分布を比較する購買行動分析モデル	38
4.1	本章の目的	38
4.2	課題と本研究への展開	39
4.3	提案手法	39
4.3.1	概要	39
4.3.2	提案手法の定式化	40
4.3.3	全ステージデータに対するユーザとアイテムの共起モデル	40
4.3.4	各ステージにおけるユーザ比較モデル構築	43

4.4	人工データによるシミュレーション実験	45
4.4.1	生成モデルの各パラメータの設定方法	45
4.4.2	人工購買履歴データの生成手順	47
4.4.3	シミュレーション条件	49
4.4.4	シミュレーション結果	50
4.5	実データ分析	53
4.5.1	分析対象データ	53
4.5.2	実験条件	53
4.5.3	提案モデルの推定結果 (潜在クラス数が10の場合)	54
4.5.4	提案モデルの推定結果 (潜在クラス数が20の場合)	56
4.6	考察	64
4.6.1	学習について	64
4.6.2	ステージ成長への施策の立案について	64
4.7	本章の結論	65
第5章	考察	66
5.1	提案手法のまとめ	66
5.2	分析者視点での手法の評価	67
5.3	利用場面の制限の視点での手法の評価	68
第6章	結論と今後の課題	70
6.1	本研究の結論	70
6.2	今後の課題と展望	71
	謝辞	73
	付録A	81
A.1	潜在クラス数の決定方法について	81
A.1.1	モデル選択基準の前提	81
A.1.2	パラメータ数	81
A.1.3	本研究における潜在クラス数の決定	82

付録B 研究業績

第1章 序論

1.1 研究背景

近年、情報技術の発展に伴い、多くの企業は顧客の行動に関する様々なデータを大量に蓄積できるようになっている。これにより、多くの企業が蓄積したデータをどのようにビジネス上の施策に活用するのかについて議論が重ねられている。特に、「どの顧客が、いつ、何を、いくらで、いくつ購入したのか」を表す ID-POS データは小売業を営む企業の多くが取得できるようになったため、顧客の購買嗜好の分析の重要性が高まっている。顧客の購買行動分析は、非常に重要なビジネス課題の1つであり、従来から様々な観点から顧客の購買行動分析の結果をマーケティングに活用する研究が盛んでいる。しかし、企業によって分析ニーズも異なり、万能の分析手法が存在する訳ではないため、新たな特性を持つデータの活用手法は分析目的に応じて開発する必要がある。

1.2 研究対象

1.2.1 Frequent Shoppers Program

小売業界においては、会員顧客との良好な関係を築くためのマーケティングツールとして Frequent Shoppers Program (FSP) [1] というシステムが広く活用されるようになっている。このシステムでは、利用や購入に対してポイントを発行し、頻繁に自社製品やサービスを購入・利用する優良顧客に対して特典や優待などを提供することができ、さらに、顧客の購買行動をより詳細に把握することで、セグメントにも活用することができる。このシステムにより、既存顧客を長期的に繰り返し購入してもらうことで、企業の売上や収益を安定させる効果が期待される。

1.2.2 会員ステージランク制

会員ステージランク制 [2],[3] は Frequent Shoppers Program の中の一つの方法が存在する。会員のある一定期間の累積購買金額や来店回数に応じて顧客のステージを定義し、ステージが上がるごとに顧客に特典や優待を提供するシステムも広く導入されている。また、特定の条件を満たすごとに顧客のステージが上位へ推移し、ポイントやクーポンなどの特典が得られる仕組みである。このシステムでは、上位会員ステージに対して高価な特典内容を設定することで、顧客への購買インセンティブを与えるマーケティングツールとなっている。会員ステージランク制のイメージは図 1.1 に示す。会員ステージランク制は現在、日本で多くの企業から注目されている。この制度を利用している企業例を図 1.2 に示す。例えば、KFC コーポレーション (飲食店)、ドン・キホーテ (スーパー)、楽天 (EC サイト)、d ポイントクラブ (ポイントサービス) などの企業がこのシステムを利用している。本研究では、会員ステージランク制を導入している小売企業に着目し、このシステムを考慮した顧客データ分析手法を研究対象とする。

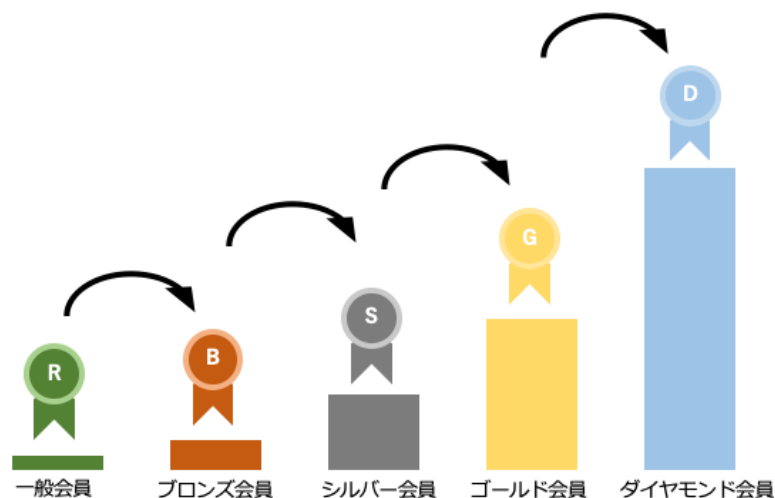


図 1.1: 会員ステージランク制

1.2.3 会員ステージランク制の特徴

図 1.3 に示すように、一般的に会員ステージランク制には次の3つの特徴がある。



図 1.2: 対象企業例

1. 顧客は新しい上位ステージに到達した際、企業から特典や優待が提供される。例えば、「Stage 2 の成長条件」に到達する際に、顧客は企業から「特典 2」を受け取ることができる。
2. 顧客のステージアップの成長条件は会員ステージごとに設定される。通常、下位会員ステージでは、成長のインセンティブを与えるために到達しやすい成長条件が設定される。
3. 「上位会員ステージほど、特典内容の価値も高く設定される」という特徴である。これら 3 つの特徴により、顧客にインセンティブを与え、より高いステージの顧客数を増やすことが会員ステージランク制の主たる目的である。

1.2.4 消費者側のメリット

本節では、会員ステージランク制の消費者側のメリットについて列挙して示す。

- 特典の付与
 - 消費者は商品を購入（サービスを利用）すればするほど、ステージが上がる。各顧客は新たなステージ到達する際に、企業から割引クーポンやショッピングポイントのなどの特典を受け取ることができる。

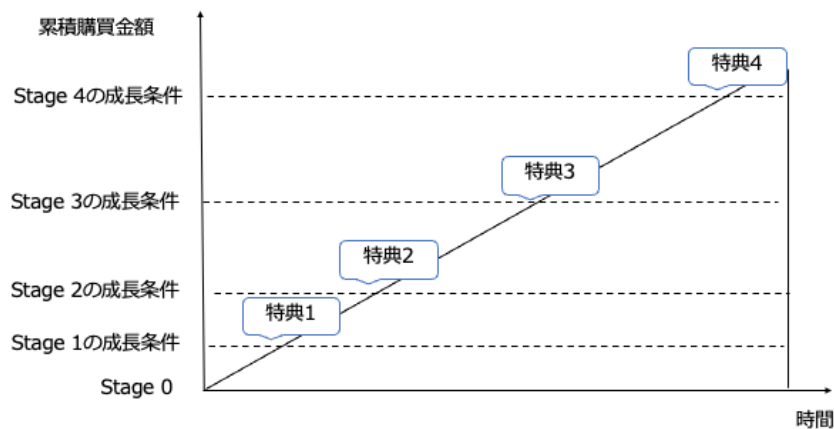


図 1.3: 会員ステージランク制の成長条件と特典の関係の設定

- 会員ステージごとの購買記録の確認
 - 会員ステージランク制度システムにより、消費者は期間内の会員ステージのステータスと過去の購買記録が確認できる。そして、今まで付与された特典の内容が確認でき、次のステージに到達するまでの計画と管理もできる。例えば、「あと〇〇円で xxx ステージに到達できる」、「今月あと〇〇回の利用で xxx ステージに到達できる」などの情報を得ることができる。

1.2.5 企業側のメリット

以下では、会員ステージランク制に対する企業側のメリットについて、列挙して示す。

- 顧客の消費心理向上
 - 会員ステージランク制度の導入により、顧客はより多くの特典をもらうためにアップグレードしたいという心理が働く可能性が高くなる。この心理によって、顧客が本来の購入するはずだった金額以上に商品を購入することにつながる。

- 効果的なビジネス施策
 - － 会員ステージランク制度の導入により、各ステージに所属する消費者の人数を把握でき、それらの情報はビジネス施策の計画に用いることができる。また、ステージに細分化されているため、ターゲットとしたい消費者に絞った施策を打つことができ、企業の運営コストを抑えることにもつながる。

1.3 研究目的

会員ステージランク制において、一般には上位層の顧客は非常に少なく、下位層の顧客が圧倒的に多いという状況がほとんどであるため、何らかの施策が望まれている。しかし、各会員ステージの顧客購買行動が把握できていない場合、有効なビジネス施策を構築することは困難である。すなわち、会員ステージランク制における有効なビジネス施策を計画するためには、企業の売上向上に繋がる優良顧客育成のための顧客購買行動分析が大変重要となる。その際、会員ステージングランク制において、購買行動分析から有効な施策を検討するため、次の2つのアプローチを考える。1つ目は上位ステージの顧客購買行動の特徴を把握することで、上位層の優良顧客の維持を目的とした施策へ結び付けるアプローチである。2つ目はステージ間の顧客の購買行動変化の把握を行うことで、下位層の非優良顧客に対し、より効果的に会員ステージの成長を促す施策を立案するというアプローチである。しかし、これらの2つアプローチのための分析手法としてモデルを構築することを考えると、会員ステージランク制に特有の問題点が生じてしまう。すなわち、下位層の顧客が圧倒的に多いため、全ての会員ステージの顧客の購買履歴データをまとめて1つのモデルを構築した場合、その結果が下位層の顧客の嗜好を反映したものとなってしまう、上位層の顧客の特徴把握や会員ステージ間の特徴比較が困難となることである。一方、各会員ステージで独立にモデルを構築した場合、得られる結果をステージ間で同じ比較基準で比較できないため、有効な施策の検討に結び付けることもできない。従って、会員ステージランク制の特徴を考慮した購買行動分析モデルの構築が望まれる。そこで本研究では、会員ステージランク制を対象とし、ステージごとの潜在的な購買行動を分析できるよ

うなモデルを構築することを目的とする。具体的には、前述した2つのアプローチを実現し、会員ステージランク制度特有の問題を解決するため、以下の2つの手法を提案する。

- 手法1 [第3章] 確率的潜在意味解析の転移学習に基づく会員ステージを考慮した購入行動分析モデル：

転移学習により、膨大な下位ステージの顧客データを上位ステージのモデル推定に再利用し、上位ステージのクラスタリング結果を改善する手法を提案する。具体的には、ステージ間の購買行動が徐々に変化することを仮定することで、隣接するステージ間の顧客購買行動で類似している部分の情報を活用する。また、データ数が多い最下位のステージから得られた信頼性が高いモデルのパラメータを再利用し、徐々に上位層に転移してモデルを構築する。この方法により、上位層モデルのクラスタリング結果が適切な解に収束しやすくなるため、精度が高くなり、より正確な顧客の購買行動を反映することが可能となる。

- 手法2 [第4章] 会員ステージ間の潜在クラス分布を比較する購買行動分析モデル：

ステージ間での顧客の購買傾向（嗜好）の差異に着目し、構築した全体モデルから得られた情報を再利用することで、ステージごとの潜在クラスの所属割合がどのように異なるかを比較する手法を提案する。具体的に各ステージのモデルを構築する際、全体ユーザのモデルから得られた情報をモデルに導入し、ステージごとのモデルに対して購買行動の比較基準を統一することで、ステージ間、クラスごとのユーザの購買行動の差異や人数の変動を比較することが可能な手法を構築する。

顧客の購買行動分析に関する研究は非常に広く行われており、従来から様々な手法やモデルが提案されてきた。これらの分析技術は、実店舗やEC（Electronic Commerce）サイト、決済サービスなど様々な業態を対象として研究が行われている。しかし、会員ステージランク制のデータ構造に基づいた顧客購買行動分析に

における研究に特化したものは非常に少ない。本研究では、会員ステージランク制を対象として精度の高い分析手法を提案しており、人工データを用いた精度評価に加え、実データを用いた購買行動分析モデルの構築を通じてステージ間の顧客の購買行動の変動を明らかにしており、ビジネス施策立案に適用可能な分析モデルの研究の発展につながるものと考えられる。

1.4 論文の構成

本論文は、6章から構成されており、次章以降の内容の概要を以下に示す。

第2章においては、一般的な顧客購買行動分析について述べる。さらに顧客購買行動手法に関する周辺研究について概要をまとめる。

第3章においては、会員ステージランク制を対象としたPLSAモデルに転移学習を導入し、下位層モデルから得られたパラメータを転移することで、上位層モデルの推定精度を向上させた手法を提案する。提案手法で抽出した潜在クラスにより顧客購買行動を分析することで、ステージ間の顧客の購買のトレンドを比較することが可能となる。人工データを生成し、シミュレーション実験を行うことで、提案手法の推定精度の側面での有効性を示す。また、提案手法を実データに適用し、ステージ間の購買トレンドについて分析を行う。

第4章においては、会員ステージ間でPLSAモデルの潜在クラスのサイズを比較可能なモデル構築法を提案する。提案手法は、全体ユーザの購買行動モデルを構築して潜在的な購買嗜好の特徴量を抽出し、全体ユーザモデルから得られた情報を再利用して各ステージのモデルを構築することで、ステージ間の顧客の購買行動の差異を比較可能としている。人工データのシミュレーション実験により、提案手法の推定精度の観点からの有効性を示し、さらに、実データによる分析実験を通じてステージ間の購買行動の変化を分析する。

第5章においては、本研究に対する全体的な観点からの考察を述べると共に、本研究で提案された2つの手法のアプローチ、新規性、解釈性、活用方法の違いについて考察を述べる。

第6章においては、本研究の結論を示し、今後の課題を述べる。

第2章 顧客購買行動分析と周辺研究

顧客購買行動分析は目的に応じて様々手法がある [4],[5],[6]。本章では、分析目的の観点から従来の顧客購買行動に関する分析手法について述べ、会員ステージランク制を対象とした顧客購買行動に関する分析手法の周辺研究について述べる。

2.1 顧客購買行動に関する分析手法

2.1.1 顧客セグメンテーション分析

顧客セグメンテーション分析とは、会社が自社の商品やサービスの売り上げを成長させるため、ビジネス対象となる市場や顧客をある特定の要素により細分化することある。施策を行う際、全体の顧客を同じビジネス施策を実施すると、本来対象ではない顧客に対して大量のコストをかけてしまう。そのため、細分化されるグループに対して最適なビジネス施策を決めようというアプローチが、顧客セグメンテーション分析である。顧客セグメンテーション分析において、従来から、トピックモデルやRFM分析などの手法は一般的な方法として存在し、さらに、下記のように様々な拡張手法が開発される。

- RFM 分析 [7]
 - RFM 分析とは分析期間内、顧客の Recency（直近の購入日）、Frequency（頻度）、Monetary（購入総額）の3つの指標で、顧客をセグメントする分析方法である。これは経営者は計算されたRFM値により、閾値を決めて顧客を区別する。そして、RFM値の高い顧客グループに対してプレミアムクーポンやポイント還元などの施策を行い、顧客のロイヤルティを維持する。また、RFM値の低い顧客グループに離反させないビジネス施策を講じる必要がある [8],[9]。

- Non-Negative Matrix Factorization (NMF)[10],[11]
 - NMFとは非負値行列因子分解であり、特に顧客の購買特徴やパターン抽出に使われる。非負値かつ意味付きの行列で対象となる特定の顧客購買行動を分析することで、得られた結果の解釈性が高い。実応用上、いくつかの手法が提案されている [12],[13].
- Principal Component Analysis(主成分分析, PCA)[14]
 - PCAはデータ内の相関や分散を考慮し、変数を新しい空間に射映し、相関のない変数で全体データのばらつきを表現する手法である。この手法はデータ削減の方法としてよく用いられる。購買履歴分析において、PCAは顧客の購買データの多次元性を削減ことで、顧客の嗜好を分析でき、第一主成分と第二主成分により、より効果的なビジネス施策が可能となる [15].
- Probabilistic Latent Semantic Analysis(確率的潜在意味解析, PLSA)[16]
 - PLSAは代表的なトピックモデルの一つであり、1999年にHofmannらが発表したモデルである。この手法はいくつかの潜在変数を仮定し、高次元データを低次元化する次元圧縮の手法である。この手法は文書分類において重要な手法であり、顧客購買行動分析でもよく使われる。
- Latent Dirichlet Allocation(LDA)[17]
 - LDAも代表的なトピックモデルの一つであり、自然言語処理から発展される。具体的に、一つの文章が複数の潜在的なトピックに確率的に生成される事象を仮定するモデルである。LDAは顧客購買データにも適用する際、購買行動が潜在的なトピック（嗜好）を割り当てることができ、各潜在クラスの特徴により、より適切なビジネス施策が可能となる。さらに、LDAにベースとして様々なモデルが拡張される [18],[19].
- Autoencoder[20]

- Autoencoderは3層構造のニューラルネットであり、教師なしモデルである。その中、入力層と出力層に同じデータを設定し、非線形な特徴抽出と時限削減によく使われる。購買履歴データにおいて、モデルの学習により、中間層で情報を圧縮し、再構築することで、有用な特徴を学習することができる。

2.1.2 バスケット分析

バスケット分析 (Basket Analysis) [21] は、アソシエーション分析 (Association Analysis) [22] の一つであり、EC サイトや小売業購買店で主に顧客がよく一緒に購入する商品のパターンを発見するために使われるデータマイニング技術である。この手法は「もし A を買うなら B も買う確率はどれくらいか」の分析を目的と設定し、商品間の関連性を分析できや顧客の購買行動を把握でき、ビジネスの機会を増やすことが可能になる。バスケット分析のに関する指標は下記の式 (2.1) ~ 式 (2.4) を示す。

$$support(\text{支持度}) = \frac{\text{商品 A と商品 B を購入した人数}}{\text{全顧客人数}} * 100(\%) \quad (2.1)$$

$$confidence(\text{信頼度}) = \frac{\text{商品 A と商品 B を購入した人数}}{\text{商品 A を購入した人数}} * 100(\%) \quad (2.2)$$

$$expectedconfidence(\text{期待信頼度}) = \frac{\text{商品 B を購入した人数}}{\text{全顧客人数}} * 100(\%) \quad (2.3)$$

$$lift(\text{リフト値}) = \frac{\text{信頼度}}{\text{期待信頼度}} = \frac{\text{商品 A の購入者の商品 B 購入率}}{\text{全顧客の商品 B 購入率}} \quad (2.4)$$

2.1.3 顧客満足度分析

顧客満足度分析は、顧客が商品やサービスにどの程度満足しているかを評価するために行う。この分析により、顧客のニーズと期待をより深く理解し、ビジネス

の改善点を特定して顧客体験を向上させることができる。従来研究で、アンケート調査は顧客満足度分析の一つの手段としてよく使われる。例えば、山本らは企業において顧客がロイヤリティを決定付ける要因の構造をアンケート調査により検証し、顧客満足度とロイヤリティの関係の妥当性を検討した [23]。野中らはレストランサービスを対象に職種における顧客接点及び顧客志向性の違いを考慮した従業員満足度モデルを構築し、職種に内部モデルの違いを明らかにし、従業員満足の変化につながる要因を把握した [24]。

2.1.4 顧客生涯価値分析

顧客生涯価値 (Customer Lifetime Value, CLV) とは、顧客関係管理 (CRM : Customer Relationship Management) [25] の一つであり、対象企業が一人の顧客からあるビジネスにもたらす利益という指標である。この分析により、企業は顧客がどの程度の長期的な価値を持っているかを理解し、それに応じてマーケティング戦略や顧客サービスの改善を決定することができる。従来から様々な観点から CLV のモデル化の研究が提案された [26],[27],[28]。

以上の手法は、顧客購買行動の理解と分析において異なる目的からサポートする。実応用上、具体的な問題やデータの特徴、タスクの目的に合わせて選択する必要がある。

2.2 会員ステージランク制を対象とした顧客購買行動に関する分析手法

会員ステージランク制の下ではステージの間で、顧客の優良性を区別しやすいという特徴がある。ここで無印良品の会員ステージランクと年間累計購買金額の関係を示す (表 2.1)。そのため、顧客のステージの成長要因の発見やステージ間の購買行動の差異の比較という観点での分析が効果的であり、これまでもいくつかの研究が行われている。例えば、酒井ら [29] は会員ステージによって優良顧客と非優良顧客を定義するとともに、判別分析による会員ステージ予測モデルを構築し、顧客が将来に離反顧客になるかどうかの要因を発見する手法を提案してい

表 2.1: 会員ステージランクと年間累計購買金額の関係 (例: 無印良品)

ステージ名 (降順)	年間累積金額 (円)	ショッピングポイント (特典)	人数 (人)
ダイヤモンド	20万円以上	1,000ポイント	1,081
プラチナ	10万円以上	500ポイント	6,230
ゴールド	5万円以上	300ポイント	29,623
シルバー	2万円以上	200ポイント	112,287
なし	2万円未満	0	574,752

る。西尾ら [30] はステージ間の差異に注目し、各顧客の潜在的な購買傾向を抽出し、購買傾向とステージ情報を考慮したうえで顧客のステージ向上のための重要商品を抽出する手法を提案した。Yang ら [31] は顧客の異なる成長パターンの施策のため、優良顧客の基準を定義し、ステージを向上させる購買行動の発見の方法を提案している。また、優良顧客が明示的に定義されていない場合に対して、土井ら [32] は RFM 指標を用いてスコアを算出し、優良顧客を定義するとともに、顧客から潜在的な優良顧客を発見する手法を提案した。

また、PLSA モデルが強力な分析手法として EC サイトにおける推薦システムや顧客のクラスタリングなど様々な分野にも応用されている。例えば、清水ら [33] は全ユーザの購買履歴データと、限られた少数のアンケートデータの両方を用いて、回答ユーザと未回答ユーザを顧客の意識を考慮しながら同時にクラスタリングする新たな購買行動分析モデルを提案した。石垣ら [34] は、PLSA をベースに、購買履歴データとアンケートデータを同時に分析し、顧客をクラスタリングするモデルを提案した。また、松寄ら [35] は EC サイトにおける購買履歴に加えて、閲覧履歴、施策実施有無を考慮した PLSA の拡張モデルを提案している。

以上のように、会員ステージランク制度の下でのデータを対象とした分析と PLSA で購買履歴分析が行われている。そして、無印良品のデータを用いて異なる切り口からの研究も存在する [36],[37]。また、異なるビジネス対象への提案手法は表 2.2 にまとめる。しかし、ステージ間で顧客クラスタを比較することを目的とした研究は行われていない。

表 2.2: 異なるビジネス対象への提案手法

著者	対象データ	手法	ビジネス対象
Tsai(2004) [38]	ID-POS 付き購買行動	Genetic Algorithm + RFM 分析	小売店
平野ら (2021) [39]	クレジットカード使用履歴 + 属性	GMM + 転移学習	カード会社
Sun ら (2021)[40]	レシート単位	LDA + Renyi Entropy	小売店
Amela Vatreš ら (2022) [41]	ID-POS 付き購買行動+ 属性	K-Means	イベント性あり購買

2.3 本章の結論

本章では顧客購買行動に関する分析手法の従来研究や周辺研究について述べた。これらの研究は、それぞれ異なる目的や手法によりビジネス施策へ非常に重要な結果を示している。また、会員ステージランク制を対象とした顧客購買行動に関する分析手法をまとめた。特に、ビジネス施策のため、会員ステージランク制の「上位層の優良顧客が少ない」、「下位層の非優良顧客が多い」というデータの特徴を考慮した研究の必要性を指摘した。この後の章で示す本研究の主たる成果はこれらの課題に対して解決を図るものである。

第3章 確率的潜在意味解析の転移学習に基づく会員ステージを考慮した購入行動分析モデル

3.1 本章の目的

本章で顧客の購買行動の分析モデルとして、確率的潜在意味解析法 (PLSA : Probabilistic Latent Semantic Analysis) を用いた手法を提案する。この手法は上位層ステージのクラスタリングの結果の向上を目的として、下位ステージの顧客購買データにより PLSA モデル構築する。そのうえで、転移学習により得られたパラメータを隣接の一つ上のステージ PLSA モデルの初期値として導入し、学習を行うアプローチを取る。PLSA モデルは構造がシンプルで柔軟に顧客の潜在的な購買嗜好の特徴を抽出する手法として、多くの従来研究でその有効性が示されている。本研究の問題設定に適しているため、実務的にビジネス施策に結びを付けるような提案が考えられる。

3.2 Probabilistic Latent Semantic Analysis

Probabilistic Latent Semantic Analysis (PLSA) [16] はトピックモデルの一つであり、ある嗜好を持ったユーザが好みのアイテムを購入する事象をユーザとアイテム共起で表現することができる。具体的に、ユーザの多様な購買嗜好を表現するため、ユーザの生起、アイテムの生起それぞれに多項分布の形を仮定し、ユーザとアイテムの間にいくつかの潜在クラスを仮定することで、潜在クラスが与えられた下でユーザとアイテムの条件付き独立を仮定している。PLSA のグラフィカルモデルを図 3.1 に示す。このモデルは言語モデルや購買履歴データを中心に多くの

適用事例を通じてその有効性が示されている [8],[42],[43],[44].

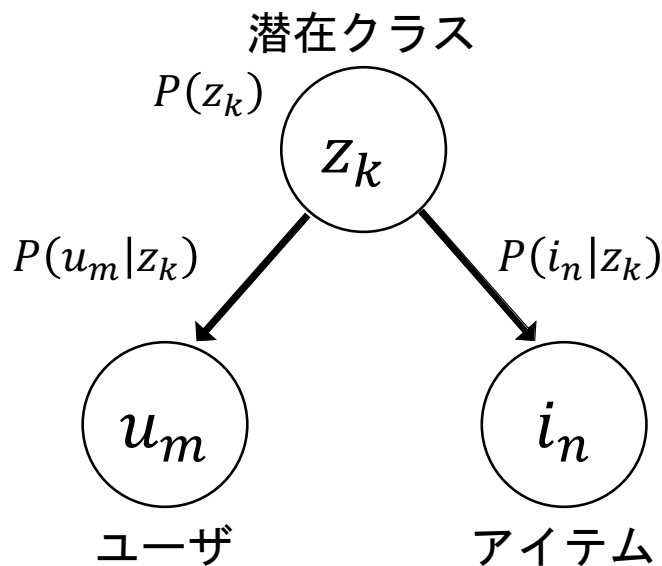


図 3.1: PLSA のグラフィカルモデル

いま，潜在クラスの下でユーザがアイテムを購入する事象を共起で表現するため， M 人のユーザ集合を $\mathcal{U} = \{u_m : 1 \leq m \leq M\}$ ， N 個のアイテム集合 $\mathcal{I} = \{i_n : 1 \leq n \leq N\}$ ， K 個の潜在クラス集合を $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$ と定義する．PLSA の確率モデルは式 (3.1) で表される．

$$P(u_m, i_n) = \sum_{k=1}^K P(z_k)P(u_m|z_k)P(i_n|z_k). \quad (3.1)$$

3.3 EM アルゴリズム

PLSA で，観測できない潜在変数が存在するという仮定を設定するため，EM アルゴリズム [45] を用い，パラメータの推定を行う．まず，E-Step では，各パラメータ $P(z_k)$ ， $P(u_m|z_k)$ ， $P(i_n|z_k)$ が既知とした上で $P(z_k|u_m, i_n)$ を計算する．続く M-Step では E-Step で算出した事後確率 $P(z_k|u_m, i_n)$ を用いて対数尤度関数 LL 最大化するために，E-Step と M-Step を繰り返すことで，パラメータを更新し，ユーザとアイテム共起モデルを推定する．

[E-Step]: ユーザとアイテムに依存するパラメータを固定したときの潜在変数 z_k の分布 $P(z_k|u_m, i_n)$ を式 (3.2) により算出する.

$$P(z_k|u_m, i_n) = \frac{P(z_k)P(i_n|z_k)P(u_m|z_k)}{\sum_{k=1}^K P(z_k)P(i_n|z_k)P(u_m|z_k)} \quad (3.2)$$

[M-Step]: E-Step で算出された $P(z_k|u_m, i_n)$ を用いてそれぞれのパラメータを算出する. 対数尤度 LL を最大化するパラメータを導出するため, 対数尤度を以下のように展開する.

$$LL = \sum_{m=1}^M \sum_{n=1}^N \log P(u_m, i_n) \quad (3.3)$$

$$= \sum_{m=1}^M \sum_{n=1}^N \log \left(\sum_{k=1}^K P(z_k) \cdot P(i_n|z_k)P(u_m|z_k) \right) \quad (3.4)$$

$$= \sum_{m=1}^M \sum_{n=1}^N \log \left\{ \sum_{k=1}^K P(z_k|u_m, i_n) \frac{P(z_k)P(i_n|z_k)P(u_m|z_k)}{P(z_k|u_m, i_n)} \right\} \quad (3.5)$$

$$\geq \sum_{m=1}^M \sum_{n=1}^N \left\{ \sum_{k=1}^K P(z_k|u_m, i_n) \log \frac{P(z_k)P(i_n|z_k)P(u_m|z_k)}{P(z_k|u_m, i_n)} \right\} \quad (3.6)$$

$$= \sum_{m=1}^M \sum_{n=1}^N \sum_{k=1}^K P(z_k|u_m, i_n) \left(\log P(z_k)P(i_n|z_k)P(u_m|z_k) - \log P(z_k|u_m, i_n) \right) \quad (3.7)$$

式 (3.7) は式 (3.5) から式 (3.6) により, Jensen の不等式を用いて展開する. LL' は式 (3.7) の右辺を変形し, 定数項を省略した式である. EM アルゴリズムでは, これを最大化する.

$$LL' = \sum_{m=1}^M \sum_{n=1}^N \left\{ \sum_{k=1}^K P(z_k|u_m, i_n) \log P(u_m|z_k)P(i_n|z_k)P(z_k) \right\} \quad (3.8)$$

さらに, 式 (3.8) を最大化するために, ラグランジュの未定乗数法を用いる. λ_k , η_k , π をそれぞれ未定乗数として $k = 1, 2, \dots, K$, ラグランジュ関数を以下の式 (3.9) のように定義する.

$$\begin{aligned}
g = LL' &+ \sum_{k=1}^K \lambda_k \left(1 - \sum_{m=1}^M P(u_m|z_k)\right) + \sum_{k=1}^K \eta_k \left(1 - \sum_{n=1}^N P(i_n|z_k)\right) \\
&+ \pi \left(1 - \sum_{k=1}^K P(z_k)\right)
\end{aligned} \tag{3.9}$$

これを偏微分して0とおくことにより、それぞれのパラメータの更新式は式(3.10)～式(3.12)のように表すことができる。

[M-Step]:

$$P(u_m|z_k) = \frac{\sum_{n=1}^N P(z_k|u_m, i_n)}{\sum_{m=1}^M \sum_{n=1}^N P(z_k|u_m, i_n)} \tag{3.10}$$

$$P(i_n|z_k) = \frac{\sum_{m=1}^M P(z_k|u_m, i_n)}{\sum_{m=1}^M \sum_{n=1}^N P(z_k|u_m, i_n)} \tag{3.11}$$

$$P(z_k) = \frac{\sum_{m=1}^M \sum_{n=1}^N P(z_k|u_m, i_n)}{\sum_{k=1}^K \sum_{m=1}^M \sum_{n=1}^N P(z_k|u_m, i_n)} \tag{3.12}$$

以上のように対数尤度関数が収束するまでE-StepとM-Stepを繰り返し行う。ここで、対数尤度の変化率が閾値以下であれば（例えば、1%など）、モデルの学習は収束したと判断する。そうでない場合は、EMアルゴリズムの更新を続ける。また、EMアルゴリズムの更新回数を事前に決めて学習を行う方法もある。

3.4 転移学習

機械学習の分野では、主なアプローチがトレーニングデータとテストデータは同じ母集団に所属すると仮定している。しかし、現実世界ではデータ分布が変化する 경우가よくあり、その際に学習したモデルを再構築する必要性が生じてしまう。一方で、モデルを再構築するため、その都度必要なトレーニングデータを収集するとコストが非常に高くなってしまう。これらの問題を解決する方法の1つとし

て、1995年、国際会議 Neural Information Processing Systems (NIPS) で転移学習 [46] という概念が提案された。この概念は、関連分野の知識や経験を利用し、新しいタスクに適用する。例えば、新しいECサイトストアがオープンしたばかりの場合、そのサイトでは、お客様の閲覧ページや購入アイテムなど購買行動データはまだ蓄積されていない。そのため、ECサイト側が顧客に対して協調フィルターリングやコンテンツベースフィルターリングを用いて正確に推薦することができない。一方で、対象ECサイトに関連する別のドメインからの大量のデータが利用可能になっている場合、例えば、他のストアで類似商品の販売データがあれば、これらのデータを用いて目標タスクに適応できる知識を取得し、推薦予測精度を高めることができる。実際に、転移学習においてというだけでなく、一般的な方法として、1つ以上のタスクから得られた知識を利用し、新しいタスクを解決することは有効なアプローチの1つである。転移学習問題においては、学習対象として2つの領域を想定している。転移学習の知識の送り手側を元ドメイン (Source Domain)、転移学習の知識の受け手側を目標ドメイン (Target Domain) と呼ぶ。過去数年間、多くのタスクは転移学習で解決されている [47],[48],[49],[50]。

3.5 課題と本研究への展開

PLSAモデルを実データを適用する際、上位層の優良顧客の維持を目的とした施策のため、高ステージの顧客購買行動の特徴を把握することが重要である。しかし、ステージごとにモデルを学習する場合、上位層の顧客の数が少ないため、顧客の購買行動の推定精度が落ちる可能性が高い。これに対して、全ステージのユーザの購買履歴データを1つの潜在クラスモデルでまとめて学習するというアプローチも考えられるが、その場合、下位層の顧客が圧倒的に多いため、推定結果が下位層のユーザの嗜好に偏ってしまう。上記の問題点を解決するため、本研究では、膨大な下位ステージの顧客データを上位ステージの分析に援用という観点に着目し、上位ステージのクラスタリング結果を改善し、施策提案に役立つ手法を提案する。

3.6 提案手法

3.6.1 概要

本研究では会員ステージランクのもとで、上位層の顧客購買行動をより精度よく推定する手法を提案する。

3.6.2 定式化

ここでは、下位層のユーザと上位層のユーザが異なる分布に従うものと仮定する。元ドメインと目標ドメインは (S) と (T) でそれぞれを表す。下位層（元ドメインのユーザ）の変数 $u_m^{(S)}$ の集合を $\mathcal{U}^{(S)} = \{u_m^{(S)} : 1 \leq m \leq M\}$ 、一つ上の上位層のユーザ（目標ドメインのユーザ）の変数 $u_{m'}^{(T)}$ の集合を $\mathcal{U}^{(T)} = \{u_{m'}^{(T)} : 1 \leq m' \leq M'\}$ と定義する。販売されているアイテムは各ステージで共通かつ二つステージの購買行動が似ているため、隣接する二つのステージに所属するのアイテムの分布が同じアイテムと仮定している。アイテム変数 i_n は、 $I = \{i_n : 1 \leq n \leq N\}$ に設定する。また、真の潜在クラスを観測することはできないため、ユーザのグループ数 V とモデルで用いる潜在クラスの数と同じ数と設定する。また、下位層のユーザと上位層のユーザ顧客からの潜在クラス数が同じ分布に属すると仮定する。元ドメインの潜在クラス変数 $z_k^{(S)}$ の集合を $\mathcal{Z}^{(S)} = \{z_k^{(S)} : 1 \leq k \leq K\}$ 、目標ドメインの潜在クラス変数 $z_k^{(T)}$ の集合を $\mathcal{Z}^{(T)} = \{z_k^{(T)} : 1 \leq k \leq K\}$ と定義する。本章の提案手法では、目標ドメイン PLSA モデルのパラメータ $P(z_k^{(T)}), P(i_n|z_k^{(T)}), P(u_{m'}^{(T)}|z_k^{(T)})$ の初期値に対して元ドメインの PLSA モデルによる $P(z_k^{(S)}), P(i_n|z_k^{(S)})$ の推定結果とランダムな $P(u_{m'}^{(T)}|z_k^{(T)})$ の値を用いる。提案手法のグラフィックモデルは図 3.2 に表す。

3.6.3 提案手法による顧客購買行動分析アルゴリズム

ここで、Low, Middle, High の 3 つのステージを想定して提案手法のアルゴリズム（図 3.3）を示す：

- Step 1: Low ステージ PLSA のモデルの学習は Low ステージユーザのものの購買履歴データを用いる。モデルの初期のパラメータはランダムな初期値に

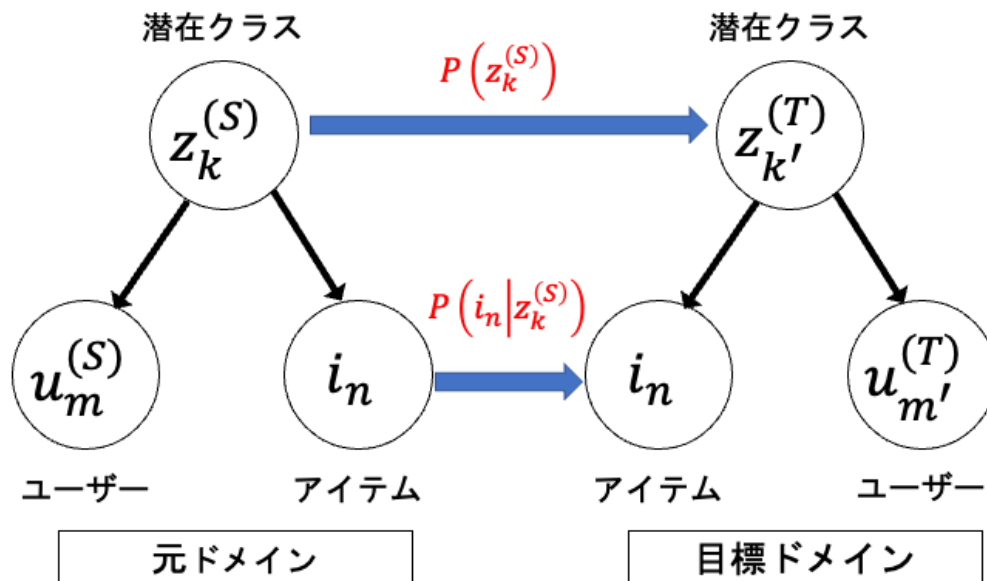


図 3.2: 提案手法のグラフィックモデル

より独立に学習する．モデルのパラメータ $P(z_k^{(Low)})$, $P(i_n | z_k^{(Low)})$ は式 (3.1)~式 (3.12) で得られる．

- Step 2: Step 1 で推定されたパラメータ $P(z_k^{(Low)})$, $P(i_n | z_k^{(Low)})$ を Middle ステージの PLSA モデル学習用のパラメータ $P(z_{k'}^{(Middle)})$, $P(i_n | z_{k'}^{(Middle)})$ の初期値として用いる．
- Step 3: Middle ステージの PLSA モデルは Middle ステージユーザの購買履歴データを用いる．パラメータ $P(z_{k'}^{(Middle)})$ と $P(i_n | z_{k'}^{(Middle)})$ が Middle ステージの PLSA モデルから得られる．
- Step 4: Step 3 の同様に, High ステージの PLSA モデルを学習する前に, Middle ステージの PLSA で得られたパラメータ $P(z_{k'}^{(Middle)})$ と $P(i_n | z_{k'}^{(Middle)})$ を High ステージの PLSA のパラメータ $P(i_n | z_{k'}^{(High)})$ と $P(i_n | z_{k'}^{(High)})$ の初期値として導入する．
- Step 5: High ステージユーザの購買履歴データを用いて High ステージの PLSA のモデルを個別独立に学習する．

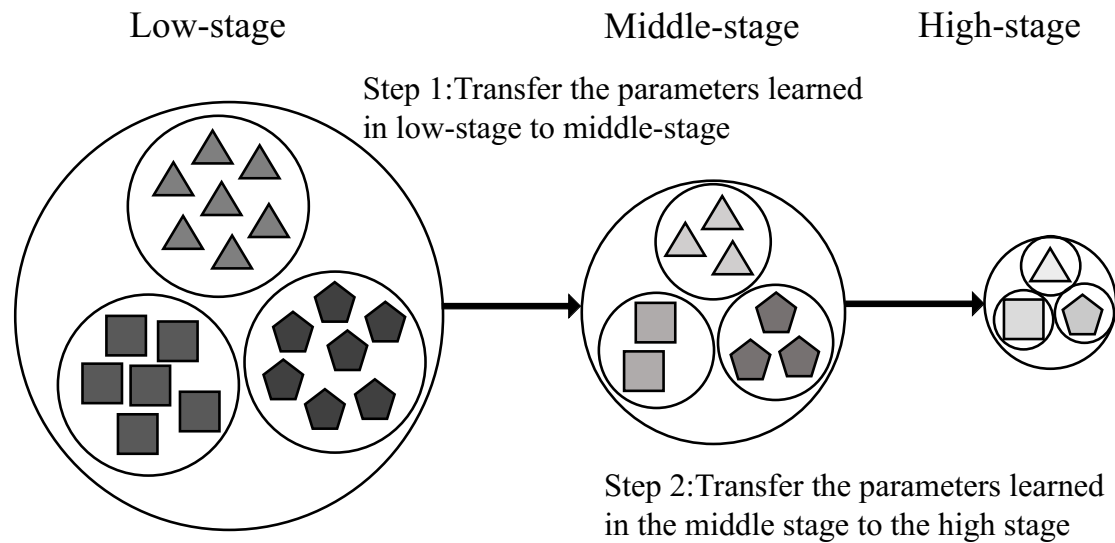


図 3.3: 提案手法のフローチャート

3.7 人工データを用いたシミュレーション実験による精度評価

実際の購買履歴データの真の統計構造を知ることは不可能であるが、本研究で提案手法の性能を検証するため、人工データを用いたシミュレーション実験を行った。無印良品の購買履歴データの特徴として、下位層の主な顧客は、単一のカテゴリーの商品のみ（例えば、文具、食品、化粧品など）を購入する傾向がある。そして、顧客のステージが上がるにつれて、多様なカテゴリーの商品に興味を持つようになる。すなわち、下位層の顧客は一つのお気に入りのカテゴリーから商品を購入する確率が高く、上位層の顧客は複数カテゴリーの商品を購入することがあると言える。ここでは、データ構造の生成を優先にするという前提で、異なるステージのユーザーがアイテムを購入するパラメーターの確率を設定し、多様な購入の推定を目的としてシミュレーションを行う。

3.7.1 グループの定義

本研究では、ユーザとアイテムを V グループに分けて、それぞれを設定する。同じグループに属するユーザはそのグループに属するアイテムを購入しやすい傾向がある（購買傾向が似ている）ことが想定される。ここで、各グループに属するユーザとアイテムとは独立に設定するため、ユーザとアイテムの数は同じである必要はない。本研究の人工データを用いたシミュレーションでは、あらかじめユーザとアイテムを V グループに均等に分割しておく。したがって、 V 個のユーザ・アイテムグループが作成することができる。すなわち、アイテムもユーザーも同じ数のグループが設定される。そして、ユーザとアイテムの数は、それぞれ M と N に設定される。したがって、各グループに属するアイテムの数は N/V であり、ユーザの数は M/V である。

3.7.2 人工データ生成手順

シミュレーション分析で顧客の購買行動の確率分布を設定するため、藤原ら [51] のデータ生成の考え方を参考とし、本研究の条件を満たすデータを生成する。本研究は、藤原らの生成モデルとは異なり、提案手法の妥当性を検証するための購買行動データの生成方法のみが必要である。真の確率 $P(z_k|c_m)$, $P(i_n|z_k)$ は未知であるため、人工データではパラメータ ω と δ として様々な値で設定する。ここで、パラメータ δ は、アイテムの購入確率を意味するのではなく、ユーザが属するクラスの購入確率の総和である。このモデルにより、購入履歴の人工データを以下のように生成する。データの構造のイメージは図 3.4 に示す。

ここでは 1 つのステージのデータを生成する。

- Step 1: ユーザ c_m をランダムに選択する。
- Step 2: V 個のクラスに対して、特定のユーザが特定のカテゴリのアイテムを購入する傾向があると仮定する。そのため、選択されたユーザ c_m あるクラスに属する確率を ω , それ以外のクラスの所属率を $\frac{1-\omega}{V-1}$ と設定する ($\frac{1}{V} \leq \omega \leq 1$)。すなわち、ユーザは特定のクラスのアイテムを購入する確率が高い状況を設定できる。

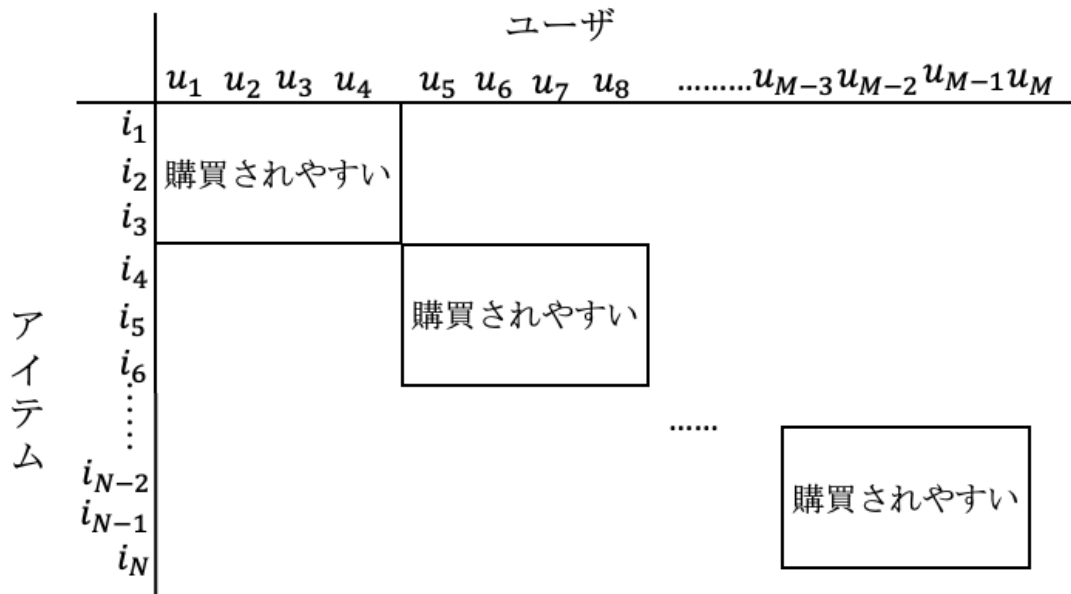


図 3.4: 人工購入履歴データの構造

- Step 3: 特定のクラスに属するアイテムの購入確率を δ と設定する. あるクラスに属するアイテムの出現確率を $\frac{V\delta}{N}$, それ以外を $\frac{V(1-\delta)}{N(V-1)}$ とする ($0 \leq \delta \leq 1$).
- Step 4: あるクラスは Step 2 で設定された確率で選択される. 次に, Step 3 である確率で選択されたアイテムを購入アイテムとする. 以上の手順により, ユーザ c_m がアイテム i_n を購入したことを購入履歴データとしてに記録される. 生成されたデータの量が最大になるまで Step 1 で繰り返す.

以上の手順により, パラメータ ω により, ユーザが属する 1 つのクラスのアイテムに対して購入傾向が強いという傾向がある購買履歴データを作成することができる. そして, パラメータ δ により, 特定のクラスのアイテムごとに購入確率をコントロールすることも可能となる. しかし, パラメータ ω が大きいと, 購買履歴データにおいて「特定の 카테고리に属するアイテムのみを購入する傾向」が強くなり, 逆に小さいと, 「多様なカテゴリに属するアイテムを購入する傾向」が強くなる. また, パラメータ δ についても同様の解釈性が可能である. この値が大きいと, ユーザは特定のカテゴリーのアイテムを購入する傾向が強いと言える.

3.7.3 シミュレーション条件

本実験では、実データの事前分析により無印良品の実データ構造の一部を参照し、人工データを作成する。ユーザランクは「Stage 0」, 「Stage 1」, 「Stage 2」, 3つのステージで構成されると仮定する。転移の元ドメインと目標ドメインの定義は表 3.1 に示す。真のグループ数は 5, アイテムの種類は 700 と設定する。「Stage 0」, 「Stage 1」, 「Stage 2」の顧客数はそれぞれ 9000, 4000, 2000 を、各ユーザの平均購入アイテム数はそれぞれ 50, 70, 110 と設定する。そして、潜在クラス数は 5 と設定する。以上の設定により学習データ数は「Stage 0」, 「Stage 1」, 「Stage 2」それぞれに対して 450000, 280000, 220000 となる。これで通常の購買履歴データと同様に、ステージが高いほど、ユーザはより多くの種類のアイテムを購入する傾向になる。同時に、ステージが高いほど、ユーザ数も少なくなる。

表 3.1: シミュレーションで元ドメインと目標ドメインの設定

	元ドメイン	目標ドメイン
Stage 名	Stage 0	Stage 1
	Stage 1	Stage 2

本章で、実際の購買行動の特徴を再現するため、ユーザは 2 つのステージの間で異なることを仮定する。そこで、 δ を固定値とし、 ω を 0.0, 0.1, 0.2, 0.3, 0.4 と変化させ、シミュレーション用の人工データを生成した。シミュレーションにおいて合計 50 個のデータセットを作成し、各データセットを用いて 10 回繰り返した。その結果の平均精度をもって効果を示した。

3.7.4 評価指標

提案アルゴリズムの性能を評価するため、KL-Divergence[52] を用いる。比較対象として「真の分布とランダムな初期値で推定させた結果の距離」と「真の分布と提案手法を用いて推定した分布との距離」になる。KL-Divergence の式は、式 (3.13) である。

$$D_{KL}(P(i, u)^{(real)} || P(i, u)^{(est)}) = \sum_{i=1}^I \sum_{u=1}^U P(i, u)^{(real)} \log \frac{P(i, u)^{(real)}}{P(i, u)^{(est)}}. \quad (3.13)$$

ここで、 $P(i, u)^{(real)}$ は各ステージの真の分布、 $P(i, u)^{(est)}$ は推定した実際の分布を示す。

3.7.5 結果

従来の PLSA モデルによる結果と提案手法による結果の比較を図 3.5 と図 3.6 に示す。棒グラフでは、縦軸の KL-Divergence の値が小さいほど推定精度が高いことを示す。そして、各 Phase で結果を比較している。ここでは、3つのケースの結果を比較している：

Case 1: 各ステージのデータを独立にランダムな初期値により PLSA モデルを学習する。

Case 2: 2つのステージのデータを組み合わせ、それぞれ独立にランダムな初期値により PLSA モデルを学習する。

Case 3: 下位ステージの PLSA モデルで学習した結果を上位ステージの初期値として用い、上位ステージの PLSA モデルを学習する。

図 3.5 は Stage 0 の PLSA モデルから得られた PLSA モデルパラメータを Stage 1 の PLSA モデルの初期値として用いた場合、従来手法と提案手法による真の分布と推定された分布の距離を示す。提案手法のパフォーマンスはランダムな初期値に基づく PLSA モデルの結果により良い。同じ分布に属するデータの数が多の場合、PLSA モデルの推定精度が向上することが分かる。

図 3.6 は Stage 1 の PLSA モデルから得られた PLSA モデルパラメータを Stage 2 の PLSA モデルの初期値として用いた場合、従来手法と提案手法による真の分布と推定された分布の距離を示す。図 3.6 の Phase 1 から Phase 3 までの状況により、パラメータの転移で従来手法より良い結果が得られていることが分かる。しかし、2回目の転移先のデータ分布が違いすぎると、負の転移が発生してしまう。即ち、

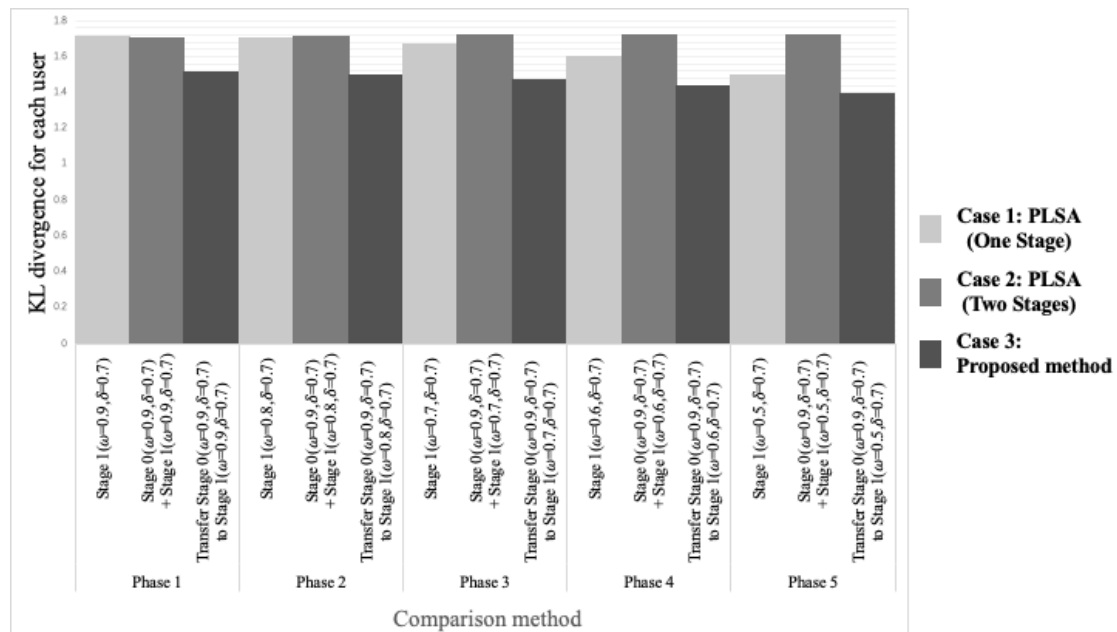


図 3.5: 結果 1: 提案手法と従来 PLSA モデルの精度比較 (Stage 0 から Stage 1 まで) : ω 変化, δ 固定.

データ間の分布の差が大きい場合、PLSA モデルの推定精度は低くなってしまいます。例えば、Phase 4 と Phase 5 の場合、データ間の分布の差分が大きいため、提案手法はうまく機能しない。そのため、実応用では、事前分析の段階で顧客のステージ間の購入傾向を確認し、各ステージの顧客の購カテゴリごとに購入された異なるアイテムの総数を比較することで、負の転移を回避することができる。

以上の分析により、提案手法の精度が高い理由は、異なるステージ間で共通に有用な特徴が類似しているため、適切なパラメータを見つけやすいことにある。

3.8 実データ分析

3.8.1 分析対象データ

本稿では、株式会社良品計画から提供頂いた会員購買履歴データを分析対象とする。顧客がより高いステージに到達した際、ショッピングギフトポイントが付与される。ここで、無印良品ブランドの会員ステージランク制は、顧客ロイヤルティを5つのステージ (Stage 0: なし, Stage 1: シルバー, Stage 2: ゴール

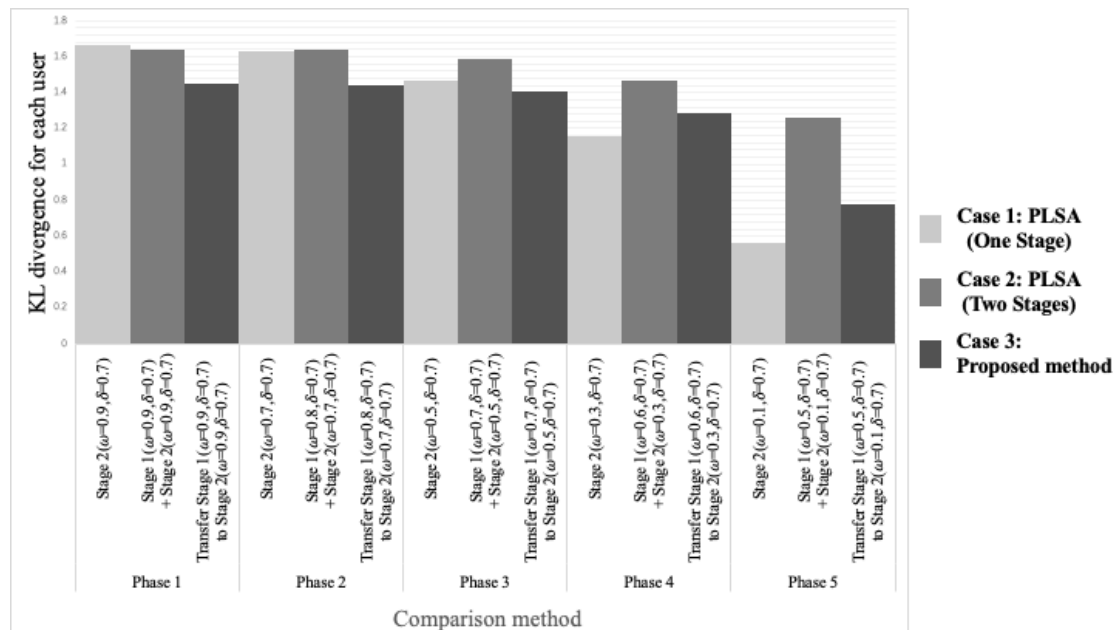


図 3.6: 結果 2: 提案手法と従来 PLSA モデルの精度比較 (Stage 1 から Stage 2 まで) : ω 変化, δ 固定.

ド, Stage 3: プラチナ, Stage 4: ダイヤモンド) で構成されている. 分析の対象期間は 2014 年 3 月 1 日~2015 年 2 月 28 日である. ただし, 毎年 2 月末にステージがリセットされ, 3 月に Stage 0 から再スタートとなる. アイテム数は 690 種類である.

3.8.2 実験結果

予備実験により, 潜在クラス数を $K = 10$ とした. そして, 無印良品の顧客購買実データを用いて, 提案手法を適用した結果を表 3.2 ~ 表 3.6 に示す. 実際のデータ分析では真の分布を観測することは不可能であるため, 人工データ分析のような絶対的な精度測定を算出し, これらの結果を比較できない. しかし, 構築したモデルにより, 全ステージの顧客購買傾向を通じて各潜在クラスから出てきたランキングの Top 10 のアイテムを分析することは可能である.

ここで, 表 3.2 ~ 表 3.6 中の ♣ 付きのアイテムは, ステージが成長しても, このクラスで重要性 (ランキング) が変化しないアイテムと定義する. そのため, こちらのアイテムは顧客ステージの成長に対して欠かせないベースアイテムとも言え

表 3.3.3: 各ステージの上位 10 アイテムの出現確率 (クラス 3, クラス 4).

各ステージの上位 10 アイテムの出現確率 (クラス 3)									
Stage 0 の出現確率: 0.087		Stage 1 の出現確率: 0.093		Stage 2 の出現確率: 0.092		Stage 3 の出現確率: 0.085		Stage 4 の出現確率: 0.097	
♣ 食品 11	0.124	♣ 食品 11	0.109	♣ 食品 11	0.101	♣ 食品 11	0.108	♣ 食品 11	0.115
食品 15	0.043	衣類・雑貨 1	0.062	衣類・雑貨 1	0.071	衣類・雑貨 1	0.072	食品 8	0.075
食品 8	0.037	食品 8	0.035	衣類・雑貨 3	0.038	衣類・雑貨 3	0.045	衣類・雑貨 1	0.054
衣類・雑貨 1	0.036	衣類・雑貨 7	0.024	食品 8	0.034	食品 8	0.041	食品 12	0.054
生活雑貨 22	0.025	衣類・雑貨 3	0.021	衣類・雑貨 7	0.034	食品 12	0.031	食品 9	0.043
生活雑貨 23	0.022	衣類・雑貨 8	0.019	食品 12	0.023	衣類・雑貨 7	0.029	衣類・雑貨 3	0.036
食品 16	0.020	衣類・雑貨 9	0.019	食品 13	0.022	食品 13	0.026	食品 13	0.036
生活雑貨 24	0.020	生活雑貨 22	0.019	衣類・雑貨 8	0.022	衣類・雑貨 6	0.024	衣類・雑貨 5	0.031
食品 17	0.019	生活雑貨 23	0.018	衣類・雑貨 9	0.022	食品 9	0.023	衣類・雑貨 6	0.027
食品 9	0.019	食品 12	0.018	衣類・雑貨 10	0.022	衣類・雑貨 8	0.022	食品 14	0.026

各ステージの上位 10 アイテムの出現確率 (クラス 4)									
Stage 0 の出現確率: 0.089		Stage 1 の出現確率: 0.090		Stage 2 の出現確率: 0.083		Stage 3 の出現確率: 0.075		Stage 4 の出現確率: 0.074	
♣ 食品 7	0.149	♣ 食品 7	0.170	♣ 食品 7	0.213	♣ 食品 7	0.225	♣ 食品 7	0.279
♣ 生活雑貨 25	0.073	♣ 生活雑貨 25	0.095	♣ 生活雑貨 25	0.082	生活雑貨 23	0.097	食品 18	0.193
食品 18	0.066	食品 18	0.061	生活雑貨 23	0.080	食品 18	0.086	食品 19	0.045
食品 25	0.043	生活雑貨 23	0.059	食品 18	0.077	♣ 生活雑貨 25	0.065	食品 20	0.038
食品 26	0.035	生活雑貨 26	0.044	生活雑貨 26	0.045	生活雑貨 26	0.042	食品 21	0.035
生活雑貨 28	0.033	食品 22	0.033	食品 19	0.035	食品 19	0.036	生活雑貨 23	0.033
食品 22	0.032	生活雑貨 4	0.031	生活雑貨 4	0.033	食品 21	0.033	食品 22	0.031
生活雑貨 26	0.031	食品 19	0.030	食品 22	0.032	生活雑貨 4	0.027	食品 10	0.030
生活雑貨 23	0.031	食品 25	0.027	食品 21	0.029	食品 22	0.025	食品 23	0.023
生活雑貨 29	0.029	生活雑貨 27	0.027	食品 23	0.027	生活雑貨 27	0.023	食品 24	0.021

表 3.4: 各ステージの上位 10 アイテムの出現確率 (クラス 5, クラス 6).

各ステージの上位 10 アイテムの出現確率 (クラス 5)					
Stage 0 の出現確率:0.044	Stage 1 の出現確率:0.040	Stage 2 の出現確率:0.040	Stage 3 の出現確率:0.041	Stage 4 の出現確率:0.036	
♣ 食品 7	★ サービスして 1	★ サービスして 1	★ サービス 1	★ サービス 1	★ サービス 1
★ サービス 1	♣ 食品 7	★ 生活雑貨 8	★ 生活雑貨 8	★ 生活雑貨 8	★ 生活雑貨 8
食品 11	★ 生活雑貨 8	生活雑貨 30	生活雑貨 30	生活雑貨 30	生活雑貨 30
食品 25	生活雑貨 30	♣ 食品 7	生活雑貨 31	生活雑貨 31	生活雑貨 31
★ 生活雑貨 8	衣類・雑貨 3	生活雑貨 31	多角化 1	多角化 1	多角化 1
生活雑貨 26	生活雑貨 31	食品 27	食品 5	食品 27	食品 27
衣類・雑貨 3	食品 27	食品 5	♣ 食品 7	衣類・雑貨 11	衣類・雑貨 11
食品 27	食品 5	食品 21	食品 21	生活雑貨 32	生活雑貨 32
食品 6	食品 29	衣類・雑貨 11	食品 28	食品 6	食品 6
生活雑貨 9	食品 21	衣類・雑貨 3	食品 29	食品 5	食品 5

各ステージの上位 10 アイテムの出現確率 (クラス 6)					
Stage 0 の出現確率:0.109	Stage 1 の出現確率:0.092	Stage 2 の出現確率:0.084	Stage 3 の出現確率:0.079	Stage 4 の出現確率:0.087	
♣ 生活雑貨 11	♣ 生活雑貨 11	♣ 生活雑貨 11	♣ 生活雑貨 11	♣ 生活雑貨 11	♣ 生活雑貨 11
食品 15	生活雑貨 33	生活雑貨 33	食品 18	生活雑貨 14	生活雑貨 14
生活雑貨 33	生活雑貨 28	食品 18	生活雑貨 33	★ 生活雑貨 22	★ 生活雑貨 22
生活雑貨 28	食品 15	生活雑貨 28	生活雑貨 14	生活雑貨 28	生活雑貨 28
食品 18	食品 18	生活雑貨 14	生活雑貨 28	生活雑貨 33	生活雑貨 33
生活雑貨 35	★ 生活雑貨 22	★ 生活雑貨 22	★ 生活雑貨 22	★ 生活雑貨 13	★ 生活雑貨 13
生活雑貨 34	生活雑貨 14	★ 生活雑貨 13	★ 生活雑貨 13	生活雑貨 9	生活雑貨 9
★ 生活雑貨 22	★ 生活雑貨 13	食品 15	生活雑貨 9	食品 18	食品 18
衣類・雑貨 1	生活雑貨 34	生活雑貨 9	衣類・雑貨 14	衣類・雑貨 12	衣類・雑貨 12
★ 生活雑貨 13	生活雑貨 9	衣類・雑貨 14	衣類・雑貨 15	衣類・雑貨 13	衣類・雑貨 13

表 3.5: 各ステージの上位 10 アイテムの出現確率 (クラス 7, クラス 8).

各ステージの上位 10 アイテムの出現確率 (クラス 7)					
Stage 0 の出現確率:0.044	Stage 1 の出現確率:0.043	Stage 2 の出現確率:0.051	Stage 3 の出現確率:0.056	Stage 4 の出現確率:0.047	
◆ 食品 11	◆ 食品 11	◆ 食品 11	★ 生活雑貨 23	★ 生活雑貨 23	★ 生活雑貨 23
0.152	0.139	0.107	0.111	0.111	0.203
食品 15	食品 30	★ 生活雑貨 23	生活雑貨 36	生活雑貨 36	生活雑貨 36
0.060	0.071	0.081	0.087	0.087	0.105
生活雑貨 33	★ 生活雑貨 23	食品 30	食品 30	生活雑貨 37	生活雑貨 37
0.049	0.055	0.079	0.073	0.073	0.068
生活雑貨 21	食品 31	生活雑貨 36	◆ 食品 11	生活雑貨 38	生活雑貨 38
0.041	0.053	生活雑貨 36	生活雑貨 38	食品 30	0.063
食品 30	生活雑貨 36	食品 32	生活雑貨 37	衣類・雑貨 16	0.061
0.035	0.052	生活雑貨 37	生活雑貨 40	食品 31	0.044
★ 生活雑貨 23	食品 32	生活雑貨 40	食品 32	生活雑貨 39	0.040
0.033	0.049	生活雑貨 37	食品 31	衣類・雑貨 17	0.038
食品 22	生活雑貨 40	生活雑貨 37	衣類・雑貨 16	食品 32	0.037
0.032	0.048	生活雑貨 37	衣類・雑貨 17	食品 31	0.033
衣類・雑貨 15	生活雑貨 37	食品 31	食品 31	食品 32	
0.029	0.044	食品 31	0.039	0.041	
0.027	0.039	衣類・雑貨 16	0.037	0.041	
食品 31	衣類・雑貨 15	衣類・雑貨 17	0.037	0.041	
0.026	0.035	衣類・雑貨 17	0.037	0.041	
食品 32	食品 22	食品 17	0.037	0.041	

各ステージの上位 10 アイテムの出現確率 (クラス 8)					
Stage 0 の出現確率:0.171	Stage 1 の出現確率:0.161	Stage 2 の出現確率:0.163	Stage 3 の出現確率:0.192	Stage 4 の出現確率:0.175	
食品 18	★ 生活雑貨 25	★ 生活雑貨 25	★ 生活雑貨 25	★ 生活雑貨 25	★ 生活雑貨 25
0.092	0.113	0.161	0.165	0.165	0.144
サービス 1	サービス 1	★ 生活雑貨 26	★ 生活雑貨 26	★ 生活雑貨 26	★ 生活雑貨 26
0.076	0.088	生活雑貨 19	生活雑貨 19	生活雑貨 41	0.089
★ 生活雑貨 25	食品 18	生活雑貨 19	生活雑貨 19	生活雑貨 42	0.051
0.059	0.067	◆ 生活雑貨 33	◆ 生活雑貨 41	生活雑貨 42	0.046
生活雑貨 33	生活雑貨 33	サービス 1	◆ 生活雑貨 33	生活雑貨 19	0.043
0.059	0.063	生活雑貨 26	生活雑貨 41	◆ 生活雑貨 33	0.043
生活雑貨 11	★ 生活雑貨 19	生活雑貨 19	生活雑貨 43	生活雑貨 43	0.040
0.054	0.054	生活雑貨 41	生活雑貨 44	生活雑貨 44	0.039
★ 生活雑貨 26	生活雑貨 19	★ 生活雑貨 41	生活雑貨 42	生活雑貨 44	0.035
0.025	0.037	生活雑貨 41	生活雑貨 47	生活雑貨 45	0.035
生活雑貨 48	★ 生活雑貨 41	生活雑貨 43	生活雑貨 47	生活雑貨 46	0.033
0.022	0.031	生活雑貨 47	生活雑貨 45	生活雑貨 46	0.032
生活雑貨 49	生活雑貨 48	生活雑貨 47	生活雑貨 44	生活雑貨 46	0.032
0.018	0.030	生活雑貨 45	生活雑貨 44	生活雑貨 46	0.032
生活雑貨 19	生活雑貨 43	生活雑貨 44	生活雑貨 44	生活雑貨 46	0.032
0.017	0.026	生活雑貨 44	生活雑貨 44	生活雑貨 46	0.032
★ 生活雑貨 41	生活雑貨 49	生活雑貨 44	生活雑貨 44	生活雑貨 46	0.032
0.016	0.025	生活雑貨 44	生活雑貨 44	生活雑貨 46	0.032

表 3.6: 各ステージの上位 10 アイテムの出現確率 (クラス 9, クラス 10) .

各ステージの上位 10 アイテムの出現確率 (クラス 9)					
Stage 0 の出現確率:0.225	Stage 1 の出現確率:0.270	Stage 2 の出現確率:0.275	Stage 3 の出現確率:0.253	Stage 4 の出現確率:0.294	
◆ 食品 15	◆ 食品 15	◆ 食品 15	◆ 食品 15	◆ 食品 15	生活雑貨 35
生活雑貨 35	生活雑貨 34	生活雑貨 34	生活雑貨 34	◆ 食品 15	◆ 食品 15
生活雑貨 34	生活雑貨 35	生活雑貨 35	生活雑貨 35	生活雑貨 35	生活雑貨 12
食品 1	食品 25	食品 25	◆ 食品 26	★ 食品 26	★ 食品 26
◆ 生活雑貨 29	◆ 生活雑貨 29	◆ 生活雑貨 29	◆ 食品 25	食品 25	生活雑貨 34
食品 25	★ 食品 26	★ 食品 26	★ 食品 26	★ 食品 26	生活雑貨 50
★ 食品 26	★ 食品 26	★ 食品 26	★ 食品 26	★ 食品 26	生活雑貨 51
★ 食品 26	★ 食品 26	★ 食品 26	★ 食品 26	★ 食品 26	生活雑貨 52
★ 食品 26	★ 食品 26	★ 食品 26	★ 食品 26	★ 食品 26	食品 25
★ 食品 26	★ 食品 26	★ 食品 26	★ 食品 26	★ 食品 26	生活雑貨 15
★ 食品 26	★ 食品 26	★ 食品 26	★ 食品 26	★ 食品 26	生活雑貨 15

各ステージの上位 10 アイテムの出現確率 (クラス 10)					
Stage 0 の出現確率:0.034	Stage 1 の出現確率:0.029	Stage 2 の出現確率:0.041	Stage 3 の出現確率:0.057	Stage 4 の出現確率:0.057	
食品 7	生活雑貨 20	生活雑貨 20	生活雑貨 20	生活雑貨 54	生活雑貨 54
生活雑貨 20	衣類・雑貨 18	生活雑貨 54	生活雑貨 54	生活雑貨 20	生活雑貨 20
生活雑貨 29	生活雑貨 54	生活雑貨 57	生活雑貨 55	生活雑貨 40	生活雑貨 40
衣類・雑貨 18	生活雑貨 57	生活雑貨 40	生活雑貨 40	生活雑貨 55	生活雑貨 55
生活雑貨 35	生活雑貨 27	生活雑貨 55	生活雑貨 57	生活雑貨 56	生活雑貨 56
食品 26	生活雑貨 40	生活雑貨 27	生活雑貨 56	生活雑貨 57	生活雑貨 57
食品 18	生活雑貨 55	生活雑貨 56	生活雑貨 27	生活雑貨 27	生活雑貨 27
生活雑貨 57	生活雑貨 5	生活雑貨 59	生活雑貨 59	生活雑貨 58	生活雑貨 58
生活雑貨 5	生活雑貨 59	生活雑貨 58	生活雑貨 58	生活雑貨 59	生活雑貨 59
生活雑貨 27	生活雑貨 37	生活雑貨 5	生活雑貨 60	生活雑貨 60	生活雑貨 60

る。そして、★付きのアイテムは、ステージが成長するにつれて、このクラスで重要性(ランキング)が高くなるアイテムであると定義する。最後、♠付きのアイテムは、このクラスでステージが上がるにつれてアイテムの重要性(ランキング)が低くなるアイテムであると定義する。表3.2~表3.6から見ると、隣接する2つのステージ間の推定確率は、ステージが変化しても大幅には変わらない。ただし、潜在クラス1と潜在クラス2では、ステージが変わるにつれてステージ間の推定確率の差が減少する傾向を示している。

次に、潜在クラスの推定確率と Top 10 のアイテムの内容の変化を比較することで、顧客の購買行動もステージごとに徐々に変化していることが分かる。例えば、クラス1、クラス2、クラス3、クラス9のランキングを見ると、顧客の購買行動が比較的安定していることが分かる。ただし、クラス4のランキングで、「食品7」の重要性は安定しているが、他の食品の重要性はわずかに変化している。クラス5では、ステージが成長するにつれて、「サービス1」と「生活雑貨8」の重要性が着実に増加している。また、クラス6における「生活雑貨11」の重要度は安定しているが、「生活雑貨22」、「生活雑貨13」の重要度が増加していることが分かる。クラス7で、ステージが上がるにつれて食品の重要性が下がり、生活雑貨の重要性は高まることが分かる。しかし、クラス8では、段階が上がるにつれて「食品18」、「サービス1」の順位が下がる一方、「生活雑貨25」、「生活雑貨26」、「生活雑貨41」の順位が上がる。同様に、クラス10の Stage 0 では、「食品7」、「食品26」、「食品18」が上位のランキングに入っていることが分かる。顧客のステージが上がるにつれて、食品の順位が下がり、「生活雑貨54」、「生活雑貨40」、生活雑貨の重要性が高くなる。そして、「生活雑貨55」、「生活雑貨59」の重要性も増えている。ここでは、顧客のステージ成長に重要なアイテムとして、ステージが上がると順位が上がる項目を想定する。これらの商品については、クーポン配布やリコメンドなどのビジネス施策を優先的に実施し、顧客の購買意欲を向上させ、次のステージへのステップアップを目指す。このほか、クラス1での「生活雑貨11」、クラス4での「生活雑貨25」、クラス5での「食品7」、クラス7での「食品11」、クラス9での「生活雑貨29」などにも注目する必要がある。これらのアイテムは、上位ステージに行くにつれて徐々にランクは下がっていくことが分かる。従って、より高い段階で同様の購買行動を持つ顧客を引き付けるために、企業は何らかのビジ

ネス施策を行うことができる。

この分析の結論として、各段階における潜在クラスの差異分析により、顧客の購買行動において隣接する2つのステージで劇的に変化するわけではないが、すべてのステージの間では変化的であることが判明した。上位ステージに比べて、下位ステージでは実際により多くのデータが取得される。シミュレーション実験と同様に、モデルのパラメータに対してランダムな初期値ではなく、下位ステージの結果を適切な初期値として使用することも可能である。多数の下位ステージ顧客によるPLSAの結果から、下位ステージでの人気商品と購買傾向が分かる。これは下位ステージモデルから隣接する上位ステージモデルに良い初期値を示し、EMアルゴリズムにより優れた局所最適解を探索する。提案手法により、より良い説明性がある結果を得るために、元ドメイン（下位層ステージ）からの知識（モデル初期値）を目標ドメイン（上位層ステージ）に移すことは可能となる。さらに、人気商品と購買傾向をステージごとに分析することができ、顧客の成長を把握し、今後の経営方針につなげることができる。実データに対しては、真の確率分布を知る術はないため、精度向上の程度を定量的に示すことはできないが、人工データに対するシミュレーションの結果から、これらの結果はより信頼ができるものと推察している。

3.9 考察

本節では、人工データを用いたシミュレーション実験と実データ分析の結果から、本章の提案手法の特性や活用法などについて考察を行う。

3.9.1 定量的な検証

シミュレーション実験により、提案手法は、ステージ間の顧客購買行動が似ている、かつ低ステージのデータサイズが大きい場合、提案手法の推定が有効であることが示された。一方、ステージ間の顧客購買行動が全く異なる場合には、提案手法は適用できないことに注意する必要がある。

3.9.2 定性的な解釈

- 無印良品のデータにおいて、表 2.1 に示すように、最下層の顧客数が圧倒的に多い。そのため、最下位ステージの顧客に対するクラスタリング分析の結果は、各データの中で最も信頼性が高い。従って、下位ステージで推定された PLSA モデルのパラメータは、隣接する上位ステージの PLSA モデルの学習の初期値として転移されることが可能となる。
- PLSA モデルのパラメータ $P(z_k)$, $P(i_n|z_k)$, $P(c_m|z_k)$ の推定は EM アルゴリズムを用いる。EM アルゴリズムを適用する場合、推定されたパラメータは尤度関数の局所解であることに注意する必要がある。そのため、PLSA モデルのクラスタリングの結果は、モデルパラメータ推定の初期値の選び方に依存する。
- データ数の多い下位層のステージで推定されたパラメータ $P(z_k)$ と $P(i_n|z_k)$ はモデル推定のパラメータの初期値としてランダムな初期値よりも各クラスの推定に対して意味を持ち、信頼性の高い初期パラメータと考えられる。そのため、シミュレーション実験により、提案手法で得られたパラメータを PLSA モデルの初期値として適用することで、データの背後にある真の顧客特性をよりよく反映することができる。
- 実データの分析結果により、ステージ間の購買行動のトレンドが把握できる。また、潜在クラスとアイテムの確率から、各ステージにおける顧客の行動傾向も把握でき、これで企業のマーケティング戦略に役立つと考えられる。

3.9.3 潜在クラス数の決定

一般に、顧客の購買履歴は、非常に複雑な現実世界の社会活動をデータとして表現する。顧客の購買履歴データ数が増えるほど、構築したモデルの精度が高くなる。そして、潜在クラス（トピックの数）が多いモデルほど、精度が高くなる。しかし、通常、学習データの数は無限に増やすことはできない。顧客数とアイテム数の掛け算で定義されるパラメータ数を有する確率モデルの推定の複雑さは、仮に大規模の学習データを用意したとしても、理想的な状況（データが無量大）から

は程遠い。一方で、有限数の潜在クラスを持つ潜在クラスモデルは、各クラスの中に同様の確率分布を持つ顧客と商品が集合的に同じであると仮定するため、実際の分析に対して効果がある。それは、人工的なデータによるシミュレーション実験で、真の潜在クラスモデルが仮定される場合、パラメータの評価が可能となる。さらに、一定数の真の潜在クラスが設定されることで、人工データを生成し、モデルの推定精度を検証できる。しかし、パラメータの対称構造、最尤推定の非一意性、およびEMアルゴリズムの特性により、推定されたクラスの数と真の潜在クラスの数と等しい場合でも、最適な推定結果が得られているとは限らない。さらに、膨大な数のユーザやアイテムに対して潜在クラス数を適切に設定することは容易ではない。その理由としては実際の分析の際、潜在クラス数が限られているとはいえ、低い解釈性を持つグループを設定し、顧客購買行動による経営分析を行うことは困難である。したがって、推定されたパラメータ(学習の結果)は、各潜在クラスの特徴を定性的に解釈する必要がある。そして、クラスの解釈可能性の観点から、最も理解しやすい潜在クラスの意味を決めて分析する。上記の理由により、本章では潜在クラスの解釈可能性に応じて、先行研究の潜在クラス数の決定方法を参照[53]し、事前分析を行い、潜在クラス数を決定した。

3.9.4 提案の有効性

本章の提案手法はPLSAモデルに基づいているため、EMアルゴリズムを用いた場合には初期値依存性という問題がある。そのため、モデル推定結果の妥当性を評価することが課題となる。一般的に、予測モデルを構築した場合、交差検証によって予測精度を評価することでモデルの妥当性を確認できる。しかし、本提案では解析モデルをこのように評価することは困難である。自然言語処理の問題設定では、言語モデルを評価するため、テストデータのPerplexityまたはCross-Entropyがよく使われる。しかし、本章の提案では顧客の嗜好を分析することが目的であり、上位層の優良顧客の数が非常に少ないため、このような評価手法は本章の研究対象となる問題には適さない。そこで、提案手法では、真の確率構造を仮定して、提案モデルの再現性を確認できる人工によるデータシミュレーション実験を行った。そして、今後の研究において、この提案手法を実際の問題に適用し、そ

の推定結果を分析することでモデルの妥当性を評価した。その結果、提案手法の推定精度の面からの有効性、並びに、実データ分析の結果の妥当性を検証することができた。

3.10 本章の結論

本章では、企業が会員ステージランク制を導入する際、顧客の購買履歴データの分析に活用するため、転移学習に基づく PLSA クラスタリングの精度を向上させる新たな手法を提案した。そして、シミュレーション実験において、パラメータ ω と δ をコントロールすることで、多くの状況の人工購買履歴データを生成し、提案手法を行った。評価の段階では、Kullback–Leibler Divergence を評価指標として用いて、設定された真の分布と推定分布の間の距離を比較した。これで提案手法により、特に初期値に基づく PLSA の上位層のクラスタリング結果の推定精度が向上することが検証した。そして、ビジネス施策のため、提案手法を実世界のデータにも適用し、企業の売上向上につながる分析を行った。また、提案手法を用いる場合、従来手法であるランダムな初期値に基づく PLSA モデルより、計算時間が短くなることが確認できた。

今後の課題としては、まずパラメータの推定においてベイズ推定を適用することであり、これにより転移学習によるクラスタリングの精度向上が期待される。そして、提案手法の汎用性を高めるため、提案方法を通常の購買データ分析における優良顧客と非優良顧客の区分に適用することが重要である。また、PLSA モデルで転移学習を適応した場合の結果の解釈可能性を定量化することも検討する必要がある。

第4章 会員ステージ間の潜在クラス分布を比較する購買行動分析モデル

4.1 本章の目的

本章では、顧客の購買行動の分析モデルとして確率的潜在意味解析法（PLSA：Probabilistic Latent Semantic Analysis）を用い、会員ステージ間でクラス分布の差異が分析できるモデルとその学習アルゴリズムを提案する。この手法は全体顧客購買データにより PLSA モデル構築し、得られたパラメータを各ステージの PLSA モデルの導入し、学習を行うアプローチである。本手法の目的はトピックモデルの一つである PLSA モデルを用いてステージ間の顧客の購買傾向変化を分析することである。しかし、PLSA などの潜在クラスモデルを実データを適用する際、ステージごとに独立にモデルを学習した場合、得られた潜在クラスを会員ステージ間で直接比較分析することができない。これは、各会員ステージで全く異なる潜在クラスが独立に作られてしまうためである。これに対して、全ステージのユーザの購買履歴データを1つの潜在クラスモデルでまとめて学習するというアプローチも考えられるが、その場合、得られた潜在クラスには全ステージのユーザが混在するため、ステージ間の顧客購買行動の変化を考察することが困難である。また、ステージが低い顧客の数が上位に比べて圧倒的に多いため、優良顧客の購買特性が学習したモデルに反映されづらいという問題もある。

上記の問題点を解決するため、本提案手法では、ステージ間のユーザの購買行動の差異に着目し、ステージ間で潜在クラスの割合がどのように異なるかを比較可能なモデルとその学習法を提案する。従来の PLSA モデルによる会員ステージ間の顧客の嗜好変化を直接的に比較できないことを提案手法で実現することで、下位

層の顧客の購買行動に対してビジネス施策に結びつけるような提案が考えられる。

4.2 課題と本研究への展開

PLSAなどの潜在クラスモデルを実データを適用する際、ステージ毎に独立にモデルを学習した場合、得られた潜在クラスを会員ステージ間で比較分析することが困難である。これに対して、全ステージのユーザの購買履歴データを1つの潜在クラスモデルでまとめて学習するというアプローチも考えられるが、その場合、得られた潜在クラスには全ステージのユーザが混在するため、ステージ間の顧客購買行動の変化を考察することが困難である。また、ステージが低い顧客の数が上位に比べて圧倒的に多いため、優良顧客の購買特性が学習したモデルに反映され難いという問題もある。上記の問題点を解決するため、本研究では、ステージ間のユーザの購買行動の差異に着目し、ステージ間で潜在クラスの所属割合がどのように異なるかを比較可能なモデルとその学習法を提案する。

4.3 提案手法

4.3.1 概要

提案モデルでは、まず、ステージ間の変化を比較し易くするため、全ステージで潜在クラスの数と同じであることを仮定し、各潜在クラスに属する顧客グループの購買行動自体は会員ステージ間で変わらず、それらの所属割合が変化するモデルを考える。すなわち、全ステージの潜在クラスのもとのアイテムの出現確率は各ステージでも保存されていることを仮定する。これにより、各ステージで潜在クラスの購買の統計的特徴が同一になり、同じ潜在クラス集合上で、ステージ間での出現確率の違いを比較分析することが可能になる。

そのための学習アルゴリズムとして、まず、全ステージの購買履歴データを用い、PLSAにより全ユーザの購買傾向モデルを構築する。全データから推定された潜在クラスの下でのアイテムの出現確率を各ステージのモデルで共通利用する。そして、各ステージのEMアルゴリズムの推定段階で、共通のアイテム出現確率

を固定しながら、各ステージモデルのユーザの出現確率に関するパラメータを推定する。すなわち、各ステージモデルにおいて、各潜在クラスのもとでのアイテムの出現確率が固定されることで、各クラスの意味（解釈）を同一に保ったまま、ユーザの各潜在クラスへの所属確率が推定されることになる。

4.3.2 提案手法の定式化

会員ステージには、 S 段階のステージを設定し、各ステージを $s \in \{1, 2, 3, \dots, S\}$ で表す。3.2 節と同様に、潜在クラスの下でユーザがアイテムを購入する事象を共起の事象として表現するため、全ステージのユーザ集合を $\mathcal{U} = \{u_m : 1 \leq m \leq M\}$ 、 N 個のアイテム集合を $\mathcal{I} = \{i_n : 1 \leq n \leq N\}$ 、全ステージ K 個の潜在クラス集合を $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$ で表し、ステージ s のもとでの K 個の潜在クラス集合も全ステージの潜在クラスと同じであるとする。ステージ s に属する M_s 人のユーザ集合を $\mathcal{U}_s = \{u_m^s : 1 \leq m \leq M_s\}$ と表す。ここで、 $M = M_1 + M_2 + \dots + M_S$ 、 $\mathcal{U}_s \subseteq \mathcal{U}$ であり、全ステージのユーザをまとめる場合、ユーザをステージの番号に従って昇順に並べるものとする。また、ステージ s における潜在クラスの生起確率と潜在クラス下でユーザの生起確率を $P_s(z_k)$ と $P_s(u_m^s|z_k)$ で表す。

提案手法の具体的な手順を以下に示す。また、各手順のイメージを図 4.1 ~ 図 4.2 に示す：

「提案手法手順 1」：全ステージデータを用い、PLSA モデルでユーザとアイテムの共起モデルを構築する。そして、構築されたモデルからパラメータ $\hat{P}(i_n|z_k)$ を推定する。

「提案手法手順 2」：手順 1 から推定されたパラメータ $\hat{P}(i_n|z_k)$ を各ステージの比較モデルに適用し、これらの値を固定しながら、各ステージパラメータ $\hat{P}_s(z_k)$ 、 $\hat{P}_s(u_m^s|z_k)$ を推定する。

4.3.3 全ステージデータに対するユーザとアイテムの共起モデル

全ステージの全データを対象としたユーザとアイテムの共起モデル（以下、全ステージユーザモデル）では、ユーザ $u_m \in \mathcal{U}$ がアイテム $i_n \in \mathcal{I}$ を購入した事象

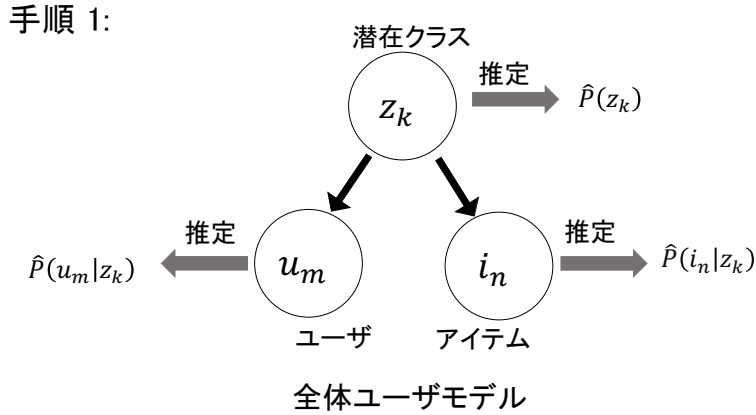


図 4.1: 提案手方法の手順 1

を (u_m, i_n) と記述する．この時，モデルの式は 3.2 節と同様に式 (4.1) ように表すことができる．そして，全ステージユーザモデルにより，EM アルゴリズムを用いて全ユーザの購買履歴データから $u_m \in \mathcal{U}$ ， $i_n \in \mathcal{I}$ ， $z_k \in \mathcal{Z}$ に対し $\hat{P}(u_m|z_k)$ ， $\hat{P}(i_n|z_k)$ ， $\hat{P}(z_k)$ を推定する．推定式は式 (4.1)~ 式 (4.1) に示す．

$$P(u_m, i_n) = \sum_{k=1}^K P(z_k)P(u_m|z_k)P(i_n|z_k). \quad (4.1)$$

EM アルゴリズム:

[E-Step]: ユーザとアイテムに関数するパラメータを固定したときの潜在変数 z_k の分布 $P(z_k|u_m, i_n)$ を式 (4.2) により算出する．

$$P(z_k|u_m, i_n) = \frac{P(z_k)P(i_n|z_k)P(u_m|z_k)}{\sum_{k=1}^K P(z_k)P(i_n|z_k)P(u_m|z_k)} \quad (4.2)$$

[M-Step]: E-Step で算出された $P(z_k|u_m, i_n)$ を用いてそれぞれのパラメータを算出する．対数尤度 LL を最大化するパラメータを導出するため，対数尤度を以下のように展開する．

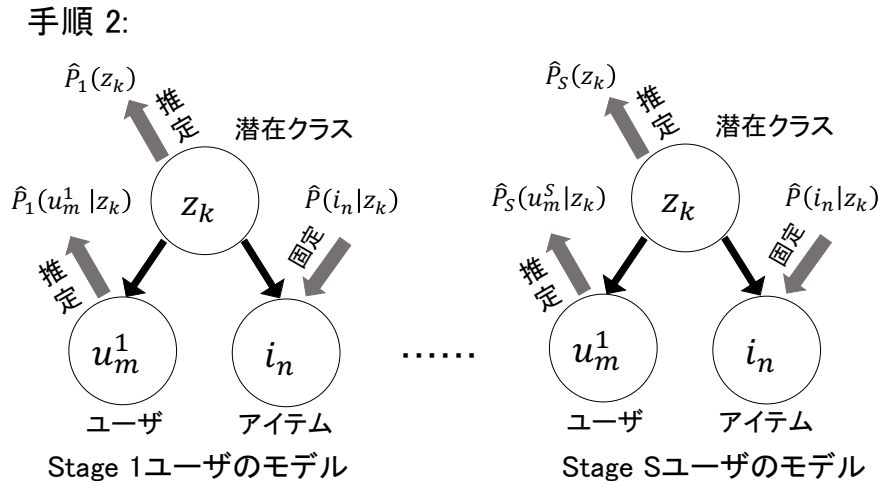


図 4.2: 提案手方法の手順 2

$$LL = \sum_{m=1}^M \sum_{n=1}^N \log P(u_m, i_n) \quad (4.3)$$

$$= \sum_{m=1}^M \sum_{n=1}^N \log \left(\sum_{k=1}^K P(z_k) \cdot P(i_n|z_k)P(u_m|z_k) \right) \quad (4.4)$$

$$= \sum_{m=1}^M \sum_{n=1}^N \log \left\{ \sum_{k=1}^K P(z_k|u_m, i_n) \frac{P(z_k)P(i_n|z_k)P(u_m|z_k)}{P(z_k|u_m, i_n)} \right\} \quad (4.5)$$

$$\geq \sum_{m=1}^M \sum_{n=1}^N \left\{ \sum_{k=1}^K P(z_k|u_m, i_n) \log \frac{P(z_k)P(i_n|z_k)P(u_m|z_k)}{P(z_k|u_m, i_n)} \right\} \quad (4.6)$$

$$= \sum_{m=1}^M \sum_{n=1}^N \sum_{k=1}^K P(z_k|u_m, i_n) \left(\log P(z_k)P(i_n|z_k)P(u_m|z_k) - \log P(z_k|u_m, i_n) \right) \quad (4.7)$$

式(4.3)は式(4.5)から式(4.6)により, Jensen の不等式を用いて展開する. LL' は式(4.7)の右辺を変形し, 定数項を省略した式である. EM アルゴリズムでは, これを最大化する.

$$LL' = \sum_{m=1}^M \sum_{n=1}^N \left\{ \sum_{k=1}^K P(z_k|u_m, i_n) \log P(u_m|z_k)P(i_n|z_k)P(z_k) \right\} \quad (4.8)$$

さらに、式(4.8)を最大化するために、ラグランジュの未定乗数法を用いる。 λ_k , η_k , π をそれぞれ未定乗数として $k = 1, 2, \dots, K$, ラグランジュ関数を以下の式(4.9)のように定義する。

$$g = LL' + \sum_{k=1}^K \lambda_k \left(1 - \sum_{m=1}^M P(u_m|z_k)\right) + \sum_{k=1}^K \eta_k \left(1 - \sum_{n=1}^N P(i_n|z_k)\right) + \pi \left(1 - \sum_{k=1}^K P(z_k)\right) \quad (4.9)$$

これを偏微分して0とおくことにより、それぞれのパラメータの更新式は式(4.10)～式(4.12)のように表すことができる。

[M-Step]:

$$P(u_m|z_k) = \frac{\sum_{n=1}^N P(z_k|u_m, i_n)}{\sum_{m=1}^M \sum_{n=1}^N P(z_k|u_m, i_n)} \quad (4.10)$$

$$P(i_n|z_k) = \frac{\sum_{m=1}^M P(z_k|u_m, i_n)}{\sum_{m=1}^M \sum_{n=1}^N P(z_k|u_m, i_n)} \quad (4.11)$$

$$P(z_k) = \frac{\sum_{m=1}^M \sum_{n=1}^N P(z_k|u_m, i_n)}{\sum_{k=1}^K \sum_{m=1}^M \sum_{n=1}^N P(z_k|u_m, i_n)} \quad (4.12)$$

以上のように対数尤度関数が収束するまでE-StepとM-Stepを繰り返し行う。ここで、対数尤度の変化率が閾値以下であれば（例えば、1%など）、モデルの学習は収束したと判断する。そうでない場合は、EMアルゴリズムの更新を続ける。また、EMアルゴリズムの更新回数を事前に決めて学習を行う方法もある。

4.3.4 各ステージにおけるユーザ比較モデル構築

各ステージのユーザモデルでは、会員ステージ s に属するユーザ $u_m^s \in \mathcal{U}_s$ がアイテム $i_n \in \mathcal{I}$ を購入した事象を (u_m^s, i_n) と記述する。本研究で対象とする事例では、

「ユーザは必ずいずれか1つのステージに属する」ものとする。よって、各ステージのユーザは重複することがない。一方で、ステージのユーザは異なるが、アイテムは全てのステージで共通である。そこで、各ステージでのユーザ出現確率が比較可能なモデルを作るため、全ステージの潜在クラスのもとでのアイテムの出現確率と各ステージの潜在クラスのもとでのアイテムの出現確率が同じであることを仮定する。従って、4.3.3節で推定された全ステージモデルから推定されたパラメータ $\hat{P}(i_n|z_k)$ を各ステージモデルの初期値として各ステージのモデルに導入する。モデルは以下の式 (4.13) のように表現することができる。

$$P_s(u_m^s, i_n) = \sum_{k=1}^K P_s(z_k) \hat{P}(i_n|z_k) P_s(u_m^s|z_k). \quad (4.13)$$

ここで、ユーザ u_m^s におけるアイテム i_n の実際の購買回数を $x(u_m^s, i_n)$ とすると、ステージ s の学習データに対する対数尤度関数 LL_s は以下の式 (4.14) のように示される。

$$LL_s = \sum_{m=1}^{M_s} \sum_{n=1}^N x(u_m^s, i_n) \log P_s(u_m^s, i_n). \quad (4.14)$$

ここで、全ステージのユーザのデータを用いて推定したい潜在クラスごとのアイテムの生起確率を固定しながら、各ステージのユーザの潜在クラスにおける出現確率を推定するため、EM アルゴリズムをベースとした以下の学習アルゴリズムを提案する。

各ステージ s のパラメータ更新ステップは以下のようになる：

提案アルゴリズム：

「**Step 1**」： 各パラメータ $P_s(z_k)$, $P_s(u_m^s|z_k)$ の初期値は同じ乱数を用いて生成する。アイテムが潜在クラスに属する確率の初期値は4.3.3節で全ステージモデルの推定による $\hat{P}(i_n|z_k)$ を用いる。

「**Step 2**」： EM アルゴリズムにより、パラメータ $\hat{P}(i_n|z_k)$ の値を固定しながら、パラメータ $P_s(z_k)$, $P_s(u_m^s|z_k)$ を更新する。

[E-Step]:

$$P_s(z_k|u_m^s, i_n) = \frac{P_s(z_k)\hat{P}(i_n|z_k)P_s(u_m^s|z_k)}{\sum_{k=1}^K P_s(z_k)\hat{P}(i_n|z_k)P_s(u_m^s|z_k)}. \quad (4.15)$$

[M-Step]:

$$P_s(u_m^s|z_k) = \frac{\sum_{n=1}^N x(u_m^s, i_n)P_s(z_k|u_m^s, i_n)}{\sum_{m=1}^{M_s} \sum_{n=1}^N x(u_m^s, i_n)P_s(z_k|u_m^s, i_n)}, \quad (4.16)$$

$$P_s(z_k) = \frac{\sum_{m=1}^{M_s} \sum_{n=1}^N x(u_m^s, i_n)P_s(z_k|u_m^s, i_n)}{\sum_{k=1}^K \sum_{m=1}^{M_s} \sum_{n=1}^N x(u_m^s, i_n)P_s(z_k|u_m^s, i_n)}. \quad (4.17)$$

ここで、この各ステージでEM アルゴリズムでは $\hat{P}(i_n|z_k)$ の値は固定し、更新しない。また、収束判定条件は 3.2 節と同様に対数尤度の変化率が閾値以下であれば（例えば、1% など）、モデルの学習は収束したと判断する。そうでない場合は、EM アルゴリズムの更新を続ける。また、EM アルゴリズムの更新回数を事前に決めて学習を行う方法もある。

4.4 人工データによるシミュレーション実験

4.4.1 生成モデルの各パラメータの設定方法

本章では、会員ステージが上がるにつれて購買行動が徐々に変化していくようなデータを疑似的に生成し、シミュレーション実験による評価を行う。提案手法で学習させた結果を分析することで、提案手法の有効性を検証すると共に、その特性を明らかにする。疑似的に生成するデータの特徴を、実際の購買履歴データの特徴に類似させるため、以下のような特徴を持ったデータを生成した。

1. 会員ステージが異なっても、ユーザの潜在的なグループ（潜在クラス）の特徴自体は保存される（会員ステージが上がった際に、全く嗜好の異なるユーザグループが突然、現れることはない）。
2. 会員ステージが上がるに従い、所属するユーザ数は少なくなる。

3. 会員ステージが上がるユーザの嗜好の多様性が増やす（様々なアイテムを購入するようになる）。

以上のような特性を有したデータを生成するために、

- 各会員ステージに対して、予め現実のユーザ数を模倣したユーザ数を割り当てる。
- 潜在クラスに相当するユーザグループを G 個設定し、これらのユーザグループごとに購買確率が高いアイテム群が異なるものとする。
- この G 個のユーザグループの特徴（各アイテムの購入確率）は会員ステージ間で不変とする。
- 会員ステージの向上と共に、これらのグループへのユーザの所属確率を、偏りの大きい分布から平坦な分布に変化させていくことで、購買の多様性が増大する状況を再現する。

以上のような特性を持った確率分布に従うデータを、各会員ステージに対し生成するための生成モデルのパラメータを以下のように定義する。

本研究の対象事例では、各ステージに対して複数の異なる購買嗜好を持つユーザが混在している。また、各ステージで、グループの数とグループに属する各アイテムの購入確率は同じと仮定する。すなわち、ステージの数 (S 個) と共通のグループの数 (G 個) を設定する。ステージごとの各グループ A_g の生起確率 $P_s(A_g)$ を $\alpha_{(s,g)}$ と表す。また、 $\alpha_{(s,g)}$ は $\sum_{g=1}^G P_s(A_g) = \sum_{g=1}^G \alpha_{(s,g)} = 1$ をものとする。各会員ステージ s と各グループ A_g の生起確率の関係を表 4.1 に示す。

表 4.2 では、アイテム i_n とグループ A_g の生起確率を定義している。ステージごとのアイテムは同じであるため、全ステージに対して同じ設定を行う。ここで、 N 種類のアイテム中の一部を人気商品としてそれぞれ均等に G 個のグループに分割し、各グループに所属するアイテム数を Q 種類とする ($N > Q$)。同じグループに所属するアイテム中の指定された一部のアイテムを購入しやすいアイテム（似た購買傾向を持つ）と仮定する。そのために、購買されやすいアイテムの確率の和を β に設定する。すなわち、1つの購買されやすいアイテムに対する確率は $P(i_n|A_g) = \frac{\beta}{Q}$ と設定

表 4.1: 各会員ステージ s における各グループ A_g の生起確率

ステージ名	A_1	A_2	...	A_G
Stage 1	$P_1(A_1) = \alpha_{(1,1)}$	$P_1(A_2) = \alpha_{(1,2)}$...	$P_1(A_G) = \alpha_{(1,G)}$
Stage 2	$P_2(A_1) = \alpha_{(2,1)}$	$P_2(A_2) = \alpha_{(2,2)}$...	$P_2(A_G) = \alpha_{(2,G)}$
...
Stage S	$P_S(A_1) = \alpha_{(S,1)}$	$P_S(A_2) = \alpha_{(S,2)}$...	$P_S(A_G) = \alpha_{(S,G)}$

する。それ以外のアイテムごとの購買確率を $P(i_n|A_g) = \frac{1-\beta}{N-Q}$ と設定する。 $P(i_n|A_g)$ は $\sum_{n=1}^N P(i_n|A_g) = 1$ を満たす。

最後に、ステージごとのグループからユーザの生起確率を定義する。ユーザ数は会員ステージが上がるほど少なくなることが一般的であるため ($M_1 > M_2 > \dots > M_S$)、ユーザ数の方は実データに近い設定でデータを生成するものとする。ステージ間の人数が異なるため、グループからユーザの生成確率も異なる。表 4.1 の各グループの生起確率により、ユーザの生起確率を一様分布だと仮定する。ここで、 $P_s(u_m^s|A_g)$ は $\sum_{m=1}^{M_s} P_s(u_m^s|A_g) = 1$ を満たす。従って、式 (4.18) により、各ステージにおけるユーザのグループの所属確率 $P_s(A_g|u_m^s)$ をベイズの定理により算出できる。さらに、各潜在推定された $P_s(A_g|u_m^s)$ でクラスに所属する人数も計算できる。

$$P_s(A_g|u_m^s) = \frac{P_s(A_g)P_s(u_m^s|A_g)}{P_s(u_m^s)} = \frac{\frac{\alpha_{(s,g)}}{M_s}}{\frac{1}{M_s}} = \alpha_{(s,g)}. \quad (4.18)$$

4.4.2 人工購買履歴データの生成手順

購買履歴の人工データは以下の流れで生成する。

人工データ生成手順：

「手順 1」： 1つのステージ s を選択する。

「手順 2」： 選択されたステージ s により、設定された $\alpha_{(s,g)}$ の確率でステージのもとで1つのグループ A_g を選択する。

表 4.2: 各グループとアイテムの生起確率

		アイテム				
		$i_1, i_2, \dots, i_Q,$ 購買されやすい アイテムグループ $P(i_n A_1) = \frac{\beta}{Q}$	$i_{Q+1}, i_{Q+2}, \dots, i_{2Q},$ $P(i_n A_1) = \frac{1-\beta}{N-Q}$...	$i_{(G-1)Q+1}, i_{(G-1)Q+2}, \dots, i_{GQ},$ $P(i_n A_1) = \frac{1-\beta}{N-Q}$	$i_{GQ+1}, i_{GQ+2}, \dots, i_N$ $P(i_n A_1) = \frac{1-\beta}{N-Q}$
グループ	A_1		購買されやすい アイテムグループ $P(i_n A_2) = \frac{\beta}{Q}$...		
	A_2			
	...				購買されやすい アイテムグループ $P(i_n A_G) = \frac{\beta}{Q}$	
	A_G					

「手順3」： 選択されたステージ s とグループ A_g のもとで、そのステージのユーザの中からランダムに $(1/M_s)$ の確率で1人のユーザを選択する。

「手順4」： 選択されたグループ A_g に対し、購買されやすいアイテムは $\frac{\beta}{Q}$ の確率で、それ以外が $\frac{1-\beta}{N-Q}$ の確率で購買されると仮定した分布に従い、1つのアイテムを選択する。そして、ユーザ u_m^s に対して、1つアイテムを生成すると、ユーザ u_m^s がそのアイテムを購入したという結果をデータに加える。手順2に戻り、ステージ s の全ユーザの総購買数（ステージ s のユーザのアイテム平均購買数 \times ステージ s のユーザ数）になるまで繰り返す。

以上の手順により、各ステージのユーザ u_m^s は、自身が所属する1つのグループに対応する「購入されやすいアイテム」に対して購買傾向が強いが、それ以外のアイテムについてもある程度の割合で購入するような購買履歴データが生成される。

4.4.3 シミュレーション条件

本稿では、ユーザのステージの成長により、グループの出現確率が変化する状況を真のモデルとして設定する。具体的には、ステージは Stage 1 から Stage 5 まで5段階、グループ数も5である ($G = 5$)。ステージにより、ユーザ数とユーザのアイテム平均購買数は株式会社良品計画の状況 (表 2.1) を参考とし、表 4.3 のように設定する。グループの特徴は表 4.4 のように設定した。具体的には、グループ A_2 , A_3 , A_4 は購買変化が小さいグループ、 A_1 がステージの成長により出現確率が小さくなるグループ、 A_5 がステージの成長により出現確率が大きくなるグループと設定した。また、 $\beta = 0.99$, $Q = 10$, アイテム種類数 N を 500 と設定する。

表 4.3: ステージごとのユーザ数とアイテム平均購買数の設定

ステージ名	人数 (人)	アイテム平均購買数 (件)
Stage 5	1,000	105
Stage 4	2,000	75
Stage 3	4,500	50
Stage 2	12,500	30
Stage 1	60,000	15

表 4.4: $P_s(A_g)$ の設定

ステージ名	$P_s(A_1)$	$P_s(A_2)$	$P_s(A_3)$	$P_s(A_4)$	$P_s(A_5)$
Stage 5	0.2	0.05	0.1	0.15	0.5
Stage 4	0.3	0.1	0.1	0.1	0.4
Stage 3	0.4	0.15	0.1	0.05	0.3
Stage 2	0.5	0.1	0.1	0.1	0.2
Stage 1	0.6	0.1	0.1	0.1	0.1

生成した人工データに対し、提案手法を適用する。各ステージのモデルの学習を行う際に、 $P_s(z_k)$ と、 $P_s(u_m^s|z_k)$ の初期値を乱数で設定する。パラメータ推定の収束条件は、対数尤度関数の値の変化率が閾値より小さくなったときとした。そして、複数回の試行を行い、対数尤度が最大になった場合の結果を用いる。

4.4.4 シミュレーション結果

ベイズ定理により、従来手法（手順1のみ）で推定されたパラメータ $\hat{P}(z_k)$ と $\hat{P}(u_m|z_k)$ を用い、式 (4.19)~式 (4.21) により、ユーザを分けて各ステージの潜在クラスの出現確率を推定した。推定結果を表 4.5 に示す。 \hat{P} は従来手法で計算された確率（パラメータ）であることを示す。

$s = 1$ のとき:

$$\hat{P}_1(z_k) = \frac{\sum_{m=1}^{M_1} \hat{P}(z_k) \hat{P}(u_m|z_k)}{\sum_{m=1}^{M_1} \sum_{k=1}^K \hat{P}(z_k) \hat{P}(u_m|z_k)}. \quad (4.19)$$

$s = 2$ のとき:

$$\tilde{P}_2(z_k) = \frac{\sum_{m=M_1+1}^{M_1+M_2} \hat{P}(z_k) \hat{P}(u_m|z_k)}{\sum_{m=M_1+1}^{M_1+M_2} \sum_{k=1}^K \hat{P}(z_k) \hat{P}(u_m|z_k)}. \quad (4.20)$$

$s > 2$ のとき:

$$\tilde{P}_s(z_k) = \frac{\sum_{m=M_1+\dots+M_{s-1}+1}^{M_1+\dots+M_s} \hat{P}(z_k) \hat{P}(u_m|z_k)}{\sum_{m=M_1+\dots+M_{s-1}+1}^{M_1+\dots+M_s} \sum_{k=1}^K \hat{P}(z_k) \hat{P}(u_m|z_k)}. \quad (4.21)$$

提案手法で人工購買履歴データから $\hat{P}_s(z_k)$ を推定した結果を表 4.6 に示す。ただし、 \hat{P} は推定された確率（パラメータ）であることを示す。

表 4.5: 従来手法による $P_s(z_k)$ の推定結果 $\tilde{P}_s(z_k)$

ステージ名	$\tilde{P}_s(z_1)$	$\tilde{P}_s(z_2)$	$\tilde{P}_s(z_3)$	$\tilde{P}_s(z_4)$	$\tilde{P}_s(z_5)$
Stage 5	0.1999	0.2000	0.1581	0.1797	0.2623
Stage 4	0.2000	0.1999	0.1589	0.1799	0.2612
Stage 3	0.2000	0.2001	0.1594	0.1804	0.2601
Stage 2	0.2001	0.2001	0.1586	0.1803	0.2609
Stage 1	0.2005	0.1999	0.1590	0.1796	0.2610

表 4.6: 提案手法による $P_s(z_k)$ の推定結果 $\hat{P}_s(z_k)$

ステージ名	$\hat{P}_s(z_1)$	$\hat{P}_s(z_2)$	$\hat{P}_s(z_3)$	$\hat{P}_s(z_4)$	$\hat{P}_s(z_5)$
Stage 5	0.1993	0.0450	0.1505	0.0750	0.5302
Stage 4	0.3012	0.0897	0.0994	0.0795	0.4303
Stage 3	0.3997	0.1353	0.0500	0.0848	0.3303
Stage 2	0.4982	0.0908	0.0998	0.0796	0.2315
Stage 1	0.5997	0.0898	0.1004	0.0796	0.1304

$$D_{KL}(P_s(A_g) \parallel \tilde{P}_s(z_k)) = \sum_{g=k=1}^K P_s(A_g) \log \frac{P_s(A_g)}{\tilde{P}_s(z_k)}. \quad (4.22)$$

$$D_{KL}(P_s(A_g) \parallel \hat{P}_s(z_k)) = \sum_{g=k=1}^K P_s(A_g) \log \frac{P_s(A_g)}{\hat{P}_s(z_k)}. \quad (4.23)$$

表 4.7: 交差エントロピー誤差の比較

	従来手法	提案手法
Stage 5	1.8076	1.3119
Stage 4	1.7205	1.3754
Stage 3	1.7831	1.3898
Stage 2	1.7997	1.3401
Stage 1	1.9433	1.2278

まず、各潜在クラスの真の生起確率と推定された潜在クラスにおける生起確率のパラメータの値を元に、事前設定のクラスと推定したクラスを対応させる。次に、交差エントロピー誤差の計算式(式(4.22)~式(4.23))により、従来手法と提案手法で推定されたモデルのパラメータ $\tilde{P}_s(z_k)$, $\hat{P}_s(z_k)$ と設定された真のパラメータ $P_s(A_g)$ の値と比較する。ここで、異なる2つのモデルにおける潜在クラス同士を推定された潜在クラスの生起確率から比較するため、 $z_k = A_g$ (潜在クラスと真のグループ) とする。すなわち、 $k = g$ と設定する。このときの比較結果を表4.7に示す。真の分布と従来手法で推定された分布のエントロピーが提案手法より大きいため、提案手法の有効性が示された。従来手法では全体ステージの影響により、ステージ間での潜在クラスの確率の変化は検出できない。また、ステージごとにモデルを独立で構築し、推定されたパラメータも異なるため、ステージ間での特徴の比較も困難である。

設定された各グループの出現確率(表4.4)と実際に推定された各潜在クラスの出現確率(表4.6)により、パラメータ $\hat{P}(i_m|z_k)$ が固定されたことで、提案手法で推

定された潜在クラスの出現確率 $\hat{P}_s(z_k)$ が設定された潜在クラスの真の確率 $P_s(A_g)$ を捉えることができるようになったと言える。すなわち、提案手法により、各ステージのパラメータの推定が可能になっている。例えば、潜在クラス1はステージの成長に従って出現確率が小さくなるクラスタである。潜在クラス5はステージの成長により出現確率が大きくなるクラスタである。他の潜在クラスはステージ成長により、購買傾向の変動が小さいクラスタである。以上により、提案手法の確率構造の推定という観点での有効性を示すことができた。

4.5 実データ分析

4.5.1 分析対象データ

本章では、株式会社良品計画から提供頂いた会員登録購買履歴データを分析対象とする。無印良品の会員ステージランクと顧客年間累計購買金額概要は表2.1に示した通りである。分析の対象期間は2014年3月1日～2015年2月28日である。本実験では、アイテムが672種類の商品カテゴリ(仮に、「アイテム1」、「アイテム2」などと示す)に分類されており、本研究ではこの商品カテゴリを用いて分析を行う。また、商品カテゴリの金額帯はおおよその金額で示す。分析対象データ数は合計約1100万件である。

4.5.2 実験条件

本研究における潜在クラス数は、主に各潜在クラスの解釈性と想定するビジネス施策の粒度という観点から適切に設定する[付録A.1]。一般に確率モデルの構造やパラメータ数については、予測精度や分類精度という観点から、情報量基準やクロスバリデーション等によって客観的に決定する場合も多いが、これらのケースの主たる目的は予測精度の高いモデルを得ることにある。本研究で対象とするクラスターリング問題では、潜在クラスとして抽出されたクラスタごとの特徴を分析し、実施可能なビジネス施策を検討するための結果を得ることを目的とするため、解釈性を重視した潜在クラス数の決定を行った。本研究では、潜在クラスの解釈性とビジネス施策検討の粒度の観点から、適用するPLSAモデルの潜在クラ

ス数は10と設定し、分析を行う。なお、潜在クラス数が増えた場合の挙動についても確認するため、併せて潜在クラス数 $K = 20$ の場合についても結果を示す。また、本研究において対象とする購買履歴では、単価が高い商品を購入することでステージが一度に2つランク以上アップするユーザは少数となっているため、これらのユーザに対して段階的なステップアップを目指した施策において対象外と判断する。そして、良品計画の会員ステージランク制度の下では、ステージが上がるにつれ、1つ上のステージへと成長するために求められる年間累積金額の増加率が大きくなる(表2.1)。ステージが成長すれば成長するほど、ユーザのステージ成長が難しくなるため、ユーザが急に低ステージから高ステージまで成長させる有効な施策を提案することは難しい。また、良品計画で扱っているメイン商品は生活用品と食品のため、企業側は定期的に商品を購入しながら、高ステージへ徐々に成長するユーザが望ましいと考えられる。本研究では、段階的にユーザへ施策を目指し、モデルを構築する際に、Stage 1 から1ランクずつ、ステップアップしてきたユーザを対象にした。

以下の節では、潜在クラス数を10とした場合と20とした場合について、それぞれ節を分けて結果を示す。

4.5.3 提案モデルの推定結果(潜在クラス数が10の場合)

ここでは、「潜在クラスごとに特徴の差異が明になること」、かつ「ステージ間の変化も分析が可能であること」の観点により、潜在クラス数を $K = 10$ とした場合の結果を示す。

潜在クラス数が10の場合の結果分析

提案手法で推定されたパラメータ $\hat{P}_s(z_k)$ の推定結果を表4.9を用いて示す。各ステージから得られるパラメータ $\hat{P}_s(u_m^s|z_k)$, $\hat{P}(i_n|z_k)$ から、ベイズ定理により、 $\hat{P}_s(z_k|u_m^s)$, $\hat{P}(z_k|i_n)$ を計算する。 $\sum_{m=1}^{M_s} \hat{P}_s(z_k|u_m^s)$ と $\hat{P}(z_k|i_n)$ の推定結果を表4.10, 表4.11示す。ここで、 $\hat{P}_s(z_k)$ は各ステージの潜在クラス出現確率の推定値である。 $\hat{P}(z_k|i_n)$ はあるアイテムの下で各クラスの出現確率である。

表 4.9, 表 4.10, 表 4.11 より, 各ステージのユーザの嗜好の特徴を把握することができる.

1. 表 4.9 より, ステージの成長によりステージ間の変化小さいが, ステージが成長すれば, ユーザの購買範囲が広がっていく.
2. 表 4.9, 4.10 より, $\hat{P}_s(z_k)$ の大きい方が人数も多い.
3. 表 4.11 より, クラス 1, クラス 2 の代表的なアイテムが他のクラスより少ない. クラス 3, クラス 4, クラス 8, クラス 9 で多様なカテゴリのアイテムを購入するユーザが, 各ステージに一定の割合で存在する.
4. ステージの成長により, クラス 4 の割合が少なくなる. クラス 3, クラス 9 では, ステージが高ければ高いほど, 所属する人数共起事象の割合が増加する.
5. 表 4.9, 表 4.11 より, クラス 9 で順位 5 までのアイテムは 1 に近い出現確率を持つため, クラス 9 の代表アイテムと言える.

ステージの成長による変化 (潜在クラス数が 10 の場合)

ステージの成長によるユーザの購買多様性を検証するため, 提案手法で推定されたパラメータ $\hat{P}_s(z_k|u_m^s)$ を用いて各ステージのユーザが各潜在クラスに所属する確率のエントロピーの平均を計算する. エントロピーの計算式を式 (4.24) に示す.

$$H_s(z) = -\frac{1}{M_s} \sum_{m=1}^{M_s} \sum_{k=1}^K \hat{P}_s(z_k|u_m^s) \log \hat{P}_s(z_k|u_m^s). \quad (4.24)$$

各ステージのユーザが各潜在クラスに所属する確率の平均エントロピーの変化を表 4.8 に示す. 表 4.8 により, ステージの成長によって各潜在クラスに所属する確率の平均エントロピーの値は大きくなっている. すなわち, ステージの成長により, ユーザの嗜好の多様性 (購買範囲) が増大していることが分かる. この様に, 提案手法により, ステージ間のユーザの嗜好の変化を比較することができる.

表 4.8: 各ステージのユーザが各潜在クラスに所属する確率のエントロピーの平均 (潜在クラス数が 10 の場合)

ステージ名 (降順)	エントロピー
ダイヤモンド (Stage 5)	0.0549
プラチナ (Stage 4)	0.0121
ゴールド (Stage 3)	0.0030
シルバー (Stage 2)	0.0009
なし (Stage 1)	0.0002

4.5.4 提案モデルの推定結果 (潜在クラス数が 20 の場合)

前節では、潜在クラス数を $K = 10$ とした場合の結果であったが、この潜在クラス数を変化させた場合の結果については、提案モデルの評価の観点からも確認しておくことは肝要である。そこで本節では、潜在クラスごとの特徴が明らかになっており、かつステージ間での特徴変化も分析できるため、潜在クラス数を $K = 20$ とした場合の結果を示す。

潜在クラス数が 20 の場合の結果分析

4.5.3 節の同様に、提案手法で推定されたパラメータ $\hat{P}_s(z_k)$ の推定結果を表 4.12 を用いて示す。各ステージから得られるパラメータ $\hat{P}_s(u_m^s|z_k)$, $\hat{P}(i_n|z_k)$, から、ベイズ定理により、 $\hat{P}_s(z_k|u_m^s)$, $\hat{P}(z_k|i_n)$ を計算する。 $\sum_{m=1}^{M_s} \hat{P}_s(z_k|u_m^s)$ と $\hat{P}(z_k|i_n)$ の推定結果を表 4.13, 表 4.14 示す。

表 4.12, 表 4.13, 表 4.14 より、各ステージのユーザの嗜好の特徴を把握することができる。

1. 表 4.12 より、ステージの成長によりステージ間の変化が大きいクラスと変化がないクラスの 2 種類がある。
2. 表 4.9, 表 4.12 より、潜在クラスが多くなると、各ステージ間の潜在クラスの詳細な変化を比較できるようになった。

表 4.9: 全ステージから推定された $\hat{P}(z_k)$ と提案手法による潜在クラス生起確率 $\hat{P}_s(z_k)$ の推定結果 (無印良品の実データのケース) (潜在クラス数が 10 の場合)

ステージ名	$\hat{P}_s(z_1)$	$\hat{P}_s(z_2)$	$\hat{P}_s(z_3)$	$\hat{P}_s(z_4)$	$\hat{P}_s(z_5)$	$\hat{P}_s(z_6)$	$\hat{P}_s(z_7)$	$\hat{P}_s(z_8)$	$\hat{P}_s(z_9)$	$\hat{P}_s(z_{10})$
All Stage(全ステージ)	0.0009	0.0003	0.1245	0.2916	0.0228	0.0653	0.1008	0.2049	0.1652	0.0238
Stage 5(ダイヤモンド)	0.0018	0.0070	0.1472	0.2552	0.0271	0.0491	0.0888	0.2147	0.1755	0.0336
Stage 4(プラチナ)	0.0010	0.0022	0.1357	0.2646	0.0275	0.0541	0.0972	0.2060	0.1799	0.0318
Stage 3(ゴールド)	0.0014	0.0014	0.1289	0.2735	0.0277	0.0587	0.0975	0.2039	0.1780	0.0290
Stage 2(シルバー)	0.0008	0.0010	0.1233	0.2861	0.0254	0.0641	0.1008	0.2042	0.1695	0.0248
Stage 1(なし)	0.0007	0.0005	0.1205	0.3057	0.0203	0.0702	0.1026	0.2051	0.1559	0.0185

表 4.10: 無印良品の実データによるステージごとと全てのユーザーが各潜在クラスにおいて出現確率 ($\sum_{m=1}^M \hat{P}_s(z_k | u_m^s)$) の推定結果 (潜在クラス数が 10 の場合)

	クラス 1	クラス 2	クラス 3	クラス 4	クラス 5
Stage 5(ダイヤモンド)	2.70	3.56	155.12	266.53	31.70
Stage 4(プラチナ)	7.59	9.54	852.29	1572.19	190.60
Stage 3(ゴールド)	32.38	43.13	3890.37	7681.13	978.99
Stage 2(シルバー)	112.69	144.07	14130.37	30029.13	3627.39
Stage 1(なし)	594.36	482.22	74417.03	161736.13	13523.83
	クラス 6	クラス 7	クラス 8	クラス 9	クラス 10
Stage 5(ダイヤモンド)	52.23	101.10	224.18	203.42	40.45
Stage 4(プラチナ)	319.94	621.48	1260.17	1180.19	216.01
Stage 3(ゴールド)	1633.44	2994.49	5906.91	5520.53	941.64
Stage 2(シルバー)	6769.46	11592.62	22533.11	20178.49	3169.67
Stage 1(なし)	39258.55	62526.76	116877.99	93217.21	12117.93

表 4.11: 各潜在クラスの代表的なアイテムの出現確率ランキング (潜在クラス数が 10 の場合)

順位	クラス 1 のアイテム	$\hat{P}(z_{1} i_n)$	価格帯	クラス 2 のアイテム	$\hat{P}(z_{2} i_n)$	価格帯	クラス 3 のアイテム	$\hat{P}(z_{3} i_n)$	価格帯
1	衣服・雑貨 79	0.9999	80300 円	生活雑貨 146	0.7905	84260~89820 円	生活雑貨 215	0.9999	4480~4510 円
2	生活雑貨 86	0.9999	3100~3170 円	生活雑貨 143	0.0428	900~2290 円	生活雑貨 36	0.9999	23210~23810 円
3	生活雑貨 109	0.9969	39920~41670 円	生活雑貨 1	0.0098	3630~3650 円	生活雑貨 199	0.9999	6010~6110 円
4	生活雑貨 146	0.2095	84260~98920 円	生活雑貨 84	0.0081	290~430 円	生活雑貨 208	0.9999	8680~9510 円
5	衣服・雑貨 245	0.0012	0~2760 円	生活雑貨 91	0.0060	40640~96190 円	生活雑貨 13	0.9999	1420~1430 円

順位	クラス 4 のアイテム	$\hat{P}(z_{4} i_n)$	価格帯	クラス 5 のアイテム	$\hat{P}(z_{5} i_n)$	価格帯	クラス 6 のアイテム	$\hat{P}(z_{6} i_n)$	価格帯
1	食品 14	0.9999	1700~1800 円	生活雑貨 178	0.9999	2060~2100 円	サーブイス 1	0.9999	23780~23810 円
2	食品 9	0.9999	850~1110 円	生活雑貨 193	0.9999	6330~6520 円	生活雑貨 14	0.9999	2320~2380 円
3	食品 71	0.9999	300~350 円	生活雑貨 48	0.9999	150410~153810 円	生活雑貨 207	0.9999	850~880 円
4	生活雑貨 213	0.9999	2760~3690 円	生活雑貨 151	0.9999	2810~2860 円	衣服・雑貨 3	0.9999	1300~1390 円
5	食品 74	0.9999	140~190 円	衣服・雑貨 344	0.9999	2340~2760 円	生活雑貨 30	0.9999	660~740 円

順位	クラス 7 のアイテム	$\hat{P}(z_{7} i_n)$	価格帯	クラス 8 のアイテム	$\hat{P}(z_{8} i_n)$	価格帯	クラス 9 のアイテム	$\hat{P}(z_{9} i_n)$	価格帯
1	生活雑貨 217	0.9999	450~460 円	食品 73	0.9999	160~330 円	生活雑貨 40	0.9999	2290~2380 円
2	生活雑貨 46	0.9999	200~280 円	食品 72	0.9999	3140~3170 円	衣服・雑貨 16	0.9999	3700~3790 円
3	生活雑貨 206	0.9999	17130~17590 円	生活雑貨 211	0.9999	6300~7220 円	衣服・雑貨 17	0.9999	770~860 円
4	生活雑貨 202	0.9999	2220~2320 円	生活雑貨 200	0.9999	1140~1200 円	生活雑貨 41	0.9999	10500~10560 円
5	食品 20	0.9999	3250~3340 円	生活雑貨 203	0.9999	300~400 円	生活雑貨 42	0.9999	2050~2100 円

順位	クラス 10 のアイテム	$\hat{P}(z_{10} i_n)$	価格帯
1	生活雑貨 194	0.9999	5180~5260 円
2	生活雑貨 205	0.9999	10490~10560 円
3	生活雑貨 173	0.9999	3940~4160 円
4	衣服・雑貨 125	0.9999	2340~2760 円
5	生活雑貨 59	0.9999	3690~4520 円

表 4.12: 全ステージモデルから推定された $\hat{P}(z_k)$ と提案手法による潜在クラス生起確率 $\hat{P}_s(z_k)$ の推定結果 (無印良品の実データのケース) (潜在クラス数が 20 の場合)

	$\hat{P}_s(z_1)$	$\hat{P}_s(z_2)$	$\hat{P}_s(z_3)$	$\hat{P}_s(z_4)$	$\hat{P}_s(z_5)$	$\hat{P}_s(z_6)$	$\hat{P}_s(z_7)$	$\hat{P}_s(z_8)$	$\hat{P}_s(z_9)$	$\hat{P}_s(z_{10})$
All Stage(全ステージ)	0.0293	0.0649	0.0349	0.0287	0.0146	0.0522	0.0162	0.1145	0.1378	0.0155
Stage 5(ダイヤモンド)	0.0478	0.1392	0.0663	0.0288	0.0371	0.0824	0.0427	0.0795	0.0648	0.0117
Stage 4(プラチナ)	0.0489	0.1407	0.0968	0.0179	0.0006	0.0384	0.0306	0.1011	0.0765	0.0192
Stage 3(ゴールド)	0.0087	0.1385	0.0683	0.0724	0.0001	0.0806	0.0597	0.0992	0.0128	0.0239
Stage 2(シルバー)	0.0104	0.0295	0.0812	0.0241	0.0326	0.0946	0.0753	0.1218	0.0276	0.0275
Stage 1(なし)	0.0005	0.0792	0.0095	0.0475	0.0405	0.0637	0.0656	0.2064	0.0820	0.0340
	$\hat{P}_s(z_{11})$	$\hat{P}_s(z_{12})$	$\hat{P}_s(z_{13})$	$\hat{P}_s(z_{14})$	$\hat{P}_s(z_{15})$	$\hat{P}_s(z_{16})$	$\hat{P}_s(z_{17})$	$\hat{P}_s(z_{18})$	$\hat{P}_s(z_{19})$	$\hat{P}_s(z_{20})$
All Stage(全ステージ)	0.0765	0.0228	0.0290	0.1154	0.0001	0.0001	0.0001	0.0774	0.0807	0.0894
Stage 5(ダイヤモンド)	0.0958	0.0723	0.0001	0.0870	0.0089	0.0166	0.0003	0.0252	0.0672	0.0263
Stage 4(プラチナ)	0.0802	0.0799	0.0183	0.0103	0.0194	0.0209	0.0002	0.0798	0.0308	0.0894
Stage 3(ゴールド)	0.1167	0.0113	0.0797	0.0031	0.0218	0.0239	0.0003	0.0649	0.0639	0.0503
Stage 2(シルバー)	0.0027	0.0543	0.0512	0.0221	0.0251	0.0252	0.0003	0.0148	0.1063	0.1733
Stage 1(なし)	0.0567	0.0536	0.0538	0.0006	0.0360	0.0251	0.0003	0.0699	0.0664	0.0086

表 4.13: 無印良品の実データによるステージごととに全てのユーザが各潜在クラスにおいて出現確率 ($\sum_{m=1}^M \hat{P}_s(z_k | u_m^s)$) の推定結果 (潜在クラス数が 20 の場合)

	クラス 1	クラス 2	クラス 3	クラス 4	クラス 5
Stage 5(ダイヤモンド)	56.12	170.16	82.22	33.20	34.74
Stage 4(プラチナ)	322.54	946.27	648.53	121.18	4.65
Stage 3(ゴールド)	245.03	4457.67	2132.75	2167.13	0.63
Stage 2(シルバー)	1128.16	3451.10	10379.31	2868.39	3277.32
Stage 1(なし)	241.66	51442.99	5617.95	26769.83	21738.37
	クラス 6	クラス 7	クラス 8	クラス 9	クラス 10
Stage 5(ダイヤモンド)	86.61	45.68	65.70	56.00	10.55
Stage 4(プラチナ)	232.66	194.36	538.85	380.16	96.45
Stage 3(ゴールド)	2348.52	1852.85	2446.75	287.46	566.57
Stage 2(シルバー)	10302.87	9190.70	11261.20	2208.40	2465.14
Stage 1(なし)	36697.21	41056.39	111473.01	35066.31	16381.47
	クラス 11	クラス 12	クラス 13	クラス 14	クラス 15
Stage 5(ダイヤモンド)	112.01	80.67	0.00	99.24	8.73
Stage 4(プラチナ)	504.74	532.71	123.95	69.34	115.73
Stage 3(ゴールド)	3377.88	370.91	2513.80	98.22	623.90
Stage 2(シルバー)	325.63	6779.66	6153.00	2898.26	2770.50
Stage 1(なし)	32955.32	35956.61	34189.91	283.68	23600.54
	クラス 16	クラス 17	クラス 18	クラス 19	クラス 20
Stage 5(ダイヤモンド)	18.93	0.41	29.91	62.37	27.75
Stage 4(プラチナ)	120.32	1.64	524.45	183.37	568.08
Stage 3(ゴールド)	637.83	10.45	2084.83	1870.02	1529.82
Stage 2(シルバー)	2560.43	43.91	1852.56	12067.87	20302.59
Stage 1(なし)	13183.21	417.62	48047.47	35536.67	4095.78

表 4.14: 各潜在クラスの代表的なアイテムの出現確率ランキング (潜在クラス数が 20 の場合)

順位	クラス 1 のアイテム	$\hat{P}(c_{1i} i_n)$	価格帯	クラス 2 のアイテム	$\hat{P}(c_{2i} i_n)$	価格帯	クラス 3 のアイテム	$\hat{P}(c_{3i} i_n)$	価格帯
1	衣服・雑貨 1	0.6141	200~2800 円	衣服・雑貨 346	0.9999	800~1100 円	生活雑貨 56	0.5568	80~10500 円
2	衣服・雑貨 2	0.4944	460~5600 円	多角化 1	0.6257	150~5000 円	衣服・雑貨 324	0.3935	930~15300 円
3	生活雑貨 1	0.4744	20~3650 円	生活雑貨 171	0.4998	1000~7000 円	生活雑貨 98	0.3839	140 円
4	衣服・雑貨 3	0.4375	90~1390 円	生活雑貨 120	0.4941	950~8240 円	生活雑貨 193	0.3748	190~6520 円
5	食品 1	0.4306	50~620 円	生活雑貨 173	0.4712	220~4160 円	生活雑貨 62	0.3708	110~9240 円

順位	クラス 4 のアイテム	$\hat{P}(c_{4i} i_n)$	価格帯	クラス 5 のアイテム	$\hat{P}(c_{5i} i_n)$	価格帯	クラス 6 のアイテム	$\hat{P}(c_{6i} i_n)$	価格帯
1	サービス 1	0.4194	30~23810 円	生活雑貨 150	0.3576	230~13200 円	多角化 7	0.9999	40~500 円
2	衣服・雑貨 1	0.3777	280~2840 円	食品 32	0.3482	50~260 円	衣服・雑貨 140	0.6314	430~3690 円
3	衣服・雑貨 2	0.3749	460~5620 円	生活雑貨 23	0.3379	280~1720 円	生活雑貨 77	0.5043	1020~50930 円
4	衣服・雑貨 3	0.3747	90~1390 円	生活雑貨 51	0.2943	670~1340 円	生活雑貨 127	0.4285	10~930 円
5	衣服・雑貨 4	0.3694	930~9510 円	生活雑貨 66	0.2418	240~33330 円	生活雑貨 143	0.4139	1390~2290 円

順位	クラス 7 のアイテム	$\hat{P}(c_{7i} i_n)$	価格帯	クラス 8 のアイテム	$\hat{P}(c_{8i} i_n)$	価格帯	クラス 9 のアイテム	$\hat{P}(c_{9i} i_n)$	価格帯
1	衣服・雑貨 310	0.9999	930~6460 円	生活雑貨 214	0.9999	930~8320 円	多角化 9	0.9999	110~170 円
2	生活雑貨 146	0.4886	5560~89820 円	生活雑貨 61	0.6667	2800~5000 円	生活雑貨 106	0.9998	28000 円
3	生活雑貨 127	0.4486	10~930 円	食品 29	0.6028	130~88790 円	多角化 8	0.8179	50~1360 円
4	生活雑貨 69	0.3749	50~5720 円	食品 13	0.4987	120~4580 円	多角化 5	0.7535	280~2690 円
5	生活雑貨 61	0.3333	2800~5000 円	食品 35	0.4890	90~6950 円	サービス 1	0.6788	30~23810 円

順位	クラス 10 のアイテム	$\hat{P}(c_{10i} i_n)$	価格帯	クラス 11 のアイテム	$\hat{P}(c_{11i} i_n)$	価格帯	クラス 12 のアイテム	$\hat{P}(c_{12i} i_n)$	価格帯
1	食品 9	0.4602	260~1110 円	生活雑貨 68	0.9999	12000~78000 円	衣服・雑貨 294	0.4307	460~2370 円
2	衣服・雑貨 266	0.3126	280~2840 円	生活雑貨 38	0.6494	100~10000 円	食品 38	0.4239	370~2160 円
3	衣服・雑貨 308	0.2113	460~12080 円	生活雑貨 150	0.4834	230~13200 円	衣服・雑貨 352	0.4096	280~2580 円
4	食品 22	0.1924	110~280 円	衣服・雑貨 32	0.4372	55050~174050 円	衣服・雑貨 237	0.3800	1910 円
5	食品 38	0.1846	370~2160 円	衣服・雑貨 181	0.4361	30~280 円	衣服・雑貨 190	0.3629	280~950 円

順位	クラス 13 のアイテム	$\hat{P}(c_{13i} i_n)$	価格帯	クラス 14 のアイテム	$\hat{P}(c_{14i} i_n)$	価格帯	クラス 15 のアイテム	$\hat{P}(c_{15i} i_n)$	価格帯
1	衣服・雑貨 79	0.5032	10~80300 円	生活雑貨 160	0.9999	1080~12190 円	生活雑貨 150	0.0086	230~13200 円
2	生活雑貨 62	0.4359	110~9240 円	食品 26	0.7285	160~330 円	衣服・雑貨 178	0.0016	280~4610 円
3	衣服・雑貨 190	0.3550	280~950 円	衣服・雑貨 137	0.6657	280~1890 円	生活雑貨 217	0.0015	10~460 円
4	生活雑貨 31	0.3374	90~3720 円	衣服・雑貨 334	0.6250	190~3690 円	衣服・雑貨 172	0.0001	380~3690 円
5	生活雑貨 142	0.2879	280~3690 円	生活雑貨 34	0.4909	40~760 円	生活雑貨 134	0.0001	17500~20000 円

順位	クラス 16 のアイテム	$\hat{P}(c_{16i} i_n)$	価格帯	クラス 17 のアイテム	$\hat{P}(c_{17i} i_n)$	価格帯	クラス 18 のアイテム	$\hat{P}(c_{18i} i_n)$	価格帯
1	生活雑貨 103	0.0795	30~1910 円	衣服・雑貨 261	0.0052	920~1900 円	衣服・雑貨 358	0.7855	460~3690 円
2	生活雑貨 143	0.0635	1390~2290 円	生活雑貨 53	0.0001	90~2600 円	衣服・雑貨 129	0.6551	930~7600 円
3	生活雑貨 87	0.0199	120~19000 円	衣服・雑貨 287	0.0001	280~3790 円	生活雑貨 103	0.5902	30~1910 円
4	生活雑貨 127	0.0042	10~930 円	衣服・雑貨 255	0.0001	430~9260 円	衣服・雑貨 261	0.5815	920~1900 円
5	多角化 4	0.0031	930~2780 円	生活雑貨 130	0.0001	280~16670 円	衣服・雑貨 126	0.5761	190~1830 円

順位	クラス 19 のアイテム	$\hat{P}(c_{19i} i_n)$	価格帯	クラス 20 のアイテム	$\hat{P}(c_{20i} i_n)$	価格帯
1	衣服・雑貨 342	0.9999	1260 円	生活雑貨 212	0.9999	50~29000 円
2	衣服・雑貨 329	0.6667	460~810 円	衣服・雑貨 131	0.9999	2780~41670 円
3	衣服・雑貨 361	0.5519	90~1890 円	衣服・雑貨 218	0.9998	930~13890 円
4	衣服・雑貨 306	0.5162	250~1480 円	生活雑貨 179	0.7330	50~2280 円
5	多角化 6	0.5000	330~460 円	生活雑貨 25	0.5957	150~20000 円

3. 表4.12 よりステージの成長によって、クラス8, クラス10, クラス13, クラス15, クラス16の割合が少なくなる。クラス1, クラス2では、全体から見ると、ステージが高ければ高いほど、潜在クラスの出現率の割合が増加している。クラス17がステージの成長による変化がないクラスである。クラス5, クラス12, クラス14では、上位層と最下位ステージのユーザしかいないクラスである。
4. 表4.12, 表4.14 より、クラス1での「衣服・雑貨1」とクラス2での「衣服・雑貨346」, 「多角化1」の3つのアイテムが0.5以上の出現確率を持つため、クラス1とクラス2の代表的なアイテムと言える。
5. 表4.14 より、クラス15, クラス16, クラス17に対してアイテムから出てきた出現確率がかなり低いため、これらのクラスは特徴的なアイテムを持たないクラスであると判断できる。すなわち、ステージ成長を問わず、ユーザが購入を続けるクラスである。

ステージの成長による変化 (潜在クラス数が20の場合)

ステージの成長によるユーザの購買多様性を検証するため、4.5.3節の同様にエントロピーの計算式は式(4.24)を用いた。

表4.15: 各ステージのユーザが各潜在クラスに所属する確率のエントロピーの平均 (潜在クラス数が20の場合)

ステージ名 (降順)	エントロピー
ダイヤモンド (Stage 5)	0.0922
プラチナ (Stage 4)	0.0216
ゴールド (Stage 3)	0.0053
シルバー (Stage 2)	0.0017
なし (Stage 1)	0.0004

各ステージのユーザが各潜在クラスに所属する確率の平均エントロピーの変化を表4.15に示す。表4.15により、潜在クラス数を20にした場合、ステージの成

長よって各潜在クラスに所属する確率の平均エントロピーの値も大きくなっている。すなわち、ステージの成長により、ユーザの嗜好の多様性(購買範囲)が増大していることが分かる。従って、ステージ間のユーザの嗜好の変化を比較することができる。

4.6 考察

本節では、本章の実験と分析の結果から、本章の提案手法の特性や活用法について考察を行う。

4.6.1 学習について

提案手法では全ステージのユーザデータから推定したモデルのパラメータ $\hat{P}(i_n|z_k)$ を各ステージのモデルに導入し、各ステージのモデルを構築した。各ステージのモデルを学習する際、EM アルゴリズムによるパラメータを推定する段階で、各ステージモデル中のパラメータ $\hat{P}(i_n|z_k)$ が固定されたことで、推定された各ステージのモデルパラメータ $P_s(z_k)$ と $P_s(u_m^s|z_k)$ は全ステージのモデルのパラメータ $\hat{P}(i_n|z_k)$ に合わせる。そのため、最終的に学習できたモデルのパラメータにより、クラスごとに各ステージ間のユーザの購買行動の違いやユーザの人数の変動を比較できるようになった。

4.6.2 ステージ成長への施策の立案について

提案手法で得られたパラメータ $\hat{P}(i_n|z_k)$ とベイズ定理によって、 $\hat{P}(z_k|i_n)$ の推定もできる。これにより、各クラスの代表的なアイテムの発見ができる。さらに、各ステージの購買行動の変動によって、ステージに成長に役を立つアイテムやステージが高い会員ユーザが好みのアイテムを発見することもできる。これらのアイテムを購入してもらえような施策を施すことによって会員のユーザの成長を促すようなアイテムであるということが示唆される。そのため、企業側からの観点から、ユーザのステージの向上を促す施策を考えることが可能である。

4.7 本章の結論

本章では、会員ステージ間でクラスタ分布の差異を比較分析できるモデルとその学習アルゴリズムを提案した。提案モデルでは、全ステージから推定されたパラメータを再利用し、各ステージのモデルを構築する。人工データでのシミュレーションと実データの分析結果より、ユーザの購買行動と到達ステージとの関係性を明らかにできることが示された。さらに、得られる結果を用い、ビジネス上における施策立案の可能性について検討した。

本提案モデルでは、会員ステージ間で、ユーザセグメントのサイズがどのように変化しているかを比較することが可能であり、様々な施策の検討へと結び付けることができる。今後の課題として、ステージが高いユーザの購買行動の要素を再利用し、低ステージのユーザを成長されるモデルを検討することなどが挙げられる。また、異なる会員ステージランク制度のデータについて実験や検証などがあげられる。

第5章 考察

本研究では会員ステージランク制のもとで顧客購買行動を分析するモデルとして、2つの手法を提案した。本章では、本研究全体を通じた考察を行う。

5.1 提案手法のまとめ

本研究では、以下に示すように2つの異なるアプローチから、2つの新たな分析モデルを提案した。

アプローチ1：上位ステージの顧客購買行動の特徴を把握することで、上位層の優良顧客の維持を目的とした施策へ結び付ける方法

提案手法1 [第3章] 確率的潜在意味解析の転移学習に基づく会員ステージを考慮した購入行動分析モデル

上位層のクラスタリング結果の精度を向上し、ステージ間の顧客購買行動のトレンドを把握可能である。

アプローチ2：ステージ間の顧客の購買行動変化の把握を行うことで、下位層の非優良顧客に対し、より効果的に会員ステージの成長を促す施策を立案する方法

提案手法2 [第4章] 会員ステージ間の潜在クラス分布を比較する購買行動分析モデル

ステージ間の比較基準を統一し、ステージ間の顧客購買行動の変化を把握可能である。

それぞれの提案手法の研究では、会員ステージランクのもとで購買行動分析への有効性及びビジネス施策への適用可能性を示すことができたその結果、以下の

ことが明らかになった。

- 全体/部分ステージの顧客購買データから潜在的な嗜好を抽出し、情報を再利用することで、対象になるステージ顧客購買行動分析が可能となる。
- 抽出された潜在的な購買行動の特徴を分析することで、各潜在クラスの特徴に解釈し、ステージ間、顧客購買行動はどのように変化するかを把握可能である。

提案手法は会社の経営者が会員ステージランク制の購買履歴分析に対して非常に有効な方法となりうるが、すべての場面に適用することはできないため、目的に応じてどのような手法を選択すべきか考慮すべき点を明確にする必要がある。そこで、本章では以下の2つの観点から提案手法の特性について考察する。(1)一つ目は分析者の視点での手法を議論する。(2)二つ目は会員ステージランク制において使用場面の制限を議論する。

5.2 分析者視点での手法の評価

本研究で提案した手法にはそれぞれ利点と欠点が存在し、分析の要望に応じて適切な手法を選ぶ必要がある。すなわち、具体的な分析の目的や状況に合わせて最も適した手法を選択することが重要である。本研究のベース手法となる確率的潜在意味解析 (Probabilistic Latent Semantic Analysis) はシンプルな構造を持ち、解釈性が高いモデルである。EM アルゴリズムにより、求めたモデルの結果は局所解に落ちるものの、得られる潜在クラスの特徴ができるため、各ステージに所属している顧客と商品の関係を把握することが可能となる。以下は分析者の視点から、2つの提案手法の特徴や活用について表 5.1 にまとめる。

表 5.1: 分析視点での提案手法の活用方法

	提案手法 1	提案手法 2
主な施策対象	上位層顧客	下位層顧客
内容	PLSA を用いた上位層ユーザのクラスタリングの結果向上手法	PLSA を用いたステージ間の購買嗜好差異比較手法
特徴	(1)PLSA によるシンプルで高い解釈性 (2) (ステージ数により) 下位層モデルの特徴量抽出と複数上位層モデル構築の多段階モデル	(1)PLSA によるシンプルで高い解釈性 (2) 全体モデルの特徴量抽出とステージごとモデル構築の二段階モデル
従来手法による改善点	会員ステージ上位層の顧客のクラスタリングの結果は保証できる	会員ステージ間の顧客の嗜好変化を直接的に比較できる
比較性	ステージ間のトレンド比較は可能	同じ基準でステージ間の変化分析は可能
施策の柔軟性	上位層顧客に実施可能	下位層顧客に実施可能
顧客嗜好の推定精度	上位ステージ層の推定精度：高	顧客全体の推定精度：高
活用方法	数の少ない上位層顧客の購買分析を効果的に行うため、高い推定精度のクラスタリング結果が求められる局面で有効	ステージ間での顧客の購買傾向 (嗜好) の差異の比較分析を行うため、解釈性が高く、かつ柔軟な施策が求められる局面で有効

5.3 利用場面の制限の視点での手法の評価

本研究は下記の 2つのアプローチに対して提案したが、利用場面の制限の視点から注意点をまとめる。

アプローチ 1：上位ステージの顧客購買行動の特徴を把握することで、上位層の優良顧客の維持を目的とした施策へ結び付ける方法

アプローチ2：ステージ間の顧客の購買行動変化の把握を行うことで、下位層の非優良顧客に対し、より効果的に会員ステージの成長を促す施策を立案する方法

- データ構造の制限：本研究で2つの提案手法とも下位層の顧客データが多い、上位層の顧客データが少ない状況に対して適用できる。しかし、提案手法1において上位層のクラスタリングの結果は下位層のステージの顧客のデータ数に依存するため、本研究対象以外データ構造に対して適用できない可能性がある。提案手法2においてステージ間の購買行動変化の注目する。具体的に全体顧客の購買行動を推定する上で、各ステージの購買行動を特にステージ制を持つデータの構造に対して制限がない。
- ステージ間の購買行動の類似性：提案手法1を実データに適用する際、ステージ間の顧客購買行動は少しずつ変化している前提で行う。もし隣接する2つのステージ間の購買行動は急激に変化する場合、提案手法1が知識転移はうまくいかない。そのため、この手法を適用できるかどうかについて、データ事前分析により、判断する必要がある。提案手法2において、全体購買行動を把握してから各ステージのモデルの推定を行うため、ステージごとに徐々に情報を転移する必要がない。そのため、ステージ間の購買行動の類似性を考慮しない。
- モデルの並列推定：提案手法1のモデル学習の段階で、下位層から構築し、得られるモデルのパラメータを徐々に一つ上のステージモデルに転移するため、ステップ数は2つで、モデル数が $n-1$ ($n \geq 2$)となる。つまり、各ステージのモデルを推定する際に、ステージごとのモデルの推定の並列計算はできない。一方で、提案手法2は全体モデルの特徴量を抽出して、ステージごとのモデルを構築するため、ステップ数は2つで、モデル数が $n+1$ となる。つまり、全体顧客のPLSAモデルを推定してから、各ステージモデルの推定の並列推定は可能となる。

第6章 結論と今後の課題

6.1 本研究の結論

本研究では会員ステージ制を対象とした顧客購買行動の分析手法として、2つの手法の提案を行った。提案手法1では、転移学習を基づいたPLSAモデルを用いて、構築した下位層会員ステージモデルからパラメータを上位層の会員ステージモデルに転移することで、上位層のクラスタリング結果を改善できることを確認した。生成した人工データを用いたシミュレーション実験において、推定精度の面から提案手法の有効性を示した。また、提案手法1を実データに適用し、ステージ間の変化とビジネス施策のための分析を行った。また、得られたステージごとの潜在クラスに所属するユーザーの購買行動を分析することで、ステージ間の変動トレンドを確認し、ビジネス施策立案に役立つ商品の発見など、有用な結果を得ることでできた。

提案手法2では二段階の潜在クラスの学習法を提案した。具体的には、全体ユーザのモデル構築の段階では、PLSAモデルを用いて全てのステージ顧客購買履歴データをまとめてモデルを構築することで、全体顧客の購買行動を表すパラメータを得た。ステージごとのモデル構築の段階では、パラメータを推定する際に、全体ユーザのモデルから得られる顧客購買行動に関するパラメータを比較基準の情報としてステージごとのモデルに導入し、そのパラメータを固定しながら、他のパラメータを推定した。提案手法1と同様に、人工データを生成して行ったシミュレーション実験により、提案手法の有効性を確認した。また、会員ステージランク制を対象とした実データを用いた実験により有効性を確認し、異なる粒度でステージ間の購買行動の変化を分析した。最後に、実データに提案手法を適用し、ビジネス施策につながる分析を行った。

本研究の結果は、提案手法によるステージ間のデータを転移や再利用し、従来

のPLSAモデルが比較できないステージ間の購買行動トレンドを比較できるようになることを示している。特に、提案手法1では、モデル推定精度の観点からステージ間のステージの購買行動は急激な変化がない場合、下位層から情報転移により、上位層の顧客購買嗜好のクラスリングの結果に対してより精度の高い上位ステージ層モデルを推定することができた。また、提案手法2では、全ステージPLSAモデルを推定する上で、それらの情報を再利用し、各ステージのPLSAモデル推定した。これで全体的な顧客購買行動を把握でき、各ステージ間の顧客の購買行動の違いを比較することができるようになった。これらの提案手法を用いることで、ビジネスシーンにおける施策立案やマーケティング戦略立案の場面で有用な示唆を得られると考えられる。

6.2 今後の課題と展望

本研究では、会員ステージランク制度の下で顧客購買行動分析に対する新しい手法を提案した。さらに、人工データの有効性と実データ分析を行った。しかし、以下の点は本研究の範囲を超えるものであり、さらなる研究の必要がある。

本研究では、購買行動に対する因果関係を明らかにするための分析はまだ行われていない。そのため、今後の第一の課題として、因果関係の観点から顧客の会員ステージランクの成長要因を分析することが挙げられる。この分析により、直接的に購買行動に影響を与える要因を特定することができるため、より効果的なマーケティング戦略を策定するための重要な情報が得られる。

二つ目として、ステージごとに異なる顧客嗜好を明らかにし、アイテムの重要性の分析を検討することも今後の課題として挙げられる。本研究でステージごとの顧客の成長に対して重要なアイテムを分析したが、実際の施策を施行する際に、顧客の嗜好を考慮し、顧客のセグメントごとのアイテムの重要性を解釈することも重要である。この分析を行うことで、より効果的な施策の立案が期待できる。

また、三つ目の課題として、他のビジネスシチュエーションデータを用いて、提案手法の有効性を検証する必要がある点も挙げられる。例えば、提案手法1に対しては、他の観点から顧客データを分けて転移学習を行うことが考えらる。一方、提案手法2に対しては評価値などの要素を含めてモデルを拡張することで、より

精度が高いモデルを構築されることも期待できる。

謝辞

本研究を行うにあたり、終始暖かい激励と御指導、御鞭撻を頂いた早稲田大学創造理工学部経営システム工学科教授・後藤正幸先生に心より感謝申し上げます。2014年に、「情報数理応用特論」という授業で後藤先生と出会いました。未熟だった私は後藤先生の授業内容にとっても興味を持ち、博士課程の進学を決め、日本留學生活の新たな扉が開きました。師匠であり父親のような存在でもある後藤先生の教えは私の人生に大きな変化をもたらしました。御多忙の中、長い間にわたり並々ならぬ御尽力、御指導を賜りました。大変お世話になりました。ここに深く感謝の意を表します。

本研究を博士論文としてまとめるにあたり、副査として御指導を賜りました、早稲田大学創造理工学部・永田靖教授、蓮池隆教授、早稲田大学データ科学センター・小林学教授には博士論文の内容について非常に有益な御指摘や助言を賜りました。深く感謝し、御礼申し上げます。

東京都市大学・メディア情報学部・情報システム学科・三川健太准教授、上智大学・上智大学理工学部・情報理工学科・理工学部情報理工学科・山下遥准教授には、研究の方向性や進め方に関わる多くの御指摘、有益な御助言を賜り、研究を進めるにあたり、大きな力となりました。心より感謝いたします。

後藤研究室博士課程で一緒に学ばせて頂いた蓮本恭輔氏、鈴木佐俊氏、山極綾子氏、清水良太郎氏、阪井優太氏、浅野正和氏、宋リンシン氏にはDゼミにて、多くの指摘や助言を頂き、深く感謝いたします。

また著者の在籍中、後藤研究室の学生諸氏には研究室やサブゼミ、合宿などでの活動を通して、交流させて頂き、研究のみならず、様々な学びを得ることができました。深く感謝し、御礼申し上げます。

本研究に対し、貴重な実データを御提供頂いた株式会社良品計画の皆様に深く感謝いたします。

最後に、お世話になった方々（王緒氏、川本卓樹氏、今西泰子氏、中西勇人氏）、大
学生活を見守ってくれた両親、多くの支えになってくれた友人（孫哲氏、Gu Xiaowei
氏、李航氏、王上氏）に心から感謝いたします。

令和6年2月1日

早稲田大学 創造理工学研究科

経営システム工学専攻

情報数理応用研究 後藤研究室

楊 添翔

参考文献

- [1] <https://www.jmrlsi.co.jp/knowledge/yougo/my08/my0811.html>, 2023.06.05 閲覧.
- [2] <https://3inc.jp/3incwp/blog/rankup/>, 2023.05.30 閲覧.
- [3] <https://cashless.arara.com/column/member-rank>, 2023.05.30 閲覧.
- [4] 小柴等, 石垣司, 竹中毅, 櫻井瑛一, 本村陽一ほか. 行動履歴データとライフスタイル調査にもとづく顧客モデル構築技術. 電気学会論文誌 C (電子・情報・システム部門誌), Vol. 133, No. 9, pp. 1787–1795, 2013.
- [5] 岩田具治, 渡部晋治, 山田武士, 上田修功. 購買行動解析のためのトピック追跡モデル. 電子情報通信学会論文誌 D, Vol. 93, No. 6, pp. 978–987, 2010.
- [6] 阿部誠, 近藤文代. マーケティングの科学: Pos データの解析. 2005.
- [7] Robert C Blattberg, Byung-Do Kim, Scott A Neslin, Robert C Blattberg, Byung-Do Kim, and Scott A Neslin. Rfm analysis. *Database Marketing: Analyzing and Managing Customers*, pp. 323–337, 2008.
- [8] Qian Zhang, Haruka Yamashita, Kenta Mikawa, and Masayuki Goto. 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.
- [9] Mahboubeh Khajvand, Kiyana Zolfaghar, Sarah Ashoori, and Somayeh Alizadeh. Estimating customer lifetime value based on rfm analysis of customer purchase behavior: Case study. *Procedia computer science*, Vol. 3, pp. 57–63, 2011.
- [10] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, Vol. 401, No. 6755, pp. 788–791, 1999.

- [11] Daniel Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*, Vol. 13, , 2000.
- [12] 蓮本恭輔, 雲居玄道, 後藤正幸ほか. 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.
- [13] Sei Okayama, Haruka Yamashita, Kenta Mikawa, Masayuki Goto, and Tomohiro Yoshikai. Relational analysis model of weather conditions and sales patterns based on nonnegative tensor factorization. *International Journal of Production Research*, Vol. 58, No. 8, pp. 2477–2489, 2020.
- [14] I.T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, 2002.
- [15] Alberto Petroni and Marcello Braglia. Vendor selection using principal component analysis. *Journal of supply chain management*, Vol. 36, No. 1, pp. 63–69, 2000.
- [16] Thomas Hofmann. Probabilistic latent semantic analysis. *arXiv preprint arXiv:1301.6705*, 2013.
- [17] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, Vol. 3, No. Jan, pp. 993–1022, 2003.
- [18] David M Blei and John D Lafferty. Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*, pp. 113–120, 2006.
- [19] Xueqi Cheng, Xiaohui Yan, Yanyan Lan, and Jiafeng Guo. Btm: Topic modeling over short texts. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 26, No. 12, pp. 2928–2941, 2014.
- [20] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, Vol. 313, No. 5786, pp. 504–507, 2006.
- [21] John R Wilke. Supercomputers manage holiday stock. *The Wall Street Journal*, Vol. 23, , 1992.

- [22] Chengqi Zhang and Shichao Zhang. *Association rule mining: models and algorithms*. Springer, 2002.
- [23] 山本祐子, 圓川隆夫. 顧客満足度とロイヤリティの構造に関する研究. 日本経営工学会論文誌, Vol. 51, No. 2, pp. 143–152, 2000.
- [24] 野中朋美, 藤井信忠, 新村猛, 高橋俊文, 貝原俊也. 顧客満足度を考慮した従業員満足度モデル—レストランにおける職種による差異の分析—. 日本経営工学会論文誌, Vol. 67, No. 1, pp. 59–69, 2016.
- [25] Muhammad Anshari, Mohammad Nabil Almunawar, Syamimi Ariff Lim, and Abdullah Al-Mudimigh. Customer relationship management and big data enabled: Personalization & customization of services. *Applied Computing and Informatics*, Vol. 15, No. 2, pp. 94–101, 2019.
- [26] Paul D Berger and Nada I Nasr. Customer lifetime value: Marketing models and applications. *Journal of interactive marketing*, Vol. 12, No. 1, pp. 17–30, 1998.
- [27] 阿部誠. Rfm データを用いた顧客生涯価値の算出—既存顧客の維持介入と新規顧客の獲得—. *マーケティングジャーナル*, Vol. 34, No. 1, pp. 73–90, 2014.
- [28] 阿部誠. 周期的な購買行動に対応した顧客の生涯価値の導出と顧客維持介入戦略への応用. *マーケティング・サイエンス*, Vol. 28, No. 1, p. 29, 2020.
- [29] 酒井拓哉, 三川健太, 後藤正幸. 会員ステージに着目した優良顧客・離反顧客の特徴分析. *経営システム= Communications of Japan Industrial Management Association*, Vol. 25, No. 3, pp. 182–187, 2015.
- [30] Yuri Nishio, Hiroaki Ito, Haruka Yamashita, and Masayuki Goto. β : a new analytical model for customer growth considering potential purchasing preferences β . *Total Quality Science*, Vol. 4, No. 3, pp. 148–159, 2018.
- [31] T. Yang, H. Yamashita, and M. Goto. A study on analysis methods of latent customer purchase behavior focused on membership stage growth. In *2019 IEEE*

- International Conference on Big Data, Cloud Computing, Data Science & Engineering (BCD)*, pp. 148–153, Los Alamitos, CA, USA, may 2019. IEEE Computer Society.
- [32] 土井千章, 石井暁, 荒木尊士, 稲村浩, 太田賢, 重野寛, 片桐雅二ほか. 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.
- [33] 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= Communications of Japan Industrial Management Association, Vol. 27, No. 2, pp. 70–76, 2017.
- [34] 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 実サービス支援のためのカテゴリマイニング技術. 人工知能学会論文誌, Vol. 26, No. 6, pp. 670–681, 2011.
- [35] 松寄祐樹, 三川健太, 後藤正幸ほか. マルコフ潜在クラスモデルに基づく ec サイトにおける施策実施効果分析に関する一考察. 情報処理学会論文誌, Vol. 58, No. 12, pp. 2034–2045, 2017.
- [36] 白井康之, 森田裕之ほか. 共進化的アプローチに基づく”らしさ”の数値化に関する研究 (特集 データ解析コンペティション: リテールマーケティングの新潮流). オペレーションズ・リサーチ= Communications of the Operations Research Society of Japan: 経営の科学, Vol. 61, No. 2, pp. 80–87, 2016.
- [37] 長尾圭一郎, 豊田秀樹, 秋山隆. チェック・インから購買までの時間を利用したマーケティング・セグメンテーション. オペレーションズ・リサーチ, Vol. 61, No. 2, pp. 71–79, 2016.
- [38] C-Y Tsai and C-C Chiu. A purchase-based market segmentation methodology. *Expert systems with applications*, Vol. 27, No. 2, pp. 265–276, 2004.

- [39] 平野洋介, 楊添翔, 雲居玄道, 阿部永, 立花徹也, 後藤正幸ほか. 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. *情報処理学会論文誌*, Vol. 62, No. 10, pp. 1704–1715, 2021.
- [40] Yi Sun, Teruaki Hayashi, and Yukio Ohsawa. A latent topic analysis framework for category-level target promotion in the supermarket. *Procedia Computer Science*, Vol. 192, pp. 2170–2179, 2021.
- [41] Amela Vatreš and Zerina Mašetić. Exploring customers’ behavior—analysing customer data, customer segmentation and predicting customers’ behavior on black friday. In *Advanced Technologies, Systems, and Applications VI: Proceedings of the International Symposium on Innovative and Interdisciplinary Applications of Advanced Technologies (IAT) 2021*, pp. 15–26. Springer, 2022.
- [42] Thomas Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 50–57, 1999.
- [43] Thomas Hofmann. Latent semantic models for collaborative filtering. *ACM Transactions on Information Systems (TOIS)*, Vol. 22, No. 1, pp. 89–115, 2004.
- [44] Gui-Rong Xue, Wenyuan Dai, Qiang Yang, and Yong Yu. Topic-bridged plsa for cross-domain text classification. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 627–634, 2008.
- [45] Arthur P Dempster, Nan M Laird, and Donald B Rubin. 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.
- [46] 神寫敏弘. 転移学習. *人工知能*, Vol. 25, No. 4, pp. 572–580, 2010.
- [47] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, Vol. 22, No. 10, pp. 1345–1359, 2009.

- [48] Bin Li, Qiang Yang, and Xiangyang Xue. Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction. In *Twenty-First international joint conference on artificial intelligence*, 2009.
- [49] Weike Pan, Nathan N. Liu, Evan W. Xiang, and Qiang Yang. Transfer learning to predict missing ratings via heterogeneous user feedbacks. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Three*, IJCAI'11, p. 2318–2323. AAAI Press, 2011.
- [50] Anastasia Krithara and Georgios Paliouras. Tl-plsa: Transfer learning between domains with different classes. In *2013 IEEE 13th International Conference on Data Mining*, pp. 419–427. IEEE, 2013.
- [51] 直広藤原, 健太三川, 正幸後藤. 閲覧及び購買行動を同時に表現するアスペクトモデルによる購買予測手法の提案. *経営情報学会誌*, Vol. 26, No. 1, pp. 1–16, 06 2017.
- [52] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, Vol. 22, No. 1, pp. 79–86, 1951.
- [53] P Liu, H Yamashita, J Iwanaga, M Taruishi, and M Goto. An analysis model based on latent class models to increase reactions to restaurant recommendation on social gourmet service. *IPSJ Journal*, Vol. 59, No. 1, pp. 211–226, 2018.
- [54] Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, Vol. 19, No. 6, pp. 716–723, 1974.
- [55] Hirotugu Akaike. Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*, pp. 199–213. Springer, 1998.
- [56] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pp. 461–464, 1978.

付録A

A.1 潜在クラス数の決定方法について

A.1.1 モデル選択基準の前提

モデル選択基準として AIC(赤池情報量基準)[54], [55] と BIC(ベイズ情報量基準)[56] などの情報量基準がよく使われている。しかしながら、これらの情報量基準の導出には前提として2つの条件がある。

1. 最尤推定量（尤度関数を最大化するパラメータ）が求まっていること。
2. パラメータの最尤推定量が漸近的に正規分布に従うこと。

以上の2つの基準を満たさない場合、AIC や BIC を適用して参考値として用いることは可能であるものの、厳密には、これらのモデル選択基準の妥当性を保証することはできない。

A.1.2 パラメータ数

PLSA モデルで、推定するパラメータ数は、(アイテム数 + ユーザ数) × 潜在クラス数となる。本研究では、膨大なユーザの数とアイテムの数があるため、モデルのパラメータが膨大になる。加えて、学習データ数が1,100万件のオーダーと非常に大規模であり、潜在クラスモデルは最尤推定量の探索が保証されていないことから、潜在クラス数が少ない（パラメータの探索空間が狭い）状況で、多くのデータへの当てはまりが良い最尤パラメータを探索することは困難である。一方、潜在クラスモデルの自由度を高める（潜在クラス数を増やす）と、最良解である最尤パラメータの探索難易度も高まると考えられる反面、学習データへの当てはまりのよいパラメータ領域が格段に広がり、それらのうちのいずれかを探索する可能性

が高まってくる。そのため、潜在クラス数を増やすと、パラメータ数が一気に増大し、対数尤度も大きく改善していくことがしばしば起こる。すなわち、非常に規模の大きい学習データへの対数尤度は、パラメータ数をかなり大きく増やしても引き続き改善され、極めて高次元で複雑なモデルの AIC や BIC が最良となってしまう。そのため、本稿で扱っているような非常に大規模な学習データに対して、AIC などのモデル選択基準を適用すると、過度に多数の潜在クラス数が選ばれてしまい、会員ステージ間の特徴比較や施策立案のための分析には適さなくなってしまう。本研究においても予備実験で、潜在クラス数 $K = 2$ から $K = 30$ までの範囲でモデル構築して確認したところ、潜在クラスの増加に従い、モデルの対数尤度と AIC も単調に増加してしまい、対数尤度や AIC が潜在クラス数の決定基準としてうまく機能しないことを確認している。

A.1.3 本研究における潜在クラス数の決定

本研究では、ユーザとアイテムの共起関係の分析に注目し、統計的特徴によってクラスタリング分析することを一つの目的としているため、あまりに学習データへの適合度を高め過ぎてしまうと、統計的特徴が微妙に異なる数多くの潜在クラスを解釈することになり、分析結果の明快な可視化や特徴の説明が難しくなってしまう。以上から、考えられる対策としては、次のようなアプローチが考えられる。

1. 適当な適合度指標を用いて選ばれる潜在クラス数が、人間による分析が比較的容易な数の潜在クラス数になるよう、学習データ数を減らして分析を行う。
2. 学習結果として推定されたパラメータから、各潜在クラスの特徴を定性的に考察し、解釈しやすい観点から（人間にとって最も特徴が把握し易いと考えられる）潜在クラス数を決めて分析を行う。サンプルサイズが 1,000 万件近いオーダーであり、これらデータ全体に厳密に細かく当てはまる統計モデルよりも、全体傾向を可視化した際の潜在クラス解釈性を確保し、分析結果に対する洞察が可能な分析モデルを採用する。

本研究では、各潜在クラスを用いてステージ成長に伴う特徴の変化を分析し、それを顧客の会員ステージ向上策の検討へと結びつけることが目的であるため、マー

ケティング施策で想定するクラスターリング粒度を考慮しつつ、分析者が結果を解釈し易く、分析しやすいという観点から潜在クラス数を探索的に決定した。本稿では、潜在クラスが $K = 10$ と $K = 20$ の場合に「潜在クラスごとに特徴の差異が明になること」、かつ「ステージ間の変化も分析が可能であること」を確認し、これらの潜在クラス数をに対する結果を示した。

付録B 研究業績

種類別	題名，発表・発行掲載誌名，発表・発行年月，連名者
論文	(筆頭)
[1]	<p>会員ステージ間の潜在クラス分布を比較する購買行動分析モデルに関する一考察</p> <p>日本経営工学会論文誌, Vol. 73, No. 2, pp. 54-69, 2022</p> <p><u>楊添翔</u>, 山下 遥, 後藤 正幸</p>
[2]	<p>Transfer learning based on probabilistic latent semantic analysis for analyzing purchase behavior considering customers' membership stages</p> <p>Journal of Japan Industrial Management Association, Vol. 73, No. 2E, pp. 160-175, 2022</p> <p><u>Tianxiang Yang</u>, Gendo Kumoi, Haruka Yamashita, Masayuki Goto</p>
[3]	<p>(共著)</p> <p>An Analytical Model of Relation between Browsing and Entry Activities on an Internet Portal Site for Job-hunting</p> <p>Total Quality Science, Vol.4, Issue 3, pp.109-118, 2019</p> <p>Yuuki Sugiyama, Takumi Arai, <u>Tianxiang Yang</u>, Tairiku Ogihara, Masayuki Goto</p>
[4]	<p>A Study of Feature Clustering Analysis based on the Hidden Layer Representation of an Autoencoder</p> <p>Total Quality Science, Vol.5, No.1, pp.11-22, 2019</p> <p>Shimpei Kanazawa, Yuuki Sugiyama, <u>Tianxiang Yang</u>, Masayuki Goto</p>
[5]	<p>商店を直線表現した TransRec モデルに基づく電子商店街の商品推薦アルゴリズム</p>

	<p>情報処理学会論文誌, Vol.62, No.2, pp.782-795, 2021 大堀 祐一, <u>楊 添翔</u>, 山下 遥, 後藤 正幸</p>
[6]	<p>中古ファッションECサイトにおける出品条件と販売価格の関係分析モデルに関する一考察 経営情報学会誌, Vol.30, No.1, pp.47-65, 2021 金澤 真平, <u>楊 添翔</u>, 後藤 正幸</p>
[7]	<p>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 情報処理学会論文誌, Vol.62, No.10, pp.1704-1715, 2021 平野 洋介, <u>楊 添翔</u>, 雲居 玄道, 阿部 永, 後藤 正幸</p>
[8]	<p>An Analytical Model of Response Interval Between Employees on Business Chat Systems Based on Latent Class Model Total Quality Science, Vol.7, No.3, pp.149-160, 2022 Fuyu Saito, Ayako Yamagiwa, <u>Tianxiang Yang</u>, Masayuki Goto</p>
[9]	<p>Evaluation of the Recommendation Effect of Individual Interventions in a Recommender System Total Quality Science, Vol.8, No.2, pp.77-88, 2023 Taichi Imafuku, Tatsuya Kawakami, <u>Tianxiang Yang</u>, Masayuki Goto</p>
講演	<p>(国際会議)</p>
[1]	<p>Transfer Learning Based on Probabilistic Latent Semantic Analysis for Analyzing Purchase Behavior Considering Customers' Membership Stages 2017 Asian Conference of Management Science & Applications (ACMSA2017), China, Fujian, 2017 <u>Tianxiang Yang</u>, Gendo Kumoi, Haruka Yamashita, Masayuki Goto</p>
[2]	<p>A Study on Analysis Methods of Latent Customer Purchase Behavior Focused on Membership Stage Growth The 4th International Conference on Big Data, Cloud Computing, and Data Science (BCD 2019), pp.148-153, 2019 <u>Tianxiang Yang</u>, Haruka Yamashita, Masayuki Goto</p>
[3]	<p>An Analysis Method Based on Customer Definition for Detecting Latent Customer Purchase Behavior Focused on Membership Stage Growth</p>

- 2019 Asian Conference of Management Science & Applications (ACMSA2019), 2019
- [4] **Tianxiang Yang**, Haruka Yamashita, Masayuki Goto
A Study of Analysis Methods for Explaining Different Important Items Focused on Membership Stage Growth
2020 INFORMS Annual Meeting, TA05, 2020
- [5] **Tianxiang Yang**, Haruka Yamashita, Masayuki Goto
Data Augmentation Based Method for Enhancing Interpretability of Biterm Topic Model
The 20th Asian Network for Quality Congress (ANQ2022), Beijing(Online), ANQ-TQS-22-044, 2022
- [6] **Tianxiang Yang**, Yuki Nishida, Haruka Yamashita, Masayuki Goto
Cross-Lingual Analysis Based on Natural Language Model to Explore Nationality Differences in Traveler Value
The 7th Asian Conference of Management Science and Application (ACMSA 2023), 2023
- [7] **Tianxiang Yang**, Hideo Suzuki, Masayuki Goto
(国内会議)
ベータ分布を導入した pLSA モデルに基づく協調フィルタリング
日本経営工学会秋季大会予稿集, pp.140-141, 金沢, 2015
楊添翔, 板垣直矢, 三川健太, 後藤正幸
- [8] 会員ステージ向上に着目した重要商品の分析手法に関する一考察
情報処理学会第79回全国大会, 7A-03, 2017
楊添翔, 山下遥, 後藤正幸
- [9] 会員ステージ制における顧客の行動分析のための転移学習モデルについて
日本経営工学会春季大会予稿集, K11, 2018
楊添翔, 雲居玄道, 山下遥, 後藤正幸
- [10] 生花ECサイトを対象とした閲覧履歴に基づく購買行動分析に関する一考察

	日本経営工学会春季大会, E23, 2021 楊添翔, 鎌形 祐志, 山極 綾子, 後藤 正幸
その他	(国際会議)
[1]	An Analytical Model of Relation Between Browsing and Entry Activities on an Internet Portal Site for Job-hunting 15th Asian Network for Quality Conference (ANQ2017), Soaltee Crowne Plaza, Kathmandu, Nepal, ICT-02, 2017 Yuuki Sugiyama, Takumi Arai, Tianxiang Yang , Masayuki Goto, Tairiku Ogi-hara
[2]	A Study on Feature Clustering Analysis by the Hidden Layer of Autoencoder 16th Asian Network for Quality Congress, Full Paper ANQ2018(JP-044), 2018 Shimpei Kanazawa, Yuuki Sugiyama, Tianxiang Yang , Masayuki Goto
[3]	An Analytical Model of Exhibition Price Change Effects on Second-Hand Fashion EC Site 2019 Asian Conference of Management Science & Applications (ACMSA2019), 2019 Shimpei Kanazawa, Tianxiang Yang , Masayuki Goto
[4]	Collaborative Filtering Based on Distributed Expression Considering Difference in Evaluation Tendencies 17th Asian Network for Quality Congress (ANQ2019), ID-223, 2019 Ryousuke Goto, Hideki Fujinami, Tianxiang Yang , Masayuki Goto
[5]	An Analytical Model of Users' Communication on Chat Systems The 20th Asia Pacific Industrial Engineering and Management Systems (APIEMS 2019), ID-208, 2019 Ayako Yamagiwa, Yuto Seko, Tianxiang Yang , Masayuki Goto
[6]	An Analytical Model of Response Interval Between Employees on Business Chat Systems Based on Latent Class Model 18th Asian Network for Quality Congress (ANQ2020), ANQ-130, 2020 Fuyu Saito, Ayako Yamagiwa, Tianxiang Yang , Masayuki Goto

- [7] A Study on Recommender System by Evaluating the Recommendation Effect of Individual Intervention
The 19th Asian Network for Quality Congress (ANQ2021), C10-P18-03(JP-040), 2021
Taichi Imafuku, Tatsuya Kawakami, **Tianxiang Yang**, Masayuki Goto
- [8] A Study on Ensemble Learning Model with Interpretability
The 19th Asian Network for Quality Congress (ANQ2021), C10-P25-03(JP-025), 2021
Taiga Yoshikawa, Ayako Yamagiwa, **Tianxiang Yang**, Masayuki Goto
- [9] An Improved Translation-based Recommendation Considering Last Purchasing Sequence
2021 IEEE 12th International Workshop on Computational Intelligence and Applications (IEEE IWCIA2021), Online, 2021
Yashu Li, Yamagiwa Ayako, **Tianxiang Yang**, Masayuki Goto
- [10] NoDust: Noisy Data Query Strategy Enhanced by Weak Supervised Learning
The 22nd Asia Pacific Industrial Engineering and Management Systems Conference (APIEMS2022), IEM051, 2022
Linxin Song, **Tianxiang Yang**, Masayuki Goto
- [11] An Extended Model of Bayesian Optimization Method Considering Input-dependent Variance for Decision Support in Business Activities
The 22nd Asia Pacific Industrial Engineering and Management Systems Conference (APIEMS2022), BM015, 2022
Taiga Yoshikawa, Yuta Sakai, **Tianxiang Yang**, Masayuki Goto
- [12] Adaptive Ranking-based Sample Selection for Weakly Supervised Class-imbalanced Text Classification
Findings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022), pp.1641-1655, 2022
Linxin Song, Jieyu Zhang, **Tianxiang Yang**, Masayuki Goto
- [13] An Analysis of Customer Characteristics using Embedding Model with Attention Mechanism Considering Auxiliary Information

- The 21st Asian Network for Quality Congress (ANQ 2023), Vietnam, JP.03, 2023
- [14] Tatsuya Ishii, Kirin Tsuchiya, **Tianxiang Yang**, Masayuki Goto
An Embedding based Analysis of Multilingual Reviews for Service Quality Improvement of Accommodations for Foreign Travelers
23rd Asia Pacific Industrial Engineering & Management System Conference (APIEMS 2023), 2023
- [15] Kanta Morimoto, **Tianxiang Yang**, Masayuki Goto
Online Flipped Conference Based Data Science Education Program and Its Educational Effectiveness in Multi-University Collaboration
The 7th Asian Conference of Management Science and Application (ACMSA 2023), 2023
Masayuki Goto, Manabu Kobayashi, Takeshi Moriguchi, Yoichi Seki, Hideo Suzuki, Takashi Namatame, Kazuhide Nakata, Aya Ishigaki, Masao Ueda, Kimitoshi Sato, Kenta Mikawa, Haruka Yamashita, Tomoaki Tabata, **Tianxiang Yang**, Ayako Yamagiwa, Yutaka Tajiri
- [16] (国内会議)
就職ポータルサイトにおける個社ページ閲覧とエントリーの関係分析モデルに関する一考察
日本経営工学会春季大会予稿集, A01, 2017
杉山裕貴, 荒井琢充, **楊添翔**, 荻原大陸, 後藤正幸
- [17] 自己符号化器の中間表現を用いた特徴分析に関する一考察
日本経営工学会春季大会予稿集, G09, 2018
金澤真平, 杉山裕貴, **楊添翔**, 後藤正幸
- [18] 評価傾向の差異を考慮した分散表現による協調フィルタリング
人工知能学会 第33回全国大会, 2P1-J-2-04, 2019
後藤亮介, 藤波英輝, **楊添翔**, 後藤正幸
- [19] 潜在表現モデルに基づくテレビ番組の魅力度要因分析モデル
人工知能学会 第33回全国大会, 2H1-J-2-01, 2019

- [20] 西村 祐樹, 金澤 真平, 楊 添翔, 後藤 正幸
購買履歴に基づくポイントカードユーザのクレジット切り替え分析モデル
第22回情報論的学習理論ワークショップ (IBIS 2019), 2019
平野 洋介, 世古 裕都, 楊 添翔, 後藤 正幸, 立花 徹也
- [21] 中古ファッションECサイトに出品価格と販売価格の関係分析モデルに関する一考察
第42回情報理論とその応用シンポジウム (SITA2019), 2019
金澤 真平, 楊 添翔, 後藤 正幸
- [22] 電子商店街における店舗情報を考慮した商品推薦に関する研究
第42回情報理論とその応用シンポジウム (SITA2019), 2019
大堀 祐一, 楊 添翔, 山下 遥, 後藤 正幸
- [23] ビジネスチャットアプリ上の会話履歴データを対象としたトピック分析モデル
第42回情報理論とその応用シンポジウム (SITA2019), 2019
山極 綾子, 世古 裕都, 楊 添翔, 後藤 正幸
- [24] ビジネスチャットアプリ上のユーザ間応答間隔分析のための潜在クラスモデル
第34回人工知能学会全国大会 (JSAI2020), 4K2-GS-3-01, 2020
齊藤 芙佑, 山極 綾子, 楊 添翔, 後藤 正幸
- [25] 外部条件を考慮した小売店における商品別売り上げの要因分析モデル
第19回情報科学技術フォーラム (FIT2020), F-008, 2020
平野 洋介, 御供 信薫, 楊 添翔, 後藤 正幸, 吉開 朋弘
- [26] LSTMモデルに基づくECサイトの閲覧履歴データ予測モデルに関する一考察
日本経営工学会秋季大会, A12, 2020
張 志穎, 楊 添翔, 後藤 正幸
- [27] クレジット切替施策検討のためのポイントカードユーザ分析モデル
日本経営工学会秋季大会, A05, 2020
平野 洋介, 楊 添翔, 雲居 玄道, 後藤 正幸, 高橋 雅信, 阿部 永, 立花 徹也
- [28] 低頻度購買商品を対象とした分散表現モデリングに関する一考察

- 日本経営工学会秋季大会, A03, 2020
山極 綾子, 楊 添翔, 後藤 正幸
- [29] 生花ECサイトの閲覧履歴データを学習する改良型Latent LSTM Allocation
モデルの提案
日本計算機統計学会 第34回シンポジウム, 2020
張 志穎, 楊 添翔, 後藤 正幸
- [30] 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の提案
日本計算機統計学会 第34回シンポジウム, 2020
平野洋介, 楊 添翔, 雲居 玄道, 阿部 永, 立花 徹也, 後藤 正幸
- [31] 潜在クラスモデルに基づくビジネスチャットアプリ上の従業員コミュニ
ケーション分析
第19回日本データベース学会年次大会 DEIM2021, 2021
齊藤 芙佑, 山極 綾子, 楊 添翔, 後藤 正幸
- [32] ECサイト上の購買行動における顧客嗜好変化の分析手法に関する一考察
日本経営工学会春季大会, A24, 2021
李 ア舒, 山極 綾子, 楊 添翔, 後藤 正幸
- [33] Self-Attention Network を用いたノードのマルチラベル分類
日本経営工学会春季大会, A20, 2021
飯塚 玲夫, 川上 達也, 楊 添翔, 後藤 正幸
- [34] 解釈性を有するアンサンブル識別機の効率的な学習法に関する一考察
日本経営工学会春季大会, A04, 2021
良川 太河, 山極 綾子, 楊 添翔, 後藤 正幸
- [35] 個別介入効果を評価する商品推薦モデルに関する考察
第35回人工知能学会全国大会 (JSAI2021), 3G2-GS-2h-05, 2021
今福 太一, 川上 達也, 楊 添翔, 後藤 正幸
- [36] アンサンブル学習の予測性を保持する単一決定木構築アルゴリズム
第35回人工知能学会全国大会 (JSAI2021), 3G1-GS-2g-05, 2021
良川 太河, 山極 綾子, 楊 添翔, 後藤 正幸
- [37] ノードのマルチラベル分類を可能にする Self-Attention Network の拡張モ
デル

- [38] 第35回人工知能学会全国大会 (JSAI2021), 3G1-GS-2g-01, 2021
飯塚 玲夫, 川上 達也, 楊 添翔, 後藤 正幸
複数の商品購買順序情報を考慮する拡張 Translation-based Recommendation
モデルの提案
- [39] 第44回情報理論とその応用シンポジウム, 1.4.3, 2021
李 ア舒, 山極 綾子, 楊 添翔, 後藤 正幸
強調データの拡張学習による Biterm Topic Model の解釈性向上法に関する
一考察
- [40] 第36回人工知能学会全国大会 (JSAI2022), 3E3-GS-2-02, 2022
西田有輝, 楊 添翔, 山下 遥, 後藤 正幸
入力依存の誤差分散を考慮したビジネス施策最適化法の提案
- [41] 第4回日本イーコマース学会全国研究発表大会, 2022
良川 太河, 阪井 優太, 楊 添翔, 後藤 正幸
多言語レビューによる宿泊施設の海外旅行者向けのサービス品質向上に関
する研究
- [42] 日本経営工学会秋季大会, A14, 2022
森本 貫太, 楊 添翔, 後藤 正幸
入力依存の分散を考慮したベイズ最適化によるビジネス施策決定モデルの
提案
- [43] 第45回情報理論とその応用シンポジウム (SITA2022), 5.4.2, 2022
良川 太河, 阪井 優太, 楊 添翔, 後藤 正幸
弱教師あり学習におけるラベル修正のためのクエリアルゴリズムに関する
一考察
- [44] 第45回情報理論とその応用シンポジウム (SITA2022), 5.4.4, 2022
宋 林シン, 楊 添翔, 後藤 正幸
顧客と商品の補助情報を考慮した Attention 機構付き Embedding モデルに
よる顧客特性分析
- 第37回人工知能学会全国大会 (JSAI2023), 1B3-GS-2-04, 2023
石井 達也, 土屋 希琳, 楊 添翔, 後藤 正幸

[45]	データの非線形性が Neural Matrix Factorization に与える影響に関する一考察 第5回日本イーコマース学会 全国研究発表大会, 2023 長命 祥吾, 松岡 龍汰, 清水 成, <u>楊 添翔</u> , 後藤 正幸
[46]	Neural Matrix Factorization に基づく予測モデルの汎化性能の改良に関する一考察 日本計算機統計学会 第37回シンポジウム (宮崎県宮崎市), 2023 長命 祥吾, 松岡 龍汰, 清水 成, <u>楊 添翔</u> , 後藤 正幸
[47]	宿泊施設を対象とした BERT と自動翻訳に基づく多言語レビューの埋め込み表現モデル 第46回情報理論とその応用シンポジウム (SITA2023), 2023 森本 貫太, <u>楊 添翔</u> , 後藤 正幸