

顧客のライフスタイルを考慮したECサイトにおける リピート購買分析

深野 剛正^{†1} 和田 昌樹^{†1} 大竹 恒平^{†1} 生田目 崇^{†1}

概要：本論文では、ファッション EC サイトの購買履歴データと一部の顧客に対して行われた意識・行動アンケート調査データを用いた初回購買顧客のリピート購買分析を行う。従来、こうした分析を行う場合は、すべての分析対象に対して共通に獲得されたデータ項目を用いて行われる。しかし実際には、本研究のように一部のデータは一部の対象に関してのみ得られている場合もある。近年、こうしたデータに対してデータを組み合わせるデータ融合技術に注目が集まっている。本研究ではこうした分析に注目し、一部の顧客から得られたライフスタイルに関するアンケート回答を全対象のライフスタイルの分析に用いる。そして、対象データ全体の共通項目からアンケートの回答が得られていない対象のライフスタイルを推定する。推定されたライフスタイルも含めて、リピート購買に関する分析を行い、リピート要因を探るとともに、その影響について考察する。

キーワード：リピート購買、ペルソナ分析、データ融合、ロジスティック回帰分析

1. はじめに

近年、インターネット利用者の増加を背景に、老若男女問わずパソコンやスマートフォンを用いたインターネットショッピングが流行している。電子商取引（Electronic Commerce, 以下 EC）市場の著しい成長により、EC サイト間や出店店舗間における顧客争奪が激化しており、差別化が必須となっている。サイト規模を拡大するにあたって、新規顧客の獲得・既存顧客の維持の2点が重要である。その中でも特に、既存顧客を維持していくことがECサイトの発展や継続には重要である。

既存顧客の定着化において最大の壁は、2回目の購買と言われている。一般に、1回目の購買から、2回目の購買に至るまでに最も顧客離反が大きいとされており、ECサイト事業者にとっては、初回の購買客を、どのようにして2回目の購買に到達させるかが課題である [1]。

しかし一方で、一定割合のトライアル購入は致し方ないため、どのような顧客に注力すれば顧客維持を効果的にできるかを理解する必要がある。

本論文は、経営科学系研究部会連合協議会主催「平成28年度データ解析コンペティション」の研究成果をもとにまとめたものである。本コンペティションではECサイトにおける顧客の購買履歴と、一部の顧客のアンケートデータが提供された。こうした一部分のデータに含まれる有益な顧客情報をいかに有効活用するかが本分析の主眼である。

2. 既存の研究と本研究の目的

前節で述べたように、自社の顧客を維持することは、多くの企業にとって重要な課題である。特に流通業を対象とした消費者の継続購買に関する分析は、過去にも数多くなされている。本研究では、特に初回購買から継続購買に至

るにはどのような条件が関係するかについて分析するが、利用するデータも踏まえて、初回購買行動の特徴並びに、生活意識を用いたモデルを提案する。以下でこうした方針に関連する既存の研究についてまとめ、さらに本研究との関連について述べる。

顧客の維持を考える上で、自社が有する顧客がどのような顧客であるかを理解することは重要である。顧客理解の際に用いられる代表的な手法の一つに、ペルソナ分析がある。ペルソナとは、対象となる製品やサービス、企業における典型的な顧客（ターゲット）を想定し、その顧客の人物像や価値観、ライフスタイルなどを設定した上で、具体的な人物像としたものである [2]。作成したペルソナを用いることで、ターゲットとなる顧客の人物像を理解・共有し、顧客満足度を高めるために必要な、製品やサービス、適切なアプローチを分析することができる [3]。

また、自社の有する顧客の価値観を適切に理解することは、個人の嗜好や消費行動を推定する要素に留まらず、CRM (Customer Relationship Management) の観点からも重要視されている。顧客の価値観の推定には、会員登録や、アンケート調査の際に得られる人口統計的データや、心理的データを用いることが主流である。また、昨今の情報通信技術の進展に伴い、SNS やマイクロブログなどのソーシャルメディアからの顧客価値観の推定を目的とした、谷田らの一連の取組みがある [4-6]。谷田らは、顧客価値観を抽出することは、マーケティングにとって、ターゲティングやセグメンテーション、あるいはポジショニングを行う際の消費者像を精神的・心理的な側面や行動から精緻に説明することを可能とし、マーケティング施策にインスピレーションを与える効果があると述べている。

抽出した顧客価値観は、ID付きPOSデータやアクセスログデータ、会員データといった企業が有する様々な消費者行動データと共に用いられる。中でも、消費者行動の推定モデルについては、レコメンデーションやサービスデザ

^{†1} 中央大学（連絡先：nama@indsys.chuo-u.ac.jp）
投稿：2017年11月30日
採択：2018年3月1日

インの領域において、様々な取組みが行われている [7-10]. また、これまでの顧客維持を目的とした研究では、このようなデータを用い、顧客の判別モデルを作成することで、優良顧客の特定 [11, 12] が試みられている.

以上で述べたように、顧客の継続購買をどのように把握するかについては、意識・行動の両面で考えることが好ましい. また、ペルソナ分析を消費者理解の一つの方法として行うことで、消費者の生活意識の意味付けを行うことが可能となる. またペルソナを変数として購買への影響を分析することで、消費者意識やライフスタイルを消費者行動と結び付けることができる. こうした分析を通して、顧客の真のニーズやウォンツに迫ることができると考えられるとともに、企業にとっては顧客の意識や行動に合わせたより適切なプロモーションを考察するための有益な情報となりと考える.

前述の通り、本論文は「平成 28 年度データ解析コンペティション」で提供されたデータを元に分析した成果をもとにしている. 本コンペティションで提供されたデータには生活価値観や購買以外の行動アンケートが含まれる. ただし、アンケート回答は 3,144 人分であり、これらのアンケート回答がない顧客が他に 100,000 人存在する. アンケート調査は主に市場での顧客意識などを分析・評価するために利用されるが、直接顧客に対してマーケティング施策を実行しようとした場合、アンケート回答者に限定された施策となってしまう. 実際は、アンケート調査を実施していない顧客の方が大多数であるので、こうしたアンケート調査の結果をすべての顧客を対象とした分析に利用できないことになる. アンケート調査は行動実績の他生活意識等、購買行動などに影響を及ぼすと考えられる重要なデータも含まれるが、データが取得できない限りは分析に用いることはできない. そこで、こうした欠点を補うために、本研究では一部の顧客に実施されたアンケート結果から抽出された生活価値観について、アンケート未実施の顧客に対して推定し、ペルソナ分析を行う. そして、推定された結果を用いてリピーター顧客の分析を行い、継続顧客となる条件を抽出する分析モデルを提案する.

3. 分析に使用したデータ

本研究では、「経営科学系研究部会連合協議会主催平成 28 年度データ解析コンペティション」で提供された、あるファッション系 EC サイトの 2015 年 4 月から 2016 年 3 月の購買データ、顧客データ、商品データ、アンケートデータを使用する. ただし先に述べた通り、アンケートは一部の顧客のみからしか得られていない. 対象の EC サイトでは、服以外にもシューズやバッグ、財布などの小物まで取り揃えている多数のショップやブランドが存在しており、ファッションのトータルコーディネートが可能である.

本研究では以下の条件に当てはまるデータを対象とし

た.

対象とした会員データ

- 期間内に 1 回以上購買を行っている顧客
- 年間の購買金額が 200 万円未満の顧客

対象者数

- アンケート回答者: 3,115 人
- アンケート未回答者: 98,335 人

また、本研究で用いたデータ期間は以下のとおりである.

データ期間

- アンケート期間: 2016 年 3 月 17 日~2016 年 3 月 23 日
- 購買観察期間: 2015 年 4 月 1 日~2016 年 3 月 31 日

4. 分析

本研究では、まず、アンケート回答や購買行動からアンケート未回答者の生活価値観を推定する. そして、行動データや顧客属性、推定した生活価値観を用いた、新規顧客から継続的に利用する顧客と離反していく顧客の判別モデルを提案する. 全体の流れは図 1 の通りであり、それぞれを次節以降説明する.

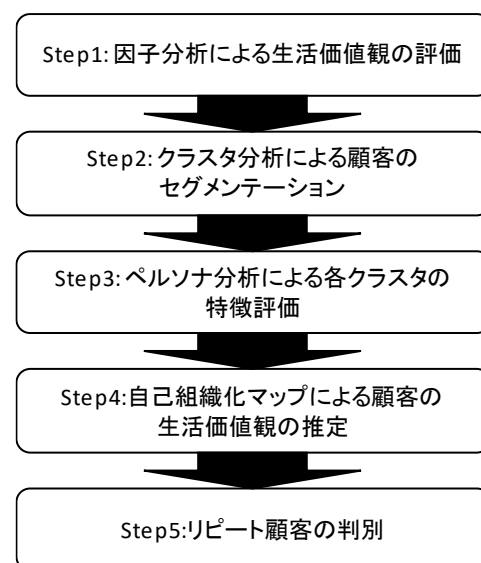


図1 分析のフレームワーク

4.1 Step 1: 因子分析による生活価値観の評価

まず、生活価値観を類型化するために、アンケート回答を用いた因子分析 [13] を行う. 用いたアンケート項目は、アンケート質問項目のうち表 1 にある「人生において重視する価値観」である. なお、回答は (1: 当てはまらない, 2: あまり当てはまらない, 3: やや当てはまる, 4: 当てはまる) の 4 段階である.

表1 アンケート項目

質問番号	人生において重視する項目
1	安定
2	感性・精神性
3	業績・名声
4	競争・勝利
5	個性
6	資格
7	所属
8	専門性
9	出世
10	富
11	有名

本研究では、試行錯誤の結果、4 因子モデルを採用した。なお、因子負荷量の推定にあたり、直交回転の一種であるバリマックス回転を行っている。分析の結果得られた因子とその因子負荷量の分析結果を表2にまとめる。

表2 因子分析によって求められた因子負荷量

人生重視価値	1	2	3	4
業績・名声	0.765	0.184	0.155	0.050
競争・勝利	0.742	0.129	0.097	-0.001
出世	0.649	0.314	0.034	0.076
有名	0.507	0.259	0.131	-0.084
富	0.411	0.130	0.157	0.256
資格	0.212	0.632	0.004	0.113
所属	0.431	0.561	-0.043	0.116
専門性	0.179	0.515	0.292	-0.147
感性・精神性	0.086	-0.018	0.676	0.180
個性	0.144	0.125	0.620	-0.139
安定	0.046	0.035	0.003	0.750
寄与率 (%)	20.694	11.171	9.134	6.746
累積寄与率 (%)	20.694	31.864	40.998	47.744

因子負荷量の値(表2の太数字の価値項目)から、第1因子から第4因子まで順に「成功重視因子」、「学習重視因子」、「個性重視因子」、「安定重視因子」と名づけた。

4.2 Step 2: クラスタ分析による顧客のセグメンテーション

因子分析で求められた因子得点を用いて、k-means 法 [14] を用いたクラスタ分析を行った。その結果を表3に示す。クラスタ数については Calinski and Harabasz の指標 [15]などを参考に5とした。

表3 分割された5つのクラスタの概要

クラスタ番号	人数	各因子得点の平均			
		1	2	3	4
1	561	-0.641	-1.003	-0.141	0.396
2	479	-0.422	-0.150	0.800	-0.865
3	565	-0.279	0.521	0.250	0.749
4	895	-0.070	0.168	-0.667	-0.340
5	615	1.263	0.301	0.254	0.125

クラスタ1から5はそれぞれ因子得点の平均から「成功・学習を重視しないクラスタ」、「個性を重視するクラスタ」、「安定を重視するクラスタ」、「個性を重視しないクラスタ」、「成功を重視するクラスタ」と名づけた。

4.3 Step 3: ペルソナ分析による各クラスタの特徴評価

クラスタ分析で求められた5つのクラスタに対しペルソナ分析をおこなう。ペルソナ分析は、顧客のプロフィールや行動を把握し、理解するために顧客のプロフィールを明らかにする分析であり、後述するリピート顧客像を評価する際の手がかりとする。本研究においては、購買データ、アンケートデータをもとにペルソナの構築を行った。以下に、各クラスタから読み取れるペルソナについてまとめる。

クラスタ1: 成功・学習を重視しないクラスタ

40代以上の女性の割合が最も多い、専業主婦クラスタである。1年を通して、バーベキューや海水浴などのイベントへの参加率が一番低い。インターネットやパソコンはあまり詳しくない。周囲との付き合いはあまり積極的に広げようとしない。

クラスタ2: 個性を重視するクラスタ

30代後半以上の男性の割合が最も高い、個性派クラスタである。他のクラスタと比べ、イベントの参加率は低めである。音楽フェスへの参加率は高い。また、アンケート項目を集計すると、既製品に飽き足らず、自分なりのカスタマイズをしたい回答者が多いことも分かった。

クラスタ3: 安定を重視するクラスタ

20代後半の女性の割合が最も多い、社交派クラスタである。パレンタインデーやホワイトデー、旅行、クリスマスパーティなど、他人と交流するイベントへの参加率が高い。

クラスタ4: 個性を重視しないクラスタ

30代後半、40代以上で6割近くを占める高年齢クラスタである。イベントについては、他のクラスタに比べて平均的な参加率となっている。

クラスタ5: 成功を重視するクラスタ

10代、20代前半が最も多い、活発的なクラスタである。バーベキューや海水浴、花火大会など、大人数で楽しむイベントへの参加率が高い。周囲との付き合いはあまり積極的に広げている。モバイル機器はいつも手放せない。

表4 独自に定義した変数の一覧

変数の種類	変数名	変数の説明
時間帯	深夜	購買時間が0～3時
	早朝	購買時間が3～6時
	朝	購買時間が6～9時
	昼	購買時間が9～15時
	夕方	購買時間が15～18時
	夜	購買時間が18～24時
季節	春	購買月が3～5月
	夏	購買月が6～8月
	秋	購買月が9～11月
	冬	購買月が12～2月
商品価格帯	G01	購買商品の平均価格が0円～1,999円
	G02	購買商品の平均価格が2,000円～2,999円
	G03	購買商品の平均価格が3,000円～4,999円
	G04	購買商品の平均価格が5,000円～8,999円
	G05	購買商品の平均価格が9,000円以上
ショップ・ブランド人気度	超人気ショップ・ブランド (G1)	1年間の累計購買商品数が上位25%
	人気ショップ・ブランド (G2)	1年間の累計購買商品数が上位25%～50%
	普通ショップ・ブランド (G3)	1年間の累計購買商品数が上位50%～75%
	不人気ショップ・ブランド (G4)	1年間の累計購買商品数が下位25%

4.4 Step 4: 自己組織化マップによるアンケート未回答顧客の生活価値観の推定

データ分析において、複数のデータを組み合わせることで個別のデータからは得られない結果を得ようとする分析手法をデータ融合と呼ぶ [16]。たとえば購買履歴とアンケート回答の2つの別々のデータにおいて、個人属性として年齢と性別はいずれのデータも得られているとする。データ融合では、これら共通に得られている項目を手掛かりに、本来別々のデータである購買履歴とアンケートを組み合わせることでアンケートから得られる意識項目と購買実態を結び付けて利用できるようにする。

本研究においては、こうしたデータ融合の概念に基づいて、一部のアンケート回答者の結果から未回答者のライフスタイルを推測する。分析にあたっては、本研究では自己組織化マップ[17]を用いた価値観推定を行った。購買行動をもとに、同様の行動をしている顧客をグルーピングし、グループごとに生活価値観を推定する。なお、マップのサイズは10×10とした。これ以上大きいと含まれる顧客数が少ないセルがあり、逆に小さくすると、一つのセルに多くの顧客が配置されてしまい、特徴づけが難しくなるからである。また、セルの隣接関係は六角格子型のマップを用いている。使用した変数は購買商品のカテゴリー・カラー・価格帯、購買ブランド・ショップ、季節・平日休日・時間帯、セール商品・予約商品、利用デバイスである。

独自に定義した変数については、表4に示す。なお、変

数はすべて区間[0,1]に基準化している。

表5は、アンケート未回答者の所属クラスを推定した概要である。所属クラスについては、推定されたセルに含まれるアンケート回答者数の多数決で決定した。アンケート回答者の人数が同数の場合は、クラス分析の結果の人数が少ないクラスに所属するものとした。

表5: 所属クラスを推定した顧客の概要

クラス番号	人数
1	12,163
2	6,642
3	13,533
4	53,928
5	12,069

4.5 Step 5: リピート顧客の判別

初回購買から半年以内に購買を行うか否かの判別モデルをロジスティック回帰分析[18]を用いて予測し正解率を評価した。対象とする顧客は、2015年4月～9月までに登録を行い、登録をしてから1回目の購買を行ったトライアル顧客とした。

なお、学習用のデータとテスト用のデータを4:1に分け、正答率の算出にはテスト用データを用いた。

説明変数には、購買した商品カテゴリー、セール期間の始

まり・終わり, 注文金額, 購買個数, 登録日から初回購買までの日数, ショップ・ブランドにおけるクラスタを用いた。また, 自己組織化マップで推定したクラスタを用いた。目的変数は, 初回購買から半年以内に購買を行うか否かの2値である。なお, グループなどの択一形式の変数については, 一つの水準を基準としてモデルから取り除いている。

まず, 分析対象のすべてのデータについて判別を行う。このとき, 所属クラスタも説明変数として用いている。なお, 個性を重視しないクラスタを基準として除いた。また, 登録日から初回購買までの日数については0日を基準とした。分析の際に, AIC (Akaike Information Criteria) を基準として変数減少法を用いた。

結果を表6にまとめる。なお, 表中の記号はそれぞれ***は0.1%有意, **は1%有意, *は5%有意, .は10%有意を表す。

このモデルによる検証用データの正答率はおよそ60%であった。この結果を見ると, いくつかの生活価値観に関するクラスタが選ばれており, 生活価値観が異なるとリピート購買の有無に影響を及ぼす変数が異なる可能性がある。

表6 全データの分析結果

説明変数	係数	有意性	オッズ比
切片	-0.910	***	
注文金額	0.000	*	1.000
購入個数	0.064	**	1.066
トップスを購買	0.360	***	1.433
パンツを購買	0.336	***	1.399
シューズを購買	0.121	*	1.129
ジャケット・アウターを購買	0.202	*	1.224
ワンピースを購買	0.271	**	1.311
セール終了日	-0.191	*	0.826
セール率が高い日に購買	0.341	**	1.406
注文金額が中価格の顧客	0.087		1.091
注文金額が弱高価格の顧客	0.119	.	1.126
登録から1日	0.377	***	1.458
登録から2~6日の間	0.601	***	1.824
登録から7~21日の間	0.922	***	2.514
登録から22~44日の間	0.829	***	2.291
登録から45日以上	0.682	***	1.978
超人気ショップ・超人気ブランドで購買する顧客	0.320	***	1.377
比較的人気なショップ・ブランドで購買	0.231	***	1.260
不人気なショップ・ブランドで購買	-0.102	.	0.903
成功・勉強を重視しない顧客	0.559	***	1.749
個性を重視する顧客	-0.327	***	0.721

そこで以下では, 生活価値観のクラスタごとにリピート購買に関するロジスティック回帰分析を行う。所属クラスタについては説明変数からは外した。変数選択については, 上記のモデルと同様 AIC を基準とした変数減少法により行った。各クラスタの結果を表7~11にまとめる。

なお, 結果の考察は次節で行う。

表7 クラスタ1の分析結果

説明変数	係数	有意性	オッズ比
切片	-0.433	.	
平均購買金額	0.000	*	1
アンダーウェアを購買	-1.233	**	0.291
帽子を購買	-0.719	*	0.487
日ごとのセール率	0.624	.	1.865
注文金額・弱高価格	-0.313		0.731
登録から2~6日の間	0.969	**	2.635
登録から7~21日の間	0.908	**	2.478
登録から22~44日の間	0.829	*	2.291
比較的人気なショップ・ブランドで購買する顧客	0.431	*	1.539
普通なショップ・人気のブランドで購買する顧客	0.916	***	2.5
不人気のショップ・ブランドで購買する顧客	1.592	***	4.914

表8 クラスタ2の分析結果

説明変数	係数	有意性	オッズ比
切片	0.170		
注文金額	0.000	*	1
水着・着物・浴衣	-15.400		0
平均購買金額	0.000	*	1
トップスを購買	0.350	.	1.418
ワンピースを購買	0.744	*	2.105
バッグを購買	-0.488	.	0.614
日ごとのセール率	1.314	**	3.721
注文金額・弱低価格	0.417		1.517
登録から7~21日の間	0.745	.	2.106
登録から22~44日の間	0.991	.	2.694
登録から45日以上	0.848	.	2.334
超人気ショップ・超人気ブランドで購買する顧客	-1.286	***	0.276
比較的人気なショップ・ブランドで購買する顧客	0.940	.	2.56
不人気のショップ・ブランドで購買する顧客	-2.001	***	0.135
不人気のショップ・普通のブランドで購買する顧客	1.169	*	3.219

表9 クラスタ3の分析結果

説明変数	係数	有意性	オッズ比
切片	0.755	***	
注文金額	0.000	.	1
購買個数	0.270	*	1.31
平均購買金額	0.000	**	1
トップスを購買	0.308	*	1.36
パンツを購買	0.405	*	1.499
バッグを購買	-0.263		0.769
セール商品を購入	-0.193	**	0.824
セール期間の始まり	0.314		1.369
登録から1日	0.867	***	2.38
登録から2~6日の間	0.698	***	2.01
登録から7~21日の間	1.204	***	3.333
登録から22~44日の間	0.729	**	2.073
登録から45日以上	0.605	**	1.831
比較的人気なショップ・ブランドで購買する顧客	-0.589	***	0.555
普通なショップ・人気のブランドで購買する顧客	-0.409	*	0.665
不人気のショップ・ブランドで購買する顧客	0.293	***	3.644
不人気のショップ・普通のブランドで購買する顧客	0.315	.	1.37

表 10 クラスタ 4 の分析結果

説明変数	係数	有意性	オッズ比
切片	1.028	***	
購買回数	0.100	***	1.106
トップスを購買	0.407	***	1.503
パンツを購買	0.445	***	1.56
シューズを購買	0.212	**	1.236
ジャケット・アウターを購買	0.368	**	1.444
ワンピースを購買	0.288	*	1.333
スカートを購買	0.204		1.227
セールの終わり	-0.227	.	0.797
セール期間の始まり	-0.225	.	0.799
大型セール日	0.337		1.401
日ごとのセール率	0.426	**	1.531
登録から1日	0.328	**	1.388
登録から2~6日の間	0.571	***	1.77
登録から7~21日の間	0.998	***	2.712
登録から22~44日の間	0.946	***	2.574
登録から45日以上	0.801	***	2.228
超人気ショップ・超人気ブランドで購買する顧客	0.332	**	1.394
比較的人気のショップ・ブランドで購買する顧客	0.280	**	1.323
普通なショップ・人気のブランドで購買する顧客	-0.179	.	0.836
不人気のショップ・ブランドで購買する顧客	-0.171	.	0.843
不人気のショップ、普通のブランドで購買する顧客	0.415	***	1.515

表 11 クラスタ 5 の分析結果

説明変数	係数	有意性	オッズ比
切片	-0.083		
平均購買金額	0.000	*	1
トップスを購買	0.473	**	1.604
シューズを購買	0.488	**	1.629
ワンピースを購買	0.370		1.447
バッグを購買	0.318	.	1.375
セール期間の始まり	0.408		1.504
登録から2~6日の間	0.426	.	1.531
登録から7~21日の間	0.463	.	1.589
登録から22~44日の間	1.065	**	2.901
登録から45日以上	0.620	*	1.859
超人気ショップ・超人気ブランドで購買する顧客	0.706	**	2.026
比較的人気のショップ・ブランドで購買する顧客	0.539	**	1.714
不人気のショップ・ブランドで購買する顧客	-0.590	**	0.555
不人気のショップ、普通のブランドで購買する顧客	-1.346	***	0.26

5. 考察

以下に、前節までの分析結果から得られる知見についてまとめる。

5.1 全体の判別モデルの考察

ロジスティック回帰分析の結果から、初回購買の商品に関しては、トップスやパンツなど全体的によく買われている商品を買うと初回リピートしやすく、水着・着物・浴衣などあまり買われていない商品を買うと、初回リピート顧客になりにくいということがわかった。

ブランドやショップに関しては、全体としては人気のブランド・ショップでの購入がリピートにつながる傾向にある。これは、まだ使い慣れていないサイトということもあり、リスク回避したいといった顧客心理が働いているのではないかと考えられる。また、セールの終わり、日ごとのセール率の変数に比べ、セール日に購買する顧客の方が、初回リピートしやすい。

なお、登録から初回購買までの日数の期間が長い方が、初回リピートする確率が高い。これは、とりあえず登録だけ済ませ、欲しい商品が発売されるタイミングやその商品が安くなるセール期間に狙いを定めたりして利用している可能性の一端を示していると考えられる。

クラスタについては、成功・学習を重視しないクラスタ（クラスタ 1）のリピート率が高くなる。これは、このクラスタに所属している顧客は平均的なクラスタと比べ、他のサイトを利用することに対して抵抗を持っているからではないかと考えられる。一方で、個性を重視するクラスタ（クラスタ 2）は、平均的なクラスタと比べ初回リピートしにくいという結果が得られた。ちなみに、ファッションに力を入れるこのクラスタでは、ファッションにおける課題に「買いたい服がない」ことを挙げていた。以上のことから、この EC サイトでは物足りずに、他の EC サイトや実店舗へ移動してしまっているのではないかと推測される。

5.2 各クラスタの考察

次にクラスタごとのモデルの結果について考察する。

どのクラスタでも、アンダーウェアや帽子など、あまり変わっていない商品を購入する顧客は、初回リピートしにくい。また、セールが盛んな時に初回購買を行った顧客は、初回リピートしやすいということがわかった。

成功・学習を重視していないクラスタ 1 では、登録から初回購買までの期間が 2~6 日の顧客が最も初回リピートしやすい。また、人気のないショップ・ブランドを購入した顧客は、初回リピートしやすいことがわかった。

個性を重視するクラスタ 2 では、全体で行った判別モデル各クラスタと比べ、正答率がとても高くなった。これは、このクラスタが最も当てはまりが大きいことを示唆していると考えられる。また、このクラスタには、初回リピートに最も特徴があると考えられる。人気のあるショップ・ブランドで購買する顧客は初回リピートしにくく、人気のないショップ・人気度が普通なブランドの商品を購入する顧客が初回リピートしやすい。このクラスタに所属する顧客は個性を重視していると推定されるので、多くの人が買っていない商品を好んで購買しているからではないかと考えられる。登録から初回購買までの期間は、22~44 日の期間である顧客が最も初回リピートしやすいことがわかった。これは、サイトを続けて利用しようと思っている新規顧客が、じっくりと欲しい商品やセールの期間を待っているためだと考えられる。

安定を重視するクラスタ 3 では、最も人気のないショップ・ブランドの商品を購入する顧客が、最も初回リピートしやすい。

個性を重視しないクラスタ 4 の結果は、新規顧客全体で作ったロジスティック回帰分析のモデルと似た結果となっ

た。それは、このクラスタが、他のクラスタと比べて最も人数が多いためと考えられる。したがって、クラスタ4は平均的なクラスタであるといえるであろう。

成功を重視するクラスタ5では、人気であるショップ・ブランドで購入した顧客が初回リピートしやすく、人気のないショップ・ブランドを購入した顧客は初回リピートしにくいとわかった。それは、このクラスタの若者の割合が少し多いためだと考えられる。登録から初回購買までの期間は、22~44日の期間である顧客が最も初回リピートしやすい。

なお、判別の精度は、クラスタ2ではおよそ80%と高かったが、それ以外のクラスタでは、全体のモデルに比べて顕著には高くなかったが、クラスタによって異なる特徴を抽出できた。

6. まとめと今後の課題

本研究では、アンケートデータを使用して顧客のクラスタ分けを行った。それを用いて、クラスタの特徴付けやアンケートデータのない顧客の所属クラスタを推定した。そのクラスタを変数とし、ロジスティック回帰分析による初回リピートの判別モデルを作成した。

また、クラスタごとに判別モデルを作成し、初回リピートの特徴を、登録から初回購買までの日数・初回購買のブランドとショップ・所属クラスタの3つから把握した。この結果から、登録から初回購買までの日数が1~3週間の期間の顧客が初回リピートしやすいこと、人気のショップ・ブランドで購入する顧客の方が初回リピートしやすいことがわかった。顧客クラスタについては、成功・学習を重視する顧客が初回リピートしやすく、個性を重視する顧客が初回リピートしにくいことがわかった。

本研究では、顧客へのアプローチ後の効果を検証していないため、実際のマーケティングが顧客定着に対して効果が表れるのか検証していく必要がある。

また、本研究で構築したモデルの予測精度は十分ではないため、新たに変数を作成するなど、精度を向上させる必要がある。

参考文献

- [1] 株式会社コアフォース, “初回リピート (2回目購入) が大切な理由” [https://wakuten.net/初回リピート \(2回目購入\) が大切な理由](https://wakuten.net/初回リピート (2回目購入) が大切な理由) (2017年7月15日閲覧)
- [2] 水本徹, “業務用製品開発におけるペルソナの活用—人物像を作るのではなく、顧客像を作る—,” 情報処理学会デジタルプラクティス, Vol. 6, No. 4, pp. 264-270 (2015).
- [3] Cooper, A., *The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity*, 2nd Ed., Sams (2004).
- [4] 谷田泰郎, 河本裕輔, 馬場彩子, “マイクロブログにおける潜在的価値観の推定,” 人工知能学会全国大会論文集, Vol. 27, pp. 1-4 (2013).
- [5] 谷田泰郎, 馬場彩子, 河本裕輔, 藤井絵美子, “価値観モデルを利用したマイクロブログ発信者の社会的類型の推定,” 言語処理学会第19年次大会発表論文集, pp. 628-631 (2013).
- [6] 谷田泰郎, “価値観マーケティングと社会知ネットワーク,” 人工知能学会論文誌, Vol. 29, No. 5, pp. 456-463 (2014).
- [7] Petteri, N., Antti, S., Andreas, F., Fabian, B. and Patrik, F., “PromotionRank: Ranking and Recommending Grocery Product Promotions Using Personal Shopping Lists,” *ACM Transactions on Interactive Intelligent Systems*, Vol. 4, No. 1, pp. 1-23 (2016).
- [8] 櫻井瑛一, 本村陽一, 安松健, 道田奈々江, 坂本和夫, “ビッグデータと確率モデリング技術を用いた自動車ディーラーにおける自動車ユーザモデルの構築,” 人工知能学会第27回「社会におけるAI」研究会予稿集, 6 pages (2016).
- [9] 井上弘子, 原田利宜, “ファジィ AHP を用いた製品レコメンドシステム” 日本感性工学会論文誌, Vol. 11, No. 2, pp. 255-263 (2012).
- [10] 西尾義英, 藤井絵美子, 安松健, “Web サイト閲覧行動のモデリングと評価,” 人工知能学会第28回全国大会予稿集, 4 pages (2014).
- [11] 田中孝昌, 濱口智大, 西郷拓海, 津田和彦, “スーパーマーケットの店舗別販売傾向と RFM 分析を利用した優良顧客分類,” 人工知能学会第20回知識流通ネットワーク研究会予稿集, pp. 1-6 (2017).
- [12] 高橋雅和, 山田隆志, 津田和彦, 寺野隆雄, “双方向リコメンデーションシステムによる小売流通マーケティングの改善手法,” 電気学会論文誌 C, Vol. 130, No. 2, pp. 317-323 (2010).
- [13] 豊田秀樹, 「因子分析入門—R で学ぶ最新データ解析」, 東京図書 (2012).
- [14] 新納浩幸, 「R で学ぶクラスタ解析」, オーム社 (2007).
- [15] Calinski, T., and Harabasz, J., “A Dendrite Method for Cluster Analysis,” *Communications in Statistics Theory and Methods*, Vol.3, No.1, pp.1-27 (1974).
- [16] 星野崇宏, 「調査観察データの統計科学」 岩波書店 (2009).
- [17] Kohonen, T., *Self-Organized Mapping*, 2nd Ed., Springer, (2001).
- [18] 豊田秀樹 (編著), 「回帰分析入門」, 東京図書, (2012).

Repeat Purchase analysis in EC Site by Considering Customer's Lifestyle

Takemasa FUKANO^{†1} Masaki WADA^{†1} Kohei OTAKE^{†1}
Takashi NAMATAME^{†1}

Abstract: In this study, we analyze repeat purchase behavior of new customers in an EC site using purchase records of all customers and questionnaire responses with respect to life style from partial customers. When we analyze using data, use only common variables that are obtained from all objects, conventionally, However, like as this study, there are some case that some variables were obtained from partial objects. For these data, data fusion technique, which integrates plural data, is focusing. In this study we use same nature technique, we use lifestyle questionnaire from partial customers to analyze all customers' lifestyle. Then, we estimate a lifestyle of each customer using above questionnaire and common variables. Moreover, we analyze repeat purchase behavior to investigate the effect of variables and to consider the effect of estimated lifestyle.

Keywords: Repeat Purchase, Persona Analysis, Data fusion, Logistic Regression

^{†1} Chuo University (Correspondence Author: nama@indsys.chuo-u.ac.jp)
Submitted: 30/11/2017
Accepted: 01/03/2018