

ファッションECサイトにおける イノベーター検出モデル —基準変数のある多種混合の項目反応モデリング—

杉山 啓太, 豊田 秀樹, 長尾 圭一郎, 磯部 友莉恵, 岡 律子

1. はじめに

マーケティング施策を考えるうえで、Rogers はイノベーションの採用態度によって消費者層を類型化するイノベーター理論 [1] を提案している。イノベーター理論ではイノベーションの普及が時系列上に正規分布すると仮定され、中でもイノベーションの採用が最も早い顧客はイノベーター (Innovators) と呼ばれる。

一般的に商品は普及にしたいが価格が下がり、市場に投入された時点の価格が最も高い。イノベーターは商品価格が最も高い時期に購買行動をとる消費者層といえ、金銭的な余裕もっていることが想定される。

また早期におけるイノベーターの購買は、後に普及が起こった結果から捉えれば普及の兆候ともいえる。したがってメーカーが商品が投入されたごく初期の段階でイノベーターの購買行動を把握できれば、メーカーにとって価値のある知見となる。

以上の点からイノベーターは優良顧客であるばかりでなく、その購買行動から商品に対する貴重な示唆を得ることができる存在である。しかしイノベーターの存在をメーカーが事前に検知することは容易ではない。なぜならイノベーターは商品が普及した時点において定義が可能となる存在だからである。イノベーターを早期に検知するためには、当該商品が普及するという前提に立たなければならない。

そこで本研究では個別の商品ではなく、毎年確実に

普及すると考えられる洋服の季節トレンドに注目し、商品の普及予測という課題を回避する。イノベーターの検出には項目反応理論を用い、推定された顧客ごとの「イノベーターらしさ」の指標であるイノベーター度を利用して予測をする。

1.1 データ説明

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データを使用する。内容は、ある EC サイトにおける 2015 年 4 月 1 日から 2016 年 3 月 31 日までの 1 年分の購買記録と 2016 年 3 月 17 日から同年 3 月 23 日までの期間に当該 EC サイトの顧客に対して行われたアンケート結果からなる。

季節トレンドに関しては当該年度に新作として発売された冬服に着目し、冬の新作商品の購買をイノベーションの採用と捉える。

商品に関するデータのうち、商品の発売日に関する情報は含まれていない。そこで本研究では、データの始まりである 2015 年 4 月 1 日以降に生じた最初の購買価格が正規の値段である商品を当該年度の新作とし、最初の購買価格が割引であった場合は旧作と定義した。

1.2 本研究の流れ

本研究の分析は大きく 2 部構成である。第 1 部では本研究における季節トレンドのイノベーターとなる顧客 (以後、イノベーター顧客) を定義する。イノベーター顧客を定義するにあたって、行動的側面と心理的側面からアプローチする。

まず行動的側面として、購買データを用いて当該年度に最も早く冬の新作を購入した顧客群を早期購買者と定義する。これはイノベーター理論の定義に即しており、早期購買者はイノベーター顧客の候補として位置づけられる。

次に早期購買者が早期に購入した新作商品とそれ以外の新作商品のシーズン中・後期の売り上げを比較す

すぎやま けいた, ながお けいいちろう, いそべ ゆりえ, おか りつこ

早稲田大学大学院文学研究科

〒 162-8644 東京都新宿区戸山 1-24-1

k-ta.sugiyama@moegi.waseda.jp

とよだ ひでき

早稲田大学文学学術院

〒 162-8644 東京都新宿区戸山 1-24-1

toyoda@waseda.jp

受付 17.7.25 採択 17.11.5

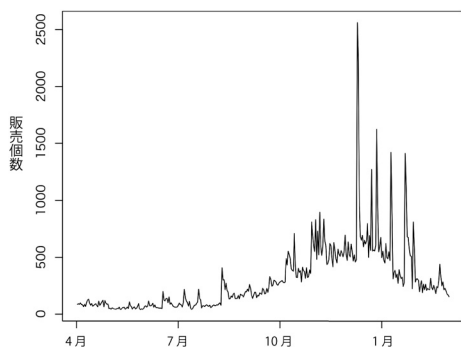


図1 冬服の購買プロット

る。仮に早期購買者に購入された商品が、後の売り上げに関してそれ以外の商品よりも売り上げが高ければ、早期購買者の購買を当該商品の普及の兆しとして期待することができる。

最後に心理的側面として、早期購買群とそれ以外の一般顧客群の2群についてアンケート項目の回答を比較する。特に差の開いた項目を見ることで、早期購買者の心理特性を浮き彫りにすることができる。そのうえで早期購買者がイノベーター顧客として十分に妥当性のある心理特性を有しているか検証する。

第2部では、イノベーター顧客を定義したうえでそれらの事前検出を目的とする。主となる分析には項目反応理論の枠組みで提案されてきたモデルを使用する。予測にあたって、早期購買期間以前のデータを利用すること、解を求めるうえで時間がかからないことに重点を置き、実用に供することのできるモデルを提案する。

2. イノベーターの定義

本節では、103,144人の顧客からイノベーター顧客を定義する。具体的には行動的側面とそれに伴う売り上げの比較、そして心理的側面を検証していく。

2.1 行動的側面から

季節トレンドにおけるイノベーター顧客として、冬の新作商品を早期に購入していることが望まれる。そこで本節ではイノベーター顧客の候補として早期購買者を定義することから始める。

経験的に冬服は最も早いもので晩夏以降に市場に投入される。このことから早期購買者は、早くて8月から9月にかけて冬服を購入する顧客群と予想がつく。図1は冬の新作商品の売り上げ個数を一年間にわたって時系列上に描画したものである。

本格的に冬服が売れ始めるのが10月以降であるのに対し、8月から9月に位置する時期の購買数に小さ

なスパイクが確認できる。このスパイクにおける詳細な注文内容を見ると、多くが割引商品であり、スパイク期間内に大量の冬服がセールで売り出されていたことが推測できる。本研究ではこのスパイクを年度の変わり目と捉え、スパイクの起きた当日から1カ月にわたる期間を早期購買期間と定義した。そして早期購買期間中に冬の新作を購入した顧客、正確には定価で冬服を購入していた顧客を早期購買者と定義する。

早期購買者に該当する顧客は2,221人である。これは全体の人数103,144人の約2.15%であり、イノベーター理論におけるイノベーターの割合である約2.5%にも近い。

2.2 売り上げの比較

次に早期購買期間に購入された冬の新作商品（以後、採用群）と、早期購買期間に購入されなかった冬の新作商品（以後、不採用群）のシーズン中・後期の売り上げを比較する。

採用群に該当する商品は1,600点、不採用群に該当する商品は40,046点であった。本研究では群間の平均値差を検証するうえで、これらの標本数の偏りを考慮した指標として効果量 (Effect Size [2]) を利用する。効果量は1群目と2群目の標本 \bar{x}_1, \bar{x}_2 と、その標本数 n_1, n_2 によって以下で定義される。

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s}$$

ただし

$$s = \sqrt{\frac{s_1^2 + s_2^2}{n_1 + n_2}}, s_k^2 k = 1, 2 = \sum_{i=1}^{n_k} (x_{k,i} - \bar{x}_k)^2$$

である。

効果量は、標準化された2群の差と解釈することができる。すなわち算出された効果量 d の値は標準偏差を単位として平均値がどれだけ離れているかを表す。たとえば $d = 1$ である場合は2群の平均値差は標準偏差一つ分離れていると解釈する。

効果量を各冬服カテゴリと冬服全体に関して求めた結果を表1に示す。なお効果量の値は正の値をとった場合に、選択商品の値が非選択商品よりも大きいと解釈する。結果から選択商品は、手袋を除くすべての商品カテゴリにおいて非選択商品よりもシーズン中・後期に売れていた。全体でも1.42の効果量が確認され、不採用商品の平均売り上げ数が2.52着であるのに対し、採用群の売り上げ数はおよそ6倍の13.87着だった。以上の結果から早期購買者の購買行動は当該商品の普及の兆しとして期待できる。

表 1 選択商品と非選択商品の効果量

商品カテゴリ	効果量
ニット／セーター	1.18
ダウンジャケット／コート	1.80
ストール／スヌード	1.50
マフラー／ショール	0.92
チェスターコート	1.25
手袋	-0.34
トレンチコート	2.73
ダブルコート	1.19
モッズコート	1.19
ムートンコート	2.73
ネックウォーマー	1.91
全体	1.42

2.3 心理的側面から

アンケートデータを利用し、早期購買者の心理特性にイノベーターとしての妥当性を検証する。検証方法として、早期購買者のアンケート回答群とそれ以外のアンケート回答群の2群について回答結果を比較し、特に差の大きい項目を考察することで早期購買者の心理特性を浮き彫りにする。2群の間には早期購買者のアンケート回答群が67人、それ以外のアンケート回答群が3,077人と人数の偏りが大きく存在している。そこでアンケートデータに関しても効果量を同様に利用し、2群の標準化した差を求める。

表2には特に効果量の大きい上位14項目を示した。効果量を算出するにあたってアンケートに含まれていた逆転項目には、効果量の値が大きければ大きいほどイノベーターに質問内容の傾向が強いことを示すように処理を施した。結果から早期購買者は以下の三つの心理特性で特徴づけられた。

2.3.1 洋服を買うお金に糸目をつけない

Q8-18・Q7-8・Q9-4・Q5・Q6-8といった質問項目から、早期購買者は洋服の購買に関してお金に糸目をつけない傾向があるといえる。最も差が見られた質問項目Q8-18では、早期購買者が季節の終盤に行われるバーゲンやセールを利用せず、季節の早期に商品を購入する顧客群であることが心理的にも示唆された。

2.3.2 洋服への高い関心

Q8-17・Q8-22・Q9-1といった質問項目から、早期購買者は洋服への関心が高い顧客であるといえる。イノベーターの購買は普及の兆しとしての側面をもち、顧客にとって魅力的な商品を見分けることが期待される。そのうえで、早期購買者が商品そのものに対して高い関心をもっていることは、早期購買者がイノベーターであるための必要条件といえる。

表 2 効果量（「質問内容」の表記は回答者に提示された実際の質問文を短く改変している。）

質問番号	効果量	質問内容
Q8-18	0.49	バーゲン／セールを利用しない。
Q7-8	0.43	服を買うお金に困らない。
Q6-28	0.39	季節が変わる時に服を購入する。
Q8-17	0.34	家での普段着にも気を使う。
Q1-5	0.32	2015年、海水浴／プールに行った。
Q3-5	0.30	個人能力を重視しない。
Q8-22	0.30	原産国や素材が気になる。
Q9-4	0.30	最近ファッションにかかる金額が増えた。
Q3-11	0.28	専門性を重視する。
Q1-11	0.27	「その他」のイベントに参加した。
Q9-1	0.26	最近ファッションへの関心が増えた。
Q5	0.25	ファッションに使える月額比率が高い。
Q6-8	0.25	洋服が安く買える時に服を買わない。
Q3-7	0.24	最先端／流行を重視する。

2.3.3 流行に敏感

Q6-28・Q3-7といった質問項目から、早期購買者は流行に敏感な顧客であるといえる。特にQ3-7は早期購買者が最先端に価値を感じる顧客であることを示しており、これはイノベーターの心理特性そのものである。

以上の考察より、早期購買者の心理特性にはイノベーターと定義するにふさわしい妥当性があると判断した。

3. イノベーター検出モデル

本節ではイノベーター顧客を事前検出するための予測モデルを提案する。予測には項目反応理論 (Item Response Theory; IRT [3, 4]) を用いる。IRTはテストを作成・評価・運用するために用いられる数理モデルであり、テスト項目への回答パターンをもとに、被験者ごとに仮定される潜在特性や項目ごとの困難度や識別力を推定することができる。

本研究では顧客の属性や購買行動をテストへの回答と捉え、その潜在特性として顧客ごとのイノベーター度を推定する。推定されたイノベーター度は、説明変数としてロジスティック回帰分析に利用し、イノベーター顧客の予測を行う。

3.1 項目反応モデル（尤度モデル）

本研究はイノベーター度の推定に用いる項目として、顧客の属性に関するデータと顧客の購買行動を集計したデータを扱う。具体的な項目は表3である。

これらはいずれも早期購買以前に収集可能であり、イノベーター顧客とそれ以外の顧客の2群において比較的大きな効果量が確認された変数である。以下の項では各項目の回答方式とそれに適用する項目反応モデルを提案する。

表3 推定に用いた項目

項目番号	項目内容
項目1	性別
項目2	商品カテゴリ数
項目3	カラーカテゴリ数
項目4	購買頻度
項目5	お気に入りSHOP登録数
項目6	年代

3.1.1 2母数ロジスティックモデル

項目1の回答パターン \mathbf{u}_1 は顧客の性別を示し、顧客 i ($1 \leq i \leq N$) (N は総顧客数) が男性であるならば $u_{i1} = 0$ を、女性であるならば $u_{i1} = 1$ とした。回答パターンとは、回答データ \mathbf{u}_1 が顧客数分格納された要素数 N のベクトルである。

2値式の項目反応モデルにはロジスティックモデルを採用する。ロジスティックモデルは項目母数の数によって種類が存在し、本研究では2母数ロジスティックモデル (Two Parameter Logistic Model [5, 6]) を用いる。

2母数ロジスティックモデルは、イノベーター度が θ_i であるような顧客 i の、困難度母数 b_1 、識別力母数 a_1 をもつような項目に対する反応を、ベルヌイ分布を用いて

$$u_{i1} \sim \text{Bernoulli}(p_{i1}(\theta_i))$$

$$p_{i1}(\theta_i) = \frac{1}{1 + \exp(-Da_1(\theta_i - b_1))}$$

のように表現する。 D は尺度因子と呼ばれ、項目反応理論の分野では $D = 1.7$ とすることでロジスティックモデルと正規累積モデルとが互いに最も近似する性質を利用する習慣がある。本研究では $D = 1.7$ とした。

3.1.2 段階反応モデル

項目2の回答パターン \mathbf{u}_2 は商品カテゴリ数を表す。商品カテゴリ数とは顧客が早期購買期間より前に購入していた商品のカテゴリ総数である。商品カテゴリ数が多い顧客ほど、さまざまなバリエーションの商品を購入している顧客といえ、イノベーターである可能性が高いと予測される。商品カテゴリ数は最大で24カテゴリ存在し、一般的な項目反応理論のカテゴリ数としては多い。そこで商品カテゴリ数には、以下の処理を施すことにした。

$$\begin{array}{ll} 0 \sim 3 \text{ 個} : u_{i2} = 1 & 4 \sim 7 \text{ 個} : u_{i2} = 2 \\ 8 \sim 11 \text{ 個} : u_{i2} = 3 & 12 \sim 15 \text{ 個} : u_{i2} = 4 \\ 16 \sim 19 \text{ 個} : u_{i2} = 5 & 20 \sim 23 \text{ 個} : u_{i2} = 6 \end{array}$$

項目3の回答パターン \mathbf{u}_3 はカラーカテゴリ数を表す。カラーカテゴリとは商品のメインカラーを18カテゴリに大別したデータであり、カラーカテゴリ数は顧客が早期購買期間より前に購入していた商品のカラーカテゴリの総数である。項目3も項目2と同様にカテゴリの数が多いため、以下のような処理を施した。

$$\begin{array}{ll} 0 \sim 2 \text{ 個} : u_{i3} = 1 & 3 \sim 5 \text{ 個} : u_{i3} = 2 \\ 6 \sim 8 \text{ 個} : u_{i3} = 3 & 9 \sim 11 \text{ 個} : u_{i3} = 4 \\ 12 \sim 14 \text{ 個} : u_{i3} = 5 & 15 \sim 17 \text{ 個} : u_{i3} = 6 \end{array}$$

項目2と項目3は共に段階的な6カテゴリの項目であり、段階反応モデル (Graded Response Model [6, 7]) を適用する。段階反応モデルでは、イノベーター度 θ_i をもつ顧客 i が項目 j ($= 2, 3$) に k ($1 \leq k \leq K$) (K は総カテゴリ数) 以上と反応する確率を $p_{ijk}^*(\theta_i)$ と定義し、これを境界特性曲線 (Boundary Characteristic Curve; BCC) と呼ぶ。BCCを2母数ロジスティックモデルによって表現すると

$$p_{ijk}^*(\theta_i) = \frac{1}{1 + \exp(-Da_j(\theta_i - b_{jk}^*))}$$

となる。またBCCの識別力母数 a_j はカテゴリ間で同値であり、 b_{jk}^* には

$$b_{j1}^* < b_{j2}^* < \dots < b_{jk}^* \dots < b_{jK-1}^*$$

の関係が成り立つ。

BCCをもとにイノベーター度 θ_i をもつ顧客 i の項目 j に対する反応は

$$u_{ij} \sim \text{Categorical}(p_{ij}(\theta_i))$$

$$p_{ijk}(\theta_i) = p_{ijk-1}^*(\theta_i) - p_{ijk}^*(\theta_i)$$

のように表現できる。 $p_{ijk}(\theta_i)$ は項目反応カテゴリ特性曲線 (Item Response Category Characteristic Curve; IRCCC) と呼ばれる。一般的に段階反応モデルをほかのモデルと並行して考察する際は、BCCよりもIRCCCのほうが解釈がしやすい。そこで本研究では b_{jk} を困難度母数と呼び、BCC固有の母数 b_{jk}^* には * を付与して困難度母数と区別する。

3.1.3 ポアソンモデル

項目4の回答パターン \mathbf{u}_4 は購買頻度を示す。購買頻度は早期購買期間より前に顧客が購入した累計数を意味する。

項目5の回答パターン \mathbf{u}_5 はお気に入りSHOP登録数を示す。お気に入りSHOPの登録数は顧客の洋服への

関心の強さを反映していると考えられる。

これらの項目は、顧客の一定時間内に成功した試行回数と捉えることでポアソンモデル (Poisson Counts Model [6, 8, 9]) を適用できる。具体的に項目 4 では顧客の購買行動を、項目 5 ではお気に入り SHOP の登録行動を成功した試行として捉えることができる。ポアソンモデルでは、顧客 i の項目 $j (= 4, 5)$ における反応は

$$u_{ij} \sim \text{Poisson}(\lambda_{ij}(\theta_i))$$

$$\lambda_{ij}(\theta_i) = \exp(\theta_i - b_j)$$

のように表現できる。

3.1.4 名義反応モデル

項目 6 の回答パターン \mathbf{u}_6 は顧客の年代を示し、顧客 i の年代は

10 代以下 : $u_{i6} = 1$	20 代前半 : $u_{i6} = 2$
20 代後半 : $u_{i6} = 3$	30 代前半 : $u_{i6} = 4$
30 代後半 : $u_{i6} = 5$	40 代以上 : $u_{i6} = 6$

のように表現される。項目反応モデルの適用を考慮するうえで、年齢の上昇や下降とイノベーター度との間に関連性を確認することができなかつた。そこで \mathbf{u}_6 は年代の区別をする名義尺度として処理をする。名義尺度に対する項目反応モデルには名義反応モデル (Nominal Response Model [6, 10]) を使用する。 \mathbf{u}_6 に対する顧客の反応と反応確率は

$$u_{i6} \sim \text{Categorical}(\mathbf{p}_{i6}(\theta_i))$$

$$p_{i6k}(\theta_i) = \frac{\exp(z_{i6k})}{\exp(z_{i61}) + \dots + \exp(z_{i66})}$$

のように表現される。 k は回答のカテゴリ番号を意味し、顧客 i が六つの年代カテゴリの中から k を選択する確率を表している。式中の z_{6k} は傾き係数 α_k と位置母数 β_k を用いて

$$z_{i6k}(\theta_i) = \alpha_k(\theta_i - \beta_{6k})$$

のように表現される。このうち位置母数 β_{6k} は、通常の項目反応理論における困難度母数 b と同値になることが知られており、共通尺度上で比較することができる母数である。

3.1.5 イノベーター予測モデル

推定されたイノベーター度 θ からロジスティック回帰分析を行う。 \mathbf{u}_7 を顧客 i がイノベーター顧客であるときに 1、そうでない場合に 0 をとるダミー変数とすると、予測式は以下のように表現できる。

$$u_{i7} \sim \text{Bernoulli}(p_{i7}(\theta_i))$$

$$p_{i7}(\theta_i) = \frac{1}{1 + \exp[-(c_1 + c_2\theta_i)]}$$

ここで c_1 と c_2 は回帰係数である。

3.2 推定にあたって

項目反応モデルは θ が所与である場合、項目反応は互いに独立であるという局所独立の仮定 (local independence assumption) を導入する。そのためイノベーター検出モデルも前提として局所独立を仮定する。また本研究ではモデルの実用性や解釈の容易さを重視し、変数間に交互作用は組み入れないこととした。

推定に用いるデータは、103,144 人の顧客データから 30,000 人分を無作為抽出して使用する。母数の推定にはソフトウェアである Stan [11] による変分ベイズ法 (Variational Bayes; VB [12]) を利用することで、短時間で多量な母数推定を行う。

3.2.1 事前分布

項目母数や潜在特性の事前分布には、豊田 [3] を参考に、以下を仮定した。

$$a, \alpha \sim \text{LogNormal}(0, 0.5)$$

$$b, b^*, \beta \sim \text{Normal}(0, 2)$$

$$\theta \sim \text{Normal}(0, 1)$$

イノベーター顧客の予測に用いる回帰係数 c_1, c_2 には、十分に広い範囲での無情報事前分布を仮定した。

3.2.2 事後分布

母数 Ψ を $\Psi = [a_1, a_2, a_3, \alpha_{61}, \dots, \alpha_{66}, b_1, b_{21}^*, \dots, b_{25}^*, b_{31}^*, \dots, b_{35}^*, b_4, b_5, \beta_{61}, \dots, \beta_{66}, \theta_1, \dots, \theta_{30000}, c_1, c_2]$ とおく。項目 j の 30,000 人の反応ベクトルを $\mathbf{u}_j (j = 1, \dots, 7)$ とおき (\mathbf{u}_7 はイノベーター顧客を示すダミー変数ベクトル)、それらのベクトルを並べたデータ行列 \mathbf{U} に関する全尤度 $L(\mathbf{U}|\Psi)$ は

$$L(\mathbf{U}|\Psi) = \prod_{i=1}^{30000} \prod_{j=1}^7 f(u_{ij}|\phi_j)$$

と表現される。ここで f は各項目の回答が従う分布関数を意味し、 ϕ_j は項目 j の項目母数とイノベーター度を表している。たとえば ϕ_1 の場合は、 $[a_1, b_1, \theta_1, \dots, \theta_{30000}]$ を意味する。

母数 $\mathbf{a} = [a_1, a_2, a_3]$, $\mathbf{b} = [b_1, b_{21}^*, \dots, b_{25}^*, b_{31}^*, \dots, b_{35}^*, b_4, b_5]$, $\boldsymbol{\alpha} = [\alpha_{61}, \dots, \alpha_{66}]$, $\boldsymbol{\beta} = [\beta_{61}, \dots, \beta_{66}]$, $\boldsymbol{\theta} = [\theta_1, \dots, \theta_{30000}]$, $\mathbf{c} = [c_1, c_2]$ の事前分布は、互いに独立であることを仮定する。

表4 母数の推定結果

母数	EAP	post.sd	95%下側	95%上側
a_1	0.72	0.02	0.69	0.75
a_2	2.41	0.11	2.21	2.63
a_3	2.29	0.09	2.13	2.47
α_{61}	0.18	0.02	0.14	0.23
α_{62}	0.05	0.01	0.03	0.08
α_{63}	0.09	0.02	0.06	0.12
α_{64}	0.13	0.02	0.09	0.18
α_{65}	0.25	0.04	0.18	0.33
α_{66}	0.10	0.02	0.07	0.13
b_1	-0.34	0.12	-0.58	-0.11
b_{21}^*	1.53	0.05	1.43	1.63
b_{22}^*	2.85	0.09	2.68	3.02
b_{23}^*	3.98	0.19	3.64	4.35
b_{24}^*	4.61	0.31	4.08	5.33
b_{25}^*	5.74	0.85	4.65	7.95
b_{31}^*	0.90	0.04	0.82	0.98
b_{32}^*	2.11	0.07	1.98	2.24
b_{33}^*	2.94	0.10	2.74	3.12
b_{34}^*	4.18	0.19	3.83	4.60
b_{35}^*	5.76	0.81	4.60	7.73
b_4	-0.49	0.03	-0.55	-0.42
b_5	-1.22	0.03	-1.27	-1.17
β_{61}	6.68	0.45	5.75	7.51
β_{62}	-3.18	0.84	-4.76	-1.49
β_{63}	-4.62	0.68	-5.82	-3.09
β_{64}	-3.08	0.57	-4.13	-1.92
β_{65}	-1.20	0.34	-1.87	-0.58
β_{66}	-4.44	0.72	-5.82	-2.96
c_1	-4.51	0.14	-4.79	-4.24
c_2	0.97	0.08	0.81	1.15

$$\begin{aligned}
 & f(\mathbf{a}, \mathbf{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{c}) \\
 &= \prod_{j=1}^3 f(a_j) f(b_1) \prod_{j=2}^3 \prod_{k=1}^5 f(b_{jk}^*) \prod_{j=4}^5 f(b_j) \\
 & \times \prod_{k=1}^6 f(\alpha_{6k}) f(\beta_{6k}) \prod_{i=1}^3 f(\theta_i) f(c_1) f(c_2)
 \end{aligned}$$

よって母数 Ψ の事後分布は、

$$\begin{aligned}
 & f(\Psi|U) \\
 & \propto L(U|\Psi) f(\mathbf{a}, \mathbf{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{c}) \\
 &= \prod_{i=1}^{30000} \prod_{j=1}^7 f(u_{ij}|\phi_j) \prod_{j=1}^3 f(a_j) f(b_1) \prod_{j=2}^3 \prod_{k=1}^5 f(b_{jk}^*) \\
 & \times \prod_{j=4}^5 f(b_j) \prod_{k=1}^6 f(\alpha_{6k}) f(\beta_{6k}) \prod_{i=1}^3 f(\theta_i) \\
 & \times f(c_1) f(c_2)
 \end{aligned}$$

となる。

表5 イノベーター度 θ の数値要約

最小値	第1四分位	中央値	平均	第3四分位	最大値
-1.64	-1.04	-0.18	0.04	0.89	3.57

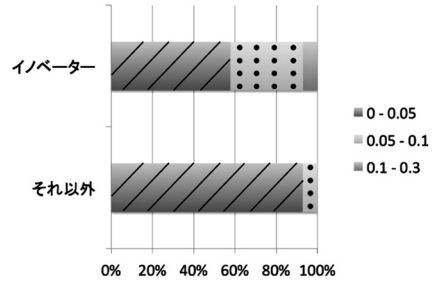


図2 判別結果

4. 結果

4.1 イノベーターの予測

各項目母数と回帰係数の推定結果を表4に示し、推定されたイノベーター度の数値要約を表5に示す。推定結果から以下のような回帰式が推定された。

$$p_{i7}(\theta_i) = \frac{1}{1 + \exp[-(-4.51 + 0.97\theta_i)]}$$

$p \geq 0.1$ をイノベーターと判別した場合の判別率は28.03%となった。これは無作為抽出した場合に期待される判別率2.15%の13倍を上回る予測精度といえる。

図2はイノベーター顧客とそれ以外の顧客の判別結果を示した帯グラフであり、右側にはそれぞれの群が該当する p の範囲を示している。それ以外の顧客の9割近くがイノベーター確率5%以下と判別されているのに対し、イノベーター顧客でイノベーター確率が5%以下と判別されるのは6割近くにとどまっている。

4.2 項目反応関数 (IRF) の描画

項目反応理論では項目反応関数 (Item Response Function; IRF) を描画することで、項目と潜在特性の関係を視覚的に考察することができる。一般的に IRF では横軸に潜在特性を配し、縦軸に項目、あるいは各カテゴリへの回答確率を配す。図3から図8は、各項目について IRF を描画した結果である。

図3の項目「性別」に関しては、イノベーター度がプラスの領域において女性の回答確率が高くなっている。このことからファッション EC サイトのイノベーターは女性である傾向がわかる。

図4と図5の項目「商品カテゴリ数」と項目「カ

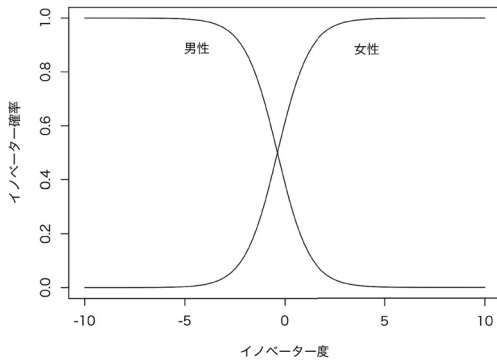


図3 「性別」のIRF

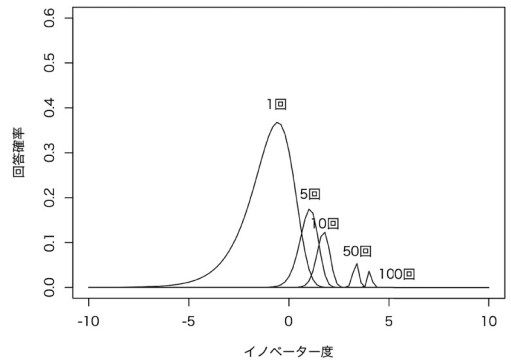


図6 「購買頻度」のIRF

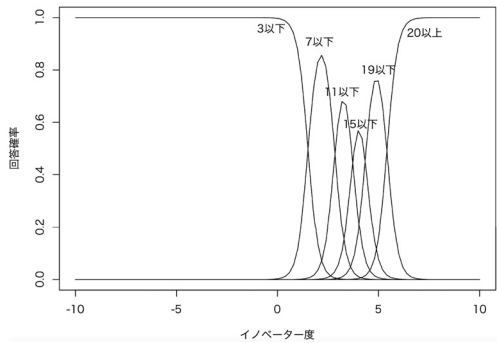


図4 「商品カテゴリ数」のIRF

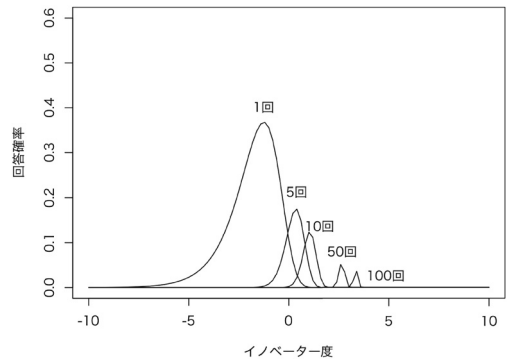


図7 「お気に入りSHOP登録数」のIRF

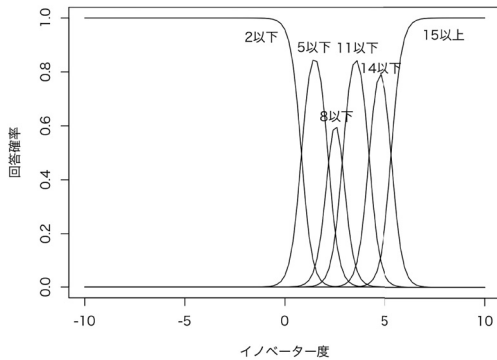


図5 「カラーカテゴリ数」のIRF

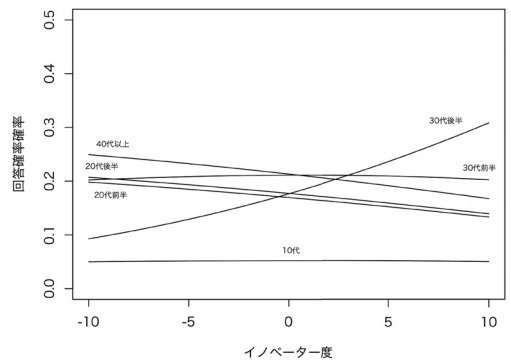


図8 「年代」のIRF

「カラーカテゴリ数」に関しては、いずれもイノベーター度が高い領域において、鋭い曲線を描いている。つまり「カラーカテゴリ数」は、イノベーター度の高い顧客をよく識別する項目といえる。

図6と図7の項目「購買頻度」と項目「お気に入りSHOP登録数」に関しては、左から試行数1回・5回・10回・50回・100回のIRFを示している。二つの項目を比較すると、「購買頻度」のIRFの方が全体的に右寄りに描かれていることがわかる。このことから1回の購買と1回のお気に入りSHOPの登録では、1回の

購買のほうが行動に踏み切るうえでのハードルが高いといえる。

図8は、項目「年代」に関してIRFを描画した結果である。全体的に穏やかな曲線を描いているが、30代後半のIRFに注目するとイノベーター度の高い領域において急激に回答確率を上げている。このことから、30代後半の顧客はイノベーター確率が高い傾向にあるといえる。

5. 結論

本研究の目的はファッション EC サイトのデータからイノベーター顧客を定義・検証することであった。イノベーター顧客を予測する際には、前提として将来普及する商品の予測という課題がある。そこで本研究では、毎年確実に普及すると考えられる「季節」に着目することで、イノベーター顧客の予測を実用的な場面で実現することができた。

イノベーター顧客を定義する際には、冬服の早期購買という行動に注目した。またアンケートデータを利用して心理的側面からも定義の妥当性を確認した。

実際の予測には項目反応理論をベースにしたイノベーター検出モデルを構築した。その結果、自然条件において顧客を抽出したときに比べて、10 倍以上高い精度でイノベーター顧客を検出することに成功した。

今後の課題として、イノベーター検出モデルの妥当性を検証するべきである。当モデルは変数の独立を前提に構築され、マーケティングの実用的見地から分析コストや解釈の容易さの点で優れている。しかし、独立を仮定せずに交互作用を組み入れたモデルなども考えられ、それらのモデルとの間で選定作業を行う必要がある。モデル選定を行うことで妥当性検証だけでなく、今回のモデルでは説明しきれなかった顧客行動を新たな知見として期待できる。

また本研究では 1 年間のデータから冬シーズンに注目して分析を行っている。イノベーター顧客の予測に有効な変数は、年度や季節によっても異なる可能性

があり、項目選定に関しては再検討の余地がある。そのためほかの年度やほかの季節においても同様の分析を行い、今後はモデルの汎用性も検証するべきである。

参考文献

- [1] E. M. Rogers, *Diffusion of Innovations*, 5th edition, Free Press, 2003.
- [2] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd edition, Routledge, 1988.
- [3] 豊田秀樹, 『項目反応理論 [入門編]』, 第 2 版, 朝倉書店, 2012.
- [4] F. M. Lord, “A theory of test scores,” *Psychometric Monograph*, **7**, pp. 1–83, 1952.
- [5] A. Birnbaum, “Some latent trait models and their use in inferring an examinee’s ability,” *Statistical Theories of Mental Test Scores*, F. M. Lord, and M. R. Novick (eds.), pp. 397–472, 1968.
- [6] 豊田秀樹, 『項目反応理論 [理論編]』, 朝倉書店, 2005.
- [7] F. Samejima, “Estimation of latent ability using a response pattern of graded scores,” *Psychometrika*, **34**, pp. 100–114, 1969.
- [8] G. Rasch, *Probabilistic Models for Some Intelligence and Attainment Tests*, Danmarks Paedagogiske, 1960.
- [9] G. N. Masters and B. D. Wright, “The essential process in a family of measurement models,” *Psychometrika*, **49**, pp. 529–544, 1984.
- [10] R. D. Bock, “Estimating item parameters and latent ability when response are scored in two or more nominal categories,” *Psychometrika*, **37**, pp. 29–51, 1972.
- [11] Stan, “Run Stan’s variational algorithm for approximate posterior sampling,” <http://mc-stan.org/rstan/reference/stanmodel-method-vb.html> (2017 年 9 月 26 日閲覧)
- [12] 藤本健治, 佐藤彰記, 福永修一, “変分ベイズ法に基づいた状態空間モデルのシステム同定,” 計測自動制御学会論文集, **48**, pp. 102–108, 2013.