

ECサイトにおける顧客の多様な探索行動を考慮した購買行動の推定

経営システム工学専攻 及川裕之

1 研究背景と目的

従来の小売店などにおけるプロモーション活動は、マスマーケティングの考えの下、顧客に対して共通プロモーションを行っていた。しかし、ECサイトにおいては取り扱う商品数やコンテンツの増加に伴い、各顧客の興味を示す商品やサイトの利用状況が多様化していることから、マスマーケティングは、サイトを利用するすべての顧客に購買を促す効力はない。そのため、ECサイトを運営する企業は、各顧客の特徴を考慮した one to one marketing によるプロモーションの策定を行っている。one to one marketing によるプロモーションを行う際に重要なのは、事前のサイト探索行動からどの顧客が近い将来に商品を購入するかを、適切に予測することである。しかしながら、顧客の行動は多様であり商品の購買時期を予測することは容易でない。

本研究はゴルフ場予約やレッスン動画の配信などの様々なサービスを配信しているゴルフポータルサイトを対象にした。また、各顧客にむけたプロモーションの示唆を行うために、各顧客の購買前における事前探索行動を考慮したうえで、近い将来にゴルフ商品の購買有無を推定する二値判別のモデルを提案する。作成したモデルから顧客ごとに商品を購入するきっかけとなる要因と、要因の重要度を抽出することを試みる。また、ONE to ONE マーケティングの考えの下、得られた分析結果から各顧客への最適なプロモーションを提案を試みる。

2 関連研究と本研究の位置づけ

消費者行動の研究分野でも利用状況や地域などの特性から消費者の異質性を考慮した研究が盛んになっている [1]。

Neelamegham ら [2] は、地域差・ジャンルの差を考慮した階層ベイズポアソンモデルを用いて国内外の新しい映画の売上予測をした。また、山口 [3] は、以前からの訪問回数の差によるビジターの異質性を考慮した階層ベイズモデルを応用してサイ

ト訪問間隔をモデル化した。

しかし、階層ベイズモデルは、逐次的に変数の選択が出来ないことから、事前に変数の選定が必要とされるが、変数の選定とモデル構築を行う分析のフレームワークを提唱しているような研究はまだ少ない。本研究では、事前に変数の選定を行い、選定された変数によりモデルを構築することで、新たな分析のフレームワークを提案する。

また、予測精度の算出している異質性を考慮した関連研究では、得られた個別パラメータを事後平均値化し、検証を行っている。また、実務では新規の顧客など、学習で得られていない未知の顧客の購買予測もする必要がある。しかし、パラメータ推定に用いていない顧客に対する判別予測をするような研究はまだ著者が知る限りは存在しない。本研究では、個々のパラメータ推定を行うが、加えて、推定した個別パラメータをモデルで学習していない顧客にも購買の有無を予測を試み、その精度を従来のモデルと比較する。

3 データの概要

本研究で用いるデータはゴルフポータルサイトを運営している企業より提供されたデータである。本研究のモデルを構築するために、ECサイトでの顧客のページ遷移を記録したアクセスログデータ、ECサイトでのゴルフ商品の購入履歴を記録した購買履歴データ、ECサイトでのゴルフ場の予約履歴を記録した予約データ、ECサイトに会員登録している顧客を記録した会員データ、ECサイトでの商品購入などに使用可能なポイントの増減を時系列で記録したポイント履歴を記録したポイント履歴データを用いる。

4 モデルの概要

本研究で商品の購買有無を推定する判別モデルを作成するために、学習期間とモデルの検証期間に分割した。学習期間は 2015 年 4 月 1 日～2016 年 3 月 31 日、検証期間は 2016 年 4 月 1 日～2016 年 4 月 30 日で、期間内のセッションごとに各デー

表 1: 本分析に用いる変数の一覧

変数	変数のカテゴリ	変数名	内容	変数の数
目的変数	---	購買有無	購買前の 1 週間のセッションを 1, それ以外を 0	---
説明変数	Web 閲覧	item_detail golf_course golf_course_detail	各コンテンツにおけるセッション時から直前 1 週間の閲覧時間と 2 週間目の閲覧時間の増加量	9×3
		lesson gear_inform compe_event	各コンテンツにおけるセッション時から直前 2 週間目の閲覧時間と 3 週間目の閲覧時間の増加量	
		my_page hot_price news	各コンテンツにおけるセッション時から直前 3 週間目の閲覧時間と 4 週間目の閲覧時間の増加量	
		campaign	セッション時から 1 か月のキャンペーンの応募数	1
	購買実績	sales_amount_first sales_amount_later sales_num_first sales_num_later	セッション時から直近 1~6 か月間の購買金額の合計 セッション時から直近 7~12 か月間の購買金額の合計 セッション時から直近 1~6 か月間の購買個数の合計 セッション時から直近 7~12 か月間の購買個数の合計	4
		rsv_amount_first rsv_amount_later rsv_flag_Sum	セッション時から直近 1~6 か月間の予約金額の合計 セッション時から直近 7~12 か月間の予約金額の合計 セッション時におけるゴルフ場の予約件数	3
		holding_points_sum holding_points_L	セッション中の顧客の合計保有ポイント セッション中の顧客の合計保有限定ポイント	2

タを作成した。

また、本研究では、EC サイトの売上を 90% 占めているカテゴリのうち、データ提供元の現場社員と重要な商品カテゴリについてヒアリング調査を行い、アイアン (単品)、アイアン (セット)、アウター、シューズ、ボール、ロングパンツ、半袖シャツを対象にそれぞれの判別モデルを作成した。

4.1 データの変数

顧客のコンテンツの閲覧時間と購買実績と予約実績ポイントの保有状況をベースに作成した変数を以下の表 1 に示す。

4.2 分析の流れ

本研究では近い将来に各顧客における商品の購買有無を判別するモデルを作成するために階層ベイズロジットモデルを用いた。提案するモデルはロジスティック回帰モデルの枠組みで表現する。(1) 式に、購買確率 p_{it} を提案モデルとして以下に示す。

$$p_{it} = \Pr\{y_{it} = 1\} = \frac{\exp\{u_{it}\}}{1 + \exp\{u_{it}\}} \quad (1)$$

(2) 式が具体的に、あるセッション t における顧客 i の購買に対する効用 u_{it} を示す。

$$u_{it} = X_{it}^T B_i \quad (2)$$

X_{it} はセッション t における顧客 i の購買有無に関する切片項を含む説明変数ベクトルであり、 B_i は

各顧客の切片項を含んだパラメータベクトルである。また、顧客の異質性を考慮するため、パラメータは顧客ごとに異なるとする。

モデルから得るパラメータを推定するために、各パラメータに事前分布を仮定し、分布から乱数発生を繰り返すシミュレーションを行う。シミュレーション方法には No-U-Turn sampler を用いた。

本分析では、パラメータベクトル B_i が (3) 式のように多変量正規分布に従うと仮定する。

$$B_i \sim MVN(\Delta, \Sigma_B) \quad (3)$$

加えて、超事前分布として (3) 式における Δ と Σ_B には、それぞれ裾の広い一様分布と逆ウィシャート分布を仮定し、それぞれの (4) 式のように表す。

$$\Delta \sim U(0, 100), \Sigma_B \sim IW(v_0, V_0) \quad (4)$$

上記のモデルを用いた分析の流れを図 1 に示す。階層ベイズモデルは、一般的な回帰モデルのように逐次的に変数選択ができないため、事前に説明変数を選定する。そこで、本分析では、ランダムフォレストを用いて、変数の重要度を算出し、判別モデルに用いる変数選択を行う。その後、階層ベイズロジットモデルを用いた提案モデルとロジスティック回帰モデルを用いた比較モデルを作成する。そして、パラメータを推定していない顧客の予測を行うために、パラメータを推定した顧客を含めて階層型クラスター分析を行い、顧客をク

ラスタリングする。最後に、各モデルを作成して得たパラメータを用いて予測精度を評価する。

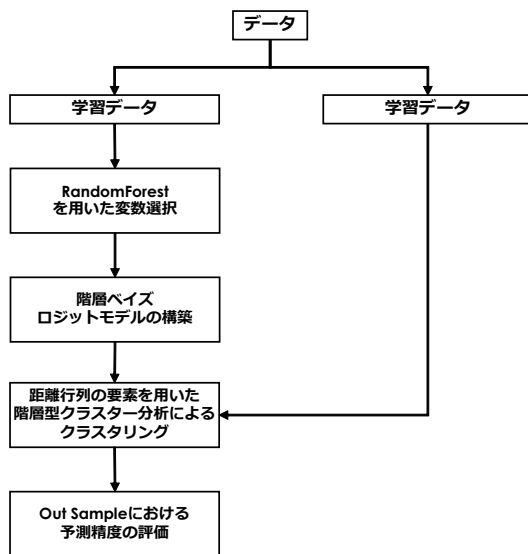


図 1: 本研究における分析の流れ

また、目的変数を購買前1週間のセッションを1、そうでないもの0と設定した。これらの2値の目的変数に対して、前節で作成した説明変数がどのような特徴を示すのかを、顧客ごとに把握する。また、本研究では、モデルの構築を行うために、4000回のIterationのうちバーンイン期間を800回とし、後半の3200回で得られた各顧客のパラメータを使用する。

4.2.1 学習データセットの作成

本研究に用いる階層ベイズロジットモデルは変数の数が多いほどパラメータ推定のためのシミュレーションの計算時間がかかる。加えて、本分析に用いる説明変数の全てが有意になるとは限らないため、モデルを作成する前にランダムフォレストを行い、結果から算出される重要度を用い、重要度の高い上位10個の変数をあらかじめ選択する。

4.2.2 検証データセットの作成

検証データにおいて顧客の購買の有無の判別する際に、パラメータを推定していない顧客がいることが考えられる。そのため、階層型クラスタ分析を用いて、パラメータを推定した顧客とそうでない顧客を含め、購買実績とゴルフのプレー実績によりクラスタリングを行う。クラスタリングの結果から、個別パラメータを推定していない顧客には、同じクラスターにおける顧客の個別パラメータの事後平均値で購買有無の判別予測をする。

5 モデルの結果

顧客の異質性を考慮した購買有無を判別する提案モデルと比較モデルの検証データに対する予測精度、各モデルの概要と結果について述べる。

本分析では、各モデルから算出される購買確率が0.5以上の場合を購買、0.5未満の場合を非購買と予測されたとし、実際の検証データに対して正しく判別出来ているかを精度として採用した。以下の表2に各カテゴリにおける提案モデルと比較モデルの正答率を示す。

表 2: 検証データに対する各モデルの正答率

カテゴリ	提案モデル	比較モデル
アイアン (単品)	0.735	0.563
アイアン (セット)	0.667	0.450
アウター	0.766	0.509
シューズ	0.728	0.565
ボール	0.534	0.219
ロングパンツ	0.672	0.531
半袖シャツ	0.654	0.278

表2から、どのカテゴリにおいても比較モデルに対する提案モデルの正答率が高いことわかる。

本研究では推定したパラメータの収束具合を評価するために、収束判定指標に \hat{R} を用いる。この指標は1.1以下の場合だとパラメータが収束していると評価できる指標である。また、パラメータの有意性を判定する指標には疑似 t -値を用いる。疑似 t -値はパラメータの事後平均値/ $\sqrt{\text{標準偏差}}$ で求めることができる。

本要旨では、表3にアイアン (単品) の結果を示す。提案モデルで推定した個別パラメータの事後平均値と括弧の左右の値はそれぞれ2.5%、また97.5%分位点である。次に \hat{R} を用いて得られた顧客数の個別パラメータのうち収束したパラメータの割合と疑似 t -値を用いて推定した顧客数の個別パラメータのうち有意と判定されたパラメータの割合を示す。比較モデルは係数と p 値を示す。

表3より、提案モデルのパラメータは十分に収束していることがわかる。

また、それぞれの分位数の間に0を含むパラメータが存在することから、購買確率に負の影響を与える顧客がいることから、購買行動に影響を与える要因の個人差が大きいことがわかる。

6 考察

本分析では、提案モデルと比較モデルで正答率を算出した結果、提案モデルが比較モデルより高

表 3: アイアン（単品）における提案モデルと比較モデルの概要

変数名	提案モデル			比較モデル	
	事後平均値	収束したパラメータの割合	有意なパラメータの割合	係数	p 値
(Intercept)	0.398 (-0.222, 0.963)	1.000	0.000	-0.019	0.778
sales_amount_first	-0.192 (-1.245, 0.729)	1.000	0.109	-0.004	0.661
sales_amount_later	-0.122 (-1.070, 0.854)	1.000	0.091	0.003	0.710
sales_num_first	-2.320 (-2.648, -1.913)	1.000	0.218	-0.101	0.228
sales_num_later	-0.539 (-0.851, -0.176)	1.000	0.151	0.010	0.920
item_detail_times_2.1	0.354 (-0.628, 1.346)	1.000	0.085	0.080	0.000 ***
item_detail_times_3.2	0.159 (-0.957, 1.265)	1.000	0.100	0.017	0.464
item_detail_times_4.3	0.129 (-1.028, 1.150)	1.000	0.063	-0.013	0.531
my_page_times_2.1	0.224 (-0.497, 0.926)	1.000	0.202	0.094	0.083
news_times_2.1	-0.027 (-0.795, 0.887)	1.000	0.202	-0.033	0.381
holding_point_sum	0.208 (-0.380, 0.799)	1.000	0.148	0.071	0.090

0.1%有意***, 1%有意**, 5%有意*, 10%有意.

い正答率を算出した。このことから、カテゴリ内でも顧客の購買行動の多様化が示唆され、ゴルフECサイトでは、各顧客にプロモーションをすることが重要である考えられる。

アイアン（単品）モデルで得られた個別パラメータで、図2のパラメータ間の散布図を作成し、カテゴリ内の各顧客の特徴を考察し、各顧客に対するマーケティング・アプローチを示唆する。

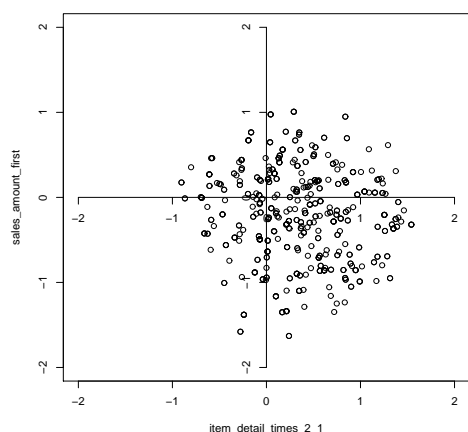


図 2: アイアン（単品）のパラメータの散布図

図2より、第四象限 (item_detail_times_2.1(+), sales_amount_first(-)) の顧客は直近の購買は少ないが、商品の閲覧が多いため、アイアン（単品）の購買を行うと考えられる。よって、直近の商品詳細ページの閲覧傾向を観察してアイアン（単品）のレコメンドをすることが効果的であると考えられる。一方で、第二象限 (item_detail_times_2.1(-), sales_amount_first(+)) の顧客は、直近で商品を購入しているが、アイアン（単品）の購買においては商品詳細ページをあまり見込まないと考えられる。よって、実店舗や他のサイトなどでクラブを閲覧して購入を検討している可能性があると考えられる。そ

のため、購買が盛んな時期にクラブ限定のクーポンなど配信して囲い込むことが効果的だと考える。

7 結論

本研究では、階層ベイズロジットモデルを用いて、顧客の多様な探索行動を考慮した購買有無の推定を行うモデルを作成した。また、分析では、パラメータを推定していない顧客に対しても購買有無を予測し、提案モデルが高い正答率を算出した。更に、得られた個別パラメータから各顧客へのプロモーションを示唆した。

本分析では、変数の選定において、カテゴリ間で同一の変数が選択された。そのため、各カテゴリの特徴を考慮した変数を作成することで、パラメータの特徴抽出や正答率向上に寄与すると考える。また、ゴルフ商品は単価の差が大きく、高価な商品の購買履歴は各顧客でぐくわずかになってしまふ。そのため、日配品などを扱うスーパーマーケットなどを対象に同様の分析を試す必要がある。

参考文献

- [1] 阿部誠. “RFM 指標と顧客生涯価値:階層ベイズモデルを使った非契約型顧客関係管理における消費者行動,” 日本統計学会誌 **37**(2) (2008): 239–259.
- [2] Neelameqham, R. and Chintagunta, P. “A Bayesian Model to Forecast New product Performance in Domestic and International Markets,” Marketing Science **30**(1) (1999): 115–136.
- [3] 山口景子. ”頻度の時間変化を考慮した階層ベイズモデルによるウェブサイト訪問行動の分析,” マーケティング・サイエンス **22**(1) (2014): 13–29.