

ECサイトにおける購買予兆発見モデルの提案

久松 俊道, 外川 隆司, 朝日 弓未, 生田目 崇

顧客が商品を購入する際の購買前の商品検索・比較行動は個人差があり、どのように購買に至るかを分析することは、小売業者にとってはマーケティング活動の新たな機会となる。従来こうした購買前行動データは取得できなかったが、ECサイトではアクセス・ログ・データにより、これらを把握することができる。本稿では、ECサイトにおける訪問者の購買前閲覧行動から購買のタイミングを予測するモデルを提案する。また、実際のアクセス・ログ・データと購買履歴データの分析結果から、購買予兆発見とマーケティング戦略について考察する。

キーワード：閲覧行動パターン、購買予兆発見、潜在クラスモデル、ロジット・モデル

1. はじめに

インターネットの普及とともに、EC (Electronic Commerce) サイトで商品が広く購買されるようになった。こうした購買チャネルの変化に伴い、購買行動のプロセスについても従来のAIDMA理論のような知覚から購買に至るモデルを拡張し、AISASに代表されるような情報収集や情報共有といった、購買前後の行動を視野に入れたプロセス・モデルが提唱されるようになってきた。なお、AISASでは「商品の認知、興味の喚起、商品の検索、商品同士の比較、購入の検討、購入、満足度の情報共有」というプロセスを経る [4]。インターネット・サイトのアクセス・ログ・データには「どのマシンから、いつ、どのページを見たのか」というデータが含まれており、さらに会員登録をしている場合には「だれが」訪問したかも把握できる。そして、訪問者の商品の検索や商品同士の比較などといった具体的な閲覧行動の把握が可能となった。したがって、アクセス・ログ・データとサイト会員の実際の商品購買データを組み合わせることで、どのように検索・比較しながら購買に至ったかという購買プロセスを分析することができる。

こうした視点から、ECサイトの運営者にとっては、アクセス・ログ・データなどの蓄積されていくデータを用いて、サイト会員の普段の閲覧行動からどのようなタイミングで購買が発生するかがわかることはマーケティング上重要な情報となりうる。サイト会員の購買状況を予測することができれば、プロアクティブなマーケティング施策を実施でき、よりの確に顧客の購買をとらえることができる。例えば、訪問者ごとに購買を事前に察知し、そのタイミングで適切な商品を勧めるといった、個別のマーケティング・アプローチが可能となる。

アクセス・ログ・データから訪問時の購買確率を推定するモデルに関する先行研究として、Moe and Fader [8] や Van den Poel and Buckinx [2] が挙げられる。Moe and Fader [8] はサイトでの「購買履歴」や「最終購買日からの閲覧回数」から購買が起きる確率を推定しており、Van den Poel and Buckinx [2] はサイトの「訪問頻度」や「購買頻度」、「ページアクセス数」などから購買が起きる確率を推定するモデルを提案している。ほかにも、小池ら [5] は、サイトへのアクセス・パターンを基に、購買パターンを解析し、重要顧客を発見するフレームワークを提案している。しかし、サイト会員ごとに商品購買前の商品検索・比較行動に違いがあり、それに伴い閲覧行動にも違いがあると考えられる。先行研究ではこうしたサイト会員の普段の閲覧行動の個人差は考慮されていない。

また、小池ら [5] の研究も、クラス判別のみを目的としており、閲覧パターンから購買行動に関する知見を誘発しようというものではない。ECサイト運営者にとってサイト会員のアクセス時の購買確率を時点ごとに推定することは、購買のタイミングを事前に知るこ

ひさまつ としみち
東京理科大学大学院工学研究科
〒162-8601 東京都新宿区神楽坂 1-3
とがわ たかし
(株) ベネッセコーポレーション
〒163-0411 東京都新宿区西新宿 2-1-1
あさひ ゆみ
静岡大学大学院工学研究科
〒432-8561 静岡県浜松市中区城北 3-5-1
なまため たかし
専修大学商学部
〒214-8580 神奈川県川崎市東三田 2-1-1

とができるため、効果的なマーケティング・アプローチにつなげることが可能となると考えられる。

以上に述べたような視点から、本稿では、ECサイトにおけるサイト会員の閲覧行動を考慮した購買予兆の発見モデルを提案する。提案したモデルを実データを用いて検証し、その効果を論じる。本稿で提案するモデルは、ECサイト運営者がサイト会員の購買タイミングを事前に把握することで、適切なタイミングでマーケティング・アプローチを行える示唆を与えることを目的とする。

2. 使用するデータの概要

本稿で使用するデータは、経営科学系研究部会連合協議会が主催した、平成23年度データ解析コンペティションにおいて、株式会社ゴルフダイジェスト・オンラインから提供された同社のECサイトに関するデータである。

データは2010年7月1日～2011年6月28日の約1年間のゴルフ用品販売ECサイトにおけるアクセス・ログ・データおよび購買データである。なお、アクセス・ログ・データにはセッションID、User ID、日付、時間、滞在時間、流入経路、入口サービス、離脱サービス、全PV（ページ・ビュー）数、受注IDが含まれていた。購買データには受注ID、User ID、受注日、商品ID、大分類、中分類、小分類、受注数量、受注金額、受注使用ポイントが含まれていた。サイト会員情報にはUser ID、性別、生年、都道府県、ハンディキャップ、メルマガ登録、会員登録日が含まれていた。

なお、提供されたデータのサイト会員数は16,116名、購買件数は4,384件、ページアクセス数は1,561,186ページであった。

3. 分析の流れ

分析対象は、全サイト会員の中から次節で説明するように前半6カ月と後半6カ月の両期間で購買実績があり、かつ購買回数が4回以上のサイト会員に絞った。これは購買予兆を発見するという目的から、ある程度の購買実績があることが必要と判断したためである。なお、1年のデータを前半と後半に分割したが、日次のセッション数と購買回数について集計したところ、1回の購買あたりのセッション数は期間を通して大きな変動はなかった。ゴルフの季節性に起因して、実際には冬季はセッション数、購買数とも減少するが、後に説明する本稿の分析においては、購買に影響を与える

要因として、個別アクセス数を用いているため、こうした季節性は分析モデルに含まれると考えられる。

次に、分析対象としたサイト会員の閲覧パターンを分類するために、サイト会員のセグメンテーションを行う。ただし、個人差を考慮するために潜在クラスモデルを用いてサイト会員の購買前の閲覧行動を用いて分類する。そして、ロジット・モデルを用いてサイト会員の閲覧行動を考慮した各セッション時の購買予兆を発見するモデルを提案する。最後に、作成したモデルを検証データに当てはめて、実際の購買状況と推定された購買確率の比較を行い、モデルを評価するとともに購買予兆の発見について考察する。

4. 分析対象サイト会員の決定

使用データの前半6カ月のデータをモデルの学習用データとし、後半6カ月のデータをモデルの検証用データとする。

ただし、学習用データ期間、検証用データ期間のいずれにおいても購買を行っており、購買回数が4回以上のサイト会員を分析対象としている。この結果、分析対象サイト会員数は259人となり、その平均年齢は約49歳であった。全サイト会員の平均年齢が約44歳であることから、購買回数が多いサイト会員は多少高齢に偏っていることがわかる。データに含まれるサイト会員数から見ると分析対象のサイト会員数は少ないが、実際の購買者は1,405人であり、そのうち645名は1回限りの購買であった。

5. サイト会員の閲覧行動把握と分類

サイト会員の購買前の閲覧行動パターンを把握し、行動パターンごとにサイト会員を分類するために、閲覧行動を変数とした潜在クラスモデルを用いてサイト会員を分類する。潜在クラスモデルとは、観測変数の背後に離散的分布を持った潜在変数を仮定し、各クラスとの類似度に基づいて所属確率を求めながらクラス分けする手法である[1]。潜在クラスモデルにおいては、サイト会員の購買前の閲覧行動を説明変数として用いる。具体的には、各サイト会員の購買前の4週間（29日～35日前）、3週間（22日～28日前）、2週間（15日～21日前）、1週間（8日～14日前）、5～7日前、3～4日前、1～2日前それぞれについての平均セッション時間、平均セッションPV数、平均アクセス数を集計して用いた。今回対象としたゴルフ用品の購買については、プレイやその予約に合わせて購買したり、ま

表 1 潜在クラス分析の BIC 値の比較

クラス数	BIC
2	58429.76
3	58195.08
4	58924.08
5	59404.42

た非計画的に購入するなどいくつかの購買生起パターンがあると想定される。したがって、購買前の閲覧パターンを把握するためにいくつかの期間に区分した。また、全体としては購買直前に閲覧回数が増えていたため、最後の1週を5~7日前、3~4日前、1~2日前とさらに細かく分類した。

クラス数が2から5について潜在クラス分析を行い、BIC (Bayesian Information Criterion) を基準として比較した結果 (表 1)、クラス数は3とした¹。

クラス数が3のときの分析結果として、各クラスの前平均セッション時間、前平均セッション PV 数、前平均アクセス数を1日平均に換算し、図 1~3 に示す。

サイト会員が各クラスに占める割合は、クラス1から順に、37%、35%、28%であった。

各クラスの特徴を以下にまとめる。

クラス1： 図 1、2 よりセッション時間やセッション PV 数は購買1週前から4週前が近づいても大きく増加することはなく、あまり変化しない。1~2日、3~4日、5~7日前のセッション時間と PV 数は購買が近づくにつれて増えている。図 3 よりアクセス数はほかの2つのクラスより多く、購買が近づくにつれて増える。したがって購買の意図と関係なく、日常的にサイトにアクセスしており、さらに購買が近づくにつれてセッションページを閲覧する量と時間も増える傾向があると言える。

クラス2： 図 1、2 よりセッション時間やセッション PV 数、アクセス数はいずれも少なく、セッション時間や PV 数は購買が近づくにつれてわずかながら増加する。したがって、購買する商品をあらかじめ決めていられる。

クラス3： 図 1、2 よりセッション時間やセッション PV 数が多く、購買が近づくにつれて増加する。1~2日、3~4日前のセッション PV 数に大きな違いはないが、セッション時間は増加している。図 3 よりアクセス数は少ない。したがって、一度のアクセスで大量にページを閲覧する傾向があり、購買直前には1

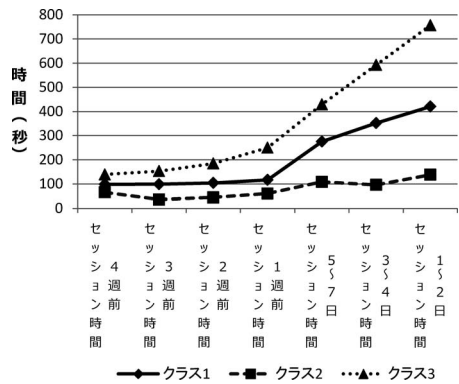


図 1 各クラスの前平均セッション時間

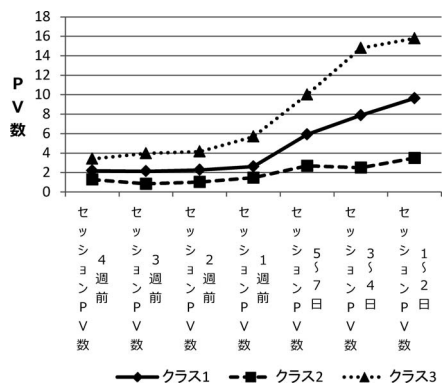


図 2 各クラスの前平均セッション PV 数

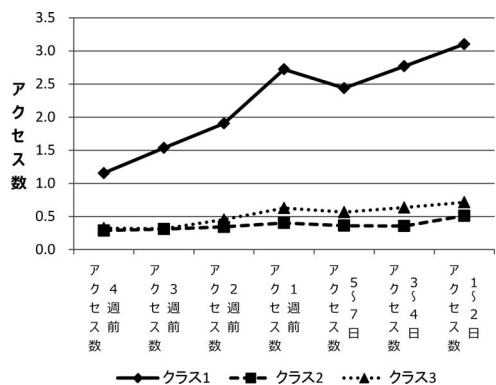


図 3 各クラスの前平均アクセス数

ページあたりにかける閲覧時間が増えると考えられる。また、クラスごとの購買カテゴリについて比較すると表 2 のようになる。クラス 1 への所属確率が最も高くなっている顧客は全体的に商品購買金額が高めである。特に、クラブやウェア (Men) の購買金額が高くなっている。クラス 2 への所属確率が最も高くな

¹ 分析には Statistical Innovations 社の Latent GOLD 4.0 を用いた。

表2 クラスごとのカテゴリ別平均購買金額 (円)

カテゴリ	クラス 1	クラス 2	クラス 3
クラブ	12399.92	9560.10	11160.58
ウェア (Lady)	539.31	506.30	880.75
ウェア (Men)	3916.03	3383.55	3631.31
用品・小物	4485.24	4996.99	5161.20
その他	371.18	95.40	0.00

ている顧客は商品購買金額が全体的に低めである。特に、クラブやウェア (Men) の購買金額は最も少ない。クラス 2 はセッション時間やセッション PV 数、アクセス数が最も少ないグループであることも考慮すると、購買が少なめ、あるいは安めの金額の商品を購買する傾向があると考えられる。クラス 3 への所属確率が最も高い顧客はウェア (Lady) や用品・小物の購買金額が最も高くなっている。したがって、本サイトでータルコーディネートをしているサイト顧客が多いことが推測される。

6. サイト会員の購買予兆発見モデル

6.1 購買モデルの作成

次に、サイト会員の購買予兆の発見をするために、確率選択モデルの一つであるロジット・モデルを基にそれぞれの顧客の各アクセス時点における購買確率を求めるためのモデルを作成する。モデル作成においてはサイト会員の普段のサイト閲覧行動も考慮するために、前述の潜在クラスモデルによって算出された各サイト会員の各クラスへの所属確率を考慮したモデルを用いる。また、本稿ではサイト会員の購買予兆の発見をするモデルを作成することを目的としている。そのため、学習用データによるパラメータ推定において、実際の購買時点だけでなく、その 1 回前と 2 回前のアクセスを「購買直前訪問」と考えてこれらに「購買」フラグを立て、これら 3 回のセッションを購買および購買直前行動として購買予兆を察知するモデルとして提案する²。

サイト会員 i の時点 t の購買確率 p_{it} を以下に示す。

$$p_{it} = \sum_{j=1}^3 \pi_{ij} \frac{\exp\{V_{ijt}\}}{\exp\{V_{ijt}\} + 1} \quad (1)$$

ただし、

π_{ij} 潜在クラス分析によって得られるサイト会員 i の

² ここで述べたように、購買直前セッションにも購買と同じ値を与えている。便宜上以降ではこれらをまとめて「購買」と呼ぶ。

クラス j への所属確率

V_{ijt} サイト会員 i がクラス j に所属する時の時点 t での購買に関する確定的効用

である。

また、確定的効用 V_{ijt} は (2) 式で表すような共変量の一次関数とする。なお共変量として、セッション時間、セッション PV 数、前回セッション時間、前回セッション PV 数、累積セッション時間、累積セッション PV 数、累積アクセス数、アクセス間隔、前回購買からの経過日数、累積購買回数、性別 (男性: 1, 女性: 0)、メールマガジン購読 (購読: 1, 非購読: 0)、年齢、サイト登録年数、アクセスした曜日 (休日: 1, 平日: 0) の 15 種類の変数を取り上げた。

$$V_{ijt} = \beta^T \mathbf{x}_t \quad (2)$$

ただし、 $\mathbf{x}_t^T = (x_{1t}, x_{2t}, \dots, x_{15t})$ である。

以上のモデルについて、最尤法によりそれぞれのクラスのこれらの変数の係数値を求める。なお、AIC を基準としたモデル選択を行い、利用する変数を決定する。

6.2 パラメータの推定結果

推定された各クラスにおける各変数のパラメータを表 3 に示す。なお、パラメータの推定においては、上述のように AIC を基準に変数増加法で変数選択を行っており、表中の“—”はその変数がモデルに含まれなかったことを示す。表 3 に示したとおり、セッション時間はどのクラスにおいても選択されなかったが、これはセッション PV 数との相関が大きかったことに起因すると考えられる。

表 3 より、クラス 1 のパラメータはセッション時間、セッション PV 数、前回セッション PV 数、累積セッション時間、累積セッション PV 数、累積アクセス数、アクセス間隔、前回購買との差、累積受注回数、性別、メールマガジン購読、年齢、アクセス曜日が変数として選択された。クラス 1 は日常的にアクセスをしていることから購買確率を説明する多くの変数が採用されたと考えられる。クラス 2 のパラメータはセッション PV 数、累積セッション PV 数、前回購買との差、性別、メールマガジン購読が変数として選択された。クラス 2 は購買前にあまり特徴的なサイト・アクセスは見られない。どの商品を購入するかをあらかじめ決めて短期的に購入に至る傾向があると考えられるので、セッション時間の変数などは選択されず、サイト会員の属性に関するものは選択されたと考えられる。クラス 3 のパラメータはセッション PV 数、前回セッショ

表 3 各クラスにおける係数の推定結果

変数	クラス 1	クラス 2	クラス 3
セッション時間	—	—	—
セッション PV 数	0.599	1.242	0.445
前回セッション時間	—	—	0.080
前回セッション PV 数	0.049	—	0.071
累積セッション時間	0.457	—	-0.135
累積セッション PV 数	-0.428	-0.835	—
累積アクセス数	-0.475	—	-0.379
アクセス間隔	0.102	—	0.034
前回購買との差	0.684	0.285	0.294
累積受注回数	0.134	—	0.395
性別	0.161	0.270	—
メールマガジン購読	0.106	0.178	-0.158
年齢	0.133	—	—
登録年数	—	—	0.111
アクセス曜日	0.051	—	0.064

表 4 モデルによる判別結果と実際の購買・非購買

	選択確率から 購買と予想	選択確率から 非購買と予想
実際に購買	A	B
実際は非購買	C	D

表 5 モデルによる的中率

	的中率 ₁	的中率 ₂
購買確率 0.9 以上で購買	94.3%	11.4%
購買確率 0.8 以上で購買	92.6%	17.5%
購買確率 0.7 以上で購買	88.8%	18.2%
購買確率 0.6 以上で購買	80.7%	15.5%
購買確率 0.5 以上で購買	63.7%	10.6%

ン時間、前回セッション PV 数、累積セッション時間、累積アクセス数、アクセス間隔、前回購買との差、累積受注回数、メールマガジン購読、登録年数、アクセス曜日が変数として選択された。クラス 3 は 1 度のアクセスで大量にページを閲覧することから、クラス 1 で選択されたような変数や登録年数といった変数が選択されたと考えられる。またクラス 1, 2 と同様にメールマガ購読が変数として選択されているが、クラス 3 のみ負の値となっている。したがって、クラス 3 はメールマガジン購読が購買確率のアップにつながらないと言える。この背景は、クラス 3 に所属するサイト会員は普段から自分で情報を収集する傾向があり、サイト側からの情報をあまり必要としないとも考えられる。

6.3 モデルの評価

本稿で提案しているモデルの検証として、提供データ期間中の後半 6 カ月の検証用データについて、得られたモデルを用いて購買確率を計算して評価する。

まず、表 4 のようにモデルによる選択確率から「購買と予想され、購買されたアクセス」(A)、「購買と予想されたにもかかわらず、非購買であったアクセス」(B)、「非購買と予想されたにもかかわらず、購買され

たアクセス」(C)、「非購買と予想し、非購買であったアクセス」(D) に 4 分割し、各アクセスを分類した。

そして (3) 式のように的中率を算出する。

$$\text{的中率}_1 = \frac{A + D}{A + B + C + D} \quad (3)$$

また、実際に予測した購買のみに着目すると以下の値も考えられる。

$$\text{的中率}_2 = \frac{A}{A + B + C} \quad (4)$$

モデルから算出された購買確率について 0.5, 0.6, 0.7, 0.8, 0.9 の確率を閾値として、その値以上で購買と判定し、その値未満では非購買であるとしたときのそれぞれの的中率は表 5 のようになった。本来は、閾値を 0.5 として購買/非購買を判定すると考えられるが、予兆発見の意味では、予兆範囲を広くとるか、もしくは狭くとるかは、意思決定者の意向ならびに状況によって異なるため、本稿では言及しない。

的中率₂については値が低いのが、これは、ほとんどが C に含まれているため、Recall (再現率) は 66.6% と低くない。的中率₂では購買確率 0.7 以上で購買とし

たときの値が最も高く、値の高い順に 0.8 以上のとき、0.6 以上のとき、0.9 以上のとき、0.5 以上のときとなった。購買確率 0.9 以上のときに購買としたときの的中率₂がこの値になったのは、(4) 式での A の値が小さいためであると考えられる。また購買確率 0.5 以上のときに購買としたときの的中率₂がこの値になったのは、(4) 式での B や C の値が大きいためであると考えられる。したがって、本稿で扱っているデータの場合、購買確率 0.7 以上で購買としたときの的中率₂が最も高いことから、購買確率が 0.7 を超えたときに購買の予兆が高まっていると考えることができる。また本稿でのモデルの的中率との比較対象として、各サイト会員それぞれの商品の平均購買間隔を基にしたものと比較を行った。具体的には各サイト会員のアクセス時の前回購買からの経過日数が各サイト会員の平均購買間隔より大きくなったら購買すると予想する方法と比較する。この場合の的中率₁は 66.7%、的中率₂は 6.6%となり、この結果本稿でのモデルのほうが的中率が良いといえる。

6.4 モデルから算出された購買確率を用いた各サイト会員別の購買予兆の発見

検証用データとした提供データ期間中の後半 6 カ月のデータを用いて、本稿のモデルから算出された購買確率の推移から検証を行う。図 4 はクラス 3 に所属のあるサイト会員のモデルから算出された各アクセス時の購買確率と購買・非購買の推移を表しているもので、図 5 はクラス 1 に所属のあるサイト会員のものである。なお、図の○が実際に購買に至ったセッションであり、△はその 2 回前までのセッションである。

図 4 を見ると、13 回のアクセスに対して 3 回の購買が行われている。図 4 に示すように、丸印およびその 1 回前、2 回前のセッションはおおむね値が大きく、購買の予兆がうまくとらえられているといえる。ただし、購買・非購買の推移を比較すると、購買確率が 0.6 より大きいときに購買をしていることが多く、このサイト会員の場合はおおむね確率が高いことに注意が必要である。

また、図 5 を見ると、このサイト会員は図 4 のサイト会員と比べて購買確率の値が大きく散らばっている。ただし、一部（15 回目のアクセス）を除くと購買セッションおよびその 1 回前、2 回前のセッションの購買確率は高く求められていることがわかる。また購買間隔が長い、例えば 1 回目のセッションから 12 回目までのセッションにおいては、購買とその直前以外は確率が幾分低く出ている。

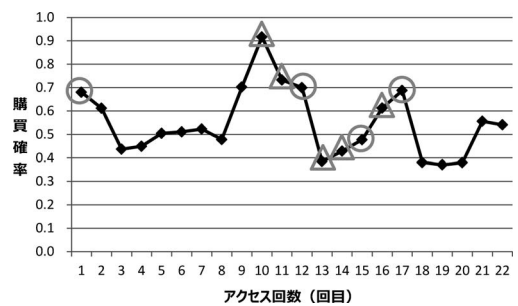


図 4 クラス 3 に所属のあるサイト会員の購買確率と購買・非購買の推移

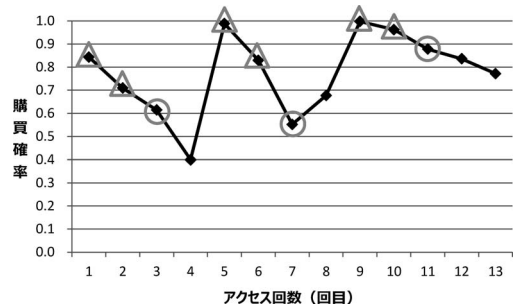


図 5 クラス 1 に所属のあるサイト会員の購買確率と購買・非購買の推移

この 2 つのケースに限れば、閾値を 0.6 程度に絞ると、うまく購買予兆をとらえることができると考えられる。そして、予兆を検出できた時点でリアルタイムにこのサイト会員に対して適切なマーケティング・アプローチをすれば有効であると考えられる。

7. 購買予兆の発見によるマーケティング施策の提案

作成したモデルから算出された購買確率から購買予兆を発見することで、さまざまなマーケティング・アプローチが可能であると考えられる。購買予兆は前述したように購買確率が 0.6 を超えたときや購買確率が急激に上昇したときなどが購買予兆といえる。購買確率が 0.6 を超える、または購買確率が急激に上昇したことが購買予兆の発見になり、このときにマーケティング・アプローチをすることが重要である。具体的にはクラス 1、2 の所属確率の高いサイト会員にはメールマガジンを送付することが挙げられる。商品の詳しい情報や閲覧している商品以外のお勧め商品のプロモーション（「この商品を買っている人はこのような商品も買っています」など）をすることが有効であると言える。クラス 1 とクラス 2 の大きな違いは、購買直前のアクセスにある。クラス 1 の 1 セッションあたりの

PV 数およびセッション時間がクラス 2 に比べて大きい、もしくは長いことから、クラス 1 に所属するサイト訪問者は購買直前にサイト上でじっくりと比較や絞り込みをしていると言える。クラス 1 のサイト会員にはいくつかの類似商品の特性の比較や口コミ情報を前面に出したプロモーションが有効といえ、クラス 2 のサイト会員には、それまでに閲覧された商品に関する詳細情報を適切に提供するような施策が求められる。逆にクラス 3 については、サイト訪問時に閲覧している商品カテゴリの中で効果的と考えられるようなアイテムをカスタマイズして表示するなど、メールマガジンのような事前の周知以外の方法のほうが有効かもしれない。

また購買確率が一定の値を超えたり、急激に購買確率が上がったタイミングで期間限定の支払金額の一部として使えるポイントを付与することで商品購買意欲をさらに刺激することができる可能性もある。

購買予兆を発見したタイミングでこれらのマーケティング・アプローチを行う利点として、サイト会員が購買を計画している商品以外の商品も同時に購買してもらえらる。購買確率が高まっているときにはサイト会員はある特定の商品について購買しようとしている可能性が高い。そこで、それ以外の商品を勧めることで、クロス・セルを期待でき、売上の拡大につながる。またモデルによって算出された購買確率から購買の予兆を発見できたということは、本来ならば何かしらの商品を購入してくれる可能性があるということである。購買確率が高くなったときに商品のプロモーションをすることで、購買確率の値から、本来ならば購買してくれるにもかかわらず、購買してくれなかったという機会の損失を防ぐこともできると考えられる。

8. おわりに

本稿では、まずサイト会員の普段のサイト閲覧行動パターンを明らかにした。そして閲覧行動パターンを考慮した購買予兆を発見するモデルをロジット・モデルをもとに作成した。本モデルについて、各変数のそれぞれのパラメータを閲覧パターンごとに求めたことで、購買確率に影響を与える変数や逆に影響を与えない変数を確認した。特に購買確率に影響を与える変数は、セッション PV 数や累積セッション時間、前回購買からの経過日数であった。そして購買予兆の発見から具体的なマーケティング・アプローチの例を挙げ、サイト会員に購入しようとしている商品の推奨や、購買

の機会損失を防ぐことを提案した。

今後の課題としては、以下のような点が挙げられる。まず、前述したように本稿では購買確率に影響を与える変数や、逆に影響を与えない変数を確認することができた。したがって、モデルに用いる変数として、購買確率に影響を与えないような変数の削除をすることや、反対に影響を与える可能性のある変数を考慮し、モデルに用いることが挙げられる。また、本稿では閲覧行動パターンの把握時やモデル作成時に、サイト会員の閲覧している商品や購買している商品については考慮しなかった。本稿で扱ったようなゴルフ用品販売の EC サイトにおいては商品の価格の幅が大きく、購買しようとしている商品によって閲覧行動パターンが異なる可能性がある。したがって、これらを考慮したモデルを作成することで、よりの確な購買行動モデルを提案でき、購買の予兆も発見しやすくなると考えられる。モデルの比較においては本稿でのモデルとサイト会員の平均購買間隔をもとにしたものを比較したが、ほかのモデルとも比較していくことでより精度の高いモデルを作成していくことも必要であると言える。

上記にも関連するが、本稿では、1 年のデータを前半と後半に分け、それぞれ学習用と検証用として利用した。ただしゴルフのシーズン並びにゴルフ・ファッションの季節性などはシーズンによっても異なることも考えられる。アクセスと購買という観点から見ると、今回のデータでは、1 購買あたりのセッション回数の通年の変化はあまり大きくなかったが、実際の購買カテゴリや個人ごとの購買アイテムのパターンなどについては、季節性の考慮が必要となると考えられる。

また、本稿では、前半期間におけるクラス判別は分析対象のすべてのサイト会員を用いた。一つの理由としては、対象とするサイト会員数がかかなり少なくなってしまい、局所的なクラスが得られやすくなった点がある。もう一つの理由として、本稿では所属クラスを決定する必要があったが、これを今回のデータで行うためには、モデル決定後に新たなサンプルの所属を決めなければならない。こうした方法については今後の課題としたい。

参考文献

- [1] 阿部誠, 近藤文代, 「マーケティングの科学—POS データの解析—」, 朝倉書店, 2005.
- [2] D. Van den Poel and W. Buckinx, “Predicting Online-purchasing Behaviour,” *European Journal of Operational Research*, **166**(2), 557–575, 2005.
- [3] 勝又壮太郎, “ウェブコンテンツ評価モデルの比較と活用—滞在時間とページビュー双方の観点から—,” マーケ

ティングサイエンス, **18**(1), 1-27, 2010.

- [4] 石井久治, 市川裕介, 佐藤宏之, 小林透, “Web アクセスログからのパターンマイニングによる購買行動の推定,” 電子情報通信学会技術研究報告. LOIS, ライフインテリジェンスとオフィス情報システム, **109**(272), 89-94, 2009.
- [5] 小池雄平, 菅谷健人, 住田潮, 高橋一樹, 平野智章, 山本浩平, “アクセスログデータに基づく重要顧客の ‘リアルタイム’ 判別,” *Department of Social Systems and Management Discussion Paper Series*, **1177**, 2007.
- [6] 野口竜司, 吉村正春, 中東優, 「Web マーケティング成功の法則 75」, 翔泳社, 2007.
- [7] 里村卓也, 金明哲, 「マーケティング・モデル」, 共立出版, 2010.
- [8] W. W. Moe and P. S. Fader, “Dynamic Conversion Behavior at E-Commerce Sites,” *Management Science*, **50**(3), 326-335, 2004.