

## EC サイトにおけるリピート顧客の予測と特徴分析

05001219 中央大学大学院 \*蘇 悦 SU Yue  
05000907 東海大学 大竹 恒平 OTAKE Kohei  
01405390 中央大学 生田目 崇 NAMATAME Takashi

### 1. はじめに

近年、インターネットの急激な普及に伴い、オンラインショッピングの利用者率は増加している。国内における EC 市場の拡大が続く一方で、EC サイト間の競争は激化している。昨今の顧客獲得競争を勝ち抜くため、EC サイト市場においてはリピート顧客の獲得に注目が集まっているが、中でも、新規顧客に対して次回以降の購買を促すようなマーケティング施策が求められている。

本研究では、EC サイトにおいて初回購買後の 1 年間に再購買した顧客をリピート顧客と定義し、リピート顧客の初回購買時の特徴を分析する。初回購買日の購買行動および探索行動の特徴を明らかにし、再購買の可能性が高い顧客の属性を探り、リピートに影響を与える要素を抽出することを目的とする。また、これらの分析を通じて、顧客の定着要因の特定を試みる。

### 2. データ概要及び分析対象

本研究では、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、以下のデータを用いて分析を行う。

- 会員登録情報データ  
登録日、性別、年齢、都道府県
- アクセスログデータ  
会員 ID 初回購買日の閲覧ページ、セッション回数などの EC サイトへのアクセス時の行動情報
- 購買データ  
販売日時、販売金額、商品カテゴリなどの購買行動に関する情報

なお本研究では、2017 年 5 月 1 日～2017 年 7 月 31 日に初回購買を行った 11,364 人の顧客を分析対象とした。ここでは、会員登録から 1 年以上経過している顧客は対象外とした。そのうち、5,613 人の顧客は初回購買日から 1 年以内に再購買を行っていた。

### 3. 分析方法

本研究では、初回購買の内容や、初回購買日のサイト閲覧情報に関する変数を作成し、ランダムフォレストにより、変数重要度を求める。次に、重要度が高い変数を用いて、ロジスティック回帰分析により、再購買を予測するモデルを構築し、再購買顧客の特徴を明らかにする。

### 3.1 ランダムフォレスト

ランダムフォレストは、多数の分類木を用いて集団学習を行う手法であり、Breiman によって提案された[2]。ランダムフォレストは、フォレスト内のすべての決定木から計算された不純度の平均的な減少量として特徴量の重要度を測定できる。重要度はランダムフォレストでのモデル構築過程で算出し、その値は各変数でモデルに与える影響力を表す。重要度の値が大きいほどその変数が予測モデルに強く寄与していることを表す。

### 3.2 二項ロジスティック回帰分析

本研究では、再購買か否かという反応が 2 値となるデータを目的変数、この反応に対していくつかの要因を説明変数として、その因果関係を分析する二項ロジスティック回帰分析を用いる。

二項ロジスティック回帰分析では、目的変数を  $y$  として、 $p$  個の説明変数  $x_1, \dots, x_p$  に対して顧客  $i$  の再購買確率  $p_i$  を式(1)のロジスティック関数で表す。

$$p_i = \frac{\exp\{\beta_0 + \sum_{j=1}^n \beta_j x_{ij}\}}{1 + \exp\{\beta_0 + \sum_{j=1}^n \beta_j x_{ij}\}} \quad (1)$$

$x_j$  : 再購買に影響を与える要因

$\beta_j$  : 各説明変数に対するパラメータ

### 4. 分析結果

本研究では、初回購買で影響を与えられと考えられる説明変数として、顧客の会員情報から 4 変数、初回購買時の購買行動から 12 変数、初回購買日のサイト閲覧行動から 80 変数の計 96 変数を作成した。モデルの作成時は、対象者を 7:3 の割合に分割し、それぞれ学習データ、テストデータとして用いた。

#### 4.1 ランダムフォレストによる変数の重要度

再購買を予測するモデルを作成する前段階として、本研究では、予測モデルに利用する変数の選択を行った。具体的には、ランダムフォレストにより作成した予測モデルの重要度を用いて計算した。表 1 に重要と評価された特徴量の上位 20 個を示す。

表 1 ランダムフォレストによる重要度

| 順位 | 変数名                  | 重要度    |
|----|----------------------|--------|
| 1  | ログイン・会員登録ページの閲覧時間    | 284.26 |
| 2  | 商品ページの閲覧時間           | 258.69 |
| 3  | 年齢                   | 253.04 |
| 4  | ショッピングカードページの閲覧時間    | 242.53 |
| 5  | 合計購買金額               | 242.20 |
| 6  | 平均ページビュー時間           | 238.32 |
| 7  | 平均セッション時間            | 236.12 |
| 8  | 平均購買金額               | 234.81 |
| 9  | 平均ページビュー数            | 233.36 |
| 10 | 初回購買日前に会員登録の有無       | 230.47 |
| 11 | アウトレットページの閲覧時間       | 170.66 |
| 12 | セッション回数              | 118.86 |
| 13 | 商品ページセッション回数         | 96.33  |
| 14 | クーポン利用の有無            | 90.69  |
| 15 | 購買件数                 | 89.53  |
| 16 | ログイン・会員登録ページのセッション回数 | 86.17  |
| 17 | アウトレットページのセッション回数    | 80.29  |
| 18 | 中古品ページの閲覧時間          | 77.80  |
| 19 | ゴルフ場予約ページの閲覧時間       | 69.82  |
| 20 | ショッピングカードページのセッション回数 | 60.28  |

#### 4.2 各説明変数のパラメータの推定結果

ランダムフォレストで得られた重要度上位 20 個の変数を用いて、再購買を予測することを目的としたロジスティック回帰分析を行った。モデルに利用した説明変数のうち、有意水準 5%において有意になった変数を表 2 に示す。

表 2 偏回帰係数の推定値 (一部)

| 変数名                  | 偏回帰係数       |
|----------------------|-------------|
| 切片                   | -0.5815 *** |
| 初回購買日前に会員登録の有無       | 1.1858 ***  |
| セッション回数              | 0.2333 ***  |
| 合計購買金額               | 0.2296 ***  |
| 平均ページビュー時間           | -0.2166 *** |
| 年齢                   | 0.2155 ***  |
| ショッピングカードページのセッション回数 | -0.1540 *** |
| クーポン利用の有無            | -0.1526 *** |
| 購買件数                 | 0.1307 ***  |
| アウトレットページのセッション回数    | 0.1255 ***  |
| 商品ページの閲覧時間           | 0.1222 **   |
| 平均購買金額               | -0.1121 *   |
| 商品ページセッション回数         | 0.0872 *    |
| アウトレットページの閲覧時間       | 0.0868 **   |

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

#### 4.3 再購買予測モデルの評価

作成した予測モデルについて、正答率、適合率、再現率、F 値を表 3 に示す。

表 3 評価指標値

|     |        |
|-----|--------|
| 正答率 | 67.22% |
| 適合率 | 67.00% |
| 再現率 | 67.87% |
| F 値 | 67.43% |

#### 5. 考察

再購買の判別モデルにおける説明変数の偏回帰係数の推定値 (表 2) のうち、幾つかの説明変数に着目し、再購買しやすい顧客の特徴について考察を行う。まず、初回購買日前日までに会員登録を行った顧客は再購買しやすいという結果が示された。会員登録後、継続的にレコメンドすることで、再購買率を促す可能性が推察される。また、セッション回数に関する変数に着目すると、ショッピングカードページのセッション回数以外の説明変数の偏回帰係数はすべて正であり EC サイトに多くアクセスしている会員程、再購買しやすいことが推察される。最後に、クーポン利用の有無について、負の偏回帰係数を有しているため、クーポンを闇雲に配信するのではなく、閲覧履歴等に基づき対象を選定することで、再購買の促進に繋がると考える。

#### 6. おわりに

本研究では、EC サイトにおける顧客の会員登録情報、初回購買日の購買行動に関する情報、初回購買日のサイト閲覧行動に関する情報を用いて、再購買の予測に取り組んだ。また、分析結果を用いたマーケティング施策の提案を試みた。

本研究では、初回購買時の情報のみを用いたが、初回購買日までのアクセス情報を考慮することで、より精度の高い再購買の予測モデルを構築することができると考える。また、分析結果を用いた施策について、さらなる分析が必要である。

#### 謝辞

本研究にあたり、データ提供と有益なコメントをいただいた、ゴルフポータル企業の方々に謝意を表します。また、本研究は JSPS 科研費 19K01945 および 17K13809 の助成を受けたものです。

#### 参考文献

- [1] 佐藤由将, 大竹恒平, 生田目崇, “EC サイトにおけるリピート顧客の特徴の分析,” 情報処理学会第 79 回全国大会講演論文集, pp. 507–508, 2017
- [2] L. Breiman, “Random Forests,” *Machine Learning*, Vol. 45, pp. 5–32, 2001.