

ゴルフポータルサイトにおけるユーザ行動データを用いた 新規顧客の定着要因の特定

北澤 晃樹[†] 生田目 崇[‡] 大竹 恒平[†]

東海大学情報通信学部経営システム工学科[†] 中央大学理工学部経営システム工学科[‡]

1. はじめに

昨今 EC サイトの市場規模が大きく成長したことにより顧客獲得競争が激化している。中でも継続的な購買を期待できる、新規顧客の定着が重要視されている。顧客獲得を目的としたクーポン配布については、そのクーポンによる割引のみにしか目がいかない場合も想定されるが、こうした顧客の継続購買は望めない。新規顧客の定着を考えた際には、初回購買からの離反を防ぐマーケティング施策が必要である。

本研究ではあるゴルフ EC サイトに会員登録をしたユーザを対象に、購買行動の分析を通じた新規顧客の定着要因の特定を目的とする。加えて、新規顧客の定着要因を考慮した継続的な購買を促すためのマーケティング施策を提案する。

2. 使用データおよび分析対象データ

本研究では、ゴルフ EC サイトから提供を受けた会員情報データ、購買商品履歴データ、ゴルフコース予約履歴データ、レビューデータ、アクセスログデータを用いた分析を行う。対象とした EC サイトではゴルフ用品の販売やゴルフコースの予約、過去に行ったゴルフプレー時のスコア管理などのサービスを提供している。

なお、データ期間は2017年7月～2018年7月の13ヶ月間であり、データ期間期首から2ヶ月を会員登録期間、会員登録後の315日間を3期間に分け、それぞれの期間で1回以上購買またはゴルフコースの予約をしている顧客を、定着顧客とした。本研究では、2017年7月、8月に会員登録を行ったユーザを対象とした。その結果、分析対象ユーザは61,692人となり、そのうち定着した顧客は3,046人であった。

3. 定着する顧客の要因分析

ランダムフォレストは複数の決定木を弱学習機として使用して組み合わせた学習器を作成す

るアンサンブル学習である。既存の研究 [1] ではアクティブユーザ推測モデルの提案を、ランダムフォレストにより行い、アクティブユーザかどうか推測している。

本研究では、定着する顧客の要因を特定するためランダムフォレストによる顧客の判別モデルを作成する。なお、ユーザの分類クラスの判別には、弱学習器による判別の多数決のうち最多のクラスを推定クラスとする。

表1 分析に使用した説明変数

データ概要	変数
会員情報データ (4変数)	性別, 年齢, 会員登録日のゴルフコース予約, 会員登録日の購買
レビューデータ (1変数)	レビュー回数 (合計)
購買商品履歴 (141変数)	購買個数 (全47種類の商品カテゴリを3段階に価格帯で分類)
アクセスログデータ (3変数)	PV数, セッション数, 平均セッション時間

目的変数は定着の有無、説明変数は表1の変数を用いてモデルを作成した。また、モデルの分類精度の確認には誤分類率である out-of-bag 推定値を利用する。

3.1 顧客の定着における変数重要度

ランダムフォレストにより定着する顧客を予測するモデルを作成した。そのモデルにおいて変数重要度を相対指標に変換したものの上位5変数を表2に示す。

表2より、PV数、セッション数といったアクセスログデータに基づく説明変数が、高い重要度を有していることが明らかになった。

次に、どの変数がどのように定着に影響しているかを確認するため、部分従属プロットにより確認する。そして特に重要かつ変化量があるものの結果を以下に示す。

Identification of new customer retention factors using user behavior data on golf portal sites

Koki Kitazawa[†], Takashi Namatame[‡], Kohei Otake[†]

[†] school of Information and Telecommunication Engineering, Tokai University

[‡] Faculty of Science and Engineering, Chuo University

表2 モデルの変数重要度の相対指標

説明変数	変数重要度
PV数	100
セッション数	79.2
平均セッション時間	28.8
会員登録日にゴルフコースの予約をしているか	22.5
年齢	7.71

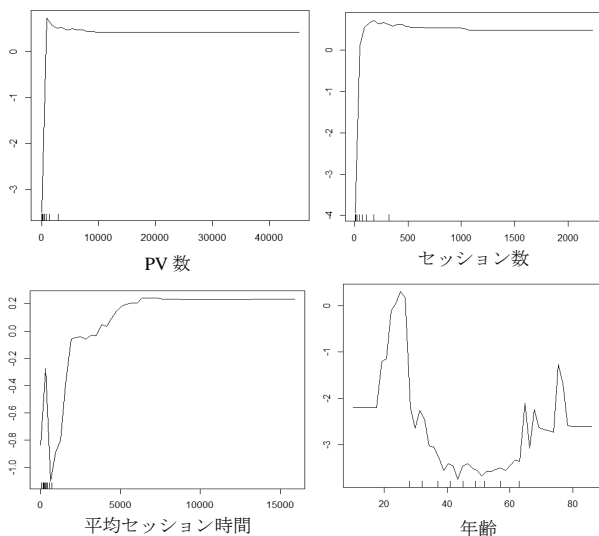


図1 ランダムフォレストの部分従属プロット

3.2 顧客の定着予測モデルの評価

構築したモデルの精度を評価するため、モデルによる予測結果を予測値と実際の値から作成したコンフュージョンマトリックス (表3) を作成し、精度、適合率、再現率、F値を算出した (表4)。

表3 コンフュージョンマトリックス

実際のクラス (Observation)	予測したクラス (Prediction)		
	定着 する	する	しない
定着 する		2,707	339
定着 しない		488	2,558

表4 モデルの評価

精度 (accuracy)	86.42%
適合率 (precision)	84.73%
再現率 (recall)	88.87%
F値 (F-measure)	86.75%

表4より作成したモデルのF値は86.75%であった。

3.3 考察

まず、アクセスログデータに関する説明変数が最も重要でありPV数、セッション数、平均セッション時間が多ければ定着しやすいという結果となった。ただし、平均セッション時間については、単調性はなく、一度離反しやすくなり、その後定着の可能性が上がっている様子がわかる。これはある程度の時間の比較検討では最終的に購入しない傾向にあると考えられる。

そしてゴルフコースの予約を会員登録日と同じ日に行っているか、が重要となった。これはゴルフコースの予約を目的に会員登録した会員はそのまま定着することが示唆される。よって、会員登録日限定のゴルフコースの予約料金を割引するクーポンの発行が効果的であると考えられる。

さらに、年齢は20代後半が1番定着しやすく、40代と50代が定着しにくいことが明らかになった。これにより、会員の利用傾向が年齢により大きく異なることが推察される。他方でログデータを使用しないモデルでは40歳前後、50代後半が定着しやすいという結果になった。これらのことから顧客に購入を即決させるため、該当の購入ページに移動したときに短時間の期間限定割引などの購入を促すためのマーケティング施策が有効だと考える。

4. まとめと課題

本研究では、作成したモデルにおいて定着した顧客のECサイトの閲覧や購買行動から定着要因を明らかにし、マーケティング施策の提案を試みた。しかし、本研究で構築したモデルはアクセスログデータから大きな影響を受けた結果となり、定着する顧客の要因をさらに詳細に特定する余地がある。今後はユーザの初回購買前後の行動や、ユーザの嗜好など本研究では考慮しなかった情報に関する変数をモデルに組み込みこむで、より詳細に定着要因を特定可能なモデルを作成できると考える。

謝辞

本研究では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] 土井千章, 川崎仁嗣, 中川智尋, 片桐雅二, 稲村浩, 太田賢, “決定木を用いたソーシャルネットワークサービス向けのアクティブユーザ推測モデルの提案”, マルチメディア、分散協調とモバイルシンポジウム 2014 論文集 2014, pp.366 -372 (2014)