

ショッピングサイトにおける優良顧客の離脱防止施策について

Measures for Keeping Loyal Customers Coming Back to EC site

笹谷 奈翁美* 坪内 孝太* 田代 昭悟* 鍛冶 伸裕* 清水 伸幸*
Naomi Sasaya Kota Tsubouchi Shougo Tashiro Nobuhiro Kaji Nobuyuki Shimizu

* ヤフー株式会社
Yahoo Japan Corporation

This paper describes our attempts to keep loyal customers coming back to EC site. Using machine-learning techniques, we develop a method for detecting the signs of customers leaving EC service. We evaluate the proposed method not only by using off-line data but also by implementing marketing champagne in actual service. This research is useful for those seeking to better understand EC service.

1. はじめに

ショッピングサイトでは、顧客を過去の注文履歴等からいくつかのセグメントに分け、それぞれに対してクーポン配布や特別なイベントへの招待などの施策を行っている。その中の目的のひとつとして、ロイヤリティの高い(注文の多い)優良顧客の離脱防止がある。優良顧客の離脱は取扱高に与えるインパクトが大きいため重要な問題となっている。

ショッピングサイトの顧客マネジメントの研究としては、Wangら[Wang 2016]や Kootiら[Kooti 2016]の研究にも見られるようにユーザの属性や行動に沿った適切な施策を講じて上げることが重要とされていた。本研究で対象とする離脱防止策は、顧客のロイヤリティに応じて別のものを利用すべきと考えられる。ユーザの属性のひとつである顧客のロイヤリティによってその離脱の原因や効果のある施策も大きく変わってくるのが容易に予想出来るためである。

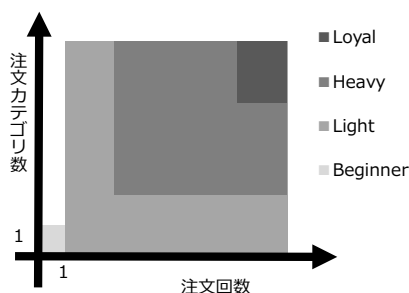
本稿では顧客のロイヤリティに応じた離脱予測モデルを構築する効果をオフラインテストで検証する。その結果、優良顧客の離脱に特化した精度の高いモデルを構築できた。そのモデルを実際のサービスで優良顧客の離脱防止施策に適用したライブテストについて述べる。

2. 顧客ランクと離脱(ランクダウン)

本研究では、Yahoo!ショッピング(以下ショッピング)の顧客を対象とする。

ショッピングでは、顧客を過去 1 年間の注文回数と注文カテゴリ数によって、Beginner、Light、Heavy、Loyal の 4 つのランクに分けている。(図 1)

図 1: 注文回数・注文カテゴリ数と顧客ランク



Loyal 顧客は、顧客数では全体の約 2%であるが、取扱高の約 16%を占める重要な顧客層である。

ショッピングでの離脱の定義は、最終注文日から 1 年間注文のない顧客である。が、今回は、最終注文翌日から 3 ヶ月間の注文回数で離脱またはランクダウン(下位のランクに遷移する)を判定することにした。理由は、利用できるデータの保存による問題や、早い時期にサービスへの施策が打てるためである。

各ランクの継続と離脱(ランクダウン)の閾値は 12 ヶ月後の離脱(ランクダウン)率が 95%以上になる値を用いる。Loyal で 3 ヶ月間の注文回数が 10 回以上の顧客は継続、3回以下の顧客は離脱(ランクダウン)とした。(表 1)

表 1: 継続と離脱(ランクダウン)の閾値

ランク	継続	離脱(ランクダウン)
Beginner	F>=1	F=0
Light	F>=1	F=0
Heavy	F>=3	F=0
Loyal	F>=10	F<=3

F は 3 ヶ月間の注文回数

3. オフラインテスト

顧客ランクごとに有効な特徴量に差があるか調べるために、モデルの比較を行った。ある1ヶ月間の注文顧客から各ランク 20000 人、正例、負例それぞれ 10000 人を抽出し学習データとした。評価データは別の1ヶ月間の注文顧客を各ランク 20000 人抽出した。正例、負例は同数である。

分類クラスは、最終注文日翌日から 3 ヶ月間の注文回数による継続と離脱(ランクダウン)(表 1)をそれぞれ正例、負例とする。特徴量は、最終注文日の前1ヶ月間の検索クエリと注文履歴から生成する。検索ワードの特徴量は、検索クエリを空白で区切ったワードの中から 5 人以上検索しているワード約 28 万種類を用いる。注文カテゴリはショッピングの商品カテゴリ第2階層までの 340 種類。注文回数は、1 回, 2 回, ..., 15 回, 16 回以上の 16 種類にカテゴリ化したもの。注文金額は、3000 円以下, 5000 円以下, 1 万円以下, ..., 10 万円以下, 10 万円超の 12 種類にカテゴリ化したものを用いる。(表 2)

比較したモデルは、顧客ランク別に検索ワードのみ、注文カテゴリのみ、注文回数・金額のみ、全てを統合した 4 つのモデルである。

表 2：モデルと特徴量

モデル	特徴量	特徴量の数
検索ワード	検索クエリを空白で区切ったワード、全体で 5 人以上利用しているワードのみ	280,108
注文カテゴリ	注文カテゴリ、ショッピングカテゴリ第 2 階層まで	340
注文回数と注文金額	注文回数と注文金額をカテゴリ化	28
統合	上記 3 つのモデルの特徴量を全て利用	280,476

学習アルゴリズムは、LIBLINEAR で提供されている L1 正則化つきロジスティック回帰を利用した。モデルの検証には 10 分割交差検定を、ペナルティ項 c は、(100, 10, 1, 0, 0.1, 0.01, 0.001, 0.0001) の範囲で Accuracy の一番高いものを採用モデルとした。

表 3：各モデルと交差検証の結果 (Accuracy)

	検索ワード	注文カテゴリ	注文回数・ 注文金額	統合
Beginner	53.68	53.40	52.07	55.09
Light	52.80	53.41	55.07	55.64
Heavy	55.07	58.85	63.75	65.02
Loyal	58.90	61.65	72.22	73.29

Beginner では注文履歴が少ないため、特徴量として検索クエリが若干有効である。一方、Heavy、Loyal では豊富な注文履歴があるため、注文履歴を特徴量としたモデルが有効であることがわかった。(表 3)

4. ライブテスト

Loyal の離脱(ランクダウン)予測は、他のランクと比べてショッピングにおける注文履歴が有効なことが前章より明らかになった。そこで、ショッピングのデータのみを特徴量として使ったモデル作成を行う。

特徴量の追加、エンコーディングの見直しを行った結果、ショッピングサイト訪問日数(6 種類)、注文カテゴリ(340 種類)、注文回数(6 種類)、平均注文金額(6 種類)の合計 358 種類の特徴量を利用したモデルの精度がよくなった。(表 4)

表 4：採用したモデルの特徴量

特徴量の種類	内容	特徴量の数
サイト訪問日数	ショッピングサイトの訪問日数をカテゴリ化したもの、log 変換したもの	6
注文カテゴリ	注文カテゴリ、ショッピングカテゴリ第 2 階層まで	340
注文回数と注文金額	注文回数と注文金額カテゴリ化したもの、log 変換したもの	12

学習アルゴリズムには、オフラインテストと同じ L1 正則化つきロジスティック回帰、モデルの検証には 10 分割交差検定を利用し、Accuracy は 84.4%であった。オフラインのテストと比較し結果が異なっているが、テストの時期に差がでること、学習に用いたデータセットが異なることなどが理由であり、同じような傾向が見られることは確認できている。特にショッピングでは、年末やボーナス前後のような時期に購買が促進されることがあり、学習期間の設定によって問題のベースラインは変化する。

施策は上記の離脱予測モデルを Loyal 顧客全体に適用し、下記 3 セグメントに分け、セグメント b に対してのみ施策を行った。

- a: 継続確率の高いと推測された顧客
- b: 離脱(ランクダウン)確率の高いと推測された顧客のうち 50%ランダムサンプリング
- c: 離脱(ランクダウン)確率の高いと推測された顧客のうち、b 以外(コントロール群)

施策内容は、次回のショッピング注文で利用できるポイント付与のメールを配信した。ポイントの有効期間は 19 日間である。

表 5 を見ると、セグメント a と b、c では 19 日間および 3 ヶ月間の注文率に有意な差があり、継続しそうな顧客と離脱(ランクダウン)しそうな顧客を高精度に特定できていることが分かる。また、ポイント付与の施策を実施したセグメント b と実施しなかったセグメント c を比較すると、ポイントの有効期限の 19 日間では大きな注文率の差が出て、その効果がその後 3 ヶ月間にも現れていることが分かる。

表 5：ライブテストの結果

セグメント	19 日間		3 ヶ月間	
	注文率	平均 注文回数	注文率	平均 注文回数
a. 継続確率の高い顧客 (施策未実施)	80.3%	3.10	98.0%	11.9
b. 離脱確率の高い顧客 (施策実施)	61.2%	2.00	90.2%	5.4
c. 離脱確率の高い顧客 (施策未実施)	54.3%	2.00	89.6%	5.3

今回適用したモデルは高精度に離脱予兆を検知でき、離脱(ランクダウン)確率の高い顧客に対して施策を行うことで、注文率を上げることができた。この施策によるアドオン効果として年間数億円の取扱高が見込まれる。

5. 結論

顧客のロイヤリティに応じた離脱予測モデルを構築する効果をオフラインテストで検証した。その結果、優良顧客は一般顧客に比べて注文情報が豊富にあるので、注文情報を用いたモデルが有効であることがわかった。優良顧客の離脱に特化した精度の高いモデルを構築し、実サービスにおいてライブテストを行い、離脱抑止効果が確認できた。

今回は優良顧客に注目したが、まだ初回購入顧客や中間層への対策が残されている。特に本稿のオフラインテストで明らかになったように、経験が浅いユーザはショッピングに関係するログがほとんどないので、他の情報を用いなければ有効な施策を打つことができず、問題の難易度も増すことが分かる。今後は、初回購入顧客等のショッピングサービスでの履歴が少ないユーザに対するアプローチを課題とする。

参考文献

- [Wang 2016] Pengfei Wang, et al.: “Your Cart tells You: Inferring Demographic Attributes from Purchase Data”, In Proc. of ACM WSDM '16, 2016
- [Kooti 2016] Farshad Kooti, et al.: “Portrait of an Online Shopper: Understanding and Predicting Consumer Behavior”, In Proc. of ACM WSDM '16, 2016