

他店舗利用状況を考慮した リードクオリフィケーション手法の提案

Proposal of Lead Qualification Method Considering the Usage Status of Other Stores

佐々木誠治¹ 後藤裕介²

Seiji SASAKI¹, Yusuke GOTO²

¹ 岩手県立大学大学院

¹ Iwate Prefectural University Graduate School

² 岩手県立大学

² Iwate Prefectural University

Abstract: In this study, we developed a lead qualification method to run effective lead nurturing campaigns for shopping complexes and EC sites with various shops. By using this method, it is possible to extract promising prospective customers and understand the characteristics of customers who use or do not use stores. We found that the model improves the accuracy by incorporating the customer's data about other store usages. Our method would help shop managers analyze the characteristics of the customers who use the store.

1. はじめに

近年マーケティングにおける顧客管理において、見込み顧客の獲得(リードジェネレーション)、有望な見込み顧客の抽出(リードクオリフィケーション)、見込み顧客の育成(リードナーチャリング)の3つの活動が行われおり、リードナーチャリングを行うためには、深い顧客理解が必要である[1]。また、従来のリードクオリフィケーション手法では、顧客の属性や行動、消費者行動の段階から顧客のスコアリング設定が行われる[2]。

ここで、見込み顧客とは、店舗を未利用かつ、店舗を利用している人の特徴を持っている人である。しかし、どんな特徴の項目が同じであれば店舗を利用するといえるのかについては分かっていないため、これを明らかにする必要がある。

複数かつ多様なテナントを抱える商業施設では、顧客によって多様な商業施設の利用の仕方が想定される。例えば、生鮮食品を扱う店舗を日常的に利用する顧客もいれば、特定のアプリ店舗のみを利用する顧客、食事を行うためにレストランを利用する顧客といったように、商業施設の利用の仕方が多様である。この顧客の多様な商業施設の利用の仕方があることによって、通常、顧客が商業施設を利用する場合には、全ての店舗を利用するのではなく、顧客の関心に基づいて限られた店舗を利用していることが考えられる。

そこで、店舗利用についての精度の高い分類器を作成する。誤判別した顧客は見込み顧客であり、その分類器のルールは特徴を示すと考えられるが、趣味嗜好が多様化した現代社会において、デモグラフィック変数やRFM変数のみに頼っての画一的な予測は難しい[3]。

本研究では、商業施設において特定店舗のリードクオリフィケーションを行うことを考える。このとき、通常の店舗との違いは、他店舗のデータが存在し、より精度の高い分類器を作成できる可能性があることや、特定の店舗を利用していない顧客に対してもアプローチが可能であることが挙げられる。

本研究では、複数のテナントを持つ商業施設やECサイトを対象として、リードナーチャリングのためのリードクオリフィケーション手法を開発した。この手法を用いることで、有望な見込み顧客の抽出と、店舗を利用する、あるいは利用しない顧客の特徴を理解することが可能となる。また、モデル構築において、他店舗利用状況変数を取り入れることにより、モデルの精度の向上やより深い顧客の特徴理解が可能になる。これにより、商業施設が店舗支援を検討する際にその店舗を利用する人の特徴を参考にして意思決定ができるようになる。

2. 提案手法

本章では提案するリードクオリフィケーション手法の説明を行う。本手法では、商業施設において、店舗のレベルの新規顧客を獲得するための見込み顧客抽出を想定し、顧客の店舗利用を分類することを考える。このとき、店舗を利用している顧客に似ている顧客は、その店舗を利用に転じやすい顧客と考え、この顧客を抽出する。この考えに基づき、店舗利用分類モデルを構築する。

本手法は以下のプロセスによって実現される。

- 1 他店舗利用状況変数を考慮した分類精度の高い分類器の作成
 - 1.1 モデル構築
 - 1.2 モデル評価
- 2 見込み顧客分析
 - 2.1 見込み顧客の抽出
 - 2.2 既存顧客の特徴抽出

本手法は、購買履歴を用いて顧客が当該店舗を利用するか、利用しないかの2値分類モデルを構築する。このとき、2値分類モデルの精度が高い場合に、分類モデルの混合行列におけるFPに該当する顧客を有望な見込み顧客とするものである。また、2値分類モデルにおける説明変数の重要度を確認し、重要度上位の変数について分類結果ごとに集計することによって、店舗を利用する顧客、利用しない顧客の特徴を理解することにつながる。

さらに本手法では、商業施設データの特性を反映し、分類モデルの精度向上するため、顧客の年齢や性別といったデモグラフィック変数や商業施設単位での総購買金額などのRFM変数に加え、他店舗利用状況変数をモデルに取り入れる。

3. 提案手法の適用

3.1 適用事例の概要

本研究で用いるデータは、複数かつ多様なテナントを抱える商業施設の購買履歴である。

- 期間：2017年02月16日～2019年07月31日
 - 件数：6863638件
 - テナント数：279店舗(一部閉店した店舗を含む)
 - 顧客数：193046人
- 基礎集計より、以下のことが分かった。
- 店舗により、顧客一人当たりの購買金額と来店頻度は異なる
 - 顧客の利用店舗数と、購買金額及び来店頻度の間に正の相関関係がある

3.2 他店舗利用状況変数を考慮した分類精度の高い分類器の作成

3.2.1 モデル構築

モデルを構築するために、表1で示すような、店舗利用フラグを目的変数とし、顧客のデモグラフィック変数①、RFM変数②、他店舗利用状況変数③④を説明変数とする。ここで、RFM変数②と他店舗利用状況変数③④については、当該店舗の情報を除いて変数を作成する。

次に、XGBoost[4]を用いて店舗利用予測モデルを作成する。実験ではPython3.8のXGBoostを用いて表2におけるmodel1~5について計算を行った。また、ハイパーパラメータチューニングは自動最適化フレームワークであるOptuna[5]を用いて、AUCの値で最適化を行い、検証はデータセットを3:7で分割し、ホールドアウト検証法を用いた。

表1 目的変数と説明変数

目的変数	shop_code_○○○	0:未利用, 1:利用済
デモグラフィック変数 ①	card_type	カードタイプ
	sex	性別
	web_member_flag	WEB会員フラグ
	age	年齢
RFM変数 ②	visit_number	来店回数
	visit_period	来店期間
	frequency	来店頻度
	payment_amount_max	最大購買金額
	payment_amount_min	最小購買金額
	payment_amount_sum	総購買金額
	payment_amount_mean	平均購買金額
他店舗利用状況変数 ③④	payment_amount_median	購買金額中央値
	payment_amount_std	購買金額標準偏差
	stores_used_number	利用店舗数
他店舗利用状況変数 ④	shop_code_101~629 (目的変数を除く)	0:未利用, 1:利用済

表 2 モデル評価の比較

店舗	model1 ①	model2 ①+②	model3 ①+②+③	model4 ①+②+④	model5 ①+②+③+④	auc_lift
625	0.696	0.919	0.920	0.920	0.924	0.005
306	0.868	0.936	0.943	0.943	0.943	0.007
150	0.668	0.888	0.892	0.887	0.896	0.008
510	0.710	0.897	0.897	0.907	0.910	0.013
114	0.769	0.879	0.877	0.896	0.893	0.014
205	0.791	0.902	0.905	0.918	0.919	0.017
507	0.770	0.918	0.920	0.938	0.935	0.017
700	0.663	0.835	0.841	0.848	0.856	0.021
131	0.598	0.740	0.745	0.795	0.797	0.057
259	0.716	0.824	0.863	0.891	0.891	0.066
平均	0.724	0.873	0.880	0.894	0.896	0.022

3.2.2 モデル評価

次に、モデルの精度について確認を行う。モデルの評価については、目的変数である店舗利用フラグにおいて、店舗利用と未利用が偏った不均衡データであることから、評価指標に AUC(Area Under the Curve)を用いる。

本研究では、レディース・メンズ、ライフスタイル、スイーツ&フードといった複数のジャンルを含んだ 10 店舗を対象とし、表 2 に示すように変数を組み合わせた場合について 10 店舗の AUC の値の平均値を比較する。

表 2 は、この 10 店舗について、表 1 に示す説明変数の組み合わせの異なる model1~model5 の AUC の値とその平均値を示している。ここで、auc_lift は model5 と model2 の AUC の値の差分を示している。表 2 により、model5 において、AUC の平均値が 0.896 と精度の高いモデルを構築できていることが分かる。

3.2.2.1 他店舗利用状況変数の寄与

次に、他店舗利用状況変数③④がモデルの精度に寄与するかの確認を行う。表 2 の auc_lift より、デモグラフィック変数①や RFM 変数②でモデルを構築した model2 の場合よりも、他店舗利用状況変数③④を取り入れた model5 場合の方が AUC の値が高いこと

が分かる。

このことから、他店舗利用状況変数③④は、モデルの精度向上に寄与しており、従来の顧客の年齢や性別、RFM 指標に加え、他店舗利用状況から顧客をセグメント分けすることができる。そのため、他店舗利用状況から顧客を特徴付けることが可能となり、より深い顧客の特徴理解につながると考えられる。

3.2.2.2 他店舗利用状況変数のモデルへの寄与が大きい店舗

次に、他店舗利用状況変数のモデルへの寄与が店舗の確認を行う。表 2 の auc_lift より、店舗 259 と店舗 131 が他店舗利用状況変数による AUC の増分が大きく、それ以外の店舗が小さいことが分かる。

AUC の増分が大きい店舗は、レディース・メンズものを扱うアパレルである店舗 259 やお土産を扱う店舗である店舗 131、AUC の増分が小さい店舗は、ライフスタイルグッズを扱う店舗 625、店舗 306、店舗 150、店舗 205、スイーツ&フードを扱う店舗 510、店舗 507、ビューティー&コスメを扱う店舗 114、カフェ・レストランである店舗 700 である。

これらのことから、店舗によって寄与の程度には差が存在するが、いずれの場合においても、他店舗利用状況変数をモデルに取り入れることによって、モデル精度の向上が見込まれる。また、利用する顧

客が多く日常的に利用される店舗においては、デモグラフィック変数や RFM 変数のみでも十分予測可能であるが、アパレルやお土産といった日常的に購買が行われない店舗においては、購買する顧客の情報が不足しやすいため、他店舗利用状況情報を用いて補うことによって、モデルの精度が向上したことが考えられる。

3.3 見込み顧客分析

本節では、2つの店舗について分析例を挙げる。ここでは、他店舗利用状況変数による AUC の増分が大きい店舗 259 と増分が小さい店舗 625 を取り上げる。両店舗についてモデルの評価を行い、次に説明変数の重要度による関連店舗及び変数の特定を行う。最後に、店舗を利用する人の特徴を説明変数の重要度上位の変数について集計し、確認を行う。

3.3.1 モデルの評価及び見込み顧客の抽出

デモグラフィック変数や RFM 変数のみで構築したモデルと他店舗利用状況変数を加えたモデルとで、AUC の値、顧客の規模、FP に該当する顧客、他評価指標の値の比較を行う。図 1, 2 は、対象店舗について、model1~model5 の ROC curve を示している。表 3 は、対象店舗について、model2 と model5 の各評価指標の結果を示している。

図 1, 2 より、店舗 259 では、model2 の AUC の値は 0.824、model5 は 0.891 と増分が大きく、model1 から model5 にいくのに従って精度が改善していくのに対して、店舗 625 では、model2 は AUC の値が 0.919、model5 は 0.924 と増分が小さく、model2 を境に精度が頭打ちとなっている。

次に、表 3 より、コンフュージョンマトリクスと各評価指標について確認を行う。店舗 259 では、テストデータの中で 57,669 人中 1,279 人が利用しており、FP に該当する見込み顧客が 472 人で、全顧客の 0.81%であるのに対し、店舗 625 では、57,669 人中 12,456 人が利用しており、見込み顧客は 8,145 人で、未利用者の 14.12%と、顧客の規模が大きく異なることが分かる。このとき、model2 と model5 を比較すると、店舗 259 では、正解率や適合率、F 値がそれぞれ向上し、再現率が低下するのに対し、店舗 259 では、正解率や適合率、F 値がそれぞれ低下し、再現率が向上していることが分かる。

これらのことから、AUC の評価において、両店舗において良好なモデルを構築できているが、デモグラフィック変数や RFM 変数及び他店舗利用状況変数は、店舗によってモデルの精度向上への寄与の程度は異なることが分かる。

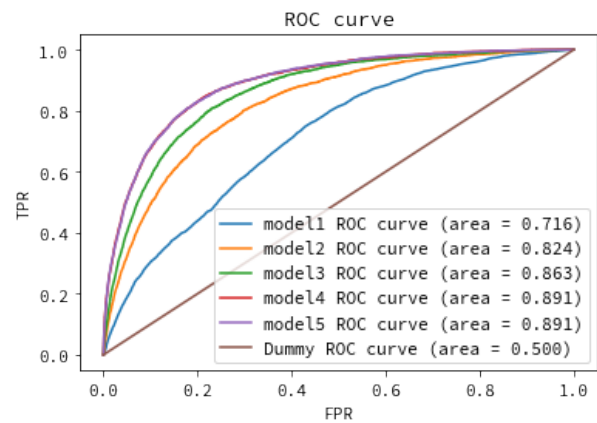


図 1 店舗 259 の ROC curve

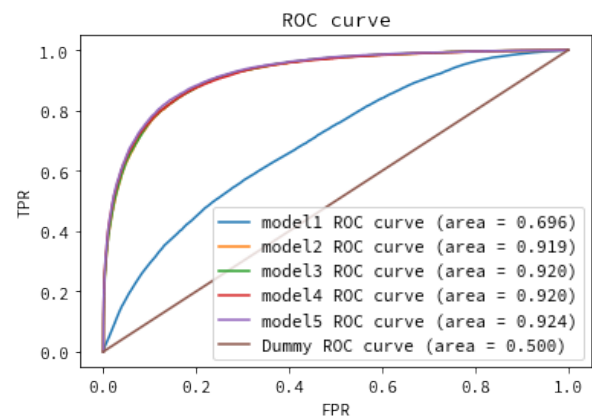


図 2 店舗 625 の ROC curve

表 3 モデル評価

	店舗 259		店舗 625	
	model2	model5	model2	model5
TP	1592	808	9808	12934
FN	2051	2835	4958	1832
FP	4392	472	2648	8145
TN	49634	53554	40227	34730
正解率	0.888	0.942	0.868	0.826
適合率	0.266	0.631	0.787	0.613
再現率	0.437	0.221	0.664	0.875
F 値	0.634	0.649	0.817	0.798

3.3.2 既存顧客の特徴抽出

次に、説明変数の重要度を確認することにより、店舗利用において重要な変数を特定し、既存顧客の特徴抽出を行う。図3, 4は対象店舗の説明変数の重要度について上位のものを示している。

図3より、店舗259においては、ビューティー&コスメを扱う店舗が6店舗あるうちの店舗115や店舗114、レディース・メンズものを扱う店舗が4店舗あるうちの店舗202や店舗225、店舗226、レディースものを扱う店舗が57店舗あるうちの店舗107や店舗174、店舗178、メンズものを扱う店舗が4店舗あるうちの店舗305、そして、店舗利用数が上位10の変数となっている。このことから、店舗259では、40代の女性を中心に利用されていることに加えて、上記の美容雑貨やアパレルの店舗を利用する顧客であることが分かる。

図4より、店舗625においては、来店回数、ライフスタイルグッズを扱う店舗が29店舗あるうちの店舗603や店舗189、店舗150、レディースものを扱う店舗が57店舗あるうちの店舗604や店舗601、カフェ・レストランが34店舗あるうちの店舗657、来店頻度、年齢、そして、スイーツ&フードを扱う店舗が15店舗あるうちの店舗618が上位10の変数となっている。このことから、店舗625では、30代の女性を中心に利用されていることに加えて、上記のライフスタイルや食品関係の店舗を利用する顧客であることが分かる。

このように、従来得られる顧客の年齢や性別の情報に加えて、重要となる他店舗の利用情報や商業施設の利用情報を見つけることが可能となる。これにより、当該店舗を利用する顧客の特徴理解や複数店舗を対象とした施策検討に活用できる可能性がある。

3.3.3 判別結果ごとの特徴

図5は、判別結果ごとに説明変数の重要度上位の変数について、顧客の平均値を示しており、図中の値は店舗の平均利用率である。これより、以下のことが分かった。

- 店舗を利用する顧客は、店舗を利用しない顧客と比較して、関連する他店舗の利用状況や商業施設の利用状況が良い傾向がある
- TP(1,1)とFP(0,1)の結果は類似している
- TN(1,0)の結果はTN(0,0)以上、TP(1,1)、FP(0,1)未満である

また、他の店舗についても同様に確認を行い、判別結果間の関係が同じであることが分かった。

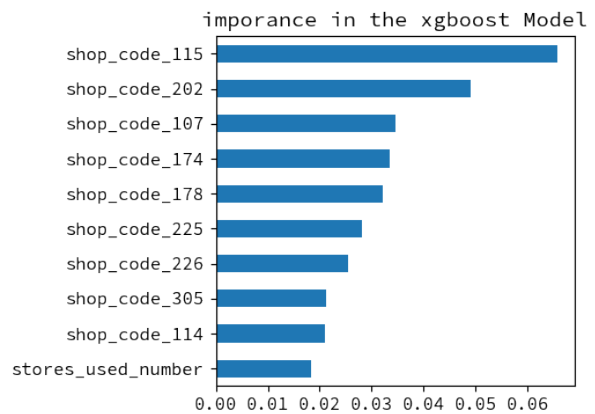


図3 店舗259の説明変数の重要度上位10

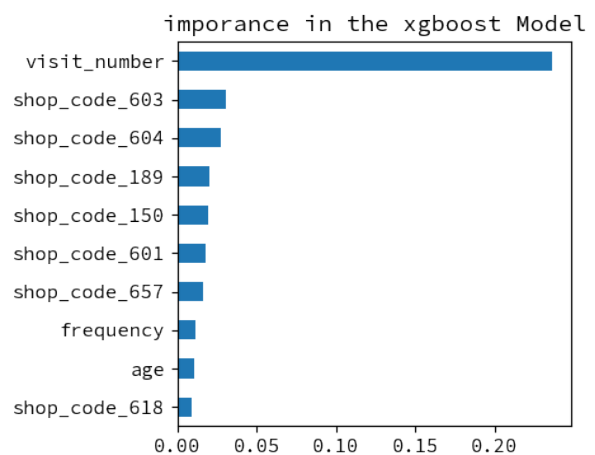


図4 店舗625の説明変数の重要度上位10

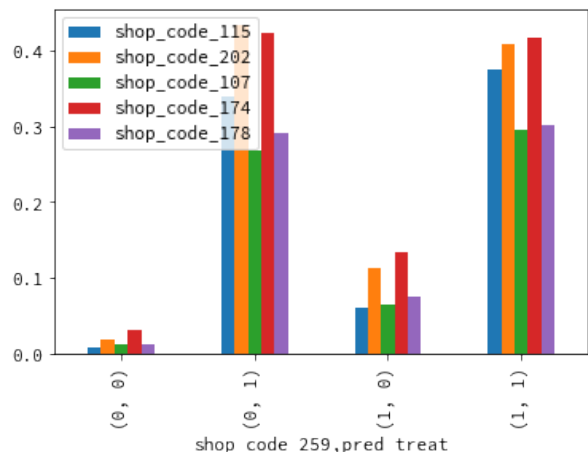


図5 店舗259の判別結果と説明変数の重要度上位5の関係

4. まとめと今後の課題

本研究では、商業施設におけるリードナーチャリングのためのリードクオリフィケーション手法を開発した。また、複数のテナントを抱える商業施設の特性を考慮するため、他店舗利用状況変数をモデルに取り入れ、その有用性の確認を行なった。また、他店舗利用状況変数が効きやすい店舗の特徴の分析を行った。これらにより、従来用いられるデモグラフィック変数や RFM 変数に加えて、他店舗利用状況変数を用いた顧客の店舗利用の予測が可能となり、より深い顧客の特徴理解が可能であることが確認できた。

今後の課題として、本研究で用いた他店舗利用状況変数は、店舗を利用したか、利用していないかの情報に留まっているが、店舗の購買金額や来店頻度といった店舗の利用の程度の情報を取り入れることにより、よりモデルの精度を向上させることが期待される。また、全ての店舗についてモデルを作成し、他店舗利用状況変数のモデルへの寄与について追加分析を行うことや、本手法によって抽出された見込み顧客に対する販促による効果検証を行うことが挙げられる。

参考文献

- [1] 伊藤克容：マーケティング管理会計の展開—顧客動向の追跡と動線設計—, 管理会計学：日本管理会計学会誌：経営管理のための総合雑誌, 26 巻, 2 号, pp.31-46, (2018)
- [2] Jonathan Turfboer Steven Blumenfeld Anna Lising Berman Alex Kinnier Daniel Yates Alexander Laskey: Solar customer acquisition and solar lead qualification, US 2015/0310463 A1, 2015-10-29.
- [3] 新美潤一郎, 星野崇宏: 顧客行動の多様性変数を利用した購買行動の予測, 人工知能学会論文誌, 32 巻, 2 号, pp. B-G63_1-9
- [4] Tianqi Chen and Carlos Guestrin: Xgboost: A scalable tree boosting system. the 22nd ACM SIGKDD International Conference, pp. 785-794, (2016)
- [5] Optuna ハイパーパラメータ自動最適化フレームワーク. <https://preferred.jp/ja/projects/optuna/>, (参照 2020 年 2 月 24 日)