

銀行口座情報を用いたクレジットカード離反予測モデルの構築

永田純平^{†1}, 須子統太^{†2}

アブストラクト: 近年, わが国ではキャッシュレス化の推進によりクレジットカード市場の拡大が続いており, カード発行会社間での顧客獲得競争が激化している. 本分析では, 銀行の発行するクレジットカードについて口座情報の紐づけが可能である点に着目し, 機械学習を用いた離反予測モデルを作成した. また, 構築した離反予測モデルの予測精度評価実験の結果をもとに, クレジットカードや銀行ビジネスのマーケティングにおける銀行データの活用可能性を考察した.

キーワード: クレジットカード, マーケティング, 離反予測, 銀行データ, 機械学習

1. はじめに

近年, わが国ではキャッシュレス化の推進によりクレジットカード市場の拡大が続いている. そのため, カード発行会社間での顧客獲得競争は激化しており, クレジットカードにおけるマーケティング施策の重要性が増している. クレジットカードにおけるマーケティング施策は主に「新規顧客の獲得」「既存顧客の利用促進」「既存顧客の離反阻止」に分けることができるが, その施策内容はカード会社により様々である.

みずほ銀行においては, 「みずほマイレージクラブカード (クレジットカード)」 (以下 MMC カード) としてクレディセゾン, オリエン特コーポレーションとの提携カードを発行している. 前述の通り, クレジットカード市場では顧客獲得競争が激化しており, カード発行会社の中には近年急速に会員数を拡大している会社も存在している. そのため, MMC カードを既に保有している顧客が別カードに流出することによる MMC カードの収益への影響は増していることが考えられる.

そこで, 本分析では MMC カードのマーケティング強化に向けた取り組みとして離反予測モデルの活用に着目する. MMC カード顧客の口座情報から, 顧客が利用する別カードの引き落とし情報を活用することで, みずほ銀行がクレディセゾンと提携して発行する MMC カード (セゾン) の顧客の離反可能性を予測する「離反予測モデル」の作成を行う. さらに, 実際のビジネスに適用する際の展望や課題等についても考察する.

なお, 本分析は 2018 年にみずほ銀行と早稲田大学との間で締結された協定等に基づいて行われているが, みずほ銀行の保有する各種データを用いた分析作業については全てみずほ銀行の職員が実施した. また, 分析作業時には顧客の氏名や住所等の情報を事前にデータセ

ットから削除することで, 個人を特定できないように措置を講じている点に留意されたい.

2. MMCカードにおける離反予測モデル

2.1 離反予測モデルの活用効果

離反予測モデルは, ある顧客が将来あるサービス・商品から離反するかどうかを事前に予測することで, 顧客の引き留め施策を実施するために用いられることが一般的である. しかし, 離反していく顧客に対し有効な引き留め施策を実施することが困難な場合も存在する.

本分析の対象とする MMC カードはみずほ銀行の取り扱う商品の一つであり, 銀行全体では MMC カード以外にも預金や投資信託など他の商品を通した顧客との接点が存在する. したがって, MMC カードの引き留めが困難な顧客においても, 他商品を通した施策を実施することで銀行サービス全体からの離反を防ぐなど, 様々な活用が期待できる.

以上のことから, MMC カードの離反予測モデルの構築は MMC カードのマーケティングだけではなく, みずほ銀行の個人顧客に対するマーケティング全体にとって大きな効果をもたらす可能性がある.

2.2 口座データの活用効果

一般的にクレジットカード会社でのマーケティングにおいては, 主に自社カードの発行時に顧客より受領する顧客情報 (年齢, 性別, 年収等) や発行カードの利用データ (利用額, 利用回数等) をもとに施策を検討・実施する. 一方, みずほ銀行での MMC カードのマーケティングでは, 顧客の口座利用情報を用いた分析が可能である. みずほ銀行の口座を給与受取口座として一時的に使用し全額を他行口座に送金するような口座の利用状況かどうか, などの条件により口座から取得できる情報に差は存在するものの, 当該口座をメイン口座として利用している顧客であれば利用している別のクレジット

^{†1} みずほ銀行デジタルマーケティング部

^{†2} 早稲田大学社会科学総合学術院

カード引き落とし額や給与振込額等を取得できる。そのため、一般のクレジットカード会社が利用可能なデータと併せて顧客の口座情報を加味した離反予測モデルの作成が可能である。さらに、一般的にクレジットカードの離反理由として考えられるのは「別カードへの乗り換え」であることから、別カードの利用状況を離反予測モデルに用いることで高い精度での離反予測が可能であると考えられる。

3. 離反予測モデルの作成

3.1 目的変数の定義

離反予測モデルの目的変数を定義するため、まず本分析における「離反」を定義する。本分析では、MMC カードの月利用額が6か月連続で0円であった顧客を「離反」した顧客、その6か月のうち初めの月を「離反した月」とそれぞれ定義した。

また、本モデルにおいては顧客の各月のデータから作成した説明変数（後述）をもとに、3か月後の月に当該顧客が「離反」するかどうかを予測する。図1に、本離反予測モデルのイメージを示す。



図1 離反予測モデルのイメージ

3.2 説明変数の定義

本分析では、離反予測に用いる説明変数を顧客属性、MMC カード情報のデータに加えて口座情報のデータからなる10変数とした（表1）。なお、本分析における「別カード」とは、みずほ銀行がクレディセゾンと提携して発行する「みずほマイレージクラブカード（クレジットカード）」以外の全てのクレジットカードのことを指す。¹

また、表1に示した説明変数のうちD~Jの算出手法

はそれぞれ以下の通りである。

- ・MMC カード利用状況：(当月の MMC カード利用額) / (過去3か月の MMC カード利用額の平均)
- ・MMC カード利用状況②：直近3か月の MMC カード利用額の平均
- ・給与の受取：直近6か月の間に当該口座で給与の受取がある場合は1，ない場合は0
- ・年金の受取：直近6か月の間に当該口座で年金の受取がある場合は1，ない場合は0
- ・新規カード有無：直近6か月の間に新規でクレジットカードの口座振替がある場合は1，ない場合は0
- ・別カード利用状況：(当月の別カード口座振替額) / (過去3か月の別カード口座振替額の平均)
- ・MMC カード利用割合：((当月の MMC カード利用額) / (当月の全クレジットカード口座振替額)) / ((過去3か月の MMC カード利用額の平均) / (過去3か月の全クレジットカード口座振替額の平均))

なお、本分析では説明変数F~Jについては説明変数として用いる場合と用いない場合とでモデルの構築実験を行い、予測精度の比較を行った。

表1 離反予測に用いる説明変数

	変数名	変数種別
A	月	MMC カード 情報
B	性別	
C	年齢	
D	MMC カード利用状況	
E	MMC カード利用状況②	
F	給与の受取	口座情報
G	年金の受取	
H	新規カード有無	
I	別カード利用状況	
J	MMC カード利用割合	

4. 評価実験

4.1 実験条件

本分析では、モデルの学習用データセットとテスト用データセットを以下のように構築した。

A) 学習用データセット

2019年10月から2020年9月の間にMMCカードの利用が確認された顧客2万人を対象に、同期間の各月の目的変数・説明変数を抽出

B) テスト用データセット

¹ 本分析では離反予測の対象をクレディセゾンとの提携カードに限定しており、同じくMMCカードであるオリエントコーポレーションとの提携カードについては「別カード」に分類している。

2020 年 10 月から 2021 年 9 月の間に MMC カードの利用が確認された顧客 1 万人を対象に、同期間の各月の目的変数・説明変数を抽出

なお、本モデルを構築するための予測手法として、「XGBoost」¹⁴「ロジスティック回帰」「ランダムフォレスト」¹⁵の 3 つを比較した。

4.2 実験結果

表 2 から表 7 に、各手法における混同行列と予測精度指標を示す。表中の正解率と F1 スコアに着目すると、XGBoost によるモデルとランダムフォレストによるモデルは比較的高い精度での予測ができているのに対し、ロジスティック回帰によるモデルは他の 2 つのモデルに比べ低い精度を示していることがわかる。

また、XGBoost やランダムフォレストによるモデルにおいて口座情報を説明変数として用いた場合、口座情報を用いなかった場合に比べ F1 スコアが若干改善した。

本実験において比較的高い予測精度を示した、XGBoost を用いたモデルでの各説明変数の重要度は図 2、図 3 の通りである（ただし説明変数「月」は省略）。なお、図 2 において青い棒グラフで表記した説明変数は MMC カードに紐づく口座情報から算出した変数であることを示す。

表 2 XGBoost モデルの混同行列

正解\予測	口座情報あり		口座情報なし	
	0 (非離反)	1 (離反)	0 (非離反)	1 (離反)
0(非離反)	6,697	2,031	6,628	2,100
1(離反)	171	871	159	883

表 3 XGBoost モデルの予測精度指標

	正解率	適合率	再現率	F1
口座情報あり	77.5%	30.0%	83.6%	44.2%
口座情報なし	76.9%	29.6%	84.7%	43.9%

表 4 ロジスティック回帰モデルの混同行列

正解\予測	口座情報あり		口座情報なし	
	0 (非離反)	1 (離反)	0 (非離反)	1 (離反)
0(非離反)	4,424	4,304	4,421	4,307
1(離反)	156	886	142	900

表 5 ロジスティック回帰モデルの予測精度指標

	正解率	適合率	再現率	F1
口座情報あり	54.4%	17.1%	85.0%	28.4%
口座情報なし	54.5%	17.3%	86.4%	28.8%

表 6 ランダムフォレストモデルの混同行列

正解\予測	口座情報あり		口座情報なし	
	0 (非離反)	1 (離反)	0 (非離反)	1 (離反)
0(非離反)	6,926	1,802	6,638	2,090
1(離反)	253	789	208	834

表 7 ランダムフォレストモデルの予測精度指標

	正解率	適合率	再現率	F1
口座情報あり	79.0%	30.5%	75.7%	43.4%
口座情報なし	76.5%	28.5%	80.0%	42.1%

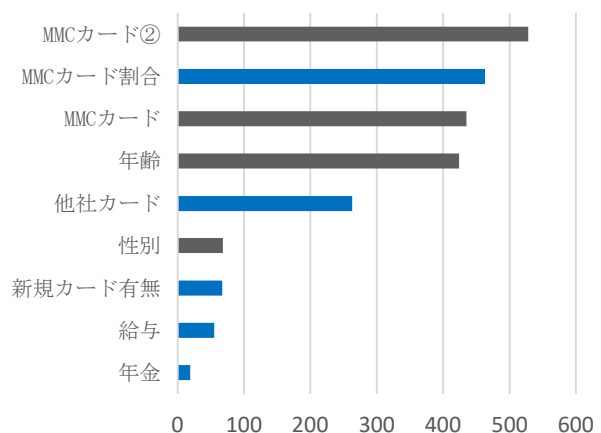


図 2 説明変数の重要度 (XGBoost, 口座情報あり)

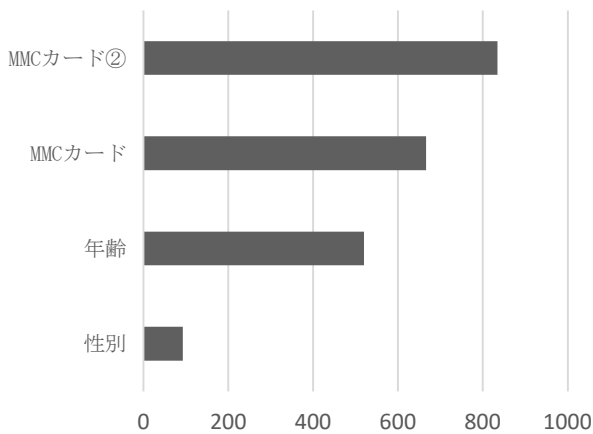


図3 説明変数の重要度 (XGBoost, 口座情報なし)

4.3 考察

まず、説明変数への口座情報の使用有無による精度差について述べる。前節で示した通り、3手法全ての場合において説明変数として口座情報を用いた場合の再現率が口座情報を用いなかった場合に比べ若干低くなった。本実験においては、モデル構築時の各パラメータのチューニングを最小限にしており、より高度なチューニングを行うことでモデルの予測精度が改善されることが考えられる。また、目的変数や説明変数の定義については比較的簡素なものを用いており、こちらについても変数の定義や算出手法を改良することで予測精度が改善されることが考えられる。

また、今回構築したモデルの予測精度評価指標のうち再現率に着目すると、XGBoostを用いたモデルでは口座情報の有無に関わらず80%を超えていることがわかる。この値は、今回構築した離反予測モデルを用いて顧客の離反予測を実施した場合に、3か月後に実際に離反する顧客のうち80%以上の顧客について「離反」と事前に予測が可能であることを示している。したがって、このモデルでは比較的高い精度で離反顧客を予測できていると言える。一方で、XGBoostを用いたモデルの場合では口座情報の有無に関わらず適合率が約30%となっており、これはモデルにより離反顧客と予測した顧客のうち実際に離反している顧客の割合が約30%に留まることを示している。よって、もしこのモデルに基づいてMMCカード顧客に対し引き留め施策を実施した場合、実施に要したコストのうち約70%は冗長であったことになる。そのため、本モデルを実際にMMCカードのマーケティング施策に用いる際には、再現率を高い水準に保ちつつ適合率を改善することが求められる。

5. ビジネス施策への活用の展望

本章では、今回構築した離反予測モデルについて、実際のMMCカードのマーケティング施策やその他の銀行の個人顧客に対するマーケティングへの活用可能性について述べる。

まず、今回構築した離反予測モデルをMMCカード(セゾン)のマーケティング施策に活用することを考える。離反予測の対象となる顧客は、直近1年間に決済履歴のある約80万人となる。今回構築した離反予測モデルのうちXGBoostを用いたモデル(口座情報あり)を用いると、評価実験において表2の通り計9,770人のうち2,902人(=2,031人+871人)が「離反」と予測されたことから、80万人の約29.7%にあたる約23万人が離反顧客と判定されることが考えられる。今回の離反予測モデルは毎月の利用状況による洗い替えを行ったうえで離反予測を行うことを想定しているため、もしこの離反が予測される顧客群に対し、引き留め施策として郵便を用いたダイレクトメール発送によるコミュニケーションを行う場合、一顧客あたりの施策コストを63円と仮定すると約1,400万円の施策費用が必要となる。しかしながら、すでに離反傾向にある顧客をこのタイミングで引き留めることは容易でなく、また今回構築した離反予測モデルの適合率が約30%と低位な段階において、毎月1,000万円を超える費用を投じることは現実的ではない。よって、今回構築した離反予測モデルを用いて引き留め施策を実施する場合には、一顧客あたりの施策コストを抑えられるメール配信を選択することが考えられる。また、離反予測モデルの適合率を改善させ、できるだけ実際に離反する顧客のみを離反顧客と判定するモデルへ改良することにより、より効果的な施策を実施できる可能性がある。

次に、離反予測モデルを個人顧客に対する銀行ビジネス全体のマーケティングに適用することを考える。前述の通り、離反予測モデルを用いることで離反顧客を事前に予測することは可能であるが、顧客はすでにMMCカードとは異なるサービスに移行しつつある段階であり、離反を引き留めることは困難ながらもみずほ銀行が提供するほかの決済サービスでニーズを充足可能な顧客が一定数存在することが予想される。例えば1か月の利用額が後日一括で引き落とされるクレジットカードよりも即時に引き落とされるデビットカードを好む場合には、みずほJCBデビットやみずほWallet搭載のSmart Debit(みずほWalletには事前チャージ式のMizuho Suicaも搭載)のプロモーション施策を機動的に実施することが考えられる。これにより、各顧客のニーズに沿ったみずほ銀行の決済サービスへの移行を促すことが可能となり、本来であれば外部に流出していた顧客をみ

ずほ銀行に留める効果が期待できる。

上記の通り、MMC カードの離反顧客モデルを構築することは MMC カードのマーケティング施策強化につながるだけでなく、みずほ銀行の提供する他の商品の顧客獲得に寄与することが考えられる。その結果として、みずほ銀行からの顧客流出を最小限に留めることが可能となり、みずほ銀行の個人顧客部門における収益拡大が期待される。また、みずほ銀行が提供するカードローンや投資信託などの他商品についても、本分析における考え方や定義を応用することで離反予測モデルを構築可能であると考えられる。さらに、みずほ銀行の全個人顧客のデータを一括で取得し予測モデルを構築することで、従来では銀行内の各商品担当部署が独立して実施していたマーケティングをフルラインで行うことも考えられる。

6. まとめと今後の課題

本分析では、みずほ銀行の発行する MMC カードの離反予測モデルを構築し、実データに対する離反予測精度の評価実験を行った。また実験結果をもとに、離反予測モデルを実際の MMC カードのマーケティング施策に活用する手法について述べた。さらに、今回提案した離反予測モデルの考え方が MMC カードだけでなくみずほ銀行の個人顧客に対するマーケティング全体に適用できる可能性についても考察した。

一方で、4 章で述べた通り今回構築した離反予測モデルは再現率については比較的高いものの適合率が低い。そのため、本モデルにより離反顧客と予測された顧客群に対し施策を実施した場合、施策コストが冗長となることから、予測精度の改善が必要であると考えられる。予測精度の改善手法としては、目的変数・説明変数の再定義や算出手法の見直し等が考えられる。特に口座情報から算出した説明変数については、離反予測モデルの予測精度向上への寄与が予想より小さく、別カードの利用状況と離反行動との関連度が高いとの仮説を裏付ける結果とはならなかった。これらの検討については、今後の課題としたい。

参 考 文 献

- [1] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). New York, NY, USA: ACM.
- [2] Ho, T. K. (1995). Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition (Vol. 1, pp. 278–282).