

データ科学

Vol.1

目次

機関誌「データ科学」発刊にあたって（松嶋敏泰）

プロジェクト報告

- 「みずほ銀行・早稲田大学」学術交流協定に基づく社内課題を用いた実践型社会人教育プロジェクト（須子統太）
- マーケティングにおける有効なアプローチの探索（岡田千果、影山京子、小林学）
- 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、須子統太）
- メールマーケティングにおける顧客の興味・関心度合いの推定と配信方式の検討（水止洋孝、大西将太、須子統太）

特集

- 空港における AI 活用事例，社会に広がる DX！AI・データサイエンスのビジネス化と直面するリスク（日立ソリューションズ・早稲田大学データ科学センターセミナーパネルディスカッションより）

2022 年 7 月

機関誌「データ科学」発刊にあたって

早稲田大学データ科学センター所長 理工学術院教授・松嶋敏泰

「現在の国勢を詳明せざれば、政府すなわち施政の便を失う。過去施政の結果を鑑照せざれば、政府その政策の利弊を知るに由なし。」

これは、早稲田大学の創設者である大隈重信が130年以上前に残した言葉で、現在の国の情勢を詳細に明らかにしなければ、政府は政治を執り行うことができない。また、過去の施政の結果と比較してみなければ、政府はその政策のよしあしを知ることができない、と統計（データ）からの意思決定の重要性を既に述べている。

データからの意思決定は、従来は、理工学などのデータが実験等で取得でき数学等で論理的に結論を導出できる（科学的方法と呼ばれる知の創造プロセスを用いた）科学分野や、データ取得や分析に多額の費用がかけられる一部の重要な意思決定でのみ行われていた。ところが、近年、情報・通信インフラの発展と、データを分析する理論や技術の進歩により誕生してきたデータ科学（データサイエンス）により、以前は不可能と思われていた広範な科学・学問領域や様々な意思決定問題においても、データを用いた科学的方法による意思決定が可能となってきている。

このデータ科学により、我々の意思決定や知の創造のプロセスが大きく変わり、社会に大きな変革が起こると捉えられている。知の創造と継承を担う拠点である大学は、この大きな流れを先導する核となることが期待されており、「早稲田大学データ科学センター」は大隈の意を源流として、2017年12月に全学的な組織として設置された（設立時の名称はデータ科学総合研究教育センター）。データ科学センターでは、教育面では様々な分野でデータを活用した意思決定ができる有用な人材の育成を目指し、研究面ではデータを基にした知の創造プロセスや意思決定プロセスの進化・深化により様々な分野における革新的研究や問題解決を目指してきた。

データ科学は、分野を問わないデータからの意思決定の科学と捉えられるメタ科学であり、学問としては体系化の途上にある発展中の学問でもあるため、上記の目的を目指したデータ科学センターの活動には様々な課題が存在している。例えば、専門の違う様々な学生に未だ寄せ集め的で学際的なデータ科学をどのように身に付けさせるか、社会における多様なデータからの様々な意思決定を大学のみで扱えるのか等の課題である。センターではそれらの課題に対して新たな試みの活動で解決を図ってきた。

まず、センターでは、全学の学生を対象として、データ科学のコアとなる考えかたを、統計学や機械学習やAI等の寄せ集め的な知識ではなく、統一した視点から身につけてもらうことを目的とした独自のデータ科学教育プログラムを構築した。この教育プログラムはオンデマンド授業をベースとした学習システムにより提供され、それぞれの学生の専門性、基礎力、目指すデータ科学の活用分野に合わせて、柔軟なカリキュラムが選択できるようにしている。さらに、その履修状況によりリテラシー級から上級までの認定も受けられる制度まで整えることができた。

センターの活動はこのような正規科目の教育プログラムの提供に限らず、テーマ、解析目的、利用データ等の条件を様々に変えて毎年全学規模で開催しているデータサイエンスコンペティション、データを用いた様々な分野の専門研究について個別相談が受けられるデータ科学研究相談など多岐に渡っている。このような正規科目提供以外の活動も加わることで、総合的なデータ科学の考え方や実践力が身に付き、様々な研究や実務の分野で活躍できる人材が育成されると考えている。

また、社会における多様なデータを用いた様々な意思決定についての教育と研究を行うためには、企業と大学の連携協力は欠くことができない。まず大学では入手できない多様なデータ、特に公的機関や企業の貴重なデータを教育研究に利用できることは大変ありがたい。このようなデータを利用可能とするため、企業と大学間の秘密保持契約のような法律的な面の対応から、研究者や学生の利用者の利便性を重要視しながらもセキュリティを保証する情報システムまで含めたトータルの仕組みとして「Waseda Integrated Research Platform (WIRP)」を構築している。

社会での様々な意思決定におけるデータ科学の活用については、学生が企業での実務体験としてそれらを学ぶ教育的インターシップも行われている。またその逆の方向として、センター独自のデータ科学教育プログラムは、専門や基礎知識のばらつきを吸収する柔軟な教育プログラムとなっているため、既に業務で培われた専門性に加えデータを活用するスキルをもった人材の企業での育成にも利用可能となると考えている。さらに社会人が実務におけるデータ活用を通じてデータ科学を実践的に学ぶプログラムも始まっており、センターの提供するこれらの社会人教育プログラムにより、社会全体を対象としてデータ科学を活用できる人材の育成も視野に入れている。

このように企業と大学が一体となった教育と研究における連携により、社会全体におけるデータ科学の浸透と発展が期待できると考えている。その連携の仕組みの一つとして、産官学連携部会、社会人教育部会、キャリア教育部会の3つの部会で構成される「早稲田大学データ科学研究教育コンソーシアム」の活動も行っている。

これらのデータ科学センターの教育と研究の定常的活動や、研究プロジェクトから様々な成果があがってきているものの、それらを蓄積し発信する仕組みは現在まで整っていなかった。正規科目などでは、オンデマンドコンテンツや教科書として、個別の研究プロジェクトにおいても学会発表や論文としてその成果が蓄積され発信されてきてはいるものの、その他の多岐にわたる活動の内容や成果についての情報の蓄積と発信は十分ではなかった。

分野を問わず社会全体に大きな影響を及ぼし、さらに発展し続けているデータ科学の教育と研究においては、ここまで述べてきたような多岐にわたる総合的で新たな試みが必要であり、データ科学センターが試みてきているそれらの情報を整理蓄積し発信することは大変有意義であると考えられる。その目的のため、早稲田大学データ科学センターでは機関誌「データ科学」をここに発刊する運びとなった。この機関誌により、まずはセンターの活動の一部の成果の発信からスタートすることになるが、徐々にセンターの多岐にわたる活動の多面的な成果を整理し発信できるよう発展していくことも考えていきたい。皆様のご協力のもと、この機関誌が直接的に間接的にデータ科学の発展に少しでも寄与できるような仕組みとして育っていくことを期待している。

「みずほ銀行・早稲田大学」学術交流協定に基づく 社内課題を用いた実践型社会人教育プロジェクト

早稲田大学データ科学センター教務主任 社会科学総合学術院准教授・須子統太

1. 「みずほ銀行・早稲田大学」学術交流協定の取り組み

早稲田大学とみずほ銀行は2018年7月13日の学術交流協定締結以来、データ科学分野における研究教育および社会貢献に関する様々な共同プロジェクトを推進してきた。このような学術交流協定はともすると実質を伴わない名ばかりの協定となってしまうこともあるが、協定締結以来、月に一度の定例会議を継続しながら常時複数の共同プロジェクトを推進し、着実に成果を積み上げてきた。例えば、社会科学的な価値の高いみずほ銀行の保有する様々なデータを早稲田大学の研究者が活用するため、個人情報等に配慮したセキュアなデータ提供環境を構築するプロジェクト[1]や、実際にそれらデータを活用した研究プロジェクト[2]などが進行している。

本稿では、それらみずほ銀行と早稲田大学の共同プロジェクトの一環として行った、実験的な社会人教育プロジェクトの取り組みについて紹介する。

2. データ科学分野における産学連携

近年、データ科学分野における大学と企業等との連携の機会は増えているが、それら連携方法を大まかに分類すると以下の図1の通りとなる。

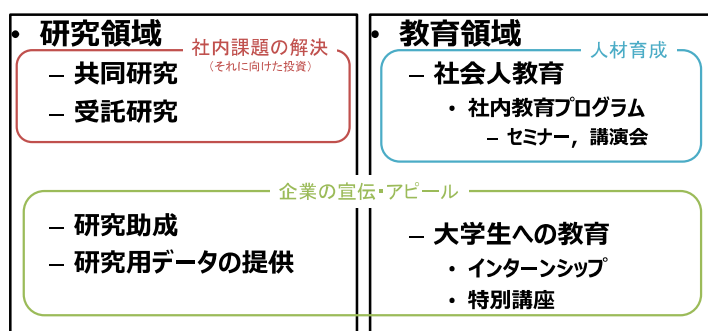


図1. データ科学分野における産学連携の方法

ここで、色で囲われた部分は「企業側から見た」目的別の分類である。この中で「社内課題の解決」を目的とした共同研究・受託研究と社内の「人材育成」を目的とした社会人教育の問題点について考える。

まず、データ科学分野における共同研究・受託研究であるが、企業内の解決したい課題と、大学研究者の研究としての興味にギャップが生じる事が多い。そのため、短期的・直接的な課題解決につなげるのが難しく企業側から見ると社内課題の解決というより、それにに向けた長期的・間接的な投資

として実施する場合が多い。もちろん、そのような長期的な投資自体には意義があるが、短期的・直接的な課題解決を期待する企業も多く、大学側側の期待とうまく折り合わない場合も少なくない。また、仮に運良く企業側と大学側の興味が合致し、短期的な課題解決に資する研究ができた場合においても、研究期間終了後の知識移転がうまくいかず、研究成果を社内で継続的に活用できない場合もある。

社会人教育については、多くの場合はセミナーなどの講義形式のプログラムが多く、データ科学に関する全般的な知識を得る事はできるが、実践的な知識が身につけにくく、社内の実課題への活用が難しい場合が多い。

そこで本プロジェクトでは、上記の受託研究・共同研究と社会人教育プログラムのそれぞれの問題点を解決する事を目的に、両者の中間的な仕組みを作る事で、データ科学分野における新しい産学連携の仕組みの実験を行った。

3. 実践型社会人教育プロジェクトの概要

今回実験した実践型社会人プロジェクトの方法はいたってシンプルである。みずほ銀行から「解決したい社内課題（テーマ）」「データ」「分析者（若手行員）」「メンター」を出してもらい、早稲田大学データ科学センターからは「指導教員」を出す。あとは大学内で行われているゼミや研究活動のように週に一度の定期ミーティングを設定し、社内課題解決に向けたデータの分析と活用方法の検討を続けた。このプロジェクトのポイントとしては、実際のデータ分析については全てみずほ銀行内で若手行員が自身の手で行い、指導教員は分析の方法や解釈等について指導するだけである点である。

このような仕組みを作ることで、若手行員が実際の社内課題・データを使ってデータ分析のスキルを実践的に学べるとともに、分析方法や結果の活用方法などの知識が社内に残り継続的に成果が活用できることを期待した。具体的には、みずほ銀行の若手行員5名を3チームに分け、3つのテーマ毎に上記プロジェクトを約1年間継続した。

4. プロジェクトの成果と今後の展望

各チームのプロジェクトの成果については、本稿の後に掲載しているプロジェクト報告書を是非ご覧頂きたい。

実際にプロジェクトを実施した結果を総括すると、各自の業務に直結するテーマを題材に分析手法等を学んだことで、データ科学的な視点や分析スキルの実践に関する教育的な効果は非常に高かったと感じた。その一方、「分析結果をそのまま実ビジネスに活用できる」という所までの成果を出すことはできなかった。その原因としては、指導教員の銀行ビジネスに対する理解が不足していた点が大きい。今後は銀行ビジネスについて精通しているメンターの関与の仕方などを改善し、教員側のビジネスに関する理解不足を補う仕組みを検討する必要があると感じた。

また、本取り組みの副次的な効果として、大学教員が実ビジネスに触れることで大学教員のスキルアップにも繋がる可能性を感じた。この指導経験を通し、大学内の実践的なデータ科学教育に対するフィードバックも期待できる。本プロジェクトは教員側の負担も決して少なくはないが、このような大学教員側のメリットもあり、新しい産学連携の方法として今後も仕組みの検討を継続してい

きたい。

[参考文献]

- [1] 小林 学, 中原悠太, 佐藤絵里子, 高谷親信, 塚本好宏, "銀行データを活用するための産学連携研究プラットフォームの構築," 情報処理学会第 84 回全国大会予稿集, pp.4-381-4-382, 2022 年 03 月.
- [2] So Kubota, Koichiro Onishi and Yuta Toyama, "Consumption responses to COVID-19 payments: Evidence from a natural experiment and bank account data," Journal of Economic Behavior & Organization, Vol. 188, pp. 1-17, August 2021.

マーケティングにおける有効なアプローチの探索

岡田 千果^{†1} 影山 京子^{†1} 小林 学^{†2}

アブストラクト: 昨今、企業ではデジタルトランスフォーメーション推進の観点でデータを活用した新しい価値やサービスの創出が加速し、金融業界でも顧客取引データやデジタル上の行動ログを活用する動きが広がりつつある。本プロジェクトではみずほ銀行の保有するデータを活用し、銀行サービスを顧客がどのような順序で、どの程度の人が利用するかを分析する。さらにサービス利用状況に対するプロモーションの効果を検証し、サービスプロモーションの施策の提案を行う。

キーワード: 銀行データ, 投資信託, iDeCo, マルコフモデル, プロモーション効果検証

1. はじめに

昨今、企業ではデジタルトランスフォーメーション推進の観点でデータを活用した新しい価値やサービスの創出が加速し、金融業界でも顧客取引データやデジタル上の行動ログを活用する動きが広がりつつある。1人1人が資産形成を行っていくことを経済政策の目標とし、個人資産増大・日本経済の成長・金融の好循環の促進のため、金融庁は「貯蓄から資産形成へ」をスローガン[1]を掲げている。みずほフィナンシャルグループにおいても、2019年度からの5ヵ年計画[2]にて「人生100年時代のライフデザインパートナー」を掲げ、リアル店舗とデジタルチャネルとの融合をはかり、お客さまの資産形成を促す施策「貯蓄から投資」プロモーションを展開している。そうした中で本プロジェクトではみずほ銀行が保有する膨大なデータをビジネス活用する可能性に焦点を当て、「貯蓄から投資」プロモーションのための対面・非対面の多くのチャネル活用と、そこから得られる顧客データを分析・活用し施策PDCAを実施することで、商品成約に至りやすい行動経路（ゴールデンルート）とそのための効果的なプロモーションを探索することを本分析の目的とする。

本プロジェクトは既存のデータを分析して資産形成商品のゴールデンルート探索を行うフェーズ1（以下Ph1）と、その結果を活用してメールによるプロモーションの効果検証を行うフェーズ2（以下Ph2）を実施した。具体的にはPh1では2018年～2020年の顧客のサービス利用状況からステータス遷移の傾向調査を実施し、マルコフモデルを用いてその後10年のサービス利用人数推移の予測を行い、投信・iDeCoを利用する人の増加傾向を示す。Ph2では、Ph1で得られた知見を用いて実

際にメールによるプロモーションを実施し、投信・iDeCoにおいて2021年12月末でのサービス利用者と未利用者に対するプロモーション効果の違いを統計的に明らかにする。最後に、これらの分析を踏まえたプロモーション方法の提案を行う。

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. 資産形成商品のゴールデンルート探索

2.1 分析目的と利用データ

みずほ銀行の顧客は金融商品を使い始めるまでに様々なフローを通過しており、行動経路の解析は商品プロモーションに対するアプローチの足掛かりになると考えられる。そこで我々はみずほ銀行が保有する顧客情報から行動を分析し、遷移確率の高い行動経路（ゴールデンルート）を探索することで、プロモーションに活用することを目的とする。Ph1で利用するデータはみずほ銀行顧客から抽出した対象者の属性情報の他、2018年12月末、2019年12月末、2020年12月末の各時点における個人向けインターネットバンキング:みずほダイレクト（以下、ダイレクト）と金融商品のサービス利用状況とした。具体的な顧客の属性情報とサービス利用状況を以下に示す。

- 顧客属性情報: 20~30代, 40~50代, 60~70代
(注: 2018年12月時点)
- サービス種別: ダイレクト, 投資信託（以下投信）, iDeCo

データは表1のようにサービス利用があれば1, 未利用であれば0で表現し、8種類のステータスに割り振っ

^{†1} みずほ銀行デジタルマーケティング部

^{†2} 早稲田大学データ科学センター

た. 例えば, ダイレクトと投信を利用している人は「ダイレクト&投信」というステータスに属する.

表 1: データ例

顧客ID	年代	2018年12月末時点			2019年12月末時点			ステータス (2018年)	ステータス (2019年)
		ダイレクト	投資信託	iDeCo	ダイレクト	投資信託	iDeCo		
A	60~70代	0	1	0	1	1	0	投信のみ	ダイレクト&投信
B	20~30代	1	1	0	1	0	0	ダイレクト&投信	ダイレクトのみ
C	40~50代	1	0	1	1	0	1	ダイレクト&iDeCo	ダイレクト&iDeCo

2.2 マルコフモデルによる分析

前節で述べた分析対象のデータを分析してステータス遷移の傾向調査を行い, マルコフモデル[3]を用いて調査後 10 年のサービス利用人数推移予測を行なう.

ステータス遷移の傾向調査では, 対象者を 3 時点 (2018 年末, 2019 年末, 2020 年末) のそれぞれにおいて 8 種類のステータスに振り分け, 各時点のステータスごとの人数の比較を行う. また, 2018 年→2019 年と 2019 年→2020 年でのステータスの遷移状況を調査し, みずほ銀行顧客の傾向を把握する.

10 年間のステータス推移予測について説明する. まず年を表す変数を y とし, y の時のサービス利用ステータス i の確率を $p_i^{(y)}$ とする. また y でステータス i の人が $y+1$ の時にサービス利用ステータス j になる条件付き確率を $q_{ji}^{(y)}$ と表現する. ここで $\mathbf{p}^{(y)} = (p_1^{(y)}, \dots, p_8^{(y)})$ と表し,

$$Q^{(y)} = \begin{bmatrix} q_{1|1}^{(y)} & \dots & q_{8|1}^{(y)} \\ \vdots & \ddots & \vdots \\ q_{1|8}^{(y)} & \dots & q_{8|8}^{(y)} \end{bmatrix} \quad (1)$$

と定義するとサービス利用状況をマルコフモデルを用いて定式化することができ, 1 年後は式(2)と表現できる.

$$\mathbf{p}^{(y+1)} = \mathbf{p}^{(y)} Q^{(y)} \quad (2)$$

従って再起的に n 年後のサービス利用状況が計算でき, 式(3)により予測が可能になる.

$$\mathbf{p}^{(y+n)} = \mathbf{p}^{(y)} Q^{(y)} Q^{(y+1)} \dots Q^{(y+n)} \quad (3)$$

本分析では 10 年後までの各ステータスにおけるサービス利用人数推移について予測を行う.

2.3 分析結果および考察

本分析の対象を 3 時点それぞれにおける 8 種類のステータスの人数シェアを表 2 に示す.

表 2: 各時点における各ステータスの人数シェア

ステータス	2018年12月末時点	2019年12月末時点	2020年12月末時点
すべてなし	47.64%	45.00%	42.94%
ダイレクトのみ	51.00%	53.43%	55.25%
投信のみ	0.24%	0.25%	0.24%
iDeCoのみ	0.01%	0.02%	0.03%
ダイレクト&投信	1.03%	1.15%	1.30%
ダイレクト&iDeCo	0.07%	0.12%	0.18%
投信&iDeCo	0.00%	0.00%	0.00%
すべてあり	0.01%	0.03%	0.06%

ここでは全体の傾向を捉えるための「ダイレクト利用の有無」と, 人数シェアの変化が大きい「ダイレクトのみからの遷移, 「ダイレクト&投信からの遷移, 「ダイレクト&iDeCo からの遷移, 「全てあり (ダイレクト&投信&iDeCo) への遷移」について結果を示す.

まずダイレクト利用の有無について傾向を見る. 2019 年から 2020 年においてダイレクト利用なしおよびダイレクト利用ありの状態を継続している人の割合を表 3 に示す. 表 3 では, 例えば 20~30 代で, 2018 年末にダイレクトなし (遷移前) で, かつ 2019 年末にダイレクトなし (遷移後) の割合が 7.6% いたことなどを表している. この表から, ダイレクト利用なしの状態を継続している人数はどの年代でも 2019 年よりも 2020 年の人数は減っており, ダイレクト未利用者の人数は年々減少傾向にあることがわかる. 一方ダイレクト利用ありの状態を継続している人数は年々増加傾向にあることがわかる.

またダイレクト未契約者と契約者それぞれの年代における人数比較を行った結果, 年代によってダイレクトの契約状況が大きく異なることが明らかとなった.

20~30 代はダイレクト契約者が未契約者の 2 倍以上

40~50 代はダイレクト契約者のほうが多い

60~70 代はダイレクト未契約者のほうが多い

以上から, 若い世代ほどダイレクトの継続利用者の割合が増加してきており, 世の中的なデジタルシフトの波に合わせてみずほ銀行のお客さまにおいてもオンラインサービス意向は若いほど加速しているという示唆が得られた.

表 3: ダイレクト有無における人数推移

年代	遷移前	遷移後	人数シェア (2018年→2019年)	人数シェア (2019年→2020年)
20~30代	ダイレクトなし	ダイレクトなし	7.6%	6.4%
40~50代	ダイレクトなし	ダイレクトなし	20.1%	19.5%
60~70代	ダイレクトなし	ダイレクトなし	17.6%	17.6%
20~30代	ダイレクトあり	ダイレクトあり	18.5%	20.1%
40~50代	ダイレクトあり	ダイレクトあり	22.5%	23.5%
60~70代	ダイレクトあり	ダイレクトあり	11.1%	11.5%

次に, 人数シェアの変化が大きい「ダイレクトのみからの遷移, 「ダイレクト&投信からの遷移, 「ダイレクト&iDeCo からの遷移, 「全てあり (ダイレクト&投信&iDeCo) への遷移」について傾向を見る.

各ステータスにおける前年末対比の増加率を表 4 に示す. 表 4 では, 例えば 2018 年末から 2019 年末のダイレクトのみにおける増加率は 104.68% で, 2019 年末から 2020 年末の増加率は 103.36% であることなどを表している.

表 4 から, 着目した 4 つのステータスどれにおいても年々増加傾向にあることがわかる.

表 4：各ステータスにおける前年末対比の増加率

ステータス	前年末対比の増加率	
	2019年末時点	2020年末時点
ダイレクトのみ	104.68%	103.36%
ダイレクト&投信	111.23%	113.00%
ダイレクト&iDeCo	181.71%	147.07%
すべてあり	226.49%	184.62%

ダイレクト&iDeCo 利用者のステータスを持つ顧客が、次の年にどのステータスになるか、その人数を表 5 に示す。表 5 より、大部分が同じステータスのままであることがわかる。また継続利用だけではなく投信を追加して「全てあり（ダイレクト&投信&iDeCo）」へ遷移する人もおり、iDeCo だけではなく他商品も興味を持つ層が一定割合存在することがわかる。

表 5：ダイレクト&iDeCo からの遷移

ステータス	ダイレクト&iDeCoからの遷移割合	
	2018年末→2019年末	2019年末→2020年末
すべてなし	0.00%	0.00%
ダイレクトのみ	0.00%	0.00%
投信のみ	0.00%	0.00%
iDeCoのみ	0.04%	0.04%
ダイレクト&投信	0.00%	0.00%
ダイレクト&iDeCo	96.60%	95.88%
投信&iDeCo	0.02%	0.00%
すべてあり	3.35%	4.07%

ダイレクト&投信から他のステータスへの遷移人数を表 6 に示す。表 6 からダイレクト&投信の 90%以上は継続利用しているが、投信をやめて「ダイレクトのみ」へ移る人数も一定数存在することがわかる。一方、iDeCo も追加して「全てあり」ステータスに遷移する人も一定割合存在しており、投信と iDeCo の同時利用をする人が増えていることも分かる。

表 6：ダイレクト&投信から他のステータスへの遷移

ステータス	ダイレクト&投信からの遷移人数	
	2018年末→2019年末	2019年末→2020年末
すべてなし	0.02%	0.01%
ダイレクトのみ	7.59%	7.87%
投信のみ	0.05%	0.09%
iDeCoのみ	0.00%	0.00%
ダイレクト&投信	91.68%	91.30%
ダイレクト&iDeCo	0.04%	0.04%
投信&iDeCo	0.00%	0.00%
すべてあり	0.63%	0.69%

「全てあり」ステータスとなった人の前年末時点のステータス内訳を表 7 に示す。表 7 から「すべてあり」ステータスの人の人数シェアのうち 40%~50%はほかのステータスからの遷移であり、上述の「ダイレクト&iDeCo」「ダイレクト&投信」から「すべてある」への

遷移傾向と一致する。

表 7：「全てあり」への遷移

ステータス	ほかのステータスからすべてありへ遷移した人数	
	2018年末→2019年末	2019年末→2020年末
ダイレクトのみ	21.26%	18.75%
ダイレクト&投信	21.29%	14.03%
ダイレクト&iDeCo	7.19%	8.61%
すべてあり	41.77%	51.18%

ここまでで注目した結果について、前節で示したマルコフモデルの式(3)を用いて、10 年後までの人数推移の予測を行う。ここで予測を行うためにはステータスの遷移確率が必要であるが、 $Q^{(2018)}$ と $Q^{(2019)}$ はデータから推定できるため、その推定結果を $Q^{(2020)}$ から $Q^{(2029)}$ までに当てはめて予測を行う。

まずダイレクト未契約・契約者数における予測の結果を図 1 に示す。図は上段がダイレクト未契約者の予測推移、下段がダイレクト契約者の予測推移を示しており、上下グラフの人数合計が各年代の合計の人数になっている。ここで $\hat{Q}^{(2018)}$ を用いた予測を青色、 $\hat{Q}^{(2019)}$ を用いた予測をオレンジ色、青とオレンジの平均を緑色で表す。

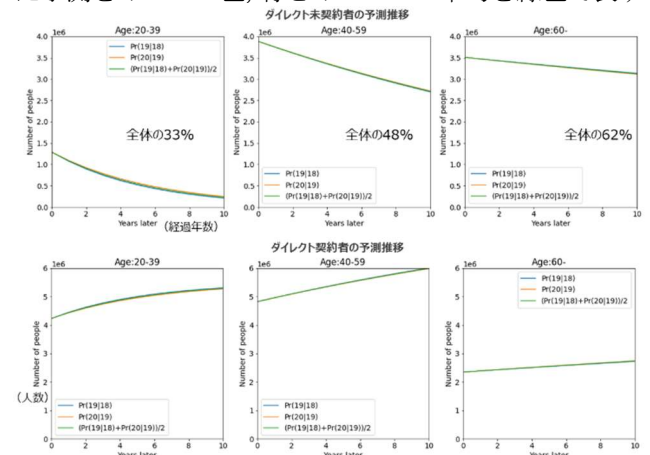


図 1：ダイレクト未契約・契約者数における予測

ダイレクト契約者（下段）の予測推移から、ダイレクト契約者は今後 10 年どの年代でも増加傾向にあることがわかる。特に、20~50 代の契約者の増加率は高く、20~30 代は 10 年後ダイレクト契約者が全体の多くを占めると予測される。一方で、60~70 代は契約者の増加は緩やかなため、ダイレクト利用者を増加させるには増加率の高い 40~50 代のうちに利用を促進する必要があると考えられる。

次にダイレクト&iDeCo 利用人数の予測結果を図 2 に示す。図 2 における増加人数の内訳として、20~30 代に対し 40~50 代は 2 倍以上となり、今後 10 年においても需要の高まりと市場規模の拡大が期待できる。

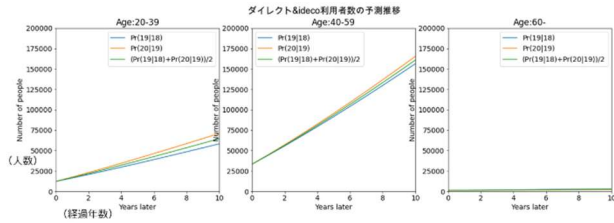


図 2：ダイレクト&iDeCo 利用人数における予測

20~30 代は 2018 年→2019 年よりも 2019 年→2020 年の増加率が高く、今後さらにダイレクト&iDeCo の人数増加は加速していくことから、今のうちに若年層への iDeCo 訴求など、啓もう活動が有効であると考えられる。

ダイレクト&投信の予測結果を図 3 に示す。どの年代でも増加傾向にあり、特に 40~50 代は他の年代に比べ 10 年後の人数の増加が著しいことがわかる。20~50 代においては、2018 年→2019 年よりも 2019 年→2020 年の増加率も高いことから、20~50 代の早い段階から投信の利用を開始してもらえ施策を行う必要があると考えられる。また、2018 年~2020 年の結果から、投信をやめてダイレクトのみへ移る人数も一定数存在するため、新規利用促進だけではなく継続して利用し続けてもらえる施策を打つことは有効と示唆される。

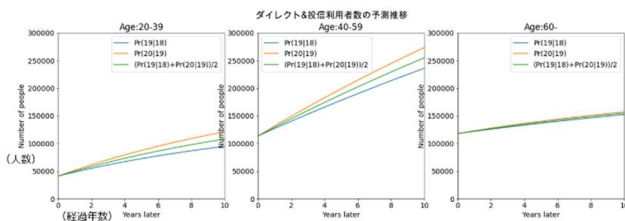


図 3：ダイレクト&投信利用人数における予測

最後に「全てあり」の 10 年後の予測結果について図 4 に示す。「全てあり」では、20~50 代について増加傾向にあり、特に 40~50 代は著しく増加する傾向にあった。また、2018 年→2019 年よりも 2019 年→2020 年の増加率が高く、今後「全てあり」の人数も年々増加していくと予測される。このことから、資産形成や老後における意識の高い 40~50 代には積極的なアプローチが有効で、投信・iDeCo になじみのない 20~30 代には金融教育的なアプローチが有効であると推察される。

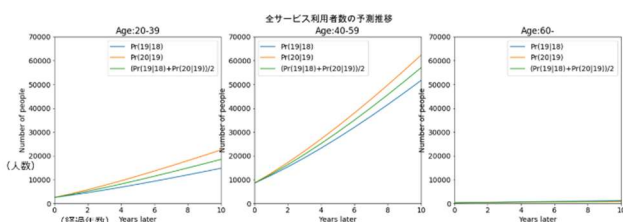


図 4：全てありにおける予測

3. 商品プロモーションの実施と有効性の評価

3.1 目的・データ

Ph1 では、投信・iDeCo を合わせて利用してステータスが「全てあり」となる人が年々増加傾向にあることが示された。特に投信と iDeCo に関して、両方のサービスの未利用者と比べてどちらかのサービスをすでに利用している人が新しくサービスを始める（つまり「全てあり」になる）際のハードルは低い可能性がある。この仮説を検証するため、iDeCo もしくはつみたて NISA（投信）に関するプロモーションメールの配信を実施した。

対象者は以下の①~③いずれかに該当する人とし、iDeCo プロモーションメールは①と③、つみたて NISA（投信）プロモーションメールは②と③を対象とした。

- ①投資口座保有かつ iDeCo 未契約者
- ②投資口座未保有かつ iDeCo 契約者
- ③投資口座未保有かつ iDeCo 未契約者（投資運用商品未保有者）

iDeCo プロモーションメールに対する対象の年齢分布を表 8 に示す。一方投信プロモーションメールに対する対象の年齢分布を表 9 に示す。各プロモーションメール対象の抽出において、年代分布がみずほ銀行顧客の年代分布と同程度になるようにした。

表 8：iDeCo プロモーションメール対象の年齢分布

ターゲット	20代	30代	40代	50代
①投信口座保有かつiDeCo未契約者	6.5%	10.2%	14.6%	18.7%
③投信口座未保有かつiDeCo未契約者	6.5%	10.2%	14.6%	18.7%
合計	100.0%			

表 9：投信プロモーションメール対象の年齢分布

ターゲット	20代	30代	40代	50代
②投信口座未保有かつiDeCo契約者	3.5%	11.2%	17.5%	17.7%
③投信口座未保有かつiDeCo未契約者	3.5%	11.2%	17.5%	17.7%
合計	100.0%			

3.2 分析の方法

本分析では、iDeCo プロモーションメール及び投信プロモーションメールを配信した対象において金融商品利用有無による反応の違いを調査する。

まず、iDeCo プロモーションメールでは対象者①と③に対してメール中に記載された URL のクリック率、及び投信プロモーションメールでは②と③に対してメール中に記載された URL のクリック率を比較することで、iDeCo もしくは投信に対する商品利用のハードルの違いを調べる。そのためにフィッシャーの正確確率検定 [4] とオッズ比を用い、金融商品利用有無における統計的有意差があるかを検証する。

3.3 分析結果および考察

iDeCo プロモーションメール、投信プロモーションメールにおける配信結果を表 10 及び表 11 に示す。ここでは配信した対象者におけるメールの開封率、並びにメール中の URL をクリックしたクリック率を示している。各メールにおいて、どちらも③の対象者より①、②とどちらかの商品をすでに利用している人のほうがクリック率が高く、投信と iDeCo に関してどちらかすでに利用している人はサービス未利用者よりも新しくサービスを始める際のハードルが低いことが示唆されている。

表 10：iDeCo プロモーションメールにおける反応

ターゲット	開封率	クリック率 (対配信)
①投信口座保有かつiDeCo未契約者	33%	2.6%
③投信口座未保有かつiDeCo未契約者	29%	1.8%

表 11：投信プロモーションメールにおける反応

ターゲット	開封率	クリック率 (対配信)
②投信口座未保有かつiDeCo契約者	33%	1.8%
③投信口座未保有かつiDeCo未契約者	26%	1.7%

また、この結果をもとにフィッシャーの正確確率検定を行った結果を表 12 に示す。結果からどちらも P 値は 1% 有意となり、①と③または②と③の間には統計的に有意な差があることが示された。

さらに、オッズ比で比較すると、①の対象者に対する iDeCo プロモーションメールは、②の対象者に対する投信プロモーションメールよりも効果があることが示唆された。以上の分析から「投信と iDeCo に関して、両方のサービスの未利用者比べてどちらかのサービスをすでに利用している人が新しくサービスを始める際のハードルは低い」と考えられ、効果的なプロモーションにつながられる可能性を示唆している。

表 12：フィッシャーの正確確率検定における結果

メール種類	P値	オッズ比	オッズ比95%の信頼区間
iDeCoプロモーションメール	2.20E-16	1.68681	1.536~1.853
投信プロモーションメール	0.00289	1.375101	1.111~1.686

4. 資産形成商品プロモーション方法の提案

本節では前節までの分析と考察を踏まえ、いくつかの前提条件の下で資産形成商品のプロモーション方法の提案を行う。ここでは資産形成商品の未利用者（新規利用者向け）へのプロモーションと、既に利用している人へのクロスセルプロモーションに分けて提案する。

なお、以下をプロモーションの前提とする。

(1) 目的

みずほ銀行既顧客のライフデザインをサポートし、

長期的に寄り添うサービスを提供する。また、個人の資産形成を促し業績に貢献する。

(2) サービス

投信（NISA 含む）・iDeCo・ダイレクト

(3) ターゲット

- A：20 代～60 代のみずほ銀行既顧客かつ資産形成商品未利用者
- B：20 代～60 代のみずほ銀行既顧客かつ資産形成商品利用者

(4) プロモーションの施策主

デジタルマーケティング部マーケティング企画部、各商品所管部（商品を専門で推進している部署）

4.1 新規利用者向け（ターゲットA）

ここでは資産形成商品の未利用者に対して、投信及び iDeCo のプロモーション方法を提案する。年代によって傾向が異なることが本分析の Ph1 からわかっているため、20～30 代と 40～50 代に分けて述べる。

<20～30 代>

(1) 現状

Ph1 の結果より、20～30 代におけるダイレクト&iDeCo は 2018 年→2019 年よりも 2019 年→2020 年の増加率が高い。また、「ダイレクト&iDeCo」、「ダイレクト&投信」両ステータスともに、20～30 代グループ人数の増加は今後加速していくと考えられる。

(2) 目的

資産形成商品未利用者が資産形成を始めない理由は様々であり、最初の一步を踏み出すまでの時間が長くハードルは高い。そのため、プロダクトアウトのコミュニケーションでは費用対効果は十分に得られないと考えられる。そこでニーズ潜在段階からコミュニケーションをはかり、資産形成ニーズが顕在化したタイミングでみずほ銀行を選んでもらうためにナーチャリングを行い、顧客育成を図る。

(3) 方法

気持ちを動かすためには、考えるより感じてもらう（「Don't think. FEEL」）ことに軸を置き、インフルエンサーや Youtuber を起用することで、銀行というイメージを壊しつつ、流行に敏感な若年層を集客する（二次拡散）。

また、コミュニケーションチャネルとして、動画や漫画に特化した各種 SNS（TikTok/Instagram 等）や動画配信サービス（Youtube 等）を活用する。

<40～50 代>

(1) 現状

本プロジェクト Ph1 の結果より、ダイレクト&iDeCo

グループの増加分の内訳として、20~30 代に対し 40~50 代は 2 倍以上となり、増加傾向として大きい。iDeCo の認知が広がってきたことと、老後がリアルになってきたことによる資産形成ニーズが高まっていると考える。また、ダイレクト&投信についても、20~50 代は増加が加速しているのに対し、60~70 代は増加が緩やかなため、早い段階から投信の利用を開始してもらうことが重要である。

(2) 目的

老後 2000 万円問題や年金問題から老後がリアルになってきたが対策方法がわからないニーズ顕在層を効率よく捕捉し、解決先の一つとして iDeCo を訴求し成約貢献する。資産形成ニーズやデジタルシフトが緩やかになる 60~70 代になる前に非対面での資産形成を促したい。また、人生 100 年時代のパートナーとしてみずほ銀行を選んでもらうために、ライフスタイルの変化が大きい（不安が大きい）この時期に、みずほ銀行への信頼感や愛着を高める。

(3) 方法

みずほ銀行が資産形成の伴走者として、ストレスを与えない有益なコミュニケーションを前提とする。コミュニケーションチャネルとして、連絡手段としての性格が強い LINE 等クローズド SNS やメールを活用し、パーソナライズされたメッセージ配信およびコミュニケーションにより資産形成ニーズが顕在化したタイミングを補足し非対面での成約を促す。

また、人生のパートナーとして伴奏し、サポートしていくため、みずほ銀行行員が出演する資産形成教育コンテンツを移動時間などの隙間時間で視聴できる短尺動画にして配信し、信頼性のある有益な情報提供および不安解消の手助けとなることで、みずほ銀行とのエンゲージメントの向上を促す。

4.3 資産形成商品クロスセル（ターゲットB）

(1) 現状

本プロジェクト Ph2 の結果より、投資運用商品をすでに扱っている人は、ほかの投資運用商品に対するハードルが低い。投信利用者への iDeCo 訴求、iDeCo 利用者への投信訴求等、積極的なセールスは有効と考える。

(2) 目的

投信利用者への iDeCo 訴求、iDeCo 利用者への投信訴

求、また、投信利用者への追加購入訴求を行うことで、クロスセル・アップセルを促進し、収益貢献する。

(3) 方法

他の資産形成商品利用のハードルが低いことがわかっているため、分かりやすいメリット訴求を行いつつ、人生 100 年時代におけるライフデザインのパートナーとして、みずほ銀行への信頼感や愛着を感じてもらい、クロスセル・アップセルを促進するために、投資運用利用サービス数または金額により、インセンティブを付与するロイヤルティプログラムを実施するなど、良質な顧客体験を提供する。

5. まとめと今後の課題

本分析では、多種多様なデータを分析・活用することにより、「貯蓄から投資」プロモーションにおける新しい施策の立案に際して定量的な裏付けを持った施策に昇華することが可能であることが分かった。

今後は、本分析を通して得られた示唆をもとに更に分析を重ね、みずほ銀行全体のサービスおよび顧客体験価値の向上につなげていく。

また、世の中の情勢や景気・相場などの外的要因や顧客の資産運用における成功体験などの定性的な側面からの分析を別の切り口として分析を重ねていくことでより精度の高いアプローチを探索していく。

参 考 文 献

- [1] 金融庁, “平成 27 事務年度 金融レポートの主なポイント”, 金融庁, 6-7, 2016
- [2] みずほフィナンシャルグループ, “5 カ年経営計画”, みずほ FG, 2019, <https://www.mizuho-fg.co.jp/company/strategy/plan/index.html>, (参 照 2021-01-17)
- [3] C. M. Bishop, パターン認識と機械学習, 丸善出版, 2012
- [4] R. A. Fisher, “On the interpretation of χ^2 from contingency tables, and the calculation of P”, Journal of the Royal Statistical Society 85 (1), 87–94, 1922.

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%

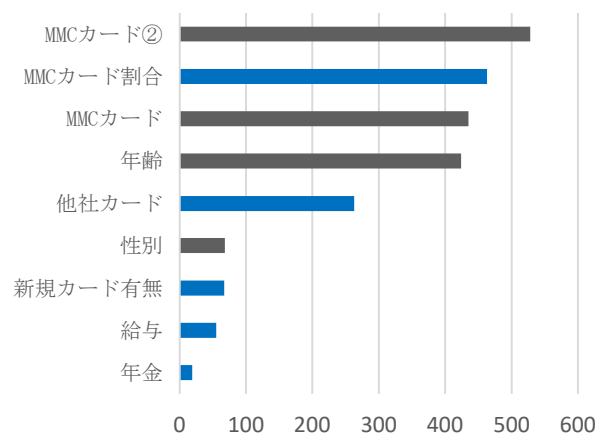


図 2 説明変数の重要度 (XGBoost, 口座情報あり)

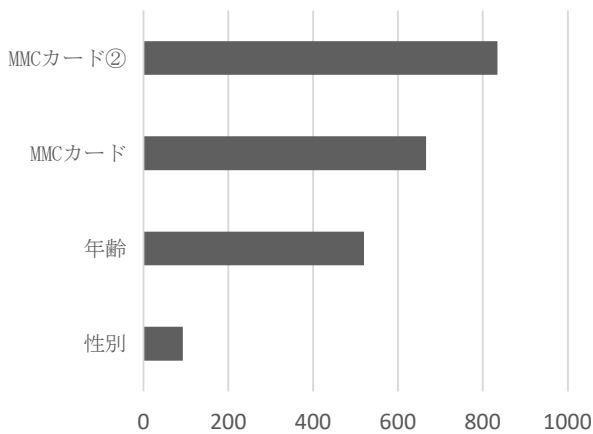


図3 説明変数の重要度 (XGBoost, 口座情報なし)

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).

メールマーケティングにおける 顧客の興味・関心度合いの推定と配信方式の検討

水止洋孝^{†1}, 大西将太^{†2}, 須子統太^{†3}

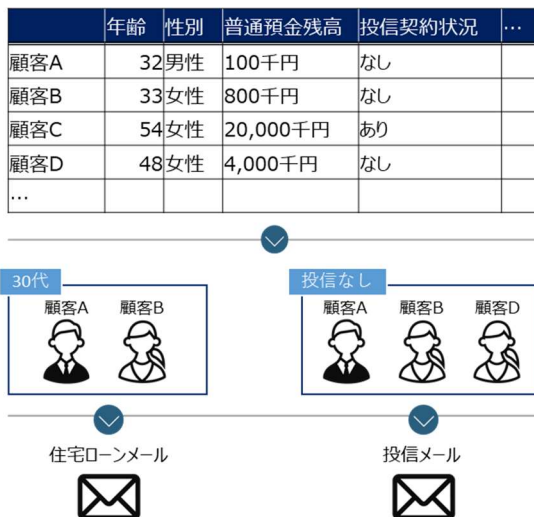
アブストラクト：電子メールを活用したメールマーケティングは主要なマーケティング手法の一つであり、みずほ銀行でも顧客との非対面コミュニケーションを設計する上で、重要な手法であると位置づけている。本分析では、顧客の状態に合わせたメールマーケティングを実現すべく、顧客の興味・関心の推定することで、興味・関心度合いと配信方法に対する反応の関係について検証した。

キーワード：メールマーケティング、ロジスティック回帰、予測モデル

1. はじめに

メールマーケティングとは電子メールを通じて顧客との非対面コミュニケーションを設計し、集客やコンバージョンの獲得へと繋げるマーケティングの手段の一つである。メールマーケティングにかかるコストは他のマーケティング手法と比較しても低く、顧客を特定した上での1対nのコミュニケーションを簡便に行えることから、大規模での自社顧客との非対面コミュニケーション設計において、非常に効果的な手法であると言える[1]。

図1. 現在のメール配信方法のイメージ



みずほ銀行においても、ダイレクトメール配信は主要な非対面チャネルと位置付けた上で、メールマーケティングに取り組んでいる。現状のみずほ銀行のメールマーケティングは図1のように、顧客の属性（性別・年齢・契約状況）等で対象者を選定し、ダイレクトメールを配

信する施策が大宗である。これらは、既存の契約者の傾向分析によって、成約確率が高いであろうと類推された集団、または戦略上、コンバージョンを狙っていくよう定めた集団をターゲティングすることが基本となっている。

しかしながら、このようなメールマーケティングは画一的かつ一方的なコミュニケーションであることにより、真に顧客便益に資するような情報提供に至っていない。近年はOMO (Online Merges with Offline) の概念が普及したことにより、よりリアルタイムに状況に合わせた非対面コミュニケーションを設計することが求められている。その中で、メールマーケティングも顧客の状態や状況に合わせて配信し、真に顧客便益に資する配信が求められ、当行としてそのあり方を目指すべき姿の一つとして位置付けている。顧客の状態や状況を考慮したメールマーケティングを構築する上で、「顧客の商品に対する興味・関心」、「顧客の嗜好性」、「リアルタイムで訪れている場所・状態」などの情報が重要であると考えられている。

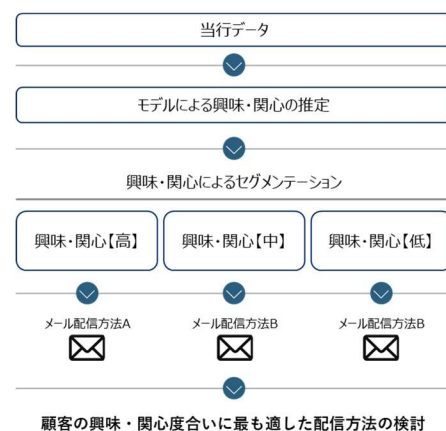


図2. 本分析のフロー図

この中でも「顧客の商品に対する興味・関心」は真に

^{†1} みずほ銀行デジタルマーケティング部

^{†2} みずほ銀行決済ビジネス推進部

^{†3} 早稲田大学社会科学総合学術院

顧客便益に資するための要素として考えられる。そこで本稿では、顧客の商品に対する潜在的な興味・関心を推定し、その興味・関心の度合いに沿ったダイレクトメール配信方法について検討を行う。具体的には、顧客の興味・関心を推定するモデルを作成し、興味・関心の度合いを数値化する。そのもとで、推定した興味・関心度合いによりセグメンテーションされたユーザーに対して、複数のメール配信方法を実験することで、興味・関心の度合いとメール配信方法との関連について考察する。(図 2)

2. 興味・関心推定モデルの作成

2. 1 興味・関心度合いの定義

本来、対象商品に対する興味・関心を直接的に測るには、アンケート調査等を実施する必要があるが、すべての顧客に対し調査を行うことは現実的ではない。そこで、本分析では対象商品の契約のしやすさを興味・関心の度合いと定義づけることで間接的に興味・関心の推定を行う。

本稿では、対象商品の成約のしやすさを予測するにあたり、目的変数を対象商品の契約有無、説明変数を顧客情報(付録, 表 7)としてモデル作成を行った。また、対象商品としては、みずほ銀行の主要商品のひとつである投資信託(以下投信)に着目し、興味・関心の度合いを推定した。

2. 2 実験データの作成

実験に使用するデータは以下の手順により作成した。

手順 1: サンプルの抽出

普通預金口座保有者の中から、ランダムに 10 万人の顧客サンプルを抽出した。

手順 2: 顧客リストに対する顧客属性の付与

手順 1 で抽出した顧客リストに対して、46 個の属性・契約状況の情報を結合した(付録, 表 7)。

手順 3: トレーニングデータとテストデータの作成

手順 2 で作成したデータの中からランダムに 10,051 サンプルを抽出し、検証用のテストデータとした。対象商品の「契約あり」と「契約なし」の比率は 1:14 であり偏りが大きいため、モデルを学習するトレーニングデータとしては、テストデータを除外したデータセットの中から、投信の「契約あり」顧客および「契約なし」顧客をそれぞれ 5,972 サンプルずつ同数ずつ抽出し、合算したデータセット利用した。

2. 3 興味・関心推定モデルの作成

対象商品の「契約あり」「契約なし」を目的変数とし

たロジスティック回帰モデル[2]を作成することで、興味・関心推定モデルとした。ロジスティック回帰モデルとしては、表 1 の全変数を説明変数として用いたモデルと、AIC 値をモデル選択基準としたステップワイズ法により変数選択を行ったモデルについて検討した。それぞれの回帰係数の推定結果は付録の表 8 と表 9 に示す。

2. 4 興味・関心の推定結果

ロジスティック回帰分析によって推定された各変数における係数を用いてテストデータ内の各顧客の投信契約有無を予測した。その際、契約の有無は判別スコア 0.5 を基準として判断した。その後、推定された値と真の値を用いて混同行列を作成し、適合率、再現率、正解率、F 値を算出した。

その結果、全変数を用いてモデルを作成した場合、正解率は 95.4% と高い水準にあるものの、適合率、再現率は 20~30% の間にあり低い水準にあることが分かった。

表 1. 全変数を用いたロジスティック回帰モデルによる予測結果

		予測	
		投信契約あり	投信契約なし
真の値	投信契約あり	79	192
	投信契約なし	336	9,443
適合率	19.0%	再現率	29.2%
正解率	94.7%	F 値	23.0%

また、ステップワイズ法により変数選択したモデルを用いた場合、適合率が 75.0%、正解率が 97.3% と高い水準になったが、再現率は 2.2% と低く実用に耐えうるモデルであるとは評価できなかった。

表 2. ステップワイズ法を用いたロジスティック回帰モデルによる予測結果

		予測	
		投信契約あり	投信契約なし
真の値	投信契約あり	88	183
	投信契約なし	424	9,335
適合率	17.2%	再現率	32.5%
正解率	94.0%	F 値	22.5%

そこで、予測精度を改善するべく、ステップワイズ法を用いたモデルに対し、判別スコアの閾値を 0.1~0.9 の間を 0.1 間隔で変更し予測を行い、各閾値における混同行列を作成した(図 3~6)。その結果、適合率は閾値の値を高くするほど高くなった一方で、再現率は閾値の

値を高くするほど低くなった。正解率については判別スコア 0.4 までは値が上昇した。また、F 値については閾値を 0.4 にすることで急激に上昇し、0.6 を堺にピークが見られた。

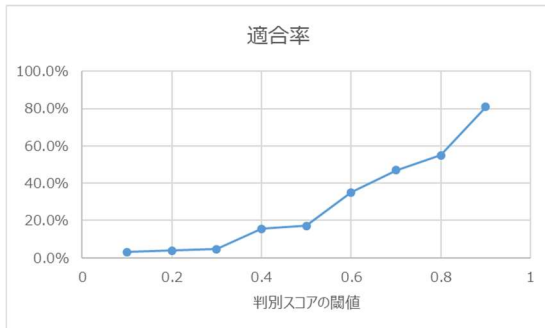


図 3. 判別スコアの閾値と適合率

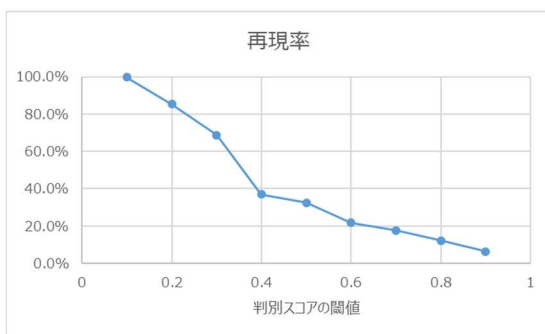


図 4. 判別スコアの閾値と再現率

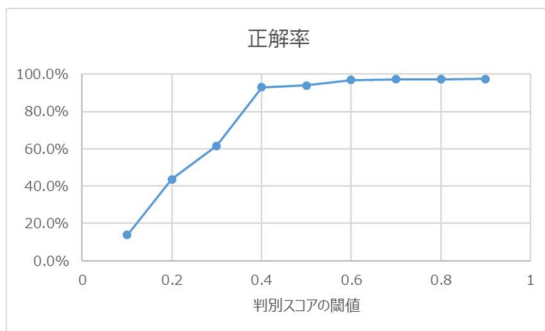


図 5. 判別スコアの閾値と正解率

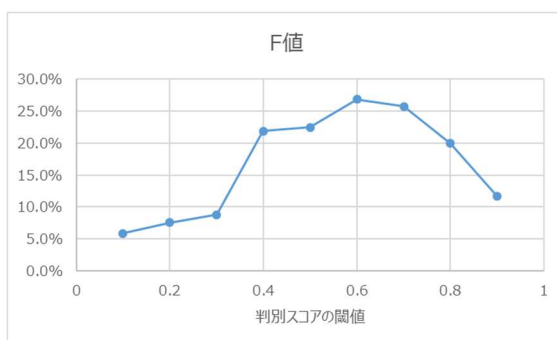


図 6. 判別スコアの閾値とF値

2. 5 考察

ロジスティック回帰分析の結果、定期預金といった貯蓄性のある商品や投資運用商品といった資産形成に関連する変数が投信の契約の有無と強く関連していると示唆された。これは、投資信託が主要な資産形成手段の一つであることから、資産形成に関心のある層が投資信託にも興味を示しているということが推察される。また、顧客セグメントやダイレクト契約の有無とも相関することが示唆されることから、顧客の資産背景、デジタル利用とも相関している可能性があると考えられる。

3. メール配信方法の検討

3. 1 配信方法の検討における方針

顧客の当行商品に対する興味・関心の度合いと異なるメール配信方法の組み合わせによる反応の違いを明確化すべく、作成したモデルを用いて顧客の投信に対する興味・関心を推定し、その度合いに応じて3つの群にセグメンテーションを実施した。各セグメントに対して2種類のメール配信方法で商品訴求を実施し、興味・関心の度合いとメール配信方法の適した組み合わせを検討した。

3. 2 興味・関心の度合いに応じたセグメンテーション

前述のステップワイズ法により変数選択したロジスティック回帰モデルを用いて、みずほ銀行のメール配信可能顧客（約5百万人）に対して判別スコアを算出し、判別スコアの値で顧客をセグメンテーションした。具体的には、判別スコアの値が0.2以下の群を興味・関心「低」、0.2～0.4の群を興味・関心「中」、0.4より高い群を興味・関心「高」としてセグメンテーション実施した。判別スコアの分布と各セグメントの関係は図7に示す通りである。

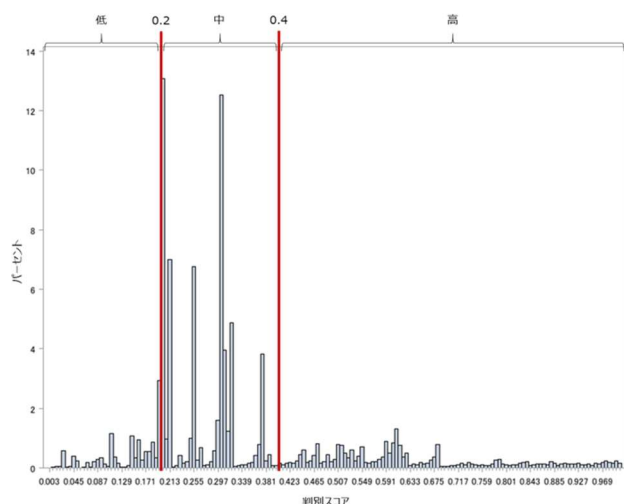


図 7. メール配信対象者における判別スコアの分布とセグメンテーション

3. 3 メール配信方法の検討

興味・関心度合いとメール配信方法との関係について実験するため、まず、意図の異なる以下の 2 種類のメールを作成した。

メール A：商品訴求の個別メール（付録・図 8）

投信に関して直接的に商品訴求するメールであり、投信の基本的な仕組みについて記載した記事および投資運用商品ページへの導線を設置している。

メール B：シナリオメール（付録・図 9）

資産運用に対する興味を喚起するメールであり、自身の資産について考えるきっかけを作る記事への導線を設置している。

このもとで、「方式Ⅰ：メール A のみを送付」「方式Ⅱ：メール B を送付後、メール A を送付」の 2 種類のメール配信方式について検討を行った。

前述の興味・関心度合いによる「低」「中」「高」の 3 つのセグメントに対し、それぞれランダムに 4 万人ずつ抽出し、更にその中の半数に対し方式Ⅰを、残りの半数に対し方式Ⅱによるメール配信を行った。メール配信は 2022 年 3 月 18 日（方式Ⅱのメール B）および 3 月 22 日（メール A）に行い、3 月 25 日までの開封率とメール内のリンクのクリック率を測定した。

どちらのメール配信方式においても商品に対する興味・関心が高いほど反応率が高くなることが想定される。一方で、方式ⅠとⅡで反応率を比較した際、興味・関心が高い群では顕著な差は無く、興味・関心が低い群ほど方式Ⅱの配信方法で反応率が高くなるという仮説を立てた。

3. 3 メール配信方法と興味・関心度合いの相関

実験の結果を表 3～6 に示す。方式Ⅰの開封率については、興味・関心「高」の顧客群が最も低く、興味・関心「高」とそれ以外の間で有意差が見られた。一方でクリック率については、興味・関心が高くなるほど値が高くなり、興味・関心「高」とそれ以外の間で有意差が見られた。

表 3. 方式Ⅰにおける各興味・関心群の反応率

興味・関心	低	中	高
開封率	36.3%	36.5%	34.9%
クリック率	0.19%	0.25%	0.38%

表 4. 方式Ⅰにおける各興味・関心群の反応率の母比率の差検定

興味・関心	高 vs 中	高 vs 低	中 vs 低
p値（開封）	0.0032	0.0014	0.6675
有意差（開封）	あり	あり	なし
p値（クリック）	0.047	0.0027	0.2698
有意差（クリック）	あり	あり	なし

また、方式Ⅱについてメール B を開封した顧客のメール A に対する反応率を計測したところ、開封率について、有意差はすべての興味・関心群の中で見られなかった。また、全ての興味・関心度合いにおいて、方式Ⅰと比較すると顕著な開封率の差が見られた。クリック率についても方式Ⅰと比較し、全体的に高い水準であった。また、クリック率が最大となるのが、興味・関心「高」であり、次点で興味・関心「低」、最後に興味・関心「中」となり、興味・関心「中」は個別メールとあまり変わらない水準であ

った。有意差についても、「高」対「低」では見られなかったが、それ以外の組み合わせでは見られた。

表 3. 方式Ⅱの 2 通目メールにおける
各興味・関心群の反応率

興味・関心	低	中	高
開封率	83.5%	82.0%	82.2%
クリック率	0.55%	0.26%	0.87%

表 4. 方式Ⅱの 2 通目メールにおける
各興味・関心群の反応率の母比率の差検定

興味・関心	高 vs 中	高 vs 低	中 vs 低
p値（開封）	0.7789	0.0847	0.4417
有意差（開封）	なし	なし	なし
p値（クリック）	0.0046	0.1037	0.0033
有意差（クリック）	あり	なし	なし

3. 4 考察

方式Ⅰ，方式Ⅱ共に興味・関心度合いによる開封率の違いに傾向はみられなかったが、クリック率については、興味・関心「高」の集団で顕著に高くなることが分かった。これは、契約有無を予測するモデルが商品に対する興味・関心を推定していることを示唆している。一方で、興味・関心「中」および「低」のクリック率について、方式Ⅰでは有意差が見られず、方式Ⅱでは有意に興味・関心「低」で高い値となることが分かった。また、方式Ⅰに対し、方式Ⅱでは興味・関心「中」を除き、顕著にクリック率が高くなった。このことから、興味・関心の低い層・高い層には方式Ⅱが適しており、それ以外の層ではどちらの配信方法でも変化がないということが示唆された。一方で、今回のセグメンテーションの閾値が適切ではなく、興味・関心「中」および「低」で逆転現象や有意差が見られないことも考えられる。従って、閾値の適切性についても今後検証の必要があると考える。

4. まとめと今後の課題

本稿では、商品の契約有無を予測するモデルを用いることで、商品に対する興味・関心の推定をした結果、おおそ興味・関心の度合いと予測値が関連しているような結果が得られた。このモデルを使用す

ることで、潜在的な興味・関心度合いに応じたコンテンツの出し分け等のパーソナライズされたマーケティングを実現することが可能になると考えている。また、今回の配信方法の検討においては、開封率・クリック率共にシナリオメールが優位であるということが分かったが、前送のメールの開封者等に送るシナリオメールの性質上、シナリオが進行する毎に対象者数が少なくなるため、成約数の獲得が困難になる。そこで今後、モデルで推定された興味・関心の度合いと反応率の相関を調べていくことにより、マーケティング目線で最も適切な配信方法を探ることが可能になると考える。

参考文献

- [1]Adobe Experience Cloud. ” メールマーケティングとは？基礎知識やメリット、効果的な手法を紹介” . Adobe Market Engage. 2021-12-06.
<https://jp.marketo.com/content/email-marketing-basic.html>
- [2] 北海道大学，東京大学，滋賀大学，京都大学，大阪大学，九州大学他. “テラシーレベルモデルカリキュラム対応教材”. 数理・データサイエンス・AI 教育強化拠点コンソーシアム. 2018.
<http://www.mi.u-tokyo.ac.jp/consortium/e-learning.html>

付録

表 7. 使用した説明変数

性別	家賃支払い口座	外貨定期預金開設日	住宅ローン契約有無	公共債残高
年金引き落とし口座	税金支払い口座	グローバル定期預金口座有無	住宅ローン実行日	保険残高
UCカード契約	学費支払い口座	グローバル定期預金開設日	カードローン契約有無	住宅ローン残高
ガス引き落とし口座	年代	保険契約有無	カードローン実行日	カードローン残高
電話料金引き落とし口座	普通預金開設日	保険開始日	その他ローン契約有無	その他ローン残高
NHK料金引き落とし口座	定期預金口座有	公共債契約有無	その他ローン実行日	給振有無
電気料金引き落とし口座	定期預金開設日	公共債契約日	顧客セグメント	
水道料金引き落とし口座	外貨普通預金口	MMC契約有無	円貨定期性預金残高	
貸金庫契約	外貨普通預金開	MMC加入日	円貨流動性預金残高	
クレジット引き落とし口座	外貨定期預金口	ダイレクト契約有無	外貨定期性預金残高	

表 8. 全変数を用いたロジスティック回帰分析結果

パラメータ	値	推定値	パラメータ	値	推定値
定数		-53.8597	グローバル定期預金口座有無	なし	44.6355
性別	男性	-0.2134	グローバル定期預金開設日		0.00223
年金引き落とし口座	なし	0.0655	保険契約有無	なし	7.8687
UCカード契約	なし	-0.2241	保険開始日		0.000437
ガス引き落とし口座	なし	0.6314	公共債契約有無	なし	7.9571
電話料金引き落とし口座	なし	-0.3379	公共債契約日		0.000493
NHK料金引き落とし口座	なし	-0.2114	MMC契約有無	なし	-0.1569
電気料金引き落とし口座	なし	-0.1102	MMC加入日		9.32E-06
水道料金引き落とし口座	なし	0.00526	ダイレクト契約有無	なし	-0.9189
貸金庫契約	なし	-1.0841	住宅ローン契約有無	なし	4.5779
クレジット引き落とし口座	なし	-0.2856	住宅ローン実行日		0.00023
家賃支払い口座	なし	-0.1898	カードローン契約有無	なし	-0.7847
税金支払い口座	なし	-0.3526	カードローン実行日		-2.07E-06
学費支払い口座	なし	0.2328	その他ローン契約有無	なし	-4.5149
年代	20代	-0.9329	その他ローン実行日		0.000265
	30代	-0.4793	顧客セグメント	超富裕層	8.807
	40代	-0.2737		コア層	-0.5796
	50代	-0.0374		ボテ層	-2.2569
	60代	0.2432		マス層	-2.7712
	70代	0.4732	円貨定期性預金残高		-0.00003
	80代	0.7127	円流動性預金残高		-3.14E-06
普通預金開設日		-0.00003	外貨定期性預金残高		3.61E-06
定期預金口座有無	なし	1.0233	公共債残高		0.000382
定期預金開設日		0.000127	保険残高		0.000013
外貨普通預金口座有無	なし	-0.5849	住宅ローン残高		0.000012
外貨普通預金開設日		0.000015	カードローン残高		-0.00026
外貨定期預金口座有無	なし	-2.3515	その他ローン残高		0.00036
外貨定期預金開設日		-0.00007	給振有無	なし	0.0572

表 9. ステップワイズ法を用いたロジスティック回帰分析結果

パラメータ	値	推定値
定数		6.9719
電話料金引き落とし口座	なし	0.253
学費支払い口座	なし	0.334
年代	10代	-2.2588
	20代	0.0337
	30代	0.1139
	40代	0.0416
	50代	0.3621
	60代	0.6289
	70代	0.593
	80代	0.5934
定期預金口座有無	なし	-1.2718
外貨普通預金口座有無	なし	-0.8395
外貨定期預金口座有無	なし	-1.1224
保険契約有無	なし	-1.7848
公共債契約有無	なし	-1.1214
MMC契約有無	なし	-0.688
ダイレクト契約有無	なし	-0.8761
カードローン契約有無	なし	-0.8687
顧客セグメント	超富裕層	0.4955
	コア層	0.4109
	ポテ層	-1.4143
	一般層	-1.9819
	オーナー	1.4233
円貨定期性預金残高		-0.00004
円貨流動性預金残高		-0.0000059
外貨流動性預金残高		-0.00003
外貨定期性預金残高		-0.000001127
公共債残高		-0.00003
カードローン残高		-0.0006



投資信託の基本的な仕組みについてご説明いたします。

[> HTMLメールが正しく表示されない方はこちら](#)



様々な金融商品の中でも、人生設計に応じた資産運用を行う手段として投資信託への関心が高まっています。投資信託について、基本的な仕組みをご説明いたします。

[詳しくはこちら](#)

今すぐ資産形成を始めたい方はこちら

みずほ積立投信



お客様のマネープランに合わせ、1,000円から始められる投資信託です。毎月決まった日に自動的に投資信託を購入。毎1,000円から積立ができます。

[> 詳しくはこちら](#)

NISA：ニーサ（少額投資非課税制度）



NISAとは、国民の資産形成を後押しするために創設された税制優遇制度で、投資から得た利益が非課税となります。お客様のニーズ・目的に応じて3種類のNISAからお選びいただけます。

[> 詳しくはこちら](#)

iDeCo：イデコ（個人型確定拠出年金）



iDeCoとは公的年金にプラスして給付が受けられる年金制度です。掛金を積み立て、預金や投資信託など自分が選んだ商品で運用した後、原則60歳以降に年金または一時金で受け取ります。

[> 詳しくはこちら](#)

保険



ご自身はもちろん、ご家族のための備えも。お客様のニーズに合わせてお選びいただける、様々な種類の保険商品をご用意しています。

[> 詳しくはこちら](#)

図 8. メール A：商品訴求の個別メールの例

MIZUHO みずほ銀行

自分のお金について知っていますか？身近な収支と将来のお金について解説！

みずほダイレクト
ログイン

> HTMLメールが正しく表示されない方はこちら

【未来想像マネー塾 1 限目】 自分のお金について知ろう！



お金に対して漠然とした不安を感じている人は少なくないと思います。そんなお金について知りたいけれど、「何から始めたら良いのか分からない」という方に未来想像マネー塾ではお金の基本を分かりやすく解説していきます。未来想像マネー塾 1 限目では自分のお金を見つめ直し、将来必要となる資金を想像できるようお手伝いします。

記事はこちらから

図 9. メール B：シナリオメール

空港における AI 活用事例，社会に広がる DX！

AI・データサイエンスのビジネス化と直面するリスク

（日立ソリューションズ・早稲田大学データ科学センターセミナーパネルディスカッションより）

（文責：早稲田大学データ科学センター 小林学）

■ はじめに

早稲田大学と株式会社日立ソリューションズ（以下、日立ソリューションズ）は、データサイエンス分野の人材育成や産学連携促進を目的とした学術交流協定を締結しており、連携の一環として、2022 年 3 月 9 日に「AI・データサイエンスのビジネス最前線と未来社会が直面するリスク」と題したオンラインセミナーを開催致しました。セミナーでは、講演①として日立ソリューションズの熊谷仁志氏から「空港の X 線検査への AI 活用事例」をテーマにご講演いただき、続いて、講演②として、本学社会科学総合学術院横野恵准教授より「（行動データを中心とした）最新のデータ倫理・ガバナンスに関する議論や研究のご紹介」についてお話をいただきました。

講演後に行われたパネルディスカッションでは、セミナータイトルである「AI・データサイエンスのビジネス最前線と未来社会が直面するリスク」について様々な角度からそれぞれの専門性を活かした非常に深い議論が行われました。ここでは、パネルディスカッションの様子をご紹介します。

パネルディスカッションのメンバーは以下の通りです。

モデレータ	早稲田大学 データ科学センター教授	小林 学
パネリスト	株式会社日立ソリューションズ	扇 健一
		米光 一也
		熊谷 仁志
	早稲田大学 社会科学総合学術院准教授	横野 恵

■ AI・データサイエンスのビジネス化におけるステークホルダーの関係性

小林：

パネル討論のテーマ「AI データサイエンスのビジネス化と直面するリスク」では、この後の議論がわかりやすいように簡単な関係図（図 1）を作りました。

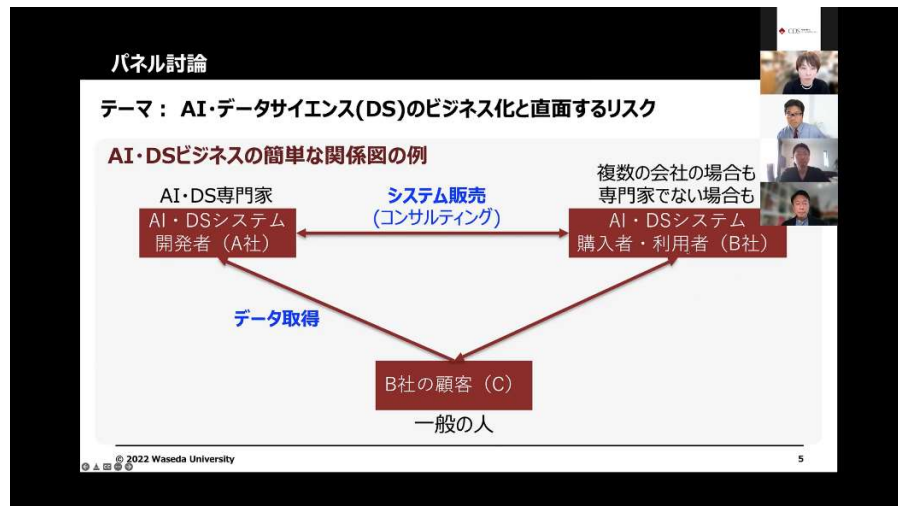


図 1

AI とデータサイエンスのビジネスでは、図 1 で示される関係性が多いように思います。図 1 ではまず、AI やデータサイエンスの専門家集団である AI データサイエンスシステムを開発する人、あるいは会社を A 社とします。また開発された AI データサイエンスシステムを導入し、利用して、企業活動をする会社を B 社としています。さらにこの B 社を利用する顧客を C としています。

図 1 では B 社は 1 社ですが、実際には複数の会社やステークホルダーで B 社が複数の会社からなるような場合もあります。また、B 社は AI やデータサイエンスの専門家でない場合も多く、専門家が作ったものを利用する立場である場合も多いです。

A 社はシステムを販売しコンサルティング業務を請け負い、AI やデータサイエンスによる分析結果を提供することもあります。そのときに一般の人 C のデータを、実はシステムのフィードバックに使ったり、あるいはシステム開発に使ったりすることがあり得る。このような関係性が成り立っているケースが多いと思います。本日はこの関係図を基に、ビジネス化や直面するリスクがこういった形であり得るのかということを考えていきたいと思います。

■ AI・データサイエンスのビジネス化の課題

小林：

先ほど熊谷様のご発表で、AI を使うときの課題として挙げていただいたものも含めて、ビジネス化の課題を図 2 に示します。

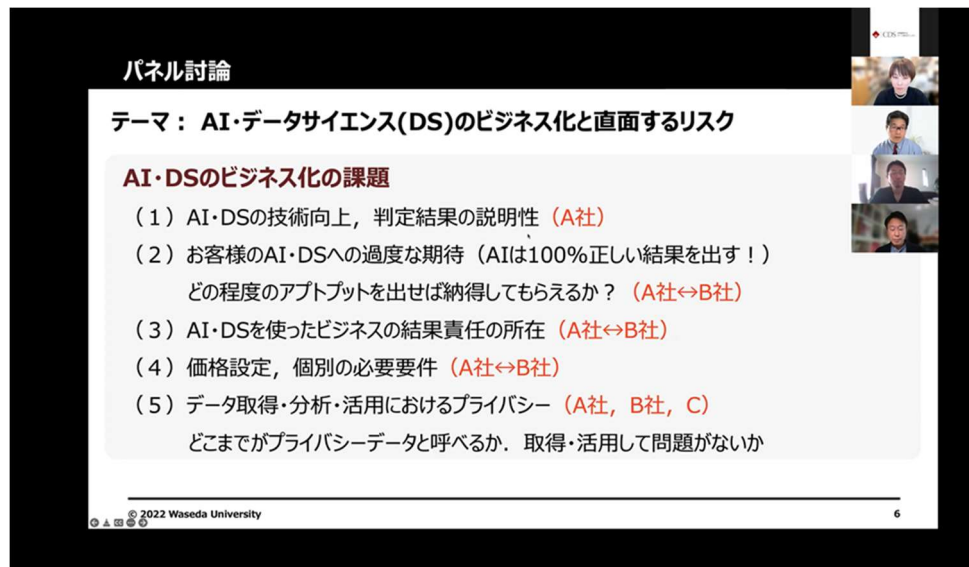


図 2

熊谷：

図 2 の (1) から (5) をそれぞれ説明します。

(1) 「AI・DS の技術向上, 判定結果の説明性 (A 社)」

実業務では, 新商品のように AI が未学習のものに追随していく必要があります, 技術向上が必要です。未知のものに対する判断は人が圧倒的に強いです。AI は学習したものしかわかりませんので, 精度向上・学習し直しが必要なこともあります。

判定結果の説明性については, 例えば, システムのテストをした際になぜ誤認識したのかなど, 主に導入時にお客様から説明を求められますが, 説明が難しいというのが現状の課題です。

(2) 「お客様の AI・DS への過度な期待 (AI は 100%正しい結果を出す!)」

これは我々も課題として捉えています。お客様の中には, 一般的な工業製品と同じ感覚で AI を捉えている方もいらっしゃいます。例えば AI で 90%検知できることをお客様に説明して導入していただこうとしても, 残りの 10%に致命的な見逃しなどがあつたらどうするんだということで議論になります。そうするとどうしても 100%でない駄目だという結論になりがちです。

(3) 「AI・DS を使ったビジネスの結果責任の所在」

現在我々が扱っている「X 線検査判定支援ソフトウェア」の AI に関しては, サービス提供元の B 社に判断の責任を負っていただいています。製品としても判定ソフトウェアではなく, 人の判定を支援するソフトウェアと説明しています。

今後技術が進化して AI が人の代替となっても, 基本的に契約で責任の所在を明確化することが必要になると考えています。現在先行している自動運転の評価の基準などが参考になるのではないかと考えています。

(4)「価格設定、個別の必要要件」

価格設定については、この AI がいくらという値付けにはなっていません。AI のシステムでは、個別の要件に応じたオーダーメイドが多いです。そのため、X 線に限ったニッチな AI 利用においても、お客様によって少しずつ条件が違っているので、学習モデルや個別のカスタマイズを行う必要があります。メーカーとしては全て同じものであれば効率的ですが、そうはなってないのが現状です。

(5)「データ取得、分析活用におけるプライバシー」

X 線の画像がプライバシーを侵すかどうかといった点は、まだ深く議論がなされていません。個人情報ではないと思いますが、どのように折り合いをつけるかというのが課題です。個人情報や、プライバシー、知的財産権などを配慮した上で、データを活用できるかできないか、お客様や関係各社と協議して契約で決定しておく必要があります。現在は、図 1 の A 社・B 社との間で X 線データに関してやりとりしているという状況です。

小林：

具体的なイメージがつかめてきました。ありがとうございます。

次の質問です。図 2 の (2) と (3) の課題について、B 社にはどのようにご説明されるでしょうか。これは、セキュリティにおける課題とも比較的近いのではと思っています。セキュリティに関してもお客様は 100%セキュリティが担保できるのか、セキュリティが破られたときにあるいは何らかの欠落があったときに、責任の所在はどちらかという点を考えると、AI・データサイエンスとセキュリティに関しては、(2)、(3)については関連があるように思っています。

セキュリティの専門家であり、企業の立場として、日立ソリューションズの熊谷さん、扇さん、米光さんにお伺いしたいと思います。まずは熊谷様から紹介いただいた事例の場合にどうしているかご説明をお願いします。

熊谷：

わかりました。(2)「AI の 100%神話」と(3)「責任の所在」について、特に(2)には、明確な答えがない問題ですが、現在、お客様にどのように説明しているかお伝えします。100%を求める方々にいくら数字を語っても響かないので、「人との比較」をご提案します。保安検査では 100%の精度で検査を行っていると思いますが、場合によっては、時間をかけないと見つからないこともあると思います。ですので、同じ条件で AI と人を並べたときに、人を 100%としたときに AI が人と同程度の精度であることを示します。

何度か人と AI の対決をしたことがあります。「正解が 10 個あるうちの何個見つけるか」というテストに対して、AI は人と同程度の結果を出すことができます。ただ、AI と人が明らかに違うのは、AI は数打てば当たる的なところがあって、正解率は高いけれど誤認識も多いという傾向があります。

それに対して、人は見逃すことはあっても誤認識が少ないという傾向があります。AI の正解率と誤認識のバランスを見ながら、人と同じくらいの精度が出るところで納得していただくこと

が多いです。同じくらいの精度であれば、実際の現場で支援できる結果を得られることが多いからです。

(3)「責任の所在」については、「完璧です」といって売れば良いかもしれませんが、そこまで過信はしていません。あくまで目的は支援という位置づけですので、B社さんに責任を持っていただきます。自動運転のレベル2と同様のイメージだと思っていただくと良いと思います。人が基本的には責任を持つといったところで折り合いをつけているのが現状です。

小林：

ありがとうございます。扇様、セキュリティにおいてはいかがでしょうか。

扇：

はい。セキュリティも同じというお話がありましたが、まさにその通りです。セキュリティに対しても昔から100%を求めている人がいました。ですが、最近では各企業においてセキュリティを担当する部署や人がおり、セキュリティは100%ではないことを理解してくださっている人がかなり多くなっているのが実情です。マルウェアやランサムウェアが猛威を振るう中で、セキュリティ対策にAIが活用されています。マルウェアの分析や攻撃の分析においては、データもかなりの量がありますのでAIの活用が欠かせません。

しかしながら、一部では、セキュリティ対策にも100%を求められるお客様もいらっしゃいます。その場合、実際の落としどころとしては98%、99%ぐらいでチューニングしてご提供しています。ただし、お客様の環境によって、そのパーセンテージはどうしても変わってきます。マルウェアはファイルですので、お客様が多く保持しているファイルがどのようなものかによって、合う分析方法が変わってきます。お客様の環境に合わせて検知率を上げるようにチューニングすることを考えます。検知率を上げて限りなく100%に近づけるチューニングというのことができますが、その分の誤検知が多くなります。100%に近づければ検知に要する手間は減りますが、誤検知したものを捌く手間は増えますね。そのため、お客様に検知率、誤検知率のバランスをチューニングしていただき、どのくらいがちょうどいいのかを試してもらう方法をとっています。

AIは人と比べてものすごいスピードで判定します。セキュリティ対策では、マルウェアや攻撃に対して人が判定していたら間に合いません。AIが貢献できる力があると思います。しかしながらAIが苦手な部分もあるため、検知の支援といったAIが得意なところをAIに任せると良いと思います。

責任の所在については、プログラムの不具合の責任は取りますが、AIの能力に関する責任は取れず、あくまで支援をする形でしかないと思っています。

お客様によってはマルウェアと判断した根拠を求めたり、誤検知をした理由の説明をもとめられたりする場合がありますが、AIがどのような方法でデータを収集して分析し、判断しているかといった仕組みをお伝えしています。AIの支援する力を認めてほしいですね。責任の所在は、結局は人になると思います。

小林：

ありがとうございます。よくわかりました。それでは米光様、お願いします。

米光：

はい。扇さんの言う通り、セキュリティも過去 15 年前ぐらいは AI と同じような課題がありましたが、この 15 年を経て、社会的な総意が取れてきていると思います。これは、「社会はミスに対してどこまで許容するのか？」という空気みたいなものが、かなり関係していると思っています。

例えば、今の AI の問題も「AI だから仕方がない」というのが、どこまで許されるのか、そういう社会かどうかというのは、(2)と(3)の点では重要なので、セキュリティと同様に、この先 15 年ぐらいで、どこかに収斂していくと個人的には思っています。

あとは、対象のシステムですね。「皆さんが受けたテスト結果を AI が判定して、本来よりも 5 点低い採点をされてしまいました」といった問題を、学生の皆さんは、どう受け止めるのか、という話です。手術などの健康に関わる場所では、ミスは 0 にしたいでしょう。AI はどの分野なら適用できるのか、どこまでは許されるかという話だと思っています。ただ、AI がセキュリティと大きく違うのは、15 年前のセキュリティの問題の時は、代替手段がありませんでした。例えばコンピューターにマルウェアが感染するという脅威に対して、対策が 100%であろうがなかろうが、感染するくらいなら 80%の対策を入れなくてはいけないという状況でした。

ところが、AI で解決できる問題は、別の解決手段もいろいろと残されていることが多く、そういう意味で日本のような慎重な判断をする文化の国は、なかなか AI が広がりにくく、逆に海外だと広がりやすいギャップをすごく感じますし、このようなカントリーリスクがあるな、ということは聞いていて思いました。

小林：

ありがとうございます。非常に明確になりました。

セキュリティに関しては代替手段もそうですが、ないとどうにもならないという状態が先にあり、100%でなくても受け入れざる得ない状態が起こって、それを受け入れる上でどこまで精度を上げるべきかという議論になっています。今の AI に関してはまだそこまでできていないので、これから議論が起こると思いますが、セキュリティの例を考えると、必要性和重要性が明らかになっていく中でおそらく社会がそれを受け入れる形になっていくのではないのでしょうか。

最後の米光様のお話で、分野ごとの許容範囲は起こってくると思いました。ただ便利なもの、要は音声認識といった、少し言葉を間違えたのは許せるけれども、手術などで間違いは許されないでしょう。そのトレードオフをどうするのか議論は残ると思いますが、必要性和重要性の議論は明確になったかなと思いました。ありがとうございました。

価格設定については個別の案件になってくると思いますので、もう 1 つの議題に移りたいと思います。

■ 行動データのプライバシー保護

小林：

本テーマは主題でもあります，プライバシーはセキュリティにも大きく関係すると思います。プライバシーデータの特に行動のデータについて考えていきたいと思います。

行動のデータや顔画像のデータなど，個人の情報が含まれるようなデータの取得・活用に関して議論をしたいです。活用すべきだという議論もあると思いますし，顔画像や個人情報を取られるのが嫌だという人ももちろんいると思います。先ほどの X 線検査の話をお聞きして，私は過去の鉄道の事件を少し調べてみました。この参考の URL（『日本の鉄道に関する事件 - Wikipedia』※1）を見ると，日本の鉄道に関する事件が一覧として出てきます。日本は世界の中では安全な国だとは思いますが，とはいえ，実は鉄道に関する事件がかなり頻繁に起こっています。

ここで大きなものを 4 つぐらいご紹介します。新しい事件ですと，京王線の刺傷事件があります。これは 2021 年 10 月 31 日に起こりました。この事件は 24 歳の男性が刃物で他の乗客を切りつけて，液体をまいて放火し，18 人が重軽傷を負ったという事件です。事件を受けて新規の全車両において防犯カメラの設置を義務付ける方針を固めたことが書かれている点です。防犯カメラを設置するだけでこういった事件を未然に防ぐことができるかという点，すこし疑問に思うこともあります。

さらにその前の 2021 年の 8 月には，小田急電鉄の小田原線の車内で無差別の刺傷事件が起こっています。乗客の 20 才の女子大生が重傷を負ったほか 10 人がけがをした事件です。2018 年には，新幹線の中で殺傷事件が起こっています。のぞみ 256 号においてお客さん 3 人に対して，なたで切りつけて 1 人を殺害したという事件や，あるいは 2015 年にはのぞみ 225 号で，自分でガソリンをかぶって火をつけて，火災が発生した事件がありました。

この 4 つの他にも先ほどの URL での一覧を見ると結構載っています。こういうことを考えていくと，安全保障上，先ほどの防犯カメラをただつけるという抑止力で済むのかという話になります。こうした鉄道に関する事件の対策を考えると，鉄道に限らず，例えばイベント会場や遊園地なども，ある意味同様に考えることができるのではないかと思います。先ほどの AI 化による手荷物検査の実施についても当然考えることができると思います。実際に鉄道を考えたときに，中国では手荷物検査，中国の駅の入り口，特に北京の地下鉄では駅で X 線の手荷物検査を行っています。これを AI 化するということも考えられます。あるいは動作認識技術・行動追跡技術などを考えて，危険人物の抽出や検知などを何らかの技術で行うことも可能になっていくだろうということで，技術的な可能性はある訳です。そうすると AI・DS によって低コストで安全性を確保できる可能性があると思います。これはほんの一例です。

議論を進めるにあたり，まずここで皆さんにアンケートにご協力いただきたいと思います。参加者の方には OK・NG をアンケートにお答えいただければと思います。

1 つ目が，AI 手荷物検査データの取得は，（手間も含めて）皆さん OK ですか，NG ですかです。次に，民間企業による顔画像や行動履歴等における個人データの取得利用は，この公共の安全のためであれば OK ですか。プライバシーということなので，顔画像や行動履歴等の個人データが OK か NG か。この 2 つをお答えいただければと思います。

【アンケート 1】

民間企業による AI 手荷物検査データ取得は「OK」 or 「NG」？

【アンケート 1 結果】

OK : 86. 4%

NG : 13. 6%

小林：

86%が OK. 結構多いですね.

では、次に、顔画像など行動履歴等の個人データの取得利用などは OK か NG かについてアンケートお答えください.

【アンケート 2】

民間企業による顔画像や行動履歴等の個人データ取得・利用は「公共の安全のためなら OK」 or 「NG」

【アンケート 2 結果】

OK : 41. 9%

NG : 58. 1%

小林：

こちらは NG の方が結構増えてきましたね.

手荷物検査が OK の方が 80%以上いて驚いているのですが、2 番目の質問の個人情報の取得の方は NG が増えてきている、そういう状況ですね. ありがとうございます. 少し、さきほどの私の言い方がきつくなってしまったのでしょうか. 事件の話を前面に出しすぎて、ミスリードしてしまったところがあるかもしれませんが、個人情報を取得するというのが、今、事件の話をしたので安全面を重視して考えたということもあろうかと思いますけれども、個人情報のデータの取得に関しての半分ぐらいは OK という話が出てきていました. この結果をもとに、パネラーの方に、ご意見を伺ってまいります.

扇：

手荷物検査、私は OK にしました. 安全確保のためにする方がいいと思います. 普通に考えて、空港、遊園地を含めいろいろな重要な施設に入るときには、検査されるのは分かっているのではないのでしょうか. コンサート会場などもそうです. ですから、軽装でいったりするのではないですか. 検査されたから何か感じるかという、私は当然だと思って受けています.

小林：

行動追跡に関してはどうですか。

扇：

行動追跡に関しては、私はNG、嫌です。これまで、マイナンバーなども漏れる漏れないという議論がありましたが、発行された瞬間から漏れました。届け先が間違ったなど。あり得ないですよ。民間企業に任せた場合にそういったことも起こり得ますので信用していません。クレジットカードなども、安全のためにセキュリティコードがカードの裏に書かれていますね、それを使えば安全になると思われていたものが、最近だと、クレジットカード番号もセキュリティコードも両方漏れていますから、全く意味がないです。というのは、民間企業では、予算や経営層の考え方で、セキュリティに関する考え・対応が変わってくるためです、いいづらいことですが、ここでは厳しく言わせていただきます。顔をハッシュ値化、非可逆性のデータに変えて保存してもらう。何かがあったときには本人の写真など画像から同じハッシュ値というのを出してマッチングさせて調べる、というぐらいが限界ではないでしょうか。

小林：

わかりました。個人情報とはそれだけだと特定が難しいデータにしておくことですね。ありがとうございます。

■ 行動データ等の取得・利用事例

小林：

ここで横野先生のご専門でいらっしゃるの、国内外の顔データや行動データ等の取得や利用等において問題になる、あるいはなったようなケースがありましたら教えていただけますでしょうか。

横野：

はい、画面共有させていただいてもよろしいでしょうか。

私達がやっているプロジェクトでそういう情報を集めていますので、ご紹介させていただきます。このWebサイト（『DATA ETHICS プロジェクト データ倫理最新情報』※2）ですが、ここにいろいろな個別トピックに関する情報を集めていて、位置情報、あるいは顔認識の技術を使った、あるいはHRTechと人事に関してこういったものを使うということについてです。

この顔認証、顔認識については、いろいろございます。基本的には直近で一番日本の中で議論されたものは、ここまでに議論に出ていたものにも関係しますが、顔認証ではなく顔認識だと思いのですけれど、JRが顔認識の機能を付けた防犯カメラを設置していたということで、これがこういう報道が出た後に、短期間のうちに中傷されたということなのですが、基本的には海外では規制の話がかなりされていて、日本では顔認識について、どういうルールのもとで使っていくのかということについて、まだ十分な議論をしていないという認識でいます。

参考になるものとしてこの中でも紹介していますが、この Web サイト（『あまりにも未熟な日本の「顔」画像利用の議論』（※3））辺りを見ていただくといいのではないかと思います。

それと先ほどのアンケートの件について、少しコメントさせていただいてもよろしいでしょうか。アンケートで、民間企業による AI 手荷物検査で、データ取得は OK・NG かという設問になっていて、これが何を意味するのかによって大分違ってくると思います。そのときに荷物検査をして、それが付随するデータとともにストックされていくのか、それに対してまたいろいろな解析を行って、このアンケート 2 にあるようなハイリスクかどうかということを判断するための何らかの解析をするのかどうかによって全く違う話になると思います。

この 2 つの話、アンケート 2 の話もそうなのですが、これはその技術としてどうかという話よりは、セキュリティに関してどういうポリシーを取るかという話が本質だと思います。

スライドを共有させていただきたいのですが、空港の話もそうですが、セキュリティに対する考え方として、ルールベースでも、最初の話の中にもあったルールベースは、データサイエンスというルールベースとは違って、セキュリティに対してルールベースで考えるかリスクベースで考えるかという話がございます。

ルールベースは、手荷物検査ですと、こういうイメージがあると思うのですが、全例に対して一律の検査や審査を行う、その中で特に問題がある慎重な対応を必要と判断されるものが検知される場合もあるが、そういう場合を除きリスクは同程度にみんな持っているというような前提から入る、それがデフォルトのあるルールベースです。

これに対してリスクベースは、個別のリスク評価がデフォルトであり、その評価に応じて異なる厳しさの検査を行うのはリスクベースということになります。

日本においてよくあることですが、技術を使ってどうしたいかということは、本当はこの大本のポリシーをどう考えるのかという問題である場合がとても多いと思います。もちろんこのルールベースとリスクベースは、両者排他的なものではなくて、ルールベースの場合にも、最初のお話あったように、経験の蓄積などそういうものによってリスクの判定が一定程度行われている、ただそれはインフォーマルな形式だし、全員に対して同じように事前にリスク評価を行うという前提ではないというところが違います。

その技術を用いるということの裏には、政治的ないし社会的な方針の原理原則の変更が実はあると思うのですが、そこに目を向けられずに、そこでどういう方針を取るのかということが曖昧にされたままに技術を導入するということにのみ着目をされてしまって、本来この部分でどういう考え方をとるかによって、その技術をどう使っていくかということも違ってくると思います。

先ほどの 100% を求める方もいらっしゃるという話がありましたが、100% が達成できたとしたら、完全にオートメイテッド・ディシジョン・メイキングしたいというご要望があるかどうか。また、リスクベースのアプローチを用いた場合には、全員に基本あるいはその幅広く個別のリスク評価をするということが発生してきて、そこでそのデータテクノロジーを用いたリスク評価の方法を導入するかどうかは、基本的にはそういう整理になると思います。そこでプロファイリングなど、あるいはその自動化された決定というように GDPR と呼ばれるようなものが方法としては選択肢となり得ますが、これは無前提に導入していいものではなく、GDPR 等ではこ

れは具体的に規制をしています。

もちろん全く使ってはいけないということではありませんが、それを使う場合には一定の条件である、あるいは場合によっては代替措置を設けて、そういったものを選択することができるようにするということが条件になっています。そこで目指しているものが何なのかということです。それが政治的、社会的な意味での方針変更に関わるものかということと、そこがクリアになった上で、その技術そのものの問題としてプロファイリングなど自動化された決定というものにそれが当たるのであれば、正当化できる範囲での利用なのかどうかということを考える必要があるのではないかと思います。

話が飛んでしまいましたが、関連することをまとめてコメントさせていただきました。

小林：

ありがとうございます。行動履歴データを含めたデータのプライバシー保護、並びにそのデータ利用に関して、非常に明確になったと思います。非常に有益ですね。こういったことを知らないでシステムなりビジネス化するときに、会社側としては、今度非常に大きなリスクになってしまうということも含めて考えないといけないということでしょうか。非常に勉強になりました。

行動履歴というところで、先ほど扇様の方でもハッシュ化するように、個人が特定できない形は必須じゃないかという点も議論としていただきましたので、誰かを特定してどうこうというよりは、やはり、どういう行動がリスクに繋がるのかということを、特定の誰かというよりは目的として、安全保障上だったら安全保障上に対してどういう形でそれを使えばいいのかという、その根本に戻ってちゃんと考えていかなければならないということですね。

■ 行動データ等の利活用に対する課題

小林：

最後に、今後おそらくスマートフォンや防犯監視カメラや、電子決済、IoT 機器等々において、データはどんどん取得されていき、さまざまなビジネスに活用され始めているのは事実だと思います。特に行動の履歴のデータのようなものに関して、皆さんのお考えをお聞きしたいと思います。

熊谷：

私個人の意見では、データが使われることは、いたしかたがなく、あとは程度の問題であろうと考えています。例えば、情報提供に見合う価値のあるサービスであればいいと思います。今でも携帯電話の位置情報など取られていると思いますが、そういうものはメタデータとして使われていて、例えば私が街中で倒れてこの人が誰だかわからないとなったときに、これは熊谷だとわかるような認証技術はありがたいと思います。ただ先ほどのお話にあったように、行動データを勝手に取られていて、誰か知らない人がそれを見てというのは気持ちが悪いです。そういうのは嫌だと思っています。

扇：

そうですね。私は、将来的には個人で全部管理できるようにしてほしいという思いがあります。情報を取ったとしても、どこに情報を与えるか、削除などのコントロールは全部個人でできるようにしてほしいと思います。そうすると今、総務省などで考えられている情報銀行という仕組みでしょうか、とても強固なシステムの中に個人情報を入れておいて、そこから、各企業がデータを取得できるかできないかを個人がコントロールするという形にすれば、許せるといったところでしょうか。

米光：

私はプライバシーよりも、データの利活用を優先したいというスタンスですが、先ほど横野さんのおっしゃった、問題のどこにフォーカスするかによって確かにアンケートの結果の捉え方は結構変わります。アンケートを取った時の話の流れだと、電車の中に監視カメラをつけて犯罪を防ぐとか、遊園地の中でも同様に犯罪を防ぐというような使われ方の話の流れの中での質問だったので、先ほどのOKと回答した42%という学生さんの割合は、結構高い数字だと思いました。どちらかと言うと、私のスタンスはマイノリティーだと思っていたので。

あと横野さんのお話の中の「100%を求める」の話は、ほぼ責任の問題（1%であってもミスが起こった時に誰がミスの責任を取るのか、という問題）のことを気にされての発言ですね。

横野：

そうですね。私としてはいろいろデータを活用して便利になっていく、イノベーションが起っていくということは、基本的には歓迎をしています。ただ、先ほども触れたのですが、それは単なる技術的な問題ではなく、基本的な価値など社会の中でのルールそのものを変えないといけないような意味を伴うものである場合もあると考えています。

顔認識の問題などはそのような側面があると思うので、基本的にはルールをきちんと作りながらやっていくことが重要で、ルールが明確化されていない中でどんどん使っていくのは、すごく不安があると思っています。

また、ルールが作れない中でどんどん先に進んでいくと、日本では起こりづらいと思うのですが、ある程度枠組みを作っておかないと実際の方が先行したときにそれがコントロールできなくなってしまうということは常に思っています。ですので、ルールを作るということも含めながら、そういったものの開発や実装を行っていくということはやはり重要だと思っています。

小林：

ありがとうございます。技術の進み具合と、データ倫理、あるいはその法整備や、ルール作りが同時進行で進んでいる訳ではないというところに大きな問題があるということですね。こういった議論そのものがおそらく重要で、こういったところからルール作りの必要性や、あるいはどういう形であれば使っていいのかというコンセンサスを社会的に醸成する必要性があるということでしょうか。

今回の議論はすぐ答えが出るものではないですが、ビジネス化する時にはこれらも含めて全体を考えて答えを導くようにしていかなるを得ないということが非常に参考になりました。

皆さんどうもありがとうございました。以上でパネル討論は終了します。参加者の皆様にもお礼申し上げます。どうもありがとうございました。

[参考文献]

※1 日本の鉄道に関する事件

出典：フリー百科事典『ウィキペディア (Wikipedia)』 2022年3月9日参照

URL :

[https://ja.wikipedia.](https://ja.wikipedia.org/wiki/%E6%97%A5%E6%9C%AC%E3%81%AE%E9%89%84%E9%81%93%E3%81%AB%E9%96%A2%E3%81%99%E3%82%8B%E4%BA%8B%E4%BB%B6)

[org/wiki/%E6%97%A5%E6%9C%AC%E3%81%AE%E9%89%84%E9%81%93%E3%81%AB%E9%96%A2%E3%81%99%E3%82%8B%E4%BA%8B%E4%BB%B6](https://ja.wikipedia.org/wiki/%E6%97%A5%E6%9C%AC%E3%81%AE%E9%89%84%E9%81%93%E3%81%AB%E9%96%A2%E3%81%99%E3%82%8B%E4%BA%8B%E4%BB%B6)

※2 DATA ETHICS プロジェクト データ倫理最新情報 2022年3月9日参照

URL :

https://dataethics.jp/resource/resource_category/06/

※3 あまりにも未熟な日本の「顔」画像利用の議論 2022年3月9日参照

URL :

https://note.com/nfi_japan/n/n171811f65949

早稲田大学データ科学センター機関紙「データ科学」第1号

発行日	2022 年 7 月 7 日
編集	須子 統太
発行者	松嶋 敏泰
発行所	早稲田大学データ科学センター 〒169-8050 東京都新宿区西早稲田 1 丁目 6-1 1 号館 3 階
