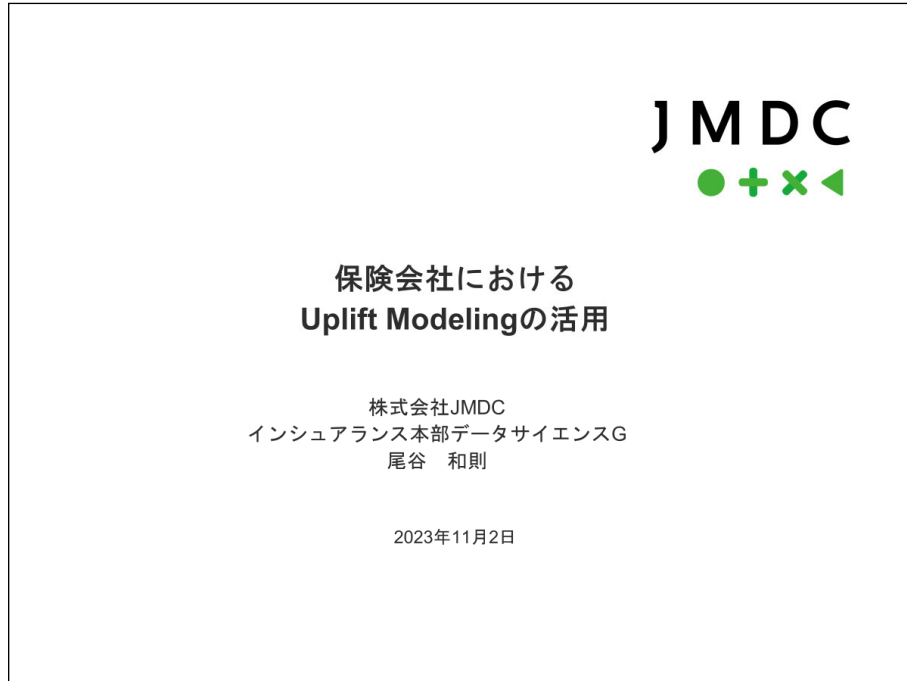


# 保険会社における Uplift Modeling の活用

JMDC 尾谷 和則君

金澤 定刻になりましたので、セッションC-2「保険会社における Uplift Modeling の活用」を開始します。 それでは、尾谷さん、どうぞよろしくお願いいたします。



尾谷 株式会社 JMDC インシュアランス本部データサイエンスグループの尾谷と申します。本日は「保険会社における Uplift Modeling の活用」という題目で、発表させていただきます。よろしくお願いいたします。

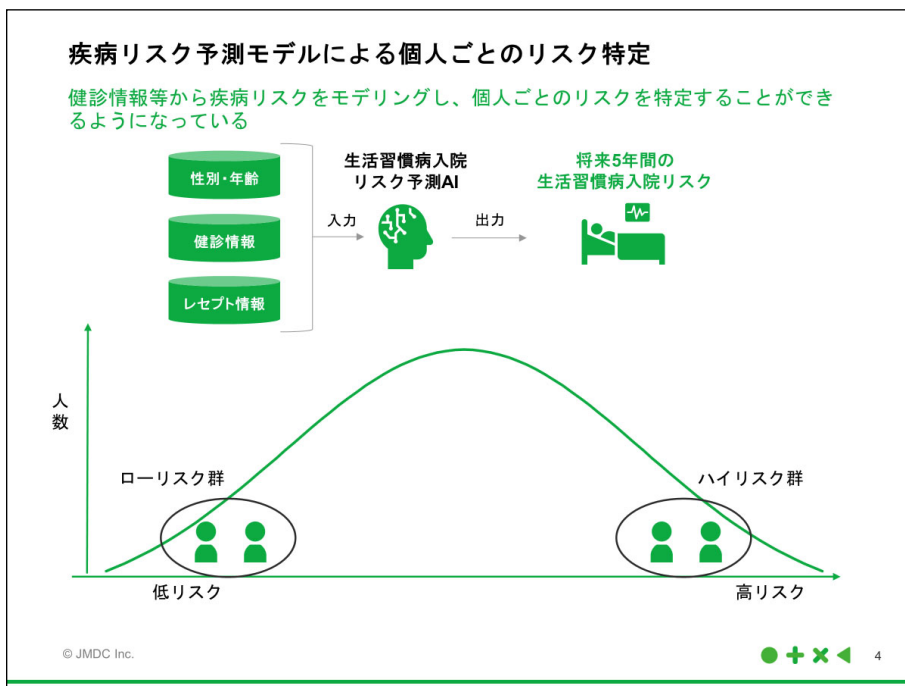
## 目次

1. 因果推論
2. Uplift Modeling
  1. Uplift Modelingとは
  2. Uplift Modelingの分析例
3. 保険会社における活用例

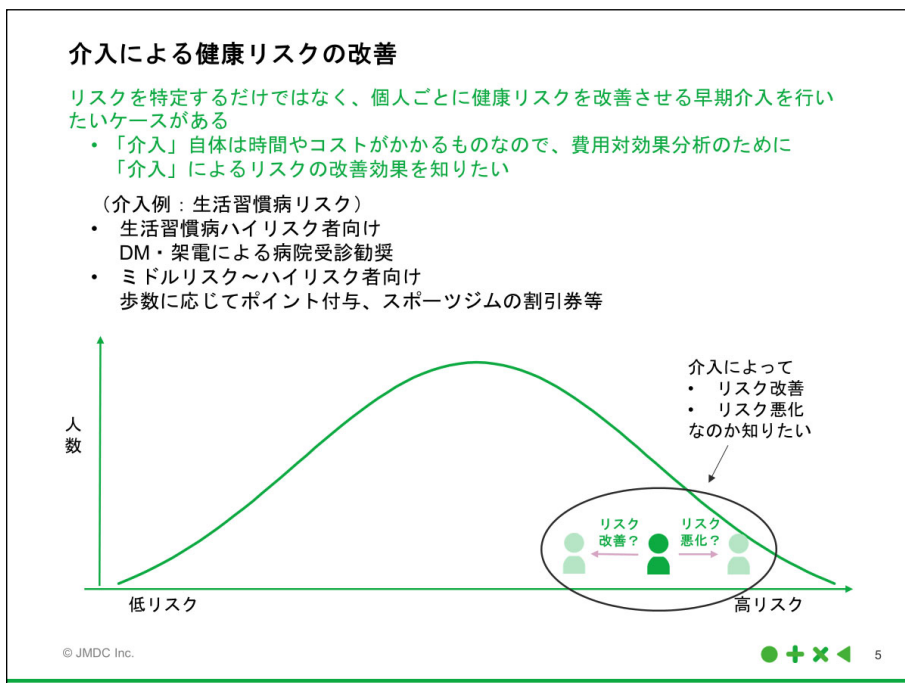
本日の発表の流れになります。まず Uplift Modeling の話をする前に、その前提知識となる因果推論について説明します。次に、因果推論に機械学習の概念を組み合わせた Uplift Modeling の原理について説明し、当社データを使った分析例をお話しします。最後に、Uplift Modeling を保険会社においてどのように活用できるかを、紹介させていただきます。

## 因果推論

それでは、まず因果推論の話からさせていただきます。



まず、どのようなときに因果推論が必要になるか説明します。昨今、保険会社でも健診情報等から疾病リスクをモデリングする技術が広まってきていると思っています。当社でも、性別・年齢、健診情報、レセプト情報を入力して将来の生活習慣病の入院リスクを予測するAIを構築しています。このようなAIによって、より高精度にハイリスク群やローリスク群を特定できるようになってきている状況です。



ただ、このようにリスクを特定して終わりではなく、このリスクを下げるということが最終的なゴールになると思います。そのために何らかの早期に介入を行います。例えば生活習慣病リスクであれば、ハイリスク者に対してDMや電話で「病院へ行ってくださいね」と受診勧奨という方法で介入を行ったり、ミド


リスク者に対しては、健康アプリでたくさん歩いたらポイントを付与する、スポーツジムの割引券を配る、といった介入によって健康改善させようという試みを行います。ただ、こういった介入には時間やコストがかかるため、介入によってどのくらい改善効果があるのかを事前に知る必要があります。この介入効果を分析するために、因果推論という考え方が必要になります。

**因果推論の概念**

施策や介入の効果量を測定するためには「因果推論」の考えが必要

- 「因果推論」とは、データに基づいて「因子」と「結果」の関係性・効果量を統計的に分析すること

主に医療やマーケティングの分野で使用されることが多い

<p>医療の例</p> 	<p>マーケティングの例</p> 
<p>薬を飲んだら病気が治る？</p>	<p>広告を出したら購入者がどのくらい増える？</p>

© JMDC Inc. ● + × ◀ 6

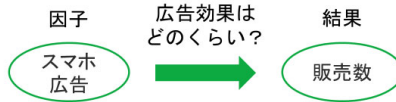
因果推論とは、データに基づいて因子と結果の関係性や効果量を統計的に分析する手法です。主に医療やマーケティングの分野で使用されることが多く、例えば医療分野では、薬を飲んだら病気が治るのかという因果関係を証明するために使用したり、マーケティング分野では、広告を出したら購入者がどのくらい増えるのかという因果の効果量を測定するために、因果推論を使用します。

## 因果推論の原理－必要なデータ－

データから統計的に介入効果を測定するためには以下のデータが必要となりそう

- ・ 介入フラグ (因子のデータ)
- ・ 介入・非介入者の結果 (結果のデータ)

例) スマホ広告を出したら販売数はどのくらい増えるか?



必要となりそうなデータ

id	スマホ広告の有無	購入有無
A	1	1
B	0	0
C	0	0
D	1	0

スマホ広告を見た群と  
スマホ広告を見ていない群で、  
平均購入数を比較すればよい?

© JMDC Inc.

● + × ◀ 7

因果推論の概念を簡単にご紹介しました。次に原理について説明いたします。因果推論は、データから統計的に介入効果を測定する手法ですが、どのようなデータが必要になるのかをお話しします。

例として、スマホ広告を出したら販売数はどのくらい増えるのかという因果推論を考えます。最低限必要となりそうなデータとして、この表のように、スマホ広告を見たか、見ていないかというフラグと、そのあと購入したか、していないかというデータが必要になります。

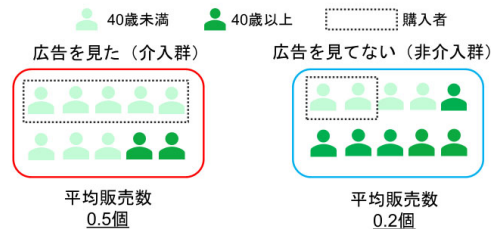
このようなデータに対して、最初に思いつく分析方法の一つとして、スマホ広告を見た群とスマホ広告を見ていない群で単純に平均購入数を比較する方法があるかと思います。

## 因果推論の原理－誤った分析例－

「介入群と非介入群の単純平均を比較」は誤っているケースがある

やっぴまいがちな誤った分析例

単純平均すると、商品のスマホ広告を見た人の群が平均販売数が2.5倍多いので、スマホ広告は販売数2.5倍の効果がある!



- ✓ スマホを使う人は若い人が多く、介入群は若い人が多い可能性がある
- ✓ 若い人は購入意欲が高いため、若い人が多いだけで介入群の販売数が多い可能性がある
- ✓ 年齢以外にも、性別や地域なども偏っている可能性あり

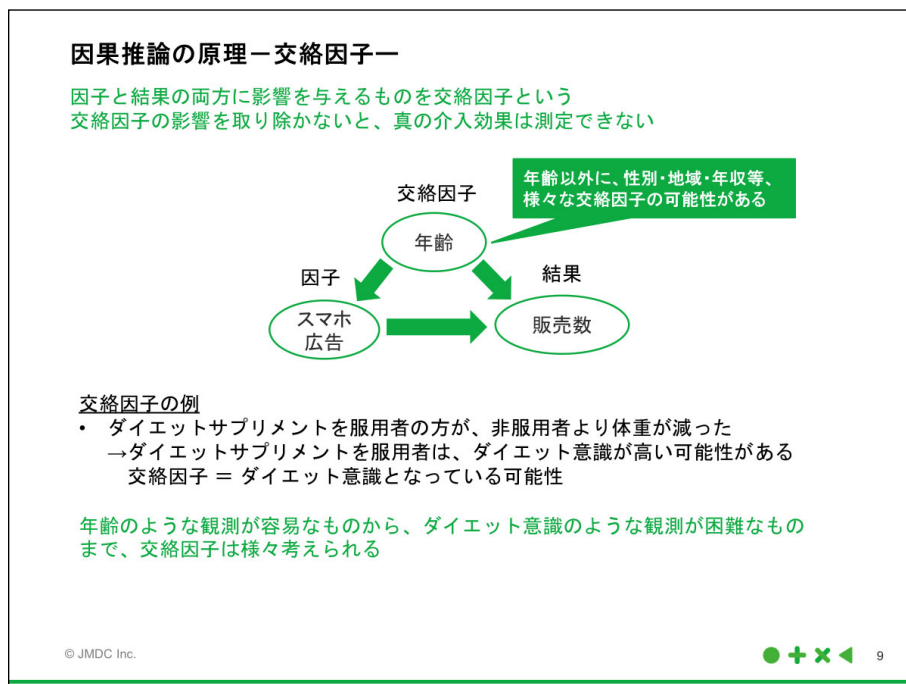
属性分布の偏りによる販売増減の影響を取り除かないと、真の広告効果は測定できない

© JMDC Inc.

● + × ◀ 8

しかし、介入群（スマホ広告を見た群）と非介入群（広告を見ていない群）で単純平均を比較することは誤っているケースがあります。具体的な例を用意しましたのでそちらを使用して説明いたします。以下の例では、単純平均したところ介入群の平均販売数が 0.5 個で、非介入群の平均販売数が 0.2 個であるため、スマホ広告は販売数に 2.5 倍の効果があると結論づけていますが、これはやっと思い間違いの誤った分析例です。

何が誤っているかといえば、スマホを使う人は若い人が多いので、介入群には若い人が多い可能性を考慮できていない点が問題です。実際に今回の例では、広告を見た群の 8 割は 40 歳未満である一方で、広告を見ていない群の 40 歳未満は 4 割です。また、若い人は購買意欲が高いと仮定すると、若い人が多いことが介入群の販売数が多い原因になっている可能性があるということです。



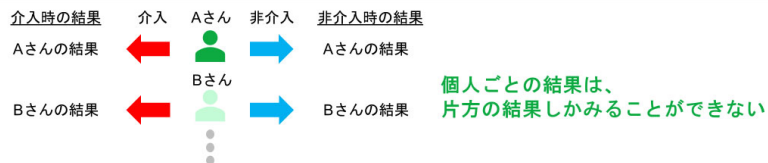
今回、年齢を例に出しましたが、年齢以外にも性別や地域なども偏っている可能性があり、このような属性分布の偏りによって販売量は増減するので、属性分布の偏りの影響を取り除かなければ真の広告効果を推定することはできません。因果推論では、この年齢のような要素を「交絡因子」と呼びます。定義上は因子と結果の両方に影響を与えるものを指します。先ほどの例では、年齢が若ければ若いほどスマホ広告を見る可能性が高くなるため、年齢は因子に影響を与え、年齢が若ければ若いほど購入意欲が高いため結果の販売数にも影響を与えており、交絡因子になっています。

この交絡因子の影響を取り除かないと真の介入効果は測定できません。他の交絡因子の例として、ダイエットサプリメントを服用している人の方が非服用者より体重が減ったという広告を目にすることがあると思います。この分析が正しく行われているか行われていないかはわかりませんが、疑わなければいけないのは、ダイエットサプリメントを服用するような人はそもそもダイエット意識が高い可能性があるのではないかということです。要は、ダイエット意識という交絡因子の存在を疑う必要があります。

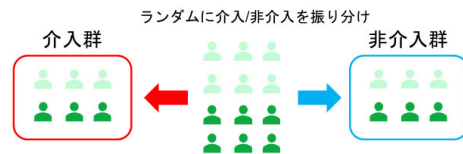
## 因果推論の原理－反実仮想とRCT－

個人ごとに介入/非介入の場合の反応を測定した結果を収集したいが不可能  
RCT（ランダム化比較試験）を行うことが理想的

### 反実仮想



### RCT（ランダム化比較試験）



- ✓ ランダムに振り分けることで、偏りのない同属性の介入・非介入のデータを収集可能
- ✓ 未観測の交絡因子の影響も取り除くことができるのがメリット

© JMDC Inc.

● + × ◀ ▶ 10

この交絡因子を取り除いた分析方法にはどのようなものがあるでしょうか。介入群と非介入群で属性の分布が異なることにより単純平均結果がゆがんでしまうので、データを準備する時点で、属性に偏りのないデータを揃えるという解決策が考えられます。Aさん、Bさんなど、すべての人に介入した結果と、介入しなかった結果を全部集めることができれば、介入群と非介入群は同じ集団となり属性分布が必ず一致するので、交絡因子の影響を考慮する必要がありません。これを「反実仮想」といいますが、このようなことは不可能です。Aさんに介入した結果と、介入しなかった結果の両方を見ることはできません。

反実仮想の代替案として一般的に使われているものが、RCT（ランダム化比較試験）です。RCTでは、被験者をランダムに介入群と非介入群に振り分けます。ランダムに振り分けることによって、介入群と非介入群で属性に偏りのない集団を構成することが可能となります。この方法の優れた点として、未観測の交絡因子の影響も取り除くことができるという点が挙げられます。

## 因果推論の原理—RCTができないケース—

RCTによるデータ取得が理想的だが、現実ではRCTができるとは限らない

### 経済的問題

実施自体に時間やコストが大きくなるのに、ビジネス上価値のある群に介入しない判断ができない  
例) 購入意欲の高い群に広告を提示しないと、機会損失してしまう

### 倫理的問題

倫理的に介入/非介入を設定することができない  
例) 喫煙の健康への影響を検証したいが、喫煙する介入群を設定できない

RCTができないときは、ランダムに介入割り当てを行わずに得られた観察データを使用して分析するしかない

© JMDC Inc.

11

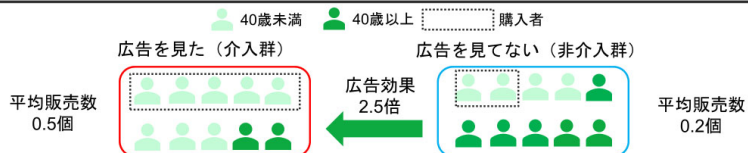
基本的には、RCTによるデータ取得が理想的と考えられていますが、現実には、RCTを実施できないケースが往々にしてあります。理由の一つに経済的問題があります。介入には時間や費用が大きくなるため、ビジネス上価値のある群に介入しないという決定を下しづらいというものです。理由の二つ目は倫理的問題です。例えば、健康に与える喫煙の影響を検証したいときに、非喫煙者に喫煙させることは許されません。

このようなRCTができない場合は、ランダムに介入/非介入の割当を行わずに得られた観察データを使用して効果検証することになります。観察データを使用する場合に、交絡因子を一定取り除くことができる手法が幾つか提案されており、本日は二つ紹介します。

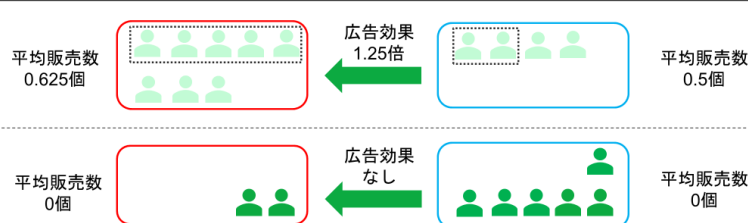
## 観察データを使用する場合—層別解析—

RCT以外の過去の観察データを使用した因果推論手法は存在  
層別解析では、交絡因子で群団を層化することで、交絡因子の影響を取り除く

### 全体で単純平均



### 年齢で層別解析



年齢で層別解析することで、年齢による販売増の影響を取り除くことができる

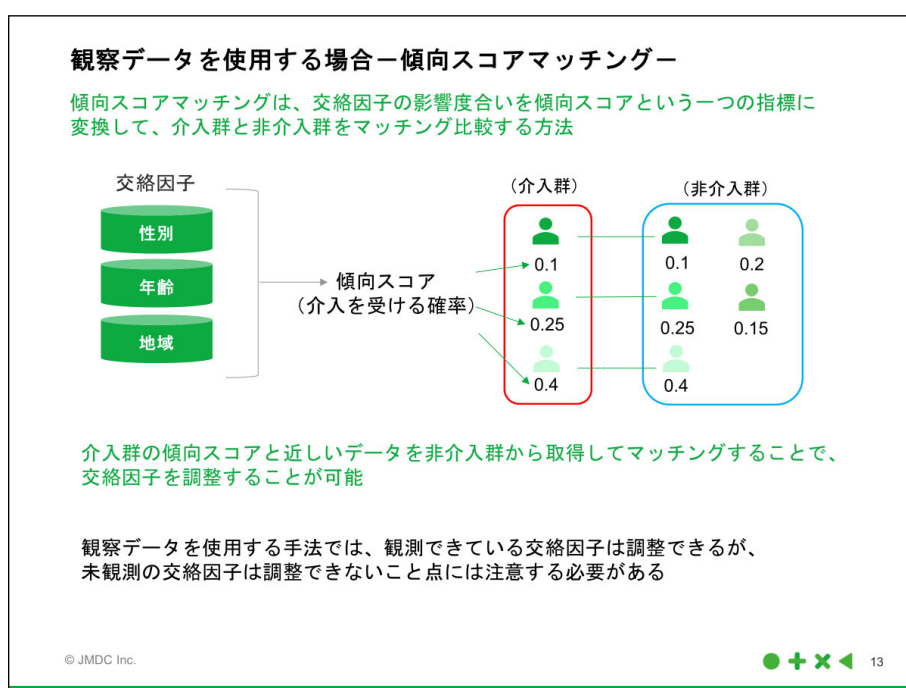
© JMDC Inc.

12



一つ目は「層別解析」です。交絡因子自体をロックして比較すれば、交絡因子の影響を取り除けるという考え方です。先ほどの、スマホ広告と販売数の例では、全体で単純平均したら2.5倍の広告効果でした。この例の分析対象者を、40歳未満と40歳以上に分けて、介入群と非介入群の平均販売数を比較します。そうすると、40歳未満の層では、広告を見た群では8人中5人が購入しており、平均販売数が0.625個となります。広告を見ていない群では4人中2人が購入しているため、平均販売数が0.5個です。40歳未満の層におけるスマホ広告の広告効果は、実は1.25倍であったことがわかります。一方、40歳以上の層では、スマホ広告を見た人も見ていない人も全く買っておらず、広告効果はなかったことがわかります。

このように、交絡因子で層別解析することで、交絡因子による販売増減の影響を取り除くことができ、真の広告効果により近づいた結果を得ることができます。



二つ目は「傾向スコアマッチング」という手法です。層別解析で、一つの交絡因子のみを使用して（例えば年齢）層別に解析したとしても、各層のサンプル数が大きく減ることはありませんが、実際には交絡因子としてさまざまな要素が考えられます（性別・年齢・地域・年収等）。これらすべてを使用して層別に分けていくと、一つの群団のサンプル数が小さくなり結果が不安定になってしまう問題が生じます。

傾向スコアマッチングでは、交絡因子の影響度合いを傾向スコア（介入を受ける確率）という一つの指標に変換して、その一次元の指標で介入群と非介入群をマッチさせて比較します。例えばこのスライドにあるように、介入群の各人の傾向スコアが0.1、0.25、0.4だったときに、非介入群から同水準の傾向スコアを持っている人を1対1でマッチングさせて比較するというを行います。傾向スコアマッチングを使うことで、交絡因子を一定調整することが可能です。

ただし、観察データを使用する手法では、観測できている交絡因子は調整可能である一方、未観測の交絡因子というものは調整し切れないという点には留意する必要があります。調整によって、真の介入効果に近づいていると考えられますが、結果については吟味する必要があることを踏まえてこのような手法を

使っていくことが望ましいと考えています。

## Uplift Modeling



### Uplift Modelingとは

© JMDC Inc. ● + × ◀ 14

ここから、因果推論に機械学習の手法を融合した Uplift Modeling についてお話ししたいと思います。

#### 個人ごとの介入効果の推定

- ここまで紹介したのは、群団に対する介入の効果量の推定方法
- 群団の中でも効果が高い人と効果が低い人が存在するため、個人ごとの効果量を推定できれば効果量の高い人に優先的に介入可能

広告を見た（介入群）	広告を見てない（非介入群）
	
平均販売数 <u>0.45個</u>	平均販売数 <u>0.40個</u>

交絡因子を取り除いたことで、群団に対する介入効果は+0.05個と分かった！  
しかし、全員が等しく+0.05個の効果があるわけではなく効果量には個人差があるはず

© JMDC Inc. ● + × ◀ 15

先に紹介した「層別解析」や「傾向スコアマッチング」は、群団全体に対する介入の効果量の推計方法です。しかし、群団の中でも効果が高い人と効果が低い人が存在すると考えられます。例えば上の例のように、交絡因子を調整した結果、平均販売数が介入群 0.45 個、非介入群 0.40 個とわかり、群団全体に対する介入効果はプラス 0.05 個と分かったとします。一方で、全員が等しくプラス 0.05 個の効果があるわ

けではありません。効果量には必ず個人差があると考えられます。例えば、若い人であればそれなりに効果はあるけれども、若くない人に関しては全く効果がないなど、その人の持つ属性によって効果量はさまざまであると考えられます。

### Uplift Modelingとは

Uplift Modelingとは、個人ごとの介入による効果量を測定できるモデリング手法  
近年、ビッグデータと機械学習の発展によって様々な手法が提案されている

- ✓ Uplift Modelingを使用すれば、スマホ広告を提示した場合の購入確率の上昇幅を予測可能
- ✓ 購入確率が上昇する人から優先的にスマホ広告を提示することが可能になる

© JMDC Inc. ● + × ◀ ▶ 16

Uplift Modelingとは、個人ごとに介入による効果量を測定することができるモデリング手法です。概念自体は昔からあるものですが、特に近年、ビッグデータ等を活用する機械学習の発展に伴い、さまざまな手法が提案されています。

イメージとしては、性別・年齢や年収・購入歴など属性情報を入力すると、個人ごとの購入確率の上昇幅が出力されます。個人単位で効果量を推定できるため、購入確率が上昇する人を選定して優先的にスマホ広告を提示するような使い方が可能です。

### Uplift Modelingの研究事例

- Uber社がCausalMLという、Uplift ModelingのPythonライブラリを公開
- その他にも、多数の会社がUplift Modelingを活用した研究を行っている

uber / causalml

Uber Engineering  
Uplift Modeling for Multiple Treatments with Cost Optimization

Uplift modeling for location-based online advertising  
Shogo Kawanaka \*\*, Daisuke Moriwaki \*\*

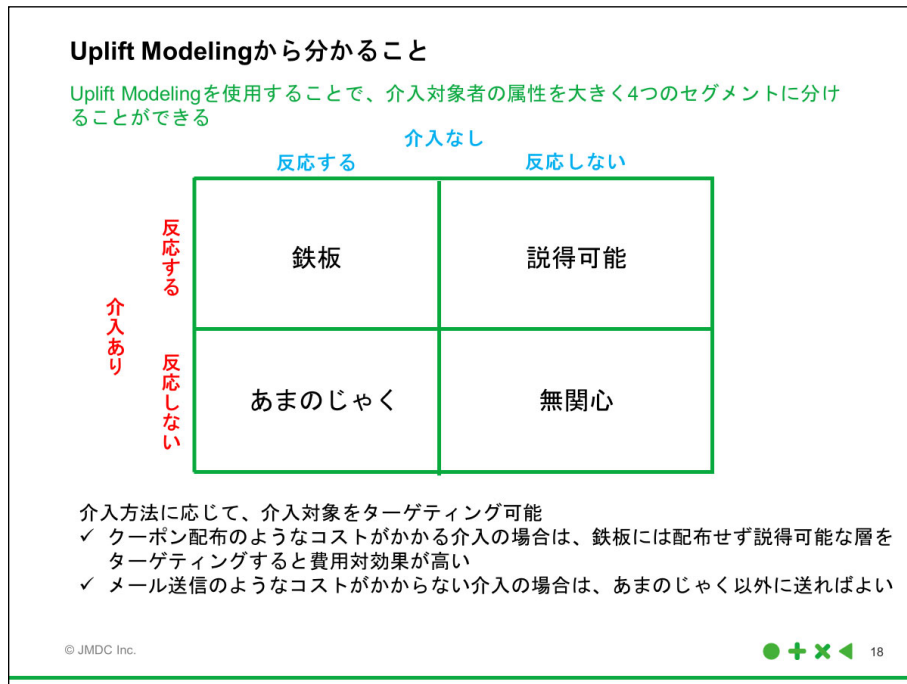
クーポンマーケティングにおける Uplift Modeling 適用の問題点と新しい評価指標  
A New Evaluation Metric and Issue of Uplift Modeling for Coupon Marketing

清水 亮洋<sup>\*1,2</sup> 富樫 陸<sup>\*1,3</sup>  
Akhito Shimizu Takashi Riku

<sup>\*1</sup>株式会社メルカリ (Mercari, Inc.) <sup>\*2</sup>北陸先端科学技術大学院大学 (Gakko University) <sup>\*3</sup>早稲田大学 (Waseda University)

© JMDC Inc. ● + × ◀ ▶ 17

先ほど Uplift Modeling としてさまざまな手法が提案されているとお話ししましたが、その具体例です。例えば、Uber 社は CausalML という Python のライブラリを開発して公開していますので、一般の方でも Uplift Modeling を使うことができ、実際にこのライブラリを使用した研究も多いです。スライドにも記載しましたがサイバーエージェントやメルカリなども論文を出しています。



また、Uplift Modeling のメリットとして、介入対象者の属性を大きく四つのセグメントに分けることができます。

一つ目は左上の鉄板層です。鉄板層は、スマホ広告を見せなくても結局購入するセグメントです。ゆえに、このセグメントにはスマホ広告を見せる意味はありません。二つ目は、一番広告を見せる意味がある群は右上の説得可能層です。スマホ広告を見せないと買わないけれども、スマホ広告を見せると買うような人です。三つ目が左下のあまのじゃく層です。広告を見せることで、逆に買わなくなってしまうような人です。介入効果としてはマイナスの集団です。最後は無関心層で、広告を見せなくても買わない人になります。

このように Uplift Modeling を使用してユーザー（介入対象者）の属性が分かることで、介入方法に応じて介入対象をターゲティングすることができるようになります。例えばクーポン配布のようにコストがかかる介入の場合、鉄板層に配布すると無駄なコストになってしまいます。クーポンを配布しても配布しなくても購入する人に配る意味はなく、説得可能層にのみクーポンを配布する方法が、最も費用対効果が高いといえます。一方、メールの一斉送信のようなコストがかからない介入であれば、あまのじゃく層以外とりあえず全体に送っておけば、誰かしら反応してくれるだろうというビジネス判断ができます。

## Uplift Modelingで必要なデータ

対象者の属性に応じて介入効果を算出するため、介入フラグ・介入結果に加えて、対象者の属性情報が必要

### 必要データ

id	介入フラグ	介入結果	属性情報			
			性別	年齢	年収	...
A	1	1	男性	31	400	...
B	0	0	女性	40	500	...
C	0	0	男性	23	300	...
D	1	0	女性	50	600	...

Uplift Modelingを使用する場合でも、交絡因子は無視できないため、RCTによるデータセットが理想的ではあるが、観察データを使用する手法も存在

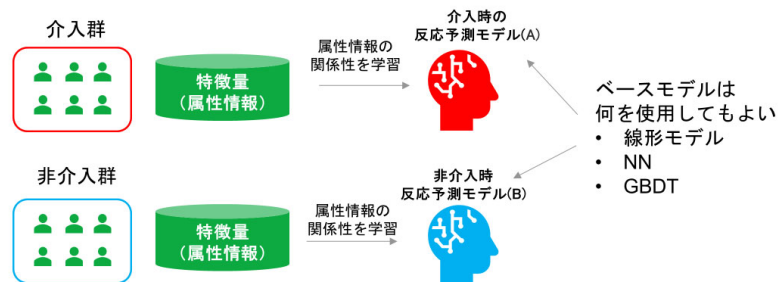
次に Uplift Modeling の原理を説明していきます。まず、必要となるデータです。スライドにあるとおり、個人ごとに、介入フラグ（介入の有無）と介入結果（購入したか否か等）に加えて、対象者の属性情報が必要です。属性情報が多く、データ量も大きいほど、より精緻なモデルを作ることができます。

注意点として、Uplift Modeling を使用する場合でも交絡因子を無視することはできません。したがって、基本的には RCT によるデータ取得が理想的ですが、観察データを使用することも可能です。

## Uplift Modelingの原理 – Meta-learner –

Uplift Modelingの一つに、Meta-learnerという手法がある  
Meta-learnerでは、T-learner、X-learner、R-learner、DR-learner等が提案されている

### T-learner (two model approach)



モデルAとモデルBを事前に学習して、  
介入効果 = モデルAの予測値 - モデルBの予測値として算出する

Uplift Modeling の一つに、「Meta-learner」という手法があります。Meta-learner は T-learner、X-learner、R-learner、DR-learner といった、さまざまな手法が提案されていますが、今回はその中の一つ、T-learner

(two model approach) を紹介します。

T-learner の発想自体はシンプルで、まずデータを介入群と非介入群に分けます。次に、介入群だけでモデルを作ります。スマホ広告の例でいえば、スマホ広告を見た群の中で、その後購入したか否かを予測するモデルを構築します。一方、非介入群に対してもモデリングします。この2モデルの予測値の差分は、介入した場合の予測購入確率から介入しなかった場合の予測購入確率の差分になっているため、それを介入効果としてみなすという手法です。

この手法のメリットは、ベースとなるモデルとして何を使用してもよいところにあります。線形モデル、Neural Network、勾配ブースティングツリーでも何でもよいので、データの属性や特性に応じたモデルを選択することができます。

ここまでが Uplift Modeling の説明になります。ここで質問を受け付けたいと思っております。よろしくをお願いします。

金澤 それでは、質問を受け付けたいと思います。会場で質問の方は挙手を、Slido で質問の方は、Slido に質問を記入してください。

質問者 A 5 ページのところで、生活習慣リスクに対して介入する例が出ていますが、他に介入が有効と見られるリスクがあったら、教えていただけないでしょうか。

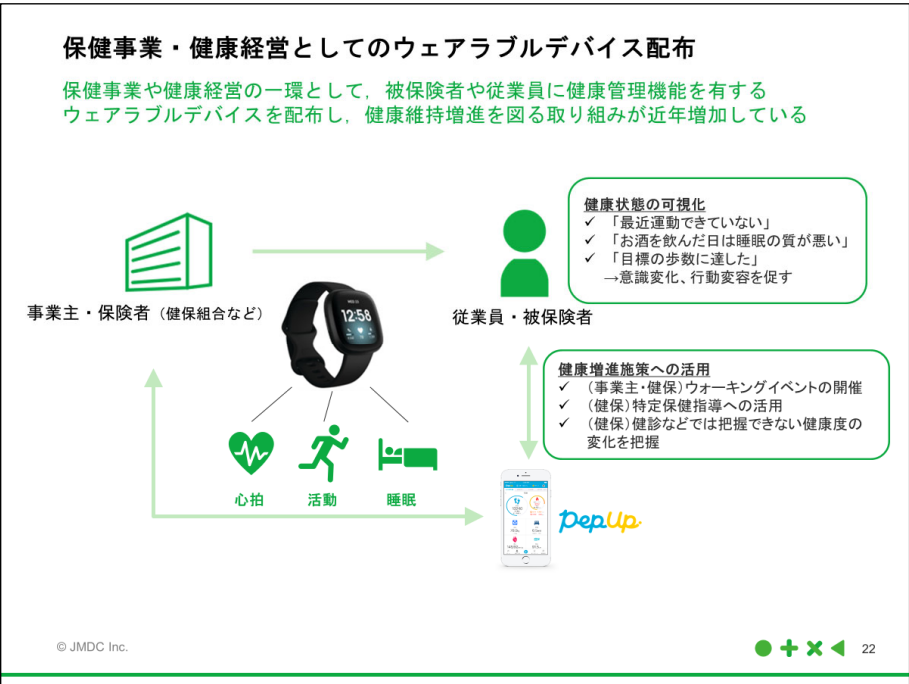
尾谷 生活習慣病以外であれば、例えばメンタル疾患などがあります。メンタル疾患や、あとは認知症リスクなど。そのような疾患は、予防的に早期介入することで、一定リスクを軽減することも期待できると思います。他には、介護リスクなども可能性はあると思います。

質問者 A ありがとうございます。

金澤 ありがとうございます。他、特になさそうですので、引き続き、次のパートに進んでいただければと思います。よろしくお願いいたします。



尾谷 では、引き続き Uplift Modeling を使った当社データでの分析例を紹介します。



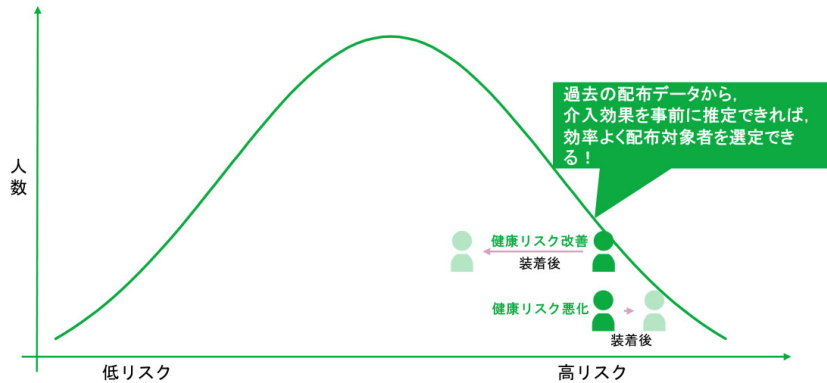
まずは、分析の背景を説明します。最近、保健事業や健康経営の一環として、被保険者や従業員に健康管理機能を有するウェアラブルデバイスを配布して、健康維持増進を図る取り組みが増加しています。このウェアラブルデバイスについてはご存じの方も多いと思いますが、心拍や活動量、睡眠量などを定量化してユーザーに可視化する機能がありますので、従業員や被保険者に対して行動変容を促せるのではないかと期待があります。



## 介入効果の推定による優先配布対象者の選定

健康リスクの高い集団には優先してウェアラブルデバイスを配布したい場合がある

- ・ 同じリスク水準でも、装着による介入効果は個人ごとに異なる
- ・ 介入効果を事前に推定することで、優先配布対象者の選定・配布台数が検討できる



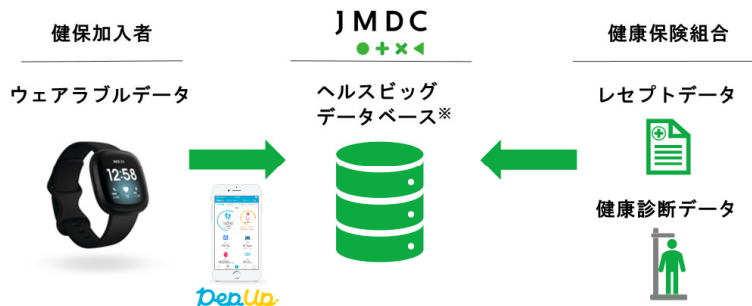
© JMDC Inc.

● + × ◀▶ 23

とはいえ、予算などの関係から、ウェアラブルデバイスを全員に配れるとは限りません。そのような場合、基本的には健康リスクの高い集団に対して、優先してウェアラブルデバイスを配布することが考えられます。しかも、このハイリスク集団の中でも、ウェアラブルデバイスの装着による介入効果は個人ごとに異なりますので、個人ごとの介入効果を事前に推定することができれば、優先配布対象者や配布台数を検討することができます。したがって、今回の分析事例における目的は、ウェアラブルデバイスを装着することによって健康リスクがどの程度改善あるいは悪化するののかという介入効果量の定量化です。

## ウェアラブルデータ×レセプト・健康診断データ

JMDCでは健保加入者向けのICTツール「Pep Up」をプラットフォームとして、fitbitのAPIを通じて連携されたウェアラブルデータと、健康保険組合から受領したレセプト・健康診断データを個人単位で結びつけて分析できる



※ウェアラブルデータとレセプト・健康診断データを個人単位で結びつけたうえで匿名加工処理を実施（健保加入者、健康保険組合双方から受領データの匿名加工利用の許諾を得ている）

ウェアラブルデバイスの装着有無による健診・レセプトデータへの影響を分析可能

© JMDC Inc.

● + × ◀▶ 24

この分析に当社のデータを利用することができます。JMDCでは、健保加入者向けのICTツール「Pep Up」



というアプリを提供しており、従業員や被保険者が装着している fitbit というウェアラブルデバイスのデータを連携することができます。当社でそのデータを匿名加工処理したうえで、ヘルスビッグデータベースに格納しています。このヘルスビッグデータベースには、健康保険組合から受領したレセプトデータや健康診断データも格納されており、このウェアラブルデータとレセプトデータ、健康診断データを結びつけて分析することが可能になっています。いつからウェアラブルデバイスをつけたのかということも分かるので、ウェアラブルデバイスの装着有無による健診レセプトデータへの影響を分析することが可能です。

**健康リスク・介入・介入結果の定義**

ウェアラブルデバイス装着による個人ごとの健康リスク変化量を調査したい  
今回、健康リスクの指標としてHbA1cを使用する

**HbA1cとは**  
健康診断の血液検査で測定する血糖値指標  
糖尿病管理のためには重要な指標で、運動や食事等の生活習慣によってコントロール可能

人間ドック 学会基準	A 異常なし	B 軽度異常	C 要経過観察	D 要治療
HbA1cの数値	～5.5%	5.6～5.9%	6.0～6.4%	6.5%～

**介入・介入結果の定義**  
介入（因子）：ウェアラブルデバイスの装着  
介入結果：HbA1cの増加有無（増加:0、非増加:1）  
※HbA1cは加齢によって増加するため、非増加で介入効果があったと判定

介入（因子） → ? → 結果  
デバイス装着 → HbA1c 増加有無

© JMDC Inc. ➡ + ✕ ◀ 25

ここからは、具体的な分析設計について説明します。今回、ウェアラブルデバイスの装着による個人ごとの健康リスク変化量を調査したいと考えています。よって、まずは健康リスクを定義しなければなりません。今回はHbA1cを使用しました。

HbA1cは健康診断の血液検査で測定する血糖値指標です。糖尿病の管理においても重要な指標になっています。この真ん中の表は人間ドック学会の基準で、5.5%までなら異常なし、5.6～5.9%であれば軽度異常、6.0～6.4%は要経過観察で、6.5%を超えると要治療となります。基本的に、HbA1cが高ければ高いほど糖尿病リスクは高く、6.5%超の場合一般的には糖尿病の疑いが強いと考えられます。

また、HbA1cは運動や食事などの生活習慣によって、一定程度コントロールすることが可能です。そのため、ウェアラブルデバイスを装着したことによって日々の生活習慣を改めるようなことがあれば、それがHbA1cの改善として表れてくると仮定して、今回の分析では健康リスクをHbA1cと定義しました。

次に、確認したい因果関係ですが、因子はウェアラブルデバイスの装着です。ウェアラブルデバイスを装着したことによって、HbA1cが増加したかしなかったかを判定するように定義しました。増加した場合は0、増加しなかった場合は1と定義しています。HbA1cは加齢によって増加するものなので、増加しなかっただけでも介入効果があったと整理しています。

## 分析対象者

### データセットの分析対象者の定義



#### 分析対象者 (N=6,710人)

- データベース
  - (株)JMDCが保有する健康保険組合由来の医療データベース
  - ヘルスケアプラットフォーム「Pep Up (ペップアップ)」
- 介入群 (装着者)
  - ウェアラブルデバイス装着前1年間と装着後1年間に健康診断を受診している者
- 非介入群 (非装着者)
  - 装着者と同時期に健康診断を受診している非装着者の中から、装着者と同年齢・性別、HbA1cが同水準の者をランダム抽出
- 装着前健診受診が2018年4月以前のデータを学習データ・検証データとし、2018年5月以降のデータをテストデータとして使用

© JMDC Inc.

● + × ◀ 26

つづいて、実際に使ったデータセットについて説明します。まずデータベースとしては、先ほどご紹介した健康保険組合の由来の医療データベースと、ヘルスケアプラットフォームの Pep Up のデータベースを使用しています。

この中から、まずは介入群（ウェアラブルデバイスを装着した群）を抽出します。ウェアラブルデバイスを装着していて、かつその前後1年間で健康診断を受診している人を介入群と定義しました。ウェアラブルデバイスを装着したことによって、HbA1cが増えたか増えていないかを見たいので、装着の前後それぞれ1年間に健康診断を受けている必要があります。

対して非介入群はウェアラブルデバイスを装着していない人ですが、単純にデータベースから抽出すると何百万人も対象になってしまい、計算コストも高くなってしまいます。そこで、非介入群はランダムサンプリングします。装着者と同時期に健康診断を受けていて、その中で年齢・性別と HbA1c が同水準の者をランダムに1対1で抽出するという方法を取っています。この年齢・性別、HbA1c が同水準の者を抽出している理由は、これらの要素が交絡因子として考えられると仮説を立てているので、あらかじめ介入群と非介入群でロックしたうえで抽出しています。

また、健診受診が2018年4月以前のデータをモデル構築用の学習データとし、2018年5月以降のデータをモデルの精度を検証するためのテストデータとしました。

## 学習データ・学習モデル

データセットのアウトカム、介入フラグ、特徴量は以下のとおり

### 特徴量（属性情報）

- 性別, 年齢
- 装着前の健診・問診項目  
HbA1c, BMI, SBP(最高血圧), DBP(最低血圧), TG(中性脂肪), TC(総コレステロール), HDLコレステロール, LDLコレステロール, GOT, GPT, GT(γ-GTP), eGFR, US(尿糖), 喫煙習慣, 運動習慣, 歩行速度, 食習慣, 飲酒, 睡眠, 生活習慣の改善, 保健指導の希望, 保健指導レベル, 体重増加, 貧血
- 装着前の医薬品の処方有無

### 介入フラグ

- ウェアラブルデバイスの装着有無

### アウトカム（介入結果）

- HbA1cの増加有無（増加：0, 非増加：1）

### モデル

- X-learner  
ベースモデル：Catboost  
傾向スコアのモデル：L1/L2正則化ロジスティック回帰モデル

© JMDC Inc.

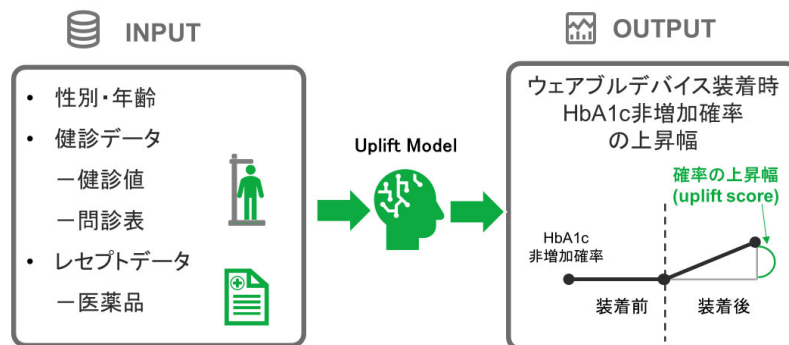
● + × ◀ 27

特徴量として性別・年齢および装着前の健診・問診項目を使っています。基本的には取得可能なすべての健診項目・問診項目を使用しています。加えて、レセプトデータから装着前の医薬品の処方有無もわかります。医薬品を約50種類のカテゴリー、糖尿病薬やメンタル弛緩薬などといったレベルに分けて、それぞれ処方されているかいないかというフラグを特徴量として使用しています。介入フラグとアウトカムについては、先ほど説明したとおりです。

Uplift Model としては X-learner を使用しました。X-learner は T-learner と異なり、モデル内部で交絡因子の調整を行ってくれます。先ほど、HbA1c や性別・年齢が同水準のものを抽出して交絡因子を一定調整した群団を作ったとお話ししましたが、観察データを使用している以上、未観測の交絡因子や調整し切れていない交絡因子は存在するので、それらを調整する目的で X-learner を使用しています。

## モデルのインプットとアウトプット

今回のモデルは、性別・年齢・健診データ・レセプトデータをインプットとして、ウェアラブルデバイスを装着した場合のHbA1c非増加確率の上昇幅を予測するモデル



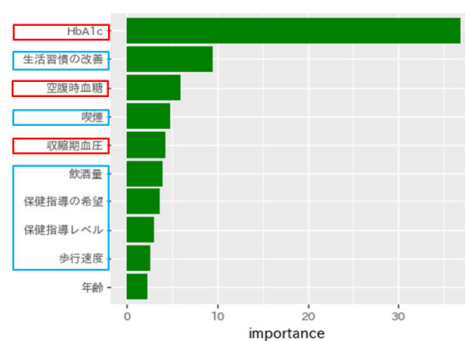
© JMDC Inc.

28

最終的に構築したモデルのまとめです。性別・年齢や健診データ・レセプトデータをインプットとして Uplift Modeling を通すと、ウェアラブルデバイス装着による「HbA1c が増加しない確率」の上昇幅が出力されます。装着しないときに対して、装着した場合にどのぐらい「HbA1c が増加しない確率」が上がるのか、その確率の上昇幅が出力されるようなモデルです。この上昇幅が大きいほど、HbA1c が増加しない確率が下がっているなので、健康リスクも下がっているということになります。

## 結果の解釈－Importance－

HbA1cや生活習慣の改善などの健診項目（赤枠）や問診項目（青枠）が介入効果の推定に寄与していることがわかる



### Importance 上位10件

テストデータを特徴量、テストデータの介入効果推定値をターゲットとして学習した考察用 CatBoost モデルの Importance

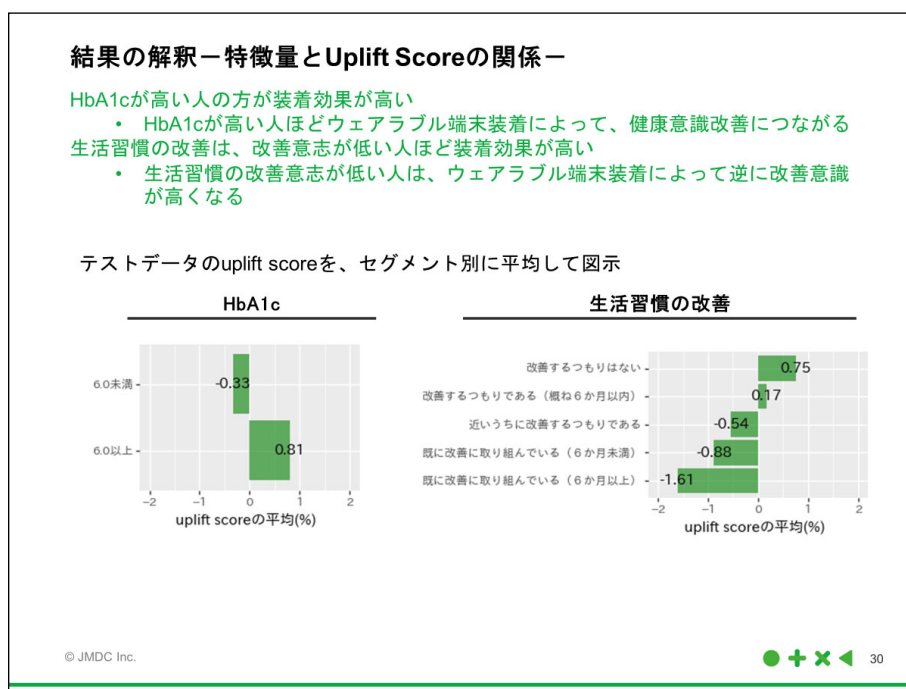
© JMDC Inc.

29

ここからは、モデルの結果の解釈について説明します。

一つ目が、今回作成したモデルがどの特徴量を重視して予測値を算出しているのか、そのトップ10です。

グラフを見ると分かる通り、HbA1c が突出しています。その次に、「生活習慣の改善」という問診項目が続きます。このように健診項目や問診項目が介入効果の推定に寄与するモデルができ上がっていることが確認できます。



続いて、特徴量が動いたときに、要は HbA1c が高ければ高いほど Uplift Score がどちら側に変動するかを確認したスライドです。モデル検証用のテストデータを推論させて Uplift Score を算出し、それをセグメント別に平均したものを図示しています。ここでは先ほどの上位 2 個の特徴量について確認しています。

例えば HbA1c であれば、HbA1c が 6.0 以上と 6.0 未満で Uplift Score の平均を比較すると、6.0 以上の方がプラス 0.81% となっていて Uplift Score が大きくなっていることが分かります。ここから分かることは、HbA1c が高い人ほど、ウェアラブルデバイスを装着することによって健康意識改善につながるということが分かります。直感的にも、HbA1c が高い人は基本的には HbA1c を下げたいという気持ちがあるので、より効果が出やすいという結果は自然です。

もう一つは、問診項目の「生活習慣の改善」です。この問診項目について補足すると、「今後、生活習慣を改善したいか」という質問項目です。回答は 5 段階あり、「改善するつもりはない」「改善するつもりである」「近いうちに改善するつもりである」「既に改善に取り組んでいる（6か月未満）」「既に改善に取り組んでいる（6か月以上）」という内容で、後ろの回答のほうがより改善意識が高いと言えます。

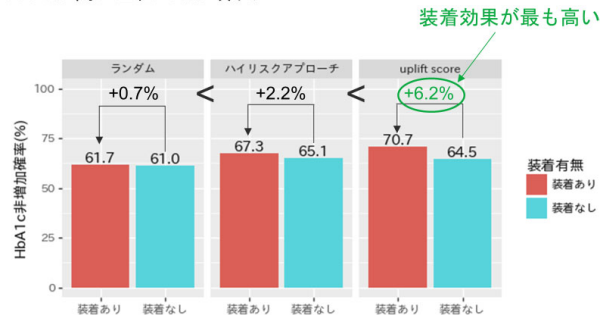
こちらの結果を見ると、一番上の「改善するつもりはない」と回答している人の Uplift Score の平均が一番高くなっています。ここから分かることは、生活習慣の改善意識が低い人ほど、ウェアラブルデバイスの装着によって改善意識が高くなっていると推察されます。

## 結果の解釈—Uplift Modelの評価—

uiplift score上位20%に介入方法が最も装着効果が高く、効率的な介入となっている

仮にテストデータ20%に介入する場合に、3種類の介入方法で装着効果を比較

- ✓ **ランダム介入**  
ランダムに全体の20%に介入
- ✓ **ハイリスクアプローチ**  
ハイリスク者の上位20%に介入 (HbA1cが高い上位20%)
- ✓ **uiplift scoreによるアプローチ**  
uiplift scoreが高い上位20%に介入



© JMDC Inc.

● + × ◀ 31

結果の解釈の最後に、この Uplift Model によってどのくらい介入効果が高い人たちを抽出できているのかを確認しました。テストデータの 20%に介入する（ウェアラブルデバイスを装着させる）場合に、3種類の方法で装着効果を比較しています。一つ目はランダム介入です。ランダムに全体の 20%に介入する方法です。二つ目は、ハイリスクアプローチです。健康リスク (HbA1c) の高い人の上位 20%に介入する方法です。最後が、Uplift Score が高い上位 20%に介入する方法です。

結果は下の図のとおりです。ランダムに介入した場合は、装着ありと装着なしの差が大体 0.7%なのに対して、ハイリスクアプローチですと 2.2%、さらに Uplift Score が上位 20%の人に介入すると、プラス 6.2%です。この 3種類の方法を比較すると、最も装着効果が高い群をモデルが抽出できていることが確認できました。

このように、Uplift Model を使えば、装着効果や介入効果が高い群に積極的・優先的にアプローチすることが可能です。

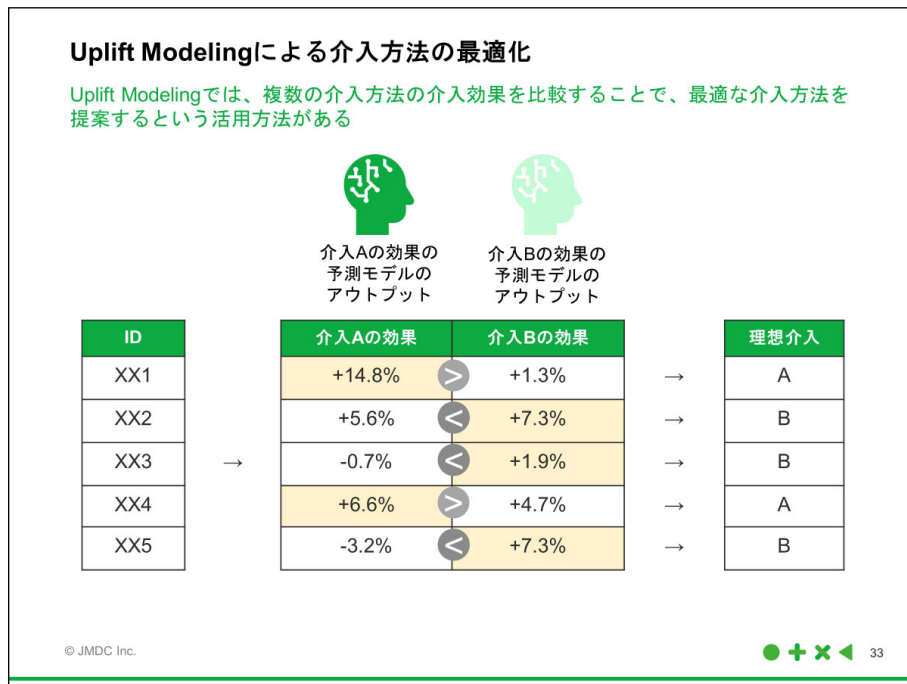
## 保険会社における活用例

© JMDC Inc.

22-22

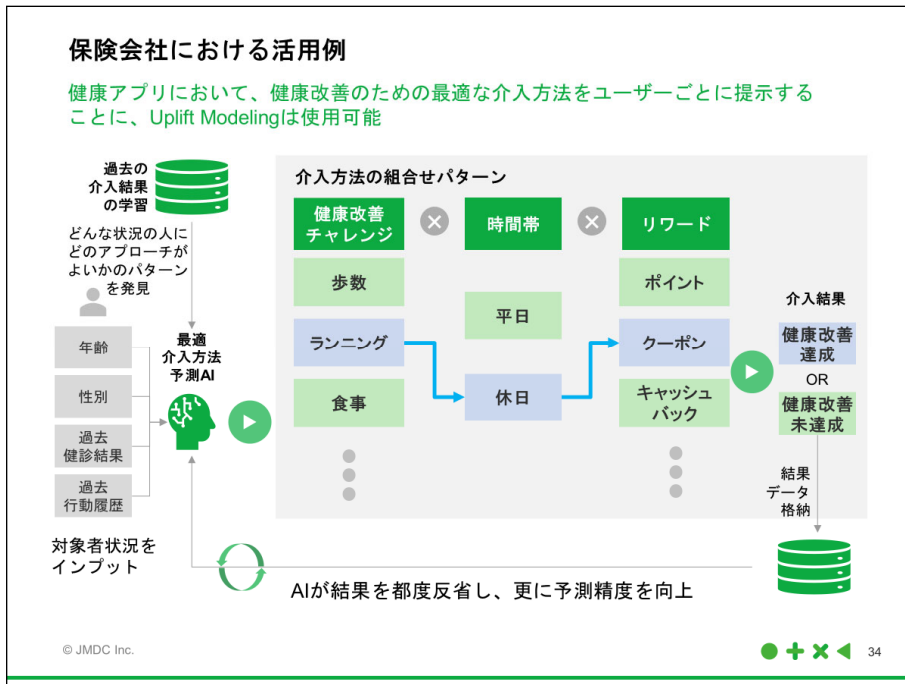
● + × ◀ 32

最後に、Uplift Modeling を実際に保険会社で活用する例を紹介したいと思います。



先ほどは一つの介入方法に対する介入効果をモデリングしましたが、複数の介入方法の介入効果を比較して、個人ごとに最適な介入方法を提案することができますようになります。例えばこのスライドの例では、介入にAとBという二つの方法があったとして、それぞれでモデリングをして個人ごとに介入Aの効果と介入Bの効果を算出します。そして、効果の大きい方を選択して介入すれば個人ごとに最適な介入ができるようになります。





この「最適な介入」の活用例として、健康アプリのようなものをイメージしていただきたいのですが、アプリから健康改善チャレンジのような目標が日々送られてくることがあります。この目標設定に際して、健康改善のために最適な介入方法をユーザーごとに提示するようなことも、Uplift Modeling を使えば可能だと考えています。

どのようなフローでモデルを構築するかを説明します。まずは、左上の「過去の介入結果」をベースにして初期的なAIを作ります。このAIは、どのような状況の人にどのアプローチがよいかというパターンを発見するAIです。このAIに年齢や性別、過去の健診結果、過去のアプリ上の行動履歴のようなものを入力した結果、右のような介入方法の組み合わせを提案してくれるAIです。例えば健康改善チャレンジの場合いろいろなチャレンジ目標があります。歩数やランニングしましょう、食事制限しましょうなどという目標です。AIにはUplift Modelingが組み込まれているので、この人であれば、例えばやることとしてはランニングが一番効果が高そうだと判断します。時間帯についても同様に、平日より休日の方がやってくれそうだと判定します。最後はリワードで、健康改善チャレンジをやったリワード（報酬）として、ポイントをももらった方がいいのか、クーポンをももらった方がやってくれるのか、キャッシュバックをももらった方がいいのかという中ならば、クーポンを配った方がこの人はやってくれそうだとこのところも判定します。

この初期的なAIに対して、さらに介入した結果を蓄積しAIにフィードバックすることで、AIがつど反省してさらに予測精度向上するという仕組みを作るといことも可能であろうと考えています。



## その他の活用例案

- ✓ 健康増進アプリの利用勧奨を行う場合に、勧奨効果の高い人をターゲットにしたリストを作成
- ✓ 健診結果の提出にインセンティブがある健康増進型保険で、健診提出していない人のうち、コールセンターや営業職員の訪問等で提出しそうな人のランク付け
- ✓ 訪問や架電による契約者とのコミュニケーションによって、どのくらい総取引量が増加するかを定量的に評価

© JMDC Inc.

● + × ◀ 35

その他の活用例案として一つ目はアプリの利用勧奨です。保険会社でも健康増進アプリなどを提供されている所があると思いますが、最初の1か月は使ってくれるけれども、その後アクティブユーザーがどんどん減っていくという課題があると聞きます。そのようなときにメールやアプリ通知などで「健康増進アプリを使ってください」などと利用勧奨を行うと思いますが、Uplift Modeling を使えばその中でも特に勧奨効果の高い人をターゲットにしたリストを作成するといったことができます。

また、健診結果の提出にインセンティブがある健康増進型保険などで、健診結果を提出していない人中でコールセンターや営業職員の人が訪問することによって提出してくれそうな人のランクづけに利用することが考えられます。

ほかにも、訪問や架電によって契約者とコミュニケーションを取ることがあると思いますが、それによって新契約がどのくらい取れるか、あるいは商品がどのくらい売れるかを定量的に評価するといったケースにも Uplift Modeling は使えると考えられます。

以上となります。最後までご清聴ありがとうございました。最後に質問タイムを設けさせていただければと思います。

金澤 それでは、質疑応答に移らせていただきたいと思いますので、会場の皆様から質問等がありましたら、挙手をお願いいたします。

ヤマカワ 住友生命のヤマカワと申します。興味深いプレゼンテーションありがとうございます。

ウェアラブルデバイスの装着効果のところで、健康改善に興味のある人にウェアラブルデバイスを装着するインセンティブがあるというバイアスがあるかもしれないと思ったのですが、そのようなバイアスも調整したうえで装着効果を出しているのでしょうか。

尾谷 非常に鋭いご指摘です。今回使用したデータは、ウェアラブルデバイスをどのような経緯で装着し

たかというところまでは取得できておりません。つまり、配布されてウェアラブルデバイスを装着したのか自発的にウェアラブルデバイスを装着したかまでは分かっていない状況です。

したがって、先ほどの結果では HbA1c が高い人ほどウェアラブル端末の装着によって健康意識改善につながると推察されたのですが、このウェアラブル端末が自発的に装着されている場合、そもそも HbA1c を下げたいという人が自発的に装着していて HbA1c が低くなっているという可能性も否定できません。これを逆因果と言います。その意味では、真に正確に分析するのであればウェアラブルデバイスを配布されたデータを取得する必要があって、そこは今後の課題にはなっています。

非常に鋭い指摘ありがとうございます。

ヤマカワ ありがとうございます。

ニシオ 住友生命のニシオと申します。プレゼンテーションありがとうございます。

お伺いしていく中で1点気になったところとして、モデリングのところと結果の解釈の間で、テストデータにおける実績データとモデルの結果の比較のところの評価ステップが入ってくるように思います。最終的な評価の解釈のところになるのですが、これはモデル結果を評価しているというところで実績データとの関係性を評価してというわけではなかったもので、間のステップがあると思っています。実際に、実務でこのようなことをするときにはその実績の評価も含めてこのモデルでよいかという議論が発生すると思うのですが、このあたりはどのような形で今回行っていますか。

尾谷 このような Uplift Modeling などの介入データの評価は非常に難しいです。本当にきちんとやるのであれば、この Uplift Score に基づいて実際に介入を行って、その後効果検証を行うようなステップでその実績を評価できると思っています。ただ今回は過去のテストデータを使用しているので、あくまでバックテストに近い結果にはなっています。その意味では正確に評価するのであれば先ほどのような方法が必要になります。

ニシオ ありがとうございます。

金澤 他、特に会場からないようでしたら、Slido のご質問を読み上げます。

一つめは、T-learner を用いて測定した介入効果について、その妥当性の確認はどのように行うのでしょうかという質問です。例えば、二つめのリスクモデルのそれぞれの汎化性能を評価する等があると思いますがご意見をお伺いできればと思います、という質問です。よろしくお願いします。

尾谷 評価方法としては、まずこの予測モデル自体の精度がどのぐらいなのかを評価することになります。したがって、例えばある介入群データで作成したモデルが他の介入群データでも、いいかえると、未知のデータに対しても精度高く反応が予測できているかが評価指標の一つになります。介入群のモデルと非介入群のモデル、この二つのモデルの精度が高ければ高いほど、基本的には精度の高い Uplift Modeling になっていると考えられます。

金澤 ありがとうございます。

それでは、ちょうど 45 分ですので本セッションはこちらで終了とさせていただきます。尾谷さん、ありがとうございました。

尾谷 ありがとうございました。