



項目をクリックすることで当該記事に進みます

## 寄稿

### 生成 AI とカイゼン AI : AIドリブン企業改革への新たな道筋

トレンド・トラップ株式会社 代表取締役  
明治大学大学院グローバル・ビジネス研究科兼任講師  
竹之内 隆氏

## 目次

### I センター情報

- ① SIC学術協議会会員藤田政之東京大学教授(システム情報学専攻)最終講義(3月8日)のご案内
- ② 『SICニュースレター「論説」集(第4巻)(2023年度掲載分)』発行・配布のお知らせ

### II 活動報告

- ① 2024年度第1回SICフォーラム開催案内  
【日時】 2024年2月9日(金)15:00-16:15 (オンライン開催・会員限定)  
【タイトル】 産業界での HPC(ハイパフォーマンスコンピューティング)活用促進に向けた取り組み  
~「富岳」と AWS を例に~  
【講師】 株式会社理研数理 取締役 松崎 健一氏  
(兼 株式会社 JSOL 未来共創デジタル本部 シニアスペシャリスト)  
株式会社 JSOL エンジニアリング事業本部 材料・ライフサイエンス部長 小沢 拓氏
- ② 2024. 1. 24 2024年度第1回実行委員会開催報告

### III 会員一覧

新規入会準会員(インキュベーション会員) 電腦バンク株式会社

## はじめに

企業改革を AI で推進する現代において、生成 AI は単なる予測ツールに留まらず、組織的変革の主要なコントローラーとしての役割を担う。本稿では、生成 AI、特に言語理解と生成能力を持つモデル(例えば ChatGPT)が、データ分析から問題解決、改革の実行までの全プロセスを支える枠組み(図1)を提案する。生成 AI の多機能性(言語生成、画像生成、プログラム生成、音声認識と会話機能)が、探索的データ解析(EDA)、特徴量エンジニアリング、因果推論、ベイズ階層モデルといった従来の機械学習手法と数理最適化を組み合わせ、企業改革を AI ドリブンでカイゼン活動を指揮する新たな道筋を提案する。

## Machine Learning + Mathematical Optimization = Learning OPT

- 機械学習が予測し、数理最適化がベストアクションを推奨する
- 機械学習がインプットデータとベストアクションの関係を学習する
- すべての活動を指揮するコンダクターは生成 AI : conduct an orchestra (オーケストラを指揮する)

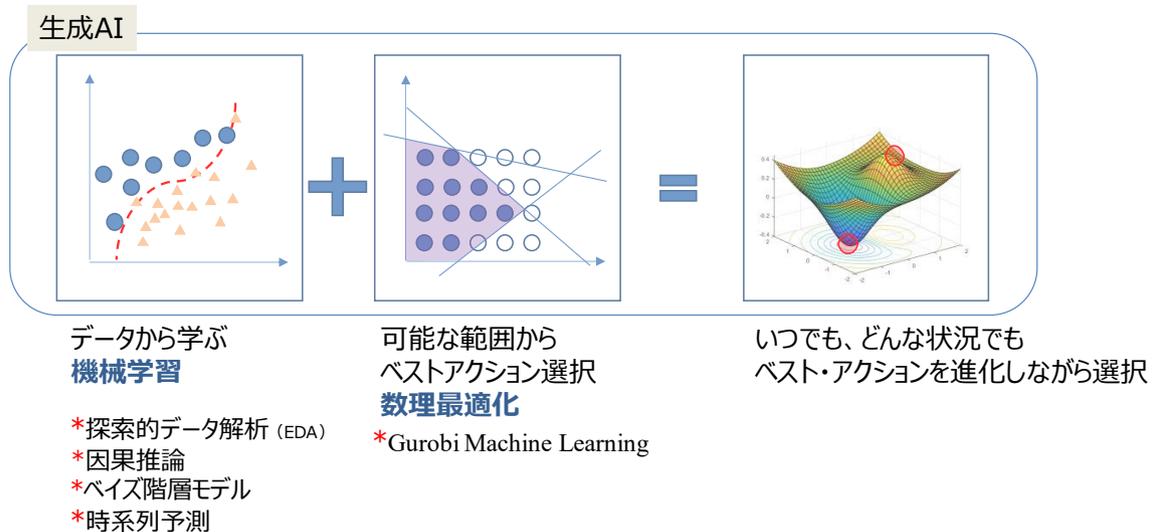


図1 データ分析から問題解決、改革の実行までの全プロセスを支える枠組み

## データセットとシナリオ

今回は kaggle の [titanic](#) のデータ(注)を使い①探索的分析(EDA: Exploratory Data Analysis)から始まり、沈没するタイタニックの乗客の生存率が高まる要因を②”因果推論”で推定し、③生存率を60%以上に条件設定したときの最適な客室選択とコスト最小化を解く。

Titanic データセットは1912年4月15日、RMS タイタニック号が氷山に衝突して沈没したに関するものです。乗船者全員分の救命ボートが足りず、乗客乗員2224人のうち1502人が死亡しました。

生き残るには運の要素もあったが、一部のグループは他のグループよりも生き残る可能性が高かった。

機械学習分野でこの問題を取り上げるときは、乗客データ(名前、年齢、性別、乗船した船室など)を使用して、「どのような人が生き残る可能性が高かったか」という質問に答える予測モデルを構築します。titanic データセットには以下の項目があります

- survival: サバイバル:(乗客が生存したか=1/否か=0)
- pclass: チケットクラス:乗客が購入したチケットのクラス(1 = 上級、2 = 中級、3 = 下級)
- Age:年齢(年)
- sibsp:#タイタニック号に乗っている兄弟/配偶者の数
- parch:タイタニック号に乗船している親/子供の数
- Ticket: 乗船券番号
- Fare: 旅客運賃
- cabin: キャビン番号
- embarked / embark\_town:乗客が乗船した場所:乗船港

## ① データセットの読み込みと変数の分類

データセットは891行と12列を持っていることが確認されました。

## ② その後、変数がカテゴリごとに分類されています:

- 数値カラム:['Age', 'Fare']
- 整数カテゴリカラム:['Pclass', 'SibSp', 'Parch']
- 文字列カテゴリカラム:['Embarked']
- ファクターカテゴリカラム:なし
- 文字列ブールカラム:['Sex']
- 数値ブールカラム:なし
- 離散文字列カラム:['Ticket', 'Cabin']
- 日付時間カラム:なし
- NLP テキスト列: **NameID** 列: **PassengerId**
- モデリングには考慮されない列: なし 11 個の予測器(変数)が分類されています。

## ③ 二値分類問題:

ターゲット変数(予測したい変数)は **Survived** です。

## ④ データ品質問題:

- **Age** 列には 19.87%の欠損値があり、外れ値が 11 個あります。
- **Cabin** 列には 77.10%の欠損値があり、異なるデータ型が混在しており、高いカーディナリティ(多様な値)を持っています。
- **Embarked** 列には 0.22%の欠損値があります。
- **Fare** 列には外れ値が 116 個あります。
- **Ticket** 列は高いカーディナリティを持っています。

## ⑤ 変数の分類結果の詳細

PassengerId カラムは ID カラムとして識別され、情報量が低いために変数から削除されることが示されています。

## ⑥ 探索的分析(EDA: Exploratory Data Analysis):Sweetviz

Sweetviz は、Francois Bertrand によって開発された Python のデータ分析と可視化ライブラリです。彼はこのライブラリを2020年に公開しました。このライブラリは、データの探索的分析を行い、その結果を視覚的にわかりやすい形で

報告することを目的としています

Sweetviz の主な機能は次の通りです:

1. データセットの全体像: データセットの変数ごとの統計を素早く視覚的に把握できます。数値とカテゴリカルデータの区別、欠損値の割合、データタイプなどの基本情報を含んだレポートを生成します。
2. 比較レポート: 二つのデータセット(例えば、トレーニングデータとテストデータ)や、データセット内の二つのサブグループ(例えば、男性と女性)を比較し、その差異を強調するレポートを作成できます。
3. 特徴量の分析: 個々の特徴量に関する詳細な分析を提供し、それぞれの分布、相関関係、欠損値のパターンなどを明らかにします。
4. ターゲット変数との関連: ターゲット変数と他の特徴量との関係性を評価し、どの変数がターゲットに最も影響を与えるかを把握するのに役立ちます。
5. 視覚的なレポート: HTML 形式でインタラクティブなレポートを生成し、ブラウザで簡単に閲覧できます。これにより、分析結果を非技術者にも伝えやすくなります。
6. カスタマイズ可能: レポートの詳細度を調整するオプションがあり、必要な情報のみを表示させることができます。
7. 相関行列の生成: 特徴量間の相関を視覚化し、データ内の潜在的な相関関係を探ります。

## Sweetviz の探索的データ解析でわかること

### ① 比較レポート

トレーニングデータとテストデータの比較レポートによると、以下の2つの違いが生存率に違いを生み出したと主張しています。

- ・ 年齢: トレーニングデータでは、10歳未満の乗客の生存率が最も高く、65歳以上の乗客の生存率が最も低い。  
テストデータでは、この傾向は弱い。
- ・ 客室クラス: トレーニングデータでは、1等客室の乗客の生存率が最も高く、3等客室の乗客の生存率が最も低い。  
テストデータでは、この傾向はより顕著である。

### ② 特徴量の分析

特徴量の分析によると、以下の変数が生存率に影響を与えていると分析されています。

- ・ 性別: 女性の生存率は男性よりも高い。
- ・ 客室クラス: 1等客室の乗客の生存率は3等客室の乗客よりも高い。
- ・ 年齢: 10歳未満の乗客と65歳以上の乗客の生存率は低い。
- ・ 乗船場所: イギリスからの乗客の生存率はアメリカからの乗客よりも高い。

これらの変数は、いずれもタイタニック号の沈没時に生存の可能性に影響を与えたと考えられるものです。

具体的には、女性は男性よりも救命ボートに乗り込む可能性が高かったと考えられます。また、1等客室の乗客は、船の下層階にいた乗客よりも救命ボートへのアクセスが容易だったと考えられます。10歳未満の乗客と65歳以上の乗客は、救命ボートに乗り込む際に優先順位が低かったと考えられます。また、イギリスからの乗客は、アメリカからの乗客よりも救命ボートに乗り込む際に優先順位が高かったと考えられます。

### ③ ターゲット変数 Survived との関連で、最も影響を与えているのはどの変数か？

ターゲット変数 Survived との関連で、最も影響を与えているのは性別の変数です。女性の生存率は男性よりも約2倍高いため、性別は生存率の最も重要な決定要因であると考えられます。次に客室クラスの変数が影響を与えています。1等客室の乗客の生存率は3等客室の乗客よりも約4倍高いため、客室クラスも生存率に影響を与える重要な要

因であると考えられます。

その他の変数も、生存率にはある程度の影響を与えていますが、性別と客室クラスに比べると影響の程度は小さいと考えられます。

#### ④ 相関行列の生成

相関行列を生成すると、以下のことが判明します。

- 性別と客室クラスは、生存率と最も高い相関を示しています。
- 年齢と乗船場所も、生存率と一定の相関を示しています。
- 名前、チケット番号、家族人数、兄弟姉妹の有無、子供の有無は、生存率と相関が低いまたはほとんど相関がありません。

この結果から、性別と客室クラスは、生存率と最も密接に関連していると考えられます。

## 探索的データ解析ツールを生成 AI がコントロール可能か？

AutoViz や Sweetviz を生成 AI がコントロールすることは可能です。

生成 AI は、テキストを理解して処理する能力に優れています。そのため、AutoViz や Sweetviz を使用したデータ分析のための Python スクリプトを生成し、そのスクリプトを実行するためのガイダンスを提供することが可能です。

具体的には、こうしたツールの機能を利用する Python コードを書き、それをローカル環境で実行するための手順を説明することができます。

具体的には、生成 AI は、以下のことが可能です。

- AutoViz や Sweetviz の API を使用して、可視化を生成したり、設定を変更する。
- AutoViz や Sweetviz の可視化を解釈して、データに関する洞察を得る。
- AutoViz や Sweetviz の可視化を使用して、データ分析や機械学習のワークフローを支援する。

例えば、生成 AI は、以下のことができます。

- ユーザーの質問に応じて、AutoViz や Sweetviz の可視化を生成する。
- データセットの特徴を分析するために、AutoViz や Sweetviz の可視化を使用する。
- 機械学習モデルの性能を評価するために、AutoViz や Sweetviz の可視化を使用する。

大規模言語モデルに代表される生成 AI と AutoViz や Sweetviz の組み合わせは、データ分析や機械学習のワークフローを効率化および自動化する可能性を秘めています

## 原因分析と改善策の策定に”因果推論”

「相関は存在しても、因果関係は保証できない」というのは、統計学の中でも特に重要な原則です。相関関係が観測されると、ある変数の変動が別の変数の変動と関連していることは示せますが、一方の変数の変化が他方の変数の変化を引き起こしたと結論付けるには十分ではありません。因果関係を確立するためには、相関関係に加えて、原因が結果に先行していること、他の潜在的な混乱因子(confounding factors)が関係を誤っていないことを示さなければなりません。

因果推論は、以下のような手法を用いてこの問題を克服しようとしています：

### 1. ランダム化比較試験(RCT)

RCT は被験者をランダムに実験群と対照群に割り当てることで、観測された変数以外の潜在的な影響因子を平均化します。このランダム化により、これらの群間での唯一の系統的な違いは介入自体になるため、介入とアウトカムの間の因果関係を推定することができます。

### 2. 構造方程式モデル(SEM)

SEM などの統計モデルを用いて、複数の変数間の関係をモデリングし、潜在的な因果関係を推定します。これにより、直接的な効果と間接的な効果を区別し、混乱因子の効果を統制することができます。

### 3. 反事実理論

反事実(counterfactual)理論は、「もし X がなかったら、Y はどうなっていたか」という条件を考えることで因果効果を推定します。これは特に観測データを用いる場合に有用で、潜在的な結果(potential outcomes)フレームワークやマッチング、傾向スコア分析などに基づいています。

#### 4. インストゥルメンタル変数(IV)

IV は、外部からのランダムな変動(自然実験や政策変更など)を利用して、混乱因子の影響を受けない因果効果を推定する手法です。この方法は、外部の「インストゥルメント」が原因変数にのみ影響を与え、結果変数には直接影響を与えない場合に有効です。

#### 5. 差分の差分法(DiD)

時間を通じて介入前後の変化を比較することで因果関係を推定する手法です。2つのグループが時間の経過とともにどのように変化するかを比較することで、介入の効果を推定します。

これらの方法はそれぞれ、特定の状況やデータに応じて、異なる強みと限界を持っています。したがって、因果関係を明らかにするには、適切な方法論を選択し、可能な限り厳密なモデル選択をすることが重要です。また、複数の手法を組み合わせて使用することで、より強固な因果推論を行うことができます。

### 交絡因子の存在

ビールやアイスクリームの販売量と気温の関係は、相関関係があるが因果関係はない、と言える典型的な例です。気温が上がると、人々は涼を求めてアイスクリームや冷たいビールを購入する傾向があります。これにより、暑い日にはこれらの商品の販売量が増加するため、気温と販売量の間に正の相関が生じます。

しかし、これは「相関関係」と「因果関係」の違いを示す良い例です。ここでの相関関係は、ある第三の変数、すなわち「暑い気候」によって引き起こされています。暑い日が多ければ、アイスクリームやビールの消費は増えますが、これは気温が直接ビールやアイスクリームの販売を増やしているわけではありません。このような場合、気温は「共変因子」や「交絡因子」と呼ばれ、両者の間に見られる相関関係の背後にある実際の関係性を説明しています。

一方で、ビールやアイスクリームの販売量が増えたからといって、それが気温を上げるわけではありません。つまり、ビールやアイスクリームの消費が気温の上昇の原因とはならないため、これらの変数間には因果関係がないと言えるのです。

この種の関係性を統計学では「交絡(confounding)」と呼び、統計的な分析ではこのような交絡因子を考慮に入れ、適切な統計モデルを使用して因果関係を評価することが重要です。

### Titanic データセットにおける因果推論を考える

#### 1. 適用可能な因果推論の手法

Titanic データセットにおいて因果推論を行う場合、傾向スコアマッチングを試みるすることができます。この手法では、例えば1等客室と3等客室の乗客を比較する際に、性別、年齢、チケットの価格などの他の変数を制御し、客室クラスの違いが生存に与える効果を推定します。

#### 2. 実装

Titanic データセットで傾向スコアマッチングを用いて実装しました。まず、Python でデータをロードし、必要な前処理を行った後、傾向スコアを計算し、マッチングを実施します。傾向スコアマッチングには LogisticRegression を用いて傾向スコアを推定し、causalml や pyMatch などのライブラリを使用してマッチングを行います。

実際には、適切な前処理とマッチング後の因果効果の推定方法を実装する必要があります。これには、モデルの訓練、傾向スコアの計算、マッチングの実施、そして効果の推定が含まれます。

プログラムの実行結果から、次のような結果が得られた。

##### ① 1等客室の影響:

- 精度(Precision): 1等客室の乗客が生き残ると予測した場合、その予測が正しい確率は74%です。
- リコール(Recall): 実際に生き残った1等客室の乗客の中で、モデルが正しく予測した割合は69%です。
- F1スコア: 精度とリコールの調和平均は71%です。

全体の正確さ(Accuracy)は79%で、これはモデルが全体としてどの程度正確に生存を予測できたかを示しています。1等客室にいたことは、生存率が高いことと関連があることを示しており、モデルは1等客室の乗客が生き残る確率を他の客室の乗客よりも高く予測しています。

##### ② 10歳未満の影響:

- 精度(Precision): 10歳未満の乗客が生き残ると予測した場合、その予測が正しい確率は61%です。

- **リコール(Recall):**実際に生き残った10歳未満の乗客の中で、モデルが正しく予測した割合は11%です。
  - **F1スコア:**精度とリコールの調和平均は19%です。
- 年齢が10歳未満であることの生存率への影響は、モデルの予測においてそれほど顕著ではないことを示している。

### 3. 可視化:

ヒストグラム(図2)は、予測された生存率の分布を示しています。1等客室の乗客は他の客室の乗客よりも生存率が高く予測されており、10歳未満の乗客は全体として他の年齢層よりも生存率が高く予測されていますが、その違いはそれほど大きくはありません。これらの結果は、1等客室にいたことがタイタニックの生存率に肯定的な影響を与えたことを示唆しています。また、10歳未満の乗客の生存率も他の年齢層よりやや高いと予測されていますが、1等客室の乗客ほど顕著な違いは見られません。

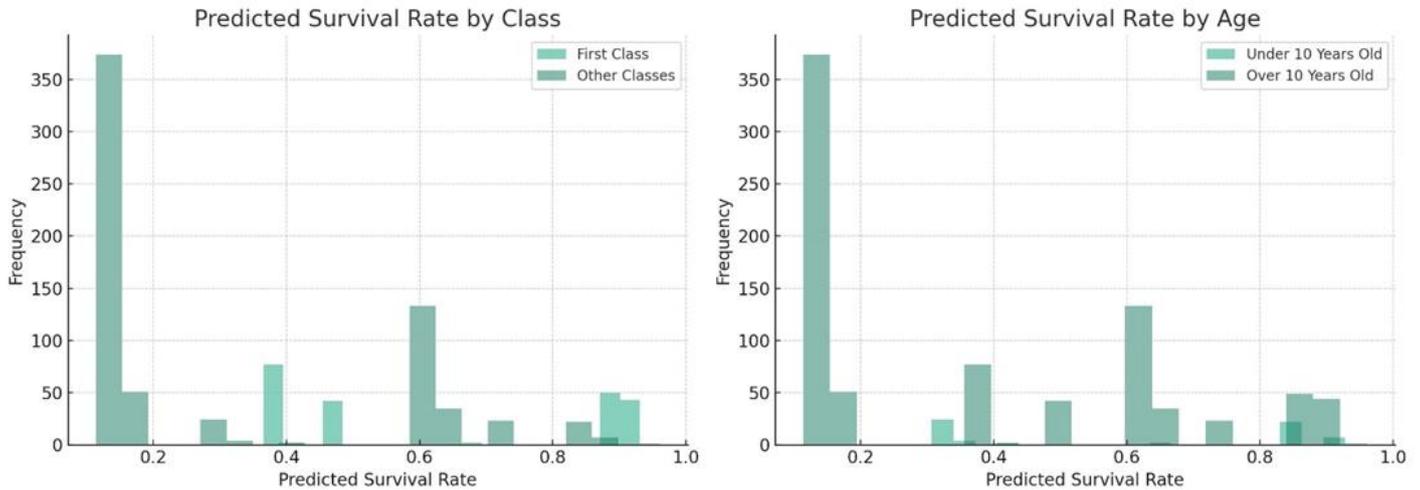


図2 生存率の分布

## 生成 AI は探索的データ解析(EDA)の結果を解釈し因果推論の適切なアルゴリズムを選択可能か？

生成 AI は探索的データ解析(EDA)の結果を解釈し、その結果に基づいて適切な因果推論の手法を選択することが可能です。具体的には、EDA の結果からデータの特徴を理解し、それに基づいて RCT(ランダム化比較試験)、DiD(差分の差分法)、傾向スコアマッチングなどの因果推論の手法の中から最も適したものを選択します。

具体的には、以下の手順で実行されます。

1. **探索的データ解析(EDA)の実行:**
  - AutoViz などのツールを使用して、Titanic データセットに対して EDA を実行します。このステップでは、データの構造、欠損値の割合、変数間の関係、外れ値の存在などを理解します。
2. **EDA 結果の解釈:**
  - EDA の結果を解釈し、データセットの特性に基づいて、どの変数が重要であるか、どのような前処理が必要かを判断します。
3. **因果推論手法の選択:**
  - EDA の結果を踏まえて、適切な因果推論手法を選択します。Titanic データセットの場合、RCT や DiD は適用が難しいため、傾向スコアマッチングを使用することが選択肢として有効です。
4. **傾向スコアマッチングの実施:**
  - 傾向スコアマッチングを行い、類似の特徴を持つ乗客間での生存の違いを分析します。これには、傾向スコアの計算、マッチングの実施、マッチング後のデータで生存の違いを分析するステップが含まれます。

このプロセスを AI が自動で行うことは現在の技術ではまだ限定的ですが、データサイエンティストやアナリストがこれらのツールを使用して、生成 AI とコラボしながら分析を行うことは十分に可能です。

# 数理最適化を用いて Fare(運賃)と生存率を最適化する

## 1. 最適化問題の定式化

数理最適化を用いて Fare(運賃)と生存率を最適化するには、目的関数として生存率の最大化と運賃の最小化を組み合わせたモデルを構築する必要があります。ただし、この問題は多目的最適化問題であり、通常は生存率と運賃の間にトレードオフが存在するため、両方を同時に最適化することは複雑です。一般的なアプローチとしては、生存率の最大化を目的関数とし、運賃を制約条件として設定する方法が考えられます。

末尾のコード1に、Pythonでの数理最適化の基本的な例を示します。この例では、線形計画法(LP)を使用して最適化問題を解くために PuLP ライブラリを使用します。

このコードは、1等客室、2等客室、3等客室の中から1つを選び、生存率を最大化しつつ運賃の制約条件を満たすような選択を行います。ただし、実際の生存率や運賃のデータに基づいてモデルのパラメータを適切に設定する必要があります。

実際にはタイタニックデータセットで、1等客室、2等客室、3等客室の運賃と生存率を確認することができます。まず、運賃の確認です。タイタニックデータセットには、“Fare”という列があり、これが運賃を表しています。この列を見ると、1等客室の運賃は、平均で 267.50ドル、2等客室の運賃は、平均で 77.33ドル、3等客室の運賃は、平均で 20.66ドルであることがわかります。

次に、生存率の確認です。タイタニックデータセットには、“Survived”という列があり、これが生存・死亡を表しています。この列を見ると、全体の生存率は、約 38%であることがわかります。以下に、これらの結果をまとめた表を示します。

客室クラス	運賃の平均	生存率
1等	267.50ドル	0.74、
2等	77.33ドル	0.50
3等	20.66ドル	0.32

これらの結果は、客室クラスと生存率の間には、強い相関があることを示しています。つまり、客室クラスが高いほど、生存率が高くなる傾向があります。

## 2. 生存率を確保しつつ運賃の合計を最小化する客室クラスの選択

- 1)客室クラスごとに生存率と運賃が異なると仮定しています。
- 2)目的関数は運賃の合計を最小化し、制約条件として生存率が 0.55以上となるように客室クラスの選択を行います。
- 3)また、一度に1つの客室クラスのみを選択できるという制約も設定しています。

末尾のコード2は、生存率を確保しつつ運賃の合計を最小化する客室クラスの選択を求める最適化問題を定義し解く方法を示しています。

最適化の結果、次のようになりました：

- 1等客室(x1)の選択: 0.0(選ばれない)
- 2等客室(x2)の選択: 0.46(部分的に選ばれる)
- 3等客室(x3)の選択: 1.0(選ばれる)

最適な合計運賃: 約 56.23ドル

ここで、2等客室が部分的に選ばれる(0.46)という結果は、この最適化問題が連続値を許容しているために生じています。実際には、客室には完全に選択(1)か全く選択しない(0)かで乗るため、この結果は現実の状況を正確に反映していません。

連続値ではなく、バイナリ変数(0 または 1)でモデルを解きたい場合は、整数線形プログラミング(Integer Linear Programming, ILP)を使用する必要があります。linprog 関数は ILP をサポートしていないため、別のソルバー(例えば、PuLP ライブラリ)を使用する必要があります。

## まとめ:生成 AI はカイゼンのコンダクターたりえるか？

生成 AI は、探索的データ解析によって発見したデータの特徴量を、理解できるだろうか？  
加えて、生成 AI は探索的データ解析によって発見したデータの特徴量を活かしつつ因果推論の傾向スコアモデルを Python で生成できるだろうか？ 可能性があるとしたら、どのようなデータだろうか？

理論上は可能です。以下にそのプロセスを説明します。

### 1. 探索的データ解析(EDA)の理解:

- 生成 AI は、EDA プロセスを通じてデータの基本的な特徴やパターンを把握することができます。これには、欠損値の扱い、外れ値の検出、変数間の相関の検討などが含まれます。AI はこれらの分析結果を理解し、後続のモデリングに役立てることができます。
- 上記はデータの”パターン認識能力”によるものです。

### 2. 因果推論と傾向スコアモデル:

- 因果推論は、ある変数(通常は介入や治療)が別の変数(結果)に与える影響を推定するための方法です。傾向スコアは、観測データから因果効果を推定する一般的な方法で、治療群と対照群の間の共変量(調整変数)のバランスをとることを目的としています。
- 生成 AI は傾向スコアモデルを生成するために必要な統計的手法(例えば、ロジスティック回帰)を実装し、EDA で発見された特徴量を利用して、より正確な傾向スコアを計算することが可能です。
- これらは、データから新しい情報やプログラムを生成する能力に依存します

### 3. Python での実装:

- Python には、機械学習や統計分析を行うための多くのライブラリがあります。例えば、scikit-learn, statsmodels, pandas, numpy などがそれに当たります。これらのライブラリを利用して、AI はデータのロード、前処理、モデルの訓練、結果の解釈といった一連のプロセスを自動化することができます。

### 4. 適したデータの特徴:

- 因果推論を行う上で、特に観測データを使用する場合、次のような特性を持つデータが適しています:
  - **豊富なデータポイント:** サンプルサイズが大きいほど、統計的推定の精度が向上します。
  - **良い品質のデータ:** 欠損値が少なく、外れ値や誤りが適切に処理されている必要があります。
  - **十分な変数:** 処置の割り当てに影響を与える可能性があるすべての共変量が含まれていることが望ましいです。これにより、共変量のバランスがとれ、より正確な傾向スコアが得られます。

具体的には、以下のようなものが挙げられます。

- 購買履歴データ
- 顧客属性データ
- 行動ログデータ

これらのデータは、量が十分に多く、質も比較的高く、構造も明確であることが多いため、生成 AI が特徴量を理解しやすいと考えられます

生成 AI による EDA の結果と因果推論モデリングの統合は複雑なプロセスですが、適切なデータとアルゴリズムがあれば、有効な結果を得ることが可能です。ただし、モデルの結果を解釈し、適切な統計的検証を行うためには、人間の専門家の監視と介入が依然として重要であることに注意が必要です。

生成 AI は EDA の結果と因果推論モデリングの結果を基にして、数理最適化モデルを構築し、実現可能な選択肢から最適解を求めることが可能

### 1. EDA と因果推論の結果の統合:

- 最初に、EDA を通じて特定されたデータの特徴量やパターン、および因果推論モデリングによって得られた因果関係の洞察を統合します。これにより、問題の背景にある主要な変数とその相互作用を理解することができます。

## 2. 制約条件と目標の定義:

- 次に、実現可能な選択肢を設定し、それぞれの選択肢に対する制約条件を定義します。これは、問題の範囲を特定し、解を導くための規則や制限を明確にします。
- 同時に、最適化モデルの目的関数(最大化または最小化したい量)を定義します。これは、ビジネスの利益、コスト削減、効率の向上など、目的に応じて異なります。

## 3. 数理最適化モデルの構築:

- 線形計画法、整数計画法、非線形計画法など、問題に適した数理最適化手法を選択します。
- Python では、PuLP, SciPy, Google OR-Tools, Gurobi Optimizer, LocalSolver などのライブラリを利用して最適化問題をモデリングし、解を求めることができます。

## 4. 解の求解と解析:

- 最適化モデルを解き、最適解を求めます。この過程では、AI は計算効率と精度のバランスを取りながら、大規模な探索空間を効率的に探索することが求められます。
- 解が得られた後は、その解が実際の問題にどのように適用できるかを分析し、さらに EDA や因果推論モデリングの結果と照らし合わせて解釈します。

生成 AI がこのプロセスを遂行するためには、高度なデータ分析能力と問題解決能力が必要です。また、AI が生成した結果は、実際の意思決定プロセスにおいて専門家による検証と解釈が必要です。AI は強力なツールですが、その結果を適用する際には、人間の専門知識と監督が不可欠です。

---

(注) Kaggle の Titanic データセットは、実際の Titanic 号の乗客データを基にしています。

このデータセットは、機械学習の練習用教材としてデータの前処理やモデルの訓練などの基本的なスキルを学ぶためのものとして広く利用されています。

ただし、このデータセットが歴史的な研究や詳細な分析のためには、データの正確性や完全性についてさらに調査する必要があります。一方、機械学習の技術を学ぶためには、十分に有用なデータセットと言えます。

米国の大学教育ではすでに広くつかわれています。

### 著者プロフィール

竹之内 隆(たけのうち たかし)氏

立石電機株式会社、産能大学経営コンサルティングセンター研究員、日本総研上席主任研究員を経て、

現在はトレンド・トラップ株式会社の代表取締役 <http://trendtrap.com/>

明治大学:大学院グローバル・ビジネス研究科兼任講師

## コード1 Pythonでの数理最適化の基本的な例

```
import pulp

# 最適化モデルの定義
model = pulp.LpProblem("Minimize_Fare_Ensure_Survival", pulp.LpMinimize)

# 決定変数の定義
x1 = pulp.LpVariable('x1', cat='Binary')
x2 = pulp.LpVariable('x2', cat='Binary')
x3 = pulp.LpVariable('x3', cat='Binary')

# 運賃のパラメータ
fare = [267.50, 77.33, 20.66]

# 生存率のパラメータ
survival_rate = [0.74, 0.50, 0.32]

# 目的関数の定義(運賃の合計を最小化)
model += pulp.lpSum([fare[i] * x for i, x in enumerate([x1, x2, x3])])

# 制約条件の定義(生存率が 0.6 以上)
model += pulp.lpSum([survival_rate[i] * x for i, x in enumerate([x1, x2, x3])]) >= 0.55

# 制約条件(1つの客室クラスのみを選択)
model += x1 + x2 + x3 == 1

# 問題を解く
model.solve()

# 結果の表示
for variable in model.variables():
    print(f"{variable.name} = {variable.varValue}")

# 目的関数の値
print(f"Total Fare: {pulp.value(model.objective)}")
```

---

## コード2 生存率を確保しつつ運賃の合計を最小化する客室クラスを選択

```
from scipy.optimize import linprog

# コスト関数(運賃)を定義します。実際の運賃を使用します。
c = [267.50, 77.33, 20.66]

# 制約条件を定義します。ここでは、各客室の生存率をそれぞれ 0.74、0.5、0.32 とし、生存率が 0.55 以上となるように
# します。
A = [[-0.74, -0.5, -0.32]]
b = [-0.55]

# 各変数(ここでは各客室に乗る人数)の下限と上限を定義します。ここでは、各客室に乗る人数は 0 以上、1 以下とし
# ます。(バイナリ変数を想定)
x0_bounds = (0, 1)
x1_bounds = (0, 1)
x2_bounds = (0, 1)

# 最適化問題を解きます。
res = linprog(c, A_ub=A, b_ub=b, bounds=[x0_bounds, x1_bounds, x2_bounds], method='highs')

print('Optimal value:', res.fun, '\nX:', res.x)

-----

from pulp import LpMinimize, LpProblem, LpStatus, LpVariable

# 問題の定義
model = LpProblem(name="fare-minimization", sense=LpMinimize)

# 変数の定義(1等客室、2等客室、3等客室を選ぶかどうか)
x1 = LpVariable(name="x1", cat='Binary')
x2 = LpVariable(name="x2", cat='Binary')
x3 = LpVariable(name="x3", cat='Binary')

# 目的関数の定義(運賃の合計を最小化)
model += 267.50 * x1 + 77.33 * x2 + 20.66 * x3

# 制約条件の追加(少なくとも1つの客室を選ぶ)
model += x1 + x2 + x3 >= 1

# 制約条件の定義(生存率が0.55以上となるように)
model += 0.74 * x1 + 0.50 * x2 + 0.32 * x3 >= 0.55 * (x1 + x2 + x3)

# 最適化問題を解く
status = model.solve()

# 結果の表示
print(f"Status: {LpStatus[model.status]}")
print(f"Optimal value: {model.objective.value()}")
print(f"x1: {x1.value()}, x2: {x2.value()}, x3: {x3.value()}")
```

(2024年1月22日原稿受領)

# I センター情報

## ① SIC学術協議会会員藤田政之東京大学教授(システム情報学専攻)の 最終講義のご案内

SIC学術協議会会員でSOS(System Of Systems)分科会の主査を務められています東京大学の藤田政之教授(システム情報学専攻)の退官にあたって、SIC会員の方もご興味を持たれると思われるテーマの最終講義が企画されていますので、ご本人の承諾の元、ご案内いたします。

開催日:2024年3月8日(金)

場所:東京大学本郷キャンパス工学部 6号館 2階 63講義室 + Zoom 配信

申込方法:参加登録は以下 URL からお願いします。ただし会場のスペースの関係から Zoom 限定での参加登録をお願いします。登録締め切り日 2月22日(木)

[藤田政之教授 最終講義・懇談会 参加登録 \(google.com\)](#)

藤田政之教授 最終講義プログラム(藤田研究室案内 HP より抜粋)

16:00-17:00

### 最終講義

司会:山内淳矢(東京大学)

専攻長挨拶:川嶋健嗣(東京大学)

タイトル: システム制御をアップデートする!

講師:藤田政之(東京大学)

17:00-18:00

### 懇談会

場所:東京大学本郷キャンパス工学部6号館3階セミナー室 AD

当日最終講義の前段で開催されるワークショップ「Fujita Festa」等のスケジュールの詳細は下記 URL をご参照ください。

[藤田政之教授 最終講義 - Fujita Laboratory \(u-tokyo.ac.jp\)](#)

以上

## ②『SICニュースレター「論説」集(第4巻)(2023年度掲載分)』発行・配布のお知らせ

2023年度(2023年1月～2023年12月)のSICニュースレターVol.5.1～Vol.5.12(Web発行)に掲載した「論説」、「新技術紹介」、非会員からの「寄稿」、および「コラム」を集約した冊子を1月22日に発行し、SIC正会員、準会員の代表者の方には既に2部配布いたしました。

入手(無料)希望の方は、SIC事務局までお申し込みください。

### 目次

#### 2023年 年頭の挨拶

SIC代表理事・センター長 浦川伸一氏(損害保険ジャパン株式会社)

#### I 論説

##### 論説1 横河電機の Digital Transformation(DX)への取組について

横河電機株式会社 常務執行役員(CIO)  
デジタル戦略本部長 兼 デジタルソリューション本部  
DX-Platform センター長 船生 幸宏氏

##### 論説2 量子コンピューティング2023

SCSK株式会社 事業投資推進本部  
本部長付 オープンイノベーション担当 杉坂浩一氏

##### 論説3 社会課題解決入門

株式会社ニューチャーネットワークス 代表取締役  
ヘルスケア AIoT コンソーシアム理事  
上智大学非常勤講師  
高橋 透氏

##### 論説4 複雑化する社会課題を解決するソーシャルデジタルツイン

富士通株式会社  
富士通研究所 コンバージングテクノロジー研究所  
ソーシャルデジタルツインコアプロジェクト  
シニアリサーチマネージャー 瀬川英吾氏

##### 論説5 東芝の DX 戦略実現に向けたスマートファクトリー化

株式会社東芝 生産技術センター グローバルモノづくり変革推進部  
部長 白須 義紀氏

#### II 新技術紹介

##### デジタル社会の実現に向けた最先端のトラスト技術

富士通株式会社 データ&セキュリティ研究所  
リサーチディレクター 鎌倉 健氏、 シニアプロジェクトディレクター 松塚 貴英氏  
シニアリサーチディレクター 藤本 真吾氏、 プロジェクトディレクター 須加 純一氏  
シニアリサーチマネージャー 引地 謙治氏、 プリンシパルリサーチチャー 中山 貴祥氏

#### III 寄稿(非会員)

##### 寄稿1 デジタル田園都市国家に向けてのシステム科学の貢献

株式会社日立製作所 研究開発グループ  
技師長 武田 晴夫氏

##### 寄稿2 大和ハウスの住宅系設計部門におけるデジタルソリューションの個人史

大和ハウス工業株式会社 技術統括本部 建設 DX 推進部  
担当部長 芳中勝清氏

##### 寄稿3 Trans-disciplinarity(TD) という考え方

特定非営利活動法人横断型基幹科学技術研究団体連合(横幹連合)  
会長 安岡善文氏

##### 寄稿4 日本における最新物流事情と日用雑貨業界の物流クライシス適応

サンスターグループ ロジスティクス研究室  
室長 荒木協和氏

#### IV コラム

コラム1 なぜ「システムマネージャー」が必要か？

東京大学・大阪大学名誉教授 木村英紀氏(SIC理事・副センター長)

コラム2 プラットフォームとシステム

東京大学・大阪大学名誉教授 木村英紀氏(SIC理事・副センター長)

#### V SIC役員一覧(2023年度)

#### VI SIC正会員一覧(2023年度)

A4約110ページ 2024年1月22日発行 無料

SICニュースレター「論説」集(第2巻)(2021年度掲載分)、(第3巻)(2022年度掲載分)も多少在庫があります。ご希望の方はSIC事務局にお問い合わせください。

## II 活動報告

### ① 2024年度第1回SICフォーラム開催案内

【日時】 2024年2月9日(金)15:00-16:15 (オンライン開催・会員限定)

【タイトル】 産業界での HPC(ハイパフォーマンスコンピューティング)活用促進に向けた取り組み  
～「富岳」と AWS を例に～

【講師】 株式会社理研数理 取締役 松崎 健一氏

(兼 株式会社 JSOL 未来共創デジタル本部 シニアスペシャリスト)

株式会社 JSOL エンジニアリング事業本部 材料・ライフサイエンス部長 小沢 拓氏

【申込方法】 参加申込はこちらから

SICイベント参加登録ページ

( <https://sysic-org.sakura.ne.jp/SICregistration.html> )

の「SICフォーラム(2024年度第1回)」の欄から申し込みください

#### 【講演概要】

JSOL 及び子会社の理研数理では、数理科学分野のアカデミアと産業界の協業によるイノベーション創出を目指した取り組みを推進しています。

本講演では、これらの取り組みの中から、「富岳」を産業界で使用するサービスメニューや事例、更には「富岳」の成果を AWS(Amazon Web Services)上で同じように利活用できるサービスメニューをご紹介します。また、これらのサービスと連携する、分子動力学領域(特に材料・ライフサイエンス)のパッケージソフトもご紹介します。

#### 【参考】

株式会社 JSOL (SIC正会員)

<https://www.jsol.co.jp/index.html>

株式会社理研数理(JSOL の子会社、理化学研究所との共同出資会社)

<https://www.riken-suuri.jp/>

以上

## ② 2024. 1. 24 15:00-17:00 2024年度第1回実行委員会開催報告

開催形式：MS-Teams によるオンライン開催

出席者数： 実行委員12名、副センター長・監事・事務局各1名、総出席者数15名

### 議題

司会 松本隆明実行委員長

#### 1. 報告事項

##### 1.1 2024年度第1回SIC戦略フォーラム開催案内

久保忠件事務局次長

「産業界での HPC(ハイパフォーマンスコンピューティング)活用促進

に向けた取り組み ～「富岳」と AWS を例に～」

2月9日(金)15:00-16:15 (オンライン開催)

##### 1.2 2023年12月19日開催の2023年度第2回理事会報告

および2023年度定例社員総会の開催時期について

出口光一郎事務局長

2023年度定例社員総会は2024年3月5日(火)開催予定

##### 1.3 戦略提言に向けた SWG の活動状況報告

- ・ 防災・レジリエンス SWG 報告((代)久保忠伴氏)
- ・ ロジスティックスSWG報告(藤野直明氏)
- ・ エネルギーSWG報告(船橋誠壽氏)
- ・ 金融 SWG(藤井紳也氏)
- ・ 科学技術 SWG、ヘルスケア SWG は提言書の最終確認の段階(木村英紀氏)

#### 2. 協議事項

##### 2.1 ウラノスエコシステム推進センター設立について

松本隆明実行委員長

浦川伸一センター長との面談結果を受けて意見交換

#### 3. その他事項

##### 3.1 準会員の募集状況

久保忠件事務局次長

2社準会員に入会予定

##### 3.2 2024年度実行委員メンバーの確認

同上

次回理事会までにフィックスし承認を得る

##### 3.3 SIC個人会員対話会開催について

木村英紀副センター長

1月31日(水)17:00-18:00 リモートにて開催予定

#### 次回、次々回の実行委員会開催予定日時

2024年度第2回実行委員会 2月20日(火) 15:00-17:00

2024年度第3回実行委員会 3月19日(火) 15:00-17:00

## Ⅲ 会員一覧

新規入会準会員(インキュベーション会員) 電腦バンク株式会社 <https://www.dn-bank.com/>

### 正会員

SCSK株式会社	NTTコムウェア株式会社
株式会社NTTドコモ	株式会社クエスト
株式会社構造計画研究所	株式会社JSOL
株式会社テクノバ	株式会社東芝
株式会社ニューチャーネットワークス	株式会社野村総合研究所
株式会社日立国際電気	株式会社日立産業制御ソリューションズ
株式会社日立システムズ	株式会社日立製作所 研究開発グループ 社会システムイノベーションセンター
損害保険ジャパン株式会社	東京ガス株式会社
東京電力パワーグリッド株式会社	日鉄ソリューションズ株式会社
日本郵船グループ株式会社MTI	ファナック株式会社
富士通株式会社	マツダ株式会社
三菱電機株式会社	横河電機株式会社
ロジスティード株式会社(旧日立物流株式会社)	

### 準会員

電腦バンク株式会社(準1)	三菱重工業株式会社 デジタルイノベーション本部(準2)
---------------	--------------------------------

(準1)インキュベーション会員、(準2)人財育成限定会員

(2024年2月1日現在:五十音順)

©SIC 2024.2

発行者: 一般社団法人システムイノベーションセンター(SIC)  
代表理事・センター長 浦川伸一  
編集者:SIC 実行委員 中野一夫 (構造計画研究所)  
事務局 〒160-0023 東京都新宿区西新宿 6-12-7 ストック新宿 B-19 号  
URL: <https://sysic.org> E-mail: [office@sysic.org](mailto:office@sysic.org) Tel.Fax:03-5381-3567