



K P M G N e w s l e t t e r

KPMG Insight

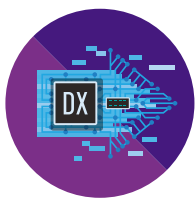
Vol.
52
January
2022



Digital Transformation

VUCA 時代に不可欠な確率的ビジネス分析





Digital Transformation

VUCA時代に不可欠な 確率的ビジネス分析

株式会社 KPMG Ignition Tokyo
DXソリューション2
マクシミリアン・シュミット/データ&アナリティクスリーダー

この10年間の機械学習とAIのブームは、コンピュータビジョン、自然言語処理、予測モデルなど多くの分野で驚くべき成果をもたらし、ビジネスコンサルティングの重要な手法としてのデータサイエンスの到来を後押ししました。しかし、単にデータサイエンスを用いただけでは、データやビジネスプロセスの不確実性の影響を見落とししたり、過小評価する可能性があります。それらを考慮することは、分析の信頼性を向上させるだけでなく、意思決定に伴うリスクを調整するためにも必要なことです。本稿では、VUCA時代におけるビジネス課題を解決するデータサイエンス手法の1つとして、因果推論(因果効果)、確率的プログラミング(統計的モデリング)、ベイズ意思決定(最適意思決定)を結びつける原理に基づいたワークフローをご紹介します。

なお、本文中の意見に関する部分については、筆者の私見であることをあらかじめお断りいたします。



マクシミリアン・
シュミット
Maximilian Schmidt

POINT 1

データサイエンスの到来

近年の機械学習とAIの進化は、ビジネスコンサルティングの手法としてのデータサイエンスの到来を後押ししている。

POINT 2

原理に基づくワークフロー

因果推論(因果効果)、確率的プログラミング(統計的モデリング)、ベイズ意思決定(最適意思決定)を結びつける原理に基づいたワークフローを活用することで、複雑なビジネス課題に対処することができる。

POINT 3

因果推論はビジネス分析に不可欠

因果推論は、行動の効果を推定する。そのため、意思決定に大きく影響を与えるビジネス分析において、因果推論はきわめて重要である。

POINT 4

ベイズ意思決定による戦略的意思決定の導出

ビジネス課題の不確実性を取り込み、選択された行動によって引き起こされるさまざまな可能性を考慮することで、複数の選択肢を得ることができる。

I ビジネス課題を解決するデータサイエンス

データサイエンスのプロジェクトは、通常、大規模なデータセットの中から関連するパターンを発見・予測し、問題を解決するという従来の機械学習に焦点を当てています。代表的なのが、過去に蓄積したデータを基に製品の販売予測を行うなどです。

しかし、実際のビジネスの課題は、従来の機械学習で解決できるものばかりではありません。売上げを伸ばすには、どの製品の価格をどのように設定すればいいのか。サプライチェーンの最適化はどうしたらいいのか。どうすれば顧客の減少を防ぐことができるか。価格設定やプロセスの最適化、顧客サービスなど、さまざまな課題を抱えています。

このような課題に対して最適解を得るには、因果推論を用いると効果的です。因果推論とは、行動(介入)と結果の間の因

果効果を推定することで、従来の機械学習よりも高精度の推測を行う手法のことです。

II 原理に基づくワークフロー

因果推論を用いたワークフローは、「因果推論による問題定式化」「確率的プログラミング¹によるモデル構築」「ベイズ意思決定²による戦略的意思決定の導出」の3つのフェーズから構成されます(図表1参照)。

フェーズ1では、まず因果推論を使用して問題を定義し、ドメインの専門知識を活用して問題内の変数間の因果関係について仮説を立てます。これにより、機械学習を用いて推定可能な関係式³が得られます。フェーズ2では、フェーズ1で得られた関係式を推定するモデルを、モデルに不確実性を組み込むことを可能にする確率的プログラミングを用いて構築します。

最後のフェーズ3では、フェーズ2で構築

した推定モデルをベイズ意思決定で用いて、クライアントと定義した目標に基づいた最適な戦略的意思決定を導出します。

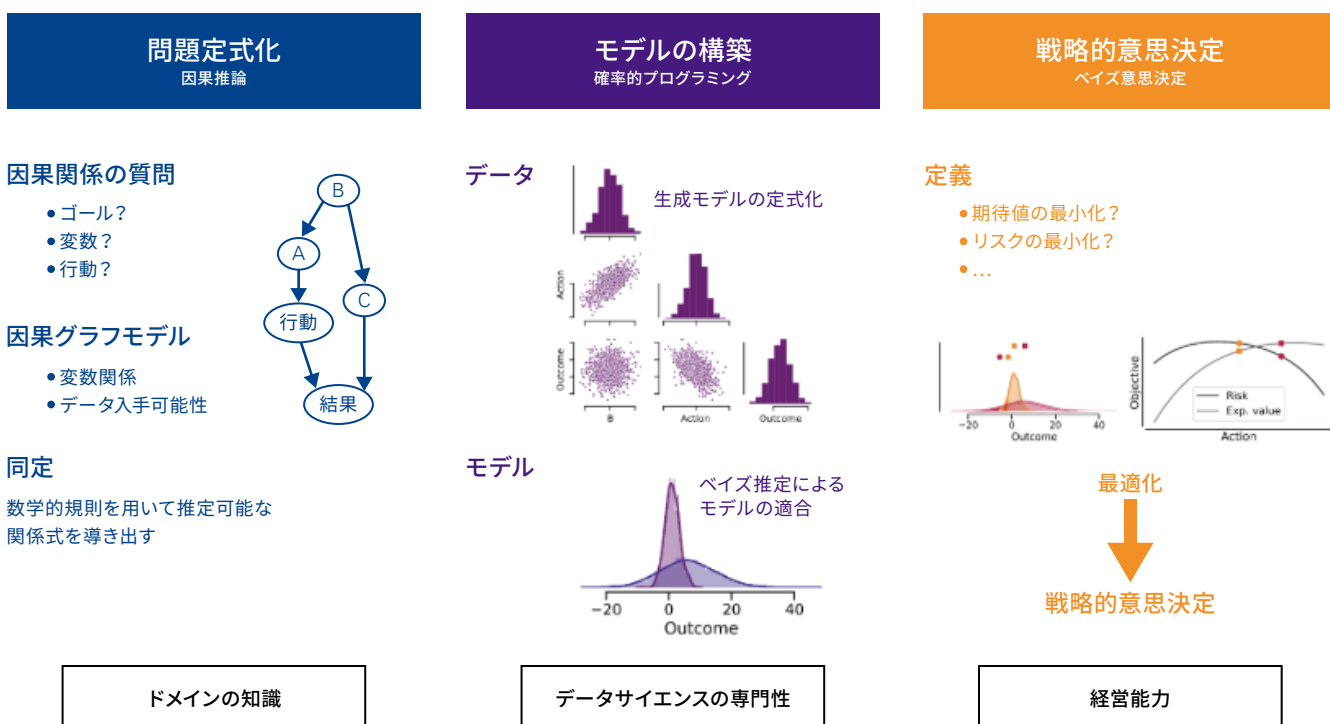
1. 因果推論による問題定式化

まず、解決すべき問題を定義します。問題はたとえば、システム内の変数は何か? どのようなアクションを取ることができるか? どの変数がクライアントによりコントロール可能か? そして、これら問題に対するゴールを、interventional distributionと呼ばれる特定の行動が選択された場合の結果の確率とします。

次に、クライアントのドメインの専門知識を活用して、システム内の変数間の因果関係に関する仮説を立てます。どの変数がどの変数に影響を与えるかを推論するというわけです。最終的に、これら仮説は因果グラフモデル(CGM)⁴というグラフ表現に要約されます。

ところで、interventional distributionを直接推定するには実験を行い、異なる行動

図表1 確率的ビジネス分析の原則的なワークフロー



出所: KPMG 作成

の影響をテストする必要があります。しかし、ビジネスのコンテキストで実際に実験するのは容易ではありません。そこで最後のステップで、interventional distributionを観測された(過去の)データのみから推定できる観測分布⁵に変換します。このステップは同定と呼ばれるもので、do-calculusを使って自動化されています。do-calculusとは、CGMを使って変換が可能かどうかを調べ、可能であれば観測分布を導出する数学的規則の集合のことです。

2. 確率的プログラミングによるモデル構築

フェーズ2は、フェーズ1で同定した推定モデルを構築します。具体的には、データがどのように生成されたかを推測し、それを確率的プログラムとして定式化します(決定論的)。

決定論的なコンピュータプログラム同

様、確率的プログラムのアルゴリズムは出力値を計算します。しかし、変数は確率的な、実際の過程に関する不確実性とデータの不確実性を捕捉する確率分布から導かれます。こうして構築された推定モデルは、フェーズ1で定義した問題、特定の行動を選んだ場合の結果の確率の推定値を導き出します。

3. ベイズ意思決定による戦略的意思決定の導出

ビジネスでは、出力変数の目的関数⁶の最適化のように、行動に対する最適な意思決定を求めようとします。ここでの目的関数とは、経営上の意思決定を反映したものです。したがって、クライアントに対して、相反するような複数の選択肢を提示することになります。たとえば、期待される結果を最大化することでよりリスク許容度を高くする選択肢と、不良な結果のリス

クを最小化することでより保守的になる選択肢といった具合です。

ベイズ意思決定では、問題の不確実性を取り込み、選択された行動によって引き起こされる可能性のあるすべての結果を考慮します。その結果、ビジネス課題に対して複数の選択肢を得ることができます。

III

ビジネス分析に因果推論は欠かせない

統計学では、観測データ中の誤った相関が全体像を歪めることがあることから、相関は因果関係ではないと言われていま。これは、状況によっては真実です。その状況とは、たとえば行動と結果の間の相互作用が、両方に影響を及ぼす追加的な要因によって重ね合わされるような場合です。

そのような場合でも、因果推論は有効

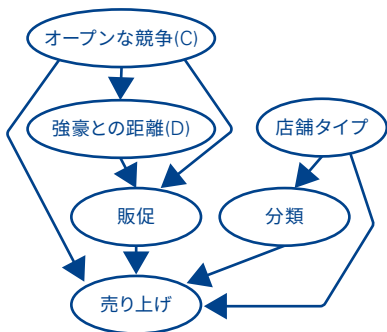
図表2 販促による売上への因果効果を検出するために適用された原則ワークフロー



因果関係の質問

$$p(\text{Sales}|\text{do}(\text{Promo}))=?$$

因果グラフモデル

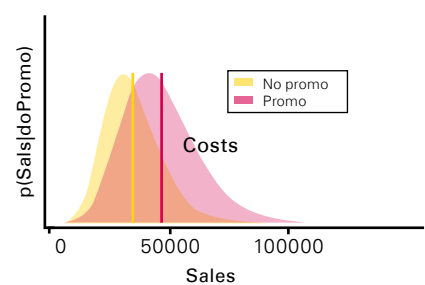
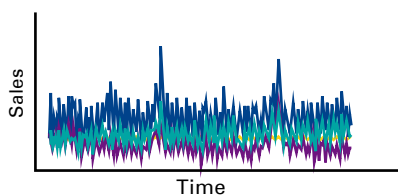


同定

$$p(\text{Sales}|\text{do}(\text{Promo})) = \sum_{C,D} p(\text{Sales}|\text{Promo}, C, D)p(C, D)$$

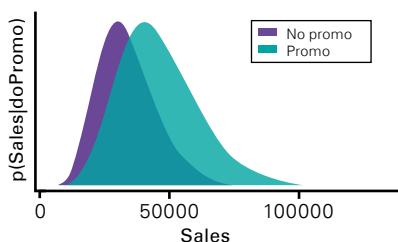
出所: KPMG 作成

データ



期待値：
販売促進による35%の売り上げ増加

モデル



です。それは、因果推論が変数間の真の因果効果を同定し、定量化を可能にするからです。従来の統計学や機械学習が連想関係に焦点を当て、「Yの値がわかっている時、Xについて何が言えるか？」のような疑問を呈するのに対し、因果推論は「YはどのようにしてXを引き起こすのだろうか?」や、「Yを値に設定すると、Xはどうなるか?」といった質問を投げかけ、行動の効果を推定します。したがって、意思決定に大きく影響を与えるビジネス分析において、因果推論はきわめて重要と言えるでしょう。

ここでは、過去に行った販売促進キャンペーンによって、その店舗の売上げが増えたかどうかを知りたいというケースで考えてみます(図表2参照)。

ワークフローのフェーズ1では、店舗近隣にある競合他社の存在が、過去にプロモーションを行うという決定に影響を与えていたことがわかります。近隣店舗からの

競争圧力がある場合、一般的に販売促進キャンペーンの実施によって売上げ増を試みる傾向があります。また、競争の激しい地域にある店舗は、他の追随を許さない店舗よりも売上げが低いという可能性も考えられます。

このように、競合他社の存在は販売促進キャンペーンを実施するかどうかの決定と販売の両方に影響を与えます。したがって、販売促進キャンペーンの販売に対する真の因果効果を見つけるには、分析において競合の存在を調整する必要があります。これは、販売に対する販売促進キャンペーンの因果効果の偏りのない視点を得るためです。

問題を体系的に定義することにより、競争が分析に与える潜在的なバイアス効果が明確で扱いやすくなるというわけです。このように、本ワークフローでは変数とその関係を明示的に定義し、分析とクライアントとのコミュニケーションの両方を容易

にします。

また、因果推論では観測されたデータを使って推定できる値を特定するので、闇雲に実験を行うようなことはありません。

IV 確率論的モデリングは、より多くの情報に基づいた意思決定を可能にする

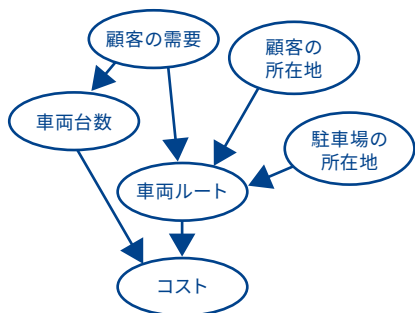
ビジネスは不確実な世界にあります。顧客の需要は時間の経過とともに変化し、取引には商品の欠陥やサプライチェーンの不安定といったリスクを伴います。また、自然災害やパンデミックのような壊滅的な出来事が発生すれば、世界全体の経済状況も激変します。世界で起きているさまざまな要因が、ビジネスに多大な影響を及ぼすのです。そのため、ビジネスの問題をモデル化する際には、これら不確実性を考慮する必要があります。

図表3 サプライチェーン内の最適な車両数を特定するために適用される原理ワークフロー



因果関係の質問
 $p(\text{Costs}|\text{do}(\text{VehicleNumber}))=?$

因果グラフモデル

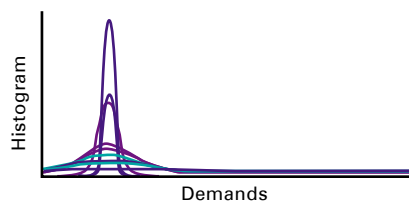


同定

$$p(\text{Costs}|\text{do}(\text{VehicleNumber})) = \int_D p(\text{Costs}|\text{VehicleNumber}, D)p(D)$$

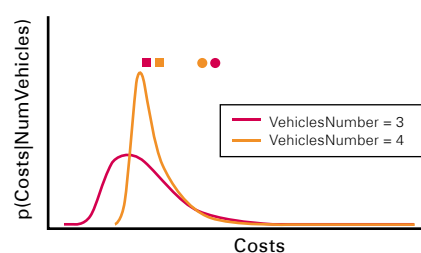
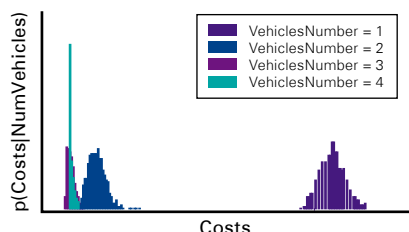
出所: KPMG 作成

データ



コスト = 車両の固定数 + ペナルティ料金
 より多くの車両 = より高い固定費
 より少ない車両 = より高いペナルティ料金

モデル



最適な車両台数は?
 3台: より低いコスト期待値
 4台: 高コストになるリスクがより低い

確率的プログラミングのフレームワークは、単純かつ体系的な方法でこれを行うことを可能にします。それは、問題における行動のための異なる選択に基づいた、起こり得る結果のための確率的予測を叩き出します。これにより、期待される結果を最適化するのか、それともより保守的にリスクを抑えて望ましくない結果が生じる可能性を最小限にするのか、というように目標を柔軟に定義することができます。

ここでは、サプライチェーンの運営に最適な台数の車両を選択するというケースについて考えてみます(図表3参照)。

クライアントは、将来の顧客の需要を満たすために十分な台数の車両を確保したいと考えていますが、その一方でメンテナンスコストを最小化したいとも思っています。この相反する利害のバランスを取るには、顧客の需要を十分に予測し、一定数の車両が存在する場合のコストの可能性を予測する必要があります。そこで、推定モデルを構築してコストの見積りを割り出し、最適な車両台数を計算します。このケースでは、3台の車両を使用すると、メンテナンスコストを節約することによってコストの期待値が低くなり、4台の車両を使用すると、台数の増加によって顧客の需要を満たす可能性が高くなることから、高コストになるリスクがより低くなります。つまり、クライアントはコストの期待値を最小化することによってリスク耐性を高めるか、高コストに直面するリスクを最小化して保守的になるか、そのどちらかを選択することができるというわけです。

本ワークフローが導出したこの結果は、私たちのアプローチが可能にする戦略的な選択です。将来のためにたった1つの起こり得るシナリオに焦点を当てるのではなく、すべての可能な未来と関連する確率を考慮することによって、これら複数の選択を明確にすることができます。その結果、クライアントにはより多くの情報に基づいた見解を提供することができます。

V 結論

本稿では、因果推論、確率的プログラミング、ベイズ意思決定の方法を結びつける原理に基づいたワークフローを紹介しました。このワークフローにより、アクションが結果に及ぼす因果関係の影響の特定、将来の展開に関する柔軟で正確な予測の作成、それら予測に基づいた戦略的な意思決定など、複雑なビジネス課題に対処することが可能となります。

- 1 データをコンピュータプログラムとして定式化することにより、データの生成モデルを構築する方法のこと。
- 2 事前の知識とデータ駆動型モデルの両方の情報によってもたらされる良い結果が起こる可能性に基づいて意思決定を行うプロセスのこと。
- 3 観測データのみから推定できる、能動的な実験を行う必要のない数式のこと。
- 4 変数間の因果関係をグラフ化したもの。変数はノードであり、因果関係は変数間の矢印で表される。
- 5 他の変数の値が観測された場合の事象の確率を定量化する確率分布のこと。
- 6 「売上高の期待値を最大化したい」などの大まかな目的を数学的に定式化したもの。

関連情報

ウェブサイトでは、クライアント企業のデジタル化に伴うビジネス変革情報を紹介しています。

home.kpmg/jp/KIT

本稿に関するご質問等は、以下の担当者までお願いいたします。

株式会社 KPMG Ignition Tokyo
DX ソリューション 2
マクシミリアン・シュミット / データ & ナリティブスリーダー

✉ maximilian.schmidt@jp.kpmg.com

KPMG ジャパン

marketing@jp.kpmg.com

home.kpmg/jp

home.kpmg/jp/socialmedia



本書の全部または一部の複写・複製・転載および磁気または光記録媒体への入力等を禁じます。

ここに記載されている情報はあくまで一般的なものであり特定の個人や組織が置かれている状況に対応するものではありません。私たちは、的確な情報をタイムリーに提供できるよう努めておりますが、情報を受け取られた時点及びそれ以降においての正確さは保証の限りではありません。何らかの行動を取られる場合は、ここにある情報のみを根拠とせず、プロフェッショナルが特定の状況を綿密に調査した上で提案する適切なアドバイスをもとにご判断ください。

© 2022 KPMG AZSA LLC, a limited liability audit corporation incorporated under the Japanese Certified Public Accountants Law and a member firm of the KPMG global organization of independent member firms affiliated with KPMG International Limited, a private English company limited by guarantee. All rights reserved. Printed in Japan.

© 2022 KPMG Tax Corporation, a tax corporation incorporated under the Japanese CPTA Law and a member firm of the KPMG global organization of independent member firms affiliated with KPMG International Limited, a private English company limited by guarantee. All rights reserved.

The KPMG name and logo are trademarks used under license by the independent member firms of the KPMG global organization.