



KPMG Newsletter

KPMG Insight

 Digital Transformation

意思決定の仕方を機械は学べるか
～長期的な便益を最大化する強化学習の手法～



Vol. **66**

May 2024



Digital Transformation

意思決定の仕方を機械は学べるか

～長期的な便益を最大化する強化学習の手法～

KPMG FAS

田中 勇輝 / マネージャー

KPMG アドバイザリー ライトハウス

小澤 友美 / シニア データサイエンティスト

「強化学習」は、囲碁で世界一のプレーヤーに勝利を収めたAlphaGoや、大規模言語モデルであるChat GPT¹の開発にも使われている機械学習の手法の1つです。強化学習は、これまでは主に限定的な領域に適用されてきましたが、今後は多方面のビジネス領域に広がっていくと考えられます。特に、データドリブンでビジネス改善を推進する際、この手法の重要性はより増していくと思われます。

本稿では、まずビジネスにおいて従来型の機械学習モデルを活用する際に陥りがちな課題について説明します。続いて、その課題を解決する手法としての強化学習の要諦と、人間の意思決定の学習プロセスを模した特徴的な学習の仕方を概説します。そのうえで、実際にホテルの客室の価格設定に強化学習を適用した実証例とその検証結果を紹介し、最後に強化学習のビジネスへの活用を今後加速させると考えられるシミュレータの必要性とその効果を解説します。

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。



田中 勇輝
Yuki Tanaka



小澤 友美
Tomomi Ozawa

POINT 1

従来の機械学習モデルの落とし穴

ビジネスの意思決定は、多岐にわたる膨大な情報に基づいた判断が求められる。回帰分析などの従来の機械学習モデルは、限定的な状況を基にした単発的な打ち手に留まり、長期的には必ずしも最善ではない可能性がある。したがって、より包括的に状況を加味したデータサイエンスの手法が求められる。

POINT 2

強化学習を活用した課題解消への試み

強化学習は、一定の期間内におけるベネフィットが最大限に達成されるように意思決定の仕方を学ぶ手法であり、単発的な改善とは根本的に異なる。具体的には、特定の状況における意思決定とその意思決定に対するフィードバックを受け取ることを繰り返して、何が成功で何が失敗かを学習する。

この手法を、ホテルの客室の価格設定に適用・検証した結果、売上を最善となる代替アプローチよりも最大5%増加させることができた。これにより、強化学習がビジネス領域においても効果があることが確認された。

POINT 3

学習に不可欠な環境を提供するシミュレータの必要性

現在、適切な学習環境を提供するシミュレータ構築が技術的に進歩してきている。シミュレータが構築できると、強化学習モデルを学習させる効果的な訓練環境となるデータを提供できるようになる。これは、強化学習をビジネスに応用する際に直面していた学習データの準備の課題が克服できることを意味し、強化学習の活用を促進すると考えられる。

I 従来型データサイエンス手法のさらなる先へ

近年、ビジネスの特定機能の支援に特化したデータサイエンスの活用や意思決定の高度化の事例が数多く見られます。たとえば、販促においてはECサイトにおける商品レコメンドのアルゴリズム、物流領域においては最適ルート探索や最適な運行スケジュールの作成などです。これらの領域では、データサイエンスの活用がいわば高度にビジネスの意思決定と絡まっており、担当領域の責任者は適切にデータサイエンスを活用できれば大きな武器を手にすることができます。

しかし、このようなデータサイエンスを使ったビジネス改善にはいくつかの落とし穴があり、それは理解されないまま利用されているか、見過ごされているようにも見受けられます。それゆえ、適用する手法の限界や課題を見極めることは非常に重要となります。ここでは簡略的にいわゆる回帰分析などの「従来型の機械学習モデル」に焦点を当てて、その課題を見ていきます（図表1参照）。

まず1つ目の課題は、従来型の機械学習モデルが出す最適解は「近視眼的」になっている、あるいはその傾向が高くなっているということです（図表2参照）。機械学習モデルは、基本的に過去データのパターンを学習することから、常に現時点にすぐ近い将来に対する予測や解が最も精度が高くなります。そのため少し先を見据えた判断というのは苦手な傾向にあります。

2つ目に、従来型の機械学習モデルは単発的で特定タイミングにおける最適解を導出しがちで、一定の期間における成果が最大となるような連続性を加味した判断というものも不得意としています。このような傾向性は、その場しのぎの判断となりやすいともいえます。

最後に3つ目の課題として、個別の領域でみると局所的なメリットはあっても、別の大きな問題を生んでしまうというケース

も見受けられます。たとえば、利益を最大化するように広告や販促活動を最適化していたはずが、リピートする可能性の低いワンタイム顧客の利用を増やしてしまい、長期的に見れば利益が目減りしたなどです。

より長期的なビジネス改善を目指し、より複合的な意思決定をしていきたい担当責任者にとって、このような課題を残したままの利用は最善とはいえません。このような特定のタイミングにおけるスナップショットの解ではなく、できる限り包括的に状況を捉えた意思決定をデータサイエンスにより強化することは可能でしょうか。注目の「強化学習」という分析アプローチは、まさにこのような問題に1つの答えを提供するアプローチだといえます。

II 意思決定の仕方を学ばせる強化学習の要諦とその効果

強化学習をあえて一言で述べるならば、「ある問題に対して長期的な便益が最大化されるように最善の意思決定の仕方を学ぶ手法」です。この手法は日々異なるビジネスの状況を踏まえ、何が失敗で、何が成功かを学習し、そのたびに何が最適かを考える人間の意思決定の学習プロセスを模しています。

ここでは、重要となるキーワードの定義をしながら、強化学習の仕組みを説明します（図表3参照）。まず強化学習では、行動や意思決定をする主体をエージェントと呼びます。このエージェントが、判断すべき内容であるタスク（環境）について意思決定を行い、それに対するフィードバックを報酬として受けることを繰り返し

図表1 機械学習モデルの全体簡略図



出所：KPMG作成

図表2 従来型の機械学習モデルの課題

課題	詳細
近視眼的	直近の精度は高いものの、遠い将来になるほど解の信頼性が下がりがやすい
単発的	継続的に連続した解ではなく、1つのタイミングにおける最適解になりがち
局所的	個別最適な解となる傾向が強く、全体最適が損なわれやすい

出所：KPMG作成

て、意思決定のルールを学習します。より詳細には、エージェントは状態と呼ばれる現在の状況を表す情報を把握したうえで意思決定を行い、その状態における行動の結果として報酬や新しい状態を引き出し、再度意思決定を行う、という過程を繰り返します。こうしてエージェントが行動する環境を探索し、意思決定とその結果のデータを経験として学習することで、意思決定ルール（方策）を改善していくというわけです。

たとえば、ホテルにおける価格設定でいえば、ホテルにおける状態は現在の空室状況ならびにその時点での部屋価格、意思決定となる行動は部屋の価格設定、報酬はその価格で予約された部屋数に応じた収入となります。このプロセスにおいてエージェントは、報酬である収入が最大化されるような価格方針の学習を目指します。

ここで着目したいのは、強化学習の目的が、報酬の累積値を最大にする意思決定ルールを獲得することだという点です。つまり、短期的には報酬が低く損失になるとしても、長期的には報酬の累積が最大になることを目指すというわけです。具体的なビジネスの文脈で言えば、一度のアクションで得られる利益が低くとも、継続的にアクションを見直すことで利益の累積が最大になるような戦略を学習するとい

うことです。このように、累積の報酬を最大化するために意思決定ルールを学習するという仕組みは、従来型の機械学習モデルとは異なる強化学習の特徴です。そして、これが従来型の機械学習モデルを活用した際の課題への1つの改善の方向性を示すものと考えられます。

強化学習は、最近では大規模言語モデルであるChat GPTの開発にも使われている学習アルゴリズムとして注目を集めている手法です。適用には難しさがあるものの、今後はより広範な範囲で徐々に広がっていくと考えられています。これまでは、ビジネス領域外ではチェスや囲碁のような複雑なゲームで人間の能力を超える結果を出す事例、ビジネス領域では自動運転やロボット制御などにおける適用例がほとんどでしたが、最近ではeコマースにおける価格設定の事例など、ビジネスのオペレーショナルな問題にも適用されるケースが出てきました。このような状況を踏まえ、ビジネスのより複雑な意思決定が必要な領域においても、その効果を検証するような状況や機運が高まってきたと考えられます。

実際、強化学習を効果的な形でビジネスに適用することが可能なかを確認するために、KPMGではホテルの客室の価格設定に強化学習を適用し、客室予約による売上の改善効果の検証を行いました。

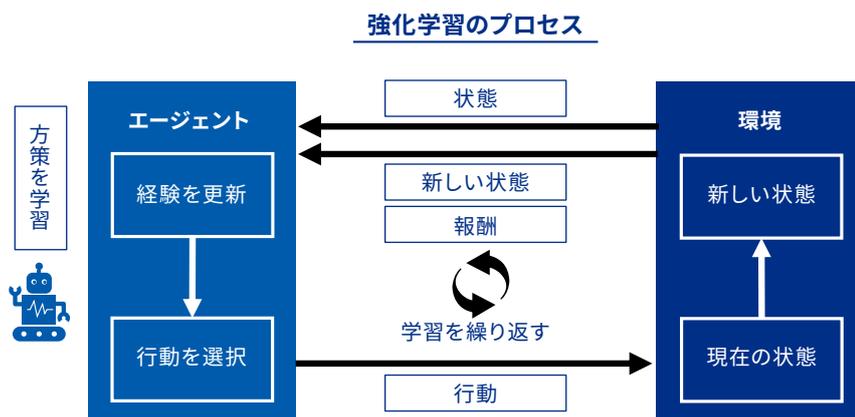
簡易化のため一定の仮定の下、客室の予約受付開始日から宿泊当日まで、空室状況ならびにその時点での客室価格に応じて、一定の頻度で客室価格を強化学習モデルが設定するような設定です。ゴールは、対象の宿泊日の予約がもたらす累積売上を最大にする価格設定戦略の学習です。学習には、DQN (Deep Q-Network) という強化学習にディープラーニングを適用した手法の祖ともいえるアルゴリズムを用いました。結果は、DQNを適用して訓練したエージェントのほうが、顧客予算を反映している固定価格よりも最大5%の売上増加を見込めることが実証できました。これは、強化学習モデルが、さまざまな価格帯とその際の客室売上を効率的に学習し、最終的に上記の固定価格よりも売上を増加させる価格とその提示タイミングを特定できるようになったことを意味します。この実証結果から、ビジネス領域における強化学習の適用は現実的になったと言えます。

III 学習に不可欠な環境を提供するシミュレータの必要性

ここまで、多くの強化学習のメリットとその効果を述べてきましたが、現在はまだ主要なアプローチとは言えません。その理由は、ディープラーニングなどと同様、学習の仕組みや背景がブラックボックス化されてしまうこと、技術面で強化学習の報酬の設定の仕方やタスクの定義方法などの難しさが挙げられます。これらに加え、ビジネスへの活用という観点で強化学習を妨げてきた最も大きな理由は、学習データの準備の難しさにあります。

機械学習モデルを訓練するには、一般に相当量のデータを要します。なかでも強化学習の枠組みでは、エージェントが環境のなかで意思決定をし、報酬または罰則の形でフィードバックを受けることによって意思決定の仕方を学習することから、膨大なデータサンプルを必要とし、最適

図表3 強化学習において、エージェントが環境と相互作用して状態、行動、報酬をやり取りする様子



出所：KPMG作成

な性能を達成するためには数十万回規模の学習が欠かせません。そのためには適切な訓練環境が必要になりますが、そこにも課題があります。理想的な訓練環境は、実際のビジネスに影響を与えることなしに、エージェントが自由に試行錯誤をして学習できることですが、現実世界でそのような環境を用意することは不可能に近いでしょう。

この課題を克服するためには、強化学習エージェント用の安全で効果的な訓練環境を提供するシミュレータを開発することが有用です。近年、この技術の発展から強化学習モデルを学習させやすくなることが見込まれており、実際に上述したホテルの価格設定の検証にも、この手法を適用しました。

シミュレータは、現実に即した形である程度詳細を簡略化しながら、eコマース、接客業、食料品販売業などのさまざまな業界の状況を仮想的に作り上げ、強化学習でエージェントを訓練するための自由度の高い環境を提供します。作り方としては、ベースとなるシミュレータに対して、領域における専門知識と利用可能な情報（商品ラインナップ、マーケティング施策、顧客行動特性など）など購買特性を反映するよう調整し、業界の販売環境を正確に再現するようなシミュレータに仕立て上げます。また、各商材の性質によって異なる

シナリオを想定することも可能です。たとえば、賞味期限のある食品には期間間に頻繁に割引が行われるという傾向を、電化製品などには新しいモデルが導入されれば流行遅れになるという傾向を、比較的自由に反映することができます。

このように作り上げたシミュレーション環境で強化学習エージェントが学習し、その結果得られた行動の結果は、シミュレータをさらに改善するために使用され、結果としてエージェントの学習の改善をもたらします。現実世界とシミュレータからのフィードバックの下で、エージェントの意思決定は絶えず改良されます。エージェントとシミュレータが運用している間、常にシミュレータ自体も改良されていくのです（図表4参照）。

このシミュレータの作り込みが容易になればなるほど、強化学習という手法も活用が進み、回帰分析のような機械学習モデルや数値最適化など比較的トラディショナルな手法の一步先に進めることが期待できます。

IV

さいごに

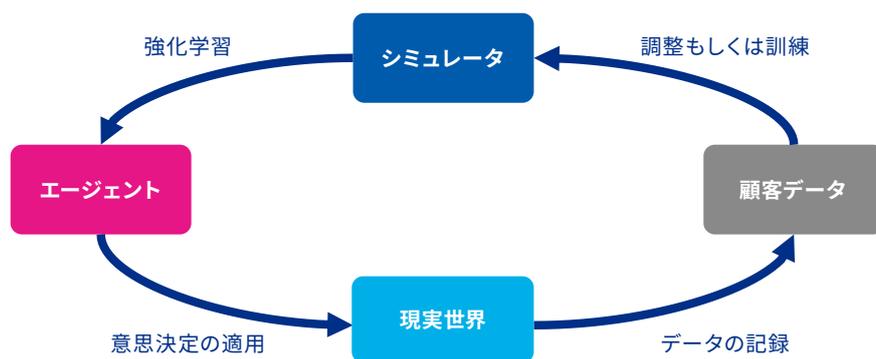
本稿では、機械学習モデルのなかでも先進的な強化学習を取り上げました。比較的トラディショナルなアプローチは長期

目線に立って意思決定をするのが難しいですが、強化学習には実際の人間のような意思決定を学ぶという仕組みがあります。強化学習は、現在もさまざまな領域で発展し続けており、新しいアルゴリズムやアプローチの開発も進んでいます。これからのビジネスにおける意思決定におけるデータサイエンス活用の1つの有効なアプローチとして、強化学習は今後、積極的に検討されていくことでしょう。

- 1 ChatGPTはOpenAI社の商標です。
- 2 ここで説明する設定は、厳密には「オンライン強化学習」と呼ばれています。

図表4 ダイナミックプライシングシミュレータ

シミュレータを用いて訓練したエージェントが、現実のビジネスで意思決定を行い、その結果として収集されたデータを活用することによりシミュレータは改良される



出所: KPMG作成

関連情報

ウェブサイトでは、プライシング（価格戦略）の最適化サービスを紹介しています。

<https://kpmg.com/jp/ja/home/services/advisory/deal-advisory/strategy/cbma/pricing.html>

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 FAS

田中 勇輝 / マネージャー

✉ yuki.a.tanaka@jp.kpmg.com

KPMG ジャパン

kpmg.com/jp



本書の全部または一部の複写・複製・転記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す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ける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。

© 2024 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.

© 2024 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.

コピーライト©IFRS®Foundation すべての権利は保護されています。有限責任 あずさ監査法人は IFRS 財団の許可を得て複製しています。複製および使用の権利は厳しく制限されています。IFRS 財団およびその出版物の使用に係る権利に関する事項は、www.ifrs.org でご確認ください。

免責事項：適用可能な法律の範囲で、国際会計基準審議会と IFRS 財団は契約、不法行為その他を問わず、この冊子ないしあらゆる翻訳物から生じる一切の責任を負いません（過失行為または不作為による不利益を含むがそれに限定されない）。これは、直接的、間接的、偶発的または重要な損失、懲罰的損害賠償、罰則または罰金を含むあらゆる性質の請求または損失に関してすべての人に適用されます。この冊子に記載されている情報はアドバイスを構成するものではなく、適切な資格のあるプロフェッショナルによるサービスに代替されるものではありません。

「IFRS®」、「IAS®」および「IASB®」は IFRS 財団の登録商標であり、有限責任 あずさ監査法人はライセンスに基づき使用しています。この登録商標が使用中および（または）登録されている国の詳細については IFRS 財団にお問い合わせください。