

次世代統合プラットフォーム MSIPによるデータ活用 —ビジネスに数理科学を—

小木 しのぶ

NTT データ数理システムが提供する MSIP は、データ活用の処理をビジュアルにワークフローの形で構築でき、PoC（Proof of Concept：実現可能かどうかの検証）からデプロイまで実現できる統合プラットフォームである。数理科学的手法を多数搭載し、ビジネスにおける課題解決を効果的に行えるだけでなく、公開されたインターフェースを用いてユーザーが機能追加も可能なオープンプラットフォームを目指している。MSIP は現在開発中であり 2021 年 9 月にリリースを予定しているが、本稿ではそれに先立ち詳細を紹介する。

キーワード：統計解析，機械学習，数理計画，統合プラットフォーム，デプロイ，オープン化

1. ビジネスの世界での数理科学

本稿の主題である統合プラットフォームの議論に入る前に、数理的技法をビジネスの世界で活用するにあたって留意すべき点を挙げる。

1.1 オーダーメイドモデリング

複雑・精緻なモデルは、恰好は良いが、それが常にビジネスの世界で実用的であるとは限らない。

ビジネスにおいてよく利用される数理的技法には、統計解析や機械学習といったデータから情報を導き出すデータ分析、ルールからその条件にあった計画を導き出す最適化、条件やその動きを定義することで実際の影響を観測するシミュレーションなど、さまざまな技術がある。近年、ビジネスに貢献する範囲は大きく広がりがつつあるが、その活用レベルはさまざまであり、効果もさまざまである。線形回帰のような単純なモデルによる予測結果が有効な場合もあるし、アンサンブルモデルによって複数のアルゴリズムを組み合わせ、投入データに応じて逐次モデリングを行い、より精度の高い予測結果を出力する複雑なモデルが必要な場合もある。

しかし、複雑なモデルだから効果の高い結果が出せるとは限らず、そのビジネス課題にふさわしい結果を出すことが、何よりも重要である。

ビジネス課題を理解したうえで、データ分析を活用すべきという点は、河本 [1] が述べているとおりであるが、さらに、モデリングそのものが実態に則していな

いとビジネス課題に適用することができない点も重要である。複雑で難しいモデルを開発し業務に適用したとなれば恰好は良いが、実際には単純なモデルのほうがフィットし、役立つことも多々ある。新手法にこだわるあまり、本来解決すべき課題を見失うこともある。

課題にフィットしたモデリングを行うためには、見た目や外聞より、その課題の本質をとらえたオーダーメイドモデリングが必要なのである。

1.2 組み合わせによる活用効果

数理科学の多くの技術が、ビジネスの世界で適用され、効果を上げている（実例については、本特集のほかの論文をご覧ください¹）が、一般的に適用初期の段階では、課題の中でも単体技術（「機械学習」や「シミュレーション」といった一つの分野という意味）で解決するところから取り組むことが多い。そうすることで、実現が比較的容易にもなることもあり、効果を測りやすい。効果が出るとなれば、ビジネス展開（運用）を行い、ブラッシュアップを続けていく。すると、単体技術の活用のみでは効果の限界に達することが多々ある。たとえば、需要予測を機械学習で行っても、それによる生産計画を手作業でできる範囲に限っていれば、予測の利点を最大限活かしているとはいえない。また、効果を最大化するはずの計画を立案できたとしても、前提条件などの変化により、想定していた効果が得られない場合もある。

より効果的な活用を考えた場合の次のステップは、複数技術の活用である。先の例でいえば、需要予測に

おぎ しのぶ

(株) NTT データ数理システム

〒160-0016 東京都新宿区信濃町 35 番地信濃町煉瓦館 1 階

¹ 数理システムユーザーコンファレンスの過去の内容は <http://www.msi.co.jp/userconf/2020/index.html> を参照

シームレスに連係させる形で数理計画技術を用いて生産計画の立案まで行うことができれば、最新の状況に合わせた計画立案を、人手を介さずに行える。すると、予測に対して最良の計画を見つけ出すことができ、状況の変化に迅速に対応できる。また、計画立案の結果をシミュレーションで検証することでロバスト性を測ることができ、効果の最大化のみを優先したりリスクの高い計画よりも、多少想定効果の減少を招いてもロバストな計画を選ぶ、といったリスクヘッジを取ることができる。

より効果的な課題解決効果を想定すればするほど、複数の技術の組み合わせは重要な課題となってくる。

1.3 ビジネス適用へのステップを踏む

技術適用初期の段階では、単体技術の活用による課題解決から始めることは上で述べたとおりである。しかし、単に技術とその利用ノウハウさえあれば、ビジネス適用のゴールである運用までをスムーズに進められるというわけでない。

課題の性質や、とりまく状況、扱うデータなどによって、その活用方法は大きく異なる。また、解決すべき課題の組み立て方も変わってくる。

ビジネス適用には大きく二つのステップがある(図1)。

- 知識発見フェーズ
- 業務組み込みフェーズ

1.3.1 知識発見フェーズ

知識発見フェーズは、技術適用の最初に取り組むサイクルである。まず手元にある情報・データを整理し、現状を把握するための分析などを行う。必要に応じて、追加でデータを収集する。長期的な課題に対して、今現在の状況で、解くべき課題、踏むべきステップを認識し、検証を行う。知識を発見していくフェーズである。

データの整理、条件の整理、データの見える化と把握、そしてモデリングへと進み、いったんは作成したモデルでの検証を行う。しかし、最初の1回でうまくいくことはまれで、改めてデータを見直し、条件を見直し、モデルを作り直し、場合によっては、方向性も見直したりする。この試行錯誤を行うことが知識発見フェーズであり、これをしっかり行うことが後にビジネスで効果を上げるためにも必要である。

そして、この段階では、試行錯誤が容易に行える、インタラクティブなツール・環境が必要になる。

1.3.2 業務組み込みフェーズ

知識発見フェーズでの試行錯誤が定着し、多くの試行錯誤が行われた結果、有用性が確認されたものにつ

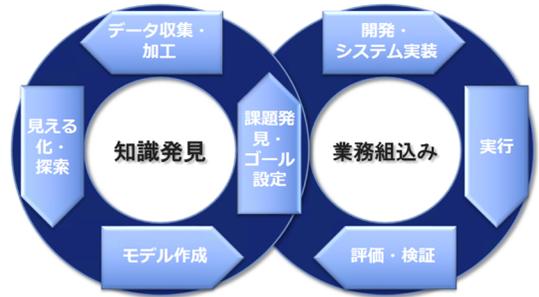


図1 知識発見フェーズと業務組み込みフェーズ

いては、運用のための業務組み込みフェーズに到達する。これは知識発見フェーズと切り離さずに、できるだけ、その知見を活かし、作成モデルもそのまま移行できることが望ましい。ここで改めてモデリングをやり直すと、知識発見フェーズで効果検証を行ったモデルとは異なるモデルになってしまう危険がある。その場合、改めて効果検証が必要になり、非常に効率が悪い。できるかぎりそのままのモデルを利用できることが効率的である。組み込む過程では、運用データを用い、システムと連係させ、業務の適用サイクルを回す。するとここで、いったんのゴールであるビジネス適用の成功を体験できる。

しかし、モデルには賞味期限がある。データの傾向の変化や、前提条件の変更が起こった場合には、リモデリングが必要になる。その際、知識発見フェーズに戻ることになるが、知識発見フェーズと業務組み込みフェーズが全く同じモデルであれば、リモデリングを行った後、再度業務組み込みフェーズに適用する際も非常にスムーズである。

2. 次世代統合プラットフォーム MSIP

1節では、ビジネスに数理科学を活用するためのポイントを述べた。

- オーダーメイドモデリング
- 複数技術の組み合わせ効果
- 知識発見と業務組み込みフェーズのステップ

これらが実現でき、数理科学をビジネスに活用するために最適な次世代統合プラットフォームMSIP (Mathematical Systems Incarnation Platform) の詳細を紹介する。

2.1 ブラウザ上でビジュアルにフローを構築

MSIPは、NTTデータ数理システムが2000年から提供しているデータマイニングツール Visual Mining Studio およびその基盤部分である Visual Analytics



図 2 ブラウザ上で動作する MSIP の様子

Platform から引き継がれたビジュアルプログラミングはそのままに、ブラウザベースで利用可能な次世代統合プラットフォームである。2021 年時点においてはオンプレミスサーバーにインストールし、クライアントマシンのブラウザから利用する形態である。サーバーは Windows 上に構築し、クライアントであるブラウザは Chrome を対象とし、OS を問わない。

2.2 オーダーメイドと拡張性

MSIP 上では、NTT データ数理システム製品をすべて合わせると、100 を超える手法やアルゴリズム（分析アイコンなど）を利用できる。これらを図 2 に示すように、画面上で組み合わせて処理フローを構築することで、第三者にも容易に説明可能なオーダーメイドなモデリングを実現できる。モジュール機能を用いることで処理を隠蔽化し、別ユーザーに対してブラックボックス化することもできる。さらに、モジュール内に含めた各種アルゴリズムに対して、パラメータの最適化を行う仕組みも搭載し、モデルパラメータのチューニングの手間を大きく省いている。

また、MSIP にはモデリングの前に行うべき前処理もアイコンとして多数搭載されているが、これらもモデリングの内部に組み込み、モデリングの一部としても利用可能な仕組みをもつ。これによって、前処理も

合わせてモデルとすることができる（図 3）ため、前処理も含めたパラメータチューニングが可能である。

そして、作成したモデルを適用する（予測などを行う）際にも、前処理を含めた形で行うことができる（図 4）。

このように、課題に合わせて多くの手法や前処理を組み合わせたオーダーメイドなモデリングが可能であることが MSIP の特徴である。

ところで、ビジネス課題は近年多様化しており、それにつれて、従来のアルゴリズムやモデリング手法だけでは解決できない課題も登場しつつある。それを受けてさらに効果的で先進的なアルゴリズムも新しく登場するであろう。すると、それらを従来の処理フローに適用し、拡張することも要求されるようになる。さらに、それらをアプリケーション化しメンテナンスを行うことや、既存の業務システムへ組み込んで日々のレポートを行うことも要求されていくだろう。これらを、統合プラットフォームのベンダーがすべて対応することには限界がある。

MSIP では、これらの拡張をユーザーが行えるよう、プラグインによる機能拡張の構造をもっている。公開されたプラグインのインターフェースを用いることで、新規のアルゴリズムも、モデリング手法も、アプリケーション化も、MSIP に組み込むことができる。つまり、

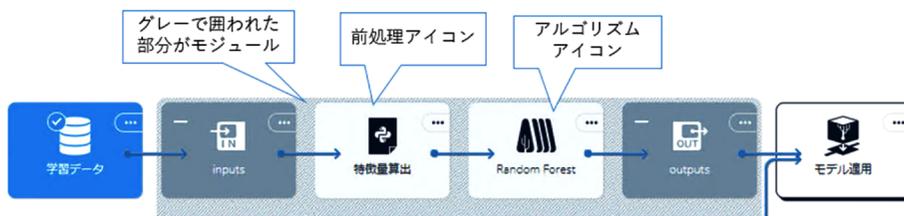


図 3 前処理を含めたモデリングモジュール

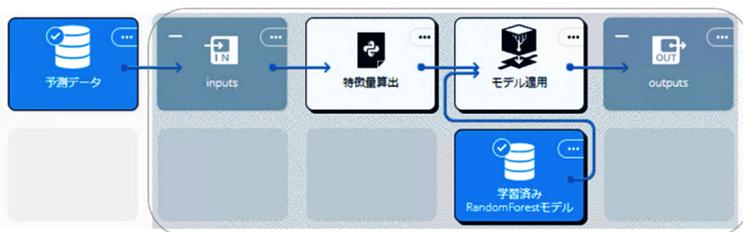


図 4 前処理を含めたモデル適用モジュール

ユーザーが、新規の分析アイコンを組み込むことや、ワークフローの代わりのユーザーインターフェースを組み込むことができるため、MSIP のもつプラットフォーム機能をフレームワークとして活用することもできるのである。

ここで、MSIP が公開するインターフェースを列挙する。

- アイコン実装プラグイン
- UI モデル実装プラグイン
- 操作のための RESTful API²
- データ処理 Python API

このほか、MSIP 以外で作成したものを有効活用できるように、アイコンの実装までは至らずとも、Python や R などのスクリプト言語をそのまま記載することで利用できるスクリプトアイコンも用意している。

2.3 複数技術の組み合わせ

統合プラットフォームである MSIP は、その名のとおり、ビジネス課題を解決するために、数理科学技術を統合し、活用するためのプラットフォームである。統計解析 (Visual R Platform)、機械学習 (Visual Mining Studio, DeepLearner, Big Data Module)、ベイジアンネットワーク (BayoLinkS)、テキストマイニング (Text Mining Studio)、数理計画 (Numerical Optimizer)、シミュレーション (S-Quattro Simulation System) のようなさまざまな技術の一つのプラットフォーム上でシームレスに連携させることができる他に類を見ない

² Web システムを呼び出すことのできる API であり、「REST」と呼ばれる設計原則に従って実装されているもの。

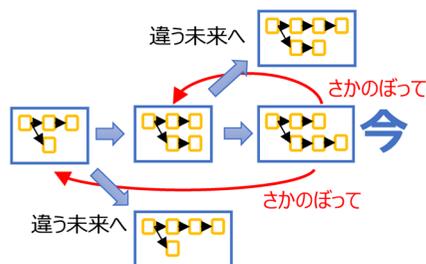


図 5 タイムシフト機能イメージ

フレームワークである。これにより、1 節で述べたとおり、単一の技術での課題解決よりも、はるかに大きな効果を得ることが期待される。

2.4 知識発見と業務組み込みサイクルの実現

知識発見フェーズでは、ビジュアルフローによるデータの可視化やモデリングが非常に重要である。特に、さまざまな処理フローを作っては試すという試行錯誤の過程において、過去の試行を保存し確認しながら実施できれば、知識発見の効率が上がる。MSIP では、これをタイムシフト機能として実現し、いつでも過去の状態に戻り、そこから別のフローを改めて作成することができる (図 5)。

MSIP は、こういった知識発見フェーズにおける利便性の高い機能を搭載しているだけでなく、知識発見フェーズで構築したモデルを、そのまま業務組み込みの運用環境で利用できるように、サーバー部分を RESTful API によって操作可能になっている。システム構成図 (図 6) を示す。

つまり、業務組み込みの段階に至っても、知識発見

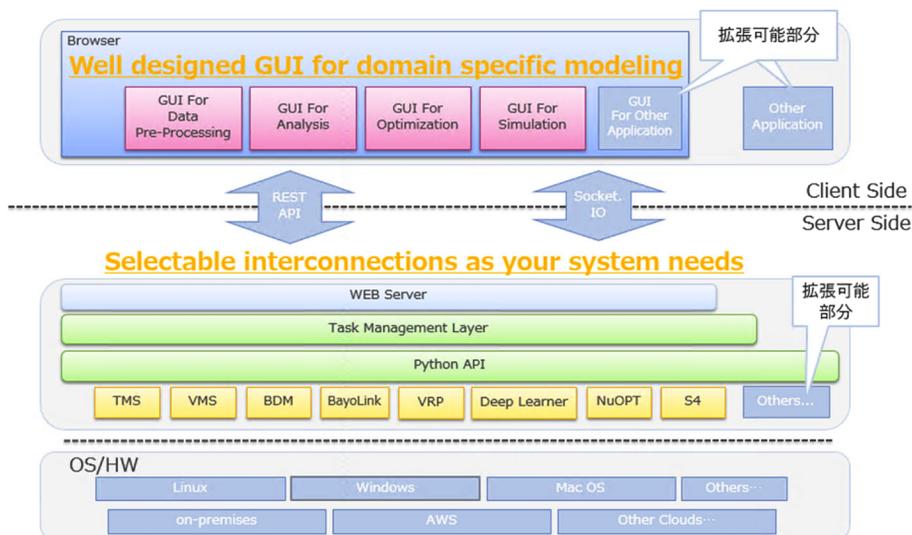


図 6 MSIP 構成図

フェーズと同じサーバーを用意し、そのモデルを移動させるだけで運用のためのサーバーを構築できる。業務組み込みのために、改めてモデリングする必要もなく、知識発見フェーズの延長で、運用でのテストや実施も可能になる。

さらに、周囲の変化に合わせてリモデリングを行う場合も、運用段階で収集したデータをもとに知識発見フェーズと同様にモデリングでき、それを運用環境に即座に反映させることができるため、スピード感をもってリモデリングを実現できる。

3. 統合プラットフォームのデータ活用事例

これまで紹介した内容を踏まえ、MSIP がどのような場面で活用されるのか、具体例を紹介する。

3.1 事例 1：コールセンターで呼損を減らす

コールセンターでの呼損（お客様が電話をかけても対応できるオペレータが足りずにお客様が待ちの状態になり、待ちきれずに電話を切ってしまうこと）を減らすことは、コールセンターの顧客満足度向上において非常に重要なポイントである。呼損を防ぐには単純にオペレータの数を増やすことが解決策ではあるが、コスト面や教育面、場所の問題などから簡単ではない。たとえ増員するとしてもできるかぎり少なく抑えたいのが当然である。そこで下記の観点が必要である。

1. 必要な日に人員を増やす
2. 必要な数の人員を増やす
3. 必要なスキルの人員を増やす

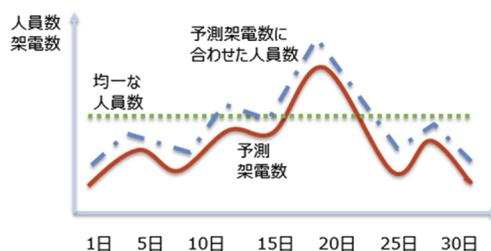


図 7 とある月の架電予測数と対応した人員数

3.1.1 必要な日に人員を増やす

コールセンターへの過去の架電数データを用いることで、架電の周期性の分析、予測モデルの構築ができる。これには統計解析や機械学習の技術を活用すればよい。たとえば統計解析における ARIMA モデルによる予測や、時系列線分表現におけるマッチング予測なども利用できる。1 か月間の日々の架電予測を立てることができれば、架電予測数の変化に合わせた人員を配置することができる。これによって呼損を防ぐことができるし、均一の人員配置をするよりは、架電が少ない時間帯には人員を減らす等、無駄な増員をせずに済む（図 7）。

しかし、予測架電数に対して、呼損を防ぐためには多めの人員配置が必要であり、この状態でも無駄は発生する。また、予測が少しでも外れると一気に呼損が増えてしまうというリスクも残る。

3.1.2 必要な数の人員を増やす

架電数予測が行えたとして、その架電数に対応可能な人員の数を揃えたとしても、予測は完璧ではなく、

どうしても揺らぎが発生し、不足や余剰が出てしまう。そこで、いつ、どの程度の揺らぎがどのくらいの確率で発生するかも合わせて予測し、人員数を調整しながら揺らぎをシミュレーションすることで、そもそも深刻な呼損が生じるのか、どこで生じるのか、それとも軽微で済むのかが検証できる。すると、深刻な呼損を防ぐための人員数を、割り出すことができる。これが複数技術の組み合わせである。MSIP 上で、予測モデルとシミュレーションをシームレスに連係させ、予測結果に対して揺らぎをシミュレーションすることでこれらが実現できる。深刻な呼損が発生する場合には手厚く、そうでない場合にはそれなりの人員配置が実現でき、より効果的な結果を得ることになる。

3.1.3 必要なスキルの人員を増やす

オペレータに関しては、数の問題のほかに、質の問題もある。問い合わせによって、回答できるスキルが異なり、数が足りていても、回答に時間がかかったり回答できなかつたりすることがある。その場合、呼損が防いでも顧客満足度を落としてしまう結果を生む。これを防ぐためには、問い合わせ内容も含めた架電予測を行い、対応可能なスキルも加味した人員配置を行うことが考えられる。このような予測モデルの構築には、問い合わせ内容という日本語の文章を分析可能なテキストマイニング技術を組み入れて、シームレスに連係する必要がある。MSIP では、テキストマイニングから予測モデルの構築、シミュレーションまで複数の技術をシームレスに組み合わせることができ、問い合わせ内容も含めた架電予測はもちろん、揺らぎを考慮したシミュレーションまで一貫して行うことができる。これにより、必要な日に必要な数、必要なスキルを把握し、無駄がなくロバストで効果的な人員配置が実現できるのである。

3.2 事例 2：異常検知ソリューションの提供

NTT データ数理システムでは、異常検知ソリューションを 2021 年 3 月にリリースする。これは、MSIP 上に、異常検知に特化したソリューションを構築して提供する。異常検知モデルを実現するには、一般的な分類モデルやクラスタリングモデル、Deep Learning などを組み合わせてモデリングを行うが、その変数の扱いやパラメータの設定にはノウハウが存在する。また、独特のデータ加工（前処理）を行ったうえでモデリングすることで、モデリングの効率が格段に上がる。異常検知ソリューションでは、複数の前処理手法を組み合わせ一つのデータ加工としたアイコンや、独自の前処理と複数のモデリング手法を組み合わせ一つのモデ

リングアイコンとしてまとめてパラメータチューニングを行えるアイコンなどを用意する。これらは MSIP 上で、ユーザーが利用可能なモジュール機能やメタパラメータ機能を活用して実現している。つまり、一般ユーザーの立場であっても、独自にこのようなソリューションを実現できるということでもある。

3.3 事例 3：製造現場ラインへの組み込み

異常検知ソリューションなどで異常検知のモデリングを実現した後、製造現場でそれを製造ラインに組み込む段階では、ラインのシステムと連係させるなどの別途開発を行うのが一般的である。このとき、作成した異常検知モデルを、開発システムに移行するのだが、場合によってはモデル自体を実装し直す必要が生じ、モデリングの際と異なる結果を引き起こす場合がある。

これを防ぐには、モデリングと全く同じものが、開発システム上で動作することが求められる。MSIP では、RESTful API によってモデリング時と全く同じ環境でモデルを動作させたり、Python プログラムの形で、開発システムに組み込むこともできるため、組み込みによるモデリングテストも不要になる。これによって数理学のビジネスでの利用効率が大きく向上することが期待できる。

3.4 事例 4：アルゴリズム開発者とのコラボレーション

数理学の世界では、日々、新しい手法や実装が生まれており、Python のライブラリといった形で提供されているのをよく目にする。しかし利用者にとって、新しい手法を利用するためにプログラミング言語の習得が必要という現実、新しい手法をビジネスで活用するための障壁にもなっている。こういった障壁をできるかぎり低くできるよう、アルゴリズム開発者が MSIP 上の分析アイコンを簡単に実装できるプラグイン機能が搭載されている。だれでも Python インターフェースを実装（ライブラリを実装する場合と同様）し、GUI を定義する JSON ファイルを用意するだけで、MSIP 上で動作するアイコンを提供できる。これにより、MSIP に搭載されている各種前処理やモデリング、数理学の数々の機能も、シームレスに連係させて利用することができるようになる。プラグイン機能で開発した MSIP 上のアイコンを、独自に販売することも、配布することも可能である。MSIP としては機能が増え、アルゴリズム開発者にとっては新たな展開の場が増えることになり、双方のコラボレーションが実現する。

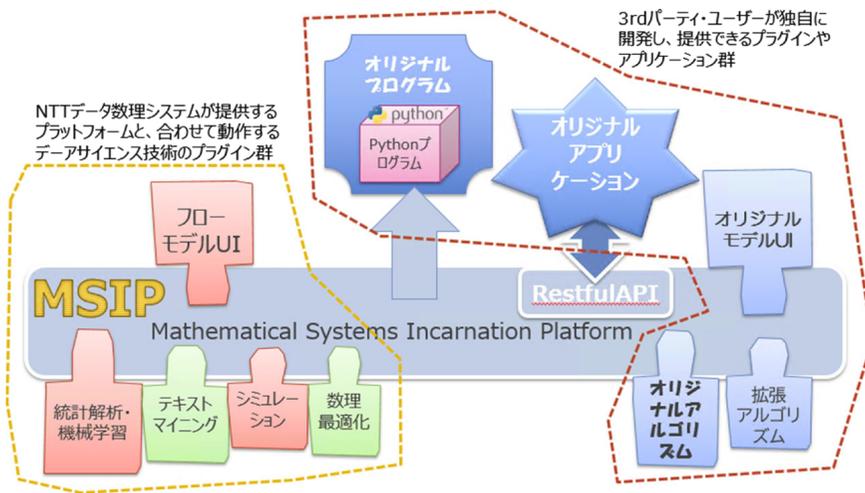


図 8 MSIP の拡張利用イメージ

4. 統合プラットフォームのあるべき姿と今後

データ活用のための統合プラットフォームは、データを用意し自身で分析するユーザーのために、利便性の高い環境を提供し、ビジネスでの活用を支援するものである。そのためには使い勝手の良さはもちろん、必要な技術を複数搭載し、運用まで見据えた仕組みをもっている必要がある。

世の中には多くの自動化ツールや分析を請け負う企業があるが、それでもやはり、自身で分析し、データ活用を進めるという行為はなくならないし、課題が複雑になり、求められるものが高度になるにつれ、課題をもつ人自身が操作可能なデータ活用のための統合プラットフォームは、より必要になっていく。

また、数理科学という進化の激しい世界においては、

少しの間立ち止まっただけで、技術は陳腐化してしまう。オープンなインターフェースを用意することにより、常に最新で効果的で役に立つ機能を世界中のいたるところから受入れ、提供していくことが、次世代統合プラットフォームとしての使命であると考えられる(図 8)。

これらの要素を実現する MSIP は、2021 年 9 月、搭載する各製品のリニューアルとともに正式リリースする。数理科学をビジネスに活用するために、必須のプラットフォームとなり、今後も進化を続けていくことを希望している。

参考文献

- [1] 河本薫, 『最強のデータ分析組織—なぜ大阪ガスは成功したのか—』, 日経 BP, 2017.