

「データサイエンティスト・オブ・ザ・イヤー」初代受賞者が語る4つの「壁」とは 大阪ガス、オージス総研、IBMに聞く、 機械学習とデータ活用を阻む「壁」の壊し方

機械学習とデータ活用を阻む「壁」を壊し、機械学習の持つ真のポテンシャルを引き出すにはどうすればいいのだろうか。大阪ガス、オージス総研、IBMに、企業のデータサイエンティスト、システムインテグレーター、分析技術の専門家というさまざまな視点の見解を聞いた。

デジタルトランスフォーメーションの進展に伴い、競合との差別化に欠かせないデータ解析、機械学習の取り組みを始めようとする企業は増加している。一方で、既に取り組みを始めた企業の中には、学習の前に、学習用のデータを準備することにさまざまな課題を感じている企業も少なくない。機械学習において重要となるのは、データの量と質だからだ。

機械学習とデータ活用を阻む「壁」を壊し、機械学習の持つ真のポテンシャルを引き出すにはどうすればいいのだろうか。大阪ガス、オージス総研、日本IBMに企業のデータサイエンティスト、システムインテグレーター、分析技術の専門家というさまざまな視点の見解を聞いた。

2018年、機械学習やデータ活用の現実

大阪ガスが機械学習の取り組みを始めた背景や代表的な事例を教えてください。

河本氏 私たちデータ分析チームは、2006年に情報通信部門に移りましたが、その頃から機械学習のトレンドが盛り上がりつつありました。そうした中、まずは社内のデータベースに完備されているデータを使って機械学習の取り組みを始めました。

そしてこの5年間は、機器や設備をネットにつないでデータを収集し、IoTデータによる故障予知や異常検知などで成果を出してきました。直近1年は、IoTの次のステップとして、画像や音声を活用し、ディープラーニングのアプローチにも挑戦しています。

代表的な事例としては、工場内の設備の画像から、腐食しているかどうか

をディープラーニングで判断する取り組みを行っています。音声についてはこれからですが、例えば風力発電の故障予知を、ディープラーニングによって振動音から検知できるようにしたいと考えています。



大阪ガス
情報通信部
ビジネスアナリティクスセンター所長
大阪大学 招聘教授 博士(工学、経済学)
河本 薫氏
日経情報ストラテジーが選出する
「データサイエンティスト・オブ・ザ・イヤー」
初代受賞者(2013年)

オージス総研では機械学習にどのように取り組んでいるのでしょうか。

山口氏 当社は、大阪ガスグループの情報システム企業として、統計解析や機械学習、ディープラーニングを手掛けており、適材適所でデータアナリティクス部隊やAI専門部隊を展開しています。

また、大阪ガスとの実績を生かし、外部へのAIソリューションの提供も

行っています。特に、機械学習・ディープラーニングの専門組織「AIテクノロジーセンター」では、センサーデータから画像、音声、テキストまで幅広く対応し、画像解析による工場の異常検知や食品会社の異物混入検査、動画解析による振る舞い検査などに取り組んでいます。

IBMから見た、昨今の機械学習活用のトレンドやIBMが携わった事例を教えてください。

山田 機械学習の活用範囲は、業種・業務を問わずさまざまな領域に広がりがつあり、「AI エニウェア」が進んできています。基本的に機械学習は人に対するレコメンドなので、仕事をしている人へのアドバイスが必要なシーンでは、あらゆる業種・業務で活用できると考えています。

例えば、人事部門では、退社するリスクが高い人材をマネジャーに通知したり、経理財務では、より早い段階で期末の売上予測を出したり、店舗業務では、優良顧客の来店を販売員にリアルタイムで通知するといった、業務改善に向けた活用事例も増えてきています。

西牧 機械学習を活用すれば、それだけで業務が最適化されると思われがちですが、一足飛びにはいかないのが実情です。機械学習の結果がどれだけ実効性を持つものなのか、現場と話し合い、理解を深める必要があります。機械学習の結果は、全てそのまま業務改善につながるわけではなく、場合によっては活用しない方がいいケースもあります。機械学習の結果と現場のビジネスルールをいかに融合させるかが活用のポイントになると考えています。



日本IBM
グローバル・ビジネス・
サービス事業
先進的アナリティクスと
最適化 技術理事
データサイエンティスト
工学博士
山田 敦

機械学習、データ分析が業務で役に立つまでのさまざまな「壁」

企業が機械学習を活用する上での課題をどのように捉えていますか。

河本氏 データ分析、機械学習が業務で役に立つまでには4つの「壁」があります。

まずは「**活用の壁**」。機械学習の活用では、最終的にビジネスに貢献することが大きな「壁」になると考えています。機械学習で高精度な予測をしても、それを実務に使うとするとうまくいかず、ビジネスにつながらないケースが多いように感じます。「データがあるから機械学習をする」アプローチではなく、「企業が抱える課題とその理由を探り、それを解決するために機械学習を活用する」アプローチが本来あるべき姿だと思っています。

また、現場スタッフの立場を理解することも課題として挙げられます。現場スタッフは、業務の中で結果責任と説明責任を果たす必要があります。それを、いきなり「機械学習の分析結果に従って行動しろ」と言うのは難しいのが現実です。私はこれを「**現場の壁**」と呼んでいます。高度な機械学習を使えば使うほど精度が上がります、結果責任は果たしやすくなりますが、ブラックボックス化して説明責任は果たしにくくなります。「精度が高ければ高いほど良い」という考えではなく、現場の説明責任にも配慮した解き方をすることが大切です。さらに、機械学習の精度指標である、捕捉率と的中率はトレードオフの関係です。一方、現場スタッフは、感覚的に許容できる捕捉率と的中率を持っています。だから、事前的中率と補足率のストライクゾーンを決めて、現場スタッフが受け入れられる解を目指すことがポイントになります。

もちろん、どれだけ頑張っても機械学習を駆使しても、目指す精度には至らない場合もあります。例えば、設備の異常検知については100%の精度を求められますが、機械学習を使って100%の精度を実現することは非常に困難です。「**精度の壁**」ですね。例えば、機械学習による異常検知の的中率が95%であった場合、5%は外れることになります。現場スタッフは100%を求めると、このままでは使えない。でも、そこで諦めるのではなく、この分析精度でも使えるように業務プロセスの方を変えられないか考えるのです。

もう1つ、機械学習の課題として見逃せないのが、分析するためのデータをいかにそろえるかという「**データの壁**」です。分析するためのデータ

が足りないと、外部から購入したり、新たに、機器をインターネットに接続してデータを収集したくなったりしますが、成果が上がるか分からない段階で、データにお金をかけるのは「リスクが高い」と言わざるを得ません。そこで大阪ガスでは、まず既存データを使って機械学習の成果を出し、その成果をさらに高めていくために、データの取得範囲や取得量を広げていくアプローチをとっています。そして、このデータ収集や整備など、データエンジニアの部分をオージス総研にお願いしています。

山口氏 データ収集・整備の部分については、大阪ガスと一緒に時間をかけて整備しました。データマートやデータウェアハウスを構築する部門を共同で設置し、メタデータをしっかり管理しています。また、収集したデータを提供する部隊や、現場スタッフのデータ活用を支援する部隊も用意しています。現在は、業務系データが中心ですが、IoT系データについても、各部門で異なるシステムを統合してデータ収集し、クレンジングができる仕組みを整備しつつあります。

西牧 私も、機械学習のためのデータを収集・整備するところは大きな課題だと感じています。河本さんがおっしゃるように、自社のデータを分析する前に、新しいデータを購入するのは非常にもったいない。社内のデータを単純に分析して成果が出なかったら、次は業務的な解釈を与えるなど、特徴量を抽出して分析すると、全く違う結果が出ることもあります。これこそが、データサイエンティストやコンサルタントの腕の見せどころであり、機械学習の効果を最大限に引き出すカギになると考えています。



オージス総研
執行役員
技術統括 技術部長
山口 健氏

「データの壁」を壊すために必要なこと

既存データを活用するということだと、メインフレームにある基幹業務データも例外ではないと思います。大阪ガスでは、メインフレームにある基幹業務データについては、メインフレーム外にコピーし、成型して機械学習やデータ分析を行っているとお聞きしました。

山口氏 メインフレームの中には個人情報も保管されており、大阪ガスとしては当然のことですが、情報漏えいを防ぐためのデータ保護を厳格に行わなければなりません。そのため、ネットワークのセグメント化や物理的にガラス張りの部屋に分離するなど外部から完全に遮断しています。また、個人情報のデータを扱う許可設定を踏まえて、メインフレーム外にデータを取り出すルールを厳格化しています。

使いやすさの側面では「どこにどんなデータがあるのか」という見取り図が分かるデータ辞書を整備して、イントラネット上に公開しました。さらに運用当初は、オーガス総研側にもヘルプデスクを作ってもらい、使いやすさの向上を図りました。また、間違った使われ方をされないように、事前に現場スタッフへの教育を行いました。併せて、非常に厳密な決裁基準を設定し、個人情報が流出しない業務フローを作っています。

河本氏 データを活用する際の課題としては、2つの側面があります。1つは、使いやすくすること。もう1つは、間違った使われ方をされないことです。

このように企業は個人情報や基幹業務データの保護にコストをかけて対応していますが、一方でデータ保護の厳格化はデータを使いやすくする際の課題にもなっていると思います。このようなデータポータビリティの課題に対し、IBMの最新メインフレーム「IBM z14」をはじめとするIBMメインフレーム(ブランド名:IBM Z)で機械学習が行えるようにしたとのことですが、その背景について教えてください。

天野 IBMメインフレームは、現在も世界中の基幹システムで使い続けられており、世界の企業データの約8割はIBMメインフレーム上で生成、保持されています。それは、長年の実績からだけでなく、他のプラットフォームでは実現できない高いレベルでの堅牢性や可用性によるものだと考えています。これは、IBMメインフレームがハードウェア、OS、ソフトウェアのアーキテクチャを一貫した思想に基づき設計、開発できるからであり、だからこそ、今まではミッションクリティカルな業務処理を中心に使用されてきました。しかし、データ活用の可能性がますます広がる今、IBM Z上の基幹業務データへのアクセスの利便性と活用の迅速化をさらに向上させるために、メインフレーム上で機械学習まで提供する必要があると考えました。

「IBM Machine Learning for z/OS」は、IBM Z上のDb2やIMS、VSAMなどのデータに対して、分散システム上のデータと同様のアクセスしやすさも提供します。「社内のガバナンスルールなどでメインフレームのデータを外部から利用できない」「メインフレームのデータを利用するのは承認手続きなどに時間がかかる」というように、「データの壁」や「組織の壁」のために基幹業務データの活用を諦めていた企業でも、メインフレーム上の基幹業務データをそのまま活用して、リアルタイムで機械学習からの洞察を活用することが可能になります。

最新IBMメインフレーム「IBM z14」は、プロセッサデザインなどの実装を、OSとソフトウェアと併せて、さらに機械学習に最適化した拡張を施しています。また、オープンソースソフトウェアのApache Sparkと連携し、豊富な機械学習アルゴリズムを備えた「IBM Machine Learning for z/OS」は、データサイエンティストにも分析初心者にもIBM Z上の基幹業務データを用いた機械学習に取り組みやすい基盤を提供します。



日本IBM
ソフトウェア事業本部
zソフトウェア事業部
ソフトウェア
ソリューションセールス
シニアITスペシャリスト
天野 恭子

メインフレーム1台で機械学習ができるようになると、企業のデータ活用はどのように変わのでしょうか。

山口氏 メインフレームは、データ分析とは異なる業務処理に使うためのマシンという位置付けで考えており、今まで機械学習をするイメージは全くありませんでした。メインフレームのデータを外部にコピーすることなく、セキュアに機械学習に活用できるのは、データサイエンティストにとっては非常にありがたいものになるかもしれません。また、IoT系の膨大なデータを分析するには、Apache Sparkなどの分散処理環境を使う必要がありますが、メインフレーム上でそれが利用でき、さらにデータスコア

リングまで行えるのであれば、機械学習によるデータ活用促進に大きく貢献するのではないのでしょうか。

今後、運用まで含めたソリューションが提供されれば、システムインテグレーターが作り込みを行う必要もなくなり、理想的な機械学習プラットフォームになると感じています。



山田 IBMメインフレームを基盤とする多くのシステムは止めることができない業務を支えています。だからこそ、今まではメインフレームから機械学習のデータを抽出する際にも、なるべくCPUに負荷をかけないように処理を行う必要がありました。機械学習の機能を伴ったIBM Zによって、メインフレームに遠慮しながら機械学習を行わなければいけないという現状を改善できるのではと考えています。

西牧 個人情報を含むデータについても、いままではガバナンスの問題などでメインフレームから切り出せず、機械学習に活用するのは難しかった。しかし、IBM Zであれば、メインフレームの中で個人情報のデータの特微量に変えることで、セキュアに機械学習に活用できるようになると考えています。

例えば、機械学習で活用できる顧客情報のデータ範囲を定義しておき、業務の中で発生する顧客情報データをIBM Zでダイレクトに分析処理にかけて結果を返すといったことも可能になると期待しています。



日本IBM
クラウド事業本部
データサイエンス・
テクニカルセールス
SPSS IT スペシャリスト
西牧 洋一郎

機械学習を目的化せず、ユーザー体験のために

最後に、これから機械学習を始める企業に向けて、それぞれの視点からのメッセージをください。

河本氏 多くの企業で機械学習への関心が高まっていますが、「機械学習で何ができるのか」を社内で再確認して、地に足の着いた取り組みをやっていくことが大切です。経営者自身が機械学習の持つ真のポテンシャルを理解して、単なる現場の業務改善だけでなく、「将来的にビジネスモデルを変革させる可能性を秘めている」ことを感じながら取り組んでほしいと思います。特に、機械学習を目的化してしまうのは最も良くない。10年、20年先のビジネスモデルを見据えて、次のステップのために機械学習を活用するのが重要です。

山口氏 現在は、「社内にデータがたまっているから機械学習をやりたい」というニーズが多く、まだまだ機械学習をする意味について啓蒙が足りていないと感じています。ベンダーの立場としては、機械学習への理解度をもっと高めて、企業が「自社の業務のどの部分に機械学習を活用できるのか」を判断し、仮説まで立てられるようにサポートしたいと思います。オーグス総研には、ニーズに合わせて、既製服からセミオーダー、オートクチュールまで、機械学習を活用するための道具立てがそろっているので、機械学習を検討している企業はぜひ相談に来てください。

西牧 経営者が機械学習の未来を認めて、覚悟を決めて取り組むことが重要です。失敗も含めて、本当にあるべきデータ活用の姿に向き合い、コミットメントすることが大切だと考えています。IBMでは、ベンダーとしてそのヒントは提供しますが、最終的に答えを見いだすのはそれぞれの企業自身です。機械学習に向けた取り組みを通じて、「どこにどんなデータがあるのか」を把握するだけでも、その企業にとって大きな価値になるはずだ。

山田 現場やお客さまの新しい体験を描き、その実現のために機械学習を適切なポイントに活用していくべきでしょう。その新しい体験を、スピード感を持って実現することが併せて大切です。そのため「データの壁」などの「壁」を素早く乗り越えられるツールがあるのであれば、積極的に利用していきたいと思います。

天野 世界中の企業や組織で毎日何十億件ものトランザクションを処理し、基幹業務データを生成、保管しているIBMメインフレームは、「業務処理の専用マシン」という既成概念を超え、いまや「業務処理とデータ活用のためのハイブリッドマシン」となっています。これまでは「データの壁」のために諦めていた最も重要な基幹業務データからの洞察を、基幹業務データが存在する場所での機械学習で導き、業種・業態を問わず、かつてない大きな飛躍をする機会として頂きたいと思います。



お問い合わせ

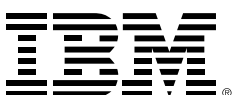
IBM アクセスセンター ☎ 0120-550-210
受付時間 9:00~17:00
(土、日、祝日を除く)

ソリューション情報

「IBM Z」で検索
https://ibm.biz/z_mainframe_jp



この冊子は、@ITに2018年2月に掲載されたコンテンツを再構成したものです。<http://www.atmarkit.co.jp/ait/articles/1802/26/news004.html>



日本アイ・ビー・エム株式会社

IBM、IBMロゴ、ibm.com、およびIBM Zは、世界の多くの国で登録されたInternational Business Machines Corporationの商標です。他の製品名およびサービス名等は、それぞれIBMまたは各社の商標である場合があります。現時点でのIBMの商標リストについては、www.ibm.com/legal/copytrade.shtml (US)をご覧ください。
©Copyright IBM Japan, Ltd. 2018 日本アイ・ビー・エム株式会社 〒103-8510 東京都中央区日本橋箱崎町 19-21
Printed in Japan January 2018 All Rights Reserved