

# ビジネスアナリティクスの実践



沼澤 優



庄司修芳



大橋俊介

## CONTENTS

- I ビジネスアナリティクスへの関心の高まり
- II 製造業におけるビジネスアナリティクスニーズ
- III ビジネスアナリティクスの実践を阻む「前処理工程」のカベ
- IV 前処理工程にどう取り組むか
- V ビジネスアナリティクスの実践に向けて

## 要 約

- 1 ビジネスアナリティクスに対する企業の期待が高まっている。ただし、ビジネスアナリティクスの導入により大きな成果を上げている企業は、ごく一部に限られるのが現状である。要因として、IT環境や組織・人材などのケイパビリティ不足が指摘されることが多いが、これは必ずしも正しくない。ビジネスアナリティクスの実践に向けては、プロセスそのものにも目を向ける必要がある。
- 2 ビジネスアナリティクスのプロセスのうち、最も注目される工程は分析の工程である。しかし、分析結果の巧拙を決めるのは分析工程そのものではなく、分析モデル作成用に質の高いデータを準備する「前処理工程」にある。
- 3 これまで、前処理工程は分析に付随する単純作業と見なされ、分析担当者に一任されてきた。一方、前処理工程の課題解決は、質の高いデータを得るためのセンサーの仕様変更や現場業務変更などを伴うため、分析担当者と関連組織が一体となった取り組みが必須となる。
- 4 これからビジネスアナリティクスに取り組む企業は、前処理工程を「分析の品質を左右する重要工程」と位置づけ、組織横断的な活動として取り組むべきである。

## I ビジネスアナリティクスへの関心の高まり

ビジネスアナリティクスとは、企業が入手可能な社内外のデータを分析することにより、これから起こり得る将来の予測や、結果を最適化するための選択肢を示し、経営や業務における意思決定の高度化や自動化を図る取り組みである。今日、センサーデバイスの小型化やワイヤレス接続の普及、通信料金の低価格化などに伴い、あらゆるモノがインターネットにつながるIoT（Internet of Things：モノのインターネット）が日常的なものとなった。これにより、企業はこれまで入手できなかった膨大なデータの収集が可能となり、分析可能な対象の幅が格段に広がったことで、ビジネスアナリティクスに対する期待が高まっている。

一般に、IoTで収集されるデータは非常に膨大なものとなる。このため、大量なデータを蓄積し高速で計算処理できるデータインフラや、AI（Artificial Intelligence：人工知能）などの高度な分析アルゴリズムを活用できる情報システムが必要となる。近年はクラウド・コンピューティングの普及により、実質的に無限ともいえるデータストレージとデータ処理能力を非常に安価で使用できるようになった。また、高度な分析アルゴリズムを具備したクラウドサービスや、無料で利用できるオープンソースの分析ソフトウェアなどが登場している。ビッグデータやAIの活用においては「分析人材がないこと」が声高に叫ばれるが、自動運転などに代表される特定の技術革新領域を除けば、こうした分析ソフトウェアを活用することで、専門の統計ス

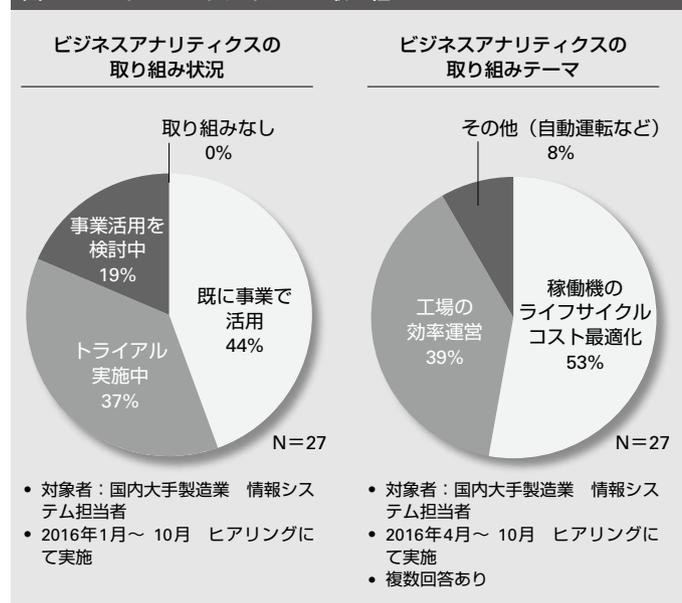
キルがなくても十分に取り組みを進めることができる。このため、かつては取り組み開始時に大きな投資を必要としたビジネスアナリティクスも、今では非常に取り組みやすいものになっている。

こうした背景を受け、本格的にビジネスアナリティクスに取り組み始める企業が増えている。その一方で、大きな成果が出たという企業はまだ多くないのが実態である。これは、企業がビジネスアナリティクスの取り組みによる効果を得るためには、人材・組織やIT環境を整備するだけでは十分でないことを示唆している。そこで本稿では、企業がビジネスアナリティクスに取り組む際の業務レベルでの課題を解説し、ビジネスアナリティクス実践のためのアプローチについて考察していく。

## II 製造業におけるビジネスアナリティクスニーズ

前述の通り、現在、ビジネスアナリティクスに関心を持つ企業は多い。図1は、野村総

図1 ビジネスアナリティクスの取り組み



合研究所（NRI）が懇意にしている製造業のIT担当に、ビジネスアナリティクスへの取り組み状況をヒアリングした結果である。ヒアリングを行ったほぼすべての企業が、ビジネスアナリティクスに関心を持っており、既に大半の企業が何らかの取り組みを開始している。

また、今回のヒアリングから国内の製造業のビジネスアナリティクスに対する取り組みの多くは「稼働機のライフサイクルコスト最適化」「工場の効率運用」の2つに分類できることが見えてきた。

## ニーズ① 稼働機のライフサイクルコスト最適化

「稼働機のライフサイクルコスト最適化」の取り組みは、エネルギーや運輸などの社会インフラ、医療関連機械、建設機械や工作機械といった産業機械の領域で先行している。よく取り上げられる代表例として、GE（ゼネラルエレクトリック）のインダストリアル・インターネットやコマツのKOMTRAXがある。

たとえばGEの場合、航空機のエンジンに搭載したさまざまなセンサーから得られる運行時のデータを解析することで、操縦時における各種機器のパラメータの最適値や最適運行ルートへの提案による燃料節約、予防保全による故障の防止、整備時間短縮による定時発着率改善などを実現している。

コマツに代表される建機や農機などのフィールド機械を扱う企業は、CBM（Condition Based Maintenance：状態基準保全）に対する期待が大きい。CBMとは稼働機のセンサー情報から得られるデータを活用して機械の劣化状態を把握し、事前に故障発生を予測して保全を行う取り組みである。現在のメンテナン

スサービスは定期訪問によるTBM（Time Based Maintenance：時間基準保全）が主流であり、実際の機械の状態にかかわらず、定期的に点検・整備を行う取り組みである。このため、本来は必要ないタイミングでの点検・整備となることが少なくないばかりか、定期点検前の突発的な機械の停止や故障などを未然に防ぐことが難しい。一方、CBMにより、機械の状態に基づいた必要最低限の点検・整備のみを行うことができれば、製品を購入した企業は整備コストを最小限に抑えつつ、突発的な設備停止による機会損失も防ぐことができる。

このように製造業がIoTデータを活用することで、従来は難しかった「製品購入後の顧客の活動プロセス」そのものに価値を提供することが可能となるため、これまでの製品販売を軸にしたビジネスモデルから新たなビジネスモデルに転換する起爆剤として、ビジネスアナリティクスへの大きな期待が寄せられている。

## ニーズ② 工場の効率運用

現在の製造業の生産現場においては、生産ラインに多くのセンサーや計器が搭載されており、さまざまなデータが取得されている場合が多い。これまで、これらのセンサーデータは、主に生産管理に活用されてきた。しかし、近年のIoTブームを受け、製造業の関心が生産リードタイム短縮や製造コスト削減、生産スケジュールの最適化といった「工場の効率運用」に向けられてきている。「稼働機のライフサイクルコスト最適化」の取り組みは高額な製品を扱う企業に多いのが現状だが、「工場の効率運用」は製品単価や分野を問わず、幅広い製造業に注目されている点特徴的である。

IoTを活用した工場運営の高度化・効率化は、ドイツにおける官民一体の「インダストリー4.0」が有名である。インダストリー4.0は、「モノづくりの革新」を目指して、ドイツの産学官が協力して2011年に立ち上げた取り組みである。この取り組みは工場の設備がネットワークを通じて情報を伝達し合い、調達、生産、物流、品質管理、生産管理を行うことで、需要の変化や生産トラブルなどの環境変化にダイナミックに対応し、工場横断でのパフォーマンスの最適化を目指すものである。IoTを活用した工場運営の高度化・効率化の取り組みは、工場の各工程における機械設備の最適制御や、工場ラインやオペレーションの最適化、さらにはAIによる完全自動化などを目指しており、実現に向けては高度なデータ分析が欠かせないものとなっている。

### III ビジネスアナリティクスの実践を阻む「前処理工程」のカベ

前章では、製造業におけるビジネスアナリティクスへのニーズが明らかであり、具体的なテーマに取り組む企業が増えていることを紹介した。一方、取り組みを開始する企業が増えているにもかかわらず、ビジネスアナリティクスの成果を上げている企業はごく一部に限られる。このギャップの要因として、IT環境や組織・人材などのケイパビリティ不足が指摘されることが多いが、分析業務そのものにも課題があると筆者らは見ている。そこでここでは、ビジネスアナリティクスの業務

プロセスそのものについて解説を加えたい。

#### 1 分析結果の巧拙を決める「前処理」

ビジネスアナリティクスの業務プロセスは、「ビジネス仮説設計」「データ収集」「蓄積」「前処理」「分析」「業務活用」の工程からなる（図2）。これらのうち、最も重要とされる工程が「分析」であるが、分析結果の巧拙を決めるのは、「分析」そのものではなく「前処理」である場合が多い。

センサー技術の発展により、これまで取得できなかったデータを「収集」できるようになり、クラウド・コンピューティングの普及により大量なデータを「蓄積」できるようになった。しかし、センサーから取得・蓄積されるIoTデータそのものは無機質な信号の塊でしかないので、そのまま「分析」に活用できるケースは稀である。たとえば、AIの代表的な手法の一つに機械学習がある。この手法は分析担当者が準備するデータをインプットして、コンピュータが一定のアルゴリズムの下で反復的に学習を行い、そこに潜むパターンを見つけ出すものである。ここで利用されるアルゴリズムが、ランダムフォレストやディープラーニングといった既に確立されたロジックである場合、最終的に得られる分析結果の優劣は分析担当者が準備するデータの質によって左右される。このため、分析担当者が分析を行う際に最も重要となるのはアルゴリズムに投入するデータの前処理工程であり、分析モデル開発に費やす時間の、実に50～90%を占めるともいわれている。

図2 ビジネスアナリティクスの実施フロー



## 2 IoTデータのスパース性との戦い

IoTデータ分析の前処理に多くの時間を要する要因は、IoTデータの「スパース性」にある。「スパース (sparse)」とは「まばらな」「わずかな」という意味の英語である。IoTの実現により、これまで取得できなかったデータを取得できるようになったことは紛れもない事実であるが、「IoTの実現により取得可能になったデータ＝分析に活用できる有意なデータ」ではないことに注意する必要がある。

IoTによって取得されるデータのほとんどは、分析に使われないものであったり、欠損値や異常値などの無効なデータを含むものであったりするため、分析担当者はこれらのデータを除去したり、有効なデータを補完したりする作業に追われる。以降では、スパース性が引き起こす前処理の難しさについて解説を加えていく。

### (1) 分析に必要なデータが見つからない

GEの航空機に搭載したジェットエンジンが太平洋を越えて飛ぶ場合、1回当たり500GB程度のデータを発生するといわれている。また、工場の場合、一つの製造ラインに数十カ所のデータ観測ポイントがあり、1工場当たり10ライン、各センサーから10ミリ秒ごとによりデータを収集すると仮定すると、1日当たりのデータ収集件数は数億件に上ると試算される。

一方、「稼働機の故障予測」や「工場ラインの生産性向上」といった分析の目的に照らすと、予測モデルなどの作成に必要なデータは、収集・蓄積されているデータのほんの一部でしかない。前述の通り、分析結果の優劣は分析アルゴリズムに投入するデータの

質に左右されるため、分析担当者はデータベース内の膨大なデータの中から分析の目的に沿ったデータのみを抽出する必要がある。

本来、ビジネスアナリティクスに取り組む際は、データの活用目的に沿って取得すべきデータを決め、機器や設備に適切なセンサーを取り付けてデータを収集・蓄積・分析する、という手順が理想である。しかし、「分析で何ができるか?」「どんな経済効果を享受できるか?」を明確にしてから取り組みを開始することが容易でない現状を鑑みると、データの活用目的に沿って取得すべきデータを定めることは難しく、「まずは集められるだけデータを集めてみた」という企業が多いのが実態ではないだろうか。このため、データベースに蓄積された膨大なデータの中から、後付けで決まった分析目的に合致したデータを探索する作業は、広大な砂漠の中から、あるかないか定かでない砂金を探すようなものであり、分析担当者に大きな負担を強いている。

### (2) 分析に必要なデータが不足している

「IoTの実現により取得可能になったデータ＝分析に活用できる有意なデータ」ではないことは前述の通りであるが、「収集・蓄積されたデータだけで、分析に必要な全データを揃えられるとは限らない」ということにも注意が必要である。

たとえば、機械学習を使って故障予測を行うモデルを導出する場合には、アルゴリズムにインプットするデータに対し、あらかじめ「設備が正常に稼働した場合のデータか」「設備に故障が発生した場合のデータか」の定義づけを行う必要がある。

これだけを聞くと簡単な作業のように思われるが、建機、農機のようなフィールド機械や自動車を扱うメーカーの場合、製造やアフターサービスの現場担当が、機械の故障情報をきめ細かく収集してデジタルデータ化できているケースは稀である。たとえば自動車の場合、運転中に事故や故障などのトラブルが発生して停車した際に、ユーザーがメーカーに直接連絡するケースはほとんどない。センサーから得られるデータを活用すれば、当該車両が稼働を停止していることは分かるが、それが「機械の故障によるものなのか」「ユーザーの不注意による事故なのか」「単に目的地に到着して駐車しただけなのか」の判断を行うのは困難である。

また、仮に自社の整備場に持ち込まれて点検・整備を行った場合であっても、点検結果や整備内容が電子化され、かつ、センサーデータと統合された情報として管理されているケースは多くない。一般に、点検結果や整備内容の情報は作業日報として管理されているが、記入フォーマットが統一されていなかったり、記載内容のレベルが不統一であったりするため、分析のインプットとしてそのまま活用できない場合が多い。このため、センサーから得られるIoTデータに対し、分析に必要な現場情報をいかにして補完するかがポイントとなるが、「そもそも手元にないデータをどう補完するか」が問われており、解決は容易でない。

### (3) 分析の実施目的が明確でない

前処理工程の作業は、ビジネス課題が特定され、必要なアルゴリズムと必要なデータの抽出条件が決まったところから始まる。一方、

ビジネス課題や分析の実施目的が曖昧なまま取り組みを開始するケースが多いことも、前処理工程を複雑化させる要因の一つといえる。

こうしたケースは、ビジネスアナリティクスの取り組みを現場に指示する経営者が、「IoTで得られるビッグデータをクラウド上に蓄積し、AIで解析すれば欲しい結果が得られる」といったビッグデータやAIに対する幻想を抱いているために生じることが多い。データ活用の目的が曖昧なままデータ分析に取り組む場合、必要情報の抽出作業や不足情報の補完作業を闇雲に行わざるを得ないため、質の高いインプットデータを準備できず、意義のある結果を得られないことが多い。NRIにデータ分析のご相談をいただく企業の中にも、それ以前の自社の取り組みで「いろいろなデータを使って、今まで見えなかったものが見えればよい」「具体的な仮説があるわけではないが、まずは何かしらの気づきが得られればそれでよい」といった曖昧なゴール設定をしてしまった結果、「成果が出ないので、外部（NRI）に頼むことにした」というケースが少なくない。ビジネスアナリティクスの取り組みを開始する際は、「AIは魔法のハコではない」というトップの正しい認識の下、明確なビジネス課題や目的設定から着手することが肝要である。

## Ⅳ 前処理工程にどう取り組むか

前章では、ビジネスアナリティクスの実践には前処理が重要であること、一方で、IoTデータのスパース性により前処理工程が非常に難しい工程になっていることについて解説した。では、業務に活用可能な分析結果を得

るには、どのような工夫が必要なのであるか。そこでここでは、前処理を進める際のポイントについて「稼働機のライフサイクルコスト最適化」「工場の効率運用」を題材に解説していく。なお、「工場の効率運用」に関しては、プロセス型生産とディスクリート型生産（組み立て加工型）で前処理のアプローチが異なるため、それぞれについて解説する。

## 1 稼働機系IoTデータ

### (1) 稼働機系IoTデータの特徴と

#### 分析時の課題

近年、自動車や建設機械、農業機械などの最終製品メーカーにおいて、エンジンなどのコアパーツにセンサーを取り付け、稼働データを取得する取り組みが増えている。これらの最終製品メーカーの製品の場合、出荷台数が数千台、数万台に上るため、すべての出荷製品から計測可能なデータのすべてを収集、蓄積しようとする、データ量が膨大なものとなる。このため、計測可能なデータをすべてリアルタイムで送信するのではなく、以下のように対象データを絞ったり、稼働機側でデータを一次加工して送信したりするなどの工夫を施しているケースが多い。

- 連続した時系列データを収集するのではなく、一定の時間間隔で定期的に取得するスナップショットのデータを収集し、蓄積する
- 走行距離や使用時間、外気温、エンジン回転数などの稼働状態の特徴を示すデータを選定し、一週間単位の集計値や平均値を稼働機側で計算の上、送信する
- エラー発生時のみ、エラー発生前後の数分間のデータを取得し、送信する

また、これらの機械は異常や故障、修理実績などの現場情報の捕捉が難しく、デジタルデータとして管理されていない場合が多い。このため、CBMのカギとなる故障予知や故障要因の解析を行う場合、肝心のデータが準備できずに頓挫するケースも少なくない。

### (2) 稼働機系IoTデータに対する前処理

稼働機系IoTデータの特徴は、機械1台1台から取得できるデータの情報量が限定的である一方、対象サンプル数が非常に多いことである。分析に利用可能なデータの種類が少なく分析の幅が狭まるにもかかわらず、データ量そのものは膨大で単純な処理でも多大な時間を要するという、まさに分析担当者泣かせのデータである。このため稼働機系IoTの前処理においては、「個々のデータの質を高めること」と「大量のデータを扱いやすくすること」の二点がポイントとなる。

#### ①個々のデータの質を高めるために

稼働機系IoTデータを活用してCBMを行う場合、いわゆる故障予測モデルを作成する必要がある。しかし、異常や故障に関する実績データが十分に取得できない場合、データベース上のIoTデータだけでは正常と故障を識別する予測モデルを作ることは難しい。そこで分析担当者は、「異常が発生したと予想される状態」をデータで表現する仮説を準備し、データベース上のIoTデータを加工して分析に資するデータを作成する必要がある。ただし、「異常が発生したと予想される状態」の仮説を分析担当者のみで準備するのは容易でないため、機械の構造をよく理解した開発系の現場メンバーや、故障や整備の実情

をよく知っているアフターサービスの現場メンバーの協力が欠かせない。

## ②大量のデータを扱いやすくするために

自動車のように出荷台数が数万台、数十万台に上る稼働機の場合、前処理の対象となるデータは膨大な量となる。このため、実施する処理自体は単純でも多大な処理時間を要するケースが多く、「ソートを行っただけなのに、処理完了までに4～5時間を要した」といった話も珍しくない。

そこで、一度に取り扱うデータのボリュームを小さくするために、一部データのみを対象としても目的に資する分析が実施できるよう、適切な単位に分析対象を切り出したり、ミリ秒単位データを分単位に集約したり、といった対応が必要となる。一般に、これらのデータの切り出しやまとめ上げの作業は分析担当者に一任されるケースが多いが、分析担当者が実施できるのは「分析の観点で見て十分なデータか」の判断のみである。このため、この場合も開発系の現場メンバーやアフターサービスの現場メンバーの協力の下、ビジネスでの活用可能性の検証に使うデータの範囲を定義することが効果的である。

## 2 プロセス型生産系IoTデータ

### (1) プロセス型生産系IoTデータの特徴と分析時の課題

プロセス型生産、すなわち連続的に流体が流れているような生産システムの場合、生産性向上のポイントは対象工程の制御にある。このため、DCS (Distributed Control System: 分散制御システム) やPLC (プログラマブルロジックコントローラー) などの制御

装置により、圧力、流量、温度やバルブ開度、液位などの各状態変数を連続的に計測している。近年では、これらの制御システムで管理する主要なデータをMES (Manufacturing Execution System: 製造実行システム) などに吸い上げることで、装置単位ではなく工場単位や工場横断のコーポレートレベルで共有されるようになるなど、活用可能なデータの質、量ともに非常にリッチなものとなってきている。

プロセス型生産の生産性向上のポイントは各工程における作業を最適制御することにあるが、操業時に運転モードを切り替えた際のステータス情報や各種モード情報などの吸い上げは不十分であることが多い。この場合、データベースに蓄積された各設備の状態変数がどのような運転モード下での値なのか分からないため、「データベース上のデータのうち、どこからどこまでが着目すべき運転状態なのか」「そのとき関連する機器はこの運転状態に対応して稼働した状態にあるか」などの情報を取得することができない。このため、プロセス型生産のデータ分析においては、分析対象となるIoTデータに運転状態 (運転モードや運転ステータスなど) をラベリングする前処理が必要となる。この前処理が不十分であると、運転状態が不明なまま分析をせざるを得ないため、分析結果に疑義が生じることにもなりかねない。

### (2) プロセス型生産系IoTデータに対する前処理

各設備の運転状態が時系列データとして管理されている場合、分析対象となるIoTデータに対する運転状態のラベリングは容易な作業

である。一方、こうしたデータが管理されていない場合は「いまAラインは運転状態にある」「パターンXで作業を実施中」「Bラインの作業が開始されたため、Cラインは準備モードに入った」といった各設備の運転状態を、状態変数間の相対関係から推定する必要がある。

具体的には、対象設備の状態変数の値やその前後関係から着目すべき運転状態を推定する。その後、分析担当者がデータ変化の幾何的形状から当時の運転モードを判別したり、SVM（サポートベクターマシン）などのパターン認識を行う機械学習モデルを使って運転モードを自動判別したりするなどして、分析対象となるIoTデータへの運転状態のラベリングを行う。これにより、単体では無機質なデータの塊でしかないIoTデータを「運転状態分析データ」としてより適切な形状に純化していく。

また、プロセス型生産の場合は上記の前処理に加え、流体の流速に乗って変数間の影響が時間的に遅れて波及する場合の遅れ時間（ラグタイム）を考慮することも必要となる。各プロセスに固有な時定数は、変数の観測周期（サンプリングタイム）よりも大きいことが通常であるため、ラグタイムがどのくらいであるかを見積もりつつ分析データを作る必要がある。ラグタイムの推定が難しい場合、複数の遅れ時間を想定し、あたかも別変数であるかのようにデータを準備して分析データを作る。この場合、モデル開発のためのインプットデータが非常に増えることになるが、「どのラグタイムの変数が最も説明力があるか」を機械学習のアルゴリズムに選択させるなどの工夫を行うことが効果的である。

### 3 ディスクリート型生産系IoTデータ

#### (1) ディスクリート型生産系IoTデータの特徴と分析時の課題

ディスクリート型生産は、個別の工程で加工された半製品を元に加工・組み立てを行って最終製品を作り上げる生産システムである。プロセス型生産は、加工対象である流体が工程の始めから終わりまで連続的に流れていくシステムであり、機械設備が密結合されているため、IoTデータを連続的なデータとして取得することができる。一方、ディスクリート型生産は加工途中の半製品が保管でき、生産順序を柔軟に組み替えることができる疎結合型の生産システムであるため、取得・蓄積されるIoTデータが非連続的なデータとなる点が特徴的である。

ディスクリート型生産のマネジメントにおいては、「いかにして工場内各工程を横断した全体最適を図り、工場全体のスループット（利益）を最大化するか」が重要となる。このため、ボトルネック工程の要因分析や、生産リードタイム短縮、最終製品の品質向上の可能性探索に資する分析への期待が大きい。これらの分析を行う場合、工程単位で管理されている非連続的なデータを組み合わせ、「当該製品は、いつ、どの工程を経て、完成品に至ったか」の一連の生産プロセスを表すデータの準備が必要である。一般に、工程単位で取得・蓄積されている非連続的なデータには、工程間の作業順序や運転状態を示すデータが含まれていないため、分析担当者は工場内の工程の流れを捕捉するデータを別途準備しなければならない。

また、加工工程のバラエティが増えると、

工程単位の製造データであっても精度が落ちるケースが多いことにも注意が必要である。最終製品のモデルチェンジやカスタマイズが多い産業の場合、半製品の品番管理が省略され「半製品の受け払い実績が取れない」「在庫数量が分からない」という企業が少なくない。この場合、分析担当者は、工程単位に取得可能なデータの精度を確認し、実施すべき前処理内容を工程単位で決定する必要がある。

## (2) ディスクリート型生産系IoTデータに対する前処理

プロセス型生産は連続性のあるデータが対象であるため、さまざまな状態変数の定量的な相関から運転状態を判断し、データにラベリングするアプローチが有効であった。一方、ディスクリート型生産の場合は、管理されているデータが工程単位に分かれているため、工程間のデータをつなぎ合わせるアプローチが必要となる。また、半製品に関する製造データは情報の精度が十分でない場合があるため、各工程の製造データを補完する作業も必要となる。

そこで、ディスクリート型生産のIoTデータの前処理は「分析目的に必要なデータが自社のデータベースに十分揃っているか」「取引先データとの連携や自社内データを活用したデータ補完によってデータが揃い得るか」の確認から着手しなければならない。特に品番マスターが各加工・組み立て工程レベルでどのように付番され、どのように管理されているかを把握することは、データの十分性を確認する上で極めて重要である。加えて、外注されている工程のスケジュールや生産実績の情報がブラックボックス化している場合が

多いため、外注品のデータの扱いがどのようになっているかを把握することも重要となる。

データにおける十分性の確認の次に必要となるのは、各工程に散在する非連続な製造データの結合である。工程間の作業順序を示すデータは、ERP (Enterprise Resource Planning) などで管理されている工場全体のマスタースケジュールを活用することが有効である。一般に、ERPの情報は、設備の実行を管理するMESや、現場設備の稼働を制御するPLCなどの制御装置と連携しており、理論的には工程間のデータのつなぎ合わせに活用できる可能性が高い。しかし、ERPの情報が前処理作業にどの程度活用できるかの見極めを分析担当者が単独で行うのは容易でないため、ERPを保守・運用するIT部門の担当者や生産技術の担当者の協力が欠かせない。

## V ビジネスアナリティクスの実践に向けて

コンピュータサイエンスの黎明期によく使われた表現に「Garbage in, Garbage out (コンピュータに組み込まれたプログラムが正しくても、与えられたデータが誤っていれば、無効な結果しか得られない)」というものがある。これはコンピュータが普及し始めた当時、「コンピュータの判断はすべて正しく、誤った入力も正してくれる」という、コンピュータ初心者の幻想を正すために作られた警句であるが、この警句は昨今のビジネスアナリティクスブームにもそのまま当てはまる。

これまで解説してきた通り、データサイエンティストは魔法使いではなく、AIなどの

分析アルゴリズムは魔法のハコではない。IoTデータの有効活用に近道はなく、企業はデータ分析の巧拙を決める「前処理工程」に地道に取り組む必要がある。

近年、ビジネスアナリティクスの業務適用や分析基盤構築を行う際のファーストステップとして、ビジネス部門の担当者と分析担当者が一体となって小規模な分析の仮説検証を行うPoC (Proof of Concept: コンセプト検証) が採用されるケースが増えている。この取り組みは、分析手法だけでなく業務適用仮説そのものを修正することも視野に入れて分析結果をビジネス的な視点で検証し、IoTデータの業務活用方針や分析システムの構築方針を検討するアプローチである。これは大変有効なアプローチであるが、分析結果の精度が前処理工程に依存することを考慮すると、PoCにおける仮説検証には、データを取得するセンサーの開発担当者や生産設備をよく知る生産技術担当者らも関与し、センサーの仕様そのもの

のや不足データの補完方法などのデータ取得・蓄積・前処理に関する仮説検証を同時並行で行うことがより効果的である(図3)。

実際にNRIがIoTデータ活用の構想作りなどを支援する際には、ユーザー部門やIT部門に加え、生産設備やデータ取得に詳しい生産技術部門の担当者に参画してもらい、PoCを実施することを勧めている。具体的には、NRIが準備する汎用的な分析基盤(図4)上に、顧客が業務で活用しているデータを提供してもらい、NRIの業務コンサルタントと分析担当者およびシステムエンジニアが、データの前処理から分析、および分析結果のビジネス適用仮説作りに至る一連のプロセスを実施する。これらの分析結果や業務適用仮説、およびデータ収集・蓄積・前処理に関する現状課題を、顧客のプロジェクトメンバーと共有し、分析やビジネス適用仮説だけでなく、データ収集・蓄積・前処理に関する新たな対応仮説も準備しながら、より実践的なIoTデ

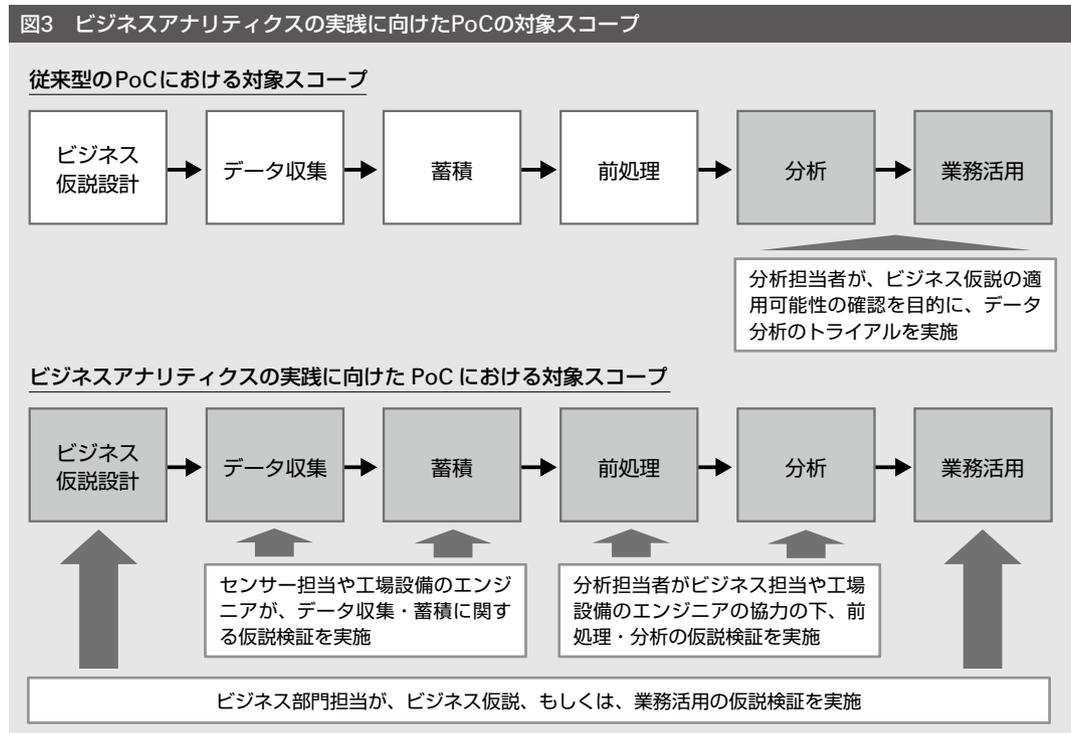
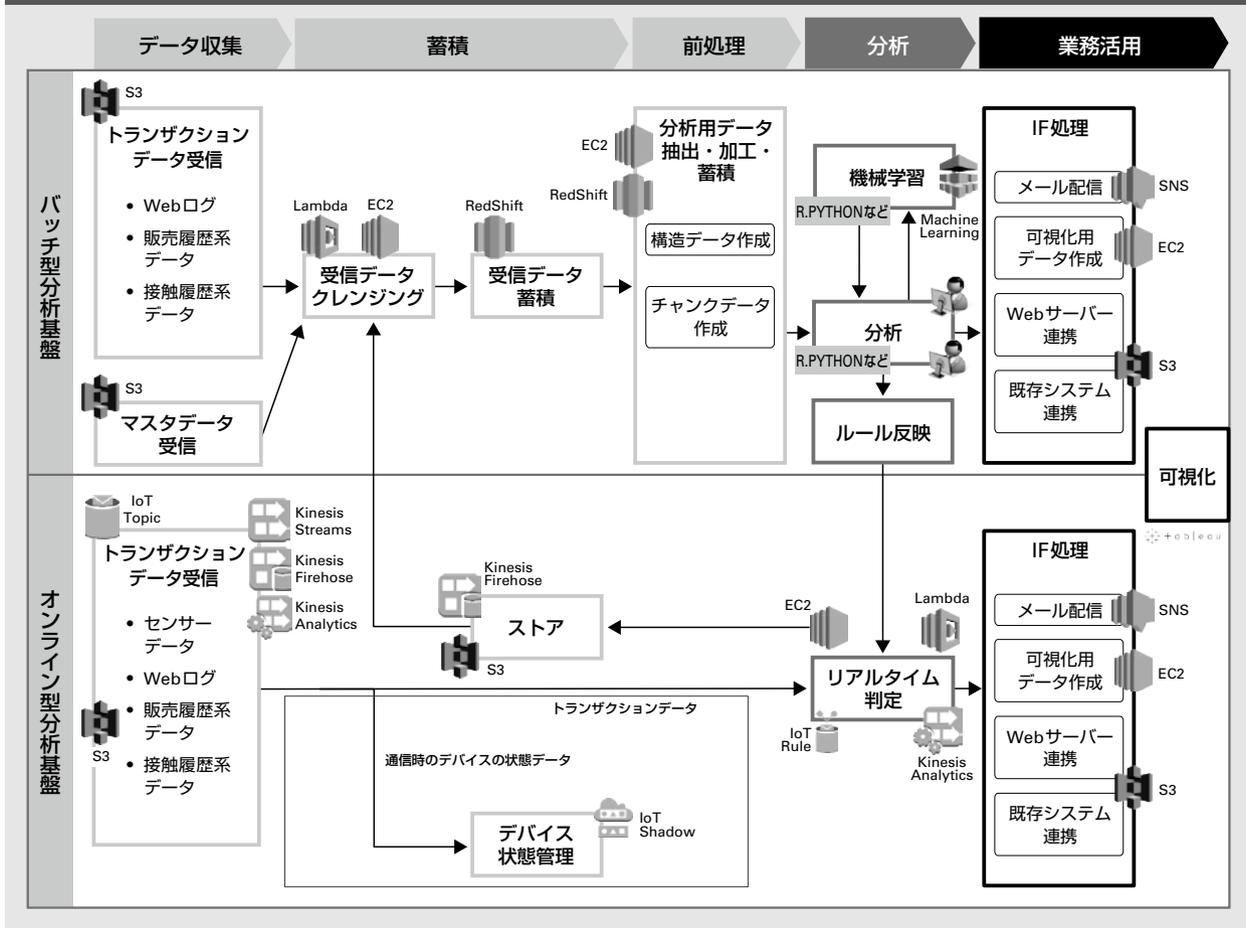


図4 NRIのビジネスアナリティクスプラットフォームの概要構成



ータ活用モデルを顧客と共に作り上げていく。本取り組みの実施により、いくつかの製造業の顧客が、自社の保有するIoTデータの活用方針を短期間で具体化でき、実際の業務適用に向けた本格検討を開始されている。

今後、デジタル技術の進展はますます進み、ビジネスアナリティクスの重要性はさらに高まっていくと予想される。これからビジネスアナリティクスに取り組む企業は、経営トップによる適切な課題設定の下、前処理工程を「分析の品質を左右する重要工程」と位置づけ、組織横断的な活動として強力に推進することが肝要になるといえよう。

著者

沼澤 優 (ぬまざわゆう)  
システムデザインコンサルティング部グループマネージャー  
専門は業務変革、ビジネスアナリティクスなど

庄司修芳 (しょうじのぶよし)  
システムデザインコンサルティング部上級コンサルタント  
専門は業務変革、ビジネスアナリティクスなど

大橋俊介 (おおはししゅんすけ)  
システムデザインコンサルティング部主任データサイエンティスト  
専門は業務変革、ビジネスアナリティクスなど