

# NTT DATA Technology Foresight 2012

## ビッグデータを活用したビジネス・アナリティクス

世界のデータ量は2年毎に倍増し、2015年には約7.9ゼタバイトに達するという試算がある。また、データの種類も、テキストや画像などの非構造化データが90%を占めると言われている。昨今、これらの大量データを分析して業務改革やサービス革新に活かすビジネス・アナリティクスへの期待が高まっている。本稿では、このような「ビッグデータ」を活用したビジネス・アナリティクスをめぐる技術動向と、それに対する当社のアプローチを紹介する。

### ビジネス・アナリティクスの技術動向

これまで、企業は社内の業務システムに格納された顧客データや取引データ等の限られた情報を活用してきた。しかし、近年のハードウェアの高性能化や新サービスの登場等により、活用できる情報が爆発的に増えつつある。その代表格は、センシング・データやライフログ・データである(図1)。センサーやスマートデバイス等の活用が拡がり、人やモノに関するデータが詳細に取得できるようになってきた。また、SNSやブログなど、個人が発信する情報も急拡大している。

ビッグデータでは、データの「量」だけでなく「多様性」と「高頻度性」にも着目する必要がある。単一ソースから取得した数値データだけではなく、複数のソースから取得したテキストや画像等の非構造化データを処理する必要がある。また、これまでは数分/数時間単位で取得していたデータが、毎秒/コマ何秒単位という非常に短い間隔で取得できるようになる。これらの特徴を持

つデータの管理や処理を、従来のデータベース技術やデータ処理技術で行うことは困難である。

このようなビッグデータを我々はどうのように活用すればよいのだろうか。ビジネス・インテリジェンス(BI)は、ビジネス環境の「見える化」とともに発展してきたが、ビッグデータの活用においては、「見える化」にとどまらず、「最適化」や「予測」までをもスコープに入れたビジネス・アナリティクス(BA)が成功のポイントとなる。

BAには、大きく3つのタイプがある。タイプ1は、データ間の関係性や規則性の発見である。例えば、顧客属性と購買履歴の因果関係を分析することによる優良顧客の見極めがある。タイプ2は、一定の制約条件下における最適解の発見である。これは、限られたリソースをいかに有効に活用するかという問題に有効であり、例えば、人員のシフトスケジューリングや発注量の最適化などがある。タイプ3は、ユーザーを理解することによる将来の予見である。これにより、一歩先回りした気の利いたサービスが提供できるようになる。例えば、金融サービスにおける不正検出や異常検知、レコメンデーションが挙げられる。

BAの実現に向け、国内外の多くのITサービス事業者は、既に大規模分散処理(Hadoop)やストリームデータ処理(CEP:Complex Event Processing)を活用した様々なソリューションを発表している。また、社内にはアナリティクスの専門チームを設置するなど、事業展開に向けた活動が着々と進みつつある。



図1 ビッグデータ活用への期待

## テクノロジーの進化

実際にビッグデータや BA に関する技術はどこまで進んでいるのであろうか。我々は、情報の増大・多様性の軸（ビッグデータの側面）と分析高度化の軸（アナリティクスの側面）で、それぞれ進化が進むものと考えている（図 2）。

既に、センサー等から発出される大量の数値データの分析と活用は進みつつある。今後、ビッグデータの側面としては、非構造化データなどのデータ多様化が進んだ後に、複数ソースのデータを重ね合わせるデータフュージョンが実現していくであろう。例えば、交通分野では、テキスト表現の交通情報や画像表現の気象情報等のデータを統合し、渋滞状況を分析する取り組みがある。ここでは、空間的粒度やデータ取得タイミングの異なるデータをいかに補完して重ね合わせるかが技術的なポイントとなる。

また、アナリティクスの側面としては、そのような多様なデータの分析技術が進み、さらに将来の予測や実世界のコントロールまでが実現されるであろう。例えば、小売店が、売れ行きに基づいて需要を予測し、入荷すべき商品と数を自動的に算出することで、実際の発注量を判断している事例がある。今後は、過去の売上実績だけでなく、周辺のイベント情報や天候などの様々な情報に基づいた高度な予測が実現していくであろう。

そして、5～10 年後には、これらの技術が統合され、従来は人間でしか成し得なかった俯瞰的な意思決定を機械が積極的に支援する世界が来ると考えている。その萌芽となる事例としては、米国防総省の先端研究プロジェクト（IARPA）における Open Source Indicators プロジェクトがある。この取り組みでは、Twitter や検索エンジンのクエリ、街角の監視用ウェブカメラなど、あらゆるデータを統合して自動解析し、大規模犯罪など重要な社会的事件の予測に役立てている。ま

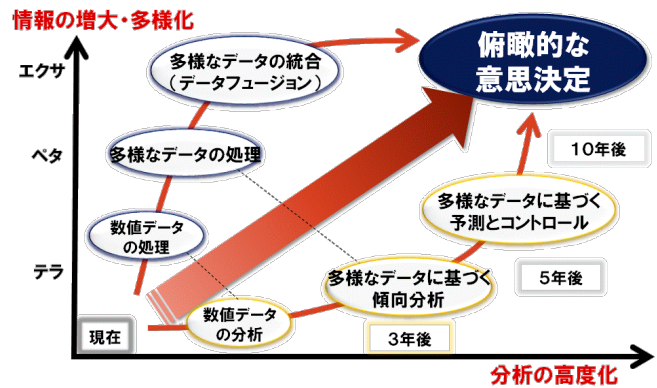


図 2 進化のステップ

だ研究段階とは言え、近い将来にはテロや大規模犯罪を未然に防げるようになるかもしれない。

## NTTデータのアプローチ

NTT データでは、ビッグデータの活用を、BI の延長として捉えている。

従来の BI は、集計分析に基づく、いわゆる「見える化」に留まっていた。しかし、「見える化」だけでは、データから高度な知見を導出するのに限界がある。NTT データでは、従来からの「見える化」と BA を含めたものを広義の BI と捉え、多様な業務分野に対するデータ分析コンサルティングの経験に基づき、BI を 4 つに分類している。全件データを即座に集計・分析するための「集計分析型 BI」、データ粒度に合わせた様々なバリエーションを分析してルールを発見する「発見型 BI」、シミュレーションや最適解を探索して業務を最適化する「WHAT-IF 型 BI」、データのリアルタイム分析による将来を先回りしたサービスの提供を行う「プロアクティブ型 BI」である。このうち、「発見型 BI」「WHAT-IF 型 BI」「プロアクティブ型 BI」の 3 つが BA の領域であり、それぞれ前述のタイプ 1、タイプ 2、タイプ 3 に相当している。

NTT データでは、前述した 4 つの BI を軸に、データ分析方法論「BICLAVIS®」を整備している。この中心にあるのが、分析パターンを目的別に分類した分析シナリオ類型である（表 1）。こ

表1 BAに関連する分析シナリオ類型

タイプ	分析シナリオ類型	分析例
発見型BI (タイプ1: 関係発見)	ターゲット型	<ul style="list-style-type: none"> <li>●購買データ、Webデータを用いた優良顧客の抽出</li> <li>●ターゲット顧客の選定(顧客ターゲット)</li> <li>●レセプト・健診データを用いた生活習慣病予備軍の抽出</li> </ul>
	リスク・スコアリング型	<ul style="list-style-type: none"> <li>●顧客特性に応じた解約リスクのスコアリング</li> </ul>
WHAT-IF型BI (タイプ2: 最適化)	予測・制御型	<ul style="list-style-type: none"> <li>●POSデータにもとづく需要予測と基準在庫の分析(自動発注)</li> <li>●システム・ログを活用した最適人員配置とシフトスケジューリング</li> </ul>
プロアクティブ型BI (タイプ3: 将来予見)	異常検出型	<ul style="list-style-type: none"> <li>●センサーデータを用いた外れ値検出</li> <li>●申請・審査業務における不正パターンの洗出し</li> </ul>
	予兆発見型 コンテキスト アウェアネス型	<ul style="list-style-type: none"> <li>●ログを用いたインシデントの予兆発見</li> <li>●Webアクセス行動にもとづくパーソナライズ</li> <li>●購買履歴やサービス利用履歴を用いたレコメンド・サービス</li> </ul>

の分析シナリオ類型に基づき、NTT データでは分析を効率的に行うためのテンプレートを整備している。

また、データウェアハウス/ビジネスインテリジェンス・ラボ®の取り組みでは、BI/BA システムの上流工程支援として、基盤製品の PoC、ツール選定支援、お客様要件に即したデモ・システムの構築によるシステム要件整理の支援といったサービスを展開している。

さらに、グローバルでの取り組みも加速している。アジア地域でのトラフィックコントロール、欧州地域とのスマートコミュニティの取り組みを進めているほか、欧米・アジアの NTT データグループ会社で構成される BI Global One Team により、知見やノウハウを結集している。このように、グローバルに拠点をもつお客様への BA サービスの展開にも積極的に取り組んでいる。

## BA の取り組み事例

当社の取り組み事例を4つ紹介する。

### ①橋りょうモニタリング [異常検出型]

橋りょうの劣化検知には、多大なメンテナンスコストがかかる。そこで、橋りょうにセンサーを設置し、大量かつリアルタイムに送信されるひずみ等のデータから異常を検知する実証システムを構築し、検証を行った。(詳細は別紙を参照)

### ②CPFRにおけるサプライチェーンマネジメントの最適化 [予測・制御型]

CPFR (Collaborative Planning, Forecasting, and Replenishment) とは、メーカーと小売が協調して商品販売計画を立案し、欠品防止、在庫削減を実現する取り組みである。この事例では、商品別に短期・中長期の需要予測を行い、それに基づき発注方式(いつ何をどれだけ発注するか)を具体化した。さらに、その効果を評価するため、欠品や納期など、現実的な制約を織り込んだシミュレーションを行った。通常、在庫発注方式の変更は多大なリスクを伴うものであり、シミュレーションによる事前評価の意義は大きい。メーカーと小売といった複数ソースのデータの組み合わせ、また在庫のリアルタイムな把握など、ビッグデータの分析の萌芽的な事例と言える。

### ③BPO におけるシフトスケジューリング [予測・制御型]

BPO(Business Process Outsourcing)とは、企業が中核業務以外の業務を外部に切り出し、業務プロセスやリソースを抜本的に見直す取り組みである。この事例では、複数の業務を処理している事務センタを対象に、各業務量の予測を行うとともに、業務の処理時限や要員スキルなど現場のリアルな制約を考慮したシフトスケジューリングを行う仕組みを構築した。大規模事業所や複雑な制約条件におけるシフトの最適化は、組合せ爆発を起こすため難易度が高いが、本取り組みにより最適なスケジュールを自動生成することが可能になり、BPO 効率の最大化を実現できた。

### ④健康保険組合の医療費削減施策 [ターゲット型]

健康保険組合では生活習慣病が重症化する被保険者の増加に伴う医療費の増加が問題となっている。そこで、現在の健康状態から、将来重症化し医療費が急増する可能性の高い、「重症化予

備軍」の被保険者を抽出し、早期に保健指導等の対策を打つ取り組みを行った。過去の大量のレセプトデータから生活習慣病の重症化及び併発パターンをデータ・マイニングにより発見し、パターンに該当する高リスク患者をリストアップして保健指導を行うことで、健康保険組合全体として医療費の抑制を実現した。

## 実現アーキテクチャ

ビッグデータを活用したビジネス・アナリティクスを実現するアーキテクチャとして、NTT データでは3つのレイヤを定義している(図3)。第1レイヤはビッグデータ基盤であり、超大量で多様な形式のデータを処理するためのエンジンである。Hadoop やインメモリ DB などの基盤ソリューションがこれに相当する。第2レイヤは、ビジネス・アナリティクス基盤であり、高度なデータ分析機能を有したライブラリや、データマイニングパッケージ等がこれに相当する。第3レイヤは、分析テンプレート/アプリケーションで、業務や分野ごとの分析テンプレートや、類型ごとの分析シナリオなどを提供する。

NTT データでは、各レイヤをトータルにサポートするとともに、特に分析を行う第3レイヤにおける付加価値提供に力を入れている。前述した「BICLAVIS®」を武器に、今後さらに多様な領域におけるノウハウを蓄積し、アナリティクスサービスの高度化を目指していく。

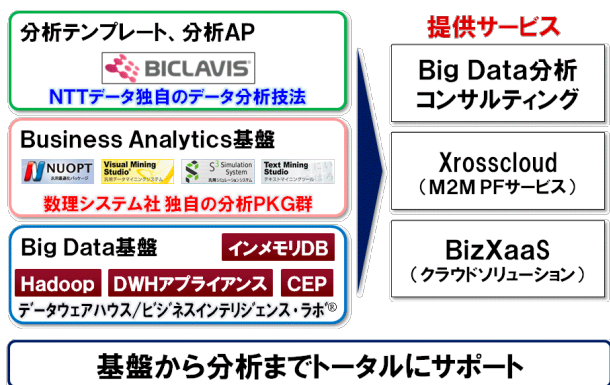


図3 実現アーキテクチャ

## 今後のビジョン

冒頭で、BA について今後の進化の方向性を示した。これらは、先に示した4つのBIそれぞれを高度化していくことで実現することができる。NTT データでは、5~10年後の俯瞰的な意思決定の実現に向け、既に具体的な取り組みを始めている。

例えば、トラフィックコントロールの分野では、ビーコンやプローブ等のセンサーデータや交通情報をインプットに大規模なマルチエージェントシミュレーションを行い、渋滞予測や信号・車線制御等を実現するための研究を行っている。

また、海運ロジスティクスの分野では、世の中で発生するインシデントにより特定海域が航行不能になる場合を想定し、インシデントの予兆を察知して多国間を航行する船舶の寄港地発着枠を確保したいというニーズもある(図4)。経済活動に影響を与える動向をいち早く察知し、リスク回避することの重要性は高い。

ビッグデータを活用したビジネス・アナリティクスは、環境の変化に合わせて社会やビジネスを最適化していくためのキーテクノロジーと言える。NTT データは、アナリティクスに関する豊富な経験、IT 基盤やツールに対する中立性、グローバルなBIソリューションの構築・展開体制を強みに、戦略的な情報活用による新たな「変革」をお客様とともに実現していきたい。

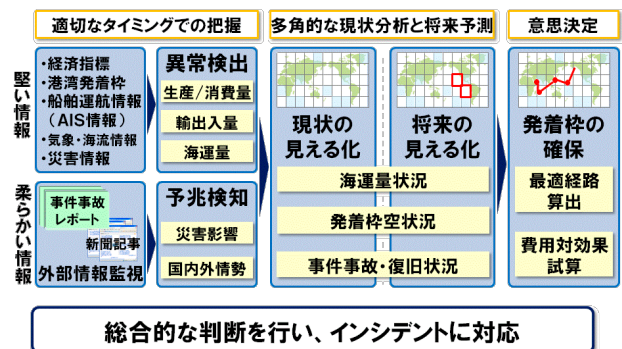


図4 海運ロジスティクスの例