

ビジネス・アナリティクスの最新動向 (1)

中川慶一郎 (なかがわ けいいちろう)
 株式会社NTTデータ数理システム
 生田目 崇 (なまため たかし)
 中央大学理工学部

1. はじめに

情報技術が急速に進化するなか、そこで発生するビッグデータとどのように向き合うかが企業経営者の大きな関心事になっている。

ビッグデータに関する論議は当初、分散並列処理により大量データの蓄積・処理を行うHadoopや時系列的に発生するストリーム・データを処理するCEP (Complex Event Processing: 複合イベント処理) のように、データをどう蓄積、検索、処理するかというIT基盤の話に終始してきた。しかし、「ビッグデータを用いて何をするのか」ということに焦点が移るにつれて、データ分析を通じてビジネスに資する洞察を引き出す「ビジネス・アナリティクス」や、それを実践する「データ・サイエンティスト」に関心が向けられるようになった。

本連載では、近年急速な発展を遂げてきたビジネス・アナリティクスの動向について4回にわたり解説する。第1回目の今回は、ビッグデータというパラダイムの本質と、そこで求められるビジネス・アナリティクスについて概説する。つづく第2回目は、中核技術である分析手法における近年の潮流についてまとめる。第3回目では、ビジネス・アナリティクスを支えるIT基盤について紹介し、最終回である第4回目はビジネス・アナリティクスの現状の問題点および今後の方向性について筆者らの考えを示す。

2. ビッグデータ

データを経営に活かすということは、何も新しいことではなく、これまでさまざまな形で議論されてきた。しかし、ここにきてビッグデータやビジネス

・アナリティクスが注目を浴びている背景には、情報の持つ価値に対する再認識と分析技術による競争優位性への期待がある [1]。

これまでにさまざまなところで言及されているように、ビッグデータには3つのV、すなわちVolume (量)、Variety (種類)、Velocity (頻度) という特徴がある。このようなデータが爆発的に増えた結果、従来のRDB (Relational Database) ではためきれない、もしくは処理しきれないといった現象が起り、ビッグデータが注目されるきっかけとなった。

しかし、ビッグデータの3つのVは、あくまでデータの特徴であり、ビッグデータの本質を理解するうえでは十分ではない。また、ビッグデータと呼ばれているデータも個々に見てみると、どういった状況で発生するデータなのか、なぜ注目されたかといった面で背景が大きく異なる。そこで、ここではビッグデータを3つの系譜に分けてみる。

1番目は従来型ビッグデータの系譜である。例えば、ウェブのアクセス・ログやPOSデータなどは、ビッグデータがブームとなる前から大規模データとして認識されてきた。このようなデータは業務システムに付随して発生するデータであり、分析目的も明確な場合が多い。

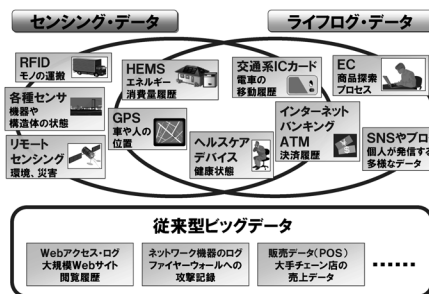


図1 ビッグデータの系譜

2番目はRFIDや加速度センサのようなセンサ機器のログ、いわゆるセンシング・データの系譜である。センサは人手の場合とは異なり、休むことなく、正確に、場合によっては劣悪な環境下でもデータの収集が可能であり、工場での異常検知、社会インフラの劣化監視や農業での生育状態の観測など、さまざまな利用が広がっている。

近年、センサ・デバイスは低価化に伴い急速に普及し、またITと融合するようになってきた。その結果、ネットワークを介してセンシング・データを簡単に収集できるようになり、ビッグデータとして認識されるようになった。

3番目はライフログ・データの系譜である。ライフログ・データについてはまだ定義すら曖昧な状況ではあるが、SNS、ブログといったソーシャルデータやECでの購買履歴など、“生活まわり”の記録と考えてよい。こうしたデータへの期待は、これまでの「行動結果の分析」から「原因や動機の分析」にまで踏み込めるようになるのではないかという点にある。ただし、ライフログ・データの活用については個人情報とも密接に関わるため、法整備を含めていまだ発展途上の分野と言える。

なお、家庭内の電力使用量をモニタリングするHEMS（Home Energy Management System）やヘルスケア・デバイスから得られるデータはセンシング・データ、ライフログ・データの両面を持つ。

3. ビッグデータ活用の発展レベル

ビッグデータを上手に活用することによって、個人、企業、社会に対して新たな価値を提供できるのではないかという期待は共通している。しかし、ビッグデータをどのように活用するかという点については、企業によって、また同じ企業でも部署によって大きく異なる。ここでは、ビッグデータの活用段階とそのビジネス領域を時間軸に合わせて4つの大きな波として整理する [2]。

第1の波は、本来業務として大規模データを処理するビジネス領域である。ここでの対象データは、例えばコンビニエンス・ストアのPOSデータや、通信会社の利用履歴であり、これらを活用した販売促進や解約防止を目的とした分析は国内外を問わずこれまでにも行われている。第1の波の特徴は、

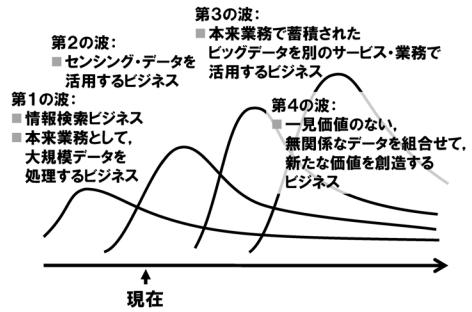


図2 ビッグデータ時代の4つの波

ビッグデータという単語が出現する前から、大規模データを活用することが、すでにビジネスに組み込まれていることである。

第2の波は、センシング・データを活用するビジネス領域である。この領域でのビッグデータ活用は、M2M（Machine to Machine）ネットワークを抜きにして考えることはできない。M2Mネットワークとは、パソコンや携帯電話だけでなく、自動車や自動販売機といった市中の機器から、橋やトンネルといった社会インフラに至るまでのすべてをネットワークにつなごうとするもので、スマート・シティやスマート・コミュニティを実現する基盤となるネットワークである。

あらゆるものをネットワークにつなげることによって、ヒト・モノ・カネの動きに関する膨大で詳細なデータをリアルタイムに収集することができるようになった。現在は、これをいかに活かすかが問われる時代といえる。

第3の波は、これまでに本来業務で蓄積したデータを二次活用して別のサービス・業務で活用するような領域である。このとき、一次活用とは業務課題解決のためのビッグデータ活用であり、前述の第1、2の波に相当する。これに対して、二次活用は上位の経営課題解決のためのビッグデータ活用である。

ここで第2の波と第3の波の違いを具体的に考えてみよう。現在、飲料の業界では多くの自動販売機がネットワークを介してデータセンタにつながっており、商品の実売本数や在庫本数、滞留時間、売れ筋などの販売情報をリアルタイムに把握することができる。自動販売機は店舗と異なり、つい最近まで“蓋を開けてみないと分からない”という状況であった。ここで、販売情報をリアルタイムに把握す

るのは業務レベルの課題であり、こうしたビッグデータ活用が第2の波すなわち一次利用である。

一方、配送担当者は、この地域、あるいはこの季節にはこのくらい売れるという経験的な情報が頭の中に入っており、それを元に商品の積み荷のバランスを調整していた。このような勘と経験に基づく暗黙知を継承していくことは極めて困難である。暗黙知を形式知に変え、組織的に知識を継承していくことは経営レベルの課題であり、そのためのビッグデータ活用が第3の波すなわち二次利用である。

最後に第4の波は、一見価値のない無関係なデータを組み合わせ、新たな価値を創造するビジネス領域である。ビッグデータ活用というと、このレベルを期待されている場合が多いように思われるが、実際にはまだ具体的な事例は乏しく、ビジネス化にはまだ時間が必要と考えられる。ただし、第4の波についても、いくつかの萌芽が現れている。

その1つが政府によるオープン・データの取り組みである。もともと、政府は公共的な立場からさまざまなデータを収集している。これらのデータを民間に開放することで、新しい産業を創出できないかの検討が始まっている。

もう1つがスマートフォンなどへの情報提供サービスである。例えば、「Google クライシスレスポンス」[3] というサービスでは、自動車メーカーとの協力によって、カーナビシステムから収集される通行実績情報を Google マップに重ね、これを参照できるようにした。その結果、災害発生時に車で被災地を目指す人々が、車が通行可能なルートを見つけることができ、たいへん高い評価を受けた。

このような情報提供サービスは、利用者の圧倒的な支持がビジネスを牽引していくということが大きな特徴と言える。また、分析という技術面から見ると、高度な分析もさることながら、マッシュアップというさまざまな情報を重ね合わせる考え方が使われている。

4. ビジネス・アナリティクスの新たな挑戦

ビッグデータは、ともすれば技術の問題として捉えられがちであり、ビジネス・アナリティクスに対する認識も、統計解析やデータマイニングを含む従来のビジネス・インテリジェンスと明確な差がある

とは言えない。しかし、ビッグデータは新しいものの見方、捉え方、考え方を与えるパラダイム・シフトであり、それに合わせてビジネス・アナリティクスも、これまでのビジネス・インテリジェンスではなしえなかった新たなチャレンジの時期を迎えているといえよう。

以下では、ビッグデータがもたらすパラダイム・シフトとそこで求められるビジネス・アナリティクスについて説明する。

4.1 目をつぶってきた事実に目を向ける

1番目のパラダイム・シフトは、これまでは把握するのは困難だとして目をつぶっていた、あるいは諦めていた事実に積極的に目を向けていこうとする“ものの見方”の変化である。

例えば、POS データは、販売という事実のデータである反面、購入者が本当に欲しい商品はそこにはなかったの、仕方なく代替品として買ったかもしれないといった、背後にある“なぜ”を考える手掛かりが欠落している。これまで、仮説検証という形で推測する、あるいは市場調査を実施することで背後を考えることはあるが、多くの場合事実データから背後を理解することは困難だとして目をつぶってしまっていた。

これに対して、最近では EC サイト内での商品の検索履歴や実店舗内で取得された動画データを活用して、購買の背景を理解しようとする試みがなされている。例えば、メニューの情報サイトの閲覧履歴とスーパーマーケットの ID 付 POS データを連携させることによって、どのような思考を経て購買に至ったか、また購買したアイテムをどのように利用、消費したかに踏み込むことができる。

「見える化」は情報分析・活用の第一歩であり、その重要性は何ら変わらない。ここでのビジネス・アナリティクスのチャレンジは非構造化データのように扱いづらいデータも積極的に活用して、一歩進んだ「見える化」を実現し、意味のある洞察を導き出すことである。

4.2 “どうにもならない時間間隔”を埋める

2番目のパラダイム・シフトは、状況を理解する際に生じる“どうにもならない時間間隔”の捉え方、また、これを超えようとする意識の変化である。

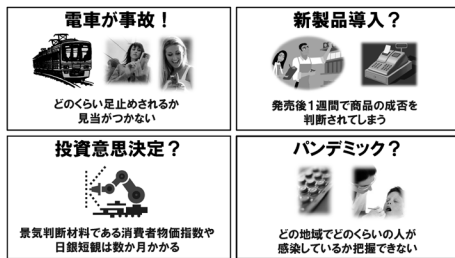


図3 時間間隔が重要な意思決定

例えば、電車の事故が発生した状況を考えてみよう。駅で足止めされた人たちの多くは、その場で事故の状況を知りたいはずであるが、公式に状況がわかるのは、ある程度時間が経ってからであろう。このような状況に直面すると、最近では多くの人がTwitter上のつぶやきから状況を推測して、迂回の判断をするのではないだろうか。

ビジネスの世界に限らず、世の中には“どうしようもない時間間隔”により、結果として本来は手段を講じるべきときに何も手を打てないという状況がある(図3)。

こうした状況では、正確さは多少犠牲にしてでもリアルタイムに状況を把握し、何とか手を打ちたい。そのために使える情報は何でも使いたいという気持ちに駆られるであろう。ビジネス・アナリティクスの新たなチャレンジはこの時間間隔を埋めることである。

インフルエンザが流行すると、タミフルの備蓄とその配分が話題になるが、公衆衛生の観点からは、インフルエンザを発症した患者数が判明する前に予防策を講ずることが肝要である。しかし、罹患者数の実数がわかってからでは遅い。Googleが提供する「インフルトレンド」[4]は、検索数とインフルエンザの流行の関係をもとに、流行の兆しをリアルタイムに知る仕組みであり、この時間間隔を見事に埋めることに成功している。

経営にとってスピードは命であり、多少精度を犠牲にしても、この“どうにもならない時間間隔”を克服できるのであれば他社に対する強力な武器となる。ここでのビジネス・アナリティクスは、予測というよりはむしろ、ゴミを含むデータや直接的には関係のないデータであったとしてもそれらを積極的に取り込んでいくことで、関連する情報から“今そ

こにある現実”をいち早くあぶり出すことである。

4.3 複雑な状況で判断する

3番目のパラダイム・シフトは、複雑な状況の中で少しでも合理的な判断を行うためにビッグデータを活用していこうという考え方の変化である。

家電製品を例に挙げると、多くの人が購入前に価格.comで値段や性能を比較したり、Amazonで他の人は何を買っているのか、実際に買った人のどんな評価をしているかといったことを調べたりするなど、インターネットから得られる情報を最大限に活用して判断している。注目すべきは、どの商品をどこで買うべきかという消費者の“課題解決のプロセス”は格段に進化しており、以前とは比較にならないほど合理的になっているということである。

もう少し複雑な状況として、取引企業の審査を考えてみる。そこでは単純な財務内容だけでなく、資本関係や取引先関係、コンプライアンスや風評など網羅的な判断が求められる。ここでのビジネス・アナリティクスは、財務内容の急激な変化や同業他社との差異を比較するといった定型化しやすい分析だけでなく、散在するビッグデータをもとに資本関係や他社との取引関係の可視化したり、ネット上の評判を集約したりするなど、アドホックな分析も組み合わせることで求められる。

現在、世界最強のチェス・プレーヤーは、スーパーコンピュータでもチェスのチャンピオンでもなく、ごく普通のPCを使ったアマチュアのチームである。2005年に行われたフリースタイルのチェスで、普通のPC3台を駆使した2人のアマチュアが、臨機応変にさまざまな手を探索することでスーパーコンピュータと組んだチェスのチャンピオンを破り、新たなチャンピオンになった。これは、複雑な状況を前にして合理的な判断を下すために、ビジネス・アナリティクスは何をすべきか、ということについて興味深い示唆を与えている。

今後ビジネス・アナリティクスが立ち向かうべきチャレンジは、かつてJ. C. R. Licklider [7]が提唱した「コンピュータによる知能増幅(Intelligence Amplification)」というコンセプトのように、分析によっていかに人間の意思決定能力を強化できるかということである。特に、現状の「見える化」だけではなく、このまま行くとどう状況が変わるか、あ

るいは、ここで行動を起こすと結果はどう変わるのかといった将来の「見える化」を実現することがポイントとなる。

危機管理のように、人間はこれまでも複雑で不確実な状況の下でも、わからないなりに何らかの判断を下してきた。一方、ITの世界でもIBMのWatson [5] やGoogleの自動運転 [6] に見られる知的な情報処理の取り組みが盛んに行われるようになってきている。

より合理的な判断を実現するためには、コンピュータはビッグデータの分析によって状況判断や意思決定に必要な材料を準備し、人間は俯瞰的、総合的な判断を担う、という役割分担に基づき、課題解決のプロセスを再構築していくことが必要となる。

5. データ・サイエンティストの育成

ビジネス・アナリティクスが注目されるに従って、それを実行する「データ・サイエンティスト」の不足が問題視されている。日本経済新聞は、向こう10年で最大25万人のデータ・サイエンティストが不足すると報じた。

一般にデータ・サイエンティストは、IT、統計分析、ビジネスの素養を兼ね備えた人材とされている。ただし、最も困難であり重視すべきは、業務に分析を組み込むデザイン能力と課題解決のためのモデリング能力の開発である。

これらの能力は実践を積むことにより鍛えられるものではあろうが、組織として系統的にデータ・サイエンティストを育成するためには、BICC (Business Intelligence Competency Center) のような情報分析活用の専門組織の立ち上げ、情報分析活用の成熟度の評価、人材像の定義、分析方法論の整備も含めた分析技術の蓄積が求められる。

こうした人材育成については、国内外のIT企業

を中心として、組織的にされるようになってきた。また、データ分析基盤を開発する企業が、大手IT企業に買収される例も数多く報道されており、単に要素技術に秀でているだけではなく、ビジネス・サイクルの中で分析を通してビジネスを全方位的に支援・機会発見していくことの重要性が問われる時代となってきたことを示していよう。

参考文献

- [1] 中川慶一郎, “ビッグデータ時代におけるビジネス・アナリティクス”, 情報未来, No. 40, pp. 18–21 (2013).
- [2] 野村総合研究所, 「ビッグデータ革命」, アスキー・メディアワークス (2012).
- [3] <http://www.google.org/intl/ja/crisisresponse/>
- [4] <http://www.google.org/flutrends/>
- [5] <http://www-03.ibm.com/innovation/us/watson/>
- [6] A. Fisher: “Inside Google’s Quest To Popularize Self-Driving Cars,” *Popular Science*, Retrieved 18 September (2013).
- [7] J. C. R. Licklider: “Man-Computer Symbiosis,” *IRE Transactions on Human Factors in Electronics*, March (1960).

略歴

中川 慶一郎 (なかがわ けいいちろう)

2000年早稲田大学大学院理工学研究科博士課程満期退学。博士(工学)。2012年株式会社NTTデータ数理システム取締役。マーケティング・エンジニアリングおよびビジネス・アナリティクスに関する研究開発、コンサルティングに従事。

生田目 崇 (なまため たかし)

1999年東京理科大学大学院工学研究科博士課程修了。博士(工学)。2013年中央大学理工学部経営システム工学科教授。マーケティング・サイエンスおよび経営科学に関する研究に従事。