

# ビッグデータビジネスの概要

鈴木 良介

2011年以降、情報・通信業界にとどまらぬ広範な事業者により「ビッグデータ」への注目が高まっている。これは、事業における付加価値の向上や、社会システムのより効率的な運用への活用が期待されるためである。ビッグデータとはどのようなものであり、なぜこのタイミングで注目されるに至ったのだろうか。国内外の先鋭的な事例を通じて、事業の革新に向けて、ビッグデータをどのように活用していくべきであるのかを検討する。

キーワード：ビッグデータ、マーケティング、プライバシー

## 1. 進むビッグデータの活用

2010年代の情報・通信活用における潮流として「ビッグデータ」への関心が高まっている。本稿では、ビッグデータの一般的な特徴および、活用事例と課題を紹介する。

ビッグデータに関する標準的な定義はなされていない。本稿ではビッグデータを、「事業に役立つ知見を導出するための『高解像』『高頻度生成』『多様』なデータ」と定義する。

ビッグデータはその名称から、データサイズの大きさに注目が集まりがちだ。また、「ビッグデータは何テラバイトか、何ペタバイトか」といった問いもよく聞かれる。しかし、データサイズありきの議論は適切ではない。経営者の立場で考えた場合、データサイズが小さかろうが大きかろうが、収益の向上など事業に役立つかどうかという点が重要である。

データサイズが大きくなることはあくまでも結果であることに留意すべきだ。「顧客に対してより高い付加価値の提供をしたい」という目的を実現するために、「個々の顧客や製品の状況を深く理解したい」「時々刻々と理解し、即時に対応したい」「多面的に検討・分析を行いたい」といった要望が生まれる。それらを充足するために、「高解像」「高頻度生成」「多様」なデータが求められる。そして、そのような特性のデータを収集してみると、データサイズが「結果的にビッグ（大容量）」となるにすぎないのである（図1）。

ビッグデータの特性のうち「高解像」とは個々の顧

客や状況に最適な対応を行うことを意味する。販売促進活動を例とすれば、「30代男性」に対して画一的な施策を講ずるのではなく、個々の顧客の趣味嗜好や過去の購買履歴に応じた施策を講ずることなどが相当する。

「高頻度生成」とは、データの取得・分析頻度を高めることを意味する。流通事業者が3カ月に一度の購買データ分析をもとに陳列棚配置の最適化を行うようなことはこれまで行われてきたが、顧客が手に取ったものの陳列棚に返してしまった商品を店内で再び推奨するようなりアルタイムでの施策は高頻度生成データの活用に相当する。

「多様」とは、データの種類が増大することを意味する。例えばこれまでであれば用いられることがなかった店舗内の防犯カメラ映像を、利用者合意のもとマーケティング目的で二次活用していくことなども考えられる。

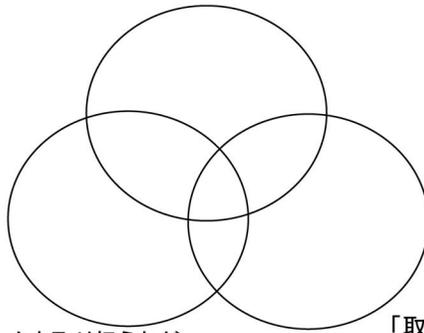
データ活用の重要性が説かれるのは、今に始まったことではない。「データマイニング」「ビジネスインテリジェンス」あるいは「センサネットワーク」や「ユビキタスネットワーク」「ライフログ」なども、似たような考え方のもとに提唱された概念だ。しかし、こうした概念の中には、市場の確立に至っていないものや、活用場面が特殊な分野に限定されているものも少なくない。では、なぜ改めて「ビッグデータ」をキーワードとして情報・通信の活用を検討する必要があるのだろうか。

背景には、2001年以降の10年間において、情報・通信の活用による「電子化・自動化」が大きく進展したことがある。これにより、解析に利用できるデータの取得や蓄積が手間的にもコスト的にも身近なものとなった。

例えば、JR 東日本の共通乗車カード・電子マネーで

すずき りょうすけ  
(株)野村総合研究所 ICT・メディア産業コンサルティング部  
〒100-0005 千代田区丸の内1-6-5 丸の内北口ビル

「取り扱うデータの解像度が高く、  
個別要素に関するデータであること」



「リアルタイムデータを取り扱うなど、  
取得・生成頻度が高いデータであること」  
(時間的解像度が高いこと)

「取り扱うデータが  
多様・非構造であること」

これらの特性を有するデータは結果として、  
「ビッグ」である(データサイズが大きい)ことが多い。

図1 ビッグデータの3つの特性 (出典：野村総合研究所)

ある Suica の提供が開始されたのは 2001 年のことだ。また、位置情報サービスの基盤となる位置データ受信機能の携帯電話への搭載も 2000 年代に大きく進展した事例の 1 つである。ブログ、ツイッター、フェイスブックなどの消費者による情報発信を促すサービスが発展し、消費者の生々しいコメントが蓄積・活用できるようになったのもこの 10 年のことにすぎない。

消費者を取り巻く情報・通信環境の成熟と同様に、企業における情報・通信活用の段階も大きく進展した。過去十数年間の情報化投資は、主として業務の「電子化・自動化」を目的としたものであった。しかし、電子化・自動化が一巡した今、その結果蓄積されたデータをどのようにして活用するか、収益につなげていくのか、という点が関心の中心となっている。このような問題意識の高まりを受けて、ビッグデータに対する関心が高まっていると考えられる。

## 2. 販売促進領域で進むビッグデータの活用

ビッグデータはさまざまな領域で活用可能だが、販売促進目的での活用が先行している。今後のデータ活用の方向性を示す事例をいくつか紹介したい。

**Shopperception の事例** EC (電子商取引) サイトでは当たり前に行われている「商品に関心を持つなど、購買に至るまでの導線分析」というデータ分析を、店舗で実現する事例も出てきた。

例えば、米国のベンチャー企業である Shopperception は、マイクロソフト社が提供する安価なモーションキャプチャシステム Kinect を用いた購買態把握ソリューションを提供している。これは陳列棚に設置されたカメラによって、どの商品が手に取られているのか、ということを機械的に記録・分析する。

このようにこれまでの POS データ、つまり「レジ通過時に把握される販売時点 (Point of Sales) データ」に加えて、POB データ (Points of Buying) の取得が容易になってきた。POB データがあれば、POS データだけでは判断不能であった「興味は持たれたけれど、購買に至らなかった商品」と「まったく興味が払われなかった商品」の区別ができ、限られた販売促進費用の最適投資などが可能となる。

**Evian Smart Drop の事例** Evian Smart Drop は、Wi-Fi 接続機能が付いた「ボタン」であり、ミネラルウォーターの販売促進を目的に作られた機器だ。外見は国内で一世を風靡した「たまごっち」に似ており、キッチンタイマーのように冷蔵庫に貼り付けることができる。ボタンを数回押すだけで、「2 リットルのペットボトル 12 本を○月○日に届けよ」という依頼ができる。2013 年にフランスで提供を開始する予定である。なお、ドバイでは「ボタンひとつでピザが届く」という類似のサービスがある。

これらは EC サイト上で提供されているワンクリッ

ク発注などに近いものだ。消費者の移ろいやすく、揮発しやすい購買意欲をリアルタイムに拾い上げることを可能とする。

**ロイズ銀行の検討事例** 個人向け金融サービスに関する事例もいくつか紹介したい。金融はそもそも情報産業であり、機微なデータの取り扱いに関する諸条件を満たすことができれば、データ活用とは相性が良い。

極端な事例として、ロイズ銀行グループが検討している取り組みを示す。顧客へ預金残高を報告する際に、現状の残高と月々に発生する支払額を踏まえて有用な商品やサービスをお進めするものだ。ロイズ銀行における個人向け銀行業務の責任者は、本サービスに関する基本的な考え方として「われわれが持つ顧客に関する詳細な情報は、これまでのように銀行のリスクマネジメント目的に限らず、顧客に対してより適切な知見を示すために用いていくことも可能である」と述べている。

**ZestFinance の事例** ZestFinance は、消費者向けローンのための与信評価アルゴリズム提供を行っている事業者であり、グーグルで最高情報責任者を務めたダグラス・メリル氏が経営する事業者である。

同社は、これまでであれば収入や家族構成、持ち家の有無など数十の変数で判断されていた与信を、より多くの種類のデータによって行う。それにより、これまでの評価指標であれば貸付対象とはし難かった顧客の一部に対しても貸付を実現するものである。例えば、「顧客接点となる金融事業者のサイトを借り手がどのように閲覧しているか」「借り手の居住地における物価の現況」「返済に遅延があった場合に、延期の申し立ての理由説明の状況」なども勘案対象にしているという。

**PHYD の事例** 自動車保険における PHYD も金融サービス高度化の事例である。PHYD は“Pay How You Drive”の略であり、運転様態に応じて保険料率を変更する。すなわち、危険な走行が多いと判断されれば保険料率が高くなる。危険度合いは、走行速度や、走行時間帯（夜間走行の割合）などにより評価される。また、「安全運転を心がければ、結果的に保険料金が安くなる」というわかりやすい動機づけによって事故の予防を促す。PHYD は、より多様なデータを用いることにより自動車保険という商材そのものを高度化した事例である。

### 3. ビッグデータ利用の多様化と高度化の進展

では、ビッグデータの利用からどんな効用を得るこ

とができるだろうか。ビッグデータ利用やビッグデータビジネスの効用、すなわち大量のデータを蓄積・分析することによって得られるメリットについては、「フィードバック先」と「リアルタイム性」の2つの視点による整理が有効である。

第一の視点は、「ビッグデータの解析を行った結果のフィードバック先が、個別（ユーザ個人など）に対するものであるか、それとも系全体（あるサービスの利用者全体など）に対するものか」である。例えば、ある携帯電話ユーザの位置データ活用を想定した場合、あるユーザの現在位置を踏まえ、そのユーザに対して当該位置に最適化された情報を配信することは、「個別フィードバック」となる。一方、あるイベント開始前に参加者が会場近くに集合しているなど、ある時間帯・ある地点に特定の属性の人がたくさん集まっていることを踏まえ、デジタルサイネージのような屋外広告を介してイベント関連情報を流すといった対応は、「系全体に対するフィードバック」となる。

第二の視点は、「取り扱うデータが、ストック（過去の蓄積）型であるかフロー（リアルタイム）型であるか」である。位置データを例に言えば、過去一年間で利用している電車の路線データに基づいてダイヤの乱れに関する情報の配信を最適化することはストック型での活用であり、いままさに通りかかっている場所や向かおうとしている場所に基づいて情報配信の最適化がなされるのがフロー型であると言える。

この例からもわかるように、ストック型であるかフロー型であるかという点は相対的なものであり、きれいに二分されるようなものではない。上の例で言えば、過去どれだけのデータまで蓄積すればストック型と呼べるのか、という明確な区切りはない。また、フロー型であればあるほど利用者にとって利便性が高いというものでもなく、分野（領域）によってはきちんと過去情報のストックを分析して結果をフィードバックした方が、ユーザに取って利用価値がある。第二の視点は、データの入手から活用までに許されるタイムラグとも言える。

以下では、上記の2つの視点で分類したそれぞれのビッグデータ活用区分について、事例を交えて概観する（図2）。

#### ①系全体フィードバック／ストック型

これまでも小規模には行われてきた一般的なデータ活用のあり方である。自社サイトを訪問した顧客がどのページを見たあとに、次はどのページを見たのか、といったデータ（動線データ）について、過去一定期

【横軸】 データの分析結果をフィードバックするまでに許されるタイムラグの長短。  
 (なにをもって「リアルタイム」とするかは、業種業態、データの利用目的によって異なる)

【縦軸】 データの分析結果が、個人々人に対してフィードバックされるのか、それとも全体最適のため(系のため)に用いられるのか?

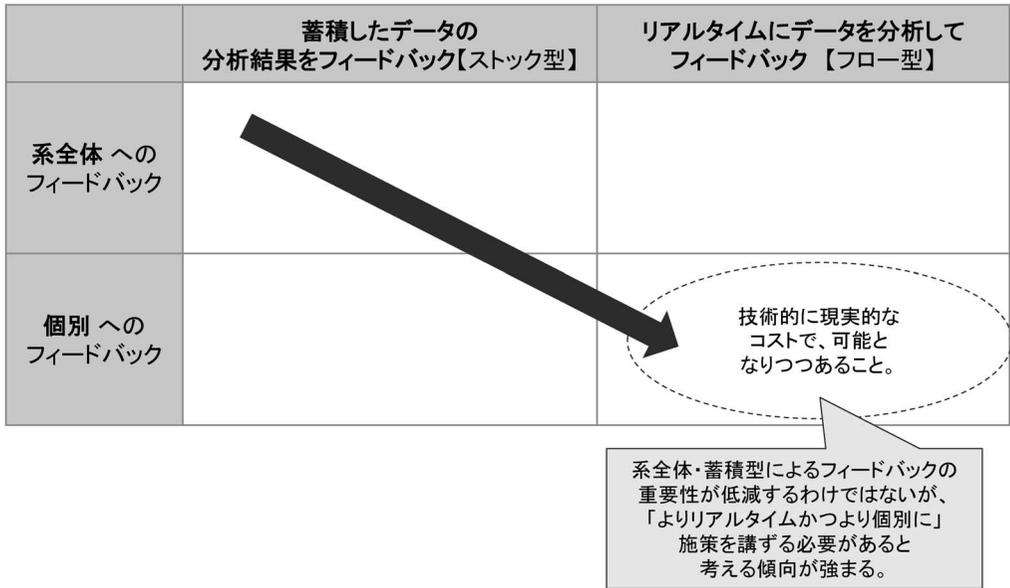


図2 ビッグデータの活用に関する整理 (出典：野村総合研究所)

間に蓄積されたデータを分析し、ユーザが必要とする情報をより容易に取得可能にするべく、サイトを改良するといった利用法である。

EC(電子商取引)ではない実際の店舗においてはPOSデータ(販売情報データ)に基づいて、商品の入れ替えや陳列棚の最適化を行っているが、そのようなデータ分析事例もこのタイプに相当する。マーケティングの世界において「スーパーでAを買う人の多くは一緒にBを買う」といった事例が語られるが、これも同様である。

#### ②個別フィードバック／ストック型

事例としては、ECサイトにおける商品推薦(リコメンデーション)が代表的である。「これまでに、〇〇を買った人にご案内しています」といったものである。以前にその人が買った商品について、「いらなくなったら売いませんか?」という案内を出すのも同様の考え方と言える。

この事例で言えば、中古品として売り出してくれそうかどうかといった判断や、売ることを促すタイミングの検討を、ECサイト利用者全体の傾向に基づいて検討したうえで、個々の利用者に対して働きかけることが想定される。

#### ③系全体フィードバック／フロー型

時々刻々と流れてくる大量のデータを収集・解析して、系全体が調和して稼働するように制御したり、あ

るいは系全体がより高い付加価値を示すようにする事例が相当する。典型的な事例が近年高い注目が集まっている、スマートグリッドやスマートシティなど、「スマート～」と呼ばれる社会システムなどに対する事例である。

例えばスマートグリッドの場合、電力を使う個別の工場・オフィスや世帯から、機器ごとの電力利用データを吸い上げ、電力網全体における需給の最適化を行うために活用される。

#### ④個別フィードバック／フロー型

事例として、ウェブサイトの閲覧状況などを踏まえた「行動ターゲティング広告」や、HFT(高頻度トレーディング)に代表される「株式売買等に係るアルゴリズム取引」が挙げられる。

「①全体フィードバック／ストック型」の重要性が低くなったわけではないが、「④個別フィードバック／フロー型」の事例が増えている。これは、近年に降って湧いてきたニーズでは必ずしもない。これまで、そもそも収集・分析対象となるようなデータが取れなかったことや、また、技術的に分析が困難であったものが、技術的な進展・普及を受けて実現可能になりつつあるのが現状と言える。

これまで、「①系全体フィードバック／ストック型」のデータ活用を行っていたもののなかには、「④個別フィードバック／フロー」としたほうが、より望まし

い場合が多々ある。そもそも実現しなかったことが、技術的に可能となりつつあるためだ。

例えば、販売している商品の売れ筋分析を想定した場合、これまでは一定期間かけて収集した品目ごとの販売実績情報を定期的に分析し、商品の採否や陳列に活かすという利用方法が一般的であった。しかしながら、個別かつリアルタイムでのデータ活用が可能となれば、店舗における在庫状況や、特定商品に対する顧客の関心（ある陳列棚の前を何度も行き来したり、手に取っているなど）をもとに、その顧客に対して商品の推薦を行うことも可能となるだろう。

インターネット広告の配信においては、すでに「④個別フィードバック／フロー型」のデータ活用が進展している。インターネット広告の分野では、複数のウェブサイトを取りまとめ、1つの広告媒体として提供するサービス（アドネットワークサービス）が普及しつつある。その中で、ある一人のネットユーザについて複数サイトの閲覧動向を踏まえた広告配信サービスが、「行動ターゲティング広告」として提供されている。

例えば、複数の新車情報サイトおよび自動車ローンのサイトを閲覧しているユーザがいれば、近々自動車購入の意欲が比較的高いと推察される。そのような閲覧者に対して、閲覧者の居住地に近い自動車販売事業者が、週末に予定している新車展示会の広告を配信するようなサービスが実現している。

## 4. ビッグデータビジネス進展に向けた課題と対応のあり方

### 4.1 最大の阻害要因は人材不足

ビッグデータの利用促進要因が増加するなかで、大きな阻害要因は「人材」不足である。ビッグデータの収集・分析・活用を主導できる人、すなわち統計学や情報科学のリテラシに富み、事業の文脈にそってデータを読むことができる人の数が不足している。

極めて大量のデータを前にしたときに、皆が「ここから何らかの有用な知見を得られるのではないか」という期待を感じるわけではない。極端なことを言えば「日本人全員の過去一年間の購買状況」といったデータが仮に利用可能であったとしても、それらのデータを前に知見導出の可能性を感じることはなく、途方に暮れてしまう人が大半だろう。

利用サイド事業者の立場としては、「それならば専門家に外注すればよいのではないか？」という発想が出てこよう。しかし、情報システム構築のための外注と同様、目的に沿った成果を得るためには、発注側に最

低限の知識・スキル・リテラシが必要であり、それがなければ何をどのように外注していいのかという判断すらつかない。外注先と円滑にやり取りするための人材にも事欠く状態というのが、一般的な利用サイド事業者の現況と思われる。

利用サイド事業者には、ビッグデータの解析ができる人がほとんどいない、という現状に鑑みると、まずは社内に、「この方向でデータ解析をしてみるか」「この部分は外部の専門家に深掘りしてもらおう」と考えられるレベルの人材を増やすための、地道な裾野拡大が求められる。

例えばPCで用いる解析ソフトのインタフェイスを、利用サイド事業者の社員にとって親しみのあるものとすることも有用であろう。統計分析に対してあまり慣れていない社員の底上げのためには、普段その社員が用いている一般的な表計算ソフトの付属機能・改良機能として、簡便な分析機能を提供することなども対応策の一例として想定される。

また、ビッグデータ活用に向けた機器やサービスの提供サイド事業者の立場としては、上のような利用者側の人材不足の状況を踏まえた形でのサービス提供が不可欠であり、それは商機とも言える。しかしながら多くの提供サイドの事業者においても、統計分析や数理モデリングを担う人材が不足している。

これに関連しては、米国における事例となるが、グーグルのチーフエコノミストである Hal Varian が、2009年にインタビューの中で「今後10年間でセクシーな職業は統計家である」と述べている。また、シリコンバレーにおいて「Hadoop が使えて、統計リテラシがある人材」に関しては、スタートアップ事業者から大手事業者まで広く募集の対象とする人気職種となるなどの動向もうかがわれ、提供サイド事業者における人材の奪い合いが既に始まっている様子が見られる。

### 4.2 次の課題はプライバシー・機密情報の不正利用対策

ビッグデータの活用が進展していく過程では、プライバシーや営業機密の取り扱いが大きな課題となる。

プライバシーに関連するデータとしては、年齢・職業・性別などの属性、趣味や嗜好に関するデータ、資産状況や健康状態に関するデータ、居住地や連絡先、あるいはコンテンツの閲覧や購買の履歴などが想定される。

例えば、GPS機能付き携帯電話から収集・蓄積された移動や行き先のデータが悪用されると、個々人の日頃の行動パターンが露見し、尾行などに悪用されてしまう恐れがある。

また、個々のデータについては、利用者自身が開示を許可したものであったとしても、複数のデータが関連付けられることによって、利用者本人が望まない事実までもが露見してしまうことも、懸念事項と言える。

また、データの取得や公開・流通が容易となる中で、データ取得者自身には悪意がなくとも、そのデータが公開されることで不利益を被る人が出てくる可能性もある。この種の懸念は、かねてより提起されているところである。例えば、『たまたま通りすがった人が、たまたま記録した内容に時刻と場所がスタンプされている。人々が、自分のもつ、自分にとっては何の意味もない情報に対して低額でアクセス権を認めるということは十分考えられる。これを利用して、時刻と場所の指定によりサーチをかける。集まった「断片」をつないで再構成すれば、欲しい情報が手に入ることになる。(中略) 結果、日常の何気ない所作すら、どのような形で収集され、誰かへの情報として流れていくかわからないという事態を招くことになるのではないか』という懸念が1999年に既に示されている。これは、把握・収集可能なデータが増大している現状においてまさに懸念されるべき事項と言えるだろう。

ビッグデータビジネスを健全に推進していくためには、これらの懸念を解消しつつ、データ分析の成果を享受できるように対策を講じていくことが必要となる。万能の施策は存在せず、個々の施策の積み重ねが求められるところであるが、近年検討が進められている対策のうち、「プライバシー保護データマイニング」がある。

これは、プライバシーや営業秘密を守りつつも、データ分析の結果得られた知見を活用しようとする際の技術的対策である。データマイニングをしつつも、一人一人のプライバシーは保護したいという一見相反するテーマについて、2000年以降、学術レベルでの検討が進められている。

本技術の活用によって、「個々人のプライバシーが暴露されることは一切ないが、それらのデータに基づく全体動向については、皆で共有することができる」といった形でのデータ活用が、技術的に可能になることが期待されており、今後の研究の進展を注視する必要がある。

制度的な対応や、技術進展による解決もちろん必要ではあるが、結局のところ消費者の「事業者によって私に関するデータが勝手に使われていないか？データの開示に見合うだけの便益が得られるか？」という疑問を払い、信頼関係を構築することが重要な点となる。

なぜならば、「気がきく」と「気持ち悪い」の領域は

不明瞭なためだ。同じような商品推奨を行ったとしても、ある事業者は「気がきく」と評価されるかもしれないが、別の事業者は「気持ち悪い」とされてしまうかもしれない。些細な表現の違いが「気がきく」と「気持ち悪い」を分けてしまうこともあるだろう。

例えば、Amazon.co.jpは「鈴木さんには〇〇という商品がおすすめ」という推奨はせず、「鈴木さんと同様に□□をお買い上げになった方は、〇〇も買うことが多いです」という推奨をしている。おすすめの根拠を明確にすることによって、いらぬ不信を抱かれることを避けていると言える。

#### 4.3 データの精度の悪さや、誤用・不適切利用に起因する課題

取得されたデータが誤っていたり、精度が良くなければ、当然にしてトラブルや事故につながる。2008年には米国にて、高校生が乗ったバスが車高よりも低いトンネルに衝突し、大破する事故が生じた。バスにはGPSと連動した頭上衝突防止ナビが設置されており、あらかじめそのバスの車高を踏まえて通行可能なルートが案内される仕組みとなっていたが、実際に案内されたルート上のトンネルはバスの車高よりも大幅に低く、事故に至った。

不適切な結果につながるのは、取得・生成されたデータが不適切である場合だけではない。正確なデータを処理・分析し、そこから得られた知見を活用する過程においても、望まない結果につながることもある。

位置データの活用に関して言えば、2010年にGoogleがGoogleマップにおけるルート案内が不適切であったとして訴えられている。これは、携帯電話のGoogleマップ上で案内されたルートを歩いていたところ、横断歩道がない道路を横断した際に車にはねられ負傷した、というものである。Googleは、本サービスの利用に際しては「交通量の多いエリアでは特に細心の注意を払って下さい。自動車を運転しているときのように、交通標識や信号に従い、歩行に向かない道路を的確に判断して下さい」という注意を掲載していたという。

ECサイトにおける商品おすすめが非常に不謹慎なものとなる恐れもある。例えば、2008年頃、あるECサイトにおけるリコメンデーションが不適切ではなかったか話題になった。ある洗剤を検索すると、一見すると関係のない入浴剤がおすすめされたのである。また、あわせて、家電製品用のタイマー付きコンセント、ポリ袋、そして「自殺」に関連する書籍などもおすすめされた。これは、いずれも当時、急増していた硫化水素を発生させて自殺をしようとした時に必要となる薬

剤・機器であった。

警察庁統計によると、2007年に29人だった硫化水素自殺による自殺者数は、2008年には1056人と急激に増加し、社会問題となった年であった。上の事例は、極端な事例に分類されるだろうが、「機械的には確かにあわせて買うことを『おすすめ』するに足るものかもしれないが、それをおすすめするのは果たして健全なことだろうか？」という問いが生ずるような状況は今後ますます増え、顕在化していくことだろう。

例えば、ソーシャルネットワークサービスの中には、「知っている人かも？」として実際の知人ではないかどうかということを示す機能を具備する事例が多いが、「おともだちかも？」として示された人物が「自分を殺しかけた元DV男」であることや、特定の嗜好を有した犯罪者であるような状況も想定される。

以上で述べてきたようにビッグデータビジネスは事業的にも技術的にも、大きな可能性を秘める一方で、さまざまな課題を包含している。産業界全般におけるクラウド利用の進展とあわせて、2010年代の情報・通信分野にとどまらぬ幅広い領域において注目すべきテーマの1つになると考えられる。

ビッグデータビジネスの健全な進展のためには、ビッグデータの活用を行うことができる人材の育成および、プライバシーに関連する情報や営業機密といった機微なデータの取り扱いに関する指針等の整備が不可欠だ。機微なデータの取り扱いへの配慮はもちろん必要ではあるが、配慮のあまりデータの活用を完全に控えてしまうこともまた、事業の推進や社会システムの整備を効率的に行ううえで不健全である。困難ではあるものの、活用と保護を両輪として進めていくことが求められる。

**付記** 本稿は筆者による [1][2] の内容に基づいて構成・執筆しました。[1]は「ビッグデータ」と呼ばれる概念についての可能性と課題についての総論であり、[2]は今後のビッグデータ活用を示唆する事例を多数紹介しています。あわせてご覧ください。

問い合わせ先：bigdata@nri.co.jp

#### 参考文献

- [1] 鈴木良介、『ビッグデータビジネスの時代』、翔泳社、2011。
- [2] 鈴木良介、『ビッグデータ・ビジネス』、日経文庫、2012。