

「ビッグデータビジネス」の興隆と 対応すべき課題

大量データの高度な活用に向けて

鈴木良介



CONTENTS

- I 「ビッグデータビジネス」の興隆の背景
- II ビッグデータで扱うデータの多様化と高度活用の進展
- III ビッグデータを活用したビジネスのあり方
- IV 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

要約

- 1 「ビッグデータビジネス」とは、昨今の革新的な情報・通信技術を活用して大量のデータ（ビッグデータ）を収集・解析することで、社会・経済の問題解決を図ることや業務の付加価値を一層高めるための事業を意味し、今後急速な立ち上がりが予想される。同ビジネスは、時々刻々と変化する消費者の行動や機器利用・業務付随データなどを、収集・蓄積・分析することで実現される。
- 2 ビッグデータの活用により、①もともとのサービス自体の魅力増大、②効率的な研究開発の実現、③適時適切な情報配信——が進むことが想定される。また、リアルタイムでのフィードバックを個別に行うことが技術的に可能となりつつあり、経営資源の適切な活用や社会インフラの効率的な整備に有用と考えられる。
- 3 ビッグデータ活用の環境は整備されつつあるものの、その活用支援を事業として成立させるための課題はいまだ多い。利用サイド事業者に対してビッグデータ活用の経済合理性の認識を促すビジネスモデルを整備することや、身近な課題である保有データの不整合の解消を入り口に導入を図ることなど、市場進展のための工夫が求められる。
- 4 ビッグデータビジネスの健全な進展には、ビッグデータを活用できる人材の育成、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整備や技術的解決が不可欠である。

I 「ビッグデータビジネス」の 興隆の背景

2010年代の情報・通信分野における大きな潮流の一つとして、「ビッグデータビジネス」の立ち上がりが予想される。

ビッグデータの明瞭な定義はいまだ存在しないが、大きく3つの観点から語られることが多い。

第1は「データサイズ」の観点である。ペタバイト（ 10^{15} ）、エクサバイト（ 10^{18} ）といった、これまでに取り扱ってきたデータサイズを大きく超えるサイズのデータは代表的なビッグデータといえる。ただし、「ペタバイト以上ならばビッグデータ」と言い切れるものではなく、従来あるデータ処理の状況下において一般的であったデータサイズを大きく上回ることをもってビッグデータと呼ばれることも多い。

第2は「データ把握・収集の頻度」の観点である。たとえ、把握・収集や処理の対象となる個々のデータサイズが極端に大きなものでないとしても、データが非常に高い頻度で把握・収集されるとするならば、その処理には従来とは異なる方法が求められる。これは、データの解像度の向上とも言い換えられる。たとえば、Webサービス提供事業者が、ユーザーのそれぞれの振る舞い（クリックなど）の背後にどのような意図があるのかを分析するためには、高いデータの解像度が求められるのは当然であるし、そのためにはデータは高い頻度で把握・収集することが求められる。

第3は「データの多様性」の観点である。定型的な数値データ、テキストデータにとど

まらず、Webサービスへのユーザーからの書き込み、膨大な数の監視カメラによって収集される映像データ、デジタルサイネージ（電子看板）を閲覧する人の顔写真、位置データ、各種のセンサーからのデータなど、多種多様なデータの活用に対する関心も高まっている。これらもビッグデータの一部を構成する重要な観点である。

このようなビッグデータの位置づけを踏まえて本稿では、「ビッグデータビジネス」という言葉を、「昨今の革新的な情報・通信技術を活用して、ビッグデータを収集・分析することにより、社会・経済の問題解決を図ることや、業務の付加価値を一層高めるための事業」の意に用いる。

過去にも、データ解析から付加価値を生み出すこのような取り組みは、さまざまな形で試みられてきた。「データマイニング」「ビジネスインテリジェンス」、あるいは「センサーネットワーク」や「ユビキタスネットワーク構想」「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であるといえる。しかし、それらのなかには市場拡大に至っていない概念や現状では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。

ビッグデータビジネスが興隆する要因は、データ解析結果を自社事業に役立てようとする「利用サイド事業者」、およびデータ解析のツールやサービスを提供する「提供サイド事業者」の双方に求めることができる。以下ではそれぞれの動向を整理する。

1 利用サイド事業者の動向

利用サイド事業者においてビッグデータを活用しようという機運が高まっている背景に

は、事業を推進する際の合理性・効率性の追求要請がより高いものとなっている点が挙げられる。

たとえば、米国には「Data is King」を標榜し、データ分析に基づく事業の効率化やサービス品質の向上を常に追求し続けている事業者がいる。苛烈さを増す競争環境のなかで、このようにより効率的で確実な施策を求める意向は、ビッグデータビジネスの促進につながるであろう。

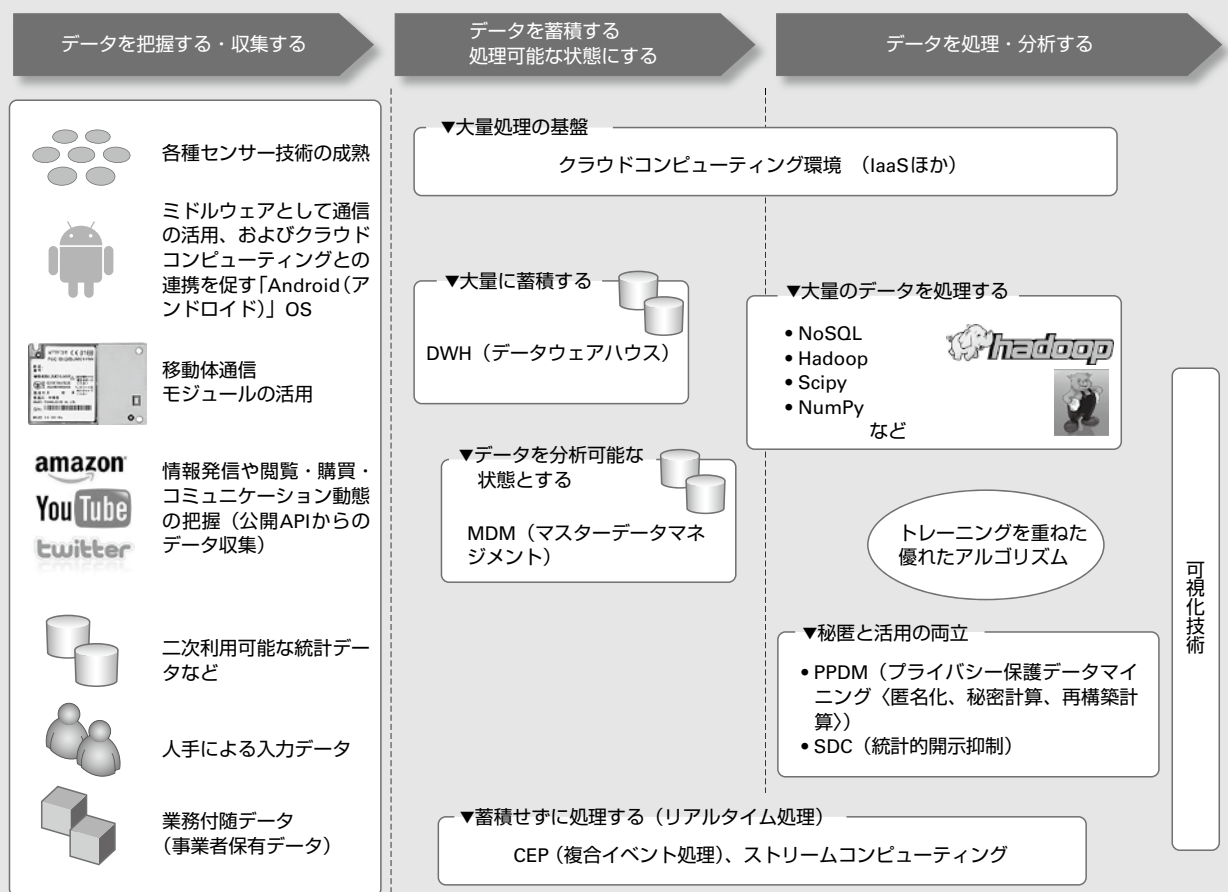
また、分析対象となりうるデータの飛躍的な増大と、分析を支援するツール類の充実も、ビッグデータビジネスを後押しする（図1）。以下、整備が進みつつあるビッグデータ活用

の環境について述べる。

(1) クラウドコンピューティングによるサーバーへのデータ蓄積

第1に、利用サイド事業者におけるクラウドコンピューティング（以下、クラウド）の利用が進みつつあり、それがビッグデータの活用を後押しするものと考えられる。なぜならば、クラウドの利用環境では、多くのデータがサーバーに集約・蓄積されるからである。サーバーにデータを集約・蓄積させることは、ビッグデータをつくるうえでの基本的な環境整備である。データの集約と蓄積に際しては、もちろん、データの整合性をどのよ

図1 ビッグデータビジネスの興隆に関連する要因と新たな動向



注) API : アプリケーション・プログラム・インターフェース、Hadoop : ハドゥープ (第1章2節参照)、IaaS : インフラストラクチャー・アズ・ア・サービス、NoSQL : ノットオンリー SQL、Numpy : ナムパイ (第1章2節参照)、OS : 基本ソフト、Scipy : サイパイ (第1章2節参照)

うに確保するかという課題を事前に解決しておく必要がある。

「データの収集と分析から自社事業に役立つ知見を導き出す」という観点は、従来から多くの事業者が持っていた関心事項であるが、「サーバーにデータが集約されること」の効用については、クラウド利用が拡大しつつある昨今にあっては特に注目すべき事項といえる。クラウド利用環境下においては、サーバー側へデータが自ずと蓄積されるためである。

(2) API公開やAndroid活用による 多種多様なデータの把握・収集

第2に、

- データの把握・収集に関する技術の進展や製品の充実が進んでいること
- 政府公開データのような二次利用が可能なデータが増大していること
- グーグルやフェイスブックなどの大規模Webサービス事業者のAPI（アプリケーション・プログラム・インターフェース）公開により、各事業者が利用可能なデータが増大していること

——が挙げられる。

スマートフォン（高機能携帯電話端末）を含む携帯電話端末の普及と高機能・高性能化は、データの把握・収集を容易にした好例である。たとえば、GPS（全地球測位システム）機能の搭載により、各端末の位置データをリアルタイムで把握することができる。電子マネー機能を具備することで決済情報の把握も可能である。さらに、スマートフォンユーザー向けにさまざまなアプリケーションソフトの開発と提供が可能となっていること

も、同ソフト開発者にとっては多くの種類のデータを把握・収集することを容易にした。

併せて、グーグルが公開しているOS（基本ソフトウェア）「Android（アンドロイド）」は、データの効率的な吸い上げが進展しつつあることの象徴的事例といえるであろう。なぜならば、Androidを活用すると、多くの電子機器が従来と比べ容易にネットワークに接続されるようになるからである。たとえば、Androidは携帯電話端末用のOSとして広く知られているが、それにとどまらず、カーナビゲーションやデジタルフォトフレーム、セットトップボックス（多機能受信機）など、各種の電子機器のOSとして利用され始めている。これにより、従来パソコンや携帯電話端末では取得が困難であったより多種多様なデータがクラウド上に蓄積され、活用にも供されるようになることが期待される。

Webサービスが開示するAPIを介して利用可能となる各種データや、自らWebサイトを運営している事業者であれば、自社サイトを訪問するユーザーのログデータなども、活用対象とすべきビッグデータの好例である。

(3) クラウドのデータ処理・分析のための 計算機資源

第3に、クラウドのようなサービスが、ビッグデータを処理・分析するための計算機資源として便利かつ安価に利用できるようになったことも、ビッグデータの活用を後押ししている。クラウドの登場は、前述した「データ集約・蓄積」のみならず、「データ処理・分析」にも影響を及ぼしている。

分析対象データと分析に向けた意欲はあるものの、そのための計算機資源をはじめ設備

投資の予算が乏しい事業者は少なくない。たとえば、Webサービスを新たに提供し始めた事業者でいえば、閲覧者が自社サイト上のどのページをどのように移動したり、どのページを長く見ていたりしたかといった「動線」を詳細に分析したいというニーズがある。このようなニーズに対しクラウドサービスは、事業者がデータ分析目的に過大な設備投資をすることも、分析に多くの時間をかけることもなく大量計算を可能とし、ビッグデータ活用の障壁を低くしたことになる。

2 提供サイド事業者の動向

ビッグデータを蓄積・分析するためのハードウェア・ソフトウェアの開発や、支援サービスを行う提供サイド事業者においても、ビッグデータビジネスに注力する意向が高まっている。なぜならば、クラウドの登場を契機に、新しい収益源を開拓する必要性が増大しているからである。

提供サイド事業者としては、IT（情報技術）ベンダーやシステムインテグレーター、通信事業者などの情報システム開発に関連する事業者が該当する。これらの事業者の多くにとって、取り扱うデータの中身にまで深く踏み込んだ理解が求められるビッグデータビジネスに関与することは容易ではない。それにもかかわらず、同事業者がビッグデータビジネスに注力してこれまでとは異なる収益源を求める背景にも、クラウドの普及がある。

従来の情報システムやITサービス市場は、クラウドの利用進展に伴い縮小することが予想される。クラウドに対する利用サイド事業者からの期待として、「情報システムにかかる総費用の低減」が大きい以上、自社での設

置・管理を前提とする情報システムからクラウドを活用する情報システムへの移行時には、市場の縮小は避けられないためである。そのため、提供サイド事業者が事業規模を維持し成長させるためには、従来は扱ってこなかったビッグデータ活用のような領域であってもそれを新規事業領域として開拓し、収益を維持・拡大することが必要になる。

もちろん、市場が全体的に縮小傾向にあるなか、既存の提供サイド事業者が講ずることのできる施策は、ビッグデータビジネスへの注力だけではない。顧客を増加させるための「グローバル展開」や、特定の大手事業者から異業種、中小事業者の市場開拓まで「顧客セグメントの拡大」を図る方向性もある。

こうしたさまざまな施策のなかでビッグデータビジネスは、「顧客単価の漸減を食い止めるための施策」と位置づけられる。提供サイド事業者が、クラウドの利用・進展に伴って縮小する情報システム部門の予算のみを収益源とするかぎり、顧客単価の増加には限界がある。提供サイド事業者は、今後、情報システム部門以外の事業部門を開拓する必要性が増す。各事業部門への訴求という観点においても、ビッグデータからの知見の導出は、汎用性と説得性のある施策と考えられる。

以下に、ビッグデータの活用を促す提供サイド事業者による製品の充実や技術開発の具体的な事例を示す。

まず、数百テラ、ペタバイトクラスのビッグデータを蓄積・活用するためのデータウェアハウスや、ビッグデータを分析可能な状態にするためのマスターデータマネジメント（MDM）関連の技術およびサービスが挙げられる。また、ビッグデータを処理する基盤

ソフトである「Hadoop（ハドゥーブ）」や、プログラミング言語「Python（パイソン）」の数理分析モジュールである「NumPy（ナムパイ）」「Scipy（サイパイ）」といった数理的な処理を、安価で強力に推し進めるツール群も充実しつつある。

また、ビッグデータを高速に処理するための技術として、「複合イベント処理（CEP：Complex Event Processing）」にかかわる技術にも留意する必要があるだろう。CEPの利用事例としては、高速での証券取引やクレジットカードの不正利用検知、情報システムの稼働状況検知などが代表的である。

これらの製品にはIBMやオラクル、日立製作所などが力を入れている。特にIBMは、かねてより本領域に注力してきたことに加え、昨年提唱している「スマータープラネット」構想のなかでもその活用事例を示すなど、本領域における存在感をより高めている。近年では上述の各社とも、関連するデータ活用事業のラインアップ整備を進めるべく、他社買収なども進めている。

次に、データの秘匿と活用を両立させるための技術として、「プライバシー保護データマイニング（PPDM：Privacy Preserving Data Mining。匿名化、秘密計算、再構築計算などの総称）」や「統計的開示抑制（SDC：Statistical Disclosure Control）」などがある。これらにかかわる研究はビッグデータのみを念頭に置いているわけではないが、ビッグデータに対しても活用が期待される。

2011年当初より、自社の新たな取り組みをビッグデータと関連づけて発表する提供サイド事業者も増加しつつあり、製品やサービスの充実や技術開発は今後より一層進展するこ

とが予想される。

II ビッグデータで扱うデータの多様化と高度活用の進展

では、ビッグデータの活用からどんな効用を得ることができるだろうか。以下、「結局のところ、ビッグデータビジネスが目的とするところは何なのか。ビッグデータはどのように活用しうるものであるのか」という点について整理したい。

1 ビッグデータで得られるメリット

ビッグデータ活用やビッグデータビジネスの効用、すなわち大量データの蓄積・分析で得られるメリットは、「フィードバック先」と「リアルタイム性」の2つの視点から整理することが有効である。

(1) フィードバック先の視点

第1のフィードバック先の視点とは、ビッグデータの解析結果のフィードバック先が、

- ①ユーザー個人などの「個別フィードバック」
- ②あるサービスの利用者全体などに対する「系全体フィードバック」

——のどちらなのかということである。たとえば、ある携帯電話端末ユーザーの位置データの活用を想定した場合、ユーザーの現在位置を踏まえ、そのユーザーに対して当該位置に最適化された情報（たとえば、そのユーザーが数時間前にWebで閲覧していた商品を、近くにある店舗が今まさに割安で販売するキャンペーンを実施しているなど）を配信することが個別へのフィードバックである。

表1 ビッグデータ活用の類型化と具体的な事例

	蓄積したデータの分析結果をフィードバック【ストック型】	データをリアルタイムに分析してフィードバック【フロー型】
系全体フィードバック	<ul style="list-style-type: none"> 製品開発 アルゴリズム改良 Web動線、陳列棚の配置最適化 ... <p style="text-align: right;">①</p>	<ul style="list-style-type: none"> スマートグリッド（次世代送電網） スマート～ ... <p style="text-align: right;">③</p>
個別フィードバック	<ul style="list-style-type: none"> EC（電子商取引）サイトにおけるリコメンデーション（商品推薦） ... <p style="text-align: right;">②</p>	<ul style="list-style-type: none"> 行動ターゲティング広告 クレジットカードの不正利用検知 HFT（高頻度トレーディング） ... <p style="text-align: right;">④</p>

何をもち「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

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でイベント関連情報を流すといった対応は、系全体フィードバックである。

(2) リアルタイム性の視点

第2のリアルタイム性の視点とは、取り扱うデータが、「ストック（過去の蓄積）型」なのか「フロー（リアルタイム）型」なのかである。上述の位置データの例にならえば、利用している電車の過去1年間の路線データに基づいて、ダイヤの乱れに関する情報配信を最適化することはストック型の活用であり、今まさに移動している場所や向かおうとしている場所に基づいて、情報配信を最適化するのがフロー型といえる。

この2つ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は相対的で、きれいに二分できるものではない。上述の例でいえば、過去のデータ（ダイヤの乱れ）をどこまで蓄積すればストック型と呼べるのかの明確な区切りはない。

また、フロー型であればあるほどユーザーにとって利便性が高まるわけでもなく、分野

（領域）によっては過去情報のストックをきちんと分析して結果をフィードバックしたほうが、ユーザーにとって利用価値があるケースもある。第2のリアルタイム性の視点は、データの入手から活用までに許されるタイムラグともいえる。

以下では、上述の2つの視点で類型化したビッグデータの活用事例を概観する（表1）。

2 ビッグデータ活用事例

(1) 系全体フィードバック・ストック型

従来から小規模には行われてきた一般的なデータ活用である。自社のWebサイトを訪問した顧客がどのページを閲覧し、次にどのページを閲覧したのかといった過去の一定期間に蓄積された動線データを分析し、ユーザーが必要とする情報をより容易に取得できるようにサイトを改良するといった利用法である。

電子商取引（EC）ではない実際の店舗では、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、こ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で語られる「スーパーでAを買う人の多くは一緒にBを買う」といった事例も同様である。

(2) 個別フィードバック・ストック型

ECサイトにおけるリコメンデーション(商品推薦)が代表的事例である。「これまで〇〇を買った人に案内しています」といったものである。以前に消費者が買った商品に対して、「いらなくなったら売りませんか」という案内も同様の考え方といえる。

この事例でいえば、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、販売を促すタイミングを、ECサイトのユーザー全体の傾向に基づいて分析・検討し、そのうえで個々のユーザーに対して働きかけるということが想定される。

(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・分析し、系全体が調和するように制御したり、あるいは系全体の付加価値をより高めるようにしたりする事例が相当する。典型が、近年、高い注目を集めているスマートグリッド(次世代送電網)やスマートシティなど、「スマート」を冠した社会システムにおけるビッグデータ利用事例である。

たとえばスマートグリッドの場合、電力需要家である工場・オフィス・世帯が個々に使用している機器の電力使用量のデータを収集し、それに太陽光や風力などを含む発電量のデータを加味することで送電網全体の需給を最適化する。

(4) 個別フィードバック・フロー型

時々刻々と変化する状況に応じて、適時適切な情報や対応を個々のユーザーなどに対して提供するような事例が相当する。具体的には、Web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「株式売買のアルゴリズム取引」が挙げられる。

3 ビッグデータの活用をより高度化・精密化する技術の進展

前項の(1)から(4)までの事例の傾向を見ると、(1)の系全体フィードバック・ストック型の重要性が低下したわけではなく、相対的に(4)の個別フィードバック・フロー型の事例が増えている。(4)は、近年になって降って湧いたニーズでは必ずしもない。収集・分析対象となるデータが、従来そもそも取得できなかったり技術的に分析困難であったりしたものが、昨今の技術の進展と普及を受けて実現可能になりつつあり、それにより新たなニーズとして捕捉されてきたのである。

従来、(1)の系全体フィードバック・ストック型には、(4)の個別フィードバック・フロー型としたほうがより望ましい場合が多い。そもそも実現したかったことが技術的に可能になりつつあるためである。

たとえば商品の売れ筋分析の場合、従来は、一定期間をかけて収集し、定期的に分析した品目ごとのデータ(販売実績情報)を、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、店舗の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たりするなど)をもとに、その顧客に対して商品を推薦することも可能となるであろう。

インターネット広告の配信においては、す

で(4)の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。複数のWebサイトを取りまとめ、それらを一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)が増加しているが、このサービスにおいて提供されている行動ターゲティング広告などが事例として挙げられる。これは、一人のネットユーザーが訪問した複数サイトの閲覧動向を踏まえ広告を配信するような技術である。

たとえば、複数の新車情報サイトや自動車ローンのサイトを閲覧しているネットユーザーの場合は自動車購入意欲が比較的高いと推察される。そのようなユーザーに向けて近くの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というようなサービスがすでに実現している。

ストック型とフロー型を分ける視点として挙げた「フィードバックのリアルタイム性」は、近年、特に携帯電話向けサービスやインターネット関連サービスできわめて重要なテーマとなっている。140文字以内の「つぶやき」を発信するサービスである「Twitter(ツイッター)」は、従来のブログやSNS(ソーシャル・ネットワーキング・サービス)サイトへの投稿と比較すると、よりリアルタイム性が高く、新たなマーケティングツールとして注目されつつある。

同様に注目されるサービスとしては、2010年以降、米国を中心に急増している位置データを活用するLBS(ロケーション・ベースド・サービス)がある。その代表的な事例には、「Google Latitude(グーグルラティチュード)」や「foursquare(フォースクエア)」がある。これらは、自動的に、あるいは自己

申告で得られた位置データに基づき、たとえば待ち合わせをしている知人に現在位置を示したり、近くの店舗から広告配信を受けられたりするといったサービスが提供されている。「友人との待ち合わせ」や「あなたが今いる場所の近くの店舗の広告」といった基本的な利用シーンを想定しても、ユーザーの現在位置に基づき配信されるこのような情報は適時適切であることが不可欠で、リアルタイム性が求められる。

また、クレジットカードの不正利用検知や情報システムに対する不正アクセス検知は、比較的古くから実用化されてきた(4)の個別フィードバック・フロー型のデータ活用事例といえる。このように情報セキュリティ強化のためにビッグデータ活用が進むことも想定される。たとえば、多数の監視カメラの映像をリアルタイムで自動解析し、異常があれば管理者に注意喚起を促すような事例などはこのタイプに相当する。このようなサービスは一部ではすでに実用化されており、今後の適用拡大が見込まれる。

Ⅲ ビッグデータを活用したビジネスのあり方

1 利用サイド事業者にとってのビッグデータ活用の経済合理性

提供サイド事業者は、利用サイド事業者のビッグデータの活用を手助けすることで収益を得ることが期待される。こうした期待に応えるには、利用サイド事業者に、「ある程度費用をかけてでも、ビッグデータの活用によってそれに見合うだけの収益向上やコスト削減が実現される」ことを認識させる必要があ

る。以下、想定される状況をいくつか例示する。

たとえば、家電製品や自動車などの一般的な機器の利用状況に関するビッグデータを収集・解析することで、次に当該機器を開発する際に適切な課題設定が可能となる。それらの機器の購入後、ユーザーはどの程度の頻度で使っているのか、どのような機能や設定を好んでいるのかといった使用状況を自動的に収集し、集まったビッグデータの分析を通じて全体の使用傾向や課題が把握できるようになる。そのような取り組みを通して、ときには、メーカーがユーザーによかれと思って装備した機能が使われていなかったり、過剰性能であったりすることが明らかになる場合もあるだろう。

一方で、想定外の使われ方をして生じた故障の原因把握や、あるいは性能を強化すべきポイントも見えてくる。こうしたデータ活用を通じて、メーカー側の見当違いの設計やむだに投じられていた開発投資を見直す効果も期待できる。

また、個々人の趣味や嗜好、行動履歴などに関するビッグデータの収集・分析は、商品のきめ細やかな推薦や広告配信の最適化など、マーケティング分野にも活用されよう。行動ターゲティング技術に象徴される「適切な人に適切な情報を配信する」という考え方は、適切な情報配信が売り上げに直結する事業部門にとっては必須の施策で、今後のビッグデータビジネスの進展が特に期待される分野である。これにより、広告および販売促進関連費用の効率化が期待される。

広告や販売促進、あるいは各種の研究開発には、日本国内だけでも数兆円規模の費用が

投じられている。利用サイド事業者にとって、「これらの費用の投資対効果は不明瞭ではないか」「より効率的な打ち手を講ずることができるのではないか」という問題意識は尽きない。これらの妥当性をビッグデータに基づき検証することには一定の価値が認められるであろう。そのことこそ、ビッグデータ活用を支援する事業者（提供サイド事業者）の商機となる。

加えて、直接的な収益につながらなくとも、機器・サービスの魅力や効率を高めるためにビッグデータを活用するという考え方も重要である。ビッグデータビジネスであっても原則は収益性の追求であるものの、短期的には「収益性よりも特定データの収集のほうが重要」という状況も想定される。これもまた、中長期的には商品の競争力が強化されるという観点において、経済合理性が期待できるものである。

たとえば、電話番号案内のようなサービスを無料で提供することを通じて、自社が開発する音声認識アルゴリズムを改善するようなケースが相当する。アルゴリズム改善を目的としたサービスでは、収益性よりもアルゴリズム改善のためのデータ収集に重きを置き、そこで改善されたアルゴリズムをサービスの競争力向上のために活用し、それにより中長期的に収益を高めるといったビジネスモデルが考えられる。

ビッグデータを活用することにより経済合理性の高い仕組みを構築できることがわかったとしても、現在稼働している仕組みをビッグデータ活用型の仕組みに入れ替えることは、一般的には抵抗が生じやすい。そのような背景もあって、新興国のインフラ整備や情

報システム、サービスの整備のなかでビッグデータを活用しようとする取り組みがある。

たとえばノキアは2000年代初頭より、ケニアをはじめとする国々で「ノキア・データギャザリングプロジェクト」を推進している。これは、社会インフラが未成熟な新興国を中心に、携帯電話を用いたインフラサービスを提供しようとするプロジェクトである。具体的には、農業関連情報（農産物の市況や気候状況など）の配信や感染症予防のためのデータ収集と配信、国勢調査の実施支援といったサービスを提供しつつ、こうした領域のビッグデータを収集し、展開可能な知見を導出しようとしている。これは、経済合理性をより打ち出しやすい国や顧客に対してビッグデータ活用型のサービスを提供しようとする好例といえる。

上述の事例からもわかるように、ビッグデータを「活用する」段階になると、情報システム部門の所掌範囲を超えて事業そのものと密接に関係する。情報・通信技術はこの数十年、利用サイド事業者の業務効率化に寄与してきた。しかし、こうした効率化を超えた新しい付加価値の提供と支援の領域においては、いまだ十分な役割を果たしきれていない。今後ビッグデータビジネスが進むことで、情報・通信技術がこの新しい領域に付加価値を提供しうるものと考えられる。

2 経済合理性を明確に示すビジネスモデルや仕組みの導入

提供サイド事業者が、利用サイド事業者に対してビッグデータ活用の経済合理性をより明確に示すには、ビジネスモデルを洗練させる必要がある。ビッグデータの活用を促す状

況変化は、前節までに論じた個別の技術や製品の進展にとどまらない。ビッグデータを「つくる」ためにデータをサーバーに効率的に集約するビジネスモデルや仕組みをつくり上げることが、ビッグデータ活用を促進するうえでは重要になる。その事例として、「通信の隠蔽」と「ゼロクリックサービス」について説明する。

● 通信の隠蔽

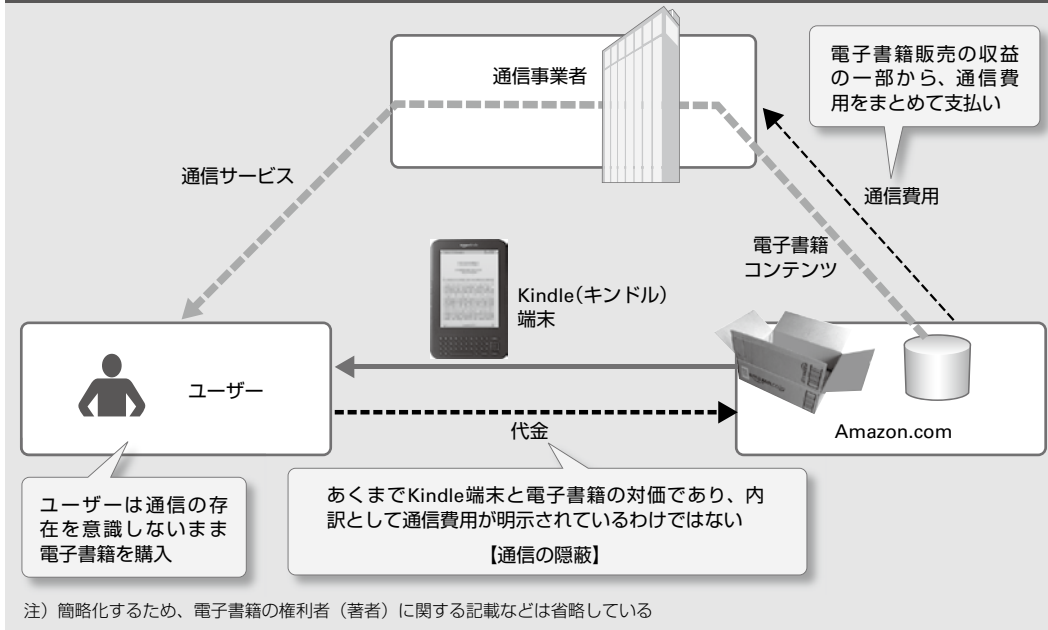
通信の隠蔽とは、ユーザーが通信サービスを受ける際に費用やリテラシー（読み書き能力）などさまざまな負担から解放され、自由自在かつ無意識にサービスが利用できるような仕組みである。言い換えれば、「通信サービス利用に伴う費用の隠蔽」や「通信サービス利用に伴う契約や設定・操作負担の隠蔽」である。

● ゼロクリックサービス

ゼロクリックサービスとは、利用者がデータをサーバーに送るための手続きを一切せずにするようにする仕組みである。通信の隠蔽と併せて提供されることで、人々の持つデータを、端末からサーバーへ集約させるプロセスにおいて大きな効果を発揮する。なぜならば、ユーザーが意識しないまま生じた通信の料金をユーザーに課すことは困難だからであり、したがって、ゼロクリックサービスは通信の隠蔽のもとで提供されるのが自然である。

通信の隠蔽の具体例としては、Amazon.com（アマゾン・ドット・コム）の「Kindle（キンドル）」端末を用いた電子書籍サービスが挙げられる（図2）。

図2 Amazon.com (アマゾン・ドット・コム) の電子書籍サービスにおけるビジネスモデル (「通信の隠蔽」の事例)



Kindle端末のユーザーは、同社のサーバーにある大量の電子書籍から希望するコンテンツを、おおむねいつでもどこからでもダウンロードして閲覧できる。このようなサービスが可能なのは、Kindle端末が携帯電話網対応の通信モジュールを内蔵し、かつ通信費用をAmazon.comが負担しているからである。ユーザーは、通信契約の必要もなければダウンロードに伴う通信費用も直接には負担しない。すなわち、ユーザーは通信サービスに伴う負担を何も感じず通信の存在すら意識することなく、読書という効用のみを手軽に享受できるのである。

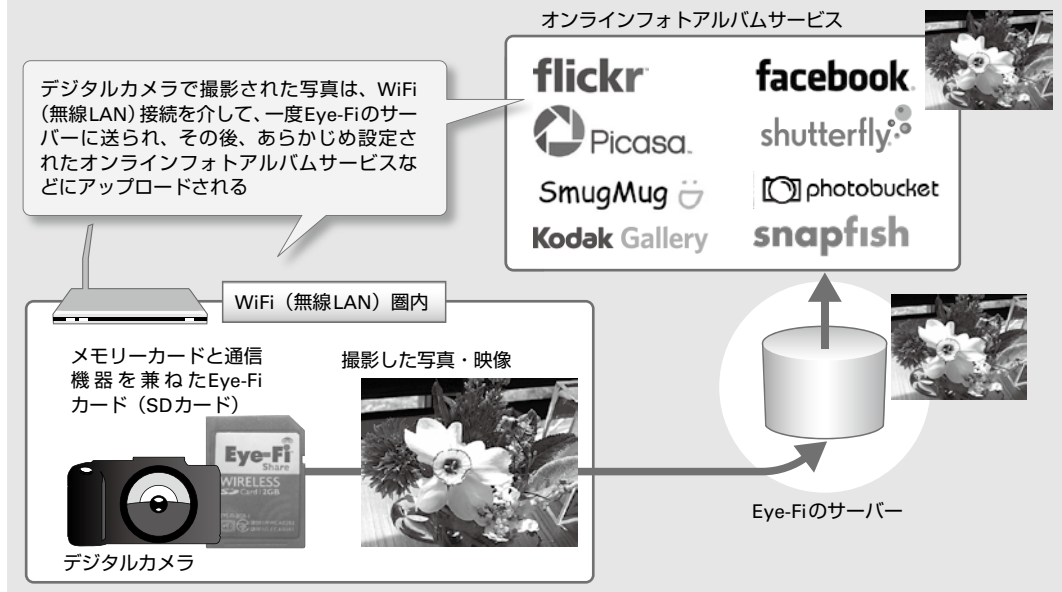
「ビッグデータをつくる」という観点により明確な事例としては、2010年から開始されたKindleの付随サービス「Popular Highlights (ポピュラーハイライト)」が参考となる。これは、ユーザー(読者)が電子書籍を読みながら文章に引いた下線の箇所を他のユーザー

と共有する仕組みである。

このサービスを導入した結果、Amazon.comは「大勢の読者が面白いと感じた文章のビッグデータ」を保持することが可能になった。「ある本のなかでみんなが面白いと感じた一節」に関する情報は、書籍販売上、あるいは新規出版計画を立案する際にきわめて有用と考えられる。このようなサービスは、通信が隠蔽された形で提供されていたために容易に開始できたといえよう。

一方、ゼロクリックサービスの事例としては「Eye-Fi (アイファイ)」が挙げられる。Eye-Fiは、Wi-Fi (無線LAN) 接続機能を備えた外部記憶媒体 (SDカード) で、デジタルカメラで撮影した写真や動画を、オンラインフォトアルバムなどに自動的にアップロードしてくれる。初期設定をすればユーザーは撮影する以外の余計な作業を一切せず、「撮影した写真が自動的にサーバーにアップロー

図3 「Eye-Fi (アイファイ)」のサービスモデル(「ゼロクリックサービス」の事例)



ドされる」という効用を得る。その結果、大量の画像データが容易にサーバーに蓄積されることになる(図3)。

クラウドと各種端末とをより円滑に連携させてビッグデータの活用につなげていくためには、通信の隠蔽やゼロクリックサービスといった仕組みを適用することが重要であることを、Amazon.comのKindle端末の電子書籍サービス、およびEye-Fiのデジタルフォトのアップロードサービスの事例で見た。

これらからもわかるように、クラウド上にさまざまなビッグデータが収集・蓄積されるためには、パソコンや携帯電話端末にかぎらず、さまざまな機器がクラウドに接続可能な状態であることが求められる。なぜならば、ある機器とその機器が取り扱えるデータとは強く結びついているため、多種多様なデータを得るには、パソコンや携帯電話端末などの汎用機器だけでは不十分だからである。

電子書籍端末やデジタルカメラ、あるいは

デジタルフォトフレームのような多種多様な機器がクラウドと連携するに当たっては、ユーザー(特に一般消費者)が積極的にそれぞれの機器に対してデータの送信を設定し、通信費用を負担することは想定しにくい。そのため、本節で紹介したような通信の隠蔽ならびにゼロクリックサービスといった観点を踏まえた仕組みづくりが必要となる。

3 データ活用環境に関する現状の不整合を踏まえた事業化

ビッグデータの利用サイド事業者に対して、提供サイド事業者がビッグデータの関連ツールやサービスを提供する場合の訴求方法としては、データ活用の不整合の改善を目的とした事業も想定される。

すでに蓄積してあるデータを活用しようとしても、「データが分析可能な状態で蓄積されていない」ケースは多い。たとえば、顧客コードや製品コードが部門やエリアによって

不統一であるような状況である。このようなデータがいくら大量にあっても「ゴミはいくら集めてもしょせんゴミ」であり、そこから有用な知見を導出するのはきわめて困難である。

整合の取れていないそうしたデータを、分析可能な状態に整理するサービスも登場している。前述のMDMがその一例である。MDMは、利用サイド事業者が取り扱うさまざまな種類のデータを一元的に管理できるように加工するサービスである。データの高度活用を提供しようとする提供サイド事業者のなかには、まずはMDMでデータの整合性を確保し、そののちに業務に活用可能な知見を導出するための仕組みを提供するなど、段階的に売り込みの強化をしようと考えているところもある。

IV 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

1 最大の阻害要因は人材不足

ビッグデータの活用が促進されるなかにあって大きな阻害要因となっているのは、「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材、すなわち統計学や情報科学のリテラシーに富む人材の数が不足している。

きわめて大量のデータを前にしたとき、全員が「ここから何らかの有用な知見を得られるのではないか」という期待を抱くわけではない。極端に言えば、「日本人全国民の過去一年間の購買状況」といったデータが仮に存在したとしても、それらのデータを前にして知見導出の可能性を感じる人は少なく、大半

の人は途方に暮れてしまうであろう。

利用サイド事業者の立場であれば、「それならば専門家に外注すればよいではないか」という発想が出てこよう。しかし、情報システム構築の外注と同様、目的に沿った成果を得るには、発注側（利用サイド事業者）に最低限の知識・スキル（技能）・リテラシーが必要であり、それがなければ何をどのように外注してよいのかすら判断できない。外注先と円滑なやり取りをする人材にも事欠くというのが、一般的な利用サイド事業者の状況と思われる。

利用サイド事業者に、ビッグデータの活用ができる人材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よう」「この部分は外部の専門家に深掘りしてもらおう」と考えることのできる人材を増やすための、地道な裾野拡大が求められる。

たとえば、パソコンで用いる解析ソフトのインターフェースを、利用サイド事業者の社員にとって親しみのあるデザインにすることも有用であろう。統計分析に慣れていない社員の底上げを図るためには、その社員がふだん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できる。

また、提供サイド事業者の立場であれば、上述した利用サイド事業者側の人材不足の状況を踏まえた形での機器やサービスの提供が不可欠であり、それは商機でもある。しかしながら、提供サイド事業者の多くも統計分析や数理モデリングを担う人材は不足している。

このことに関連しては、米国の事例となるが、グーグルのチーフエコノミストであるハル・ヴァリアン（Hal Varian）氏が2009年のインタビューのなかで、「今後10年間でセクシーな職業は統計家である」と語っている^{注1}。また、シリコンバレーでは、「Hadoopが使える統計リテラシーがある人材」は、創業したばかりのベンチャー事業者から大手事業者まで広く募集対象とされ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いてすでに人材の奪い合いが始まっている様子が見て取れる。

2 次の課題はプライバシー・機密情報の不正利用対策

ビッグデータビジネス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・性別・職業などの属性、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・連絡先、あるいはコンテンツの閲覧および購買の履歴などが想定される。

たとえば、GPS機能付き携帯電話端末から収集・蓄積された個人の移動情報や行き先のデータが悪用されると、個々人の日常の行動パターンが他人に知られ、尾行などに悪用されてしまうおそれがある。

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ても、複数のデータが関連づけられることで利用者本人が望まない事実までもが知られてしまうことも懸念事項といえる。

さらに、データの取得や公開・流通が容易

になるなかで、データ取得者自身には悪意がなくても、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。こうした懸念はかねてより指摘されている。たとえば、「たまたま通りすぎた人が、たまたま記録した内容には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、欲しい情報が手にはいることになる。（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」^{注2}という懸念が、1999年にすでに示されている。これは、把握・収集可能なデータが増大している現状にあって、まさに懸念されるべき事項といえるだろう。

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。これには万能の施策はなく、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、以下に、「環境オプトイン」と「プライバシー保護データマイニング（PPDM）」の2つの概要を示す。

● 環境オプトイン

データの取得に関する制度的な対策として環境オプトインに関する研究が進められている。環境オプトインとは「特定の環境において、誰もが『この事業者は、プライバシーに

配慮して、安心安全にパーソナル情報を取りあつたこと』が判る合意形成」^{注3}とされている。

たとえば、ある店舗に設置された監視カメラの映像を分析してマーケティングに用いるような場合には、店舗の入り口にその旨をあらかじめ掲示しておくべきとする考え方である。この考え方はいまだ検討段階にあるが、こうしたことへの制度的な対策は必要になるだろう。

● プライバシー保護データマイニング

次に、ビッグデータの活用段階にも求められる対策が想定される。一例に、データ分析の結果得られた知見をプライバシーや営業機密を守りつつも活用しようとする際の技術的対策として、プライバシー保護データマイニングに関する研究がある。データマイニングをしつつも、一人ひとりのプライバシーは保護したいという、一見相反するようなテーマについて、2000年以降、学術レベルでの検討が進められている^{注4}。

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共有できる」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。

以上、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、高いチャレンジ要素（課題）を包含しており、産業界全般で進むクラウド利用と併せて、2010年代の情報・通信分野における注目すべきテーマの一つになると

考えられる。

ビッグデータビジネスの健全な進展には、ビッグデータを活用できる人材の育成、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整備や技術的解決が不可欠である。ただし、機微なデータの取り扱いへの配慮は必要ではあるものの、配慮のあまり、ビッグデータの活用を全く控えてしまうこともまた、事業の推進や社会システムを効率的に整備していくうえで不健全である。困難ではあるものの、ビッグデータ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。

注

- 1 Steve Lohr “For Today's Graduate, Just One Word: Statistics,” *The New York Times*, August 5, 2009
- 2 松王政浩「複合現実感技術は何をもたらすのか」『情報倫理学研究資料集 IV』日本学術振興会未来開拓学術研究推進事業電子社会システム部門「情報倫理の構築」プロジェクト、京都大学文学研究科、広島大学文学研究科、千葉大学文学部、2002年6月
- 3 日本情報処理開発協会「パーソナル情報の利用のための調査研究報告書」2011年3月
- 4 J. ヴァイタヤ、Y. M. ズー、C. W. クリフトン著、嶋田茂、清水将吾訳『プライバシー保護データマイニング』シュプリンガー・ジャパン、2010年

著者

鈴木良介（すずきりょうすけ）

ICT・メディア産業コンサルティング部主任コンサルタント

専門は情報・通信業界の事業者や行政機関などに対する調査・コンサルテーション。情報・通信の効率的かつ安全な活用に関し、クラウドコンピューティングやビッグデータ活用の観点から検討している