

# ビッグデータの活用と 個人情報保護

特集

1

## 広がるビッグデータの 流通・利活用と課題



佐藤 一郎 Sato Ichiro 国立情報学研究所・副所長、教授

総合研究大学院・情報学専攻教授を兼任。専門はシステムソフトウェア。個人情報保護法改正などの情報関連の法制度に関する政府委員など。



### はじめに

ビッグデータという言葉とともに、多様なデータの利活用が試みられ、データの経済的価値が上がるとされています。その結果、データを現代の石油と呼ぶことさえあります。そのデータの中でも消費者一人一人にかかわるデータ、つまりパーソナルデータは経済価値を生み出すと信じられ、政府や企業において、その利活用が実施または検討されています。これは消費者からすると、本人が知らないところで勝手にパーソナルデータが使われてしまったり、プライバシー侵害などにもつながることになります。さらに流出・漏えいなどの不安も広がることになります。

本稿では消費者保護の観点から、パーソナルデータを中心にビッグデータの利用や流通の状況と今後について概説します。なお、パーソナルデータというとその流出・漏えいを心配する人が多いのですが、ビッグデータの普及は仮に流出・漏えいがなくても、大きなリスクがあり、それを理解することが重要です。

### ビッグデータとは

ビッグデータという言葉がはやる以前から、

消費者に関するパーソナルデータの利活用は行われていました。しかし、ビッグデータにかかわる技術により、その利活用の仕方が大きく変わってきています。

さて、ビッグデータとは大量かつ多様なデータとそれを処理するシステムの総称です。何ができるのかという観点で説明すると、以前はコンピューターの処理能力的に難しかった大量データの分析や、多様な種類のデータを組み合わせた分析ができるようになります<sup>\*1</sup>。この結果、これまでは一握りの消費者だけを抽出して分析したり、各消費者のごく一部の断片情報だけを扱ったりするケースが多かったのですが、ビッグデータにかかわる技術を駆使することにより、多数の顧客を抱える小売事業者が、顧客一人一人の購買履歴を詳細に分析できるようになっています<sup>\*2</sup>。

例えば Amazon の強みは商品推奨ですが、そのしくみは図のようなものです。まず全顧客と全商品に関する巨大な表を作ります。次に各顧客について購入した商品やウェブページを見た商品などの度合いに応じて数値を入れていきます。そして商品に関する数値が似ている顧客同士は類似した購買行動があるとして、一方の顧客が購入した商品やウェブページを見た商品

\*1 コンピューターの処理性能やアルゴリズムの効率性が急激に向上したわけではなく、多数のコンピューターを協調させることで、大量かつ多様なデータを処理する方法が発達した。  
\*2 小売事業者が消費者行動を細かく分析するようになった背景には、消費者が商品を購入する前に、その商品を製造または販売する企業の宣伝よりも、その商品を既に入力した消費者のクチコミ情報を信用するようになったために、従来のマスのようなマーケティングは難しくなり、個々の消費者の行動を細かく分析しなければならなくなったことが挙げられる。

特集1 広がるビッグデータの流通・利活用と課題

を、もう一方の顧客に推奨していきます。当然、Amazon の場合、顧客数も商品数も多いので、前述の表は膨大な大きさとなり、ビッグデータの技術がなければ処理できません。

図 商品推奨（協調フィルタリング）のしくみ



商品購入やウェブ表示実績に応じて数値を与える。1番目の顧客と3番目の顧客は数値の並びが似ているので、互いに類似した購買行動を取ると推測して、1番目の顧客の購買品を3番目の顧客に推奨し、またはその逆を推奨する。

### 無料サービスの裏側

ビッグデータにかかわる技術でパーソナルデータを駆使しているのが、通称、GAFAと呼ばれる企業群（Google、Apple、Facebook、Amazon）でしょう。

Googleの場合、個々のユーザーの検索履歴やウェブの閲覧履歴、視聴した動画などの大量のパーソナルデータを集めています。それを分析することで、ユーザーの特性や関心を把握し、それぞれに応じた広告を表示することで広告効果を高め、大きな広告収入を得ています\*3。さらに新しいサービスを開発・提供することで、さらにパーソナルデータを集めるというビジネスモデルです。

Googleをはじめ、大手ウェブ検索サービスでは、1回の検索で100台程度のサーバを使っているはずですが、それなりのコストがかかりますが、無料で提供できるということは、それに見合う利益があるからであり、その源泉がパーソナルデータです。ユーザーが検索したキーワードによって、ユーザーの関心事や状況がみえてきます。例えば病名を検索すれば、その検索事業者は、本人またはその家族がその病気にかかっている可能性があるかと推定して、それにかかわる広告を表示していきます。同様のこと

はネット通販にもいえます。例えば、ある消費者ががんに関する書籍とカツラを購入していれば、その消費者は抗がん剤を投与されていると推定して商品推奨に反映してきます。

また、Facebookに限らず、国内外のソーシャル・ネットワーキング・サービス（SNS）の大半は、ユーザーが書き込んだ文章の分析はもちろん、そのユーザーと友人になっているユーザーとの関係についても詳細な分析をし、表示する広告を選ぶことで広告効果を上げています。

消費者が無料でウェブ検索やメール、地図、SNSなどの便利なサービスを楽しむのは、見返りに自らのパーソナルデータを提供しているからだということを忘れてはいけません。

### パーソナルデータと種別

パーソナルデータは次の3種類に分類されます。

#### ①主体的提供データ

本人が自ら提供する情報。氏名、住所、電話番号などが相当する。

#### ②観測データ

センサーやカメラなどで個人を観測することによって得られる情報。

#### ③推定データ

ある個人に関する断片的な情報に他人や何らかの統計情報を組み合わせ、その個人がどういった人かを推定するデータ。

さて10年ほど前は、パーソナルデータというと①主体的提供データが中心でしたが、近年、Internet of Things (IoT) という言葉で代表されるように、カメラはもちろん、各種センサーから取得された心拍や体温、位置、移動経路など、容易に取得・収集した大量かつ多様なデータを、自動的に分析できるようになっています。

例えば多数の個人に関する顔画像と氏名などの組み合わせに関するデータベースがあれば、防犯カメラに写った人物が誰であるか分かる確率が高くなっているのが実情です。さらにカメラやセンサーは小型化が進んでおり、消費者は行動を観測されていることに、気づきにくく

\*3 Googleの収入の95%以上は広告関連といわれている。

なっています。

一方、日本の法制度ではパーソナルデータのうち、個人情報とされる範囲は個人情報保護法を通じて保護されることになってはいますが、保護対象は個人情報<sup>\*4</sup>の利用に関するものにとどまり、カメラやセンサーなどによる個人情報の取得に関しては、ほとんど規制がないことは理解しておくべきです。

そしてビッグデータと密接な関係があるのは③推定データ（プロファイリングデータとも呼ばれます）です。例えば SNS などでは、ビッグデータにかかわる技術を駆使することにより、ユーザー本人が入力したデータ（①に相当）よりも、SNS 事業者により推定されたそのユーザーに関する推定情報の量のほうが多いとされます。その推定情報には、そのユーザーと他のユーザーとの関係、過去の書き込みなどから推定したユーザーの性格や行動などが含まれ、SNS の場合、事業者にもよりますが、ユーザー1人当たり数ギガバイトの量になることもあります。

### プロファイリングされる消費者

ここで問題となるのは、プロファイリングデータの精度が低く、しばしば間違っただプロファイリングにより、消費者の権利利益の侵害が起こり得ることです。

例えば SNS 事業者の場合、ユーザーの書き込みなどから、そのユーザーの性格や関心事を推定していきますが、SNS への書き込みはそのユーザーの断片的な情報に過ぎません。その結果、SNS の書き込みに乱暴な文言が頻繁に含まれていれば、そのユーザーは攻撃的な性格・行動の持ち主と推定されがちです。さらに断片的な情報だけでは不十分な場合、事業者は外部情報で補いますが、その外部情報となるのは、そのユーザーと類似した書き込みをする他のユーザーに関する情報や、「類は友を呼ぶ」という仮説からそのユーザーの友人の情報

です。しかし、あるユーザーの書き込みが他のユーザーの書き込みと類似しているからといって、両ユーザーの性格や関心事が同じとは限りません。また不幸にして外部情報に含まれる他のユーザーが仮に特異な性格や関心事の持ち主だと、そのユーザーも同じ性格と関心事の持ち主として推定されます。さらにプロファイリングによるデータを外部情報としてプロファイリングするので、間違いが増幅されやすいのが現実です。

問題なのは、こうした間違っただプロファイリングは個人の権利利益に影響を与えるにもかかわらず、消費者本人はプロファイリングが存在することも、自分がどのようにプロファイリングされているのかも知らず、間違っただプロファイリングの修正を求めることも困難だということです。

### 格付けされる消費者

ビッグデータにおいて、消費者は常にデータとして扱われる存在、つまりデータを通じて評価され、処遇されています。

例えばアメリカでは、企業に就職するときに、クレジットカードレポート（引き落とし日に残高不足やローン支払い延滞などを記載したもの。カード会社が発行する）の提出が求められることは少なくありません<sup>\*5</sup>。このレポートはある種の消費者の格付けとして利用され、スコアが低ければ就職や賃貸借契約などに影響します。また、中国でも企業による消費者の格付けは進んでおり、そのスコアは与信や金利優遇への利用はもちろん、最近、中国で普及しているQRコードなどを利用した簡易決済における不正抑止にも使われています。さらに、中国では企業による消費者の格付けと政府による国民信用データベースの連動が進み、格付けスコアは公共サービス利用可否にも利用され、一方で政治的な行動がそのスコアに影響するという報道もあります<sup>\*6</sup>。

<sup>\*4</sup> 個人情報の範囲は個人情報保護法によるが、簡単にいうと、誰であるかが分かる情報と、外部情報との照合により誰であるかが分かる情報となる。また2016年の同法改正により、一部識別子や生体情報も個人情報に加えられた。

<sup>\*5</sup> レポートの提出は任意ということになっているが、現実には提出を拒否すれば採用候補者から外されることもあり得る。

<sup>\*6</sup> Christina Zhao: 「14億人を格付けする中国の『社会信用システム』本格始動へ準備」、Newsweek 日本版(2018年5月2日) <https://www.newsweekjapan.jp/stories/world/2018/05/14-8.php>

格付けは正しいとは限りません。ある人の出身地や育った場所が、たまたま犯罪率や貧困率が高い場所とすると、その人も同様に扱われるおそれがあり、格差がますます広がります。

今後はIoTにより取得される個人行動にかかわるデータを含めて、取得・利用されるパーソナルデータの範囲は広がり\*7、至るところでプロファイリングされるでしょう。また、SNSの書き込み内容や購買履歴で与信審査が行われることになるでしょう\*8。その場合、プロファイリングが間違っている場合だけでなく、若かりし日に勢いでSNSに書いた乱暴な言葉ひとつで、消費者は実情よりも高い保険料を算定されたり、支払い能力が疑問視されて金利が高めに設定されたりしかねません。SNS上の友人にローン延滞者や犯罪歴のある者が含まれば、同様に扱われるリスクがあり、リスクを避けるために友人を選び、そのことが差別や社会的排除につながる可能性があります。また消費者のどんな行動が格付けに影響するかは明らかになっているとは限らず、格付けの仕方次第で消費者の行動を恣意的に方向づけることも不可能ではありません。

EUではこの問題を解決するために、「忘れられる権利」\*9が導入されていますが、日本においては法制度が追いついていないのが実情です。

## まとめ

ビッグデータを活用することにより事業者が実現したいことは、事業の効率化です。例えば消費者のパーソナルデータを駆使して広告を選別することは、消費者にとっては関心のない広告を見なくてすむというメリットもありますが、事業者にとっては広告効果を上げることに他なりません。

このとき留意することは、効率化と公平性は両立するとは限らないということです。無償サービスが広告収入によって賄われている場

合、事業者にしてみれば購買余力の高い層に広告を出したいのは当然です。その場合、逆に購買余力の低い層は広告効果がないとなると、これまで便利に使っていた無償のサービスが提供されなくなる可能性があり、ますます格差が広がる危険性をはらんでいます。

SNSを含めたサービスにおいて、数多くの広告を見させるためには、消費者がそのサービスを利用する時間を延ばせばよく、そのためには消費者にとって不快な情報、例えばその消費者の関心事や主義主張に合わない情報を見せないようにします。消費者にとっては心地良いことのように思えますが、偏った情報しか表示されないことになり、これは昨今問題となったSNS上のフェイクニュースの遠因になっています。これを防ぐには、消費者も自らのパーソナルデータがどのように扱われているのかを気に留めておくべきです。

また、前述のGAFAsのように、パーソナルデータを収集・分析して事業に生かし、さらに新しいサービスを開発・提供することでパーソナルデータを集めるというビジネスモデルが展開される状況においては、パーソナルデータが特定の事業者へ集中するという問題が起きています。過度な集中は事業者間の競争を阻害しますし、消費者の事業者への依存を高めることから、長期的には望ましい状態とはいえません。

EUは、2018年5月に施行されたGDPR(EU一般データ保護規則)などの法規制を通じて、特定の企業へのパーソナルデータの集中を規制する施策を打ち出しています。日本においては2017年5月施行の改正個人情報保護法により、世界の個人情報保護制度に追いついた状況であり、今後も世界情勢に合わせて、企業の活動の活性化と消費者保護の両立を図っていかねければなりません。

\*7 位置情報が取得できれば、今後は消費者が広告を見た後に実際に来店した場合、高額の広告料が支払われるようなケースも出てくるだろう。

\*8 近年、アメリカ「ゼスト・ファイナンス」、ドイツ「クレディテック」をはじめ、欧米諸国でソーシャルメディアの情報を活用して消費者向けローンを提供するベンチャー企業が生まれている。

\*9 インターネット上に掲載された自分の個人情報を削除できる権利。2012年1月、EUは個人情報保護を定めた「一般データ保護規則案」を公表。この第17条に「忘れられる権利 right to be forgotten」が明文化された。